UKGovLD

the UK Government Linked Data
Working Group
data.gov.uk/linked-data/UKGovLD

Revision of UK Public Sector URI Guidance

Summary of Survey Responses

Version 1
12 Feb 2013

Authors:	Stuart Williams (Epimorphics ltd)	12/02/2013
Review:	Alexander Coley (Defra)	

Introduction

This document presents a summary of responses and recommendations for work arising from a survey[1] devised to inform the process of revising the data.gov.uk URI Sets document "Designing URI sets for the UK public sector"[2] first published by the CTO Council of the UK Cabinet Office in 2009.

Invitations were sent to 44 individuals across the UK public sector, including government departments and agencies, devolved administrations, local authorities and trading funds. Invitations were also extended to IT suppliers involved in the publication of public sector linked data and to individuals in the EU investigating linked data publishing.

17 responses were received.

Executive Summary

This section summaries the main points of agreement and difference expressed in survey responses. It summarises the obstacles given in survey response to the adoption of linked data publish practices and it summarises additional comments made by respondents at the end of the survey.

Points of General Agreement

- Background, definitions, framework and principles are all found useful especially for people new to the topic.
- Whilst there is support for continued use of the current **{sector}.data.gov.uk** based patterns most response recognise a need for more flexibility to:
  - address the needs of trading funds, local-authorities and devolved administrations;
  - better enable organisation/administration of shared URI spaces.
- Lack of an effective governance framework and accessible sector oriented stakeholder has made it difficult for new-comers to engage, even to know how and with whom to engage.
- With the exception of spatial-objects (**../so/..**) the use of **../def/..**, **../id/..**, **../doc/..** and **../data/..** in sub domain URI patterns is broadly adopted and generally understood.
- Additional technical guidance, possibly in separate documents, is sought on:
  - vocabulary discovery, selection and use
  - dataset registration, discovery and metadata
  - versioning and change over time
  - use of named graphs for data management

Points of difference

- Response behaviours for 'things' (**../id/..**) URI. Most responses support continued alignment with W3C TAG guidance (current and future). One response proposes an alternate approach that avoids 303 redirections and advocates the use of 200 responses and distinct URI for 'thing' and 'description of thing'.
- One response argues that "... *open public data... of a national scope, even if managed from a trading fund, should be published at data.gov.uk* ..."

---

[1] https://www.surveymonkey.com/s/Y8KVYKR
[2] http://www.cabinetoffice.gov.uk/sites/default/files/resources/designing-URI-sets-uk-public-sector.pdf

Barriers to the adoption of link data publishing

The main obstacles cited in responses are:

- Costs, especially maintenance. Difficulty in articulating benefits to budget holders
- Lack of infrastructure and process available to potential data publishers seeking to publish at data.gov.uk
- Commitment to persistence, longevity and quality and the need for supporting infrastructure.
- Plain English documentation - the technical language of much writing is inaccessible to 'newcomers'.
- Lack of high-level leadership an buy in.
- Lack of compelling examples - failure to communicate benefits.

Additional comments

Survey response additional comments mention

- Need for publisher to focus on the usability of a data source
- Integration of "Designing URI sets for the UK public sector" and its companion "Designing URI sets for Location".
- Need for executive briefing material.
- the practical difficulties of implementing linked data the face of 'simpler' JSON and XML based approaches better aligned with developers current skill sets.

Recommendations

It is evident that there a number of audiences (policy makers, management, data owners and publishers, IT services...) with varying needs. People wrestling with building a case to publish data and people wrestling with the practicalities if actually doing so.

Rather than revising a single document we should consider dividing the material into a suite of related documents:

- An overarching guide or map of the document suite to act as an entry point for all comers. This should also contain any aspects that bridge between technical and policy guidance eg. Principles, Framework and Principles.
- Repackage and expand background and motivational material targeted on policy makers and case builders. Extend to cover governance (policy aspects) framework - and the 'publishing commitment' (which should be expressed in metadata - and could be less than 'forever'). This should also include explicitly identifying status of URIs (for example as experimental dataset within the metadata)
- Factor out the core technical guidance on URI patterns into a relatively compact reference guide accompanied by a more practically oriented "How-To" guide that is backed by examples from actual deployments.
  The core patterns need to be revised to accommodate
    - non-data.gov.uk base URI

- - - Patterns organised around larger grained units of publication eg. datasets or collections in-order enable organisationally independent URI with relatively coarse-grained, but reconfigurable, request routing (proxy and/or redirect) to the infrastructure operated by the 'current' custodian of the data.
  - Possibility converge URI patterns for location with generic pattern document.
- Develop separate technical guidance on:
  - Dataset metadata for expressing provenance, data quality, scope and coverage.
  - Versioning wrt to datasets and data items whose properties vary with time (including the use of named graphs).
  - Approaches to dataset discovery.
  - Data management using Named Graphs.
  - Publishing linked data for use by developers.
    - Listing patterns, developer formats - LDA or alternates.
    - File based, Triplestore based and RDB adapter based publication
    - Recommended Vocabularies and their use (possibly by reference to existing tutorials) Org, Cube, FOAF, SKOS... (others)

The biggest areas of challenge remain demonstration of value and generating sufficient critical mass that publishing linked data becomes self-sustaining. There is a question in the background about the sustainability of linked data publishing. There is reticence to depend on it because of lack of confidence in its sustainability. This leads to a lack of reuse and duplication that undermines the potential for linked data generate savings through sharing and reuse. Guidance alone cannot address that challenge.

Detailed Summary

The following subsections present summaries of each response made for each section of the survey. In some cases a summary is followed by a more detailed review of comments from the responses received. The final subsection simply contains the final additional comments offered by some respondents.
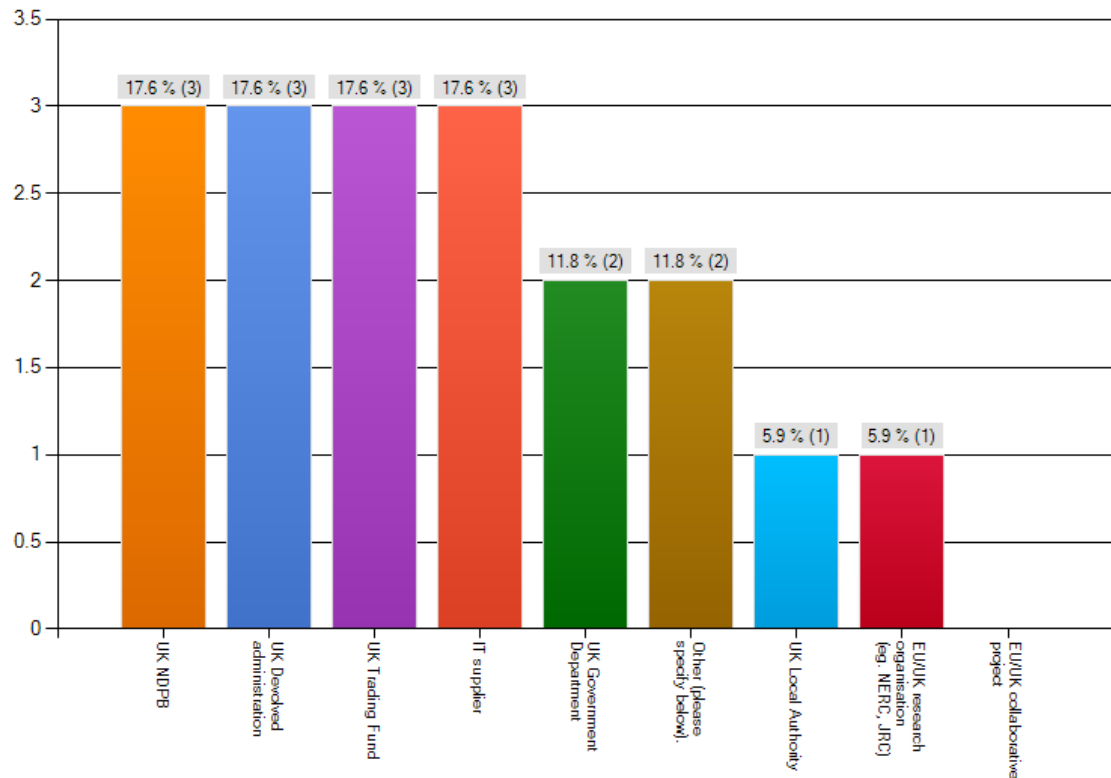
Survey Demographic

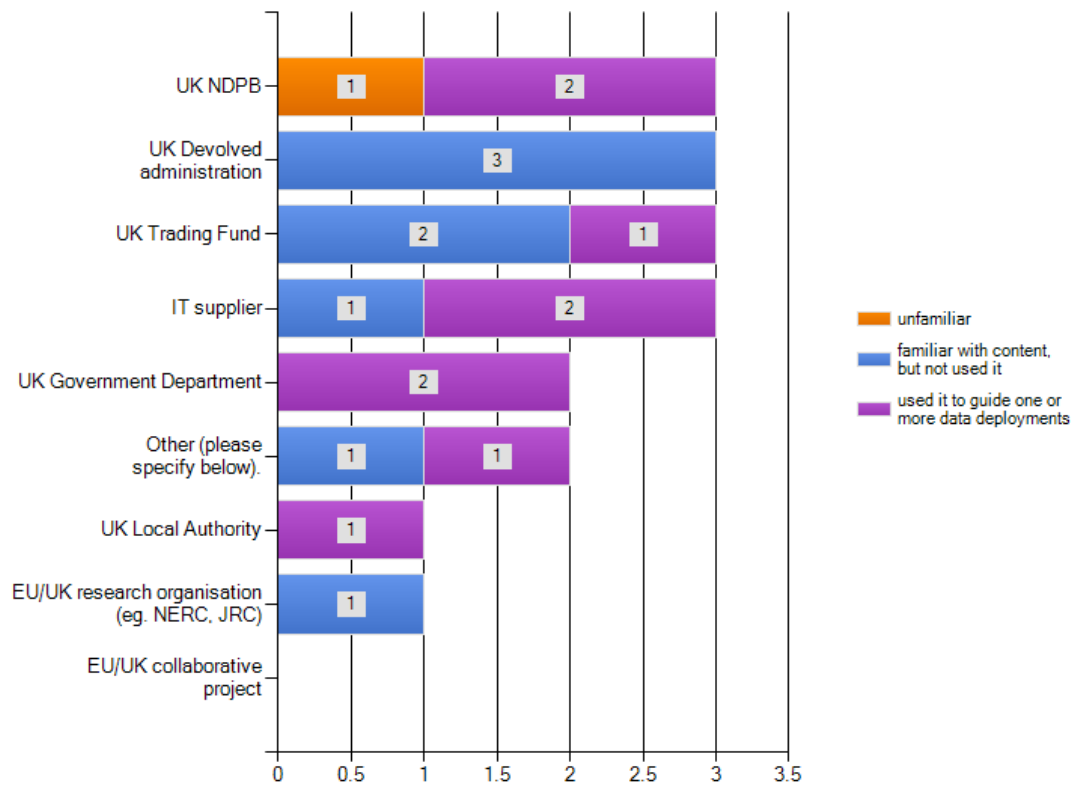Invitations to participate in the survey were send to 44 participants spanning:

- UK Government Departments (7)
- Devolved Administrations (6);
- Local Authorities(4)
- Other UK Public Bodies (NDPB's, BBC, ODI TSB) (5);
- Trading Funds(7);
- UK Academia (2)
- EU Administration (4)
- IT suppliers(8)
- Other(1)

The survey received 17 non-empty responses. The charts below show the spread of organisation represented by in responses and their levels of awareness and use of the existing guidance.

Response Demographic

Category	Value
UK NDPB	17.6 % (3)
UK Devolved administration	17.6 % (3)
UK Trading Fund	17.6 % (3)
IT supplier	17.6 % (3)
UK Government Department	11.8 % (2)
Other (please specify below).	11.8 % (2)
UK Local Authority	5.9 % (1)
EU/UK research organisation (eg. NERC, JRC)	5.9 % (1)
EU/UK collaborative project	

Awareness and Use of current Guidance

Legend:
- unfamiliar
- familiar with content, but not used it
- used it to guide one or more data deployments

Utility of the current guidance

"What aspects of the current guidance have you found most and least useful (some of which are listed below)?

- **Motivation/Background**
- **Definitions, Framework and Principles**
- **Domains name and Pattern guidance**
- **Formats and behaviours**
- **Change and Time**
- **Governance**

What would you most like to see added/extended or removed?"

17 open ended responses:

     unfamiliar (1), familiar but not used (8), used to guide one or more deployment (8)

Summary (Utility)

It is clear that the material on motivation**, background, definitions, framework and principles** are regarded as useful by all. These aspects seem to be of most use to people coming to public-sector linked data publishing for the first time. These aspects are less useful for the more experienced practitioners and who have a need for more technically oriented guidance on:

- **URI patterns** covering datasets and non data.gov.uk situations - trading funds, devolved administrations and local-authorities)
- Dataset discovery, URI set and dataset metadata
- Change over time
- Use of named graphs in data management

There is a widely perceived need to improve the coverage of crucial area of **Governance**, crucial because it is a point where policy and technology meet. Ultimately it is an expression of public-sector intent with respect to sustainable linked-data publishing.

More detail follows below.

General

Across all responses, **Motivation/Background, Definitions, Framework and Principles and Domain name and pattern guidance** were found useful - and particularly so for those that have not yet used or attempted to use the guidance in a deployment.

More examples throughout are requested in several responses.

One response requests more introductory material on the focussing on business value and cost reduction over transparency:

> *"...the place of linked data technologies as a means of **reducing cost** and **enhancing productivity** within an organisation ("self re-use"????!) - as opposed to just a transparency discussion."*

Governance

The material on **Governance** was called out as least useful in several responses and in need of more development. The need for senior management awareness of what it means to be a custodian of a set of URIs was mentioned:

> " Governance is a key area - currently the lack of an underpinning governance framework does not make starting in this area as easy as it should be."

> "... governance: given that in the end all things change, I believe the paper should acknowledge the need to ensure that "being the custodian of a set of URIs" is something that is highlighted to senior management so that it isn't forgotten during organisational change - or simply relegated to be considered alongside re-hosting a website."

> "... as a UK Trading Fund we would like the guidance to reflect the ability to publish persistent identifiers within our own domain-name & governance rather than being required to use (only) data.gov.uk. Given the need to support multiple governance domains, a discussion of how one might discover "authoritative" URI sets (e.g. those endorsed for use by data.gov.uk - including those published outside the data.gov.uk domain) would be useful.

Patterns

The guidance on **Domain name and URI patterns** was generally found to be useful. However there are requests for more examples and for patterns to be extended to address the needs of trading-funds and devolved administrations outside of the scope of data.gov.uk in order to serve the needs of trading funds, devolved administrations and local authorities.

There is also a suggestion that:

> " The guidance could also improved around how to choose suitable keys or identifiers to use within URIs. Perhaps referencing patterns such as Shared Keys, Patterned URIs, etc. "

Dataset/URI set Discovery and Metadata

The need for the guidance to cover the **Discovery of authoritative URI sets and Datasets** was mentioned. One response remarks on a seemingly artificial distinction between URI sets and a more general notion of dataset:

> "However I find this to be somewhat artificial, typically when undertaking a data conversion and publishing exercise we are working with datasets. When decomposed, elements of those datasets might map to the "URI Set" concept, **but there's currently no guidance on how to describe datasets, or what URI patterns to use**. I'd argue that many of the quality criteria for URI Set really apply to the notion of a dataset. I'd like to either seem some greater discussion on the relationship between URI Set and Dataset, or perhaps the notion of URI set replaced with the notion of datasets and sub-sets, which would align with VoiD."

There is an implicit appeal there for more guidance on the publication of **URI set/Dataset metadata**. The need for a concrete guidance on metadata vocabularies is mentioned in other comments.

"Some recommendations on how to point to a bulk download of a set. It is cumbersome to have to get e.g. one bus stop at a time. Likewise, a way to declare what is available in some sort of manifest."

Change over Time

Change over time is recognised as a complex and important topic that is not adequately addressed in the current guidance.

"This is a very complex area and I've yet to see any convergence on approach for handling it properly. While I don't expect that the document would be able to cover all options, I think having more discussion and examples would be useful. Where identifiers are derived from existing literal keys, then the governance of the original identifier ought to be taken into account when considering the stability and change characteristics of the URI. Postcodes, for example, are not stable, over time they can be re-used."

Use of Name Graphs for Data management

One response notes:

*"An important step in our publishing approach, which we have found quite useful, but isn't touched upon in the guidance, is to **use named graphs for organising data** - and in particular we tend to associate a dataset with a single named graph. Dataset metadata usually goes in a separate named graph. An ontology has a named graph etc. I appreciate there is a variety of viable approaches here - but some discussion on sensible approaches to use of graphs for data management might be useful."*

Current Patterns and Practice

"Current data.gov.uk guidance takes a thematic-sector approach to the top level partitioning of the data.gov.uk URI space intended to establish persistent (or at least semi-persistent) URIs that remain stable in the face of organisational change - eg.

- **http://education.data.gov.uk,**
- **http//:transport.data.gov.uk,**
- **http://environment.data.gov.uk.**

a) **To what extent do you think this approach: - has been successful - should be continued and promoted - should be grandfathered and replaced by some other approach (if so - what?)**

b) **Do you have any comments on both the governance and the practicalities of data publishing in such multi-tenanted URI spaces?"**

17 open-ended responses: unfamiliar (1), familiar but not used (8), used to guide one or more deployment (8)

Summary (Current patterns and practice)

Responses are generally supportive of the continued use of the **{sector}.data.gov.uk** approach, but not to the exclusion of the use of URIs outside that scope. There is perceived benefit to be had through reuse of common patterns under a more flexible root structure.

The lack of effective governance frameworks is a significant issue in creating impedance to engagement, through real or apparent lack of ownership, leadership and established practices for sharing the URI space of a sectored domain.

Practical difficulties highlighted include:

- Administering a shared URIs space, both from the point of view allocating URI for use and from the point-of-view of routing requests for information to authoritative sources.
- Transitioning the provision of services between suppliers
- Choosing the right sector - there is not always a clear right answer to what sector should be used in a given case (independent of whether the expression of a sector in a URI occurs in domain name or somewhere within a path).

More detail follows below.

Governance

Most responses recognise and understand the intent of the current patterns to create organisationally independent persistent reference URI.

However, responses also highlight **a lack of effective governance**, which serves to create impedance for 'newcomers' due to lack of a model for engagement with any kind of {sector} stakeholder group; and to ultimately undermine confidence in the intended persistence of the due to the apparent lack of ownership (by anyone).

> *"However, there has been a lack of visibility (and perhaps ownership) of these domains - only 'location' and 'environment' domains seem to be actively in use. Furthermore, it is not clear*

how a data publisher requests a segment of the multi-tenanted (sub-) domain. This process, along with identification of the governing authority for each domain, needs to be made clear."

"I've never been clear what the final set of top-level domains will be, and who gets to administer them. I can see that having a department/organisation independent domain(s) for central government data is encourage stability, but in practice the existing domains are either effectively "orphan", or they're managed by a particular department anyway."

"a) The approach feels right - however, we have not had the leadership to set up 'sector' sub domains and make them the responsibility of a lead department/organisation as per the described governance arrangements. We should have a complete set of these themes, perhaps aligned to the 'sector open data panels' that have been set up. b) version 2 of the document has more to say on the governance of Sector, and then the governance of Sets within a Sector."

"The proposed approach is conceptually the preferred approach. The practicalities however are extremely difficult to manage, as it requires an active custodian and stewardship to manage (and promote) the multiple tenants that are a feature of each domain. This is what we have failed to achieve so far. If such arrangements cannot be secured, then alternative approaches will have to be allowed for, with the top level partitioning reflecting what is realistically deliverable."

Flexibility

The guidance is scoped to **{sector}.data.gov.uk** URI and as such fails to meet the needs of entities such as devolved administrations, trading funds and local authorities.

"This approach seems to work for non-trading fund bodies. However, trading funds will require/desire branded URI patterns"

Having said that, much of the logical structure of the current guidance could be reused in 'lower' order parts of the patterns adopted for use under other governance regimes.

"The generic URI-pattern guidance for UK public sector can be expressed as: ***http://{base}/{stem}/{vocabulary}/{reference}[:{version}]*** *Where: - **{base}** is the subdomain within which the data is published plus any fixed path segments (e.g. 'environment.data.gov.uk') - **{stem}** indicates the category of entity (e.g. 'thing'* ***/id/,*** *'document'* ***/doc/,*** *'definition'* ***/def/,*** *'data'* ***/data/,*** *'spatial object'* ***/so/*** *etc.) - **{vocabulary}** enables the identifiers to be partitioned semantically (e.g. 'bathing-waters') - **{reference}** is an identifier that is unique within the scope of the preceding URI path - [optional] **{version}** defines a particular version of the resource..."*

Practicalities

Practical issues that are raised by responses include:

- Administration of a shared URI space and the routing of requests to an authoritative source.

"When two or more organisations want to publish data in the same domain, partitioning up the URI space in a sensible way, and setting up and maintaining the proxying to relevant servers becomes more complex than if the two organisations had simply had separate domains."

- The need to able to transition the services that deliver data between different suppliers consistent with a general notions that any contracts for such service need to be capable of being retendered from time to time:

   *"I'd vote to consider a move to a common sub-download, e.g. **id.gov.uk**, and have that operate as a purl type service that would allow anyone in government to register new sets within that space. ....**Transitioning services between suppliers needs to be possible** and that requires mapping not just entire domains, but individual sets to different hosts."*

- Choosing the relevant sector:

   "I guess the issue may be in having to make an arbitrary decisions when a URI set may potentially fall into more than one categorisation."

   "The thematic approach I believe is useful and is consistent with modern thinking on data structures. My greatest concern is how can I use this to design my URIs at the local level because there seems to be very little guidance on this, at least at NI level."

   "...from discussions within the EC, we can see some problems with this approach, e.g. - what should be the level of granularity for the sector (e.g. environment, climate-change or emission-trading)? - how to guarantee persistence of URIs when the responsibility for a certain sector gets split up between different institutions (e.g. DG Environment at one point was split into DG Environment and DG Clima)?"

Thematic Sectoring

"How could the guidance be revised to cater better for trading-funds, devolved-administrations, local-authorities and other public sector bodies e.g. the {sector} concept could be retained further down the URI path: http://{stem}/{sector}/.... and the remainder of the practices adopted 'underneath' that revised root. The significance of a given {sector} may or may not be aligned across public-sector bodies - enabling larger or smaller groupings to participate in community-of-practice and/or governance activities.

a) Do you have any thoughts on the prospect of such cross (UK) public alignment and shared practices?

b) If you represent a non-central government body (trading-fund, local authority, devolved administration, other public sector body - eg. BBC) would you welcome guidance that could be adopted on a voluntary basis?"

13 open ended responses:

 unfamiliar (0), familiar but not used (5), used to guide one or more deployment (8)

Summary (Thematic Sectoring)

There appears to be willingness to consider **voluntary** alignment around the concept of thematic sectors independent of where it expressed within a URI (domain name or path segment). However, at least one response questions whether the 'sector' notion is quite so universally applicable.

More detail follows.

General

The principal tension appears to be between identifiers being aligned with organisational structures in contrast to identifiers aligned with thematic sectors. Without effective stakeholder groups providing governance over a given sector it is difficult for organisations to align URIs used in the publication of data on purely thematic basis.

> " The practicalities are such that structures are likely to remain quite strongly linked to individual organisational ownership. I have access to "data.wales.gov.uk" and can manage and operate that. If I am publishing education data, I can put that under that domain, with no dependencies elsewhere. Therefore, that is what I am most likely to want to do. The guidance needs to reflect on those realities and work out what can sensibly be achieved in terms of common approaches."

> "For trading funds, I would like to see the guidance permitting the 'minting' of identifiers within their own (sub-) domain. Given the nature of a trading fund, it is essential that we retain a coherent "brand-message"; this is reinforced by the use of our own domain-name in publication of data, content and services. I think that the URI-pattern guidance could be adopted 'beneath' the trading-fund's (sub-) domain."

However there is a willingness to use/reuse sectoring within more organisationally aligned URI patterns where it makes sense.

"We would be comfortable with putting {sector} within our URIs, where the object we are identifying relates to a specific sector. We would be happy to see guidance that we can adopt on a voluntary basis."

However the contra view was also expressed:

"Given the broad nature of activities undertaken by trading funds, use of the {sector} URI-segment feels inappropriate."

There is some level of acceptance that trading funds and devolved administrations will need to be able to publish data within domains they wholly control:

" I think [I] accept that data.ordnancesurvey.co.uk, data.metoffice.gov.uk etc are acceptable domains and don't try to lever them into a fixed pattern"

However, that acceptance is not universal:

"I feel that URIs that are open public data, and of a national scope, even if managed from a trading fund, should be published at data.gov.uk. We should encourage links to be made to data.gov.uk URI(s) in preference to those rooted at a trading organisation."

Sub-domain URI patterns

"Current data.gov.uk URI guidance advocates the use of some top-level patterns below the top-level sectors.

- **/def/{vocab}/{term} for vocabulary and terminology definitions**
- **/id/{concept}/{instance} for things (people, places, events)**
- **/doc/{concept}/{instance} for reference data about things (and documents like things in general)**
- **/data/{dataset}/... for datasets and data items**
- **/so/... for spatial-objects (INSPIRE)**

a) **Are patterns like this useful:**
   - **in establishing and communicating behavioural expectations of individual URI;**
   - **in organising URI spaces beneath some common 'root' or 'authority';**
   - **in appealing to the intuitions of data publishers and data consumers (end-users and application developers).**
b) **Are there more categories of subjects items that it would be useful to indicate through additional top-level patterns?**
c) **Do these patterns hit the right balance in terms of appealing to a users intiutions and maintaining a level of opacity:**
   - **Semi-opaque e.g.: /id/bathing-water/ukc2102-03600**
   - **Fully-opaque e.g.: /3F281000-E95A-11d2-886B-00C04F869F04**
d) **Outside of 'pure' {sector}.data.gov.uk space, are these sub-patterns useful in the context of data-publishing by trading-funds, devolved-administration, local-authorities and other public-sector bodies operating outside of "*.gov.uk" space."**

14 open ended responses:

> unfamiliar (1), familiar but not used (5), used to guide one or more deployment (8)

Summary (Sub-domain URI patterns)

Responses are generally supportive in principle of the utility and value of common URI patterns beneath some root (or stem) URI with a domain. Most find the current cluster of **.../id**, **.../doc**, **.../def** and **.../data** relatively clear and intuitive. **.../so** remains a source of potential confusion and debate - whether spatial-objects are 'thing'-like tending toward '**.../id**'-like behaviour or whether they are more 'document'-like tending toward '**.../doc**'-like behaviour.

One response questions the current top-level partitions and provides a sizable description of an alternative that should be consider as the details of the guidance are revised.

The issues connected with multi-tenanted sub-domains and the routing of requests re-surfaces with suggestions that some reorganisation of the pattern components would make it easier to administer at servicing of requests - both for the allocation of segments of URI space, and for data arising from those allocations.

Semi-opaque patterns are preferred over fully opaque patterns. It is also suggested that the guidance be extended to discuss appropriate keys for using URI construction.

General

Responses show general support for the usefulness of the high level idioms associated with these top level patterns and particularly the top level indicators such as /id, /doc/, /def etc though not necessarily slavish alignment with the exact patterns beneath. General alignment with the spirit.

There is near universal support in the responses for semi-opaque, somewhat mnemonic URI patterns; and support for some level of indicative high-level typing to be apparent in published URI. This is accompanied with a caution:

> *"We believe that mnemonic codes should be used in URIs to facilitate their creation and give make them more appealing to users. However, they should not be used to encode semantics or structure/relationships into the URI."*

> *"Use of patterns is very useful. Semi-opaque, constructable URIs are better than fully opaque, particularly where they can be constructed from public, shared keys. I'd like to see the patterns used in individual datasets documented, e.g. as a formal URI templates."*

However, one response describes the core patterns as problematic:

> *" A fully-opaque URI (e.g. a GUID) would inhibit use - simply because humans are not well suited to using long alpha-numeric strings! Personally, I think that the use of /id/, /doc/, /so/, /def/ (etc.) is problematic. Predominantly, this pattern appears to be used to distinguish between "real" things and "information" things."*

and goes on to field a fairly sizable alternative arrangement worth deeper consideration in the construction of revised guidance.

Another response re-iterates that the current patterns make organization and administration of the space more problematic - largely because they tend toward the creation of multi-tenanted domains:

> "Such patterns are useful. They give guidance to publishers and so reduce the burden of designing a first publication and help consumers in given some expectation of the kind of thing to expect at each pattern. They are not helpful in terms or organization the URI space under one authority. If we move to a registry-based approach for e.g. location.data.gov.uk then it would be significantly better to have patterns more like: ../authority/theme/def/... ../authority/theme/id/... etc. So that one group within one authority can handle all the different types of information for a given theme. I.e. move the def/id/doc/data/so patterns down a (flexible) number of levels."

Spatial object's continue to be a source of discussion/debate/confusion in opposing directions:

> "... although I'm personally not so convinced that "so" is particularly distinct from "doc"."

> "... though I still wonder if the /so/ is confusing and could simply be done using the usual /id/ and /doc/ approach."

> "Furthermore, my experience is that it is _REALLY_ (!) confusing trying to describe the difference/sameness between /id/, /so/ and /doc/ entities!!! It would be much simpler to simply treat _all_ information resources (including spatial objects) as directly resolvable and use a #fragment pattern to distinguish between the real-world resource and the information resource that describes it."

> "I personally don't think spatial objects are information objects...at least not quite in the same way that documents are...maybe" [in response to the premise in the question that follows].

Some responses suggest that some definitions may be more locally scoped to use with a particular data set (e.g. a particular data cube) and their organisation should be more closely aligned with that dataset and its data structure definition (if any):

> "The use of a top-level /def/ pattern is a little limiting: often some vocabulary elements may be relevant to a specific dataset (e.g. terms in DataCube) so these ought to be scoped at a lower-level. I'd argue that /dataset or similar ought to be added to the list."

A seemingly orthogonal comment mentions the practice of creating local holdings (copies) of datasets:

> "Its easier for us to have our own copy of them and manage that copy, than it is to try and manage a "live" link through to OS's linked data set of LSOA boundaries. - the granularity of the data is a concern. I would appreciate guidance on how linked data can work effectively with quite granular data (e.g. the GCSE points score for 2012 for school X)"

The point at issues seems to be whether that is a literal copy or whether its subject URI are recast with a 'local' domain so that references make direct reference to the copies rather that the live data of the original publisher (all modulo some permissive data licensing).

Response Behaviours

"W3C TAG is reviewing its guidance wrt to the infamous httpRange-14 issue.

a) What are your thoughts on the continued use of 303 status code redirections used in HTTP response for /id URI (and in some case for /so URI though this is being rethought - spatial-objects are information objects)?

- **Continue the practice (at least for now)?**
- **Abandon the practice without reservation!**
- **Re-align with any fresh TAG guidance?**
- **Something else?"**

10 open ended responses:

unfamiliar (0), familiar but not used (4), used to guide one or more deployment (6)

The chart below summarises the responses with respect to the options presented in the survey question:

Response behaviours for so-called "non-information" resources.
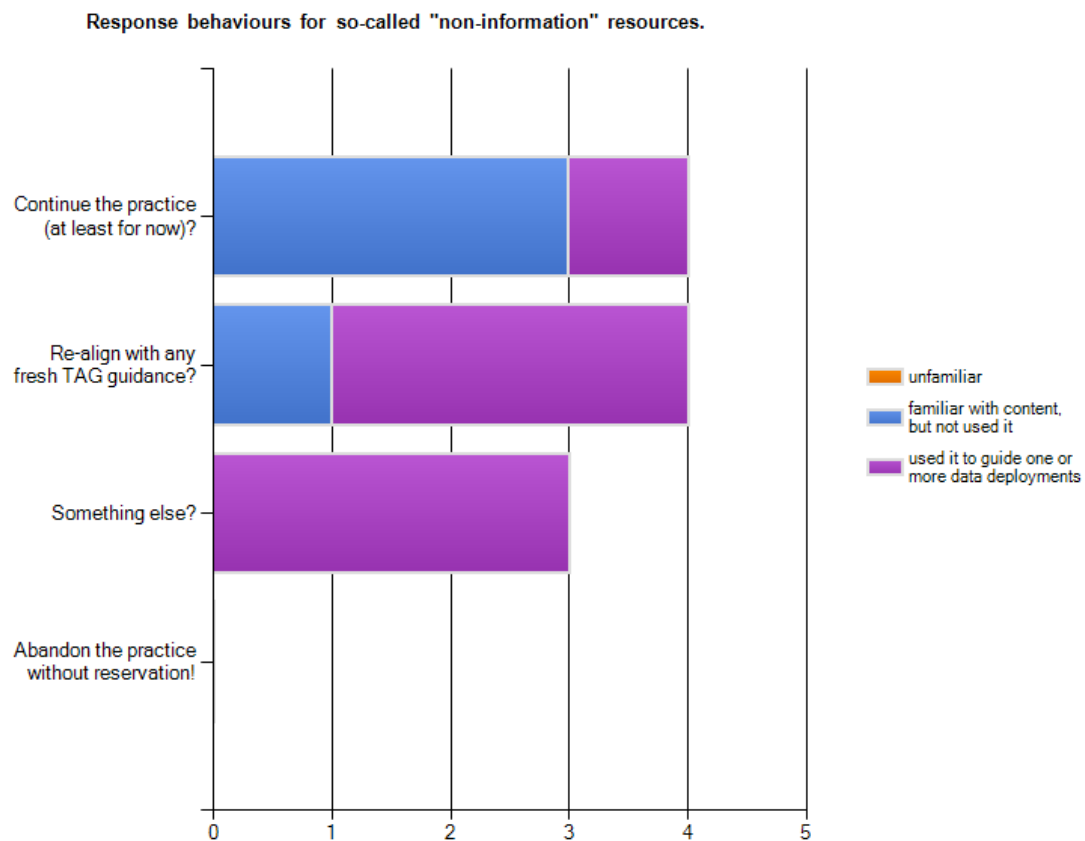
Summary (Response Behaviour)

The general sentiment of the survey responses is to continue current practice and to re-align with any fresh guidance from the W3C TAG.

However as significant number(3) of responses advocate other approaches. Of these, one venture no further opinion; one is aspirational:

"If we can find a way to remove unnecessary complexity to what we link to, and what representations that are dereferenced, then I would be keen to support that. The distinction between real-world and information, can be a barrier to adoption, and I have not come across a practical situation when the distinction was necessary, or not obvious."

and one is more substantive:

"The fresh TAG guidance is a fall back not a replacement to 303. It still says that you should have two URIs (whether 303ed or hash-based). Given the pain that OWL 2 punning causes (me at least) then simply adopting punning as an answer doesn't appeal. **If you need two URIs then have them, if you don't then don't.** *Don't pun.* **For simply serving individual resources we should simply abandon 303.** *We have, as far as I know, precisely zero cases in data.gov.uk where we wished to make different statements about the doc and the id resources. We have plenty of cases where doc URIs have been accidentally sent in emails and presumably used in code. For serving paged lists of resources as in the LDA then there is a role for separation so you can describe the page (and the forward/back links) separately from the subset of resources on that page. However, I think paging behaviour is probably out of scope for the URI guidance. Or is it?"*

Change and Versioning

To what extent do you think the revised guidance should discuss approaches to change over time and versioning of reference data? e.g. the head teacher for a school or the location of a monitoring point may change from time to time.
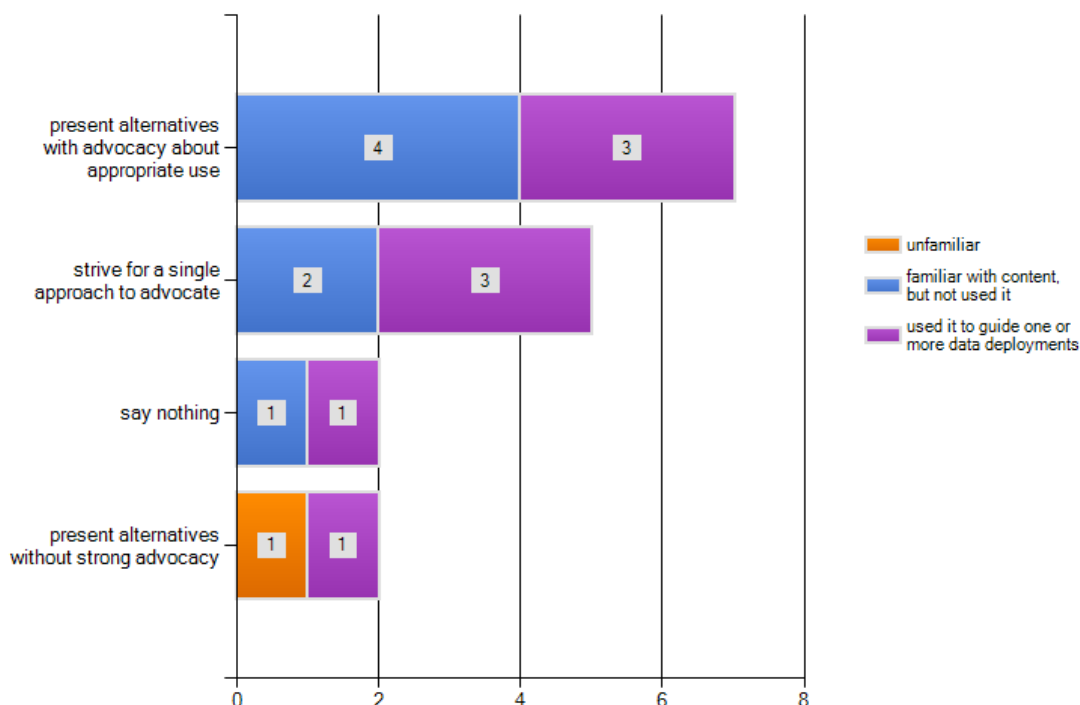
10 open ended responses:

unfamiliar (0), familiar but not used (4), used to guide one or more deployment (6)

Summary (Change and Versioning)

Of the 'say nothing' responses, one argues that the topics is important, but that it should be addressed independently of the URI guidance:

> "*There is a strong case for an advice document which lays out the different approaches (which include data cube, VersionedThing, non-monotonic update, named graphs, out of band versioning). However, there are simply too many different situations (both in terms of data and in terms of infrastructure constraints) to provide a definitive single approach. Such an advice document might be within scope of the URI work but it should not be part of the URI guidance itself.   The guide could, and probably should, include recommendations for *if* your approach to versioning involves versioned URIs then here is the preferred pattern to use for such versioned URIs.*"

The other argues:

> "*This issue concerns the evolution over time of data, which is a separate issue from the evolution over time of URIs for identifying things. Relationships of different versions of things depend on the rules used by data providers. These relationships should be captured explicitly and not encoded in the URIs. This issue should therefore be left out of the guidance.*"

The 5 "strive for a single approach to advocate" responses do not argue the case further beyond:

> *"The guidance currently raises this issue and shows some approaches."*

However, the topic is generally held to be complex and probably more complex than is evident in the guidance as it stands.

The remaining responses acknowledge the importance of the topic and fall in the "present alternatives" with(7) and without(2) "advocacy" categories and argue for more discussion.

Governance

"Do you have any thoughts on governance arranges to ensure the quality, completeness and persistence of published data?"

11 open ended responses:
unfamiliar (0), familiar but not used (4), used to guide one or more deployment (7)

Summary (Governance)

There is a clear tension between a need to provide assurances to a data consumer or user about the quality, freshness, completeness, persistence, accuracy, provenance of some dataset and placing obstacles in the way of a willing publisher.

> *"I think that if a URI set is endorsed by UK public sector, the governance authority must establish that it is fit for purpose. More assessment is necessary to determine what the criteria should be. As a minimum, a data publisher should accept a set of terms & conditions when "registering" their content indicating that they accept the need to persist the data."*

Responses focus mostly on the need to provide:

- metadata with a dataset (or URI set) so that data users can make their own assessments
- active monitoring for 'outage's and for datasets that have become stale.
- dataset discovery
- dataset quality assessment (how-to)
- the nature of the publishing commitment

One response states:

> *"This should be a focus for the document. Although it currently lists quality criteria, we have no examples of the 3 levels of governance that version 2 of the document defines."*[3]

while others comment:

---

[3] Work on a revision to this published guidance has previously been started and it is expected that new material created for that work in progress will be adopted by any process that arises from this survey. However, the version 2 mentioned in the comment was not widely available in a citable form at the time of this survey, hence reference was made to the original version 1 publication instead.

"There's been plenty written on this topic, the essentials cover clear licensing, provenance and publication metadata, links to background and method documentation, etc. I'd suggest that thought needs to be given to how reliability of the data is managed over time."

and

"Guidelines on how to assess quality would be useful, as well as guidelines/encouragement to increase quality. It's hard enough getting people to start publishing linked data without giving them new hoops to jump through."

Barriers

"Are there any obstacles or omissions in the guidance that make it unattractive to publish URI sets?"

10 open ended responses:

       unfamiliar (0), familiar but not used (5), used to guide one or more deployment (5)

Summary (barriers)

The main obstacles cited in responses are:

- Costs, especially maintenance. Difficulty in articulating benefits to budget holders
- Lack of infrastructure and process available to potential data publishers seeking to publish at data.gov.uk
- Commitment to persistence, longevity and quality and the need for supporting infrastructure.
- Plain English documentation - the technical language of much writing is inaccessible to 'newcomers'.
- Lack of high-level leadership an buy in.
- Lack of compelling examples - failure to communicate benefits.

How-To Guidance

What aspects of the linked data publishing process would you most like any practical "How-to" companion documentation to cover? Please tick as many as apply. Please suggest anything else that should be address.
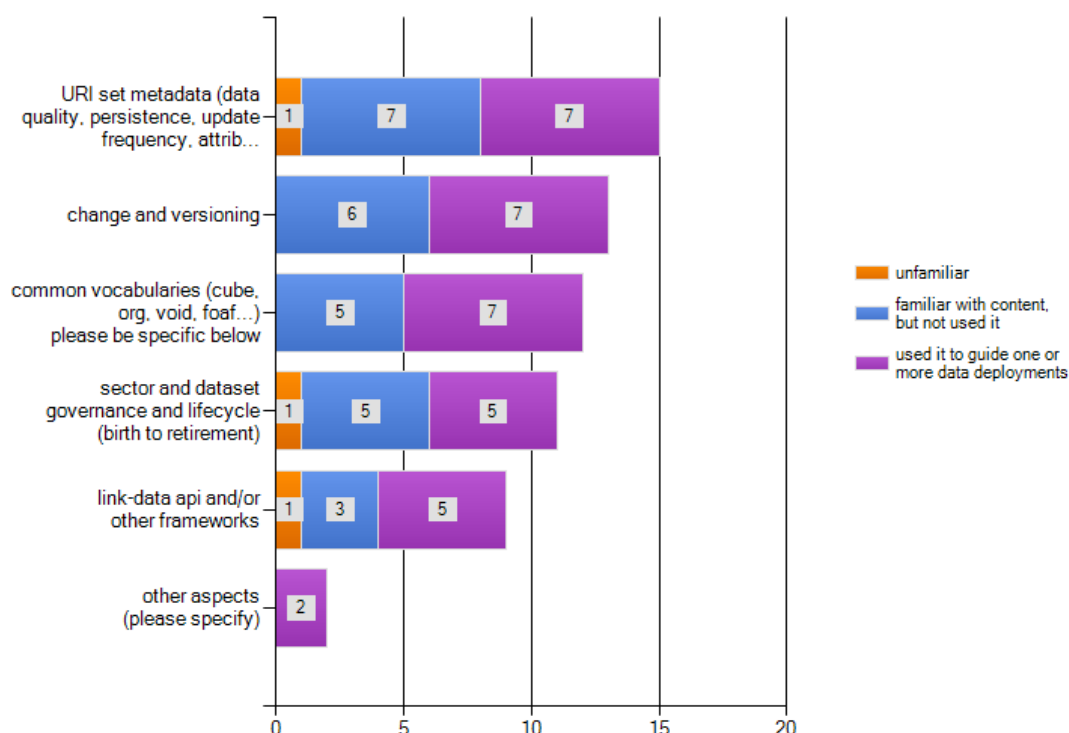
8 open ended responses:

unfamiliar (0), familiar but not used (4), used to guide one or more deployment (4)

Summary (How-To Guidance)

The dominant request from the responses is for practical guidance on the selection and use of common vocabularies both for data (data cube, Org, FOAF, SKOS etc) and metadata (DC, VoID...).

Practical guidance on Change over Time and Versioning is also sought, however, give the complexity of the topic and responses from earlier questions, it may be best to factor substantive discussion of change and versioning into a separate document (or documents) in the interests of keeping the core URI guidance documents focussed.

Under 'other aspects' responses(2) suggest:

"Other: registration of datasets within UK public sector (e.g. @ data.gov.uk) ... see works of Reynolds & Tandy regarding Registries for UK Location."

"...Guidance on first-choice/starting point ontologies would be good. Detailed worked examples of best practice...."

Additional Comments

"Please use the box below express other requirements or thoughts to inform the process of revising the URI sets guidance documents."

5 substantive responses were given to this question

"Publishers should be encouraged to give some thought to the usability of the source data. As the BBC have shown, Linked Data isn't just for developers, it can be a useful part of a service. So as well as encouraging use of URIs and machine-readable formats, publishers ought to think about the utility of their linked data pages, e.g. for someone coming from a search engine. Schema.org is increasingly important, so including relevant embedded metadata ought to be referenced."

"Ways to discover public URI sets, and their quality criteria. Ways to list URIs within a URI set."

"We sometimes found it a bit confusing to have two separate documents for URI patterns in general and for URI patterns for geographic information. Maybe both documents could be somehow combined/integrated? It would also be good to ensure that all the (very useful) examples of the UK location document are consistent with the general guidelines (e.g. /so/ vs. /doc/ vs. /id/ URIs)."

"As well as obvious necessary technical guidance there needs to also be an idiots type guide or clear non technical executive summary which also indicates the benefits in order for us to get senior management level buy into the concept."

"I'm sorry that I've not completed the details of many of the questions given earlier. What I'm reflecting on is where my own team's publication of data has arrived at. We have embraced the open data practices - building our web applications on top of a tier of open data. In practice however, the open data tier has been delivered using primarily JSON and to a certain extent XML, at the end of URLs. This has been inexpensive and simple to implement and has not required re-training, re-equipping or re-skilling of the web and information management staff in my team. We have already built a number of proofs of concept that demonstrate the re-use of these data resources, outside of the interfaces they were implemented to support. The challenge for Linked Data - and speaking from our own experience of doing a linked data solution - is that it compares quite unfavourably with the approach outlined above. We <could> have implemented our projects using Linked Data. They would however, have cost significantly more to deliver (in numerous ways). I would be happy to contemplate such costs, if the benefits were there to outweigh them. That however, is the problem and is perhaps where the biggest amount of work is needed. We published some linked data, but found no-one, other than ourselves was really interested! There wasn't the market demand for it. There's demand for open data, but not specifically Linked Data. We have focussed on meeting the demand for that open data in the most cost-effective way. I recognise that this is properly the work of the ODI to deliver, so I'm not expecting to emerge from this work, but it provides a context........ I

would contend that until there is significant experience out there in terms of publishing linked data, it will be difficult to judge what the "right" guidance is going to be. We need to learn from mistakes, and I'm not sure we've made sufficient mistakes yet to be able to learn things properly. That is why I don't feel able to comment on the previous chapters."