



UNIVERSIDADE DO MINHO

MESTRADO INTEGRADO EM ENGENHARIA INFORMÁTICA

DISSERTAÇÃO

Plano de trabalhos

Ano Letivo 2019/2020

Identificação

Curso: Mestrado Integrado em Engenharia Informática

Orientando/Aluno: José Carlos Lima Martins (a78821@alunos.uminho.pt)

Orientador: Professor José Carlos Leite Ramalho, do Departamento de Informática da Universidade do Minho

Local de Trabalho: Departamento de Informática, Universidade do Minho, Braga

Tema

CLAV: API de dados e Autenticação

Resumo

O CLAV é um projeto nacional financiado pelo *Simplex*. O Objetivo deste projeto é a classificação e a avaliação de toda a documentação circulante na administração pública portuguesa. Desta forma as entidades públicas têm uma ferramenta por forma a saberem quando determinada documentação deve ser eliminada ou arquivada.

Nesta dissertação, a API do CLAV irá ser documentada em *Swagger*, parametrizando, descrevendo e adicionando exemplos para as várias rotas.

Ainda nesta dissertação, vão ser adicionadas mais possibilidades de exportação à API como o CSV, XML e RDF para além do JSON já existente.

Quanto à autenticação, pretende-se continuar os trabalhos de integração na aplicação principal da autenticação do Authentication.gov desenvolvido pela AMA usando SAML 2.0 com as extensões que a AMA considera obrigatórias. Ainda de um ponto de vista da autenticação, pretende-se também proteger a API de dados com múltiplos níveis de acesso.

Outro ponto que se pretende estudar, no âmbito da autenticação, é o da criação de um API Gateway e verificar se este componente pode auxiliar na simplificação da estrutura final da aplicação.

Por fim, mais tarde, pretende-se integrar o CLAV no backbone de serviços da AMA designado por iAP.

Enquadramento

São cada vez mais os governos e organizações que definem políticas e estratégias para a disponibilização de dados abertos nos domínios da ciência e da Administração Pública. Da mesma forma têm sido promovidas políticas para a transformação digital na Administração Pública portuguesa com o objetivo da otimização de processos, a modernização de procedimentos administrativos e a redução do consumo de papel. Esta transformação tem sido conseguida com a desmaterialização de processos, a promoção da adoção de sistemas de gestão documental eletrónica e da digitalização de documentos destinados a ser arquivados [2].

Com estes objetivos em mente, a DGLAB (Direção-Geral do Livro, dos Arquivos e das Bibliotecas) apresentou a iniciativa da Lista Consolidada para a classificação e avaliação da informação pública. Esta Lista Consolidada serve de referencial para a construção normalizada dos planos de classificação e tabelas de seleção das entidades que executam funções do Estado [2].

Por forma a tornar disponível a utilização da Lista Consolidada, a DGLAB em colaboração com a Universidade do Minho (e financiada pelo *Simplex*) procedeu ao desenvolvimento do projeto CLAV (Arquivo Digital: Plataforma modular de classificação e avaliação da informação pública) [3].

A plataforma CLAV disponibiliza a Lista Consolidada (ontologia com as funções e processos de negócio das entidades que exercem funções públicas) associada a um catálogo de legislação e de organismos. Esta informação é disponibilizada em formato aberto para a integração nos sistemas de informação das entidades promovendo a interoperabilidade através da utilização de uma linguagem comum (Lista Consolidada) usada no registo, na classificação e na avaliação de informação pública. A plataforma também viabiliza a desmaterialização dos procedimentos associados à elaboração de tabelas de seleção tendo como base a Lista Consolidada e ao controlo de eliminação e arquivamento da informação pública através da integração das tabelas de seleção nos sistemas de informação das entidades alertando-as quando determinado documento deve ser arquivado ou eliminado. [3].

A ontologia está armazenada numa base de dados de triplos RDF (Resource Description Format) e o seu modelo foi descrito em OWL (Ontology Web Language). A ontologia pode ser explorada através da linguagem/protocolo SPARQL (SPARQL Protocol and RDF Query Language) bem como integrada com a LOD (Linked Open Data) ou com a iniciativa do governo português dados.gov.pt [4].

Está a ser desenvolvida a API de dados por forma a permitir a processos ou aplicações aceder aos dados sem a intervenção humana para além claro de suportar a plataforma CLAV. Esta interface é desenvolvida seguindo uma metodologia REST. As operações são acessíveis através de URLs e respondem com informação em JSON, onde após o final desta dissertação a mesma informação poderá ser obtida também em CSV, XML e RDF. A ideia da API de dados é no futuro ser possível criar novas aplicações usando a API de dados do CLAV [4]. Para tal é essencial que a API de dados possua uma boa documentação para que futuros programadores/utilizadores consigam compreender bem como utilizar a API. Com esse objetivo em mente, nesta dissertação será realizada a documentação da API em *Swagger* explicando as várias operações possíveis da API.

Apesar de o projeto ter em mente a disponibilização aberta de informação pública é necessário existir algum controlo quanto à alteração, adição e eliminação da informação presente na Lista Consolidada (ontologia) para que a mesma se mantenha consistente e correta. Há, portanto, a necessidade de controlar os acessos à API protegendo as rotas com múltiplos níveis de acesso restringindo as operações que o utilizador pode realizar consoante o seu nível. Desta forma garante-se que apenas pessoal autorizado pode realizar modificações na Lista Consolidada. A criação de Tabelas de Seleção (para futura integração nos sistemas das entidades públicas) é da

mesma forma controlada visto que é mantida uma cópia na ontologia.

Este controlo de acesso obriga a existir formas de autenticação. Uma delas, Authentication.gov (Autenticação.gov), criada pelo Estado português, permite a autenticação dos cidadãos portugueses nos vários serviços públicos [1] entre os quais, a Segurança Social, o Serviço Nacional de Saúde e a Autoridade Tributária Aduaneira. Como tal, e sendo este um projeto do Governo Português, pretende-se que seja possível a autenticação no CLAV através do Authentication.gov.

Por outro lado, devido à complexidade que o controlo de acesso e a autenticação adicionam pretende-se estudar se a criação de um API Gateway permite a simplificação da comunicação entre interface/utilizadores e a API.

Objetivos

- Documentação da API de dados
- Adição de mais formatos de exportação à API de dados
- Integração do Authentication.gov
- Proteção da API de dados
- Estudar a criação de um API Gateway
- Integração do CLAV no iAP

Métodos de trabalho/investigação

- Reuniões semanais com o orientador
- Reuniões periódicas com a equipa da Torre do Tombo (DGLAB - Direção-Geral do Livro, dos Arquivos e das Bibliotecas)

Calendarização

- **Estudo do problema.** Pesquisa e compreensão do que já se encontra desenvolvido.
- **Documentação em *Swagger*** da API de dados do CLAV
- **Adição de outros formatos de exportação** à API tais como CSV, XML e RDF
- Continuação do trabalho de **integração** da autenticação **Authentication.gov** usando SAML 2.0 com extensões consideradas obrigatórias pela AMA
- **Proteção da API** de dados com múltiplos níveis de acesso
- **Estudo da criação de um API Gateway** bem como se este componente auxilia na simplificação da estrutura da aplicação
- **Integração do CLAV** no backbone de serviços da AMA designado **iAP**
- **Escrita da dissertação**

	2019			2020						
	Out.	Nov.	Dez.	Jan.	Fev.	Mar.	Abr.	Mai.	Jun.	Jul.
Estudo do Problema										
Documentação em Swagger										
Adição de formatos de exportação										
Integração do Authentication.gov										
Proteção da API										
Estudo da API Gateway										
Integração do CLAV no iAP										
Escrita da dissertação										

Referências

- [1] AMA. *Autenticação.gov*. Acedido em 24/09/2019. URL: <https://autenticacao.gov.pt/fa/Default.aspx>.
- [2] Alexandra Lourenço; José Carlos Ramalho; Maria Rita Gago; Pedro Penteado. *Plataforma CLAV: contributo para a disponibilização de dados abertos da Administração Pública em Portugal*. Acedido em 24/09/2019. Jul. de 2019, p. 1. URL: http://eprints.rclis.org/38643/1/Plantilla_EDICIC2019Barcelona-PT-COM_Penteado_v08.pdf.
- [3] Alexandra Lourenço; José Carlos Ramalho; Maria Rita Gago; Pedro Penteado. *Plataforma CLAV: contributo para a disponibilização de dados abertos da Administração Pública em Portugal*. Acedido em 24/09/2019. Jul. de 2019, p. 2. URL: http://eprints.rclis.org/38643/1/Plantilla_EDICIC2019Barcelona-PT-COM_Penteado_v08.pdf.
- [4] Alexandra Lourenço; José Carlos Ramalho; Maria Rita Gago; Pedro Penteado. *Plataforma CLAV: contributo para a disponibilização de dados abertos da Administração Pública em Portugal*. Acedido em 24/09/2019. Jul. de 2019, p. 21. URL: http://eprints.rclis.org/38643/1/Plantilla_EDICIC2019Barcelona-PT-COM_Penteado_v08.pdf.

Assinaturas

Orientador (José Carlos Leite Ramalho)

.....

Orientando/Aluno (José Carlos Lima Martins)

.....