

Federated Process Mining

Federated Event Log Generation for Process Discovery across Organizational Boundaries

Bernardo Nugroho Yahya

Industrial & Management Engineering

Hankuk University of Foreign Studies, South Korea



한국외국어대학교
HANKUK UNIVERSITY OF FOREIGN STUDIES

Agenda

1 • The event-log assumption

2 • How is the data distributed? - Examples

3 • Collaborative Process Mining

4 • When does the Federation become necessary?

5 • Protect – Correlate - Federate

6 • Takeaways

Federated Process Mining: Exploiting Event Data Across Organizational Boundaries

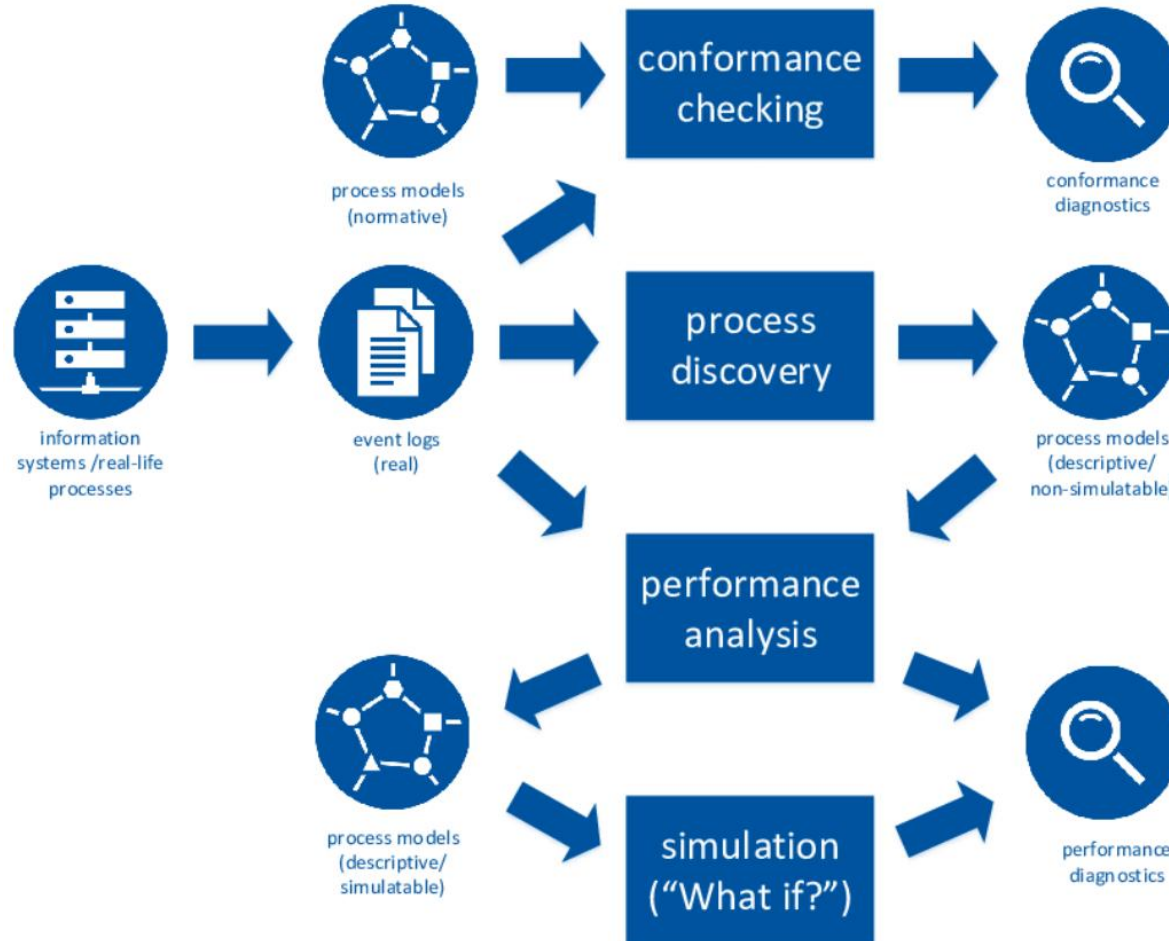
Wil M.P. van der Aalst

Process and Data Science (Informatik 9), RWTH Aachen University, Aachen, Germany and

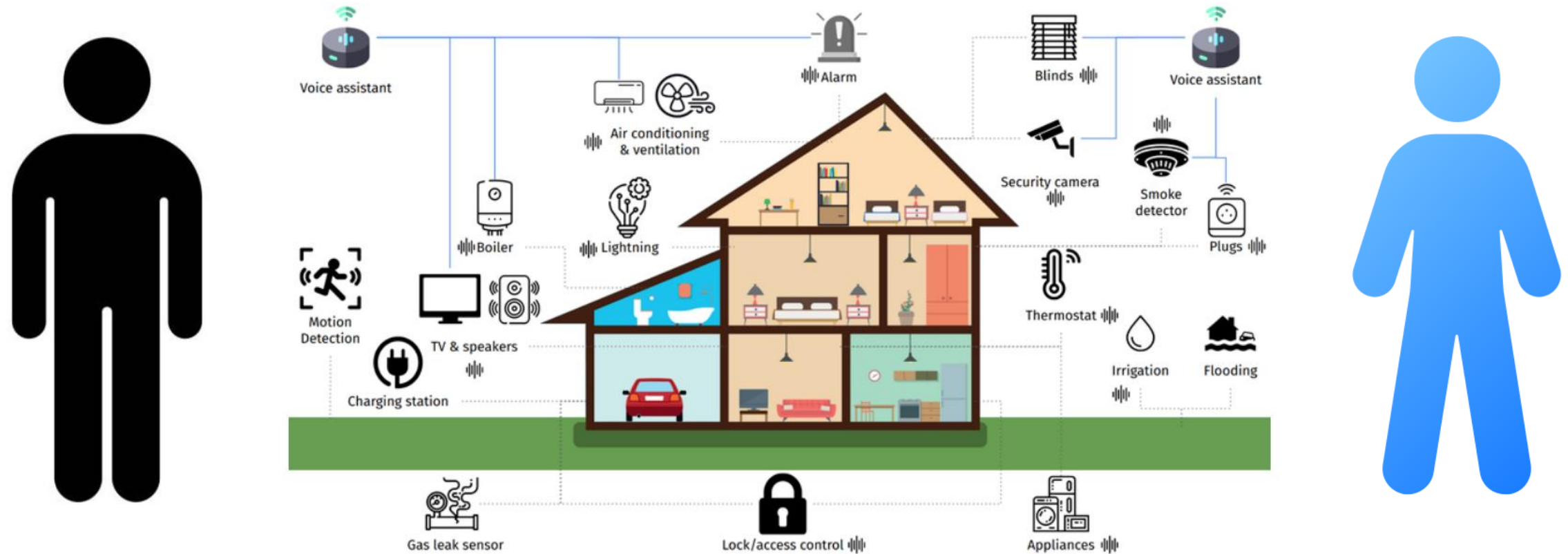
Fraunhofer-Institut für Angewandte Informationstechnik (FIT), Sankt Augustin, Germany

Email: wvdaalst@pads.rwth-aachen.de

The Event Log Assumption



Smart Home Environment - Example



Smart Home Environment - Example

Alternative Case Notions



Event	Subject	Timestamp	Activity	Case: Subject	Case: Subject + Day
E01	S1	2026-08-24 07:00	Wake Up	S1	S1-0824
E02	S1	2026-08-24 07:05	Toileting	S1	S1-0824
E03	S1	2026-08-24 07:20	Prepare Breakfast	S1	S1-0824
E04	S1	2026-08-24 08:00	Leave Home	S1	S1-0824
E05	S1	2026-08-25 06:50	Wake Up	S1	S1-0825
E06	S1	2026-08-25 06:55	Toileting	S1	S1-0825
E07	S1	2026-08-25 07:10	Prepare Breakfast	S1	S1-0825
E08	S1	2026-08-25 07:50	Leave Home	S1	S1-0825
E09	S2	2026-08-24 06:30	Wake Up	S2	S2-0824
E10	S2	2026-08-24 06:40	Prepare Breakfast	S2	S2-0824
E11	S2	2026-08-24 07:00	Toileting	S2	S2-0824
E12	S2	2026-08-24 07:40	Leave Home	S2	S2-0824
E13	S2	2026-08-25 07:10	Wake Up	S2	S2-0825
E14	S2	2026-08-25 07:15	Toileting	S2	S2-0825
E15	S2	2026-08-25 07:45	Leave Home	S2	S2-0825

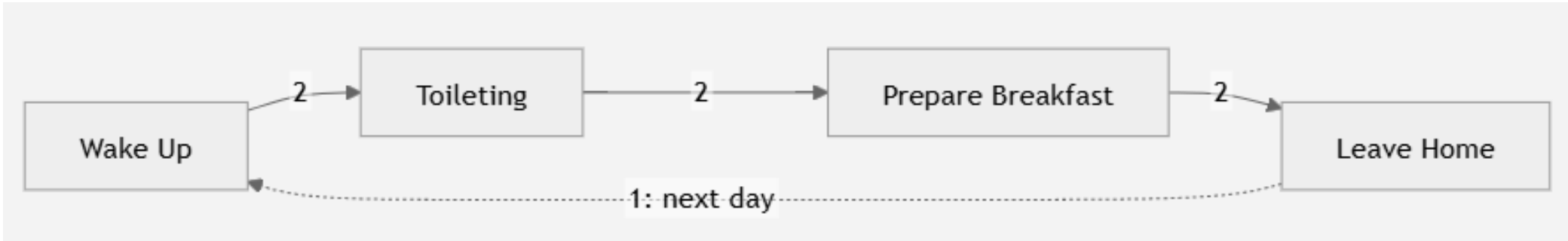
Smart Home Environment – CaseID (Subject)



Subject 1

Wake Up → Toileting → Prepare Breakfast → Leave Home
→ Wake Up → Toileting → Prepare Breakfast → Leave Home

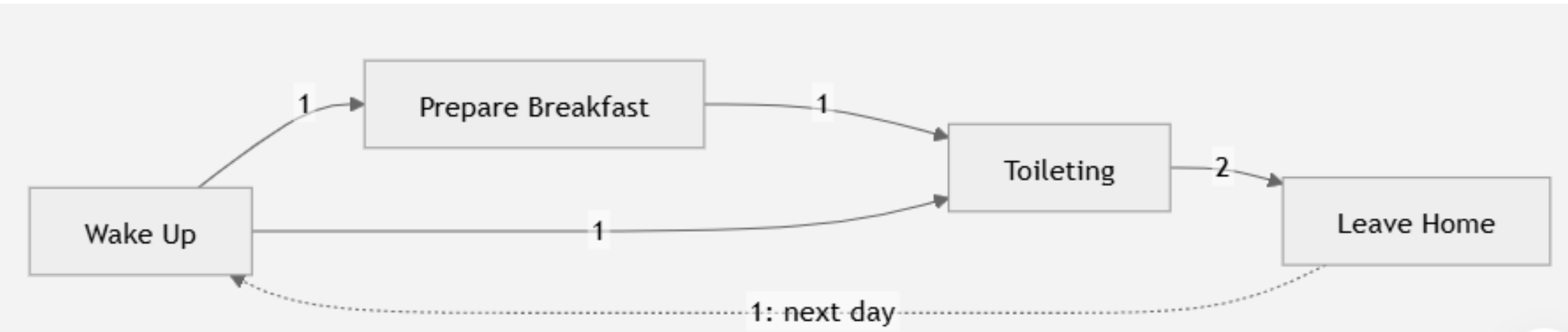
Directly-Follows Graph (DFG)



Subject 2

Wake Up → Prepare Breakfast → Toileting → Leave Home
→ Wake Up → Toileting → Leave Home

Directly-Follows Graph (DFG)



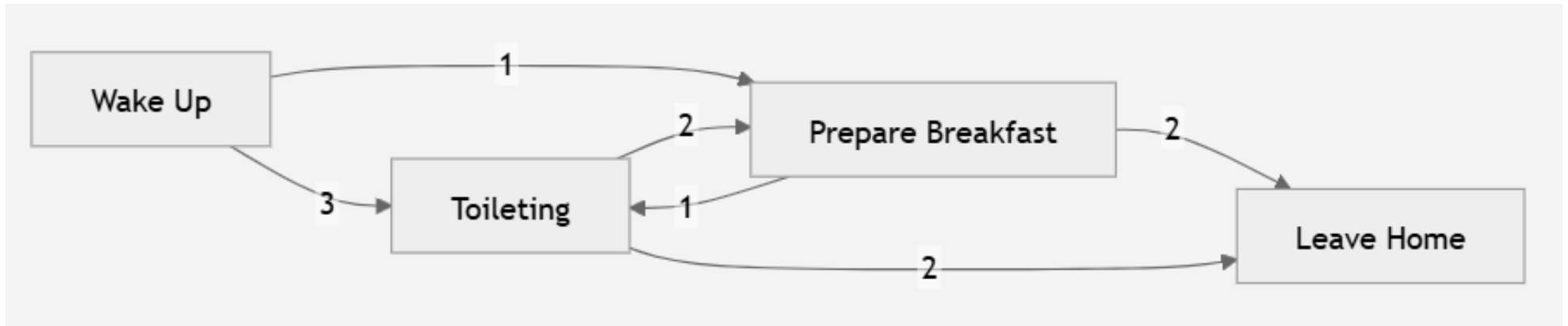
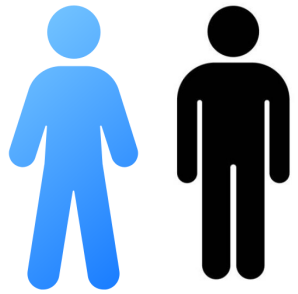
Smart Home Environment – CaseID (Subject + Day)

S1-0824: Wake Up → Toileting → Prepare Breakfast → Leave Home

S1-0825: Wake Up → Toileting → Prepare Breakfast → Leave Home

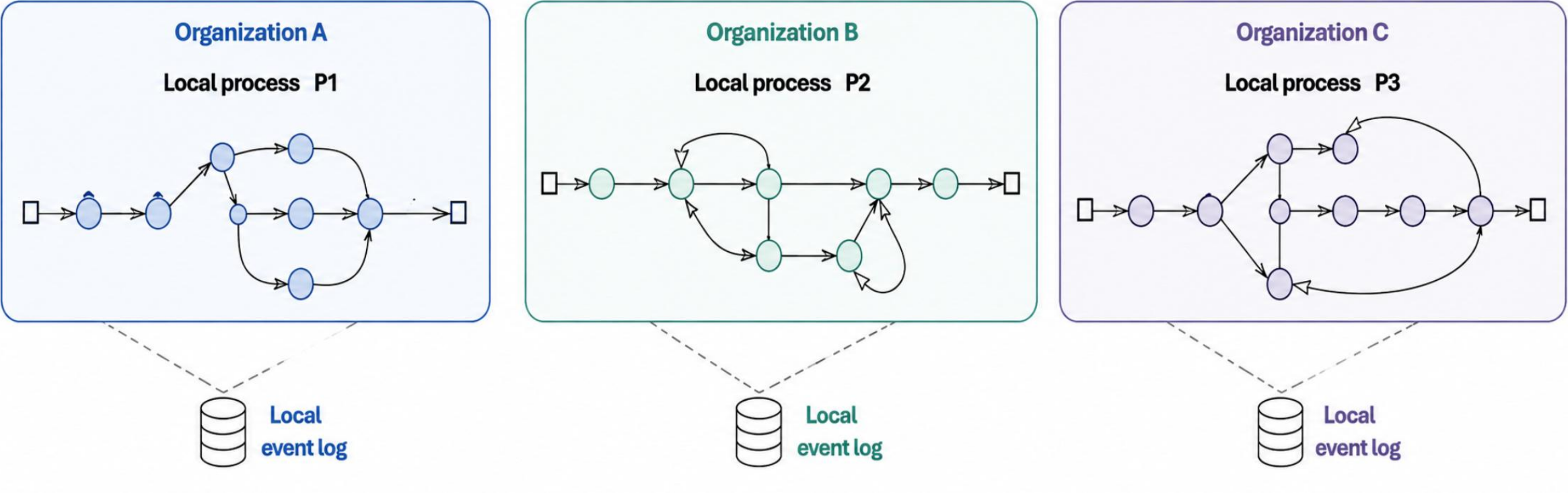
S2-0824: Wake Up → Prepare Breakfast → Toileting → Leave Home

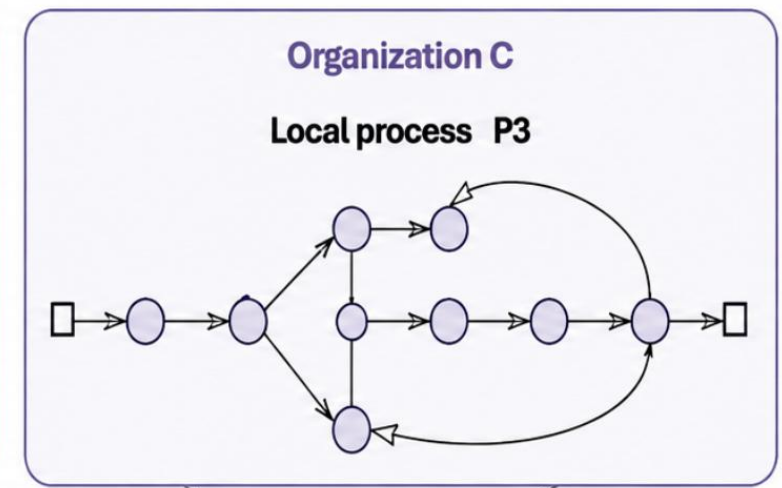
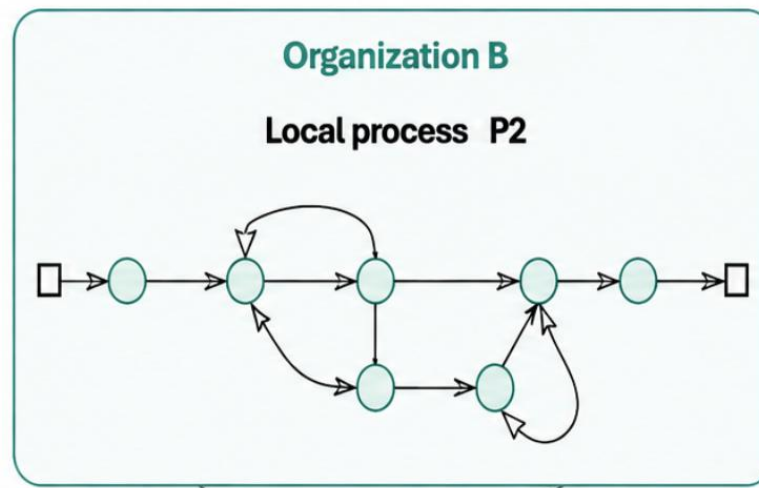
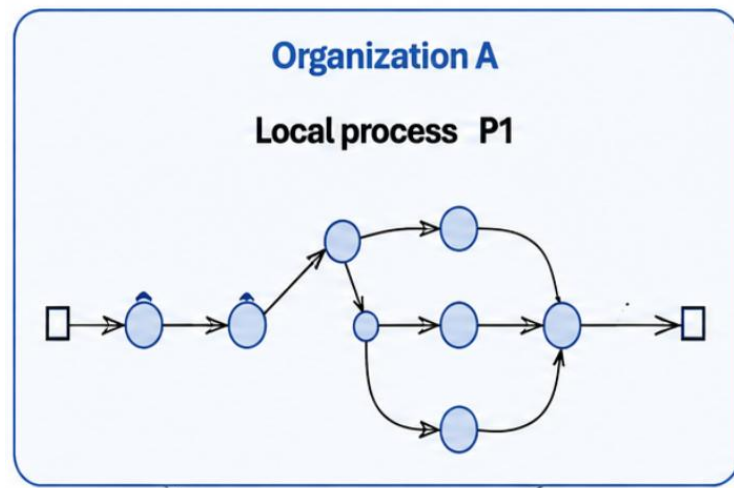
S2-0825: Wake Up → Toileting → Leave Home



The events remain unchanged, but the selected case notion changes the process view. OCPM avoids committing to one flat case notion upfront.

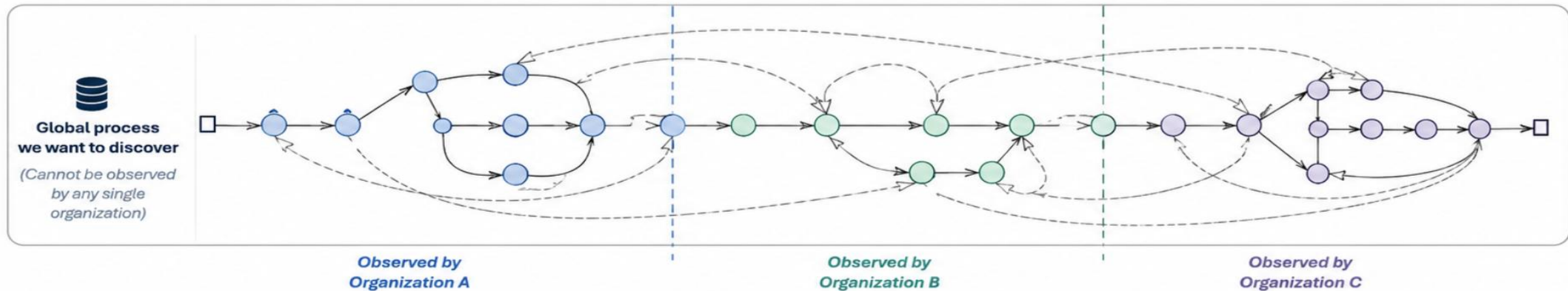
Each organization has its own local process





But some process questions require evidence from more than one organization.

(End-to-end flow, handovers, waiting times, rework across boundaries, bottlenecks, compliance, etc.)



Relational Conditions Influencing Data Sharing

Conflict

- Organizations have competing interests, goals, or priorities.
- Concerns about privacy, misuse, and competitive disadvantage limit data sharing.
- Higher conflict usually results in restricted or highly protected information exchange.

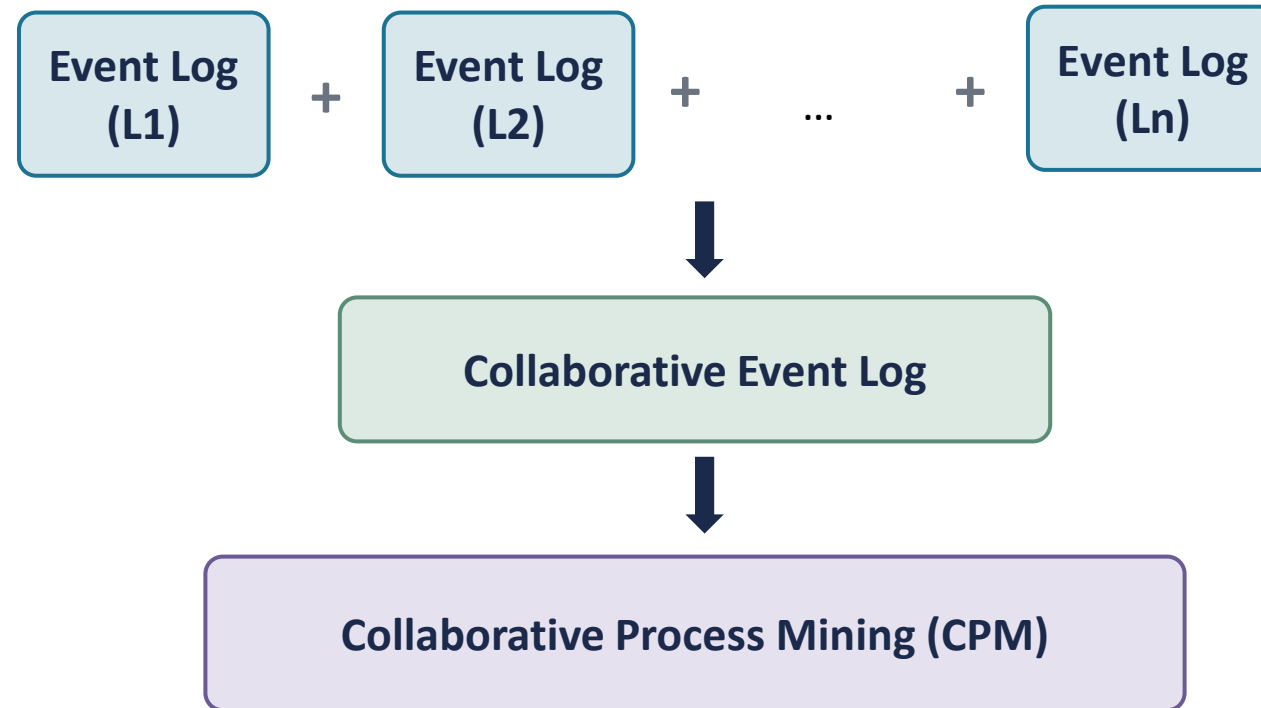
Cooperation

- Organizations work together to achieve shared operational objectives.
- They exchange relevant data to coordinate activities and improve process visibility.
- Cooperation enables joint analysis while each organization retains its autonomy.

Commitment

- Organizations are willing to maintain a stable, long-term relationship.
- They consistently contribute agreed-upon data and resources.
- Strong commitment supports trust, continuity, and sustainable cross-organizational process mining.

When Central Sharing Is Acceptable



If data can be centrally shared under acceptable governance, collaborative Process Mining may be sufficient.

When Does Federation Become Useful?



Organizations still need cross-boundary process insight — but complete raw-log pooling is undesirable, restricted, impractical, or unnecessary.

Privacy	Confidentiality	Commercial sensitivity
Contractual boundaries	Governance	Regulatory constraints
Data ownership	Infrastructure limits	Unnecessary exposure

What “Federate” Actually Means

FEDERATE *Connect or combine independent entities while keeping them locally controlled.*

FEDERATED *Connected and coordinated, but still locally controlled.*

“Federate = collaborate without fully centralizing.”

Federated Learning

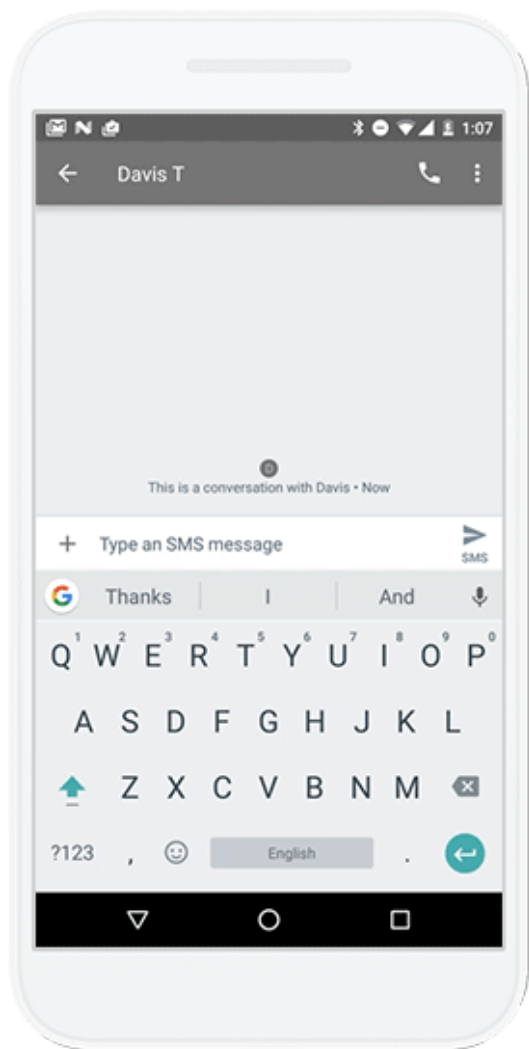
Federated learning (FL):

- a decentralized machine learning method
- it trains a shared AI model using data from many devices or servers
- without centralizing the raw data itself.

Update Model



Federated Learning

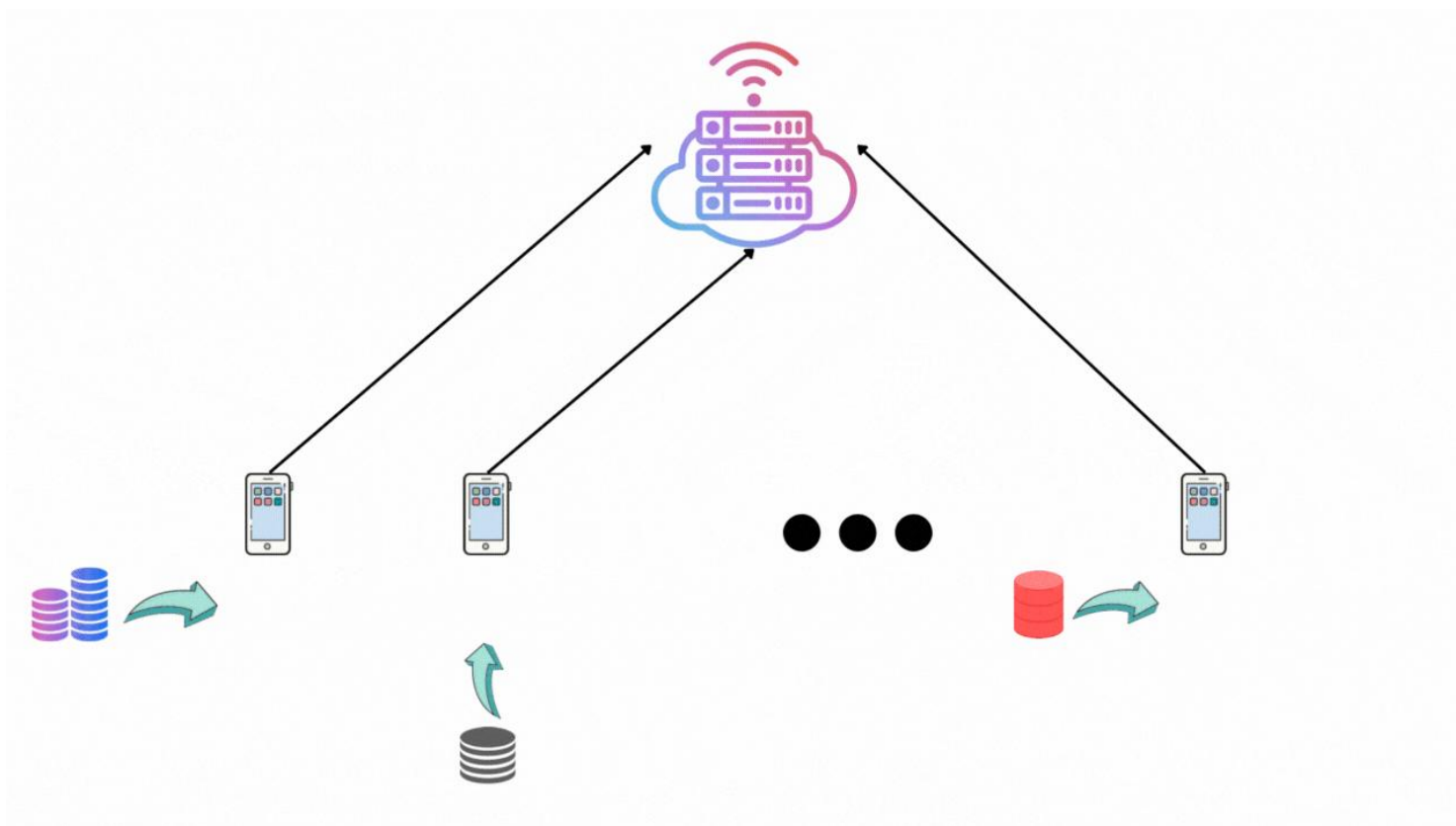


FEDERATED LEARNING FOR MOBILE KEYBOARD PREDICTION

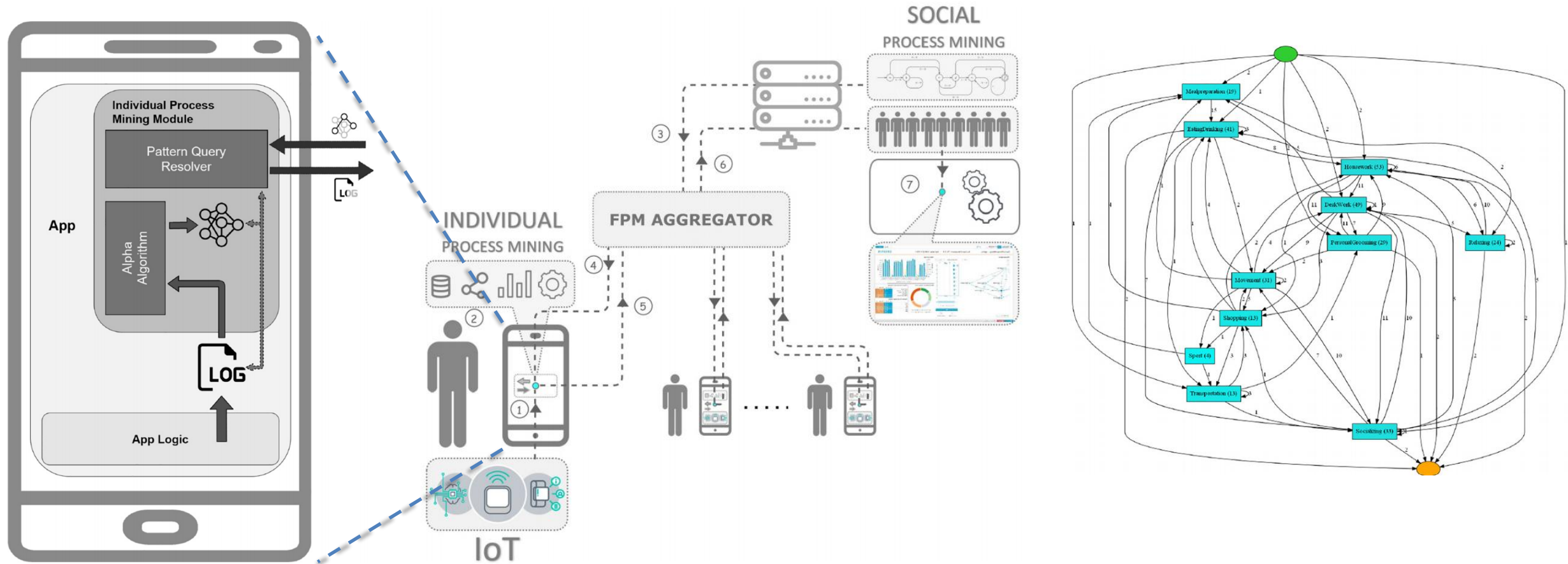
*Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays
Sean Augenstein, Hubert Eichner, Chloé Kiddon, Daniel Ramage*

Google LLC,
Mountain View, CA, U.S.A.

{harda, kanishkarao, mathews, swaroopram, fsb
saugenst, huberte, loeki, dramage}@google.com



Federated Process Mining – Human Behavior

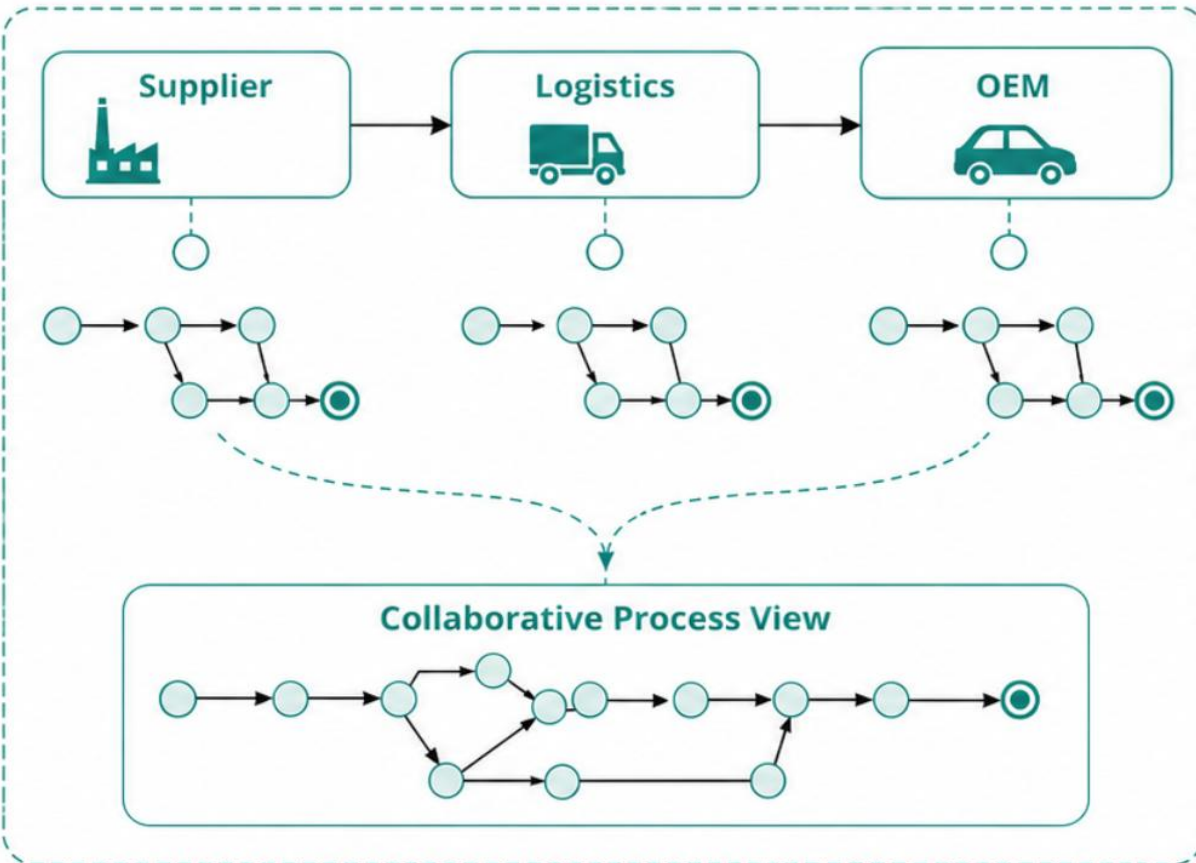


J. Rojo, J. Garcia-Alonso, J. Berrocal et al. (2022), SOWCompact: A federated process mining method for social workflows, Information Sciences

Federated Process Mining – Supply Chain

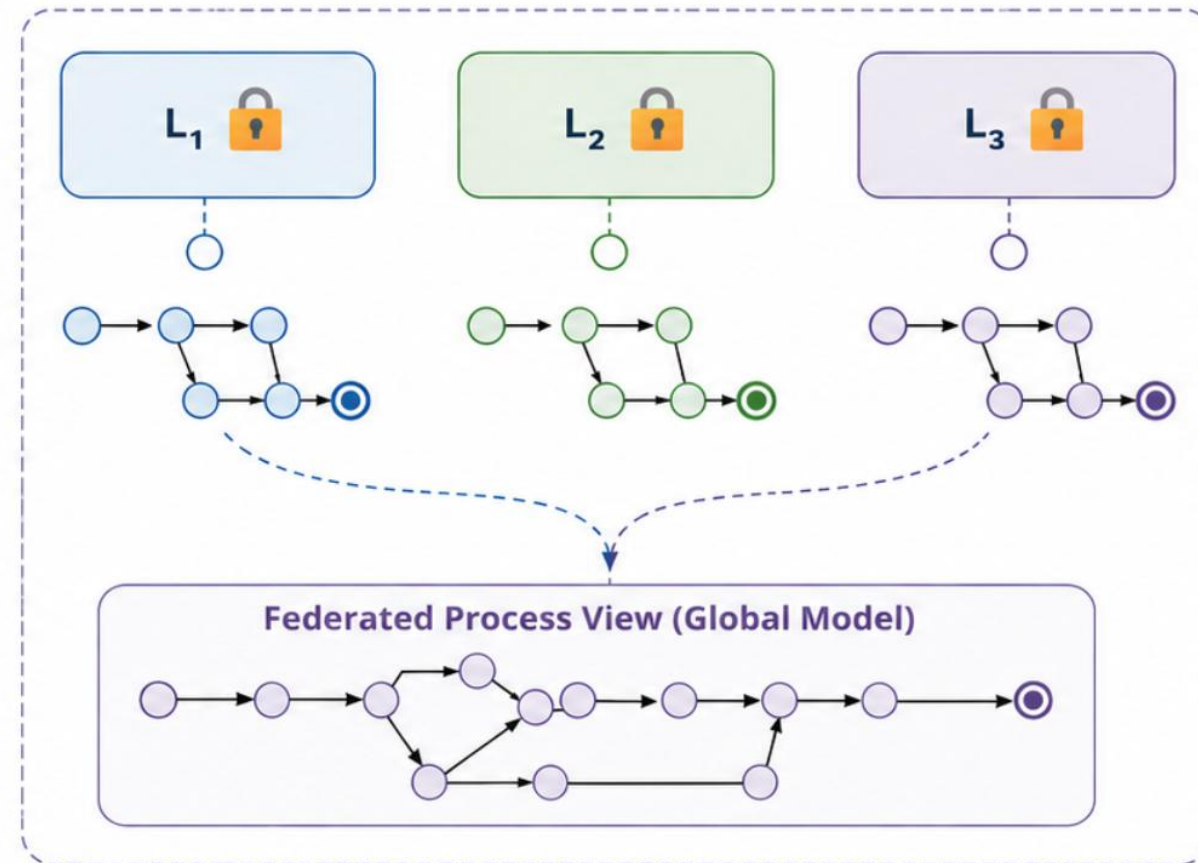
COLLABORATIVE PM

The process is executed collaboratively across organizations.



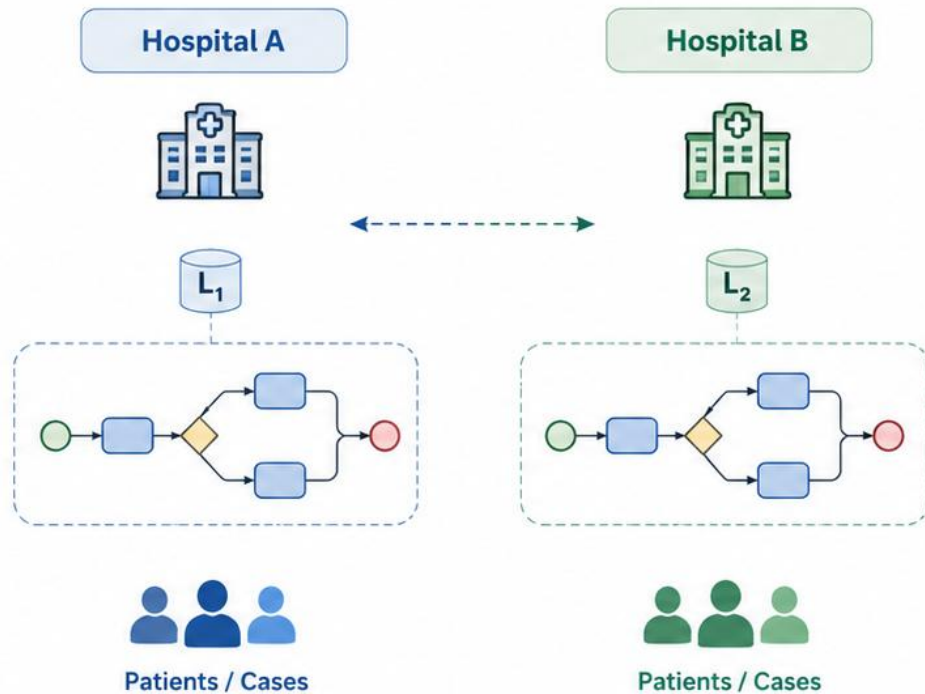
FEDERATED PM

Event data remains local but we analyze the process together.



Federated Process Mining – Data Distribution Types

HORIZONTAL

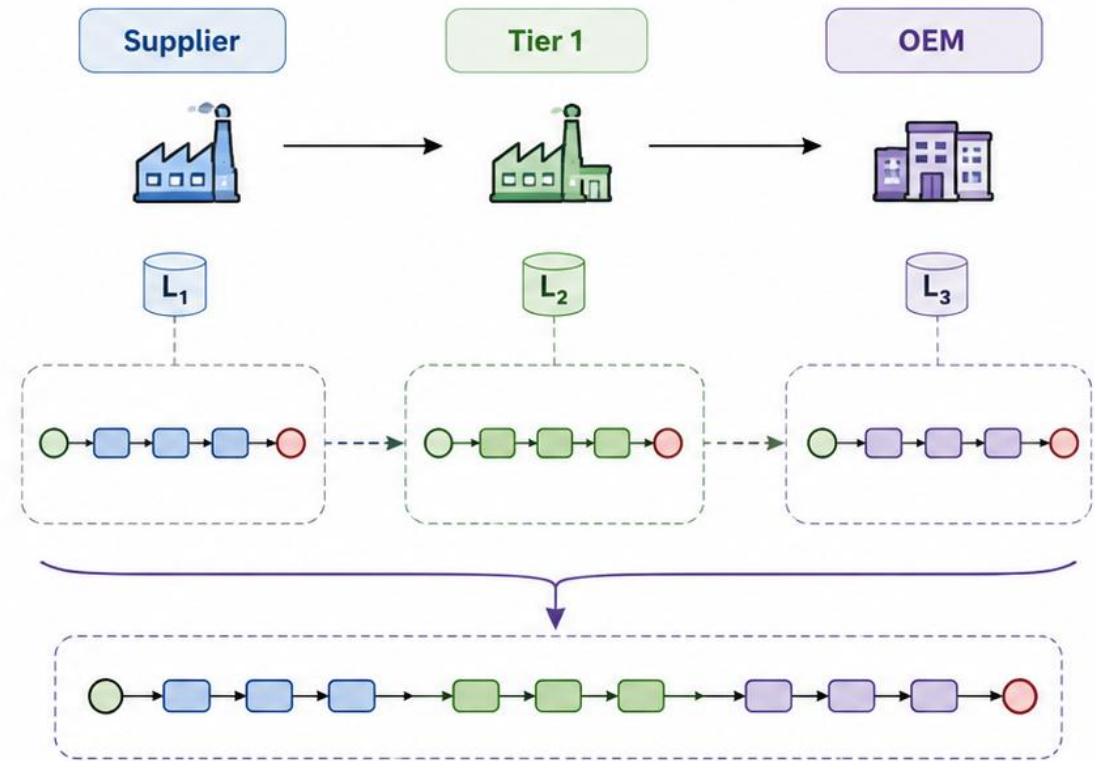


Same / similar process



Different cohorts

VERTICAL



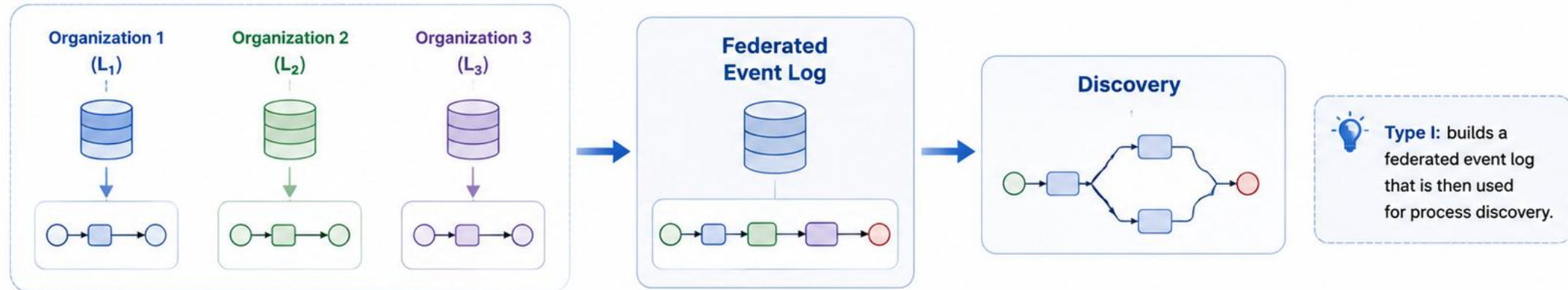
One end-to-end process



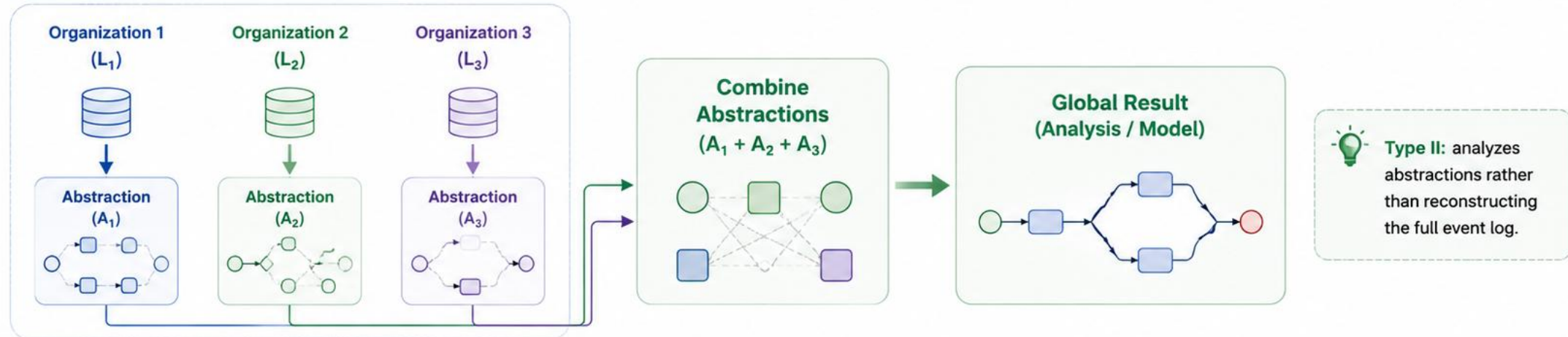
Distributed stages

Federated Process Mining – Federation Types

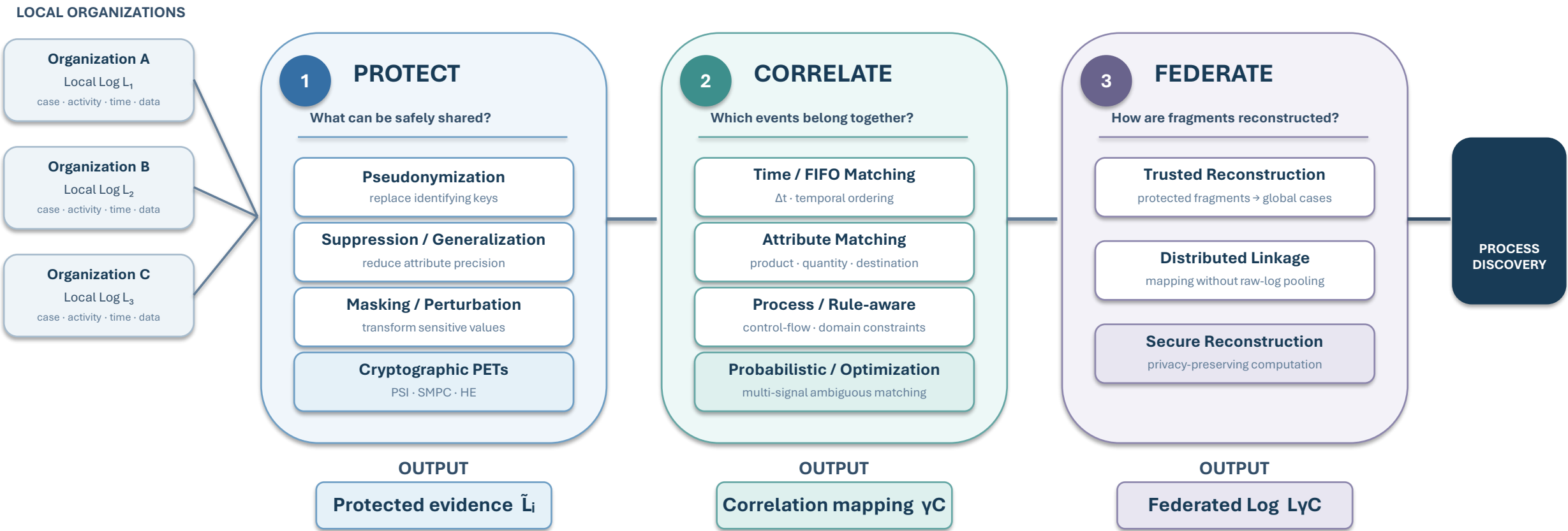
Type I Federated event log is constructed and used for discovery.



Type II Local logs are abstracted. Abstractions are combined for analysis.



Protect - Correlate - Federate



Protect - Suppression

Case	Activity	Partner	Destination
S101	Ship	Customer X	Busan
S102	Ship	Customer Y	Ulsan
↓	↓	↓	↓
S101	Ship	[REMOVED]	Busan
S102	Ship	[REMOVED]	Ulsan

Remove information that is not necessary for the collaborative task.

Protect - Generalization

Original	Generalized
10:07:32	10:00–11:00
Busan Port Terminal 3	Busan
Product: P31-2026-A	Product family P31
Quantity :203	Quantity 200–250

More general values can reduce disclosure—but also reduce correlation precision.

Protect - Masking and pseudonymization

Masking

Customer X → C***** X
Order 123456 → 12*****

Pseudonymization

Order 123456 → X7F9A
Local identifier is replaced by a token.

Important: replacing identifiers does not automatically create a shared cross-organizational identifier.

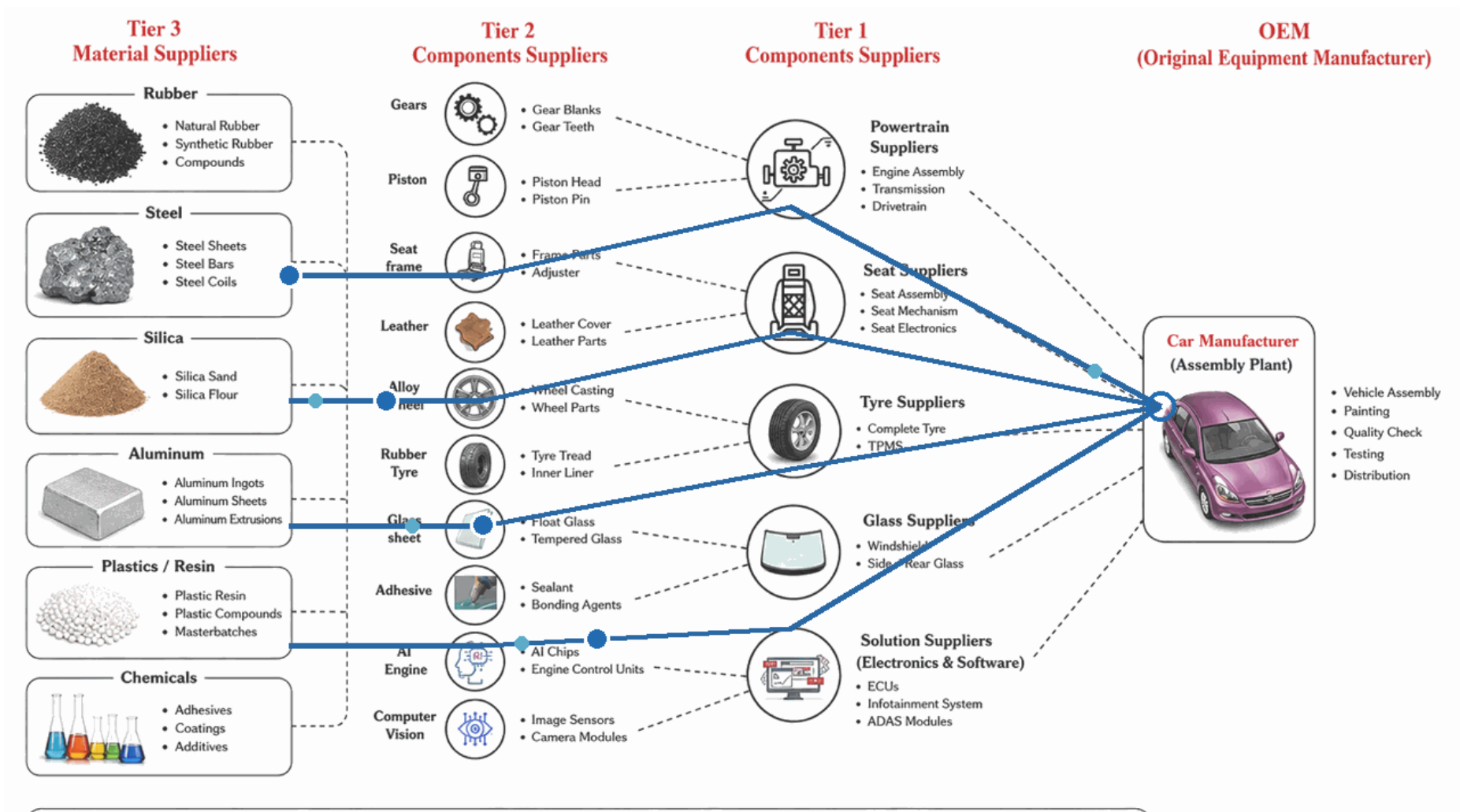
Protect - Swapping

Case	Product	Destination
S101	P31	Busan
S102	P44	Ulsan
↓ swap selected values	↓	↓
S101	P31	Ulsan
S102	P44	Busan

Preserve some distributions while breaking direct associations.

Swapping may preserve aggregate statistics but distort case-level process relations. It is therefore unsuitable when accurate event correlation is required.

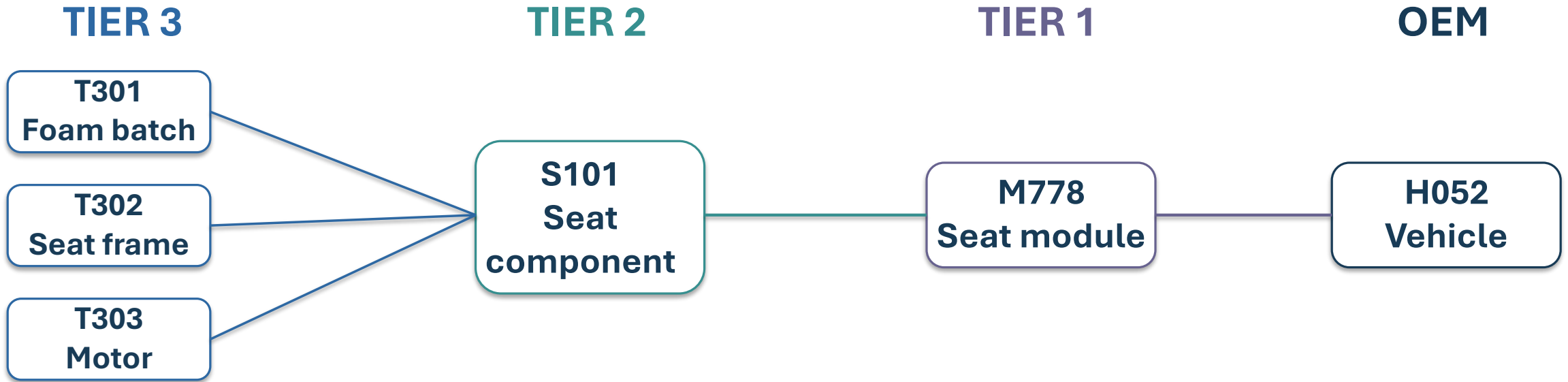
Cross-Organizational Event Correlation



No single organization observes the complete end-to-end process.

Why case correlation is difficult

Local case IDs may refer to different objects at different production stages.



T301 ≠ S101 ≠ M778 ≠ H052 — and that is expected.

The correlation problem is to infer which local cases/events contribute to the same end-to-end process lineage.

Correlation evidence may come from time, attributes, process rules, or multiple signals.

Time / FIFO Matching: a transparent baseline



Why use it?

- Simple and explainable
- Requires minimal shared evidence
- Useful as a research baseline

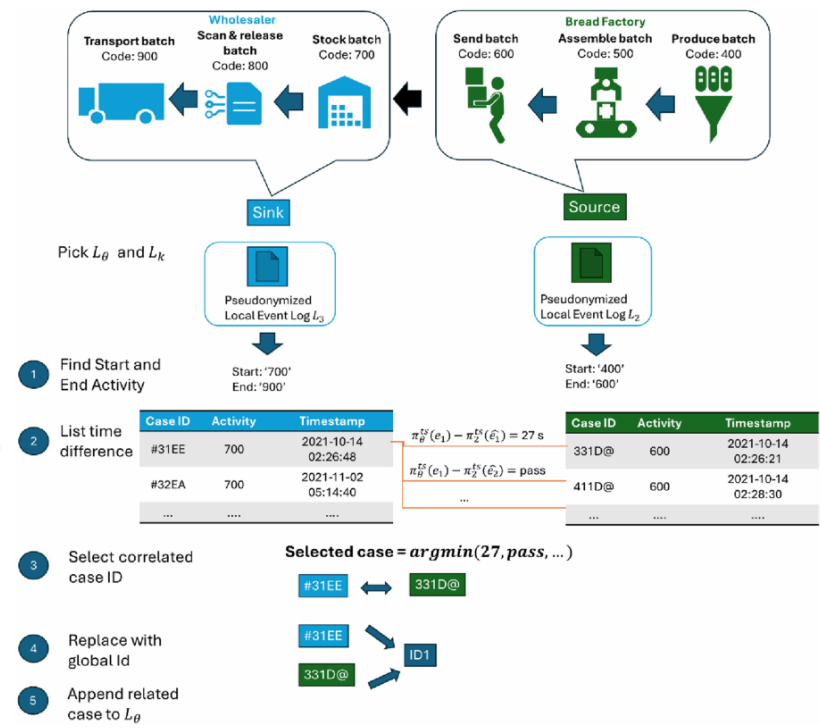
Assumption: temporal order is a useful proxy for cross-organizational linkage.

Limitation: overtaking, batching, parallel transport, rework, and variable delays can violate the ordering assumption.

Research progression:

FIFO baseline → attribute/rule-aware matching → probabilistic or optimization-based correlation.

Event Correlation with FIFO



2

List time difference

Case ID	Activity	Timestamp
#31EE	700	2021-10-14 02:26:48
#32EA	700	2021-11-02 05:14:40
...	...	...

$\pi_\theta^{ts}(e_1) - \pi_2^{ts}(\hat{e}_1) = 27\text{ s}$

$\pi_\theta^{ts}(e_1) - \pi_2^{ts}(\hat{e}_2) = \text{pass}$

...

Case ID	Activity	Timestamp
331D@	600	2021-10-14 02:26:21
411D@	600	2021-10-14 02:28:30
...	...	...

Beyond FIFO: event correlation is a design choice

Method	Evidence	Advantage	Limitation
Time / FIFO	timestamp	simple, interpretable	ordering assumptions
Attribute-based	product, qty., destination	stronger evidence	attribute disclosure
Process / rule	domain constraints	semantic consistency	application-specific
Probabilistic	multiple signals	handles ambiguity	calibration / complexity
Optimization	global constraints	consistent assignment	computational cost
Secure matching	protected features	privacy	crypto / communication cost

Time-based matching is a baseline—not the endpoint of event-correlation research.

From correlations to a federated event log

Supplier local log

S101
Prepare 09:00
Ship 10:00

Manufacturer local log

M778
Receive 10:12
Inspect 10:30

Logistics local log

L52
Pickup 13:30
Deliver 14:10

Protected lineage links: Linkage Level S101 ↔ M778 ↔ L52

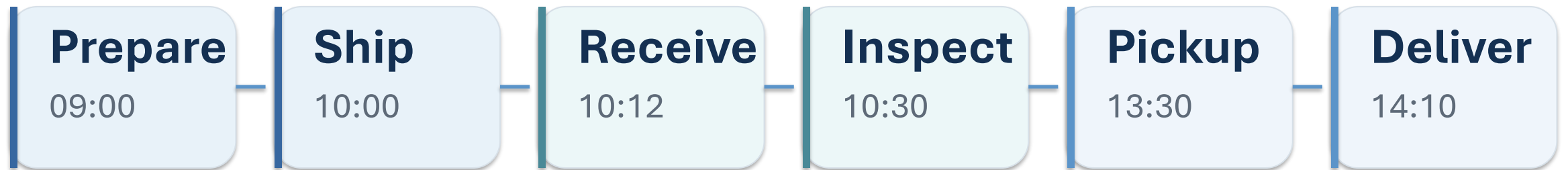
Logical representation of a federated event log

Global Case	Activity	Time	Organization	Local Ref.
G1	Prepare	09:00	Supplier	S101
G1	Ship	10:00	Supplier	S101
G1	Receive	10:12	Manufacturer	M778
G1	Inspect	10:30	Manufacturer	M778
G1	Pickup	13:30	Logistics	L52
G1	Deliver	14:10	Logistics	L52

- The table represents a logical analytical view. It does not imply that all raw records are physically pooled in one repository.
- A protected materialized view may be constructed under agreed governance, while the original local logs remain under local control.

Construct a global / federated trace

Global Case G1



Local IDs can remain local. The federated representation stores only the protected linkage needed to reconstruct an analyzable end-to-end case.

Now ordinary Process Mining can operate on the federated view

Federated Event Log — Process Discovery — End-to-End Model

Supplier

Manufacturer

Logistics

Prepare → Ship → Receive → Inspect → Pickup → Deliver

The analytical payoff is visibility across organizational boundaries.

But evaluate the reconstruction—not only the final model

Correlation quality

- Correct links
- Missed links
- False links
- Ambiguous matches

Privacy / disclosure

- What was released?
- What was generalized?
- What can be inferred?

Process quality

- Fitness
- Precision
- Generalization
- Usefulness

A plausible process model built from wrong correlations can still be wrong.

Framework

Framework	What it is	Role in an FPM prototype
Flower	Federated learning framework	Client / server coordination, secure aggregation
FedML	Federated learning platform	Communication, experiment orchestration
PyTorch distributed / custom	Low-level distributed computation	Custom local computation & aggregation
PM4Py	Process-mining library	Local discovery / conformance — the PM component, not federation infra

Takeaways - Summary

The next frontier is not only federated event logs, but federated process evidence.

1

A complete and correctly correlated event log cannot always be assumed.

2

Federation is needed when cross-boundary insight is required but full centralization is unacceptable.

3

Vertical FPM requires protection, correlation, and reconstruction—not aggregation alone.

Takeaways - Open Research Opportunities

Federated Object-Centric Process Mining

Privacy-aware event abstraction

Uncertainty-aware correlation

Distributed event knowledge graphs

Adaptive edge / cloud Process Mining

Privacy–utility–fidelity trade-offs

These are research opportunities, not established methods.

Protect locally.

Correlate securely.

Federate globally.

Discover collaboratively.

Federated Event Log Generation for Process Discovery across Organizational Boundaries

Thank you

Bernardo Nugroho Yahya
Industrial & Management Engineering
Hankuk University of Foreign Studies, South Korea

Email: bernardo@hufs.ac.kr



한국외국어대학교
HANKUK UNIVERSITY OF FOREIGN STUDIES