

Universidade do Minho

Escola de Letras, Artes e Ciências Humanas

Helena Carolina Pacheco Faria

Espólio de Eurico Thomaz de Lima:

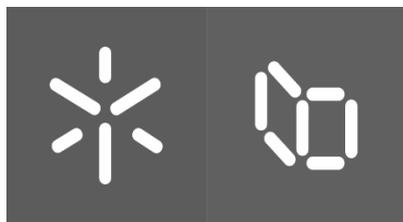
criação de uma edição eletrónica da correspondência

Espólio de Eurico Thomaz de Lima: criação de uma edição eletrónica da correspondência

Helena Carolina Pacheco Faria

UMinho | 2022

setembro de 2022



Universidade do Minho

Escola de Letras, Artes e Ciências Humanas

Helena Carolina Pacheco Faria

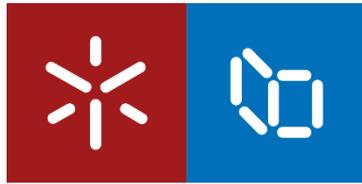
Espólio de Eurico Thomaz de Lima:

criação de uma edição eletrónica da
correspondência

Relatório de Projeto
Mestrado em Humanidades Digitais

Trabalho efetuado sob a orientação de:

Professora Doutora Idalete Maria Silva Dias
Professor Doutor José João Antunes Guimarães Dias Almeida



Universidade do Minho

Escola de Letras, Artes e Ciências Humanas

Helena Carolina Pacheco Faria

Espólio de Eurico Thomaz de Lima:

criação de uma edição eletrónica da
correspondência

Relatório de Projeto
Mestrado em Humanidades Digitais

Trabalho efetuado sob a orientação de:

Professora Doutora Idalete Maria Silva Dias
Professor Doutor José João Antunes Guimarães Dias Almeida

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho



Atribuição-NãoComercial-Compartilhalgual CC BY-NC-SA

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Agradecimentos

À minha família. Pelo amor, carinho e apoio incondicional.

Aos meus amigos. Por me motivarem a ser e fazer melhor.

Aos meus orientadores. Pela paciência e acompanhamento.

À minha tuna. Por todos os momentos que partilhamos e pelos que virão.

À Tânia. Por ser o meu pilar.

Declaração de Integridade

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho

Espólio de Eurico Thomaz de Lima: Criação de uma edição eletrónica da correspondência

Resumo

Este projeto tem como principal objetivo contribuir para a criação de uma edição eletrónica da correspondência do espólio de Eurico Thomaz de Lima, espólio pertencente à Universidade do Minho. Para assim ser possível, a correspondência necessita de tratamento eletrónico prévio. A correspondência terá de passar, então, por processos de anotação estrutural dos mesmos.

Os processos de transcrição terão como base a metalinguagem de anotação eletrónica Extensible Markup Language (XML), aliada às diretrizes de anotação da Text Encoding Initiative (TEI) e também as propostas do projeto Digital Archive of Letters In Flanders (DALF), sendo esta uma das referências notáveis no tratamento digital de documentos epistolográficos.

PALAVRAS CHAVE:

Eurico Thomaz de Lima; preservação digital; edição eletrónica; Text Encoding Initiative; Linguagem XML.

Eurico Thomaz de Lima's Asset: Creation of an electronic edition of the correspondence

Abstract

This project's main objective is to create an electronic edition of the correspondence from the assets of Eurico Thomaz de Lima, an asset belonging to the University of Minho. For this to be possible, correspondence requires prior electronic treatment. Correspondence will then have to go through structural annotation processes.

The annotation processes will be based on the Extensible Markup Language (XML) electronic annotation metalanguage, combined with the annotation guidelines of the Text Encoding Initiative (TEI) and also the proposals of the Digital Archive of Letters in Flanders (DALF) project, one of the notable references in the digital treatment of epistolographic documents.

Keywords:

Eurico Thomaz de Lima; digital preservation; electronic edition; Text Encoding Initiative; XML language.

Índice Geral

DIREITOS DE AUTOR.....	III
AGRADECIMENTOS	IV
DECLARAÇÃO DE INTEGRIDADE.....	V
RESUMO	VI
ABSTRACT	VII
ÍNDICE DE FIGURAS.....	IX
CAPÍTULO 1 – INTRODUÇÃO, CONTEXTUALIZAÇÃO DO PROJETO, OBJETIVOS E METODOLOGIA.....	10
1.1 Introdução ao projeto.....	10
1.2 Contextualização do projeto	11
1.3 Objetivos do Projeto	13
1.4 Metodologia	14
CAPÍTULO 2 – PRESERVAÇÃO DIGITAL DE DOCUMENTOS EPISTOLOGRÁFICOS	16
2.1 Introdução.....	16
2.2 Uso de Padrões e Normas de Preservação digital	18
CAPÍTULO 3 – O ESPÓLIO EPISTOLOGRÁFICO DE EURICO THOMAZ DE LIMA	19
3.1 Estrutura macro do Espólio de Eurico Thomaz de Lima	20
3.2 ESPÓLIO EPISTOLOGRÁFICO.....	21
3.2.1 Descrição da estrutura macro do Espólio.....	21
3.2.2 Segundo nível da descrição da estrutura macro	24
3.2.3 Descrição da estrutura micro	28
CAPÍTULO 4 – ANOTAÇÃO DA CORRESPONDÊNCIA	30
4.1 Text Encoding Initiative e DALF.....	30
CAPÍTULO 5: TEI HEADER E DTD.....	35
5.1 Metainformação: descrição da TeiHeader	35
5.2 Processo de validação do documento XML	42
CAPÍTULO 6: CONCLUSÃO	52
REFERÊNCIAS BIBLIOGRÁFICAS	53
ANEXO 1.....	55
ANEXO 2.....	59

Índice de Figuras

Figura 1: Representação visual do repositório digital.....	20
Figura 2: Organização da correspondência por caixas.....	21
Figura 3: Eurico Thomaz de Lima com Waldemar Henrique	22
Figura 4: Eurico Thomaz de Lima com Miecio Harzowski, Eugen Szenkar, João Itiberê da Cunha, Alda Caminha e Maria Sá Earp.....	22
Figura 5: Certidão de nascimento de E.T.Lima - frente e verso.....	22
Figura 6: Desenho de Eurico Thomaz de Lima por Manuel Gaspar	23
Figura 7: 1º prémio "Papoila de Oiro"	23
Figura 8: Excerto de partitura de Eurico Thomaz de Lima, "Variações Vimaranenses".....	24
Figura 9: Exemplo de envelope sem carta - frente e verso.....	25
Figura 10: Exemplo de parafernália - programa de concerto enviado juntamente com uma carta de Olga Pedrário	26
Figura 11. Carta nº155, pertencente à caixa 2.Mudança de remetente: Ernestina Santos Thomaz de Lima	27
Figura 12: Corpo da carta n. º 155	28
Figura 13: Recorte de jornal enviado juntamente com a carta n. º 60	40
Figura 14: Exemplo de pagela e oração	41
Figura 15: Partitura "Chô em Oitavas" de Arnaldo Rebello	42
Figura 16: Parte do DTD - elemento raiz.....	43
Figura 17: Secção <titleStmt>	43
Figura 18: Validação da secção <titleStmt>.....	44
Figura 19: Secção <publicationStmt>	44
Figura 20: Validação da secção <publicationStmt>	45
Figura 21: Secção <sourceDesc>.....	45
Figura 22: Validação da secção <sourceDesc>	46
Figura 23: Secção <physDesc>	46
Figura 24: Validação da secção <physDesc>	47
Figura 25: Secção <text>, subsecção <envelope>.....	48
Figura 26: Selo "Tudo pela Nação" - frase de António Oliveira de Salazar	48
Figura 27: Validação da secção <envelope>.....	49
Figura 28: Secção <body>.....	50

Figura 29: Validação da secção <body>51

Capítulo 1 – Introdução, contextualização do projeto, objetivos e metodologia

1.1 Introdução ao projeto

Ao longo dos tempos, temos visto um grande crescimento tecnológico que veio melhorar certos aspetos da vida humana. Também nas Humanidades se pode aliar o avanço tecnológico, proporcionando maior facilidade de acesso, tanto na investigação como na qualidade de ensino. Hoje em dia, são cada vez mais as pessoas que preferem o suporte digital ao suporte impresso porque carregar um e-book na mala é menos incomodativo que carregar um livro de capa dura, por exemplo. Assim, para corresponder à constante evolução envolvente, surge uma nova área humanística que permite manter as humanidades atualizadas através da preservação de toda a informação presente num documento impresso.

A Preservação Digital permite salvaguardar todo o tipo de documentação, entre os quais, documentos patrimoniais, como os documentos trabalhados neste projeto, e de interesse para a comunidade, bem como auxiliar todas as áreas de investigação e ensino, devido ao fácil acesso garantido. Com a preservação digital, é também possível criar arquivos digitais acessíveis a toda a comunidade. Os objetivos desta nova área centram-se em manter a longevidade dos documentos, a sua integridade intelectual e a fácil acessibilidade aos mesmos sem apresentar barreiras informáticas. A preservação digital permite, assim, construir uma versão eletrónica de um determinado documento onde é possível localizar, extrair e manipular a informação constante do texto. Grandes volumes de documentos digitais podem ser perdidos devido às falhas de planeamento e a insuficiência de conhecimentos sobre os efeitos da obsolescência tecnológica. (Santos & Flores, 2015)

Apesar de todo o avanço tecnológico e da vasta procura de documentos criados no mundo digital, ainda não existem muitos métodos de preservação e os que existem não garantem uma eficácia a longo prazo. Tanto a obsolescência tecnológica como o uso inadequado dos computadores podem levar a perda dos registos relevantes, deixando assim imensas lacunas nos acervos arquivísticos. Da mesma forma, a falta de procedimentos de segurança de preservação ameaça a autenticidade e a confiabilidade dos documentos digitais, pois estes registos são facilmente alterados, duplicados, convertidos e até mesmo falsificados sem deixar vestígios aparentes.

Assim, o projeto centra-se na criação de uma edição eletrónica da correspondência do espólio de Eurico Thomaz de Lima, espólio pertencente à Universidade do Minho. Este espólio é uma fonte de conhecimento e de património histórico, cultural e musical uma vez que documenta a vida e percurso profissional de um compositor de renome português. Este projeto pretende, ainda, contribuir através do processamento eletrónico dos documentos para a criação de uma ontologia.

1.2 Contextualização do projeto

Eurico Thomaz de Lima foi um célebre compositor, pianista e pedagogo português, filho de António Thomaz de Lima e de Maria Ernestina Santos Lima. Nasceu a 17 de dezembro de 1908 na ilha de São Miguel, Açores, e faleceu na cidade da Maia a 8 de junho 1989, deixando para trás uma vida cheia de prémios, destacando o 1º premio “Papoila de Ouro”, recebido em 1941 (Lessa, 2007). Este primeiro prémio é um dos objetos que se pode encontrar no espólio estudado. Thomaz de Lima ingressou no Conservatório Nacional de Música onde teve vários artistas como mentores entre eles, Alexandre Rey Colaço e Vianna da Mota no piano, Luís de Freitas Branco em Estética e História da Música e Hermínio do Nascimento no Solfejo (Gonçalves, 2005). Estes artistas são referidos várias vezes na correspondência do espólio onde se pode compreender a existência de uma grande cumplicidade entre os mesmos e o compositor, Eurico Thomaz de Lima.

Foi convidado pela Sociedade Nacional de Música da Câmara de Lisboa para uma série de concertos. O primeiro em 1926 e os restantes em 1929, 1931, 1935, 1937 (num recital com obras do compositor) e em 1939 (num recital com obras de F. Chopin). A 10 de Novembro de 1932, com apenas 24 anos, apresentou pela primeira vez um recital exclusivamente com obras da sua autoria cujas críticas o consagraram como compositor. Com o seu sucesso, Eurico Thomaz de Lima conseguiu levar a sua carreira além-fronteiras, sendo celebrenemente conhecido e aclamado no Brasil, depois de realizar duas digressões, 1949 e 1952, onde atuou nas mais importantes salas de concerto do país.

Tendo a sua vida dividida entre Portugal e Brasil, o compositor desenvolveu amizades nos dois países que pode ser comprovada depois da análise da correspondência do espólio. Eurico Thomaz de Lima manteve correspondência com personalidades do panorama musical brasileiro, em particular com a compositora Olga Pedrário e a cantora Alma Cunha de Miranda. Uma

componente relevante na sua vida profissional foi a sua dedicação à pedagogia musical, tendo lecionado em várias academias de música, nomeadamente a Academia de Amadores de Música de Lisboa, Academia Mozart do Porto, Academia de Música e Belas Artes da Ilha da Madeira, Academia Parnaso (Gonçalves, 2005). Uma parte significativa da sua obra e ação tem as crianças como público-alvo. Verifica-se uma preocupação com a progressão em termos de complexidade da composição à medida que a criança vai adquirindo mais conhecimento. Como referido em Gonçalves (2005), o folclore enquanto património cultural assume um papel preponderante.

Este projeto propõe contribuir para a criação do espólio digital da correspondência de Eurico Thomaz de Lima. O desenvolvimento do espólio digital em questão assenta na realização de uma série de fases interligadas entre si, começando pela transcrição e edição das cartas, fase coordenada pela Professora Doutora Anabela Barros, com a colaboração dos alunos do Mestrado em Humanidades Digitais da Universidade do Minho, ano letivo 2020/2021. A segunda fase, desenvolvida no âmbito do presente projeto, consiste no processamento eletrónico dos documentos epistolográficos de acordo com as diretrizes da Iniciativa de Codificação Textual, *Text Encoding Initiative (TEI)*, e a metalinguagem de anotação, *Extensible Markup Language (XML)*. A terceira fase, em desenvolvimento por Tânia Rego, estudante do mesmo mestrado, consiste na criação de uma ontologia que represente o conhecimento presente nas cartas, partindo dos derivados da fase anterior. A escolha deste projeto advém do interesse pessoal em documentos antigos e na sua contextualização, seja ela histórica ou cultural, que levaram à existência dos mesmos. Estes documentos são património cultural, para além de desvendarem pormenores biográficos de Eurico Thomaz de Lima, permitem o estabelecimento de várias ligações em termos musicais, mas também em termos históricos e culturais e, como tal, devem ser preservados para poderem ser visualizados pela comunidade e pesquisados por públicos vários, desde investigadores da área das Ciências Musicais a investigadores interessados nas relações entre músicos portugueses e brasileiros, na relação entre Eurico Thomaz de Lima e o Estado Novo, etc.

Paralelamente, a recolha de informação através da correspondência servirá de base para a criação de uma ontologia de informação sobre Eurico Thomaz de Lima, de forma a preservar o património que a correspondência demonstra. Este projeto encontra-se a ser desenvolvido pela Tânia Rego que também teve um papel ativo no processamento da correspondência.

Este projeto vem aliar os conhecimentos de edição digital adquiridos ao longo do Mestrado em Humanidades Digitais e também o gosto pessoal pela preservação digital de documentos epistolográficos que contribuem para o crescimento cultural da sociedade.

No âmbito do presente projeto foram sujeitas ao processamento eletrônico setenta cartas do espólio. As três fases acima referidas permitirão tornar acessível uma inestimável fonte de informação sobre a vida pessoal e profissional de Eurico Thomaz de Lima.

1.3 Objetivos do Projeto

O objetivo principal deste projeto é o processamento eletrônico das transcrições da correspondência de Eurico Thomaz de Lima de forma a torná-las mais acessíveis para toda a comunidade, através da sua disponibilização *online*. As edições podem também servir de objeto de estudo para futuras investigações. Este projeto apresenta também objetivos gerais, tais como a preservação do património cultural e musical e a contribuição para a criação de um arquivo digital pesquisável do espólio. Definidos os objetivos gerais, o projeto tem também como objetivos específicos:

- i. Familiarização com o espólio físico através de visitas ao Edifício dos Congregados onde o Espólio se encontra depositado;
- ii. Categorização das cartas por: cartas de amigos, cartas de compositores/músicos, cartas de família, cartas de alunos, entre outros;
- iii. Recolha de informação de elementos e relações desconhecidas através da análise do espólio;
- iv. A anotação estrutural e semântica das cartas, de e para o compositor, utilizando a metalinguagem de anotação eletrónica *Extensible Markup Language (XML)*, seguindo as diretivas de anotação da *Text Encoding Initiative (TEI)* e as propostas de anotação do Projeto *Digital Archive of Letters in Flanders (DALF)*;
- v. Criação de um livro de registo com os metadados das cartas, incluindo elementos vários que acompanham as mesmas;
- vi. Criação de um esquema de metadados que servirá de base para a elaboração das fichas bibliográficas das cartas em formato eletrónico;

- vii. Criação de um esquema de validação *Document Type Definition (DTD)* que permita a validação dos documentos *XML*.

Com a recolha de informação de elementos e relações desconhecidas na correspondência foi possível compreender melhor o percurso de Eurico Thomaz de Lima, visto este não ter sido muito estudado na comunidade portuguesa. Este projeto encontra-se paralelamente ligado com o projeto da Tânia Rego, que se encontra a criar uma ontologia das relações pessoais e profissionais do compositor, o que permite então criar uma espécie de “Euricopedia”¹.

Importa ainda referir que para além das cartas, também foi definido como objetivo propor esquemas de metadados para os envelopes e elementos que acompanham as cartas, tais como as pagelas, os postais, os poemas, os programas de concertos, as partituras e outros elementos, uma vez que estes elementos também contêm informação relevante para a descoberta de informações e relações entre as muitas peças do espólio, como remetentes, dedicatórias, agradecimentos e património histórico, cultural e musical.

Um objetivo mais específico deste projeto é a utilização da metalinguagem de anotação eletrónica XML, uma vez que é uma linguagem descritiva e semântica que proporciona liberdade ao autor para criar um modelo híbrido de anotação para os documentos do espólio que se baseie em propostas da TEI e em anotações customizadas de acordo com as especificidades dos documentos do espólio.

1.4 Metodologia

A elaboração deste projeto centrou-se nas seguintes etapas:

1. Fotografar o espólio de correspondência;
2. Proceder a uma análise macro do espólio de correspondência;
3. Analisar o espólio ao nível do microtexto com o intuito de elaborar um modelo de etiquetagem XML;
4. Elaborar de um documento XML composto por uma secção TEI Header composta por metadados e uma segunda secção com o texto do documento epistológico anotado conforme a TEI e o projeto DALF;
5. Criar um documento DTD que valide os documentos XML criados para todas as cartas;

¹ A ideia da criação de um género de “Euricopedia” foi proposta pelo Professor José João Almeida.

6. Contribuir para criar *scripts* para a disponibilização da versão eletrónica da correspondência.

Para a realização deste projeto foi necessária uma deslocação ao Edifício dos Congregados para fins de investigação e fotografia das cartas que ainda não tinham sido devidamente digitalizadas. Assim, foi possível aumentar o número de cartas a serem transcritas de forma a tornar este projeto mais rico e também para dar um contributo à UC lecionada pela Professora Anabela Barros no âmbito da qual, juntamente com os alunos do 1º ano do Mestrado em Humanidades Digitais, ano letivo 2020/2021, já haviam sido feitas transcrições de cerca de metade do espólio epistolográfico.

Para iniciar o processo de etiquetação e dar início à segunda etapa do projeto, procedeu-se a uma análise detalhada da correspondência para ser possível criar um modelo XML que servisse como base para a etiquetação de todas as cartas. É de notar que as cartas contêm propriedades que são comuns a (quase) todas, mas também propriedades únicas que as distinguem e que podem tornar o processo de criação de modelo XML ainda mais desafiante, uma vez que é necessário, nestes casos, propor etiquetas próprias do projeto. Neste ponto, refira-se o exemplo da existência no espólio de cartas soltas sem sobrescrito. Neste caso, foi necessário adaptar o esquema de metadados para etiquetar documentos epistolográficos com e sem sobrescrito. Verificou-se, ainda em relação aos sobrescritos, que as anotações propostas pela TEI não são suficientemente específicas, sendo necessário procurar outros modelos que abordassem esta vertente. Aqui surge o projeto DALF que oferece as propostas adequadas para a etiquetação dos envelopes.

A terceira etapa consistiu na análise das cartas ao nível do microtexto, ou seja, abordagem que incide sobre as informações do conteúdo, tais como entidades nomeadas, com vista à criação no futuro de um arquivo digital pesquisável. O levantamento das informações a anotar no conteúdo é essencial para a definição das etiquetas TEI-XML e DALF-XML a utilizar no esquema de anotação.

Passando para a quarta etapa, foi necessário proceder à criação de um documento DTD (documento que define a estrutura de anotação de um documento XML), e consequente validação dos documentos XML criados.

Na quinta e última etapa do projeto, as cartas anotadas em formato XML e validadas pelo documento DTD foram submetidas a um processo de *scripts* Python e CSS para assim ser criado um ficheiro PDF com as cartas em formato eletrónico.

Capítulo 2 – Preservação digital de documentos epistolográficos

2.1 Introdução

O crescimento tecnológico e a constante evolução das Humanidades caminham lado a lado na progressão da preservação digital e na criação de arquivos digitais. O texto é um dos meios de preservação de informação histórico-cultural, sociológica e biográfica, servindo de fonte de informação para fins de investigação nas mais variadas áreas de estudo. Paralelamente, as tecnologias têm vindo a ocupar um lugar de destaque na vida das pessoas. Atualmente, torna-se difícil imaginar uma sociedade sem tecnologias especialmente de informação e comunicação. Nota-se uma preferência pelo uso de documentos digitais devido à facilidade de acesso e manuseamento, sendo assim cada vez mais necessário preservar património textual em formato eletrónico e digital.

Assim, para a realização deste projeto, é necessário compreender a definição de edição semi-diplomática e edição interpretativa. A edição semi-diplomática tem um papel menos interventivo que a edição interpretativa visto que se trata da edição do manuscrito conforme ele se apresenta e o editor intervém no sentido de desenvolver abreviaturas e identificá-las. As edições semi-diplomáticas preservam as características essenciais dos documentos, assim como a ortografia, a pontuação, a acentuação, a translineação, entre outros. Já a edição interpretativa é uma edição mais elaborada e crítica de um texto de testemunho único (monotestemunhal), onde o editor para além de transcrever o texto, corrige os erros ortográficos oferecendo ao público um texto mais apurado pois desenvolve notas filológicas, linguísticas, históricas e outras que se tornem pertinentes para quem a edição se destina. Estas notas permitem revelar elementos relevantes para a compreensão do texto assim como o seu contexto histórico, social e cultural, uma vez que as edições interpretativas de documentos são mais direcionadas a um público que não é especialista na área.

As notas contêm informação sobre mudança de tinta, de punho, caracteres apagados/modificados seja nas entrelinhas como nas margens, entre outros (Keller & Gonçalves, 2018). Marengo e Cambraia (2016) afirmam que, além de tornar o texto acessível ao público leitor através da facilitação de leitura, não se devem desprezar a sistematicidade da metodologia para a concretização da mesma.

Existe uma necessidade crescente de resposta à edição de documentos epistolográficos e consequente preservação para serem acessíveis ao longo do tempo, uma vez que o documento

digital possui uma variedade de complexidades e especificidades que podem comprometer a sua autenticidade e acesso a longo prazo.

Desta forma, precisamos também compreender a definição de preservação digital e para tal destaca-se o seguinte conceito:

La preservación digital puede definirse como el conjunto de los procesos destinados a garantizar la continuidad de los elementos del patrimonio digital durante todo el tiempo que se consideren necesarios. La mayor amenaza para la continuidad digital es la desaparición de los medios de acceso. No puede decirse que se han conservado los objetos digitales si, al ya no existir los medios de acceso, resulta imposible utilizarlos. El objetivo de la preservación de los objetos digitales es mantener su accesibilidad, es decir, la capacidad de tener acceso a su mensaje o propósito esencial y auténtico. (Weeb, 2003)

Weeb reforça a importância da acessibilidade dos documentos digitais e a sua autenticidade, como também a contínua possibilidade de consulta a longo prazo. Defende também que os atributos dos documentos como, autenticidade, integridade, fidedignidade e utilização devem ser mantidos durante o processo de preservação. Assim, conseguimos compreender que a preservação digital consiste no conjunto de processos responsáveis por garantir o acesso continuado à informação digital durante longos períodos de tempo sem que a integridade e autenticidade dos documentos digitais seja comprometida.

Existem duas vertentes da preservação digital, por um lado, temos a preservação digital como a preservação da informação contida num suporte através da digitalização da imagem e por outro, temos a preservação digital enquanto armazenamento, manutenção e acesso ao documento digital a longo prazo.

É reconhecida a importância da preservação digital partindo do pressuposto de que é necessário desenvolver repositórios digitais que sejam fiáveis e que assegurem a preservação da informação, garantindo assim a sua acessibilidade no futuro.

É de notar que a preservação digital é um desafio com características de urgência na sociedade moderna e, por ser frágil a nível de hardware e software, é necessário aplicar estratégias de preservação de documentos digitais, uma vez que sem estes não existiriam garantias de acesso,

confiabilidade e integridade dos documentos. Porém, existe uma falta de métodos comprovados que assegurem a existência e a continuidade do acesso à informação digital.

2.2 Uso de Padrões e Normas de Preservação digital

Na preservação digital é necessário usar padrões de maneira a reduzir a diversidade de formatos digitais e reduzir também o risco de obsolescência tecnológica. Estes padrões podem ser de facto ou de jure. Os padrões de facto são aprovados por órgãos regulamentadores, mas são amplamente utilizados enquanto os padrões de jure são aprovados e regulamentados por um órgão competente como por exemplo a *International Organization of Standardization (ISO)* podendo aceder às suas especificidades.

Assim, um formato adequado para a preservação deve ser:

- (i) - Livre de qualquer restrição de uso;
- (ii) - Ter código aberto e documentado;
- (iii) - Ser amplamente utilizado;
- (iv) - Multiplataforma;
- (v) - Preferencialmente sem compressões ou de compressão sem perdas;
- (vi) - Configurar uma norma emitida por um organismo oficial

Para além de padrões, a preservação digital é composta por dois tipos de procedimentos: os procedimentos estruturais e os procedimentos operacionais. Os procedimentos estruturais (de manutenção) referem-se à fase inicial da preservação, nomeadamente a infraestrutura dos documentos, as normas que o acompanham, a escolha de padrões, financiamento necessário e a definição dos metadados. Os procedimentos operacionais (procedimentos de preservação dos componentes) são aqueles que se referem à preservação física, lógica e intelectual dos documentos digitais através de ações de migração (transferência periódica de material digital de uma dada configuração de hardware/software para uma outra, ou de uma geração de tecnologia para outra subsequente), emulação (utilização de software especial, designado vulgarmente por Emulador, que é capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software, inicialmente incompatível) e o encapsulamento. (Santos & Flores, 2015).

A preservação digital deve seguir normas e modelos de referência para a criação de um recurso utilizável por outros sistemas, bem como um recurso sustentável ao longo dos tempos. Uma das normas mais conhecidas e maioritariamente usada em projetos de edição e preservação digital é o modelo de referência *OAIS (Open Archive Information System)*. Este modelo determina o âmbito do repositório em função da comunidade de utilizadores assim como controla a informação de forma a atingir objetivos da preservação de longo prazo e ainda assegura que a informação é legível pela comunidade de utilizadores sem estes terem de recorrer à fonte de produção de informação (Saramago, 2004).

Uma vez que cada documento é único e apresenta problemas específicos, seja de leitura ou de edição, é compreensível que as normas existentes para cada tipo de edição não sejam de todo eficazes, isto é, não consigam englobar os múltiplos fenómenos que se encontram nos mais variados documentos. Até ao momento, não é possível dizer que existe uma estratégia 100% eficaz, visto que não existe uma que possa ser aplicada a todos os tipos de documentos digitais. Esta problemática levanta assim a necessidade crescente de explorar as estratégias já existentes.

Segundo o *Arquivo da Web Portuguesa (AWP)*, um formato necessita de conter os seguintes critérios para ser considerado um formato adequado: (i) isento de direitos legais que restrinjam a sua utilização; (ii) uma norma emitida por um organismo oficial; (iii) documentado abertamente através de uma especificação pública e livre; (iv) amplamente usado; (v) lido e escrito por múltiplas plataformas de software, incluindo código aberto; (vi) não comprimido ou comprimido sem perdas de informação (Ferreira, 2011).

Capítulo 3 – O espólio epistolográfico de Eurico Thomaz de Lima

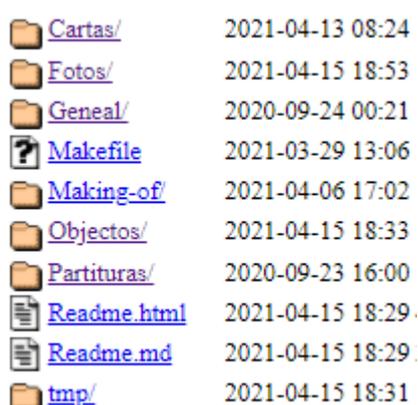
A partir das visitas realizadas à Sala Eurico Thomaz de Lima no Edifício dos Congregados em Braga é possível proceder à identificação de três níveis de análise do espólio: um primeiro nível

macro que procura descrever os elementos que compõem o espólio na sua totalidade; um segundo nível macro que incide sobre o espólio epistolográfico e um terceiro nível de análise de natureza micro com foco específico nos metadados, na estrutura e no conteúdo do objeto-carta, bilhete postal, cartão de visita, etc. Os subcapítulos que se seguem visam fornecer uma análise detalhada destes níveis.

3.1 Estrutura macro do Espólio de Eurico Thomaz de Lima

Neste projeto, foi reconhecido desde cedo a necessidade de preservar todos os documentos, cartas, postais, pagelas, prémios, fotografias, entre outros documentos que pertencem ao espólio de Eurico Thomaz de Lima. Estes documentos são uma fonte rica de informação de um compositor português, sendo assim considerado património musical e cultural. Estes documentos dão-nos entrada direta para a vida pessoal e profissional do compositor, informações sobre concertos e recitais que deu, alunos de quem foi professor nas escolas e academias de música, entre muitas outras informações de interesse para fins de investigação.

Do levantamento feito até à data de redação deste relatório, o espólio de Eurico Thomaz de Lima pode ser dividido em quatro partes, sendo elas, correspondência, partituras, álbuns de fotografias e objetos. A correspondência é composta por vários tipos de documentos desde envelopes, bilhetes postais, pagelas, recortes de jornais e programas de concertos que descrevem concertos musicais realizados pelo compositor, assim como, as obras interpretadas pelo mesmo.



 Cartas/	2021-04-13 08:24
 Fotos/	2021-04-15 18:53
 Geneal/	2020-09-24 00:21
 Makefile	2021-03-29 13:06
 Making-of/	2021-04-06 17:02
 Objectos/	2021-04-15 18:33
 Partituras/	2020-09-23 16:00
 Readme.html	2021-04-15 18:29
 Readme.md	2021-04-15 18:29
 tmp/	2021-04-15 18:31

Figura 1: Representação visual do repositório digital

3.2 Espólio Epistológico

3.2.1 Descrição da estrutura macro do Espólio

A correspondência encontra-se dividida por 4 caixas no total. A primeira caixa contém 95 cartas (C001 - C095); a segunda caixa contém 106 cartas (C096 - C202); a terceira contém 72 cartas (C203 - C275) e a quarta caixa contém 34 cartas (C276 - C310). A todos estes documentos foi atribuído um número identificativo para, assim, ser mais fácil o seu manuseamento, organização e localização. Para além das 34 cartas, a quarta caixa contém 3 fotografias de papéis e capas nas quais as cartas estavam guardadas antes de serem organizadas.

 Arrumar/	2021-04-06 17:02	-
 Cx1-MEP113-C001a_C095/	2021-03-29 10:10	-
 Cx2-MEP999-C096-C202/	2021-04-06 17:10	-
 Cx3-MEP108-C203-C275/	2020-12-19 15:31	-
 Cx4-MEP112-C276_C310/	2020-09-24 14:43	-
 meta-cartas/	2021-04-15 13:09	-

Figura 2: Organização da correspondência por caixas

Para além das pastas com a correspondência, no repositório digital encontra-se ainda uma pasta com o título “Arrumar” que contém cartas ainda por arquivar e, também, a pasta “meta-cartas” com toda a metainformação das cartas. Nesta pasta, encontra-se um catálogo manual com informação sobre toda a correspondência do espólio como, por exemplo, remetente e destinatário, data, dimensões da carta e uma pequena sinopse da mesma. Embora a autoria do catálogo seja desconhecida, importa referir o contributo precioso do mesmo para a realização do presente projeto.

Passando para cada uma das restantes pastas da Fig. 1, a pasta “Fotos” contém álbuns de fotografias de Eurico Thomaz de Lima onde se destacam alguns artistas mencionados nas cartas, como, por exemplo, Alma Cunha de Miranda, António Boto com um autógrafo dedicado a Eurico Thomaz de Lima, Mieczysław Horszowski, Eugen Szenkar, João Itiberê da Cunha, Alda Caminha, Maria Sá Earp, Waldemar Henrique, a sua esposa Ângela Thomaz de Lima, entre outros. Encontra-se também uma reportagem fotográfica de Eurico Thomaz de Lima da sua primeira digressão ao Brasil, em 1949, onde relata os momentos importantes da mesma com destaque para as fotografias representadas nas Figuras 3 e 4.



Figura 3: Eurico Thomaz de Lima com Waldemar Henrique



Figura 4: Eurico Thomaz de Lima com Mieczysław Horzowski i, Eugen Szenkar, João Itiberê da Cunha, Alda Caminha e Maria Sá Earp.

Ao nível macro, os álbuns fotográficos, para além da própria fotografia, também são álbuns comentados pelo próprio compositor e a conjugação da fotografia com o texto fornece informação importante para a etiquetação. Na pasta “Geneal” (Genealogia), encontra-se a certidão de nascimento de Eurico Thomaz de Lima com informação sobre local e data de nascimento, os pais,

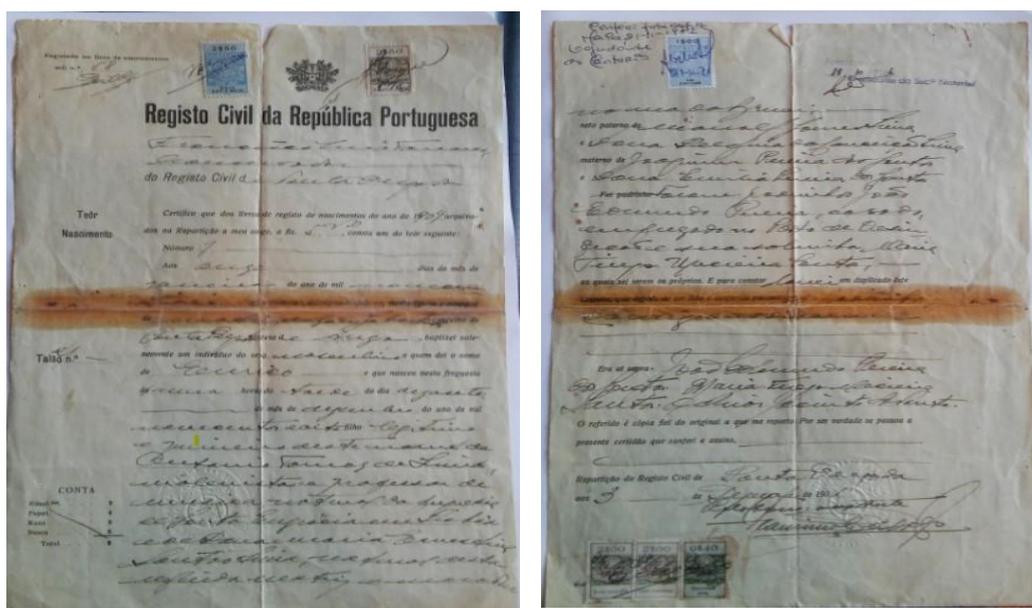


Figura 5: Certidão de nascimento de Eurico Thomaz de Lima - frente e verso

avós e padrinhos, como se pode observar na Figura 5.

Para além de todos estes documentos, encontram-se ainda no espólio alguns diplomas de Eurico Thomaz de Lima, como o diploma de sócio honorário da Casa do Porto e o diploma de sócio honorário do Liceu Literário. Também se encontra um desenho de Eurico Thomaz de Lima, cuja autoria é de Manuel Gaspar. Neste desenho, Eurico Thomaz de Lima encontra-se debruçado sobre o piano, o que era, sem dúvida, uma prática constante na vida do compositor.



Figura 6: Desenho de Eurico Thomaz de Lima por Manuel Gaspar

Além destes elementos, é possível encontrar também o 1º prémio “Papoila de Ouro”, prémio de composição da Emissora Nacional, recebido no concurso “Jogos Florais”, em 1941 (Lessa, 2007).



Figura 7: 1º prémio "Papoila de Ouro"

Por último, no espólio foram encontradas 12 partituras de Eurico Thomaz de Lima, de onde se destacam as obras “Marianita”, “Dança Portuguesa”, “Variações Vimaraneses” e “Senhora Quintaneira”.

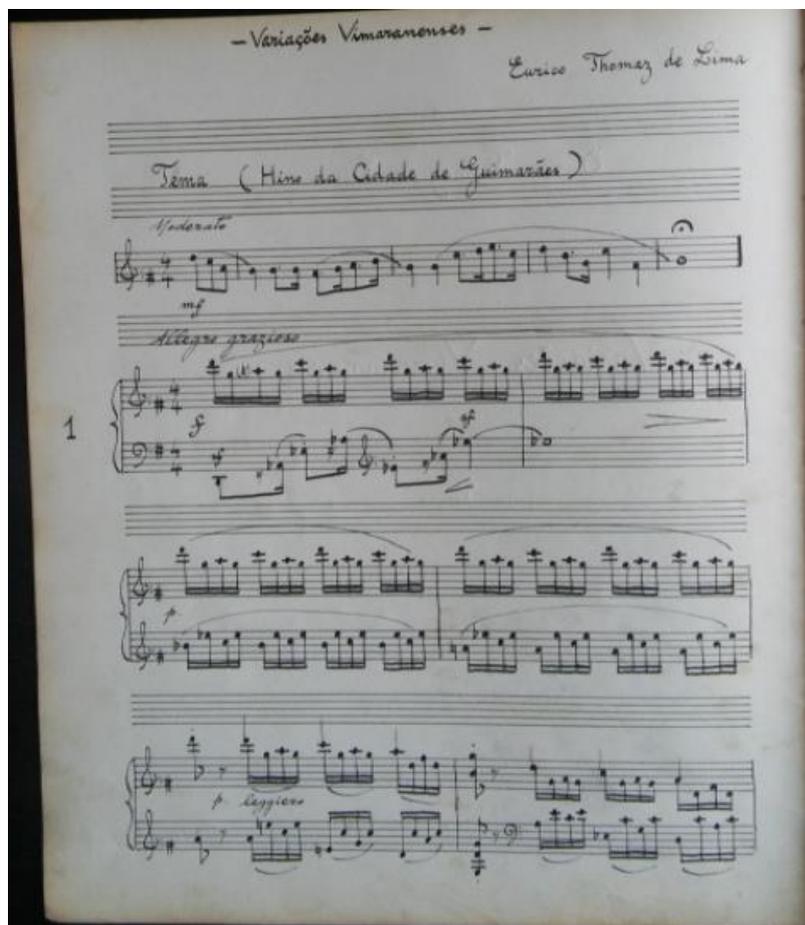


Figura 8: Excerto de partitura de Eurico Thomaz de Lima, "Variações Vimaraneses"

Uma vez descrito o espólio e depois da análise macro do mesmo, segue-se então uma análise do segundo nível da estrutura macro do espólio epistolográfico onde serão descritos os objetos carta, bilhetes postais, cartões de visita e também os tipos de parafernália presentes na correspondência do espólio.

3.2.2 Segundo nível da descrição da estrutura macro

A correspondência do espólio contém diversos tipos de documentos, tais como cartas, bilhetes postais, cartões de visita, recortes de jornais, partituras, poesias e dedicatórias que também precisam de ser transcritas e anotadas. Maioritariamente, as cartas fazem-se acompanhar de envelope, sendo este um elemento precioso para os metadados da correspondência, porém existem casos em que 2 cartas se fazem acompanhar por apenas 1

envelope ou então não existe envelope a acompanhar a carta. Dentro desta problemática, destaca-se uma série de cartões de visita de Berta Alves de Sousa dirigidos a Eurico Thomaz de Lima que estão todos agrupados no mesmo envelope. Estas situações constituem um desafio para o registo eletrónico da organização física dos documentos do espólio.

Os envelopes são objetos de informação valiosos para os metadados da correspondência, uma vez que fornecem informações valiosas tais como a morada do remetente e do destinatário, a data da redação da carta, se o envelope tiver selo, este também poderá conter informações relevantes sobre a cidade ou época vivida na altura, informações que contribuem para o processo de autenticação da correspondência. Existem também situações em que apenas temos presente no espólio o envelope que envolvia a carta. Tomando como exemplo um envelope da caixa 4, numerado de 307, que apresenta destinatário que é Eurico Thomaz de Lima e a morada do mesmo, Rua da Cedofeita, 29 1º - Porto, mas não apresenta remetente nem a respetiva morada, como pode ser visto na figura 9.

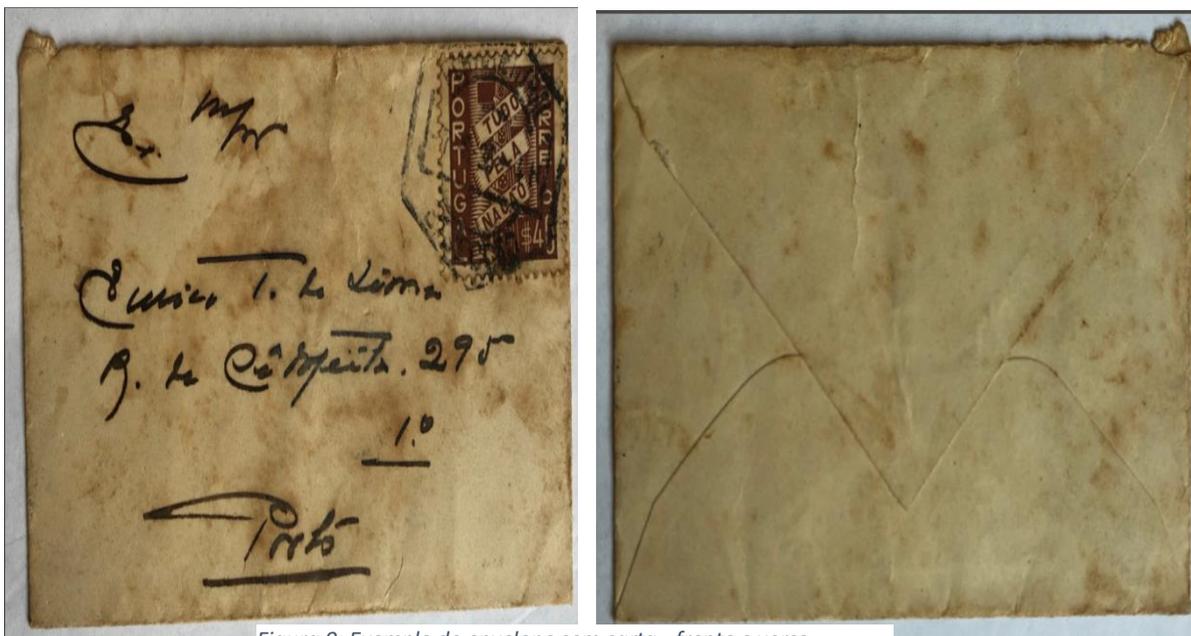


Figura 9: Exemplo de envelope sem carta - frente e verso

Outra característica da correspondência deste espólio é a existência de parafernália que pode surgir das mais variadas formas como fotografias, partituras, programas de concertos e também recortes de jornais. Tomando o exemplo da carta nº36, enviada por Olga Pedrário a Eurico Thomaz de Lima, a compositora envia juntamente com a carta um programa de concerto onde teve uma atuação. Neste programa encontra-se informação que permite identificar factos

biográficos e ligações entre entidades, músicos, instituições, obras, entre outras, tal como, a data e o local do concerto, os artistas presentes, as músicas interpretadas, o pianista, a ordem das obras e comentários às obras interpretadas.

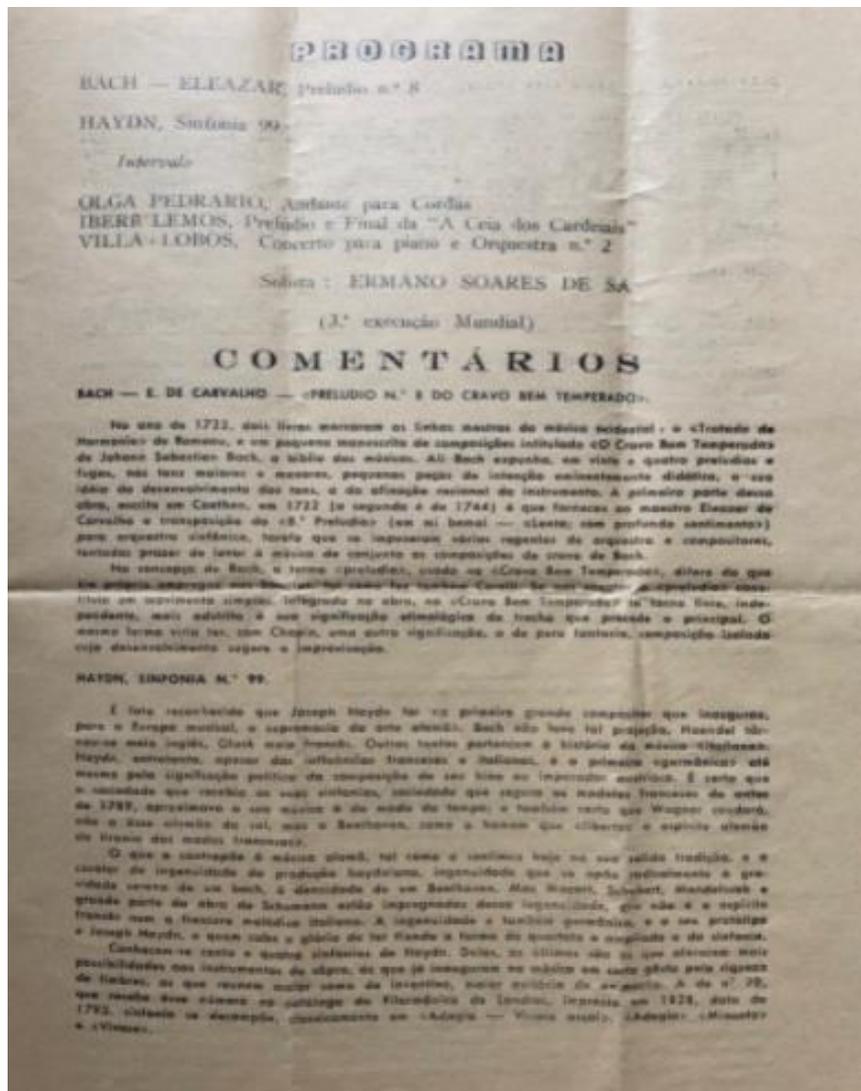


Figura 10: Exemplo de parafernália - programa de concerto enviado juntamente com uma carta de Olga Pedrário

Para além de duas cartas no mesmo envelope, existem também casos de cartas escritas por duas mãos, isto é, uma pessoa termina com uma saudação final e uma outra pessoa inicia uma outra secção da carta com uma saudação inicial. Tomando como exemplo a carta n.º 277, o remetente inicial é António Thomaz de Lima, pai de Eurico Thomaz de Lima, mas no final da mesma é possível ler um *postscript* (p.s) onde se verifica a mudança de remetente que passa a ser a mãe, Ernestina Santos Thomaz de Lima, que escreve a Eurico Thomaz de Lima. Na figura abaixo, segue-se o exemplo da carta n.º277 onde é possível observar a mudança de remetente e também uma mudança na caligrafia.

Mudança de remetente: Ernestina Santos Thomaz de Lima

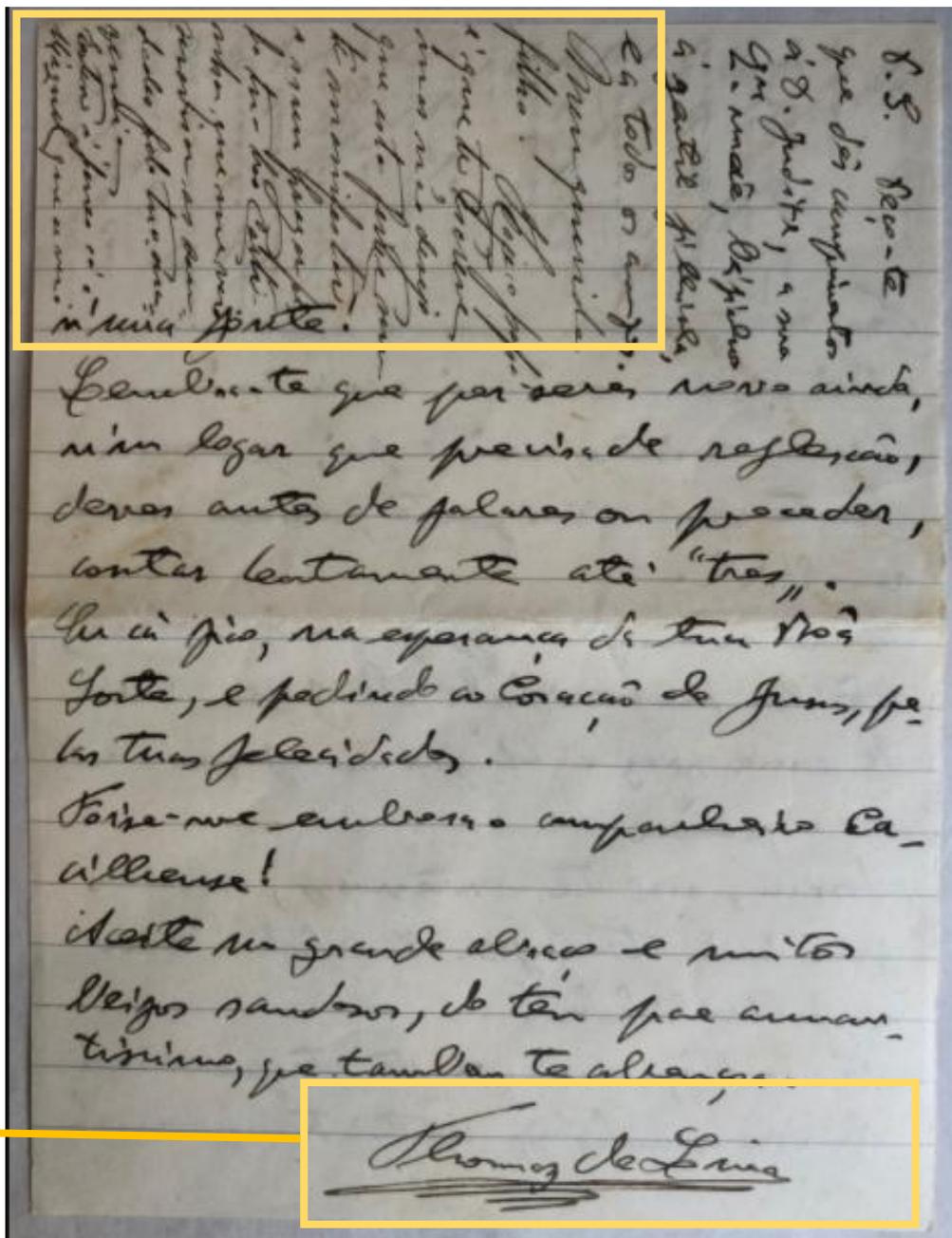


Figura 11: Carta n.º 277 - exemplo de carta com dois remetentes

3.2.3 Descrição da estrutura micro

Compreende-se por estrutura micro, todos os elementos que compõem o texto da carta, sendo estes, o cabeçalho normalmente composto por data e local da redação da carta, uma saudação inicial, o corpo da carta, despedida e possíveis *postscripts* (*p.s.*). Neste subcapítulo será feita uma descrição de todas estas partes com exemplos que a suportam.

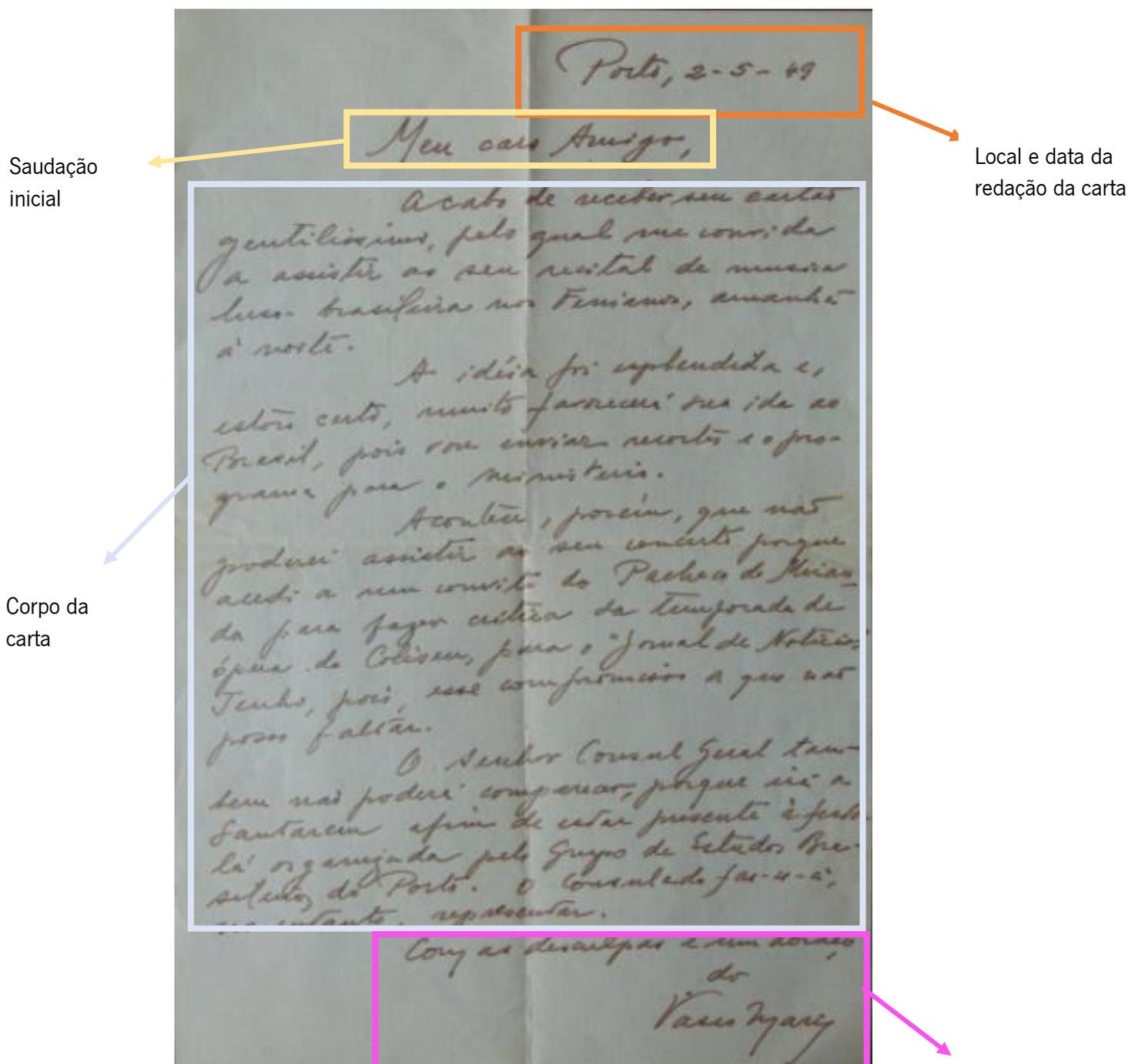


Figura 12: Corpo da carta n.º 155

Despedida e assinatura do remetente

Como é possível observar na figura 13, a carta nº 155 inicia-se com o local e a data da redação da mesma, seguida de uma breve saudação inicial que pode tomar a forma de “Meu caro...”, “Meu querido...”, “Caro...”, entre outros.

Segue-se então o corpo da carta que corresponde à mensagem em si, aquilo que o remetente quer transmitir ao destinatário. Neste exemplo, o remetente agradece o convite que lhe foi endereçado para o recital de música luso-brasileira, lamentando o facto de não poder assistir.

No final da carta, dá-se o momento de despedida onde normalmente o remetente envia os seus cumprimentos tanto ao destinatário como aos familiares do mesmo e deixa uma assinatura.

Feito o reconhecimento dos elementos constituintes da carta, segue-se no próximo capítulo a descrição do processamento eletrónico do espólio e as dificuldades encontradas bem como a solução considerada mais favorável para cada uma dessas dificuldades. Também no próximo capítulo será relatado o processo de anotação com a metalinguagem de informação XML, com especial destaque para as diretrizes da TEI e as propostas do projeto DALF. Importa referir que a anotação descritiva pressupõe, necessariamente, uma fase de análise detalhada da estrutura e do conteúdo do tipo de documento em questão.

Capítulo 4 – Anotação da correspondência

4.1 Text Encoding Initiative e DALF

No processo de transcrição e anotação é necessário adotar normas que tornem o documento transcrito um documento válido e bem formado. Assim, nos processos de transcrição e anotação eletrónica da correspondência do espólio de Eurico Thomaz de Lima foram utilizadas as diretrizes da Iniciativa de Codificação Textual TEI, uma vez que esta norma é a referência incontornável no âmbito da preservação digital de material nas áreas das Ciências Humanas. Paralelamente, foram também utilizadas normas do projeto Digital Archive of Letters in Flanders (DALF), uma vez que estas normas oferecem propostas para elementos do espólio como envelopes e parafernália que serão abordados posteriormente.

O propósito da TEI é providenciar normas para a formação e gestão, em formato digital, de todos os tipos de documentos tais como manuscritos, documentos arquivísticos e muitos outros (Burnard, L. 2014).²

A TEI segue a sintaxe da linguagem de anotação XML, propondo etiquetas que visam descrever a estrutura de um documento eletrónico e os seus elementos, sendo então reconhecido como um padrão internacional utilizado para representar um leque variado de tipologias de texto, tais como poesia, recursos lexicográficos, manuscritos, textos orais, textos performativos, entre outras tipologias. A TEI pretende assim ser um formato normalizado para troca de informação, fornecer linhas orientadoras para apresentar os textos em formato digital e ser utilizada para representar todos os géneros de texto (Guerreiro, 2014).

Por outro lado, a TEI não oferece propostas de etiquetagem de certos elementos que são importantes para este projeto como, por exemplo, a ocorrência de envelopes e de parafernália que acompanham as cartas. Assim, foi necessário recorrer a outras propostas que cobrissem esta necessidade de anotação, tendo-se adotado as propostas de anotação do projeto Digital Archive of Letters in Flanders (DALF).³

² As normas da *Text Encoding Initiative (TEI)* podem ser consultadas no site oficial em <https://tei-c.org/>

³ As normas do projeto *Digital Archive of Letters in Flanders* podem ser consultadas no site <https://ctb.kantl.be/project/dalf/dalfdoc/index.html>

"DALF is an acronym for "Digital Archive of Letters in Flanders". It is envisioned as a growing textbase of correspondence material which can generate different products for both academia and a wider audience, and thus provide a tool for diverse research disciplines ranging from literary criticism to historical, diachronic, synchronic, and sociolinguistic research. The input of this text base will consist of the materials produced in separate electronic edition projects. The DALF project can be expected to stimulate new electronic edition projects, as well as the international debate on electronic editions of manuscripts. In order to ensure maximum flexibility and (re)usability of each of the electronic DALF editions, a formal framework is required that can guarantee uniform integration of new projects in the DALF project. Therefore, the project is from the start aimed at adherence to international standards for electronic text encoding." (Vanhoutte, & Van den Branden, 2003)

Neste processo, a metalinguagem de anotação utilizada foi a *Extensible Markup Language (XML)*. Esta linguagem foi escolhida, uma vez que possibilita a anotação de partes constituintes de um texto de acordo com a função semântica que as mesmas desempenham na estrutura e no contexto textual (Guilherme et al., 2006, p.113).

4.2 Processo de anotação da secção <text>

Depois de toda a análise feita ao espólio e à correspondência do mesmo, iniciou-se então o processo de anotação utilizando, como já foi mencionado anteriormente, as normas de transcrição propostas pela TEI e também pelo projeto DALF. A TEI é composta por blocos de informação principais: o cabeçalho <teiHeader> que integra todos os dados bibliográficos referentes ao documento original e à versão digital e o texto <text> que contém a carta em si, incluindo as anotações descritivas.

A secção <text>, a ser tratada neste capítulo, aborda todo o texto que está a ser transcrito na carta. Esta secção é composta por subpartes, nomeadamente, a secção inicial da carta etiquetada com <opener>, sequência de parágrafos <p> e a secção final da carta etiquetada com <closer>. Uma vez que a TEI não propõe formas de anotação para a saudação inicial e final da carta, foram então utilizadas as normas do projeto DALF.

Iniciando o processo de anotação, surge o elemento <opener> que engloba todos os elementos indicativos de início da carta, tais como, a data, o local da redação da mesma e a saudação. No entanto, nem todas as cartas se fazem acompanhar deste elemento, uma vez que nem todos os remetentes seguem a mesma ordem no momento de escrita. Assim, a etiqueta

<opener> pode conter as seguintes informações: <date> para a anotação da data; <placeName> para a localização da redação da carta e <salute> para a saudação inicial.

```
<opener>
  <date when="28_12_1956">28/XII/1956</date>
  <abbr expand="Excelentissimo Senhor">Exmo. Snr</abbr>
  <occupation> <abbr
expand="Professor">Prof.</abbr></occupation>
  <persName>Eurico Tomás de Lima</persName>
</opener>
```

O elemento <closer> contém todos os elementos indicativos de despedida, nomeadamente a saudação final, a assinatura e a data. Novamente, nem todas as cartas se fazem acompanhar destes elementos e ainda existem outros elementos que podem ser facultativos como é o caso do *postscript* - <ps>.

```
<closer>

<p>
  Os mais cordiais e gratos cumprimentos de <lb/>
  <signed>
    <persName>Berta Alves de Sousa</persName>
  </signed>
</p>
</closer>
```

Seguidamente, e depois de analisadas as partes mais importantes da secção <text>, surgem os elementos utilizados para a anotação da estrutura dos documentos, tais como os parágrafos - <p>; o início da página <pb/>; as quebras de linha <lb/>; *postscript* <ps> e a anotação do conteúdo, como variantes ortográficas, formas variantes e abreviaturas <choice>.

Numa edição eletrónica de um documento, o objetivo é ser o mais fiel possível ao documento original e documentar a estrutura do mesmo conforme esta se apresenta. Assim, existem elementos que se podem fazer acompanhar de atributos qualificativos e de informação adicional como, por exemplo, a indicação do número de início de página. Desta forma, ao iniciar a anotação da carta faz sentido indicar qual a página do documento original que está a ser anotada. O elemento <pb/> faz-se acompanhar de um atributo *n* (*number*), por exemplo, <pb *n*="2"/>. Quando uma página termina e outra se inicia é necessário anotar essa mudança inserindo <pb *n*="3"/>, e assim sucessivamente até ao final da carta. O valor entre aspas corresponde ao número da página do documento original.

```
<pb n="2"/>
```

```
<p> a bom recato. O soneto da<persName>Florbela <lb/> Espanca</persName> -  
<occupation>poetisa</occupation> que residiu <lb/> em  
<placeName>Matosinhos</placeName> - vai dar uma <lb/> .... de grande interesse ao ....  
<lb/> trabalho pois, creio, é o único <lb/> exemplo indica que existe na <lb/> obra poética  
desta grande alentejana.
```

As quebras de linha representadas por <lb/> estão inseridas no elemento <p> - parágrafo. Por norma, o elemento <lb/> deve aparecer no momento em que uma nova linha começa.

Tendo em conta a altura da escrita das cartas, é necessário ter em atenção as mudanças ortográficas que foram decorrendo ao longo dos anos. Assim, e para anotar essa diferença, recorre-se à tag <choice>. Esta etiqueta permite registar a forma original da palavra (<orig>) e a forma corrigida ou regularizada (<reg>) de acordo com a ortografia contemporânea. Desta forma, é sempre possível extrair informação a partir da forma ortográfica atual e facultar ao investigador acesso às duas versões. Veja-se o exemplo abaixo que demonstra a utilização da etiqueta <choice> no caso da palavra “expontaneidade”.

```
carta tão viva de emoção e
```

```
<choice><orig>expontaneidade</orig><reg>espontaneidade</reg></choice>
```

```
<choice><orig>actuar</orig><reg>atuar</reg></choice>
```

Uma das características da correspondência do espólio é a existência de bastantes abreviaturas tendo-se optado por expandir as mesmas de forma a proporcionar maior facilidade na leitura do documento. Recorreu-se à etiqueta `<abbr>` que é a norma proposta pela TEI para este caso que é acompanhada pelo atributo `expand`

```
<abbr expand="Excelentíssimo">Exmo</abbr>
```

Nas abreviaturas, as letras do final da palavra podem aparecer em sobrescrito, sendo estas anotadas com o elemento `<sup>`

```
<abbr expand="Excelentíssimo">Ex<sup>mo</sup></abbr> Senhor <lb/>
```

Foram também anotadas entidades mencionadas nas cartas como o nome próprio de artistas (`<persName>`), lugares (`<placeName>`) e instituições (`<institution>`). Uma vez que também se faz referência ao título de obras do compositor e a nomes de jornais e revistas, foi criada uma etiqueta própria do projeto visto não existirem outras propostas por parte das normas já referidas. Assim, o título das obras, jornais e revistas são anotados com a etiqueta `<work>` que se faz acompanhar do atributo `type`. É este atributo que classifica a categoria do que queremos anotar.

O elemento `<work>` pode aparecer na transcrição das cartas como:

```
<work type="newspaper ">
```

```
<work type="newspaper">Jornal de Viseu </work>
```

```
<work type="music ">
```

```
<work type="music">Toccata</work>
```

```
<work type="poem ">
```

```
<work type="poem">Mors-Amor</work>
```

A etiquetagem semântica do microtexto desempenha um papel importante para o conhecimento do espólio, permitindo explorar novos caminhos de pesquisa. Importa referir que o projeto de criação de uma ontologia a partir da correspondência de Eurico Thomaz de Lima, desenvolvido por Tânia Rego, depende da existência dos documentos etiquetados ao nível do microtexto.

Capítulo 5: TEI Header e DTD

5.1 Metainformação: descrição da TeiHeader

O cabeçalho TEI contém metadados que são valiosos e necessários para a anotação das cartas. Este cabeçalho contém documentação de descrições bibliográficas sobre o texto anotado (informação obrigatória), decisões editoriais tomadas no processo de etiquetação (informação recomendável), descrição detalhada de elementos textuais de carácter não-bibliográfico relevantes e um registo das alterações realizadas no documento eletrónico. Assim, de acordo com as diretrizes, o cabeçalho TEI divide-se em quatro secções: <fileDesc>; <encodingDesc>; <profileDesc> e <revisionDesc>.

```
<teiHeader>  
  
  <fileDesc>  
  
    (..)  
  
  </fileDesc>  
  
  <encodingDesc>  
  
    (...)  
  
  </encodingDesc>  
  
  <profileDesc>  
  
    (...)  
  
  </profileDesc>  
  
  <revisionDesc>  
  
    (...)  
  
  </revisionDesc>  
  
</teiHeader>
```

Das quatro secções propostas pela TEI, apenas a primeira secção <fileDesc> é obrigatória, sendo esta a secção de foco que vai ser abordada neste capítulo. O <fileDesc> abarca a descrição bibliográfica do documento eletrónico e do documento original, e pode ser dividido em subsecções, nomeadamente, <titleStmt>, <publicationStmt> e <sourceDesc>.

```
<fileDesc>

    <titleStmt>...</titleStmt>

    <publicationStmt>...</publicationStmt>

    <sourceDesc>...</sourceDesc>

</fileDesc>
```

Na subsecção <titleStmt> é importante incluir o título (<title>) atribuído ao ficheiro eletrónico de forma a diferenciar o documento eletrónico da carta do documento original. Segue-se o autor da carta (<author>) e, por último, a indicação dos colaboradores responsáveis pelas várias fases de processamento do documento eletrónico (<respStmt>), nomeadamente as fases de transcrição das cartas para formato eletrónico e de anotação estrutural e semântica dos documentos e a data de realização das respetivas tarefas. Esta informação permite manter o histórico do processamento efetuado. Tomemos como exemplo o esquema de anotação da carta n.º 196:

```
<fileDesc>

    <titleStmt>

        <title>Carta de Evaristo de Campos Coelho a Eurico Thomaz de Lima:
versão eletrónica</title>

    <author>Evaristo de Campos Coelho</author>

    <respStmt>

        <resp>Transcrição efetuada por</resp>

        <name>Helena Faria</name>

        <date>21/07/2021</date>

    </respStmt>

    <respStmt>
```

```

        <resp>Anotação conforme as diretrizes TEI e do DALF Project- Digital Archive
of Letters in Flanders</resp>
        <name>Helena Faria</name>
        <date>21/07/2021</date>
    </respStmt>
</titleStmt>
</fileDesc>

```

Na subsecção <publicationStmt> encontra-se informação sobre as autoridades responsáveis, por um lado, pelo documento original, neste caso a Professora Elisa Lessa, curadora do espólio físico no Edifício dos Congregados, por outro, o Professor José João Almeida, responsável pelo espólio eletrónico. Este projeto trata-se de um projeto conjunto entre o Departamento de Música da Universidade do Minho, o Departamento de Informática e o Centro de Estudos Humanísticos da Universidade do Minho como é referido no exemplo abaixo. Para fins de investigação e de acesso ao espólio, tanto físico como eletrónico, é importante especificar os direitos associados à utilização dos documentos através da anotação <availability>.

```

<publicationStmt>

  <authority>
    <p>Departamento de Música-Universidade do Minho</p>
    <p> Departamento de Informática </p>
    <p> Centro de Estudos Humanísticos da Universidade do Minho</p>
  </authority>
  <idno>C196</idno>
  <availability status="restricted">
    <p>*contacto da professora Elisa Lessa para repositório físico*</p>
    <p>*contacto do professor José João para repositório digital*</p>
  </availability>
</publicationStmt>

```

Passando para a secção <sourceDesc>, este elemento abarca toda a informação referente ao documento original tendo como referência as anotações propostas pelo projeto DALF. Esta

secção contém, assim, um outro elemento designado de <letDesc> (*letter description*) que, por sua vez, pode ser dividido em quatro partes fulcrais para o processamento, sendo elas, <letIdentifier> (letter identifier), <letHeading> (letter heading), <physDesc> (physical description), <envOcc> (envelope occurrence) e <letContents> (letter contents). O <letDesc> contém informação sobre o documento original, tal como a localização física dos documentos e número de identificação anotado pelo elemento <letIdentifier>; o autor; a data, destinatário e data do documento anotado pelo elemento <letHeading>; qual o tipo de documento a ser anotado e quantidade de páginas e dimensões do mesmo que corresponde ao elemento <physDesc>, ocorrência de envelope através do elemento <envOcc> e, ainda, um resumo do documento anotado pelo elemento <letContents>. Com base neste resumo, o utilizador poderá decidir se pretende consultar e/ou aceder ao documento em questão. Vejamos abaixo o modelo de anotação proposto para a secção <sourceDesc>, tomando como exemplo a carta nº 196, de Evaristo de Campos Coelho, datada a 5 de janeiro de 1968.

```
<sourceDesc>

  <letDesc>
    <letIdentifier>
      <country>Portugal</country>
      <region>Minho</region>
      <settlement>Braga: Edifício dos Congregados Universidade do Minho</settlement>
      <collection>Espólio Eurico Thomaz de Lima</collection>
      <idno>C196</idno>
    </letIdentifier>
    <letHeading>
      <author attested="yes">Evaristo de Campos Coelho</author>
      <addressee attested="yes">Eurico Thomaz de Lima</addressee>
      <placeLet attested="yes"></placeLet>
      <dateLet attested="yes">05_01_1967</dateLet>
    </letHeading>
    <physDesc>
      <type>Bilhete Postal</type>
      <support>1 página</support>
      <extent>
```

```

        <dimensions>
            <height units="mm">148</height>
            <width units="mm">105</width>
        </dimensions>
    </extent>
</physDesc>
<envOcc occ="yes" />
<letContents>
    <p>Refere-se à atuação de Eurico Thomaz de Lima transmitida na Emissora
Nacional </p>
    <note>Etiqueta facultative que inclui informação adicional ao resumo do conteúdo da
carta</note>
</letContents>
</letDesc>
</sourceDesc>

```

Uma outra proposta do projeto DALF, essencial para a realização deste projeto, consiste na anotação da parafernália uma vez que, contém informação adicional e relevante para as demais ligações.

Segundo o projeto DALF, o termo ‘parafernália’ designa todos os elementos que não são originados no momento de redação da carta, são enviados em conjunto com a mesma. Tendo em conta a riqueza do espólio epistolográfico de Eurico Thomaz de Lima no que diz respeito a elementos que acompanham as cartas, propõe-se, no âmbito deste projeto, um modelo de etiquetagem para estes documentos. Foi possível identificar os seguintes tipos de documentos que funcionam como parafernália:

- i. Partituras;
- ii. Recortes de jornais;
- iii. Poemas;
- iv. Pagelas;
- v. Fotografias;
- vi. Dedicatórias;

- vii. Postais festivos;
- viii. Programas de concerto e musicais;
- ix. Entre outros.

Segue-se abaixo vários exemplos da parafernália encontrada ao longo da análise do espólio.

Na carta n.º 56, enviada pela soprano-ligeiro, Alma Cunha de Miranda a Eurico Thomaz de Lima, encontra-se anexado um recorte do jornal “Voz de Portugal”, datado a 2-9-1960, onde se faz menção a um Recital de Folclore no Liceu Literário Português, do qual Eurico Thomaz de Lima era sócio honorário. Este recital foi realizado aquando do 92.º aniversário da instituição. O recital deu-se por volta das 21 horas e foram interpretadas algumas obras das quais se destacam, da primeira parte, “Minha mãe me deu um lenço” de Alexandre Rey Colaço, “Eterna Canção” de Júlio Dantas e António Vianna e “Marianita” de Eurico Thomaz de Lima. Da segunda parte destacam-se obras como “Rosa dos Ventos”, “Rolinha” e “Poema do Negro”. Também no recorte, a autora fez uns comentários sobre o artigo onde reforça que se esqueceram de mencionar o belíssimo pianista, Murillo dos Santos, que se encontra na imagem ao lado da Soprano.



Figura 13: Recorte de jornal enviado juntamente com a carta n.º 60

Passando para a carta n. °60, enviada também por Alma Cunha de Miranda a Eurico Thomaz de Lima, a parafernália assume a forma de uma pagela de Nossa Senhora da Glória do Outeiro, venerada no Rio de Janeiro, Brasil. No verso da pagela, encontra-se uma oração e uma mensagem em que a remetente deseja que a Santa proteja a saúde do compositor e o seu lar.



Figura 14: Exemplo de pagela e oração

Recorrendo ao exemplo da carta n.º.67, enviada a 15-3-1956, o remetente volta a ser a soprano Alma Cunha de Miranda que envia a Eurico Thomaz de Lima uma partitura de um compositor brasileiro, Arnaldo Rebello, que a acompanhou nas suas atuações. A partitura tem como título “Chôro de Oitavas” e está classificada para o 8º ano do programa da Escola Nacional de Música. Envolvendo a partitura encontra-se ainda uma dedicatória de Rebello a Eurico Thomaz de Lima. Este tipo de parafernália demonstra uma elevada troca de impressões, ideias e documentos musicais entre colegas e amigos do compositor o que torna ainda mais rica a informação extraída da mesma.

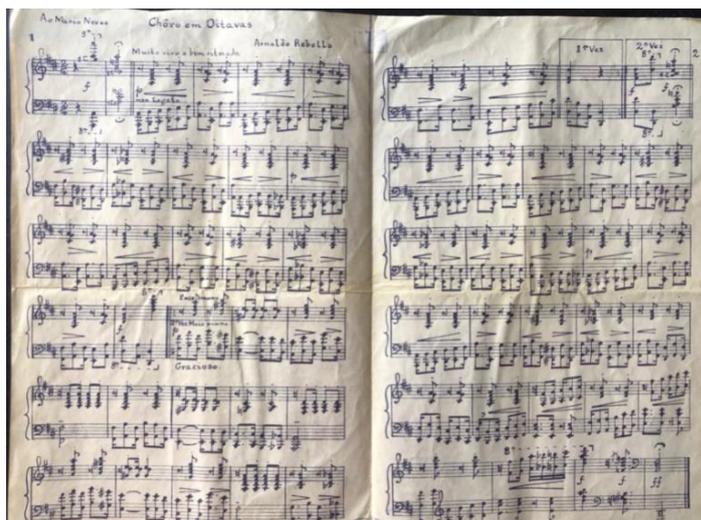


Figura 15: Partitura "Chôro em Oitavas" de Arnaldo Rebello

Este cabeçalho segue a mesma estrutura para todas as cartas, mas a informação anotada é diferente conforme as necessidades de cada documento, uma vez que, todas as cartas são únicas e podem conter elementos diferentes das outras.

5.2 Processo de validação do documento XML

O processo de validação dos documentos epistolográficos etiquetados de acordo com as normas XML-TEI pressupõe a criação de um documento DTD cuja funcionalidade é declarar as etiquetas e os atributos que devem ser utilizados para anotar a estrutura e o conteúdo das cartas, a ordem em que esses elementos deverão aparecer e o seu estatuto de elementos e atributos facultativos ou obrigatórios. O modelo DTD assume particular relevância para o processamento dos documentos XML em termos de localização e extração de informação e de integração de novos colaboradores que venham contribuir para a etiquetagem dos documentos do espólio

Tomando como exemplo a carta n.º152, de Fernando Lopes Graça a Eurico Thomaz de Lima, a 14 de novembro de 1948, segue-se uma análise por blocos dos elementos que compõem o DTD, para assim ser mais fácil compreender todo o processo de validação dos documentos eletrónicos. Em primeiro lugar, focar-se-á a metainformação incluída na secção TeiHeader do documento XML, passando, de seguida, a abordar as propostas de etiquetagem para os sobrescritos e, por último, as propostas para a estrutura e o conteúdo das cartas.

Tal como referido no Capítulo 5 e como é possível ver nas declarações da Figura 16, o documento XML é composto por duas secções principais, a secção <teiHeader> e a secção <text>.

O cabeçalho <teiHeader> é composto pela secção <fileDesc>, que, por sua vez, contém 3 subsecções obrigatórias, nomeadamente, <titleStmt>, <publicationStmt> e <sourceDesc>.

```
<!ELEMENT TEI.2 (teiHeader,text)>
<!ELEMENT teiHeader (fileDesc)>
<!ELEMENT fileDesc (titleStmt,publicationStmt,sourceDesc)>
```

Figura 16: Parte do DTD - elemento raiz

A primeira declaração do documento DTD acima explicita que o elemento raiz do documento XML é a etiqueta <TEI.2>. A vírgula nas declarações DTD funciona como operador de sequência, definindo, assim, a ordem obrigatória dos elementos declarados.

Como representado na Figura 17, a secção <titleStmt> começa com o título atribuído ao ficheiro eletrónico <title>, com menção específica de que se trata da versão eletrónica da carta transcrita e anotada. Segue-se a identificação do autor da carta por via da etiqueta <author>. Por último, a referência às tarefas desenvolvidas, os respetivos responsáveis e as datas de execução das mesmas.

```
<titleStmt>
  <title>Carta de Fernando Lopes Graça a Eurico Thomaz de Lima:
  versão eletrónica</title>
  <author>Fernando Lopes Graça</author>
  <respStmt>
    <resp>Transcrição eletrónica efetuada por</resp>
    <name>Helena Faria</name>
    <date>22/04/2021</date>
  </respStmt>
  <respStmt>
    <resp>Anotação conforme as diretrizes TEI-XML e do
    projeto DALF - Digital Archive of Letters in Flanders</resp>
    <name>Helena Faria</name>
    <date>22/04/2021</date>
  </respStmt>
</titleStmt>
```

Figura 17: Secção <titleStmt>

A declaração DTD abaixo, correspondente ao XML da Figura 17, define o conteúdo dos elementos <title>, <author>, <resp> como sendo do tipo #PCDATA (*Parsed Character Data*), ou seja, texto que será validado. Na Figura 18, o conteúdo dos elementos <name> e <date> é definido como sendo conteúdo misto, conteúdo este composto por texto intercalado com elementos vários. Embora os elementos <name> e <date> da secção <respStmt> sejam compostos apenas por conteúdo textual (#PCDATA) sem intercalação de outros elementos, o facto destes elementos

ocorrerem no próprio conteúdo da carta intercalados pelos elementos <choice>, <abbr> (abbreviation), <lb/> (line break), <pc> (*punctuation*) e apenas ser possível definir cada elemento uma única vez no DTD, estas declarações validam todas as ocorrências destes elementos no documento XML.

```
<!ELEMENT titleStmt (title,author,respStmt)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT respStmt (resp,name,date)>
<!ELEMENT resp (#PCDATA)>
<!ELEMENT name (#PCDATA|choice|abbr)*>
<!ELEMENT date (#PCDATA|pc|lb|abbr|choice)*>
```

Figura 18: Validação da secção <titleStmt>

Na declaração de conteúdo misto, os elementos intercalados são separados pela barra vertical (|) e o asterisco indica que os elementos são facultativos, uma vez que podem estar ausentes ou ocorrerem uma ou mais vezes.

Na secção do <publicationStmt> são anotadas as entidades responsáveis pelo espólio físico e arquivo digital, contactos dos mesmos, o número identificativo da carta e os termos de utilização do documento <availability>

```
<publicationStmt>
  <authority>
    <p>Departamento de Música-Universidade do Minho</p>
    <p>Departamento de Informática e Centro de Estudos
Humanísticos-Universidade do Minho</p>
  </authority>
  <idno>C152</idno>
  <availability status="restricted">
    <p>*contacto da professora Elisa Lessa para repositório
físico*</p>
    <p>*contacto do professor José João para repositório
digital*</p>
  </availability>
</publicationStmt>
```

Figura 19: Secção <publicationStmt>

No DTD correspondente ao XML da Figura 19, define-se o conteúdo da secção <availability> como uma sequência de parágrafos <p>. A declaração do elemento <p> na Figura 20 representa o conteúdo de todas as ocorrências deste elemento nos documentos do espólio já etiquetados. Intercalado com o texto, os parágrafos podem conter entidades nomeadas, como por

exemplo, nomes de pessoas, nomes de instituições, nomes de locais, nomes de obras, datas, moradas, profissões, entre outras, abreviaturas, listas.

```
<!ELEMENT publicationStmt (authority,idno,availability)>
<!ELEMENT authority (p+)>
<!ELEMENT p
(#PCDATA|choice|lb|work|persName|placeName|institution|date|occupation|pc
|pb|signed|abbr|num|list|quote|time|address|ref|sup| addrLine|date|p)*>
<!ELEMENT idno (#PCDATA)>
<!ELEMENT availability (p+)>
<!ATTLIST availability status NMTOKEN #REQUIRED>
```

Figura 20: Validação da secção <publicationStmt>

Para além dos elementos, é obrigatório declarar os atributos num documento DTD, utilizando a palavra-chave ATTLIST (*attribute list*). No documento XML, o elemento <availability> é complementado pelo atributo 'status' que especifica os termos de utilização do documento para fins de investigação. É necessário definir se o atributo é obrigatório '#REQUIRED' ou facultativo '#IMPLIED'.

Como representado na 21, o documento original é descrito na secção <letDesc>, composta por um primeiro bloco de informação <letIdentifier> que documenta a localização do documento original em termos geográficos e no âmbito de determinada coleção. O segundo bloco <letHeading> contém informações relevantes para a catalogação do documento epistolográfico, nomeadamente, o autor, o destinatário, o local e a data.

```
<sourceDesc>
  <letDesc>
    <letIdentifier>
      <country>Portugal</country>
      <region>Minho</region>
      <settlement>Braga: Edifício dos Congregados
Universidade do Minho</settlement>
      <collection>Espólio Eurico Thomaz de
Lima</collection>
      <idno>C152</idno>
    </letIdentifier>
    <letHeading>
      <author attested="yes">Fernando Lopes Graça</author>
      <addressee attested="yes">Eurico Thomaz de
Lima</addressee>
      <placeLet attested="yes">Lisboa</placeLet>
      <dateLet attested="yes">14_11_1948</dateLet>
    </letHeading>
```

Figura 21: Secção <sourceDesc>

Os elementos da secção <letHeading> são complementados pelo atributo 'attested' que permite ao anotador especificar se a informação se encontra de forma explícita no documento em si. Se for o caso, o valor do atributo 'attested' é 'yes', conferindo aos dados um alto grau de confiabilidade. Se a informação for obtida por outras vias, o valor do atributo 'attested' é 'no'.

```
<!ELEMENT sourceDesc (letDesc)>
<!ELEMENT letDesc (letIdentifier,letHeading,physDesc,envOcc,letPart*,
letContents)>
<!ELEMENT letIdentifier (country,region,settlement,collection,idno)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT region (#PCDATA)>
<!ELEMENT settlement (#PCDATA|choice|abbr)*>
<!ELEMENT collection (#PCDATA)>
<!ELEMENT letHeading (author,addressee,placeLet,dateLet*)>
<!ELEMENT addressee (#PCDATA)>
<!ATTLIST addressee attested NMTOKEN #REQUIRED>
<!ELEMENT placeLet (#PCDATA)>
<!ATTLIST placeLet attested NMTOKEN #REQUIRED>
<!ELEMENT dateLet (#PCDATA)>
<!ATTLIST dateLet attested NMTOKEN #REQUIRED>
```

Figura 22: Validação da secção <sourceDesc>

Na Figura 22, a declaração do elemento <letDesc> especifica todos os blocos de informação que compõem a secção <letDesc> no documento XML. De seguida, abordar-se-á os restantes blocos, nomeadamente, <physDesc>, <envOcc> e <letContents>.

```
<physDesc>
  <type>carta</type>
  <support>1 página</support>
  <extent>
    <dimensions>
      <height units="mm">214</height>
      <width units="mm">274</width>
    </dimensions>
  </extent>
</physDesc>
<envOcc occ="yes" />
<letContents>
  <p>Carta a questionar Eurico se quer aderir à
Federação Internacional dos Compositores e Musicólogos Progressistas</p>
  <note>Caso deseje, deve informar o mais rápido
possível</note>
</letContents>
</letDesc>
</sourceDesc>
```

Figura 23: Secção <physDesc>

Uma vez que a carta n.º152 não contém parafernália, o bloco de informação respetivo não está anotado na secção <physDesc> do documento XML. No entanto, o documento DTD deverá definir a possibilidade da ocorrência destes elementos. Na Figura 24, a declaração do elemento <physDesc> define o tipo (<type>) de documento epistológico (carta, bilhete postal, cartão de visita), seguido de informação sobre a paginação, seguido de dois elementos facultativos, propriedade indicada pelo asterisco que funciona como operador de frequência, nomeadamente, a extensão do documento original (<extent>) no que respeita as suas dimensões e os elementos de parafernália.

```

<!ELEMENT physDesc (type,support,extent*,paraphernalia*)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT support (#PCDATA)>
<!ELEMENT extent (dimensions)>
<!ELEMENT dimensions (height,width)>
<!ELEMENT height (#PCDATA)>
<!ATTLIST height units NMTOKEN #REQUIRED>
<!ELEMENT width (#PCDATA)>
<!ATTLIST width units NMTOKEN #REQUIRED>
<!ELEMENT paraphernalia (paraphList)>
<!ELEMENT paraphList (paraphItem)>
<!ELEMENT paraphItem (paraphDesc)>
<!ATTLIST paraphItem id CDATA #REQUIRED>
<!ELEMENT paraphDesc (p)>
<!ELEMENT envOcc EMPTY>
<!ATTLIST envOcc occ NMTOKEN #REQUIRED>
<!ELEMENT letPart (letHeading)+>
<!ELEMENT letContents (p*,note*)>
<!ELEMENT note (#PCDATA|p|work)*>

```

Figura 24: Validação da secção <physDesc>

O elemento <envOcc> (*envelope occurrence*) encontra-se definido no modelo DTD como elemento vazio (*EMPTY*) por ser um elemento que não abarca conteúdo, indicando a (in)existência de um sobrescrito. Se o valor do atributo 'occ' for 'yes', somos remetidos para a respetiva informação na secção <text> do documento XML, como representada na Figura 25.

```

<text>
  <envelope>
    <envPart side="front">
      <address type="receiver">
        <addrLine><abbr expand="Excelentíssimo">Exmo</abbr>
        Senhor <lb/>
          <persName>Eurico Thomaz de
        Lima</persName></addrLine><lb/>
          <addrLine>Rua Egas Moniz,6-1º</addrLine><lb/>

```

```

        <addrLine><placeName>Porto</placeName></addrLine>
    </address>
</envPart>
</envelope>

```

Figura 25: Secção <text>, subsecção <envelope>

A representação XML da informação nos sobrescritos segue a proposta de etiquetação do projeto DALF. Existe a distinção entre a parte da frente e do verso do envelope por via do elemento <envPart> (*envelope part*) e do atributo 'side' que admite os valores 'front' e 'back'. Por norma, cada parte é composta pelo elemento <address> complementado pelo atributo 'type' utilizado para diferenciar o remetente (*sender*) do destinatário (*receiver*). O sobrescrito da carta n.º 152 não contém informação no verso. Os sobrescritos do espólio da correspondência de Eurico Thomaz de Lima apresentam outros elementos de interesse para fins de investigação em várias áreas, tais como, estudos filatélicos e histórico-sociais.

A título de exemplo e recorrendo à carta n.º 100, enviada por Olímpia Doria a Eurico Thomaz de Lima, a 16-4-1936, realça-se o selo que consta do envelope que contém a frase "Tudo Pela Nação", uma frase utilizada por António de Oliveira Salazar durante o Estado Novo.



Figura 26: Selo "Tudo pela Nação" - frase de António Oliveira de Salazar

Dada a importância da correspondência na vida das pessoas ao longo do século XX, os sobrescritos eram utilizados pelo Estado como meio de comunicar com a população.

```

<!ELEMENT envelope (envPart+)>
<!ELEMENT envPart (address)*>
<!ATTLIST envPart side NMTOKEN #REQUIRED>
<!ELEMENT address
(#PCDATA|addrLine|abbr|lb|name|institution|street|settlement|num|sup)*>
<!ELEMENT street (#PCDATA|num|lb|abbr|placeName)*>
<!ATTLIST address type NMTOKEN #IMPLIED>

```

```

<!ELEMENT addrLine
(#PCDATA|occupation|lb|abbr|persName|num|settlement|choice|institution|street|signed|date|work|placeName)*>
<!ELEMENT occupation (#PCDATA|choice|abbr|lb|pc)*>
<!ELEMENT lb EMPTY>
<!ELEMENT abbr (#PCDATA|sup)*>
<!ATTLIST abbr expand CDATA #REQUIRED>
<!ELEMENT sup (#PCDATA|lb|choice|num)*>
<!ELEMENT persName (#PCDATA|lb|pc|choice|abbr)*>

```

Figura 27: Validação da secção <envelope>

Como é visível das declarações DTD representadas na Figura 27, para além das informações que por norma aparecem na morada, este projeto propõe anotar também abreviaturas (<abbr>) e nomes de instituições (<institution>).

Como descrito no Subcapítulo 4.2, a secção <body> do documento XML é composta por três subsecções principais: <opener>, um ou mais parágrafos (<p>) e <closer>. Adicionalmente o início de cada página é marcado pela etiqueta <pb/> (*page beginning*), complementada com o atributo 'n' para indicar o número da página. Importa referir que é obrigatório utilizar o atributo 'when' com o elemento <date> com o objetivo de fornecer uma forma regularizada das datas que aparecem das mais variadas formas nos documentos. Esta proposta facilitará a extração de informação a partir de datas.

```

<body>
  <pb n="1"/>

  <opener>
    <placeName>Lisboa</placeName>
    <date when="14_11_1948">14 de Novembro de 1948</date>
    <salute>
      <p>Meu caro Amigo:</p>
    </salute>
  </opener>
  <p>
    Tem esta por fim inquirir de si se desejará dar a <lb/> sua
    adesão à <choice><orig>projectada</orig><reg>projetada</reg></choice>
    <institution>Federação Internacional dos Compositores e Musicólogos
    Progressistas</institution>, cuja criação está prevista nos <lb/> termos
    consignados no <work type="manifesto">Manifesto do <event>II Congresso
    Internacional <lb/> de Compositores e Críticos Musicais</event></work>,
    de <placeName>Praga</placeName>, publicado no <lb/> nº59 da <work
    type="journal">Vértice</work>, de <date when="-_07_1949">Julho deste
    ano</date>
  </p>

```

```

    <p> Em caso afirmativo, queira comunicar-me a sua adesão <lb/>
    com a urgência possível, para incluir o seu nome entre os dos <lb/>
    outros aderentes protugueses da <institution>Federação</institution>
    </p>

    <closer>
        <p>
            Os melhores cumprimentos do <lb/>

            <signed>
                <persName>Fernando Lopes Graça</persName>
            </signed>
        </p>
    </closer>
</body>

```

Figura 28: Secção <body>

Embora a carta n.º152 não apresente *post scriptum*, a declaração do elemento <body> no DTD define a secção facultativa <postscript>. A declaração do elemento <p> é abrangente, definido como conteúdo misto, para abarcar todas as entidades nomeadas e propriedades relevantes para o processamento eletrónico. Com base na carta em análise, refira-se, em particular, o elemento <choice> que, verificando-se a existência de palavras cuja ortografia não corresponda à ortografia contemporânea, permite anotar a forma ortográfica original (<orig>) e a forma regularizada (<reg>), possibilitando, assim, as duas formas de pesquisa.

Note-se, ainda, a importância de etiquetagem de instituições (ex. Federação Internacional dos Compositores e Musicólogos Progressistas), as obras (ex. Manifesto do II Congresso Internacional de Compositores e Críticos Musicais) e os eventos (ex. II Congresso Internacional de Compositores e Críticos Musicais) referidos nos documentos, o que permitirá estabelecer uma rede de ligações entre indivíduos (ex. Fernando Lopes Graça e Eurico Thomaz de Lima), as instituições, as obras e os eventos mencionados.

```

<!ELEMENT text (envelope*,body)>
<!ELEMENT body (pb, opener, (p+, pb?, p*, pb?)*,
closer*,p*,postscript*,noteStmt*)>
<!ELEMENT p
(#PCDATA|choice|lb|work|persName|institution|date|occupation|pc|pb|
placeName|signed|abbr|num|list|quote|time|address|ref|sup|addrLine|
date|p)*>
<!ELEMENT choice (orig,reg)>
<!ELEMENT orig (#PCDATA|lb|pc)*>
<!ELEMENT reg (#PCDATA|lb|pc)*>

```

```

<!ELEMENT pb EMPTY>
<!ATTLIST pb n NMTOKEN #REQUIRED>
<!ELEMENT opener
(#PCDATA|address|lb|placeName|date|salute|occupation|persName|abbr|pc|cho
ice|signed|num|time|p)*>
<!ELEMENT placeName (#PCDATA|lb|choice|abbr|pc)*>
<!ELEMENT salute (p+)>
<!ELEMENT work
(#PCDATA|lb|placeName|persName|pc|abbr|num|sup|date|event)*>
<!ATTLIST work type NMTOKEN #REQUIRED>
<!ELEMENT institution
(#PCDATA|lb|abbr|placeName|pc|name|persName|choice)*>
<!ELEMENT event (#PCDATA|lb)*>
<!ELEMENT pc (#PCDATA)>
<!ELEMENT closer (p+, date?)>
<!ELEMENT signed (persName)>
<!ELEMENT postscript (#PCDATA|label|p|institution)*>
<!ELEMENT noteStmt (note)>

```

Figura 29: Validação da secção <body>

Este modelo DTD foi testado em algumas cartas, sendo que, à medida que as restantes cartas vão sendo etiquetadas, o modelo será adaptado para incluir novas entidades e propriedades relevantes para o processamento. Este modelo representa, assim, um ponto de partida de análise e anotação eletrónica dos documentos que compõem o espólio epistológico de Eurico Thomaz de Lima.

Capítulo 6: Conclusão

Este projeto conta pretende contribuir para a transcrição e anotação da correspondência do espólio epistolográfico de Eurico Thomaz de Lima, através da criação de um modelo XML (Extensible Markup Language) seguindo as normas de anotação da TEI (Text Encoding Initiative) e do projeto DALF (Digital Archive of Letters in Flanders).

Para a realização deste projeto foi necessário proceder a uma análise detalhada do Espólio da correspondência de Eurico Thomaz de Lima a ser trabalhado de forma a compreender os elementos que o compõem, as suas estruturas e o respetivo conteúdo. Esta análise detalhada é pré-requisito para o processamento eletrónico dos documentos em TEI-XML. Durante esta análise, foram encontradas algumas dificuldades, tais como, a elegibilidade de algumas palavras que tornou o processo de transcrição mais demorado, a existência de cartas sem envelope ou de envelopes que continham duas cartas para as quais não existiam normas de anotação. Assim, foi necessário encontrar soluções adequadas para o projeto em questão.

No âmbito deste projeto foi possível criar um modelo de validação DTD que servirá de referência para a validação das restantes cartas que serão no futuro etiquetadas. O trabalho do processamento eletrónico poderá ser utilizado para criar outros derivados, como a ontologia da correspondência de Eurico Thomaz de Lima.

Referências Bibliográficas

BAGGIO, C. C. & FLORES, D. (2013). *Documentos digitais: Preservação e Estratégias*. Revista do Instituto de Ciências Humanas e da Informação.

BALCKY, L., F. (2011). *O Arquivo na Era Digital*. (Dissertação de Mestrado), Universidade Nova de Lisboa.

Burnard, L. 2014. *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*. Marseille: OpenEdition Press. doi:[10.4000/books.oep.426](https://doi.org/10.4000/books.oep.426)

CAMBRAIA, C. N. (2005). *Introdução à Crítica Textual*. São Paulo: Martins Fontes.

CAMBRAIA, C. & MARENGO, S. (2016). *Estudo socioterminológico da variação/mudança em manuscritos militares dos séculos XVIII e XIX*. Interdisciplinar: Revista de Estudos em Língua e Literatura. 11. 203-224.

DALF <https://ctb.kantl.be/project/dalf/dalfdoc/index.html>

Guerreiro, D. (2014, Junho 1). *TEI: Text Encoding Initiative #TEI* Bibliotecas e Humanidades Digitais [TEI: Text Encoding Initiative #TEI – Bibliotecas e humanidades digitais \(hypotheses.org\)](https://hypotheses.org)

GUILHERME, A. R., GOMES, M. & TAVARES, L. (2006). *A interpretação do ilegível e do ininteligível na edição da correspondência privada*, Universidade de Lisboa.

GONÇALVES, C., (2005). *Obras para a infância de Eurico Thomaz de Lima: os duetos para piano*, Universidade do Minho

KELLER, T. & GONÇALVES, A. F. C. (2018). *Edição diplomática e comentários paleográficos de manuscritos do final do século XIX, da cidade de Santa Maria*, Universidade Federal de Santa Maria, Brasil.

LESSA, E. (2007). *Eurico Thomaz de Lima e a Imprensa Brasileira: Um Caso Feliz de Recepção Musical*. Revista Música, 12, 165-174 <https://doi.org/10.11606/rm.v12i0.61765>

SANTOS, H. M. & FLORES, D. (2015) *Estratégias de preservação digital para documentos arquivísticos: uma breve reflexão*. Universidade Federal de Santa Maria, Brasil.

SILVA, C. S. G. (2015). *Transcrição da correspondência dirigida a Camilo Castelo Branco: reflexões sobre princípios e normas de codificação eletrónica* (Relatório de estágio). Retirado do repositório da Universidade do Minho. <http://hdl.handle.net/1822/44299>.

TEI <https://tei-c.org/>

WEEB, C. (2003). *Diretrizes para la preservación del patrimonio digital*.
[10.22201/iibi.0187358xp.2012.56.33014](https://doi.org/10.22201/iibi.0187358xp.2012.56.33014)

Anexo 1

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI.2 SYSTEM "etl.dtd">
<TEI.2>
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Carta de Fernando Lopes Graça a Eurico Thomaz de Lima:
versão eletrónica</title>
      <author>Fernando Lopes Graça</author>
      <respStmt>
        <resp>Transcrição efetuada por</resp>
        <name>Helena Faria</name>
        <date>21/07/2021</date>
      </respStmt>
      <respStmt>
        <resp>Anotação conforme as diretrizes TEI e do DALF
Project- Digital Archive of Letters in Flanders</resp>
        <name>Helena Faria</name>
        <date>21/07/2021</date>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <authority>
        <p>Departamento de Música-Universidade do Minho</p>
        <p>Centro de Estudos Humanísticos-Universidade do
Minho</p>
      </authority>
      <idno></idno>
      <availability status="restricted">
        <p>*contacto da professora Elisa Lessa para repositório
físico*</p>
        <p>*contacto do professor José João para repositório
digital*</p>
      </availability>
    </publicationStmt>
    <sourceDesc>
```

```

<letDesc>
  <letIdentifier>
    <country>Portugal</country>
    <region>Minho</region>
    <settlement>Braga: Campus dos Congregados
Universidade do Minho</settlement>
    <collection>Espólio Eurico Thomaz de
Lima</collection>
    <idno>C152</idno>
  </letIdentifier>
  <letHeading>
    <author attested="yes">Fernando Lopes Graça</author>
    <addressee attested="yes">Eurico Thomaz de
Lima</addressee>
    <placeLet attested="yes">Lisboa</placeLet>
    <dateLet attested="yes"></dateLet>
  </letHeading>
  <physDesc>
    <type>carta</type>
    <support>1 página</support>
    <extent>
      <dimensions>
        <height units="mm">214</height>
        <width units="mm">274</width>
      </dimensions>
    </extent>
    <paraphernalia>
      <paraphList>
        <paraphItem id="obj1 ou part1">
          <paraphDesc>
            <p></p>
          </paraphDesc>
        </paraphItem>
      </paraphList>
    </paraphernalia>
  </physDesc>
  <env0cc occ="yes" />
<letContents>

```

```

        <p>Carta a questionar Eurico se quer aderir à
Federação Internacional dos Compositores e Musicólogos Progressistas</p>
        <note>Caso deseje, deve informar o mais rápido
possível</note>
        </letContents>
        </letDesc>
        </sourceDesc>
        </fileDesc>
</teiHeader>
<text>

        <envelope>
            <envPart side="front">
                <address type="receiver">
                    <addrLine><abbr expand="Excelentíssimo">Exmo</abbr>
Senhor <lb/>
                    <persName>Eurico Thomaz de
Lima</persName></addrLine><lb/>
                    <addrLine>Rua Egas Moniz,6- <num type="ordinal"
value="1">1o</num></addrLine><lb/>
                    <addrLine><placeName>Porto</placeName></addrLine>
                </address>
            </envPart>

        </envelope>

        <body>
            <pb n="1"/>

            <opener>
                <placeName>Lisboa</placeName>
                <date when="14_11_1948">14 de Novembro de 1948</date>
                <salute>
                    <p>Meu caro Amigo:</p>
                </salute>
            </opener>

```

```
<p>
    Tem esta por fim inquirir de si se desejará dar a <lb/> sua
    adesão à <choice><orig>projectada</orig><reg>projetada</reg></choice>
    <institution>Federação Internacional dos Compositores e Musicólogos
    Progressistas</institution>, cuja criação está prevista nos <lb/> termos
    consignados no <work type="manifesto">Manifesto do <event>II Congresso
    Internacional <lb/> de Compositores e Críticos Musicais</event></work>,
    de <placeName>Praga</placeName>, publicado no <lb/> nº59 da <work
    type="magazine">Vértice</work>, de <date when="-_07_1949">Julho deste
    ano</date>
    </p>
    <p> Em caso afirmativo, queira comunicar-me a sua adesão <lb/>
    com a urgência possível, para incluir o seu nome entre os dos <lb/>
    outros aderentes portugueses da <institution>Federação</institution>
    </p>
    <closer>
        <p>
            Os melhores cumprimentos do <lb/>

            <signed>
                <persName>Fernando Lopes Graça</persName>
            </signed>
        </p>
    </closer>
</body>
</text>
</TEI.2>
```

Anexo 2

```
<!ELEMENT TEI.2 (teiHeader,text)>
<!ELEMENT teiHeader (fileDesc)>
<!ELEMENT fileDesc (titleStmt,publicationStmt,sourceDesc)>
<!ELEMENT titleStmt (title,author,respStmt+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ATTLIST author attested NMTOKEN #IMPLIED>
<!ELEMENT respStmt (resp,name,date)>
<!ELEMENT resp (#PCDATA)>
<!ELEMENT name (#PCDATA|choice|abbr)*>
<!ATTLIST name type NMTOKEN #IMPLIED>
<!ELEMENT date (#PCDATA|pc|lb|abbr|choice)*>
<!ATTLIST date when NMTOKEN #IMPLIED>
<!ELEMENT publicationStmt (authority,idno,availability)>
<!ELEMENT authority (p+)>
<!ELEMENT p
(#PCDATA|choice|lb|work|persName|institution|date|occupation|pc|pb|placeN
ame|signed|abbr|num|list|quote|time|address|ref|sup| addrLine|date|p)*>
<!ELEMENT idno (#PCDATA)>
<!ELEMENT availability (p+)>
<!ATTLIST availability status NMTOKEN #REQUIRED>
<!ELEMENT sourceDesc (letDesc)>
<!ELEMENT letDesc (letIdentifier,letHeading,physDesc,envOcc,letPart*,
letContents)>
<!ELEMENT letIdentifier (country,region,settlement,collection,idno)>
<!ELEMENT country (#PCDATA)>
<!ELEMENT region (#PCDATA)>
<!ELEMENT settlement (#PCDATA|choice|abbr)*>
<!ELEMENT collection (#PCDATA)>
<!ELEMENT letHeading (author,addressee,placeLet,dateLet*)>
<!ELEMENT addressee (#PCDATA)>
<!ATTLIST addressee attested NMTOKEN #REQUIRED>
<!ELEMENT placeLet (#PCDATA)>
<!ATTLIST placeLet attested NMTOKEN #REQUIRED>
<!ELEMENT dateLet (#PCDATA)>
<!ATTLIST dateLet attested NMTOKEN #REQUIRED>
```

```

<!ELEMENT physDesc (type,support,note*,extent*,paraphernalia*)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT support (#PCDATA)>
<!ELEMENT extent (dimensions)>
<!ELEMENT dimensions (height,width)>
<!ELEMENT height (#PCDATA)>
<!ATTLIST height units NMTOKEN #REQUIRED>
<!ELEMENT width (#PCDATA)>
<!ATTLIST width units NMTOKEN #REQUIRED>
<!ELEMENT paraphernalia (paraphList)>
<!ELEMENT paraphList (paraphItem)>
<!ELEMENT paraphItem (paraphDesc)>
<!ATTLIST paraphItem id CDATA #REQUIRED>
<!ELEMENT paraphDesc (p)>
<!ELEMENT envOcc EMPTY>
<!ATTLIST envOcc occ NMTOKEN #REQUIRED>
<!ELEMENT letPart (letHeading)+>
<!ELEMENT letContents (p*,note*)>
<!ELEMENT note (#PCDATA|p|work)*>
<!ELEMENT text (envelope*,body)>
<!ELEMENT body (pb, opener, (p+, pb?, p*, pb?)*,
closer*,p*,postscript*,noteStmt*)>
<!ELEMENT envelope (envPart+)>
<!ELEMENT envPart (address)*>
<!ATTLIST envPart side NMTOKEN #REQUIRED>
<!ELEMENT address
(#PCDATA|addrLine|abbr|lb|name|institution|street|settlement|num|sup)*>
<!ELEMENT street (#PCDATA|num|lb|abbr|placeName)*>
<!ATTLIST address type NMTOKEN #IMPLIED>
<!ELEMENT addrLine
(#PCDATA|occupation|lb|abbr|persName|num|settlement|choice|institution|st
reet|signed|date|work|placeName)*>
<!ELEMENT occupation (#PCDATA|choice|abbr|lb|pc)*>
<!ELEMENT lb EMPTY>
<!ELEMENT abbr (#PCDATA|sup)*>
<!ATTLIST abbr expand CDATA #REQUIRED>
<!ELEMENT sup (#PCDATA|lb|choice|num)*>
<!ELEMENT persName (#PCDATA|lb|pc|choice|abbr)*>

```

```

<!ELEMENT num (#PCDATA|choice|sup)*>
<!ATTLIST num type NMTOKEN #REQUIRED>
<!ATTLIST num value NMTOKEN #REQUIRED>
<!ELEMENT choice (orig,reg)>
<!ELEMENT orig (#PCDATA|lb|pc)*>
<!ELEMENT reg (#PCDATA|lb|pc)*>
<!ELEMENT pb EMPTY>
<!ATTLIST pb n NMTOKEN #REQUIRED>
<!ELEMENT opener
(#PCDATA|address|lb|placeName|date|salute|occupation|persName|abbr|pc|cho
ice|signed|num|time|p)*>
<!ELEMENT placeName (#PCDATA|lb|choice|abbr|pc)*>
<!ELEMENT salute (p+)>
<!ELEMENT work
(#PCDATA|lb|placeName|persName|pc|abbr|num|sup|date|event)*>
<!ATTLIST work type NMTOKEN #REQUIRED>
<!ELEMENT institution
(#PCDATA|lb|abbr|placeName|pc|name|persName|choice)*>
<!ELEMENT event (#PCDATA|lb)*>
<!ELEMENT pc (#PCDATA)>
<!ELEMENT closer (p+, date?)>
<!ELEMENT signed (persName)>
<!ELEMENT postscript (#PCDATA|label|p|institution)*>
<!ELEMENT label (#PCDATA|abbr)*>
<!ELEMENT list (item)>
<!ELEMENT item (work|placeName|persName|abbr|choice)*>
<!ATTLIST list rend NMTOKEN #REQUIRED>
<!ELEMENT quote (#PCDATA|lb|choice|occupation|pc|date|abbr)*>
<!ELEMENT noteStmt (note)>
<!ELEMENT ref (#PCDATA|persName|lb|placeName|date)*>
<!ELEMENT time (#PCDATA|sup)*>
<!ATTLIST time when NMTOKEN #REQUIRED>

```