Modularized Anonymization of Mobility Data

Anonymity — **Data Quality**
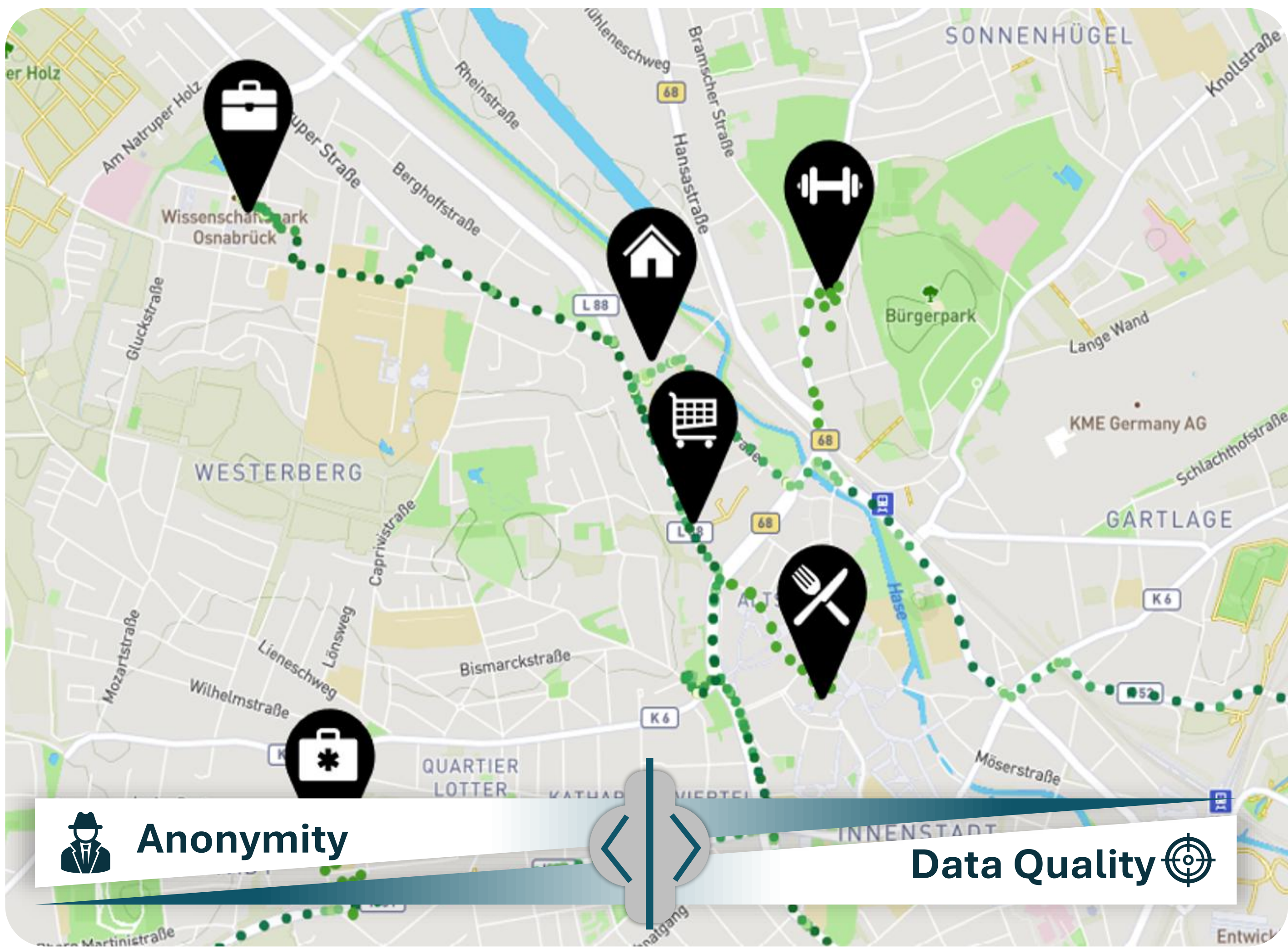
Privacy and Mobility Data

A Trip is Worth a 1000 Words

Personal movement data contains sensitive information that is not relevant for traffic analysis, such as :

- Place of residence and work
- Places of residence of acquaintances, friends, and relatives
- Shopping behavior
- Means of transportation
- Doctor visits (medical history)

This is why purpose-specific data reduction and anonymization are necessary. This provides results such as:

- ten people traveled from origin A to destination B

Anonymization

ⓘ While data reduction strengthens anonymity, it can also significantly degrade data quality and, in some cases, render the data unusable. Managing this trade-off is essential.

Anonymization Toolkit
Enabling Modular Usage

Problem

- Anonymization depends on the use case and can be diverse and multi-stage
- Processes are often complex, opaque and hidden in algorithms
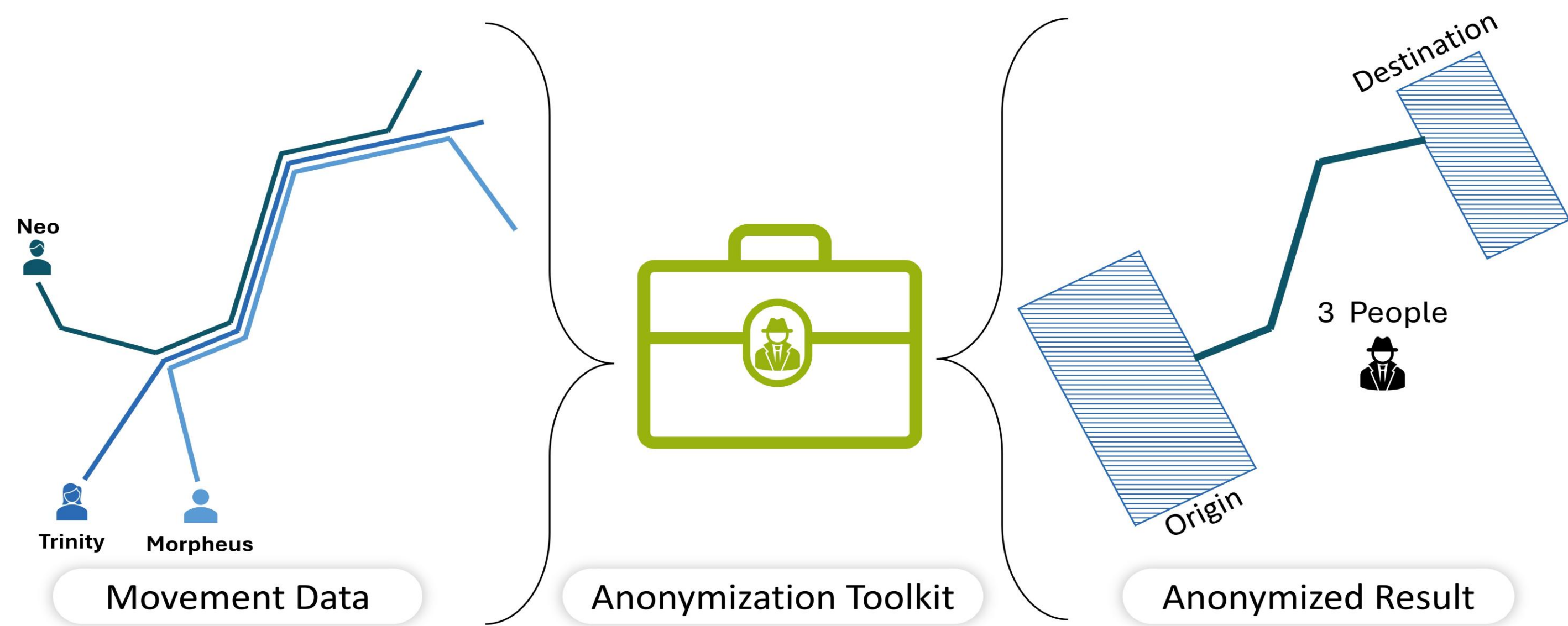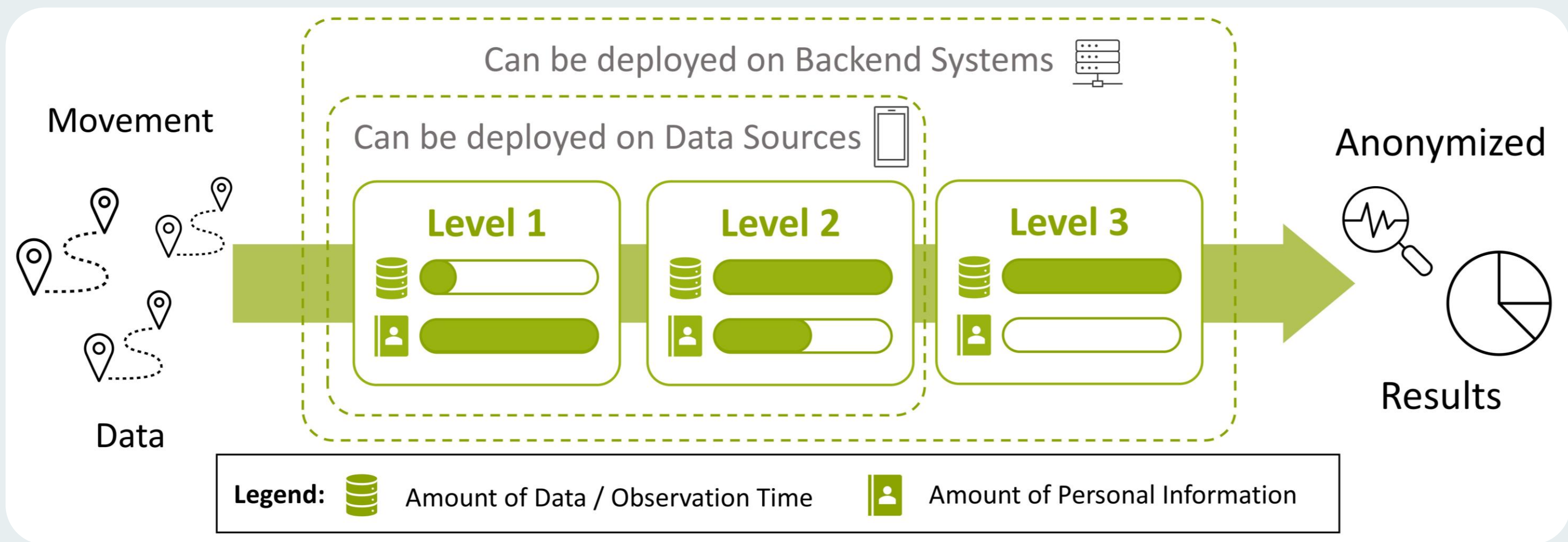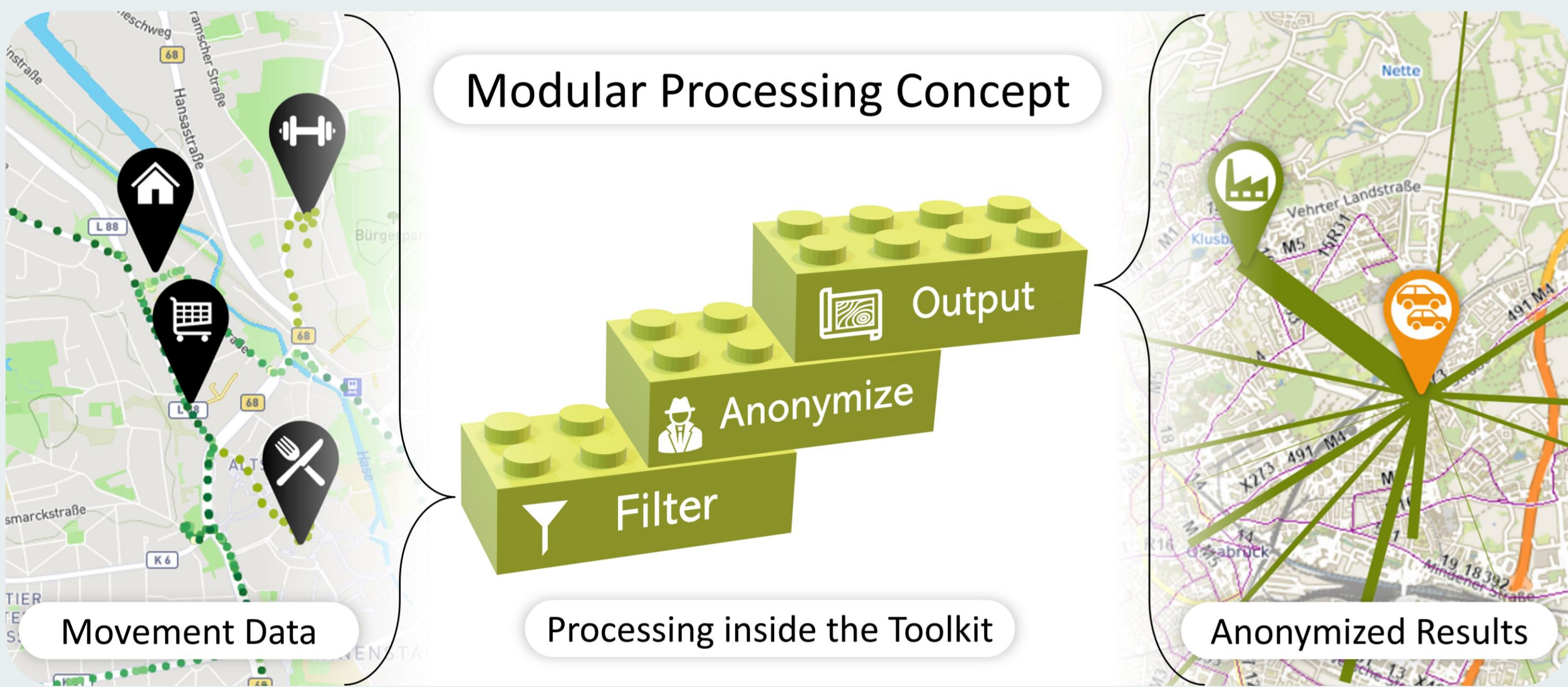- Mobility data ("trips"/"trajectories") should be anonymized close to the source

Solution

- Modular Anonymization Toolkit
- Tools and evaluation methods as modules for recurring anonymization techniques

Architecture

- Module coupling concept enables uniform, system- and language-independent communication
- Modules are portable and can be used on different platforms
- Complexity level model defines the application levels of the modules.

Level 1: Single-trip operations (e.g., filtering)

Level 2: Multiple trips by a single person (e.g., pattern analysis)

Level 3: Multiple trips by multiple people (e.g., k-anonymization)

Modules of levels 1 and 2 can be deployed on the edge to be applied close to the data source before any data transfer.

Modules of Level 3 group and aggregate data to generate anonymized datasets.

Modular Processing Concept

Movement Data — Processing inside the Toolkit — Anonymized Results

Filter → Anonymize → Output

Movement Data — Can be deployed on Backend Systems — Can be deployed on Data Sources — Level 1 — Level 2 — Level 3 — Anonymized Results

Legend: 🗄 Amount of Data / Observation Time   👤 Amount of Personal Information

Processing in Modular Flows
Step by Step, Modular per Use Case

Node-RED

- The anonymization of data is a complex, multi-step process that requires combining different methods depending on the use case
- The Anonymization toolkit is built for users with varying levels of technical knowledge and designed to be as transparent and understandable as possible
- Graphical flow editor based on the low-code tool Node-RED
  - Anonymization methods as individual building blocks
  - Users can select, string together, and connect them
  - The result is a so-called "flow" → a sequence of anonymization steps

- Modules generate an intermediate datasets for before-and-after comparisons and further analysis in geodata visualization tools
  - Simplified parameterization of individual modules
  - Increased comprehensibility of the effects of individual modules on the data at hand

- Additional traffic-specific modules:
  - District Mapping
  - Time Slot Mapping
  - Common Places Analysis/Filtering
  - Differential Privacy
  - K-Anonymity
  ...

Movement Data — Anonymization Toolkit — Anonymized Result

Publications

- Christian Fries, Jan Kettler, Thorsten Paßfeld, Ralf Tönjes and Christian Kray, „Quantifying the Impact of k-Anonymizing Trajectory Data on Common Traffic Management Use Cases", 33rd ACM SIGSPATIAL, Minneapolis, USA, 2025 (accepted)

Contact

Ralf Tönjes, Thomas Hupperich, Jan Kettler, Christian Fries
intelligent-pendeln@hs-osnabrueck.de

intelligent pendeln
www.hs-osnabrueck.de/iip