Making Thesauri Work for the Semantic Web

Alistair Miles

Brian Matthews

Michael Wilson

Alistair Miles
Rutherford Appleton Laboratory

Overview

- Introduction
- Thesaurus Interchange Format (TIF)
- Thesaurus Tools

Alistair Miles
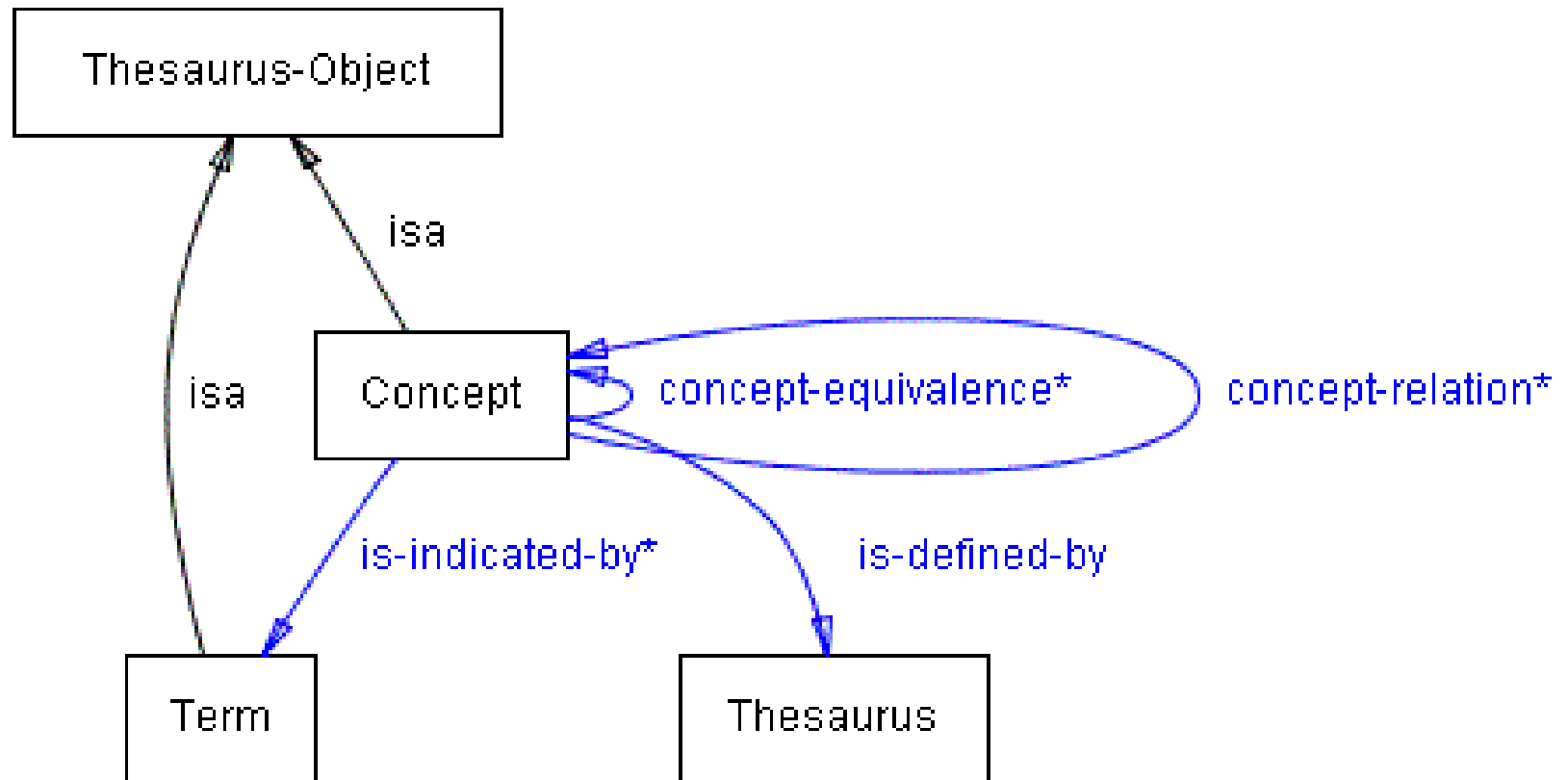Rutherford Appleton Laboratory

SWAD-Europe

- Semantic Web Advanced Development for Europe
  - Research & Development
  - Demonstration & Dissemination
- Partners:
  - CCLRC
  - HP Labs
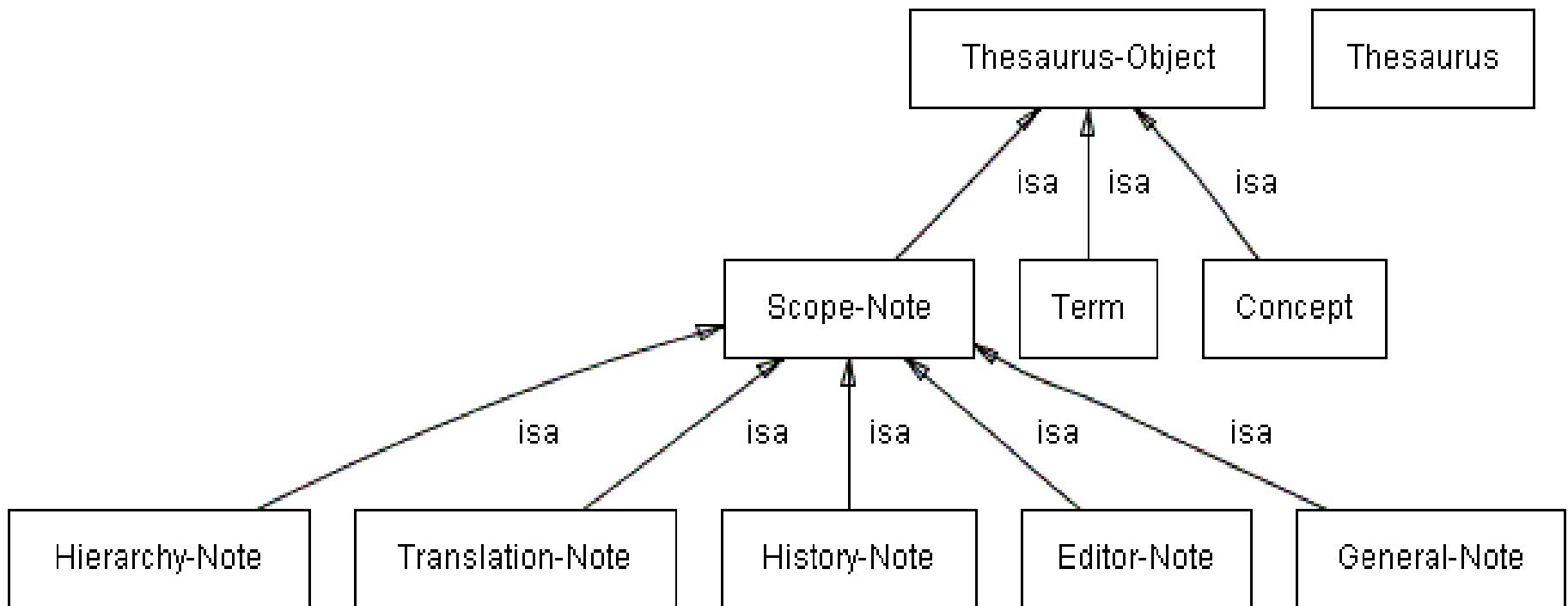 - ILRT
  - ERCIM (INRIA)
  - Stilo

Alistair Miles
Rutherford Appleton Laboratory

TIF for the Semantic Web

- RDF Schema
- Concept-based
- Extensible

Alistair Miles
Rutherford Appleton Laboratory

TIF Overview

Alistair Miles
Rutherford Appleton Laboratory

TIF Classes

TIF Built-In Properties …

Alistair Miles
Rutherford Appleton Laboratory

... Concept-Relation

Alistair Miles
Rutherford Appleton Laboratory

... Concept-Equivalence

Alistair Miles
Rutherford Appleton Laboratory

CCLRC
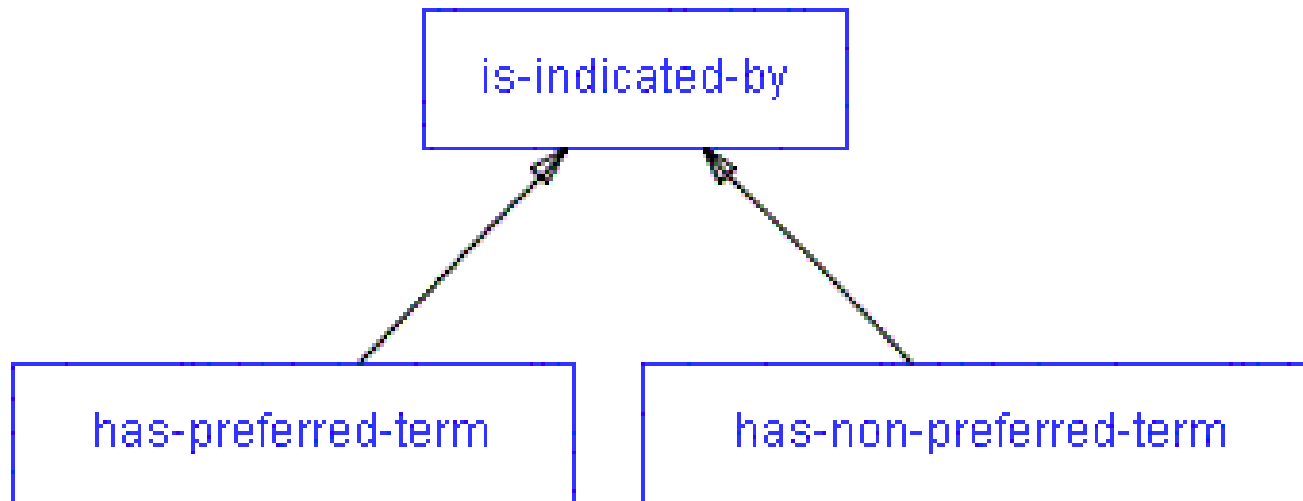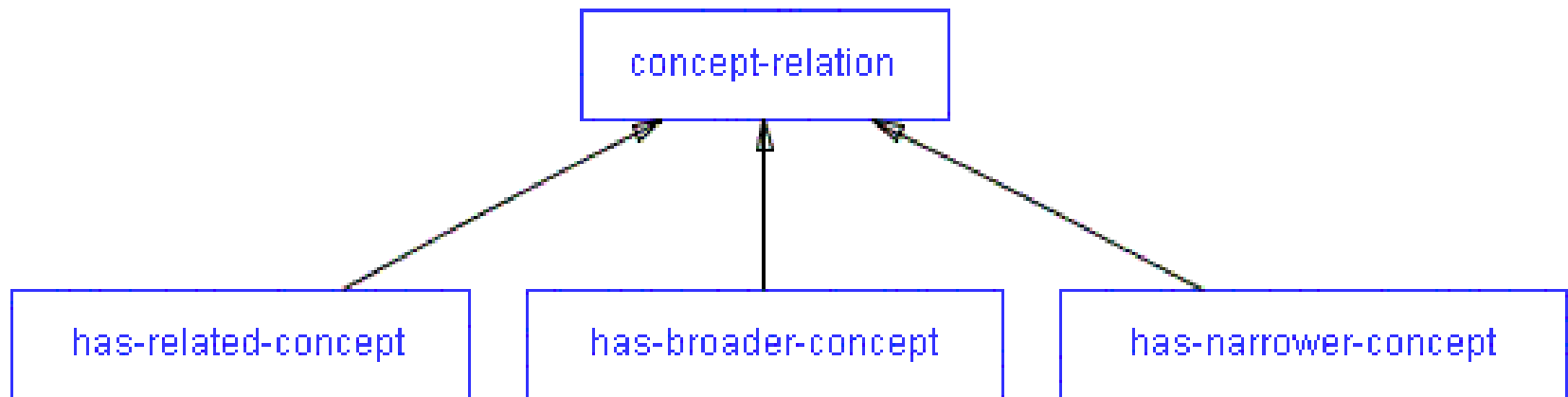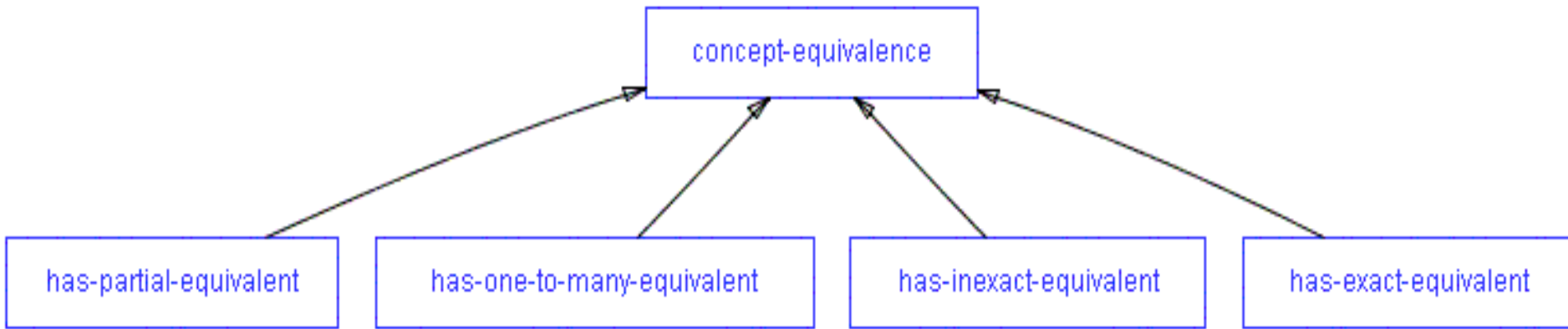
```
<rdf:RDF   xmlns="http://www.w3c.rl.ac.uk/2003/07/22-tif#"
           xmlns:dc="http://purl.org/dc/elements/1.1/"
           xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xml:base="http://www.rl.ac.uk/sophia/">


    <Thesaurus rdf:about="GCL">
        <dc:title>Government Category List</dc:title>
        <dc:description>List of categories used for government metadata.</dc:description>
    </Thesaurus>


    <Term rdf:about="GCL?term=A-levels">
        <value>A-levels</value>
    </Term>


    <Concept rdf:about="GCL?concept=607">
        <code>607</code>
        <preferred rdf:resource="GCL?term=Firearms"/>
        <nonPreferred rdf:resource="GCL?term=Gun%20licensing"/>
        <nonPreferred rdf:resource="GCL?term=Handguns"/>
        <broader rdf:resource="GCL?concept=564"/>
        <related rdf:resource="GCL?concept=598"/>
        <related rdf:resource="GCL?concept=827"/>
        <inexact rdf:resource="GCL?concept=111"/>
        <partial rdf:resource="GCL?concept=222"/>
    </Concept>

</rdf:RDF>
```
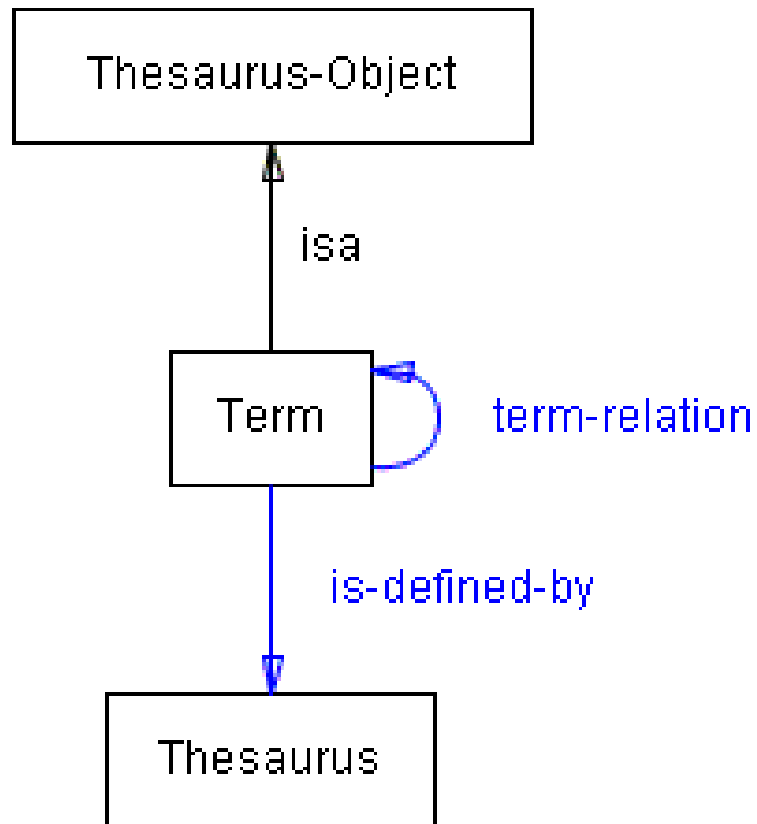
Alistair Miles
Rutherford Appleton Laboratory

Term-Based Extension

- TIF-Simple

Alistair Miles
Rutherford Appleton Laboratory

TIF-S Built-In Properties

Alistair Miles
Rutherford Appleton Laboratory
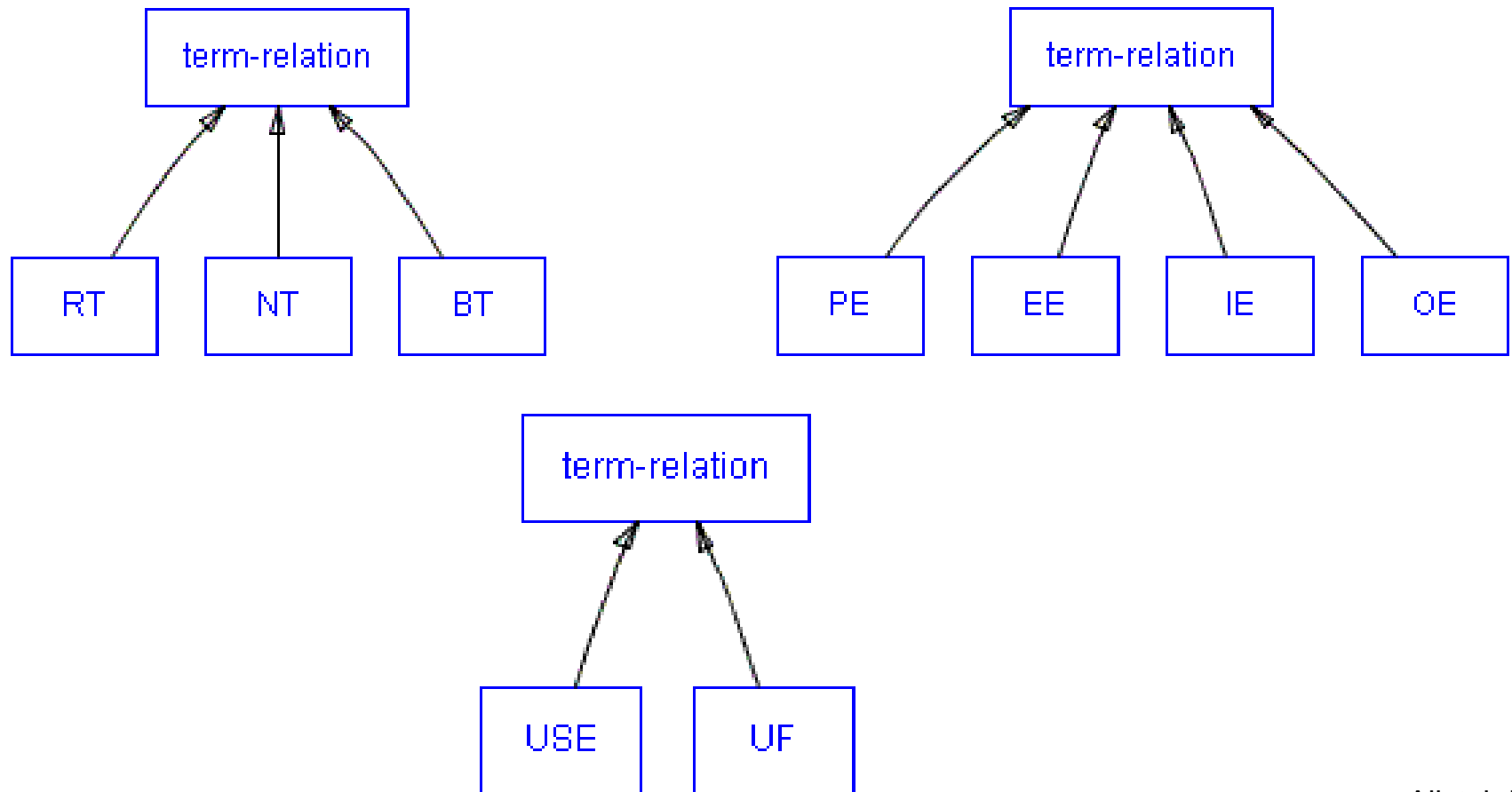
Generating TIF-S from TIF

- TIF-S as derived relations
- Mapping rules
- Use Closed World Machine (CWM)

Alistair Miles
Rutherford Appleton Laboratory
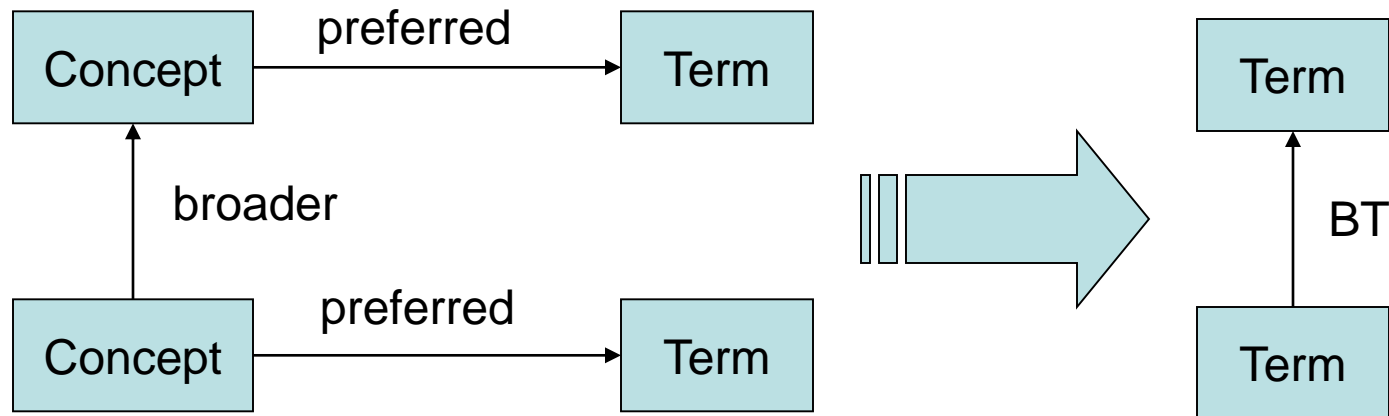
TIF Summary

- RDF Schema
- Concept-based
- Mono- or Multi- lingual
- Extensible framework of properties
- Term-based extension (TIF-S)
- Exploit strengths of both models
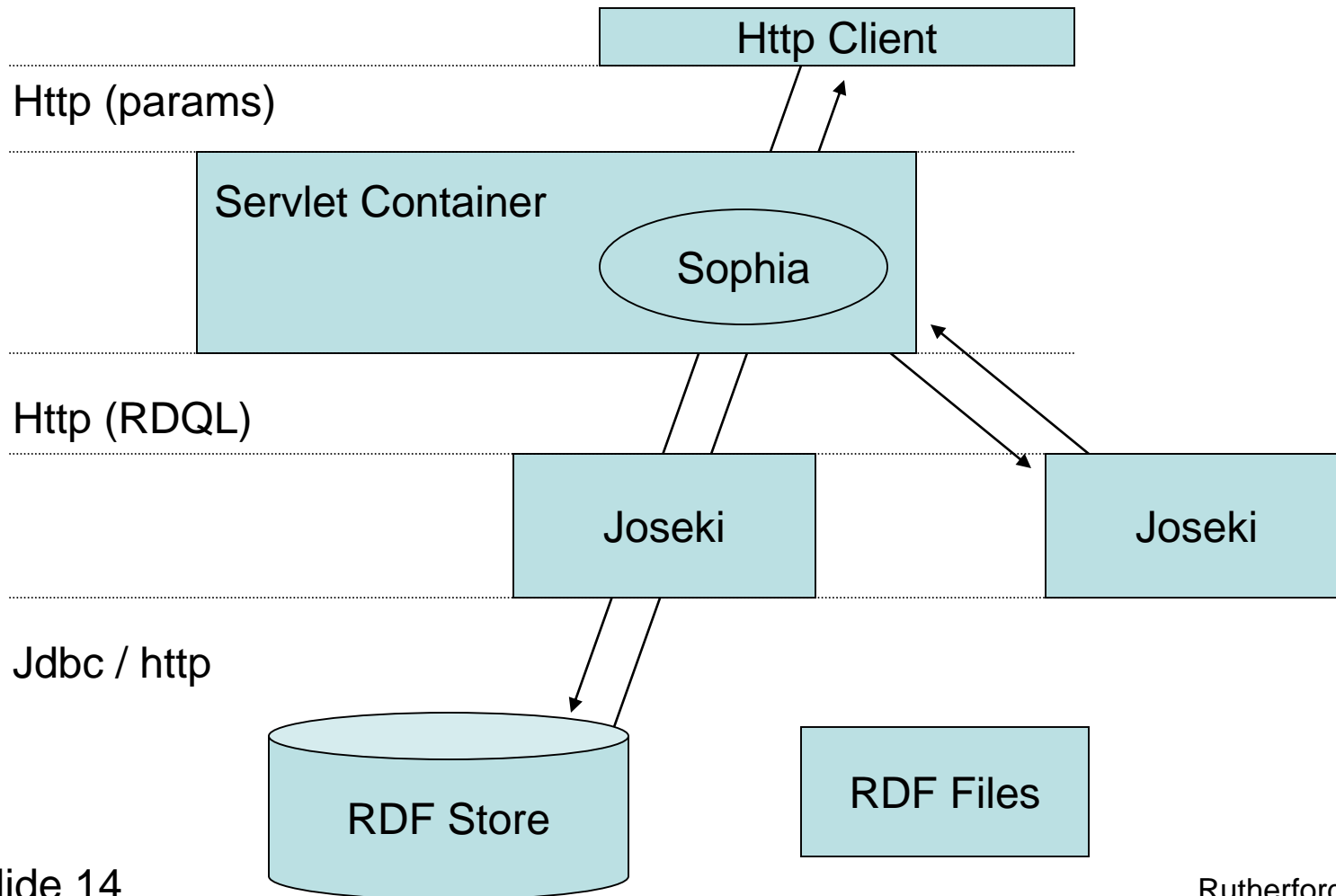
Alistair Miles
Rutherford Appleton Laboratory

Thesaurus Tools

- Low-level
  - Http server: 'Sophia*Lite*'
  - Java API: 'Thes'
- User-level
  - Thesaurus browser: 'Sophia*Multilingua*'

Alistair Miles
Rutherford Appleton Laboratory

Sophia*Lite*
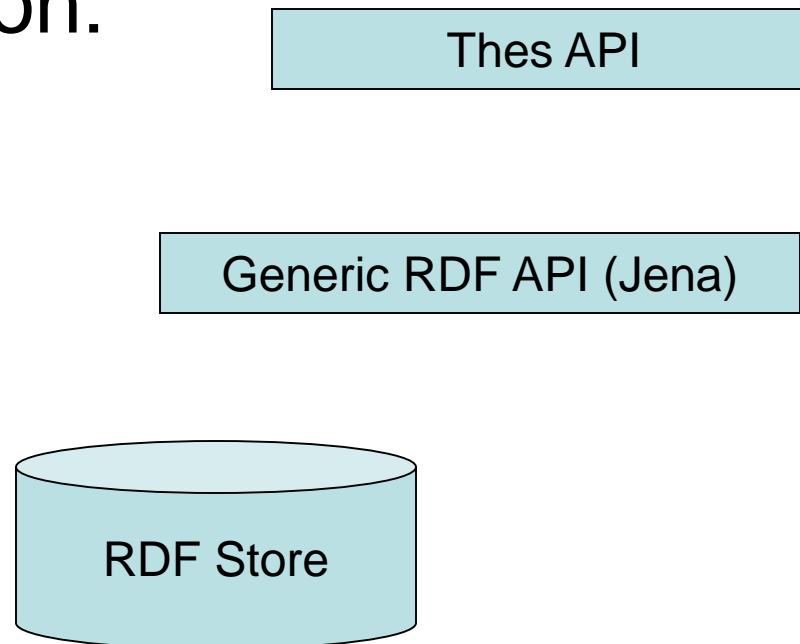
- Http thesaurus server
- Http get parameters
  ?concept
  ?term
  ?data   [topconcepts / topterms / allconcepts / allterms]

- Return RDF thesaurus data

Sophia*Lite* Architecture

Http Client

Http (params)

Servlet Container

Sophia

Http (RDQL)

Joseki

Joseki

Jdbc / http

RDF Store

RDF Files

Slide 14

Alistair Miles
Rutherford Appleton Laboratory

Thes Java API

- <u>Link to UML</u>

- Implementation:

Thes API

Generic RDF API (Jena)

RDF Store

Alistair Miles
Rutherford Appleton Laboratory

Sophia*Multilingua*

- Thesaurus Viewer
- Demo
- Leverage SW tools for powerful lightweight apps

Alistair Miles
Rutherford Appleton Laboratory

Future Work

- Using OWL for validation and inference
- Thesaurus web service
- Publish datasets
- More user-level applications
  - Use existing thesauri to enhance web experience
- Ontology migration

Alistair Miles
Rutherford Appleton Laboratory

Information

- Alistair Miles: a.j.miles@rl.ac.uk

- TIF on the web:

http://www.w3c.rl.ac.uk/SWAD/thesaurus/tif/tif.html

- SWAD-Europe Thesaurus Activity:

http://www.w3c.rl.ac.uk/SWAD/thesaurus.html

Alistair Miles
Rutherford Appleton Laboratory