Metadata Extraction and Register for Enterprise Information Architecture in the Brazilian House of Representatives

Mariana Baptista Brandt
Câmara dos Deputados, Brazil
mariana.brandt@camara.leg.br

Keywords: enterprise information architecture; metadata register; metadata identification; House of Representatives (Brazil).

Abstract

This paper presents part of the Enterprise Information Architecture of the Brazilian House of Representatives, which aims to model the information of the strategic business processes and integrate it to their information systems. This procedure intends to be part of the institution's Enterprise Architecture.

The extraction of business metadata is one of the most important parts of the information modeling - the Information Architecture methodology used by the institution. It starts by analyzing the business process using process mapping and modeling with Business Process Management methodology. The information analyst joins the business area and the process management teams in the activity of business mapping and modeling so as to get to know the business and identify the information produced and consumed during the business process.

The information modeling team follows the business mapping and modeling meetings. Depending on the business process, the information modeling team can follow more steps of the process management team to understand the business information. During the business mapping, the team identifies the procedures, activities, information flows and documents needed to the business execution. The business mapping diagram produced is used by the information analyst, who identifies, from each activity, where there is important information input or output. This creates the document “Business Process Information Map”, that shows the information and documents used and produced in the business process for business acknowledge and metadata extraction. The process modeling implements adjusts and improvements for the efficacy and efficiency of the business. After it’s done, the information modeling team check if there are any changes in the business information and update it if necessary.

The next step consists in the metadata identification from the documents and information of the process analysis. The business area which is responsible for the process provides copies of all the documents identified in the process to the information modeling team, who analyses the documents for metadata extraction. The documents can be from any kind: manuals, reports, IT systems screens, checklists, administrative processes, orientation guides, etc. The metadata extracted from these documents are registered in a repository. For each metadata, are also registered its attributes: description, data steward, access mean, standard entry, format rule, responsible for first entry in the system, business rule, access level and if it’s part of an open data dataset. These attributes came from the analysis of the information policies of the institution (Information Content Management Policy, Digital Preservation Policy, Index Policy).

Publishing Policy,\textsuperscript{4} Information Security Policy,\textsuperscript{5} Freedom of information Law\textsuperscript{6} and it’s regulation\textsuperscript{7}).

These attributes bring important information about the metadata and can be about the metadata itself or resulting of the relationship between the metadata and the business process. It means that for a determined business process, the metadata can have, for example, an access level different from another business process, among other differences related to the business process. The metadata attributes are: description, data steward and format rule. These are the attributes that identifies the metadata as unique and can’t be altered according to the process. If there is a need of changing in these attributes, it should be analyzed if there’s a need of creating a new metadata. The other attributes are related to the business process in which the metadata is in.

After identifying the business metadata and representing it with all its attributes, the business area is required to validate it. The information modeling team provides orientations about how the validation must be done and offers help to this task if necessary.

When the metadata are validated they are used for other activities of the enterprise information architecture, such as: information governance, information retrieval requirements, management information needs report and the information architecture diagram. The last one is a diagram that includes metadata, data stewards and information technology systems that are part of the business process. This documentation, in addition to the terminological part developed in parallel by another team (business glossary, taxonomies and thesauri) makes what is called Information Architecture Model. This model must be a guideline to the development of IT solutions and information management tools for the business process.

The register of the metadata and its attributes in a repository makes possible to have a general vision of the institution’s information. Besides, it promotes the governance because it shows who the data stewards are. The metadata reuse in more than one business process is another possibility provided by the registration in a repository. The data steward has the authority to decide about the information content characteristics and attributes and its business related metadata.

The metadata mapping and its reuse in different business processes and information systems allows improvements in the information management and information quality, because it avoids non controlled redundancy and inconsistencies. It also highlights the data steward, who must warrant information authenticity, integrity, accuracy and security, and who will be accountable when one of these criteria is not observed.

The enterprise information architecture aims to organize and integrate the business processes information to its IT systems, contributing to the institution’s enterprise architecture. The metadata are its more representative element and provide the improvement of information access and information quality.

\textsuperscript{5} Ato da mesa 47/2012. Available at: <http://www2.camara.leg.br/legin/int/atomes/2012/atodamaesa-47-16-julho-2012-773827-norma-cd.html>.
\textsuperscript{6} Available at: <http://www2.camara.leg.br/legin/fed/lei/2011/lei-12527-18-novembro-2011-611802-norma-pl.html>
\textsuperscript{7} Ato da mesa 45/2012 Available at: <http://www2.camara.leg.br/legin/int/atomes/2012/atodamaesa-45-16-julho-2012-773823-norma-cd.html>.