RIF-CS and CERIF alignment study: All that and a service registry...?

Emma Tonkin

tonkin@cs.bris.ac.uk

Pisa, 7 April 2014   Twitter: @emmatonkin

Introduction

"Two households, both alike in dignity
In fair Verona, where we lay our scene
From ancient grudge break to new mutiny
Where civil blood makes civil hands unclean…"
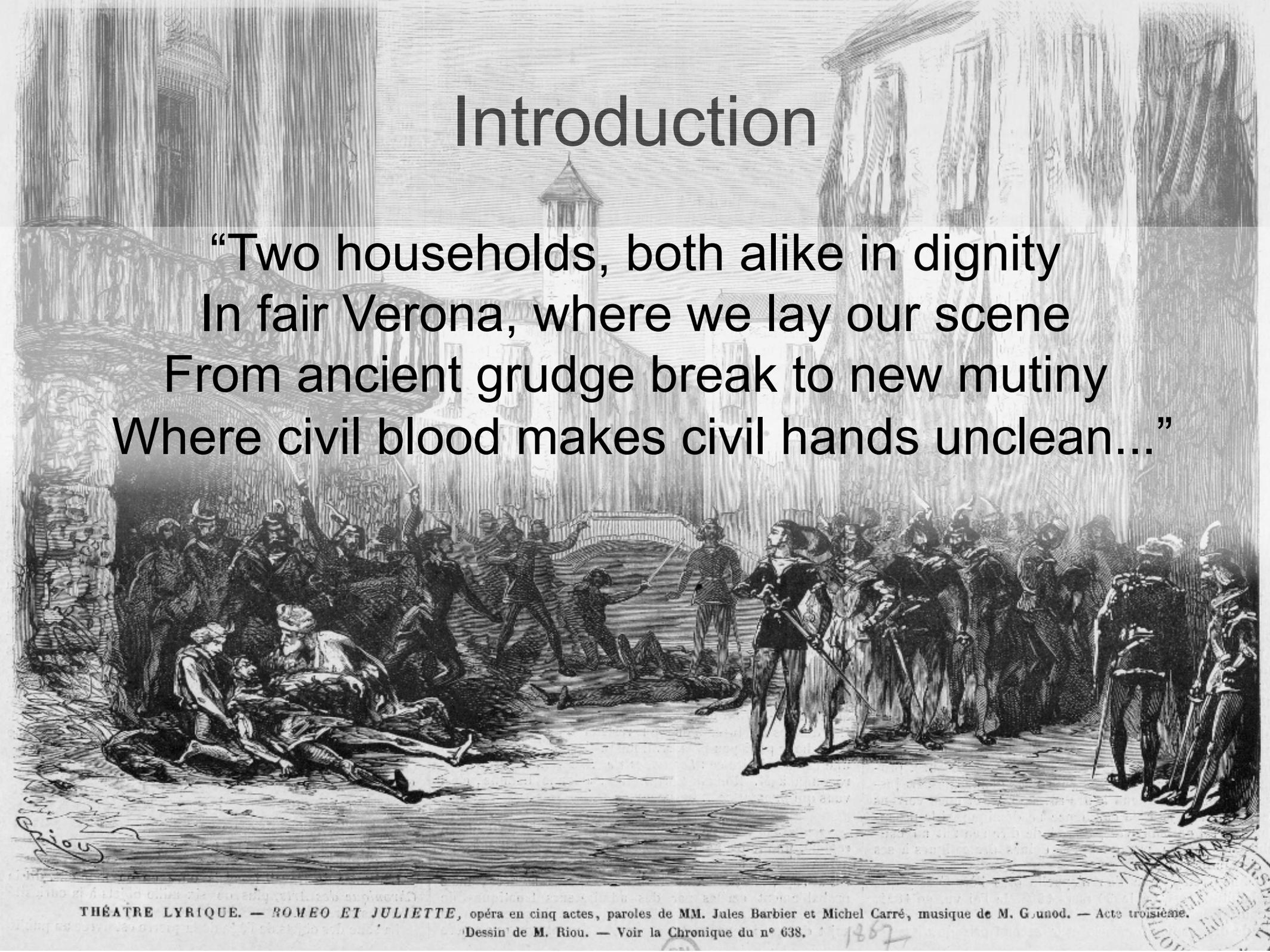
Australian National Data Service: RIF-CS

Resource Interchange Framework – Collections and Services

CERIF

Common European Research Information Format

Pictorially...

vs

Image attribution
Chris McKenna (Thryduulf)

The 'upshot'

- Different regions have solved the publication of research outputs (including data, associated works, papers etc) in different ways

- Family histories of these specific standards:

  – IESR + OCKHAM + ANDS → RIF-CS

  – CERIF91 (flatfile) → CERIF 2000 (Database) → CERIF 2006 (Semantic + XML exchange) → CERIF 1.5 (2012) (federation) → CERIF 1.6 (2013) (datasets) [Dvorak, Jan (2013). Contextual Metadata]

What links these standards?

- A perceived similarity in use case

- RIF-CS: "aims to provide a central resource describing the disparate research [data] collections made available by researchers and research organisations throughout Australia."

- CERIF: "a standard for managing and exchanging research data, i.e. information about researchers, projects, outputs and funding that arises from the research process. It provides a data model that can be used to describe the research domain"

- Perspective shift: Service model vs data model

Initial reaction

- It is said that:
  "You can put a boat on a ship, but you can't put a ship on a boat."

- By comparison, we might guess that:
  "You can put RIF-CS in CERIF, but you can't put CERIF in RIF-CS."

- You know what they say about assumptions.

- Let's test it: [when] can these standards be mapped?

Designing a study

- Query: *can these standards be mapped?*

- Approach:

  - Choose a use case

  - Explore with reference to both standards

  - Review the outcome

Use case: Registering and mapping services

- A university has developed a large collection of data, along with a number of research outputs.

- "The university wants to submit their data collection, *including the service through which it can be accessed*, to a 'Research Data Commons', thus *publicising* the existence of their research data collections. Using RIF-CS it is possible for the university to automatically exchange a detailed, machine-readable description of a collection or set of collections with the Data Commons."

Simple to encode in RIF-CS

```xml
1  <?xml version="1.0"?>
2  <registryObjects xmlns="http://ands.org.au/standards/rif-cs/registryObjects"
3   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
4   xsi:schemaLocation="http://ands.org.au/standards/rif-cs/registryObjects
5   http://services.ands.org.au/documentation/rifcs/schema/registryObjects.xsd">
6   <registryObject group="The Australian Partnership for Sustainable Repositories">
7    <key>au.edu.apsr.a1</key>
8    <originatingSource>
9     http://devl.ands.org.au/deployment/cosi/orca/register_my_data
10   </originatingSource>
11   <activity type="program">
12    <name type="abbreviated">
13     <namePart>APSR</namePart>
14    </name>
15    <name type="primary">
16    <namePart>
17     The Australian Partnership for Sustainable Repositories
18    </namePart>
19    </name>
20    <location>
21     <address>
22      <electronic type="url">
23       <value>http://www.apsr.edu.au</value>
24      </electronic>
25     </address>
26    </location>
27    <relatedObject>
28     <key>au.edu.apsr.a2</key>
29     <relation type="funds"/>
30    </relatedObject>
31    <description type="brief">The APSR Project aims to establish a centre of excellence for the
```

(etc.)

This example supposes:

- That there exists a program [a type of activity, defined in the vocabularies as 'a system of activities intended to meet a public need']. The name of the programme is given in full and abbreviated forms alongside a description and information regarding funding. An actionable URI is given.

- That there exists a project with a specified start [and end] date, with a given description and identifier.

- That there exists a collection with a given identifier (here a handle.net identifier), name, URL, a number of supported access methods and an identified managing authority

- That there exists a person (a type of party)—note that the vocabulary documentation allows 'person' to be 'an identity assumed by one or more human beings. The person has a given first and lastname, has a management responsibility over identified objects, and a set of 'existenceDates' (introduced as part of RIF-CS v1.3.0, in late 2011).

Mapping the example to CERIF

- Partially achieved

- Absent the grim detail, in short -

- Concepts related to service registries presented difficulties

- ...although workarounds exist

- Each standard carries the heritage of its origins

Discussion (I)

- We have seen that RIF-CS and CERIF circa 2012 do not (entirely) map in this case

- They can undoubtedly be mapped in some scenarios

- There are ways to inject missing information into CERIF (notably, via Dublin Core)

- CERIF may extend and embrace this functionality (if it hasn't already!)

"There can be only one"

- Short response: This statement is false.

- Concise response: Why doesn't context get more airtime in standards development?

- Long response...

Evolution

Source: XKCD.com/927
See http://xkcdsucks.blogspot.it/2011/07/comic-927-alternative-title-was-things.html for an opposing (mildly profane) viewpoint

Digression

'What's that you're holding?'

Shawn looked down. 'It's the Lancrastian Army Knife,' he said.

Granny peered closer. 'What's the curly thing?' she said.

'Oh, that's the Adjustable Device for Winning Ontological Arguments,' said Shawn. 'The King asked for it.'

'Works, does it?'

'Er... if you twiddle it properly.'

'And this?'

'That's the Tool for Extracting the Essential Truth from a Given Statement,' said Shawn.

– Terry Pratchett, *Carpe Jugulum*

Discussion (II)

- Toolbox vs swiss-army-knife

- Use case affects form *and* function *and* philosophy

- Form, function and philosophy affect implementation

- Implementation affects practicality and performance, not to mention price
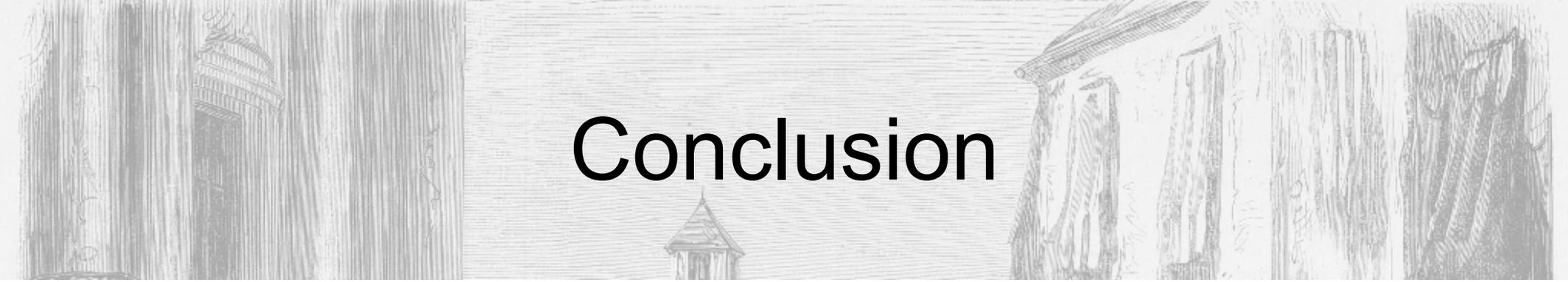
Conclusion

- Publishing vs *publicising*

- Data model vs service model

- Worthy, potentially complementary (but differing) goals

p.s.

NOUS N'AVONS PAS LES MÊMES VALEURS.

...and that's okay