Towards a Common Description Vocabulary for Industrial Datasets

Christian Mader, Steffen Lohmann and Sören Auer

Smart Descriptions & Smarter Vocabularies (SDSVoc)

Amsterdam, November 30th 2016

Fraunhofer
Retaining Sovereignty over Data in Industrial Communication
Smart Services and Industry 4.0
A Network of Trusted Data
Goals of the Industrial Data Space (IDS)

- Security
- Data exchange

- Decentral Approach
  Distributed architecture

- Sovereignty
  over data and services

- Data Governance
  “rules of the game”

- Network
  of platforms and services

- Economies of scale
  Networking effects

- Open Approach
  Neutral and user-driven

- Trust
  Certified participants
Main Components and Interactions

IDS Architecture

- Connectors provide data endpoints to access Individual resources, data dumps, service calls,…
- Broker helps finding endpoints and establish bilateral connections
- Appstore provides packaged applications for usage in Connectors
- Vocabularies capture domain-specific semantics for establishing a common language between partners
Linked Data and Vocabularies in the IDS
The Role of RDF Vocabularies and Metadata

- **IDS vocabulary** defines how to describe IDS components and data endpoints based on **Linked Data technologies**
- Ensures **interoperability** between communication partners **without a central authority**
- Drives **search and retrieval** for endpoints and data (describing data and service offerings)
- Essential for a component to be **IDS-compliant**
- Basis to establish and enforce **security**
- Metadata **accompanies** any IDS communication
Establishing secure and traceable communication

**Metadata on the Protocol Level**

- Transfer metadata and payload data simultaneously
- Protocol header contains metadata:
  - Source & Destination Endpoints
  - ID Token
  - Usage policy (access restrictions, time-to-live,...)
Establishing secure and traceable communication

Metadata on the Protocol Level
Building a Data and Service Marketplace

IDS Vocabulary for Discovery

- Metadata Vocabulary describes the connector configuration as Linked Data
- Announced to Broker on connector deployment
- Broker provides search interface (UI, SPARQL, …)
- Multiple Brokers possible
Building a Data and Service Marketplace
IDS Vocabulary for Discovery
Supporting Bilateral Communication between Partners
Leveraging Metadata for Direct Communication

- Communication works without Broker if Connector is known
- No SPOF
- Each connector runs a Web server
- Connector serves same configuration information as found on Broker
Supporting Bilateral Communication between Partners
Leveraging Metadata for Direct Communication
Supported Functionalities and Requirements

Current State of the IDS Vocabulary


- IDS Vocabulary is able to express
  - Endpoint origin and basic metadata (provider, service contact…)
  - Coverage and content (timely, spatial, topic)
  - Data formats
  - Policies (e.g., pricing, permissions, expiry)

- Aligned to existing vocabularies
  - DCTerms, DCAT, PROV-O, ODRL, DQV, IANA, …
Classes and Alignment to Existing Vocabularies
Current State of the IDS Vocabulary
Exemplary Industrial Use Case
Metadata and Vocabularies – Practical Application
Current Tasks

- Extend with **service description** capabilities
  - URIs for parameter descriptions
  - Protocol bindings (WSDL-like)
- Include **protocol header specific properties**, e.g.,
  - Message payload digest
  - Authentication token
Open Questions

Next Steps and Future Work

- Vocabularies describing communication protocols? (e.g., typical header fields)
- Vocabularies for annotating
  - Industry branch (e.g., NAICS,...)
  - Content taxonomies (e.g., financial, research,...)
  - Standards (GS1,...)
  - Algorithms (MD5,...)
- Implementation & Software Integration
  - APIs needed to read/write/modify description documents
  - rml.io, Pinto, XSPARQL,...
Towards a Common Description Vocabulary for Industrial Datasets

Christian Mader, christian.mader@iais.fraunhofer.de