Batch Process Development - An Information Supply Chain

Alistair Gillanders  
Senior Consultant – Batch Processing  
AEA Technology Engineering Software  
56 W Main Street, Suite 211  
Newark, DE 19702, USA  
Tel: (302) 369-0773  
Fax: (302) 369-0877  
Email: Alistair.Gillanders@software.aeat.com

KEY WORDS

process development, new product introduction, uncertainty, risk, information management, decision support, software, S88, CAPE, STEP

ABSTRACT

This paper reviews the information requirements throughout the development lifecycle across different sectors of the batch industries. It then discusses the challenges facing software seeking to enable smarter working through this lifecycle. Aspects considered include the creation, analysis, manipulation and retrieval of process information in ways that enable more effective decision support, eliminate transcription and work cooperatively allowing users to choose “the right tool at the right time”. Some of the issues addressed are:

- Uncertainty and data quality/rigor  
- Risk vs. consequences including the business context  
- Resource constraints…addressing “what should we do next?”  
- Multi-disciplinary synergies and frictions (especially chemist & chemical engineer)  
- Supporting/competing standards…where does ISA/S88 fit?  
- The differing needs of simulation, design and recipe execution systems

Benefits from an effective IT environment are discussed covering both tactical (direct cost and time savings) and strategic (“picking the winners”, time to market, more efficient, robust and flexible processes) benefits.
The paper draws conclusions regarding the classes of software needed, the essential nature of fully supporting information sharing, the impact of existing and developing standards, the current “state of the art” available and some ideas on how the future might look.

**THE PROCESS OF PROCESS DEVELOPMENT**

In order to define effective heuristics for batch process development it is important to understand the business and technical context. In other words you must understand both your process development lifecycle and what principle factors influence it. This statement will appear obvious to most of you, however it does challenge some commonly held perspectives.

It is generally accepted that whilst there are many common features to the batch processes that result from process development activities (this is what makes S88 possible), the actual requirements of process development differ significantly from one industry sector to another and from one business to another. However it is instructive to consider where these differences really arise.

Figure 1 offers a view of a process development lifecycle that most practitioners will find familiar. As it implies the variety in process development usually arises not from fundamental differences in the development process so much as differences within the business and technical context.

There are various business issues that define this context. Examples might include:

- the nature of any “new” product or process
  - novel product vs line extension
- where and how it has been identified
  - research based pharma vs generic
  - in-house research vs licensing vs toll contract
- the timescale available for development
  - 6 months toll contract vs 10 years for a complex new drug
- the sustainable profit margin
  - new paint vs cure for cancer
- internal technical and resource constraints
  - must fit into existing plant vs greenfield site
  - familiar chemistry vs low in-house knowledge
  - small technical team vs large multi-disciplinary team
- external factors outside your direct control
  - regulatory implications for a simple commodity product vs a sterile injectable drug
  - competitor new product release cycles
  - a process largely pre-defined by your customer
The existence of this context doesn’t change the underlying lifecycle. It does affect the significance you attach to the various stages through the lifecycle. It may make you more or less willing to iterate around the stages, or consider parallel development tracks. In short the context influences the way you make business decisions about how best to work through the lifecycle for your product, for your company within your industry sector.

AN INFORMATION SUPPLY CHAIN

We are all familiar with the concepts of a product supply chain with issues such as being demand driven, process velocity, sustainable capacity and management of inventory levels. Whilst this integrated approach successfully spans the manufacturing and supply organizations it does not comfortably include product and process research and development. However the concept of a supply chain can be very usefully applied to process development where the real added value product is process knowledge (3), an “Information Supply Chain” in fact.

In a product supply chain the objective is the efficient delivery of actual product within the context of a business vision. In an information supply chain the objective is the efficient delivery of process knowledge to enable investment decisions that make the business vision reality.

Like any supply chain there are raw materials, staged processes that add value to the basic data (data repositories/knowledge bases) and a need to plan for the effective use of resources to grow the knowledge base. The objective being to the grow knowledge base at an appropriate rate to deliver the right quality product at the right time with minimum risk.

Most of the key concepts of a product supply chain have an equivalent within an information supply chain. For example:

- The customer is the product supply chains who need cost effective new processes and/or products to remain competitive.
- Development speed through metrics such as development lead time (2) are akin to process velocity.
- Combine this with the number of development projects that can routinely be progressed at these rates and you are looking at your sustainable capacity.
- Ensuring that you only invest sufficient resources to grow the knowledge base at the right pace is like managing inventory levels. Too high an inventory and you have increased costs (including risk costs if a project fails, pharmaceuticals is probably the highest risk with around 1 in 10 projects resulting in a product on the market but they are not alone). Too low an inventory and you are exposing yourself to other risks through decisions based on insufficient knowledge that could have been gathered.

SOME IMPLICATIONS

Thinking about process development as a supply chain challenges the traditional “technology-transfer” approach to managing information through the development lifecycle. More importantly it greatly emphasizes the need for well-defined and managed repositories for process knowledge that can comfortably grow as process knowledge develops. This requires that all stakeholders can access and contribute to the repository. Furthermore to avoid the
substantial low-value activities and the risk of errors that transcription implies software tools and systems that use and create data must support these knowledge repositories. This creates a number of technical IT challenges because it becomes essential that software tools, as well as users, effectively share their information.

**Questions Before Answers**

We are not seeking knowledge for the sake of knowledge. The goal is good business decisions at the right time so the knowledge must be accessible by users of different disciplines in a way that they are comfortable with, where they need it, when they need it. Effective information repositories and the concept that data is only worth something when it is used are explored more later on.

But this is not just about reducing effort for low value information. It is important to invest in the activities that offer the most valuable information for the current status of development. This is usually about minimizing risk as early as possible...focusing not just on what needs to be done, but what is most valuable to do next?

**Decisions and Risk**

The lifecycle may be thought of as a sequence of activities with milestones or decision points along the way. Many companies now implement this specifically with concepts like stage-gate management of development projects (4). This needs care since a poorly chosen milestone can actually slow up development by creating queues but it does offer powerful opportunities for control and feedback. The key decision at each milestone is go/iterate/no go and is based on the potential return on further investment and the probability of that return being realized.

\[
\text{Development Value} = f_n(\text{Cost, Potential Profit, Risk})
\]

The decision is complicated since additional resources could improve any of these factors and also the factors are not independent. More cost might reduce risk but will reduce overall profit for example, this might be justified for a high margin pharmaceutical but dubious for a paint. Currently evaluating the “development value” is usually fairly subjective, to do so in a comparable way for many projects on different products at different stages in the lifecycle is particularly challenging. So ways of helping assess and compare the various factors that influence a decision for your business are likely to be valuable...so called “Decision Support” methodologies.

Put simply the same information about process options may result in different decisions for a different business context, and that is as it should be. Consequently any information supply chain must be optimized to create and track those factors important to your business at the right time and with the confidence required. But measuring many of these factors is difficult at
identifying reliable indicators is therefore an important pre-cursor to improving the quality of the decision making process.

**DATA, INFORMATION AND KNOWLEDGE**

*The Nature of Data*

To learn how we might move toward a full implementation of an information supply chain we need to understand the data we are working with. This can be done usefully at a high level to establish some key concepts that can be obscured by a detailed view. In particular concepts such as data ownership, uncertainty, quality and availability are important.

In classic database theory we learn that one source of data is the route to a bright new world. In reality data ownership is usually a lot more complex. Who “owns” the price of water in your company? Understanding how many sources you need to consult to answer that fairly innocuous sounding question can be instructive in itself. If you consider the properties of a new chemical entity under development who defines those? Does the “ownership” remain in one place through the lifecycle or does it change? If you accept that multiple data sources are unavoidable for many classes of data (materials information being one of the more challenging in this regard) how do you manage clashes? If you have four different sources for the density of Acetone at 20°C and they don’t all agree…which one is right?

A lifecycle necessarily means we are starting with minimal data and growing our knowledge as we progress. This means there is certain to be much uncertainty, particularly early in the lifecycle, but we still need to make sensible decisions. This is generally managed by either doing without some information until it is justified to invest resources in gathering or measuring additional data, or we estimate it. For example, how many reaction steps have you measured proper kinetic data for? For those you have measured did you do a full kinetic study, or was there some estimation at some stage of the project? Assuming a rate equation, Arrhenius kinetics, rate constants doubling with a 10°C rise in T then deriving rate coefficients so the reaction takes about the same as in the lab is both common and frequently more than adequate to satisfy the issue at hand.

In other words we are often able to make decisions with good confidence using data that is far from “perfect” in quality. The risks are that we put more confidence in the results than the quality of the input data justifies, or that someone else subsequently uses the data or results without being aware of the quality of the information they are using. It is not wrong to have estimated kinetic data for initial heat and mass balances for a new process; it may be dangerous to use that data as the main basis for a safety case for a highly exothermic system. This introduces the need for meta-data (or data about data) with information on data quality (e.g. assumptions made) vital to ensuring good decisions can continue to be made when the information is shared for others to use. So our information supply chain must track the changing quality of data as well as the data itself if it is to effectively support decision making.

As alluded to previously data is the foundation of knowledge but knowledge is only derived if the right data is available to the right person at the time they need it, where they need it and in a way they can interpret it. Value is only added to your business if this knowledge is used to
make good decisions. Data is not enough, knowledge is not enough, and good decision making is not enough…all three are needed.

This means storing data in a manner that protects the integrity of this key asset and yet supports robust and reliable ways of interacting with it. Obviously this must include the flexibility to retrieve data to answer a wide variety of questions most of which we probably haven’t thought of yet. We ideally need to organize the data within the context of both the natural structures of a process and the ongoing lifecycle so we can interrogate the knowledge by a variety of familiar routes (10). This is the role of a well-designed DBMS (database management system)...what type of DBMS to use is discussed shortly.

**The Role of Standards**

There are a number of relevant standards that aim to influence the way we implement data management solutions within the process industry in general and the batch industry in particular. The most important standards effecting the batch industry are probably ISA/S88 (5), ISA/S95 (6) and STEP (7) although all of these have spun off other influential initiatives such as PISTEP, CAPE-OPEN, pdXi, PIPPIN, EPISTLE, POSC-CEASAR etc.

The difficulty is that, although these standards agree on many fundamentals, they are often not congruent in their proposed data structures. This may be reflective of the main industry partners involved early in their evolution and is probably inevitable given the complexity of the task. Even “just” classifying equipment types is a challenge and one where the batch industry arguably has the greatest complexity.

Difficulties aside, identifying, specifying and agreeing these common fundamentals is very valuable for all parties. However problems can arise when these standards seek to widen their scope to influence implementation as well as concept. Implementations are always going to be influenced to a degree by the nature of the parent system. For example the equipment information needed by a CAD system differs from that needed by a simulation system or a control system. A more productive approach, one that is now fairly well recognized, is to agree on standards for interchange rather than storage...even here it is necessary to be cautious not to overly constrain implementation flexibility.

ISA/S88.02 is a good example of this approach. The result is the potential for common interfaces supporting easily configurable “mapping layers”. This would enable data to flow between systems in any direction without constraining unduly the flexibility of an individual system to structure its internal data appropriately.

**CLASSES OF SOFTWARE**

There are four generic classes of software system that are needed to successfully implement an information supply chain, and hence derive the benefits:

- Software applications that add value to basic data – e.g. experimental design and analysis tools, process simulation tools, equipment selection and design tools etc.
- Information storage and retrieval systems – these are the core data management technologies and need, as far as possible, to guarantee the integrity of the repositories
independently of the tools that use and add to them. Once invalid data is in…finding it and fixing any consequential damage rapidly becomes unmanageable.

- Information visualization tools – these are the key to converting raw data into knowledge and include a wide range of analysis, trending and decision support technologies.
- Software interfacing tools to share process knowledge both within the information supply chain and to external systems ultimately including seamless transfer to the manufacturing systems within the product supply chains.

For an information supply chain to offer maximum benefits the components you choose need to be more than just best in class…they must also work cooperatively with the rest of the team. It does not matter how technically great a tool is, if the results are not effectively tracked and disseminated so they can be used for making decisions across the business…you might as well not bother. How often have mistakes happened when the information to avoid them was known but locked away in someone’s filing cabinet or on their PC? The days of the stand-alone tool that can ignore the world around it are rapidly fading.

**STATE OF THE ART**

So what is the current status of components that can help deliver a working information supply chain? This is a non-trivial question however there are some general observations that can help in selection of appropriate technologies and partners for building a strong information system for the future.

One of the key aspects of any knowledge repository is clearly the system utilized for the long-term persistent storage of all the information (the DBMS introduced earlier). In this area there is an ongoing debate regarding object oriented vs relational technologies. Fortunately this debate is nearing a consensus. The tentative conclusions are that, whilst object oriented systems are optimal for developing software, the tried and trusted relational model is still the most appropriate for storage. The challenge for the future is to develop systems that integrate these two approaches successfully (8). Although the next generation of development tools that will benefit from this integration are not yet with us there are several commercial and open-system groups (8,9) already trying to define a stronger interface between the data and application layers (the relational and object views).

Regarding the need for managed replication and wide area sharing of information we are more fortunate. Many of these problems have already been tackled by solutions in the wider commercial world and we can learn from these. Indeed scalable solutions combining powerful messaging and data hub technologies are commercially available, well proven and already in use by many major process industries albeit not always for technical systems yet. This approach requires that participating systems can both send and react to messages for sharing data. This of necessity means technical systems need to become network aware, this re-iterates the need for software components that are more than stand-alone solutions.

**CONCLUSION**

This paper has proposed a way of looking at the information and system needs of the process development process that allows us to learn from body of work aimed at improving the manufacturing and supply process. That is to consider process development as an Information
Supply Chain. In doing this a variety of fundamental system characteristics become desirable with the need to share information across both software systems and users at the core.

The benefits that implementing this could offer are substantial and accrue from both direct tactical benefits and potentially much greater strategic benefits. Examples of tactical benefits would include productivity enhancement, elimination of transcription errors, capital and revenue savings on specific projects. Examples of strategic benefits include cross-discipline team empowerment, increased development speed, reduced time to market, enhanced learning from past projects, more focused use of resources at the right time and earlier and better decision making.

Early adopters of these technologies are seeing dramatic benefits already and the tools are only starting down the integration road (e.g. 80% reduction in time required for some development activities with increased quality and knowledge delivered (4)). As existing and new tools embrace the need to share information effectively these early benefits will grow significantly.

REFERENCES

5. ANSI/ISA-S88 - Batch Control Standards Parts 1 and 2
6. ANSI/ISA-S95
7. STEP: ISO 10303 - Industrial automation systems and integration - Product data representation and exchange