Chapter 5
Process Discovery

All truths are easy to understand once they are discovered;
the point is to discover them.
Galileo Galilei (1564–1642)

The previous chapters showed how to create a BPMN model. This chapter goes further by showing how to create models that are both correct and complete. To this end, one needs to thoroughly understand the operation of a business process, and one needs to possess the technical skills to represent it in an appropriate BPMN model. These two types of skill are hardly ever unified in the same person. Hence, multiple stakeholders with different and complementary skills are typically involved in the construction of a process model.

This chapter presents the challenges faced by the stakeholders involved in the lead-up to a process model. Then, we discuss methods to facilitate effective communication and information gathering in this setting. Given the information gathered in this way, we show step by step how to construct a process and what criteria should be verified before a process model is accepted as an authoritative representation of a business process.

5.1 The Setting of Process Discovery

Process discovery is defined as the act of gathering information about an existing process and organizing it in terms of an as-is process model. This definition emphasizes gathering and organizing information. Accordingly, process discovery is a much broader activity than modeling a process. Clearly, modeling is a part of this activity. The problem is though that modeling can only start once enough information has been put together. Indeed, gathering information often proves to be cumbersome and time-consuming in practice. Therefore, we need to first define a setting in which information can be gathered effectively. In order to address these issues, we can describe four phases of process discovery:

1. Defining the setting: This phase is dedicated to assembling a team in a company that will be responsible for working on the process.

M. Dumas et al., *Fundamentals of Business Process Management*,
DOI 10.1007/978-3-642-33143-5_5, © Springer-Verlag Berlin Heidelberg 2013

2. Gathering information: This phase is concerned with building an understanding of the process. Different discovery methods can be used to acquire information on a process.
3. Conducting the modeling task: This phase deals with organizing the creation of the process model. The modeling method gives guidance for mapping out the process in a systematic way.
4. Assuring process model quality: This phase aims to guarantee that the resulting process models meet different quality criteria. This phase is important for establishing trust in the process model.

Typically, one or several *process analysts* are responsible for driving the modeling and analysis of a business process. Often, the process analyst is not familiar with all details of the business process. The definition of the setting of process discovery is critical since it helps the process analyst to secure the commitment of various domain experts for providing information on the process. These domain experts have to cover the relevant perspectives on the process. Therefore, different *domain experts* should be involved. A domain expert is any individual who has intimate knowledge about how a process or activity is performed. Typically, the domain expert is a process participant, but it can also be a process owner or a manager who works closely with the process participants who perform the process. Also suppliers and customers of the process can be considered as domain experts. The involved domain experts should jointly have insight into all activities of the process. It is the task of the process owner to secure the commitment and involvement of these persons. In the following, we will focus on the relationship between process analyst and domain expert in order to illustrate three challenges of process discovery.

5.1.1 Process Analyst Versus Domain Expert

One fundamental problem of process discovery relates to the question of who is going to model the business process. This problem is illustrated by the following exercise.

Exercise 5.1 Consider the following two tasks, and explain their difference:

- The task of modeling the process of signing a rental contract in your city.
- The task of modeling the process of getting a license plate for your car in Liechtenstein as a foreign resident.

The point of this exercise is to emphasize a potential difference in knowledge about processes. If you have already acquired some knowledge of mapping processes with BPMN by the help of this book, you will be able to create an initial process model for the rental process. The reason is that you not only have modeling

Table 5.1 Typical profile of process analyst and domain expert

Aspect	Process Analyst	Domain Expert
Modeling Skills	strong	limited
Process Knowledge	limited	strong

knowledge but also some knowledge about the domain of renting a flat in your city. The case is likely to be different for getting a license plate for a car in Liechtenstein as a foreign resident. There are only few foreign residents living in Liechtenstein, our colleague Jan vom Brocke being one of them. Most other people would not know how this process works. If we are supposed to model a process that we do not know, we have to gather an extensive amount of information about it in order to understand how it works. That is exactly the situation we are typically facing in practice: a process needs to be modeled, we have the required modeling skills, but we have only limited knowledge of the corresponding domain.

In order to describe the typical modeling situation in a plastic way, we distinguish between the role of a *process analyst* and the role of a *domain expert*. In a real modeling project, we have one or a few process analysts and several domain experts. Process analysts and domain experts have complementary roles in the act of process discovery as well as different strengths as shown in Table 5.1. The process analyst is the one who has profound knowledge of business process modeling techniques. A process analyst is familiar with languages like BPMN and skilled in organizing information in terms of a process diagram. However, process analysts have typically a limited understanding of the concrete process that is supposed to be modeled. For this reason, they depend upon the information being provided by domain experts. Domain experts have detailed knowledge of the operation of the considered business process. They have a clear understanding of what happens within the boundaries of the process, which participants are involved, which input is required, and which output is generated. On the downside, the domain expert is typically not familiar with languages like BPMN. In some companies, domain experts even refuse to discuss process models and diagrams, because they feel more comfortable by sticking to natural language for explaining what is happening in the process. As a consequence, domain experts often rely on a process analyst for organizing their process knowledge in terms of a process model.

At this stage, it has to be emphasized that the difference in modeling skills of process analysts and domain experts only results from different exposure to practical modeling and modeling training. Many companies use training programs for improving the modeling skills of domain experts. Such training is a prerequisite for modeling initiatives where process participants are expected to model processes on their own. On the other hand, there are consulting companies that specialize in a particular domain. It is an advantage when process analysts of consultancies can be assigned to modeling projects who are experts in modeling and have at least a certain level of domain expertise.

5.1.2 Three Process Discovery Challenges

The fact that modeling knowledge and domain knowledge is often available in different persons in a modeling project has strong implications. It gives rise to three essential challenges of process discovery, namely fragmented process knowledge, thinking in cases, and lack of familiarity with process modeling languages.

Exercise 5.2 An online book retailer faces a problem with its order process in terms of processing time. In order to identify the root cause of the problem, the company decides that every team involved with the order process should model its part of the process. Why could this approach be problematic?

The first challenge of process discovery relates to *fragmented process knowledge*. Business processes define a set of logically related activities. These activities are typically assigned to specialized participants. This has the consequence that a process analyst needs to gather information about a process not only by talking with a single domain expert, but with several domain experts who are responsible for the different tasks of the process. Typically, domain experts have an abstract understanding of the overall process and a very detailed understanding of their own task. This makes it often difficult to puzzle the different views together. In particular, one domain expert might have a different idea about which output has to be expected from an upstream activity than the domain expert actually working on it. Potential conflicts in the provided information have to be resolved. It is also often the case that the rules of the process are not explicitly defined in detail. In those situations, domain experts may operate with diverging assumptions, which are often not exactly consistent. Fragmented process knowledge is one of the reasons why process discovery requires several iterations. Having received input from all relevant domain experts, the process analyst has to make proposals for resolving inconsistencies, which again requires feedback and eventually approval of the domain experts.

The second challenge of process discovery stems from the fact that domain experts typically *think of processes on a case level*. Domain experts will find it easy to describe the activities they conducted for one specific case, but they might have problems responding to general questions about how a process works in the general way. Process analysts often get answers like "you cannot really generalize, every case is different" to such a question. It is indeed the task of the process analyst to organize and abstract from the pieces of information provided by the domain expert in such a way that a systematically defined process model can emerge. Therefore, it is required to ask specific questions about what happens if some task is completed, what if certain conditions do or do not hold, and what if certain deadlines are not met. In this way, the process analyst can reverse engineer the conditions that govern the routing decisions of a business process.

The third challenge of process discovery is a result of the fact that domain experts are typically *not familiar with business process modeling languages*. This observation already gave rise to the distinction of domain experts and process analysts. In this context, the problem is not only that domain experts are often not trained to

create process models themselves, but also that they are not trained to read process models that others have created. This lack of training can encumber the act of seeking feedback to a draft of a process model. In this situation it is typically not appropriate to show the model to the domain expert and ask for corrections. Even if domain experts understand the activity labels well, they would often not understand the sophisticated parts of control flow captured in the model. Therefore, the process analyst has to explain the content of the process model in detail, for example by translating the formal notation of the process model to a natural-language description with the same meaning. Domain experts will feel at ease in responding to these natural-language explanations, pointing out aspects that need modification or further clarification according to their understanding of the process.

5.1.3 Profile of a Process Analyst

The skills of a process analyst play an important role in process discovery. Expert process analysts can be described based on a set of general dispositions, their actual behavior in a process analysis project, and in terms of the process resulting from their efforts.

Exercise 5.3 You are the manager of a consulting company, and you need to hire a person for the newly signed process analysis project with an online book retailer. Consider the following two profiles, who would you hire as a process analyst?

- Mike Miller has ten years of work experience with an online retailer. He has worked in different teams involved with the order process of the online retailer.
- Sara Smith has five years of experience working as a process analyst in the banking sector. She is familiar with two different process modeling tools.

Research on expertise in the general area of system analysis and design has found that there are certain personal traits that are likely to be observed for expert process analysts. Apparently, there seems to exist a set of certain personal dispositions that help in becoming an expert in process analysis. One of the ways to describe personality is the so-called *Five Factor Model* developed in psychological research. In essence, this model contains the dimensions openness (appreciating art, emotion, and adventure), conscientiousness (tendency to self-discipline, achievement and planning), extraversion (being positive, energetic, and seeking company), agreeableness (being compassionate and cooperative), and neuroticism (being anxious, depressed and vulnerable). These factors have also been studied regarding their connection with expert analysts. These experts appear to be strong both in terms of conscientiousness and extraversion. Indeed, process discovery projects require a conscientious planning and coordination of interviews with various domain experts in a limited period of time. Furthermore, process discovery projects are sometimes subject to enterprise-internal politics in situations where the agenda of different process stakeholders is not thoroughly clear or where stakeholders might fear losing

their position. In such an environment, it is valuable to have an energetic and extraverted process analyst involved who is able to create a positive atmosphere for working on the project.

Process discovery in general belongs to the category of ill-defined problems. This means in the beginning of a process discovery project, it is not exactly clear who of the domain experts have to be contacted, which documentation can be utilized, and which agenda the different stakeholders might have in mind. The way how expert analysts navigate through a project is strongly influenced by experiences with former projects. Therefore, there is a strong difference between the way how novices and expert analysts conduct problem understanding and problem solving. In terms of problem understanding, it has been observed that expert analysts approach a project in terms of what are the things that need to be achieved. Novices lack this clear goal orientation, and try to approach things in a bottom-up way. This means, they often start by investigating material that is easily accessible and talk to persons that readily respond. Experts work in a different way. They have an explicit set of triggers and heuristics available from experiences with prior projects. They tend to pay specific attention to the following aspects:

- Getting the right people on board. If you need to talk to a given process participant, make sure their immediate supervisor and the one above them is on board and that the process participant knows that their hierarchy backs their involvement in the process discovery effort.
- Having a set of working hypotheses on how the process is structured at different levels of details. In order to progress with the project, it is important to have a short and precise set of working hypotheses, which they step-by-step challenge. Prepare a extensive set of questions and assumptions to be discussed in workshops or interviews.
- Identifying patterns in the information provided by domain experts. These can be utilized for constructing parts of a process model. Such pieces of information typically refer to specific control structure. For instance, statements about certain activities being alternative, exclusive, or subject to certain conditions often point to the usage of XOR-gateways. In a similar way, statements about activities being independent of another, or sometimes being in one or another order, often suggest concurrency. For their knowledge of such patterns, it is often easy for expert analysts to sketch out processes.
- Paying attention to model aesthetics. Models have to look nice to be engaging to a wide audience. This does not only help to have a resulting model that is easy to understand by stakeholders, but also valuable throughout the process of creating the model. Experts also use the right level of abstraction. For example, you should not show a super-detailed model to an executive-level manager. The importance of layout is apparent from the fact that expert analysts often take half of the time while creating a model for repositioning its elements in a meaningful way.

5.2 Discovery Methods

As we have now a rough idea of the tasks that a process analyst has and which capabilities and limitations he has to keep in mind when interacting with domain experts, we turn to different techniques for gathering information about a process. In general, we distinguish three classes of discovery techniques, namely evidence-based discovery, interview-based discovery, and workshop-based discovery. There are strengths and limitations, which we will discuss subsequently.

Exercise 5.4 Imagine you would be assigned the task of modeling the process of how a book order is processed by your favorite online book retailer. How can you systematically gather the required pieces of information about this process?

5.2.1 Evidence-Based Discovery

Various pieces of evidence are typically available for studying how an existing process works. Here, we discuss three methods: document analysis, observation, and automatic process discovery.

Document analysis exploits the fact that there is usually documentation material available that can be related to an existing process. However, there are some potential issues with document analysis. First, most of the documentation that is available about the operations of a company is not readily organized in a process-oriented way. Think of an organization chart, for instance. It defines the departments and positions, it is helpful to identify a potential set of process stakeholders. Such material can help to structure phases of a process. For example, in case of our online book retailer, it might reveal that the sales department, the logistics department and the finance department are likely to be involved with the book order. Second, the level of granularity of the material might not be appropriate. While an organization chart draws rather an abstract picture of a company, there are often many documents that summarize parts of a process on a too fine-granular level. Many companies document detailed work instructions for tasks and work profiles for positions. These are typically too detailed for modeling processes. Third, many of the documents are only partially trustworthy. For a process discovery project, it is important to identify how a process works in reality. Many documents do not necessarily show reality. Some of them are outdated and some state how things should work idealistically, and not how people conduct them in reality. The advantage of document analysis is that a process analyst can use them to get familiar with certain parts of a process and its environment, and also to formulate hypotheses. This is helpful before talking to domain experts. On the downside, a process analyst has to keep in mind that documents do not necessarily reflect the reality of the process.

If we use *observation* as a method of discovery, we directly follow the processing of individual cases in order to get an understanding of how a process works. The process analyst can either play the active role of a customer of a process or the

passive role of an observer. As part of the *active customer role*, the process analyst triggers the execution of a process and records the steps that are executed and the set of choices that are offered. For instance, in the case of the online book retailer, the analyst can create a new book order and keep track of which activities are performed at the side of the retailer. This provides a good understanding of the boundaries of the process and its essential milestones. However, the analyst will only see those parts of the process that require interaction with the customer. All backoffice processing remains a black box. The role of a *passive observer* is more appropriate for understanding the entire process, but it also requires access to the people and sights where the process is being worked on. Usually, such access requires the approval of the managers and supervisors of the corresponding teams. Furthermore, there might be a potential issue with people acting differently, because they are aware of being observed. People usually change their behavior under observation in such a way that they work faster and more diligently. This is important to be kept in mind when execution times have to be estimated. However, discovery based on observation has the advantage that it reveals how a process is conducted in reality today, which is in contrast to document analysis that typically captures the past.

A third option of *automatic process discovery* emerges from the extensive operational support of business processes provided by various information systems. Automatic process discovery makes use of event logs that are stored by these information systems. Such event data have to be recorded in such a way that each event can be exactly related to three things: an individual case of the process, a specific activity of the process, and a precise point in time. If these three pieces of information are available in the event logs, then automatic process discovery techniques can be used to reconstruct the process model, for example for the online book retailer. Since this approach shares some characteristics with data mining, where meaningful information is extracted from fine-granular data, these techniques of automatic process discovery are subsumed to the research area of process mining. The advantage of automatic process discovery is that event logs capture the execution of a process very accurately including information about execution times. A limitation is though that some log information can be misleading. This may be the case if a system crashes such that logs are not stored correctly. These failure types relating to a flawed storage of event logs are summarized with the term noise. Furthermore, the models resulting from process mining may not be directly understandable. Process behavior can be very complex, such that the generated models are hardly readable. In such a case, the logs have to be filtered or clustered for getting models that help understanding the process.

5.2.2 Interview-Based Discovery

Interview-based discovery refers to methods that build on interviewing domain experts about how a process is executed. With these methods, we have to explicitly take into account the challenges of process discovery, namely the fact that process

knowledge is scattered across different domain experts, that domain experts typically think in terms of individual cases, and that domain experts are often not familiar with business process modeling languages. This has implications for how the interviews can be scheduled, and which phases and iterations are required.

Exercise 5.5 Consider that the order process of your favorite online book retailer has ten major activities that are conducted by different persons. How much time do you need approximately for creating a process model that is validated and approved by the process owner? Make appropriate assumptions.

We have mentioned that process knowledge is typically fragmented due to specialization and division of labor. For this reason, interviews have to be conducted with various domain experts involved in the process. As the process analyst might not yet understand the details of the involvement of different domain experts, it might be required to discover the process step by step. There are two strategies available for scheduling interviews: starting backwards from the products and results of the process and starting at the beginning by proceeding forward. Conducting interviews in a forward way permits to follow the flow of processing in the order of how it unfolds. This is particularly helpful for understanding which decisions are taken at a given stage. However, following the processing in a backward way has also advantages. People working in a process require certain input to be available for conducting their work, and this perspective makes it easy to consider what has to be achieved before a specific activity can be conducted. Both perspectives, the downstream and the upstream perspective are important when interviewing domain experts. With each interview partner, it must be clarified which input is expected from prior upstream activities, which decisions are taken, and in which format the results of an activity are forwarded to which subsequent party.

The discovery challenges emphasize that the expertise of the process analyst is required for abstracting information on how individual cases are executed in order to construct meaningful process models. Typically, the process analyst gathers information about the process in interviews and later organizes the material offline before constructing an initial process model. As a consequence, interviewing a domain expert is often conducted in different iterations. After an initial interview, the process analysts creates a draft process model, which is then discussed with the domain expert in terms of correctness and completeness. Here, it is important to ask what happens if something goes wrong or how unexpected cases are handled. This feedback interview usually triggers another round of rework. In some cases, the second feedback round leads to an approval of the process model. In other cases, a third feedback round is required for checking the reworked process model again. Often, domain experts feel more comfortable with free-form interviews where they can discuss the process at a level of detail that they find appropriate. Structured interviews, in contrast, can create a feeling of running through a checklist, with the effect that domain experts hold back important information that they are not explicitly asked for.

It is a strength of interview-based discovery that the interview situation provides a rich and detailed picture of the process and the people working in it. It has the

potential to reveal inconsistent perceptions that different domain experts may have on how a process operates. It also helps the process analyst to understand the process in detail. However, it is a labor-intensive discovery method. Several iterations are required for arriving at a point where domain experts feel comfortable with how a process is described in a process model.

One recurrent pitfall of interviews is that when asked how a given process or activity is performed, the interviewee tends to describe the normal way of processing. Thus, exceptions tend to be left aside. In other words, the interview ends up covering only the "sunny-day" scenario. One way to prevent this pitfall is to reserve time during the interview to focus on the "rainy-day" scenarios. Questions that can be used to spark discussion on the rainy-day scenario are: "How did you handle your most difficult customer?", "What was the most difficult case you have worked on?". This technique allows one to uncover variations or exceptions in the process that, while not necessarily frequent, have a sufficient impact on the process to be worth documenting.

5.2.3 Workshop-Based Discovery

Workshop-based discovery also offers the opportunity to get a rich set of information on the business process. Although this is not always the case, the setting can be organized in such a way that the contributions to the discussion are immediately used to model the process. In contrast to interviews, it not only involves more participants, but also a bigger set of roles. Additional roles are required for facilitating the discussion and for operating the process modeling tool. The *facilitator* takes care of organizing the verbal contributions of the participants. The *tool operator* is responsible for directly entering the discussion results into the modeling tool. Several *domain experts* also participate, as much as the *process owner* and the *process analyst*. The involvement to this extensive set of persons requires diligent preparation and scheduling. Furthermore, the process will not be sketched out in detail in only one session. It can be expected that three to five half-day sessions are required.

At the start of a process discovery effort, when there is not yet information available for modeling the process, it can be beneficial to take a more lightweight and participative approach to organizing the workshops. One technique to engage the workshop participants in the discovery effort is by asking workshop participants to build a map of the process using sticky notes. The facilitator starts with a pad of sticky notes. Each sticky note is meant to represent a task or event. The group starts to discuss how the process typically starts. The facilitator then writes the name of the (supposedly) first task or event into a sticky note and posts it on the wall. Then the facilitator asks what can happen next. The participants start mentioning one or more possible tasks. The facilitator writes these activities in new sticky notes and starts posting these on the wall, organizing them for example from left to right or top to bottom to capture the order of the activities. At this stage no lines are drawn between the tasks and no gateways are discovered. The purpose of this exercise is to build

Table 5.2 Relative strengths and limitations of process discovery methods

Aspect	Evidence	Interview	Workshop
Objectivity	high	medium-high	medium-high
Richness	medium	high	high
Time Consumption	low-medium	medium	medium
Immediacy of Feedback	low	high	high

a map of activities and their temporal ordering. Sometimes, participants disagree on whether something is one task or two tasks. If the disagreement cannot be resolved, the two tasks can be written as two sticky notes and these two related sticky notes are pasted next to each other. The facilitator also needs to pay attention to the fact that the tasks being posted should be at the same level of detail. When people start mentioning small micro-steps, like "putting the document on a fax machine" the facilitator should try to lift the level of abstraction. In the end, this exercise leads to a rough map that the process analyst can take as input for constructing an initial model.

Exercise 5.6 Consider the following two companies. Company A is young, founded three years ago, and has grown rapidly to a current toll of 100 employees. Company B is owned by the state and operates in a domain with extensive health and security regulations. How might these different characteristics influence a workshop-based discovery approach?

Workshop-based process discovery requires an organized facilitation and an atmosphere of openness. In terms of facilitation, the facilitator has to ensure that the parole is balanced between the different participants. This means on the one hand restricting the speech time of talkative participants. On the other hand, more introverted participants should be encouraged to express their perspective. An atmosphere of openness is helpful for having everybody participate. This aspect is influenced by the culture of the company. In organizations with a strongly emphasized hierarchy, it might be difficult for domain experts to express their view openly if their supervisor is present. If creativity and independent thinking is appreciated in the company, the participants are likely to feel at ease with discussing issues. It is the responsibility of the facilitator to stimulate a constructive workshop interaction in both cases. In this case, workshops have the potential to resolve inconsistencies directly with all involved parties.

5.2.4 Strengths and Limitations

The different methods of process discovery have strengths and limitations. These can be discussed in terms of objectivity, richness, time consumption, and immediacy of feedback (see Table 5.2).

- Objectivity: Evidence-based discovery methods typically provide the best level of objectivity. Existing documents, existing logs and observation provide an unbiased account of how a process works. Interview-based and workshop-based discovery both have to rely on the descriptions and interpretations of domain experts who are involved with the process. This bears the risk that those persons may have perceptions and ideas of how the process operates, which may be partially not correct. Even worse, the process analyst is also at risk that domain experts might opportunistically hide relevant information about the process. This can be the case if the process discovery project happens in a political environment where groups of process stakeholders have to fear loss of power, loss of influence, or loss of position.
- Richness: While interview-based and workshop-based discovery methods show some limitations in terms of objectivity, they are typically strong in providing rich insights into the process. Domain experts involved in interviews and workshops are a good source to clarify reasons and objectives for why a process is set up as it is. Evidence-based methods might show issues that need to be discussed and raise questions, but they often do not provide an answer. Talking to domain experts also offers a view into the history of the process and the surrounding organization. This is important for understanding which stakeholders have which agenda. Evidence-based discovery methods sometimes provide insight into strategic considerations about a process when they are documented in white papers, but they hardly allow conclusions about the personal agendas of the different stakeholders.
- Time consumption: Discovery methods differ in the amount of time they require. While documentation of a company and a particular process can be easily made available to a process analyst, it is much more time-consuming to conduct interviews and workshops. While interview-based discovery suffers from several feedback iterations, it is difficult to schedule workshops with various domain experts on short notice. Automatic process discovery often involves a significant amount of time for extracting, reformatting, and filtering of event logs. Passive observation also requires coordination and approval time. Therefore, it is a good idea to start with document analysis, since documentation can often be made accessible on short notice.
- Immediacy of feedback: Those methods that directly build on the conversation and interaction with domain experts are best for getting immediate feedback. Workshop-based discovery is best in this regard since inconsistent perceptions about the operation of a process can be directly resolved by the involved parties. Interviews offer the opportunity for asking questions whenever process-related aspects are unclear. However, not all issues can be directly resolved with a single domain expert. Evidence-based discovery methods raise various questions about how a process works. These questions can often only be answered by talking to domain experts.

Since each discovery method has strengths and limitations, it is recommended to utilize a mixture of them in a discovery project. The process analyst typically starts with documentation that is readily available. It is essential to organize the project in such a way that the information can be gathered from the relevant domain experts

in an efficient and effective way. Interviews and workshops have to be scheduled during the usual work time of domain experts. Therefore, they have to be motivated to participate and involved in such a way that it is the least time-consuming for them. Once issues arise about specific details of a process, it might be required to turn back to evidence-based discovery methods.

Exercise 5.7 In what situations is it simply not possible to use one or more of the described discovery methods?

5.3 Process Modeling Method

Modeling a process in the discovery phase is a complex task. Therefore, it is good to follow a predefined procedure in order to approach this task in a systematic way. One way to do so is to work in five stages, as follows:

1. Identify the process boundaries
2. Identify activities and events
3. Identify resources and their handovers
4. Identify the control flow
5. Identify additional elements

5.3.1 Identify the Process Boundaries

The identification of the process boundaries is essential for understanding the scope of the process. Part of this work might have been done already with the definition of a process architecture. Technically, this means we need to identify the events that trigger our processes and those that identify the possible process outcomes. For example, let us consider again the order fulfillment process that we modeled in Chap. 3. We observe that this process is triggered by the receipt of a purchase order from the customer, and completes with the fulfillment of the order as an outcome. These two events mark the boundaries of this process. Accordingly, we use a start message event and an end event in BPMN to represent them. If our process would have had negative outcomes, we would have modeled these via terminate end events.

5.3.2 Identify Activities and Events

The goal of the second step is to identify the main activities of the process. The advantage of starting with the activities is that domain experts will clearly be able to state what they are doing even if they are not aware of working as part of an overarching business process. Also documents might explicitly mention activities,

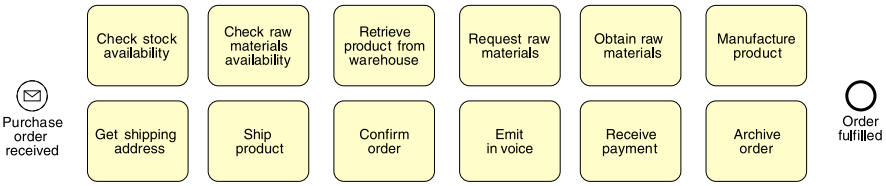

Fig. 5.1 The main activities and events of the order fulfillment process

for instance a set of work instructions. In the case of the order fulfillment process, this stage may lead to a set of activities for which the order and routing conditions are not yet defined. In this step, we also need to identify the events that occur during the process, which we will model with intermediate events. Figure 5.1 lists the 12 activities of our example.[1] Note that this initial set of activities and events may undergo revisions, e.g. more activities may be added as we add more details into our model. If the process is too complex, we suggest to focus on the main activities and events only at this stage, and add the others at a later stage when a deeper understanding of these elements and their relations has been gained.

5.3.3 Identify Resources and Their Handovers

Once we have defined the set of main activities and events, we can turn to the question of *who* is responsible for them. This information provides the basis for the definition of pools and lanes, and the assignment of activities and events to one of these pools and lanes. At this stage, the order of the activities is not defined yet. Therefore, it is a good step to first identify those points in the process where work is handed over from one resource to another, e.g. from one department to the other. These handover points are important since a participant being assigned a new task to perform, usually has to make assumptions about what has been completed before. Making these assumptions explicit is an essential step in process discovery. Figure 5.2 shows the set of activities and events of the order fulfillment process now being assigned to pools and lanes. The sequence flows indicate handover points. The handover points also help to identify parts of the process which can be studied in isolation from the rest. These parts can be refined into sub-processes with the help of the involved stakeholders. For example, in the order fulfillment process the acquisition of raw materials (cf. Fig. 4.19) could be handled in isolation from the rest of the process, since this part involves the suppliers and personnel from the warehouse & distribution department.

[1] For simplicity, we only consider one supplier in this example, so for instance there is only one activity "Request raw materials" instead of "Request raw materials from Supplier 1" and "Request raw materials from Supplier 2".

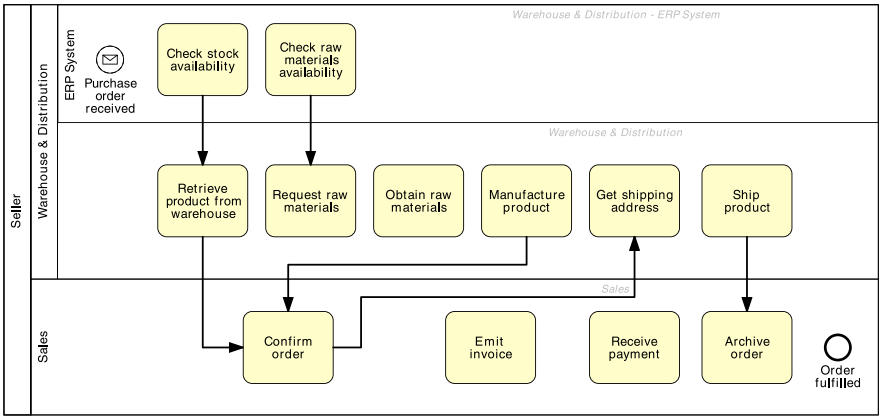

Fig. 5.2 The activities and events of the order fulfillment process assigned to pools and lanes

5.3.4 Identify the Control Flow

The handover points define an initial structure for the control flow. In essence, control flow relates to the questions of *when* and *why* activities and events are executed. Technically, we need to identify order dependencies, decision points, concurrent execution of activities and events and potential reword and repetition. Decision points require the addition of (X)OR-splits, and relevant conditions on the alternative sequence flows. Rework and repetition can be modeled with loop structures. Concurrent activities that can be executed independently from each other are linked to AND gateways. Event-based splits are used to react to decisions taken outside the process. If we have modeled more than one business party in the previous step via the use of multiple pools in this step we also need to capture the exchange of information between the various pools via message flows. Figure 5.3 shows how order constraints are captured by control-flow arcs in the order fulfillment process. Here we can see that the handovers that we identified in the previous step have now been refined in more elaborate dependencies.

Exercise 5.8 What is the relationship between the type of a gateway and the conditions of the subsequent arcs?

5.3.5 Identify Additional Elements

Finally, we can extend the model by capturing the involved artifacts and exception handlers. For the artifacts, this means adding data objects, data stores and their relations to activities and events via data associations. For the exception handlers, this means using boundary events, exception flows and compensation handlers. As we mentioned in Chaps. 3 and 4, the addition of data elements and exceptions, depends

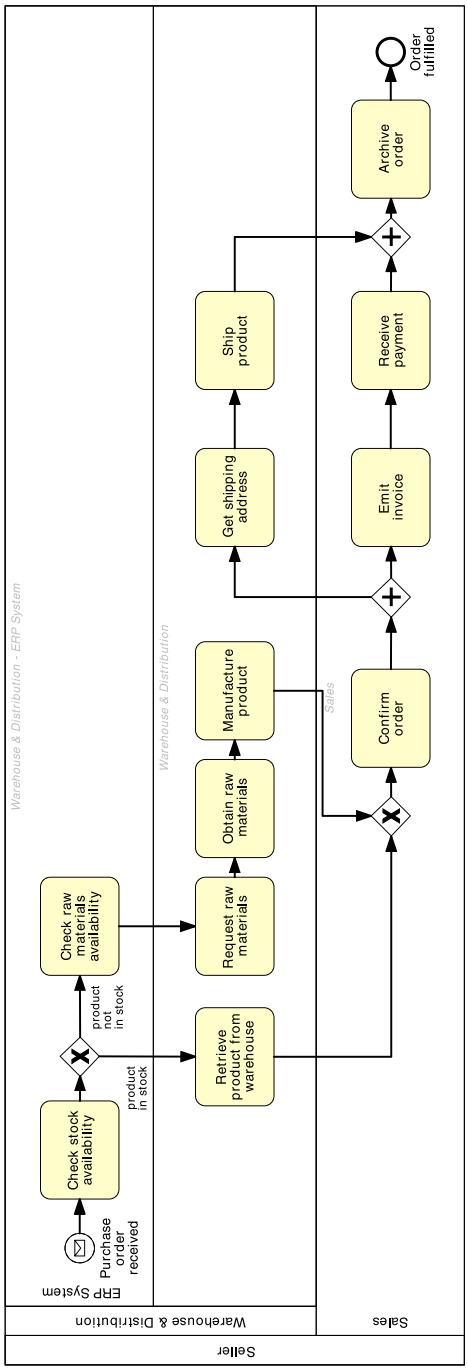

Fig. 5.3 The control flow of the order fulfillment process

on the particular modeling purpose. For example, if the process is meant to be automated, it is desirable to explicitly capture data and exception aspects. In this step we may also add further annotations to help support specific application scenarios, for instance, if the model is used for risk analysis or for process cost estimation we may need to add risk and cost information. In general, which elements to be added depends upon the particular application scenario.

In this section we illustrated a method for constructing a business process model via a number of incremental steps. In a scenario where multiple business parties are involved, an alternative option is to start with a choreography diagram first, and then incrementally refine this diagram into a collaboration diagram. In this case, we use the choreography diagram to identify the resources first, and model each of them via a pool. Next, inside each pool we model those events and activities that handle handover of information between parties (i.e. send and receive activities, message and signal events); we can derive these elements from the activities of the choreography diagram. We can then continue with Step 2 of the above method by adding the other internal activities. Next, in Step 3 we model the inner resources within each party using lanes, and then continue with the rest of the method as normal.

5.4 Process Model Quality Assurance

Process discovery involves at least a process analyst and various domain experts. Since gathering information and organizing it in a process model is often done in a sequential way, and not simultaneously, there is a need for various steps of quality assurance. Here, we focus on syntactic, semantic and pragmatic quality. Figure 5.4 shows that verification is used to achieve syntactic quality, validation provides semantic quality, and certification ensures pragmatic quality. Modeling guidelines and conventions help to ensure a good quality right from the start.

5.4.1 Syntactic Quality and Verification

Process models constructed in process discovery projects typically have to adhere to syntactical rules and guidelines. *Syntactic quality* relates to the goal of producing a process model that conforms to these rules. First of all, this means that the content of the model should comply with the syntax as defined by the process modeling language in use. For instance, in BPMN it is not allowed to draw a sequence flow across the boundaries of pools. BPMN defines an extensive set of syntax rules. Following these rules helps to make sure that a process model can always be interpreted. Beyond that, many companies define *guidelines* in order to guarantee consistency and comparability of process models, which we will discuss below.

Verification essentially addresses formal properties of a model that can be checked without knowing the real-world process. In the context of process model

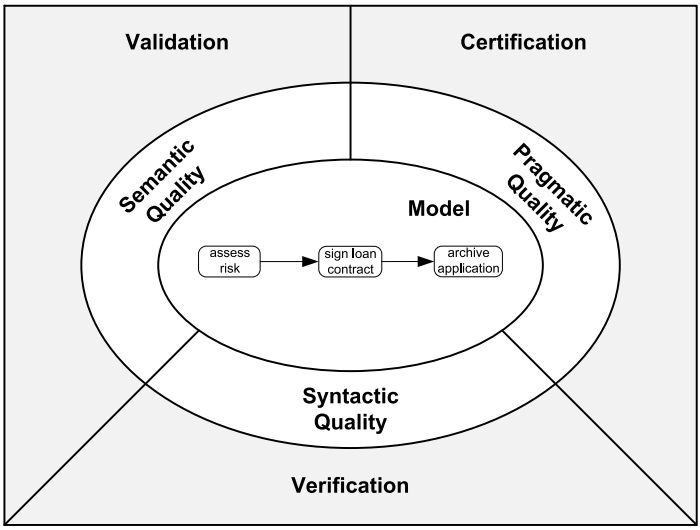

Fig. 5.4 Quality aspects and quality assurance activities

verification, structural and behavioral correctness can be distinguished. *Structural correctness* relates to the types of element that are used in the model and how they are connected. For instance, an activity should always have an incoming and an outgoing arc and every element should be on a path from a start event to an end event of the process model. Such properties can often be checked quite easily by inspecting the graph-based structure of the process model. *Behavioral correctness* relates to potential sequences of execution as defined by the process model. It is a general assumption that a case should never be able to reach a deadlock or a livelock. This is the case when the *soundness* property holds (see Chap. 3). Common sound and unsound process fragments are depicted in Fig. 5.5. Verification properties such as soundness can be checked after a process model is created. Alternatively, a process modeling tool can enforce that a model is correct by design. This can be achieved by allowing only edit operations on the model that preserve structural and behavioral correctness.

Exercise 5.9 Have a look at Fig. 5.5. Explain what exactly is going wrong in the unsound process model fragments.

5.4.2 Semantic Quality and Validation

Semantic quality relates to the goal of producing models that make true statements about the considered domain, either for existing as-is processes or future to-be processes. The particular challenge of a semantic quality assessment is that the process

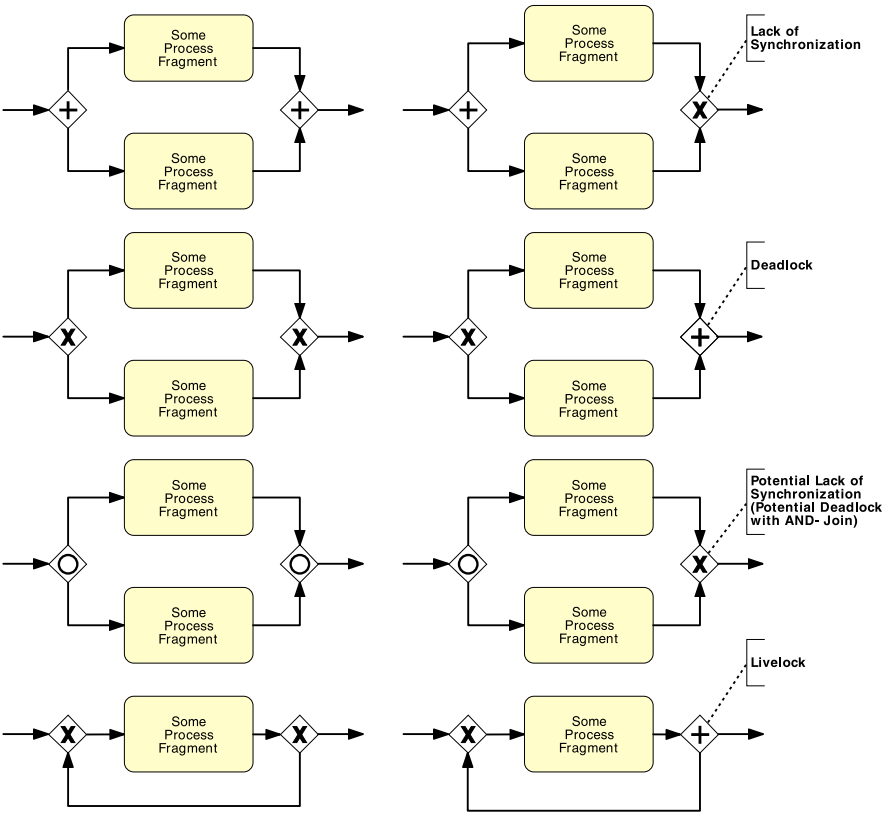

Fig. 5.5 Common sound and unsound process fragments

model has to be compared with the real-world domain of a particular business process. This means there is no set of formal rules that can be used to easily check semantic quality. Whether the process model at the center of Fig. 5.4 is of good semantic quality can only be assessed by talking to people involved in the process and by consulting documentation.

Validation deals with checking the semantic quality of a model by comparing it with the real-world business process. There are two essential aspects of semantic quality: validity and completeness indexCompleteness. *Validity* means that all statements included in the model are correct and relevant to the problem. Validity can be assessed by explaining domain experts how the processing is captured in the model. The domain expert is expected to point out any difference between what the model states and what is possible in reality. *Completeness* means that the model contains all relevant statements on a process that would be correct. Completeness is more difficult to assess. Here, the process analyst has to ask about various alternative processing options at different stages of the process. For example, the model in Fig. 5.3 is still missing all data elements and exception handlers. It is the job of

the process analyst to judge the relevance of these additional elements. This judgment has to be done against the background of the modeling objective, which the process analyst should be familiar with. Let us consider an example to understand the difference between validity and completeness. If the process model for loan assessments expresses that any financial officer may carry out the task of checking the credit history of a particular applicant while in practice this requires a specific authorization, the model has an issue with semantic quality (invalid statement). If the task of checking the credit history is omitted then it has a semantic problem due to incompleteness. Validation can be supported by techniques like simulation or interviews. Alternatively, there are tools that provide truthfulness by design. This is, for instance, achieved by building a process model from the logs of an information system, as we will see in Chap. 10. In practice, process models often require the *approval* from the process owner. This approval is a special validation step, since it again refers to the correctness and completeness of the process model. Beyond that, the approval of the process owner establishes the normative character of the process model at hand. As a consequence, the process model can now be archived, published or used as an input for process redesign.

5.4.3 Pragmatic Quality and Certification

Pragmatic quality relates to the goal of building a process model of good usability. The particular challenge of pragmatic quality assessment is to prognosticate the actual usage of a process model beforehand. Accordingly, this aspect very much focuses on how people interact with a model. Whether the process model at the center of Fig. 5.4 is of good pragmatic quality can, for instance, be checked by testing how well a user understands the content of the model.

Certification is the activity of checking the pragmatic quality of a process model by investigating its usage. There are several aspects of usability including understandability, maintainability, and learning. *Understandability* relates to the fact how easy it is to read a specific process model. *Maintainability* points to the ease of applying changes to a process model. *Learning* relates to the degree of how good a process model reveals how a business process works in reality. There are several characteristics of a model that influence usability including its size, its structural complexity, and its graphical layout. Certification can be conducted using user interviews or user experiments. Alternatively, there are rules that strive to provide usability by design. This can be achieved, for instance, by following design rules on the structure of the process model. There are two essential checks for understanding, maintainability and learning. The first one relates to the consistency between visual structure and logical structure. Figure 5.6 and Fig. 5.7 show the same fragment of the order fulfillment process model. The second model is a rework of the first one in terms of layout. Here, the element positions have been changed with the aim to improve the consistency between visual structure and logical structure. The second check is concerned with meaningful labels. It is important that activities and other

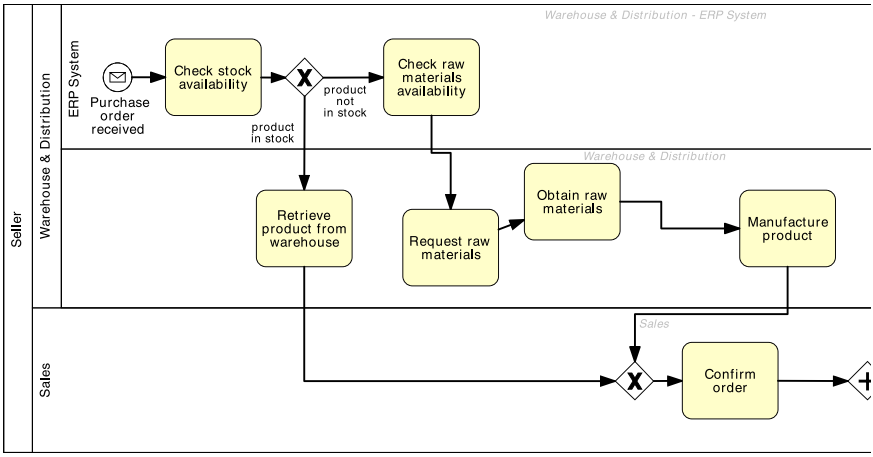

Fig. 5.6 Extract of the order fulfillment process model with bad layout

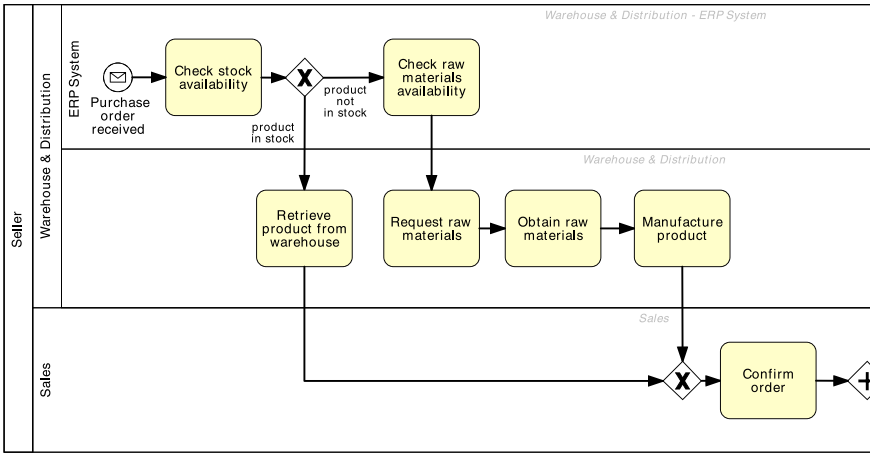

Fig. 5.7 Extract of the order fulfillment process model with good layout

elements have labels that follow specific naming conventions, as those presented in Sect. 3.1.

5.4.4 Modeling Guidelines and Conventions

Modeling guidelines and *conventions* are an important tool for safeguarding model consistency and integrity for bigger modeling initiatives with several people involved. The goals of such guidelines and conventions are to increase readability

and comparability in order to facilitate efficient model analysis. A guideline document typically covers naming conventions for processes, tasks and events, modeling conventions for layout and usage of tasks, events, lanes and pools, and a restriction of the set of elements. *Naming conventions* usually recommend or enforce the verb-object style for labeling of activities and suitable styles for other elements as discussed in Sect. 3.1. Too technical and too generic verbs should be avoided. *Modeling conventions* define element usage and layout. Layout for BPMN models is typically defined with a horizontal orientation. Usage of pools may be enforced for each model along with corresponding message flow. The detail of how start and end events are captured can be specified as well. All the conventions often come along with rules how to capitalize or what to reference in the elements names, e.g. using a glossary. Finally, *restrictions* can be defined in order to simplify the set of elements of BPMN. Such restrictions are recommended to increase understanding of models also by non-expert users.

One set of guidelines that has recently been proposed are the so-called Seven Process Modeling Guidelines (*7PMG*). This set was developed as an amalgamation of the insights that were derived from available research. Specifically, the analysis of large sets of process models by various researchers have identified many syntactical errors as well as complex structures that inhibited their interpretation. The guidelines that are part of 7PMG are helpful in guiding users towards mitigating such problems. The guidelines are as follows:

G1: Use as few elements in the model as possible. The size of a process model has undesirable effects on the understanding of process model and the likelihood of syntactical errors. Studies have shown that larger models tend to be more difficult to understand and have a higher error rate.

G2: Minimize the routing paths per element. For each element in a process model, it is possible to determine the number of incoming and outgoing arcs. This summed figure gives an idea of the routing paths through such an element. A high number makes it harder to understand the model. Also, the number of syntactical errors in a model seems strongly correlated to the use of model elements with high numbers of routing paths.

G3: Use one start and one end event. Empirical studies have established that the number of start and end events is positively connected with an increase in error probability. Models satisfying this requirement are easier to understand and allow for all kinds of formal analysis.

G4: Model as structured as possible. A process model is structured if each split gateway matches a respective join gateway of the same type. Block-structured models can be seen as formulas with balanced brackets, i.e., every opening bracket has a corresponding closing bracket of the same type . Unstructured models are not only more likely to include errors, people also tend to understand them less easily. Nonetheless, as discussed in Chap. 4.3, it is sometimes not possible or not desirable to turn an unstructured process model fragment (e.g. an unstructured cycle) into a structured one. This is why this guideline states “as structured as possible”.

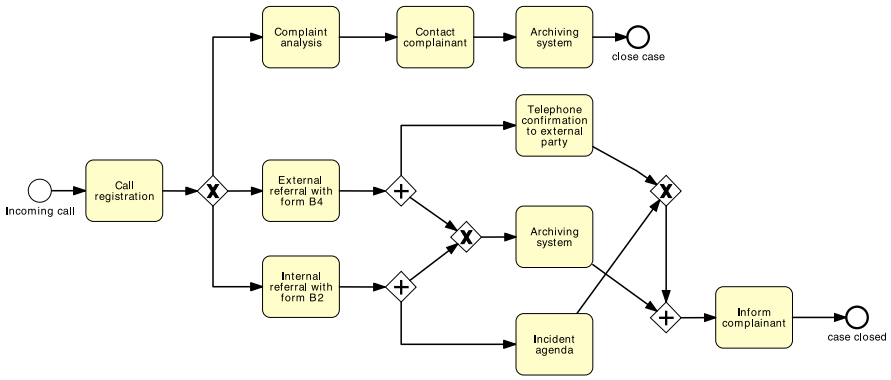

Fig. 5.8 A complaint handling process as found in practice

G5: Avoid OR-gateways. Models that have only AND-gateways and XOR-gateways are less error-prone. This empirical finding is apparently related to the fact that the combinations of choices represented by an OR-split are more difficult to grasp than behavior captured by other gateways.

G6: Use verb-object activity labels. A wide exploration of labeling styles that are used in actual process models, discloses the existence of a number of popular styles. From these, people consider the verb-object style, like "Inform complainant", as significantly less ambiguous and more useful than action-noun labels (e.g. "Complaint analysis") or labels that follow neither of these styles (e.g. "Incident agenda").

G7: Decompose a model with more than 30 elements. This guideline relates to G1 that is motivated by a positive correlation between size and errors. For models with a more than 30 elements the error probability tends to climb sharply. Therefore, large models should be split up into smaller models. For example, large sub-components with a single entry and a single exit can be replaced by one activity that points to the original sub-component as a sub-process.

The need for guidelines like 7PMG is emphasized by the structure of process models we have seen in practice. Figure 5.8 shows a simplified version of a complaint handling process model of one of our industry partners. A complaint is triggered by a phone call by a complaining customer. It is decided whether the complaint can be handled or whether it has to be referred to an internal or external party. An external referral leads to a telephone confirmation to the external party. An internal referral is added to the incident agenda. If no referral is needed, a complaint analysis is conducted and the complainant is contacted. In either case, the complaint is archived and the case is closed.

Exercise 5.10 Consider the process model of Fig. 5.8. Explain which 7PMG guidelines point to potential for improvement. Remodel the process based on your observations.

The 7PMG are but one of the available sets of modeling guidelines. Moreover, it is a tentative set in the sense that research in the area of process model quality is rapidly developing. Many of its guidelines are already applied in practice and have been discussed as state-of-the-art at previous places in this book. For example, in Sect. 3.1 the beneficial verb-object labeling convention was already mentioned. Also, the threshold to start decomposing process models as in guideline G7 was discussed in Sect. 4.1. What is special about the 7PMG is that these guidelines have a strong empirical basis and, as such, transcend the knowledge of individual modelers. As insights develop further, it seems both likely and favorable that the set will be updated and expanded.

5.5 Recap

This chapter described how to proceed in the different phases of process discovery. In essence, we defined four stages of process discovery, namely defining the setting of process discovery, applying different discovery methods for gathering information about the process, stepwise modeling the process, and finally addressing different aspects of quality assurance.

The definition of the setting of process discovery has to take into account the different characteristics and complementary skills of process analysts and domain experts. While process analysts are skilled in analyzing and modeling processes, they often lack detailed domain knowledge. In contrast, domain experts have typically limited modeling skills, but detailed understanding of the part of the process they are involved with. This implies three challenges of process discovery. First, different domain experts have an understanding of only a part of the process. Different partial views have to be integrated in process discovery. Second, domain experts tend to think in cases and not on the general process level. The process analyst has to abstract from these cases. Third, domain experts often have difficulties in understanding process models. Therefore, the process analyst has to guide the domain expert in reading the model for getting feedback.

Different methods of process discovery can be used ranging from evidence-based methods to interviews and workshops. Evidence-based methods typically provide the most objective insight into the execution of the process. However, the immediacy of feedback is low and the richness of the insight can be mediocre. Interviews can be biased towards the perspective or opinion of the interview partner, but reveal rich details of the process. The interview situation offers the chance of direct feedback and clarification. Workshops can help to immediately resolve inconsistent perspectives of different domain experts. On the downside, it is difficult to have all required domain experts synchronously joining. It is recommended to utilize a mixture of the methods that reflects the specifics of the discovery project.

We then defined a process modeling method including six steps. First, the boundaries of the process have to be defined in terms of start and end events. Second, the essential activities of the process have to be identified. Subsequently, we need to determine the handovers between different persons and departments. Once that aspect

is clarified, we can sketch out the details of the control flow. Then, routing conditions and intermediate events have to be added. Finally, additional perspectives and annotations can be included.

In the last sections of this chapter, we discussed different measures of quality assurance. First, we emphasized verification as a tool for assuring syntactical correctness. Then, we discussed how validation helps to establish semantic quality. Finally, we presented different aspects of pragmatic quality and described how they can be certified.

5.6 Solutions to Exercises

Solution 5.1 In case you are supposed to map the process of signing a rental contract in your city, it is likely that you have some experience with this process, either from renting a flat yourself, or from stories from your friends, or from you or your friends giving a flat for rent. Assuming you have already studied the chapters on process modeling, you have both domain expertise and process modeling expertise. This is an uncommon situation. Most often, you face situations like mapping the process of getting a license plate for your car in Liechtenstein as a foreign resident. This is a process for which you would unlikely have domain knowledge. Process discovery typically brings you as a process analyst into an environment that you do not know in detail beforehand. Process discovery is concerned with understanding the process under consideration and also the domain surrounding it.

Solution 5.2 An advantage of having teams modeling processes themselves is first that a lot of process models can be created in a short span of time. It is critical though that these teams possess the required skills in process modeling. According to the third challenge of process discovery, domain experts typically do not have process modeling skills and feel uncomfortable with the modeling task. Furthermore, domain experts often think in cases (second challenge) and lack the process perspective to generalize. Finally, there is the risk that the results from such a modeling initiative might be fragmented and difficult to integrate. It is typically the responsibility of the process analyst to integrate the fragmented perspectives.

Solution 5.3 Domain knowledge can be very helpful for analyzing processes. It helps to ask the right questions and to build analogies from prior experience. On the other hand, the skills of an experienced process analyst should not be underestimated. These skills are domain-independent and relate to how a process discovery project can be organized. Experienced process analysts are typically very skilled in scoping and driving a project into the right direction. They possess problem-solving skills for handling various critical situations of a process discovery project. There is clearly a trade-off between the two sets of skills. It should be ensured that a certain level of process analysis experience is available. If that is not the case for the applying domain expert, the process analyst might be preferred.

Solution 5.4 As a customer, we would have to rely mostly on evidence-based process discovery. We can place exemplary orders and study the different processing options for them. In relation to these placed orders, we could also contact the customer help desk and inquire details of the process that we cannot directly observe. If we were assigned to a process discovery project by the process owner, we would get access to the domain experts within the company. In this case, we could also use interviews and workshop-based discovery methods.

Solution 5.5 This process contains ten major activities that are executed by different persons. We can assume that there will be a kickoff meeting with the process owner and some important domain experts on day one. One day might be required to study available documentation. An interview with one domain expert can take from two to three hours, such that we would be able to meet two persons per day, and document the interview results at night time. Let us assume that we meet some persons only once while we seek feedback from important domain experts in two additional interviews. Then, there would be a final approval from the process owner. This adds up to one day for the kickoff, one for document study, five days for the first iteration interviews, and further five days if we assume that we meet five experts three times. Then, we need one day for preparing the meeting for final approval with the process owner, which would be on the following day. If there are no delays and scheduling problems, this yields $2 + 5 + 5 + 2 = 14$ work days as a minimum.

Solution 5.6 Before starting with process discovery, it is important to understand the culture and the sentiment of an organization. There are companies that preach and practice an open culture in which all employee are encouraged to utter their ideas and their criticism. Such organizations can benefit a lot from workshops as participants are likely to present their ideas freely. In strictly hierarchical organizations, it is necessary to take special care that every participant gets an equal share of parole in a workshop and that ideas and critique are not hold back. It might be the case that the young dynamic company has a more open culture than the company with extensive health and security regulations. This has to be taken into account when organizing a workshop.

Solution 5.7 There are various circumstances that may restrict the application of different discovery methods. Direct observation may not be possible if the process partially runs in a remote or dangerous environment. For instance, the discovery of a process of an oil-producing company for pumping oil from an oil rig to a ship might belong to this category. Then, there might be cases where documentation does not exist, for example when a startup company, which has gone through a period of rapid growth wants to structure its purchasing process. Lack of input may also be a problem for automatic process discovery based on event log data. If the process under consideration is not yet supported by information systems, then there are no data available for conducting automatic process discovery. In general, interviews are always possible. It might still be a problem though to gain commitment of domain experts for an interview. This is typically the case when the process discovery

project is subject to company-internal politics and hidden agendas. Workshop-based discovery can be critical in companies with strong hierarchy which have a culture of suppressing creative thinking of their staff.

Solution 5.8 The type of the gateway has to be consistent with the conditions of the subsequent arcs. If there is an XOR-split, then the conditions on the arcs have to be mutually exclusive. If there is an OR-split, then the conditions can be non-exclusive. If an AND-split is used, there should be no conditions on the arcs.

Solution 5.9 Four unsound fragments are shown with the following problems:

- The lack of synchronization relates to an AND-split followed by an XOR-join. In this case, the two tokens created from the AND-split are not synchronized XOR-join, potentially leading to the duplicate execution of activities downstream.
- A deadlock occurs, for instance, if an XOR-split is followed by an AND-join. As the XOR-split creates a token only on one of its outgoing arcs, the AND-join requiring a token on each of its incoming arcs gets stuck waiting for a second token to arrive.
- In case there is an OR-split followed by an XOR-join, we potentially get a lack of synchronization. This depends upon the conditions of the OR-split. If only one token is generated from it, the process can proceed correctly. If multiple tokens are generated, there is a lack of synchronization. In the same vein, there is a potential deadlock if the OR-split is followed by an AND-join.
- A livelock can occur in an inappropriate loop structure. Here, there is an XOR-join used as an entry to a loop, but the loop exit is modeled with an AND-split. This has the consequence that it is never possible to leave the loop. Each time the AND-split is reached, it creates one token exiting the loop, but also another token that stays within the loop.

Solution 5.10 The process model reveals several problems. Several elements with the same name are shown twice (end event and archiving activity), therefore G1 is violated. Also the control structure is very complicated, violating G4 asking for a

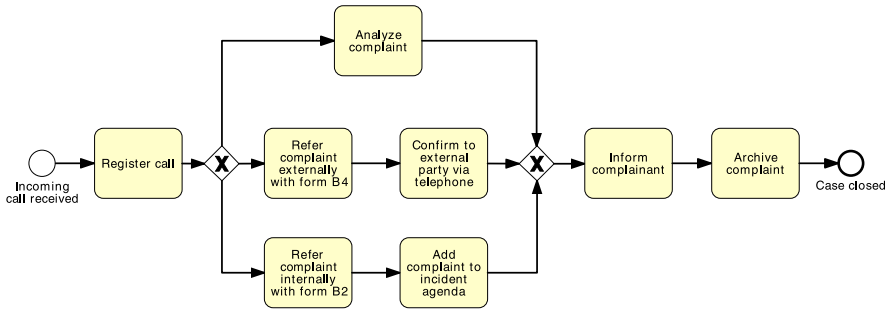

Fig. 5.9 The complaint handling process reworked

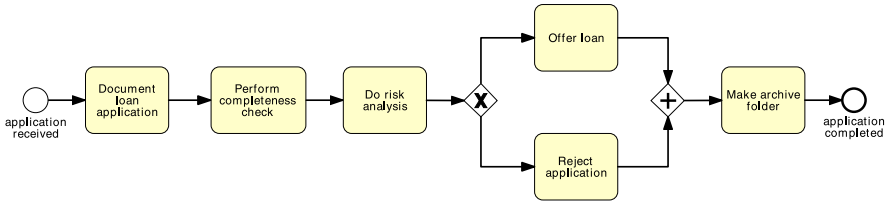

Fig. 5.10 A loan application process

structured model. Finally, several activities do not follow the naming conventions of G6.The model can be reworked and simplified to the one shown in Fig. 5.9.

5.7 Further Exercises

Exercise 5.11 Imagine you are responsible for the human resources department of a leading consultancy. Which characteristics would you check when hiring new process analysts?

Exercise 5.12 As responsible for human resources department of a consultancy, how would you develop the skills of your junior process analysts?

Exercise 5.13 How would you as a process analyst prepare for an interview with a domain expert?

Exercise 5.14 Analyze the loan application process model of Fig. 5.10 for soundness-related problems.

Exercise 5.15 Look up tools on the internet that offer a soundness check for process models.

Exercise 5.16 Consider again the loan application process model of Fig. 5.10. What are indications that it would not be complete?

Exercise 5.17 Have a look at the activity labels of the loan application model of Fig. 5.10 and propose improved labels where appropriate.

Exercise 5.18 Have a look at the process model of Fig. 5.10 showing a sales campaign process for one of our industry partners. Describe which 7PMG guidelines can be used to improve this model. Have a look at the process model of Fig. 5.11 showing a sales campaign process for one of our industry partners

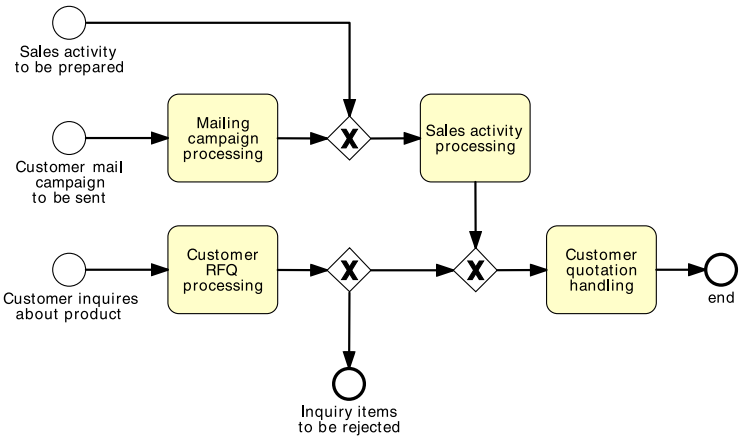

Fig. 5.11 A sales campaign process

5.8 Further Reading

The general topic of process discovery is well covered in the book on workflow modeling by Sharp and McDermott [86]. This book gives detailed advice on all phases of process discovery, specifically data gathering and workshop organization. Other practical advice is summarized by Verner [101] and by Stirna, Persson, and Sandkuhl [88]. Interview techniques are widely discussed as a social science research method for instance in the book by Berg and Lune [7] or the book by Seidman [85].

Frederiks and van der Weide [18] discuss the skills required from process analysts, particularly when engaging in process discovery efforts. In a similar vein, Schenk, Vitalari and Davis [83] and Petre [66] discuss the capabilities that expert process analysts (as opposed to novice ones) generally display when engaging in process discovery.

In this chapter, we emphasized "manual" process discovery techniques, wherein process models are manually constructed based on data collected from various process stakeholders by means of interviews, workshops and related techniques. As mentioned in Sect. 5.2.1, there is also a whole range of complementary techniques for automatic discovery process models from event logs. These automatic process discovery techniques are part of a broader set of techniques for analyzing event logs, collectively known as process mining [94]. We will discuss several process mining techniques later in Chap. 10.

The modeling method introduced in Sect. 5.3 revolves around the discovery of activities and control-flow relations between activities. This family of approaches is usually called *activity-based modeling* [68]. An alternative approach to process modeling is known as *artifact-centric modeling* [59] or *object-centric modeling* [68]. In artifact-centric modeling the emphasis is not on identifying activities, but rather *artifacts* (physical or electronic objects or documents) that are manipulated within a given process. For example, in an order-to-cash process, typical

artifacts are the purchase order, the shipment notice and the invoice. Once these artifacts have been identified, they are analyzed in terms of the data that they hold and in terms of the phases they go through during the process. For example, a purchase order typically goes through the phases *received*, *accepted*, *manufactured*, *shipped* and *invoiced*. These phases and the transitions between these phases are called the artifact lifecycle. The main emphasis in artifact-centric process modeling is put on identifying these artifact lifecycles. Several industrial applications of artifact-centric process modeling have shown that it is quite suitable when discovering processes that exhibit significant amounts of variation, for example variation between business units, geographical regions or types of customer as discussed for example by Caswell et al. [59] and Redding et al. [68].

The quality of conceptual models in general, and of process models specifically, has received extensive attention in the research literature. The Sequal framework introduced by Lindland, Sindre, and Sølvberg adapts semiotic theory, namely the three perspectives of syntax, semantics and pragmatics, to the evaluation of conceptual model quality [45]. An extended version of this framework is presented by Krogstie, Sindre, and Jørgensen [42].

Validation and verification of process models has also received extensive attention in the literature. Mendling [51] for example provides numerous pointers to related research. The verification of Workflow nets specifically is investigated by van der Aalst [93] who connects soundness analysis of process models with classical Petri nets notions of liveness and boundedness.

The 7PMG guidelines discussed in Sect. 5.4.4 are by Mendling, Reijers, and van der Aalst in [53]. These guidelines build on empirical work on the relation between process model metrics on the one hand and error probability and understandability on the other hand [50, 54, 55, 63, 73, 74]. Specifically, the impact of activity label quality on process model understanding is investigated by Mendling, Reijers, and Recker [52]. Another set of modeling guidelines are the Guidelines of Process Modeling by Becker et al. [5].

As a complement to process modeling guidelines and conventions, it is useful to also keep in mind potential pitfalls to be avoided in process modeling projects. For example, Rosemann [78, 79] draws a list of 22 pitfalls of process modeling, including a potential lack of strategic connection, l'art pour l'art, to name but a few. His bottom line is that modeling success does not directly equate with process success.