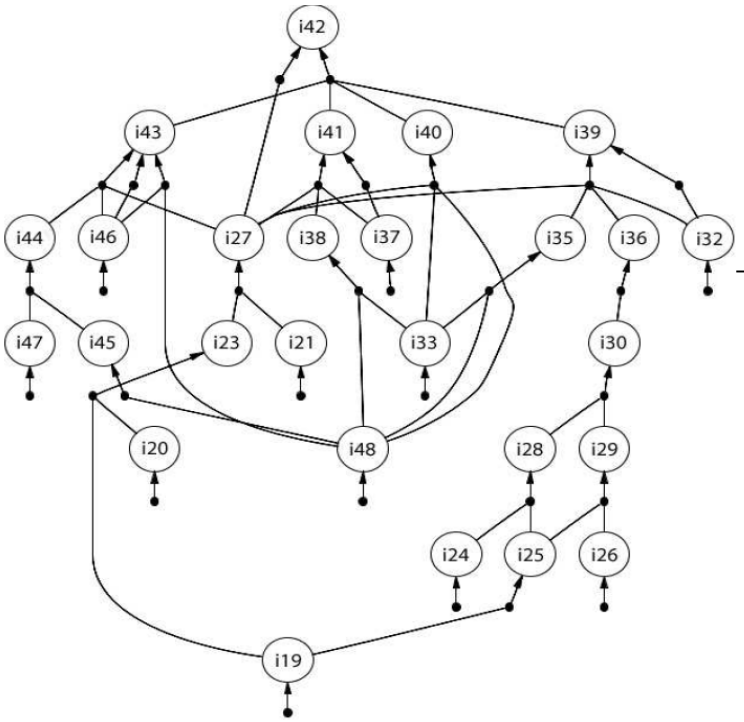
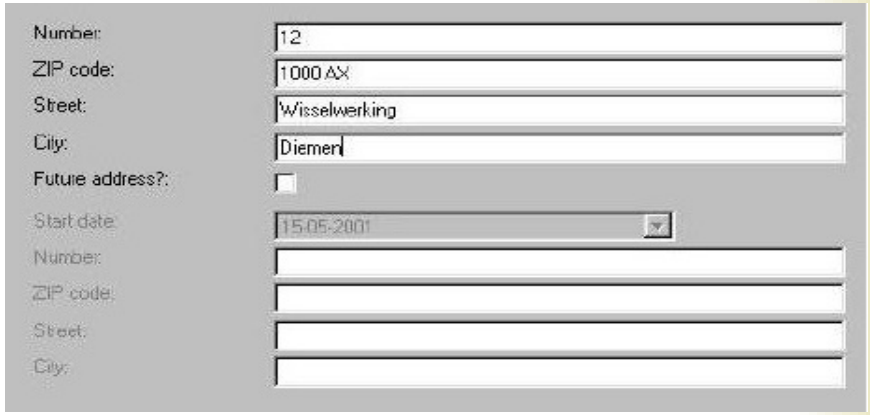
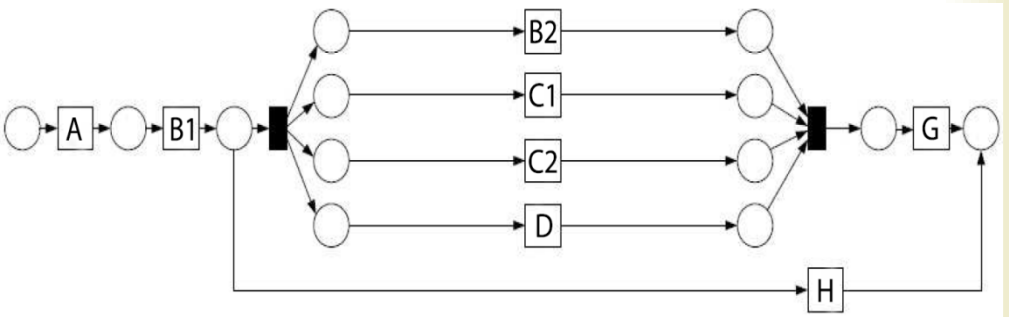
Composing Workflow Activities on the Basis of Data-Flow Structures

Authors:
Han van der Aa
Hajo Reijers
Irene Vanderfeesten

From Data to Workflow







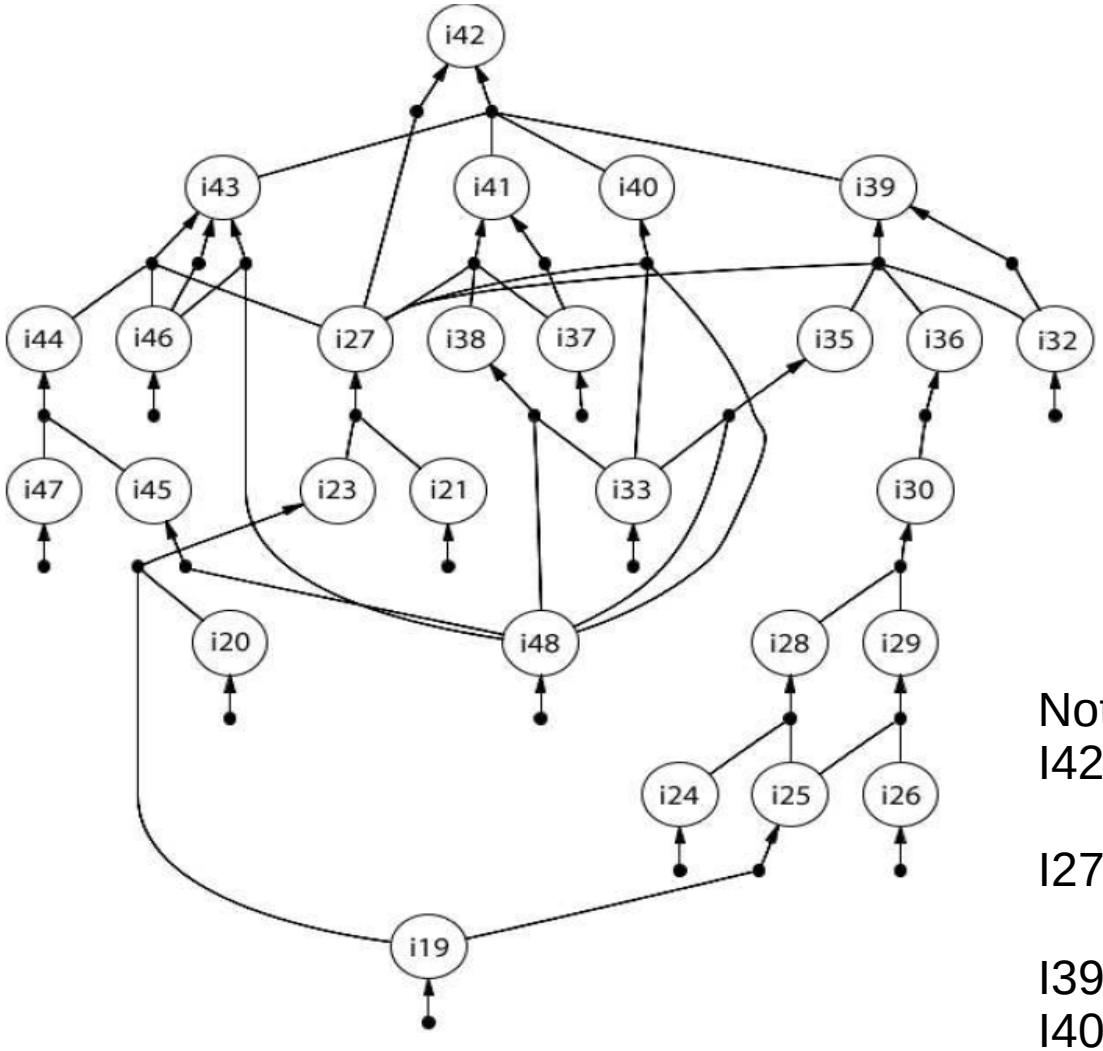
Outline

- Data-flow to workflow
- Preliminaries
- Motivating example
- Activity Composition
- Composition Guidelines
- Tool support
- Validation
- Conclusions

Product Data Model

- Operationalize ideas based on Product Data Model (PDM) stemming from Product-Based Workflow Design
- A PDM captures the structure of elementary data-processing steps that comprise a workflow
- Directed graph that consists of:
  - Data elements: pieces of information (or data)
  - Operations: elementary data processing steps
    - An operation has zero or more input data elements
    - An operation has exactly one output data element

Student Grants Example



Notable elements:
I42: total amount of student grant

I27: eligibility to receive grant

I39: amount of supplementary grant
I40: amount of basis grant
I41: amount of loan
I43: amount of tuition credit

Workflow Activity

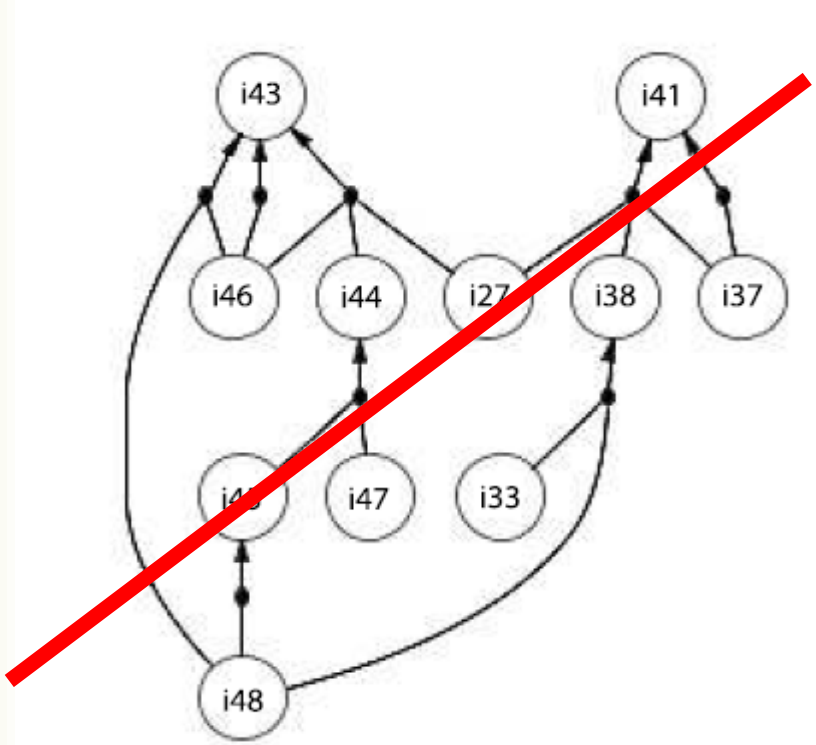
- A logical piece of work within a workflow
- Executed by a resource (e.g. an employee)
- Comprises a number of elementary data processing steps

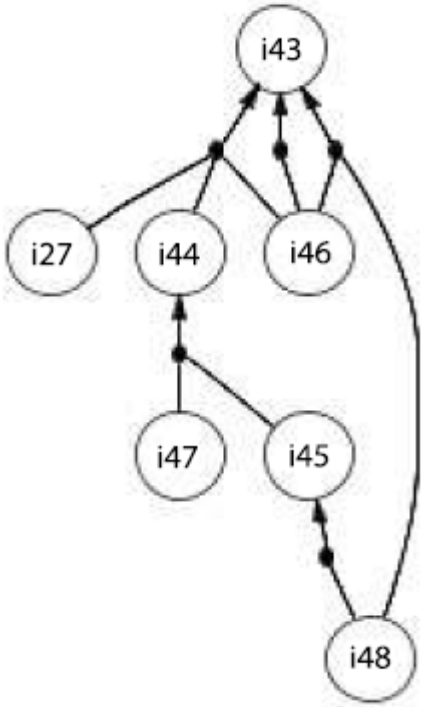
- Example --- determine a person's age:
  1) Retrieve person's date of birth
  2) Check today's date
  3) Calculate age

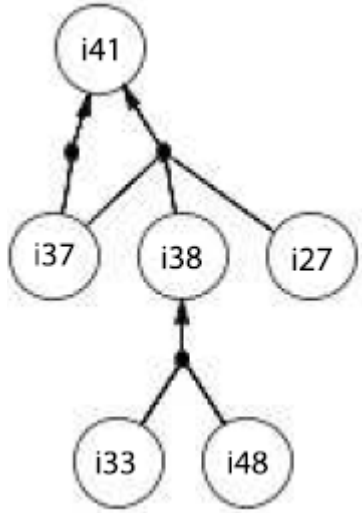
Activity Composition

- *Activity composition* is the act of grouping together elementary data processing steps into workflow activities.
- Proper composition:
  - Ensures activities are of the right granularity (hand-overs versus flexibility)
  - Creates activities that are meaningful for employees
  - Improves understandability of process models (quick overview)
- Improper composition yields unfavorable effects

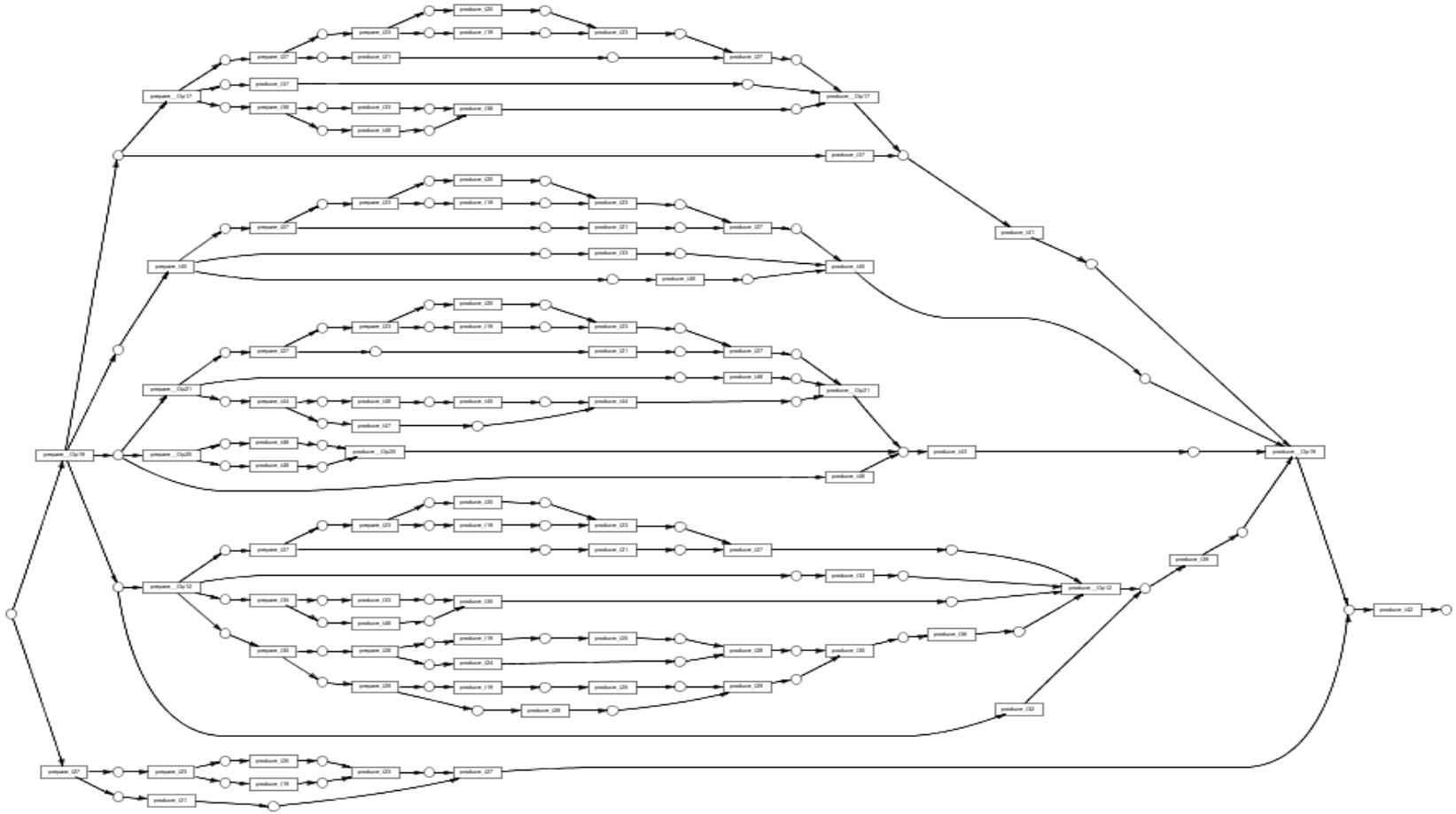
Example Activities



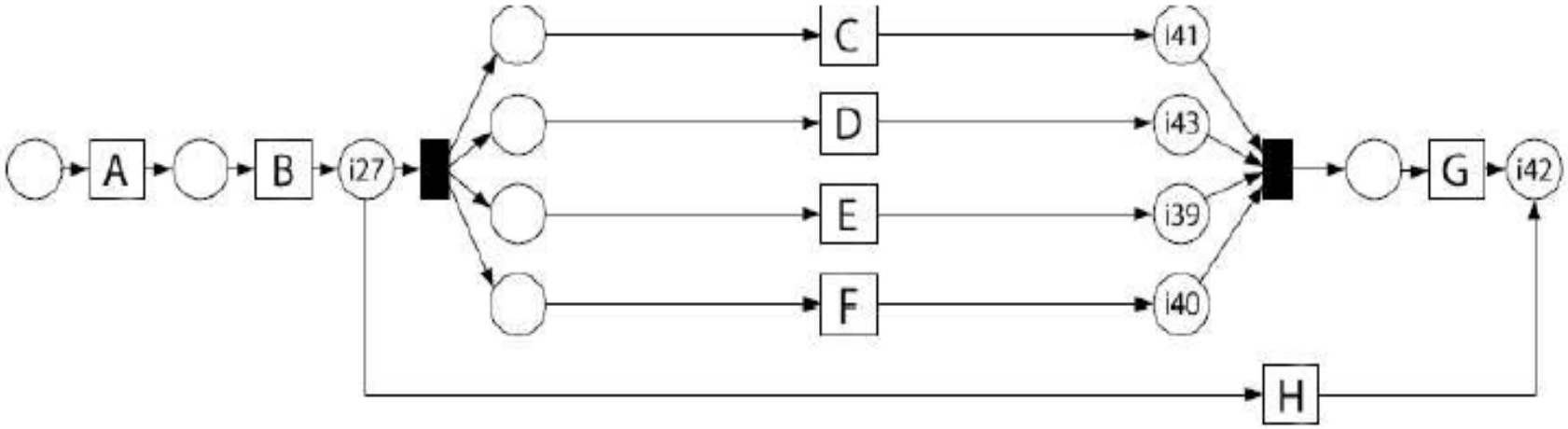




Process Model – Before Composition



Process Model – After Composition



Motivation

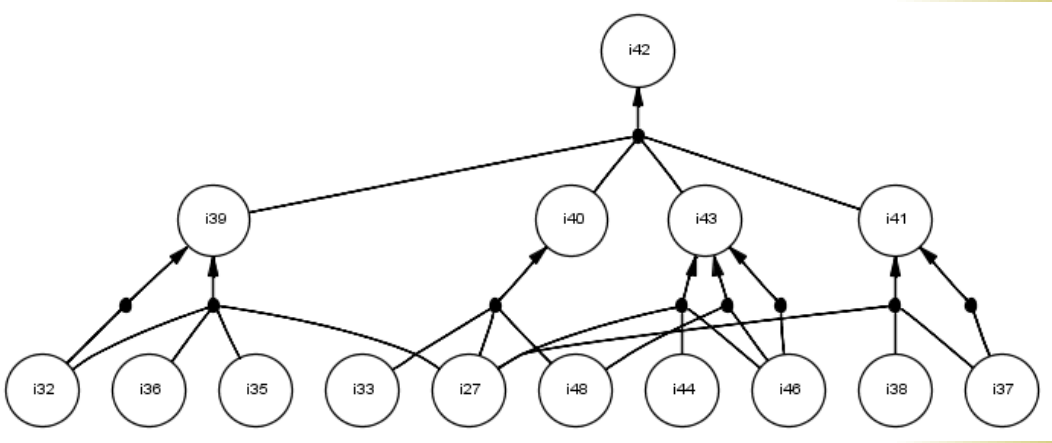
- Despite importance of activity composition, no guidelines or tool support are available
- Goal: support the task of activity composition so that it can be performed in a time-efficient manner, irrespective of case knowledge and level of expertise
- Achieved through definition of composition guidelines
- Focus on structural data-flow relations

Data Element Importance

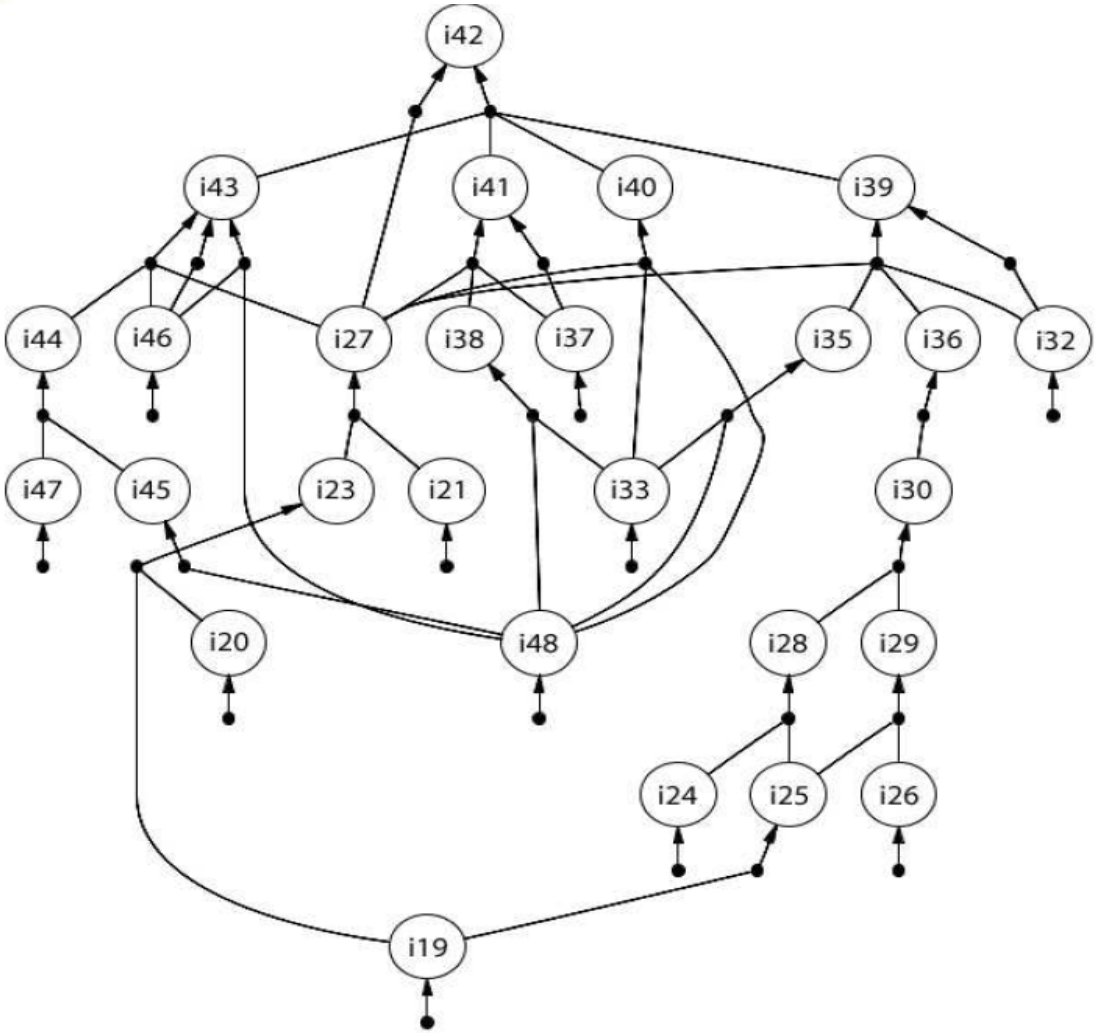
- Not all data elements in a workflow are of equal importance
  - i27: eligibility to receive a grant
  - i28: income of applicant's father
- Composition guideline 1:
  Activities should work towards the production of an important data element.
- Proposition 1:
  Important data elements in a PDM can be identified based on five structural patterns.

Data Element Importance - Patterns

1) Root Data Element
2) Leaf Data Elements
3) Conditional Data Elements
4) Equal-level Data Elements
5) Reference Data Elements



Data Element Importance (2)



1. *Root data element*
i42: total amount of student grant

2. *Leaf data elements*
11 derived from application

3. *Conditional Elements*
i39: amount of tuition fee credit
i41: amount of suppl. grant
i43: amount of loan

4. *Equal-level data elements*
i40: amount of basic grant

5. *Reference data elements*
i19: date of request
i27: eligibility to receive a grant
i33: living situation
i48: type of education

Semantic Relatedness

- Composition guideline 2:

Activities should consist of operations that are semantically related to each other.

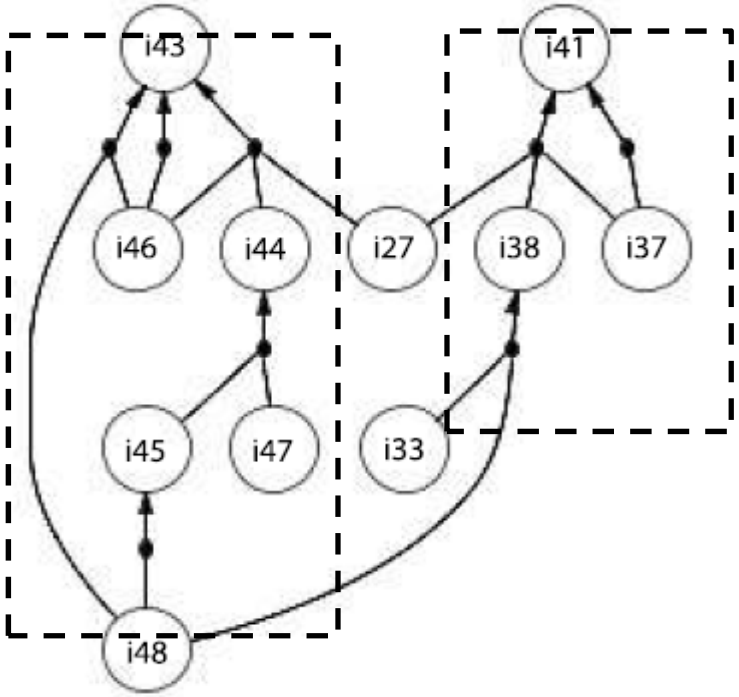
Definition (*Associated Data Element*)*:*

The Associated Data Element of an operation is the unique important data element (IDE) for which there exists a path in the PDM from the operation to that data element, such that this path does not contain any other IDEs.

- Proposition 2:

A semantically coherent activity is an activity that consists of a set of operations that have the same associated data element.

Semantic Relatedness (2)



ID	Description
i37	Requested amount of loan
i38	Maximum amount of loan
i41	Amount of loan assigned
i43	Credit for tuition fees assigned to applicant
i44	Maximum amount of credit for tuition fees
i45	Tuition fees of educational institution
i46	Applicant has requested credit for tuition fees
i47	Tuition fees declared by law

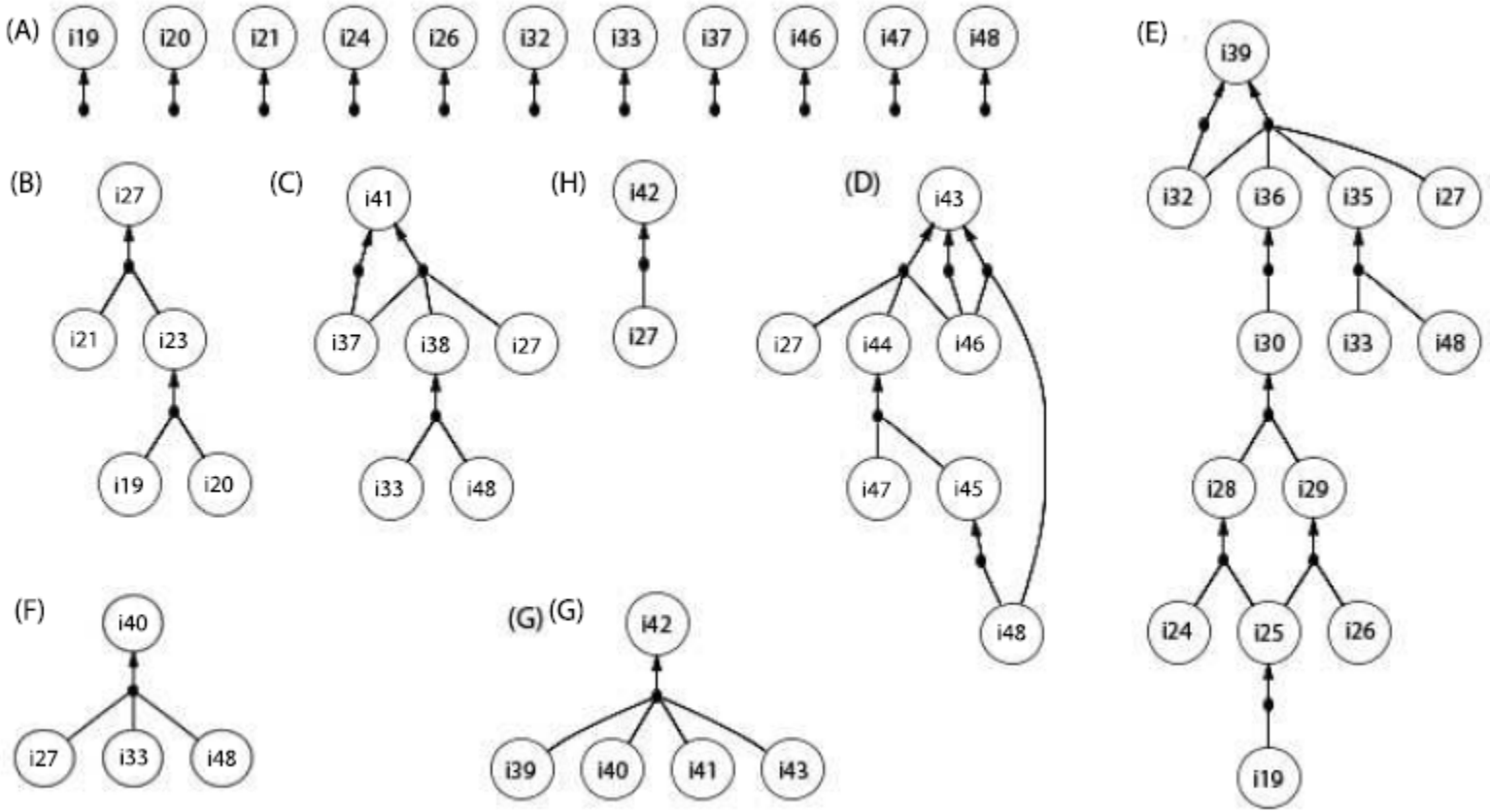
Well-Designed Activities

- Proposition 3:

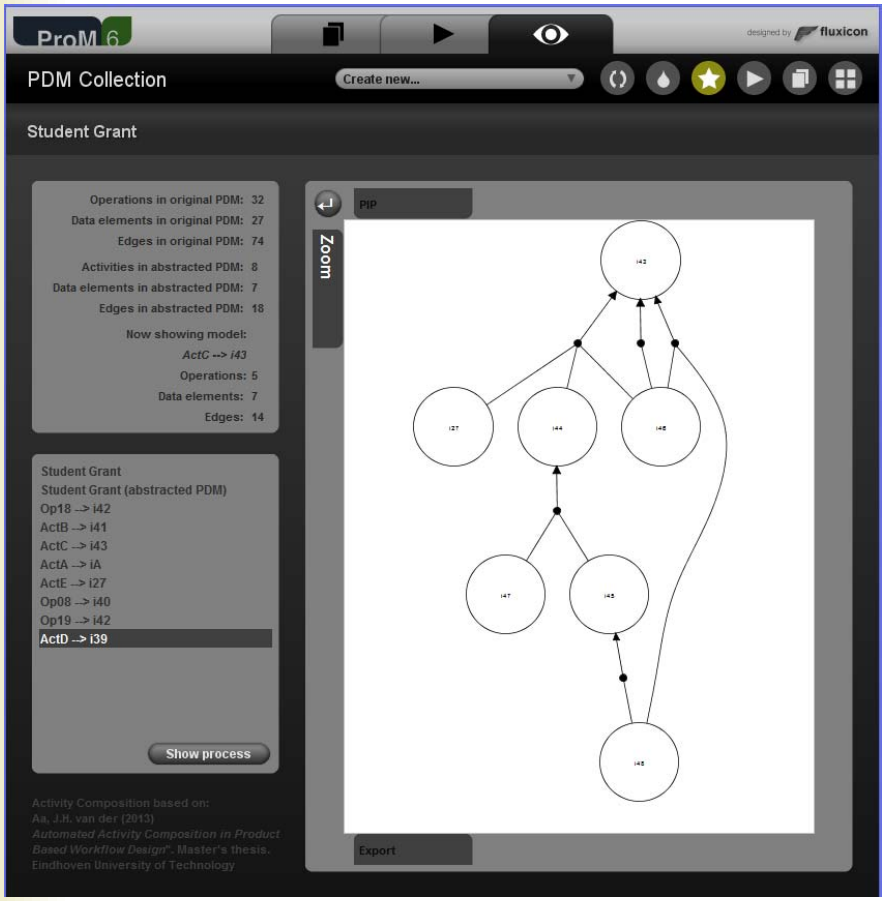
Well-designed activities work towards the production of an important data element and consist of semantically related operations.

Involved notions can be identified based on purely structural properties → approach can be automated

Well-Designed Activities (2)



Tool Support



Available at:

- www.promtools.org/prom6/nightly
- PDMAggregation package

Validation

Preliminary validation of propositions:

- 5 different workflows (4 of real-life business processes)
- 11 activity designs created by experienced modellers
- Assumed that experienced modellers adhere to our guidelines

Validation (2)

- Proposition 1 (important data elements):
  - Precision: 0.90, Recall: 0.80
- Proposition 3 (well-designed activities):
  - Jaccard-index: 0.73, Rand-index: 0.94
- Implies that modellers also take other factors into account
- Expert solutions:
  - Tend to have finer granularity
  - Include other constructs
- Generally generated designs are good approximations of manual designs
- Causes of important differences clearly identifiable

Conclusion

- Introduced fundamental composition guidelines
- Lack of support for activity composition addressed
- Activities should work towards important data element and consist of semantically related operations
- These properties can be identified based on structural data-flow relations
- Preliminary validation justifies the guidelines in the context of existing business processes
- Incorporation of information beyond structure presents interesting opportunities for future research