



# Scout

Design and architecture of a novel  
preservation watch system

Luís Faria

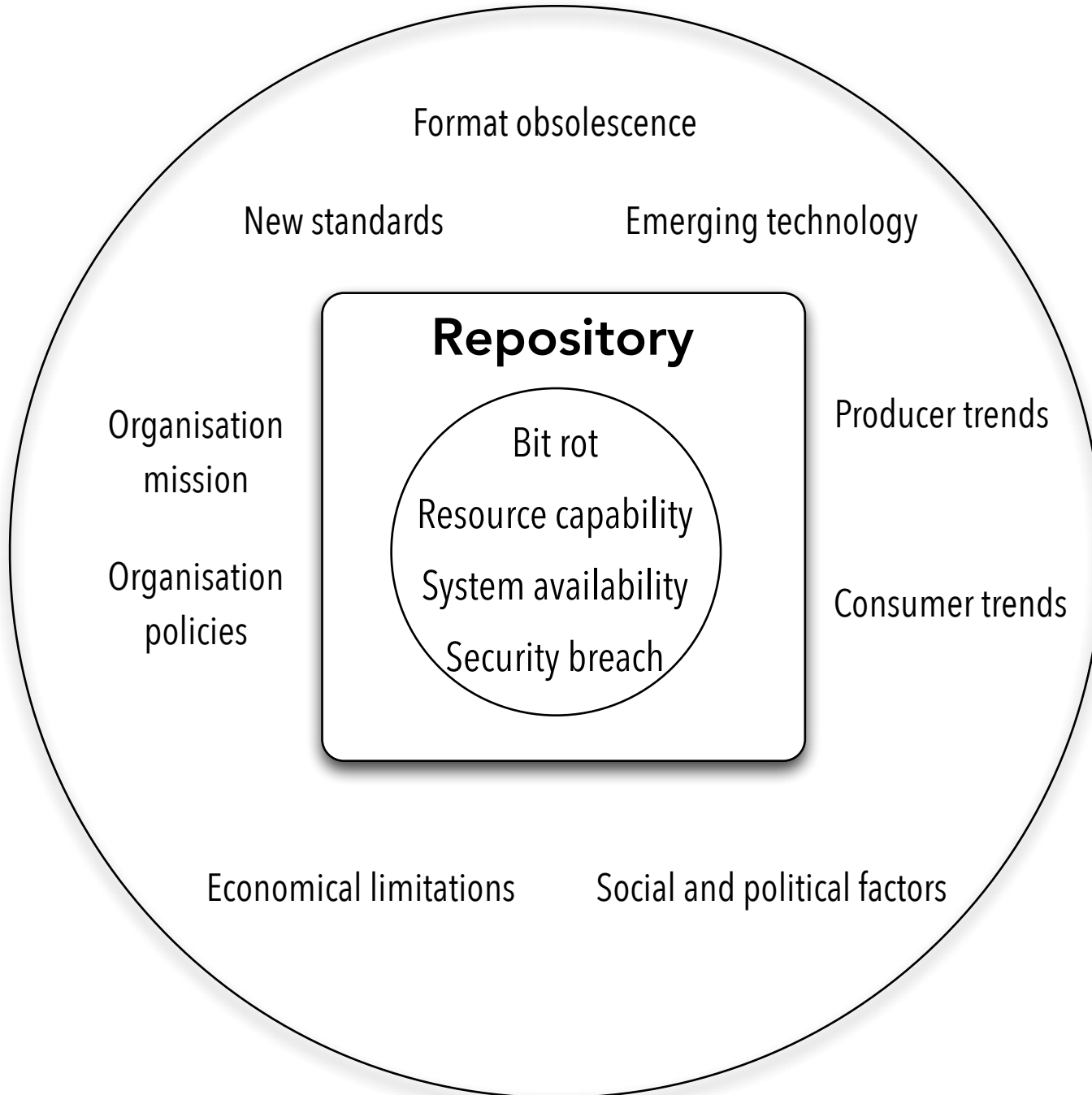
[lfaria@keep.pt](mailto:lfaria@keep.pt)

## Outline

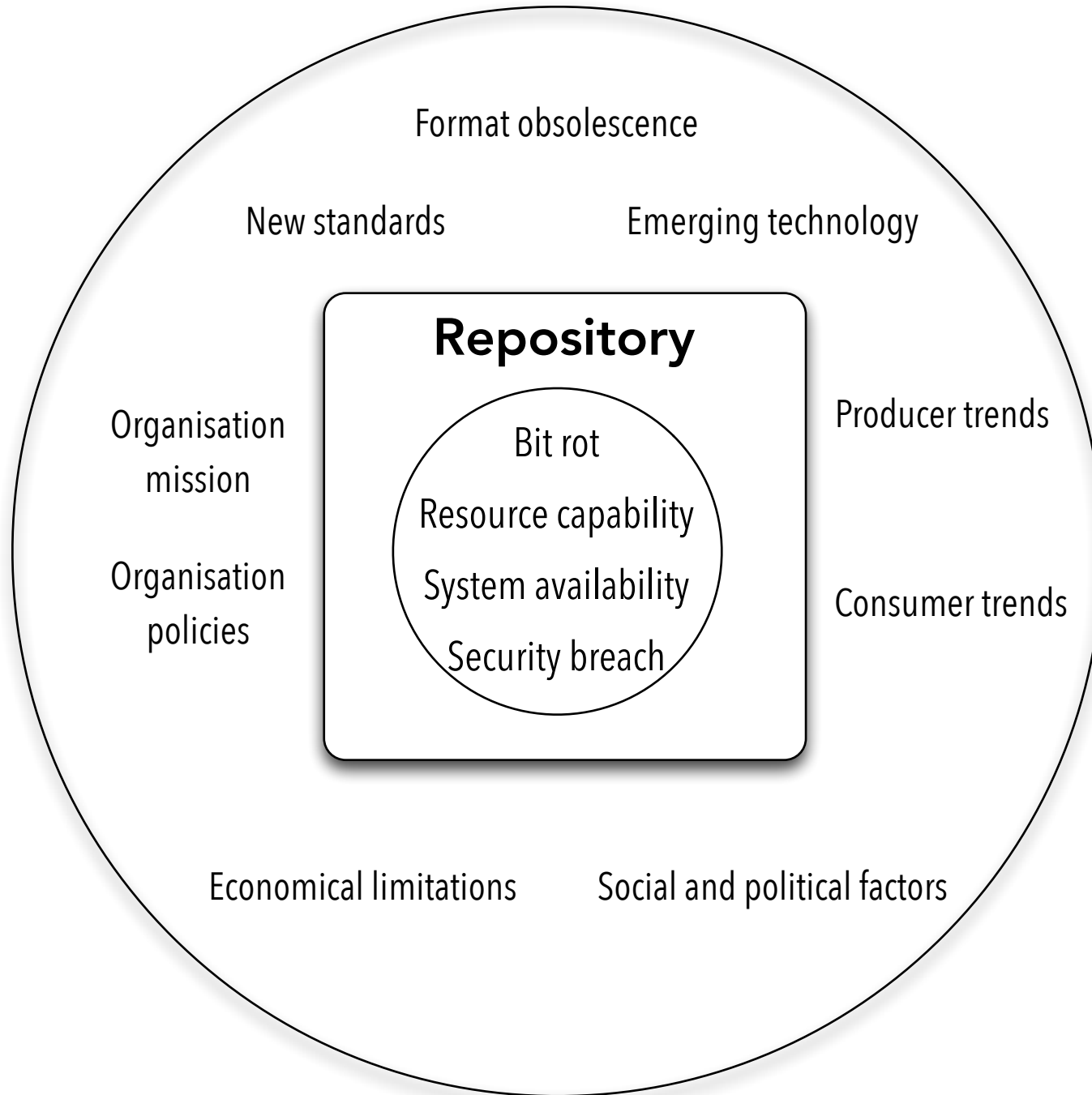
- Preservation monitoring
  - Why and what is needed
  - State of the art
- A novel approach
  - Methodology
  - Architecture
  - How can you participate?

# Preservation monitoring

# Why do we need monitoring?



# Why do we need monitoring?



**Risks**

**Opportunities**

## State of the Art

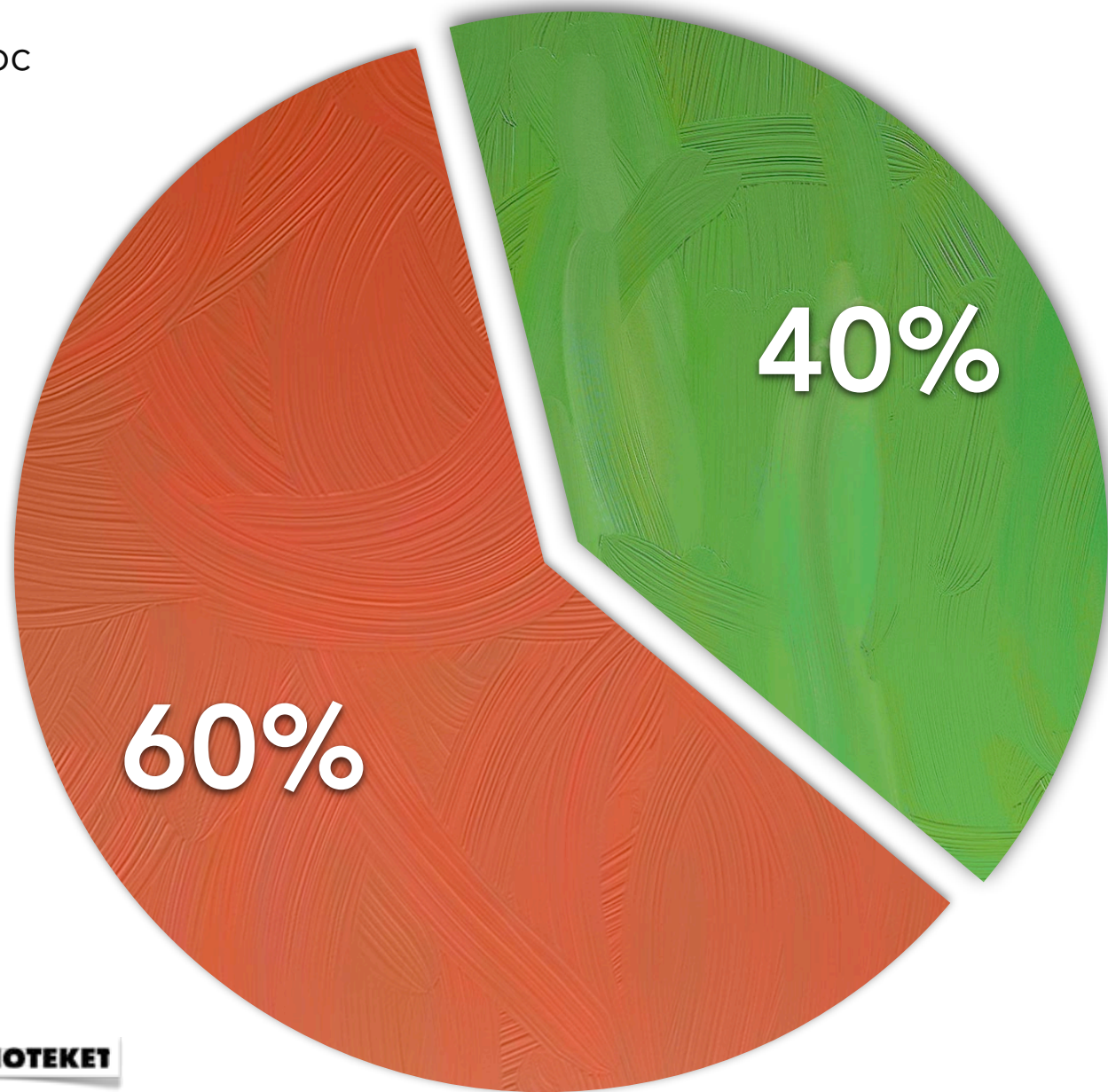
- Digital Format Registries
- AONS
- Technology watch reports

## State of the Art

- Digital Format Registries
  - Lack of coverage
  - Statically-defined generic risks
  - Lack of structure in risks
  - Focus on format obsolescence
- AONS
  - Total dependency on format registries
- Technology watch reports
  - Machine unreadable

# Risk Assessment

- Yes but manual and adhoc
- None



Survey on:



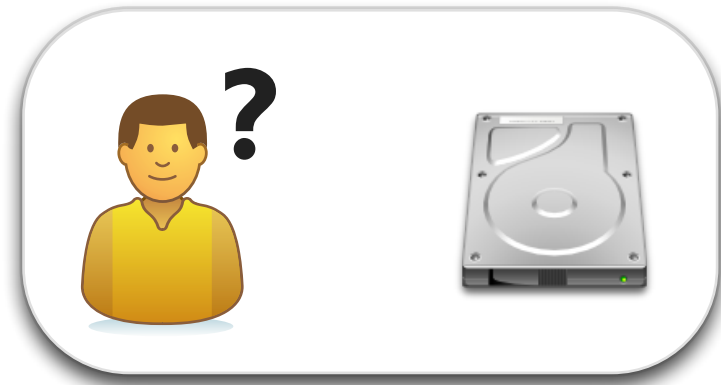


## What is needed?

- We need data!
  - From anywhere and everywhere
  - Sharing
- Usability & Scalability
  - Structured data
  - Controlled vocabulary

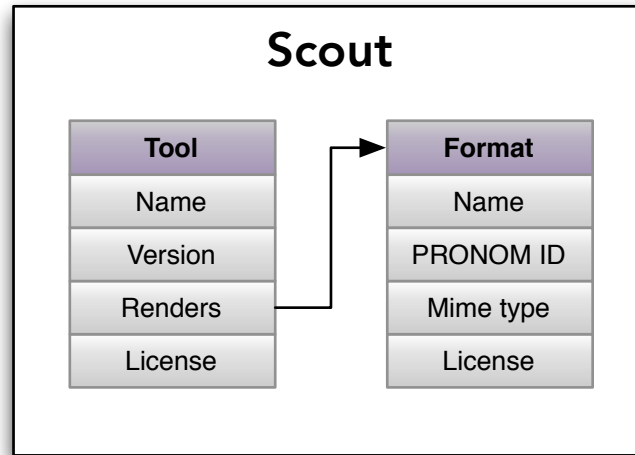
# Scout

A novel approach



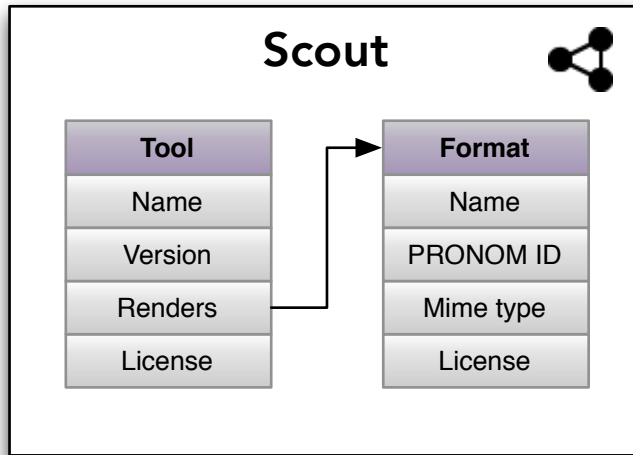


?



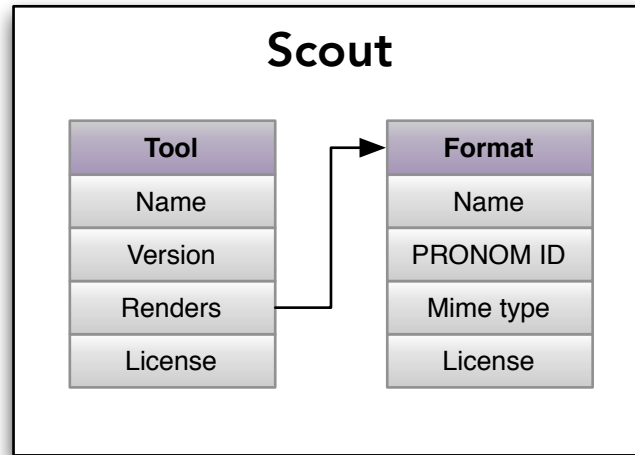


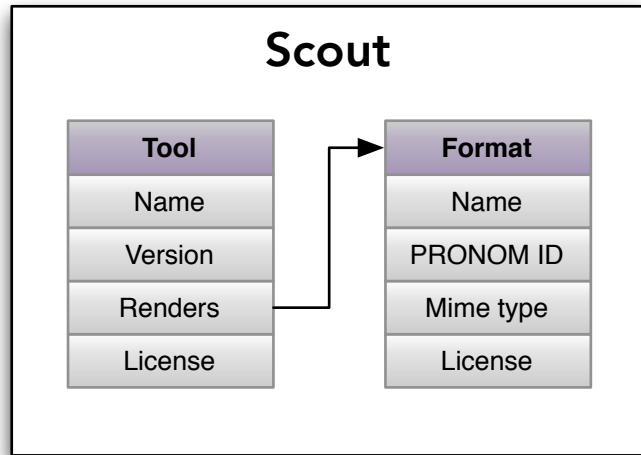
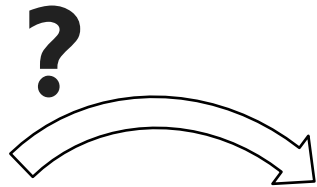
?





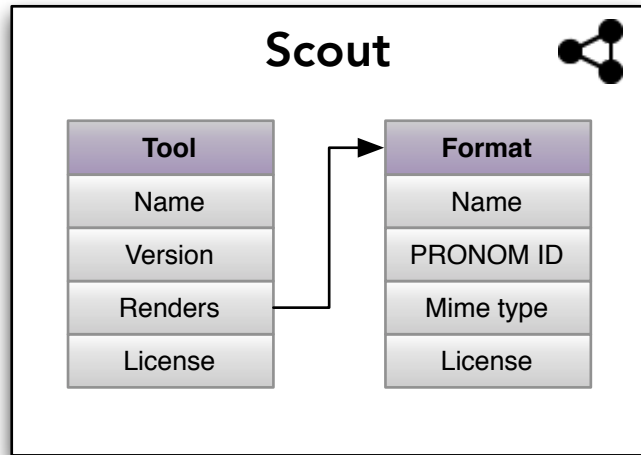
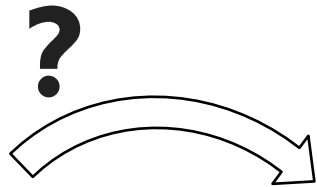
?





**PRONOM**

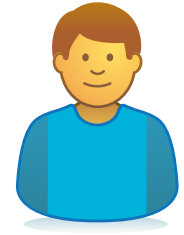
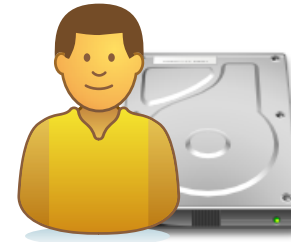
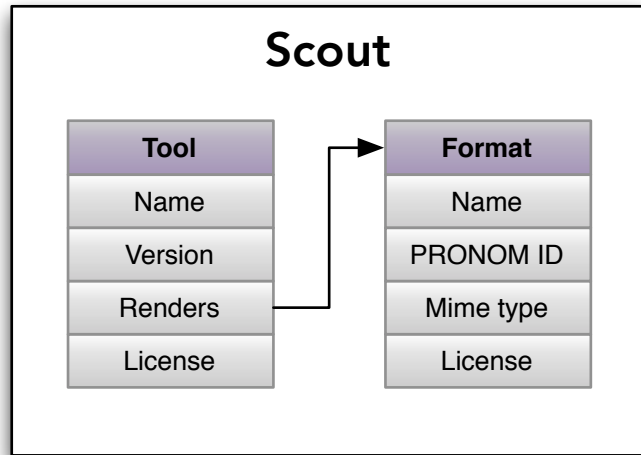
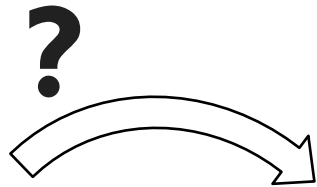




**PRONOM**







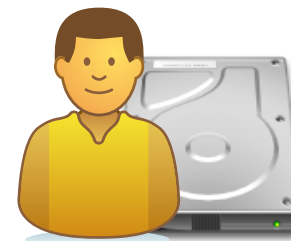
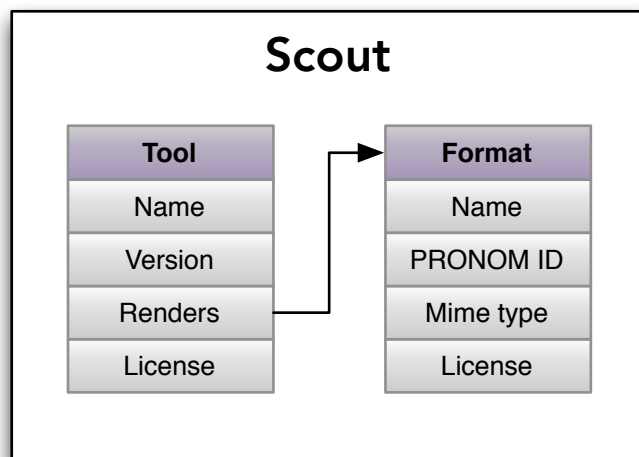
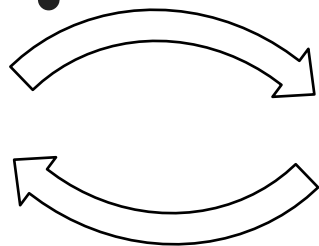
ConversionSoftwareRegistry

**PRONOM**



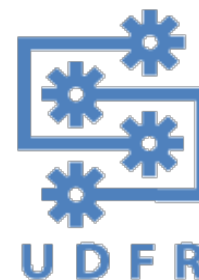


?



ConversionSoftwareRegistry

**PRONOM**



## Methodology

- Survey practitioners on information to be monitored
- Create structured data model
- Find sources of information (try to automate)
- Define notification triggers
- Frequent monitoring of sources of information
- Frequent assessment of triggers and notification

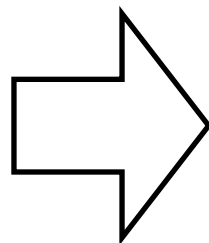
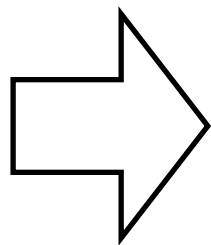
## Example questions

- Are there any tools that can render the format X?
- Is my repository the only one that has format Y?
- Are my preservation plans still valid?
- Are my repository policies being enforced?

## Information Sources

- Format registries & software catalogues
- Digital repositories & web archives
- Organizational objectives
- Experiments
- Simulation
- Human knowledge

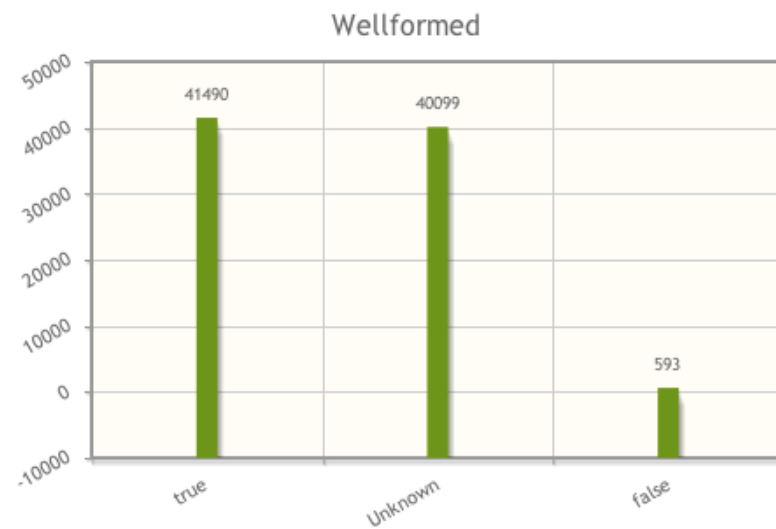
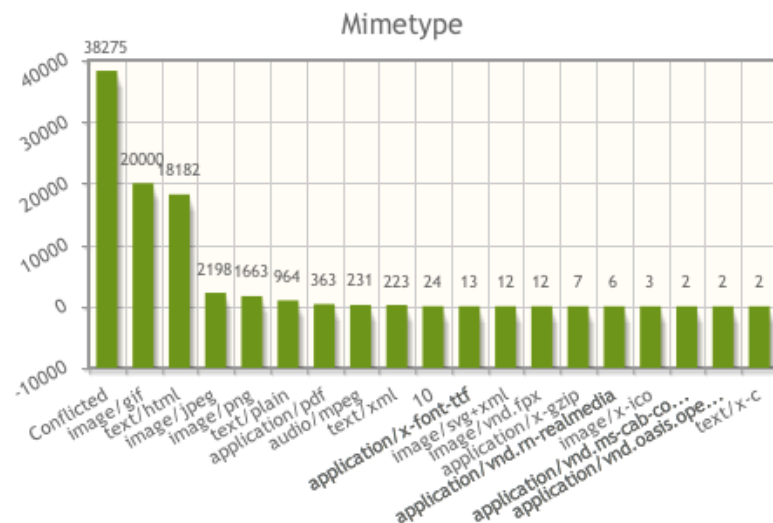
# C3PO Content profile tool



Characterization

Reports:

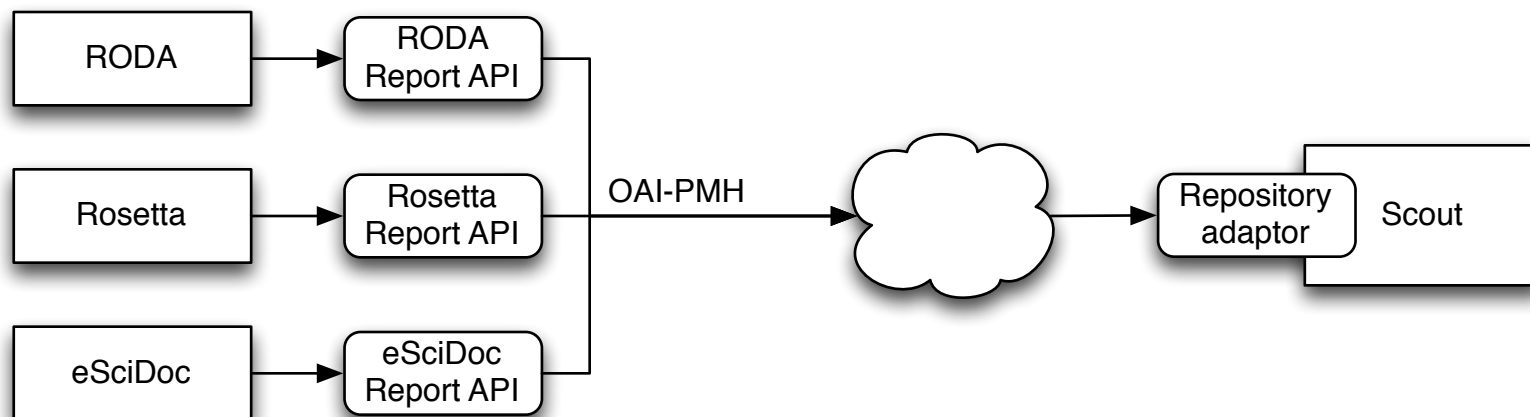
- Aggregation
- Analysis
- Representative datasets



<https://github.com/peshkira/c3po>

Petar Petrov <[petrov@ifs.tuwien.ac.at](mailto:petrov@ifs.tuwien.ac.at)>

# Repository events adaptor



- OAI-PMH with PREMIS
- Normalize events
- Fine-grain events
- History
- Events example
  - Ingest started/ended
  - Representation downloaded
  - Plan executed

# Define triggers

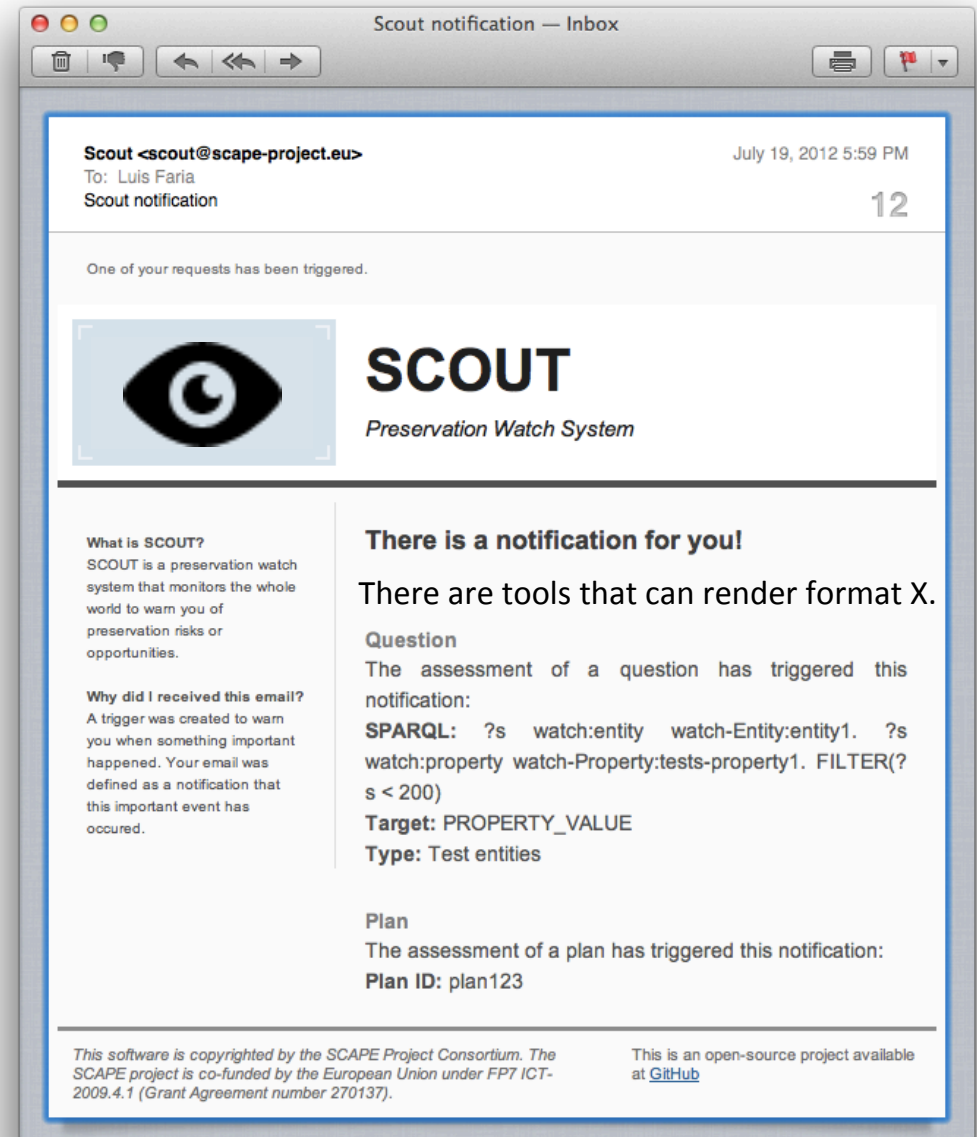
- Notify me when there are tools that can render the format X.

```
SPARQL Help
SELECT ?s WHERE { ?s rdf:type watch:Entity .
  ?s watch:type ?type .
  ?type watch:name "tools"^^xsd:String .
  ?value watch:entity ?s .
  ?value watch:property ?property
  ?property watch:name "renders"^^xsd:String .
  ?value watch:value "format X"^^xsd:String
}
```



# Receive notifications

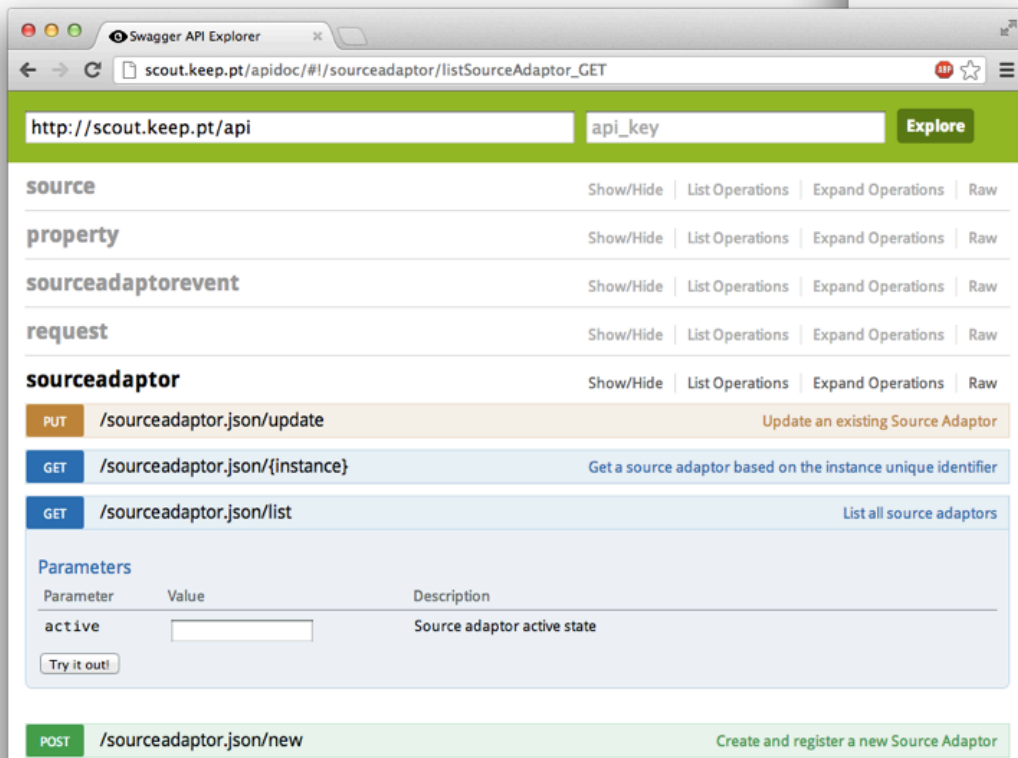
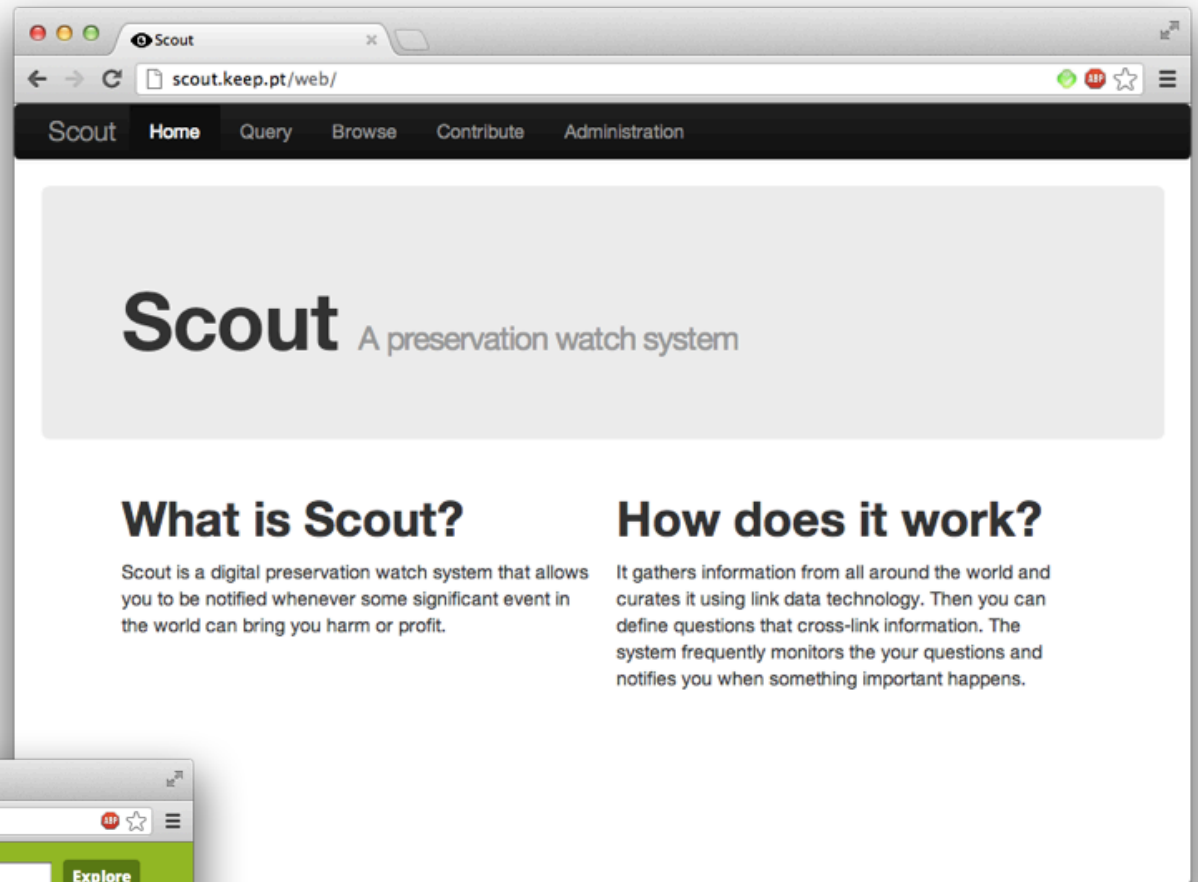
Email



HTTP Push API

# Interfaces

Web page



REST API

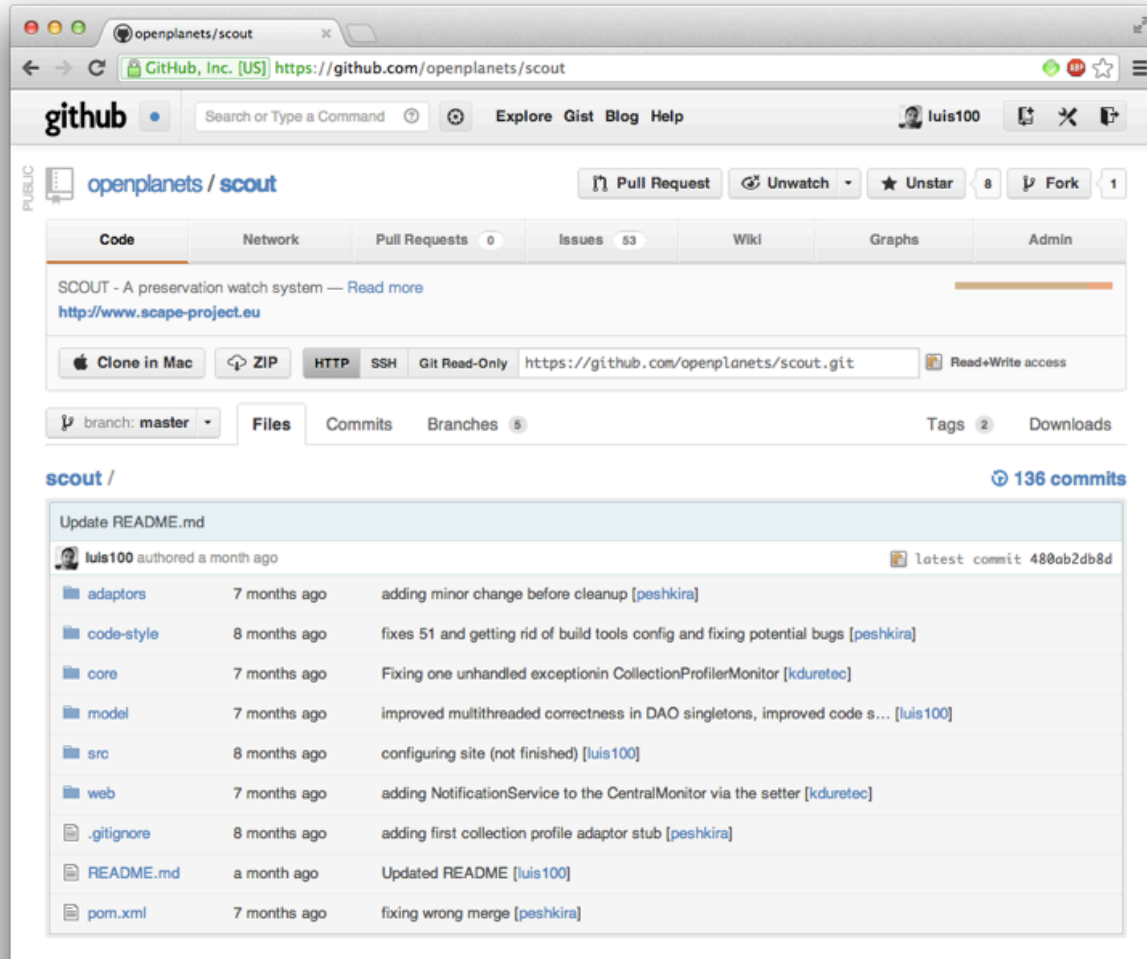
## How to be a part of Scout

- Join the surveys
  - Send your email to me <[lfaria@keep.pt](mailto:lfaria@keep.pt)>
- Integrate your content
  - Send your content profile with C3PO
  - Send repository events with Report API (soon)
- Contribute with information (soon)
  - Use Scout form for manual input of knowledge

# Scout



SCALable Preservation Environments  
<http://www.scape-project.eu>



Watch us on  
GitHub!

<https://github.com/openplanets/scout>

Luís Faria <[lfaria@keep.pt](mailto:lfaria@keep.pt)>

**KEEP**SOLUTIONS  
University of Minho SPIN-OFF

<http://www.keep.pt>