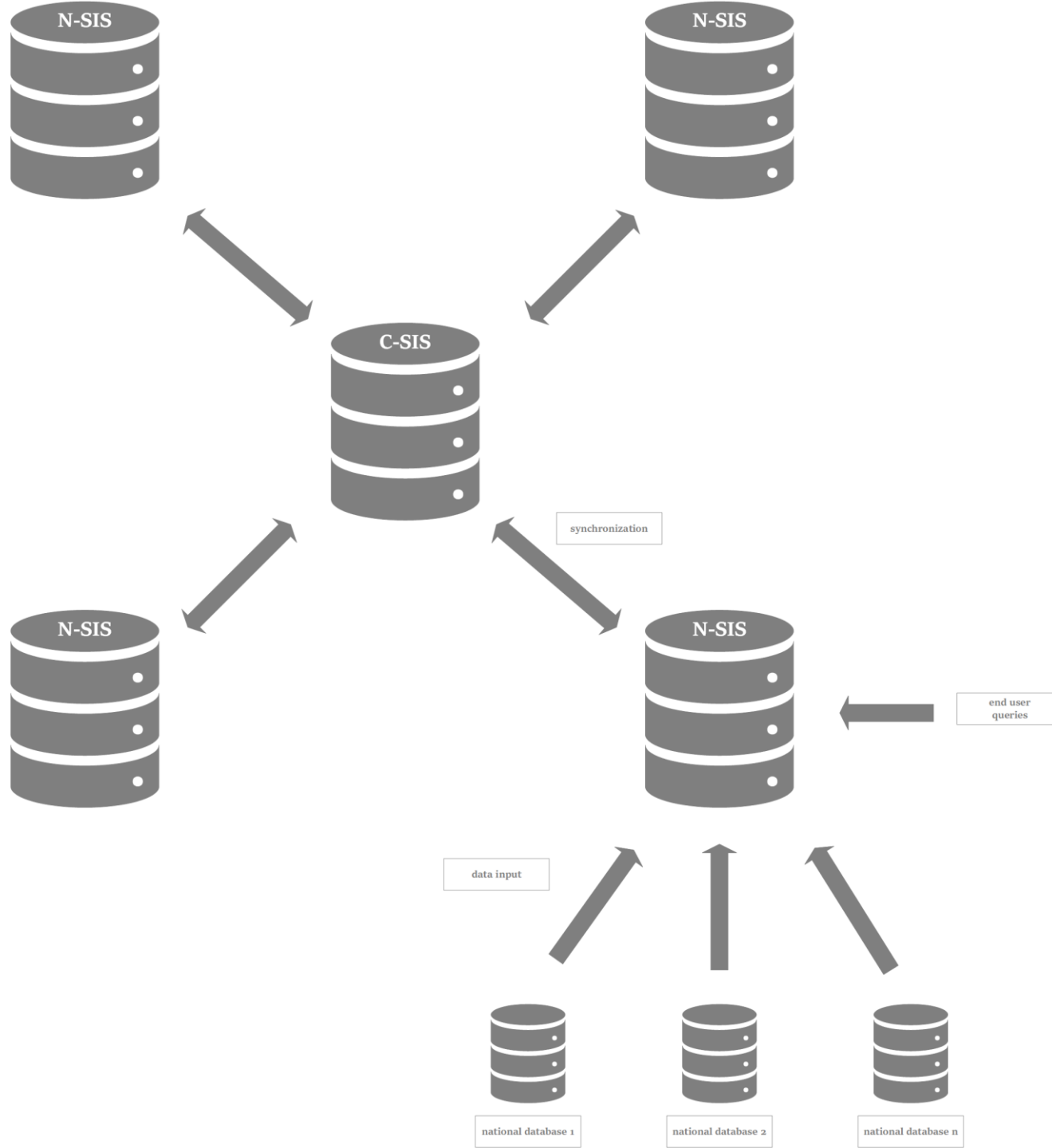


Data management at scale

Lessons from the Schengen Information System

Prof. Dr. Matthias Leese

eu-LISA Industry Roundtable, 10 September 2026



17 763 405 971

SEARCHES BY MEMBER STATES AND JHA AGENCIES IN 2025



+18%

COMPARED TO 2024



94 605 202

DATA STORED BY 31 DECEMBER 2025



+2%

COMPARED TO DECEMBER 2024



364 984

HITS REPORTED IN 2025 BY MEMBER STATES



-8%

COMPARED TO 2024

SIS 2025 statistics (source: eu-LISA)

Argument and structure



- proposition: good data management at scale means preserving the meaning and usability of data as they move across contexts
- four practical challenges:
 - 1. making data
how are real-world objects, persons, and decisions translated into structured data?
 - 2. maintaining data
how are records validated, corrected, updated, and interpreted over time?
 - 3. moving data
what happens when information travels between organisations and jurisdictions?
 - 4. using data
how do search systems, interfaces, algorithms, and user practices shape what becomes visible?

1. Making data



- producing data means to translate the real world into:
 - categories
 - fields
 - identifiers
 - classifications
- key question:
will a representation still work once it leaves the context in which it was created?
- take-away:
good data management at scale is not only about whether a record is correct – it also means thinking about whether others can retrieve, interpret, and use it

2. Maintaining data



- dealing with data issues often requires human expertise:
 - inspection of warnings
 - investigation of inconsistencies
 - interpreting apparent anomalies
- automated validation $\neq$ contextual validity
 - a formally valid record may still be misleading or outdated
 - an apparent anomaly may be perfectly legitimate once the underlying case is understood
- take-away:
good data management at scale combines technical validation with human judgement and organisational knowledge

3. Moving data



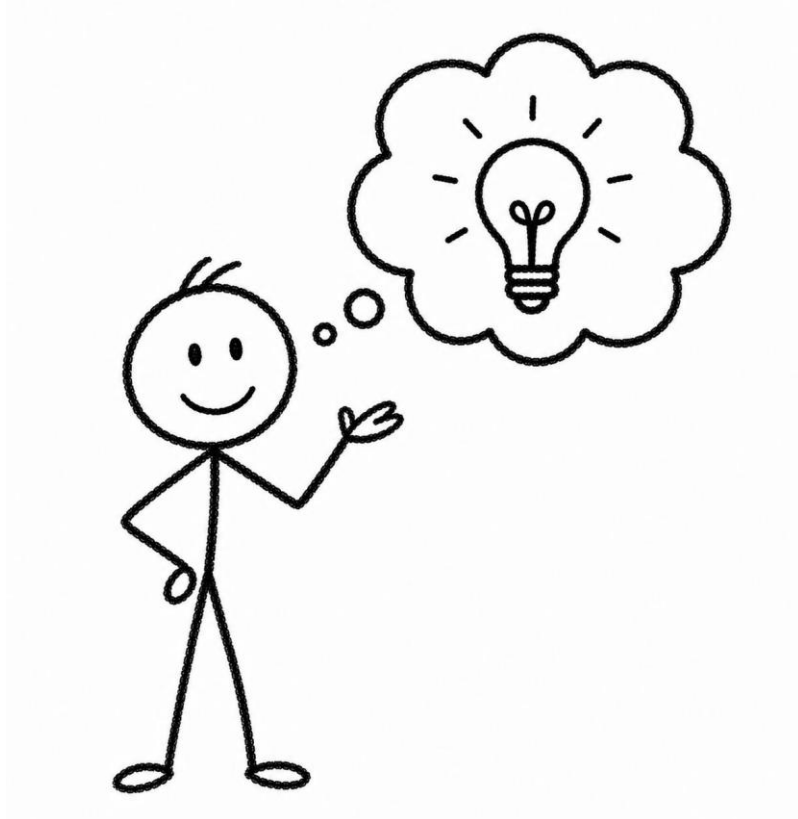
- large-scale information systems need to reconcile two principles:
 - commonality
 - subsidiarity
- technical interoperability:
can systems exchange data?
- practical interoperability:
can receiving users understand and act upon the information?
- take-away:
good data management at scale needs to preserve/create intelligibility as data move between contexts

4. Using data



- search queries determine which stored records become visible
 - users often search with incomplete or uncertain information:
- too restrictive → false negatives
 - relevant records may not be found
- too permissive → false positives
 - too many potential matches may be returned
- take-away:
good data management at scale needs to ensure findability

Implications



- use cases for AI based on proper understanding of records
- do we know:
 - where data came from?
 - which organisational practices produced them?
 - how categories changed over time?
 - where national or organisational variation exists?
 - which forms of uncertainty remain embedded in the data?
- AI cannot reconstruct context that was not preserved
 - this makes technical, organisational, and procedural provenance increasingly important
- AI readiness should therefore be understood as a stress test for data management at scale