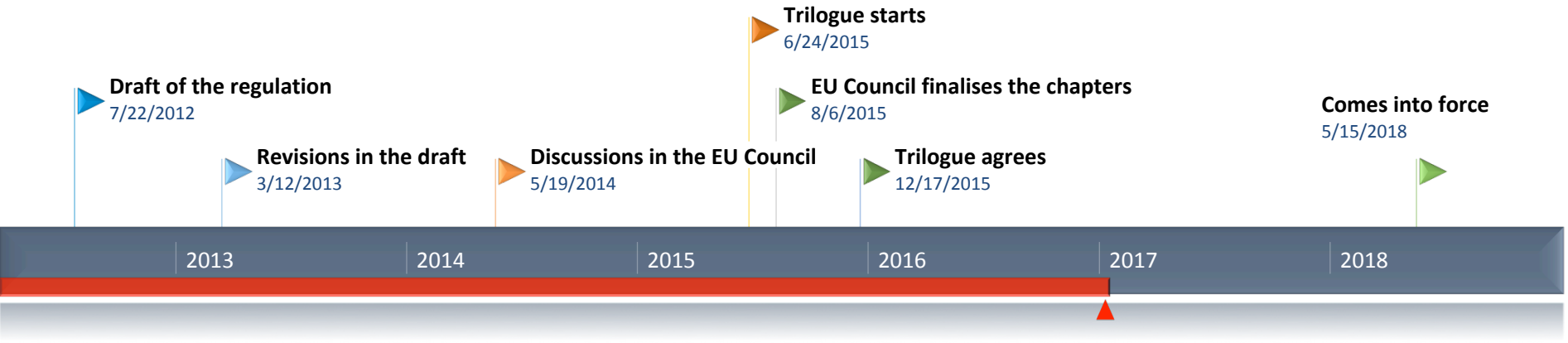
Horizon 2020
European Union funding
for Research & Innovation

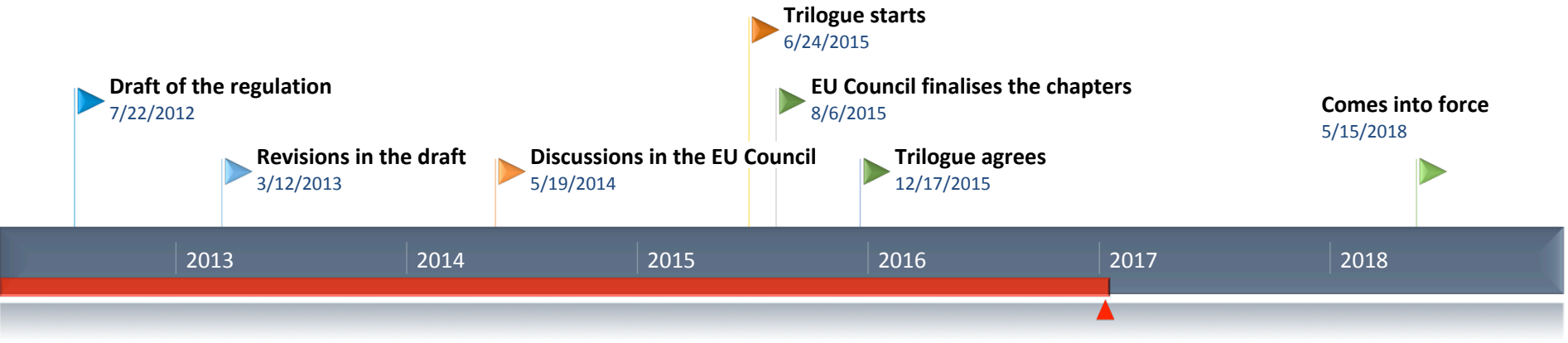
Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance

H2020-ICT-2016-1 Big Data PPP: privacy-preserving Big Data technologies (ICT-18-2016) call

Technological problem - General Data Protection Regulation supporting consent and transparency



Technological problem - General Data Protection Regulation supporting consent and transparency

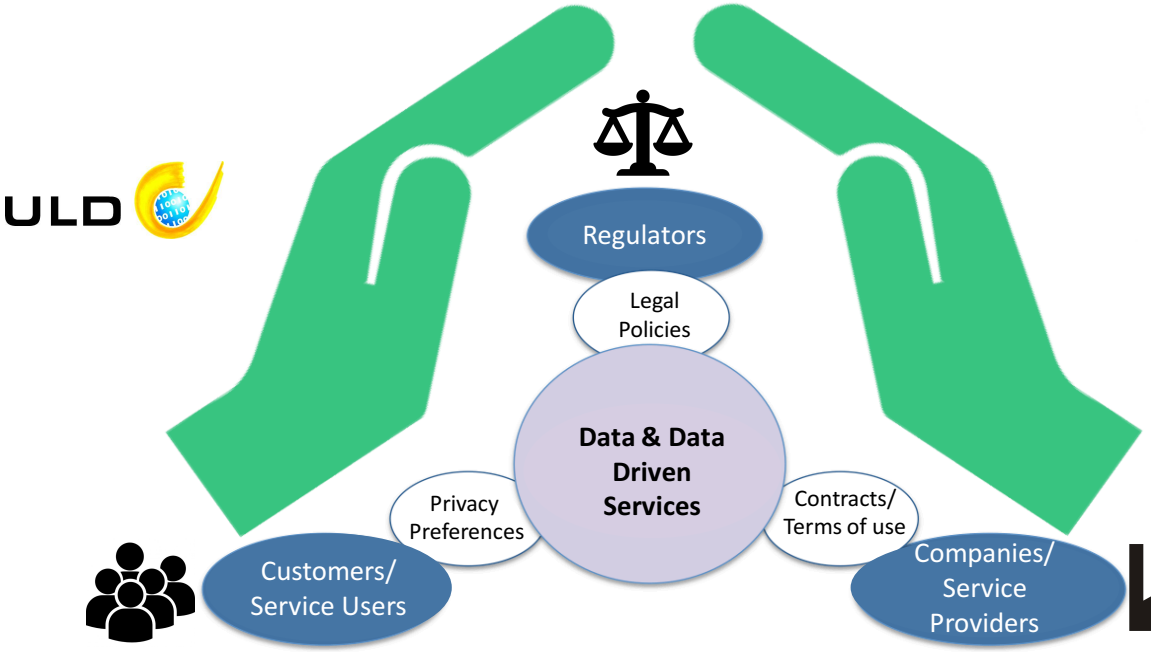


Companies whose business models rely on personal data

Data subjects who would like to declare, monitor and optionally revoke their (often not explicit) preferences on data sharing

Regulators who can leverage technical means to check compliance with the GDPR

Technological problem - General Data Protection Regulation supporting consent and transparency



ULD

the answer company™ THOMSON REUTERS®

Deutsche Telekom

proximus

W3C®

ERCIM European Research Consortium for Informatics and Mathematics

WU WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

TenForce The Pragmatic Company

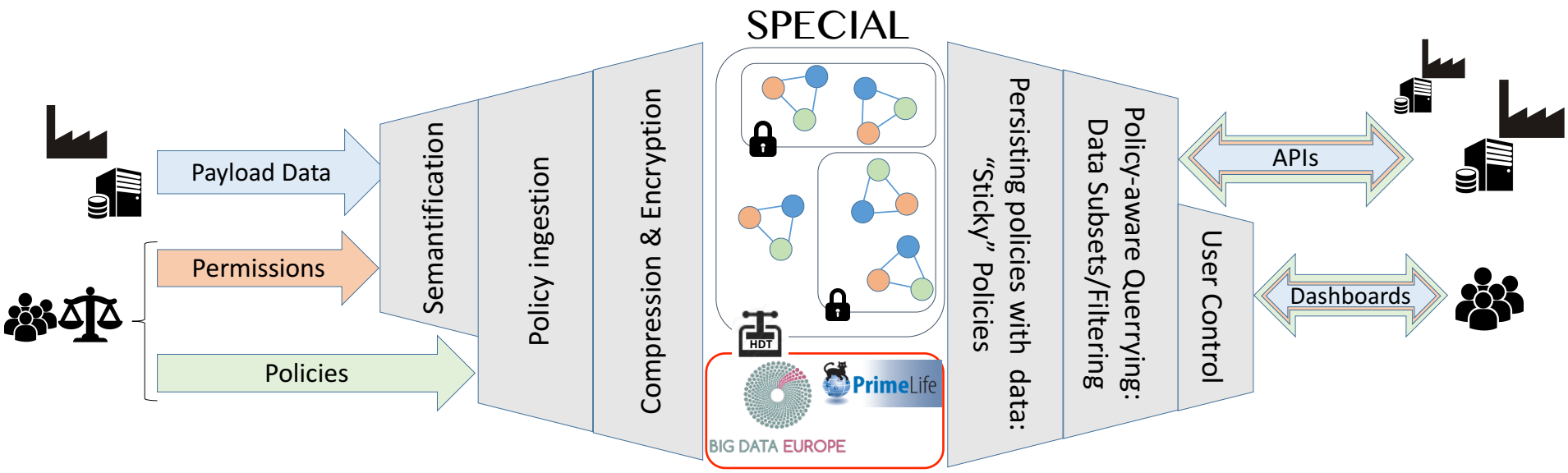
CeR ICT

Technische Universität Berlin

Technological problem - General Data Protection Regulation supporting consent and transparency

- Policy management framework
  - ❖Gives **users control** of their personal data
  - ❖Represents **access/usage policies** and **legislative requirements** in a **machine readable format**
- Transparency and compliance framework
  - ❖Provides information on how data is **processed** and with whom it is **shared**
  - ❖Allows data subjects to take **corrective action**
- Scalable policy-aware Linked Data architecture
  - ❖Build on top of the Big Data Europe (BDE) platform **scalability and elasticity mechanisms**
  - ❖Extended BDE with **robust policy, transparency and compliance protocols**

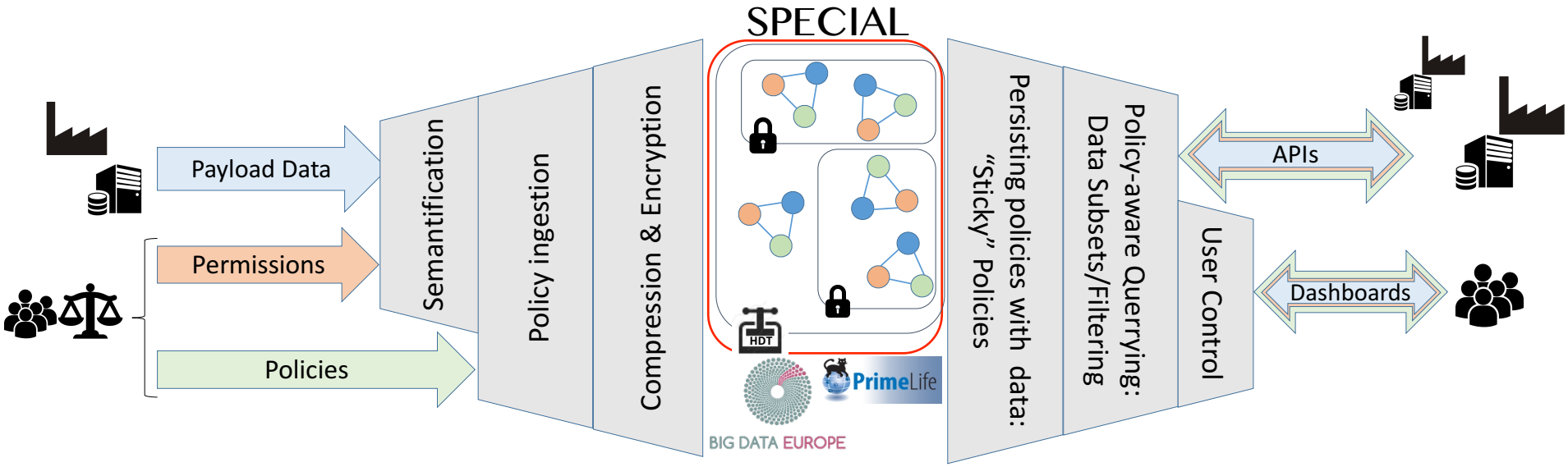
Software components - Foundations



BIG DATA EUROPE PrimeLife

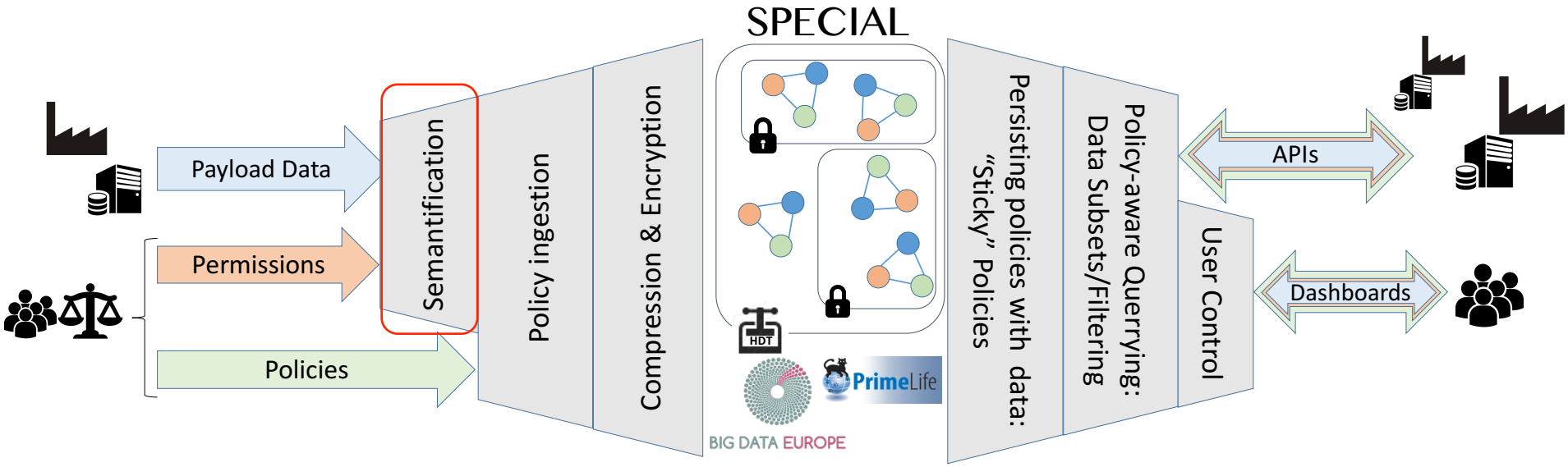
- **Big Data Europe** scalability and elasticity
- **PrimeLife** policy languages, access control policies, release policies and data handling policies

Software components - Foundations



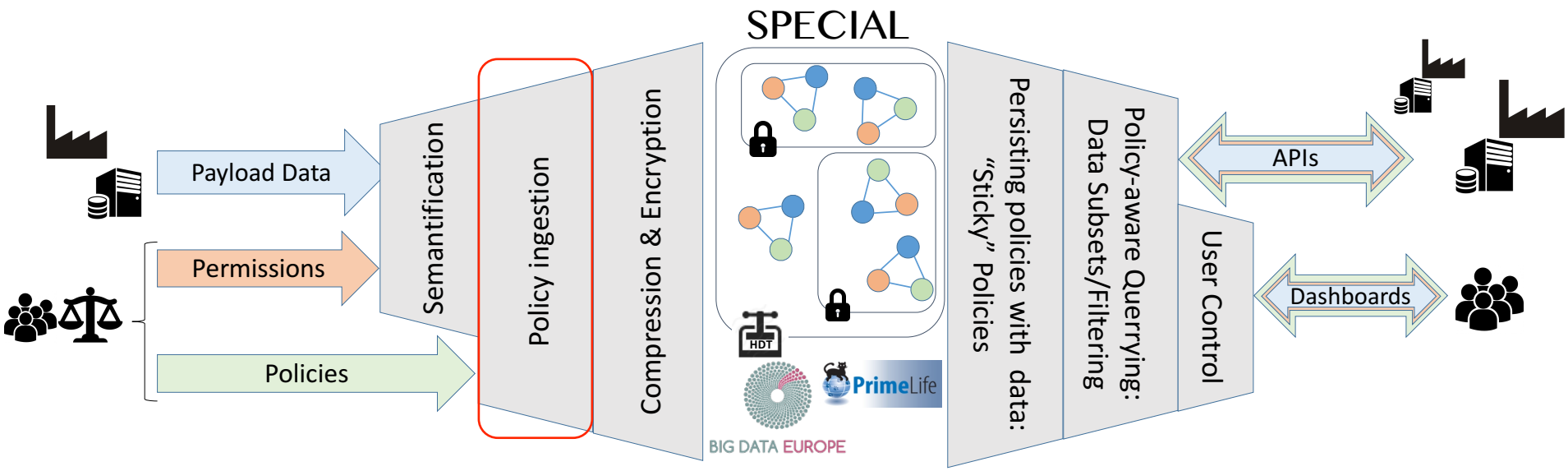
- SPECIAL uses the **Linked Data paradigm**
- All data items are identified by **globally unique identifiers** (i.e. Internationalised Resource Identifiers (IRI's))
- By using HyperText Transfer Protocol (HTTP) IRI's **everything is potentially linkable**

Software components - Semantification



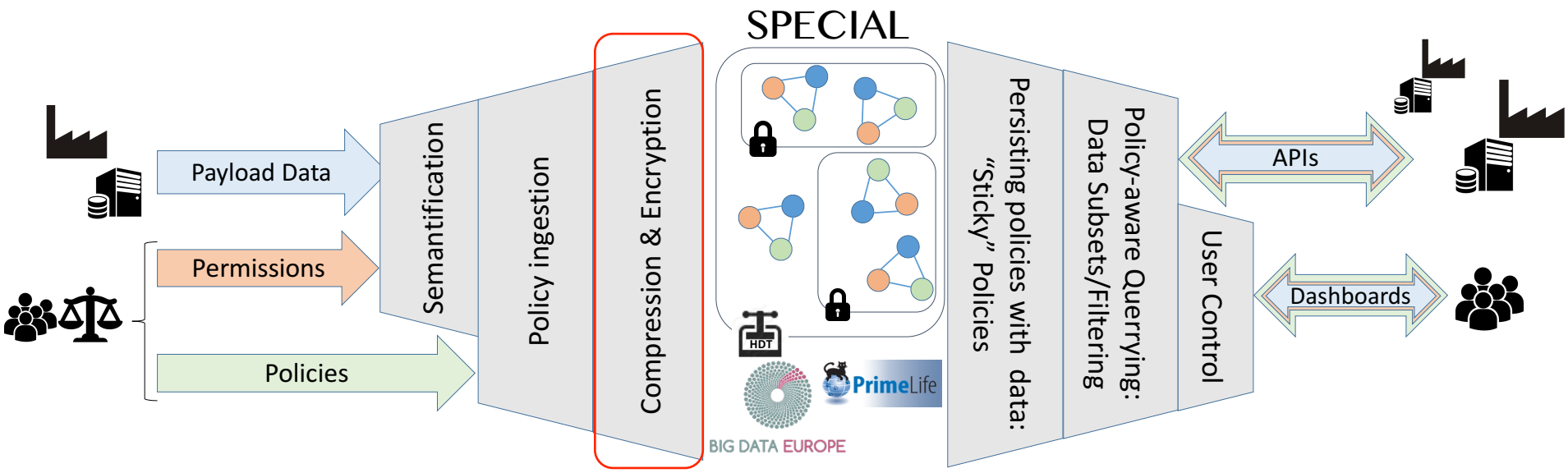
- IRI's allow SPECIAL to make semantic assertions (access/usage constraints) on the data items using **Linked Data annotations**
- Legacy systems can be integrated via **transformation middleware**

Software components - Policy Ingestion



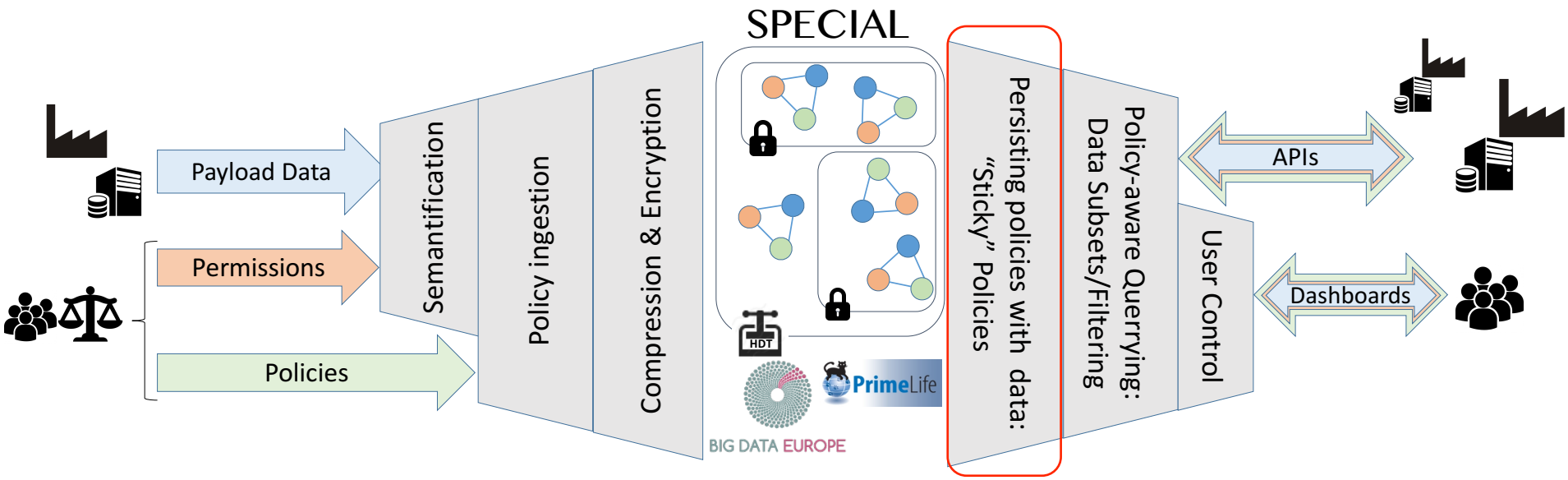
- Record **context** information and **access/usage** constraints
- Handle a broad **variety of sources and formats**
- Take a **privacy-by-design** approach and allows for conscious decisions about data collection and data (re)use

Software components - Compression & Encryption



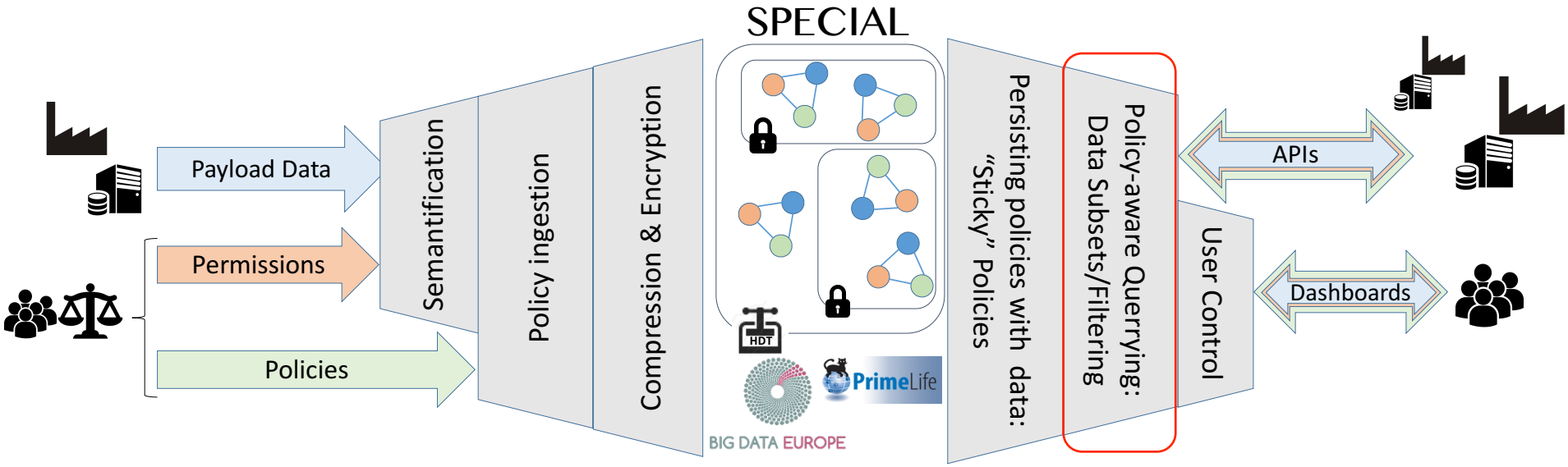
- When sharing data or query results information is **securely stored and exchanged**
- Enable efficient **queryable encryption** based on **compressed** RDF data

Software components - Sticky Policies



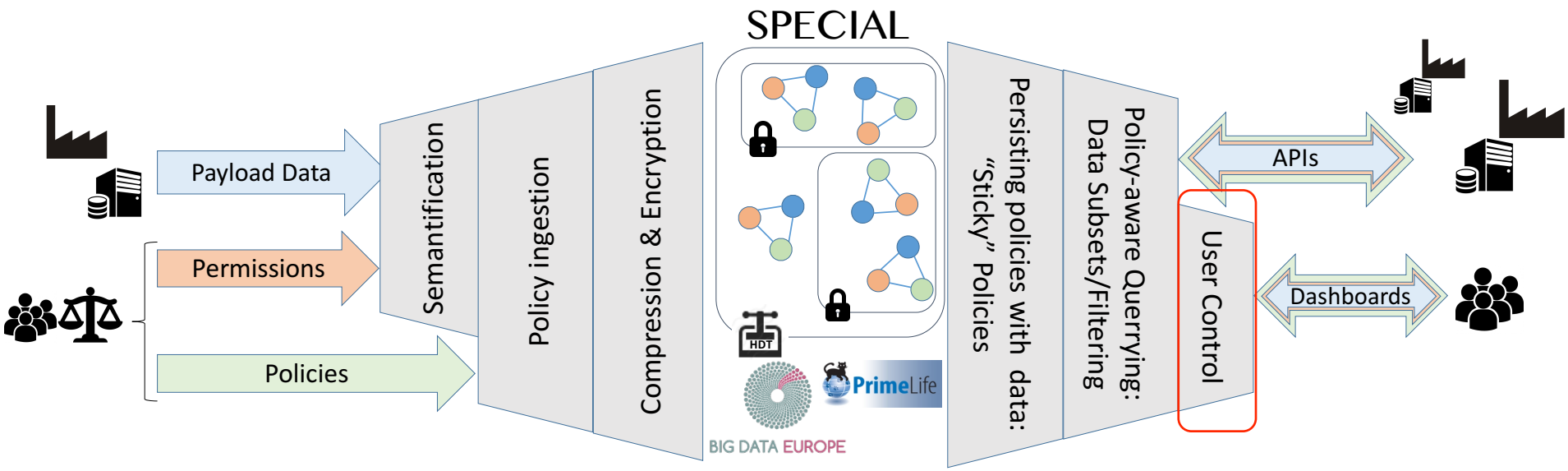
- **Data sharing** can be done along data value chains in a way that includes the policy information
- Gluing policy information to the payload data persistently, even across company borders, is called "**sticky policies**"
  - ❖Data protection **constraints**
  - ❖Other **limitations** and **obligations**

Software components - Policy-aware Querying



- Categorise and subdivide data through annotations into **sensitivity categories/levels** or based on **fine-grained user-policies**
- Policy aware **aggregation** and **anonymisation** techniques
- Recording of the sharing event in a manner that supports **non-repudiation**

Software components - User Control



- Interactive Dashboard
  - ❖Display **highly relevant information** to the user based on context
  - ❖Map what the users sees to their **entire Linked Data graph**
  - ❖Investigate how semantified data can cater for **better informed consent**
- Relieve the burden of policy management via **Templates**
- Support **versioning**, **revocation**, and **forgetting** functionality

Adversaries & Additional input

- Challenges
  - ❖Provide synthesised linked graph data (linked to existing open data sets) and challenge users to **reconstruct those encrypted graphs**
  - ❖Develop simulated synthesised policies and datasets and derive challenges to **retrieve and re-construct unauthorised information** from our system
- Workshops
  - ❖Discuss limitations and possible **additional challenges**
  - ❖Derive **challenges that can not be tested automatically** e.g. policies that cannot be enforced by automated means need to be protected by (legal) contracts
- Additional Input
  - ❖ICT-18-2016 and ICT-14-2016 projects
  - ❖Privacy & Us (Privacy & Usability) https://privacyus.eu/, Data markets Austria https://datamarket.at/, etc…
  - ❖W3C standardisation activities

Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance

Technical/Scientific contact
Sabrina Kirrane
Vienna University of Economics and Business
sabrina.kirrane@wu.ac.at

Adminsitrative contact
Philippe Rohou
ERCIM W3C
philippe.rohou@ercim.eu