CodeSharing

CodeSharing:
a simple API for disseminating our TEI encoding

Martin Holmes
University of Victoria

The need

- Most TEI users are "self-taught or learned by doing" (Marjorie Burghart & Malte Rehbein 2012).
- Users need "a source for a compendium of examples suitable for inductive learning" (Stella Dee, forthcoming).
- We have TEI By Example, we have the TEI Guidelines, we have the TEI Wiki, we have Cheatsheets...and yet...
- ...you can never find enough good examples of how to encode what it is you actually want to encode.

The audience

- Novice encoders
- Researchers into encoding practices

Novice encoders...

- ...most often need to search the documents in their own project

- ...are not keen on this sort of thing:

  ```
  //div[@type='chapter']
  ```

- ...and especially not on this:

  ```
  <div\s+[^>]*type\s*=\s*"\s*chapter\s*"[^>]*>
  ```

Novice encoders

- May not have read access to the full repository
- Need easy access to example encoding
- Have no XPath or XQuery skills
- Find regular expressions overwhelmingly difficult
- Find the Guidelines intimidating and Guidelines examples often not directly relevant.

Researchers into encoding practices

- TEI Council: "Before we make a decision on this, we should find out what people actually do when they use <thing>."
- We ask on TEI-L.
- We look at projects we know (although most do not share their XML).
- We end up with unrepresentative samples.

Lou Burnard to TEI-L 2013-06-18:

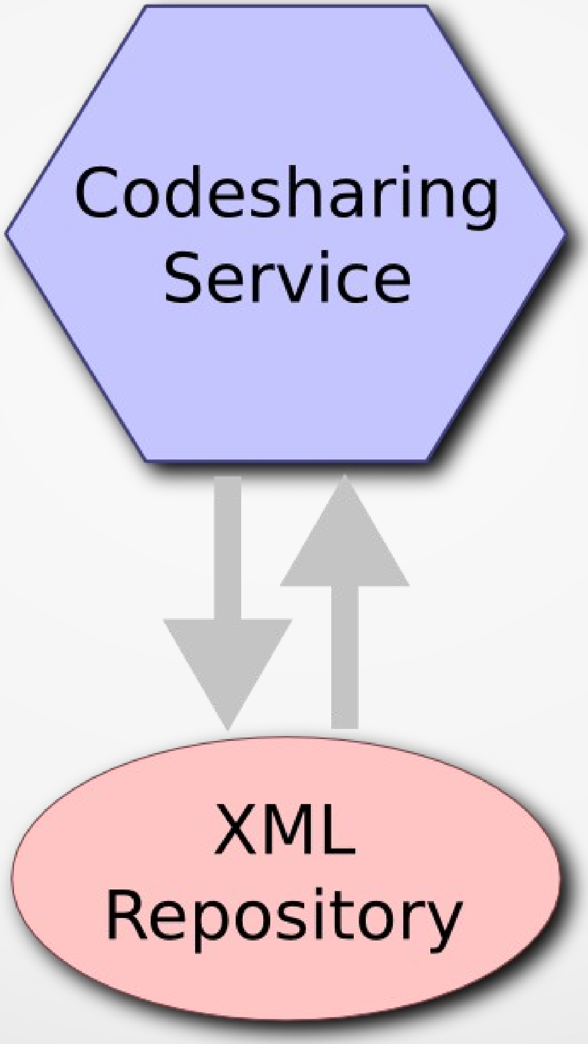
I've been asked by the TEI Council to CONSULT THE COMMUNITY about how the @corresp attribute is used, if indeed it is. The Guidelines are, to be fair, a bit vague on this particular topic ("indicates correspondence of any kind") so it is possible that divergent practices have developed...

DO YOU USE @CORRESP?

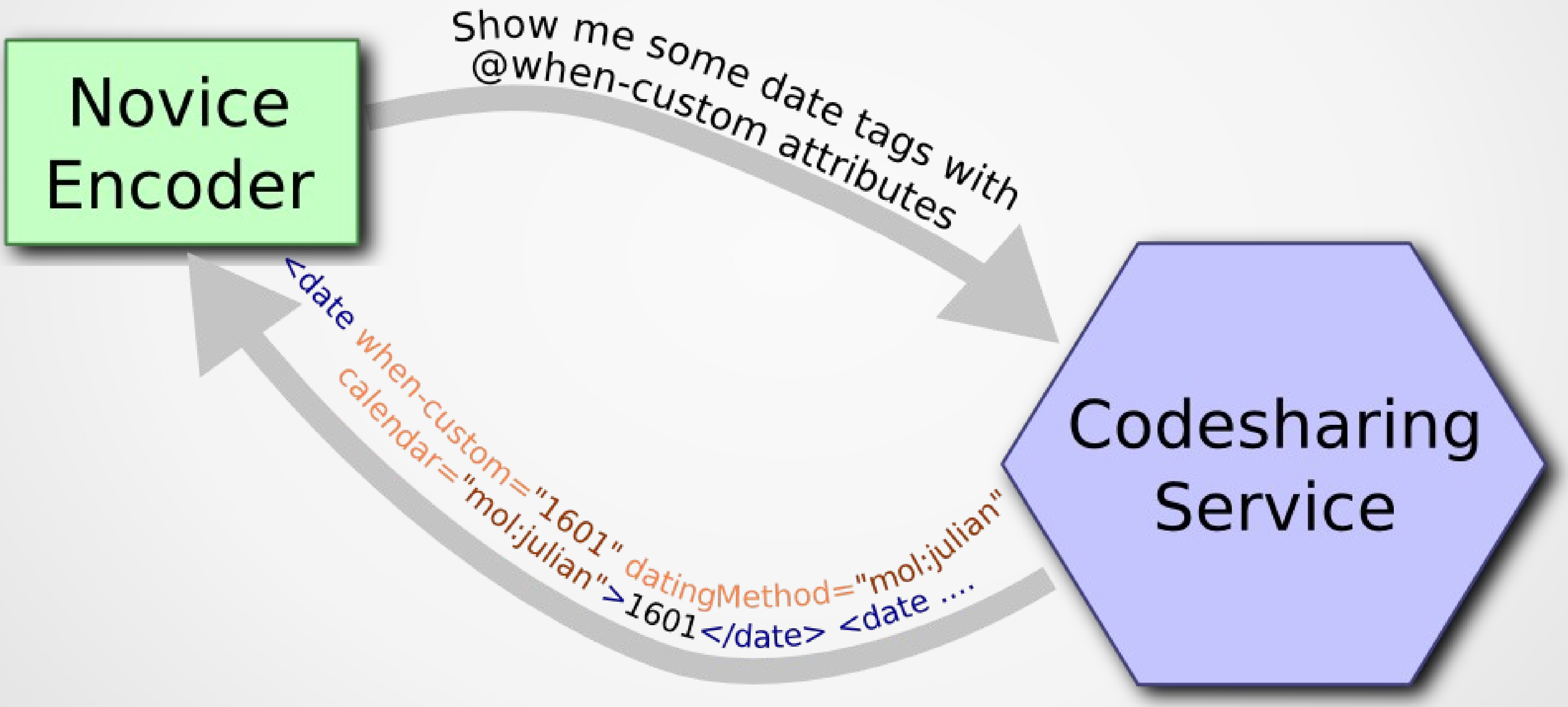
IF SO, FOR WHAT? (please supply an example or two!)

- 10 responses on-list
- 13 examples (+ some links)

The basic idea



The basic idea



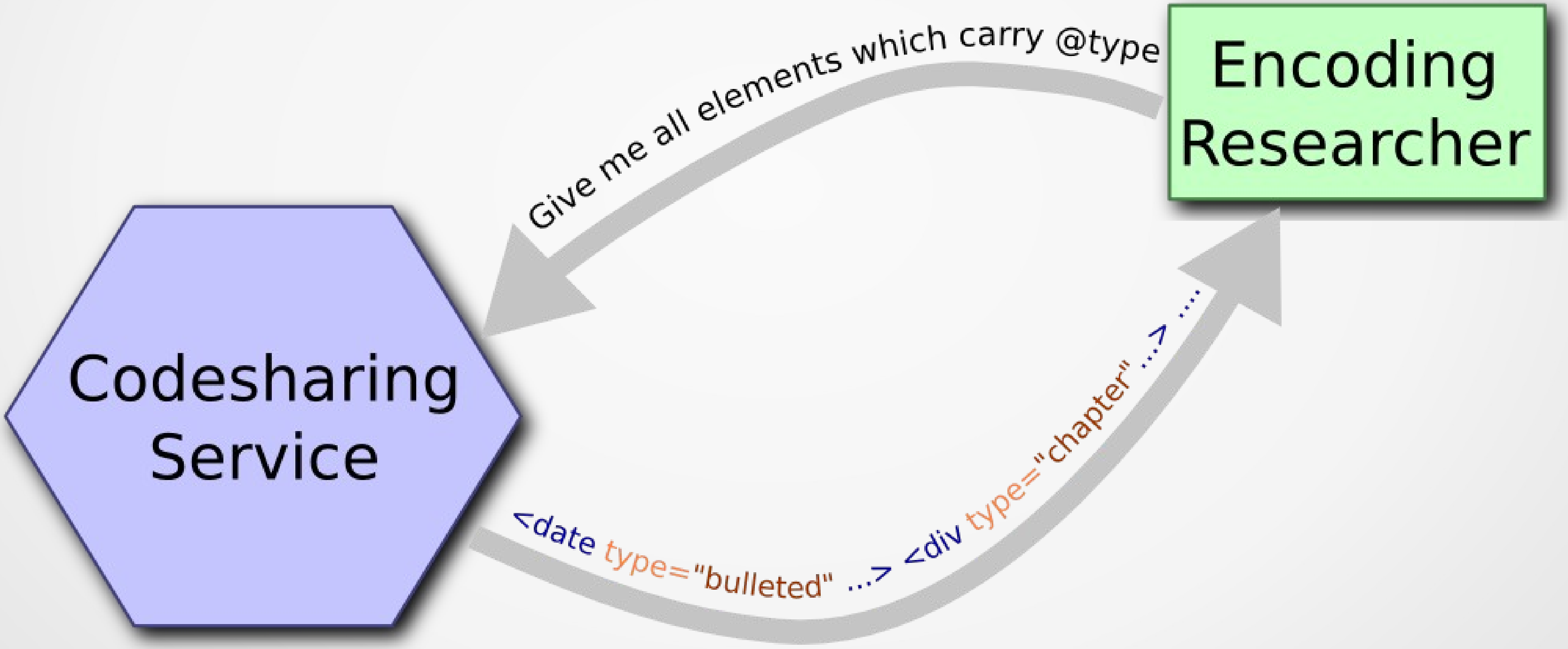
Novice Encoder

Show me some date tags with @when-custom attributes

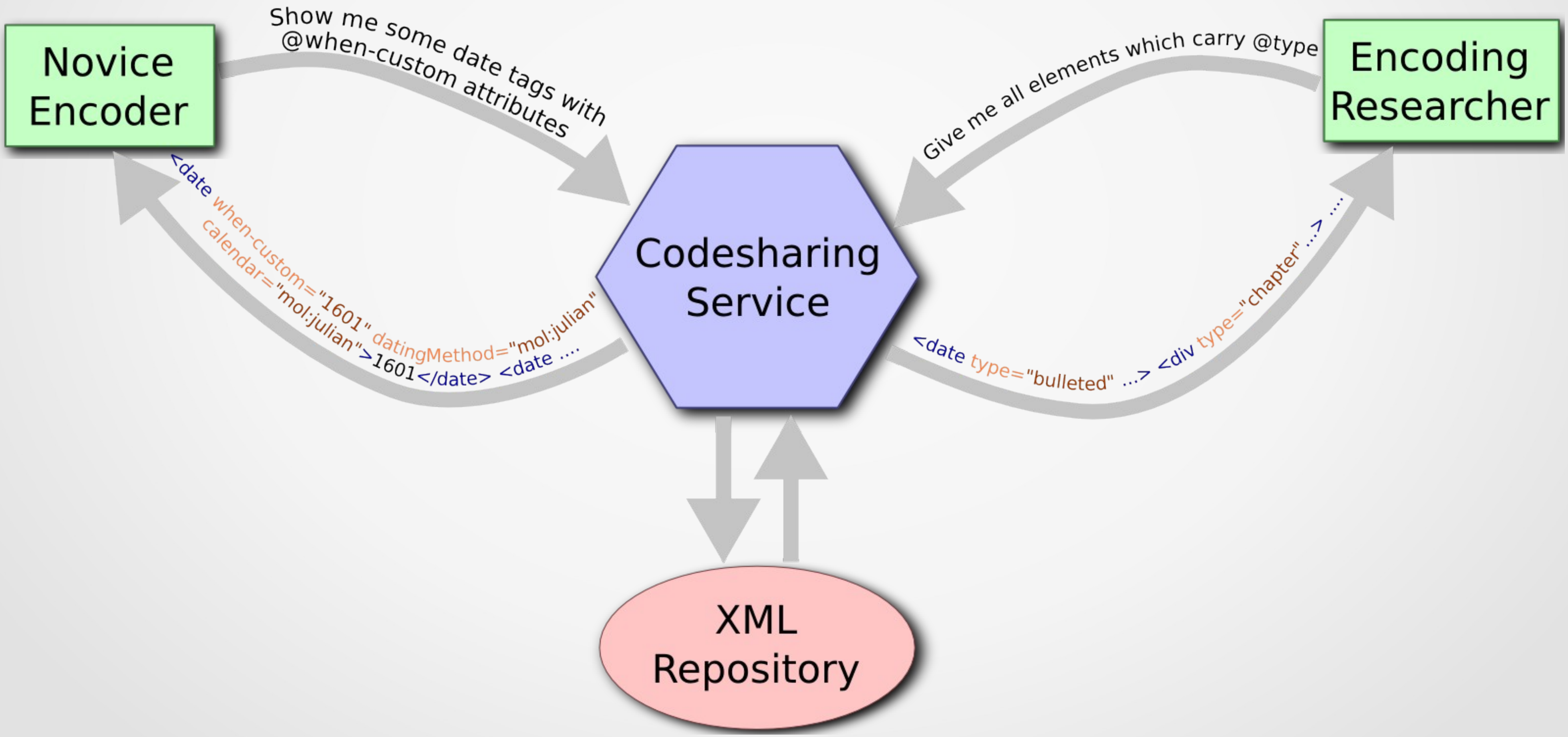
Codesharing Service

<date when-custom="1601" datingMethod="mol:julian" calendar="mol:julian">1601</date> <date ....

The basic idea

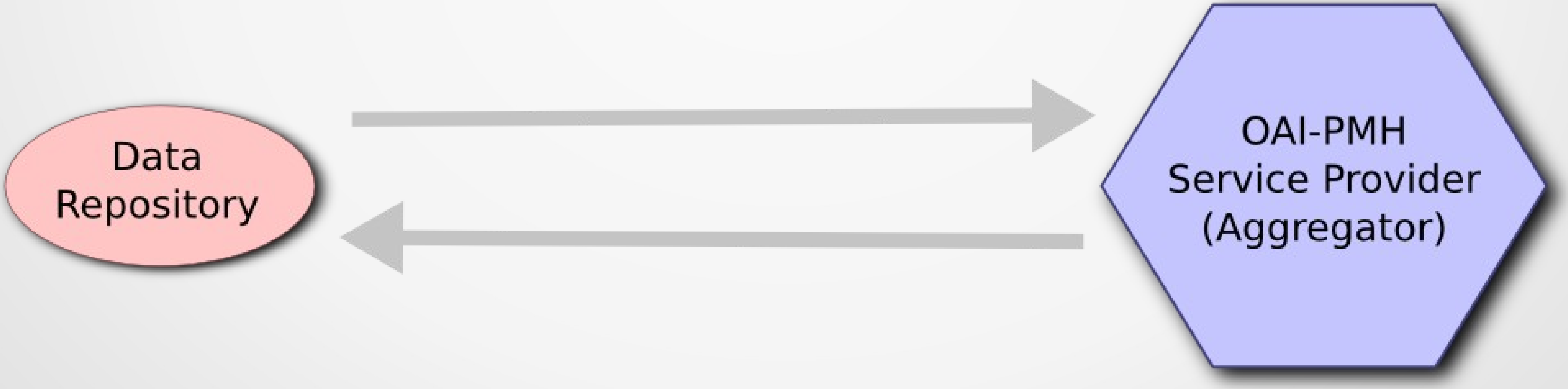

Encoding Researcher
Give me all elements which carry @type
Codesharing Service
<date type="bulleted" ...> <div type="chapter" ...> ....

The basic idea

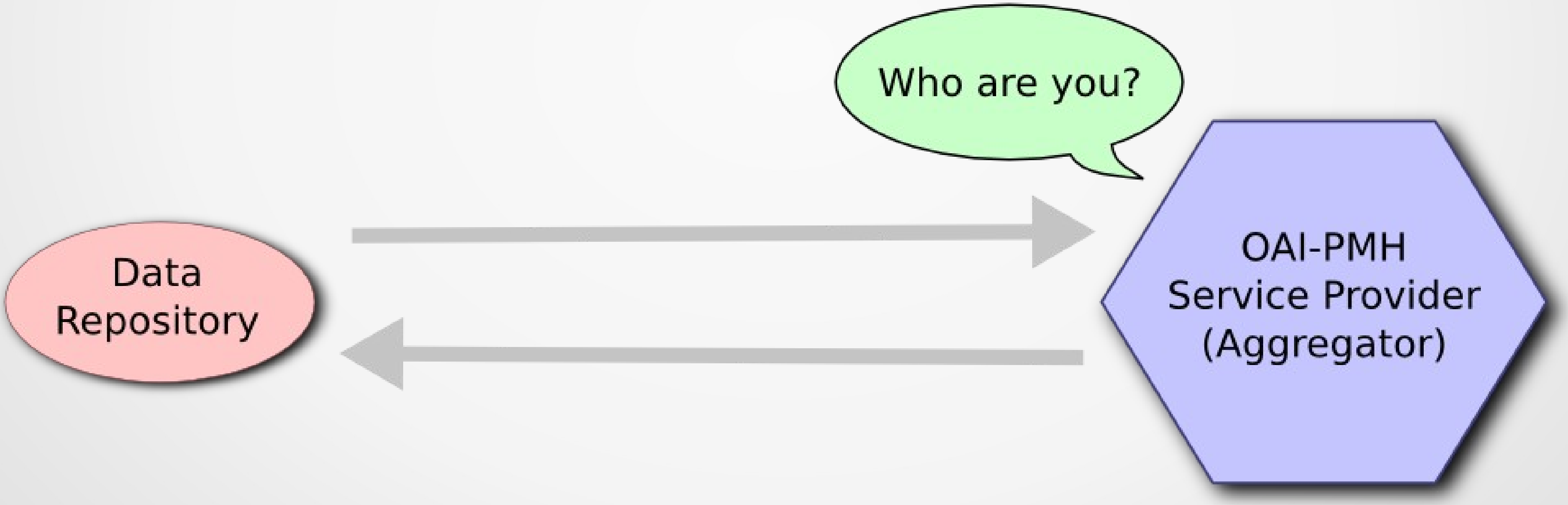

Novice Encoder
Show me some date tags with @when-custom attributes
<date when-custom="1601" datingMethod="mol:julian"
calendar="mol:julian">1601</date> <date ....
Codesharing Service
Give me all elements which carry @type
Encoding Researcher
<date type="bulleted" ...> <div type="chapter" ...> ....
XML Repository

The model: OAI-PMH

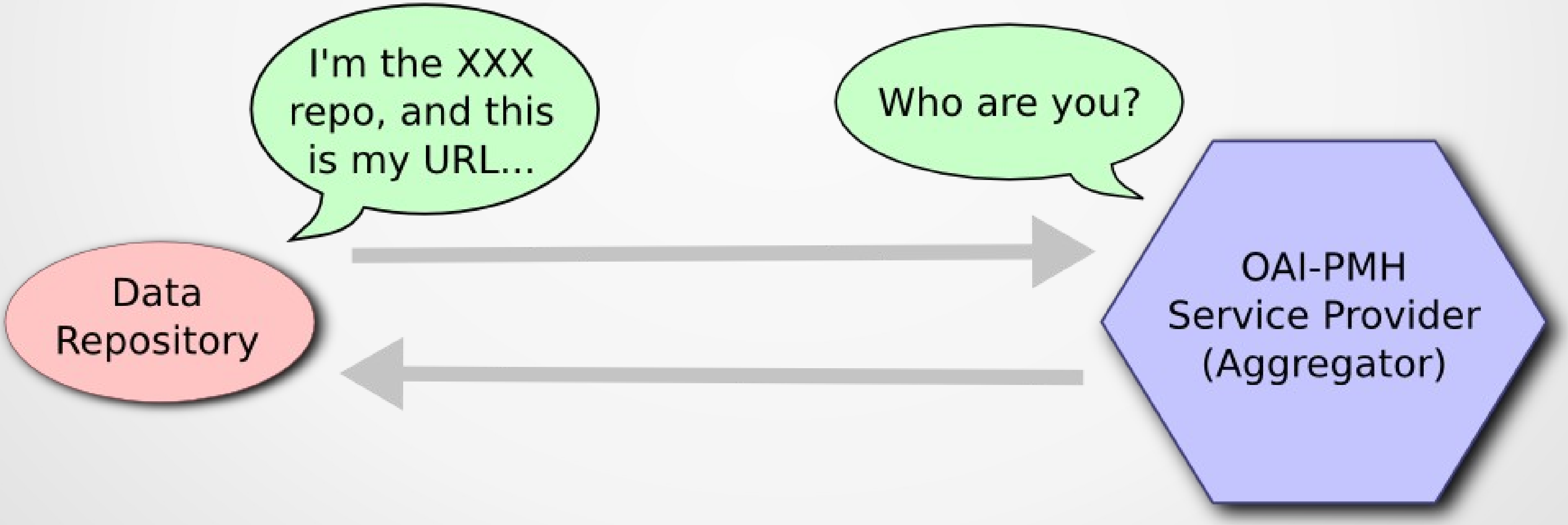
The Open Archives Initiative Protocol for Metadata Harvesting



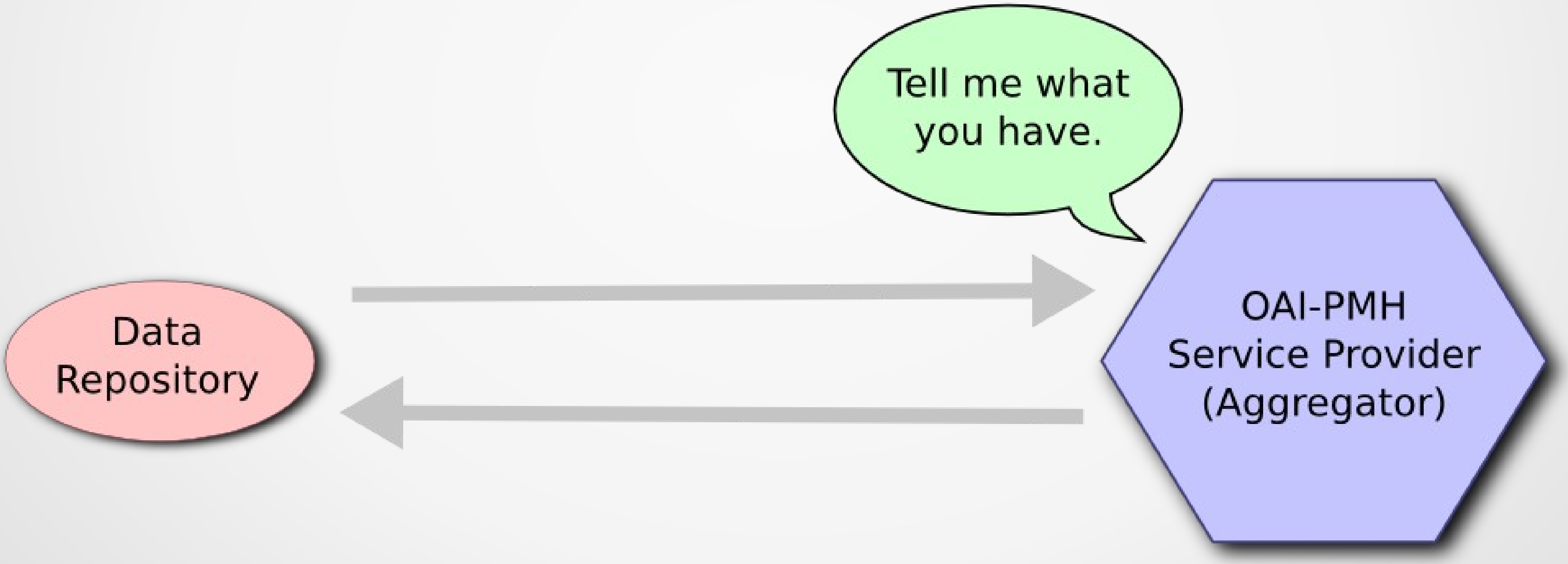
The model: OAI-PMH



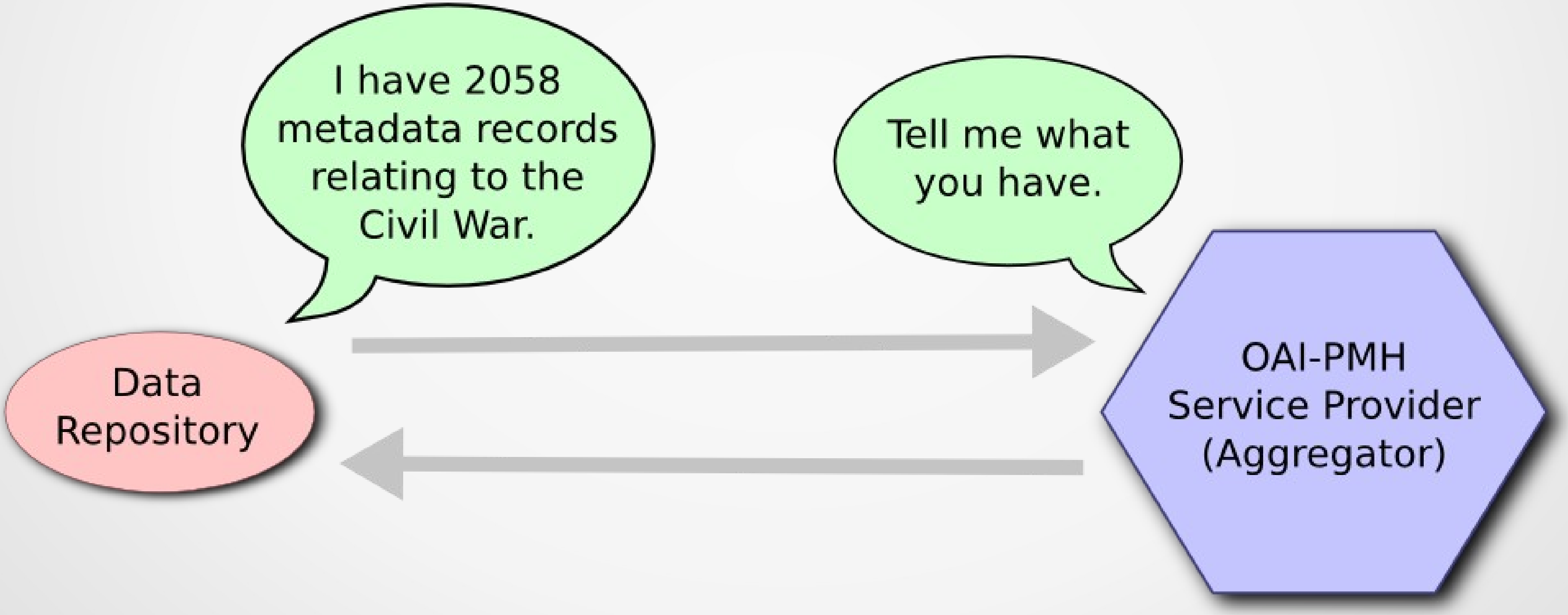
The model: OAI-PMH


I'm the XXX repo, and this is my URL...
Who are you?
Data Repository
OAI-PMH Service Provider (Aggregator)

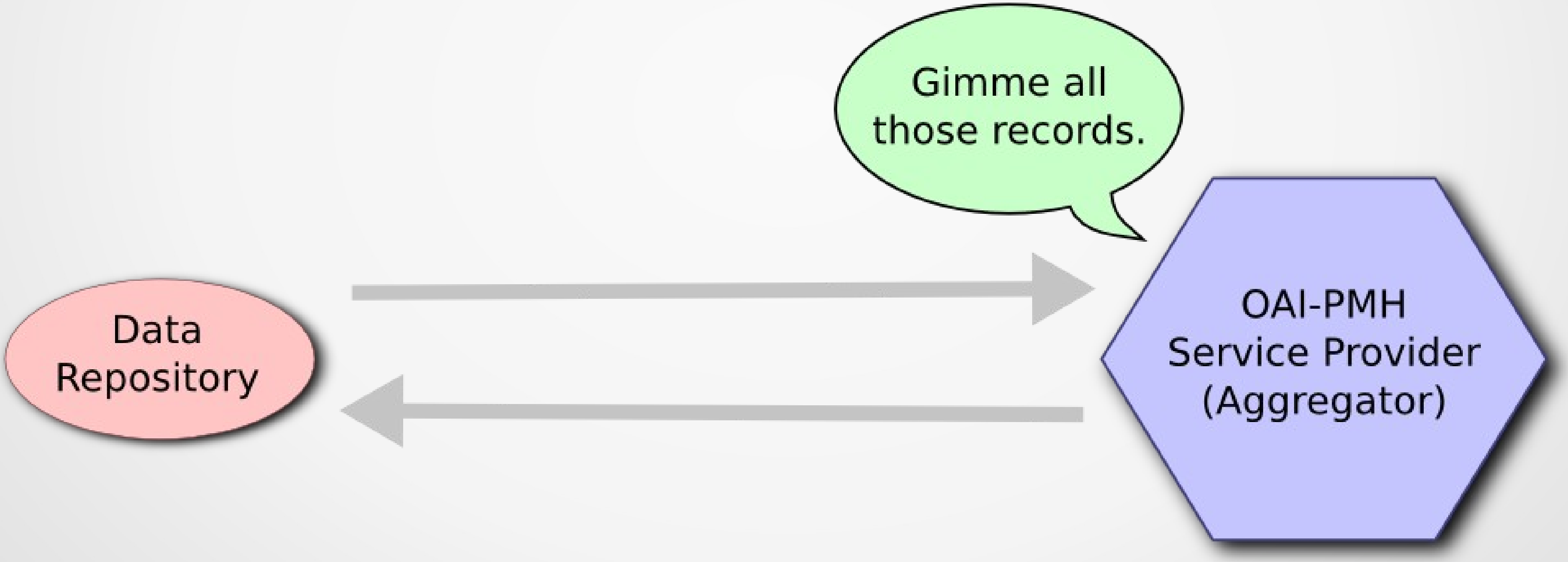
The model: OAI-PMH



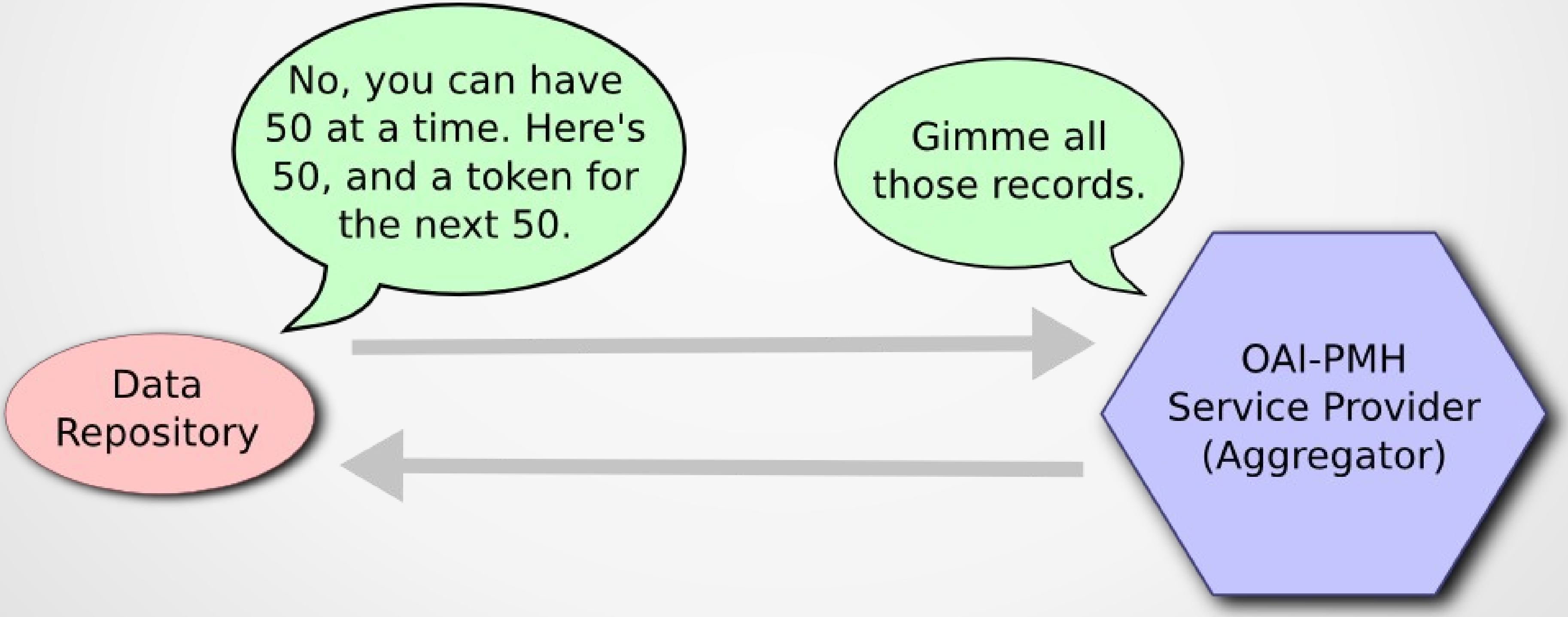
The model: OAI-PMH


I have 2058 metadata records relating to the Civil War.
Tell me what you have.
Data Repository
OAI-PMH Service Provider (Aggregator)

The model: OAI-PMH



The model: OAI-PMH


No, you can have 50 at a time. Here's 50, and a token for the next 50.
Gimme all those records.
Data Repository
OAI-PMH Service Provider (Aggregator)

The model: OAI-PMH

?verb=Identify

Who are you?

?verb=ListSets

?verb=ListIdentifiers

Tell me what you have.

?verb=ListRecords

?verb=GetRecord

Gimme all those records.

The model: OAI-PMH

```
?verb=
          Identify
          ListMetadataFormats
          ListIdentifiers
          ListSets
          ListRecords
          GetRecord
```

- http://bcgenesis.uvic.ca/oai.xq?verb=Identify

The model: OAI-PMH

- http://bcgenesis.uvic.ca/oai.xq?verb=ListRecords&metadataPrefix=oai_dc
- Resumption token:

```
<resumptionToken completeListSize="7153">
   from:2010-01-01;until:2020-01-01;set:;next:21
</resumptionToken>
```

The nitty-gritty of CodeSharing

```
codesharing?verb=
    identify
    listElements
    listAttributes
    listNamespaces
    listDocumentTypes
    getExamples
```

The nitty-gritty of CodeSharing

```
codesharing?verb=getExamples
    &elementName=
    &attributeName=
    &attributeValue=
    &namespace=
    &wrapped=
    &maxItemsPerPage=
```

The nitty-gritty of CodeSharing

- Response:
  - a TEI XML file
  - results in the <body>
  - settings and resumption URIs in the <front>
- http://mapoflondon.uvic.ca/codesharing?verb=listElements
- http://mapoflondon.uvic.ca/codesharing?verb=listAttributes
- http:...?verb=getExamples&elementName=date

A sample implementation

- The sample implementation is running here:

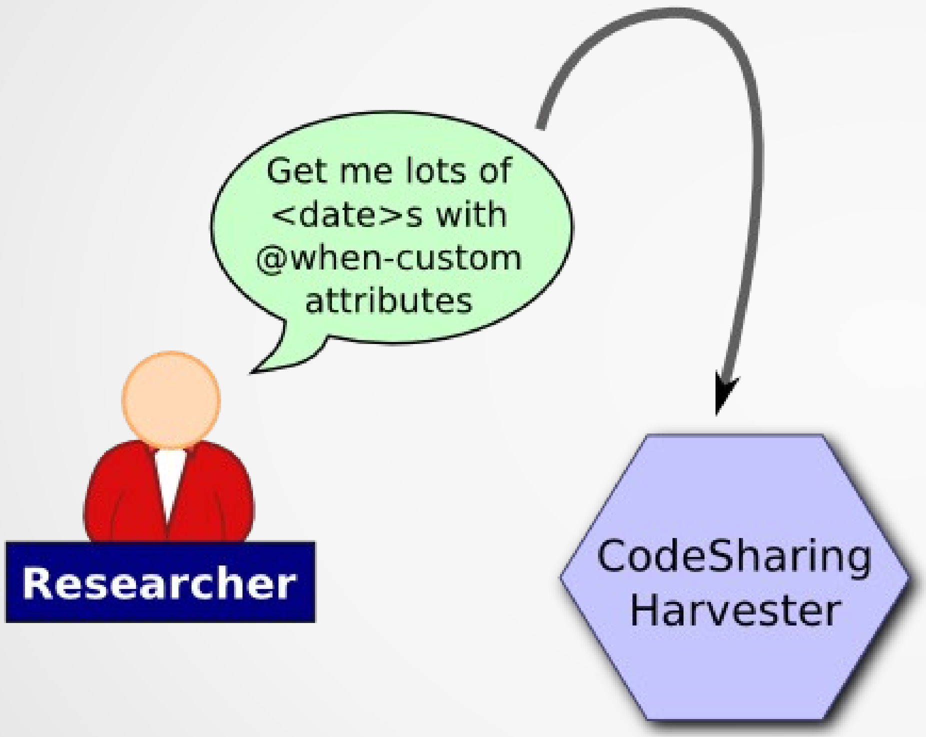
  http://mapoflondon.uvic.ca/codesharing

- There is a human-friendly interface here:

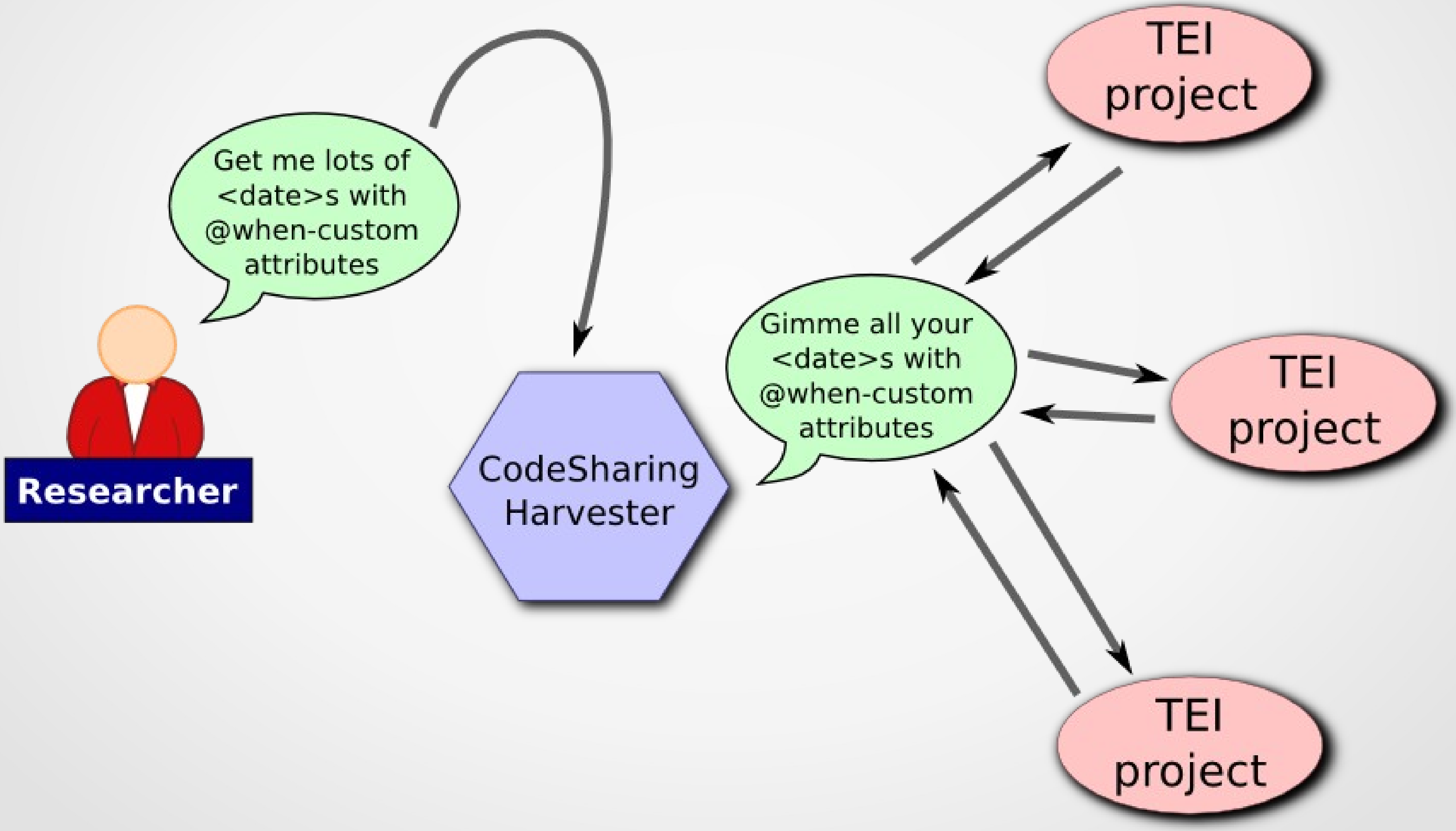
  http://mapoflondon.uvic.ca/codesharing.htm

Imagining a harvester

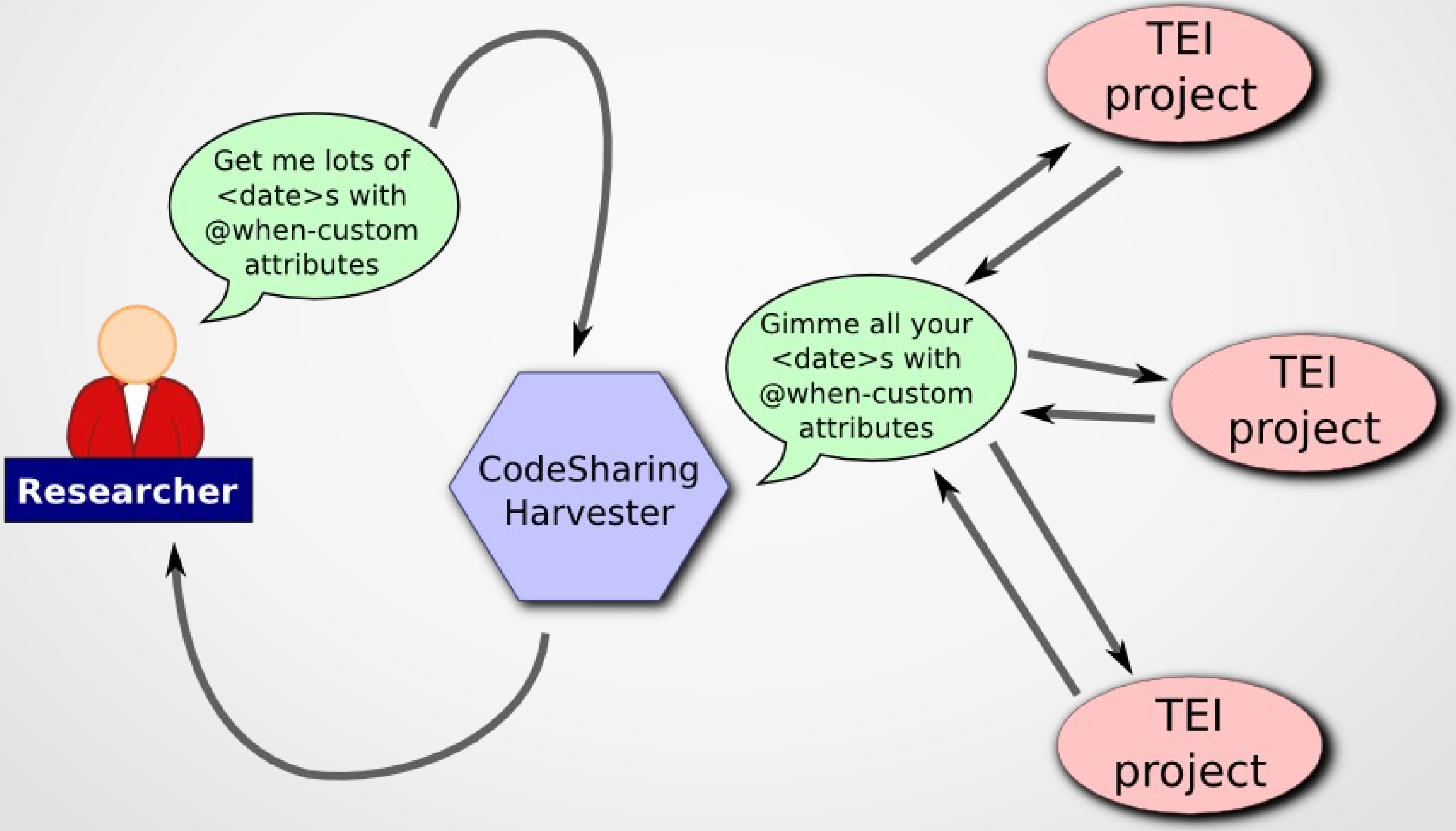
Imagining a harvester



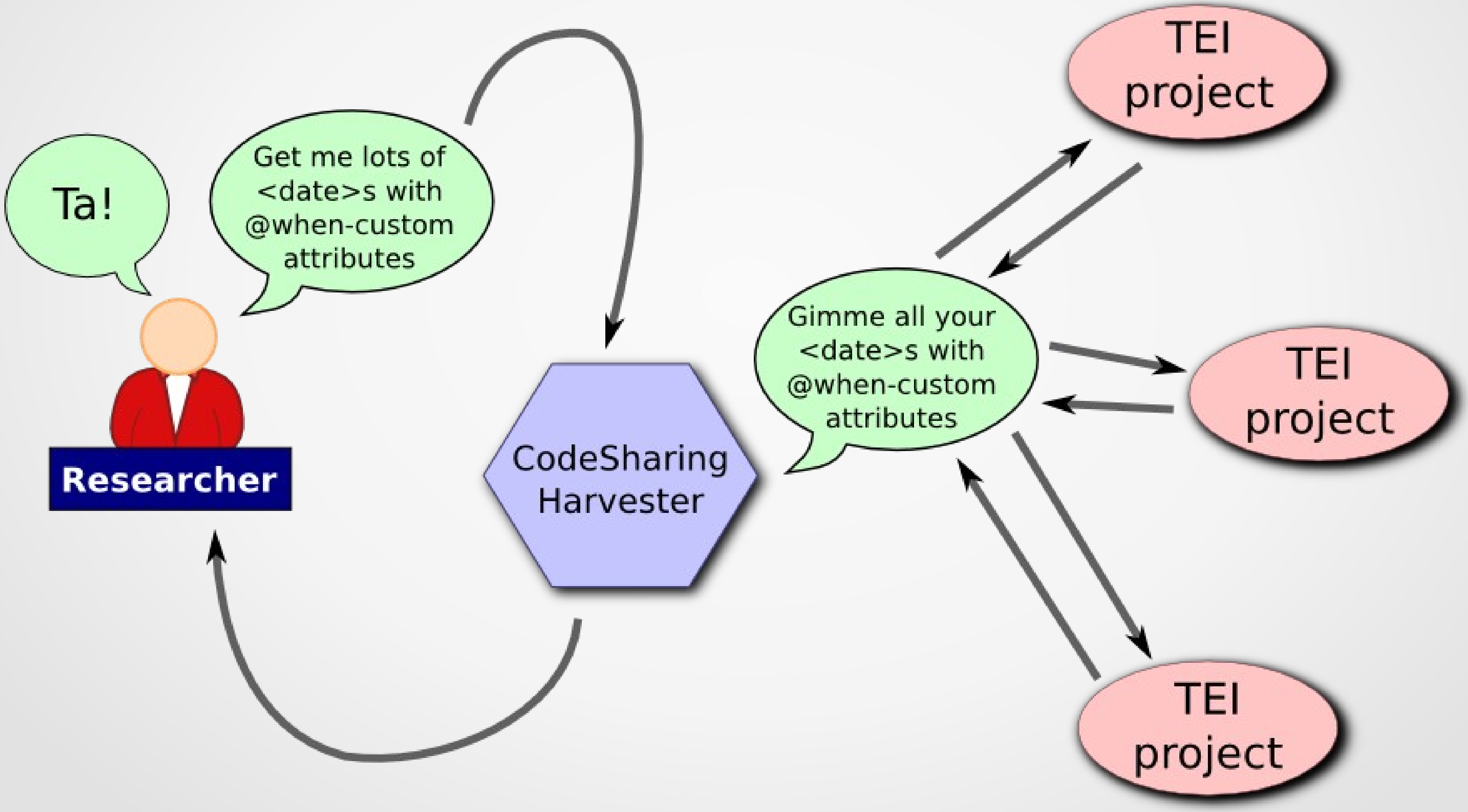
Imagining a harvester


Get me lots of <date>s with @when-custom attributes
Researcher
CodeSharing Harvester
Gimme all your <date>s with @when-custom attributes
TEI project
TEI project
TEI project

Imagining a harvester


Get me lots of <date>s with @when-custom attributes
Researcher
CodeSharing Harvester
Gimme all your <date>s with @when-custom attributes
TEI project
TEI project
TEI project

Imagining a harvester


Ta!
Get me lots of <date>s with @when-custom attributes
Researcher
CodeSharing Harvester
Gimme all your <date>s with @when-custom attributes
TEI project
TEI project
TEI project

The state of play

- Working implementation of the repository side in XQuery 3.0 for eXist
- Working XSLT for the human-friendly interface
- Complete semi-formal specification for the protocol
- BUT no work on a harvester yet.

Changes since last year

- Addition of document type filtering (optional in API)
- Rewrite of the sample implementation (performance improved over 10-fold)

The future: implementation

- Implementations in more projects using eXist.
- Creation of a plugin version for eXist?
- Listing of projects providing CodeSharing on the TEI wiki.
- Ports to other frameworks (BaseX; pure Saxon XSLT; PHP with Zorba...).
- Creation of a harvester application.
- Response compression?
- XPath support?

Where to find it

- Current specification: http://mapoflondon.uvic.ca/codesharing_protocol.xhtml
- SourceForge repo: https://sourceforge.net/projects/codesharing/
- Sample implementation (XML): http://mapoflondon.uvic.ca/codesharing
- Sample implementation (human-friendly interface): http://mapoflondon.uvic.ca/codesharing.htm
- Calling potential collaborators!