Development of Robust Traceability Benchmarks

Xiaofan Chen
John Hosking
John Grundy
Robert Amor

THE UNIVERSITY OF AUCKLAND NEW ZEALAND Te Whare Wānanga o Tāmaki Makaurau
SWINBURNE UNIVERSITY OF TECHNOLOGY
Australian National University

Traceability links

- Inter-relationships between artifacts

Links	Correct	Incorrect
Captured	True positive (True links)	False positive (Incorrect links)
Fail to be captured	False negative (Missing links)	True negative

What is a Traceability Benchmark?

- A standard test or set of tests employed to compare the performance of traceability recovery techniques [1].
- Four components
  - **Dataset**
  - **Tasks**
  - **Answer sets**
  - **Measures**

Barriers

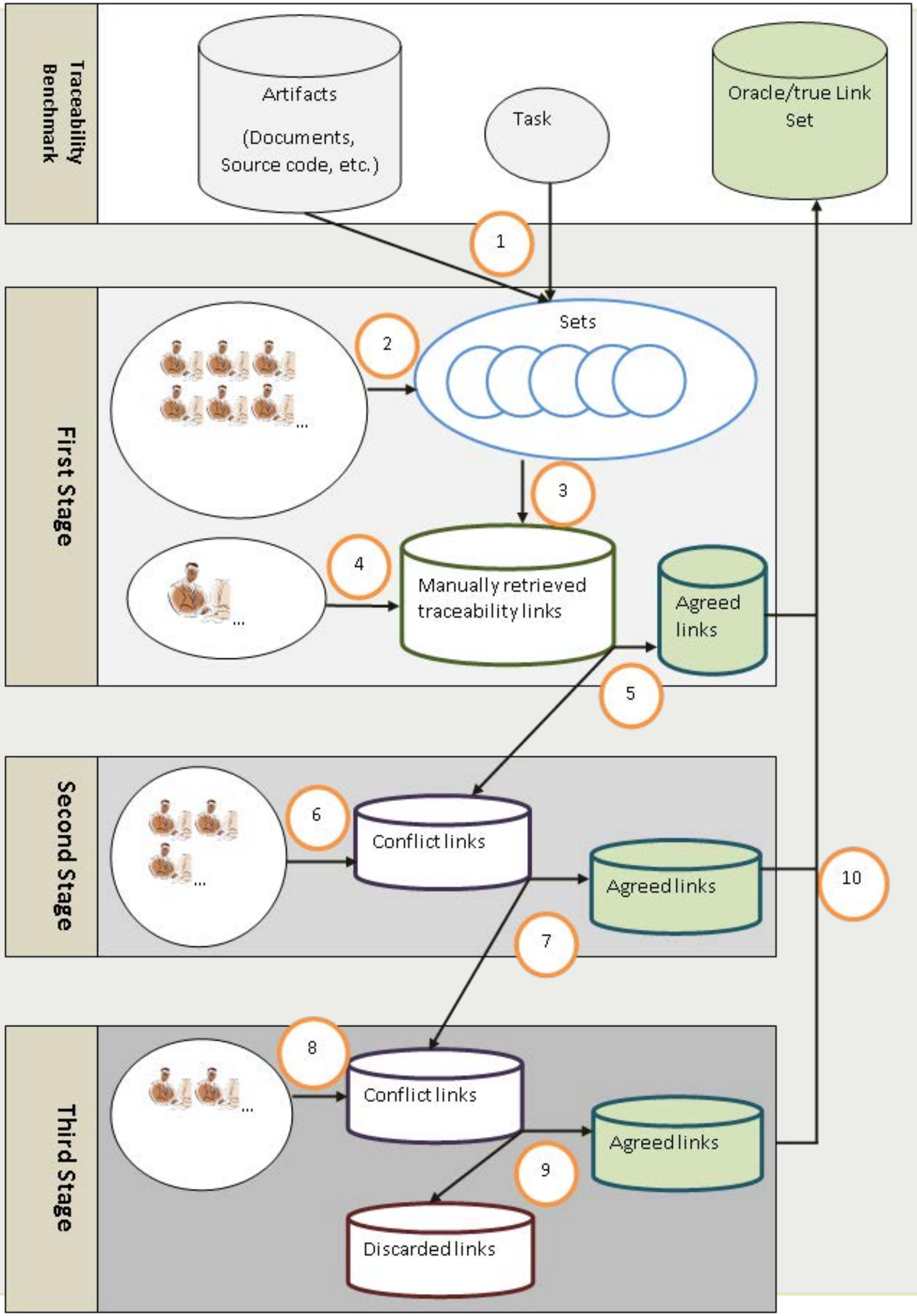
- The lack of publicly available benchmarks
- The diversity of traceability issues
- The difficulty of manually building benchmarks

Issues

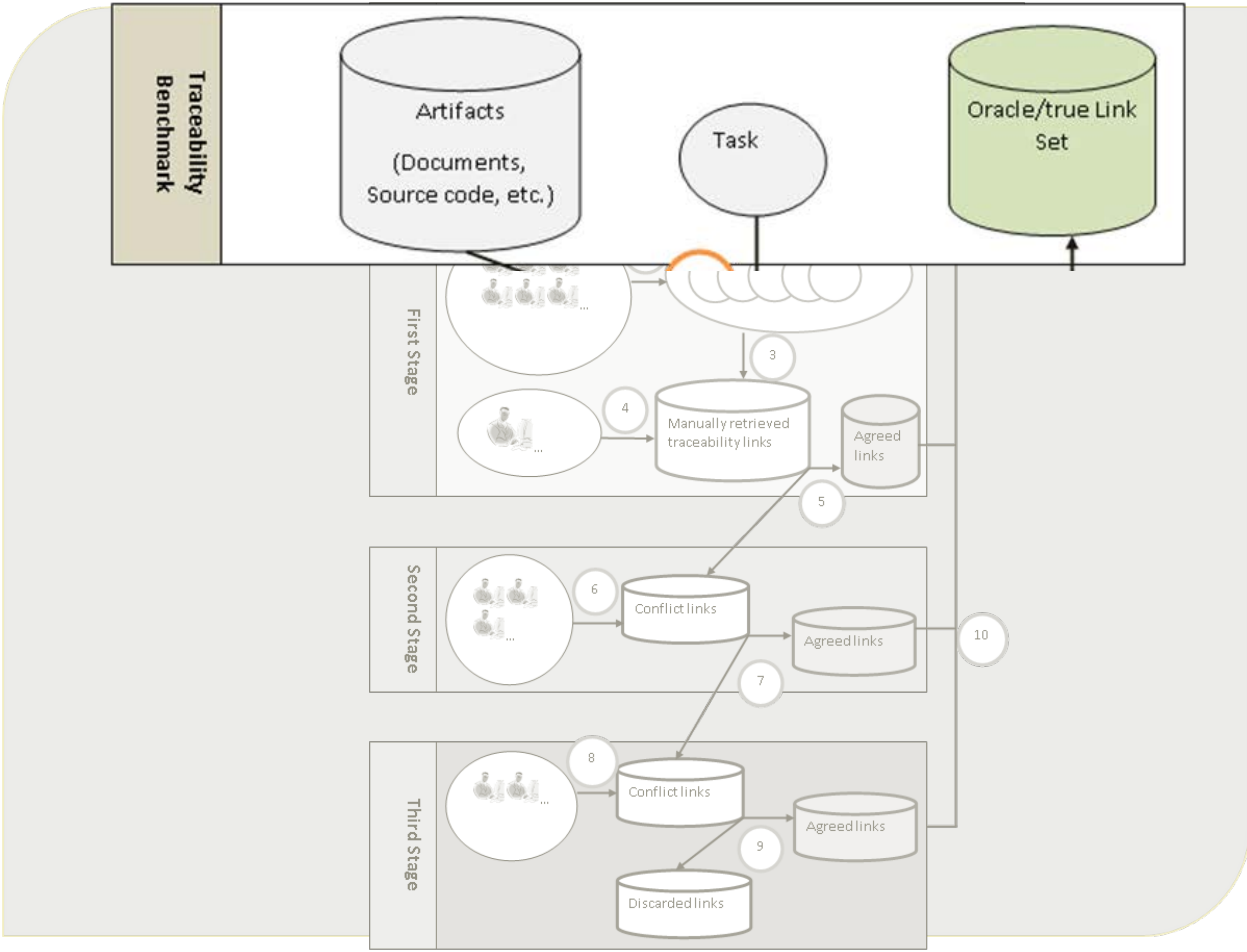
- Three key issues:
  - how to find an appropriate dataset
  - how to manually identify correct links
  - how to verify links are correct or not.
- No guidelines to assist researchers in developing traceability benchmarks.

Our approach

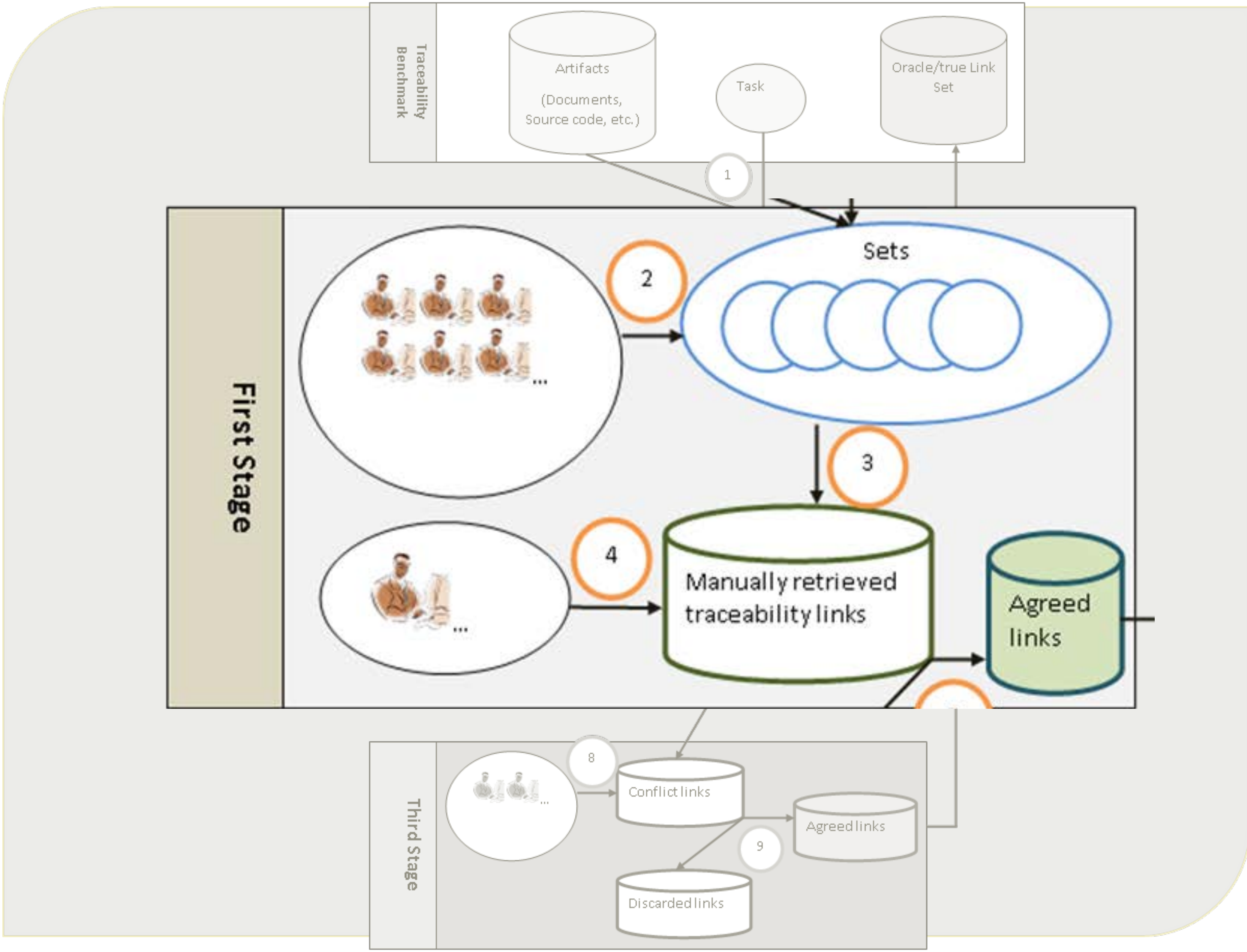
- Five steps to build a benchmark
  - Task identification
  - Artifact selection
  - Project selection
  - True link set development
  - Evaluation metrics
    - Precision, recall, F-measure



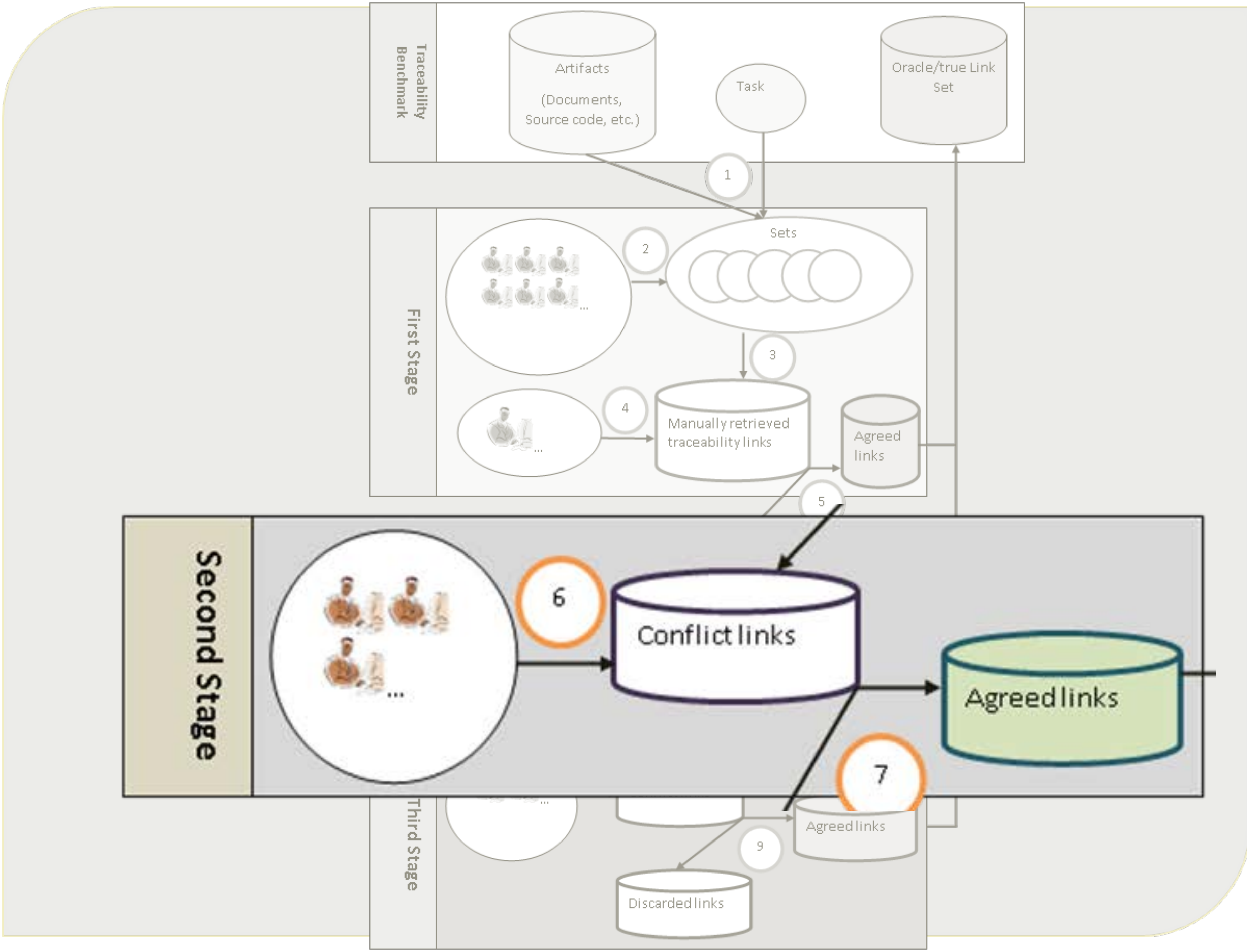
Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
1
First Stage
2
Sets
3
4
Manually retrieved traceability links
Agreed links
5
Second Stage
6
Conflict links
Agreed links
10
7
Third Stage
8
Conflict links
Agreed links
9
Discarded links

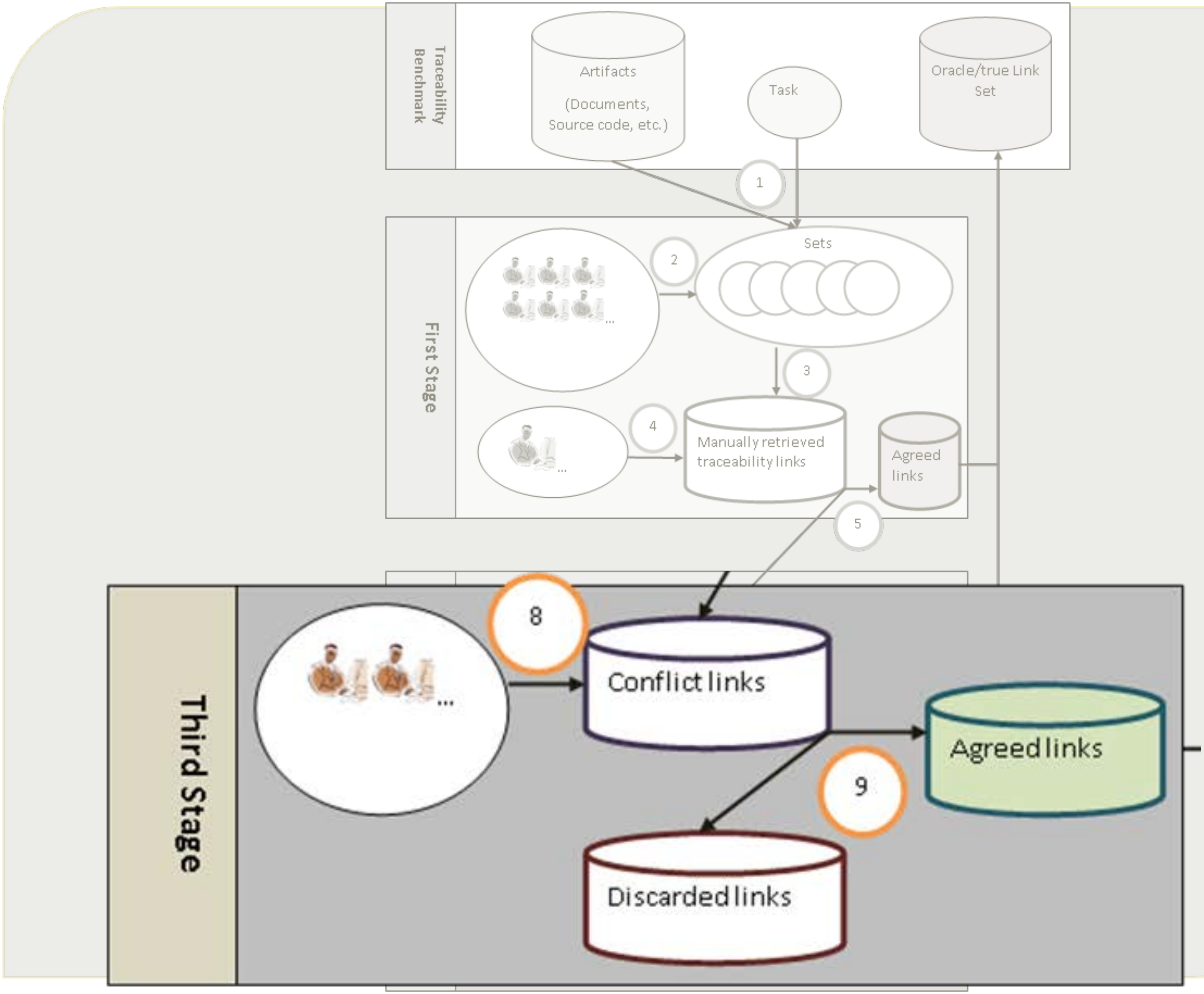


Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
First Stage
3
4
Manually retrieved traceability links
Agreed links
5
Second Stage
6
Conflict links
Agreed links
10
7
Third Stage
8
Conflict links
Agreed links
9
Discarded links

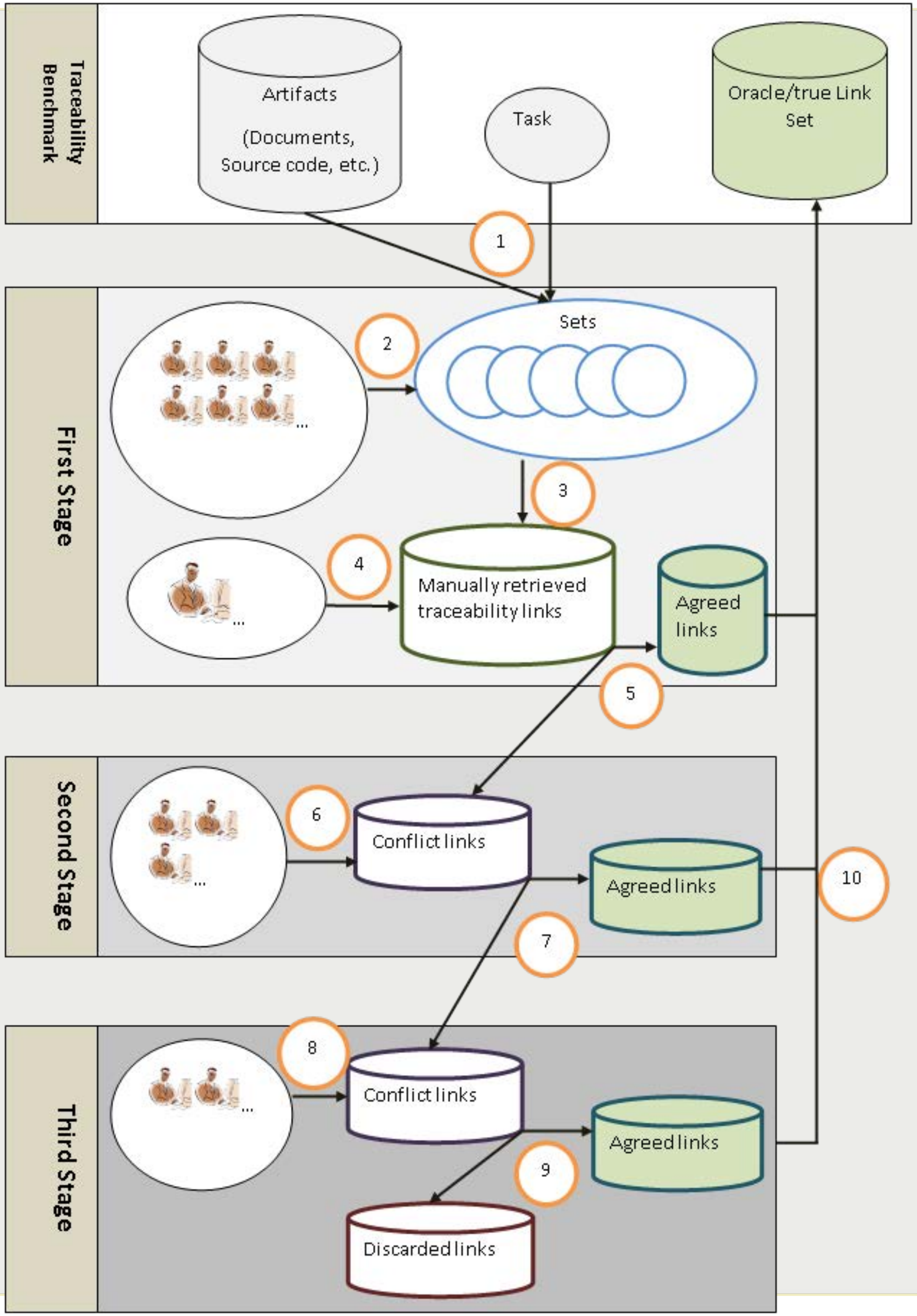


Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
1
First Stage
2
Sets
3
4
Manually retrieved traceability links
Agreed links
Third Stage
8
Conflict links
Agreed links
9
Discarded links


Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
1
First Stage
Sets
2
3
4
Manually retrieved traceability links
Agreed links
5
Second Stage
6
Conflict links
Agreed links
7
Third Stage
Agreed links
9
Discarded links



Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
1
First Stage
Sets
2
3
Manually retrieved traceability links
4
Agreed links
5
Third Stage
8
Conflict links
Agreed links
9
Discarded links



Traceability Benchmark
Artifacts
(Documents, Source code, etc.)
Task
Oracle/true Link Set
1
First Stage
2
Sets
3
4
Manually retrieved traceability links
Agreed links
5
Second Stage
6
Conflict links
Agreed links
10
7
Third Stage
8
Conflict links
Agreed links
9
Discarded links

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

$$\Pr[e_i] = \frac{\sum_{j=1}^{m_i} \Pr(x_j)}{m_i} \times \Pr(y_i)$$

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

$$\Pr[e_i] = \frac{\sum_{j=1}^{m_i} \Pr(x_j)}{m_i} \times \Pr(y_i)$$

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

$$\Pr[e_i] = \frac{\sum_{j=1}^{m_i} \Pr(x_j)}{m_i} \times \Pr(y_i)$$

Probability of errors

- Assumptions:
  - Errors are dependent on the type of participant
  - Links are independent
  - Errors made on links are independent
- Formula:

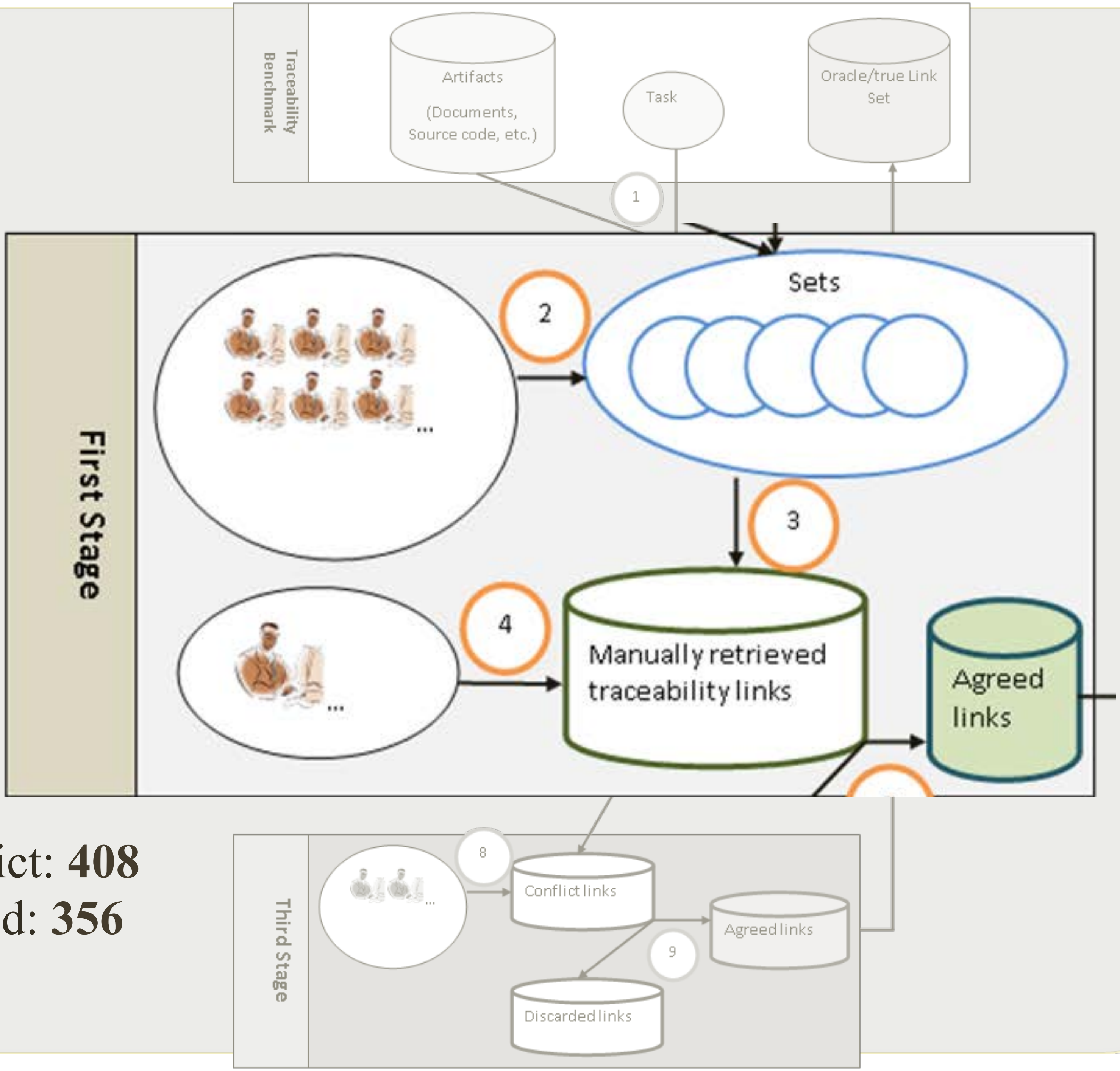
$$\Pr[E] = \frac{\sum_{i=1}^{k} (n_i \times \Pr[e_i])}{N}$$

$$\Pr[e_i] = \frac{\sum_{j=1}^{m_i} \Pr(x_j)}{m_i} \times \Pr(y_i)$$

$$P(x) = \frac{n!}{(n-x)!x!} \times p^x \times q^{n-x}$$

Case study

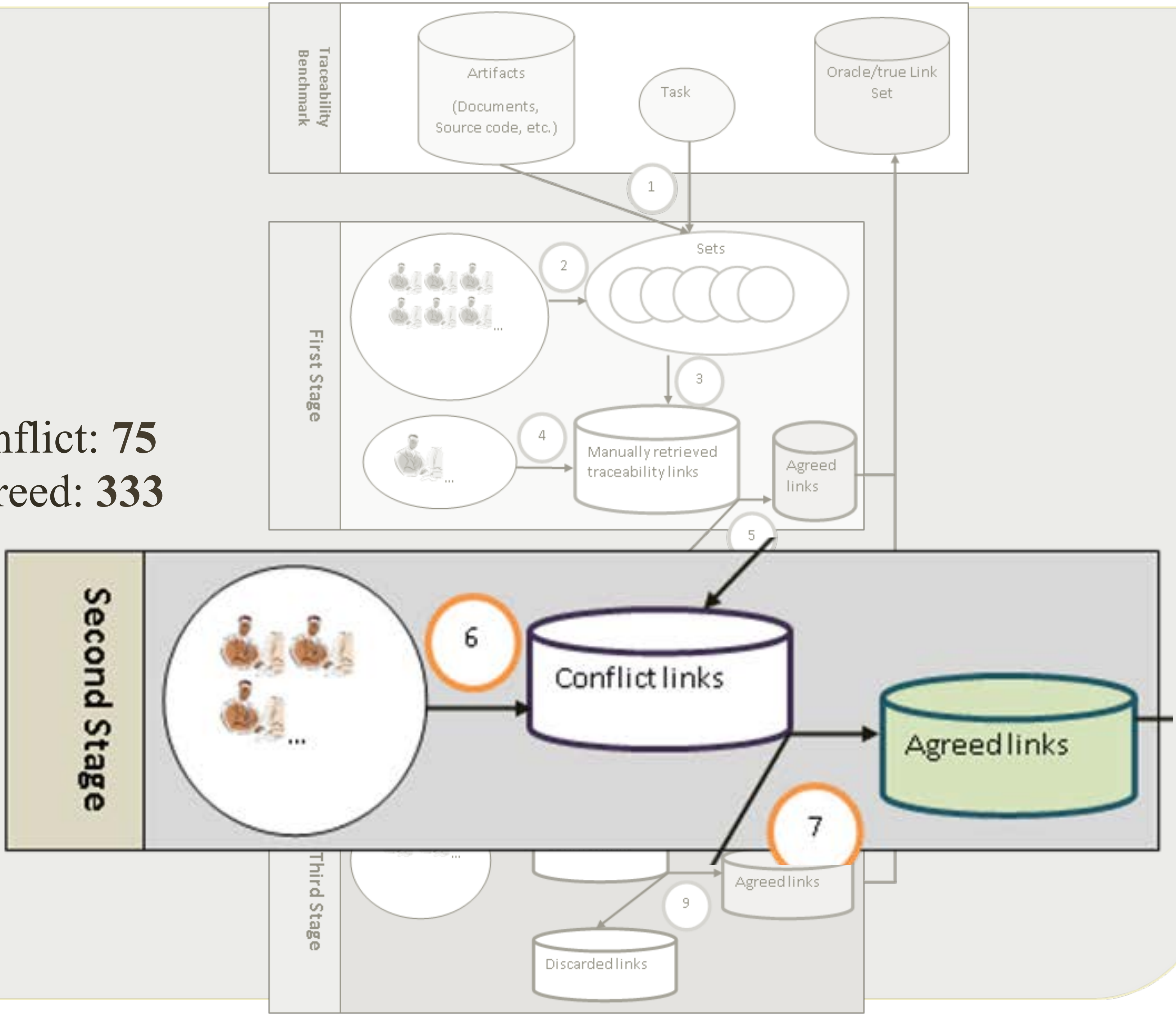
JDK 1.5		**#classes / sections**
Java packages	java.awt, javax.naming, and javax.print packages	249
PDF files	***JPS_PDF.pdf***: Java™ Print Service API User Guide	68
	dnd1.pdf: Drag and Drop subsystem for the Java Foundation Classes	41
	jndispi.pdf: Java Naming and Directory Interface™ Service Provider Interface(JNDI SPI)	73
	Total sections:	182



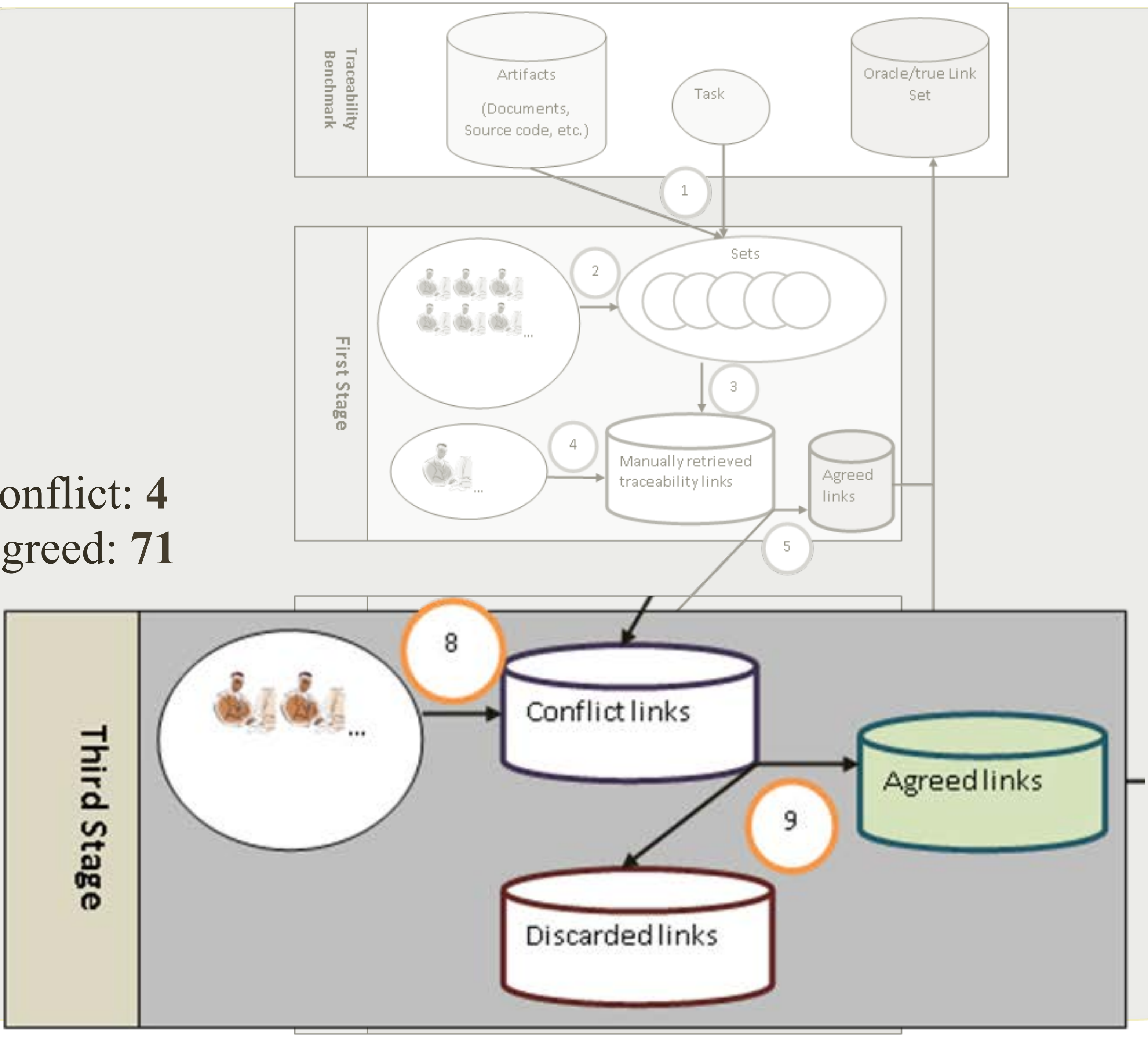
Conflict: **408**

Agreed: **356**

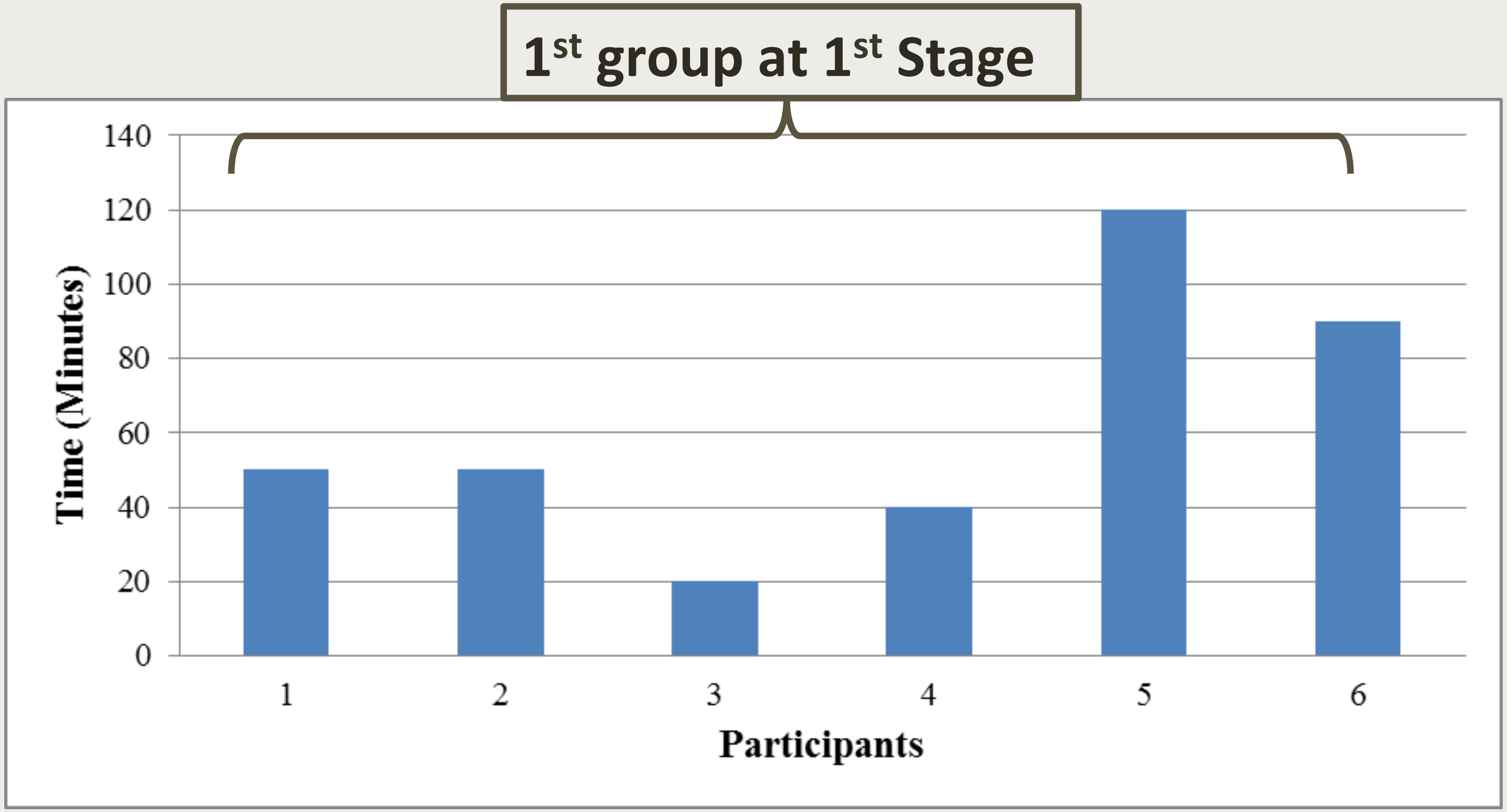
Conflict: **75**
Agreed: **333**



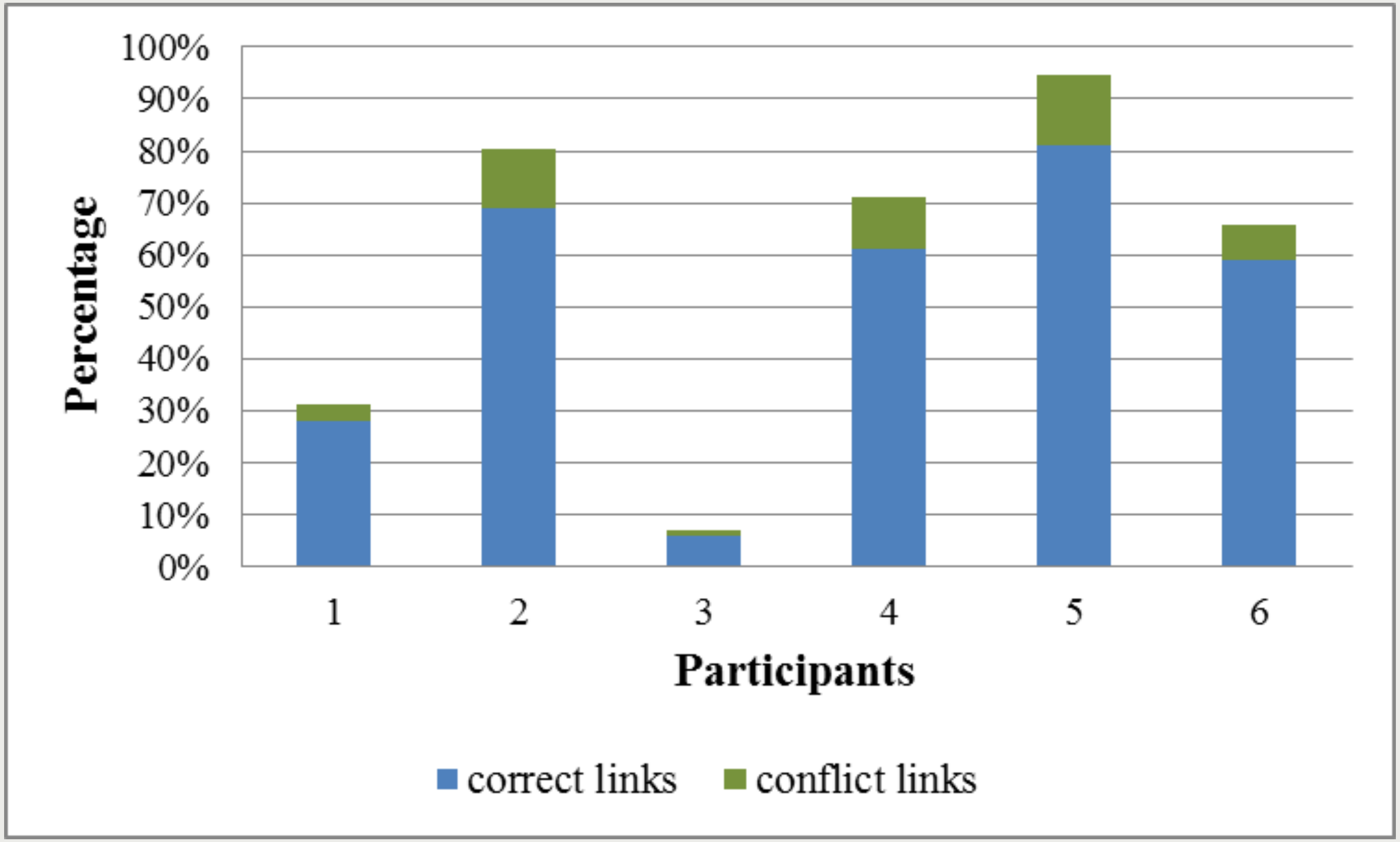
Conflict: **4**
Agreed: **71**



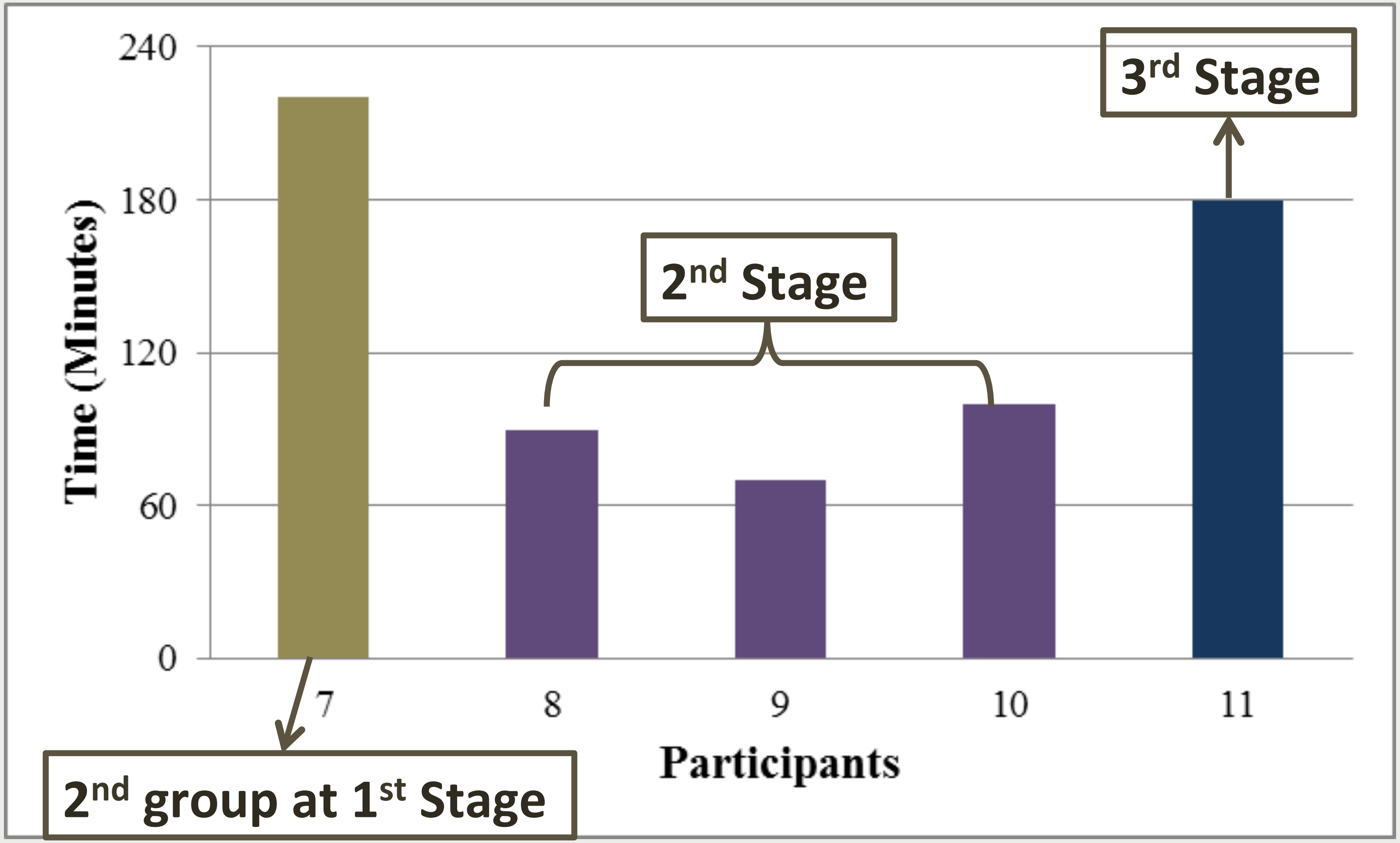
Time taken by the first six participants at the first stage



Links captured by the first six participants



Time taken by the rest of the participants in verifying links



Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links ($n=R-C$)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n$ * 10% (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	5	30	0.2	0.955821
	5	145	21	124	0.2	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
	$\sum_{j=1}^{m_1} \Pr(x_j) \Big/ m_1 = \sum_{j=1}^{7} \Pr(x_j) \Big/ 7 = 0.881222357$					
	$Pr(y_1) = 0.0225$ for $y_1 \geq 2$, where $n=2$, $p=(0.2+0.1)/2=0.15$					
2nd stage	8	272	53	219	0.2	0.9999765
	9	272	123	149	0.2	0.9996381
	10	272	87	185	0.2	0.9998751
	$\sum_{j=1}^{m_2} \Pr(x_j) \Big/ m_2 = \sum_{j=1}^{3} \Pr(x_j) \Big/ 3 = 0.9998299$					
	$Pr(y_2) = 0.04$ for $y_2 \geq 2$, where $n=2$, $p=(0.2+0.2)/2=0.2$					
3rd stage	11	75	4	71	0.05	0.0641971
	$\sum_{j=1}^{m_3} \Pr(x_j) \Big/ m_3 = \sum_{j=1}^{1} \Pr(x_j) \Big/ 1 = 0.0641971$					
	$Pr(y_3) = 0.0266119$ for $y_3 \geq 3$, where $n=5$, $p=(0.2+0.1+0.2+0.2+0.05)/5=0.15$					

Retrieved Links (*R*)	Conflict Links (*C*)	Agreed Links (*n=R-C*)
10	1	9
84	12	72
18	2	16
35	5	30
145	21	124
117	12	105
764	408	356
272	53	219
272	123	149
272	87	185
75	4	71

Stage	Particip	…ample Error …obability for …rticipant (*p*)	*Pr*(x_j) for $x_j \geq n$ * 10% (round to the nearest integer)
1st stage	1	2	0.8657823
	2	2	0.9940038
	3	2	0.8592625
	4	2	0.955821
	5	2	0.9990852
	6	2	0.9969877
	7	1	0.497614
2nd stage	8	2	0.9999765
	9	2	0.9996381
	10	2	0.9998751
3rd stage	11	05	0.0641971

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links ($n=R-C$)	Example Error Probability for Participant (p)	$_i$) for ≥ n * 10% nd to the nearest er)
1st stage	1	10	1	9	0.2	57823
	2	84	12	72	0.2	40038
	3	18	2	16	0.2	92625
	4	35	5	30	0.2	5821
	5	145	21	124	0.2	90852
	6	117	12	105	0.2	69877
	7	764	408	356	0.1	7614
2nd stage	8	272	53	219	0.2	99765
	9	272	123	149	0.2	96381
	10	272	87	185	0.2	98751
3rd stage	11	75	4	71	0.05	41971

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links ($n=R-C$)	Example Probabili... Participa...	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	5	30	0.2	0.955821
	5	145	21	124	0.2	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
2nd stage	8	272	53	219	0.2	0.9999765
	9	272	123	149	0.2	0.9996381
	10	272	87	185	0.2	0.9998751
3rd stage	11	75	4	71	0.05	0.0641971

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links ($n=R-C$)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	5	30	0.2	0.955821
	5	145	21	124	0.2	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
2nd stage						
3rd stage						

$$\sum_{j=1}^{m_1} \Pr(x_j) \Big/ m_1 = \sum_{j=1}^{7} \Pr(x_j) \Big/ 7 = 0.881222357$$

$Pr(y_1) = 0.0225$ for $y_1 \geq 2$, where $n=2$, $p=(0.2+0.1)/2=0.15$

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links ($n=R-C$)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	5	30	0.2	0.955821
	5	145	21	124	0.2	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
2nd stage						
3rd stage						

$$\sum_{j=1}^{m_1} \Pr(x_j) \Big/ m_1 = \sum_{j=1}^{7} \Pr(x_j) \Big/ 7 = 0.881222357$$

$Pr(y_1) = 0.0225$ for $y_1 \geq 2$, where $n=2$, $p=(0.2+0.1)/2=0.15$

$$\Pr[e_i] = \left(\sum_{j=1}^{m_i} \Pr(x_j) \Big/ m_i\right) \times \Pr(y_i)$$

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links (n=R-C)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	5	30	0.2	0.955821
	5	145	21	124	0.2	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
2nd stage	8	272	53	219	0.2	0.9999765
	9	272	123	149	0.2	0.9996381
	10	272	87	185	0.2	0.9998751
3rd stage						

$$\sum_{j=1}^{m_2} \Pr(x_j) \Big/ m_2 = \sum_{j=1}^{3} \Pr(x_j) \Big/ 3 = 0.9998299$$

$Pr(y_2) = 0.04$ for $y_2 \geq 2$, where $n=2$, $p=(0.2+0.2)/2=0.2$

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links (n=R-C)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1^{st} stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	35	[illegible]	[illegible]	[illegible]	0.955821
	5	145	[illegible]	[illegible]	[illegible]	0.9990852
	6	117	12	105	0.2	0.9969877
	7	764	408	356	0.1	0.497614
2^{nd} stage	8	272	53	219	0.2	0.9999765
	9	272	123	149	0.2	0.9996381
	10	272	87	185	0.2	0.9998751
3^{rd} stage						

$$\Pr[e_i] = \left(\sum_{j=1}^{m_i} \Pr(x_j) \Big/ m_i\right) \times \Pr(y_i)$$

$$\sum_{j=1}^{m_2} \Pr(x_j) \Big/ m_2 = \sum_{j=1}^{3} \Pr(x_j) \Big/ 3 = 0.9998299$$

$Pr(y_2) = 0.04$ for $y_2 \geq 2$, where $n=2$, $p=(0.2+0.2)/2=0.2$

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links (n=R-C)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
2nd stage						
3rd stage	11	75	4	71	0.05	0.0641971

$$\sum_{j=1}^{m_3} \Pr(x_j) \Big/ m_3 = \sum_{j=1}^{1} \Pr(x_j) \Big/ 1 = 0.0641971$$

$Pr(y_3)$ = 0.0266119 for $y_3 \geq 3$, where n=5,

p=(0.2+0.1+0.2+0.2+0.05)/5=0.15

Stage	Participant	Retrieved Links (R)	Conflict Links (C)	Agreed Links (n=R-C)	Example Error Probability for Participant (p)	$Pr(x_j)$ for $x_j \geq n * 10\%$ (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
2nd stage						
3rd stage	11	75	4	71	0.05	0.0641971

$$\sum_{j=1}^{m_3} \Pr(x_j) \Big/ m_3 = \sum_{j=1}^{1} \Pr(x_j) \Big/ 1 = 0.0641971$$

$Pr(y_3)$ = 0.0266119 for $y_3 \geq 3$, where n=5,

p=(0.2+0.1+0.2+0.2+0.05)/5=0.15

$$\Pr[e_i] = \left(\sum_{j=1}^{m_i} \Pr(x_j) \Big/ m_i\right) \times \Pr(y_i)$$

$$\Pr[E] = \frac{\Pr[e_1]\times n_1 + \Pr[e_2]\times n_2 + \Pr[e_3]\times n_3}{N}$$

Stage	Participant	Retrieved Links (*R*)	Conflict Links (*C*)	Agreed Links (*n*=*R*-*C*)	Example Error Probability for Participant (*p*)	*Pr*(x_j) for $x_j \geq n$ * 10% (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	[illegible]	[illegible]	[illegible]	[illegible]	.955821
	5	[illegible]	[illegible]	[illegible]	[illegible]	.9990852
	6	[illegible]	[illegible]	[illegible]	[illegible]	.9969877
	7	764	408	356	0.1	0.497614
						$\sum_{j=1}^{m_1} \Pr(x_j) \Big/ m_1 = \sum_{j=1}^{7} \Pr(x_j) \Big/ 7 = 0.881222357$
						Pr(y_1) = 0.0225 for $y_1 \geq 2$, where *n*=2, *p*=(0.2+0.1)/2=0.15
2nd stage	8	272	53	219	0.2	0.9999765
	9	272	123	149	0.2	0.9996381
	10	272	87	185	0.2	0.9998751
						$\sum_{j=1}^{m_2} \Pr(x_j) \Big/ m_2 = \sum_{j=1}^{3} \Pr(x_j) \Big/ 3 = 0.9998299$
						Pr(y_2) = 0.04 for $y_2 \geq 2$, where *n*=2, *p*=(0.2+0.2)/2=0.2
3rd stage	11	75	4	71	0.05	0.0641971
						$\sum_{j=1}^{m_3} \Pr(x_j) \Big/ m_3 = \sum_{j=1}^{1} \Pr(x_j) \Big/ 1 = 0.0641971$
						Pr(y_3) = 0.0266119 for $y_3 \geq 3$, where *n*=5, *p*=(0.2+0.1+0.2+0.2+0.05)/5=0.15

$$\Pr[E] = \frac{\Pr[e_1]\times n_1 + \Pr[e_2]\times n_2 + \Pr[e_3]\times n_3}{N}$$

***Pr[E]*=0.027** for *E*>=*N**10%

- N = 760
- n_1 = 356, n_2 = 333, and n_3 = 71

Stage	Participant	Retrieved Links (*R*)	Conflict Links (*C*)	Agreed Links (*n*=*R*-*C*)	Example Error Probability for Participant (*p*)	$Pr(x_j)$ for $x_j \geq n$ * 10% (round to the nearest integer)
1st stage	1	10	1	9	0.2	0.8657823
	2	84	12	72	0.2	0.9940038
	3	18	2	16	0.2	0.8592625
	4	[illegible]	[illegible]	[illegible]	[illegible]	.955821
	5	[illegible]	[illegible]	[illegible]	[illegible]	.9990852
	6	[illegible]	[illegible]	[illegible]	[illegible]	.9969877
	7	764	408	356	0.1	0.497614
	$\sum_{j=1}^{m_1} \Pr(x_j) \Big/ m_1 = \sum_{j=1}^{7} \Pr(x_j) \Big/ 7 = 0.881222357$					
	[illegible] *p*=(0.2+0.1)/2=0.15					
2nd stage	8	[illegible]	[illegible]	[illegible]	[illegible]	0.9999765
	9	[illegible]	[illegible]	[illegible]	[illegible]	0.9996381
	10	[illegible]	[illegible]	[illegible]	[illegible]	0.9998751
	[illegible] $(x_j) \Big/ 3 = 0.9998299$					
	Pr(y₂) = 0.04 for *y₂*≥2, where *n*=2, *p*=(0.2+0.2)/2=0.2					
3rd stage	11	75	4	71	0.05	0.0641971
	$\sum_{j=1}^{m_3} \Pr(x_j) \Big/ m_3 = \sum_{j=1}^{1} \Pr(x_j) \Big/ 1 = 0.0641971$					
	$Pr(y_3) = 0.0266119$ for $y_3 \geq 3$, where *n*=5, *p*=(0.2+0.1+0.2+0.2+0.05)/5=0.15					

Actual error rates for participants

Stage	Participant	Retrieved Links (*R*)	Incorrect Links (*W*)	Actual Error Probability (*p*=*W*/*R*)
1st stage	1	10	1	0.1
	2	84	5	0.05952
	3	18	1	0.05556
	4	35	3	0.08571
	5	145	6	0.04138
	6	117	6	0.05128
	7	***764***	***31***	***0.04058***
2nd stage	8	272	3	0.01103
	9	272	8	0.02941
	10	272	6	0.02206
3rd stage	***11***	***75***	***2***	***0.02667***

$$Pr[E] = 0.0012 \text{ for } E \geq N*5\%$$

Cost-quality tradeoffs

- Workload allocated to each participant
- Number of participants verifying a link
- Knowledge of the traced project of each participant

Problems

- The difficulty of determining whether two elements are related.
- How much workload is suitable for a participant to undertake?
- The difficulty of recruitment.
- The scalability of benchmarks.

Threats to validity

- False positive links may be included.
- Some links may be harder to identify.
- True links may fail to be included.
- The case we used is a small fraction of the JDK1.5 system.
- May show different probability error results for different systems and participants.

Future work

- Extend the JDK1.5 benchmark
- Explore other probability distributions

Summary

- Five steps to build a benchmark
- Rigorous identification and verification strategies
- A formula to compute the probability error
- JDK1.5 benchmark is available from:
  - http://tinyurl.com/7l3ohe4.

References

1) S. E. Sim, S. Easterbrook, and R.C. Holt, "Using benchmarking to advance research: a challenge to software engineering," ICSE 2003, pp. 74-83

2) J. Cleland-Huang, A. Czauderna, A., Dekhtyar, O. Gotel, J. H. Hayes, E. Keenan, G. Leach, J. Maletic, D. Poshyvanyk, Y. Shin, A. Zisman, G. Antoniol, B. Berenbach, A. Egyed, and P. Maeder, "Grand challenges, benchmarks, and TraceLab: developing infrastructure for the software traceability research community," TEFSE 2011, May, Waikiki, USA

3) Chen, X., Hosking, J.G., Grundy, J.C. and Amor, R. Development of robust traceability benchmarks, 2013 Australasian Conference on Software Engineering (ASWEC 2013), Melbourne, Australia, July 2013, IEEE CS Press.

4) Chen, X., Hosking, J.G. and Grundy, J.C. Visualizing Traceability Links between Source Code and Documentation, 2012 IEEE International Symposium on Visual Languages and Human-Centric Computing, Innsbruck, Austria, Sept 30-Oct 4 2012, IEEE CS Press.

5) Chen, X. and Grundy, J.C. Improving Automated Documentation to Code Traceability by Combining Retrieval Techniques, In proceedings of the 26th IEEE/ACM International Conference on Automated Software Engineering, Nov 6-10 2011, IEEE Press.

6) Chen, X., Hosking, J.G. and Grundy, J.C. A Combination Approach for Enhancing Automated Traceability, New Ideas and Emerging Results Track, In Proceedings of the 2011 International Conference on Software Engineering (ICSE2011), Honolulu, Hawaii, USA, May 21-28 2011.