5. Content Management Framework

State Portal provides information and transactional government services to citizens, Businesses and overseas people,) based on the content contributed by various government entities and third parties. These sources contribute content in different forms and formats, structured or unstructured, with or without metadata using different taxonomies and vocabularies. State Portal need to organize content and deliver information to end users as per their expectations.

Content undergoes different phases in its lifecycle such as authoring, publishing, archival etc. Each of the phases of the content life cycle need to be supported by Policies, Tools, processes and/or procedures. State Portal should define content policies and processes on the lines of Content Management Policies described in "Content Framework for the National Portal of India" document (http://india.gov.in/cfw) from National Portal Secretariat. Suitable content management policies and publishing workflows should also be defined.

State Portals should take care of "Content Formats", "Content Sources" and "Content Contracts" before they start authoring content. A brief description of these and content life cycle phases are provided in the following sections.

To make content sharable, easily discoverable, reusable, and to maintain the currency of content, taxonomies should be used and metadata should be associated.

To manage set of activities during the content life cycle, State Portals should essentially make use of Content Management System (CMS). Expected functionality of a CMS is detailed in the following sections. CMS apart from the functionalities described under section "Content Management System" should also support all functionalities detailed under different sub sections of this section.

5.1. Content Taxonomy

State Portal needs to provide intuitive navigation and easy discovery of information, which require better means of resource description. Content taxonomies are good methodologies to achieve better resource description. Taxonomies provide a means for categorizing or classifying information within a reasonably well-defined associative structure, in which each term in taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.

Content displayed on State Portal includes data from relational data, documents, digital assets, XML, web pages, web services, discussion groups, etc. By tagging such resources with relevant terms from the taxonomy, we enable search and retrieval of those information assets, thereby, making users to reach the content they need in a fast manner.

The purposes of content taxonomies are to create content, classify content and discovery content so that right information is found on the portal at the right time. Personalize content delivery also requires content taxonomy.

Taxonomies in State Portal will exist at least at two levels namely Portal Taxonomy and Departmental Taxonomy. CMS should support managing multiple taxonomies and should be able to normalize different taxonomies to portal taxonomy.

Separate taxonomies should be maintained for each type of content for example content repository for application forms, Content repository for schemes, content repository for policies, etc.

Taxonomies for services, forms, document for each department could be different. Taxonomy management should be a part of CMS

5.2. Content Metadata

Any resource used as a part of State Portal should have metadata. This metadata can be based on the current elements of the Dublin Core as defined in the Dublin Core Metadata Element Set, Version 1.1: Reference Description [DC], and contain definitions for the following properties:

	Element	Description
1	Title	A name given to the resource
2	Creator	An entity primarily responsible for making the content of the resource
3	Subject	The topic of the content of the resource
4	Description	An account of the content of the resource
5	Publisher	An entity responsible for making the resource available
6	Contributor	An entity responsible for making contributions to the content of the resource
7	Date	A date associated with an event in the life cycle of the resource
8	Type	The nature or genre of the content of the resource
9	Format	The physical or digital manifestation of the resource
10	Identifier	An unambiguous reference to the resource within a given context
11	Source	A reference to a resource from which the present resource is derived
12	Language	A language of the intellectual content of the resource
13	Relation	A reference to a related resource
14	Coverage	The extent or scope of the content of the resource
15	Rights	Information about rights held in and over the resource

Information using the Dublin Core elements may be represented in any suitable language (e.g., in HTML meta elements). However, RDF is an ideal representation for Dublin Core information.

Metadata is classified into mandatory, optional and extended. All content eligible for publishing and in the content repository should have mandatory metadata. Optional metadata enhances the discoverability of content. Extended metadata can be used by State Portals to manage content within the context of the state. Metadata should be used for both structured as well as un-structured data.

Every resource (content) in the repository whether published or un-published should have associated metadata and this metadata should be complete in terms of the mandatory elements. Content repository should contain the content along with its metadata file.

Metadata capture should start from the content source itself. Workflow (authoring, publishing and Archival) should be integrated with metadata management so that required metadata can be captured along with the workflow.

Metadata should make use of terms defined in the taxonomy only so that restricted vocabularies are used.

Metadata should be managed using XML. Metadata and taxonomies should be in Unicode.

Once the content is classified and organized, content that can be managed using XML should be made structured. Government Services will have contact details. Instead of embedding the contact details with the services description content resource, contact details can be handled as a structured data. This also avoids putting the entire content on workflow if there is a change only in the contact details. Presentation can compose information by stitching the unstructured data "service description" and structured data "Contact details" which is in XML format. "Contact directory" should be XML based and be treated as structured content.

Metadata should be effectively used to manage content life cycle also by including extended metadata elements like "expiry date", "next review due on", etc.

5.3. Content Formats

Based on the State Portal taxonomy, state portals should establish formats for each part of the content displayed on the portal. Content formats that can be considered are textual content, graphical content, applications, widgets, audio & video content, tickers, tabular format, downloadable documents. Based on the assessment of the user needs & concerns and constraints on the content, each section or group of content at a minimum should have a primary format and a secondary format. In essence, each content component should have alternate formats for delivery. Primary format will be first choice for content delivery where as secondary format can be used when data in primary format cannot be made available. Also State Portals should establish acceptable file formats for each content format.

Content can have differences between formats at source, storage format, delivery formats. At a minimum, following issues should be taken in consideration when deciding formats

Content delivery formats
a. Minimum user interface capabilities
b. Size of the content (in Kbs) vis-à-vis the general bandwidth availability of users
c. Usability of the content –readability, printability
d. Type of content – static, relatively static or dynamic

Content Storage formats
a. Reduce the size in storage (in Kbs)
b. Transformability of content between storage and delivery formats
c. Enhanced discoverability of the content by external entities like national portal, other State Portals, department website, etc

Illustrative formats are given below:

Example Content	Source format	Storage format	Primary Delivery format	Secondary delivery format
Budget Speech by railway minister	Audio	Audio	Audio	Speech transcript
Population growth trend	Tabular or records	records in database	Graphical	Tabular
Social welfare scheme details	Text	Text	Text	Downloadable document
Local Weather	Content feed	No need to store	Same as source format	Text

Example Content	Source format	Storage format	Primary Delivery format	Secondary delivery format
Application for a drivers license	Form	Form	Fill able forms	Downloadable document

The content formats that are described in the National Portal content framework document should be used as guidelines.

5.4. Content Sources

It may not be practical for State Portal team to author entire content of the entire State Portal. Hence it is very essential to identify the sources of content within and outside the state government. Information architecture which clearly articulates the organization and structure of State Portal should be made use of to identify the content sources. Identifying the content sources is also essential from the point of view of ascertaining the content ownership so that content delivered to end users is accurate and is in the right context. Content origin should have the content ownership

Illustrative content and their sources are given below:

Type of Content	Source of Content
Agricultural Commodities prices	Marketing Research and Information Network, Directorate of Marketing & Inspection, Department of Agriculture & Cooperation
Social welfare scheme details	State department of Social Welfare
Application for a drivers license	RTO, State Transport Department
State Education Board Examination results	State Education Boards
Handbook of state for the year	State Gazetteer Department

5.4.1. Content Contracts

From the long term sustenance perspective content contracts will be established by the content provider and content consumer to ensure the currency and authenticity of the content. Establishing content contracts will go a long way as private establishments may solicit content from State Portal and third party content can be solicited by State Portal. Further these contracts can be enforced using available technologies like service contracts and access control mechanisms. At a minimum, content quality assurance including content metadata should be part of such contracts. This will also resolve the issues of content ownership. Content contracts should include the modalities of content exchange including format, frequency, violations.

5.5. Content Management System

Content management systems should have the following (but not limited to) features or capabilities:
a. Content storage or repository management
b. Content authoring and publishing
c. Content delivery
d. Content exit and archival
e. Content repurposing for different audience and for different interfaces
f. email notifications for automated content edits and reviews
g. Native content conversion to web formats (States to Specify)
h. Provide searching and retrieving of content based on various criteria
i. Browser based interface

j.   Support for setting up development, staging and production environment setup in the platform

5.5.1. Content Authoring

Content authoring refers to the process of creating content for State Portals. Content can be created by fetching content from the different available sources (government entities, third party, etc. Content authoring is done through an authoring tool.

Authoring tools should support the creative effort of authors by providing templates, images, layouts, etc. Content authors should be able to create a unique content resource either by linking and uploading documents (in PDF) or keying in text to create content resource. Should support creation of templates for content creation and enable authors to select input content from different sources.

Tools should have built in agents to pull content from different sources including content feeds and present the same to authors in required formats. For this purpose, authors should be able to configure content resources and the frequency at which the content from the resource should be pulled. Authors should also be able to specify template for a given type of content.

Authoring tools should be integrated with the content repositories to enable authors to choose from available content and create new content. Authoring tool should be able to produce output (content resources) in different formats namely HTML, XML, PDF. Authoring tools should support spell check and language correction options.

Authoring tools should be integrated with metadata management tools so that content authors can attach metadata to the content they author.

Content creators need ways to provide alternate versions of media resources for use by different kinds of device and delivery context, while maintaining the same information semantics.

Content feeds should be handled using XML and RSS technologies

5.5.2. Content Publishing

Authored content will be staged in the CMS before it is released for delivery. Authored content will be routed through various stages (modification, approval, moderation, and publishing) of workflow and finally gets published into content repository. CMS should support creating and configuring workflows based on categories of content. Metadata of the content may get modified through the workflow. Content that is approved to be published should be released into content repository only after making sure that all the mandatory metadata is created for a piece of a content.

Content publishing of State Portal should follow the guidelines specified in "Content Framework for the National Portal of India" prepared by National Portal Team

5.5.3. Content Delivery

The content used on State Portal includes text, graphics, and images, audio and video resources. Content delivery refers to aggregating content in a structured format from content repositories and displaying the same to the end user. Content delivery component of CMS should have capabilities to pick up content as available in the content repositories, structure the information for presentation and associate default style and display the information for usage.

_____

Content delivery should have specific templates for each component of content categories. It should have repository of styles and should be able to create and modify styles.

It should manage content validity and presentation. It should be able to set up "content review, exit and archival" policies in the content delivery component of CMS. It should have notification capabilities so that notifications and alerts can be set up. Content delivery should send notifications /reminders/ alerts based on the policies that are set up.

Please refer the relevant sections of "Content Framework for the National Portal"

5.5.4. Content Storage Management

State Portal handles both structured and un-structured content. Un-structured content includes web pages, documents, images, audio files, video files, RSS feeds, etc. Structured content should be stored in a Relational Database Management System (RDBMS) where as un-structured content should be managed by a content repository.

CMS should have content repositories to handle unstructured content. Content repositories should be flexible enough so that content delivery can work on multiple repositories which are using different taxonomies to deliver information as required. Content repositories should also store templates, formats, etc.

CMS should be able to store content for delivery across multiple devices (channels)

5.5.5. Content Exit and Archival

Information delivered on the State Portal would lose relevance beyond the validity period. Content which is found non relevant should be archived. Validity period may vary across content categories. Archived content will be stored in a repository before the same can be permanently deleted. Metadata associated with the content should be effectively used to manage content validity and further archival.

State Portals should have content archival policies on the lines of guidelines detailed in "Content Framework for the National Portal "

_____