CONSEQUENCE

www.consequence-project.eu

Sharing Scientific Data: Scenarios and Challenges

Benjamin Aziz[2], Shirley Crompton[1] and Michael Wilson[2]

[1]STFC Daresbury Laboratory, UK
[2]STFC Rutherford Appleton Laboratory, UK

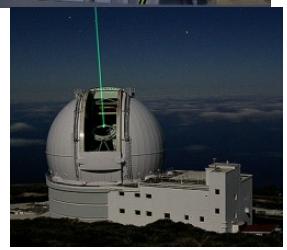
W3C Workshop on Access Control Application Scenarios
Abbaye de Neumunster, Luxembourg
17 November 2009

Science & Technology Facilities Council
e-Science

Science and Technology Facilities Council (STFC)

Developing and supporting science facilities for the disciplines of particle physics, astronomy, microelectronics, etc. And recently medical research






e-Science Centre

- High performance services and applications to support scientific research

- Research and development programme includes:
  - Large-scale intensive computing based on Grids
  - Semantic Web technologies
  - **Scientific data management, including sharing & preservation**

Enabling Grids for E-sciencE

NGS National Grid Service

Science & Technology Facilities Council
e-Science

Objectives of this Research Work

- Current situation is one of two extremes:
  - Mostly no sharing (when scientific studies are commercially funded)
  - Loose data sharing (when studies publicly funded)
- But also:
  - Gap between high-level data sharing agreements (DSAs) with academics, funding parties and industrials, and the low-level security mechanisms enforcing these DSAs
- Facilitate flexible data sharing and usage of scientific data across different administrative domains
- Provide a top-to-down process in which not only system administrators and developers are involved, but also scientists, lawyers and project managers (i.e. stakeholders)

CONSEQUENCE

www.consequence-project.eu

EU FP7 Project Consequence (2008-2010)

Project Statement

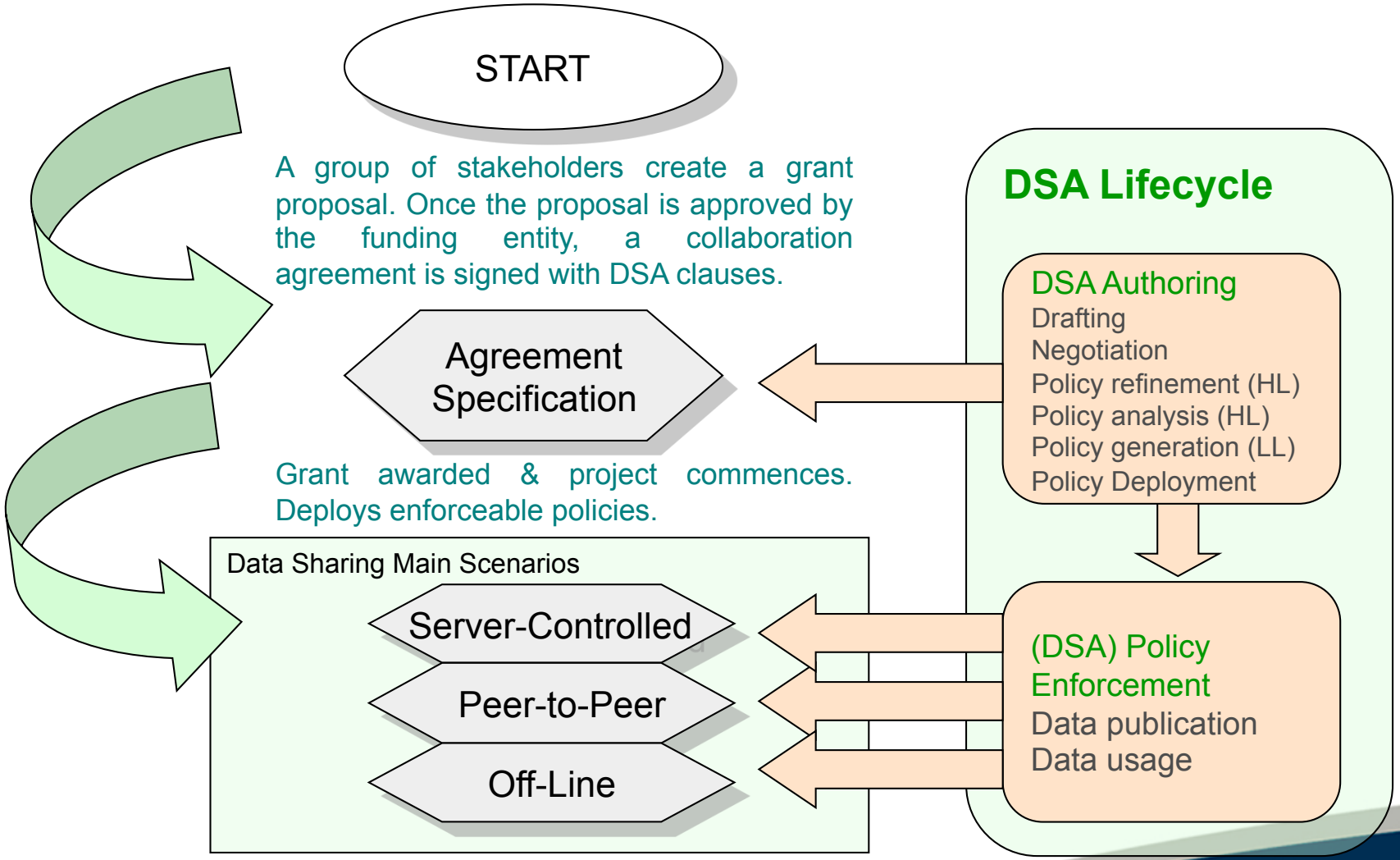
- The aim is to fill in the gap between high-level business security requirements and low-level security mechanisms
- Provide a framework for context-aware data-centric security where data can be shared and processed across administrative domains in a fine-tuned manner

Partners

- **BAE** (Test-bed owners, DSA Management)
- **STFC** (Test-bed owners, DSA Management)
- **HP** (research, DSA authoring and mapping)
- **CNR** (research, formal analysis)
- **Create-Net** (research, policy mapping)
- **Imperial College London** (research, enforceable policies)
- **EMIC** (Coordinators, security infrastructure)

Science & Technology Facilities Council
e-Science

Scientific Data Sharing Lifecycle

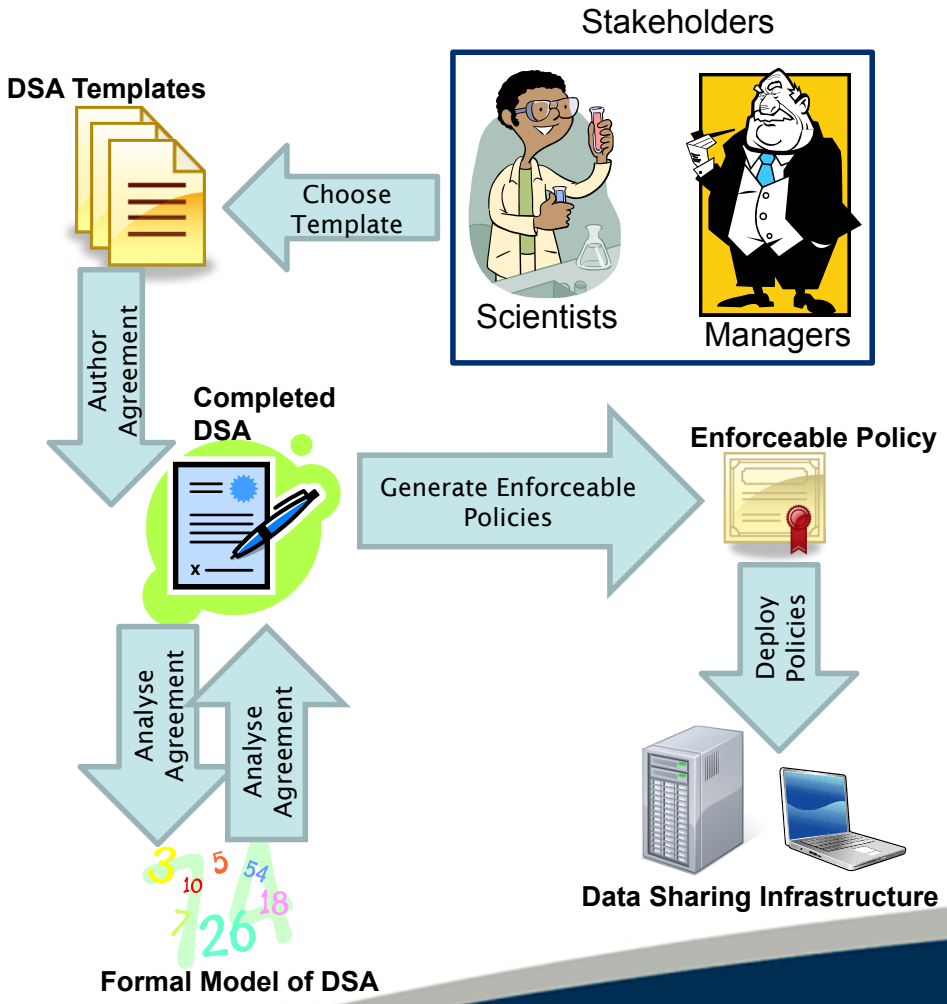


Main Scenarios and Requirements

- **Scenarios**
  - DSA and Policy Management
  - Server-based Data Accessing
  - Peer-to-Peer Data Sharing
  - Offline Data Usage
- **Requirements**
  - Stakeholders are able to manage risk
  - Traditional access control for publicly and commercially funded studies
  - Usage control leading to and controlling derived data
  - Context-awareness including location, date and time
  - Fine-grained control of different parts of a data file
  - Offline access and usage control

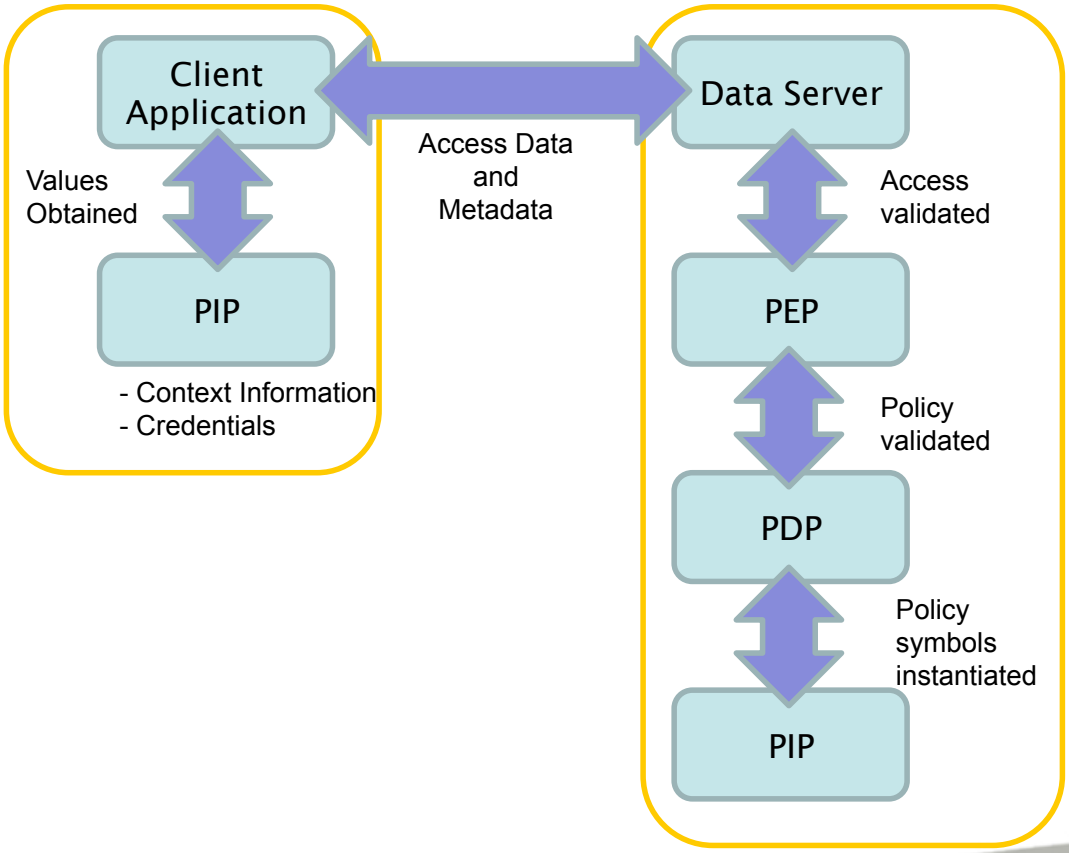
Data Sharing Agreements and Policy Management

- Main challenges:
  - Mapping of DSAs written in natural languages to:
    - formal policy languages (e.g. based on process algebra) for analysis and verification,
    - enforceable policy languages (XACML-based) for deployment
  - Feedback from the analysis step to the DSA authoring in a manner sensible to the stakeholders
  - Long-standing collaborations may have evolving or new DSAs over long timescale due to derived data



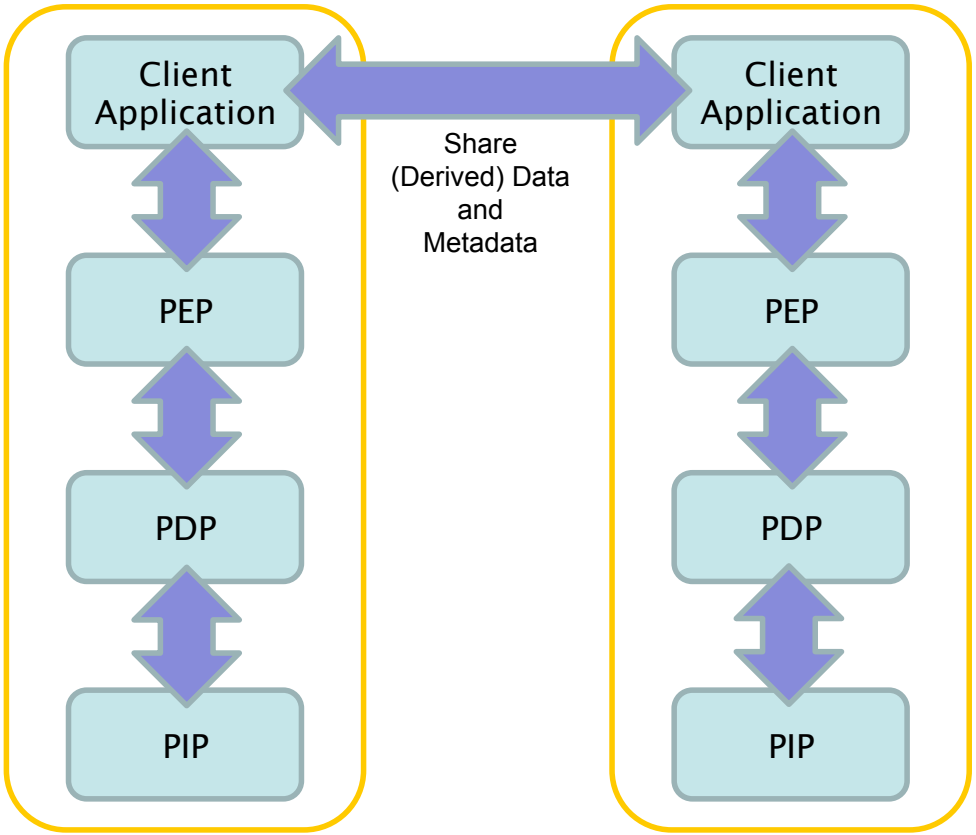
Server-based Data Accessing

- Main challenges here are the classical enforcement of access control policies
  - Context-awareness
    - Client permitted within certain locations, times and dates
  - Fine-grained access to data files based on metadata
    - E.g. releasing a experiment's image but not the results data and experiment conditions behind it



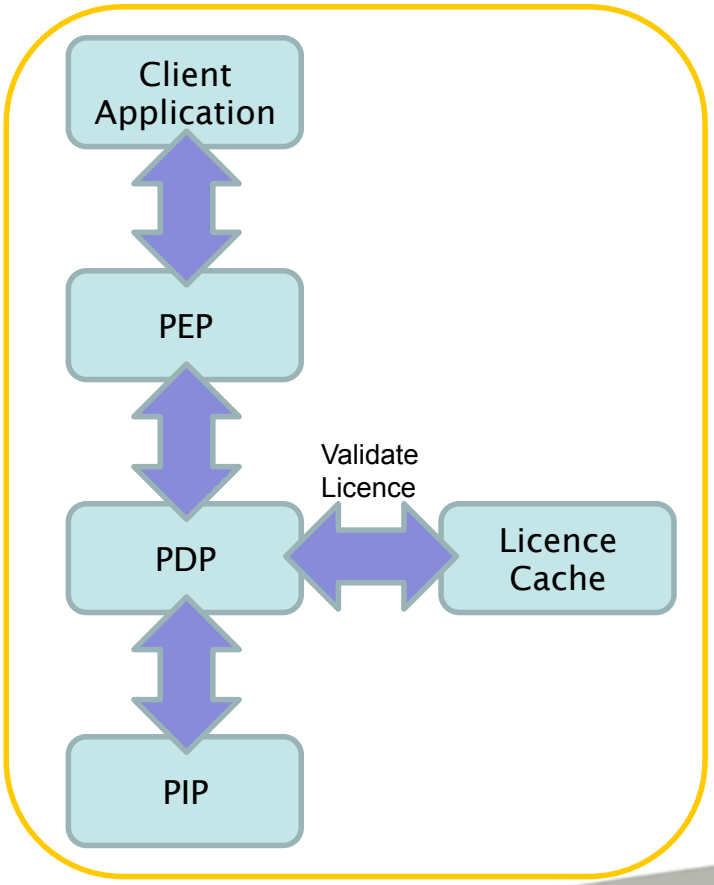
Peer-to-Peer Data Sharing

- Main challenges here include:
  - Context-awareness
    - Where and when are the two clients?
  - Derived data
    - Viewing
    - Transformation
    - Visualisation (i.e. view + transform)



Offline Data Usage

- Main challenges here are:
  - Use of caching at client-side as a form of sticky policies
  - Enforcing usage control based on original policy
    - Enterprise Digital Rights Management



Conclusion

- This paper presented a number of scenarios for scientific data sharing based on high-level data sharing agreements
- The work is carried out within project Consequence, which aims at providing a process for mapping high-level agreements to low-level mechanisms
- The main challenges posed by these scenarios included classical access and usage control, context-awareness, fine-grained control and offline usage