

Persistent Reference on the Web

Jonathan Rees
Creative Commons
W3C TAG F2F 20 October 2010

Outline of the memo you read

- Persistence and persistent reference
- Gambling / trust
- Threat model
- Prevention and remediation tactics
- Case studies

Our choices

- Embrace pluralism
- Advocate for 'Just Say HTTP' - **convincingly**
- Status quo

Persistent references in scholarly publishing

- Long reference (persistent)

Wilson ET, Wong J, Gribble PL (2010)
Mapping Proprioception across a 2D Horizontal Workspace.
PLoS ONE **5**(7): e11851.

- Compact reference (~persistent)

doi:10.1371/journal.pone.0011851

- Hybrid reference (persistent; clickable)

Wilson ET, Wong J, Gribble PL (2010)
Mapping Proprioception across a 2D Horizontal Workspace.
PLoS ONE **5**(7): e11851.

[doi:10.1371/journal.pone.0011851](https://doi.org/10.1371/journal.pone.0011851)

[this hyperlinks to <http://dx.doi.org/10.1371/journal.pone.0011851> !]

What this becomes with 'Just say HTTP:'

- Long reference (persistent)

Wilson ET, Wong J, Gribble PL (2010)
Mapping Proprioception across a 2D Horizontal Workspace.
PLoS ONE **5**(7): e11851.

- Compact reference (?persistent; clickable)

<http://dx.doi.org/10.1371/journal.pone.0011851>

- Hybrid reference (persistent; clickable)

Wilson ET, Wong J, Gribble PL (2010)
Mapping Proprioception across a 2D Horizontal Workspace.
PLoS ONE **5**(7): e11851.

<http://dx.doi.org/10.1371/journal.pone.0011851>

W3C's example: 'Just Say HTTP'

- Long reference (persistent)

XML Schema Part 1: Structures. W3C Recommendation, World Wide Web Consortium, 2 May 2001.

- Compact reference (?persistent; clickable)

<http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>

- Hybrid reference (persistent; clickable)

XML Schema Part 1: Structures. W3C Recommendation, World Wide Web Consortium, 2 May 2001. This version is <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>.

Just say HTTP: position would be...

“The TAG encourages the prudent creation and use of http: URIs in situations where compact persistent identifiers are needed. {? Steps are being/have been taken to make some http: URIs worthy of everyone’s confidence.} The community is urged to trust {certain http: URIs} and may have confidence in them because {reasons}.”

Scheme-agnostic position would be...

“The TAG encourages the use of URIs in situations where compact persistent identifiers are needed. Creators of persistent identifier systems should publish, via IETF, methods to map p.i.’s to equivalent http: URIs. Web client developers are encouraged to implement all mappings so registered.”

Status quo would be...

ISSUE-50

State:

OPEN

Opened on:

2005-03-15

worse than either alternative.

**Just Say HTTP:
What would be involved?**

Why don't they Just Say HTTP?

Sources of distrust

- ‘Authoritative meaning’ of http: URIs is defined by protocol behavior, not by a social contract (compare URI and URN schemes)
- Domain names are authoritatively defined by IANA and registrars, who do not have persistence as a core value (compare handle system, IETF)
- “Scruffiness” of Web

W3C *says* http: URIs can be trusted...

“The XML representation of schema components uses a vocabulary identified by the namespace name <http://www.w3.org/2001/XMLSchema>.”

- XML Schema Part 1: Structures Second Edition, 2004

... but it's not believed

- Existing normative specs that define the meaning of particular http: URIs attempt to override the authority of RFC 2616 and IANA...
- as long as 200 responses are in harmony with the spec, most of us don't care...
- but having two different legitimate "corrects" that might disagree leads important constituencies to mistrust all http: URIs.

Arguing for trust...

- Obtain agreement that some specs can sometimes override authority of http: protocol in determining “authoritativeness” (correctness)
- Declare that some specs can sometimes override the authority of IANA...

- *or* -

- “fix” IANA by obtaining agreement that certain parts of domain name space are to be as persistent as, say, URN scheme names

Risks

Just say HTTP: risks

- Too hard to influence IANA (ICANN)
- Too hard to convince others to trust it
- We might build it and not have anyone come

Scheme-agnosticism risks

- Poor infrastructure uptake (unclickable links)
- ‘Dividing the web’

Status quo risks

- Exploding demand (ORCID, URIs in data sets, archived RDF, ...) will lead to more and more kludges and inconsistencies
- Inconvenience to users
- Growing divide between mainstream publishing/archiving/library world and Web

Relatively easy issues - don't let them distract you

- What would DOIs and handles be as URIs
- Identifiers for non-documents
- Metadata best practice
- Persistence best practice
- Relation between the resource `<urn:x:y>`
and the resource `<http://x-urn.org/y>`