

# Links refer, URIs don't

Henry S. Thompson  
5 Aug 2011

Copyright © 2011 [Henry S. Thompson](#)



## Table of Contents

1. [Acknowledgements](#)
2. [The official story of URIs](#)
3. [URI syntax: some terminology](#)
4. [The nature of resources](#)
5. [Official story summary](#)
6. [A narrow view. . .](#)
7. [What's the problem, then?](#)
8. [What's a presentation?](#)
9. [Web practice: alternative representations](#)
10. [Resource, representations, alternatives](#)
11. [The nature of the work of art](#)
12. [FRBR and the Web](#)
13. [Time-varying resources](#)
14. [Indexicals, cont'd](#)
15. [Taking presentation seriously](#)
16. [Theory of presentations, cont'd](#)
17. [Presentations and persistence](#)
18. [Local context](#)
19. [Back to local context](#)
20. [Not so unique. . .](#)
21. [Local context for URIs](#)
22. [Elaborating context](#)
23. [Compositionality: past its sell-by date?](#)
24. [RDF already isn't compositional](#)
25. [Two reasons why URIs aren't names](#)
26. [Conclusions](#)

## 1. Acknowledgements

I owe much of what I understand about the Web to Tim Berners-Lee, along with Dan Connolly, Harry Halpin, Jonathan Rees and Larry Masinter

Brian Cantwell Smith has hugely influenced my approach to all things computational

- He first brought the flight name examples to my attention over 30 years ago

- And stimulated me to think harder about the URI/link distinction just a few months ago. . .

## 2. The official story of URIs

There is a moderately clear official consensus about URIs

- Explicitly specified in IETF's [RFC 3986](#)
- Summarised and extended in W3C's [WebArch](#)

'URI' stands for

### Uniform

There's a standard generic syntax and (partial) semantics

### Identifier

URIs are identifiers, that is, names (*not* addresses)

### Resource

What it is they identify -- anything at all

## 3. URI syntax: some terminology

We will need to refer to parts of URIs as they are written:

http://www.ltg.ed.ac.uk/~ht/identity/ComponentGraph.svg?.....#.....

scheme | domain | path | query (optional) | fragment

More examples:

```
http://www.ed.ac.uk/
ftp://ftp.funet.fi/pub/mirrors/perl/
mailto:ht@inf.ed.ac.uk [that's a bit weird. . .]
file:///D:/Documents/HTalks/Inaugural/slides.html [this talk]
http://localhost/ht/Documents/HTalks/Inaugural/slides.html
[likewise]
http://www.ltg.ed.ac.uk/~ht/travel.html#travel
http://maps.google.co.uk/
maps?q=10+Crichton+Street,+Edinburgh&layer=c&
sll=55.944586,-3.187494&cbp=13,332.96,,0,1.84&cbll=55.944532,-3.18738&
hl=en&ssp=0.006295,0.006295&ie=UTF8&hq=&hnear=10+Crichton+St,+Edinburgh
ll=55.944532,-3.18738&spn=0.000012,0.006727&t=m&z=17&
vpsrc=0&panoid=e8EPc2P5vtC6b-oE8SlGiw
https://mail.google.com/mail/?ui=2&shva=1#inbox
```

## 4. The nature of resources

At first 'resource' seems like a vacuous label

- If anything at all can be a resource
- Knowing something is a resource tells you nothing

But consider the word **referent**

- Anything can be a referent
- When/if someone/something refers to it!

So, being a resource is not about some *intrinsic* property of something

- It just means, being identified by a URI

This is the first hint of my preferred methodology in this talk

- Look for parallels with natural language
- And get quick and cheap victories by appealing to relevant insights from Philosophy of Language

## 5. Official story summary

URIS identify resources

Resources can be anything at all

Accessing a URI may yield a representation of (the current state of) its referent

- But some resources have no representation
- And others have more than one
  - All accessible from the same URI via content negotiation

Only information resources have representations

## 6. A narrow view. . .

Not before time, I need to emphasise that there's a lot I've ignored so far

- The Web browser is not the only user agent/client
- `http` is not the only URI scheme
- Human beings are not the only users

## 7. What's the problem, then?

That's all pretty much OK, isn't it?

- A bit complex

- Uses words as technical terms in not completely obvious ways
- Some of the concepts are a bit soft/under-specified

No, it's not OK

In practice

- Too many people either don't know or don't care
  - Virtually all users
  - Most developers
- The Web has moved on
  - Many background assumptions/motivations are no longer as important as they once were
  - Or are simply not true at all
- Some important things have always been missing
  - The status of presentations
  - The difference between URIs, links and link assertions

## 8. What's a presentation?

A browser goes beyond just retrieving a representation

- From representation to, shall we say, **presentation**

### Representation

```

Metadata:
Content-type:
application/xhtml+xml

Data:
<!DOCTYPE html PUBLIC "...
  "http://www.w3.org/...
<html xmlns="http://www...
<head>
  <title>5 day forecast for
    Oaxaca</title>
</head>
<body>
  <h2>Today's forecast</h2>
  <p>Hot and windy</p>
  ...
</html>

```

Determines

### Presentation



## 9. Web practice: alternative representations

Almost no use is made of the official mechanisms for distinguishing alternative representations of the same resource

For example, to get a Spanish version of a weather report from a widely used weather site

- You request a report explicitly specified to be in *English*
  - Via the official-but-almost-always-ignored mechanism
  - Which *is* ignored
  - And overridden by an *ad hoc* cookie specifying Spanish
- And the desire for mobile-appropriate presentation would likewise not be sent in the officially-approved way
  - But rather encoded in the URI itself

Developers often ignore/skate over the official story about resources and their representations

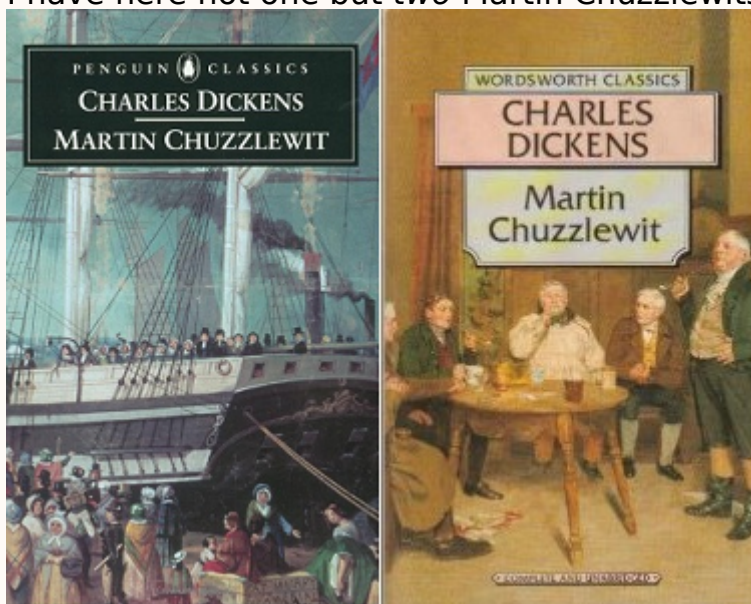
- In favour of a much more tightly linked one-to-one connection

## 10. Resource, representations, alternatives

This space is much richer and more complex than the current theory allows for

The Information Scientists have struggled with the ontology of created works for a long time

- Just what do we need to deal with when we consider, for example, Charles Dickens's novel *Martin Chuzzlewit*
- I have here not one but *two* Martin Chuzzlewits



- Are they the same book?
  - Yes and no, right?
  - Same author, same story
  - Same language
  - Different publishers, different editions
  - Different physical objects

## 11. The nature of the work of art

IS have developed a detailed story about all this

- Known as FRBR
  - Functional Requirements for Bibliographic Records

FRBR has a *four*-level ontology

**Work** "a distinct intellectual or artistic creation"; "an abstract entity"

**Expression** "the intellectual or artistic realization of a **work** in the form of [some] notation", e.g. "variant texts"; first, second, . . . editions; English original vs. French translation

**Manifestation** "the physical embodiment of an **expression** of a **work**"; "all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form"

**Item** "a single exemplar of a **manifestation**"; "a concrete entity"

## 12. FRBR and the Web

Not all of the FRBR architecture maps directly onto the Web situation

- But getting clear where it does
- And where it doesn't
- Would be of use to both communities

Some people in the IS community, notably Allen Renear, have made a start on this

- But it needs a lot more work

## 13. Time-varying resources

Clearly there's a lot of it about

- But the official story about it is hopelessly vague
- In particular, with respect to what counts as acceptable (with respect to the Social Contract) variation

The Philosophy of Language offers a possible way in

Consider English words such as *this*, *here* and *tomorrow*, as well as *you* and *I*.

- Linguists call this class of words **indexicals**

On one well-thought-of Philosophy of language account, an indexical such as *now* has

*Simplified version  
of Davidson's  
original story*

- a single **meaning**
- but multiple **interpretations**

The meaning is fixed, the interpretation varies according to the context of use

## 14. Indexicals, cont'd

The meaning of *now* is something like "the time at which the utterance containing the word is made"

- It is the same for every ordinary use of the word

But the interpretation is, well, whatever time it happens to be when the word is used

- and this of course changes all the time

More generally, the meaning of an indexical can be understood as a function from contexts (that is, contexts of utterance) to interpretations

In a formal approach to all this, indexical meaning really is just an **index**

- Because context of utterance is modelled as a sequence
  - E.g. `<[Henry S. Thompson], {[a lot of folk]}, 2011-10-05:11:08:05, . . . >`
- So on this account the meaning of *now* is pretty much just  $\lambda(c)c_3$

The parallel with time-varying resources is clear

- But more work is needed to see if that observation can be cashed in
  - In terms of a useful formal story
- Some aspects of the functional account of presentation are a tiny step in this direction

## 15. Taking presentation seriously

For the vast majority of web developers, it's all about what I've been calling presentation

- The user experience

Maybe we should make that more central in our theory

Here's a sketch of how we might start

A **browser** implements a function from a URI and a browser state to a URI plus the substance of a request:  $B_1(U, S_B) \rightarrow \langle U, \sigma \rangle$

A **server** implements a function from a URI plus a substance from a request and a server state to a substance in a response:  $S(U, \sigma, S_S) \rightarrow \sigma'$

A **browser** also implements a function from a substance from a response, a browser state and a web state to a presentation:  $B_2(\sigma, S_B, S_W) \rightarrow \pi$

Compose all three of these, and a browser implements a function from URIs to presentations (with dependencies on three kinds of state: browser, server and web):  $B(U) \rightarrow \pi$

People responsible for URIs know this

- They understand what they are doing in terms of this story
- Their goal is precisely to effect the connection between a particular URI and a particular presentation

## 16. Theory of presentations, cont'd

**B2** (and thus **B**) depends on Web state

- because e.g. rendering an HTML page may involve appeals to **B1** ◦ **S** for stylesheets and scripts
  - As well as to the whole of **B** for frames, embedded images, etc.

In principle this recursion could fail to terminate

- In practice this is extremely unusual
- Although I have heard anecdotal evidence that crawlers do sometimes fall into unintended recursion tar pits

## 17. Presentations and persistence

The stability over time of **B** for some URI is then what people are talking about when they discuss **persistence**

We can outline some interesting classes of expected behaviour in these terms

1. Once a particular instance of **B** is established, there is *no* expectation that the resultant presentation will change in any way
  - URIs which present as images, audio and video are often in this category
2. Once a particular instance of **B** is established, there is an expectation that the resultant presentation will not change
  - but there is a recognition that errors may need to be corrected
  - URIs which present as some form of transcription of documents created outside the Web are typically in this category
  - as are W3C Recommendations
3. URIs whose presentations change only in accordance with a published versioning policy
4. A newspaper home page
5. My home page
6. A 'live' news blog



## 18. Local context

Arguably, much of the focus of Web architecture discussions to date has been misplaced

- That is, the significance of URIs on their own has been exaggerated

Words, after all, don't mean anything

- When considered as simply a sequence of characters or sounds

They have to be *used* before we can talk about their meaning

- And we have to know *how* they're used

Consider the following character sequence as if found on a piece of paper in an otherwise featureless bottle on a desert island beach:

chat

Conversation? Cat? Something else altogether? There's no way to tell without more. . .

J'ai vue un chat  
Lets have a chat

Or consider (spoken)

The word Colin has five letters  
My friend Colin has five bicycles

It's words *in use* that have meaning

- So not just the external context
  - As for indexicals, above
- But also the local, linguistic, context

## 19. Back to local context

Ordinary language names function within specific linguistic contexts

What a name *means* depends on the details of the surrounding utterance or sentence

Consider what at first seems like a very ordinary (imaginary) name: *EZY386* (short for *easyjet 386*)

Here are some example uses of this 'simple' name

*EZY386 will depart from gate E17 at 2010* [announcement]  
*Just arrived on EZY386* [text message]  
EZY386 flies from Stansted to Avalon  
EZY386 is easyJet's 3rd most popular flight to Avalon  
I prefer EZY386 to EZY387  
EZY386 has an 102% on-time record  
EZY386 was cancelled yesterday  
EZY386 was delayed because of a problem with one of its engines

## 20. Not so unique. . .

So *EZY386* isn't so simple after all

- We seem to be happy to use/understand it to mean a wide range of things
- Up and down some kind of specific/generic or abstractness scale
- Based on clues from the local linguistic context

People are smart

- Computers are dumb
- And natural language understanding is an unsolved problem

None the less it might be that this kind of flexibility in the use and understanding of names would help us with our theories about URIs

- By widening the focus a bit
  - from URIs as such
  - to URIs in a local context

## 21. Local context for URIs

So, for example, maybe the resource/representation distinction is like this

- That is, in at least some cases, different local contexts for a URI
- Signal different points on an abstraction-like scale

So apparently non-canonical behaviour gets brought back into line

For example, contrast these two local contexts:

```
<a href="http://www.example.org/index.html">. . .
```

```
Subject: http://www.example.org/index.html  
Predicate: http://purl.org/dc/elements/1.1/creator  
Object: http://www.example.org/members/1234
```

The cognoscenti will recognise the infamous `httpRange-14` problem

From this perspective, maybe 200 OK should be more flexible

- And no longer mean "Here is a representation of your URI's referent"
- But rather "Here is a representation useful for that URI as you are using it"

## 22. Elaborating context

To put this in a way that connects up with the discussion of indexicals

- Stop talking about a URI's (single) referent
- But rather about a URI-in-context's referent
- Elaborating the 'context' in "Meaning as a function from context to interpretation"
  - to include the local as-it-were linguistic context
  - or, alternatively but equivalently, to include intent
    - That is, whos's asking/why are you asking

Something like this seems preferable to just throwing up our hands

- And declaring the SemWeb and the OFW to be two different Webs
  - With two different core constituents
  - Which just happen to look the same
  - And be called URIs

## 23. Compositionality: past its sell-by date?

Phenomena such as syncategorematicity, indexicality and the specific/generic distinction

The lion ate the antelope The lion is a fearsome beast
---

All undermine, in one way or another, strict compositionality

- Typically by some form of delayed binding
  - Hacked into compositionality by packaging up some kind of discriminator for 'later' use

Absent a principled alternative to compositionality, maybe such a move is needed for URIs

- That is, giving up on treating URIs as completely-context-independent identifiers

## 24. RDF already isn't compositional

The naive model-theoretic semantics of RDF graphs is alleged to be compositional

- The URIs labelling vertices and edges are interpreted as individuals in the model
  - Independent of context

But it's not clear this is really the case

- `rdfs:seeAlso`
- `rdfs:definedBy`

These seem to refer not to individuals, but to collections of RDF

*Thanks to Jonathan Rees for bringing these to my attention*

## 25. Two reasons why URIs aren't names

1. In natural language, names are easily discoverable
  - That is, we know what to call things
    - This is why even the Kripkeans can't dispense with descriptions
  - There is *nothing* corresponding to this for URIs
    - Except search engines!
    - Do I hear you say "aha, the extended mind"?
2. Most natural language names are not 'actionable'
  - Most URIs are

## 26. Conclusions

`httpRange-14` points to a deep failure of registration between theory and practice on the Web

Both theory (standard) and practice will have to change to reconcile them

The traffic is not all one way: I think Philosophy of Language has to reconsider the whole matter of compositionality