An Information Architecture for the UK Public Sector

A report from the Public Sector Information Domain of the CTO Council's cross Government Enterprise Architecture.

Draft – for adoption by the CTO Council Delivery Group

Version 0.1f

April 2011

Change Control

Version	Date	Editor	Note
0.1a	01/12/2009	Paul Davidson, CIO Sedgemoor District Council	From a pre-workshop survey, and workshop in October 2009, followed by presentations and debate at the Information Domain.
0.1b	14/12/2009	Paul Davidson	Feedback from first draft. For circulation to the Information Domain.
0.1c	07/01/2010	Paul Davidson	Feedback from December 09 Information Domain Meeting
0.1d	21/06/2010	Paul Davidson	Re-alignment to ICT Strategy
0.1e	23/03/2011	Paul Davidson	Alignment to Information Domain Information Management Principles
0.1f	01/04/2011	Paul Davidson	Alignment to Government ICT Strategy – March 2011

Contents

Introduction

1. This paper is a part of the family of policy and guidance documents associated with the cross-Government Enterprise Architecture (xGEA). It sets out the context, scope, themes and principles of the UK Public Sector Information Architecture (PSIA) and plans for its development and onward maintenance and management.

2. The UK Public Sector Information Architecture supports many goals including

    - increased information sharing towards transforming government services
    - promoting information re-use to reduce waste and duplication
    - the release of non-personal public data to support transparency, accountability, and the creation of new economic and social value.
    - better decision-making.
    - economy, efficiency, and effectiveness

3. The architecture is aligned to the Information Management principles devised by the CTO Council's Information Domain. The architecture is comprised of standards organised into seven themes in a secure and trusted framework that supports federation and multiple jurisdictions.

    - **Semantics** - the meaning of information
    - **Syntax** - the format of information
    - **Quality** - the confidence to re-use information
    - **Use Rights** - the right to use information for a defined purpose
    - **Authentication** - who is using information
    - **Transport** - how to move information
    - **Governance and Assurance** - the behaviour and culture to protect and exploit information

4. The architecture will be supplemented with tools that enable programs to assess their use of information standards across each of the themes.

5. Most of the information architecture is drawn from existing policies and established good practice, while some is new thinking. The information architecture provides a cohesive story that walks through how public sector information is both protected and exploited, and signposts where greater detail, and associated tools and services are available.

Why do we need an Information Architecture?

Applying UK Public Sector Information Management Principles.

6. The CTO Council's Information Domain has developed a set of Information Management Principles for the UK Public Sector, supported by explanations and illustrations of each, and references to existing policy and guidance.

7. The Information Architecture provides a framework in which information can be described and assessed against those principles.

• Information is a valuable asset
• Information Is business driven
• Information is fit for purpose
• Information should be reused
• Information operations are minimised
• Linked Information is more useful
• Information always has an accountable owner and a source
• Information is governed by law & regulation
• Information has a lifecycle
• Public have a right to public data
• Citizens have a right to their data
• Governance of information is critical
UK Public Sector Information Management Principles

8. Much of the aspiration towards transforming the UK Public Sector relies upon the ability to confidently and seamlessly share and re-use information across the many departments, agencies and authorities that serve individuals, organisations and communities. 'Public Open Data' is also being published to promote transparency and to provide information that enables people to participate in, and challenge, decisions from the public sector.

9. These aspirations bring together the disciplines of

- Operational Data Integration
- Public Data Publishing
- Knowledge Management
- Statistics, Analytics and Insight

10. The Information Architecture aims to provide unifying standards and frameworks across these otherwise disparate disciplines, leading to greater re-use.

The Government ICT Strategy

11. The Government ICT Strategy[1], published March 2011, emphasises the need and role of a Public Sector Information Architecture.

> Managing information effectively and appropriately is essential to the delivery of secure, seamless and efficient operational services. It provides the basis for informed decision making and the platform upon which performance can be measured. Modern, knowledge-based service delivery underpinned by effective information architecture and open standards will support government to build more transparent, trusted and efficient information exchange processes. The Government will develop an information strategy that is supported by an architecture framework which will underpin the design of government's new information systems..

The current state of Information architecture across the UK Public Sector.

12. Most government departments and agencies have developed information models that guide how their information is managed, deployed, and secured. These models often contain the same concept, (e.g. Person, Event, Circumstance, Place and so on), or similar patterns (e.g. Patient in Hospital, Learner at School, Driver of a Car).

13. The e-Government Interoperability Framework (eGIF), has provided the leadership to harmonise the formats by which data is passed across government, however, there is little to connect the data models from each department or agency, with those that are developed elsewhere in government. In particular, there is no top-level conceptual model that can be universally adopted so that the meaning of information can be confidently passed on, as data is exchanged.

14. As a result, information can be used and re-used within a department or sector, but does not 'travel well' when it has to cross boundaries. This can lead to duplication where the information held by one department or agency, is either not visible, or not trusted by another.
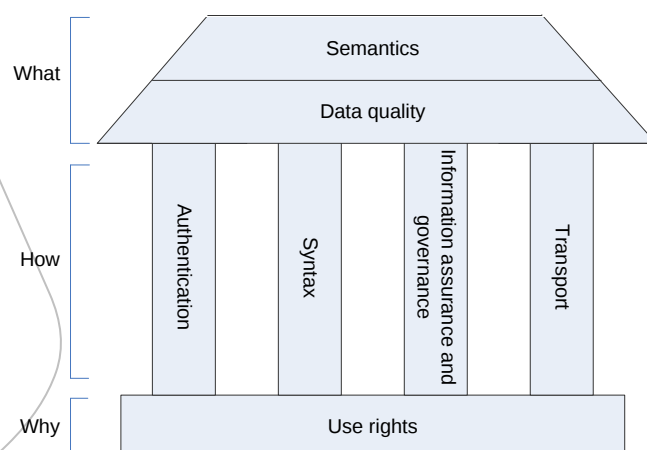
The Architecture

Themes for describing Information

15. The themes of the information architecture, address critical questions about information, its use and movement:

---

[1] Government ICT Strategy - http://www.cabinetoffice.gov.uk/resource-library/uk-government-ict-strategy-resources

- *What is the information?*
  - ○ **Semantics** - the meaning of information
  - ○ **Quality** - the confidence to reuse information
- *Why is the information to be exchanged?*
  - ○ **Use Rights** - the right to use information for a purpose
- *How is the information to be exchanged?*
  - ○ **Syntax** - the format of information
  - ○ **Authentication** - who is using information
  - ○ **Transport** - how to move information
  - ○ **Information Governance and Assurance** - the behaviour and culture to protect and exploit information

16. To illustrate these themes …

If you were passed the data **19630128**, you would not feel any wiser as to how to make use of it. However, if you were told …

This is a Date of Birth of a Person	**Semantics**
The first four characters are the year, the next two characters are the month, the final two characters are the day.	**Syntax**
This data was verified by sight of a Birth Certificate, during a personal visit, by a trusted organisation.	**Quality**
This data is being provided as a part of Claim for a Welfare Benefit using a specified Legal Gateway.	**Use Rights**
The data is provided by, and passed to organisations within the Public Sector who have identified themselves to a level appropriate to the confidentiality of the data	**Authentication**
The data has been passed over a secure network with appropriate quality of service such that the data has not been interfered with since it was sent.	**Transport**
The recipient of the data undertakes to protect it and use it only for prescribed purposes.	**Governance and Assurance**

… then the data becomes useable, and re-useable.

Documents as containers of Information

17. Documents are used to store and communicate information. Documents can be

- Structured | Consisting only of defined data values and definitions conforming to a known system and schema. Capable of being machine-readable.

- Semi-structured | Containing some data values and definitions which could be 'marked-up' to become machine read-able.

- Unstructured | Intended to be human-readable.

18. Documents bring together a collection of information at a point in time, and become useable when their context is properly understood. The context of a document is typically defined using a common format of metadata, that then applies to the whole document.

19. Data is the individual values and statements that can be contained in a document, and become useable when the concepts that they refer to are defined and linked in Logical Data Models.
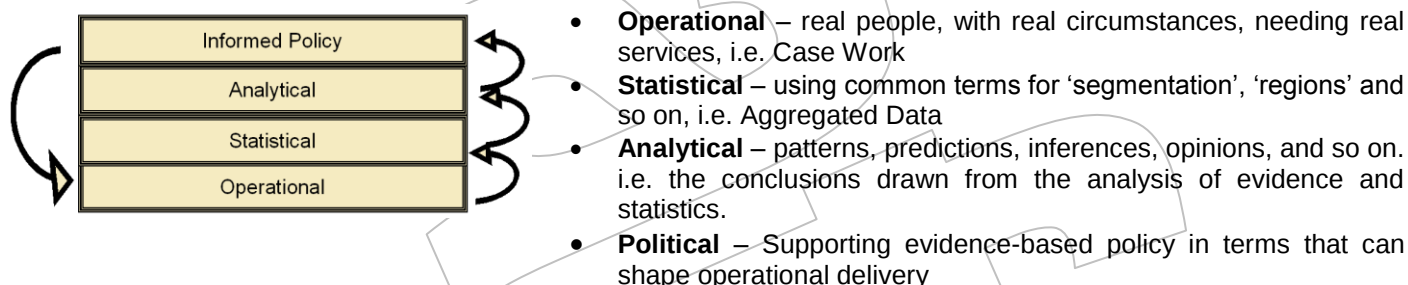
20. Characteristics of Documents and Data

An Information Architecture for the UK Public Sector

Documents	Data
Physical representation of a collection of information at a point in time, for example • word processing, spreadsheets • messages • web pages	Values and statements about subjects. Identifiers and Controlled Vocabularies
Contains statements (Structured, Semi-Structured, Unstructured).	Definitions of • entities • attributes • relationships
Discovered and evaluated via document level metadata, for example subject, format, creator, rights, quality, status	Meaning conveyed via • Logical Data models • Ontology
Search-able Share-able	Query-able Link-able

21. The Public Sector Information Architecture is designed to provide the means to describe both documents and data over the seven themes.

Using and Re-using Information across organisational boundaries
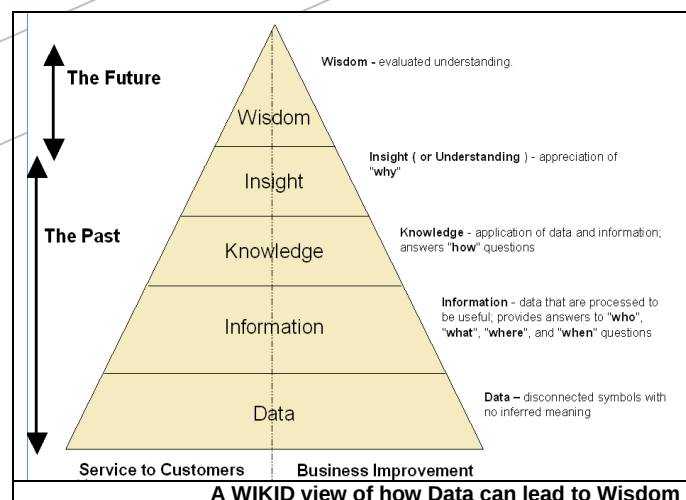
22. The standards and assets necessary to describe information across the seven themes, will be separately considered across four information contexts, such that each context can support and inform the next.

- Informed Policy
- Analytical
- Statistical
- Operational

- **Operational** – real people, with real circumstances, needing real services, i.e. Case Work
- **Statistical** – using common terms for 'segmentation', 'regions' and so on, i.e. Aggregated Data
- **Analytical** – patterns, predictions, inferences, opinions, and so on. i.e. the conclusions drawn from the analysis of evidence and statistics.
- **Political** – Supporting evidence-based policy in terms that can shape operational delivery

23. The public sector information architecture will provide the information standards to demonstrate that decisions are taken, based on conclusions that are reached, from statistics that are gathered, from cases that have occurred.

24. The four contexts above map well to the 'WIKID' view of deriving value from data.

25. The Information Architecture demonstrates how Knowledge Management is supported by effective Information Management.

A WIKID view of how Data can lead to Wisdom

Defining and Building the Information Architecture

Designs and Patterns

26. The Public Sector Information Architecture (PSIA) provides design guidance for

	PSIA Report	Status as at March 2011
Using the same identifier to refer to the same 'thing'.	Designing URI Sets for the UK Public Sector.	Version 2.0b
Communicating the definitions of terms used.	Defining Concepts and Classes for the UK Public Sector	Under Development
Data Modelling		Under Consideration
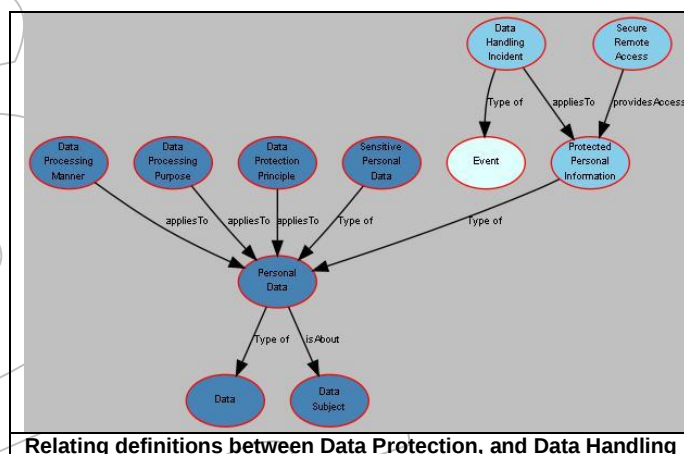
27. These designs are then the basis of patterns to describe information in each of the seven themes. Therefore the Information Architecture can be expressed as a series of connected Ontology, with URIs naming instances of specific information assets.

28. Where appropriate, this can support a 'Master Data' approach to managing key data sets that can be confidently reused across the public sector.

29. Information standards for a defined purpose or scenario may then be derived from these patterns and expressed using the most appropriate form e.g. an XML Schema, a SKOS Vocabulary, etc.

30. A potential example from the 'Governance and Assurance' theme illustrates how definitions from 'Data Protection' (dark blue), can be related to 'Data Handling' (light blue), and to universal definitions (white).

31. This ability to 'read-across' from various sources of definitions enables instances of information to be handled appropriately and effectively, by machine-readable links and associations.

Relating definitions between Data Protection, and Data Handling

The CTO Council working with Stakeholders

32. This document sets out actions to define and build the Information Architecture. For each theme, the Information Domain, of the CTO Council's 'Cross Government Enterprise Architecture' will identify and work with a set of stakeholders and programmes.

33. The architecture will define a set of standards and assets across the seven themes. Some will be developed and maintained centrally, whilst others will be devolved to representative groups of each sector (e.g. education, taxation or transport). Accordingly, The ICT Strategy identifies the following types of standard.

> The CTO Council will centrally manage only the standards that are required across a number of organisations and that are not specific to a particular business area (for example, education, taxation or transport). Accordingly, three types of standard have been identified:
>
> - **universal**: fundamental standards that are required by all public sector organisations (for example, XML)
> - **common**: standards used across multiple business domains (for example, champions), and
> - **local**: where responsibility is held by local domains/businesses/regions.

34. The governance of these standards and assets will ensure that each inherits from its parent, and that all stakeholders are properly engaged to ensure fitness for purpose, and re-use of existing standards and assets.

Defining standards for the 7 Themes

Semantics – the meaning of information

35. To pass information between sectors, each party will need to agree and communicate the definitions of the terms used, and use the same identifiers and vocabulary to refer to the same 'thing'. To share information with the public, those definitions and identifiers will need to be further associated with globally recognised sources, although these may be less authoritative and accurate. These 'Key Reference Data and Definitions' come together to underpin a 'Web of Linked Public Sector Data'

36. Some data entities commonly recur across many public sector information scenarios. The Public Sector Information Architecture will provide logical models and patterns to be incorporated into documents and messages, thus building on the existing 'Government Data Standards Catalogue' (GDSC).

	Pattern
Who?	Person
	Organisation
Where?	Location
	Address
What?	Subject
Why?	Condition
When?	Period

Syntax – the format of information

37. For information to pass between machines, it must be in a recognised format and structure. Information may be 'rendered' in a range of recognised formats that suit the requirements of a sector or information context. For example, the financial sector makes use of XBRL[2], and the Analytical context suits SDMX[3]. Where it is valuable, the same information can also be made available in formats that are not sector or context specific, for general re-use.

38. Where information is exchanged to support a tightly coupled process, a specific XML schema may be developed. Conversely, where information is published for general re-use, a 'Linked Data' approach may be more appropriate.

39. The Public Sector Information Architecture (PSIA) will build on the Standards and Architecture Framework (SAF), to recommend the formats in which information can be published, syndicated and exchanged, as determined by its purposes, contexts, and audiences.

40. Specific syntaxes will be published within the Government Data Standards Catalogue (GDSC) where the Public Sector Information Architecture proposes patterns for common data entities, e.g. Person.

Quality - the confidence to re-use information

41. The ability to express the quality of data assumes greater importance when data is to be shared. It is necessary to know whether data is fit for any proposed new purpose, and to be able to make judgements about which data to accept when there are contradictions from different sources. The standards to define data quality will differ for each of the four information contexts, and could be applied to a document, data set, or an individual assertion.

42. The provenance of data is of particular importance when inviting re-use beyond the public sector. Where such data is further combined or aggregated, that provenance data should remain firmly associated.

43. The Public Sector Information Architecture will publish a set of patterns, metadata and vocabulary to express data quality and provenance for each information context.

Rights - the right to use information for a purpose

---

[2] XBRL – Extensible Business Reporting Language - http://www.xbrl.org/Home/

[3] SDMX – Statistical Data and Metadata Exchange - http://sdmx.org/

44. The right to use information can be described in terms of intellectual property via licensing, and data protection via legal gateways and consent.

45. The Public Sector Information Architecture will publish a set of patterns, metadata and vocabulary to express information rights, and 'actionable' licenses.

46. Describing information in this way will define the purpose for which it has been collected, and the terms upon which it can be re-used.

Authentication - who is using information?

47. Authentication standards can be described in terms of the risk to a loss of confidentiality, and the measures to assure the identity of the person or device claiming information rights. A 'Trust Framework' defines the relationships between information users, identity providers, identity brokers, and information services.

48. The Public Sector Information Architecture will publish a set of patterns, metadata and vocabulary to define the terms used when expressing authentication requirements and solutions, and how they relate to each-other.

49. Describing information in this way will enable a potential re-user to judge if they are a person who has the necessary credential to state that they are empowered by their organisation to act in a role which has a defined purpose to use it.

Transport - how to move information

50. The standards by which information is transported can be described in terms of 'Availability', 'Integrity', and 'Confidentiality' requirements. The requirements of transporting information ranges from individual items to support closely coupled processes, through to bulk transfers of sensitive data.

51. The Public Sector Information Architecture will publish a set of patterns, metadata and vocabulary to define the terms used when expressing the characteristics of a network and how that applies to the nature of information to be transported..

52. Describing information in this way will enable a potential re-user to judge if they have access to an appropriate transport mechanism over which a piece of information may be moved.

Governance and Assurance - the behaviour and culture to protect and exploit information

53. Data controllers will need to be assured that organisations that re-use their data will continue to handle it appropriately. A series of Risk Assessment techniques and Information Governance accreditations have emerged within sectors, together with individual data sharing agreements.

54. The Public Sector Information Architecture will publish a set of patterns, metadata and vocabulary to define the terms used when expressing Information Governance and Information Assurance.

Managing the Public Sector Information Architecture
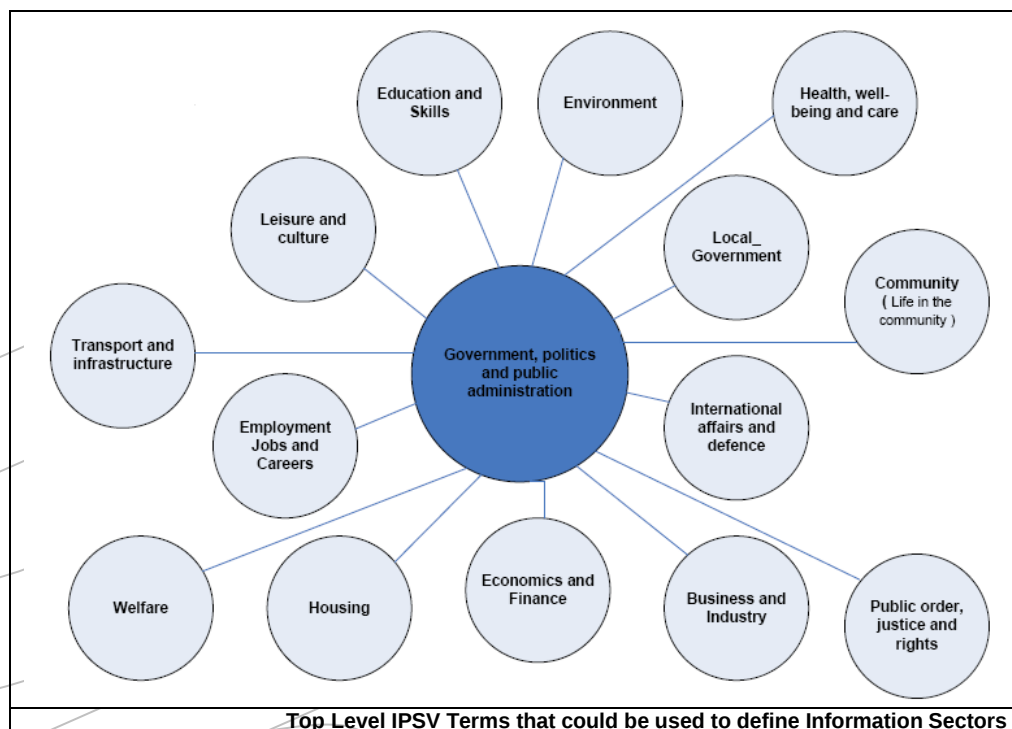
Federation of Standards – who defines what information?

55. The standards, methods, and content, that make up the Public Sector Information Architecture will require governance and management at
   - Universal
   - Core / Common
   - Local
   … levels; operating to consistent methods and quality assurance standards.

56. Taking the top-level of the 'Integrated Public Sector Vocabulary' (IPSV) provides an example of how information standards could be allocated over the sectors of the Public Sector.

57. A further governance layer needs to be defined to take account of jurisdictions.

58. Governing standards and information in sectors is a move away from departmental structures, to logical structure.

Top Level IPSV Terms that could be used to define Information Sectors

59. The Public Sector Information Architecture will complement an 'Open Standards Review Board' governing 'Universal' and 'Common' standards, with links into standards setting bodies for each sector.

60. In particular, this approach encourages sectors to define and govern standards for their stakeholders, and to make them available for use by other sectors, while inheriting common definitions from the centre.

61. A model infrastructure has been defined by the CTO Council's Information Domain to contain, manage and govern the standards that underpin the architecture. The components of the model infrastructure cover 'Registries', 'Directories', 'Repositories', and Tools.