

# Data Sharing Agreements & Enterprise Rights Management

Jan 2011, London

Michael Wilson (STFC)

#### Structure of the talk



- Data Sharing Agreements (DSA)
- The Consequence Approach
- STFC Scenario
- Defining a DSA vocabulary and policies
- Walk through of a use case
- Summary of Main Advances

## DSA – roles & objectives



#### Data Producers:

- Need to ensure that data is
  - Used for the intended purposes only
  - Distributed to those who should receive it
  - Used only by those who should use it

#### Data Consumers:

- Need to ensure that data is supplied
  - to quality
  - to time
- DSA between pairs or groups of enterprises (e.g. a supply chain)
- DSA currently exist as paper MoU or Agreements
- Need for technical enforcement of DSA

#### Structure of DSA - 13/19 sections



- Definition of terms
- Parties to the agreement
- Purpose of the agreement
- Period of agreement
- Justification for access
- Description of the data
- Data Quality labelled
- Description of the data users
- Method of data access or transfer
- Location of data and custodial responsibility

- Restrictions on Data Use.
- Derived data
- Dissemination to third parties
- Confidentiality
- Disposal of data
- Administration of the Agreement
- Breaches to the agreement
  - Applicable Law
- Signature

#### **Badly drafted DSA clauses**



#### Redundant constraints:

5.8 Data or information acquired by one party from the other under the Agreement shall not be disseminated commercially or otherwise disclosed, transmitted or sold to any organizations other than the parties to the Agreement ...

#### Vague intentions can't be technically enforced:

5.15 The parties agree to acquire and use data and information in a manner that respects the scientific value of the data...

Environmental Data Sharing AGREEMENT Between Environment Canada, Pacific and Yukon Region and British Columbia Ministry of Water Land and Air Protection, April 29, 2004

#### Well drafted DSA clauses



- 5.9 Quality controlled data or information acquired by one party from the other under the Agreement may be disseminated or otherwise disclosed or transmitted without cost to any individual or organization without the need for written consent from the party that owns the information providing that (1) the data or information are not modified, (2) the jurisdictional ownership is clearly acknowledged and (3) the release does not compromise the privacy or commercial competitiveness of the data source;
- 5.10 Raw data or draft information acquired by one party from the other under the Agreement may not be disseminated or otherwise disclosed to any individual or organization other than the parties to the Agreement without the written consent of the party that owns the data or produced the information;
- 5.12 The Agreement authorizes the production, dissemination and sale of derivatives by the parties.

#### Language of agreements



- Default nobody can do anything
- Authorisations
  - Access authorisation
  - Usage authorisation
- Prohibitions
  - Constraints within the authorisations
- Obligations
  - Timing
    - Pre-obligations obliged to do something before an action
    - Post-obligations obliged to do something after an action
  - Trigger event
    - Authorisation events (e.g. When a user accesses a file)
    - General events (e.g. When a security level changes in the context)

#### Context of the data access & usage



#### Who

users or users in which roles can access & use data

#### What

What data or data types does the agreement apply to

#### Where

What location can the access & usage take place

#### When

What time can the access & usage take place

#### How

- What functions in which software tools can be used on the data
- Technological enforcement of DSA requires trusted providers of context information

### Consequence approach



- An Enterprise Rights Management (ERM) solution
- Combine:
  - Access control
  - Usage control (UCON)
  - Digital Rights Management (DRM)
- To support policies that can be technologically enforced
- And that can be drafted by managers
- In an expressive language for DSA policies
- Which calls on trusted context authorities when evaluated

### **Access & Usage Control**



- 2 stages: Credential Authentication and Action Authorisation.
- When a credential is presented to a controller it compares the credential to an access control list, grants or denies the presented request.
- Conventional access control actions are provide discrete access to a resource: read, write, delete, download.
- Conventional access control requires access to the controller and access control list – usually on-line.
- Usage control (UCON since 2002) extends the set of actions to address mutability (filter, merge, modify) to derive a new data set from the original.
- UCON is finer grained than access control and requires ongoing authorisation often off-line. To control fine grained usage control the digital media are usually encrypted.

## **Digital Rights Management**

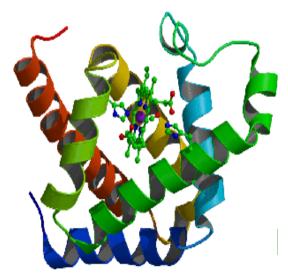


- DRM technologies control use of digital media by preventing access, copying or conversion to other formats by end users.
- The usually authorise playing or viewing of digital media on specific devices or by specific users.
- They usually operate through the publication of a license with the digital media which the player/controller can match to the specific device or specific user when a request is made to open the digital media.
- Licenses usually have an expiration date. Device and user identity credentials are usually provided by the host platform.
- Licences, the digital media and the player/controller are stored together so that off-line access is supported.
- To control off-line access the digital media is usually encrypted.

## STFC Data Sharing Scenario



- •A Research Council awards a grant to a consortium to **study an enzyme** that is essential to the lifecycle of HIV virus.
- •The project will pave the way for the development of new drugs
- •The consortium includes researchers from several universities, and a commercial partner.
- •STFC ISIS facility beamline is used to:
  - collect neutron diffraction patterns of an enzyme bound to the interacting host proteins
  - Users will download data and analyse it locally
- STFC store the experimental data
- Policies in DSA between STFC & universities state who has access to it, where and when ?



#### STFC Scenario DSA Policies



#### Authorisation Policies

- 1 Before the end of the three year embargo period, access to the experimental data is restricted to the principal investigator and co-investigators.
- 2 After the embargo period, the experimental data may be accessed by all users.
- 3 Beamline scientists can access image data related to or produced on their experimental station.
- Prohibition Policy
- 4 Access to numerical data should be denied to users which country is either Iran or North Korea.

#### STFC Scenario DSA Policies



- Authorisation Obligation Policies linked to policy 2
- 5 BEFORE a public user downloads data THEN the system must notify the principal investigator.
- 6 AFTER a public user reads some data THEN the system must log the event.
- Usage Policies prohibition
- 7 IF a user does not use **eCat** THEN that user **cannot** transform experimental data into derived data.
- Declarations

Agreement start date and end date Identity of trusted context authorities Duration of off-line usage license

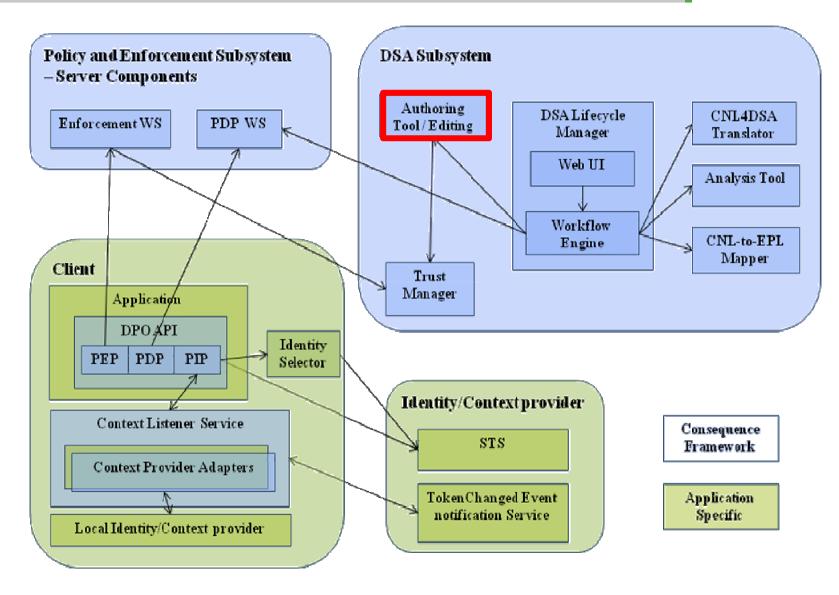
#### What do users have to do?



- Define the ontology and DSA user vocabulary:
  - Protege V3.1.1, owl V1
- Define the DSA policies
  - DSA Authoring tool
- Publish application terms
  - Metadata Manager
- Map application terms to DSA user vocabulary
  - CNL/EPL config file GUI one per domain/party
- Define STS for context variables: UID, location, (time)
- Define services to provide events (e.g. security level)
- Define services to perform obligation actions & action handlers,
- Application client, server

### **Consequence Architecture**





## Define the ontology & vocab.

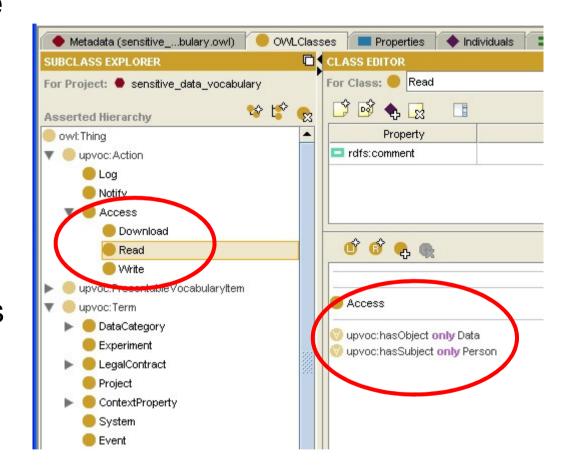


Ontology is defined in the standard language OWL Standard tool Protege V3.1.1

Define application actions: system & user

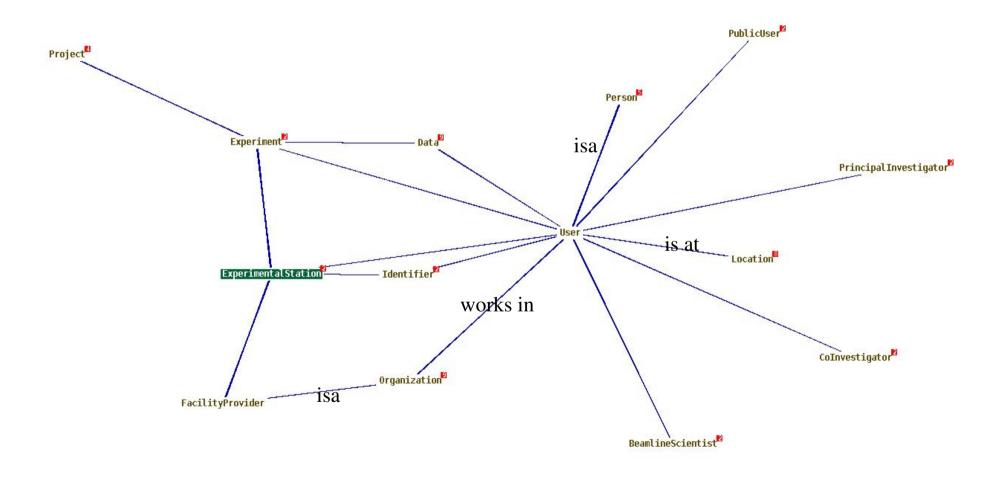
Define terms & properties relating terms

Vocabulary is defined as presentable vocabulary subset of larger ontology.



## Visible vocabulary and hidden ontology

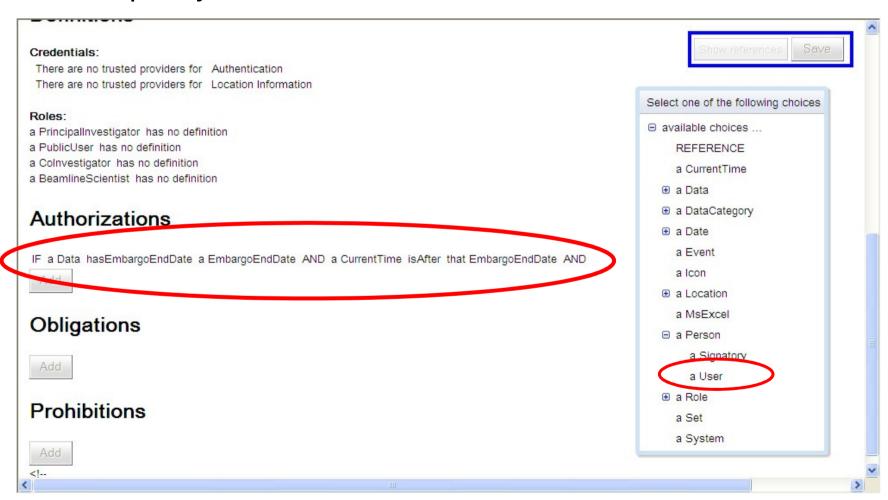
**Experimental Station** is hidden from the user vocabulary but is used in the ontology to define visible terms.



#### Define the DSA policies



The authoring tool presents the words which could be put next into a policy



## **DSA Authoring - references**



IF a Data hasPrincipalInvestigator a Set AND a User hasRole a PrincipalInvestigator AND that User in that Set THEN that User CAN Access that Data
IF a Data hasCoinvestigator a Set AND a User hasRole a Colnvestigator AND that User in that Set THEN that User CAN Access that Data
IF a Data has as data category a ImageData AND that Data hasBeamlineScientist a Set AND a User hasRole a BeamlineScientist AND that  User in that Set THEN that User CAN Read that Data
IF a Data hasEmbargoEndDate a EmbargoEndDate AND a CurrentTime isAfter that EmbargoEndDate AND a User hasRole a PublicUser THEN that  User CAN Access that Data
IF a Data has as data category experimental data AND  Add
Obligations
IF a User hasRole a PublicUser AND a NexusFile has as data category experimental data AND that NexusFile hasPrincipalInvestigator a  Set AND a User hasRole a PrincipalInvestigator AND that User in that Set THEN a System MUST Notify that User
IF a User hasRole a PublicUser AND a NexusFile has as data category experimental data THEN AFTER that User Read that  NexusFile THEN a System MUST Log a Event  □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □
Add
Prohibitions
IF a Data has as data category a NumericalData AND a User hasRole a PrincipalInvestigator AND that User hasLocation a NorthKorea THEN that User CANNOT Access that Data
IF a Data has as data category a NumericalData AND a User hasRole a PrincipalInvestigator AND that User hasLocation a Iran THEN that User CANNOT Access that Data
Add

## **DSA Authoring**



a BeamlineScientist has no definition ▶ Change	Show references	Save	
Authorizations	now references	Sure	
IF a Data hasPrincipalInvestigator a Set AND a User hasRole a PrincipalInvestigator AND that User in that Set THEN that User CAN Data	Access that	<b>o</b> 1.	1
F a Data hasCoinvestigator a Set AND a User hasRole a Colnvestigator AND that User in that Set THEN that User CAN Access th	at Data	Ø 1.	1
IF a Data has as data category a ImageData AND that Data hasBeamlineScientist a Set AND a User hasRole a BeamlineScientist AN User in that Set THEN that User CAN Read that Data	ID that	<b>o</b> 1.	1
IF a Data hasEmbargoEndDate a EmbargoEndDate AND a CurrentTime isAfter that EmbargoEndDate AND a User hasRole a PublicUs User CAN Access that Data  Add	ser THEN that	<b>o</b> 1	ı
Obligations			
F a User hasRole a PublicUser AND a NexusFile has as data category experimental data AND that NexusFile hasPrincipalInvestigator Set AND a User hasRole a PrincipalInvestigator AND that User in that Set THEN a System MUST Notify that User	ra 💩	<b>1</b> 1	⇒
IF a User hasRole a PublicUser AND a NexusFile has as data category experimental data THEN AFTER that User Read that  NexusFile THEN a System MUST Log a Event  Add	٥	<b>†</b> ↓	⇒
Prohibitions			
IF a Data has as data category a NumericalData AND a User hasRole a PrincipalInvestigator AND that User hasLocation a NorthKorea User CANNOT Access that Data	THEN that	s 1	ĻΨ

## Comparing English, CNL & EPL



#### English:

Any registered user can access any data after its embargo period has passed.

Controlled Natural Language

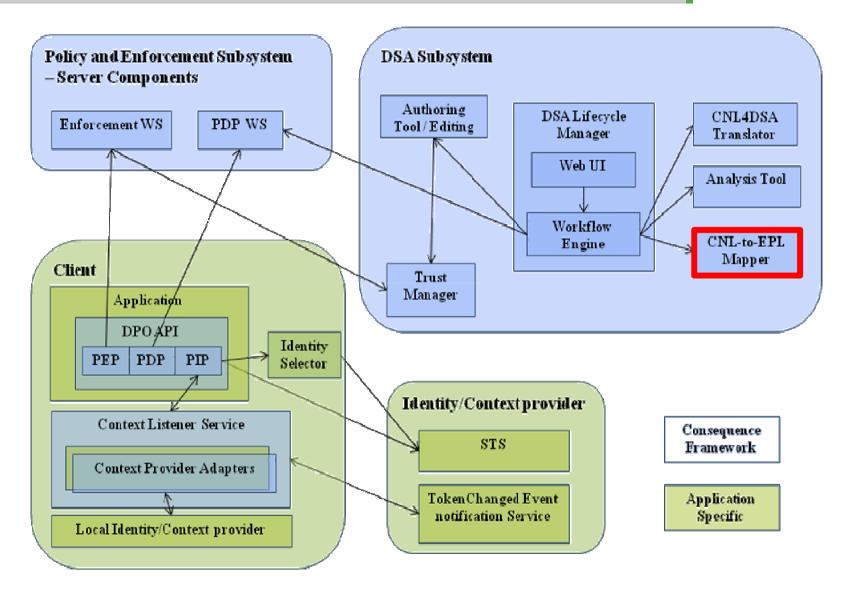
IF a Data hasEmbargoEndDate a EmbargoEndDate
AND a CurrentTime isAfter that EmbargoEndDate
AND a User hasRole a PublicUser
THEN that User CAN Access that Data

Enforceable Policy Language:

```
authorization a1read = allow read(),
target( data_category=="experimental_data" ) to any
  when( currentTime() > object.embargo_end_date);
authorization a1generateGraph = allow generateGraph()
target ( data_category=="experimental_data" ) to any
  when ( currentTime() > object.embargo_end_date);
```

### **Consequence Architecture**

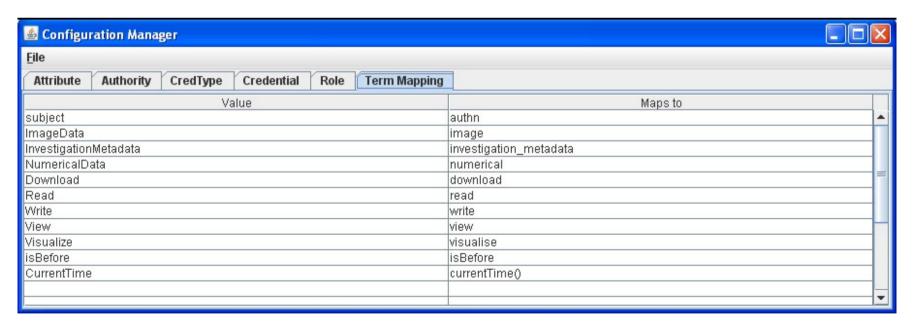




#### Mapping to the executable language

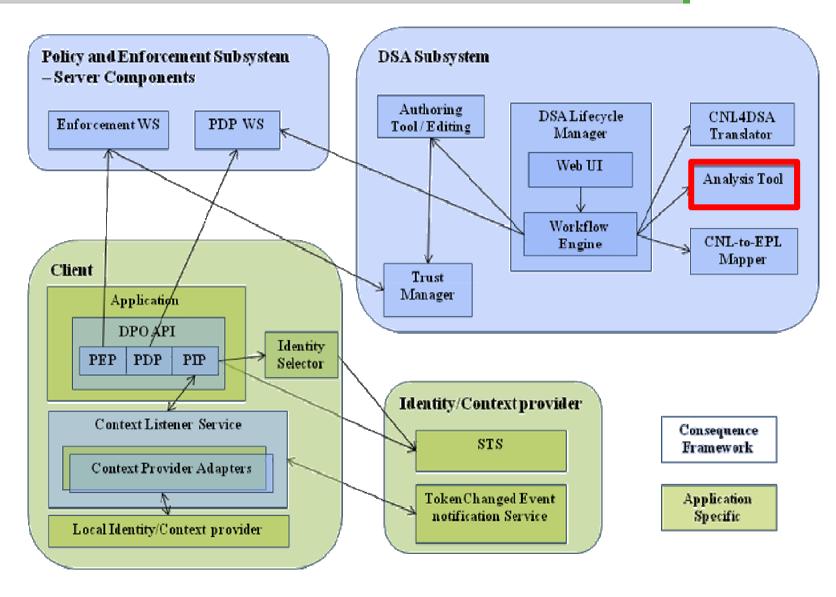


- Terms from the ontology that appear in the controlled natural language (CNL) must be mapped to terms used in the application which appear in the enforceable policy language (EPL).
- A GUI is provided for users to define the mapping.



### **Consequence Architecture**





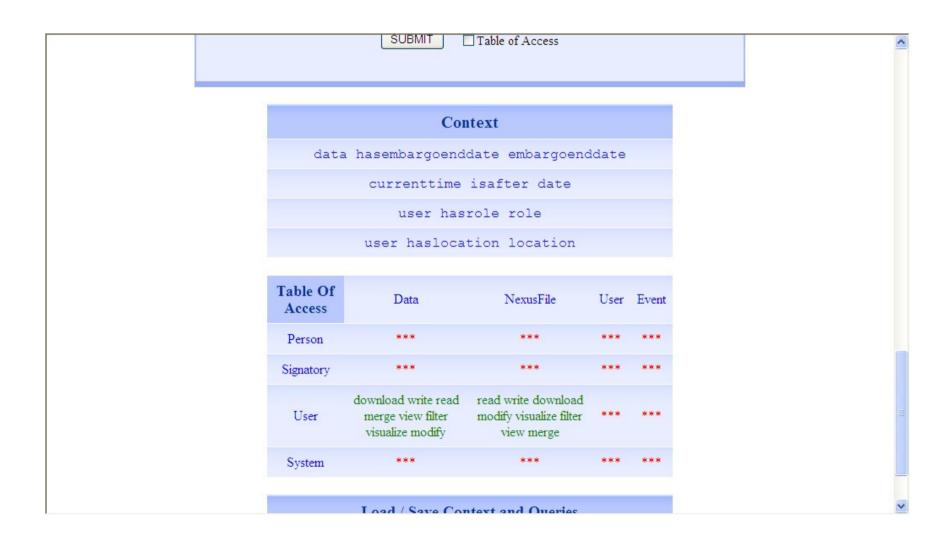
## **DSA Analysis**





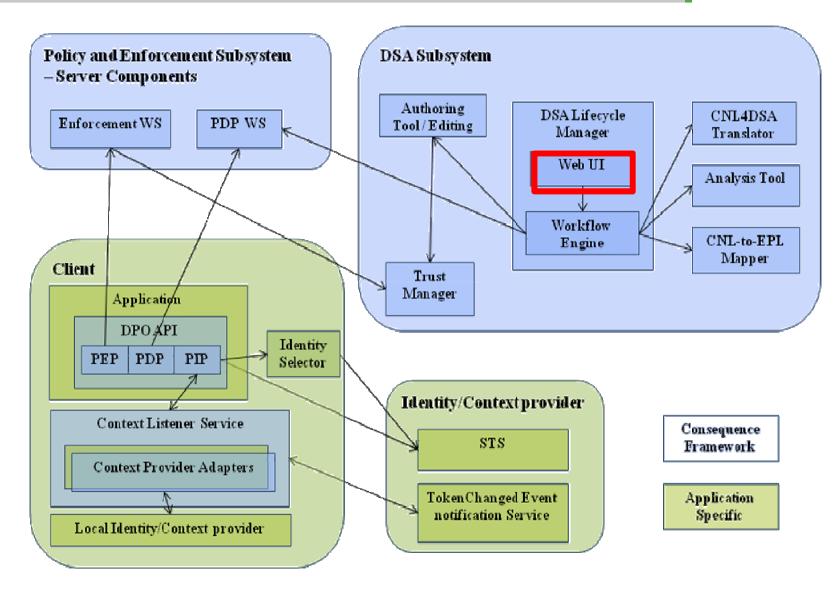
## **DSA Analysis**





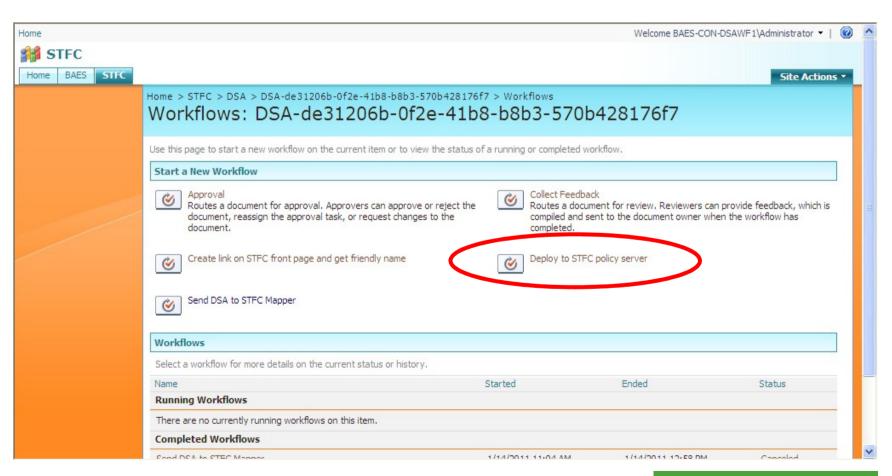
### **Consequence Architecture**





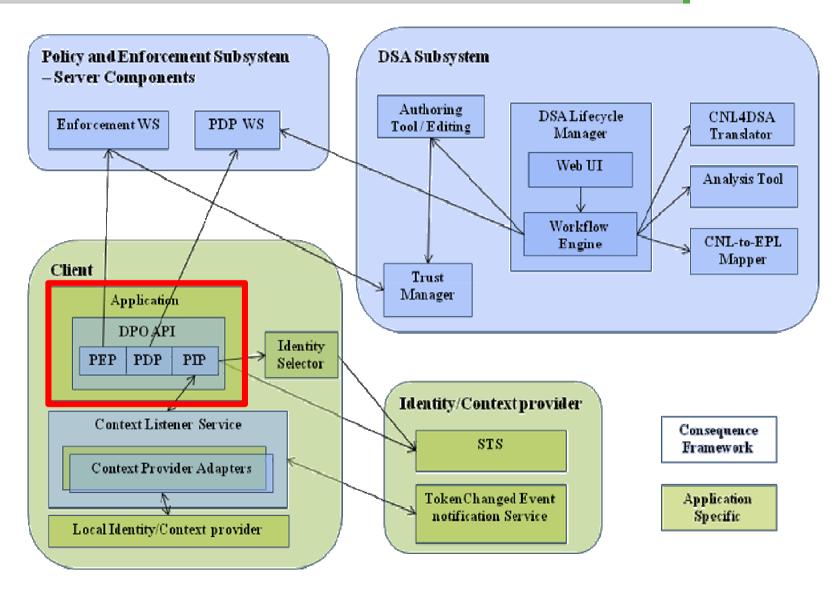
## **DSA** deployment





### **Consequence Architecture**



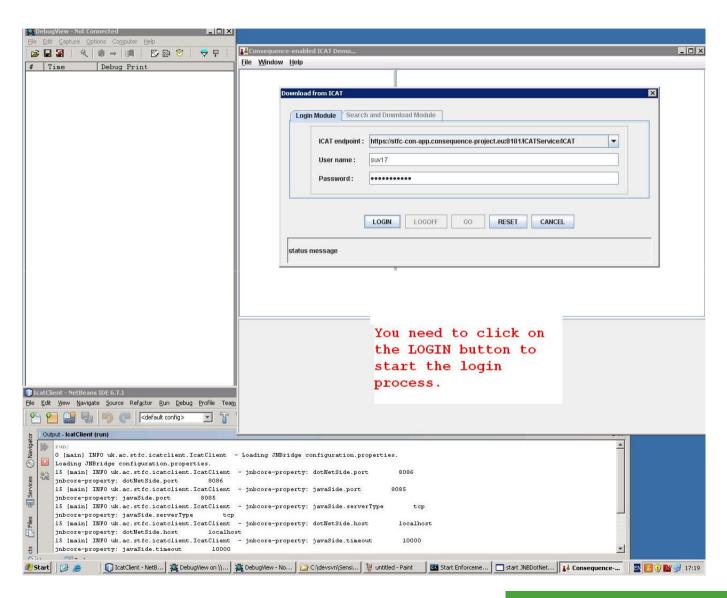


## ICAT client application - login



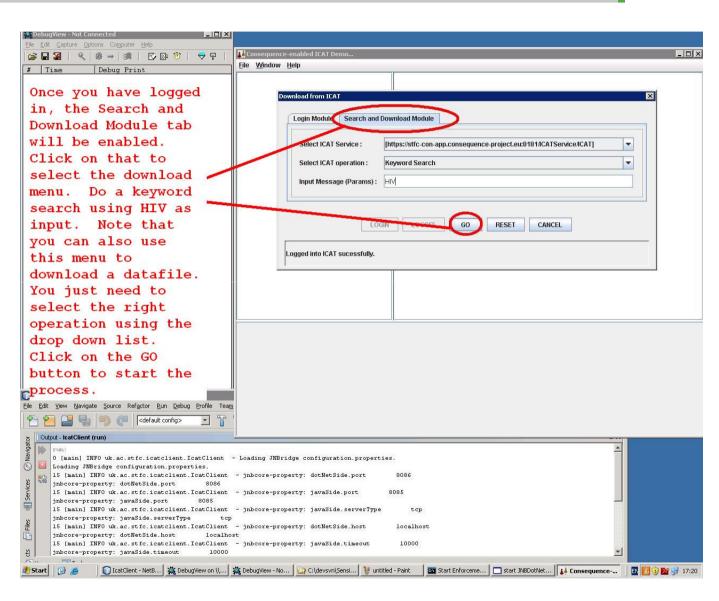
The user identifies themselves by logging into the application.

The server can now determine their role in different projects: PI, CoI, BS, PublicUser.



### Keyword search for data



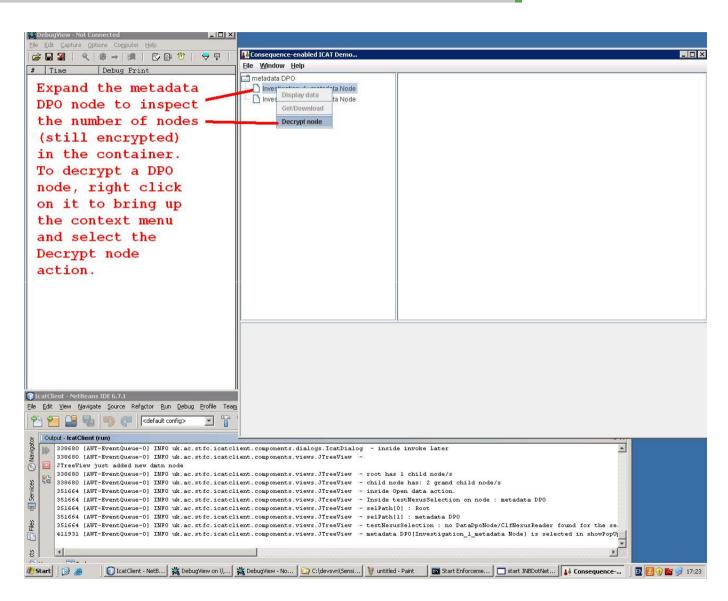


### **Encrypted Metadata results**



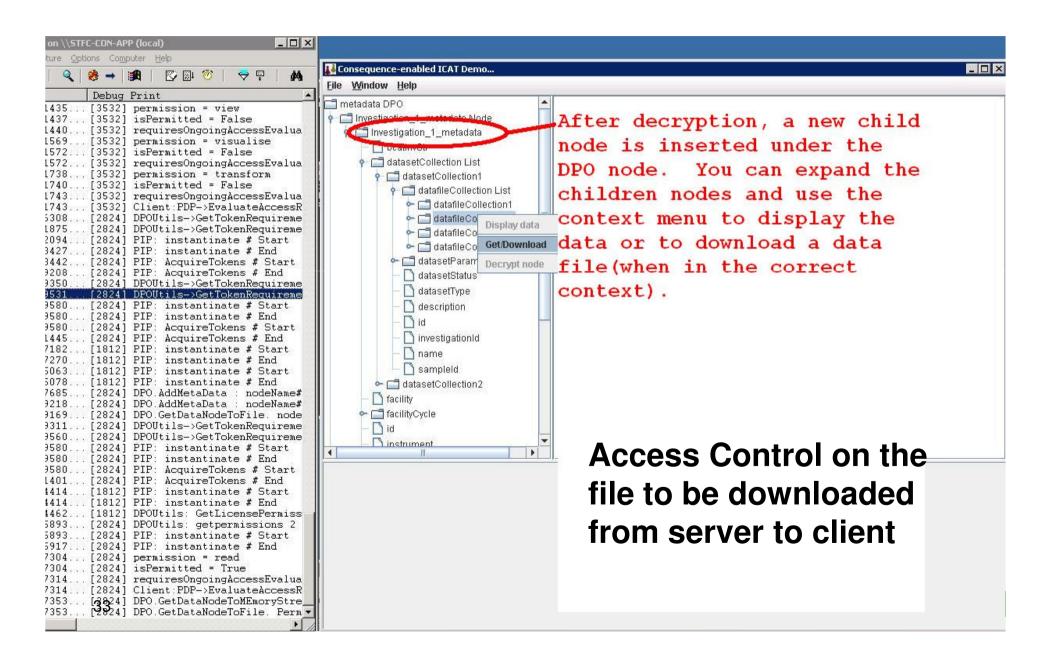
Metadata as well as data is protected by policies

Only metadata from those datasets which the client is authorised to see is shown



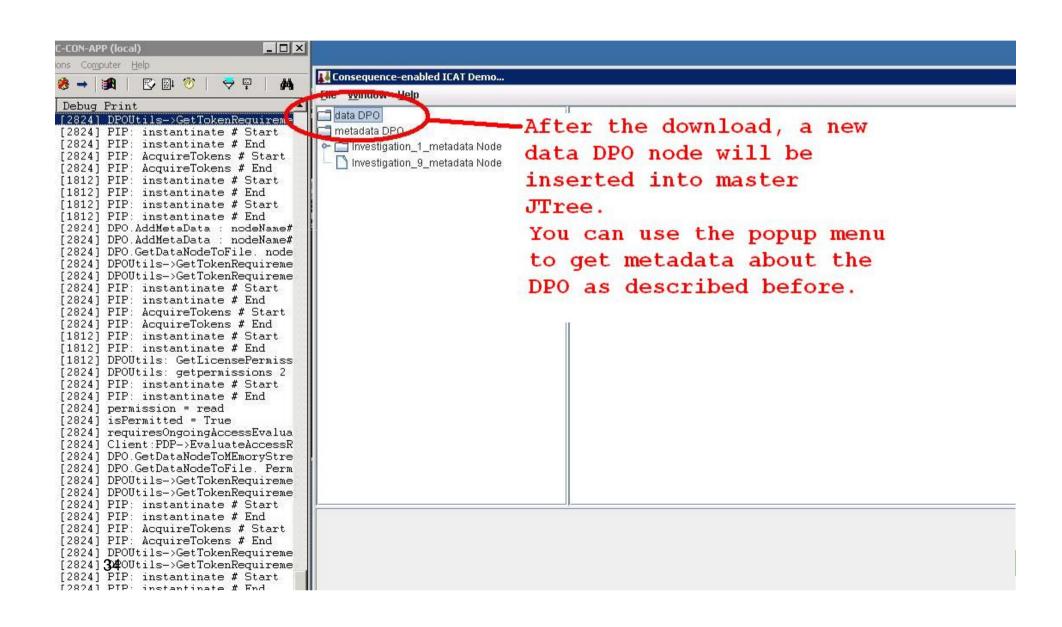
#### Select data file to download





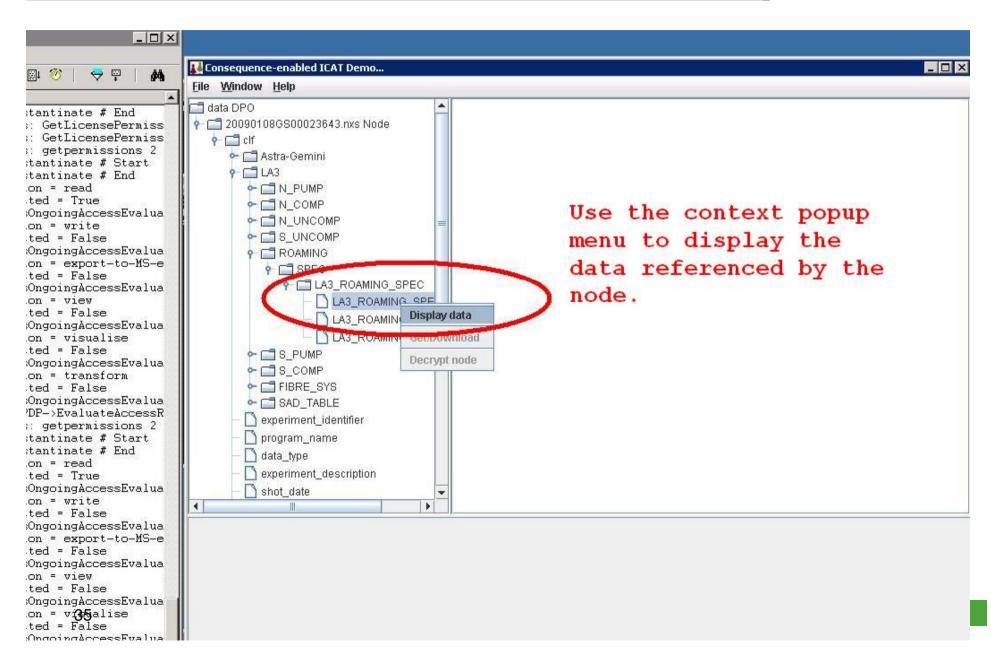
# Navigate file (DPO Object )to find data





#### Select data in Nexus file to view



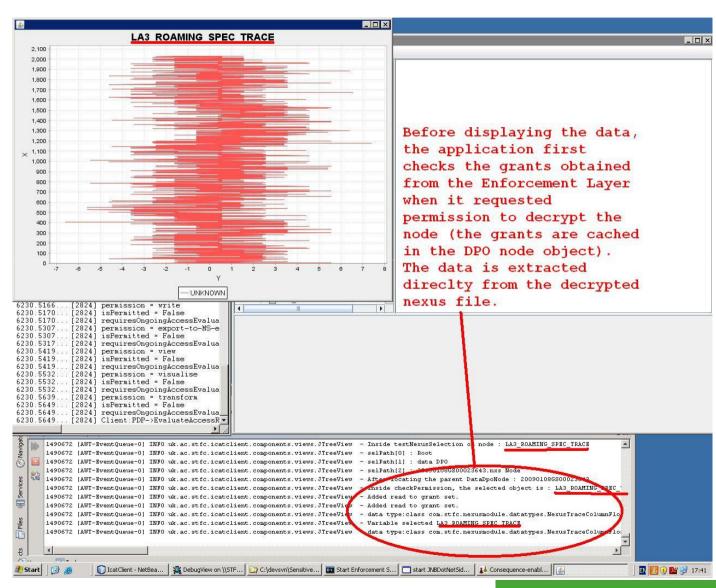


## Usage control: Visualising data



The default display shows a visualisation of a data trace rather than the actual numerical data.

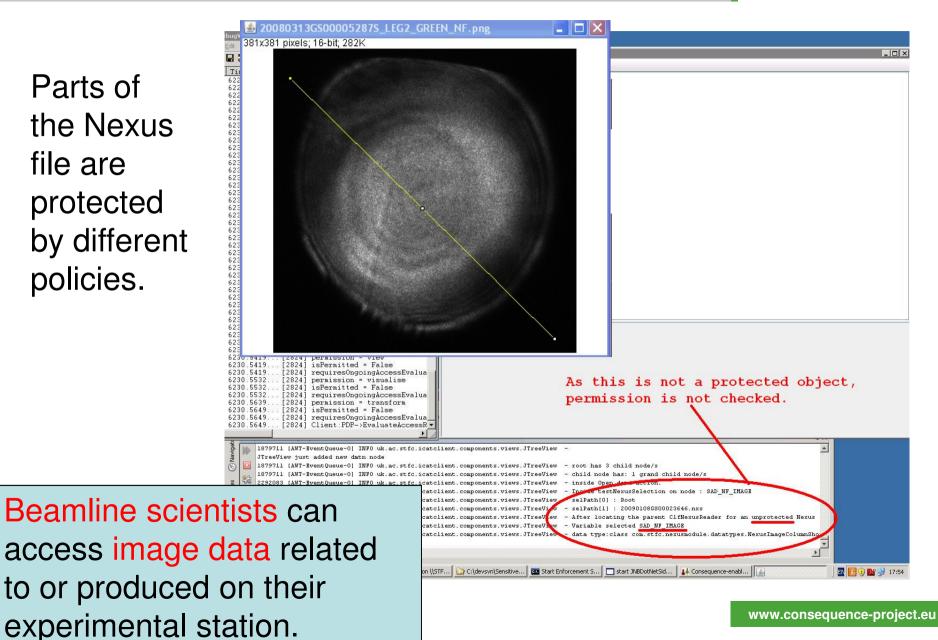
Usage control is applied to the requested action.



### Policies apply within files

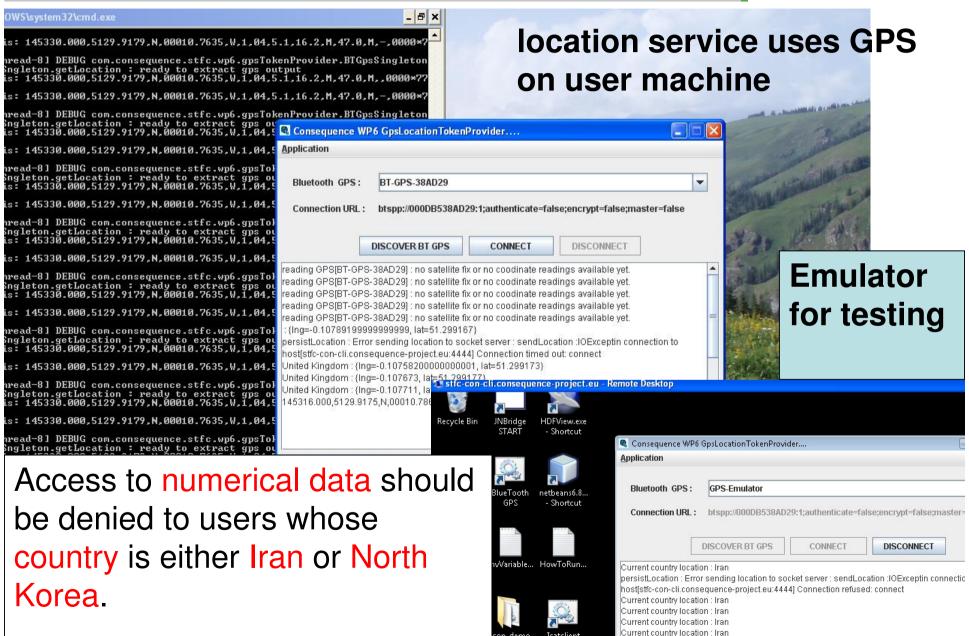


Parts of the Nexus file are protected by different policies.



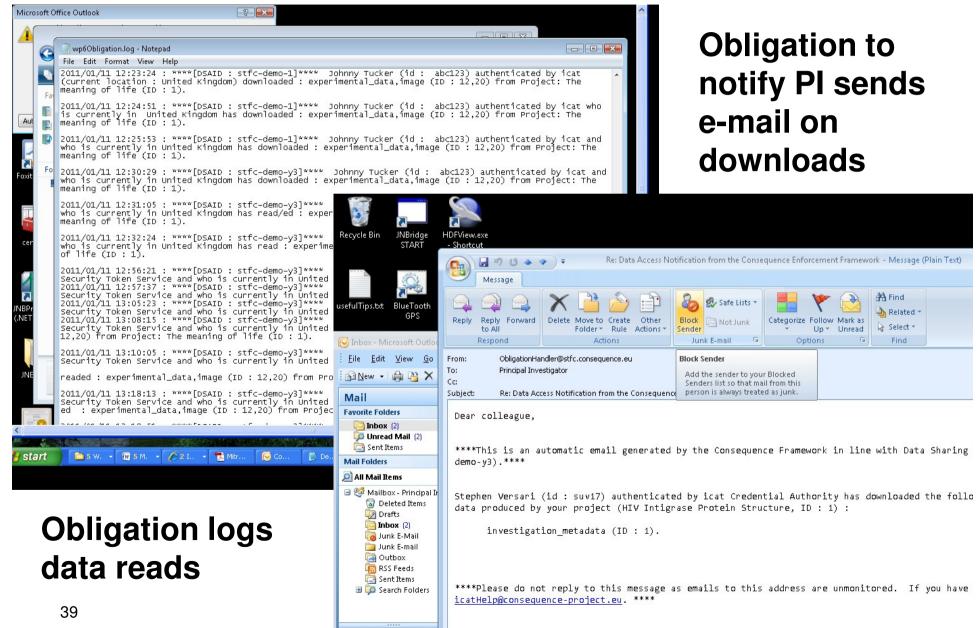
#### **Context service sets location**





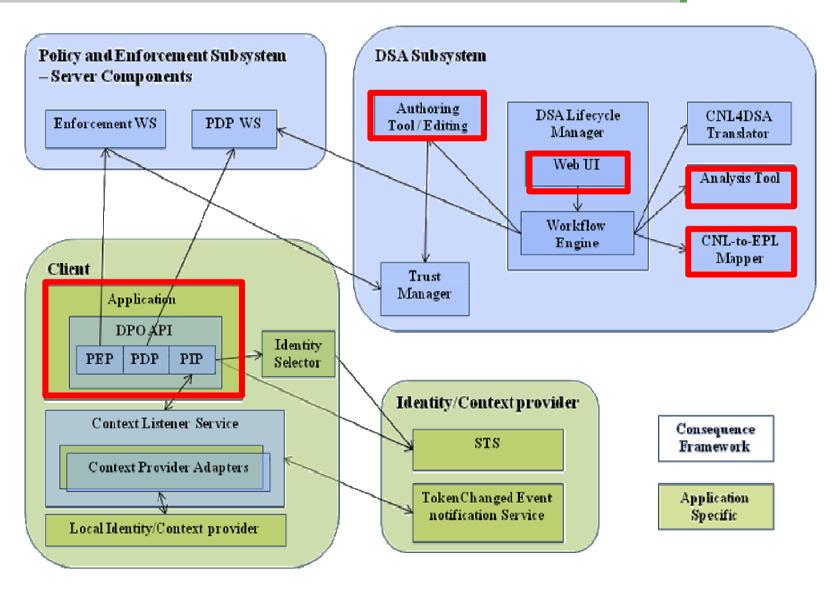
#### **Obligations**





#### **Consequence User Interfaces**





#### **Main Advances**



- Integrating usage control and digital rights management technologies
- Integrating separate advances in the expressiveness of policy languages into a single, usable demonstration
- Semantic driven DSA authoring for managers not techies
- Generic language technology with UI ontology, authoring tool,
   CNL/EPL mapper, metadata manager
- Integration of software engineering formal analysis methods video later
- Integration of mainstream commercial applications video later
- For research innovations see next talk by Dr Emil Lupu

## **Questions?**

