



Metadata and Data Standards for e-Governance Applications

-Approach Paper-

Government of India
Ministry of Communications and Information Technology
Department of Information Technology
National Informatics Centre

<http://egovstandards.gov.in>

Metadata and Data Standards for e-Governance Applications

Prepared by: Dr.(Ms.)Meenakshi Mahajan

Reviewed by: Deputy Director General NIC

Approved by: Director General

**No part of this document shall be reproduced without prior permission of
Director General, National Informatics Centre**

Table of Contents

Amendment Log	4
1. Introduction	5
2. Background Information	5
3. Benefits of Metadata	6
4. Terms of Reference	7
5. Issues related with Metadata and Data Standards.....	7
5.1 DATA STANDARDS	7
5.2 METADATA STANDARDS	8
6. Approved Standards at National & International Level	9
7. Conclusion.....	10

Amendment Log

Version No.	Date	Change Number	Brief Description	Sections Changed
0.1	19 -12-05	NA	First Version	All

1. Introduction

With the exponential growth of internet, the web has become the largest information space, the world has ever known. One of the reasons for the rapid growth of web are the standard information architectures adhered to by web developers and the language, HTML, within which structures are expressed is overly simple, straightforward to learn and rapid to implement. But at the same time, it is not scalable for the further long term development of networked information. To have more accurate retrieval of networked information and better management of digital objects, rigorous description and construction of data is required—the key to this is *Metadata*.

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information. Whether or not particular information functions as data or metadata is a matter of context or perspective--- what is metadata to one person or application can be data to another.

2. Background Information

Government possesses one of the largest repositories of information products and services. Public access to these services is uncoordinated, cumbersome, complicated, slow, paper-based and most often very confusing because majority of government services involves the collaborative effort of several related departments and various tiers of government. As a result, the user has to go from department to department for availing different services. To help citizens

avail all kinds of services at Single Window Facilitation Centres, the key enabler is the e-Governance Interoperability Framework (eGIF).

The e-GIF defines the technical policies and specifications, governing information flows across government and the public sector. It is primarily supported by the successful communication between two or more information systems. Successful communication depends on mutually agreeable syntax, semantics and compatible communication mechanism.

Semantic interoperability requires that the precise meaning of exchanged information is understandable across applications. The communicating parties/information systems are required to avoid the use of different terms for similar concepts, or use similar terms to mean different things. For this, they have to follow agreed upon standards for data and metadata. *Semantic interoperability consists of two sets of standards – one for data and other for metadata.*

3. Benefits of Metadata

The fact that Metadata facilitates interoperability of e-Governance applications, is endorsed by the benefits which metadata provides:

- i. Helps people find information
- ii. Helps identify content
- iii. Matches User and Content
- iv. Exploits Government Information

4. Terms of Reference

- (i) To identify which resources need Metadata
- (ii) To identify which data elements (mandatory as well as non-mandatory) to use
- (iii) To identify which encoding schemes to be used—existing ones or new ones
- (iv) To device a mechanism for common codes of data for Legacy Applications
- (v) To define integrated data, metadata standards and related operational specifications—arrive at common codes of data for computerization in Government.
- (vi) To develop a metadata framework and devise metadata standards
- (vii) Forms re-engineering and Change Management needed in connection with the data standards

5. Issues related with Metadata and Data Standards

5.1 Data Standards

Achieving data coherence across government means that government organisations need data schemas that have been agreed upon for use throughout the government. As the government databases have evolved in a largely decentralised manner, it is required to have Government Data Standards

Catalogue; it is a data dictionary to be operated across Government that holds information about data items widely used across Government.

The national data dictionary would be produced in several stages, beginning with a section covering core data elements that run across all government departments (e.g., name, address, etc.), and a section dedicated to each functional area of government activities (such as rural development, agriculture, etc) that covers the data elements unique to that area.

The Data standards may be expressed as a structured list of data items with descriptions, field lengths and formats, and UML models.

An agreed set of data standards is a key element in the development of interoperable applications. Adoption of data standards for use across government will enable easier, more efficient exchange and processing of data. It will also remove ambiguities and inconsistencies in the use of data.

5.2 Metadata Standards

In order to evolve the Metadata standards, it is first necessary to understand the Metadata Schemes. These are the sets of metadata elements designed for a specific purpose such as describing a particular type of Information source. The definition or meaning of the elements themselves is known as the semantics of the scheme.

Metadata standards will cover the core set of 'elements' that contain data needed for the effective labeling of the resource, so that retrieval and management of the resource across all of governments' information systems become easier.

Metadata standards provide a way for information resources in electronic form to communicate their existence and their nature to other electronic applications (e.g. via HTML or XML) or search tools and to permit exchange of information between applications.

6. Approved Standards at National & International Level

Many different metadata schemes are being developed in a variety of user environments and disciplines. Some of the most common ones are as follows:

- (i) Dublin Core
 - simple, concise and used to develop Web based documents and describe the Web resources)
- (ii) The Text Encoding Initiative (TEI)
 - (Guidelines for marking up electronic texts such as novels, plays, poetry and primarily to support research in the humanities)
- (iii) Metadata Encoding and Transmission Standard (METS)
 - (Guidelines for describing complex digital library objects)

and many more

Many countries in Europe and around the world for e.g. the UK, Ireland, Denmark, Finland, Australia, New Zealand, and Canada have adopted the Dublin Core (<http://dublincore.org/>) approach as part of their e-Governance framework, or are in the process of doing so, whereas the US government and a few other countries are following GILS (<http://www.gils.net/>) approach.

In India, the Andhra Pradesh Government has initiated the projects for Common Coding Structure for Computerization in Government. Also it is working for Metadata Standards and Forms reengineering.

7. Conclusion

With reference to Indian scenario, where the information is accessed by the common man at the central, state, district, block and at the Panchayat level, it becomes all the more necessary to define standards for metadata and data to achieve the eGIF.

The approach towards developing these standards should be based on the following premises:

- i. A detailed view of Legacy applications
- ii. Identification of Data Elements-Mandatory and Non Mandatory
- iii. Metadata for Content Management
- iv. a hierarchy of standards ensuring that international standards are used wherever possible.
- v. National standards could be used to cover items specific to the country, and new standards are developed where none exists.
- vi. A minimum set of standards that are relevant for system interconnectivity and information exchange should be selected. These standards should be open in the sense that they are publicly available and well documented.
- vii. Moreover, standards having wide market support should be chosen to reduce cost and risk.