

---

# The Print Model

---

Disciplina de Procedência de Dados e  
Data Warehousing

Profa. Dra. Cristina Dutra de Aguiar Ciferri  
[cdac@icmc.usp.br](mailto:cdac@icmc.usp.br)

---

# Outline

- Motivation
- The PrInt Model
  - Theoretical basis
  - Architecture
  - Operations

# Research Issue

## ■ Motivation

- there are a crescent number of integration applications in which **updates** on heterogeneous **sources** are **not allowed**

How to avoid **decision retaking** by **tracking** actions about integration decisions and automatically **reapplying** them in subsequent integration processes?

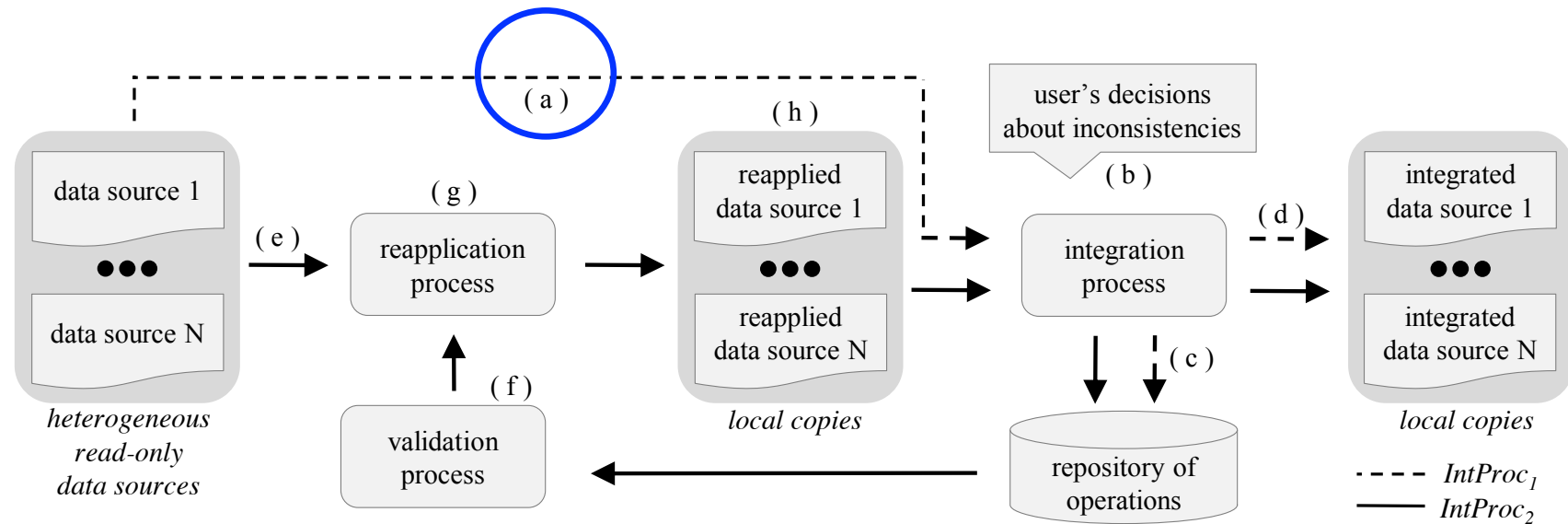
# The PrInt Model

Provenance model to support integration processes

## ■ Theoretical basis

- ❑ it incorporates **data provenance** to improve the integration process
  - set of metadata that allows for the identification of sources and transformations applied to data
- ❑ it focuses on **instance** level integration
- ❑ it is applied to integration processes where the sources are **read-only**
- ❑ it provides a **strict reproduction** of decisions among distinct integration processes

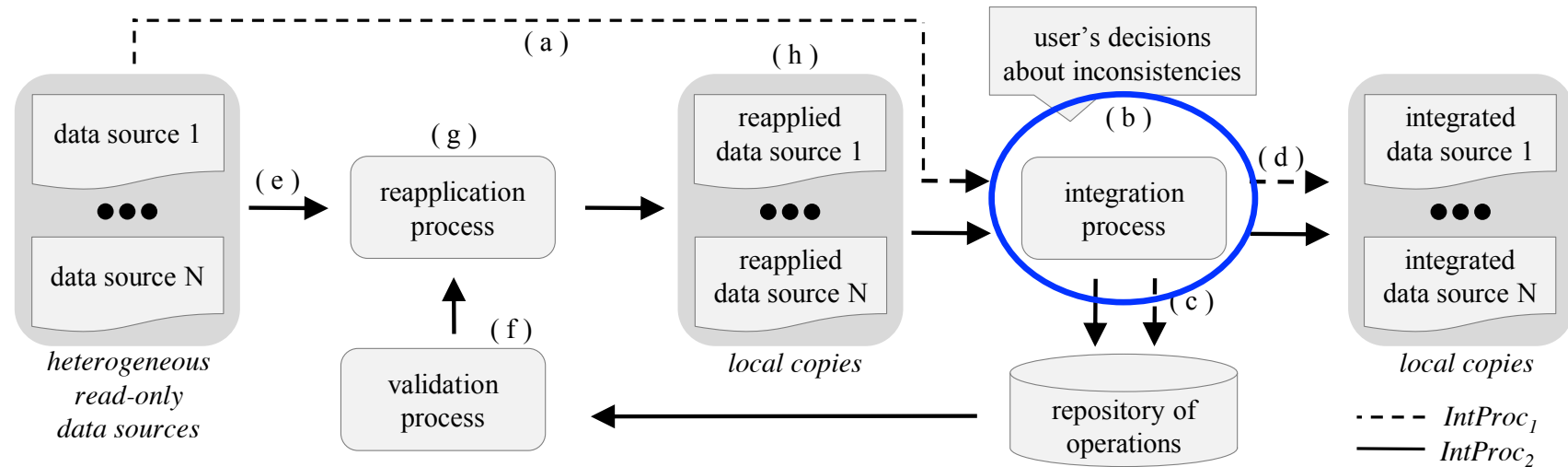
# Architecture



## ■ Integration scenario

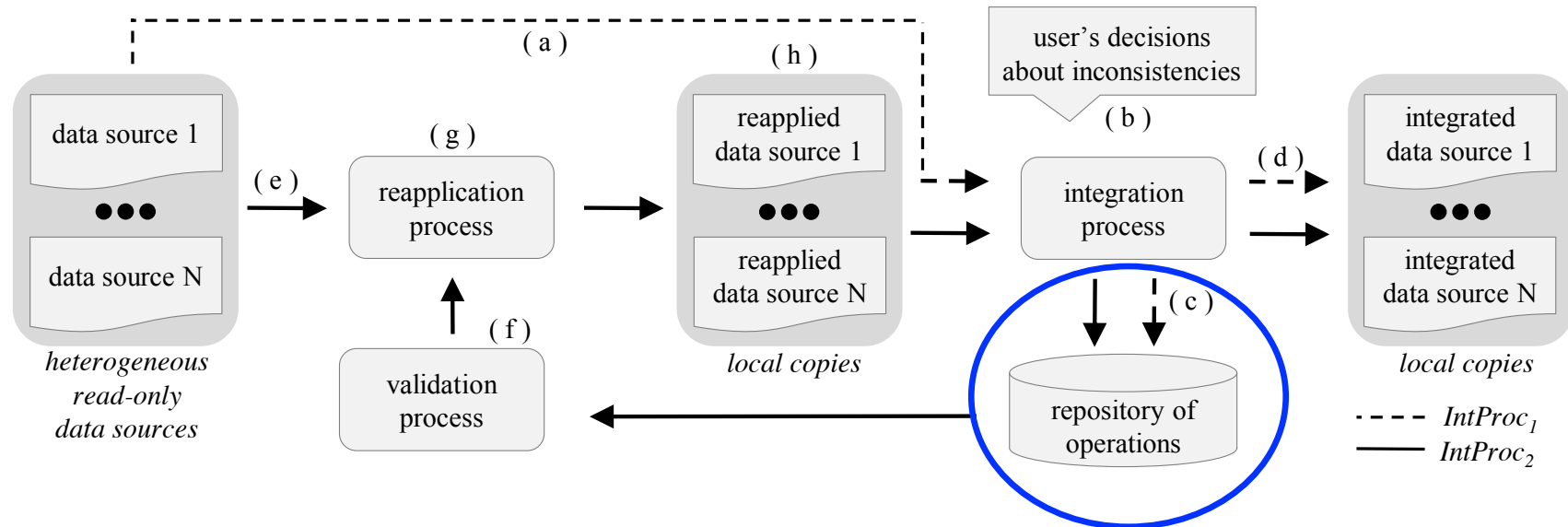
- a) a first integration process, *IntProc<sub>1</sub>*, is executed by receiving as input several sources

# Architecture



- Integration scenario
  - b) inconsistencies are solved

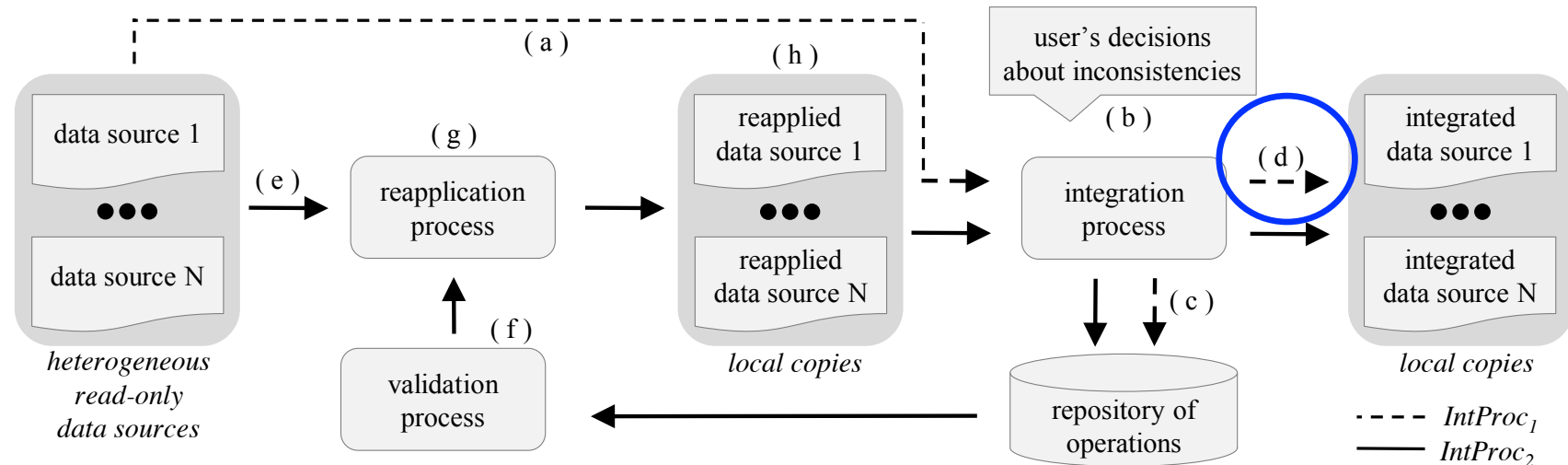
# Architecture



## ■ Integration scenario

- c) decisions are stored as operations in a repository

# Architecture

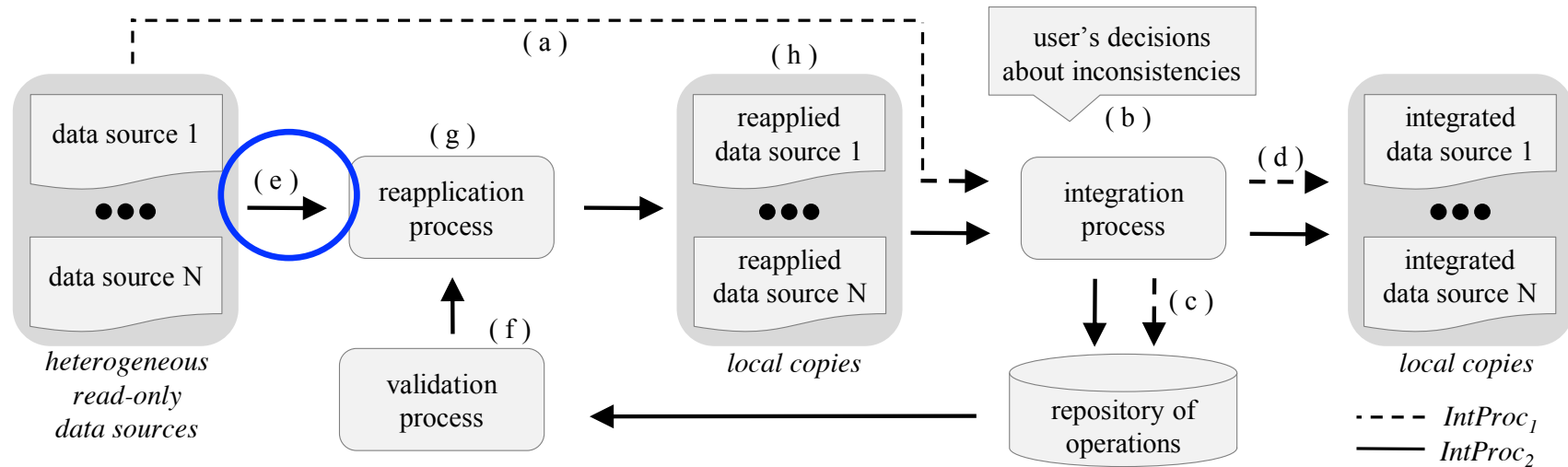


## ■ Integration scenario

- d) consistent copies of sources are locally stored as read-only copies



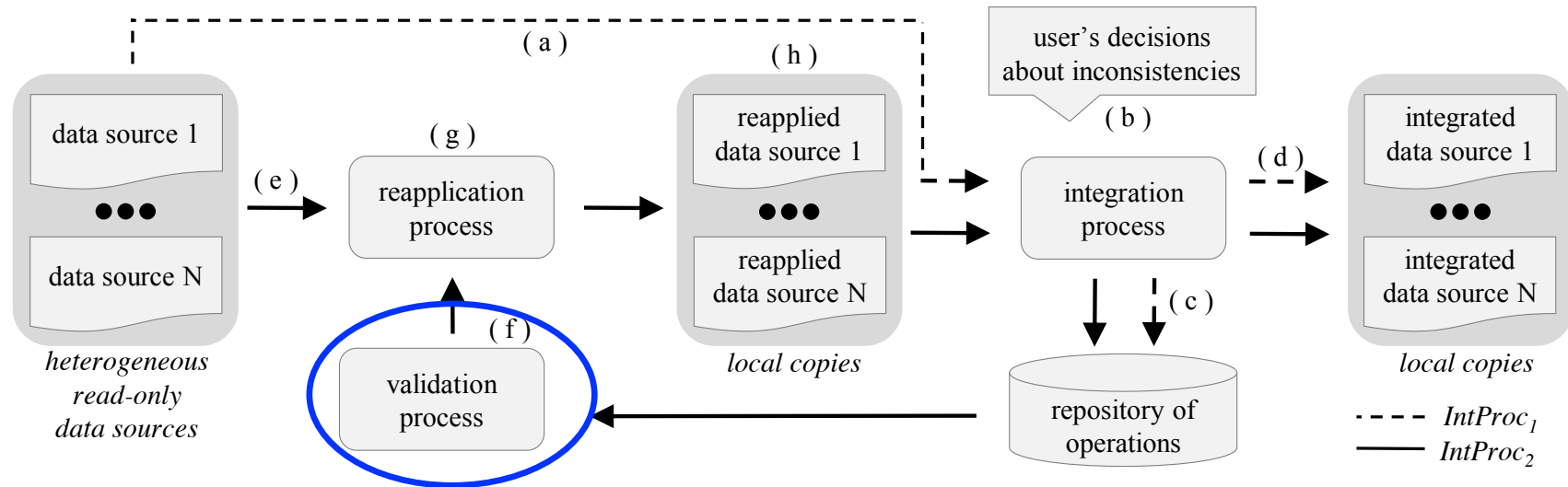
# Architecture



## ■ Integration scenario

e) a second integration process, **IntProc<sub>2</sub>**, is started

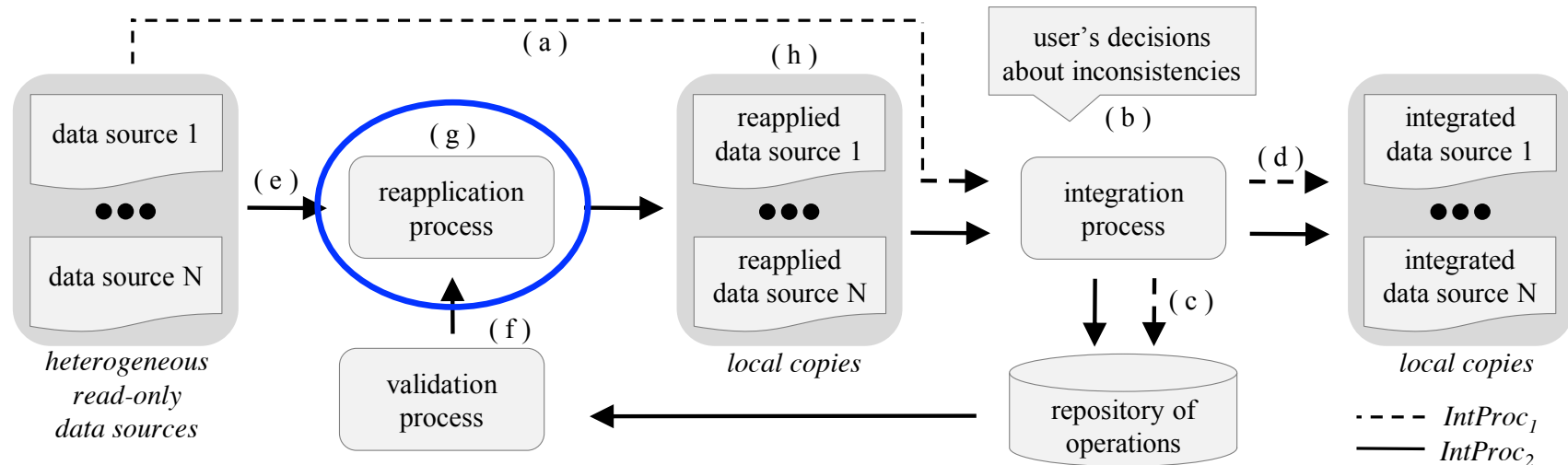
# Architecture



## ■ Integration scenario

- f) before reapplying the operations, PrInt **validates** the repository of operations

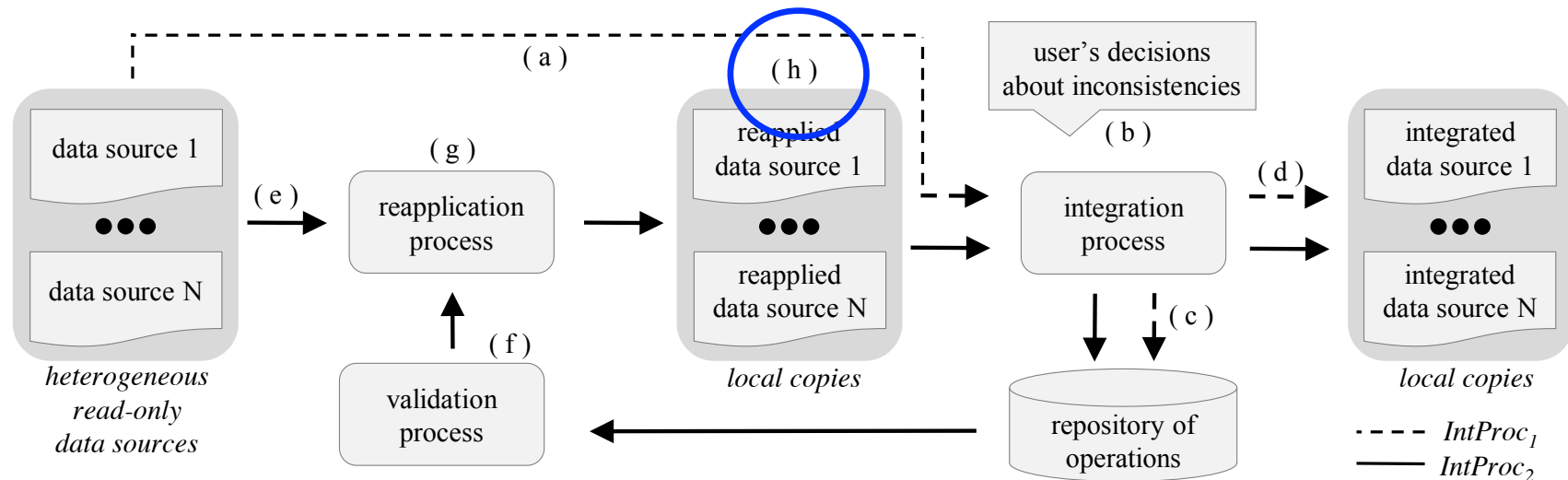
# Architecture



## ■ Integration scenario

- g) Print reapplies all valid operations, thus reproducing integration decisions taken in IntProc<sub>1</sub>

# Architecture



## ■ Integration scenario

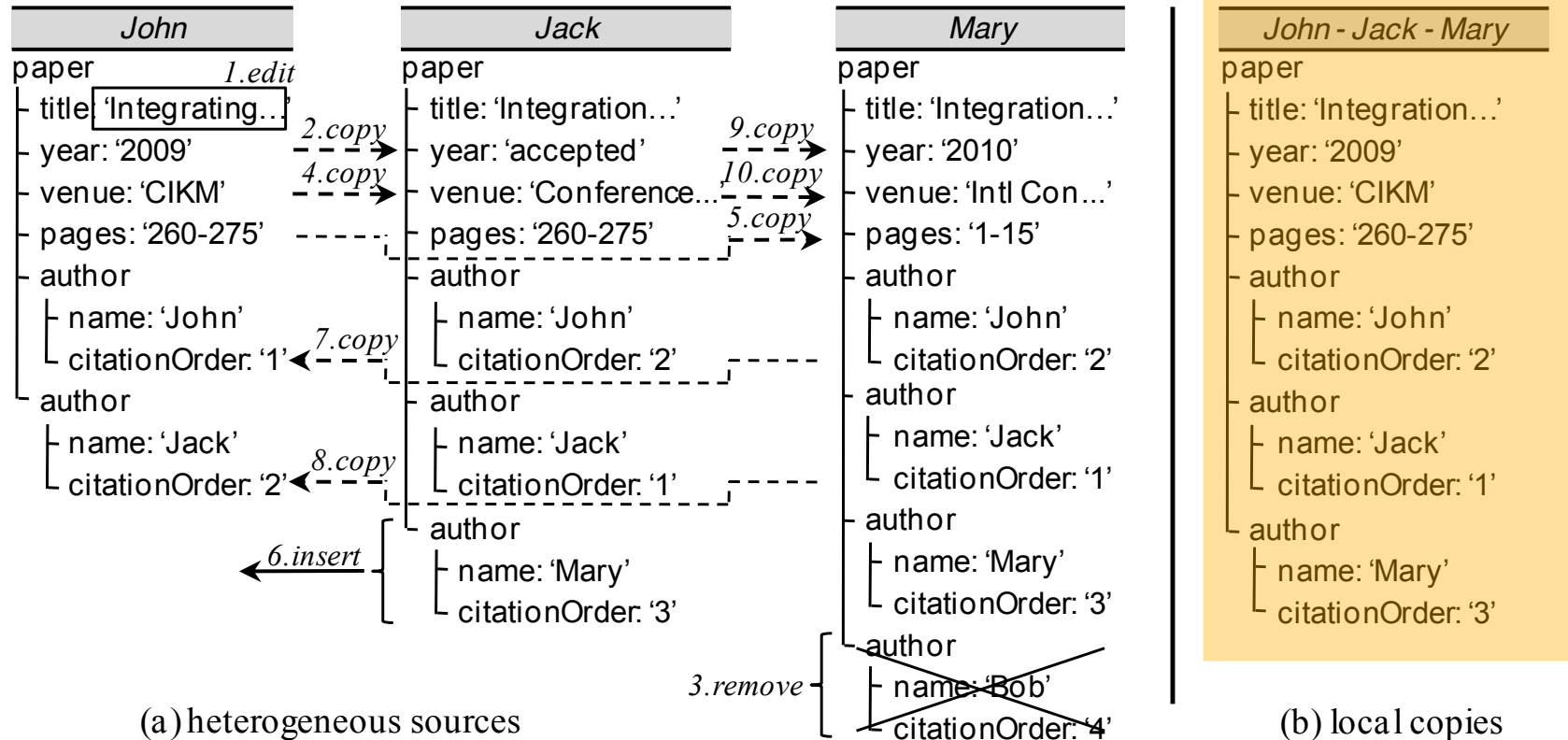
- h) intervention for solving inconsistencies in  $IntProc_2$  are limited to those new inconsistencies originated from source updates since  $IntProc_1$

---

# Main Characteristics

- Storage of integration decisions in a repository
  - based on operations of **edit**, **copy**, **insert** and **remove**
  - management of transitive and overlapping operations
- **Validation** and **reapplication** of operations
  - the VRT method

# Example of Data Integration

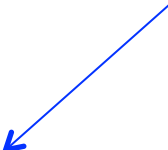


# Repository of Operations

id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

an integer that identifies an operation

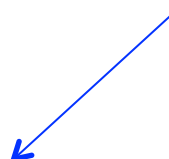


id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...



# Repository of Operations


the type of the operation



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

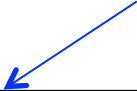
source that provides the value



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

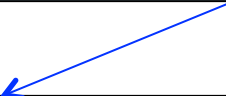
source that is the target of the operation



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

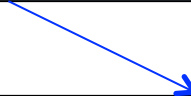
values of key attributes of the object



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

attribute name involved in the operation



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Repository of Operations

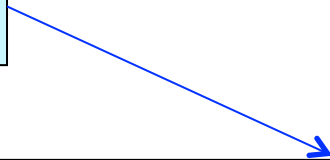
origin's attribute value



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

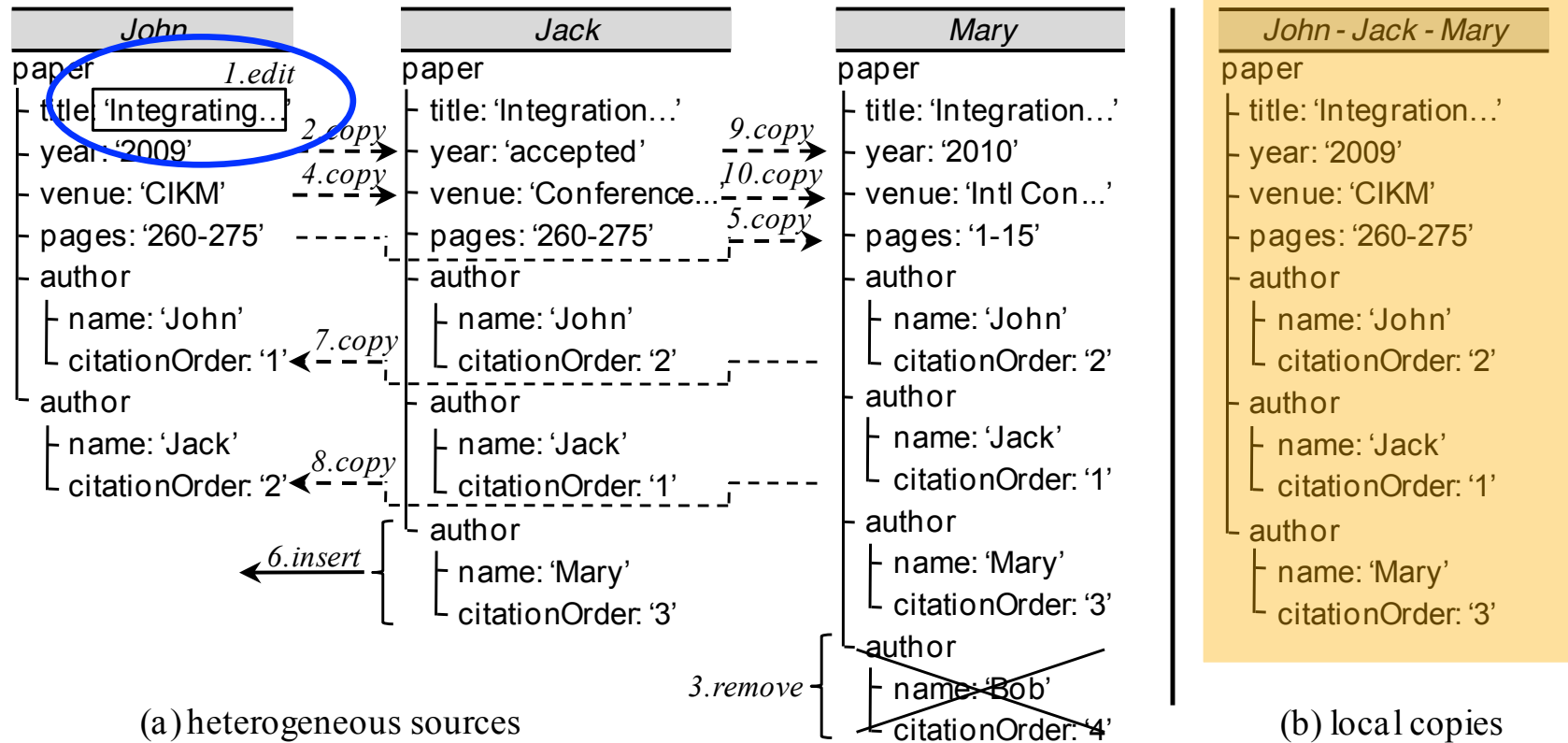
# Repository of Operations

original target's attribute value



id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

# Example of Data Integration



the value of paper's title for John is **edited** from 'Integrating...' to 'Integration...'

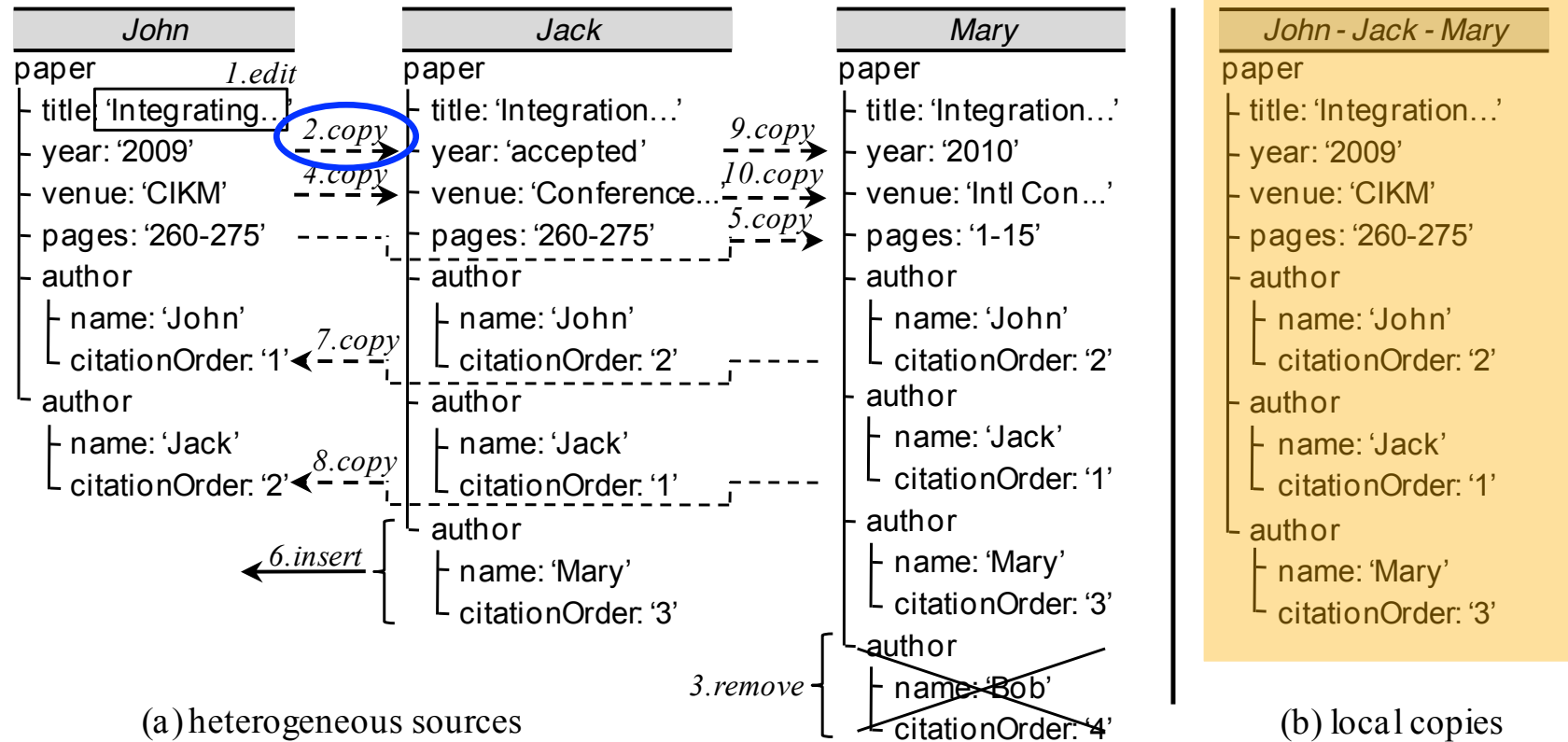


# Well-defined Edit Operation

id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

- ❑ *originValue* is the value obtained from input
- ❑ *origin* is set to null to indicate an external action
- ❑ *originValue* and *targetValue* should have different values

# Example of Data Integration



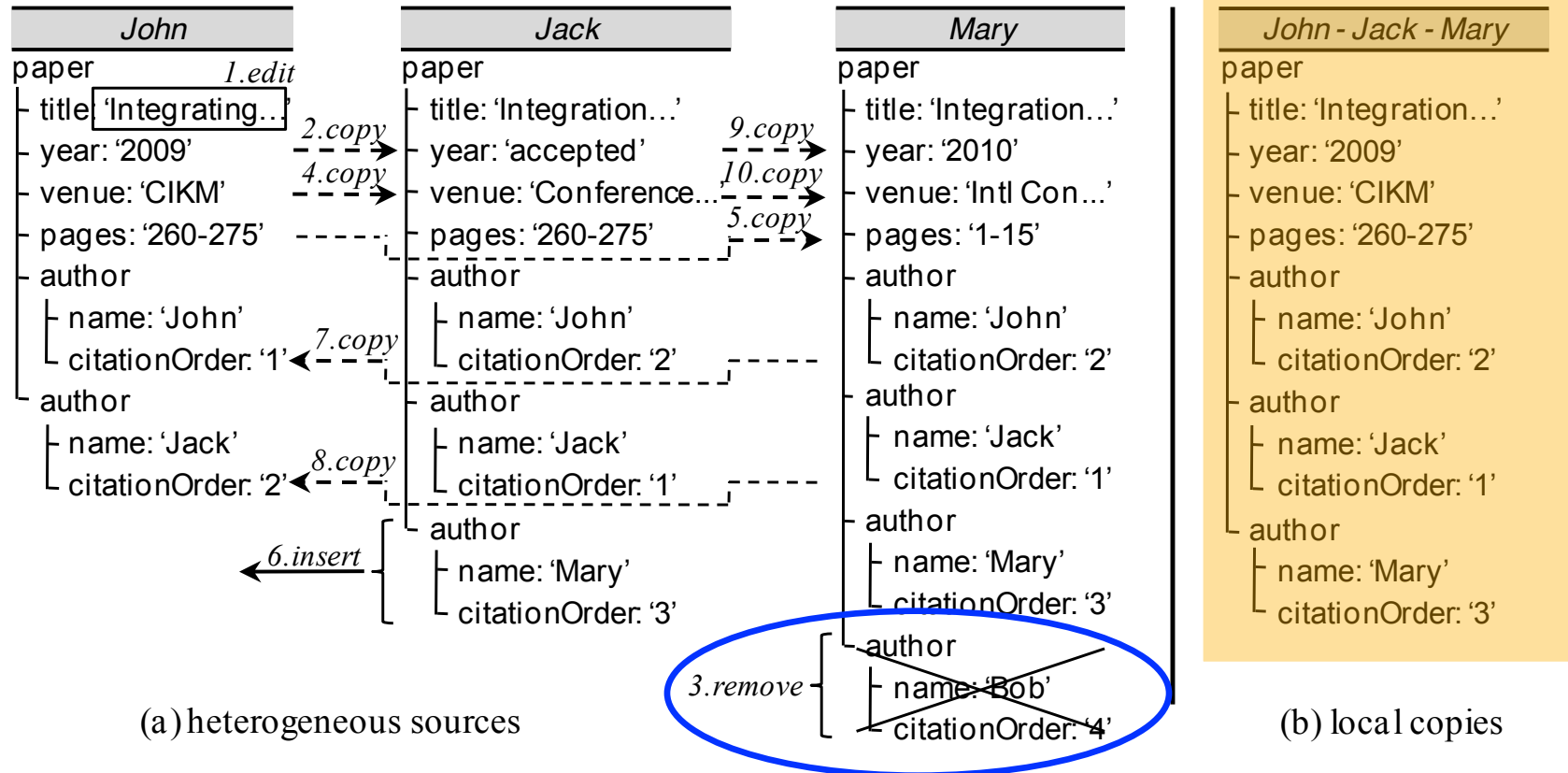
paper's year is copied from John to Jack

# Well-defined Copy Operation

id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

- ❑ *originValue* and *targetValue* should have different values

# Example of Data Integration



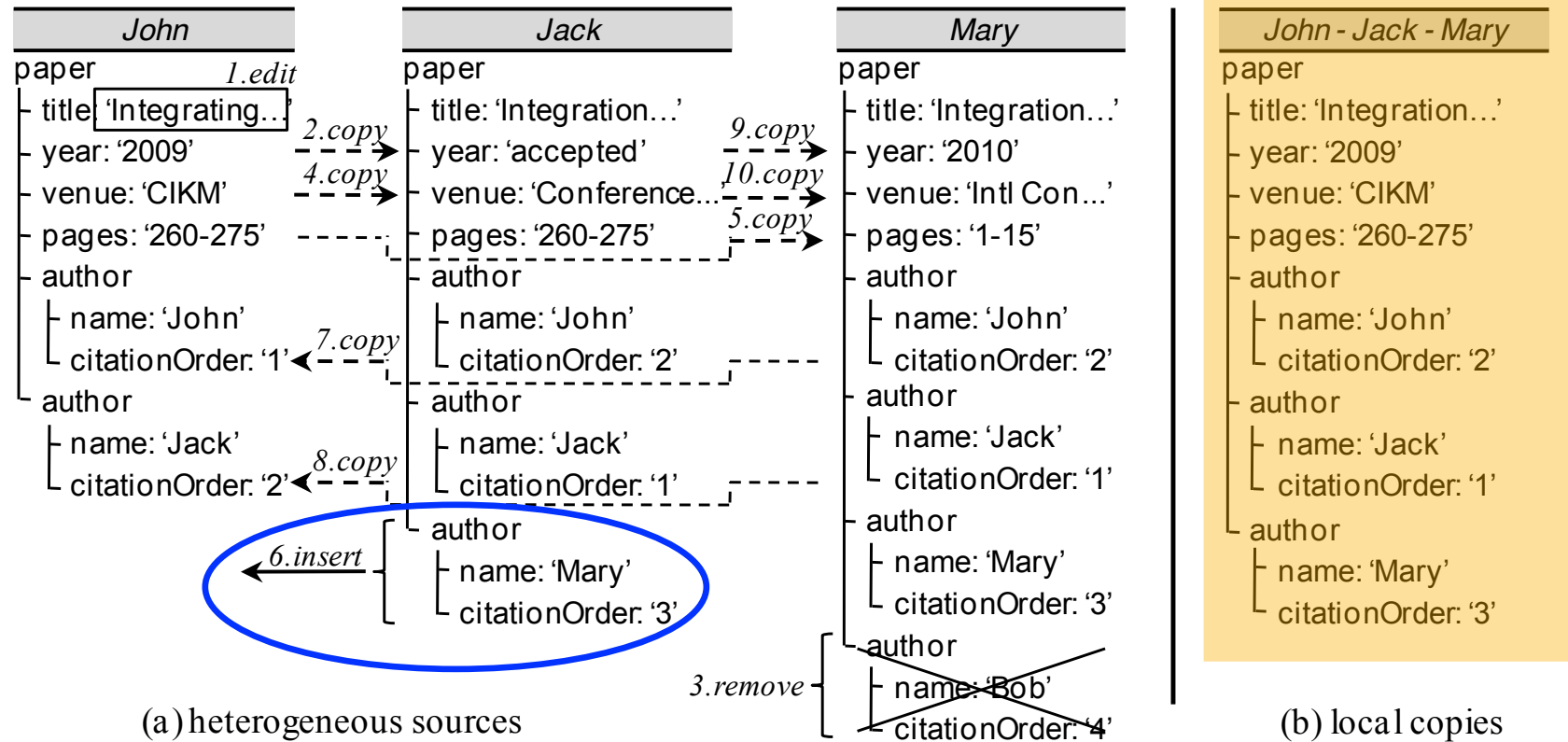
author 'Bob' is removed from Mary

# Well-defined Remove Operation

id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

- mapped to a remove operation and several edit operations
- *origin*, *objAtt*, *originValue* and *targetValue* are set to null

# Example of Data Integration



author 'Mary' from Jack is **inserted** into John

# Well-defined Insert Operation

id	op	origin	target	objKey	objAtt	originValue	targetValue
1	ed	null	John	title=Integrating...	title	Integration...	Integrating...
2	cp	John	Jack	title=Integration...	year	2009	accepted
3	ed	null	Mary	title=Integration.../name=Bob	citationOrder	null	4
4	rm	null	Mary	title=Integration.../name=Bob	null	null	null
5	cp	John	Jack	title=Integration...	venue	Conference...	CIKM
6	cp	John	Mary	title=Integration...	pages	260-275	1-15
7	in	Jack	John	title=Integration.../name=Mary	null	null	null
8	cp	Jack	John	title=Integration.../name=Mary	citationOrder	3	null
9	cp	Mary	John	title=Integration.../name=John	citationOrder	2	1
10	cp	Mary	John	title=Integration.../name=Jack	citationOrder	1	2
11	cp	Jack	Mary	title=Integration...	year	2009	2010
12	cp	Jack	Mary	title=Integration...	venue	Conference...	Intl Con...

- mapped to an insert operation and several copy operations
- *objAtt*, *originValue* and *targetValue* are set to null