e-Government Interoperability Framework (e-GIF)

Government Data Standards Catalogue

Volume 1 – General Principles

EXECUTIVE SUMMARY

The e-Government Interoperability Framework (e-GIF) mandates the adoption of XML and the development of XML schemas as the cornerstone of the government interoperability and integration strategy. A key element in the development of XML schemas is an agreed set of data standards. The Government Data Standards Catalogue sets out the rationale, approach and rules for setting and agreeing the set of Government Data Standards (GDS) to be used in the schemas and other interchange processes. It also contains the standards agreed to date. These standards are also recommended for data storage at the business level.

The Catalogue comprises 3 volumes:

Volume 1, this Volume, sets out the general principles, i.e. the rationale, approach and rules for setting standards

Volume 2 sets out the Data Types standards

Volume 3 sets out the Data Items standards

Contents

1   INTRODUCTION

The e-Government Interoperability Framework (e-GIF) mandates the adoption of XML and the development of XML schemas as the cornerstone of the government interoperability and integration strategy. A key element in the development of XML schemas is an agreed set of data standards, and these Government Data Standards (GDS) are to be used in these schemas and other interchange processes. They are also recommended for data storage at the business level.

Rationale

> The adoption of data standards for use across government will enable easier, more efficient exchanging and processing of data. It will also remove ambiguities and inconsistencies in the use of data. Inevitably the migration to these new standards may appear at the outset to be costly and time-consuming to some parts of government. However this burden should be outweighed by reduced development costs through the use of the agreed XML schemas that use these standards. It will also be easier and cheaper to use these standards from the outset in systems development rather than making changes during the life of the systems.

Objectives

> These standards apply to all systems that are mandated in the e-GIF and are for use in all other public sector interfaces. Compliance with these standards should follow the e-GIF Compliance rules.

Scope

> The Catalogue is used to record the agreed standards used in all XML Schemas, and other electronic interchanges of data involving the public sector, developed to support the e-GIF. The standards are defined at a logical (business) level and not at a physical database storage level, and do not document display properties. However it is recommended that they be used for specifying data storage at the business level.

Ownership

> The Office of the e-Envoy is the owner of the Catalogue, and as such is responsible for managing the change process, and for ensuring that the standards are used in all future developments.

Stakeholders

> The development and maintenance of this Catalogue is the joint responsibility of the Government Processes Group and the Government Schemas Group.

Control

> The Catalogue will be maintained using the UK GovTalk's Request for Comments (RFC) and Request for Proposal (RFP) change control procedures.

Verification

Because of differing legislative constraints it is important for some government organisations to know to what level data has been verified. Where these levels of verification occur they will be detailed in the specific standards. Other organisations that do not have requirements for all these levels will only use the lowest level that is appropriate to them.

Changes Forecast

The list of standards currently shown in Volumes 2 and 3 are not the definitive list of all government standards. Further ones will be added after they have been agreed. It is likely to be quite a while before this Catalogue can be said to contain the comprehensive list of all government standards.

The Catalogue will be held on UK GovTalk and the standards will be held there in a database with links to and from the XML schemas where they are used, and also to any external standard, e.g. W3C, on which the government standard is based.

The standards may be extended in due course to cover clerical and display formats.

2 PRINCIPLES

2.1 Precedence of Data Standards used in the Catalogue

When a standard is required for a piece of data it must be recognised that standards may already exist and that they should be used whenever possible. Standards bodies exist at several levels e.g. International, EU, British, and also at several committee levels, e.g. W3C, OECD. Also some standards deal with interchange, some with storage and some with display characteristics, or any combination of these.

Determining which is the right standard to adopt depends on a full understanding of the rationale behind the standard and the situation in which it was developed. It will therefore be necessary when considering whether to use an existing standard, to understand the community in which that standard has been agreed, its date of agreement and its relevance to the piece of data under consideration.

As a general rule of thumb the policy is to use International standards if they exist, but only if they are appropriate and up to date.

2.2 Naming Conventions

A full business name will be assigned to each data standard using the following format:

OBJECT-QUALIFIER(S)-DESIGNATOR

where:

- OBJECT is a keyword that describes the main object/entity/concept to which the standard relates, eg. TAXPAYER, CLAIMANT, AGENT. This subject word is omitted where the name of the standard is otherwise sufficiently clear as to render a prefix superfluous;

- QUALIFIER(S) is a qualifying word(s) used to describe the standard uniquely. Each word used is itself meaningful to the standard being described. The order of the words is in decreasing order of importance from left to right;

- DESIGNATOR is a keyword that designates the class or category of data to which the standard belongs, eg. name, number, code etc. This designator word is at the end of names that are made up of several words.

e.g. TAXPAYER-REFERENCE-NUMBER

2.2.1 Standard Designators

The following standard designators are used in this Catalogue:

CODE   Data, maintained by the user, which readily identifies an occurrence of an object class. There will normally be a screen dialogue or batch process to enable the user to input changes.

DESCRIPTION   Alphanumeric data that describes a specific item.

NUMBER	Alphanumeric data that is used for identification purposes, e.g. a reference number.
QUANTITY	An expression of the number of units of anything except money.
RATE	An expression of a proportion or measure of value, eg. a percentage.
TEXT	Alphanumeric data in free format that is not descriptive.
TYPE	A piece of data that categorises an object e.g. TAXPAYER TYPE.

Any of the Data Types (see Section 2.4.1) may be used as standard designators e.g. TIME, DATE

2.2.2 Use of Acronyms within Data Standard Naming

The Rules set out below are to be used when constructing business or data names.

- Avoid the use of abbreviations/acronyms wherever possible as they impede readability.

- If it is necessary to use an abbreviation or acronym because of space limitations refer to Appendix A which contains a list of those abbreviations/acronyms used to date in this Catalogue. It can be seen from these that there should be only one permissible abbreviation for each word/phrase, but it is possible for an abbreviation to have more than one meaning. Should this lead to confusion in any particular situation then the words/phrases should be explained in full.

- If the word, term or name required to be shortened appears in Appendix A then the abbreviation shown beside it must be used. If it does not then follow the guidelines below in constructing your own:

  ❖ only words longer than 8 characters should need to be abbreviated

  ❖ use commonly used English abbreviations in preference to business ones

  ❖ use business abbreviations in preference to computer terms, which are in turn preferable to made-up ones

  ❖ if a common abbreviation is not available then make one up, but it must be meaningful and unambiguous

  ❖ abbreviations should generally appear in upper case

- If it is necessary to create a new abbreviation then first start by removing the vowels from the word, and then if necessary progressively remove consonants from the middle of the word.

2.3.　Context of Data Standards

The types of standards addressed by this Catalogue are at the Business level and not the Physical level, i.e. they do not cover how the data is actually stored within computer systems.

2.4　Scope of Data Standards

The standards found in this Catalogue are Customer Focused and not internal to any specific government organisational unit (e.g. Personnel Number, Cost Centre)

The standards adopted are set at two levels:

2.4.1　Data Type　(see Volume 2)

These include those generic items of data that can be associated with a variety of different entities/objects, e.g. date, name, address, telephone number.

2.4.2　Data Item　(see Volume 3)

These include the more specific occurrences of data that are not covered by the above-mentioned data types, e.g. National Insurance Number, VAT Number, Driving Licence Number.

There will on occasion be a need for a data type to be taken down to a more significant specialisation, e.g. Date - Death Date, where extra validation is required for that specific occurrence.

It is not the intention to overly complicate this Catalogue by defining too much at the specialisation level, i.e. the data item level. It is intended to be used for the most significant data items that have an importance across Government in general.

2.5　Change Control

Changes to this Catalogue will be managed using the UK GovTalk RFC and RFP processes. For more details see www.govtalk.gov.uk

The identification of new standards for inclusion in the Catalogue will generally emerge through the work of the Government Processes Group and the Government Schemas Group. But the opportunity exists for anyone to propose a new standard, or a change to an existing one, through the RFP process on UK GovTalk. The Office of the e-Envoy will manage these requests and feed them into the work of these two groups.

3 DATA STANDARD TEMPLATE

Each data standard will be documented using the following template:

Name: *The full name of the Data Type/Data Item (in accordance with the Data Naming Standards as defined in Section 2.2).*

Definition: *A simple but unambiguous definition of the Data Type or Item*

Business Format: *The required format of the data from the business perspective. This will include the minimum and maximum number of characters if appropriate, and the structure of the data type/item, e.g. for National Insurance Number the structure is AANNNNNNA where A is an alpha character and N is a numeric character. Where Alpha is specified it refers to the full character set available through a standard UK keyboard excluding numerics.*

XML Schema ID: *The identifier of the XML schema where the data standard is used. It is expected that a standard will only be used in one schema and all government schemas will be held on UK GovTalk. The XML schema will show the pattern, i.e. the size and mask, of the standard.*

Validation: *Generic for Types and specific for Items. The validation rules to be applied for acceptance of data (e.g. first alpha character must be A, B or C).*

Values: *List of the acceptable values (e.g. Male, Female)*

Default Value: *For any list of values, the default value to be used unless otherwise stated*

Owner: *Name(s) of those Departments/Agencies/Other Bodies who own this standard*

Based on: *Derivation of standard (e.g. BACS, ISO, BSI, BSEN, W3C)*

Verification: *Steps taken to establish the correctness of the Data Type or Item. (For some Departments, the stringency with which data collected has been checked needs to be known prior to making use of that data – e.g. has date of birth been verified by checking Certificate of Baptism or Birth Certificate. These different levels of verification will be detailed here. Other Departments who do not have the same requirements for these levels will use only the lowest level when passing information).*

Comments: *Additional notes*

Version :   *The version number of this standard*

Date: *The date this version was agreed as a Government Data Standard*

APPENDIX A   ABBREVIATIONS/ACRONYMS

Detailed below are the abbreviations and acronyms used within this Catalogue:

Banks Automated Clearing System	BACS
British Standard European Norm	BSEN
British Standards Institute	BSI
British Telecom	BT
Crown Court Criminal Justice System	CCCJS
Customs & Excise	HMCE
Department	DEPT
Department of Social Security	DSS
European Union	EU
Employment Service	ES
Government	GOVT
Government Data Standard(s)	GDS
Internet Engineering Task Force	IETF
Inland Revenue	IR
International Organisation for Standardisation	ISO
National Insurance Number	NINO
Not Applicable	N/A
Number	NO
Office of e-Envoy	OeE
Office for National Statistics	ONS
Organisation for Economic Co-operation and Development	OECD
Request for Comment	RFC
Request for Proposal	RFP

Standardised Industrial Classification	SIC
Subscriber Trunk Dialling	STD
To be determined	Tbd
Telephone	TEL
United Kingdom	UK
United Kingdom Passport Agency	UKPA
Value Added Tax	VAT
World Wide Web Consortium	W3C
Extensible Markup Language	XML