TECHNISCHE
UNIVERSITÄT
DARMSTADT

Dealing with Heterogeneous Data in Publish/Subscribe Systems: The Concept-Based Approach

Mariano Cilia

Databases and Distributed Systems Group

Traditional Publish/Subscribe

- Provides asynchronous communications
- Dynamic number of producers and consumers
- It naturally decouples consumer and producer
  - making them anonymous to each other
- Communication through a mediator (NS)
- Event/notification definition
  - global data structure (global repository)

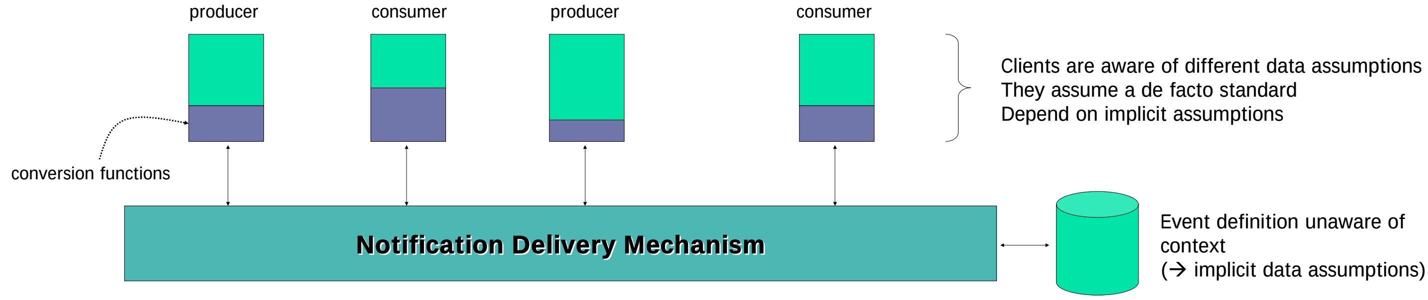
Pub/Sub = data exchange among loosely-coupled applications

But an homogeneous data context is globally assumed, which is a unrealistic assumption for open loosely-coupled systems

Data Exchange Issues

- Data from different sources/components is represented differently
  - different organizations/departments use different units and representation formats
- Context information is usually left implicit and consequently it is lost when crossing component or institutional boundaries
  - (date) 7/11/2003 Which one is the month?
  - (price) 200 Currency? €?, U$S?...
- Data from different apps needs to be interpreted by applications
  - no cultural assumption!

➔ To process events in a semantically meaningful way, explicit information about semantics of data is required



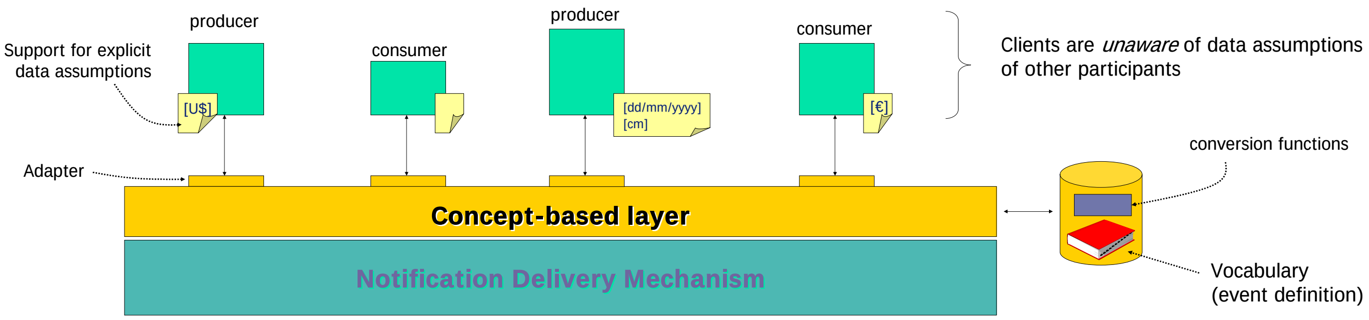
Clients are aware of different data assumptions
They assume a de facto standard
Depend on implicit assumptions

Event definition unaware of context
(→ implicit data assumptions)

The Concept-based Pub/Sub Approach

- Provide a higher level of abstraction to describe the interests of publishers and subscribers
- Events represented by using Ontologies (common vocabulary)
  - common interpretation basis for data and events
  - organized as infrastructure- and domain-specific ontologies
- Subscribers and Publishers can specify their assumptions
  - *Price* < 100 [€]
  - *DeliveryDate* := 7/11/2003 [dd/mm/yyyy]
- Allow Ontology relationships for subscriptions

- This approach focuses on data integration aspects
  - does not deal with msg routing strategies, transactions, QoS, etc which are delegated to the underlying delivery mechanism
- Data integration
  - implicit assumptions are made explicit (semantic context)
  - conversion functions are now part of the infrastructure
- The notification service delivers **ready-to-process** data to subscribers
  - no further (transformation) processing is needed



Clients are *unaware* of data assumptions of other participants

Vocabulary
(event definition)

Contributions

- Richer msg data structures
  - not restricted to flat messages
- Explicit definition of context information
- Conversion functions are part of the infrastructure
  - avoiding code scattering among participant applications
- Subscriptions now include consumer's desired context
  - automatic conversion of data according to desired context
  - delivery of ready-to-process notifications

- More powerful subscription language
  - use of ontology relationships (e.g. specialization)
- Empowers autonomy of participants
- Built as a layer that can run on top of different academic/commercial notification services
- Supports pub/sub interactions in open loosely-coupled systems