Dynamic Intra- and Inter-Enterprise Collaboration Using an Enhanced Multidatabase Architecture

Cristian Pérez de Laborda · Christopher Popfinger · Stefan Conrad

popfinger@cs.uni-duesseldorf.de

Department of Computer Science

University of Düsseldorf, Germany

HEINRICH HEINE
UNIVERSITÄT
DÜSSELDORF

Outline

- Requirements analysis
- Architecture
- Characteristics
- Implementation
- Link Patterns
- Current and future work

Preliminaries

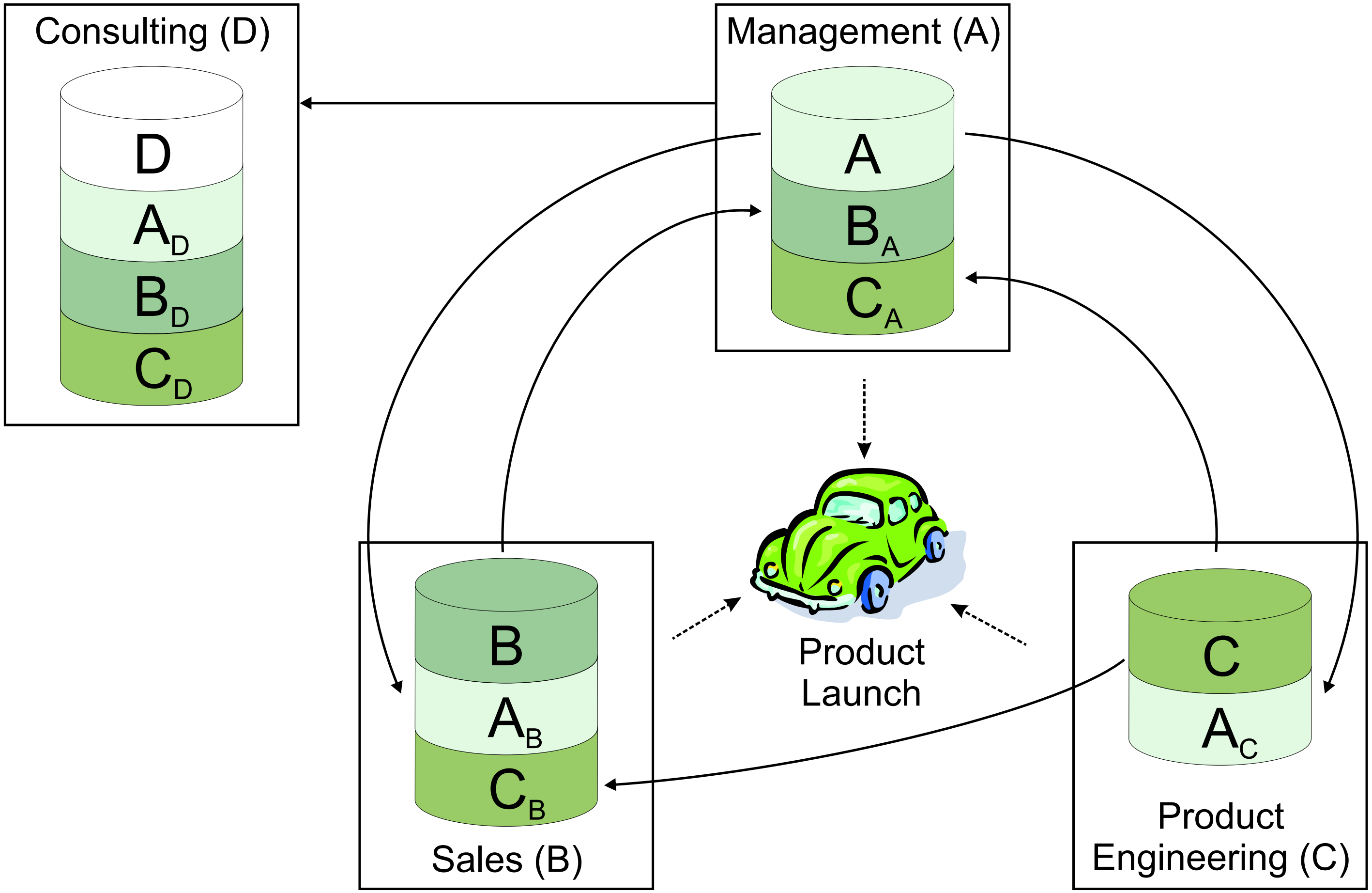
- Local data sources typically raised autonomously
- Data sources fit the special needs of the local users
- Local autonomy preserves data ownership:
  - Correctness
  - Consistency
  - Up-to-dateness
- Logical and physical heterogeneity

Intra- and Inter-Enterprise Collaboration

- Shared access to information among
  - cooperating departments of a single company
  - (temporary) collaboration of multiple companies (virtual corporations)
- Requirements
  - Integration of legacy data sources
  - Information system applicable to data policy
  - Preservation of data ownership
  - No central authority
  - *Support of volatile data sources*

⇒ Evaluation of a push-based replication strategy

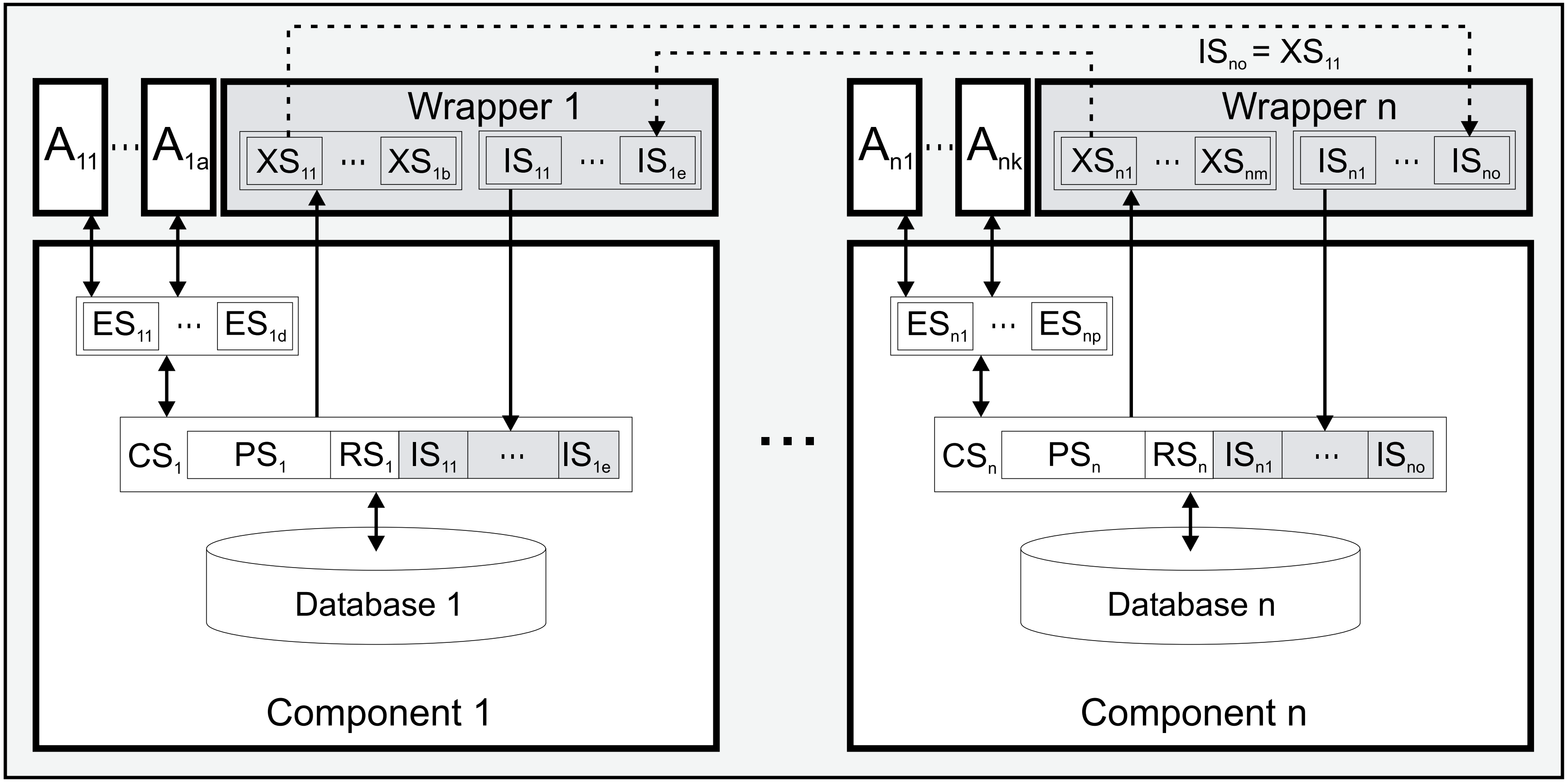
Scenario


Consulting (D)
D
A_D
B_D
C_D
Management (A)
A
B_A
C_A
Product
Launch
B
A_B
C_B
Sales (B)
C
A_C
Product
Engineering (C)

Basic Functionality

1. Administrator makes a subset of its data accessible
2. Remote peers subscribe to specific part of the data
3. Publisher creates individual delivery schedule for each subscriber
4. Initial transfer of schema and data
5. Propagation of data modifications to the subscribers according to delivery schedules
6. Subscribers can join or leave at any time

Architecture



Characteristics

- Autonomy and Heterogeneity
  - Dynamic interconnection of autonomous and heterogeneous data sources
  - Feasible trade-off between local autonomy and reasonable degree of information sharing
  - Peers can choose level of participation (which data is imported/exported)
  - Specific arrangements among partners required
- No central authority
  - No centralized data storage
  - No central event broker (publish/subscribe systems)
  - No single-point-of-failure

Characteristics (II)

- Wrapper organized as P2P system
  - Data exchange exclusively negotiated pairwise
  - Network of self-responsible peers
  - Volatile peers
- Standardized Exchange Format: Common representation of heterogeneous data (based on OWL)
- Local integration
  - No global schema
  - Individual (local) integration strategies
  - Definition of integrity constraints and index structures on imported data

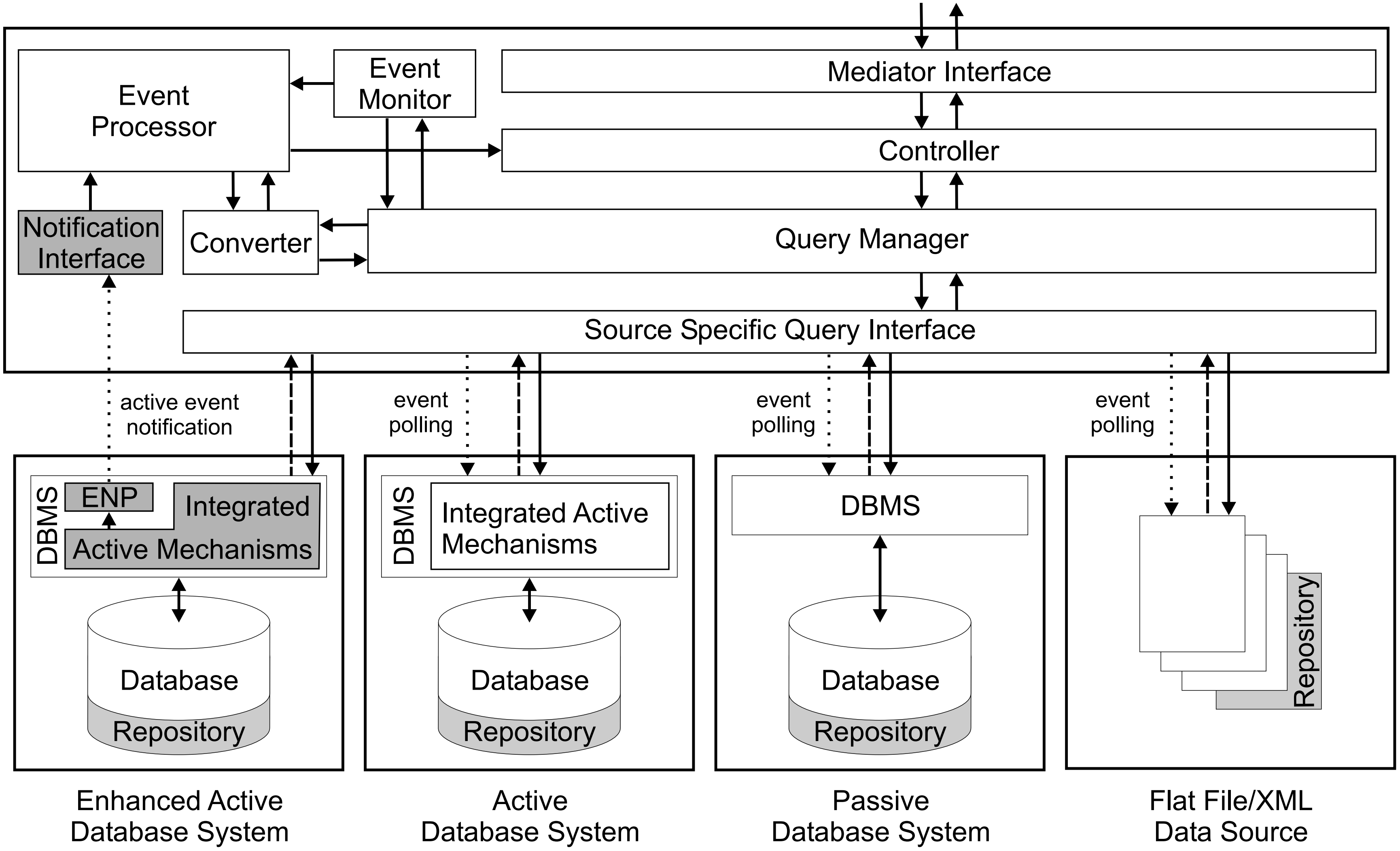
Characteristics (III)

- Replication
  - Improving data availability and query performance
  - Lazy single master replication: one master and multiple *read-only* replicas
  - Individual refreshment strategies: immediate, periodical, aggregated,...
  - Export of imported data
    - Direct updates: Updates are replicated from the master to the slaves directly
    - Cascading updates: Updates are replicated from an imported replica on a slave to another slave

Characteristics (IV)

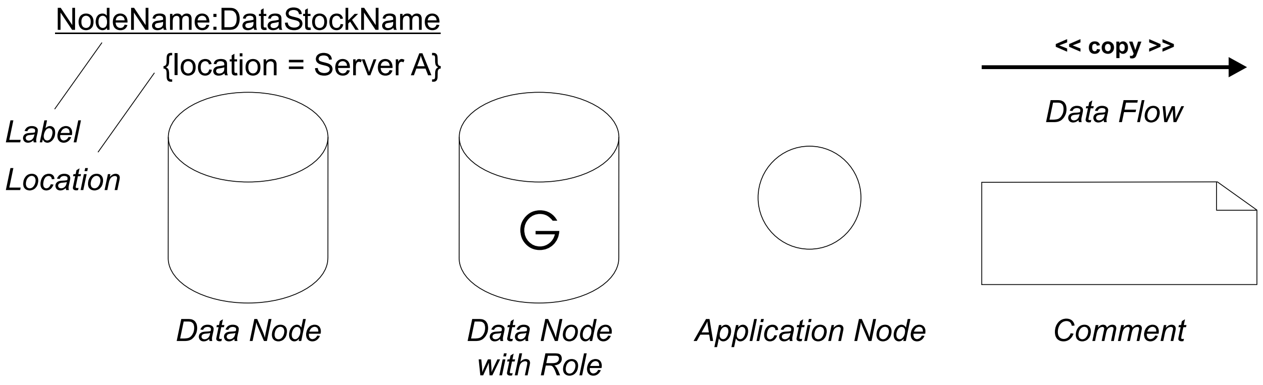
- Push-based protocol
  - Publish/Subscribe concepts
  - Push-based propagation of data (and schema) modifications to subscribers
  - Propagation queues on publisher:
    - Individual queues for each subscriber
    - Resend updates after network breakdown or failure of the target system
  - Wrapper with event detection subsystem

Wrapper Component



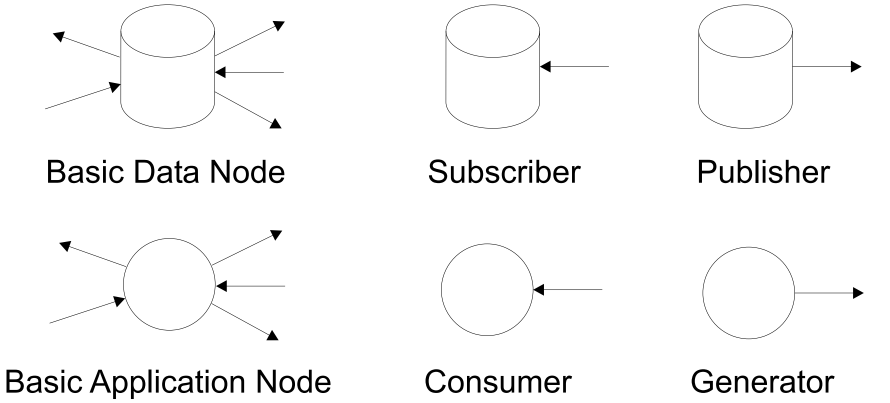
Link Patterns

- Composition of information platform either planned or evolved dynamically
- *Link Patterns* for the description and modeling of data flows in the information system
- Based on UML syntax



Link Patterns (II)

Elementary Link Patterns



Basic Data Node

Subscriber

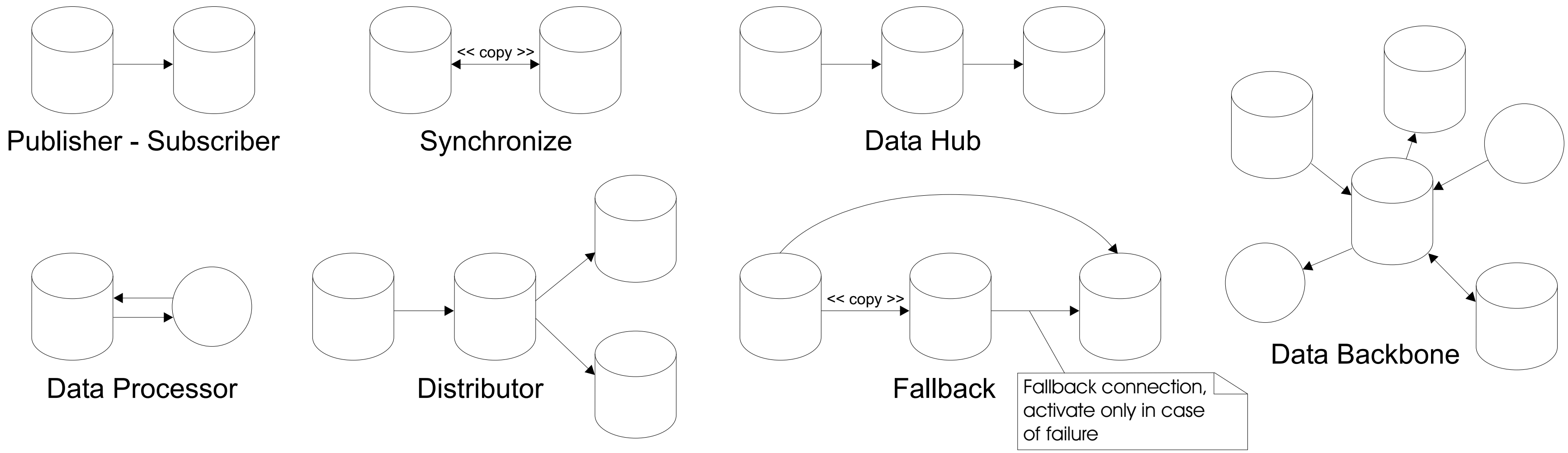
Publisher

Basic Application Node

Consumer

Generator

Data Independent Link Patterns



Publisher - Subscriber

Synchronize

Data Hub

Data Backbone

Data Processor

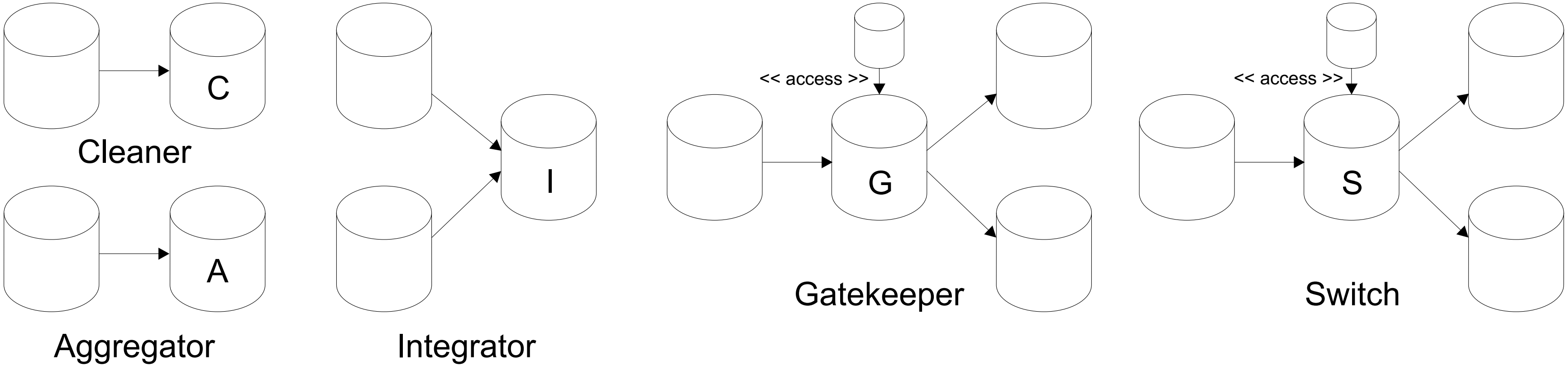
Distributor

Fallback

Fallback connection, activate only in case of failure

Link Patterns (III)

Data Sensitive Link Patterns



Current and Future Work

Current Work

- Implementation of wrapper component for relational sources
- Definition and evaluation of several exchange formats
- Analysis of network traffic and scalability

Future Work

- Support of semi-structured and unstructured sources
- Access Control and Security
- Compression
- Tools for administration, monitoring, and performance analysis

Thank you for your attention!

popfinger@cs.uni-duesseldorf.de

HEINRICH HEINE
UNIVERSITÄT
DÜSSELDORF