

# Mapping the NRC Dataflow Model to the Open Provenance Model

Natalia Kwasnikowska  
and  
Jan Van den Bussche

Hasselt University and Transnational University of Limburg  
Belgium



IPAW  
June 17-18, 2008



# Overview

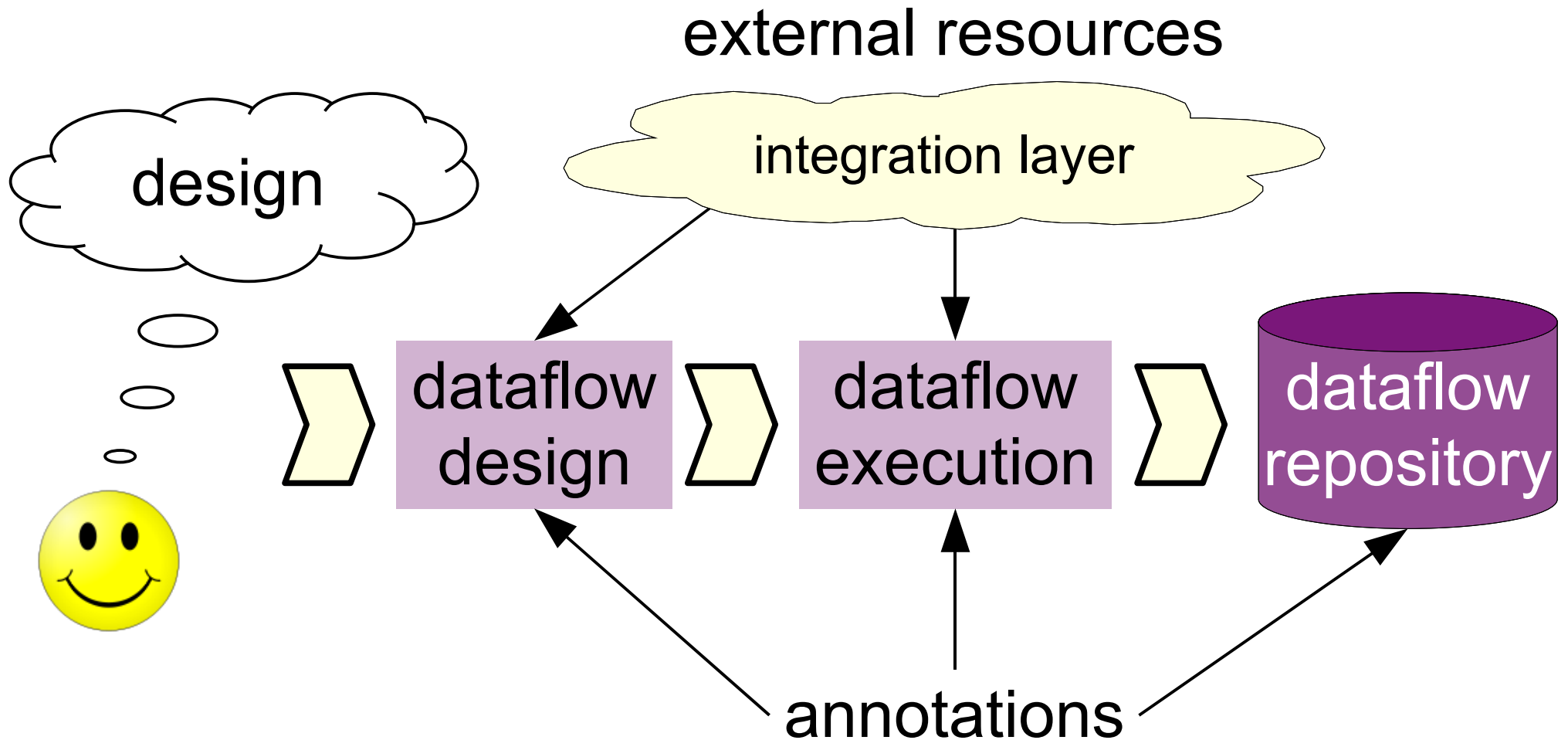
- Example
- Nested Relational Calculus (NRC) Dataflow Model
- Open Provenance Model (OPM)
- Mapping of NRC dataflow executions to OPM

# Example

Given a set of **raw data** produced in a proteomics experiment, generate a list of possibly identified **proteins**



# Nested Relational Calculus Dataflow Model



# Dataflow Design

```
dataflow identify (data: TMSdata): ProteinCandidateList is  
  let list := for x in data return  
    < id: x.id, spectra: extract(x.file) >  
  in validation(search_1(list), search_2(list), search_3(list))
```

```
dataflow search (data: TMSextracted): AAList is  
  for x in data return  
    < id: x.id, aalist: for y in x.spectra return dbSearch(y) >
```

# Base and Complex Types

dataflow *identify* (data: TMSdata): ProteinCandidateList is  
let list := for x in data return  
    ⟨ id: x.id, spectra: *extract*(x.file) ⟩  
in *validata* { ⟨ id: **Number**, spectra: { **TMSfile** } ⟩ } (list))

dataflow *search* (data: TMSextracted): AAList is  
for x in data return  
    ⟨ id: x.id, aalist: for y in x.spectra return *dbSearch*(y) ⟩

# Service Names

dataflow *identify* (data: TMSdata): ProteinCandidateList is  
let list := for x in data return  
    ⟨ id: x.id, spectra: *extract*(x.file) ⟩  
in *validation*(*search\_1*(list), *search\_2*(list), *search\_3*(list))

dataflow *search* (data: TMSextracted): AAList is  
for x in data return  
    ⟨ id: x.id, aalist: for y in x.spectra return *dbSearch*(y) ⟩

# Late Binding of Service Names

```
dataflow identify (data: TMSdata): extract candidateList is  
let validate in data return  
  < id: x.id, spectra: extract(x.file) >  
in validation(search_1(list), search_2(list), search_3(list))
```

```
dataflow search (data: TMSextracted): AAlist is  
for x in data return  
  < id: x.id, aalist: for y in x.spectra return dbSearch(y) >
```



# Late Binding of Service Names

```
dataflow identify (data: TMSdata): extract candidateList is  
let validate in data return  
  < id: x.id, spectra: extract(x.file) >  
in validation(search_1(list), search_2(list), search_3(list))
```

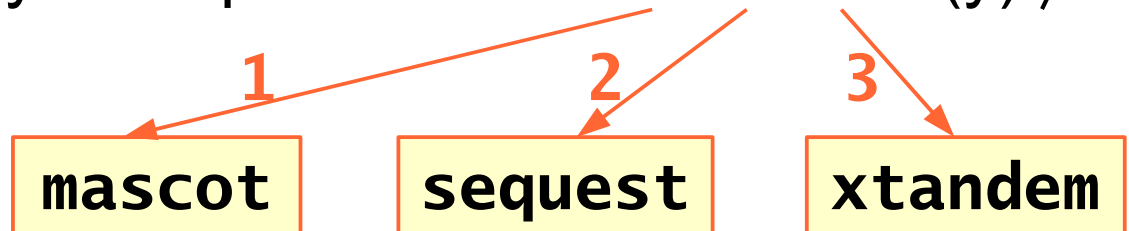
```
dataflow search (data: TMSextracted): AAList is  
for x in data return  
  < id: x.id, aalist: for y in x.spectra return dbSearch(y) >
```

1 2 3

# Late Binding of Service Names

```
dataflow identify (data: TMSdata): extract candidateList is  
let validate in data return  
  < id: x.id, spectra: extract(x.file) >  
in validation(search_1(list), search_2(list), search_3(list))
```

```
dataflow search (data: TMSextracted): AAList is  
for x in data return  
  < id: x.id, aalist: for y in x.spectra return dbSearch(y) >
```



# Execution of *identify*(tmsInput)

dataflow *identify*(data: TMSdata): ProteinCandidateList is

let list := for x in data return

    ⟨ id: x.id, spectra: *extract*(x.file) ⟩

in *validation*(*search\_1*(list), *search\_2*(list), *search\_3*(list))

## tmsOutput

### tmsInput

<i>id</i>	<i>file</i>
1	rawVial10
2	rawVial11
⋮	⋮
42	rawVial51
⋮	⋮
55	rawVial64

<i>protID</i>	<i>prob</i>	<i>evidence</i>		
protein <sub>1</sub>	99	<i>peptide</i>	<i>score</i>	<i>spectrum</i>
		pep <sub>1</sub>	9	spectrum <sub>vial10,5</sub>
		pep <sub>2</sub>	7	spectrum <sub>vial23,2</sub>
		⋮	⋮	⋮
protein <sub>2</sub>	96	<i>peptide</i>	<i>score</i>	<i>spectrum</i>
		pep <sub>8</sub>	9	spectrum <sub>vial51,3</sub>
		pep <sub>2</sub>	8	spectrum <sub>vial10,5</sub>
		⋮	⋮	⋮
⋮	⋮	⋮		

# Run of *identify*(tmsInput)

<i>subexpression</i>	<i>assignment</i>	<i>value</i>				
let	data = tmsInput	tmsOutput				
for	data = tmsInput	tmsExtracted				
<i>validation</i>	data = tmsInput list = tmsExtracted p1 = tmsResult_1 p2 = tmsResult_2 p3 = tmsResult_3	tmsOutput				
⟨id, spectra⟩	data = tmsInput x = ⟨id: 2, file: rawVial10⟩	<table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial10,1</sub> spectrum<sub>vial10,2</sub> spectrum<sub>vial10,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial10,1</sub> spectrum <sub>vial10,2</sub> spectrum <sub>vial10,3</sub>
<i>id</i>	<i>spectra</i>					
2	spectrum <sub>vial10,1</sub> spectrum <sub>vial10,2</sub> spectrum <sub>vial10,3</sub>					
<i>extract</i>	data = tmsInput x = ⟨id: 2, file: rawVial10⟩ raw = rawVial10	<table border="1"> <tbody> <tr> <td>spectrum<sub>vial10,1</sub> spectrum<sub>vial10,2</sub> spectrum<sub>vial10,3</sub></td> </tr> </tbody> </table>	spectrum <sub>vial10,1</sub> spectrum <sub>vial10,2</sub> spectrum <sub>vial10,3</sub>			
spectrum <sub>vial10,1</sub> spectrum <sub>vial10,2</sub> spectrum <sub>vial10,3</sub>						
⋮	⋮	⋮				
<i>extract</i>	data = tmsInput x = ⟨id: 55, file: rawVial64⟩ raw = rawVial64	<table border="1"> <tbody> <tr> <td>spectrum<sub>vial64,1</sub> spectrum<sub>vial64,2</sub></td> </tr> </tbody> </table>	spectrum <sub>vial64,1</sub> spectrum <sub>vial64,2</sub>			
spectrum <sub>vial64,1</sub> spectrum <sub>vial64,2</sub>						
<i>search_1</i>	data = tmsInput list = tmsExtracted	tmsResult_1				
<i>search_2</i>	data = tmsInput list = tmsExtracted	tmsResult_2				
<i>search_3</i>	data = tmsInput list = tmsExtracted	tmsResult_3				

# Runs of *search*(tmsExtracted)

	<i>subexpression</i>	<i>assignment</i>	<i>value</i>																			
<b><i>subexpression</i></b>	for x	data = tmsExtracted	tmsResult_2																			
	⋮	⋮	⋮																			
let for																						
<b><i>validation</i></b>	⟨id, alist⟩	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>id</i></th> <th colspan="2"><i>alist</i></th> </tr> <tr> <td></td> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td></td> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>id</i>	<i>alist</i>			<i>spectrum</i>	<i>results</i>		spectrum <sub>vial11,1</sub>	match1 match2		spectrum <sub>vial11,2</sub>	match1 ⋮ match5		spectrum <sub>vial11,3</sub>	match1
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>id</i>	<i>alist</i>																					
	<i>spectrum</i>	<i>results</i>																				
	spectrum <sub>vial11,1</sub>	match1 match2																				
	spectrum <sub>vial11,2</sub>	match1 ⋮ match5																				
	spectrum <sub>vial11,3</sub>	match1																				
⟨id, spectra⟩																						
<b><i>extract</i></b>	for y	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1							
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,1</sub>	match1 match2																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
spectrum <sub>vial11,3</sub>	match1																					
⋮																						
<b><i>extract</i></b>																						
<b><i>search_1</i></b>	<b><i>dbSearch</i></b>	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,2</sub>	match1 ⋮ match5											
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
<b><i>search_2</i></b>		y = spectrum <sub>vial11,2</sub>																				
⋮	⋮	⋮	⋮																			
<b><i>search_3</i></b>	data = tmsInput list = tmsExtracted		tmsResult_3																			

# Runs of *search*(tmsExtracted)

	<i>subexpression</i>	<i>assignment</i>	<i>value</i>																			
<b><i>subexpression</i></b>	for x	data = tmsExtracted	tmsResult_1																			
	⋮	⋮	⋮																			
let for																						
<b><i>validation</i></b>	⟨id, alist⟩	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>id</i></th> <th colspan="2"><i>alist</i></th> </tr> <tr> <td></td> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td></td> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>id</i>	<i>alist</i>			<i>spectrum</i>	<i>results</i>		spectrum <sub>vial11,1</sub>	match1 match2		spectrum <sub>vial11,2</sub>	match1 ⋮ match5		spectrum <sub>vial11,3</sub>	match1
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>id</i>	<i>alist</i>																					
	<i>spectrum</i>	<i>results</i>																				
	spectrum <sub>vial11,1</sub>	match1 match2																				
	spectrum <sub>vial11,2</sub>	match1 ⋮ match5																				
	spectrum <sub>vial11,3</sub>	match1																				
	⟨id, spectra⟩																					
<b><i>extract</i></b>	for y	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1							
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,1</sub>	match1 match2																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
spectrum <sub>vial11,3</sub>	match1																					
	⋮																					
<b><i>extract</i></b>		data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,2</sub>	match1 ⋮ match5											
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
<b><i>search_1</i></b>	<b><i>dbSearch</i></b>	x = spectrum <sub>vial11,2</sub> y = spectrum <sub>vial11,2</sub>																				
<b><i>search_2</i></b>	⋮	⋮	⋮																			
<b><i>search_3</i></b>	0	⋮	⋮																			
	list = tmsExtracted																					

# Runs of *search*(tmsExtracted)

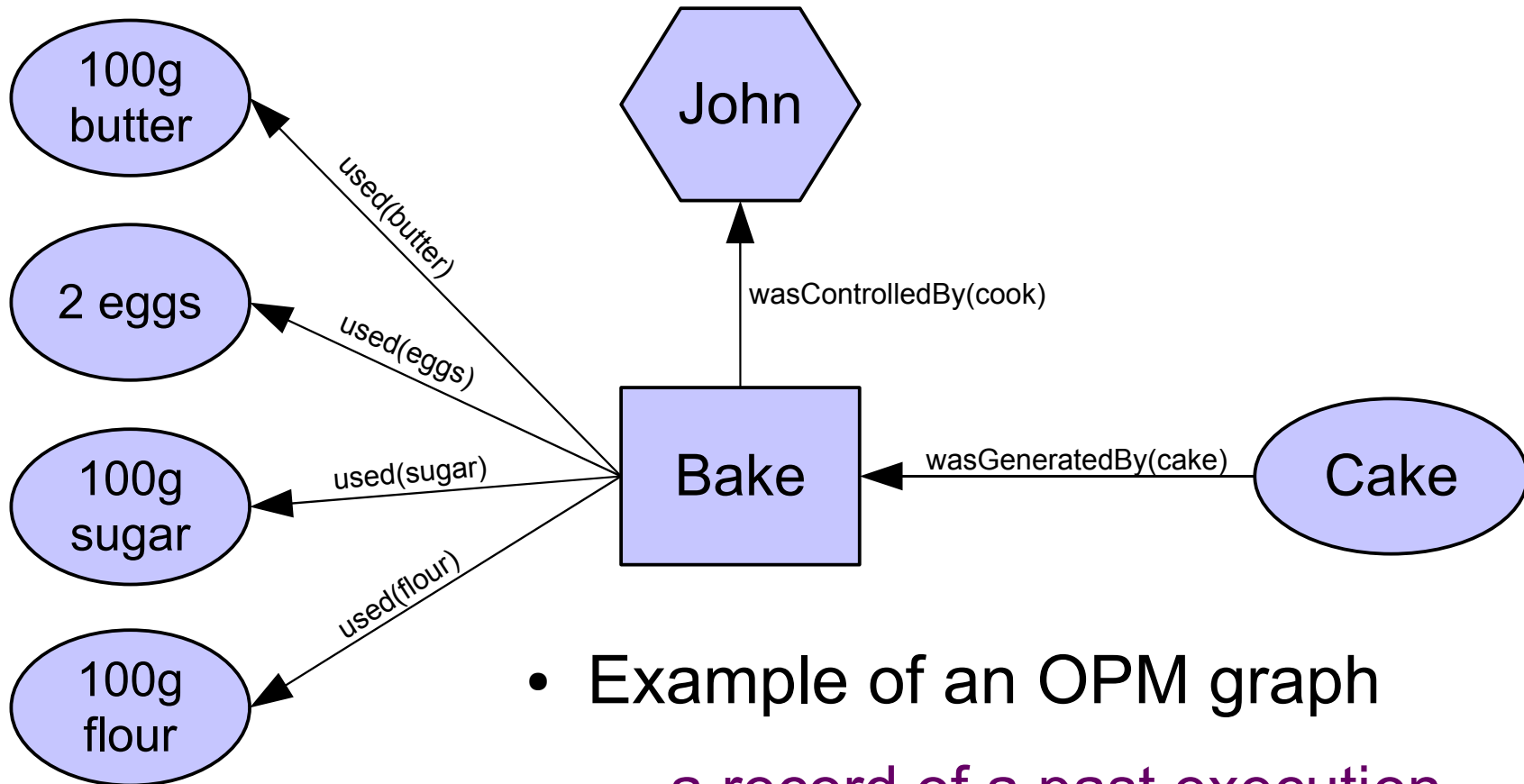
	<i>subexpression</i>	<i>assignment</i>	<i>value</i>																
<b><i>subexpression</i></b>	for x	data = tmsExtracted	tmsResult_1																
	⋮	⋮	⋮																
let for																			
<b><i>validation</i></b>	⟨id, alist⟩	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>alist</i></th> </tr> </thead> <tbody> <tr> <td></td> <td> <table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table> </td> </tr> </tbody> </table>	<i>id</i>	<i>alist</i>		<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1
<i>id</i>	<i>spectra</i>																		
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																		
<i>id</i>	<i>alist</i>																		
	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1										
<i>spectrum</i>	<i>results</i>																		
spectrum <sub>vial11,1</sub>	match1 match2																		
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																		
spectrum <sub>vial11,3</sub>	match1																		
	⟨id, spectra⟩																		
<b><i>extract</i></b>	for y	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1				
<i>id</i>	<i>spectra</i>																		
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																		
<i>spectrum</i>	<i>results</i>																		
spectrum <sub>vial11,1</sub>	match1 match2																		
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																		
spectrum <sub>vial11,3</sub>	match1																		
	⋮																		
<b><i>extract</i></b>		data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,2</sub>	match1 ⋮ match5								
<i>id</i>	<i>spectra</i>																		
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																		
<i>spectrum</i>	<i>results</i>																		
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																		
<b><i>search_1</i></b>	<b><i>dbSearch</i></b>	x = 2 y = spectrum <sub>vial11,2</sub>																	
<b><i>search_2</i></b>	⋮	⋮	⋮																
<b><i>search_3</i></b>	⋮	⋮	⋮																

# Record of a Past Execution of *identify*(tmsInput)

	<i>subexpression</i>	<i>assignment</i>	<i>value</i>																			
<b><i>subexpression</i></b>	for x	data = tmsExtracted	tmsResult_1																			
	⋮	⋮	⋮																			
let																						
for																						
<b><i>validation</i></b>	⟨id, alist⟩	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>id</i></th> <th colspan="2"><i>alist</i></th> </tr> <tr> <td></td> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td></td> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td></td> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>id</i>	<i>alist</i>			<i>spectrum</i>	<i>results</i>		spectrum <sub>vial11,1</sub>	match1 match2		spectrum <sub>vial11,2</sub>	match1 ⋮ match5		spectrum <sub>vial11,3</sub>	match1
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>id</i>	<i>alist</i>																					
	<i>spectrum</i>	<i>results</i>																				
	spectrum <sub>vial11,1</sub>	match1 match2																				
	spectrum <sub>vial11,2</sub>	match1 ⋮ match5																				
	spectrum <sub>vial11,3</sub>	match1																				
	⟨id, spectra⟩																					
<b><i>extract</i></b>	for y	data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,1</sub></td> <td>match1 match2</td> </tr> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> <tr> <td>spectrum<sub>vial11,3</sub></td> <td>match1</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,1</sub>	match1 match2	spectrum <sub>vial11,2</sub>	match1 ⋮ match5	spectrum <sub>vial11,3</sub>	match1							
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,1</sub>	match1 match2																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
spectrum <sub>vial11,3</sub>	match1																					
	⋮																					
<b><i>extract</i></b>		data = tmsExtracted <table border="1"> <thead> <tr> <th><i>id</i></th> <th><i>spectra</i></th> </tr> </thead> <tbody> <tr> <td>2</td> <td>spectrum<sub>vial11,1</sub> spectrum<sub>vial11,2</sub> spectrum<sub>vial11,3</sub></td> </tr> </tbody> </table>	<i>id</i>	<i>spectra</i>	2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>	<table border="1"> <thead> <tr> <th><i>spectrum</i></th> <th><i>results</i></th> </tr> </thead> <tbody> <tr> <td>spectrum<sub>vial11,2</sub></td> <td>match1 ⋮ match5</td> </tr> </tbody> </table>	<i>spectrum</i>	<i>results</i>	spectrum <sub>vial11,2</sub>	match1 ⋮ match5											
<i>id</i>	<i>spectra</i>																					
2	spectrum <sub>vial11,1</sub> spectrum <sub>vial11,2</sub> spectrum <sub>vial11,3</sub>																					
<i>spectrum</i>	<i>results</i>																					
spectrum <sub>vial11,2</sub>	match1 ⋮ match5																					
<b><i>search_1</i></b>	<b><i>dbSearch</i></b>	x = 2 y = spectrum <sub>vial11,2</sub>																				
<b><i>search_2</i></b>	⋮	⋮	⋮																			
<b><i>search_3</i></b>	⋮	⋮	⋮																			

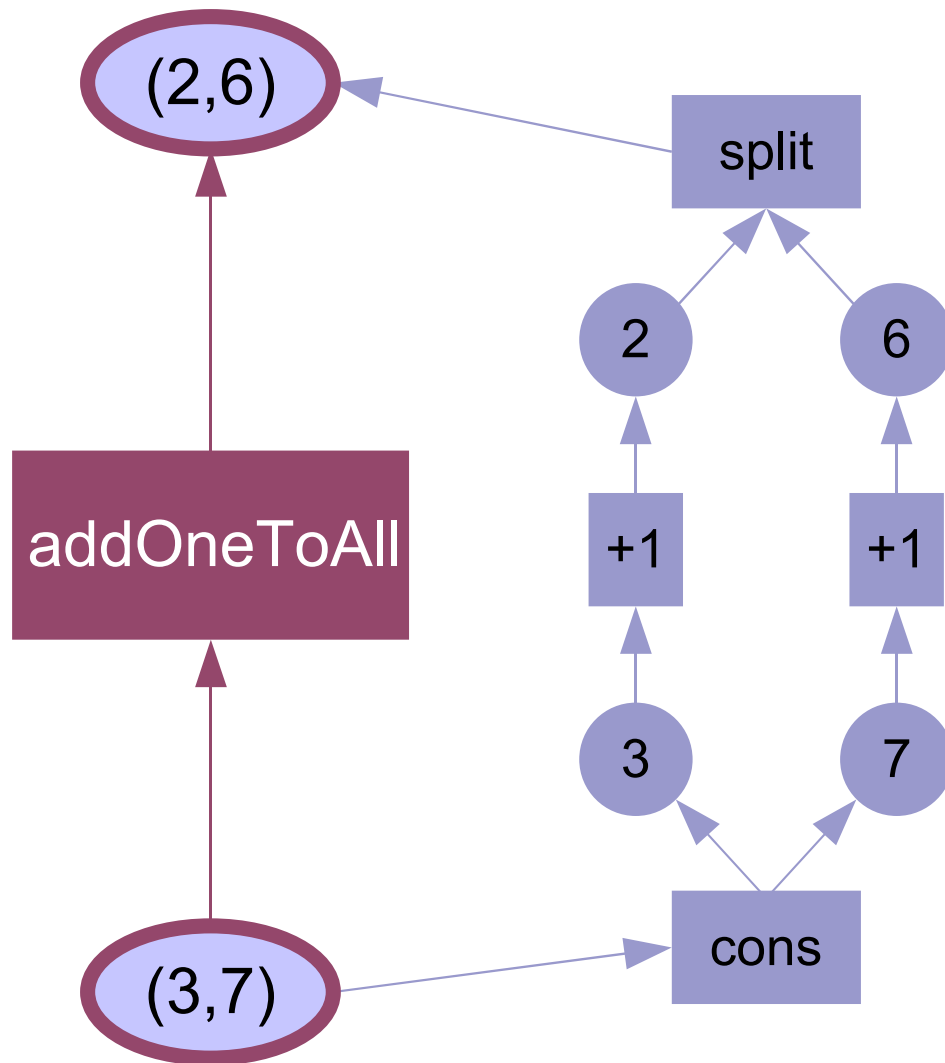


# Open Provenance Model



- Example of an OPM graph
  - a record of a past execution
  - 3 types of nodes
  - 3 types of edges (5 in the model)

# Open Provenance Model

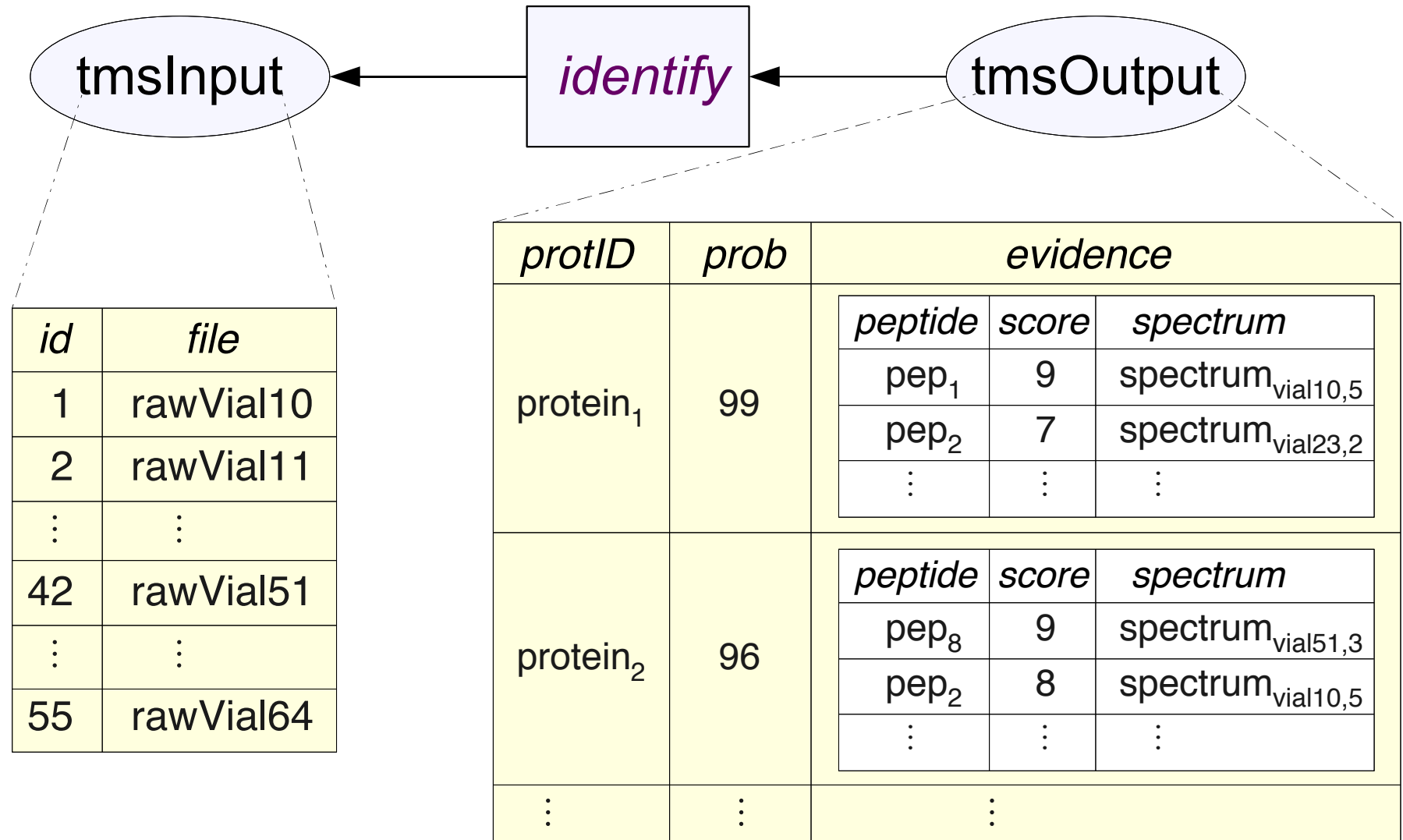


Alternate **accounts** of  
a past execution

# A Trivial OPM Graph for *identify*(tmsInput)



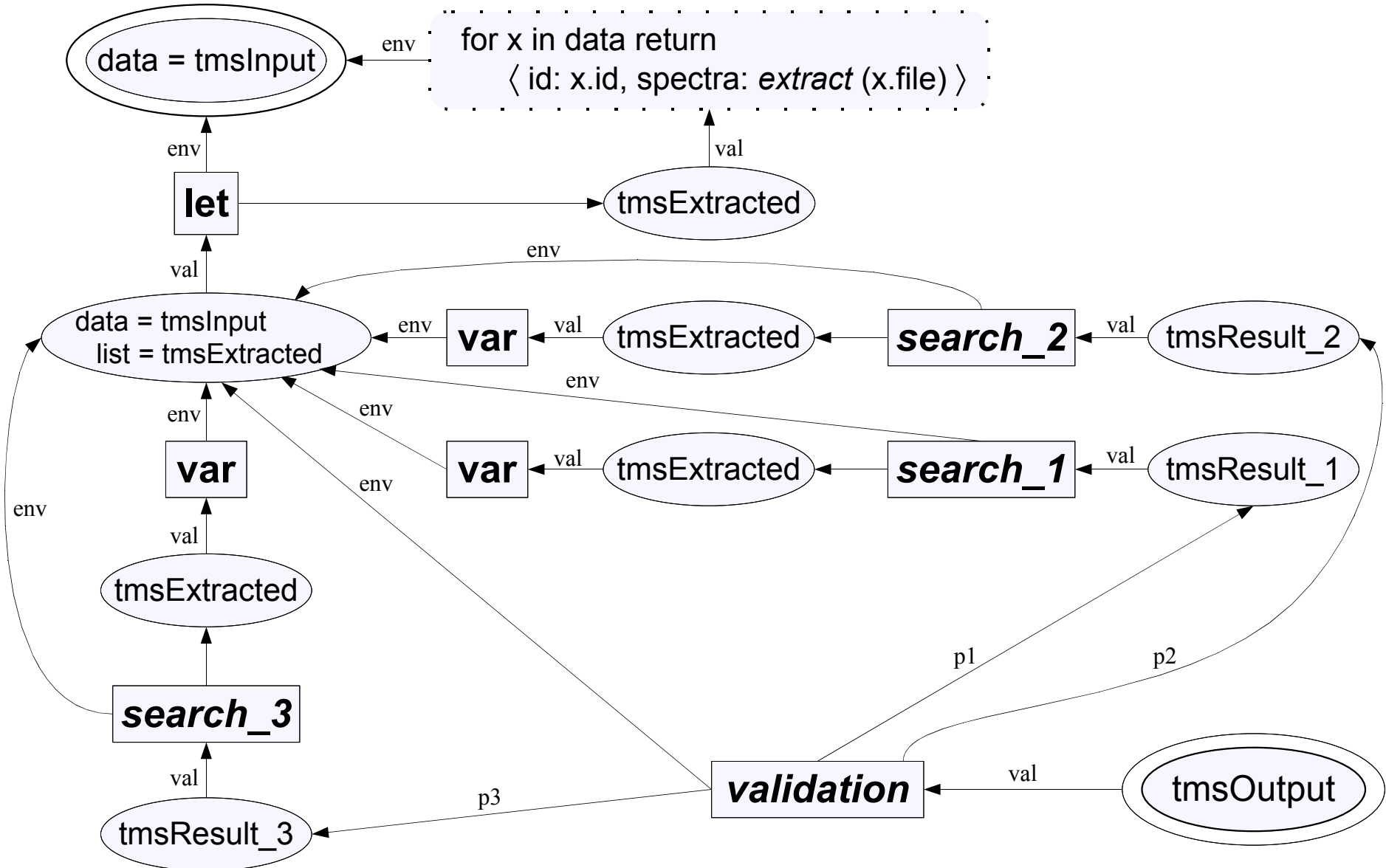
# A Trivial OPM Graph for *identify*(tmsInput)



```

dataflow identify(data: TMSdata): ProteinCandidateList is
  let list := for x in data return
    < id: x.id, spectra: extract(x.file) >
  in validation(search_1(list), search_2(list), search_3(list))

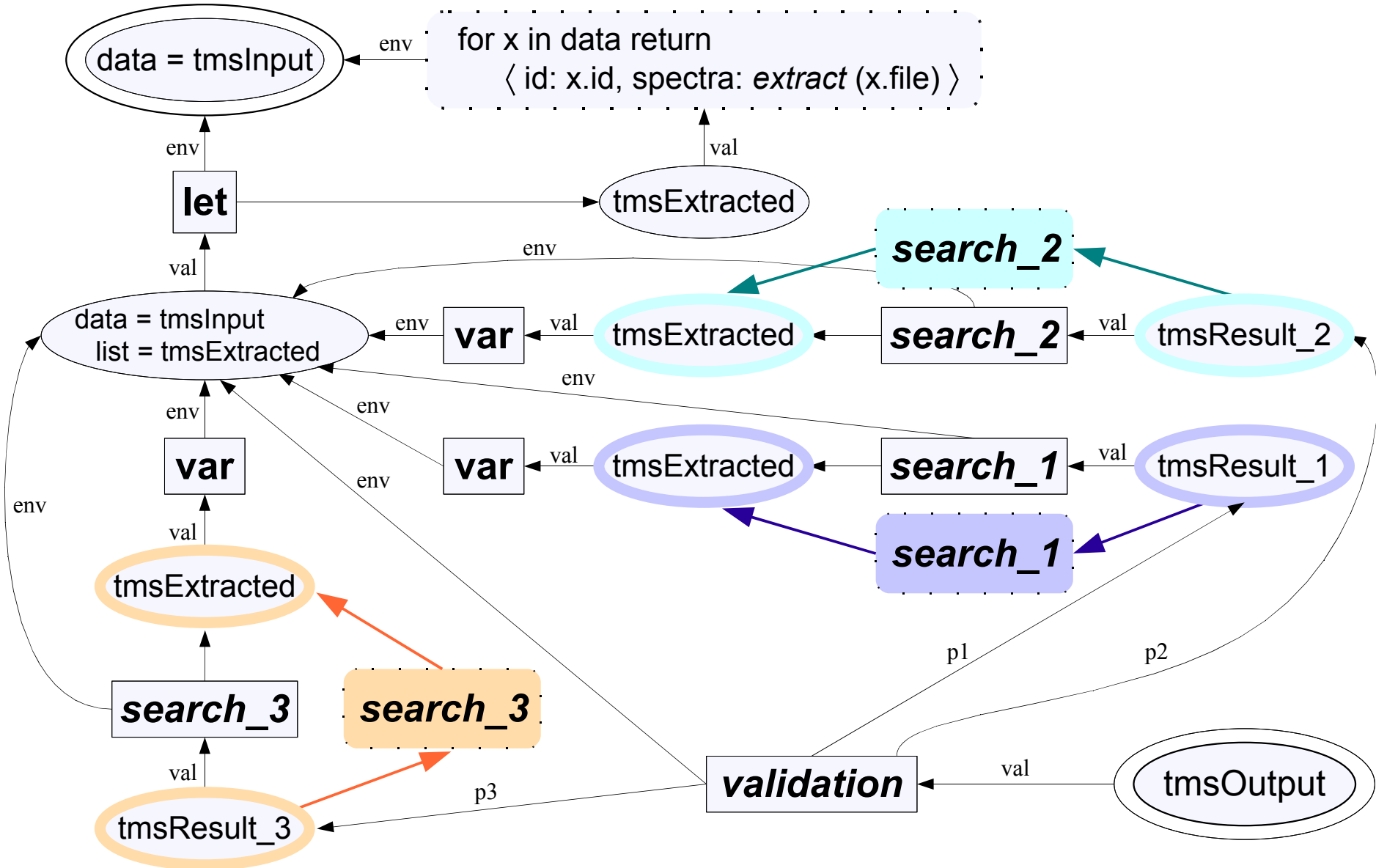
```



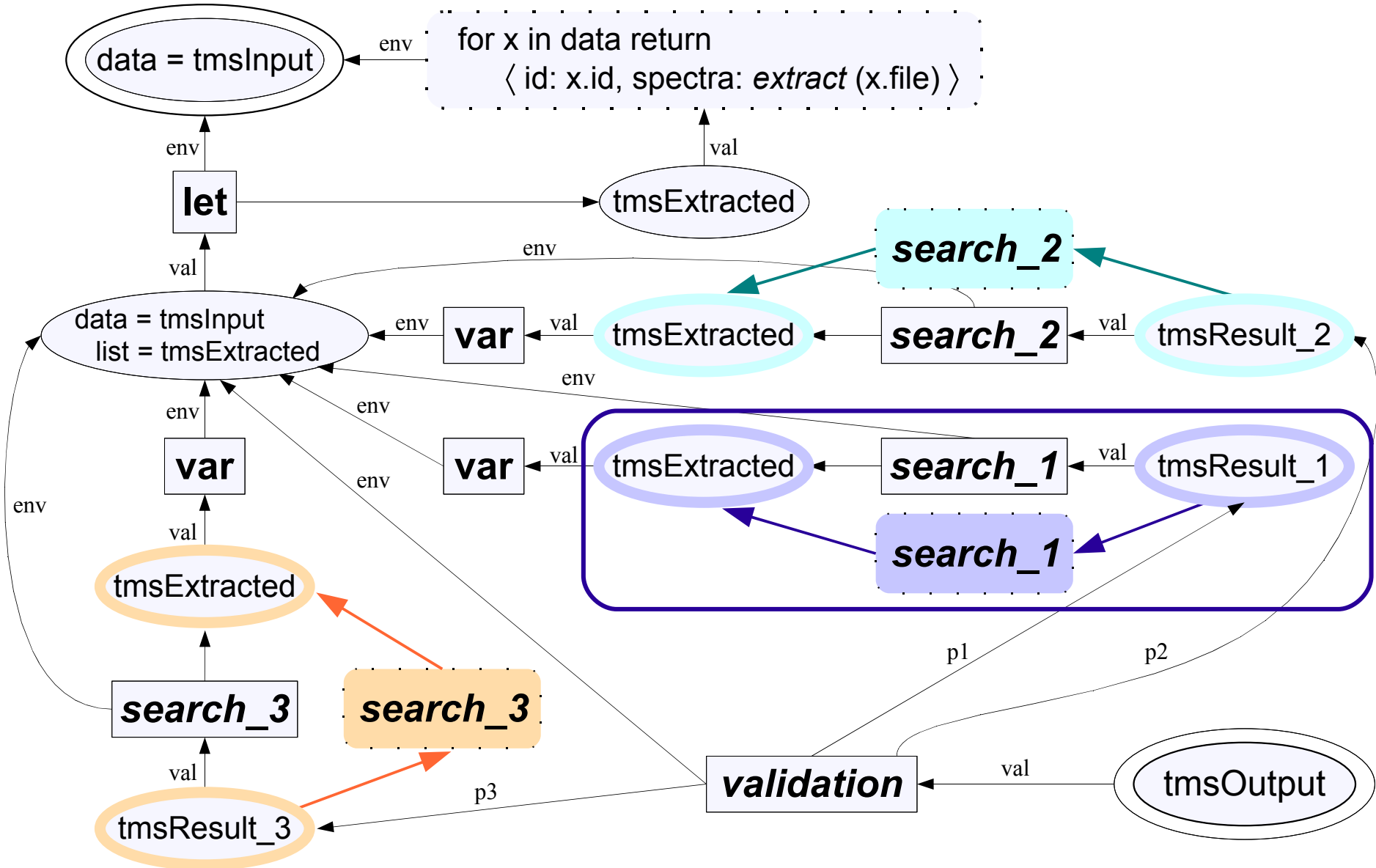
```

dataflow identify(data: TMSdata): ProteinCandidateList is
  let list := for x in data return
    < id: x.id, spectra: extract(x.file) >
  in validation(search_1(list), search_2(list), search_3(list))

```

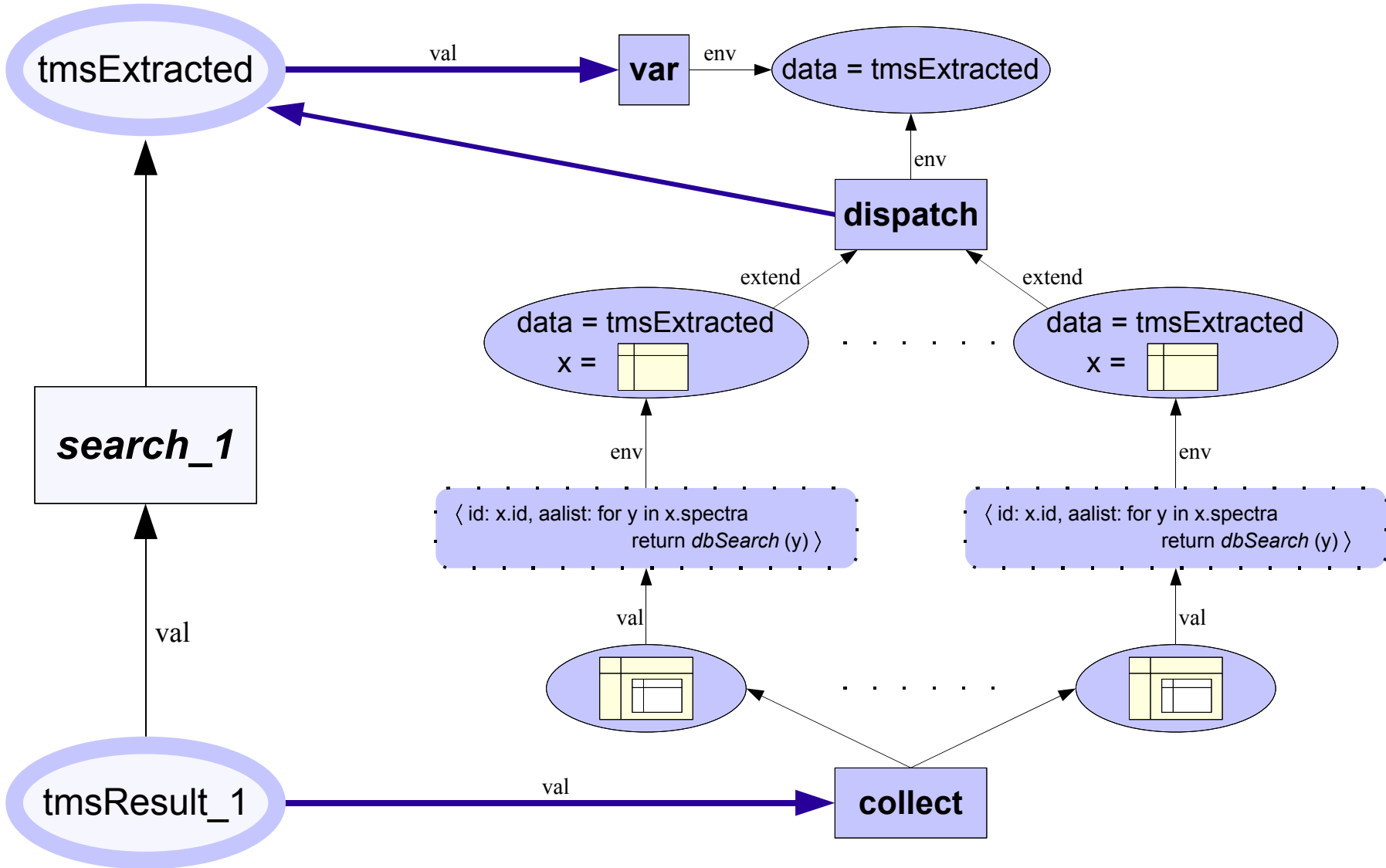


dataflow *identify*(data: TMSdata): ProteinCandidateList is  
 let list := for x in data return  
     ⟨ id: x.id, spectra: *extract*(x.file) ⟩  
 in *validation*(*search\_1*(list), *search\_2*(list), *search\_3*(list))



# Accounts

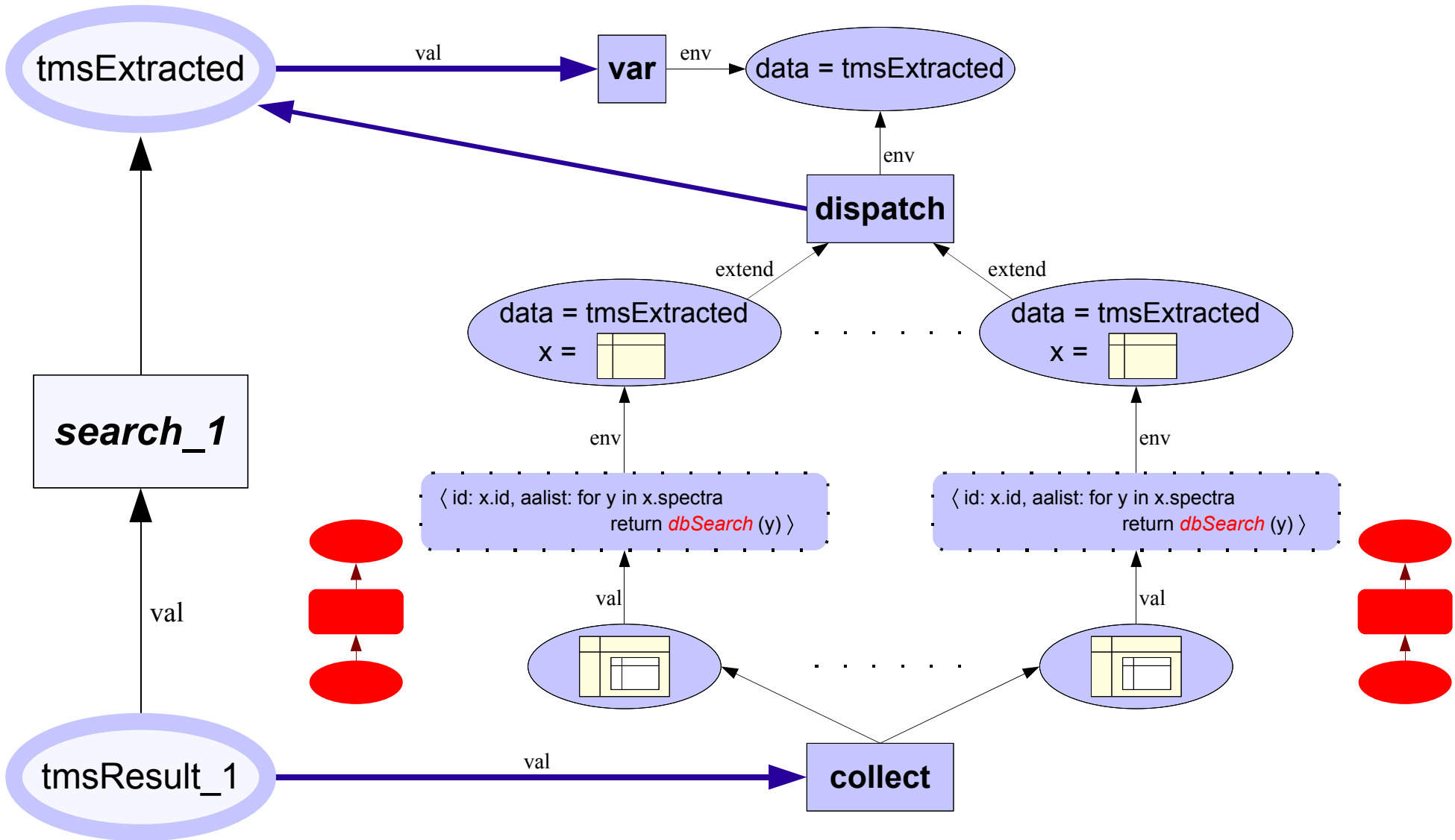
dataflow *search* (data: TMSextracted): Aalist is  
 for x in data return  
 < id: x.id, aalist: for y in x.spectra return *dbSearch* (y) >



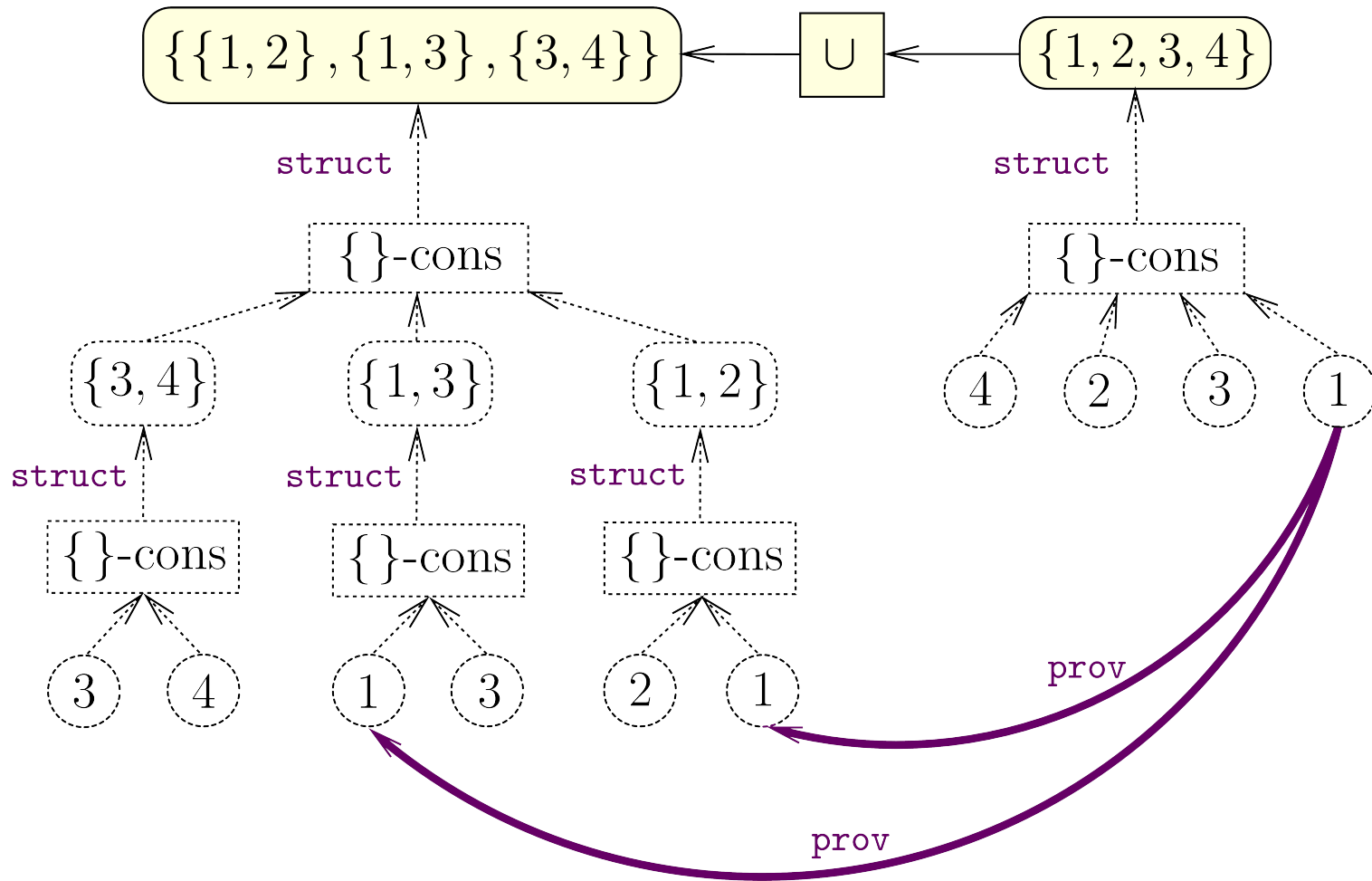


# Accounts

dataflow *search* (data: TMSextracted): Aalist is  
 for x in data return  
 < id: x.id, aalist: for y in x.spectra return *dbSearch* (y) >



# Subvalue Provenance



# Conclusions and Future Work

- Mapping provides validation for both models
  - NRC Dataflow Model
  - Open Provenance Model
- OPM graphs provide a basis for visualisation of
  - NRC dataflow executions
  - subvalue provenance
- Which OPM graphs can be mapped to NRC runs?
- Actual implementation of the mapping
  - **nrcflow** prototype (design stage)