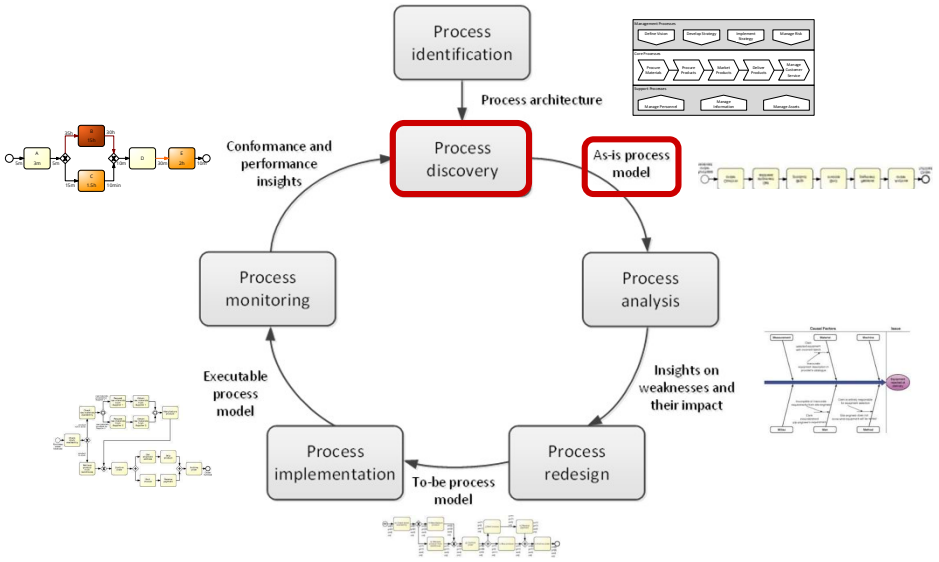
Process Discovery

Spring 2021 - MAJU

Nauman H. Ansari

1

Process Modeling in the BPM Lifecycle



2

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. This means that process discovery is a much broader activity than process modeling. Clearly, modeling is a part of this activity.
- 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 tasks of process discovery:

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

2. Gathering information: concerned with building an understanding of the process. Different discovery methods can be used to acquire information on a process.

done together

3. Conducting the modeling task: deals with the actual creation of the process model. A modeling method gives guidance for mapping out the process in a systematic way.

4. Assuring process model quality: aims to guarantee that the resulting process model meets different quality criteria. This task is important for establishing trust in the process model.

3

Who is involved?

Process analyst

Domain expert

4

Exercise 5.1

You are the manager of a consulting company and you need to hire a person for the newly signed BPM project with an online bookstore. 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-to-cash 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 languages and with several modeling tools.

5

Solution 5.1

- 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 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 assured that a certain level of process modeling analysis experience is available. If that is not the case for the applying domain expert, the process analyst would be preferred.

6

6

Exercise 5.2

Consider the order-to-cash process of your preferred online bookstore and assume you have access to three internal resources: a customer relationship manager (who handles sales and reclaims), a warehouse worker (who looks after shipments), and a financial officer (who raises invoices and collects payments). As a process analyst, what questions do you need to ask these domain experts to be able to obtain a complete and systematic view of this process?

Hint. Think of the different exposure to this process that the three resources have and of the possible conditions, process outcomes, and exceptions that they may have experienced while executing this process.

7

Process Discovery Methods

8

Process discovery methods

1. **Evidence-based**
   - Document analysis
   - Observation
   - Automated process discovery

2. **Interview-based**

3. **Workshop-based**

Choose one or more on the basis of context and budget

9

Document Analysis

Documents point to existing roles, activities and business objects:

- Process descriptions (ideal scenario)
- Internal policies
- Organization charts
- Employment plans
- Quality certificate reports
- Glossaries and handbooks
- Forms
- Work instructions…

Could be used to gather information before approaching domain experts.

Potential issues:

- May not be process-oriented and trustworthy
- May require abstraction or refinement

10

Observation

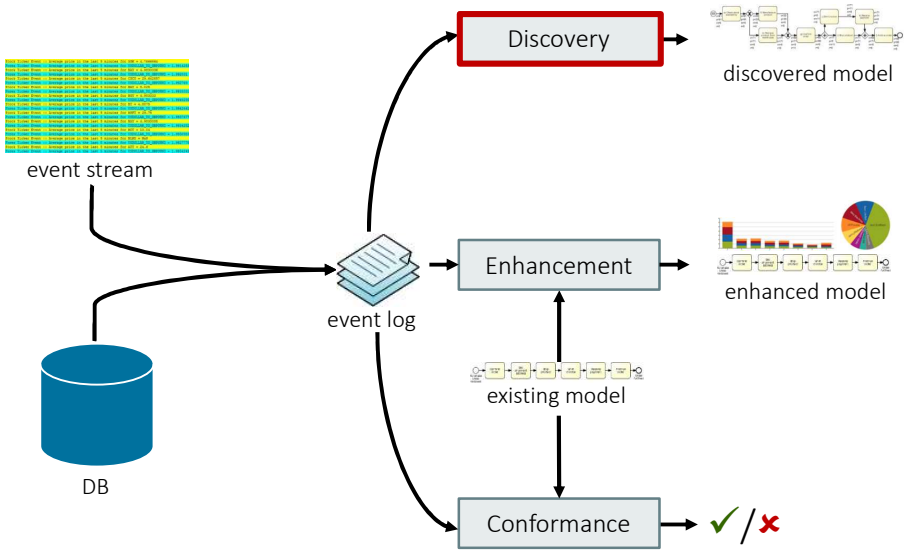
- Follow directly the execution of individual process instances, then abstract from instance to process level:
  - **Active role**: play a specific role, e.g. customer
  - **Passive role**: observe participants and their environment
- Trace business objects in the course of their lifecycle

Potential issues:

- Active role: no big picture
- Passive role: participants' bias

11

Automated process discovery and Process mining



12

Automated discovery: Minimum data requirements

- Activity name and timestamp
- Reference to case id

case id	event id	properties				
		timestamp	activity	resource	cost	...
1	35654423	30-12-2010:11.02	register request	Pete	50	...
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400	...
	35654425	05-01-2011:15.12	check ticket	Mike	100	...
	35654426	06-01-2011:11.18	decide	Sara	200	...
	35654427	07-01-2011:14.24	reject request	Pete	200	...
2	35654483	30-12-2010:11.32	register request	Mike	50	...
	35654485	30-12-2010:12.12	check ticket	Mike	100	...
	35654487	30-12-2010:14.16	examine casually	Pete	400	...
	35654488	05-01-2011:11.22	decide	Sara	200	...
	35654489	08-01-2011:12.05	pay compensation	Ellen	200	...

Additional information:

- Activity resource, cost
- Case attributes (e.g. customer reference, type of case…)

13

Example – Process Discovery through Document Analysis

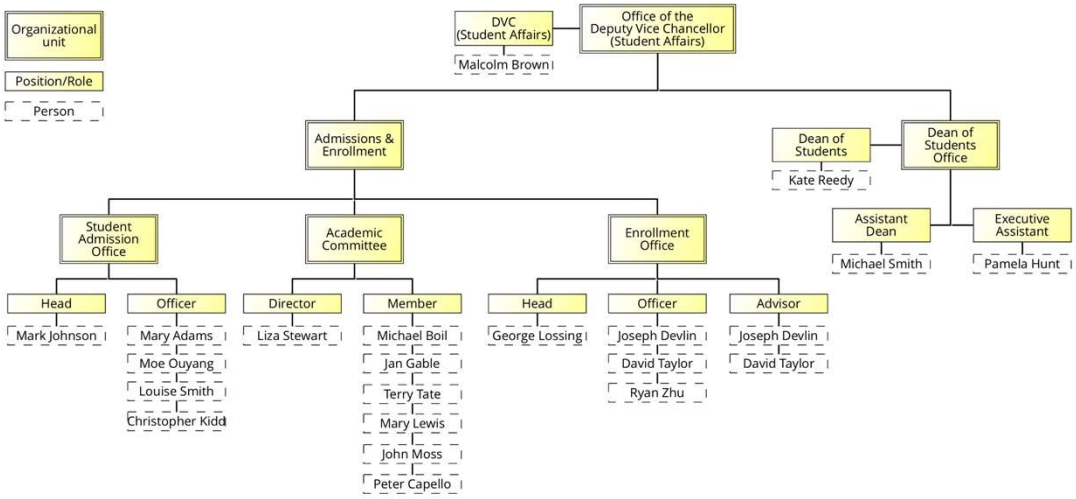
As a process analyst working for the University of Newtown, you have been engaged by Mark Johnson, the process owner of the student admission process, in a project that aims at improving this process. In order to model the as-is process, you start by collecting relevant information about this process. The available documentation includes the organization chart of the Office of the Deputy Vice-Chancellor (DVC) for Student Affairs where Mark's team sits, the UML class diagram of the student admission system which supports this process, and a set of relevant organizational policies that you extract from the university's Web pages.

Based on this documentation, formulate initial hypotheses on how the student admission process works. Next, identify the relevant domain experts to interview and their supervisors whom you should seek approval from.

14

Example 5.3 (cont'ed)

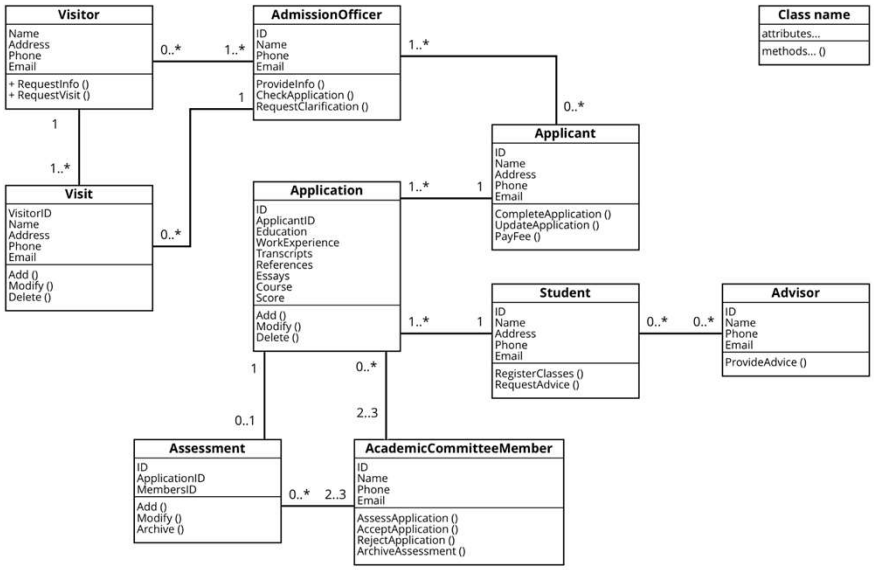
Organizational chart of the Office of the DVC (Student Affairs)



15

Example 5.3 (cont'ed)

Extract of the UML class diagram of the student admission system



16

Example 5.3 (cont'ed)

Organizational policies for student admission

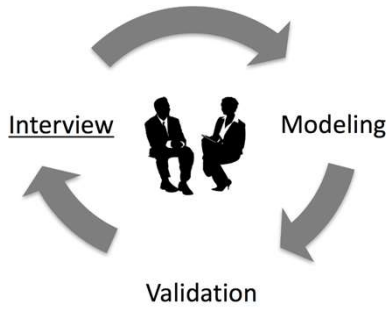
- *An applicant is admitted if:*
  - *their prior education is consistent with the study area of the selected course*
  - *the submitted essay is not plagiarized and is of good quality*
  - *the score of the prior degree is at least 70 out of 100 (standard 100-point scale)*
  - *the two reference letters are satisfactory.*
- *Successful applicants must accept the offer within four weeks from notification.*

17

Interview-based Discovery

- Interview-based discovery aims at interviewing domain experts to inquire about how a process is executed.
- We can use two strategies for conducting an interview:

(i) starting from the process outcomes (e.g., an order being fulfilled), we work our way backwards until we reach the process triggers (e.g., the receipt of a purchase order); or
(ii) starting from the triggers, we proceed forward until we reach the process outcomes.



- When conducting an interview, it is more effective to balance between a structured and a free-form interview approach. For example, considering a 1-h interview, one may spend the first 45 min to go through a list of predefined questions

18

Interview-based Discovery - Example

After collecting relevant information on the student admission process, you interviewed some representatives for the two roles involved in this process: Mary Adams and Louise Smith as student admission officers, and Peter Capello as a member of the academic committee (Mark Johnson, the process owner, confirmed that the enrollment office is not involved in this process).

19

Workshop-based Discovery

- Gather all key stakeholders together
- Participants interact to create shared understanding
- Typically one process analyst (facilitator), multiple domain experts, process owner may also attend
- May be software-supported – a model is directly created during the workshop (typically a separate role – tool operator)
- Model is used as a reference point for discussions

20

Example: Any difference in discovery?

Consider the following two companies:

- **Company A** is young, founded three years ago, and has grown rapidly to a current toll of one hundred employees.
- **Company B** is owned by the government and operates in a domain with extensive health and security regulations.

How might these different characteristics influence a workshop-based discovery approach?

Before starting with process discovery, it is important to understand the **culture** and the sentiment of an organization.

21

Discovery methods: strengths and weaknesses

Method	Strength	Weakness
Document Analysis	• Structured information • Independent from availability of stakeholders	• Outdated material • Wrong level of abstraction
Observation	• Context-rich insight into process	• Potentially intrusive • Stakeholders likely to behave differently • Only few cases
Automatic Discovery	• Extensive set of cases • Objective data	• Potential issue with data quality and level of abstraction
Interview	• Detailed inquiry into process	• Requires sparse time of process stakeholders • Several iterations required before sign-off
Workshop	• Direct resolution of conflicting views	• Requires availability of several stakeholders at the same time

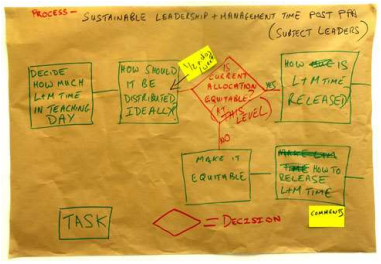
22

Process Modeling Method

23

Organizing the Gathered Material (Stepwise method to conduct the modeling)

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 (e.g. data objects, different types of events, exception handling…)



24

1. Identify the process boundaries

What are the process triggers?

Purchase order received

What are the possible outcomes (positive/negative)?

Positive outcome: order fulfilled
Negative outcome: order rejected

Which perspective do we assume?

Seller

What artifacts are required as input and output to the process?

Input: Purchase order
Output: Invoice, Shipment notice

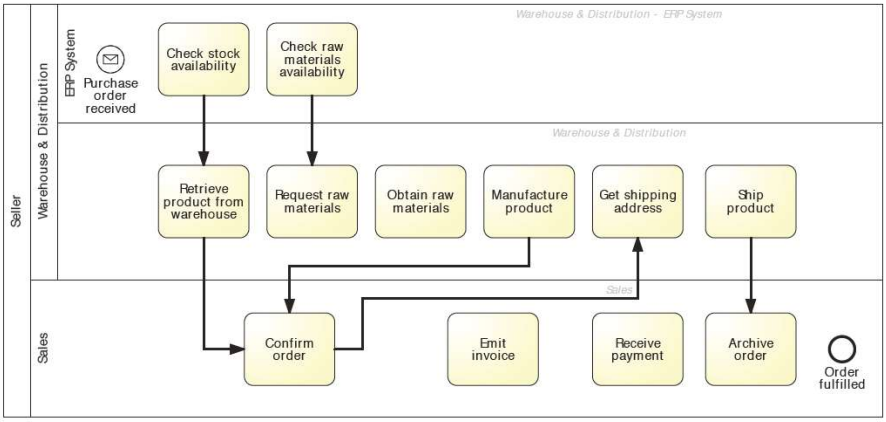
25

2. Identify Activities and Events

The goal of the second step is to identify the main process activities and intermediate events.

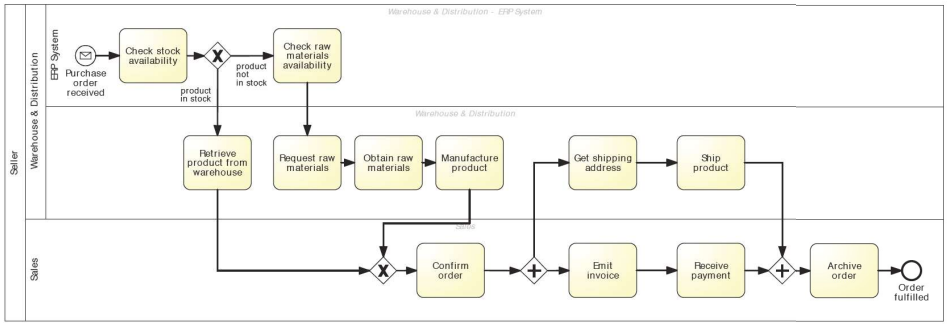
26

3. Identify Resources and Handoffs



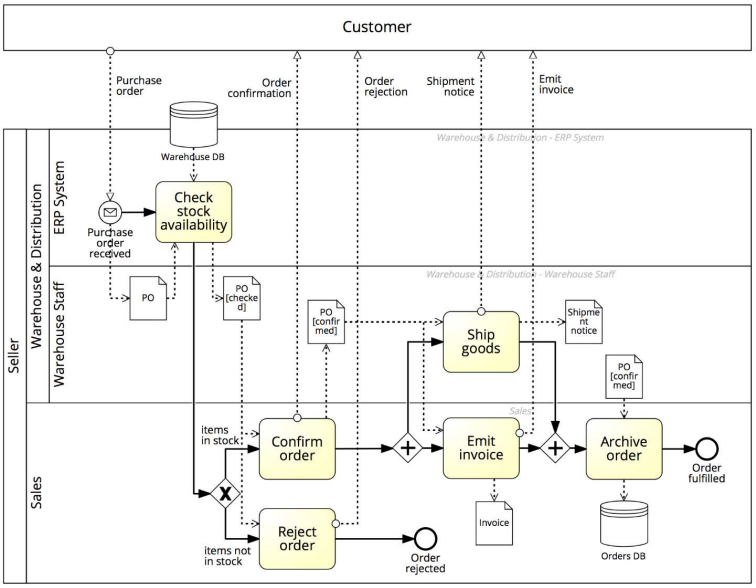
27

4. Identify Control Flow



28

5. Identify additional elements



Process Modeling Method Exercise

- Exercise 5.7: Identify the process boundaries for the procure-to-pay process de- scribed in Exercise 1.7 of Chapter 1 (page 30).
- Exercise 5.8: Identify the main activities and events for the procure-to-pay process.
- Exercise 5.9: Using the description of the procure-to-pay process, first identify the involved resources; next, assign the activities and events you obtained in Exercise 5.8 to these resources; and finally identify the handoffs.

Solution 5.7

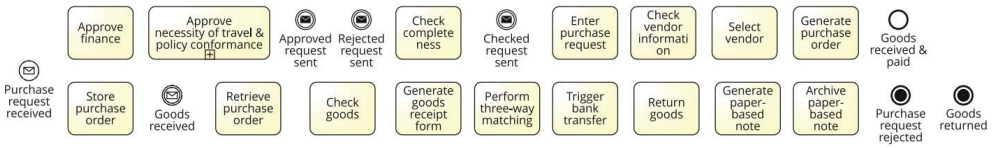
- We take the perspective of the company and consider the employee as the customer.
- Accordingly, we identify one start event, namely “Request for purchase received”,;
- and three end events, namely “Goods received & paid” (positive outcome), “Purchase request rejected”, and “Goods returned” (negative outcomes).

31

31

Solution 5.8

- We identify 16 main activities and four intermediate events.
- As for the activities, given that the approval for the necessity of purchase and for the conformance to the company’s policies are done by the same supervisor, we can scapture these two approvals with a collapsed sub-process.
- As for the events, we use three intermediate throwing message events to communicate the results of the checks done by the supervisors and by the purchasing department back to the employee, and one intermediate catching message event to model the receipt of the goods.



32

32

Solution 5.9

- We identify one pool for the employee, one for the vendor, and one for our company.
- The latter includes the following lanes: supervisor, purchasing department, enterprise system, accounts payable office, and goods receipt department.
- In the lane for the supervisor we add a text annotation to specify that a four-eye principle applies to the two approval activities ("Approve finance" and "Approve necessity of purchase & policy conformance"). Activity "Archive paper-based note" is performed by both the purchasing department and by the accounts payable office.

33

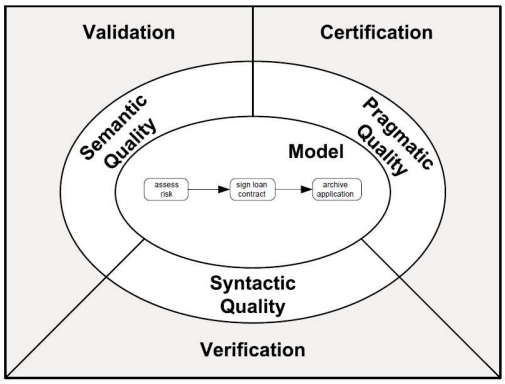
33

Process Model Quality Assurance

34

Process Model Quality Assurance

- Validity
- Completeness



- Understandibility
- Mantainability
- Learning

- Structural correctness
- Behavioral correctness

35

Syntactic quality: Verification

Syntactic quality relates to the conformance of a process model to the syntactic rules of the modeling language used.

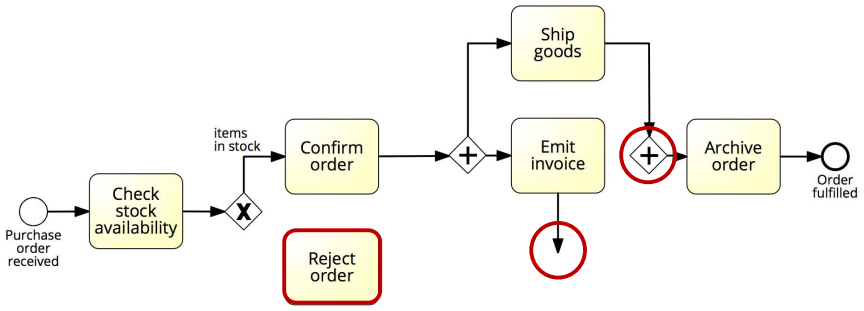
Two types of syntactic rules: structural rules and behavioural rules.

A model is of high syntactic quality if it is syntactically correct:

- Structurally correct +
- Behaviorally correct

36

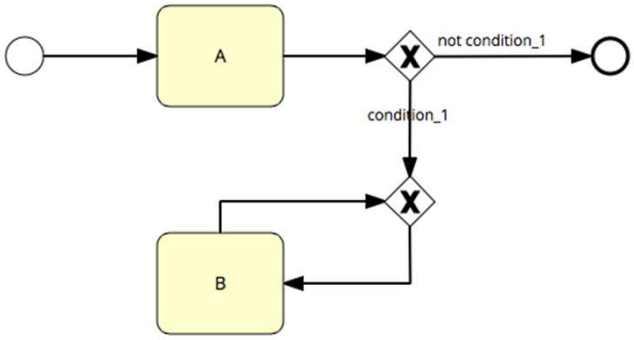
Example: structural correctness



37

Example: Behavioral correctness (no option to complete)

If condition_1 is true, the instance cannot complete and activity B will be repeated forever (livelock)



Note: this model is **structurally incorrect**, because B is not on a path to the end event

38

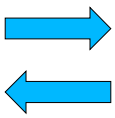
Semantic quality: validation

Semantic quality relates to the adherence of a process model to its real-world process.

Validation is the activity of checking the semantic quality of a model by comparing it with its real-world business process.

A model is of high semantic quality if it is semantically correct:

- Valid (all *model* instances are correct and relevant) +
- Complete (all possible *process* instances are covered)



Domain Expert Process Analyst

39

Exercise 5.14

Consider the following process description.

A company has two warehouses that store different products: Amsterdam and Hamburg. When an order is received, it is distributed across these warehouses: if some of the relevant products are maintained in Amsterdam, a sub-order is sent there; likewise, if some relevant products are maintained in Hamburg, a sub-order is sent there. Afterwards, the order is registered and the process completes.

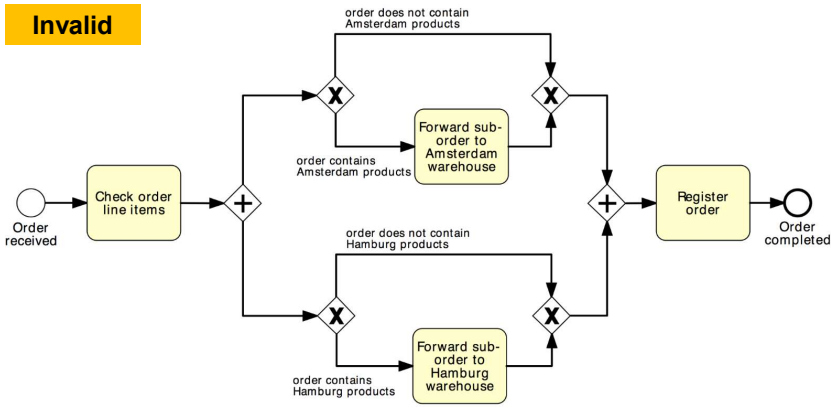
7

40

Exercise 5.14 (cont'ed)

What can we say about the semantic quality of this model?

Invalid



It is **not possible** that products are neither in the Amsterdam nor in the Hamburg warehouse.

7

41

Pragmatic Quality: Certification

Pragmatic quality relates to the usability of a process model

Challenge = anticipate the particular usage of the model

Usability:

- Understandability: how easy it is to read and comprehend the model
- Maintainability: how easy it is to apply changes
- Learning: how good a model reveals how its corresponding process works in reality

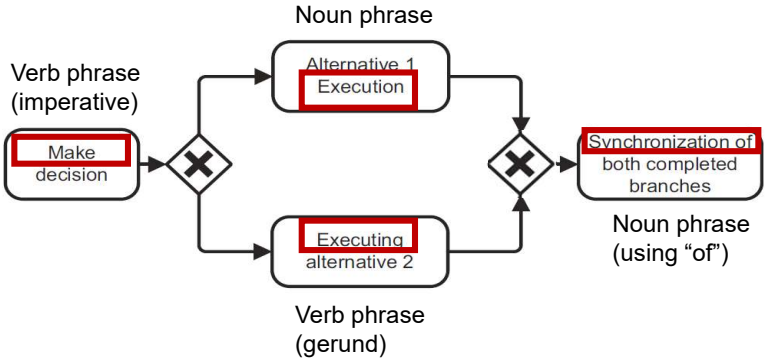
Model characteristics that influence usability include size, structural complexity and layout

42

42

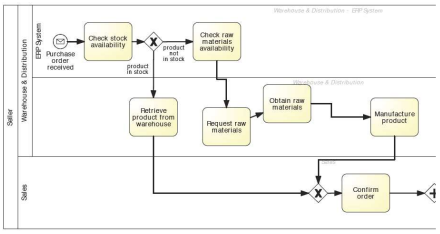
Is this process model of good pragmatic quality?

Different labeling styles

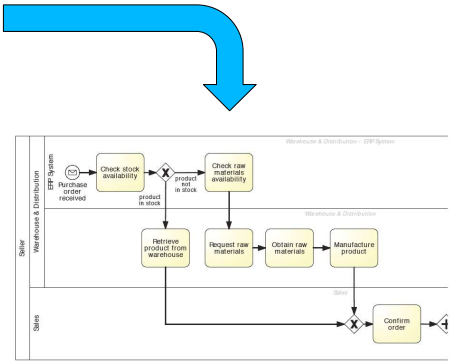


43

Pragmatic Quality: Layout



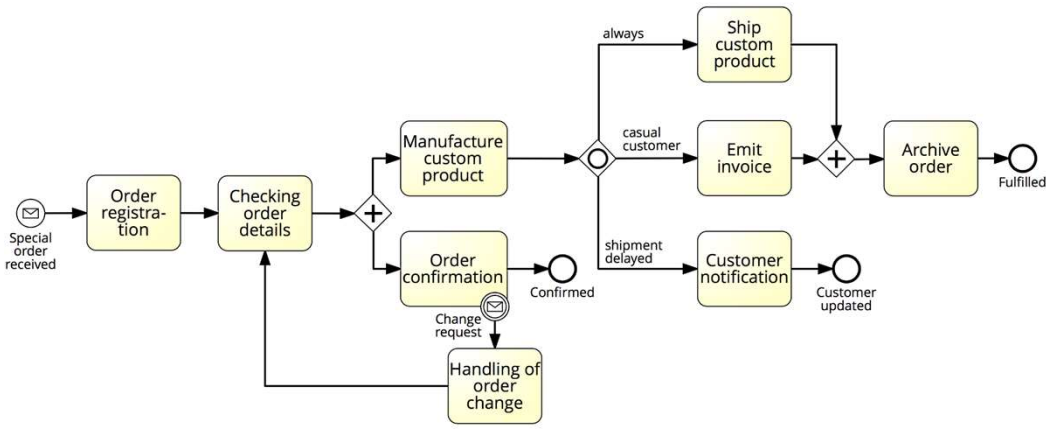
Models must look nice



44

Exercise 5.16

Is the process model below of good pragmatic quality? If not, how can it be improved?



45

Solution 5.16

- This model employs different labeling styles. For example, activities “Order registration” and “Checking order details” follow the action-noun style, while “Ship customer product” and “Emit invoice” follow the verb-object style.
- Moreover, the label of events “Confirmed” and “Fulfilled” lacks a reference to a business object (the order).
- The same applies to the boundary message event “Change request”, which in addition lacks the past-participle verb “received”.
- **To improve the pragmatic quality of this model we need to homogenize the various labeling styles, e.g., using a verb-object style for activities and an object-verb style for events.**
- The layout of this model is consistent with a left-to-right orientation, so there is no need to re-layout the model.

46

46