European Symposium on Computer Aided Process Engineering – 15
L. Puigjaner and A. Espuña (Editors)

Pharmaceutical Informatics: A Novel Paradigm for Pharmaceutical Product Development and Manufacture

Chunhua Zhao, Girish Joglekar, Ankur Jain, Venkat Venkatasubramanian*, G.V.Reklaitis
{chunhua, gjogleka, jain18, venkat, reklaiti}@ecn.purdue.edu
Institute for Advanced Pharmaceutical Technology
Purdue University
West Lafayette, IN, USA

Abstract

The compelling drivers for the pharmaceutical industry are minimizing the time between a drug's discovery and its delivery to the marketplace, and maintaining high productivity in the manufacturing processes. During a product's lifecycle many complex decisions must be made to achieve these goals. To better support the development and manufacturing processes at each stage, we have proposed a new paradigm to facilitate the management and transfer of data, information and knowledge. A prototype for the informatics paradigm has been successfully developed. As a first step, the feasibility of the proposed informatics framework is demonstrated using a case study based on a pilot plant operation.

Keywords: Pharmaceutical informatics, Drug development and manufacture, Process development tools, Information modeling and integration

1. Pharmaceutical Informatics: the grand vision

The major tasks involved in bringing a pharmaceutical product to the market are outlined in Figure 1. Under ideal circumstances, each stage operates in isolation and generates process knowledge according to certain expectations. The knowledge is then passed on to the next stage at the appropriate time. However, in reality considerable amount of "starting over" occurs due to the changes in overall direction as more knowledge is gained at each stage. The information flows between the various stages of this process tend to be voluminous and there is high level of uncertainty in the underlying technical and commercial data. In addition, the technical specifications and information base which must be developed to support production are very extensive and the regulatory oversight requirements are very demanding. The current industrial response to all these challenges is sub-optimal. There are many *individual islands of automation and no* comprehensive, integrated, environment that links these islands. Therefore, practitioners must make do with a limited computer-based assistance to acquire, manage, analyze and interpret complex product and processing information with enormous amounts of human intervention. This increases the inefficiencies,

* Author to whom correspondence should be addressed: venkat@ecn.purdue.edu

uncertainties, costs, delays, and product quality concerns all along the product development.

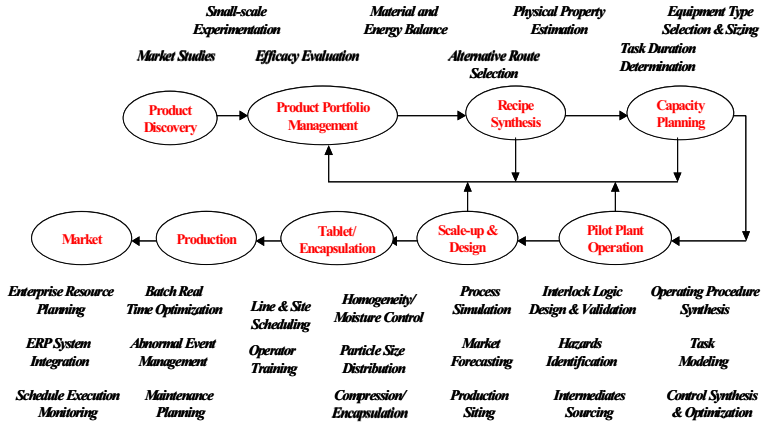

Figure 1. Major tasks in the product development and manufacture

We propose a new paradigm for pharmaceuticals development and manufacture that can effectively address the challenges faced by the industry. In this paradigm, a web-based informatics infrastructure is developed to support the key decisions spanning the entire process, including product portfolio selection, capacity allocation decisions, process monitoring and quality control, production planning and scheduling, process safety analysis, and supply chain management. The data/information/knowledge flows in the entire process development are modeled. We call this infrastructure *Pharmaceutical Informatics* (PI). In this novel paradigm, we adopt an information-centric view instead of current application-centric view as discussed in next section.

2. Information Centric View

At present, the scope and representation of a process are limited to the requirements of a specific stage, and are often influenced by the tools used in that stage. Therefore, only a partial representation of the process is required. For example, a simulation tool does not need to have material safety related data for its execution, but this kind of information is crucial to process safety analysis. Furthermore, different tools are structured in different ways to present the same process information. The lack of a coherent and unified process view results in islands of information with virtually no provisions for the computer assisted exchange of knowledge between these tools. We call this paradigm application-centric view.

The most important aspect of the proposed informatics infrastructure is the separation between the underlying process information and the tools which use the information. Information is central to the infrastructure in the new paradigm, not the tools that use the information. Various applications become providers as well as consumers of the information. We believe that only by dramatically changing the current application-

centric view to information-centric view it is possible to develop the vision we have proposed in previous section.

Thus, in the new paradigm, instead of entering the process specification into individual tools, the task here is to *describe* the information. Instead of encoding process specifications in objects in a particular programming language or tool specific constructs, the process information should be *explicitly described*. A repository of process information will be at the core of the informatics infrastructure. The repository will cover a wide range of information blocks such recipes, equipment configurations, experimental data, plant operation data and so on. Furthermore, the information is separated from its presentation, contrary to what is currently implemented in many process development tools. Though new to pharmaceutical process development, this separation of tools, information, and presentation have been very well studied and adopted in other disciplines, for example, it is very easy to make analogy with business logic, data, and presentation in business process.

The explicit description of domain concepts and relationships between these concepts is known as *an ontology*. The recent developments in the field of ontology have created new software capabilities which greatly facilitate implementation of the Pharmaceutical Informatics infrastructure. Ontology is defined as an explicit specification of a conceptualization (Gruber, 1993). The shared understanding is the basis for a formal encoding of the important entities, attributes, processes and their inter-relationships in the domain of interest. Ontologies can be used to describe the semantics of the information sources and make the contents explicit, thereby enabling integration of existing information repositories, either by standardizing terminology among the different users of the repositories, or by providing the semantic foundations for translators. Web Ontology Language (OWL), recommended by W3C, is used as the language to specify the ontologies. OWL can formalize a domain by defining classes, properties of these classes, and relations between them. OWL can also define individuals and assert properties about them, and furthermore reason about these classes and individuals to the degree permitted by the formal semantics of the OWL language.

As a very first step towards the information-centric paradigm, we demonstrate that it is feasible to explicitly describe the information related to pharmaceutical development process, and it is achievable for various application tools to access the information. So the goal here is not to construct a comprehensive ontology for the information. We demonstrate the feasibility of the proposed framework using a pharmaceutical pilot plant operation.

3. Process Ontologies

We first create ontologies for the domain of process recipe information. As the first step, the existing standards related to process information and literatures on process ontologies were reviewed because they summarize a common view of the process information from industrial practitioners. The standards can be grouped into three categories: equipment specifications (AP231, FIATECH, STEP), process engineering

computations (CAPE-OPEN), batch recipe specifications (ISA S88, S95). Although there is no single standard that has been accepted in the industry and that can adequately describe a pharmaceutical process, the review provided a good basis for developing the required ontologies. Whenever appropriate, the same terminology and keywords were used in the new ontologies. For process engineering area, OntoCAPE has been developed define conceptual data models (Yang, 2004). However, this extensive development is targeted for continuous process design and simulation. At present it does not accommodate the general recipe, the most important concept in pharmaceutical processes.

We use Protégé (see URL), an ontology and knowledge-base editor, for creating ontologies and Racer (see URL), a Description Logic reasoning system, as the inference engine to check data consistency and validity. Complex rules for checking data consistency are processed using Jess (Ernest J. Friedman-Hill, 2003), which is a rule engine and scripting environment.

The higher level concepts in the process ontology include Material, Equipment, Reaction, Separation as well as RecipeElement under which Operation and UnitProcedure are defined. Other important concepts include Stream and Port. OWL provides the capability to describe complex relations between the concepts in the process information. For example, *hasUnitProcedure* property in Operation and *hasOperations* in UnitProcedure are inverse functional relations. The *hasOperation* property in Port has cardinality restriction of 1. The inference engine can clarify the hierarchy of classes, and furthermore, check consistency to ensure the completeness and validity of the process information based on predefined consistency rules. As an example, a consistency rule is used to require that the *hasUpstreamPort* and *hasDownstreamPort* of a Stream be different.

4. Pharmaceutical Informatics Infrastructure

Figure 2 shows the overall architecture of the new paradigm. Process ontologies are first created as well as the logic embedded in the rules to ensure the completeness and validity of the process information. Based on the ontologies, instances of the concepts and relations are created for a particular process using the web based information management facility. A web-based interface for information management was developed because the web is the natural environment for the use of the infrastructure with such a wide scope. Additionally, development of thin-client applications in web environment has become feasible due to the recent advances in web technologies. The process information repository is in OWL format or related databases. Given the diversity of the tools that will access the process information, a middle layer consisting of controller, adaptor or translator is created for the tools as well as the web interface to access the repository.

In this work, we investigated and critically assessed various technologies in order to make this infrastructure work, including the link between web interface to the information repository and different scenario for tools to access the information. A Web

based process information management prototype has been developed. The current effort is to use the interface to create a specific process based on the ontologies. The design guideline is to minimize the code to be written in order to create the presentation. Thus XSLT is used to link the OWL files to the presentation, and XForms is used to define forms for various process information. SVG is used as the format to generate recipe network or PFD from process information. A portal is also created to manage the process information.

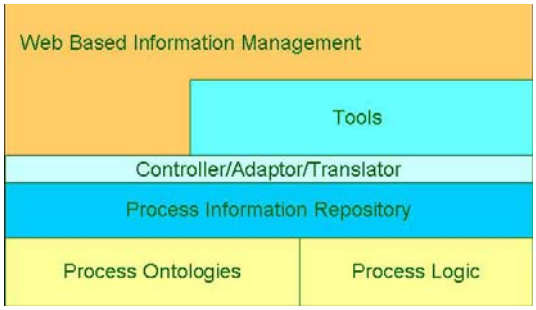

Figure 2. Pharmaceutical Informatics infrastructure

The software tools that will use the process information defined in OWL can be grouped into the following three types:

1. Tools that have native interface for the ontologies, that is, tools that are developed based on the ontologies defined in this work
2. Tools that have the ability to read or import process information from databases or XML
3. Tools that use proprietary input and output formats.

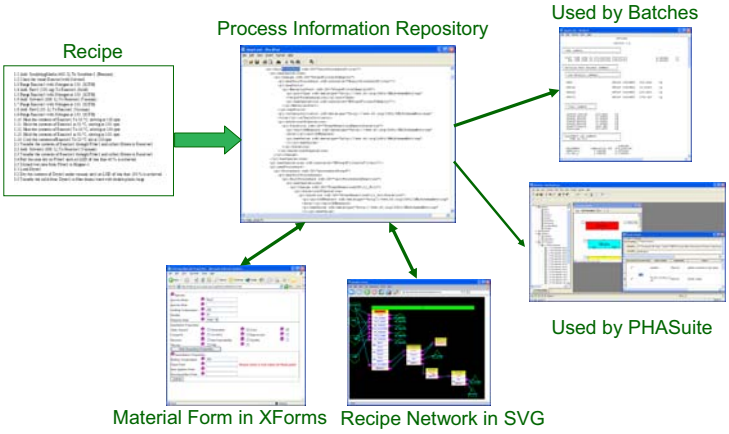

Figure 3. Interactions between the process repository and application tools

A simple pharmaceutical API process has been used to demonstrate the technical feasibility of the infrastructure. The interactions between the process repository and software tools are shown in Figure 3. The process consists of four batch unit procedures: Reaction, Filtration, Drying1 and Drying2, involving 20 operations. All the three scenarios for tools to consume the process information in OWL are successfully

implemented. As an illustration of a category 1 tool, Java classes were created using Kazuki (see URL) and to analyze the details of a given operation. As an illustration of a category 2 tool, the repository was mapped into XML read into PHASuite (Zhao, 2002) for safety analysis. To illustrate the use of a category 3 tool, a Batches (Batch Process Technologies, 2003) simulation model of the process was created using Kazuki and Jena Framework. We also demonstrate a representation layer of the repository using Web interface adapting XML based web technologies, including XForms for entering information, and SVG for generating graphic views.

We are currently working on integrating information generated from tools to the information repository. From what we have shown, the process information can be sufficiently described using OWL, which can act as a process information repository. The open and explicit description of the information provides a solid foundation for better use, sharing, and integrating process information.

5. Summary and Future Research Plan

In this paper, we proposed a novel paradigm to facilitate transfer of data, information and knowledge using ontological models to better support the development and manufacturing processes. The main concept behind the proposed pharmaceutical informatics infrastructure is the separation of process information from the tools that use this information, and furthermore, the information is described explicitly. A prototype of the informatics infrastructure has been developed for process information in the pilot plant stage, to demonstrate the feasibility of the framework. Major components of this infrastructure were identified. Process ontologies have been created, and methodologies for using the ontologies to describe the process information and share information for tools are discussed. Work is in progress to further expand and implement the informatics infrastructure to the entire pharmaceutical development process and its various tasks and decision levels.

References

Batch Process Technologies, Inc., 2003, Batches Reference Manual, W. Lafayette, IN.

Ernest J. Friedman-Hill, 2003, Jess 6.1 manual. Sandia National Laboratories.

Gruber, T. R., 1993, International Workshop on Formal Ontology. Italy.

Golbreich, C. and Imai, A., 2004, 7th International Protégé Conference.

Haarslev V., Möller R., 2001, Lecture Notes in Computer Science.

Jena Semantic Web Framework: http://jena.sourceforge.net/

Kazuki: http://projects.semwebcentral.org/projects/kazuki/

Protégé. http://protege.stanford.edu/

OWL Overview: http://www.w3.org/TR/owl-features/

SWRL: http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/

Tim Berners-Lee, James Hendler, Ora Lassila, Scientific American, May 2001.

Yang A., W. Marquardt, 2004, ESCAPE 14.

Zhao, C., 2002, Ph.D. Thesis. Knowledge Engineering Framework For Automated HAZOP Analysis, Purdue University.

Acknowledgements

The authors gratefully acknowledge funding support from The Indiana 21st Century Research and Technology Fund.