

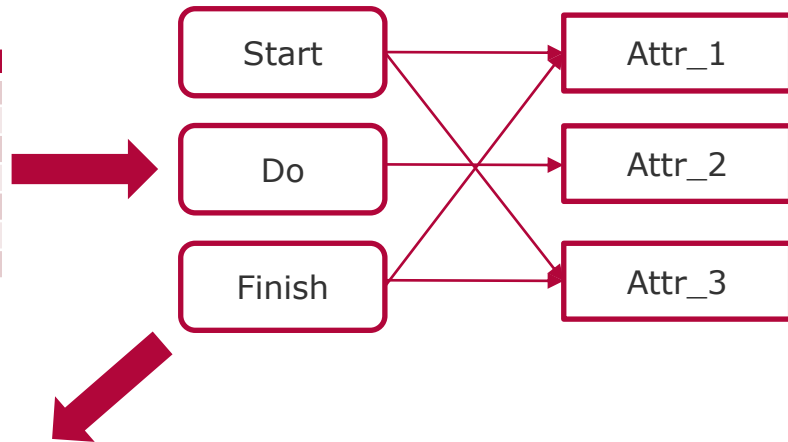


Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPI @ HPI
Summer 2021

Recap - Model Discovery on Single Event Log

Case	Activity	Time	Attr_1	Attr_2	Attr_3
1	Start	1	True		hello
1	Do	2		9.001	
2	Start	3	False		hi
2	Do	4		8.999	
1	Finish	4	False		bye
2	Finish	5	True		see you
3	Start	6	False		howdy!



Bano, Dorina & Weske, Mathias. (2020). Discovering Data Models from Event Logs

1

Attr_1
Attr_3

2

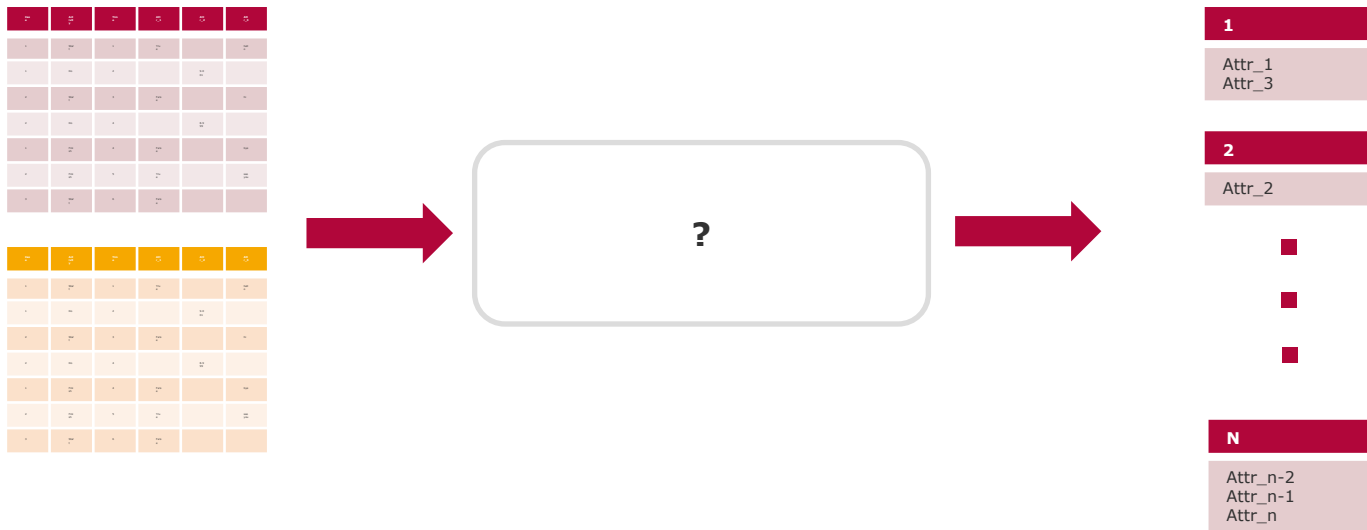
Attr_2

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 2

Recap - Problem Definition

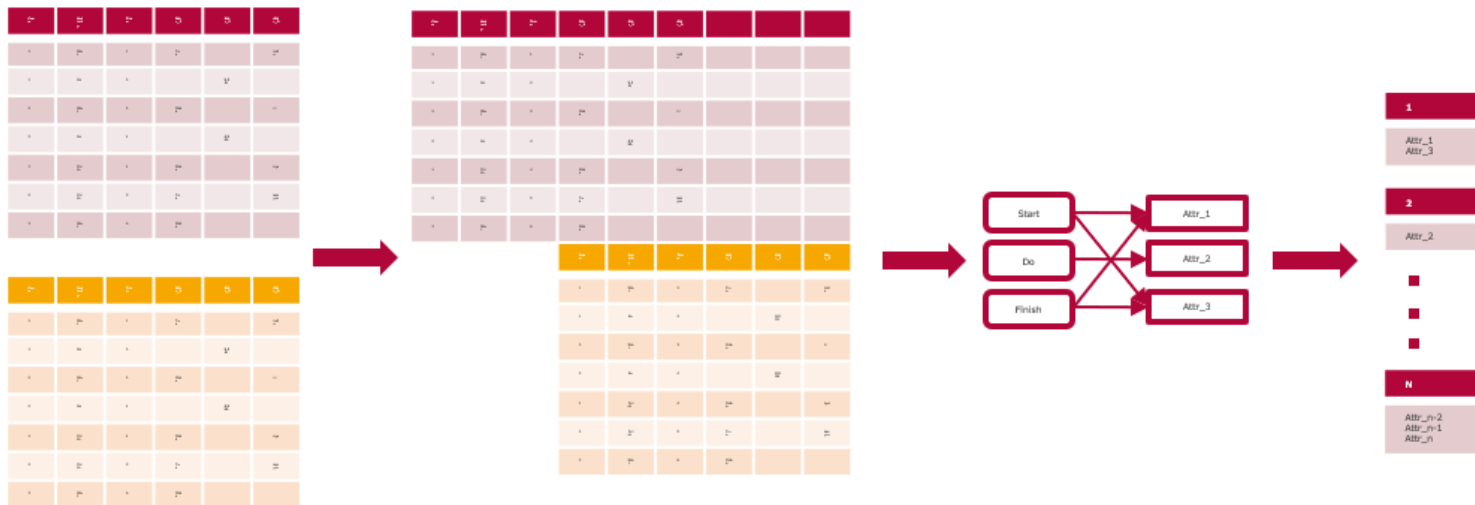
- Multiple event logs => One data model



Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 3

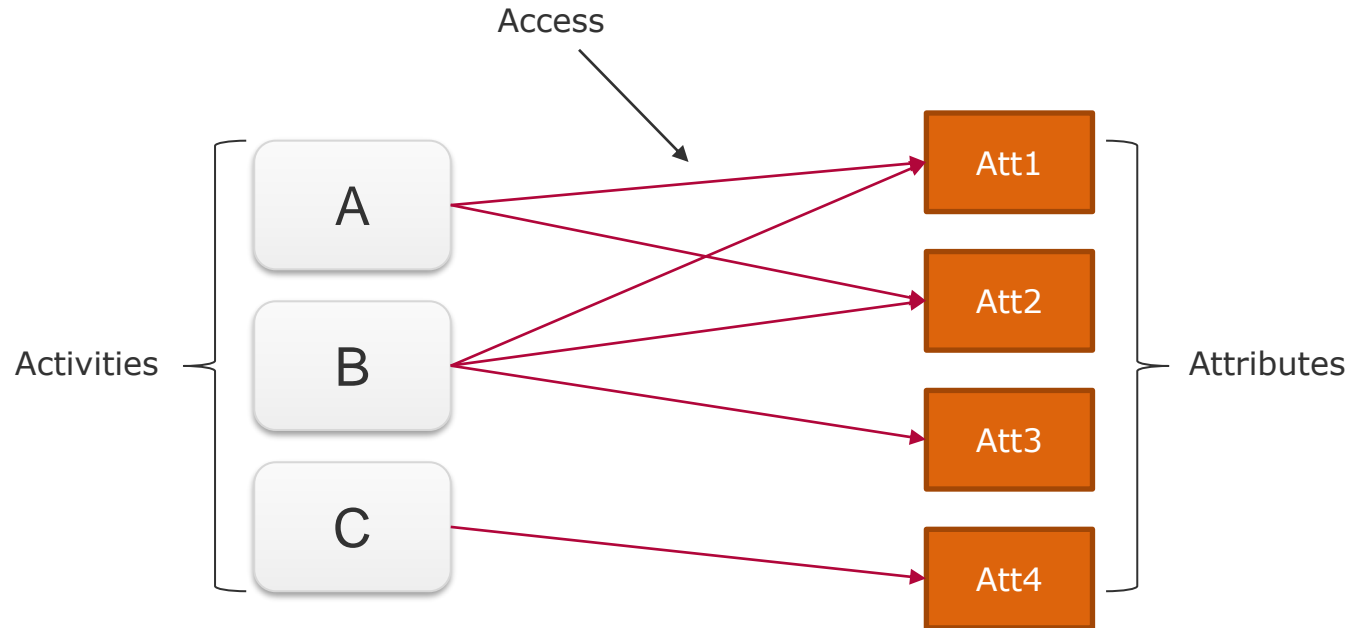
Recap – „Naive“ Approach



Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 4

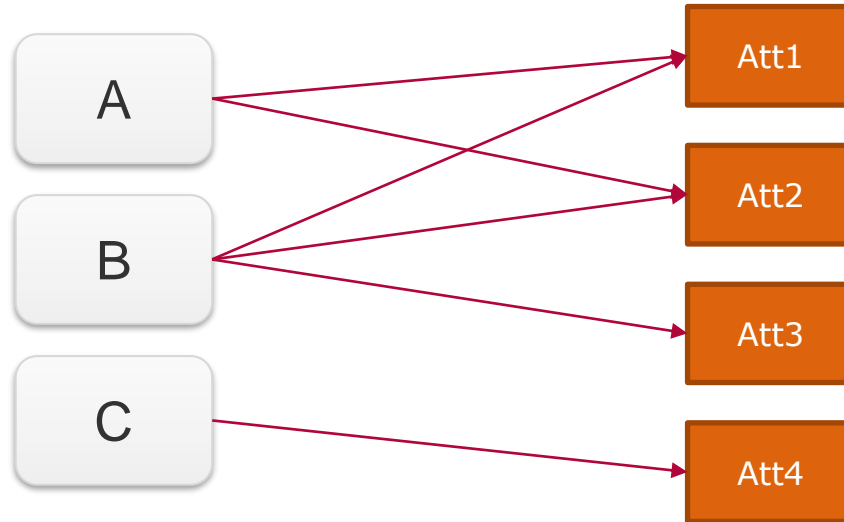
A2A Diagrams - Example



**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 5

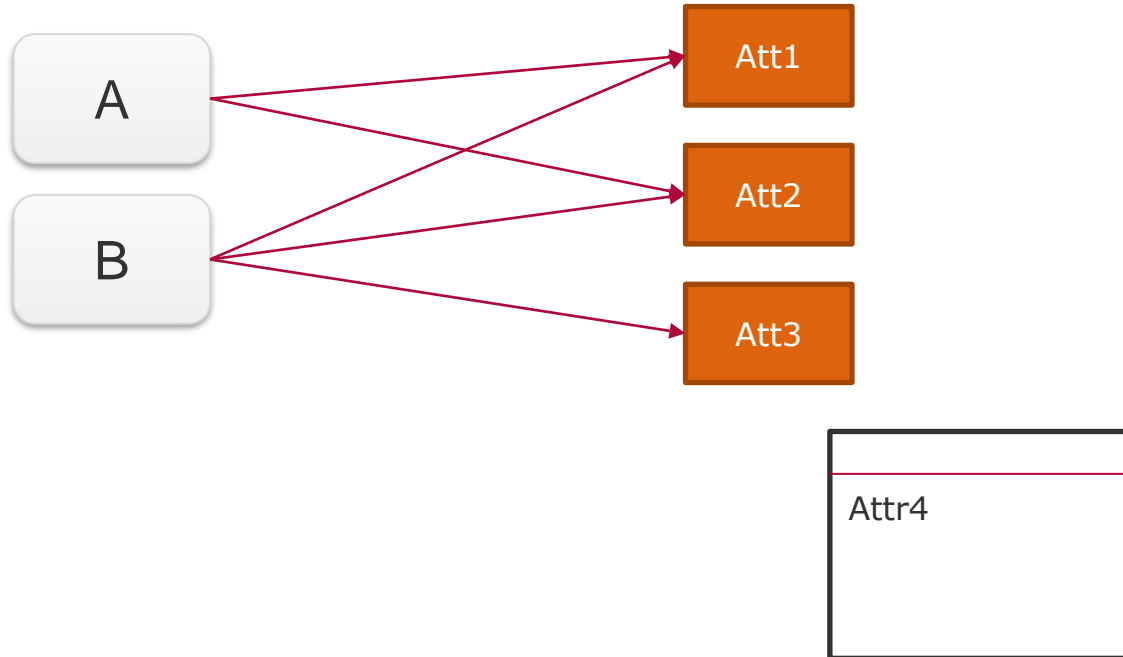
A2A Diagrams - Example



**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 6

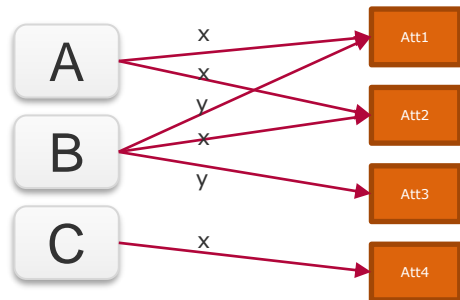
A2A Diagrams - Example



**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 7

A2A Diagrams – Data Structure

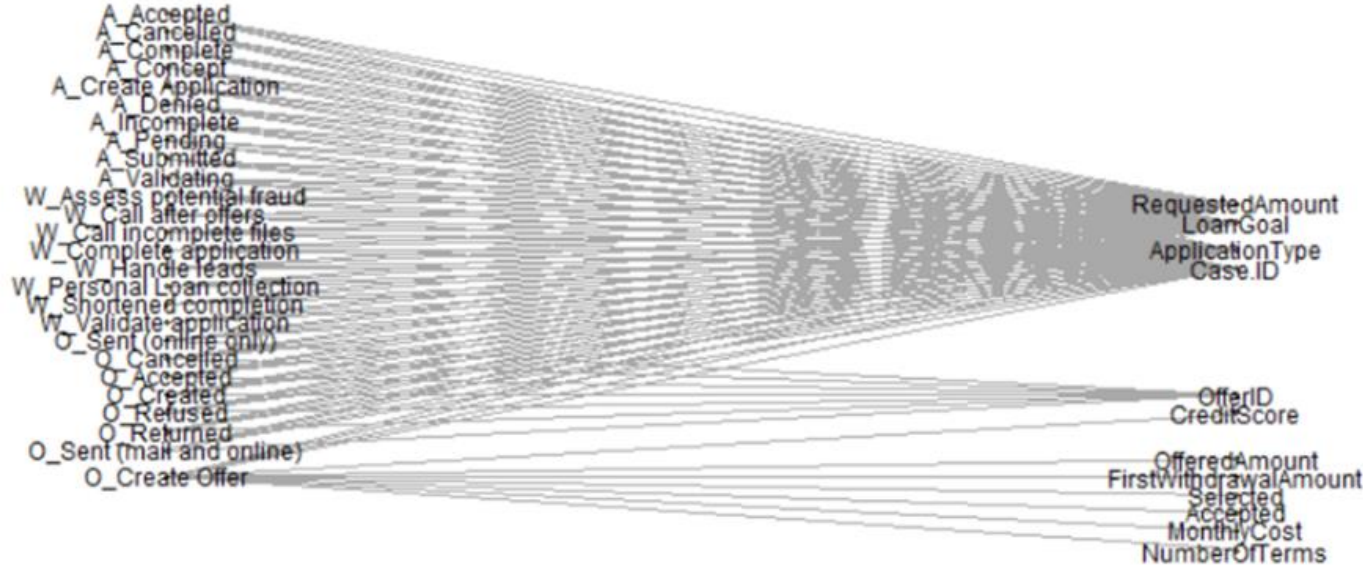


	Att1	Att2	Att3	Att4
A	x	x	0	0
B	y	x	y	0
C	0	0	0	x

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **8**

A2A Diagrams – Real World Example



A2A diagram for one BPI challenge 2017 event log

Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 9

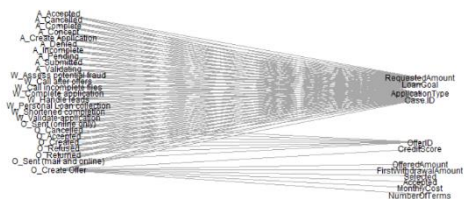
Intermission: Dataset – BPI Challenge 2017

- Two event logs
 - OfferApplications: 1160405 events, 13 attributes
 - Offers: 193849 events, 11 attributes

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **10**

A2A Diagrams – Data Structure



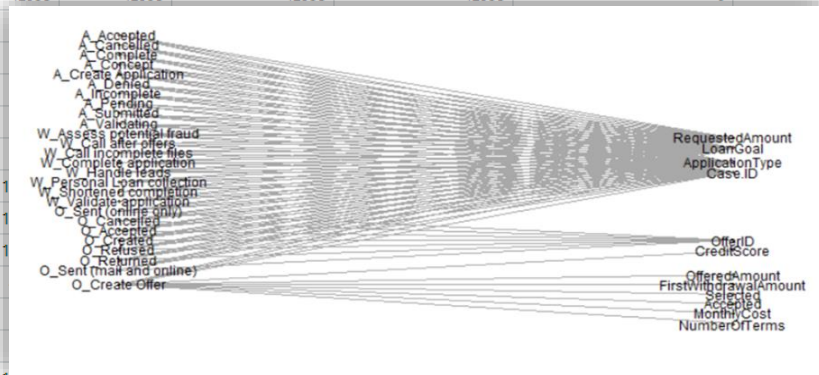
Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
1 A_Accepted	31509	31509	31509	31509		0	0	0	0	0	0	0
2 A_Cancelled	10431	10431	10431	10431		0	0	0	0	0	0	0
3 A_Complete	31362	31362	31362	31362		0	0	0	0	0	0	0
4 A_Concept	31509	31509	31509	31509		0	0	0	0	0	0	0
5 A_Create Application	31509	31509	31509	31509		0	0	0	0	0	0	0
6 A_Denied	3753	3753	3753	3753		0	0	0	0	0	0	0
7 A_Incomplete	23055	23055	23055	23055		0	0	0	0	0	0	0
8 A_Pending	17228	17228	17228	17228		0	0	0	0	0	0	0
9 A_Submitted	20423	20423	20423	20423		0	0	0	0	0	0	0
10 A_Validating	38816	38816	38816	38816		0	0	0	0	0	0	0
11 O_Accepted	17228	17228	17228	17228		0	0	0	0	0	0	17228
12 O_Cancelled	20898	20898	20898	20898		0	0	0	0	0	0	20898
13 O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
14 O_Created	42995	42995	42995	42995		0	0	0	0	0	0	42995
15 O_Refused	4695	4695	4695	4695		0	0	0	0	0	0	4695
16 O_Returned	23305	23305	23305	23305		0	0	0	0	0	0	23305
17 O_Sent (mail and online)	39707	39707	39707	39707		0	0	0	0	0	0	39707
18 O_Sent (online only)	2026	2026	2026	2026		0	0	0	0	0	0	2026
19 W_Assess potential fraud	3000	3000	3000	3000		0	0	0	0	0	0	0
20 W_Call after offers	190750	190750	190750	190750		0	0	0	0	0	0	0
21 W_Call incomplete files	165736	165736	165736	165736		0	0	0	0	0	0	0
22 W_Complete application	129796	129796	129796	129796		0	0	0	0	0	0	0
23 W_Handle leads	43771	43771	43771	43771		0	0	0	0	0	0	0
24 W_Personal Loan collection	22	22	22	22		0	0	0	0	0	0	0
25 W_Shortened completion	238	238	238	238		0	0	0	0	0	0	0
26 W_Validate application	193648	193648	193648	193648		0	0	0	0	0	0	0

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 11

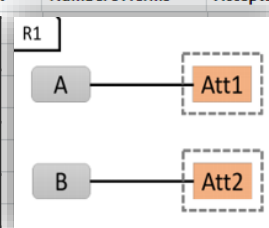
Underlying Model – Which Attributes Belong Together?

	Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
1	A_Accepted	31509	31509	31509	31509	0	0	0	0	0	0	0	0
2	A_Cancelled	10431	10431	10431	10431	0	0	0	0	0	0	0	0
3	A_Complete	31362	31362	31362	31362	0	0	0	0	0	0	0	0
4	A_Concept	31509	31509	31509	31509	0	0	0	0	0	0	0	0
5	A_Create Application	31509	31509	31509	31509	0	0	0	0	0	0	0	0
6	A_Denied	3753	3753	3753	3753	0	0	0	0	0	0	0	0
7	A_Incomplete	23055	23055	23055	23055	0	0	0	0	0	0	0	0
8	A_Pending	17228	17228	17228	17228	0	0	0	0	0	0	0	0
9	A_Submitted	20423	20423	20423	20423	0	0	0	0	0	0	0	0
10	A_Validating	38816	38816	38816	38816	0	0	0	0	0	0	0	0
11	O_Accepted	17228	17228	17228	17228	0	0	0	0	0	0	0	17228
12	O_Cancelled	20898	20898	20898	20898	0	0	0	0	0	0	0	20898
13	O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
14	O_Created	42995	42995	42995	42995	0	0	0	0	0	0	0	42995
15	O_Refused						0	0	0	0	0	0	4695
16	O_Returned						0	0	0	0	0	0	23305
17	O_Sent (mail and online)						0	0	0	0	0	0	39707
18	O_Sent (online only)						0	0	0	0	0	0	2026
19	W_Assess potential fraud						0	0	0	0	0	0	0
20	W_Call after offers						0	0	0	0	0	0	0
21	W_Call incomplete files						0	0	0	0	0	0	0
22	W_Complete application						0	0	0	0	0	0	0
23	W_Handle leads						0	0	0	0	0	0	0
24	W_Personal Loan collection						0	0	0	0	0	0	0
25	W_Shortened completion						0	0	0	0	0	0	0
26	W_Validate application						0	0	0	0	0	0	0



Underlying Model – Which Attributes Belong Together?

Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A					0							
B	0	0	0	0	x	0	0					
C					0							
D					0							
E					0							
O_Accepted	17228	17228	17228	17228								17228
O_Cancelled	20898	20898	20898	20898								20898
O_Create Offer	42995	42995	42995	42995		42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995		0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695		0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305		0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707		0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026		0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000		0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750		0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736		0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796		0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771		0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22		0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238		0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648		0	0	0	0	0	0	0

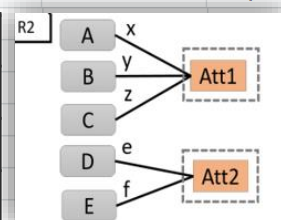


Rule 1

Underlying Model – Which Attributes Belong Together?

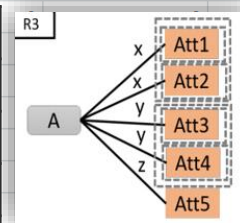
Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A					0							
B	0	0	0	0	x	0	0					
C	0	0	0	0	x	0	0					
D	0	0	0	0	x	0	0					
E					0							
O_Accepted	17228	17228	17228	17228								17228
O_Cancelled	20898	20898	20898	20898								20898
O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995		0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695		0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305		0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707		0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026		0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000		0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750		0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736		0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796		0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771		0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22		0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238		0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648		0	0	0	0	0	0	0

Rule 2



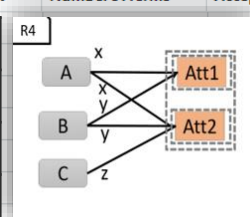
Underlying Model – Which Attributes Belong Together?

Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A			0	0	0							
B	0	0	x	x	x	0	0					
C			0	0	0							
D			0	0	0							
E			0	0	0							
O_Accepted	17228	17228	17228	17228								17228
O_Cancelled	20898	20898	20898	20898								20898
O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995		0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695		0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305		0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707		0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026		0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000		0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750		0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736		0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796		0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771		0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22		0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238		0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648		0	0	0	0	0	0	0



Underlying Model – Which Attributes Belong Together?

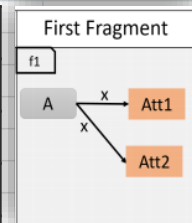
Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A	z	0	0	m	n	z	0					
B	y	y	x	x	x	0	z					
C	0	x	0	z	0	y	x					
D	x	0	0	0	y	z	0					
E	0	y	0	x	0	0						
O_Accepted	17228	17228	17228	17228								17228
O_Cancelled	20898	20898	20898	20898								20898
O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995	42995	0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695	4695	0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305	23305	0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707	39707	0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026	2026	0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000	3000	0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750	190750	0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736	165736	0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796	129796	0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771	43771	0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22	22	0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238	238	0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648	193648	0	0	0	0	0	0	0



Rule 4

Underlying Model – Which Attributes Belong Together?

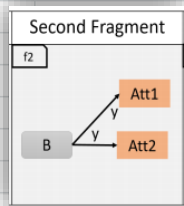
Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A	z	0	0	m	n	z	0	0	0	0	0	0
B	y	y	x	x	x	0	z	0	0	0	0	0
C	0	x	0	z	0	y	x	0	0	0	0	0
D	x	0	0	0	y	z	0	0	0	0	0	0
E	0	y	0	x	0	0	0	0	0	0	0	0
O_Accepted	17228	17228	17228	17228			0	0	0	0	0	17228
O_Cancelled	20898	20898	20898	20898			0	0	0	0	0	20898
O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995	42995	0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695	4695	0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305	23305	0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707	39707	0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026	2026	0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000	3000	0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750	190750	0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736	165736	0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796	129796	0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771	43771	0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22	22	0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238	238	0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648	193648	0	0	0	0	0	0	0



Rule 4

Underlying Model – Which Attributes Belong Together?

Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
Activity	Att1	Att2	Att3	Att4	Att5	Att6	Att7					
A	z	0	0	m	n	z	0	0	0	0	0	0
B	y	y	x	x	x	0	z	0	0	0	0	0
C	0	x	0	z	0	y	x	0	0	0	0	0
D	x	0	0	0	y	z	0	0	0	0	0	0
E	0	y	0	x	0	0	0	0	0	0	0	0
O_Accepted	17228	17228	17228	17228			0	0	0	0	0	17228
O_Cancelled	20898	20898	20898	20898			0	0	0	0	0	20898
O_Create Offer	42995	42995	42995	42995		42995	42995	42995	42995	42995	42995	0
O_Created	42995	42995	42995	42995		0	0	0	0	0	0	42995
O_Refused	4695	4695	4695	4695		0	0	0	0	0	0	4695
O_Returned	23305	23305	23305	23305		0	0	0	0	0	0	23305
O_Sent (mail and online)	39707	39707	39707	39707		0	0	0	0	0	0	39707
O_Sent (online only)	2026	2026	2026	2026		0	0	0	0	0	0	2026
W_Assess potential fraud	3000	3000	3000	3000		0	0	0	0	0	0	0
W_Call after offers	190750	190750	190750	190750		0	0	0	0	0	0	0
W_Call incomplete files	165736	165736	165736	165736		0	0	0	0	0	0	0
W_Complete application	129796	129796	129796	129796		0	0	0	0	0	0	0
W_Handle leads	43771	43771	43771	43771		0	0	0	0	0	0	0
W_Personal Loan collection	22	22	22	22		0	0	0	0	0	0	0
W_Shortened completion	238	238	238	238		0	0	0	0	0	0	0
W_Validate application	193648	193648	193648	193648		0	0	0	0	0	0	0



Rule 4

Offers – A2A Matrix

	Activity	Case.ID	MonthlyCost	Selected	ApplicationID	FirstWithdrawalAmount	CreditScore	OfferedAmount	NumberOfTerms	Accepted	OfferID
1	O_Accepted	17228	17228	17228	17228	17228	17228	17228	17228	17228	17228
2	O_Cancelled	20898	20898	20898	20898	20898	20898	20898	20898	20898	20898
3	O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
4	O_Created	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995
5	O_Refused	4695	4695	4695	4695	4695	4695	4695	4695	4695	4695
6	O_Returned	23305	23305	23305	23305	23305	23305	23305	23305	23305	23305
7	O_Sent (mail and online)	39707	39707	39707	39707	39707	39707	39707	39707	39707	39707
8	O_Sent (online only)	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **19**

OfferApplications – A2A Matrix

	Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID
1	A_Accepted	31509	31509	31509	31509	0	0	0	0	0	0	0	0
2	A_Cancelled	10431	10431	10431	10431	0	0	0	0	0	0	0	0
3	A_Complete	31362	31362	31362	31362	0	0	0	0	0	0	0	0
4	A_Concept	31509	31509	31509	31509	0	0	0	0	0	0	0	0
5	A_Create Application	31509	31509	31509	31509	0	0	0	0	0	0	0	0
6	A_Denied	3753	3753	3753	3753	0	0	0	0	0	0	0	0
7	A_Incomplete	23055	23055	23055	23055	0	0	0	0	0	0	0	0
8	A_Pending	17228	17228	17228	17228	0	0	0	0	0	0	0	0
9	A_Submitted	20423	20423	20423	20423	0	0	0	0	0	0	0	0
10	A_Validating	38816	38816	38816	38816	0	0	0	0	0	0	0	0
11	O_Accepted	17228	17228	17228	17228	0	0	0	0	0	0	0	17228
12	O_Cancelled	20898	20898	20898	20898	0	0	0	0	0	0	0	20898
13	O_Create Offer	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	0
14	O_Created	42995	42995	42995	42995	0	0	0	0	0	0	0	42995
15	O_Refused	4695	4695	4695	4695	0	0	0	0	0	0	0	4695
16	O_Returned	23305	23305	23305	23305	0	0	0	0	0	0	0	23305
17	O_Sent (mail and online)	39707	39707	39707	39707	0	0	0	0	0	0	0	39707
18	O_Sent (online only)	2026	2026	2026	2026	0	0	0	0	0	0	0	2026
19	W_Assess potential fraud	3000	3000	3000	3000	0	0	0	0	0	0	0	0
20	W_Call after offers	190750	190750	190750	190750	0	0	0	0	0	0	0	0
21	W_Call incomplete files	165736	165736	165736	165736	0	0	0	0	0	0	0	0
22	W_Complete application	129796	129796	129796	129796	0	0	0	0	0	0	0	0
23	W_Handle leads	43771	43771	43771	43771	0	0	0	0	0	0	0	0
24	W_Personal Loan collection	22	22	22	22	0	0	0	0	0	0	0	0
25	W_Shortened completion	238	238	238	238	0	0	0	0	0	0	0	0
26	W_Validate application	193648	193648	193648	193648	0	0	0	0	0	0	0	0

New Rule: Multilog Access

„Each set of cells accessed by multiple event logs is treated as representing the same data model class.“

	Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID	ApplicationID
11	O_Accepted	34456	17228	17228	17228	17228	17228	17228	17228	17228	17228	17228	34456	17228
12	O_Cancelled	41796	20898	20898	20898	20898	20898	20898	20898	20898	20898	20898	41796	20898
13	O_Create Offer	85990	42995	42995	42995	85990	85990	85990	85990	85990	85990	85990	0	42995
14	O_Created	85990	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	85990	42995
15	O_Refused	9390	4695	4695	4695	4695	4695	4695	4695	4695	4695	4695	9390	4695
16	O_Returned	46610	23305	23305	23305	23305	23305	23305	23305	23305	23305	23305	46610	23305
17	O_Sent (mail and online)	79414	39707	39707	39707	39707	39707	39707	39707	39707	39707	39707	79414	39707
18	O_Sent (online only)	4052	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026	4052	2026

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 22

New Rule: Multilog Access

„Each set of cells accessed by multiple event logs is treated as accessed by the same data model class.“

	Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID	ApplicationID
11	O_Accepted	34456	17228	17228	17228	17228	17228	17228	17228	17228	17228	17228	34456	17228
12	O_Cancelled	41796	20898	20898	20898	20898	20898	20898	20898	20898	20898	20898	41796	20898
13	O_Create Offer	85990	42995	42995	42995	85990	85990	85990	85990	85990	85990	85990	0	42995
14	O_Created	85990	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	85990	42995
15	O_Refused	9390	4695	4695	4695	4695	4695	4695	4695	4695	4695	4695	9390	4695
16	O_Returned	46610	23305	23305	23305	23305	23305	23305	23305	23305	23305	23305	46610	23305
17	O_Sent (mail and online)	79414	39707	39707	39707	39707	39707	39707	39707	39707	39707	39707	79414	39707
18	O_Sent (online only)	4052	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026	4052	2026

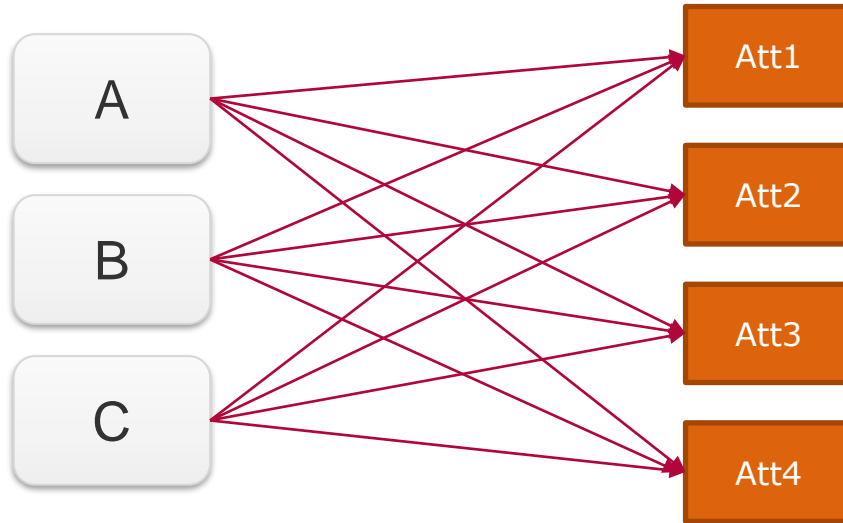
C1
Case.ID
FirstWithdrawalAmount
NumberOfTerms
Accepted
MonthlyCost
Selected
CreditScore
OfferedAmount
OfferID

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 23

New Definition: Fully Connected A2A

„A2A diagram/matrix in which **all** activities access **all** attributes“



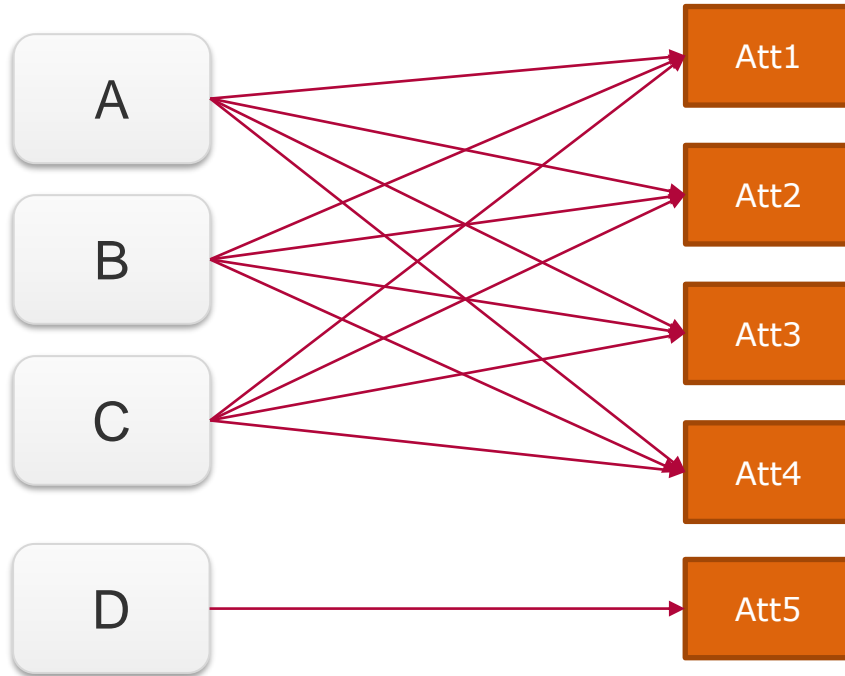
	Att1	Att2	Att3
A	x	y	z
B	x	y	x
C	x	x	y

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **24**

New Definition: Partially Fully Connected A2A

„A2A diagrams / matrices in which a **subset** of activities always accesses a **subset** of attributes“



	Att1	Att2	Att3	Att4
A	x	y	z	0
B	x	x	x	0
C	x	x	x	0
D	0	0	0	y

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **25**

New Rule: (Partially) Fully Connected A2A

- Attributes belonging to partially fully connected a2a subsets are treated as belonging to one data class

	Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID	ApplicationID
11	O_Accepted	34456	17228	17228	17228	17228	17228	17228	17228	17228	17228	17228	34456	17228
12	O_Cancelled	41796	20898	20898	20898	20898	20898	20898	20898	20898	20898	20898	41796	20898
13	O_Create Offer	85990	42995	42995	42995	85990	85990	85990	85990	85990	85990	85990	0	42995
14	O_Created	85990	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	85990	42995
15	O_Refused	9390	4695	4695	4695	4695	4695	4695	4695	4695	4695	4695	9390	4695
16	O_Returned	46610	23305	23305	23305	23305	23305	23305	23305	23305	23305	23305	46610	23305
17	O_Sent (mail and online)	79414	39707	39707	39707	39707	39707	39707	39707	39707	39707	39707	79414	39707
18	O_Sent (online only)	4052	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026	4052	2026

c2

LoanGoal
ApplicationType
RequestedAmount
ApplicationID

c1

Case.ID
FirstWithdrawalAmount
NumberOfTerms
Accepted
MonthlyCost
Selected
CreditScore
OfferedAmount
OfferID

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 26

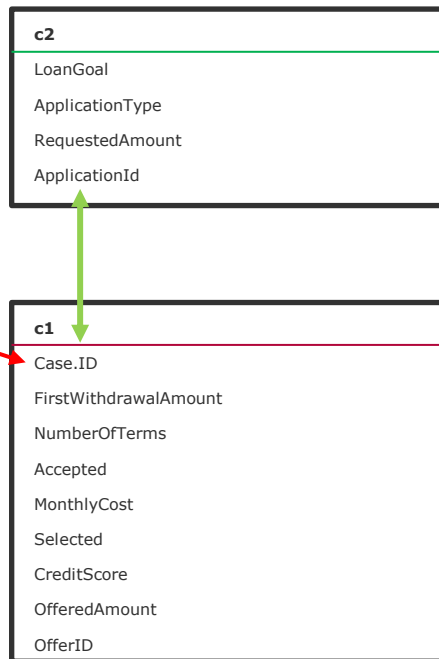
BPI 2017 Challenge Data Definition vs. Result

For all applications, the following data is available:

- Requested load amount (in Euro),
- The application type,
- The reason the loan was applied for (LoanGoal), and
- An application ID.

For all offers, the following data is available:

- An offer ID,
- The offered amount,
- The initial withdrawal amount,
- The number of payback terms agreed to,
- The monthly costs,
- The creditscore of the customer,
- The employee who created the offer,
- Whether the offer was selected, and
- Whether the offer was accepted by the customer.

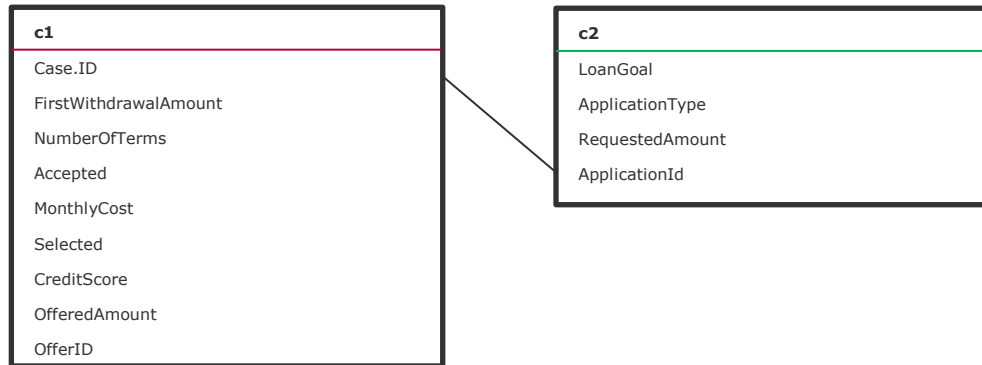


<https://www.win.tue.nl/bpi/doku.php?id=2017:challenge>

Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 27

Limitation - Key Matching



In performance experiments on five diverse datasets, our algorithm reaches an average recall of 88% and 91% in primary keys and foreign keys discovery, respectively. We show with an experiment that without knowledge of primary keys (which is assumed in related work), the performance of foreign key discovery is much worse, indicating the necessity to discover primary keys in advance or simultaneously. We compared precision and recall with the state-of-art algorithms.

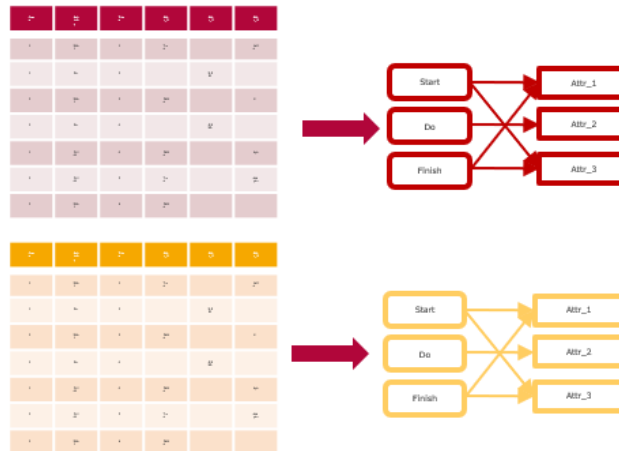
Jiang, L., Naumann, F.
Holistic primary key and
foreign key detection. *J Intell
Inf Syst* **54**, 439–461 (2020)

Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **28**

Proposed Algorithm

- A2A for all EL
- Multilog-Rule + Combine (Union)
- Fully Connected Rule
- Rules 1-4 as normal



**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart 29

Proposed Algorithm

- A2A for all EL
- Multilog-Rule + Combine (Union)
- Fully Connected Rule
- Rules 1-4 as normal

Activity	Case.ID	LoanGoal	ApplicationType	RequestedAmount	FirstWithdrawalAmount	NumberOfTerms	Accepted	MonthlyCost	Selected	CreditScore	OfferedAmount	OfferID	ApplicationID
11 O_Accepted	34456	17228	17228	17228	17228	17228	17228	17228	17228	17228	17228	34456	17228
12 O_Cancelled	41796	20898	20898	20898	20898	20898	20898	20898	20898	20898	20898	41796	20898
13 O_Create Offer	85990	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	85990	42995
14 O_Created	85990	42995	42995	42995	42995	42995	42995	42995	42995	42995	42995	85990	42995
15 O_Refused	9390	4695	4695	4695	4695	4695	4695	4695	4695	4695	4695	9390	4695
16 O_Returned	46610	23305	23305	23305	23305	23305	23305	23305	23305	23305	23305	46610	23305
17 O_Sent (mail and online)	79414	39707	39707	39707	39707	39707	39707	39707	39707	39707	39707	79414	39707
18 O_Sent (online only)	4052	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026	4052	2026

t1
Case.ID
<u>FirstWithdrawalAmount</u>
<u>NumberOfTerms</u>
Accepted
<u>MonthlyCost</u>
Selected
<u>CreditScore</u>
<u>OfferedAmount</u>
OfferID

1	2	3	4	5	6
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1



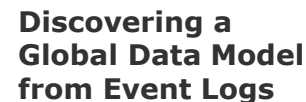
1	2	3	4	5	6
1	1	1	1	1	1
2	1	1	1	1	1
3	1	1	1	1	1
4	1	1	1	1	1
5	1	1	1	1	1
6	1	1	1	1	1



**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **30**

-
- A diagram of a fully connected feedforward neural network. On the left, there are three light gray rounded rectangular nodes labeled 'A', 'B', and 'C' stacked vertically. On the right, there are four orange rounded rectangular nodes labeled 'Att1', 'Att2', 'Att3', and 'Att4' stacked vertically. Every node on the left is connected to every node on the right by a red line, representing a fully connected architecture.



Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **31**

Proposed Algorithm

- A2A for all EL
- Multilog-Rule + Combine (Union)
- Fully Connected Rule
- Rules 1-4 as normal

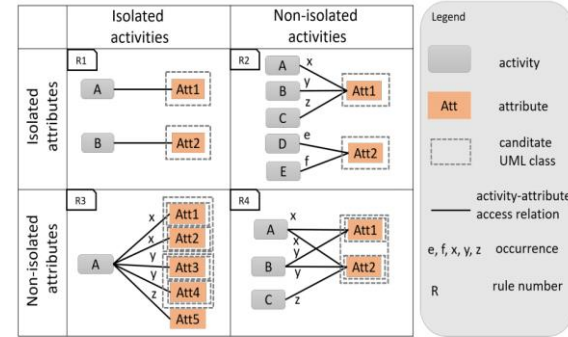
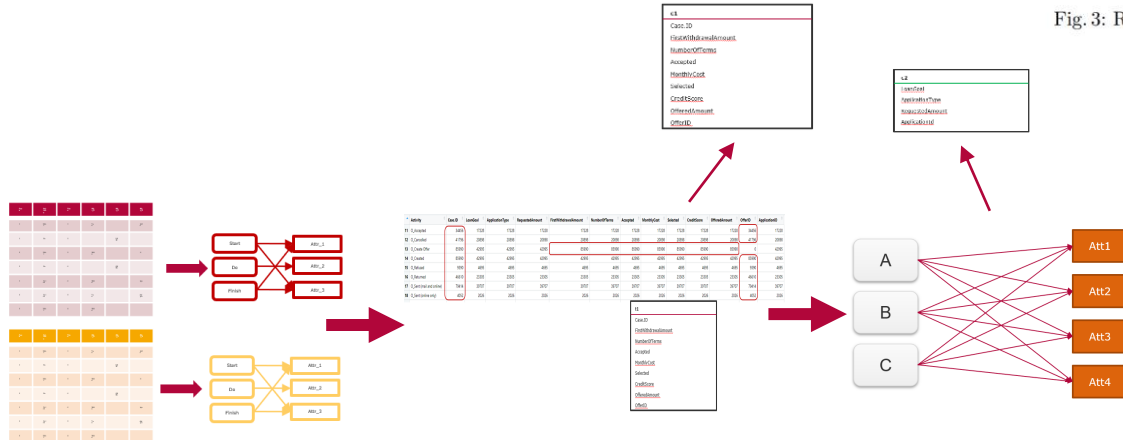


Fig. 3: Rules for deriving data model classes



Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer 2021
Chart 32

Proposed Algorithm

- A2A for all EL
- Multilog-Rule + Combine (Union)
- Fully Connected Rule
- Rules 1-4 as normal



Discovering a Global Data Model from Event Logs

Kim-Pascal Borchart
BPT @ HPI Summer 2021
Chart 33

Dataset Overview – BPI Challenge 2020

- Dataset of five event logs
 - Permits: 86581 events, 170 attributes
 - DomesticDeclarations: 56437 events, 7 attributes
 - InternationalDeclarations: 72151 events, 20 attributes
 - PrepaidTravelCost: 18246 events, 19 attributes
 - RequestForPayment: 36796 events, 11 attributes

**Discovering a
Global Data Model
from Event Logs**

Kim-Pascal Borchart
BPT @ HPI Summer
2021
Chart **34**