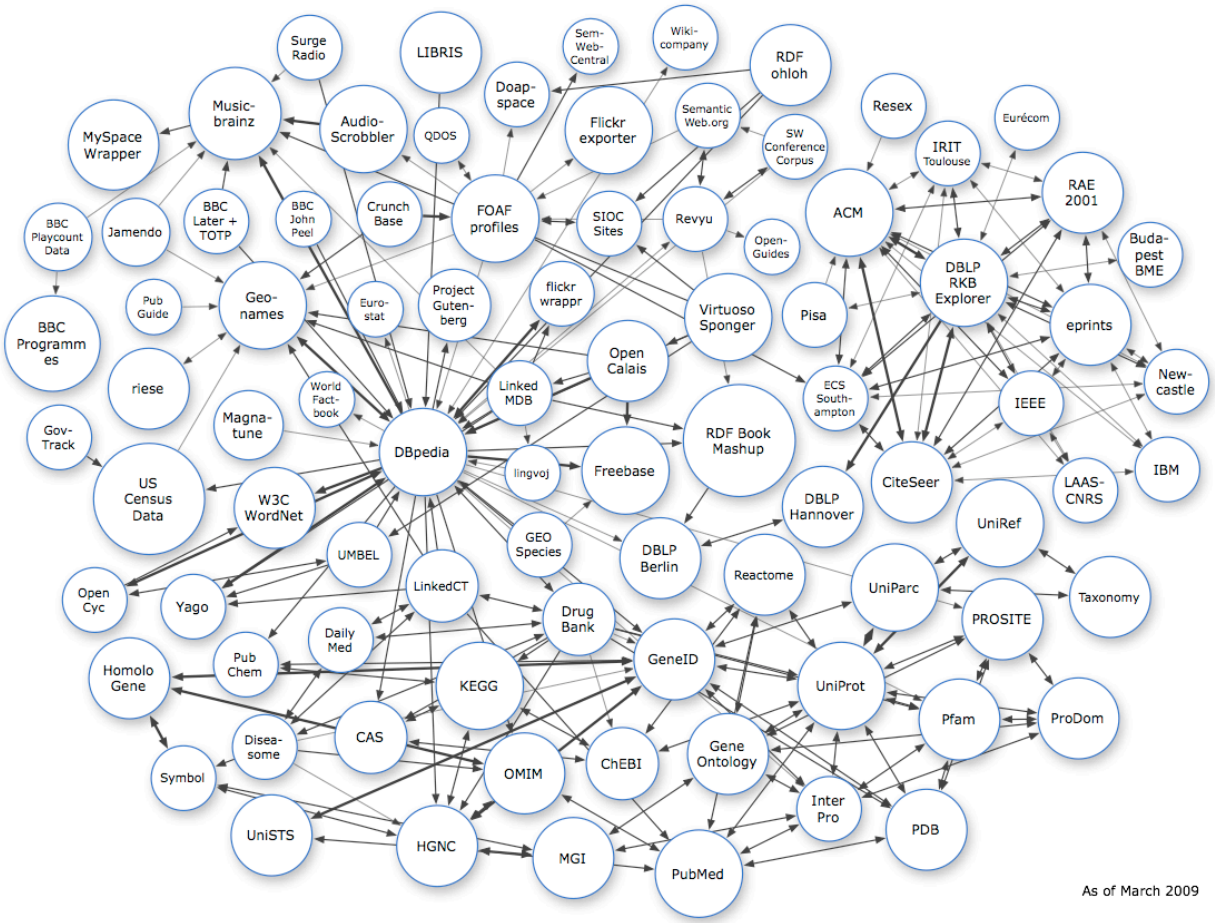
Policy-Aware Pipes - Accountability in Mashup Service of Linked Data

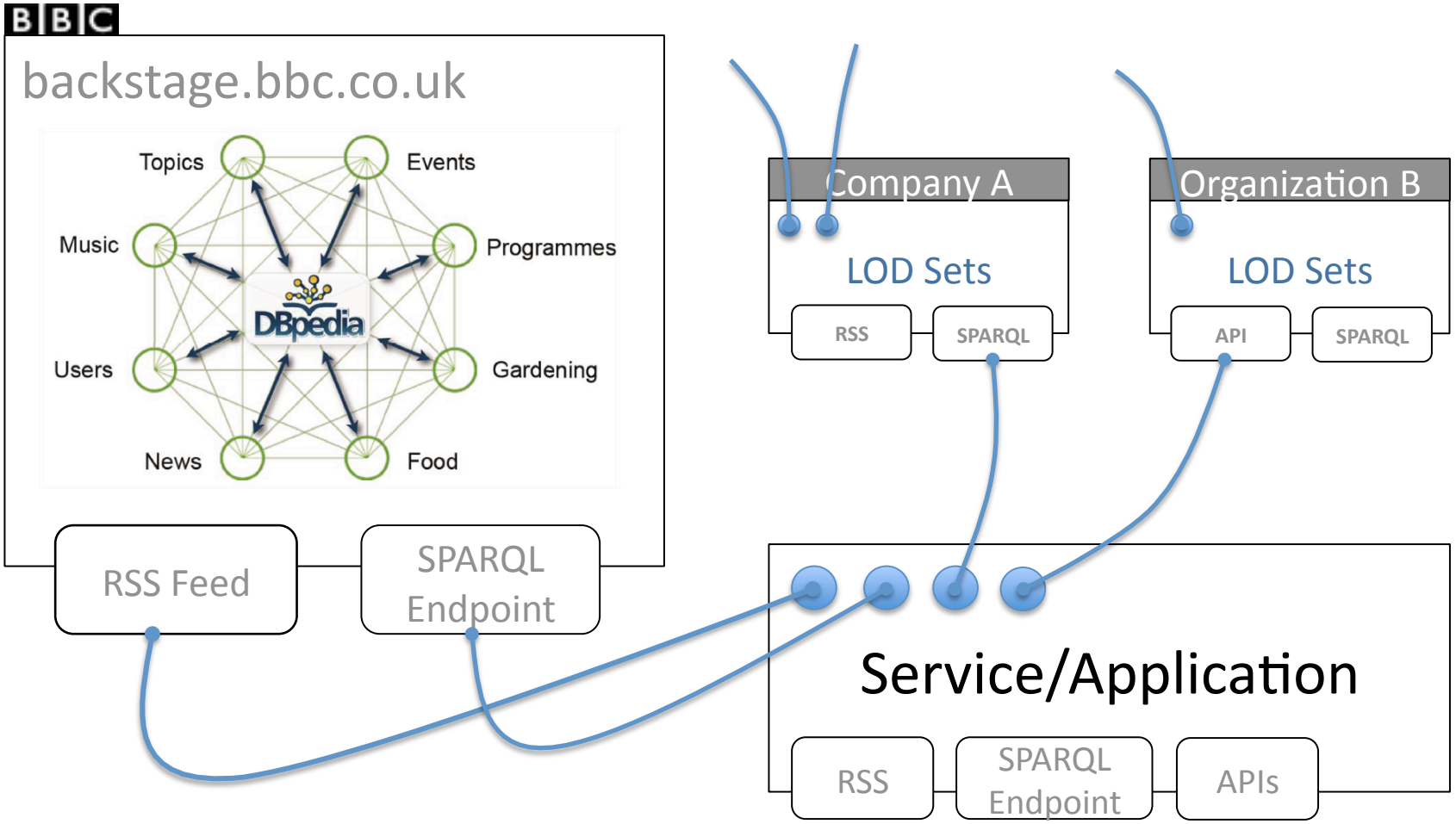
Fuming Shih

MIT CSAIL
Decentralized Information Group

LOD Datasets



In Practice …


BBC
backstage.bbc.co.uk
Topics
Events
Music
Programmes
DBpedia
Users
Gardening
News
Food
RSS Feed
SPARQL Endpoint
Company A
LOD Sets
RSS
SPARQL
Organization B
LOD Sets
API
SPARQL
Service/Application
RSS
SPARQL Endpoint
APIs

Scenario(1)

- Joe is a pop-music fan and a Web expert
- He creates a site called ``music-meme.org''
  - Up-to-date articles and news about pop music
  - Publishes RSS feed
  - Music notes: mashup information and reviews
  - Joe reads policies stated on BBC's website
    - http://welcomebackstage.com/about/faqs/#canimakemoney

Joe's Checklist of BBC's Policy

- You must not: Charge users for accessing your work that contains or uses BBC Content
  - Sell applications that use or incorporate BBC Content
  - Re-edit or re-contextualize BBC Content in any way that is illegal, is or is likely to bring the BBC into disrepute or is otherwise inappropriate
- However, you own the Intellectual Property rights in and to your application and nothing prevents you being able to sell or commercially license your application and make money from it.

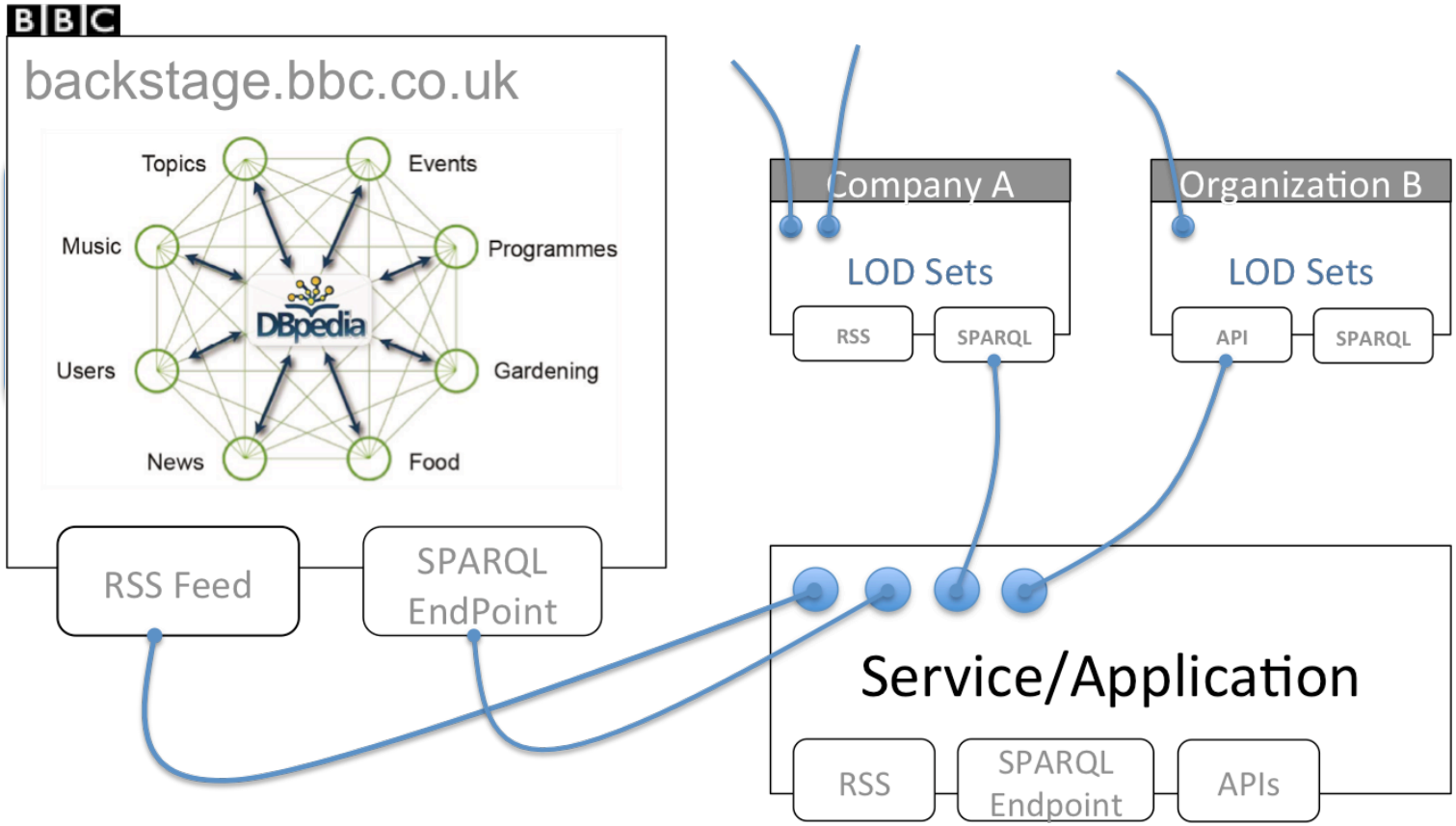
Scenario(2)

- Bob is a software engineer of website Alpha Music Box, an online music store.
  - Charges users for downloading songs
  - Bob implements music review by obtaining feeds from music-meme.org
  - Eventually, albums with better reviews → better sales!
    - Bob's assumes everything is free for reuse from music-meme.org
    - Use of music reviews in a commercial mashup possibly infringes BBC's policy

Observed Problems

- Policy is not machine readable from the source
- Policy should be propagated as well as data in the mashup
- More than just access control - information accountability needed
  - assertions of intended usage on data (e.g. Creative Commons)
  - help people who tend to comply with the policy
  - increase trust

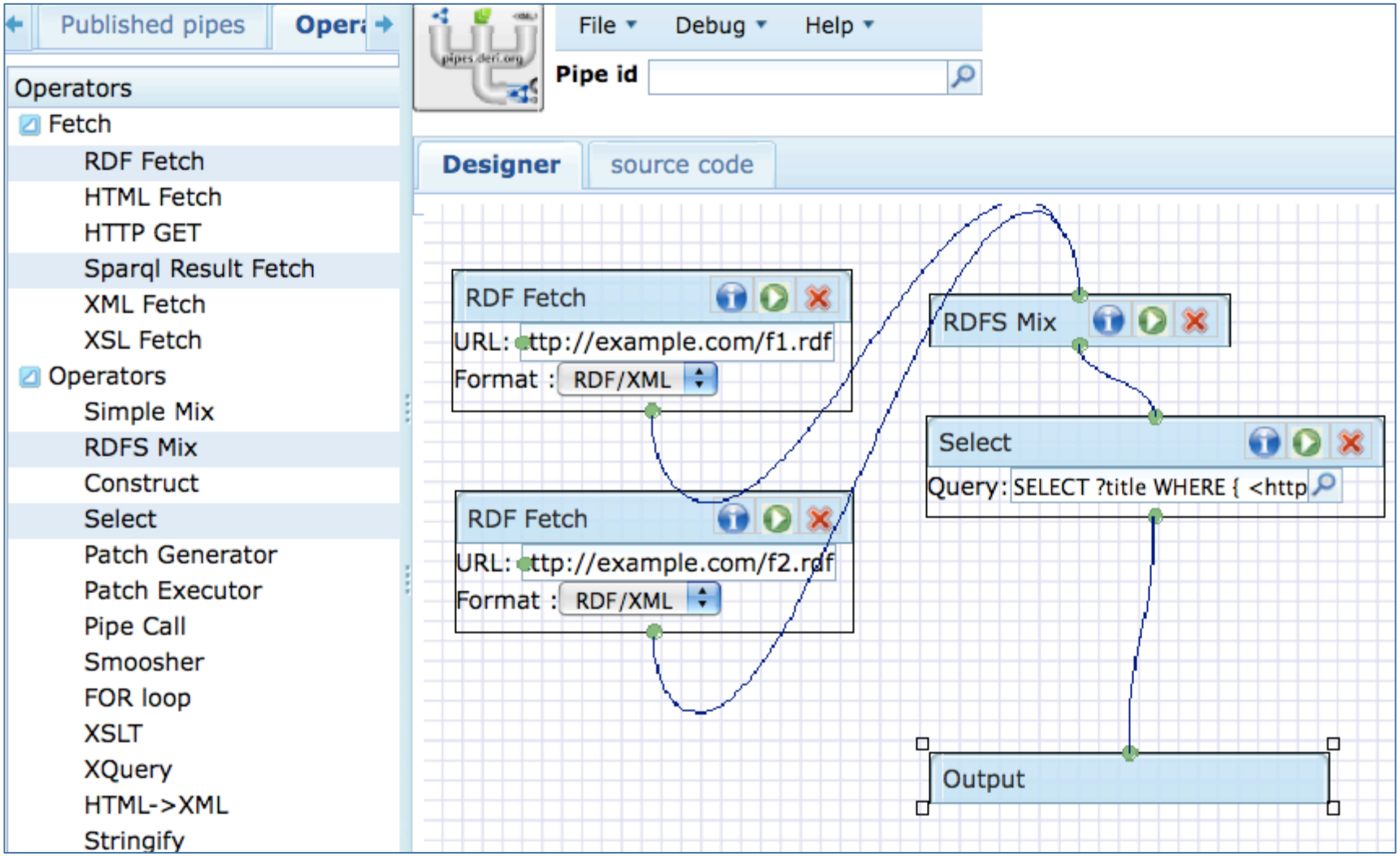
Requirements of Policy Support for Information Accountability



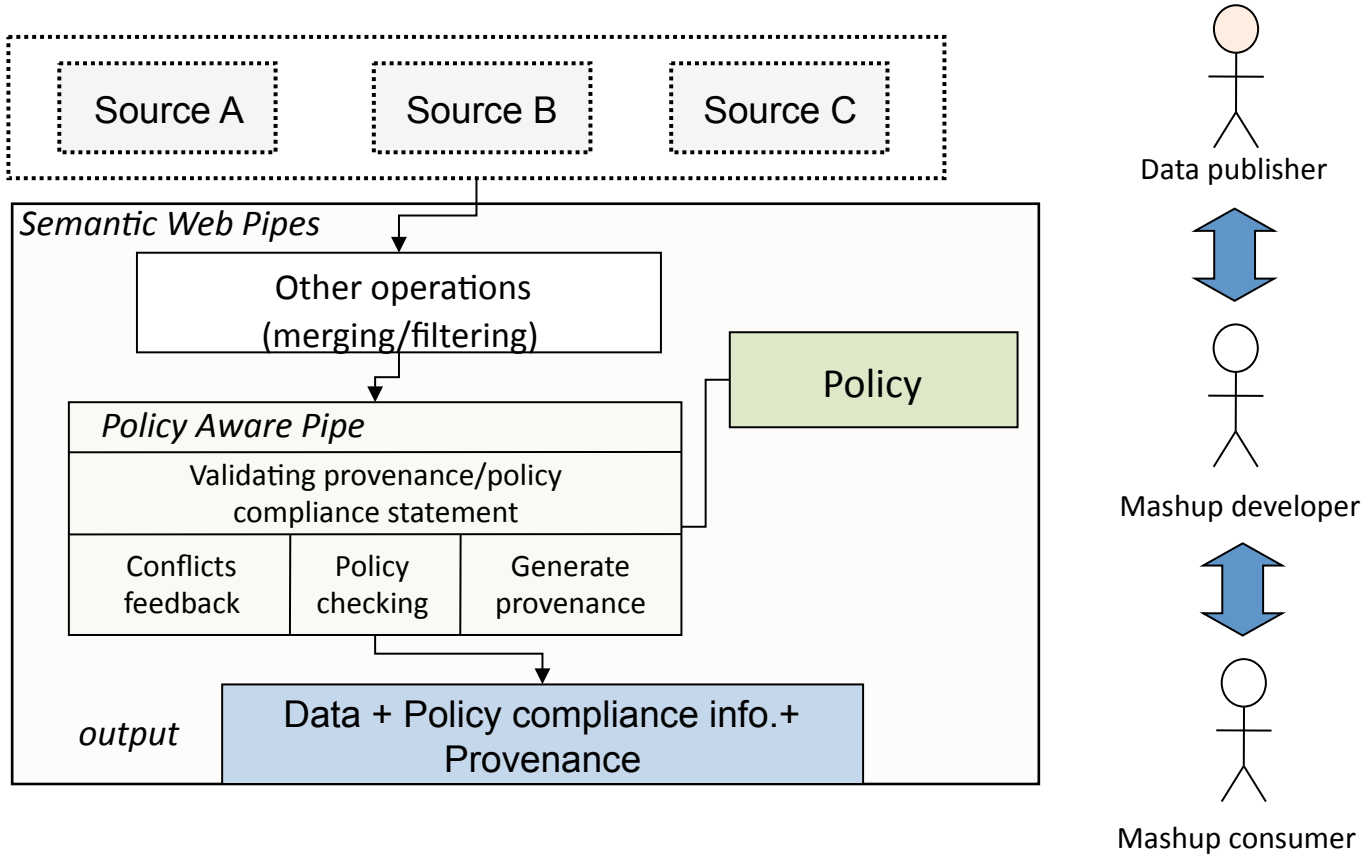
Pipes

- Rapid prototyping of web content
  - Concepts of the workflow system
  - Graphic editor + built-in operators/filters
  - Easy for casual users to reuse web content
    - Fetch data from multiple sources
    - Apply built-in filters and integrate data
    - Output as web pages or another data stream
  - Deri's Semantic Pipes(Linked data) → Policy Aware Pipes (policy+)

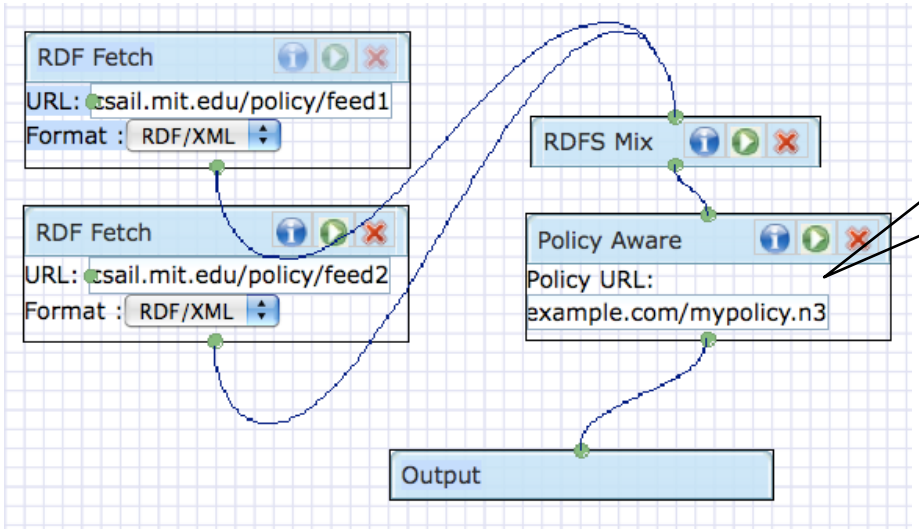
Deri's Semantic Pipes



Policy Aware Pipes



Editors & Debugging



log	type	timestamp	message
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_660	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Profile of http://foolme.csail.mit.edu/wsprofleB#profile is conflicted with policy :http://foolme.csail.mit.edu/policy/policyA#wspolicy"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_661	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Rejecting profile with author:http://www.w3.org/People/djweitzner/foaf#DJW"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_662	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Error]Some merged web service is conflicted with the policy of channel:http://meerkat.oreillynet.com/?_fl=rss1.0"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_679	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Profile of http://foolme.csail.mit.edu/wsprofleA#profile is conflicted with policy:http://foolme.csail.mit.edu/policy/policyB#wspolicy"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_682	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Rejecting profile with subject:news"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_684	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Error]Some merged web service is conflicted with the policy of channel:http://johnson.freenet.net/?_fl=rss1.0"

Scenario Walkthrough

- Ontologies to
  - Policy ontolo

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ws: <http://foolme.csail.mit.edu/ns/wslite#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

ws:profile      rdf:type  rdfs:Class;
                rdfs:label "Web service profile"^^xsd:string .

ws:author       rdf:type                rdf:Property;
                rdfs:subPropertyOf      dc:creator;
                rdfs:label              "web service author"^^xsd:string .

ws:subject      rdf:type                rdf:Property;
                rdfs:subPropertyOf      dc:subject.
```

```
@prefix rdf: <http://www.w3.org/1999/02
@prefix rdfs: <http://www.w3.org/2000/0
@prefix xsd: <http://www.w3.org/2001/X
@prefix p: <http://foolme.csail.mit.edu/n
@prefix dc: <http://purl.org/dc/elements/
@prefix foaf: <http://xmlns.com/foaf/0.1>

p:wspolicy  rdf:type  rdfs:Class;
            rdfs:label "Web service polic

p:creator       rdf:type            rdf:P
                rdf:subPropertyOf   dc:creator;
                rdfs:domain         rdfs:Resource.

p:rejectSubject  rdf:type                   rdf:Property;
                 rdfs:subPropertyOf         dc:subject.

p:rejectPerson  rdf:type                    rdf:Property;
                rdfs:subPropertyOf          foaf:Person.
```

Data Publisher – BBC Backstage

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns="http://purl.org/rss/1.0/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cc="http://creativecommons.org/ns#”
  xmlns:ws="http://foolme.csail.mit.edu/ns/ws#"

  <channel rdf:about="http://www.example.com/?_fl=rss1.0">
    <title>Some Data Source Feed</title>
    <link>http://www.example.com</link>
    <description>A data source with policy and service profile</description>
    <cc:license rdf:resource="http://www.creativecommons.org/licenses/by-nc-nd/3.0/us/" />
    <cc:morePermissions rdf:resource="http://foolme.csail.mit.edu/policy/policyA#wspolicy" />
                <ws:profile rdf:resource="http://foolme.csail.mit.edu/policy/wsprofileA#" />
                …
  </channel>
```

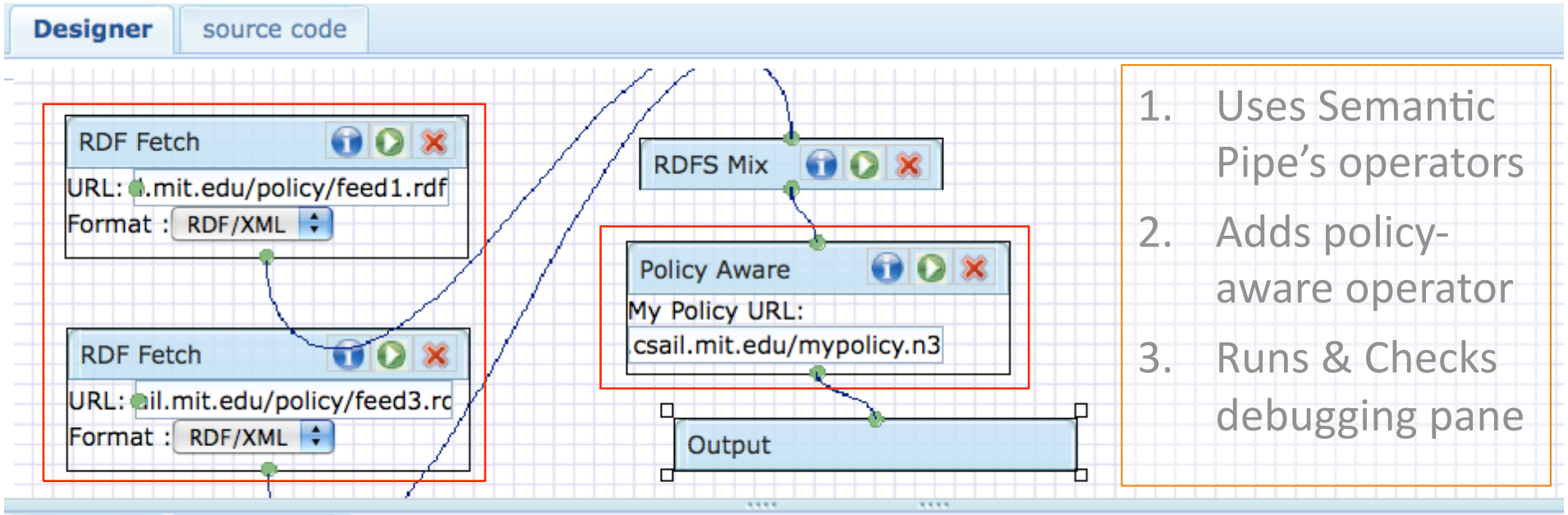
```
<>  dc:title  "BBC Backstage Data Policy.
:wspolicy  a   p:wspolicy ;
          p:rejectSubject  "commercial"^^xsd:string ;
          p:creator   http://backstage.bbc.co.uk/;
```

Mashup Developer

- Joe uses PAP to construct his mashup
  - Check policy compliance
    - Use each source's policy to check all other data sources' profile
    - Use it's own policy to check all data sources
  - Generate provenance data
  - Publish RSS feed of music-meme.org

```
<>  dc:title   "Music-meme Profile".
p:profileA  a  ws:profile ;
        ws:subject  "community"^^xsd:string ;
        ws:author   http://dig.csail.mit.edu/People/Joe#I;
```

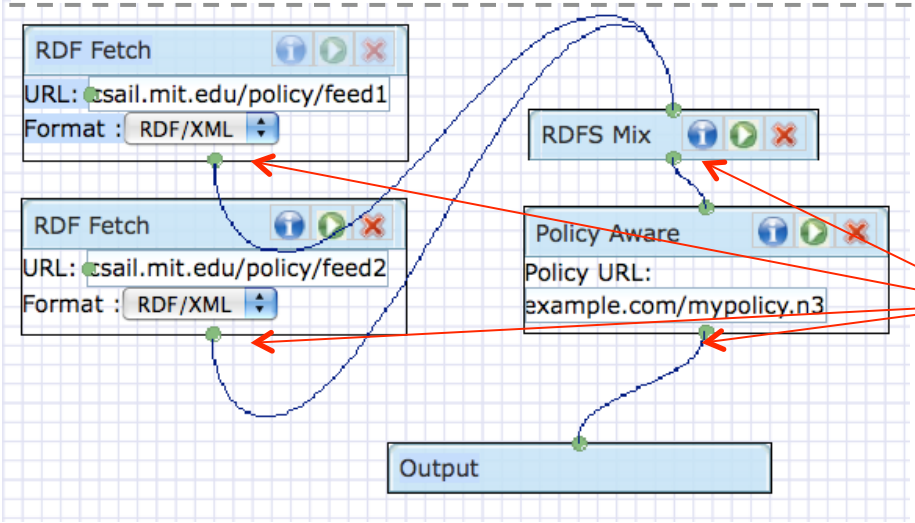
Using PAP in Semantic Pipes



log	type	timestamp	message
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_874	"POLICY"	"2009-10-21T16:18:07-04:00"	"[Success]All ws sources comply with the policy of channel:http://meerkat.oreillynet.com/?_fl=rss1.0"
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_885	"POLICY"	"2009-10-21T16:18:07-04:00"	"[Success]All ws sources comply with the policy of channel:http://www.w3.org /2000/08/w3c-synd/home.rss"
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_885	"POLICY"	"2009-10-21T16:18:07-04:00"	"Validate, done!"
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_911	"POLICY"	"2009-10-21T16:18:07-04:00"	"[Success]Your service comply with the policy of channel:http://meerkat.oreillynet.com /?_fl=rss1.0"
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_917	"POLICY"	"2009-10-21T16:18:07-04:00"	"[Success]Your service comply with the policy of channel:http://www.w3.org/2000/08 /w3c-synd/home.rss"
http://foolme.csail.mit.edu/policy /log/log2009-10-21#16-18-07_921	"POLICY"	"2009-10-21T16:18:07-04:00"	"Merge my rss feed, done"

Provenance generation

```
<rdf:Description rdf:about="http://music-meme.org/policy/pro2009-07-10#13-10-25_317">
   <rdf:type rdf:resource="http://foolme.csail.mit.edu/ns/provenance#annotation"/>
    <provenance:outpout rdf:resource="http://music-meme.org/sparqlendpoint/"/>
    <provenance:timestamp>2009-07-10T13:10:25-04:00</timestamp>
    <provenance:sparql>
          "SELECT ?rev ?name ?title WHERE { ?rev a rev:Review; rev:reviewer ?person;
                foaf:primaryTopic ?record. ?record dc:title ?title; foaf:maker ?artist.
                ?artist foaf:name ?name.}"
          </sparql>
    <provenance:source rdf:resource= "http://api.talis.com/stores/bbc-backstatge"/>
   <provenance:compliant-with rdf:resource="http://api.talis.com/policy#wspolicy"/>
</rdf:Description>
```



PAP logs all the operations in the Semantic Pipes, saves logs as triples.

Mashup Consumer

- Bob uses PAP to build another mashup from Joe's *music-meme.org's* data feed

log	type	timestamp	message
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_660	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Profile of http://foolme.csail.mit.edu/wsprofleB#profile is conflicted with policy :http://foolme.csail.mit.edu/policy/policyA#wspolicy"
http://foolme.csail.mit.edu/policy /log/log2009-07-13#00-00-32_661	"POLICY"	"2009-07-13T00:00:32-04:00"	"[Alert]Rejecting profile with author:http://www.w3.org/People/djweitzner/foaf#DJW"

```
<rdf:Description rdf:about="http://music-meme.org/policy/pro2009-07-10#13-10-25_317">
   <rdf:type rdf:resource="http://foolme.csail.mit.edu/ns/provenance#annotation"/>
    <provenance:outpout rdf:resource="http://music-meme.org/sparqlendpoint/"/>
    <provenance:timestamp>2009-07-10T13:10:25-04:00</timestamp>
    <provenance:sparql>
          "SELECT ?rev ?name ?title WHERE { ?rev a rev:Review; rev:reviewer ?person;
                foaf:primaryTopic ?record. ?record dc:title ?title; foaf:maker ?artist.
                ?artist foaf:name ?name.}"
          </sparql>
   <provenance:source rdf:resource= "http://api.talis.com/stores/bbc-backstatge"/>
   <provenance:compliant-with rdf:resource="http://api.talis.com/policy#wspolicy"/>
</rdf:Description>
```

Conclusion

- PAP integrates policy into mashup development
  - Different policy support to different actors in the mashup cycle
- Introduce information accountability to mashup environment
  - Transparency of policy about data usage
  - Increase trust between mashups

Questions?