Data Policy Manager

Willem Elbers (MPI-PL)

Adil Hasan (SIGMA/University of Oslo)

3rd EUDAT Conference

EUDAT

Date: 25 September 2014

Agenda

- Introduction
  - Objectives
  - Policies
  - Architecture
  - Security aspects
- Demo
  - Two scenarios
- Discussion

Objectives

- allow a Community Manager (CM) to specify data management policies

- allow a Community Manager (CM) to manage those policies via a web-portal

Data Policy Manager Portal

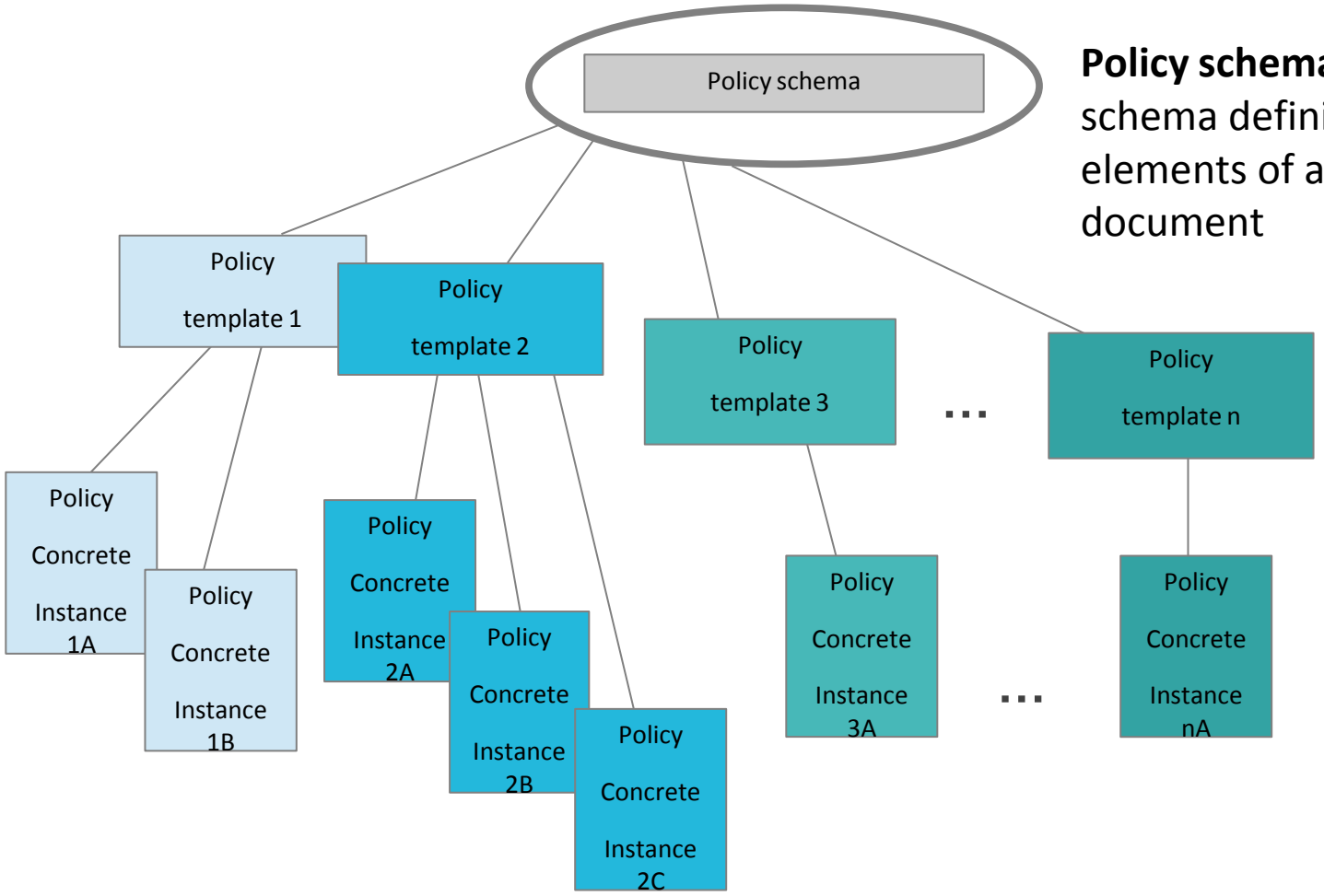
By using the portal the CM can:

- Select policies (replication, integrity checking, retention, etc…)
- Hand-pick the sites involved in the policy
- Visualize the history of the issued policies on a searchable table
- Visualize the history of policy execution

The created policies are in an abstract way, i.e. independent from any technology or technology format (e.g. iRODS rules).

EUDAT

Policy hierarchy



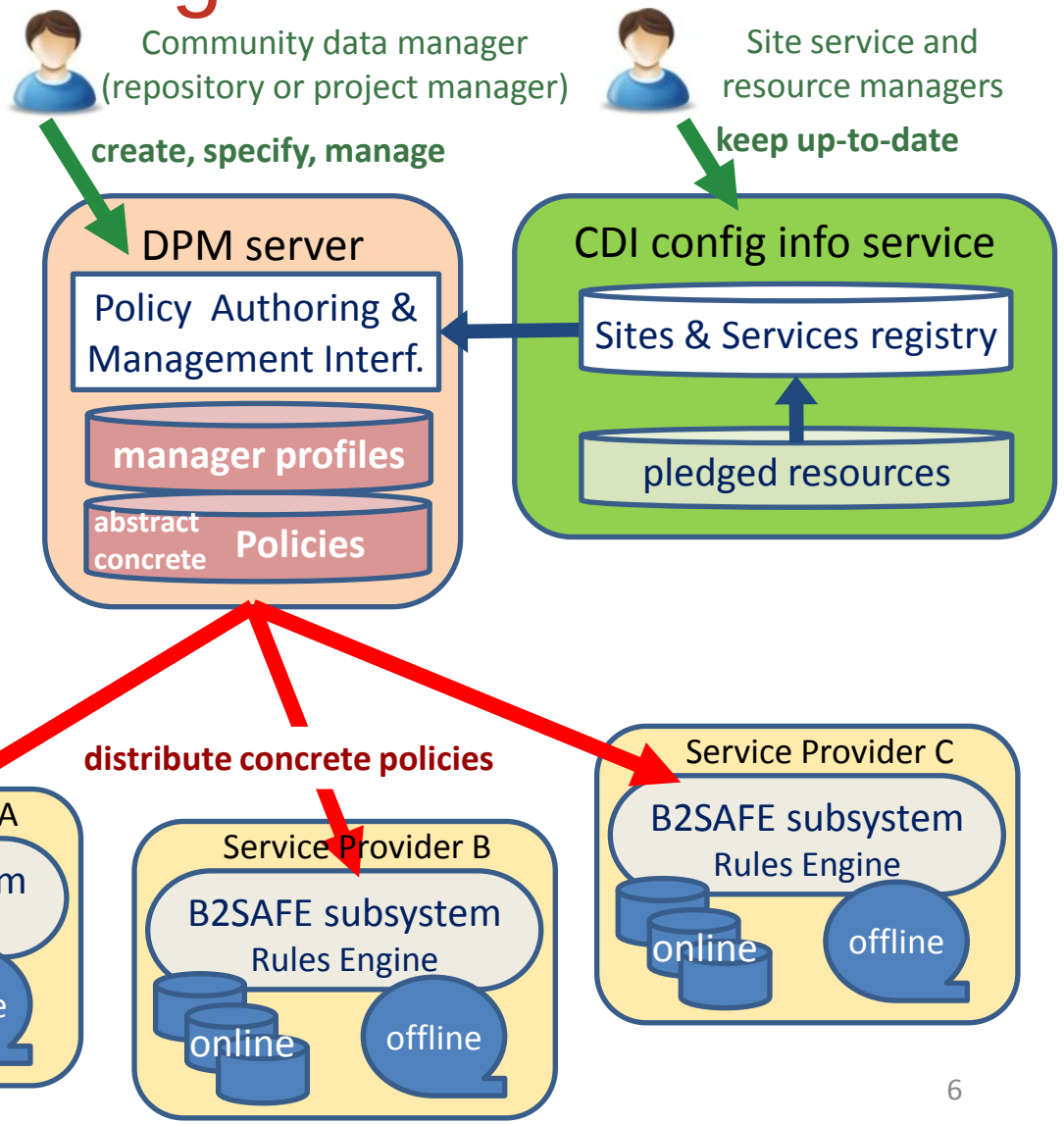
Policy schema: the xml schema defining the elements of a policy document

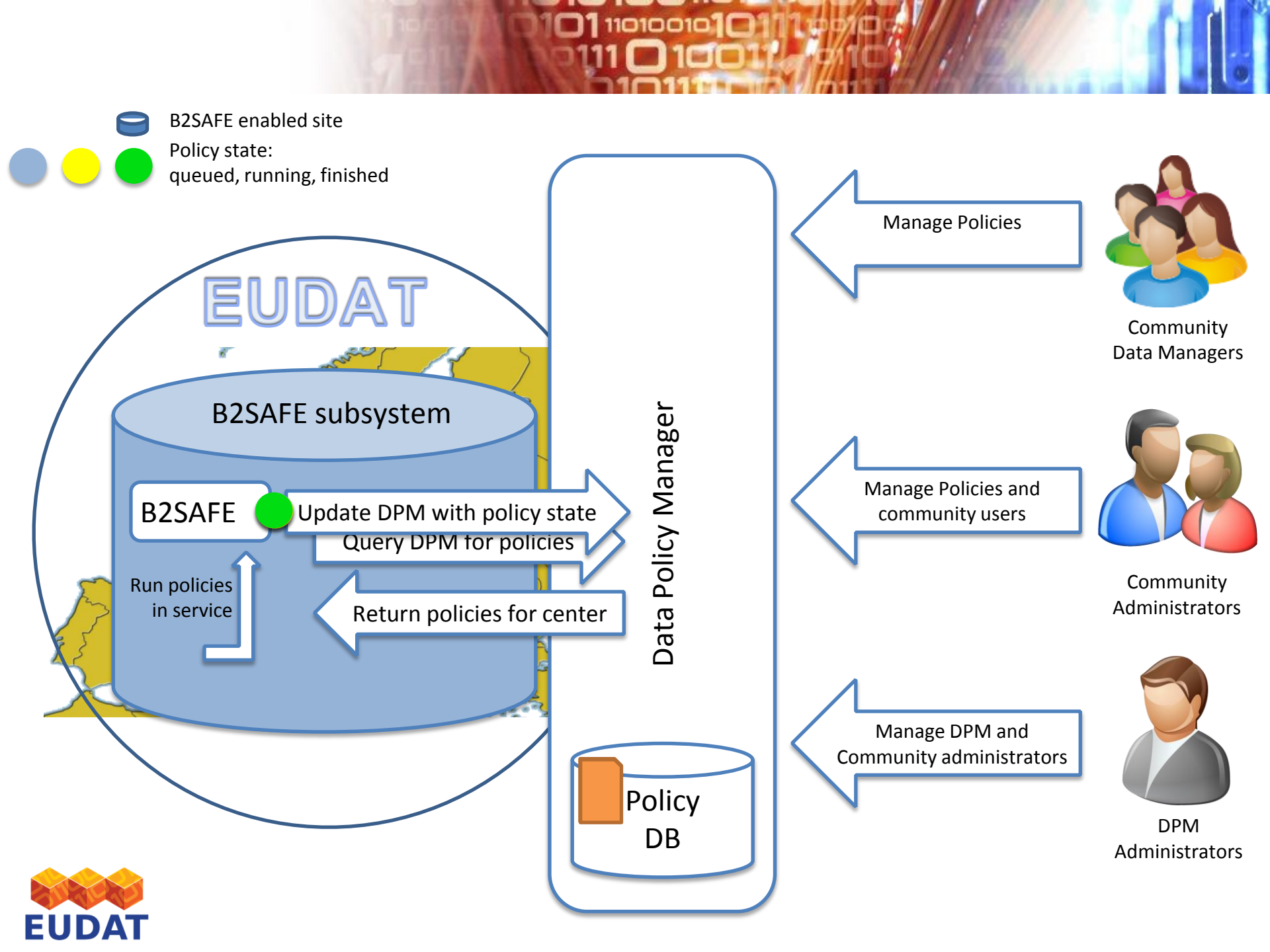
Data Policy Manager Architecture

Data Management Policies specified for community projects to be implemented by B2SAFE subsystems running at the different sites.

Policy examples:

- *replication rules regarding pledged and required storage resources*
- *rules for PID assignment, digital object linking*
- *rules for data retention*
- *rules for data integrity checkings*




B2SAFE enabled site
Policy state:
queued, running, finished
EUDAT
B2SAFE subsystem
B2SAFE
Update DPM with policy state
Query DPM for policies
Run policies
in service
Return policies for center
Data Policy Manager
Policy
DB
Manage Policies
Community
Data Managers
Manage Policies and
community users
Community
Administrators
Manage DPM and
Community administrators
DPM
Administrators
EUDAT

Security Aspects

- Authentication and authorization
  - EUDAT AAI infrastructure (currently crowd SSO)
  - Role based authorization

- Policy
  - Creation:
    - scoped to administrative domain
  - Execution:
    - Scoped to administrative domain
    - Initiated by the data center

EUDAT

Upcoming developments

- Continue development
  - Hardening of the service
  - New features

- Policy work
  - Extend support for B2SAFE replication scenarios
  - Continue work on integrity verification policies
  - Start work on new policies

EUDAT

Demo

- 2 scenarios demonstrated
  - Policy management which includes creation, execution, modification and deletion
  - Policy user administration
    - Very, very early and rudimentary. Allow a user to register as a member of a community, approve and deny the user

Scenario 1

- Community manager wishes to ***create*** a policy to replicate data from site A to B. Later the manager needs to ***update*** the policy to include new data. At the end of the project the manager wants to ***delete*** the policy

Scenario 2

- A new community manager ***joins*** a project and needs to manage policies for that project. Later the community manager ***leaves*** and is removed from the project

EUDAT

DPM Team

Website
https://dpm-eudat.norstore.uio.no/

Resources on github
https://github.com/EUDAT-B2SAFE/B2SAFE-DPM

Mailing list
eudat-tf-policymanager@postit.csc.fi>

Maria Francesca Iozzi (SIGMA/UiO)
Nikolai Vazov (SIGMA/UiO)
Thierry Toutain (SIGMA/UiO)
Adil Hasan (SIGMA)

Willem Elbers (MPI-PL)
Claudio Cacciari (CINECA)
Mark van de Sanden (SurfSARA)
Elena Erastova (RZG)
Michal Jankowski (Poznan)

EUDAT

Abstract Policy template

Policy template: the policy document which defines a policy process, but without specific parameters. Therefore the tasks are defined, but without input/output parameters. For example:

```
<dataset>
  <collection id="0">
    <persistentIdentifier type="PID"></persistentIdentifier>
  </collection>
</dataset>
<actions>
  <action name="replication onchange"
    <type>replicate</type>
    <trigger>
      <action>modify object</action>
    </trigger>
    <targets>
      <target id="0">
        <location xsi:type="irodsns:coordinates">
          <irodsns:site type="EUDAT"></irodsns:site>
          <irodsns:path></irodsns:path>
          <irodsns:resource></irodsns:resource>
        </location>
      </target>
    </targets>
  </action>
```

Define data sets

Policy type

Define action

Target descriptons

Concrete Policy description

Policy instance: the policy document which defines a policy process, but with specific parameters.

```
<dataset>
    <collection id="0">
      <persistentIdentifier type="PID">
              11100/6c8ac19e-c982-11e2-b3cb-e41f13eb41b2
      </persistentIdentifier>
    </collection>
  </dataset>
  <actions>
    <action name="replication onchange">
      <type>replicate</type>
      <trigger>
        <action>modify object</action>
      </trigger>
      <targets>
        <target id="0">
          <location xsi:type="irodsns:coordinates">
            <irodsns:site type="EUDAT">CINECA</irodsns:site>
            <irodsns:path>/path/to/destination</irodsns:path>
            <irodsns:resource>defaultResc</irodsns:resource>
          </location>
        </target>
      </targets>
    </action>
  </actions>
```