

# The Challenge of Process Discovery

**Laury Verner**  
CTO, ProActivity Inc.

## Abstract

Capturing and understanding your organization's business processes is a critical step in any business process initiative, but one that is often performed poorly, if at all. The history of business is replete with discovery failures and successful but extraordinarily painful efforts at process discovery. These failures are generally the result of a mix of causes: perceived lack of value, insufficient resources, faulty methodology, or inadequate tooling. This paper discusses the concept of process discovery. It explains why discovery is needed, compares several methodologies in current use, and describes how discovery can yield valuable insights for automation and business improvement initiatives.

## Prologue

To motivate the ideas we will be discussing, we'll begin by relating three true stories of companies who faced business process discovery challenges.

### Story 1

Recognizing that they wanted to be a more "process focused" enterprise, a New England Bank decided to do something about it. They selected a team of their best business analysts and chartered them to interview employees and capture the bank's current processes. As the summer began, the team was enthusiastic and was making remarkable progress. As the weeks progressed, however, more and more time was going into managing documents, drawing and re-drawing diagrams, reviewing, correcting, checking interdependencies, and ensuring consistency. Despite the wealth of documentation produced, it was hard to get a sense of the big picture and the team was unable to draw actionable conclusions. There was little time left for analysis and by the end of the summer the team disbanded. Few meaningful insights for the enterprise were achieved. The project was killed by its inability to scale.

### Story 2

Over time, processes grow to reflect the organizational structure of a company. Take the case of a large computer company. Its processes for web content management were a mirror of its divisional structure. They had a separate web-content management process for each division. This came to be recognized as a problem, not only in terms of cost efficiency but also in terms of quality. This became embarrassingly clear to management when a competitor's product was featured on the company website. A team was formed to correct this problem. They were faced with a threefold challenge: how to discover the five parallel processes; how to compare these processes and converge them to a single company-wide best-of-breed process; and, finally, how to generate the detailed functional requirements for the web-content management application.

### Story 3

All companies are facing the requirements of the new Sarbanes-Oxley law, which makes CEO's and CFO's personally accountable for the accuracy of their company's financial statements. They need to build controls into their organization to guarantee that the numbers are accurate. One high-tech company we know of hired an army of highly paid consultants to create Word documents to describe their processes and prescribe the required controls.

While this undoubtedly reassures management that they are complying with the Sarbanes-Oxley requirements, many of these processes will change in the next year or two. Because their business processes have been documented in an inflexible, unstructured format, making even minor changes will be a major challenge. Unfortunately, the documentation will have to be substantially re-done, much to the delight of the consulting firm.

## Part I: The Need for Discovery

### *Business Process Analysis*

The purpose of Business Process Analysis (BPA) is to enable organizations to model and analyze their business processes. Armed with this information, business analysts can gain insight into complex business problems and begin to make recommendations for how solve them. Here are some typical examples of business problems for which a business process approach is relevant:

1. Diagnosing the root cause for a known process problem, such as finding out why the warranty process takes so long.
2. Finding unknown weaknesses and bottlenecks in existing processes, such as the claims process (insurance) or the warranty process (manufacturing)
3. Understanding the interrelations and integration of hundreds of data and document components, such as in the clinical trials process (pharmaceutical industry)
4. Creating standard processes for supply chain interactions (e.g., application of SCOR methodology)
5. Converging multiple parallel processes, performed by different departments, to a single enterprise-side standard process
6. Preparing for package implementation, specifically to perform a gap analysis of the new package relative to current business requirements
7. Generating functional requirements that can be given to developers of a new custom-built system
8. Designing the business logic of a process that will be automated using commercial BPMS products

In addressing all these challenges, discovery enables the analyst to:

- Understand how the current process really works—not a simplified or management-level overview, but clearly laying out each step and interaction in detail. This information will be needed to ensure completeness and accuracy when the analyst begins to design an improved process or introduce digitization.
- Find out what is wrong with the current process, including the weak spots, bottlenecks, manual steps and redundancies. The idea is to “X-Ray” the process as a basis for diagnosis.
- Have a base line for comparison (how much better is To Be relative to As Is, ROI analysis)

### *Business Process Lifecycle*

Business process solutions follow a simple lifecycle:

**Discover→Analyze→Design→ Develop→Deploy→Operate→Maintain**

Discovery, the first step, consists of capturing the organization’s existing (As Is) processes. Once these processes have been discovered, we analyze them to identify and diagnose their weaknesses. Based on this diagnosis, we design new (To Be) processes that correct these weaknesses. We may at this point perform further analysis, to compare the To Be processes with the As Is processes and create a case for action. If the analysis indicates that the benefits will

justify the costs, the organization will then develop and deploy the To Be processes. Finally, the organization will operate the processes in the production environment and maintain them. The To Be process has become the new As Is process.

### ***Definition of Discovery***

Why should discovery present a challenge? Certainly, if a company is performing a business processes, they must already know how it works! But this is just not the case, in practice. Their process knowledge is tacit. It exists in the minds of those individuals who actually participate in the process. Each participant has a local view of the process. Unfortunately, no single participant has a complete global view of the process from end to end. So our definition of Discovery is to transform the organizational understanding of current business processes from tacit or implicit to explicit. An explicit process will communicate the structure and details of business processes in such a way that everyone can understand them and make appropriate decisions.

### ***What do we Discover in Discovery?***

#### **The Process Space**

We do not generally conduct discovery in order to discover a single process. A high-level value chain like “book to build” is really a complex network of interrelated processes. It is important to understand the ways that processes feed one another. For example, the order management process feeds data to the billing process. The discovery requirement is therefore to identify all the relevant processes and to understand their integration points. One effective way to organize processes is to create a hierarchy. In this way, higher level processes can be decomposed into lower level processes, in a series of layers. At the lowest layer of the hierarchy are the activities—individual process steps that are not decomposed.

#### **Process Topology**

Given a process, we first must know the individual process steps, or activities, that constitute the process. Second, we need to know the shape of the process, that is, the flow logic. This includes the process entry and exit points, sequential flow of activities, decisions, forks and joins. This information is essentially visual and we need a visual rendition to understand the topology.

#### **Process Attributes**

The visual rendition of a process does give us all the information that we need. It is important to understand the detailed attributes of the process and the attributes of each activity in the process. Without this information, it is impossible to perform simulation or to create analytical reports.

Following are examples of attributes of the process as a whole:

- Process Owner: Who is the business owner of the process?
- Purpose: What is the ultimate purpose of the process?
- Customer: On whose behalf is the process performed?
- SME: What is our source of information for the discovery?

Following are examples of attributes of each activity in the process:

- Roles: Who performs the activity?
- Resources: What tools are used?
- Data: What business data structures does the activity consume and produce?
- Duration: What is the touch time and total time of the activity?
- Description: What is done in the activity?
- Rules: What business rules govern the performance of the activity?
- SME: What is our source of information for the discovery?

In addition to these standard attributes, it may be necessary to create custom attributes that are related to the business problem being solved, such as geographical location, automation characteristics, or cost estimates.

## Part II: Why Discovery is Hard

### *The Classical Workshop*

A large, conservative insurance company needed to revamp their customer acquisition process to keep up with more nimble competitors. They hired a well-known consulting firm to help them develop a new solution. The first step was to discover the existing process, step-by-step, to understand all the functions performed by people and by their legacy systems. This was done in a classical workshop setting. A lead business modeler facilitated the workshop. She and ten business and IT experts and met every day for a month, sketching the process, listening to war stories, and arguing about what gets done at every step. During each session a junior analyst had the job of taking notes, which was itself a challenge because of the fast pace of the sessions. At the end of the month, a set of drawings with text narration was produced, impressively filling a large three ring binder.

### *Causes of Failure*

This workshop was remarkably successful. Many are not. Some are canceled prematurely. The reasons are varied:

- A credibility gap is sometimes the problem. Sometimes managers simply do not believe that discovery of current processes can be done efficiently, effectively, and on time.
- It is often impossible to get all the key business people freed from their daily responsibilities to participate in the workshop for the time that is necessary. The workshop is not viewed as top priority, unless an executive sponsor is demanding it. If a key person is missing, you may not be able to reconstruct that part of the process.
- Some participants describe aspects of the process quite clearly and confidently, but inaccurately. Other participants may be reluctant to challenge him. This is especially true if the participant is a manager, having only a high-level view of the process, with many gaps and omissions. This phenomenon of the “lead liar” is familiar to many from politics.
- Other participants may be fearful of describing aspects of the process that are, in reality, embarrassingly dysfunctional or inept. Some participants may fear that their reputations would suffer. Departmental politics may rear its ugly head.
- It is hard to capture the necessary detail about the process and it takes too long to hold the entire group. Eventually people are exhausted. They do not have the stamina or the time to keep going and so the workshop comes to an end.
- Even if the necessary detail were available, it is hard to do the bookkeeping. There are numerous linkages of data and metadata with the process and the task keeps getting harder and harder as more of the process is discovered.
- Documenting the processes, with all the required views, data, drawings, and analytical reports is a major production challenge, requiring a multiple levels of review and approval, to identify missing information, enforce consistency, and keep information up to date. Document management and publication can easily consume one-third to one-half the total effort.

### *Why Not Skip Discovery?*

The mantra of the early 1990's was Business Reengineering. It taught that we should not bother with discovery: it takes too long, provides sketchy information of dubious value, and anyway the current process is probably so broken that it would be better to start with a clean slate.

The experiences of the last five years have shown this thinking to be false in many cases. While it is true that some processes must be worked on *ab initio*; in most cases, the current processes embody procedures and data that are critically important. The To Be process must start with the As Is process to ensure completeness and correctness.

### Part III: Discovery Methodologies

#### *Discovery Roles*

The key players in process discovery are:

- The Sponsor
- The Subject Matter Expert (SME)
- The Analyst

The sponsor, typically a business executive, charters the Business Process Analysis project, deciding its scope and the goals. The sponsor explains the current business problems, identifies processes that need to be discovered, and establishes the timetable for the overall project. The sponsor is critically important to ensure the success of the discovery, because a strong sponsor will be able to make the required resources and business information available to the analysts.

The SME's are the sources of information about the process. A SME can be a business manager or a worker on the front line, who performs process tasks day after day. A SME can also be an IT professional, especially when the process is heavily automated.

The analyst is typically a person with some industrial engineering background but may also come from the IT organization. The analyst is responsible for compiling, organizing, analyzing, and presenting the information gathered from the SME's. Discovery will typically require multiple sessions between the analysts and the SME's.

#### *Discovery Sessions*

Discovery sessions can be approached in a variety of different ways. We will classify the main approaches as follows:

- Centralized vs. Distributed
- Top-Down vs. Bottom Up
- Free Form vs. Structured

#### **Centralized vs. Distributed**

In the centralized methodology the Analyst assembles multiple SME's together for a series of workshop sessions. Facilitating the group, the Analyst constructs the processes, piece by piece. This approach has many advantages. Properly run by a skilled Analyst, the workshop can uncover aspects of the process from the interaction of the SME's. The analyst can "fill in the missing pieces" in real time. As we have seen above there are many disadvantages, including the logistical problems of assembling the right people (especially if they are managers), the lack of time, and the phenomenon of the lead-liar. Another disadvantage is that the quality of the discovery sessions depends critically on the skill and experience of the Analyst.

In the distributed methodology the Analyst obtains information from SME's in separate discovery sessions. In effect, the Analyst captures a series of "process fragments," which are portions of an entire end-to-end process. An important advantage of the distributed methodology is that it is inherently democratic—the Analyst will get everyone's vote, which means that information will surface that might be suppressed in the group setting.

The disadvantage of this approach is obviously that all this disparate information must be fit together. The key challenge is how to resolve the inconsistencies between SME's.

### **Top-Down vs. Bottom Up**

A classic question of methodology is whether to work top down or bottom up. The top down approach to discovery begins with large scale processes and divides them into smaller scale processes. For example, we might start with the "Order to Cash" process and decompose it into a series of constituent processes, like Manage Customer Contact, Enter Order, Process Order, Bill Customer, and Receive Payment. Each of these is still a very large process that will need to be decomposed further. In this way we arrive at a hierarchy of processes. Ultimately we reach a process at the activity level, in which simple non-decomposable activities are performed. The main advantage of this approach is that it ensures breadth of scope. We know that the analyst is covering all the key areas that need to be addressed. The disadvantage is that we may miss details. The top-down methodology will not find activities or processes unless we know in advance about their existence. We could therefore end up with inaccuracies or missing steps.

The bottom up approach begins with the SME's themselves. The SME's offer information about what they do in a process to the analysts. The Analysts gather this local information and assemble it into a process context. Of course the main advantage of this bottom up approach is richness of detail. You are getting the information directly from the horse's mouth. There are, however, serious disadvantages. How do you know when you have interviewed enough SME's? So, completeness may be a problem. Another problem is more subtle. How does the analyst create a process context from all this local information? It is the task of fitting the details into a comprehensible set of processes.

### **Free Form vs. Structured**

In the free form style of discovery, the SME's informally describe the process, its issues, examples, war stories, etc. After the raw information is captured, the analyst will structure this information into a more appropriate format, for example into visual diagrams. The advantage is that the SME is completely uncensored and can free associate about the process. Unexpected and possibly valuable insights may emerge. The disadvantage is the enormous amount of time required to structure this information and the meager signal-to-noise ratio. In addition, it will be nearly impossible to ensure consistency in the level of information received from one SME to the next.

In the structured style, the SME responds to a pre-defined set of questions, structured and organized to ensure consistency. This information can be provided interactively, with the analyst guiding the SME, or the SME may answer the questions offline. The SME may answer the questions in a form or may construct a diagram of the process. The advantage of a structured approach is that you ensure consistency and completeness of detail. However, the analyst may miss valuable insights that do not fit into the form.

### ***Critical Needs in Tooling***

To make the discovery methodology work effectively, proper tools are required. Here we highlight two critical capabilities that the tooling should provide.

### **Validation**

In all approaches we have been discussing, there is the underlying problem of accuracy: How do you know the information you have discovered is consistent and complete? This is especially problematic in the distributed approach to discovery. For example, one SME describes performing an activity and then sending the Purchase Order to a second SME for the next step in the process. Suppose that the second SME does not acknowledge receiving the Purchase Order. The analyst must resolve this inconsistency. A variant of this occurs when the second SME acknowledges receiving a "PO." The problem is now to decide if "PO" is the same as a "Purchase

Order.” In either case, the analyst must first know that the inconsistency exists. Correction will often require going back to the SME for supplementary information. The tool must therefore be capable of finding and classifying all inconsistent and incomplete information in our processes. It must then enable the analyst to correct such problems easily and quickly. Without this validation, the processes will lack true integrity.

### **Integration**

In many of these approaches, the analyst will face the challenge of fitting process fragments into a coherent whole. This is especially true in the bottom up style of process discovery. For example, the analyst may receive hundreds of process drawings that need to be fit together into an end-to-end process. This is clearly a job for a tool. The capability that the tool must provide is to understand how process fragments connect with one another and then integrate the fragments into a coherent process that can be analyzed.

### ***Methodology Recommendations***

Methodology is not religion. There is no single correct discovery methodology that is guaranteed to work on all business problems in all organizations. With that caveat in mind, we will now describe an approach to discovery that we have found to be effective in many situations. The approach is a hybrid. It consists of the following elements:

1. Start by establishing a top down structure in the form of a process hierarchy. This hierarchy will serve as the guiding context in which to ensure completeness of scope.
2. The analysts should perform bottom up interviews with SME’s using a structured interview technique. This will provide richness of detail.
3. After validating the information and integrating the process fragments, the analyst should schedule centralized reviews of the end-to-end processes with the SME’s and the key stakeholders. These reviews will verify accuracy and ensure alignment.

It is important to point out that this recommended approach assumes that the tooling supports the critical requirements for process validation and integration.

## **Part IV: Gaining Insight from Discovery**

During discovery, we amass an impressive array of facts about the organization’s current processes, but we need to bear in mind that the purpose of discovery is to provide insight into the underlying business problems that motivated the discovery. This insight must serve as the foundation for a diagnosis of the As Is processes and as a basis for the design of the To Be processes.

We therefore need to tap into the knowledge base generated in discovery. From this knowledge base we need to create work products that furnish this insight. These work products present the knowledge in structures and formats that highlight different aspects of the processes. The two most important types are visualization work products and analytical work products.

### ***Visualization Work Products***

Process visualization means the presentation of process knowledge in a diagrammatic form, typically consisting of blocks linked by arrows.

## *Types of Process Diagrams*

### **Activity Flow Diagrams**

Activity Diagrams illustrate the process topology. They display activities as blocks or icons interconnected by arrows. As one looks from left to right (or from top to bottom), one sees how the process starts, how the flow of work progresses, the decision points at which the process changes, and finally how the process comes to an end. This is the best known type of process diagram.

### **Information Flow Diagrams**

Information Flow Diagrams focus on the information exchanged between activities. They show where business data structures are used in the process flow and how they change state. These diagrams are typically elaborations of activity flow diagrams.

### **Resource Diagrams**

Resource Diagrams focus on the people or systems that perform the activities in the process. They show the interfaces between these resources. For example, it will show where one system interconnects with another system, or where there is a handoff from one role to another role.

## *Process Views*

Diagrams provide the greatest insight if they can be viewed ways that provide different perspectives. Following are several examples of views that can be superimposed on the diagram types just described:

### **Hierarchical View**

Processes can be defined at different level of resolution. The hierarchical view allows one to display a process as a flow of sub-processes. Instead of each block representing an activity, it represents a sub-process, or collection of interrelated activities.

### **End to End View**

Sometimes what is needed is to view the elementary activities, instead of the sub-processes. In this case, we need to show the end-to-end flow of the process. A process may be composed of 6 sub-processes, each containing 8 activities. The end to end view would therefore display 48 activities.

### **Swim-lane Views**

It is quite useful to visualize a process structured into a series of horizontal or vertical swim-lanes so that each swim-lane represents activities performed by a specific Role. Thus, the first swim-lane might contain activities performed by the Customer; the second swim-lane might contain activities performed by the Customer Service Rep; and so on. It is often helpful to change the swim-lane schema. For example, we might want the swim-lanes to represent organizations instead of roles. Or we might want the swim-lanes to represent technologies employed. Or we might want the swim-lanes to represent the locations in which the activities are performed. Or we might want the swim-lanes to represent Value Added and Non-Value Added activities, and so forth.

### **Path Analysis**

This is a technique of visualization in which a specific path through the process is highlighted, for example, displayed in red. An important special case is the critical path; viz., the path that statistically determines the overall cycle time of the process.

### Highlighting

In this technique, all activities that meet a specific criterion are shown in a contrasting color; for example, displayed in green. An important example is to highlight all activities that are manual or all activities that are non-value added.

Obviously these views can be combined for greater insight. We may choose to view a process organized into swim-lanes by role and to highlight all manual activities. Or we may want to display a hierarchical view, showing process interconnection in which swim-lanes are based on process owner. Another important technique is to compare two processes by showing them in swim-lanes. For example, the upper swim-lane shows Process 1 and the lower swim-lane shows Process 2. The user can easily see the similarities as well as the differences between the processes

### Analysis Work Products

In visualization we considered one process at a time. This is not sufficient for many reasons. It is often required to extract information from many processes and integrate this information into a report. A simple example is a job description. To define the job for the Customer Service Rep (CSR), it is necessary to know all the activities that for which the CSR is responsible. Therefore, we need a report that will extract from all the processes, only the activities performed by the CSR. We cannot get this from any single process.

The analyst needs the ability to generate reports that extract information from the knowledge base, organize this information, filter it based on defined criteria, and perform various calculations. These reports can be used for diagnosing the As Is and for comparing the To Be with the As Is.

Here are a few examples of analytical reports:

1. Job Description: What are the activities for which a role is responsible?
2. Functional Specification: What are the functional requirements of a system?
3. Cost Analysis: What is the labor cost of a process?
4. Cycle Time Analysis: What is the end-to-end duration of a process?
5. Touch Point Report: What are all the touch-points between the Customer and the Customer Service Rep?
6. Handoffs: What are all the handoffs between roles or between systems? What information is communicated at each point?
7. Gap Analysis: What are all the gaps between the As Is and the To Be processes?

In many cases, deterministic analyses are not sufficient and we need to simulate the process. Simulation is used to identify bottlenecks in the process, predict the behavior of the process under different loads, estimate resource utilization, and analyze cycle time and throughput in a dynamic, stochastic way that reflects reality much more accurately than deterministic reports.

### Static and Dynamic Views

The required process views and analytical reports may not be known at the beginning of the discovery. It is therefore useful if the tooling allows these views and reports to be generated dynamically. Without this capability, the analyst must draw each static diagram. For example, drawing each swim-lane diagram individually would represent a considerable effort. It is time-consuming and can lead to errors. Ideally, diagrams are automatically generated directly from the knowledge base. This approach is called "parametric" since the user instructs the software on the required parameters of the view (swim-lanes, information flow, highlighting, etc.) and the software produces the required view based on the data in the knowledge base.

Similarly, the user should be able to define any report and generate the report directly from the knowledge base. The report can be run many times. Each time it is run, it extracts the data dynamically from the knowledge base and therefore represents current process definitions.

### **Communication**

The analyst will need to communicate and share the facts and insights of process discovery. This can be done in different ways. One way is to publish work products in document form and email them to stakeholders. Another way is to publish work products in HTML form and make them available over the web. This will allow the stakeholders to access process diagrams and analytical reports from their browser. Ideally, the publication can be personalized, so that users are given the views that are directly relevant to their projects and positions. Portal technology is particularly attractive for this purpose as the portal provides the central point of communication in the enterprise.

## **Part V: Lessons Learned**

Process discovery is a key element in Business Process Analysis. With the right methodology and appropriate tooling, discovery can provide a clear, complete, accurate, and actionable foundation for process design and optimization. In view of the preceding discussion, we can summarize the key points as follows.

- Sponsorship is Key

It is important to enlist the active cooperation and participation of many individuals in the organization. Therefore you must secure the appropriate level of sponsorship.

- Scoping

Don't try to boil the ocean. Begin by focusing on a key business area within the organization, with a clearly defined problem, objective, and timeframe. Later you can expand the scope.

- Start with the End in Mind

Be aware of the business problem, the scope, and the goal that the sponsor defined at the start of the discovery project. This will prevent you from acquiring facts that may be irrelevant to the problem at hand.

- Methodology

Choose a methodology that is appropriate for the organization and for your problem. Our recommendation is the hybrid Top Down + Bottom Up. It provides the overall business context as well as capturing the process details. Whatever methodology you choose must be thoroughly understood, agreed to, and clearly communicated to your sponsor and to all participants in the process activities.

- Engage Stakeholders

During the interviews and analysis, make sure you have interviewed the key owners and participants in the processes. Be sure to communicate with upstream and downstream process owners as well. Ensure that the IT Group participates as required to capture data and system attributes. Review end-to-end processes and key analytical results with the relevant stakeholders for their review and validation.

- Tooling

Tooling is critically important. Your choice of business process discovery tool should be based upon your methodology and upon the work products you want to create. Don't try this with simple office products (Visio and Word). Select the product that provides the tooling capabilities (e.g., validation and integration) that support your methodology.

---

**Laury Verner** is CTO of ProActivity, a Boston-based software company that specializes in business process discovery, analysis, and design. Prior to that, Laury was a partner at CSC Consulting. Laury has a Ph.D. from Johns Hopkins University and a B.A. from the University of Pennsylvania. He can be reached at [Lauryv@proactivityinc.com](mailto:Lauryv@proactivityinc.com).