

Plataforma M51-CLAV: o que há de novo?

Alexandra Lourenço

Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB),
alexandra.lourenco@dglab.gov.pt

José Carlos Ramalho

Universidade do Minho, jcr@di.uminho.pt

Maria Rita Gago

Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB),
m-rita.gago@dglab.gov.pt

Pedro Penteado

Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB),
pedro.penteado@dglab.gov.pt

Palavras-chave: Avaliação suprainstitucional, classificação funcional, Lista Consolidada, Macroestrutura Funcional; ontologia; plataforma modular de classificação e avaliação.

1. Introdução e Método

A presente comunicação tem como objetivo a apresentação do “Projeto M51-CLAV – Arquivo digital: Plataforma modular de classificação e avaliação da informação pública”, desenvolvido pela Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB), em parceria com a Universidade do Minho, no âmbito do Aviso N.º 02/SAMA2020/2016, para dar cumprimento à Medida 51 do Simplex + “Arquivo digital” (cf. <https://www.simplex.gov.pt/>). Pretende-se apresentar os primeiros resultados, principalmente os relativos às duas etapas iniciais: Levantamento de requisitos de requisitos funcionais e não funcionais e Modelação e análise de informação.

A plataforma visa incrementar a utilização pelas entidades da Administração Pública (AP) de um referencial comum para a classificação e avaliação da informação, que permita o desenvolvimento de instrumentos de natureza transversal a aplicar em contexto organizacional (Plano de classificação / Tabela de seleção). O Projeto M51-CLAV, enquanto solução para os problemas de estruturação da informação pública, integra dois componentes:

1) A estrutura conceptual para a organização e avaliação da informação pública - baseada no referencial Lista Consolidada para a classificação e avaliação da informação pública (LC), disponibilizada em agosto em http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2017/08/Lista-Consolidada_Versao-1.1_Divulgacao.xls.

A LC é uma estrutura hierárquica de classes que representam as funções, subfunções e processos de negócio executados pela Administração Pública, contemplando a sua descrição e avaliação.

Este referencial é o culminar de um caminho que aglutina vários projetos:

- Projeto MEF “Macroestrutura Funcional”;
- Projeto “Harmonização de classes de 3.º nível em planos de classificação conformes à MEF”;
- Projeto ASIA “Avaliação Suprainstitucional da Informação Arquivística”.

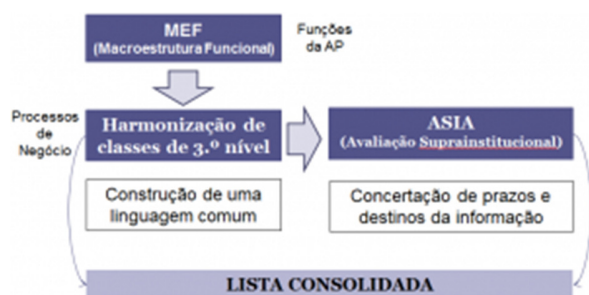
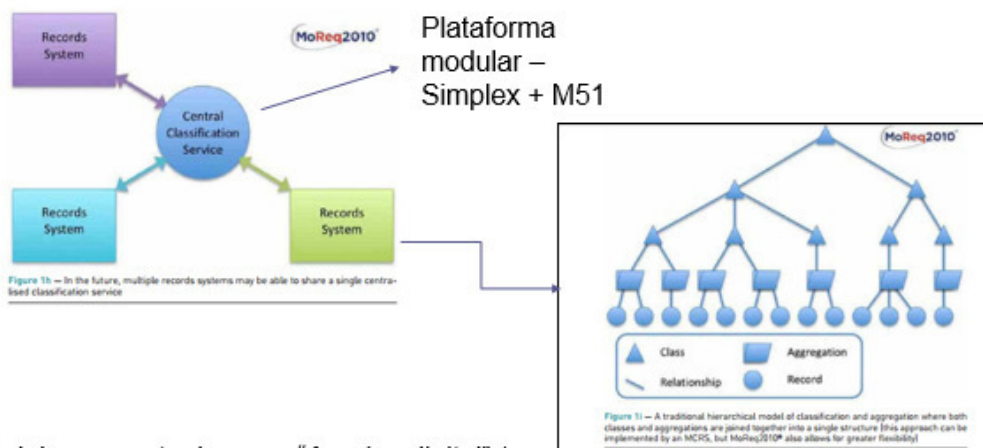


Figura 1 – Projetos que originaram a Lista Consolidada

2) A estrutura tecnológica que permite centralmente gerir e disponibilizar a estrutura conceptual, de acordo com o MoReq2010, agregando logicamente a informação produzida pelas diversas entidades públicas, bem como desmaterializar o serviço de avaliação da informação pública e o serviço de controlo de eliminação de informação arquivística pública (por exemplo, a submissão de propostas de tabela de seleção e o envio de autos de eliminação, para apreciação da DGLAB).

Esta estrutura tecnológica terá capacidade de integração com sistemas de gestão de documentos (*record systems*) ou com sistemas orientados ao negócio (*business systems*), fornecendo e recebendo componentes relativos à classificação e à

avaliação. Possibilita que as organizações apliquem esquemas hierárquicos normalizados para a organização da informação, garantindo a interoperabilidade semântica, ao mesmo tempo que assegura a construção de uma memória de Estado.



Modelo conceptual para o “Arquivo digital” /
Plataforma modular – componente A
(adaptada do Moreq2010)

Figura 2 – A plataforma M51-CLAV como serviço central para a classificação e avaliação

A plataforma, ao disponibilizar novos produtos e serviços e ao simplificar e desmaterializar procedimentos, vem alterar substancialmente a situação existente, abaixo descrita, substituindo-a por um novo mecanismo de interação.

O “Arquivo Digital: Plataforma modular para a classificação e a avaliação da informação pública” deverá disponibilizar:

- Produtos e serviços e orientados à Administração Pública, atualmente inexistentes, como:
 - Referencial Lista Consolidada para a organização e avaliação da informação pública (ontologia), contemplando a representação das funções do Estado e dos processos de negócio da AP, das relações entre estes, com identificação das entidades responsáveis pela preservação da informação e do ciclo de vida desses processos;
 - Tabelas de seleção, derivadas da LC, com o ciclo de vida da informação associada aos processos de negócios (prazos de conservação / retenção administrativa e destino final da informação pública) para implementação em contexto organizacional ou pluriorganizacional;
- Desmaterialização e simplificação dos serviços de:

- Avaliação da informação produzida pela AP: atividades de recolha de dados, submissão, análise e decisão sobre as propostas, procedimento atualmente sujeito a uma complexa tramitação.
- Controlo de eliminação de documentos, através de submissão de autos diretamente na plataforma. Esta submissão é obrigatória para todas as entidades públicas que pretendam eliminar documentos / informação produzidos, de acordo com o Decreto-Lei n.º 447/88 e portarias de gestão de documentos subsequentes.

O desenvolvimento do Projeto M51-CLAV assenta em várias fases de trabalho, nomeadamente:

1) Levantamento de requisitos funcionais e não funcionais, que decorrerá em 2 etapas:

- 1ª etapa (2017): definição dos requisitos necessários ao desenvolvimento da plataforma no que diz respeito à disponibilização do referencial Lista Consolidada e ao controlo da eliminação da informação pública ativa.

- 2ª etapa (2018): definir-se-ão os requisitos necessários para o desenvolvimento da plataforma no que diz respeito à disponibilização de Portarias de Gestão de Documentos (PGD's) ainda em vigor e de Relatórios de Avaliação de Documentação Acumulada (RADA's), bem como ao controlo da eliminação da informação pública de uso não corrente.

2) Modelação e análise de informação (2017).

3) Implementação de sistema de suporte à ontologia, que decorrerá em 2 etapas:

- 1ª etapa (2017): implementação do sistema de suporte ao referencial LC para a informação ativa.

- 2ª etapa (2018): implementação do sistema de suporte às PGD's e RADA's (nos casos aplicáveis) para a informação de uso não corrente.

4) Implementação de sistema de controlo da eliminação da informação pública:

- 1ª etapa (2017): implementação do sistema de suporte ao controlo da eliminação informação ativa.

- 2ª etapa (2018): implementação do sistema de suporte ao controlo da informação de uso não corrente, de que resultarão aplicações web para a disponibilização deste serviço.

5) Testes, avaliação de usabilidade e implementação de melhorias

- 1ª etapa (2017): teste do sistema de suporte ao controlo da eliminação informação ativa.

- 2ª etapa (2018): teste do sistema de suporte ao controlo da eliminação informação não ativa.

6) Estudo com proposta legislativa para um novo enquadramento da aplicação de instrumentos transversais de classificação e avaliação, bem como dos produtos e serviços disponibilizados pela plataforma modular.

7) Disseminação do "Arquivo Digital: Plataforma modular para a classificação e a avaliação da informação pública".

2. Resultados e Discussão

2.1. Levantamento de requisitos

O trabalho desenvolvido na etapa do Levantamento de requisitos funcionais e não funcionais relativos à Plataforma modular alicerçou as etapas subsequentes, tendo permitido:

1/ Identificar os requisitos da ontologia acima referida, de que se destaca a definição das entidades do modelo ontológico, bem como dos que são necessários à sua gestão;

2/ Identificar os requisitos necessários à desmaterialização do procedimento de avaliação e de eliminação, incluindo procedimento de submissão de novas propostas;

3/ Identificar serviços e produtos com os quais será necessário fazer interface como fornecedor de serviços e para a interoperabilidade, como por exemplo: Regulamento Nacional de Interoperabilidade Digital (RNID); Plataforma de Interoperabilidade na Administração Pública (iAP), Portal Dados.gov, Linked Open Data (LOD);

4/ Identificar e caracterizar os mecanismos de interação entre as entidades públicas e o cidadão para disponibilização e caracterização de informação pública (em articulação com o Portal do Cidadão e Portal Dados.gov);

5/ Mecanismos de autenticação previstos nas políticas públicas de modernização administrativa.

Nesta fase do trabalho, desenvolvida pela DGLAB e pela Universidade do Minho, foram recolhidos complementarmente contributos junto de futuros utilizadores, nomeadamente os organismos da AP, e ainda, junto de fornecedores de *software* de gestão de documentos de arquivo (*records system*), em duas reuniões organizadas para esse efeito.

2.1.1. Requisitos para as entidades do modelo ontológico e respetiva gestão

Na fase inicial do levantamento de requisitos, foram identificadas como entidades (classes abstratas) do modelo ontológico: Lista Consolidada, Tabela de Seleção, Auto de Eliminação, Termos de Índice, Legislação, Organização, Tipologia de Organização, Utilizador e Logs.

Lista Consolidada: É uma estrutura hierárquica de classes criada para a classificação dos processos de negócio (PN's) da AP, constituída por 4 níveis:

1º nível: representação das funções da AP;

2º nível: representação das subfunções da AP;

3º nível: representação dos PN's executados pela AP;

4º nível: representação de subdivisão dos processos de negócio para efeitos de avaliação. Esta classe apenas foi definida nos casos em se revelou necessário atribuir diferentes prazos e destinos finais à materialização informacional de distintas etapas dos PN's.

Tabela de Seleção: Gerada a partir da Lista Consolidada, representa um subconjunto da LC em que aos atributos de cada nodo são acrescentados novos atributos que permitam guardar informação relativa a prazos de conservação administrativa, forma de contagem desses prazos e destino final a aplicar à documentação/informação produzida.

Auto de Eliminação: É da competência de uma organização, que tenha uma PGD, organizar, coordenar e executar todas as ações de eliminação segundo os parâmetros estabelecidos pela PGD.

Termo de Índice: Vocabulário controlado definido de forma incremental na plataforma, que poderá ser associado a cada um dos processos/classes com o objetivo de facilitar a pesquisa/consulta ou permitir uma nova forma de navegar na informação da LC. A partir destes termos será gerado um índice para acesso direto às classes/processos.

Legislação: A plataforma guardará um catálogo da legislação de suporte à LC e às tabelas de seleção. Este catálogo será gerido de forma incremental.

Organização: Entidades da AP que intervêm no PN na qualidade de dono ou de participante, com consequências ao nível da conservação dos documentos.

Tipologia de Organização: Para facilitar a gestão, dentro da plataforma, das relações entre organizações e processos introduziu-se mais este conceito, que é um agregador de organizações que partilham determinadas características. Exemplos: Agrupamentos de Centros de Saúde, Cartórios Notariais.

Utilizador: Este conceito representa os utilizadores da plataforma que pertencerão a um determinado perfil que determinará o seu nível de permissões e consequentemente as operações e funcionalidades a que terão acesso.

Logs: Todas as ações realizadas na plataforma ficam inscritas num registo de logs. Este funcionará como um histórico de operações e servirá de suporte a estatísticas de utilização e até mesmo de responsabilização pelas operações realizadas. Nesse registo fica guardada a data, o tipo de operação e o utilizador que a realizou.

Foram considerados como requisitos de gestão:

- Identificação de requisitos para a gestão da Lista Consolidada, prevendo as respetivas operações de manutenção e relações entre as classes;
- Identificação de requisitos para a gestão das Tabelas de Seleção, prevendo os requisitos do modelo e das operações de manutenção a ter em conta;
- Identificação de requisitos para a gestão dos Autos de Eliminação, para as operações de manutenção e respetivo modelo;
- Identificação de requisitos para a gestão dos Termos de Índice, do respetivo modelo e operações de manutenção;
- Identificação de requisitos para a gestão da Legislação, com identificação do respetivo modelo, das operações de manutenção e das relações entre a Legislação e as outras entidades do modelo;
- Identificação de requisitos para a gestão das Organizações, atributos do modelo, operações de manutenção sobre a entidade Organizações e relações entre estas e outras entidades do modelo;
- Identificação de requisitos para a gestão das Tipologias de Organização, respetivo modelo e operações de manutenção.

Como forma de complementar os requisitos acima descritos, foram identificadas questões a que a plataforma deverá responder, podendo estas ser transversais ou relativas às diversas relações estabelecidas entre as instâncias das entidades da ontologia.

2.1.2. Requisitos para a desmaterialização do procedimento de avaliação e de eliminação

Iniciou-se ainda a identificação dos requisitos para a desmaterialização do procedimento de avaliação e de eliminação.

2.1.3. Requisitos para a interface e a interoperabilidade

Para garantir a interoperabilidade, o Projeto necessitou de efetuar o alinhamento com o Regulamento Nacional de Interoperabilidade Digital, aprovado pela RCM n.º 91/2012, que define as especificações técnicas e formatos digitais a adotar pela AP.

Considerou-se que, neste contexto, aplicam-se os seguintes domínios do RNID:

- a) Formatos de dados, incluindo códigos de caracteres (neste caso, o formato Unicode UTF-8), formatos de som e imagens (fixas e animadas), audiovisuais, dados gráficos e de pré-impressão;
- b) Formatos de documentos (estruturados e não estruturados) e gestão de conteúdos, incluindo gestão documental (por exemplo, para a documentação serão usados os formatos XML e PDF);
- c) Tecnologias de interface web, incluindo acessibilidade, ergonomia, compatibilidade e integração de serviços (no caso, no que diz respeito a tecnologias Web a utilizar: HTML, CSS, JavaScript, XSLT, XML);
- d) Especificações técnicas e protocolos de integração, troca de dados e orquestração de processos de negócio na integração interorganismos (por exemplo, para a troca de informação entre processos serão usados web services baseados em SOAP e SAML).
- e) Especificações técnicas e protocolos de comunicação em redes informáticas;
- f) Especificações técnicas de segurança para redes, serviços, aplicações e documentos.

O Projeto teve ainda a preocupação de garantir a disponibilização de serviços via iAP, enquanto plataforma de integração transversal à Administração Pública, quer para consumo de *webservices* disponibilizados por outras entidades, quer para disponibilização de *webservices* para outras entidades (exemplo: Dados.gov).

Em concreto, a integração com a iAP far-se-á colocando a M51-CLAV como fornecedora de serviços nessa plataforma. Neste momento, pode ser feita através de *webservices* SOAP assíncronos com utilização de *WS-addressing*. Até ao fim do ano, estará também disponível a integração via REST.

Neste momento, a plataforma é constituída por quatro grandes serviços:

- Plataforma de Integração (PI);
- Fornecedor de Autenticação (FA);
- Plataforma de Pagamentos da Administração Pública (PPAP);
- Gateway de SMS da Administração Pública (GAP).

Para o M51-CLAV são particularmente relevantes a PI e o FA. A PI é importante porque o M51-CLAV terá de constituir-se como um novo fornecedor de serviços nessa plataforma e o FA porque será utilizado como serviço de autenticação.

A integração com o portal Dados.gov far-se-á disponibilizando a ontologia da plataforma em formato RDF.

Foram ainda definidos, em fase inicial, os requisitos necessários para potenciar a integração da plataforma M51-CLAV no LOD ("Linked Open Data"), de modo a permitir, entre outros, a exportação de vocabulários ontológicos em Resource Description Framework (RDF). Foram ainda considerados requisitos que implicam a modelação de dados em SKOS ("Simple Knowledge Organization System") e OWL ("Ontology Web Language").

2.1.4. Requisitos para a autenticação

Quanto aos mecanismos de autenticação previstos, procedeu-se à identificação de requisitos para a utilização do Cartão de Cidadão e da Chave Móvel Digital, enquanto mecanismos de autenticação e identificação eletrónica através do Fornecedor de Autenticação. A iAP já tem na sua plataforma de integração um fornecedor de autenticação que presta este serviço.

2.2. Modelação e análise

Com a fase de modelação e análise de informação, pretendeu-se:

1/ Definir o modelo abstrato de dados do sistema de informação de suporte à plataforma modular. Partiu-se da análise dos requisitos levantados na atividade anterior, para efetuar a especificação formal do modelo com todas as entidades, atributos e relações necessárias.

2/ Produzir a especificação das interfaces com sistemas externos: iAP, Dados.gov, Linked Open Data (LOD).

3/ Decidir sobre as tecnologias e normas que serão usadas na implementação do sistema, muitas delas dependentes das opções tomadas na especificação do modelo.

A modelação e análise de dados é usada para definir e verificar os requisitos de dados necessários para suportar os processos de negócios dentro do contexto de um sistema de informação organizacional.

Existem três tipos diferentes de modelos de dados que poderão ser produzidos na evolução dos requisitos para a implementação do sistema final.

Os requisitos são inicialmente registados como um modelo de dados conceptual, que é essencialmente um conjunto de especificações independentes da tecnologia e é usado para discutir os requisitos iniciais com as partes interessadas.

O modelo conceptual é então traduzido para um modelo de dados lógico, que documenta estruturas dos dados que podem ser implementados. A implementação de um modelo conceptual de dados pode exigir vários modelos de dados lógicos.

O último passo na modelação é transformar o modelo de dados lógicos num modelo de dados físicos que organiza os dados de forma a responder da melhor maneira aos requisitos de acesso, desempenho e armazenamento. A modelação define não apenas elementos de dados, mas também as suas estruturas e as relações entre eles.

2.2.1. Justificação da escolha de um modelo ontológico

Atendendo à natureza hierárquica de algumas entidades, como o são a Lista Consolidada e as Tabelas de Seleção, e à grande quantidade e complexidade de relações existentes entre as entidades abstratas, sobre as quais irá ser necessário raciocinar e extrair informação, o modelo encontrado para dar solução ao problema foi o ontológico.

O modelo ontológico baseia-se em relações binárias muito simples: X relaciona-se com Y pela relação r. Estas relações são normalmente representadas por triplos: (X, r, Y) em que:

- X: é o sujeito da relação e corresponde normalmente à instância de uma classe;
- r: é a relação, também designada por predicado;
- Y: é o objeto da relação e pode corresponder à instância de uma classe ou de um valor dum tipo primitivo (inteiro, decimal, tempo, string, etc).

Quando Y corresponde a uma instância de classe dizemos que estamos perante uma relação ou "Object Property". Quando Y é um valor dizemos que estamos perante um atributo ou "Data Property".

Com estes triplos é possível modelar praticamente tudo. No entanto, há que acautelar algumas situações, como aquelas em que é necessário recorrer a atributos compostos. O atributo não é um valor simples, mas um conjunto de valores (tuplo). Para resolver este caso, foi criada uma nova classe: Atributo Composto. As suas instâncias corresponderão a atributos compostos em que os vários valores que os compõem aparecerão como atributos destas instâncias (casos das "Notas de aplicação" e das "Notas de exclusão" referentes às classes da LC).

2.2.2. Modelo da Lista Consolidada

A realização de operações sobre a Lista Consolidada será controlada pelo nível de acesso de cada utilizador. Todas as operações que alteram o estado da LC só estarão disponíveis para utilizadores com acesso de administração.

A LC, como já referido atrás, é uma hierarquia de classes e processos. Todas as classes e processos são constituídos pelos seguintes atributos:

Código: Código identificador atribuído pela DGLAB aquando da criação da classe;

Título: Atributo textual que contém o título pelo qual a classe será reconhecida;

Descrição: Atributo textual que contém uma indicação do âmbito e do que constitui a classe, como por exemplo o seu início, etapas e fim, no caso de processos de negócio;

Notas de Aplicação: Atributo textual que descreve em que situações deve ser usada a classe. Podem ser genéricas (comuns a todas as organizações) ou específicas (próprias de uma ou mais organizações). No caso de serem específicas, terão de ter uma relação com as organizações em causa. Por este motivo, este atributo foi promovido a atributo composto;

Notas de Exclusão: Atributo textual que descreve em que situações a classe não deverá ser usada. Podem ser genéricas (comuns a todas as organizações) ou específicas (próprias de uma ou mais organizações). Podem ainda remeter para outras classes. Nesse caso, o texto descritivo irá conter uma ou várias referências que se pretendem transformar em hiperligações ativas na interface da plataforma. Devido a poderem ser específicas e, também, poderem remeter para outras classes, este atributo foi também promovido a atributo composto.

As classes de nível 3, relativas a processos de negócio, têm mais dois atributos:

Tipo de Processo: Pode ser "comum" ou "específico";

Processo Transversal: Sim ou Não (se "Sim", intervirão como participantes no processo outras organizações).

As classes de nível 3 ou de nível 4 (quando estas existirem) têm ainda os seguintes atributos adicionais:

Prazo de conservação administrativa (PCA): prazo durante o qual o documento deve ser preservado (pode ser alterado ao longo do tempo, mas todos os valores terão de ser guardados e associados às datas em que estavam em vigor). Devido à sua estrutura, este atributo foi também promovido a atributo composto;

Justificação do prazo de conservação: justificação para o prazo de conservação indicado. Pode decorrer do tipo de relação com a entidade Legislação ou do tipo de relação entre processos de negócio (de síntese, complementar, de sucessão ou cruzada). Tem relação com a Legislação em 90% dos casos e tem de estar associada ao fator tempo. Devido à sua estrutura, este atributo foi também promovido a atributo composto;

Destino final (DF): apresenta um valor fixo - conservação, eliminação e conservação parcial. Neste último caso, é preciso guardar a percentagem do que se conserva. Devido à sua estrutura, este atributo foi também promovido a atributo composto;

Justificação do destino final: justificação para o destino final; pode decorrer do tipo de relação com a entidade Legislação ou do tipo de relação entre processos de negócio (de síntese, complementar, de sucessão ou cruzada);

No modelo, quando um processo de negócio (terceiro nível) se subdivide em quartos níveis, os atributos PCA, Justificação do PCA, DF e Justificação do DF apenas devem estar preenchidos ao nível mais baixo.

2.2.3. Modelo das Tabelas de Seleção

Uma Tabela de Seleção (TS) é um subconjunto da LC, e por isso segue o mesmo modelo ao qual acrescem alguns atributos adicionais:

ResponsávelDF: identificação do responsável pelo cumprimento do destino final;

Status: a TS pode estar num de 4 estados possíveis:

1. Proposta,
2. Validada pela DGLAB,
3. Publicada,
4. Revogada.

Acresce ainda os atributos e relações relativos à identificação do diploma de suporte à TS.

2.2.4. Modelo dos Autos de Eliminação

Um Auto de Eliminação (AE) é um registo comprovativo do abate documental. É composto pelos seguintes atributos:

Código: identificador único, atribuído pelo sistema;

Data: data do auto, se este for gerado no sistema será a data do sistema à hora da geração do auto;

Organização-prod: organização produtora, que está associada à PGD;

Organização-resp: organização responsável pela eliminação. Pode corresponder à organização produtora ou à que lhe sucedeu;

Responsável: pessoa responsável pela criação do auto. Corresponde ao utilizador que estiver autenticado no sistema para este efeito;

eAutorizadoPor: relação entre o AE e o diploma legal, que deverá estar registado no catálogo legislativo da plataforma, e que autoriza a ação. Pode ser uma PGD ou um despacho, no caso do RADA;

Um AE pode referenciar várias agregações. Cada agregação é definida pelo seguinte conjunto de atributos:

Título: título da agregação (pasta de arquivo);

Código: referência;

Data-inicial: data inicial da agregação;

Data-final: data de encerramento da agregação (necessária para a contagem do prazo de conservação administrativa);

Classe: classe de nível 3 ou nível 4 respetiva (quando existente) de que depende a agregação.

2.2.5. Modelo dos Termos de Índice

O modelo é constituído pelos seguintes atributos:

Código: código identificativo do termo. Pode ser uma forma abreviada desde que se garanta que não há colisões;

Termo: atributo textual que contém o termo designativo.

2.2.6. Modelo da Legislação

O modelo é constituído pelos seguintes atributos:

Código: código identificativo do diploma legal. Pode ser uma forma abreviada desde que se garanta que não há sobreposições entre os códigos;

Título: atributo textual que contém o título do elemento legislativo e que, nalguns casos, descreve o âmbito e conteúdo;

Data: atributo que contém a data do elemento legislativo (entrada em vigor);

Tipo: atributo textual que contém o tipo do elemento legislativo;

Número: atributo textual que contém o número do elemento legislativo;

Link: atributo textual que contém o endereço Web para o documento original.

2.2.7. Modelo da Organização

O modelo é constituído pelos seguintes atributos:

Código: código identificativo da organização. Pode ser uma sigla desde que se garanta que não há sobreposições entre os códigos;

Nome: atributo textual que contém o nome designativo da organização;

Sigla: atributo textual que contém a sigla da organização.

2.2.8. Modelo da Tipologia Organização

O modelo é constituído pelos seguintes atributos:

Código: código identificativo da tipologia de organização. Pode ser uma sigla desde que se garanta que não há sobreposições entre os códigos;

Nome: atributo textual que contém o nome designativo da tipologia de organização;

Descrição: atributo textual opcional que contém uma memória descritiva e resumida da tipologia de organização;

Organizações: lista de organizações enquadradas nesta tipologia (opcional).

3. Especificação

A *Ontology Web Language* (OWL) é um conjunto de linguagens de representação de conhecimento para especificação de ontologias. As ontologias são uma maneira formal de descrever taxonomias e sistemas de classificação, definindo essencialmente a estrutura do conhecimento para vários domínios: os substantivos, que representam classes de objetos e os verbos, que representam as relações entre os objetos. As ontologias são tipicamente muito flexíveis, uma vez que se destinam a representar informações provenientes de todos os tipos de fontes de dados heterogêneas. Quando se fala numa solução tecnológica baseada em OWL, a Web Semântica está inerente a todo o processo.

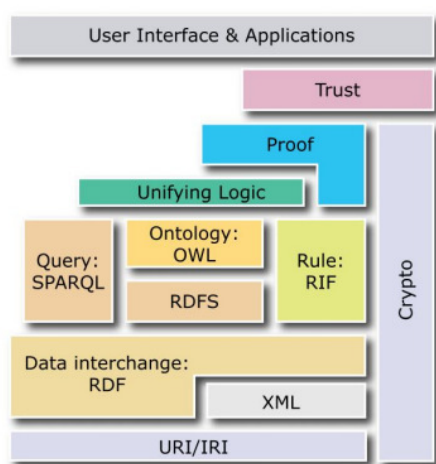


Figura 3- Pilha tecnológica da Web Semântica

Não há uma maneira correta de desenvolver uma ontologia. Neste Projeto seguiu-se uma metodologia iterativa: começou-se com uma versão básica à qual se foram

juntando elementos à medida que se foram detalhando os requisitos. Seguiram-se os seguintes princípios, fundamentais para a tomada de decisão:

- Não há uma maneira correta para modelar um determinado domínio; há sempre várias alternativas viáveis. A melhor solução depende normalmente da aplicação que se pretende obter no final;
- O desenvolvimento de uma ontologia é um processo iterativo;
- Os conceitos da ontologia devem aproximar-se muito de objetos, físicos ou lógicos, e de relações dentro do domínio específico. Os primeiros correspondem normalmente a substantivos e os últimos a verbos, nas frases que descrevem o domínio de interesse.

Resumindo, a decisão sobre o objetivo da ontologia e quão detalhada ou generalista irá ser, condicionará muitas das decisões de modelação que deverão ser tomadas. Perante várias alternativas possíveis, ter-se-á de escolher aquela que é melhor para o Projeto, que é mais intuitiva, mais extensível e mais fácil de manter e evoluir. É preciso lembrar que uma ontologia é um modelo de uma parte da realidade do mundo e os conceitos na ontologia devem refletir isso.

3.1. Representação gráfica da informação presente na ontologia

O levantamento de requisitos levado a cabo na primeira etapa do Projeto foi bastante exaustivo e os conceitos relevantes foram identificados no ponto correspondente à fase anterior.

Recorreu-se a ferramentas de representação gráfica da informação como forma de análise e validação dos requisitos identificados. Neste caso, essa análise foi feita recorrendo a duas amostras de dados: a classe 100 e a classe 150 da Lista Consolidada.

Esta análise foi feita com a ajuda da ferramenta Web "Word Clouds". Para cada amostra foi gerada uma lista de palavras e respetivas frequências e a respetiva *tagcloud* em formato de imagem. As duas imagens apresentam-se a seguir nas figuras 4 e 5.

Da análise das figuras, pode-se constatar que há grandes diferenças entre estas duas classes. Na classe 150 da LC há muitas mais relações e a ênfase está colocada aí. Na classe 100, a ênfase está nos conceitos.

As *tagclouds* geradas vieram confirmar os conceitos e as relações já enunciados no levantamento de requisitos.

A Figura 6 representa uma imagem do sistema de exploração visual que se tem vindo a usar para validar o trabalho de modelação. Neste exemplo, a rede encontra-se centrada na classe 150.10.700. À esquerda pode ver-se o grafo de relações desta classe e, à direita, o painel informativo com os atributos da classe.

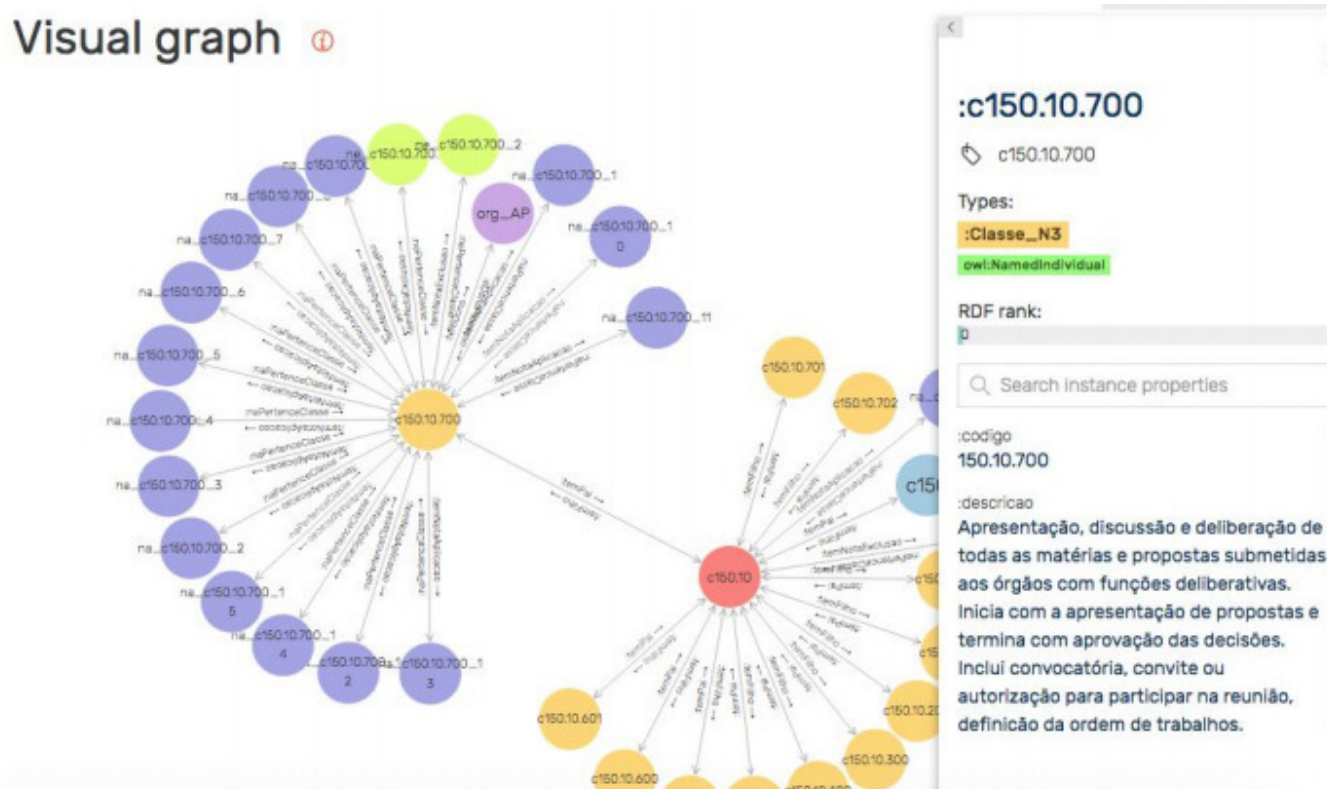


Figura 6 – Classe 150.10.700, seus atributos e relações

4. Conclusões

Nos últimos anos a Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB) desenvolveu vários projetos com vista à construção de referenciais para a classificação e avaliação da informação do Estado, que facilitassem a elaboração de planos de classificação e tabelas de seleção nos organismos públicos e tornassem mais eficiente e económico a gestão documental. A atual Lista Consolidada para a

classificação e avaliação da informação pública (LC), disponível no sítio web daquela entidade, é tributária dos referidos projetos e dos seus resultados.

Na sequência destes projetos, a DGLAB, em parceria com a Universidade do Minho, desenvolve atualmente o “Projeto M51-CLAV – Arquivo digital: Plataforma modular de classificação e avaliação da informação pública”, que visa constituir-se como a grande fonte de disponibilização da LC, bem como a grande via de interação dos organismos públicos com a entidade, com vista à apresentação de propostas de tabelas de seleção ou de eliminação da informação pública,

No atual Projeto, foram desenvolvidas as etapas de Levantamento de requisitos funcionais e não funcionais e de Modelação e análise de informação, cujos contornos, apresentados com algum nível de detalhe neste artigo, foram essenciais para a etapas que se seguem. Atualmente, o trabalho continua a ser desenvolvido, encontrando-se agora a decorrer, simultaneamente, as etapas relativas à Implementação de sistema de suporte à ontologia de informação ativa e ao Controlo e eliminação de informação ativa, que beneficiaram dos resultados aqui indicados.

Pretende-se, deste modo, contribuir para a concretização da Plataforma modular de classificação e avaliação da informação pública M51-CLAV, que constitui uma inovação para a Administração Pública, automatizando e simplificando um processo anteriormente complexo.

5. Referências

LOURENÇO, Alexandra; et al. (2013) - Orientações básicas para o desenvolvimento dos 3^{os} níveis de planos de classificação conformes à Macroestrutura funcional [Em linha]. Lisboa: DGLAB]. [Consult. 2 Out. 2017]. Disponível na Internet: http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/2013_Orient-3-niveis_PC-MF.pdf

LOURENÇO, Alexandra; RAMALHO, José Carlos; PENTEADO, Pedro (2015) - Uma ontologia para os processos de negócio da Administração. In: Atas do 12.º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas. [Em linha]. Lisboa: BAD. [Consult. 2 Out. 2017]. Disponível na Internet: https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/1453/pdf_89

LOURENÇO, Alexandra; PENTEADO, Pedro (2016) - Simplex +: o que precisamos para além da Medida 51? [Em linha]. Lisboa: BAD]. [Consult. 2 Out. 2017]. Disponível na Internet:

<https://www.bad.pt/publicacoes/index.php/arquivosmunicipais/article/view/1535/1465> .

LOURENÇO, Alexandra; PENTEADO, Pedro (2016) - Avaliação Suprainstitucional da Informação Arquivística (ASIA): documento metodológico. [Em linha]. Lisboa: DGLAB]. [Consult. 2 Out. 2017]. Disponível na Internet: http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2016/03/ASIA_Doc-metodologico2016-03-10.pdf .

LOURENÇO, Alexandra; PENTEADO, Pedro (2015) - A caminho da ASIA – Avaliação Suprainstitucional da Informação Arquivística. In: Atas do 12.º Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas. [Em linha]. Lisboa: BAD. [Consult. 2 Out. 2017]. Disponível na Internet: http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/1458/pdf_90 .

LOURENÇO, Alexandra; PENTEADO, Pedro (2014) – Una estrategia para mejorar el acceso y la reutilización de la información pública en Portugal: el papel de la interoperabilidad semántica. In Girona 2014: Archivos e Industrias Culturales. [Em linha]. Girona: ICA. [Consult. 2 Out. 2017]. Disponível na Internet: <http://www.girona.cat/web/ica2014/ponents/textos/id200.pdf> .

Nota: Deve ser ainda consultado o sítio Web da DGLAB onde a Lista Consolidada e os documentos metodológicos para a sua elaboração, bem como os conteúdos dos projetos que a antecederam, estão disponíveis.