

# PRESTO

---

*A WWW information architecture for legislation and public information systems*

## Slogan

“All documents, views and metadata at all significant levels of granularity and composition should be available in the best formats practical from their own permanent hierarchical URIs.”

## Technologies

- **P** – Public, Permanent URLs<sup>1</sup>
- **REST** - Representation, State Transfer<sup>2</sup>
- **O** – Object-oriented<sup>3</sup>

## Advantages

- Incremental: encourages mash-up<sup>4</sup> approach and infrastructure for AJAX<sup>5</sup> and APP<sup>6</sup>
- Low-risk: fits as layer on top of existing information systems
- Based on COTS systems: no exotic or complicated technologies
- Based on standards and key web concepts: not widely comprehended<sup>7</sup>
- Vendor neutral
- Maximize current investment: formats and data
- Pragmatic

## Details

A WWW information system that conforms to the PRESTO architecture has the following characteristics:

- Public, permanent URIs for all documents. Following REST principles, there may be multiple representations of the resource available in different formats and notations, with normal access to the URI retrieving the document in the most browsable human readable format: typically HTML or PDF. A *resolver* is used to map from the public, permanent URL to the system-dependent URL for that resource. The public, permanent URL has no system-dependent or format-dependent features. The public, permanent URI represents the idea of the document as the user thinks of it.

For example, <http://www.eg.gov/legislation/2008/ChildProtectionAct1945/>

---

<sup>1</sup> See <http://www.purl.org>

<sup>2</sup> See <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>

<sup>3</sup> See <http://en.wikipedia.org/wiki/Object-oriented> notably on encapsulation

<sup>4</sup> See <http://2006.xmlconference.org/programme/presentations/61.html>

<sup>5</sup> See <http://en.wikipedia.org/wiki/AJAX>

<sup>6</sup> See [http://en.wikipedia.org/wiki/Atom\\_\(standard\)](http://en.wikipedia.org/wiki/Atom_(standard))

<sup>7</sup> See <http://www.w3.org/Provider/Style/URI>

- Metadata, indexes and views of the resource are available as sub-resources of the document's public, permanent URI, using hierarchical URIs. Following object-oriented principles, methods are tightly coupled with data objects according to class: in a REST system, this coupling is reflected by the use of generic URI steps under the document's URI. Generic subresources (yet to be finalized) include
  - Index – an index
  - toc – a table of contents
  - metadata – metadata in whatever format is available
  - dc-metadata – the Dublin Core metadata
  - types – a list of the available representations using MIME content type. These content types can be used as subresources for retrieving particular representations other than the default representation

For example, <http://www.eg.gov/legislation/2008/ChildProtectionAct1945/text/xml>

- Within the document, every significant grain of information would be readily apparent to the reader of a document also has a public, permanent URI. The existence of the URI is independent of the system's ability to produce a precise representation of that resource: however it allows other systems, for example annotation and collaboration systems, a name for that grain of information which those systems can then use for linking, mashing-up, etc. For example, <http://www.eg.gov/legislation/2008/ChildProtectionAct1945/PartI/Section3/Para2>
- Representations, metadata, indexes and views of the subresource are in turn available as subresources. The key concept here is that the representation of a resource is the "best-fit". For example, if the document is only available as an old scanned PDF file, then the representation available for the subresource would be the whole PDF file. If the document is available as an indexed XPS file, the representation available might be the specific pages that the grain of information fits on. If the document were a legacy word processing document, brought into Office Open XML, then the representation available might contain the particular information at a paragraph level of granularity with some context paragraphs. If the document were a semantic-rich SGML document, the representation might be the exact information required.
- There can be resource identifiers for composite information. For example, say the public, permanent URI <http://www.eg.gov/legislation/2008/ChildProtectionAct1945/PartI/Section3/Para2> might refer to an amended text, which is stored as an RTF archive document plus a second document specifying the change. The representation returned might be a compound document (using MIME multi-part or Atom mechanisms or Open Packaging conventions) containing as good a fit of (converted) RTF pages plus the relevant pages for the amendments. In such a case, if exact enough metadata concerning how to merge the two were not available, a human would have to figure out the effective text by reading. However, because there is a public, permanent URI for the amended text, it can be used by other systems, for example for cross referencing.