

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO

Tratamento e disponibilização de arquivos de anatomia patológica: modelo de enriquecimento de metainformação

Ana Carolina Ramalho Sampaio



Mestrado em Ciência da Informação

Orientador: Prof. Doutor Gabriel David

Coorientadora: Prof. Dr^a. Maria Fernanda Gonçalves

22 de outubro de 2025

Tratamento e disponibilização de arquivos de anatomia patológica: modelo de enriquecimento de metainformação

Ana Carolina Ramalho Sampaio

Mestrado em Ciência da Informação

Aprovado em provas públicas pelo Júri:

Presidente: Prof^ª. Doutora Carla Teixeira Lopes

Arguente: Prof^ª. Doutora Maria Teresa Pereira

Vogal: Prof. Doutor Gabriel de Sousa Torcato David

22 de outubro de 2025

Resumo

Os arquivos clínicos dos hospitais mais antigos constituem um património de enorme valor informacional. Contudo a efetivação desse valor para efeitos de prestação de cuidados e para investigação é condicionada por várias dificuldades de acesso, com muitos registos em papel, falta de normalização de termos e dificuldades de identificação.

O presente estudo visa contribuir para a definição de uma estratégia de digitalização e enriquecimento de metainformação descritiva que melhore o acesso continuado ao Arquivo de Anatomia Patológica do Hospital Universitário São João. Este arquivo é composto por relatórios histológicos e citológicos, produzidos desde 1962, existindo uma reconhecida necessidade de melhoria no acesso à informação contida nos mesmos. Neste contexto, partindo de uma análise das necessidades informacionais dos profissionais do Serviço de Anatomia Patológica, abordam-se aspetos críticos de pesquisa e recuperação da informação deste arquivo, destacando-se o carácter parcial dos instrumentos de pesquisa que permitam recuperar a informação deste universo documental, considerando a dinâmica evolutiva do vocabulário e termos de pesquisa utilizados ao longo do tempo.

Com o intuito de promover o acesso integrado e abrangente ao arquivo de relatórios de anatomia patológica, a presente dissertação questiona de que forma pode o enriquecimento de metainformação descritiva contribuir para a definição de uma política de tratamento técnico do arquivo de anatomia patológica do HUSJ. Com base numa abordagem de investigação-ação, foram realizadas entrevistas a patologistas e investigadores do Serviço de Anatomia Patológica, que permitiram um conhecimento e exploração da dinâmica informacional do serviço e do comportamento de pesquisa dos profissionais. Paralelamente, tendo por base amostras de relatórios representativas deste acervo, bem como instrumentos de trabalho de suporte à produção e uso dos relatórios, foi realizada uma análise do vocabulário controlado SNOMED CT e da linguagem natural em uso ao longo do tempo, de modo a identificar constrangimentos técnicos e terminológicos à recuperação da informação. Com base nas mesmas amostras, foram realizados testes de aplicação de ferramentas de reconhecimento ótico de caracteres e transcrição, no sentido de apoiar a definição da proposta de estratégia de enriquecimento de metainformação descritiva.

Os resultados apresentados permitirão apoiar o HUSJ na concretização de uma estratégia de tratamento massivo deste arquivo, incluindo-se na proposta contributos relativos à sistematização de descritores terminológicos a utilizar, aplicação de OCR para extração de termos e inclusão de elementos de metainformação descritiva alinhados com vocabulários controlados no Repositório Clínico Digital.

Palavras-chave: Arquivos clínicos; Anatomia Patológica; Metainformação descritiva; Vocabulários controlados; OCR; Repositório Clínico Digital.

Abstract

The clinical archives of oldest hospitals constitute a heritage of enormous informational value. However, the realisation of this value for the purposes of healthcare provision and research is conditioned by various difficulties of access, with many records on paper, a lack of standardisation of terms, and difficulties of identification.

This study aims to contribute to the definition of a strategy for digitising and enriching descriptive meta-information that improves continued access to the Pathological Anatomy Archive of Hospital Universitário São João. This archive consists of histological and cytological reports produced since 1962, and there is a recognised need to improve access to the information contained in these reports. In this context, based on an analysis of the information needs of the professionals of the Pathological Anatomy Service, critical aspects of research and retrieval of information from this archive are addressed, highlighting the partial nature of the research tools that allow the retrieval of information from this documentary universe, considering the evolutionary dynamics of the vocabulary and search terms used over time.

With the goal of promoting integrated and comprehensive access to the pathological anatomy report archive, this dissertation questions how the enrichment of descriptive meta-information can contribute to the definition of a technical treatment policy for the HUSJ pathological anatomy archive. Based on an action research approach, interviews were conducted with pathologists and researchers from the Pathological Anatomy Service, which allowed for an understanding and exploration of the informational dynamics of the service and the research behaviour of professionals. At the same time, based on representative samples of reports from this collection, as well as work tools to support the production and use of reports, an analysis of the SNOMED CT controlled vocabulary and natural language in use over time was carried out in order to identify technical and terminological constraints to information retrieval. Based on the same samples, tests were carried out on the application of optical character recognition (OCR) and transcription tools, in order to support the definition of the proposed strategy for enriching descriptive meta-information.

The results presented will enable HUSJ to implement a strategy for the mass processing of this archive, including contributions relating to the systematisation of terminological descriptors to be used, the application of OCR for term extraction, and the inclusion of descriptive meta-information elements aligned with controlled vocabularies in the Digital Clinical Repository.

Keywords: Clinical records; Pathological Anatomy; Descriptive meta-information; Controlled vocabularies; OCR; Digital Clinical Repository.

Objetivos de Desenvolvimento Sustentável (ODS) da ONU

Os Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas fornecem uma estrutura global para alcançar um futuro melhor e mais sustentável para todos. Inclui 17 objetivos para enfrentar os desafios mais urgentes do mundo, incluindo pobreza, desigualdade, alterações climáticas, degradação ambiental, paz e justiça. Para este projeto, os ODS aplicáveis são descritos na tabela 1.

Tabela 1: ODS da ONU aplicáveis ao projeto

ODS	Meta	Contribuição	Indicadores e métricas de performance
3	3.8 Alcançar a cobertura universal de saúde, incluindo o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais seguros, eficazes, acessíveis e de qualidade.	Digitalização, indexação e disponibilização estruturada dos relatórios de anatomia patológica no Repositório Clínico Digital, permitindo acesso rápido e seguro à informação para apoio ao diagnóstico e investigação.	• Número de relatórios digitalizados e indexados. • Tempo médio de recuperação da informação.
9	9.5 Reforçar a investigação científica, melhorar as capacidades tecnológicas e incentivar a inovação, aumentando substancialmente o número de profissionais de investigação e desenvolvimento.	Implementação de tecnologias de OCR, mineração de texto e vocabulários controlados para enriquecer a metainformação, melhorando a infraestrutura digital hospitalar.	• Número de ferramentas OCR testadas. • Nível de interoperabilidade com outros sistemas hospitalares.
16	16.6 Desenvolver instituições eficazes, responsáveis e transparentes em todos os níveis.	Definição e aplicação de política formal de tratamento técnico do arquivo, garantindo gestão documental eficiente, rastreabilidade e preservação de longo prazo.	• Existência de política formal aprovada. • Percentagem de documentos com metainformação descritiva estruturada.

Fonte: Adaptado de United Nations Regional Information Centre, 2025.

Agradecimentos

Agradeço ao professor Gabriel David por aceitar orientar esta dissertação, pela paciência e pelos contributos enriquecedores ao longo desta investigação. Agradeço imensamente à professora Fernanda Gonçalves, que possibilitou esta experiência no Arquivo do HUSJ e que tanto fez para que esta investigação fosse possível. Sou grata a professora por me acompanhar em todos os momentos, pelas conversas, pela orientação e por partilhar a sua trajetória de trabalho e pesquisa, que enriqueceu a minha visão sobre a gestão arquivística e sobre a atuação de uma profissional admirável.

Agradeço à Dra. Joana pela disponibilidade sempre que precisei de apoio para desenvolver este trabalho, pelo aprendizado sobre o RCD, a gestão de metadados e a realidade do Serviço de Arquivo. Agradeço também à Dra. Marta pelos momentos de partilha, pelas conversas e pelo carinho. Agradeço a toda a equipa do Arquivo, que foi muito acolhedora, ajudou-me em diversos momentos e fez-me rir bastante durante estes meses de convívio.

Agradeço à Maria José por ter sido o meu pilar no Serviço de Anatomia Patológica, por todo o conhecimento partilhado e por todo o tempo que disponibilizou para conversar sobre o trabalho e sobre a vida. Agradeço ao Joel, que dedicou o seu tempo a longas conversas e por partilhar seu conhecimento sobre os sistemas do Serviço de Anatomia Patológica, parte essencial desta investigação. Agradeço a todos os patologistas que contribuíram nas entrevistas, que partilharam informações valiosíssimas e despertaram em mim grande curiosidade pela área da anatomia patológica.

Agradeço muito à minha equipa de trabalho pela compreensão e apoio nesta fase em que estive ausente da biblioteca em diversos momentos, pelas conversas e por suportarem ouvir-me falar tanto sobre esta investigação.

Agradeço às amigadas que construí neste mestrado, que se tornaram verdadeiros amigos. Vocês foram essenciais na rotina, nos momentos bons e difíceis, dentro e fora da FEUP. Sem vocês, este mestrado certamente teria sido menos divertido.

Agradeço à minha família, que, mesmo distante, esteve sempre presente durante todo este tempo. Agradeço à minha mãe, por sempre me acolher como ninguém, por ser uma gestora e profissional da informação incrível e por conversar por horas comigo sobre esta investigação. Agradeço ao meu pai, por apoiar com entusiasmo os meus projetos, os meus sonhos e tudo a que me proponho realizar. Aos meus irmãos, que sempre me recordam qual é o meu propósito nesta vida, vocês são tudo. Às minhas tias, que sempre incentivaram os meus estudos e me deram apoio para ser quem sou. Às minhas avós, que caminham ao meu lado em todos os momentos.

Agradeço à minha esposa, por ser uma investigadora incrível que me inspira diariamente e me faz crescer. Obrigada por me ouvir tanto, por ser o meu apoio, por todas as nossas conversas académicas e por me dar o privilégio de viver tudo isto ao seu lado.

Ana Carolina Ramalho

*"No silêncio e no recôndito do pensamento,
nasce a claridade da compreensão quando a palavra
se transforma em termo"*

Emilia Currás

Índice

1	Introdução	1
1.1	Motivação e questão principal	2
1.2	Objetivos	3
1.3	Estrutura da dissertação	4
2	Tratamento, recuperação e reutilização de arquivos de anatomia patológica	5
2.1	O valor dos arquivos de anatomia patológica para a prestação de cuidados e investigação	6
2.2	Análise da evolução terminológica em anatomia patológica	18
2.3	Técnicas de enriquecimento de metainformação descritiva	23
3	Estratégia de investigação	29
3.1	Metodologia de investigação	29
3.2	Plano de investigação	31
3.3	Técnicas de pesquisa e análise dos dados	33
4	A dinâmica informacional do arquivo de anatomia patológica do HUSJ	35
4.1	Uso e recuperação da informação no contexto do Serviço de Anatomia Patológica	35
4.2	O arquivo de anatomia patológica do HUSJ	40
4.3	Caracterização dos relatórios citológicos e histológicos	57
5	Análise da atividade informacional dos utilizadores e arquivo	64
5.1	Comportamento de pesquisa dos patologistas e investigadores	64
5.2	Sistematização de termos descritores em anatomia patológica	73
5.3	Testes com ferramenta OCR nos relatórios	77
5.3.1	Transkribus	77
5.3.2	Landing AI	82
5.3.3	RODA	84
5.4	Proposta de política de tratamento técnico	86
6	Conclusões e trabalho futuro	91
6.1	Contributos	91
6.2	Trabalho futuro	93
	Referências	94
A	Guião de entrevista	98
B	Termo de consentimento	102

C Ferramentas para análise das entrevistas no NVIVO	105
D Modelo de metadados atual do RCD	107
E Formulário de registo de lotes - SPD	111

Lista de Figuras

2.1	Abordagem pós-custodial na Ciência da Informação: perspectiva sistêmica.	10
2.2	O uso da IA na transformação do acesso e da interpretação dos arquivos.	17
2.3	Ciclo de vida de um objeto de informação.	24
3.1	Fases da investigação-ação.	30
4.1	Linha temporal de sistemas e métodos no Serviço de Anatomia Patológica. . . .	36
4.2	Diagrama de sequência para o <i>use case</i> "Pedido de documento retrospectivo de anatomia em contexto de urgência".	43
4.3	Processo de preparação e digitalização - SPD.	46
4.4	Ambiente de preparação de ativos - SPD.	47
4.5	Exemplo de relatório de 2002 com desenhos referentes à peça.	48
4.6	Exemplo de relatório de 1988 - frente.	49
4.7	Exemplo de relatório de 1988 - verso.	50
4.8	Exemplo de relatório de 2002.	51
4.9	Exemplo de relatório complementar de 2002.	52
4.10	Campos de preenchimento de metainformação para processos de internamento - SPD.	55
4.11	Campos de preenchimento de metainformação para processos de ambulatório - SPD.	56
4.12	Processo de produção de relatórios histológicos no Serviço de Anatomia Patológica.	58
5.1	Perfil dos participantes.	66
5.2	Fontes de pesquisa referidas nas entrevistas.	69
5.3	Diagrama de processo de pesquisa para encontrar o termo controlado.	70
5.4	Diagrama de processo de pesquisa por termos para levantamento de casos.	72
5.5	Modelos em português na plataforma Transkribus.	78
5.6	Teste com relatório digitado com correção manual - Transkribus.	79
5.7	Teste com relatório digitado sem correção - Transkribus.	79
5.8	Transcrição de campos para preenchimento no relatório - Transkribus.	80
5.9	Teste com relatório datilografado - Transkribus.	80
5.10	Teste com relatório pós correção - Transkribus.	81
5.11	Teste com relatório manuscrito - Transkribus.	81
5.12	Teste inicial Landing AI	83
5.13	Análise de imagem no Landing AI	83
5.14	Extração de dados no Landing AI	84
A.1	Cards dos vocabulários controlados.	101
C.1	Áreas por número de entrevistas (itens) codificadas.	105

E.1	Formulário de registo de lotes de documentos ativos.	112
-----	--	-----

Lista de Tabelas

1	ODS da ONU aplicáveis ao projeto	iii
2.1	Interrogações de pesquisa em bases de dados científico-técnicas.	6
2.2	Referencial teórico sobre tratamento, recuperação e uso de arquivos de anatomia patológica.	11
2.3	Referencial teórico sobre tratamento, recuperação e uso de arquivos de anatomia patológica.	12
2.4	Sistematização de potenciais contributos com o tratamento do arquivo para a Anatomia Patológica.	14
2.5	Terminologias e vocabulários controlados da área da saúde.	21
3.1	Desenho do processo da investigação em 5 fases.	32
4.1	Análise crítica das observações sobre uso de terminologias e pesquisa.	38
4.2	Análise crítica das observações sobre marcas de produção.	39
4.3	Análise crítica das observações sobre circuito de produção e uso.	39
4.4	Natureza e dimensão do arquivo de anatomia patológica.	42
4.5	Análise crítica descritiva dos relatórios - Amostra 3.	59
4.6	Análise crítica descritiva dos relatórios - Amostra 4.	60
4.7	Análise crítica descritiva dos relatórios - Amostra 5.	61
5.1	Caracterização da amostra de entrevistados.	65
5.2	Importância e finalidade das anotações manuscritas.	67
5.3	Prática de anotações manuscritas x digitais.	67
5.4	Sistematização de termos de morfologias - SNOMED CT x SNOMED SAPJ.	75
5.5	Comparação entre ferramentas de OCR aplicadas ao arquivo do HUSJ	86
5.6	Proposta de elementos adicionais no RCD para metainformação descritiva em relatórios citológicos e histológicos.	89
C.1	Codebook NVIVO.	106
D.1	Esquema de metadados do RCD (56 elementos).	108

Lista de Acrônimos

AP	Anatomia Patológica
BVS	Biblioteca Virtual em Saúde
CDA	<i>Clinical Document Architecture</i>
CI	Ciência da Informação
DeCS	Descritores em Ciências da Saúde
EHR	<i>Electronic Health Records</i>
GAN	<i>Generative Adversarial Networks</i>
HTR	<i>Handwritten Text Recognition</i>
HUSJ	Hospital Universitário São João
IA	Inteligência Artificial
ICD	Classificação Internacional de Doenças
MCDT	Meio Complementar de Diagnóstico e Terapêutica
MeSH	<i>Medical Subject Headings</i>
NLM	<i>National Library of Medicine</i>
OCR	<i>Optical Character Recognition</i>
OMS	Organização Mundial da Saúde
RCD	Repositório Clínico Digital
SNOMED CT	Systematized nomenclature of medicine clinical terms
SNS	Sistema Nacional de Saúde
SONHO	Sistema Integrado de Informação Hospitalar
SPD	Suporte de Preparação e Digitalização
SPMS	Serviços Partilhados do Ministério da Saúde
TGS	Teoria Geral dos Sistemas

Capítulo 1

Introdução

A gestão dos arquivos de natureza clínica representa um desafio substancial para as instituições de saúde, destacando-se o volume considerável de informação acumulada ao longo de décadas. Os arquivos hospitalares desempenham um papel crucial na prestação de cuidados de saúde e na investigação científica, além de oferecerem um elevado potencial para reutilização em diversos contextos. Entre os arquivos hospitalares, destacam-se não apenas os processos clínicos dos utentes, mas também aqueles de natureza específica, geralmente geridos a nível departamental. Exemplos notáveis incluem os arquivos de imunohemoterapia, medicina de reprodução e anatomia patológica, cada um com suas particularidades de gestão e valor intrínseco para uso e reutilização.

No Hospital Universitário São João (HUSJ), o arquivo de anatomia patológica¹ preserva relatórios histológicos e citológicos produzidos desde 1962 até os dias de hoje, constituídos por documentos em suporte de papel e suporte digital. A gestão deste acervo, crucial tanto para o suporte aos cuidados clínicos quanto para a pesquisa científica, enfrenta desafios relativos à organização, preservação e acesso, bem como um elevado potencial de reutilização. Os relatórios armazenados em papel estão distribuídos por diversos locais e organizados conforme critérios tradicionais de uso e pesquisa, tais como tipo, ano e número de exame. No caso dos relatórios manuscritos, o acesso à informação contida nos mesmos implica o manuseio do documento físico, ao passo que no caso de relatórios nado-digitais se identifica, em determinados anos, a existência de notas manuscritas em cópias impressas, constituindo originais múltiplos.

Apesar da generalização da produção de relatórios nado-digitais, bem como da implementação da requisição eletrónica, a existência de documentos que nascem em formato digital e posteriormente são impressos e anotados manualmente, resultando em originais múltiplos, são aspetos que constituem atualmente entraves à utilização eficiente desses registos.

Efetivamente, a existência de relatórios provisórios e definitivos, que carregam informações adicionais ou modificadas, dificultam a definição de um único documento como original definitivo, obrigando a refletir quanto ao processo de produção e uso da documentação ao longo do tempo para potencializar a sua recuperação. A existência de anotações manuscritas e a ausência

¹No decorrer desta dissertação, a nomenclatura utilizada para a anatomia patológica enquanto serviço estará com letras maiúsculas e, na qualidade de área temática, estará em letras minúsculas. O mesmo se aplica para arquivo enquanto serviço ou na qualidade de conjunto de documentos.

de metainformação estruturada pesquisável dificultam a recuperação da informação, restringindo o seu potencial de uso para decisão clínica e para investigação científica.

Neste contexto, tendo o HUSJ em funcionamento o Repositório Clínico Digital (RCD) de suporte ao tratamento e digitalização de documentação clínica, identificam-se constrangimentos no tratamento do arquivo de anatomia patológica, os quais dificultam o aproveitamento do potencial de seu uso face ao volume de informação. Os **problemas** identificados são respectivamente:

- A inexistência de uma estratégia de enriquecimento da metainformação descritiva dos relatórios de anatomia patológica que permita maximizar a sua disponibilização e utilização para os clínicos e investigadores. Essa falta implica a ausência de uma sistematização de termos em vocabulários controlados adequados para a produção de metainformação descritiva a associar aos documentos em causa;
- A abrangência temporal alargada (1962-2020), que dificulta o processo de padronização dos termos descritores e riqueza de metainformação descritiva;
- As dificuldades no tratamento, digitalização e indexação em massa dos relatórios citológicos e histológicos manuscritos e datilografados, que exigem esforço financeiro e de trabalho manual para sua realização.

Neste contexto, este estudo propõe investigar soluções no domínio da Ciência da Informação, com o intuito de aprimorar o tratamento, a organização e a disponibilização desses relatórios. É de interesse intrínseco desta investigação fomentar uma gestão documental que maximize sua utilidade no contexto hospitalar universitário, estabelecendo conexões entre a informação arquivística e as necessidades de seus múltiplos utilizadores.

1.1 Motivação e questão principal

A necessidade de disponibilizar o arquivo de anatomia patológica para o acesso do corpo clínico e da comunidade de investigadores, com capacidade de recuperação efetiva da informação, é o principal fator motivador desta pesquisa. O processo técnico adequado para alcançar essa recuperação é a indexação, que tem como papel crucial a preparação de uma representação do conteúdo temático dos documentos.

A importância da investigação científica no campo da ciência da informação, particularmente neste contexto, decorre de uma necessidade de compreender de forma abrangente os processos envolvidos na construção e organização de arquivos de anatomia patológica, que são cruciais para várias aplicações. É essencial que esse processo multifacetado seja entendido não apenas em relação a origem da informação que está a ser arquivada, mas também à luz dos contextos específicos que cercam sua criação e posterior utilização, bem como das metodologias empregadas para a recuperação dessa informação.

Esta pesquisa concentra-se na terminologia de indexação que facilita o armazenamento e a recuperação eficazes desses arquivos, ao mesmo tempo em que leva em consideração todos os outros

fatores intrínsecos que estão fundamentalmente ligados à natureza da própria informação, que à luz de Buckland, 1991 e Capurro et al., 2003, são nomeadamente sua natureza contextual e interpretativa, sua relevância, sua dualidade subjetiva e objetiva, sua temporalidade e transformação, a dependência de suporte e representação.

Nesse sentido, esta pesquisa considera o contexto técnico-científico à época da produção dos documentos em consideração, reconhecendo que os termos descritivos aplicados a esses documentos não são construções imutáveis, estando sujeitos a mudanças ao longo do tempo. A informação será tratada visando o seu desempenho central no avanço da sociedade e no suporte a decisões, ou seja, o que marca a sua abordagem é o objetivo de contribuir para a tomada de decisão no contexto hospitalar universitário.

O processo envolvido na construção desses arquivos inevitavelmente traz à tona considerações críticas relacionadas aos mecanismos de recuperação da informação, especificamente no que diz respeito à maneira pela qual a informação é descrita e indexada. Neste contexto, esta investigação abrange uma compreensão sobre as atuais necessidades específicas dos utilizadores e seus comportamentos de pesquisa, incluindo o corpo clínico e investigadores, destacando assim as implicações práticas associadas à utilização das informações neles contidas.

Nesse sentido, a **questão de investigação** que suporta esta pesquisa é: De que forma pode o enriquecimento de metainformação descritiva contribuir para a definição de uma política de tratamento técnico do arquivo de anatomia patológica do HUSJ?

As hipóteses levantadas para resposta à questão de investigação, com base em estudos como dos autores Smith et al., 2011; Ehsani et al., 2008; Burger et al., 2016; Souza, 2021, apontam para o uso de vocabulários controlados, tecnologia de reconhecimento ótico de caracteres (OCR) e de mineração de texto como ferramentas capazes de melhorar significativamente o processo de indexação e recuperação de informações em relatórios de anatomia patológica. São precisamente estas hipóteses que foram propostas para exploração no âmbito do presente estudo, apresentando-se em seguida os objetivos definidos.

1.2 Objetivos

Com referência ao exposto anteriormente, o objetivo principal desta pesquisa contribui para **Definir uma política de tratamento técnico do arquivo de anatomia patológica do HUSJ.**

Com intuito de alcançar o objetivo principal, os objetivos específicos determinados são:

1. Propôr uma estratégia de enriquecimento da metainformação descritiva dos relatórios citológicos e histológicos integrados no Repositório Clínico Digital (RCD);
2. Enquadrar o comportamento de pesquisa e as necessidades informacionais dos patologistas e investigadores em relação aos relatórios citológicos e histológicos;
3. Realizar um levantamento e sistematização de termos de indexação baseados em vocabulários controlados, tendo por base o estudo da evolução da terminologia utilizada nos relatórios ao longo do tempo;

4. Explorar o potencial do Reconhecimento Ótico de Caracteres (OCR) e da mineração de texto para o enriquecimento de metainformação e identificação automática de termos, tendo por base o tratamento de uma amostra de relatórios.

1.3 Estrutura da dissertação

Para além da introdução, em que se apresenta a motivação, a questão principal e os objetivos, esta dissertação contém mais 5 capítulos.

No Capítulo 2, é apresentada a revisão da literatura acerca do valor dos arquivos de anatomia patológica, seja para a prestação de cuidados ou para investigação, incluindo-se a análise sobre a evolução terminológica em anatomia patológica e as técnicas de enriquecimento de metainformação descritiva.

No Capítulo 3 é descrita a estratégia de investigação utilizada, sendo detalhada a metodologia, o plano de investigação e as técnicas de pesquisa e análise de dados.

No Capítulo 4 é apresentada a caracterização do objeto de estudo, bem como as atividades e os utilizadores que estão inseridos neste contexto de investigação.

No Capítulo 5 são apresentados os resultados das entrevistas e dos testes com ferramentas de OCR realizados que resultam na proposta de tratamento técnico do arquivo e do enriquecimento de metainformação descritiva.

No Capítulo 6 são apresentadas as considerações finais e sugestões para trabalhos futuros.

Capítulo 2

Tratamento, recuperação e reutilização de arquivos de anatomia patológica

Este capítulo descreve o processo de levantamento bibliográfico, resultados obtidos e referências de literatura identificadas como relevantes. A revisão da literatura segue uma abordagem narrativa de conceitos e práticas no âmbito da Ciência da Informação (CI), com enfoque na arquivística, tendo como domínio de aplicação a medicina, mais especificamente a anatomia patológica. Tem ainda a função de mapear pesquisas e aplicações em torno dos vocabulários controlados, específicos para a ciência da saúde, abordando-se o estado da arte acerca do enriquecimento de metainformação descritiva. Por metainformação descritiva entende-se, como definido pelo EAD (*Encoded Archival Description*) e exemplificado pelo modelo ISAD(G), um conjunto de elementos utilizados para a apresentação de informação bibliográfica e descritiva, cujo propósito é fornecer suporte à recuperação da informação. Alguns exemplos deste tipo de elementos podem-se realçar o título, datas, história biográfica, história custodial, âmbito e conteúdo, onde se encontra a indexação, existência e localização dos originais e cópias, bem como outras características do documento.

A secção 2.1 é dedicada à caracterização do arquivo de anatomia patológica e sua importância para o corpo clínico e para fins de investigação. A secção 2.2 toca-se na identificação de vocabulários controlados utilizados na anatomia patológica, com o objetivo de realizar um levantamento de termos descritores utilizados atualmente e dos termos cronologicamente anteriores a estes. Já a última secção 2.3 tem como foco as informações relacionadas ao tratamento e enriquecimento de metainformação descritiva, contando com análise de estudos de caso em relação ao tema em questão.

As referências bibliográficas apresentadas nesta revisão de literatura foram selecionadas a partir de um processo de pesquisa não sistemático. Durante o percurso foram utilizadas bases científicas capazes de satisfazer a qualidade da informação que exige a pesquisa, além de outras fontes igualmente fiáveis. As fontes utilizadas foram base de dados de informação científico-técnicas Pubmed e Scopus, bibliotecas universitárias, repositórios institucionais (Universidade do Porto); pesquisa por citações e sugestões bibliográficas de especialistas.

Em relação às pesquisas realizadas nas bases de dados científicas, as interrogações, filtros e resultados obtidos podem ser consultados na tabela 2.1. A tabela 2.2 ilustra os estudos mais atuais e que apontam tópicos relacionados à investigação. Para além destes, outros autores são reconhecidos durante a revisão de literatura.

Tabela 2.1: Interrogações de pesquisa em bases de dados científico-técnicas.

Base de dados fonte	Query	Nº de resultados	Nº de selecionados
Pubmed	(medical records[All Fields] AND optical character recognition[All Fields]) AND ((ffrft[Filter]) AND (fft[Filter]))	14	2
Pubmed	(Archives[All Fields] AND Pathology[All Fields] AND medical records[All Fields]) AND ((ffrft[Filter]) AND (fft[Filter]))	42	3
Pubmed	(medical records) AND (optical character recognition) AND ((y_5[Filter]) AND (ffrft[Filter]) AND (fft[Filter]))	11	2
Scopus	TITLE-ABS-KEY (digital records) AND SUBJAREA (medi) AND (LIMIT-TO (DOCTYPE , ar))	169	1
Scopus	Pathological Anatomy AND medical records AND PUBYEAR > 2018 AND PUBYEAR < 2026	210	3
Scopus	medical records AND Optical Character Recognition AND PUBYEAR > 2018 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE , ar))	59	2
Scopus	(controlled vocabulary OR thesaurus OR ontology) AND (pathology) AND (evolution OR development OR change OR progression) AND (standardization OR classification OR nomenclature OR terminology) AND PUBYEAR > 2018 AND PUBYEAR < 2026	13 667	5

2.1 O valor dos arquivos de anatomia patológica para a prestação de cuidados e investigação

Como forma de dar início ao tema, buscam-se diferentes perspectivas na literatura quanto à arquivística e a arquivologia. Tendo presente que a arquivologia aborda a informação de forma orgânica, com a intenção de tornar evidência algo que ocorreu em alguma altura da história. Esse objetivo desdobra-se em abordagens teóricas e técnicas a serem aplicadas no intuito de coletar, organizar, armazenar, preservar e recuperar as informações contidas nos arquivos. Nesse sentido, segundo Fonseca, 2005, a arquivologia, na perspectiva brasileira, é a ciência que se preocupa em gerir e preservar de forma eficiente a documentação de uma instituição, de forma a salvaguardar o património histórico e informacional.

No contexto português, a arquivística se enquadra como uma disciplina no âmbito da CI e uma das suas características principais é a atuação do arquivo enquanto sua natureza funcional.

No que diz respeito aos arquivos, Silva; Ribeiro et al., 1998 referem que "arquivo é um sistema (semi-)fechado de informação social materializada em qualquer tipo de suporte, configurado por dois factores essenciais - a natureza orgânica (estrutura) e a natureza funcional (serviço/uso) - a que se associa um terceiro - a memória - imbricado nos anteriores".

Nestas duas visões, brasileira e portuguesa, que se diferem no processo de construção conceptual enquanto ciência ou disciplina, nesta investigação entende-se que a perspectiva de recuperação e uso se encontram em ambas. Portanto, o contributo enquanto uma ciência ou disciplina, que têm como objetivo a promoção do acesso à informação, é enriquecedor para esta pesquisa, uma vez que o paradigma em questão está dedicado ao tratamento e disponibilização da informação, com foco em suprir a necessidade informacional do utilizador.

Com base nos conceitos referidos, é crítica a importância que um arquivo apresenta para a comunidade designada que o utiliza e para a instituição que o detém, no que se refere a capacidade informacional histórica que está contida. Essa importância, aplicada a área da saúde, soma-se à potência que um arquivo hospitalar detém de servir dados clínicos e informação capazes de identificar padrões de doença de uma comunidade ao longo dos anos, os tratamentos utilizados em cada período da história, bem como contextos sociais dos utentes, volume de atendimentos e outras informações valiosas para diferentes objetivos.

A expressão "dados clínicos" refere-se, neste estudo, a informação dos utentes que são utilizadas para fins de identificação, considerando informações administrativas e registos clínicos. À luz do que é exposto por Bradley et al., 2024, os dados podem ser encontrados tanto de forma estruturada, quando estão devidamente identificados por etiquetas ou campos que indicam ou requerem uma estrutura específica, ou de forma não-estruturada, quando são apresentados em texto livre por uma narrativa de consulta clínica ou relatórios.

Em uma visão da arquivística mais centrada na diplomática, Duranti, 1994 refere que "A razão pela qual eles (os arquivos) puderam servir a tantas finalidades é que os materiais arquivísticos, ou registros documentais, representam um tipo de conhecimento único: gerados ou recebidos no curso das atividades pessoais ou institucionais, como seus instrumentos e subprodutos, os registros documentais são as provas primordiais para as suposições ou conclusões relativas a essas atividades e às situações que elas contribuíram para criar, eliminar, manter ou modificar".

A informação contida nestes registos é o que os torna valiosos e isto pode ser justificado a partir das propriedades intrínsecas à informação. Segundo Le Coadic, 1996, as propriedades destacam a natureza multifacetada e seu papel central nas interações humanas e na sociedade. Para além disso, é um recurso que cumpre funções específicas e possui atributos que tornam a informação central no processo de produção e disseminação do conhecimento. As propriedades da informação, de acordo com Le Coadic, 1996, são as seguintes:

1. O carácter relacional, que identifica a informação como algo que se concretiza no contexto de uma relação entre emissor e um recetor. Essa relação é mediada pela intencionalidade e pelo significado atribuído à mensagem;

2. O caráter utilitário, que atribui a informação um valor prático, sendo assim, é útil na resolução de problemas e tomada de decisões. Esse valor está diretamente associado à capacidade de atender às necessidades informacionais dos utilizadores;
3. O caráter subjetivo, que se refere a dependência da informação de um recetor que a processe, a partir de conhecimentos culturais, cognitivos e sociais. Essa propriedade implica dizer que a informação tem um teor subjetivo e interpretativo;
4. O caráter cultural, responsável por descrever que a informação se encontra imersa em um contexto cultural que a molda por todo seu ciclo de vida, desde a sua produção, durante a sua organização e até à sua disseminação. Nessa perspectiva, a importância dos sistemas de gestão estão na preservação do contexto que promove o significado;
5. O caráter dinâmico, que enfatiza que a informação não é estática, ela tem a capacidade de evoluir com o tempo e pode gerar novas informações a partir de processos interpretativos e de reutilização em contextos específicos;
6. O caráter cumulativo, que identifica a capacidade de acumulação ao longo do tempo, que forma uma base de conhecimento para registo histórico e evolução;
7. O caráter social, que pontua o desempenho crucial na interação social, que facilita a comunicação, a cooperação e a formação de redes de conhecimento.

Entretanto, Silva; Ribeiro, 2000, após refletir e definir a informação social como "um conjunto estruturado de representações codificadas (símbolos, significantes) socialmente contextualizadas e passíveis de serem registadas num qualquer suporte material e/ou comunicadas em tempos e espaços diferentes", apontam 6 propriedades da informação, sendo respetivamente:

1. Estruturação pela ação (humana e social): a informação não existe de forma espontânea. É produzida, organizada e estruturada pela ação dos sujeitos em contextos sociais. Essa estruturação reflete práticas culturais, históricas e cognitivas, o que significa que cada informação carrega marcas do coletivo que a produz;
2. Integração dinâmica: a informação não é estática, integra-se a outros conjuntos informacionais e contextuais de modo dinâmico. Ela adquire novos significados conforme é associada a diferentes usos, situações e interpretações, permanecendo em constante atualização;
3. Pregnância: refere-se à capacidade da informação de gerar novos sentidos, interpretações e conhecimentos. Ou seja, a informação tem potencial de fecundidade semântica, podendo suscitar leituras diversas de acordo com o contexto sociocultural e o horizonte do sujeito;
4. Quantificação: a informação pode ser medida, registada e avaliada em termos quantitativos, seja pelo volume de dados que representa, seja pelo número de ocorrências e registos que produz. Essa dimensão é relevante, por exemplo, nos processos de gestão da informação;

5. Reprodutividade: uma vez codificada e registada, a informação pode ser reproduzida indefinidamente em diferentes suportes e formatos, mantendo a sua essência, ainda que o meio material possa variar. Essa característica sustenta a sua difusão e acessibilidade;
6. Transmissibilidade: a informação possui a capacidade de ser comunicada e partilhada em diferentes tempos e espaços. Essa propriedade permite a circulação intergeracional e intercultural, assegurando a continuidade da memória social.

As duas últimas propriedades atribuem à informação a característica de algo que passa por um processo, ou seja, é criada, difusa, organizada, armazenada, recuperada e utilizada. Desta forma, enquanto as propriedades propostas por Le Coadic, 1996, que muito se assemelham as definidas por Silva; Ribeiro, 2000, enfatizam a dimensão multifacetada da informação e a sua relevância para os processos sociais. Esta investigação foca-se na abordagem de Silva; Ribeiro, 2000, que compreende a informação como fenómeno social estruturado pela ação humana, dinâmico e capaz de gerar, reproduzir e transmitir sentidos ao longo do tempo. Esta perspetiva revela-se mais adequada para a análise seguinte, que se orienta por uma visão sistémica da informação.

A clássica Teoria Geral dos Sistemas (TGS), desenvolvida por Bertalanffy, 1968, propõe uma abordagem interdisciplinar para compreensão de sistemas mais complexos, com intuito de permitir a sua aplicação em diferentes áreas do conhecimento. Destaca-se nesta pesquisa, esta preocupação com uma visão integrada, em uma perspetiva de que a informação é gerada por diversos agentes, com diferentes origens, componentes e destino.

A TGS tem influência profunda na CI, pois permitiu a compreensão de sistemas de informação como estruturas organizadas e interdependentes. Para além disso, influenciou a forma de pensar a organização do conhecimento, ou seja, a criação e uso de sistemas classificatórios, vocabulários controlados, e outras ferramentas técnicas compreendem sistemas estruturados, formando um modelo sistémico de recuperação da informação.

Durante esta pesquisa muito será citado o modelo de organização do conhecimento de Lancaster, responsável por descrever a indexação e recuperação da informação como um sistema dinâmico. Lancaster, 2004 referencia em sua obra que os membros de uma comunidade designada tem como objetivo utilizar uma base de dados para satisfazer a sua necessidade informacional e descreve que a política de indexação, bem como a exatidão da indexação, sendo essa etapa dividida entre análise conceptual e tradução, são os fatores principais que influenciam diretamente o sistema de recuperação da informação.

O paradigma científico e pós-custodial, que tem como centralidade a informação social como objeto de conhecimento, defendido na Ciência da Informação, implica algumas abordagens que se diferem no seu processo tradicional defendido na perspetiva da Biblioteconomia e na Arquivologia. Segundo Ribeiro, 2008, sem abandonar completamente o tecnicismo, a preservação e as preocupações de acesso das teorias e práticas tradicionais, esse paradigma visa trazer alguns aspetos fundamentais que o caracterizam. Na figura 2.1 é possível perceber que a visão holística é um dos pontos principais nessa perspetiva pós-custodial da CI. Essa abordagem é crucial para o tratamento de arquivos, bibliotecas e seus respetivos sistemas em diversas áreas.

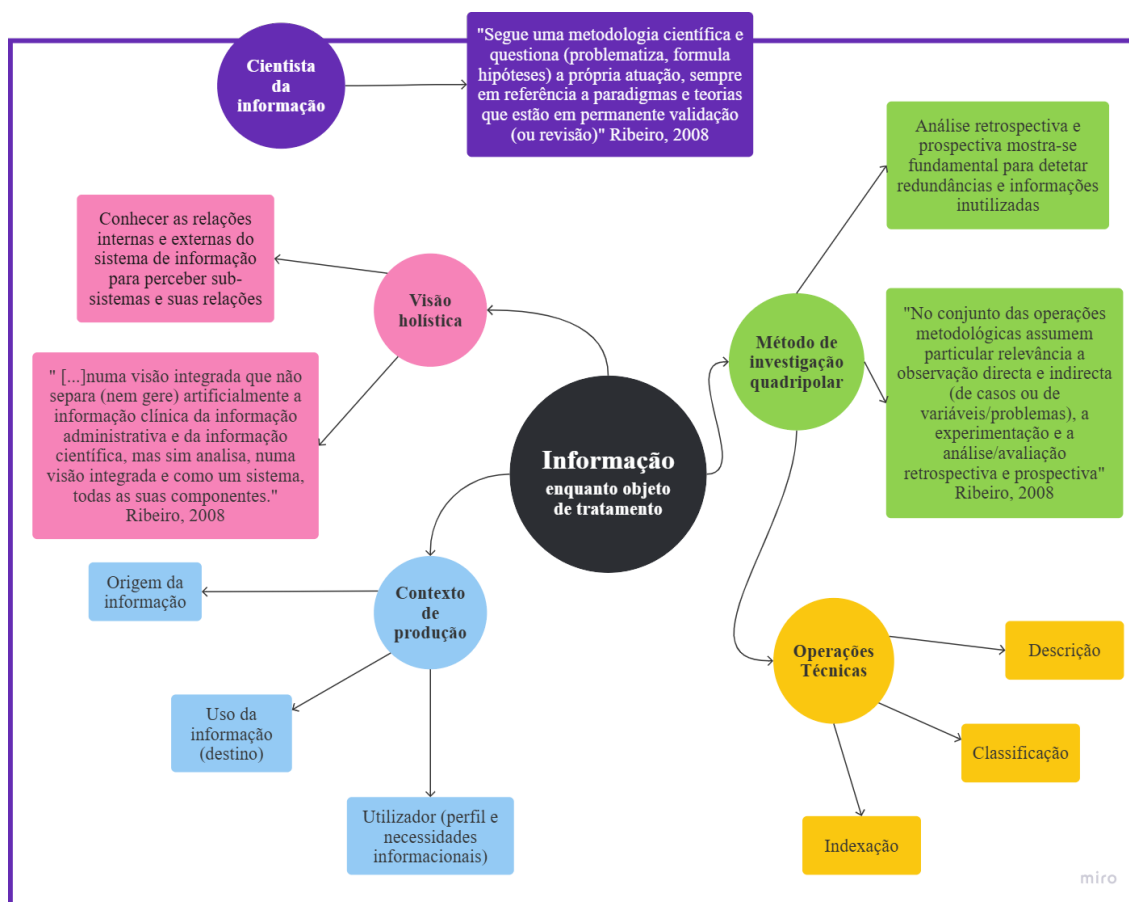


Figura 2.1: Abordagem pós-custodial na Ciência da Informação: perspectiva sistêmica.

Fonte: Desenvolvido a partir de Ribeiro, 2008

No seguimento da perspectiva citada anteriormente, relacionando-a com o tratamento dos arquivos de anatomia patológica de forma a potenciar o seu uso, contexto alvo desta investigação, pode ser observado em diferentes pesquisas científicas e com aplicações tecnológicas. O quadro 2.2 aborda os principais autores que contribuíram para a construção teórica desta investigação.

Tabela 2.2: Referencial teórico sobre tratamento, recuperação e uso de arquivos de anatomia patológica.

Tema	Referência	Métodos/Abordagem	Principais achados e contribuições	Citação direta
2.1 Valor dos arquivos para cuidados e investigação	Casparie et al. (2007)	Criação e manutenção de sistema nacional descentralizado, com codificação padronizada, transferência eletrônica de extratos de laudos, integração com registros de câncer e screening populacional.	Rede cobre 64 laboratórios, 42 milhões de registros; permite compartilhamento descentralizado imediato, controle de qualidade contínuo, integração com registros nacionais, suporte à assistência clínica, pesquisa, epidemiologia e biobancos; uso em pesquisas epidemiológicas, padronização SNOMED, controle de acesso via criptografia.	“The PALGA system is a unique, highly automated pathology archiving and communication tool...providing considerable support to pathologists in their daily practices...[it] supports the Dutch cancer registration, national screening organizations, and investigators wishing to address research questions in various domains, including epidemiology, pathology, and molecular biology of disease.”
	Ehsani et al. (2008)	Integração de base digital a sistemas clínicos e operatórios; uso para consultas retroativas e pesquisa populacional.	Bases organizadas oferecem economia e ganho científico; permitem análises retrospectivas e aumento do poder estatístico em estudos.	“Databases of pathology specimens can have enormous benefits for clinical use and long-term population-based research...highly advantageous for retrospective analyses.”
	Bondi et al. (2010)	Implementação de sistema PACS regional com scanners de lâminas, organização hierárquica de casos, integração com plataformas de ensino à distância e reuniões de consenso.	Solução integrada para promoção da qualidade e controle em oncologia; facilita compartilhamento regional; suporte a testes de proficiência, ensino, referência diagnóstica online; promove auditabilidade, interatividade e consulta remota. Digitalização facilita avaliação conjunta e reprodutibilidade diagnóstica, além de registrar dados estruturados, níveis de acesso e recursos para ensino.	“This project provided an integrated environment for digital pathology, spanning quality control, education, and reference: cases are catalogued in an extensive and systematic digital archive, easily accessible, with diagnosis discussed and validated by experts, used as diagnostic reference tool (casuistic atlas online).”
	Smith et al. (2011)	Avaliação de arquivo histórico para uso em estudos epidemiológicos e moleculares; ligação com iniciativas como Shared Pathology Informatics Network.	Arquivos de AP são fonte abundante de material clinicamente anotado; essenciais para revisão de casos, estudos de recorrência e inscrição em ensaios clínicos.	“Tissue repositories maintained by pathology departments represent an abundant resource of clinically annotated human specimens.”
2.2 Evolução terminológica em anatomia patológica	Castro et al. (1977)	Implantação de processamento eletrônico de dados no INCa; utilização simultânea de diferentes sistemas (ICD-8, MOTNAC, SNOP, Anderson) para codificação de diagnósticos histopatológicos; criação de listas simplificadas e índices alfabéticos.	Introduziu no Brasil a padronização terminológica em AP com múltiplos sistemas combinados, estruturando um banco de dados de mais de 150 mil diagnósticos; destacou desafios de uniformidade de nomenclatura e necessidade de interoperabilidade.	“O ideal... seria a utilização de um único sistema de codificação, tanto para neoplasias como para patologia não neoplásica, no caso o SNOP. (...) Contudo, a necessidade de fornecer às Autoridades de Saúde informações mais genéricas nos moldes do ICD-8... nos levam à utilização obrigatória de dois sistemas simultâneos de codificação.”

Tabela 2.3: Referencial teórico sobre tratamento, recuperação e uso de arquivos de anatomia patológica.

Tema	Referência	Métodos/Abordagem	Principais achados e contribuições	Citação direta
	Ehsani et al. (2008)	Conversão automática de diagnósticos atribuídos por patologistas para formato OMS usando lista mestre de termos e regras de priorização.	Padronização terminológica aumenta a consistência e facilita integração com bases externas.	“Converted... to the standardized World Health Organization (WHO) format... with a master list of key site and diagnostic terms, and prioritization rules.”
	Burger et al. (2016)	Revisão de técnicas de NLP para codificação de laudos; histórico desde codificação SNOP/SNOMED até aprendizado de máquina e sistemas baseados em regras.	Evolução mostra transição de simples busca de termos para métodos avançados com extração numérica e ontologias; desafios incluem ausência de datasets compartilhados para comparação.	“Early attempts to extract encoded data... consisted of automated retrieval of terms... represented as SNOMED codes... later, more advanced approaches... using NLP techniques.”
2.3 Técnicas de enriquecimento de metainformação descritiva	Ehsani et al. (2008)	Digitalização de 30 anos de registros (30 mil casos) com OCR e programas Java para extração de dados clínicos e diagnósticos; integração com LIS; mapeamento para formato OMS.	Modelo prático de conversão de acervo físico em base digital pesquisável; filtragem e padronização aumentam a reutilização para pesquisa e apoio clínico.	“High-throughput automated scanning of the paper archives was employed... filtered through a multi-tiered process... converted... to the standardized World Health Organization (WHO) format.”
	Ehsani et al. (2008)	Tagging automático de diagnósticos e localização anatômica; associação com laudos cirúrgicos e dados de imagem; verificação manual redundante.	Anotações e enriquecimento com termos-chave aumentam organização e pesquisa orientada por critérios complexos.	“Each entry... was automatically tagged with key neuropathologic diagnostic and site-of-occurrence terms.”
	Smith et al. (2011)	Digitalização de ~120 mil relatórios (1956–1979) via scanner de alta capacidade, OCR, indexação em PDF; definição de “integridade do arquivo” e avaliação de qualidade de blocos e lâminas.	Criação de acervo eletrônico pesquisável com baixo custo e prazo rápido; integridade do arquivo é essencial para valor científico; avaliação de viabilidade de testes moleculares em material antigo.	“We share our experience with scanning approximately 120,000 pathology reports from 1956 to 1979 into an electronically searchable archive, with a \$9,000 budget, completed in 6 weeks.”
	Burger et al. (2016)	Uso de NLP para extrair dados estruturados, detectar negações e identificar casos; métodos incluem correspondência de termos, ML probabilístico e sistemas baseados em regras.	Técnicas geram alto desempenho na extração de dados clínicos relevantes; permitem metadados enriquecidos para pesquisa e vigilância epidemiológica.	“Common approaches were word/phrase matching, probabilistic machine learning and rule-based systems... good performances... are reported for all of these.”
	Souza (2021)	Mapeamento e harmonização entre linguagem natural em PEPs e terminologias de referência; uso de PLN para extração de termos e identificação de sinonímia/ambiguidade.	Ampliação e enriquecimento de terminologias clínicas; criação de corpus léxico computacional em português, útil para padronização e interoperabilidade.	“Essa amostra demonstrou a riqueza de termos da Terminologia de Interface, o que será útil no enriquecimento da Terminologia de Referência.”

Os arquivos de anatomia patológica são compostos por documentos com dados clínicos do doente e, a depender do local, contam também com as lâminas referentes aos exames, as amostras. A pesquisa realizada por Smith et al., 2011, mostra que à medida que os arquivos de anatomia patológica se tornam clinicamente irrelevantes, devido a sua antiguidade elevada, os custos para o armazenamento tornam-se maiores e a justificação para seu tratamento é crucial para manter a informação valiosa contida nestes documentos. Portanto, o projeto mostra uma forma de minimizar os custos relacionados ao armazenamento e tratamento, bem como tornar o arquivo mais atrativo aos investigadores. Para além de propor uma estrutura para o tratamento das amostras, o projeto também compartilha as experiências de digitalização de cerca de 120000 relatórios desde 1956 a 1979.

Outro projeto de referência foi realizado na Rede de Saúde da Universidade de Toronto por Ehsani et al., 2008, onde foi criado um banco de dados neuropatológico pesquisável a partir de arquivos em papel, utilizando-se a digitalização automatizada de alto rendimento. O OCR foi realizado nesse processo como forma de converter as imagens digitalizadas em texto, ou seja, para extração de texto para um formato editável. O banco de dados foi projetado para permitir uma fácil integração com outros já existentes para vínculo dos documentos a informações relacionadas. Notam-se vantagens para análises retrospectivas e aprimoramento do tamanho das amostras, uma vez que a partir de mais de 30.000 registos inseridos no banco de dados, foi possível disponibilizar a informação de forma mais efetiva para clínicos e comunidade de investigadores.

Já o estudo realizado por Carrajo et al., 2008 desenvolveu um software para atuar como um portal único para a equipa do hospital acessar os dados clínicos dos doentes de forma eficiente. Esse software classifica a informação segundo uma série de critérios de forma que a eficiência se tornou muito melhor em comparação aos métodos tradicionais com uso de papel. A transição dos papeis para a aplicação de registos médicos eletrónicos (EMR) possibilitou o acesso a documentos estruturados e padronizados.

Os estudos realizados no âmbito dos arquivos de anatomia patológica mostram que o arquivo fornece apoio informacional em diferentes problemas clínicos, que os autores Straus et al., 2019 destacam em sua obra. O quadro 2.4 produzido a partir da obra, referencia principalmente os campos diretamente ligados a anatomia patológica e exemplifica a utilidade da informação nesses aspetos.

Tabela 2.4: Sistematização de potenciais contributos com o tratamento do arquivo para a Anatomia Patológica.

Campos	Utilidade
Etiologia/risco	Identificar causas ou factores de risco para doenças
Diagnóstico de doenças	Monitorar a frequência de doenças, causa suas manifestações clínicas e como usar esse conhecimento na classificação de doenças dos doentes
Diagnóstico diferencial	Quando considerar as possíveis causas dos problemas clínicos do paciente, como seleccionar aqueles que são prováveis e responsivo ao tratamento
Testes diagnósticos	Seleccionar e interpretar testes de diagnóstico para confirmar ou excluir um diagnóstico, com base na sua precisão, exatidão, aceitabilidade, segurança e despesa
Prognóstico	Estimar o curso clínico provável do doente ao longo do tempo e prever possíveis complicações do distúrbio
Prevenção	Reduzir a chance de doença, identificando e modificando factores de risco e como diagnosticar a doença precocemente na triagem

Fonte: Adaptado de Straus et al., 2019

Como referido na tabela 2.4, o uso dos dados para prestação de cuidados é um dos aspetos mais importantes em relação aos documentos arquivísticos clínicos. A pesquisa realizada por Smith et al., 2011, mostra que estes arquivos contam com recursos inestimáveis para a investigação clínica, o que permite a análise da progressão da doença, compreender as condições e desenvolver tratamentos direcionados com base em dados históricos. Estes arquivos podem não só melhorar a compreensão em relação as doenças, como facilitar a descoberta de marcadores que podem levar a diagnósticos precoces e abordagens de medicina personalizada.

Segundo Gonçalves et al., 2011, em um estudo realizado anteriormente no HUSJ sobre a reutilização de informação clínica para apoiar a definição de uma política de gestão da informação da saúde, a otimização do uso dos registos clínicos pode contribuir para a redução de exames repetitivos, uma vez que a completude dos registos ajuda o clínico a perceber todo o percurso do doente. Pode também ajudar na melhoria na continuidade dos cuidados, apoiar a tomada de decisões e uma melhor coordenação dos cuidados em diferentes sectores hospitalares. A autora também levanta outros aspetos relevantes para os quais a digitalização contribui, sendo eles a segurança dos registos de doentes do HUSJ, bem como a acessibilidade facilitada a registos anteriores para capacitação dos clínicos na revisão de casos de estudo.

O acesso aos exames e diagnósticos retrospectivo de um mesmo doente é essencial para a anatomia patológica, pois esses registos tem a capacidade de sustentar novas análises ou diagnósticos. Por outro lado o arquivo da anatomia sempre teve uma organização mais cuidada pelo suporte ao dia a dia dos serviços. O facto de serem de menor dimensão do que os arquivos de processo clínico permitiu essa organização continuada no tempo.

Para além do uso dos dados para a prestação de cuidados aos doentes, é de se destacar o uso

para o ensino e investigação. Nesse aspecto, o tratamento dos arquivos de anatomia patológica para esse fim pode capacitar a comunidade de investigadores a identificar padrões de doenças da população a qual se atende, verificar a frequência de doenças por temporada, ou seja, realizar estudos epidemiológicos mais efetivos, bem como rastrear tratamentos efetivos já realizados. Ou seja, o uso para investigação ajuda a desenvolver a prevenção e evolução da ciência no que diz respeito a área da saúde.

A exploração dos dados clínicos, como os relatórios de anatomia patológica, tem uma importância elevada quando levados em consideração os avanços científicos que podem ser gerados a partir desses estudos. Existe atualmente uma questão em relação a produção de dados sintéticos em detrimento do uso dos dados reais existentes nos arquivos hospitalares em relação a produção massiva e uso de dados sintéticos na saúde.

A questão está no porquê produzir dados sintéticos em massa se existem décadas de registros hospitalares que podem ser tratados para fins de investigação. Os benefícios da produção de dados sintéticos são claros, como a facilidade de lidar com dados que não pertencem a nenhum titular, ou seja, garante a conformidade com as regulamentações e a privacidade dos utentes. E ainda o aumento do volume de dados de treinamento, que se mostra útil nos casos em que o acesso a dados reais é restrito ou insuficiente para as pesquisas mais amplas.

Os dados sintéticos são geralmente produzidos por GANs (*Generative adversarial networks*), que consistem em duas redes neurais que "enganam" uma a outra para extrair a melhor informação possível. Essa ferramenta é utilizada para criar dados sintéticos a partir de dados reais Chen et al., 2021. Em uma GAN, gerador é uma rede neural responsável por produzir um dado falso, que tem por objetivo enganar a rede neural discriminadora crítica, que por sua vez tenta distinguir o que real e o que é falso. À medida que esse processo evolui, o gerador aprende com as distinções realizadas pelo discriminador. Dessa forma, o gerador cria cada vez falsificações mais realistas. Essa técnica tem sido largamente utilizada para gerar imagens realistas de lesões na pele, imagens de raio-x e imagens histológicas, como demonstrado no estudo realizado pelos autores.

Entretanto, a disponibilização de dados reais em um repositório hospitalar, pensado para o acesso de clínicos e investigadores, torna capaz a pesquisa para identificar doenças recorrentes da comunidade em causa da região que se atende. É possível perceber mais das necessidades reais dos doentes que utilizam os serviços, ou seja, ter descobertas reais para problemas reais. A produção de mais dados em contextos em que a informação já existe, e apenas precisa de tratamento para estar disponibilizada, faz com que exista uma duplicação dos dados e consequentemente um volume muito maior de dados e documentos a serem processados, organizados, armazenados e inseridos em sistemas que possibilitem a recuperação da informação.

Apesar dos potenciais tecnológicos para gestão e reutilização de arquivos aqui levantados, os sistemas de informação existentes em contexto hospitalar têm que evoluir para que essa exploração seja possível, deixando de ser apenas repositórios de informação e virando ativos estratégicos capazes de gerar conhecimento a partir dos dados armazenados.

Nessa linha, Gouveia, 2025 argumenta que a integração da inteligência artificial e da chamada “memória profunda” pode transformar de forma decisiva o acesso e a interpretação dos arquivos.

Trata-se de uma perspectiva que vai além da automação por si só, permitindo explorar relações complexas, padrões e contextos que os métodos tradicionais dificilmente alcançam, mas que exige também vigilância ética, transparência e uma reavaliação das práticas arquivísticas fundamentais. O mapa mental 2.2 ilustra as quatro dimensões da integração da IA em arquivos, que são respectivamente:

- Para além da automação, rumo à argumentação;
- Ética e transparência como elemento central;
- Redefinição de acesso e descoberta;
- O Arquivo do futuro e o futuro do arquivo.

A memória profunda, ao revelar camadas mais profundas de significado, não só amplia a compreensão do passado e do presente, mas também reconfigura o papel do arquivista, que passa a atuar como curador digital, mediando a interação entre algoritmos e utilizadores. Nesse sentido, a incorporação de tecnologias de IA em arquivos hospitalares não deve ser vista apenas como uma oportunidade técnica, mas como um desafio interdisciplinar que combina computação, arquivística e clínica para extrair valor real e sustentável da informação.

Como forma que compreender tópicos essenciais para o desenvolvimento de um SI capaz de capacitar os utilizadores a reutilizar a informação através de uma recuperação efetiva, foi realizado um levantamento bibliográfico em relação à evolução terminológica em anatomia patológica.

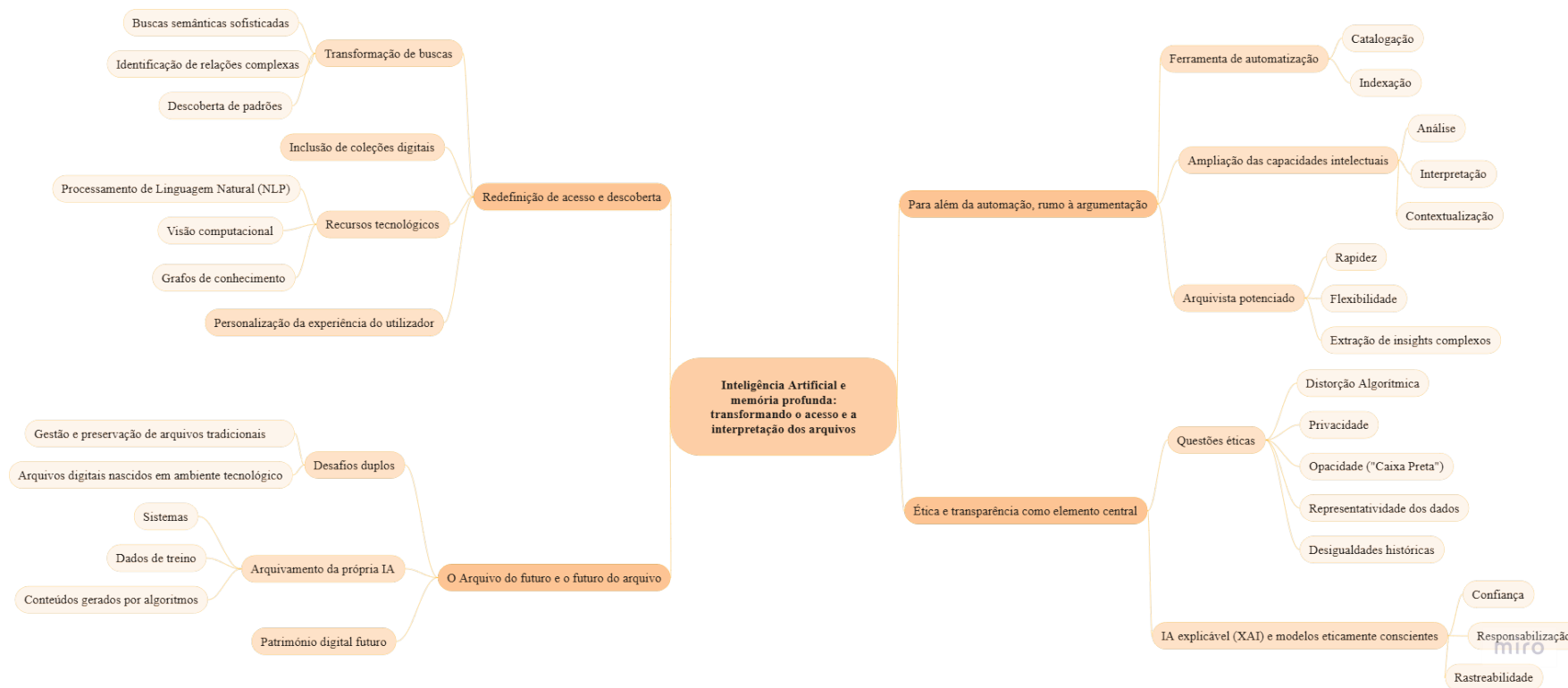


Figura 2.2: O uso da IA na transformação do acesso e da interpretação dos arquivos.
Fonte: Desenvolvido a partir de Gouveia, 2025.

2.2 Análise da evolução terminológica em anatomia patológica

Dada a importância dos arquivos clínicos para a prestação de cuidados e para a investigação, faz-se necessário haver uma representação e recuperação, a partir do tratamento da informação, para "construir interfaces entre os acervos de documentos, as informações e seus usuários"(Alvarenga, 2003). Para alcançar esse objetivo, um requisito é a padronização terminológica em saúde. Os dados, estruturados e controlados, disponíveis por meio de vocabulários controlados e terminologias clínicas, são ferramentas capazes de proporcionar esta padronização de termos para registo de dados clínicos.

O vocabulário controlado é uma linguagem desenvolvida para promover significado a organização da informação e composta por termos que podem ser organizados em uma estrutura relacional ou alfabética. O vocabulário controlado constitui-se, em sua forma mais simples, de uma lista de termos a serem empregados no processo de representação temática de um documento, ou seja, na indexação (National Information Standards Organization, 2010). Segundo Cunha et al., 2008, sua principal função é compatibilizar a linguagem empregada pelos pesquisadores à linguagem adotada em um sistema de informação, elaborada por um indexador.

Segundo Currás, 1995, vocabulário controlado “é uma lista de termos elaborada para identificar o assunto ou os assuntos de um documento com especificidade bastante para permitir sua recuperação rápida e eficaz”. Essa ferramenta fornecerá meios para organizar a informação, para: i) promover a tradução da linguagem natural para uma linguagem controlada para indexação e recuperação da informação; ii) viabilizar a consistência em relação ao formato dos termos; iii) indicar a relação semântica entre os termos; iv) fortalecer as hierarquias de um sistema de navegação, ou seja, para criação de etiquetas, para apoiar os utilizadores na localização da informação, e por fim v) promover a recuperação da informação por meio de pesquisa da informação (National Information Standards Organization, 2010).

Vale ainda abordar que o controle terminológico não está apenas preocupado com um simples gestão de palavras. As palavras na verdade são criadas a partir de um conceito e a partir delas surgem os **termos** (conceito > palavra > termo). Os termos, diferente das palavras, são "como uma unidade de pensamento que se configura em nosso cérebro, ao se colocar os neurônios em atividade pelos impactos de informação concebida do exterior, armazenadas para o uso posterior e que se expressam por palavras, de acordo com uma linguagem especializada e normalizada para cada parcela do conhecimento"(Currás, 1995). Ou seja, um termo possui apenas um significado e a palavra pode conter vários significados, a depender o contexto em que está inserida. Um termo é a expressão do significado de um conceito para comunicação entre pessoas de uma mesma área do conhecimento.

Os vocabulários controlados são de 3 tipos: esquemas de classificação bibliográfica, listas de cabeçalhos de assunto e tesouros. "Nas classificações, o arranjo alfabético é secundário, na forma de um índice que remete ao arranjo principal, que é hierárquico. No tesouro, o arranjo explícito dos termos é alfabético, mas existe uma estrutura hierárquica implícita, incorporada à lista alfabética por meio de remissivas. A tradicional lista de cabeçalhos de assuntos é similar ao tesouro por

ser de base alfabética, mas difere dele porque incorpora uma estrutura hierárquica imperfeita e por não distinguir claramente as relações hierárquicas das associativas"(Lancaster, 2004). Como exemplo de classificações, tem-se a Classificação Decimal Universal e a Classificação Decimal de Dewey. Uma das listas de cabeçalho de assunto mais conhecidas é a *Library of Congress Subject Headings* e o próprio *Medical Subject Headings*. Já os tesouros mais conhecidos são *Art & Architecture Thesaurus (AAT)* e o tesouro da UNESCO.

Segundo a International Organization for Standardization, 2013, delineia ainda que o objetivo principal de um tesouro é facilitar tanto o indexador quanto o pesquisador na seleção de terminologia preferencial idêntica ou combinações das mesmas para denotar um assunto específico. Consequentemente, um tesouro é meticulosamente estruturado para aprimorar a navegabilidade, garantir uma cobertura terminológica abrangente dentro de um domínio específico e padronizar os termos descritores, o que viabiliza uma melhor recuperação da informação.

Segundo Lancaster, 2004, para aumentar a probabilidade de recuperação bem-sucedida de um documento durante o processo de pesquisa, é essencial reconhecer que a implementação de estratégias de indexação, nomeadamente na fase de tradução, com base em um vocabulário controlado cuidadosamente selecionado é de extrema importância. Ao utilizar essa ferramenta metodológica, a eficiência geral da recuperação de informações é significativamente aprimorada, pois os utilizadores que realizam pesquisas com os termos controlados predeterminados terão uma probabilidade substancialmente maior de localizar as informações necessárias que procuram.

Os vocabulários controlados têm cinco funções críticas no domínio da organização e recuperação de informações: eles fornecem mecanismos de tradução, garantem a consistência em vários documentos e consultas, indicam relações entre conceitos, auxiliam na definição de campos específicos de investigação e facilitam a recuperação de materiais relevantes. Entretanto, essa ferramenta exige também que o utilizador saiba como pesquisar por ela ou utilizá-la de forma correta em seu próprio sistema de pesquisa. Esse aspecto é o que dificulta uma utilização plena de vocabulários controlados, muitas vezes eles ficam restritos ao uso de profissionais da informação em atividades específicas.

Levando-se em consideração ainda que os termos são tangíveis, porém não imutáveis, com o passar do tempo os termos descritores são modificados tanto na sua forma de escrita, quanto na sua organização hierárquica. Esse fenômeno ocorre em qualquer área do conhecimento, que se dá pela evolução da língua, mas também pelos avanços científicos que reestruturam os esquemas terminológicos e/ou hierárquicos. Portanto, um vocabulário controlado pode sofrer diversas mudanças terminológicas ou estruturais que requerem uma revisão para mantimento da padronização. Isto quer dizer que quando os termos são modificados, é crucial que suas versões anteriores sejam mantidas, como uma referência cronológica. Porém, estas mudanças quando constantes, tanto do termo quanto na estrutura do vocabulário controlado, geram instabilidade para sua utilização, pois demandam o esforço na reciclagem do aprendizado de forma mais constante.

Segundo Ghosh, 2017, a anatomia patológica, nascida entre os séculos XVI e XVII, com o contributo de Giovanni Battista Morgagni, introduziu as relações entre manifestações clínicas e estruturas anatómicas. Esse fator evidenciou a necessidade de estudos sistemáticos e classificação

das doenças com base nas observações realizadas. A necessidade de uma terminologia adequada e normalizada na anatomia patológica teve início com a classificação das doenças e das estruturas anatómicas. Os vocabulários controlados pioneiros foram conduzidos pela Organização Mundial da Saúde (OMS) e outros organismos internacionais para criar uma nomenclatura consistente e capaz de classificar as doenças de forma padronizada para todo o mundo, como a Classificação Internacional de Doenças (CID).

No Brasil, a evolução terminológica em anatomia patológica esteve fortemente vinculada a padronização e informatização iniciados na década de 1970. Um marco importante que ocorreu neste processo foi o trabalho desenvolvido na Seção de Anatomia Patológica do Instituto Nacional do Câncer (INCa), descrito por Castro et al., 1977. Nesta época, o setor já acumulava um acervo de mais de 150 mil relatórios histológicos registrados manualmente desde 1946, entretanto a recuperação dessa informação contida nos relatórios mostrava-se lenta e imprecisa, devido ao grande volume de casos, à diversidade de termos utilizados e à ausência de nomenclatura uniforme.

A implantação de um processamento eletrônico de dados trouxe a necessidade de incluir sistemas terminológicos e de codificação padronizados. O estudo do INCa analisou os principais vocabulários disponíveis na época, sendo eles a Classificação Internacional de Doenças (ICD-8), o Manual de Nomenclatura e Codificação de Tumores (MOTNAC), a *Systematizes Nomenclatura of Pathology* (SNOP) e a Nomenclatura do Hospital Anderson.

Cada um deles apresentava vantagens e limitações, o ICD-8 tinha foco estatístico e não atendia às necessidades específicas clínicas e diagnósticas da patologia. O MOTNAC permitia classificar tumores, mas ficava restrito às neoplasias. Já o SNOP era o sistema mais abrangente, contemplando topografia, morfologia, função e etiologia, o que possibilitava descrever tanto lesões neoplásicas como não neoplásicas. O SNOP, de 1965, é a versão que deu origem ao *Systematized Nomenclature of Medicine* (SNOMED) de 1974.

Diante deste cenário, o INCa optou por utilizar dois sistemas simultaneamente, o MOTNAC, para codificar neoplasias e o SNOP para patologias não neoplásicas. Para tornar mais simples as atividades cotidianas, os autores elaboraram listas de códigos adaptadas e um índice alfabético em português, de forma a reduzir ambiguidades. Essa experiência pioneira permitiu estruturar um banco de dados informatizado com estatísticas detalhadas sobre tumores primários e metastáticos, topografias, morfologias e lesões não neoplásicas, contribuindo tanto para a gestão da informação em saúde quanto para o ensino e pesquisa.

Dessa forma, o trabalho de Castro et al., 1977 representa um momento importante na história da terminologia da anatomia patológica no Brasil, marcada pela transição da gestão manual para a informatização de acervos, introduzindo a prática da codificação sistemática e evidenciando os desafios da interoperabilidade entre os diferentes sistemas terminológicos/vocabulários controlados. Ao mesmo tempo que buscava harmonizar padrões internacionais, como o ICD, SNOP e MOTNAC, também promoveu adaptações locais, refletindo a necessidade de articulação entre o rigor dos vocabulários e da indexação/codificação com a usabilidade prática na realidade do serviço.

Tabela 2.5: Terminologias e vocabulários controlados da área da saúde.

Nome do modelo	Responsável pelo desenvolvimento	Escopo
Terminologia Anatomica (TA)	Federative Committee on Anatomical Terminology (FCAT)	Termos da Anatomia em inglês e latim
TS22220:2011	International Organization for Standardization (ISO)	Identificação dos temas em cuidados de saúde
21090:2011		Tipos de dados padronizados para o intercâmbio de informações
13606		Descrição de alto nível dos modelos de informação clínica
23940 (ContSys)		Processos de cuidados de saúde para a continuidade dos cuidados
14155		Investigações clínicas
IDMP		Produtos medicinais
OpenEHR	Fundação openEHR	Especificação para modelos interoperáveis de informação clínica
SNOMED CT	SNOMED International (antiga Organização Internacional para o Desenvolvimento de Normas de Terminologia da Saúde)	Ontologia para representar o registo de saúde eletrónico, abrangendo termos clínicos padronizados
ICD-10 / ICD-11	Organização Mundial de Saúde (OMS)	Classificação de doenças para uso clínico e epidemiológico
ICF		Classificação de funcionalidade, deficiência e saúde
ICHI		Classificação de procedimentos de saúde
INN		Nomes genéricos de substâncias farmacêuticas
ATC		Classificação de ingredientes de medicamentos
MeSH (Medical Subject Headings)	National Library of Medicine (NLM)	Tesouro especializado em ciências da saúde
DeCS (Descritores em Ciências da Saúde)	Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME)	Tesouro especializado em ciências da saúde, com categorias exclusivas sobre Ciência e Saúde, Vigilância em Saúde, Saúde Pública, Homeopatia, e Medicinas Tradicionais, Complementares e Integrativas

Nas décadas seguintes, a trajetória iniciada pelo INCa encontrou continuidade com a evolução das terminologias internacionais. No âmbito dos vocabulários controlados destacam-se alguns mais utilizados no âmbito das ciências da saúde, como visto na tabela 2.5. O primeiro a ser citado, um vocabulário controlado internacionalmente reconhecido como aquele que provê uma lingua-

gem comum aos sistemas de informação clínica, amplamente utilizado na anatomia patológica, o SNOMED International.

Esse vocabulário controlado multilingue é um dos mais precisos na área da saúde e está em contínuo desenvolvimento de acordo com as diversas necessidades informacionais, com atenção as mudanças no contexto da informação clínica de saúde para o ambiente digital National Library of Medicine, 2025a. Ele permite ainda mapear outros sistemas de codificação, como o CID-10, o que facilita a interoperabilidade entre os sistemas e a qualidade na descrição dos registros eletrônicos de saúde (EHR). É possível verificar sua eficiente utilização em diversos estudos como o Souza, 2021. Entretanto, também demonstra lacunas quanto a tradução e efetividade para a língua portuguesa.

Destaca-se também a Classificação Internacional de Doenças (ICD), que consiste em um padrão global para informações de diagnóstico de saúde. Essa classificação é padronizada por códigos e é amplamente utilizada por profissionais da saúde, investigadores e gestores para padronizar diagnósticos, monitorizar epidemiologia e facilitar a comparação de dados de saúde entre países.

"O ICD é importante porque proporciona uma linguagem comum para o registro, a notificação e a monitorização das doenças. Isto permite ao mundo comparar e partilhar dados de uma forma consistente e normalizada - entre hospitais, regiões e países e ao longo de períodos de tempo. Facilita a recolha e o armazenamento de dados para análise e tomada de decisões baseadas em provas"(World Health Organization, 2025). O ICD se mostra fundamental para registros clínicos e prontuários, estatísticas de mortalidade de morbidade e definição de políticas públicas de saúde.

O MeSH (*Medical Subject Headings*) consiste em um tesouro especializado em ciências da saúde produzido pela *National Library of Medicine* (NLM). Ele é organizado por hierarquias e é utilizado para indexar artigos na Pubmed e MEDLINE (National Library of Medicine, 2025b). O tesouro multilingue DeCS (Descritores em Ciências da Saúde) é um vocabulário estruturado que foi criado pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, conhecido pelo nome BIREME, para servir como uma linguagem única na indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos, e outros tipos de material, assim como para ser usado na pesquisa e recuperação de assuntos da literatura científica nas fontes de informação disponíveis na Biblioteca Virtual em Saúde (BVS) como LILACS, MEDLINE e outras. Esse vocabulário foi desenvolvido a partir do MeSH com o objetivo de permitir o uso de terminologia comum para pesquisa em múltiplos idiomas, proporcionando um meio consistente e único para a recuperação da informação.

Quando se coloca o quesito saúde, os vocabulários controlados mostram-se ainda mais requisitados para a indexação de documentos nesta área. Os diferentes termos, muito famosos pela sua complexidade e especificidade, trazem um desafio de descrição temática de documentos da área da saúde, tanto para seu uso interno ou interdisciplinar. Para além desse fator, segundo Shortliffe et al., 2014, um problema crucial que exige uma alta precisão e exaustividade na descrição de documentos clínicos textuais é a imprecisão dos termos utilizados, uma vez que a linguagem utilizada não reflete necessariamente termos controlados. Ou seja, o uso de linguagem natural, ampliada pela variedade da linguagem clínica não controlada, dificulta uma descrição padronizada dos documentos clínicos.

Nesse quesito, os relatórios de anatomia patológica são um verdadeiro desafio, uma vez que são concebidos em sua maioria em texto livre e em linguagem natural, ou seja, com dados clínicos não-estruturados, e ainda com anotações manuais como complemento durante a análise da peça. Para além desse fator, a estrutura dos relatórios citológicos e histológicos diferem com o passar dos anos e depende da especialidade de requisição do relatório. Os dados clínicos não-estruturados serão, neste estudo, a oportunidade para coleta de termos descritores capazes de indexar por assunto/temática do documento.

2.3 Técnicas de enriquecimento de metainformação descritiva

O enriquecimento da metainformação descritiva é crucial por diversos motivos, sendo um deles a representação de objetos digitais, neste caso documentos arquivísticos de anatomia patológica, e consequentemente colaborar com a melhoria da recuperação da informação. A afirmação tem como referência o que foi evidenciado por autores como Gilliland, 2008; Haynes, 2018; Méndez et al., 2014 em suas obras. O uso de metainformação tem como funções principais, segundo Gilliland, 2008, 6 tópicos principais que acompanham o ciclo de vida de um objeto de informação, sendo eles:

- Criação, multiversionamento, reutilização e recontextualização de objetos de informação;
- Organização e descrição;
- Validação;
- Busca e recuperação;
- Utilização e preservação;
- Destinação.

O objeto de informação é o objeto conceptual da preservação. Um objeto de informação é percebido como informação significativa à interpretação de um objeto de dados por meio de sua representação associada. O objeto de dados é uma instanciação tecnológica de um objeto de objeto de informação, composto de um ou mais fluxos de bits e dependente de um ambiente técnico específico para fornecer acesso. Informação de representação é o conjunto de informações necessárias para interpretar um objeto de dados como um objeto de informação significativo, ou um componente de um ambiente técnico que oferece suporte à interpretação desse objeto (como uma ferramenta de software ou plataforma de hardware).

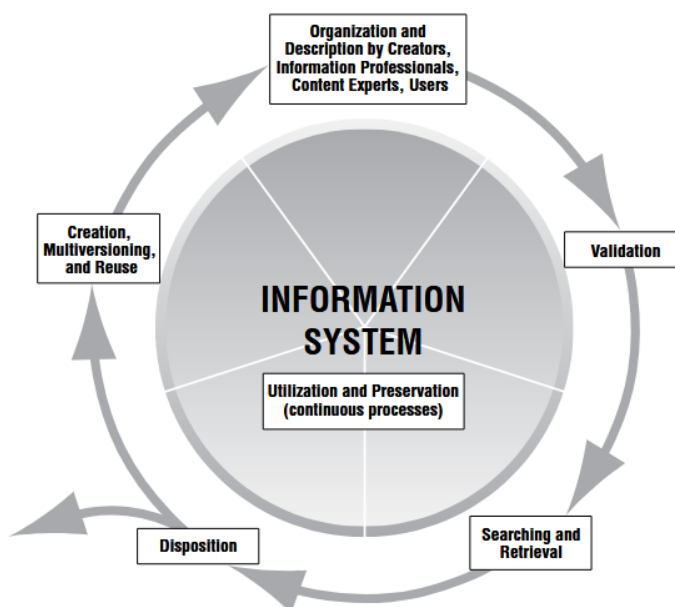


Figura 2.3: Ciclo de vida de um objeto de informação.

Fonte: Gilliland, 2008

A figura 2.3 representa o ciclo referenciado pela autora, com cada uma das fases e o uso da metainformação nesse processo. Estas funções visam, em suma, ampliar o acesso à informação e à própria metainformação, bem como reter o contexto ligado ao objeto, expandir o seu uso, potencializar o desenvolvimento do sistema de informação que o gere, preservar o objeto e/ou sua descrição ao longo do tempo com mudanças das plataformas tecnológicas, com foco na interoperabilidade entre sistemas, e assegurar questões legais que podem estar associadas ao objeto, como direitos de autor.

O uso de algumas tecnologias podem impulsionar fases deste ciclo, como na organização e descrição e na busca e recuperação. Com a evolução tecnológica, esse processo já não necessita ser feito apenas manualmente, apesar de necessitar de revisão por um ser humano de forma contínua. Algumas tecnologias a se destacar são:

- Reconhecimento Ótico de Caracteres (OCR), que consiste em uma tecnologia usada para converter diferentes tipos de documentos, como imagens digitalizadas, fotos de documentos ou PDFs, em textos editáveis. O OCR identifica os caracteres que aparecem na imagem e transforma-os em texto digital que pode ser pesquisado e editado. Ele é muito usado para digitalização de documentos físicos e automação de processos.
- Mineração de texto, refere-se à utilização de métodos automáticos para explorar grandes volumes de texto com o objetivo de identificar padrões, tendências e informações relevantes. Técnicas de mineração de texto são amplamente usadas para analisar mensagens em redes sociais, opiniões de clientes, notícias, e extrair conhecimentos úteis desses dados não estruturados.

- Processamento de Linguagem Natural (NPL) é uma área da inteligência artificial que cria métodos para que computadores compreendam, interpretem, processem e gerem linguagem humana, seja falada ou escrita. O PLN está presente em ferramentas como tradutores automáticos, chatbots, assistentes virtuais e sistemas de análise de sentimentos.
- *Machine Learning* consiste em algoritmos capazes de aprender a partir de dados, identificando padrões e realizando previsões ou decisões automaticamente. O machine learning é utilizado em detecção de fraudes, recomendações de produtos, reconhecimento de voz, reconhecimento facial e classificação de informações.

O estudo de Ramdhani et al., 2021, se concentrou no Reconhecimento de Entidades Nomeadas (NER), que consistem em uma técnica utilizada para identificar e extrair informações importantes de documentos em texto. Esse processo ajuda a categorizar entidades específicas, como o nome, data, local e outras entidades, ou seja, metainformação descritiva. Para tal, os autores utilizaram o OCR como ferramenta para extração de dados no texto, verificaram que a precisão não é perfeita, mas que a correta preparação dos documentos diminui as chances de falha no processo de reconhecimento ótico.

Já o estudo realizado por Burger et al., 2016, analisa diversas abordagens de NPL aplicadas a relatórios de patologia, demonstrando que a extração automática de dados estruturados a partir de textos livres pode melhorar significativamente o acesso e a organização da informação. Nesse estudo foram identificadas técnicas como a correspondência de palavras, os sistemas baseados em regras e o aprendizado de máquina probabilístico, mais conhecido por *machine learning*.

Outro estudo que destaca-se nessa área foi realizado por Ehsani et al., 2008 que descreve um projeto de digitalização e organização de registros patológicos arquivados, utilizando o OCR como ferramenta para conversão dos documentos em papel para arquivos pesquisáveis. A abordagem adotada incluiu a digitalização de registros, a extração automática da metainformação (número de caso, data, diagnóstico, por exemplo), a estandarização e filtragem de dados com base na OMS e a criação de um banco de dados integrados para facilitar as buscas e estudos epidemiológicos.

Outra forma de promover o enriquecimento da metainformação descritiva, complementar ao uso do OCR, é a mineração de texto. De acordo com Nguyen et al., 2020, esse processo ajuda na extração automática de características patológicas de relatórios da patologia clínica digitalizados. O OCR foi aplicado para converter os documentos em PDF para texto editável, com algumas técnicas para correção dos erros, e dessa forma aplicar a mineração de dados. A metodologia aplicada incluiu o uso de expressões para definir padrões de busca para extrair termos relacionados a leiomiomas, endometriose e adenomiose. A verificação da efetividade foi realizada a partir da extração manual e comparação da concordância entre os termos.

Em junho de 2023, foi apresentado o modelo português de transcrição automática de manuscritos com uso da plataforma **Transkribus**. O projeto "Transcrever os Processos da Inquisição Portuguesa", tem como objetivo criar um modelo de transcrição automatizada de documentos manuscritos portugueses dos séculos XVI ao XIX, para transcrever os processos da inquisição portuguesa, dentre 1536 até 1821.

O Transkribus é uma plataforma que conta com inteligência artificial, especificamente algoritmos avançados de *machine learning*, para reconhecimento de texto, análise de imagens e identificação estrutural de documentos históricos manuscritos. Desenvolvido por pesquisadores e arquivistas, o software oferece reconhecimento de texto com alta precisão, suporte multilíngue e ferramentas para treinar modelos específicos de caligrafia. "O Transkribus oferece tecnologia OCR de ponta para converter documentos manuscritos em texto digital, suportando uma ampla variedade de scripts históricos e modernos"(READ-COOP, 2022).

As suas funcionalidades principais são (i) reconhecimento automatizado de textos em mais de 20 idiomas, (ii) ferramentas de análise de *layout* para identificar estruturas complexas em documentos e (iii) capacidade de exportar resultados em formatos como XML e texto simples. Para além dessas, como referido por READ-COOP, 2022, destaca-se também algumas funcionalidades avançadas, como a análise do *layout*, que deteta automaticamente blocos de texto, imagens e elementos gráficos, o que facilita a segmentação de páginas com estruturas complexas.

O Transkribus se diferencia de outras ferramentas de OCR por ser especializado em documentos históricos e manuscritos, os quais geralmente não são bem aceites por ferramentas comuns de OCR, que tem como foco o processamento de documentos modernos. A plataforma permite que sejam inseridos ficheiros em .png, .jpg, .tif, .pdf dentre outros e realiza uma leitura e transcrição com uso de OCR a partir de modelos já existentes ou através de treinamento de modelos específicos.

Segundo a página web do projeto, TraPrInq, 2025, a paleografia digital foca em dar acesso público aos arquivos e aos documentos digitais por meio desse processo. A ferramenta automatizada por meio de HTR (Handwritten Text Recognition) possibilitou a transcrição com sucesso desses documentos que foram insatisfatoriamente transcritos por OCR comum. O modelo tem uma margem de erro de aproximadamente 5% em transcrições automáticas.

O enriquecimento da metainformação descritiva desempenha um papel fundamental no aumento da acessibilidade, reutilização e recuperação da informação dos arquivos. Como demonstrado através dos estudos citados, a integração de tecnologias avançadas como o OCR, NER e NPL contribuem significativamente para a gestão de metainformação. A literatura destaca ainda a importância de manter a integridade contextual ao mesmo tempo que expande-se o uso destes objetos digitais, garantindo a interoperabilidade entre sistemas.

Apesar do potencial demonstrado por diversas tecnologias para processamento automatizado de arquivos, como o OCR e mineração de texto, o uso de inteligência artificial (IA) no âmbito da saúde carece de regulamentação clara, dado o risco inerente ao tratamento de dados sensíveis da população em causa.

Em Portugal, O Serviço Partilhado do Ministério da Saúde (SPMS) recomenda boas práticas ao uso da IA no setor, considerando níveis de risco variados associados às diferentes aplicações em saúde. O documento publicado SPMS (2025) aborda o impacto do Regulamento da Inteligência Artificial (Regulamento (UE) 2024/1689) nos sistemas de saúde, descrevendo requisitos e recomendações para o seu uso ético. O documento cita as possibilidades de transformação

e otimização para fins de diagnóstico, estimativas de sucesso de tratamentos, apoio ao cuidado de pacientes e investigação clínica a partir do uso de ferramentas tecnológicas disruptivas.

Destacam-se benefícios operacionais, como a automação de atividades administrativas para diminuição da carga de trabalho dos profissionais de saúde em 10-30%, permitindo que se concentrem em atividades de maior valor clínico e interação direta com os pacientes. Esse fator colabora para um ambiente de saúde mais seguro e produtivo, o que diminui o risco de erro humano nas atividades rotineiras. O documento cita ainda o benefício do apoio à decisão clínica, com a IA agindo como apoio na análise de imagens médicas. Em Portugal, projetos nacionais que utilizam IA já afetam positivamente o estudo da fragilidade em idosos (FRAILCARE.AI) e a codificação clínica por *machine learning* (Codificação Previsional).

O segundo projeto citado chama atenção, uma vez que, "apresenta-se como uma funcionalidade de apoio ao codificador durante o exercício da codificação clínica em ICD-10-CM/PCS através da identificação e apresentação de códigos mais prováveis. Com este objetivo, desenvolve-se um algoritmo de aprendizagem automática, com a capacidade de analisar linguagem natural e identificar os códigos com base nas notas de alta e outros textos clínicos, oriundos do SClínico Hospitalar"(Núcleo de Saúde Digital Global e Relações Internacionais e Unidade de Cibersegurança, 2025).

Dentre as vantagens, identifica-se ainda a melhoria da qualidade dos dados, onde o Espaço Europeu de Dados de Saúde (EEDS) visa garantir que dados da saúde, como histórico clínico, imagens médicas e relatórios, por exemplo, sejam partilhados de forma interoperável e acessível em toda Europa. Uma base de dados bem estruturada e de qualidade é crucial para que os algoritmos de IA proporcionem resultados fiáveis e relevantes.

O EEDS foi recentemente regulamentado, em fevereiro de 2025, e em seu regulamento consta que "tem por objetivo criar o Espaço Europeu de Dados de Saúde, a fim de melhorar o acesso e o controlo pelas pessoas singulares relativamente aos seus dados de saúde eletrónicos pessoais no contexto dos cuidados de saúde, bem como atingir com maior eficácia outros fins relacionados com a utilização de dados de saúde eletrónicos nos sectores dos cuidados de saúde e da prestação de cuidados que possam beneficiar a sociedade, designadamente a investigação, a inovação, a elaboração de políticas, a preparação e resposta a ameaças sanitárias, incluindo a prevenção e combate de futuras pandemias, a segurança dos doentes, a medicina personalizada, as estatísticas oficiais ou as atividades de regulamentação"(Parlamento Europeu e Conselho da União Europeia, 2025).

No entanto, esses desenvolvimentos trazem também desafios importantes, como citado por Núcleo de Saúde Digital Global e Relações Internacionais e Unidade de Cibersegurança, 2025, sendo alguns deles a privacidade e segurança dos dados, a falta de transparência e questões relacionadas a supervisão por humanos. O balanceamento dos conjuntos de dados utilizados para treino, validação e teste podem induzir um modelo ao erro ou a reprodução de discriminação por género, sexualidade ou etnia, além de haver uma falta de transparência quanto aos modelos prontos e finalidade dos dados utilizados em produtos existentes no mercado.

A qualidade dos dados é outro fator que exige uma correta representatividade e completude,

uma vez que este é um aspeto fundamental para o bom funcionamento dos sistemas de IA. O documento aponta níveis de risco, sendo o inaceitável e proibido por lei aqueles sistemas que utilizam o reconhecimento facial em espaços públicos e sem consentimento ou IA para manipulação de comportamentos.

Adicionalmente, estudos na área da Ciência da Informação refletem preocupações similares no domínio da ética da informação e da IA aplicada à gestão documental, como analisado no artigo "Visualização de Inteligência Artificial e memória profunda: transformando o acesso e a interpretação dos arquivos". Gouveia, 2025 destaca como as tecnologias emergentes podem transformar o acesso a acervos históricos. Contudo, ressalta-se a necessidade de estratégias éticas e robustas para lidar com o risco de manipulação interpretativa, perda de contexto e comprometimento da integridade documental.

Esta convergência entre regulamentação, aplicação tecnológica e reflexão ética evidencia a importância de definir políticas claras e adaptadas ao contexto hospitalar. No caso da anatomia patológica, o uso da IA deve ser cuidadosamente avaliado quanto ao nível de risco, garantindo transparência, proteção de dados, rastreabilidade e governança documental adequada. Os resultados dos trabalhos revisados sublinham a necessidade de supervisão humana mesmo em um processo completamente automatizado, apesar dos avanços tecnológicos que reduzem os esforços manuais. Além disto, a aplicação de metodologias específicas, incluindo a mineração de texto e modelos de transcrição automática, como o Transkribus, mostram o potencial para melhorar este processo moroso de digitalização, transcrição e correção de OCR. Em última análise, esta revisão ilustra que a evolução contínua dos tratamentos técnicos em arquivos de anatomia patológica não só potencia a preservação da informação, mas também viabiliza a investigação e prestação de cuidados de forma célere.

Capítulo 3

Estratégia de investigação

Identificando-se como objeto de estudo desta investigação o arquivo de anatomia patológica do HUSJ, o mesmo tem como foco as dificuldades atualmente retratadas no tratamento e disponibilização em larga escala dos documentos, considerando relatórios citológicos e histológicos, datados desde 1962 até 2020. Como mencionado no capítulo 1, essa dificuldade prende-se com diversos aspectos que são abordados no âmbito da investigação, como a diferença de formatos de suporte e a falta de um padrão de estrutura e classificação dos documentos, colocando entraves ao processamento e acesso efetivo e ágil à informação contida nesses documentos por parte de clínicos e investigadores, bem como a ausência de um modelo de enriquecimento de metainformação descritiva.

3.1 Metodologia de investigação

No seguimento do que foi enquadrado enquanto objetivos da pesquisa e da informação de estado da arte levantada a partir das referências bibliográficas, a metodologia desta dissertação caracteriza-se como **investigação-ação, de abordagem qualitativa, de natureza aplicada e com objetivo explicativo**. A abordagem epistemológica a ser desenvolvida, por sua vez, está mais alinhada com o pós-positivismo, uma vez que existem padrões que devem levar em conta todo o contexto social em que a pesquisa está inserida. De referir a preocupação de replicação da mesma, sendo possível aplicar o modelo e os passos seguidos nesta investigação em outras realidades similares.

A metodologia selecionada tem como característica chave a melhoria do processo de enriquecimento da metainformação descritiva no RCD, a partir do tratamento e disponibilização do arquivo de anatomia patológica, com a participação da comunidade designada na evolução da pesquisa. O estudo de Almeida et al., 2019 aponta a "possibilidade de uso da pesquisa-ação em estudos com propósitos diversificados na Ciência da Informação, a exemplo da elaboração de tesouros, tutoriais, modelos, políticas de desenvolvimento de coleções, programas de competências em informação e políticas de informação".

Este estudo mostra o crescente uso da investigação-ação no âmbito da Ciência da informação a partir da análise de 29 artigos publicados nesta área de estudo e com esta metodologia entre 2005 e 2018. A partir deste estudo, entende-se que com essa metodologia é possível desenvolver pesquisa e estimular a comunidade envolvida a buscar soluções para os problemas encontrados, levantar dados, analisá-los e obter resultados, aplicando os conhecimentos da Ciência da Informação e buscando contributos para a evolução da própria ciência a partir das experiências práticas.

Segundo Bradbury, 2015, a investigação-ação tem como etapas a identificação do problema, o planeamento, a ação, a observação, a reflexão e por fim a revisão e replaneamento, sendo essa última etapa a referência para o processo cíclico que é proposto pela metodologia em causa. Koe-rich et al., 2009, acrescenta o processo mais detalhado que também compreende a coleta de dados e a sua análise, como mostra a figura 3.1.

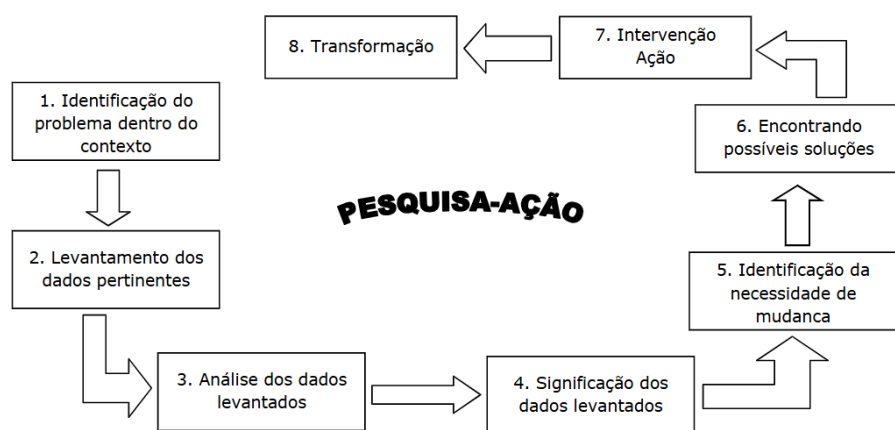


Figura 3.1: Fases da investigação-ação.

Fonte: Retirado de Koerich et al., 2009, p. 718

De acordo com Bradbury, 2015, conhecimento é gerado não apenas pelo pesquisador, mas também por todos os participantes em causa, sendo este um aspeto crítico desta pesquisa. Apesar de ser essencialmente técnica no que tange a indexação pela perspetiva da CI, a pesquisa visa compreender e considerar as dificuldades sentidas no Serviço de Anatomia Patológica na integração dos documentos em papel ao RCD e a recuperação da informação.

Portanto, a investigação-ação mostra-se a metodologia mais adequada, no sentido que o comportamento de pesquisa dos clínicos e investigadores norteiam o processo técnico em causa. Sem a participação ativa desses utilizadores, no contexto desta pesquisa, o processo técnico perde o seu valor, uma vez que a indexação visa principalmente a recuperação efetiva da informação para sanar a necessidade informacional do utilizador em relação aos relatórios.

Ainda segundo a obra previamente citada, o primeiro princípio referido na investigação-ação contemporânea é o "eu" como elemento relacional, e é explicado na metodologia pelo facto de estar em constante troca com os recursos sociais, emocionais, históricos e cognitivos. Nesse sentido, o que se pretende é compreender e considerar, para o desenvolvimento do processo técnico, a perspetiva do utilizador com intuito de caracterizar a necessidade informacional e como o utilizador a

interpreta.

O segundo princípio da investigação-ação contemporânea é referente aos sistemas visarem a sua integridade ao longo do tempo. Nesse sentido, o que se pretende é compreender o contexto técnico-científico à época da produção dos documentos em consideração, reconhecendo os processos de tratamento os quais eram desenvolvidos, como forma de identificar a forma de indexação dos documentos e os termos utilizados. A ideia é buscar a completude, de forma a desenvolver e aplicar uma estratégia que tenha em consideração todo o trabalho já realizado para tratamento dos documentos de anatomia patológica e desenvolva uma melhoria no processo de forma cíclica.

E o último princípio da investigação-ação contemporânea é a contribuição prática como característica primacial, que remete a explicação inicial sobre a escolha da metodologia em causa. Vale ainda inferir que apesar da metodologia utilizada ter como foco a solução de problemas emergentes reais, esta investigação se preocupa com fatores para além de apenas a digitalização dos relatórios em papel, tocando-se na capacidade de tirar partido deste manancial de documentação na sua atividade.

Estes princípios visam levar as pessoas a refletirem sobre a sua prática, explorando o relacionamento entre a teoria e a prática, além de focar na redução de problemas e na identificação de melhorias para o contexto. Para isto, envolve-se as pessoas no processo de mudança, ligando a investigação, a ação e avaliação. Em contrapartida, segundo Karim, 2001, a investigação-ação pode ser limitante em alguns aspetos, como a dependência da relação entre pesquisador e participante, a dificuldade de generalizar resultados para outras realidades devido a especificidade do contexto e a dependência da colaboração de participantes para o desenvolvimento da investigação.

3.2 Plano de investigação

Com base no processo realizado na investigação-ação, apresenta-se o desenho de pesquisa na tabela 3.1, sendo de referir que para o plano de ação proposto, o período de realização desta investigação compreende os meses de fevereiro a setembro de 2025. Todas as fases determinadas têm relação direta com as técnicas de coleta de dados e os objetivos específicos, os quais contribuem em conjunto para a satisfação do objetivo geral. As técnicas de coleta de dados foram definidas a partir daquelas que são habitualmente utilizadas na investigação-ação, segundo Traqueia et al., 2021. O princípio cíclico nesta investigação, devido ao prazo limitado de realização, é aplicado apenas nas fases 4 e 5, para verificação da efetividade da ferramenta com OCR para recuperação da metainformação descritiva e a eficácia de todo o processo aplicado à recuperação e precisão da informação no RCD.

A partir da execução de cada fase, pretende-se extrair resultados capazes de satisfazer os respetivos objetivos específicos recordados na tabela 3.1.

Tabela 3.1: Desenho do processo da investigação em 5 fases.

Objetivo geral	Definir uma política de tratamento técnico do arquivo de anatomia patológica do HUSJ				
Objetivos específicos	Propôr uma estratégia de enriquecimento metainformação descritiva dos relatórios citológicos e histológicos no Repositório Clínico Digital (RCD)				
		Realizar um levantamento e sistematização de termos de indexação baseados em linguagens controladas, tendo por base o estudo da evolução da terminologia utilizada nos relatórios ao longo do tempo	Explorar o potencial do OCR no enriquecimento de metainformação e identificação automática de termos, tendo por base o tratamento de uma amostra de relatórios	Enquadrar o comportamento de pesquisa e a necessidade informacional dos clínicos e investigadores em relação aos relatórios citológicos e histológicos	
Coleta de dados	Fase 1	Fase 2	Fase 3	Fase 4	Fase 5
	Levantamento de dados sobre o processo de tratamento dos documentos com os profissionais dos Serviços de Arquivo e de Anatomia Patológica. Preparação das amostras para a investigação	Análise dos relatórios de Anatomia Patológica. Recolha e definição de termos descritores a partir de vocabulários controlados e comparação com termos cronologicamente respetivos	Leitura com OCR e avaliação da efetividade da ferramenta	Entrevistas aos clínicos e investigadores para compreender a necessidade informacional e o comportamento de pesquisa	Teste de pesquisa no RCD com clínicos e investigadores para avaliar o nível de recuperação e precisão da informação pesquisada
Técnica de pesquisa	Observação participativa, não-estruturada e naturalística	Pesquisa documental	Observação sistemática	Entrevistas semi-estruturadas	Observação sistemática
Fonte de dados	Processo de tratamento documental – Dra. Joana Gomes; Visita ao Serviço de Anatomia Patológica – Maria José e Joel	100 relatórios citológicos e histológicos de anatomia patológica e SNOMED CT	Transkribus e 100 relatórios citológicos e histológicos	9–12 Clínicos e investigadores	9–12 Clínicos e investigadores e RCD versão teste
Dados	Notas descritivas e reflexões	Metadados, notas descritivas, notas reflexivas, termos descritores identificados, termos controlados	Tempo de leitura, nº de linhas transcritas, nº de caracteres corretamente lidos	Transcrições	Termo de pesquisa, número de resultados, precisão dos resultados antes e depois do enriquecimento de metainformação
Ferramentas de tratamento dos dados	Excel	Excel	Excel	NVIVO	Excel
Tratamento dos dados	Análise temática	Análise documental e de conteúdo	Avaliação estatística descritiva (média, percentual de erro, taxa de acerto)	Análise temática	Análise descritiva e avaliação qualitativa de usabilidade
Resultado esperado	Quadro-resumo	Tabela comparativa com descrição dos relatórios e respetivos termos controlados	Gráficos e tabelas com métricas de desempenho	Gráficos, tabelas e quadros com comparações, destaques de falas e dados pertinentes	Quadro comparativo com níveis de precisão e recall de pesquisa com síntese da perceção do utilizador

Fonte: Adaptado de Santos et al., 2023.

A **fase 1** refere-se ao enquadramento, para melhor percepção do contexto atual de como os profissionais da informação procedem com o tecnicismo de descrição dos documentos e digitalização dos mesmos no RCD. A partir dessa etapa foi possível compreender como a metainformação é atribuída aos relatórios, bem como os relatórios são inseridos no sistema.

A **fase 2** é a análise documental e recolha e definição dos termos descritores a partir dos vocabulários controlados previamente pesquisados na revisão de literatura, SNOMED. Após a definição dos termos, foi realizada uma comparação cronológica com os respetivos termos utilizados nos relatórios selecionados como amostra para o tratamento técnico.

A **fase 3** é dedicada à leitura dos documentos com recurso à tecnologia OCR para extração da informação e indexação, sendo este exercício realizado no ambiente de teste do RCD. Para esta pesquisa é considerada uma amostra de 10 documentos por período temporal, ou seja, 100 relatórios citológicos e histológicos. Os documentos amostrais foram selecionados a luz do estudo realizado por Smith et al., 2011, onde a continuidade temporal era critério de seleção, mas diferentemente do outro critério utilizado no estudo em relação à boa condição do papel, as marcas de produção dos relatórios são mais importantes para este contexto. A partir da análise dos resultados obtidos, foi realizada uma avaliação para validar a eficácia do uso do OCR como ferramenta de identificação de metainformação, pretendendo-se utilizar uma análise da precisão, relevância e tempo de processamento.

A **fase 4** conta com entrevistas aos clínicos e investigadores e visa compreender o comportamento de pesquisa e a necessidade informacional dos utilizadores alvo desta pesquisa. Inclui a realização de entrevistas semi-estruturadas, com base em um questionário guião A, tendo os entrevistados a abertura para contribuir com tópicos para além das perguntas propostas. Dessa forma a qualidade da pesquisa é garantida, uma vez que a investigação-ação pressupõe uma participação mais ativa da comunidade designada. Nessa fase, são retomadas as fases anteriores, considerando o que foi dito durante as entrevistas e adicionando o contributo dos patologistas em relação a terminologias que se tornaram obsoletas ao longo dos anos e a estruturação dos relatórios.

A **fase 5** inclui o teste de pesquisa no RCD, em ambiente de teste do repositório, procurando-se clínicos e investigadores para avaliar o nível de recuperação da informação, bem como a precisão do resultado obtido. As duas últimas fases visam responder ao objetivo específico de exploração do OCR no enriquecimento da metainformação e identificação automática de termos. De acordo com a metodologia selecionada para a pesquisa, a fase 5 não se caracteriza como o fim do processo, dado que se trata de um processo cíclico de constante revisão.

3.3 Técnicas de pesquisa e análise dos dados

Com objetivo de descrever os processos delineados para coleta e análise de dados durante a investigação, este subcapítulo conta com conjunto de técnicas definidas a partir da literatura científica levantada, à luz de Bradbury, 2015, Lakatos et al., 2017 e Traqueia et al., 2021. Como sinalizado no quadro 3.1 do desenho da investigação as técnicas utilizadas são:

- Observação participativa, não estruturada e naturalística;

- Pesquisa documental com documentos escritos primários;
- Entrevistas semi-estruturadas.

Para descrever e armazenar os dados coletados a partir da observação participativa, na fase 1, é utilizado um protocolo de observação com base em Creswell et al., 2018. O protocolo consiste em um quadro com data, notas descritivas e notas reflexivas, que após análise e limpeza dos apontamentos, deu origem às tabelas 4.1, 4.2 e 4.3. A observação conta com algumas questões orientadoras sobre processos específicos que são previamente formuladas para orientar a observação no contexto da investigação e no decorrer do processo, pretende-se elucidar outras questões associadas ao contexto de investigação que venham a surgir.

A pesquisa documental segue o processo de leitura documentária, sustentada pelo processo de indexação realizado na CI. "Os modelos de leitura documentária têm um papel fundamental para a indexação, pois contribuem para o direcionamento da leitura documentária, análise de assunto, identificação e seleção de conceitos, além de orientar a metacognição do indexador nas variadas tipologias textuais, segundo suas estruturas e temáticas específicas"(Fujita et al., 2020). Nesse sentido, o processo de leitura documental é o primeiro passo para a indexação e elaboração de resumos, como Lancaster, 2004 também afirma em sua obra.

A técnica de análise documental também visa o conteúdo, uma vez que identifica e levanta principalmente os termos descritores utilizados em cada documento. O protocolo utilizado para a análise tem como objetivo o apontamento de características (i) identificativas, (ii) estruturais e (iii) terminológicas. Os três aspetos visam a compreensão organizacional, contextual e descritiva dos documentos.

As entrevistas foram conduzidas com suporte do guião, que se encontra no apêndice A e foram analisadas com a ferramenta transkriptor para transcrição e correção das entrevistas. A fase de análise qualitativa foi realizada no NVIVO e foi utilizado o COLAB para gerar gráficos.

Nos próximos capítulos, 4 e 5, é possível verificar os resultados obtidos a partir da metodologia aplicada e a discussão que os envolve, sendo abordadas a contextualização e a caracterização do arquivo de anatomia patológica do HUSJ, bem como o processo e resultados obtidos de cada fase planeada.

Capítulo 4

A dinâmica informacional do arquivo de anatomia patológica do HUSJ

O arquivo de anatomia patológica do Hospital Universitário São João (HUSJ) representa um repositório crucial de documentos clínicos que sustentam a prática diagnóstica e assistencial na área da anatomia patológica. A gestão e organização desses documentos, que abrangem relatórios citológicos e histológicos produzidos ao longo de décadas, enfrenta desafios significativos devido ao seu volume expressivo, diversidade de formatos, dinâmicas de produção e uso da informação ao longo do tempo, colocando constrangimentos na recuperação da informação. Este capítulo visa apresentar uma análise detalhada da dinâmica informacional deste arquivo, abordando aspectos desde a sua estrutura física e digital, passando pelos processos de preparação, digitalização e indexação, até ao uso e recuperação da informação em contexto clínico e de investigação. Destacam-se as atividades de produção e uso dos relatórios, as dificuldades e evoluções tecnológicas ocorridas ao longo do tempo no Serviço de Anatomia Patológica do HUSJ.

4.1 Uso e recuperação da informação no contexto do Serviço de Anatomia Patológica

Como forma de dar início a compreensão do uso e da recuperação dos relatórios, realizados no serviço de Anatomia Patológica, foi realizada uma visita ao Serviço para conversa com a assistente técnica e o técnico de informática do Serviço de Anatomia Patológica. A partir desta visita foi possível compreender a linha temporal de sistemas e métodos adotados no Serviço, ilustrados na figura 4.1.

A gestão de dados e relatórios na Anatomia Patológica do Hospital São João passou por uma evolução significativa, desde processos totalmente manuais até a informatização e a integração com sistemas hospitalares centrais como o JOne e o SONHO. A partir das observações participativas com a assistente técnica e do informático, foi possível desenhar uma linha temporal representada na figura 4.1 que ilustra essa evolução.

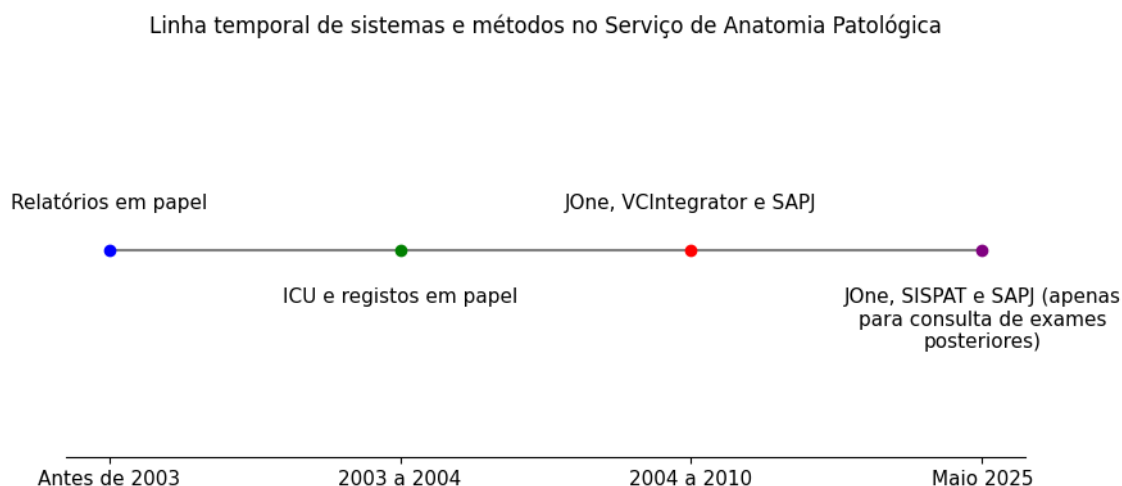


Figura 4.1: Linha temporal de sistemas e métodos no Serviço de Anatomia Patológica.

No período desde 1959 a 2003, antes da informatização, o serviço contava apenas com relatórios em papel e entrega manual. Antes da implementação de sistemas eletrônicos como o ICU, os relatórios eram exclusivamente em papel. Após serem "protocolados" (fechados e disponibilizados), eram impressos diariamente e um funcionário da Anatomia Patológica levava-os fisicamente aos diversos serviços do hospital. Para garantir o registo, existia um livro de registo onde a pessoa que recebia os exames assinava, atestando a entrega.

A organização e acesso aos exames eram feitos através de um código de exame (ex: H25/327 para histologia de 2025, exame nº 327), que era a "chave primária" e sempre presente em todos os exames. Paralelamente, existiam fichas catalográficas por nome de doente, onde se anotavam os números dos exames que o doente tinha realizado ao longo dos anos. Essas fichas permitiam recuperar os relatórios físicos para consulta do histórico do doente.

Existia um processo paralelo ao serviço, mas que neste período era responsabilidade do mesmo, o Registo Oncológico. Para este registo era utilizado o ICD-O (International Classification of Diseases-Oncology, a versão ICD para doenças oncológicas). Este registo era usado para monitorizar a sobrevivência de pacientes com cancro, o que envolvia o envio de cartas periódicas a pacientes ou mesmo a padres de aldeias para obter informação sobre o estado de saúde. Este processo de Registo Oncológico era muito anterior à informatização completa.

Em 2003 deu-se início à informatização do Serviço de Anatomia Patológica, a qual começou com uma patologista conceituada e muito interessada em tecnologia e sistemas. Inicialmente, os dados do doente eram transferidos via disquete para o computador. Os dados iniciais não vinham do SONHO (repositório de dados do doente do hospital), sendo preenchidos manualmente. O informático continuou e desenvolveu o software a partir do que já existia em colaboração com a doutora e a assistente técnica e coordenadora do Serviço.

Apesar da informatização e mais tarde a produção de relatórios em formato nado-digital, as notas manuscritas, de suporte ao trabalho de análise da peça ou por diversos motivos que serão melhor abordados no capítulo seguinte, podem ter uma existência e relevância verificadas apenas

por confrontação com os documentos em papel. Este é um dos factores diferenciais neste arquivo, pois conforme o sistema evoluiu, as notas manuscritas ficaram dependentes apenas do documento físico. A pertinência destas notas para os patologistas implica a importância da sua informatização em larga escala para viabilizar o acesso sem a necessidade de análise dos documentos em papel.

Entre 2003 e 2004 houve um período de transição para o ICU, antes da coexistência JOne/VCIntegrator. A Anatomia Patológica começou a enviar dados para o ICU (mais tarde renomeado para VCIntegrator). Numa primeira fase, os relatórios finais eram enviados para os clínicos tanto em papel quanto para o ICU em simultâneo. Posteriormente, deixou-se de enviar em papel e os dados passaram a ser enviados apenas para o ICU. O ICU (e mais tarde VC Integrator) foi desenvolvido pela faculdade de medicina da Universidade do Porto. Durante esta fase, o software desenvolvido pelo informático (não batizado, por isso refere-se como SAPJ), gerenciava os relatórios internamente e era responsável por enviar os dados para o ICU.

A partir de 2004, criou-se o JOne, um software do hospital, desenvolvido por uma equipa interna do Hospital São João. Nesta fase, o SAPJ passou a enviar os relatórios tanto para o VCIntegrator (antigo ICU) quanto uma cópia para o JOne em simultâneo. Ao receber uma requisição, os clínicos usavam o número do JOne. O SAPJ dava entrada no exame usando esse número para importar informações do doente do SONHO para o seu sistema. O SAPJ utilizava o SNOMED, e ao longo do tempo foram criados códigos personalizados (ex: para quadrantes da mama) que não existiam na versão SNOMED da altura, ou seja, códigos que foram adaptados à realidade do serviço. A tradução do SNOMED para português foi feita pelos internos do serviço de Anatomia Patológica, o que ajudou na utilização dos códigos, mas afetou a padronização da linguagem controlada que codificava os relatórios.

A partir de 13 de maio de 2025, a Anatomia Patológica começou a utilizar um novo sistema informático, o SISPAT. Esta é uma plataforma de gestão global de todos os dados e imagens gerados no laboratório de Anatomia Patológica, permitindo a aquisição, gestão, visualização e análise de todas as etapas do exame, desde a entrada no serviço à saída do diagnóstico. Este sistema se encontra em processo de implementação, portanto a perceção do sistema durante a produção desta investigação reflete um período de adaptações intensas por parte do secretariado e dos patologistas. Desde maio de 2025 utiliza-se o JOne, SISPAT e SAPJ, sendo o antigo sistema interno apenas para consulta de exames anteriores de doentes que ainda não foram migrados para o novo sistema.

Com o SISPAT, os relatórios da Anatomia Patológica já não são enviados para o VCIntegrator, sendo agora enviados exclusivamente para o JOne. Este novo sistema trabalha com o SNOMED CT, sua versão mais recente. Os clínicos já estão a usar o SISPAT para macroscopia¹ e microscopia, e os novos códigos (incluindo para topografia) já estão a ser utilizados.

O SAPJ ainda está em uso para os casos antigos. Relatórios feitos ou com entrada dada no programa antigo continuam a ser processados nele. A necessidade de tratar os dados antigos do SAPJ

¹Segundo Varma et al., 2024, no contexto da patologia, a análise macroscópica refere-se ao exame de amostras que podem ser vistas a olho nu, por oposição ao exame microscópico, que requer um microscópio para ver pequenos pormenores. Trata-se de avaliar o aspeto geral, o tamanho, a forma e a cor das amostras de tecido.

(que usavam termos e códigos SNOMED diferentes) para que sejam compreendidos no contexto do SISPAT (que usa o SNOMED recente) é crucial. A ideia é que o hospital abandone o VCIntegrator e utilize apenas o JOne. O SClínico do hospital (que tem conexão com o sistema nacional) também consulta o JOne. Esta linha temporal ilustra a complexidade da transição digital, as diferentes etapas de integração e a importância da gestão dos dados e da sua metainformação para interoperabilidade em quaisquer dos sistemas utilizados no hospital, com intuito de resguardar a informação e viabilizar o seu acesso a longo prazo.

A partir da observação realizada no Serviço, com anotações presentes nas tabelas 4.1, 4.2 e 4.3 de protocolo de observação, enriquecida pelo contributo de ambos os participantes que já fazem parte do Serviço há muito anos, foi possível compreender que o vocabulário controlado utilizado é de facto o SNOMED, mas em uma versão demasiado antiga e que foi modificada pelos próprios como forma de adequar os termos a realidade do Serviço. Ou seja, os termos utilizados não refletem o que oficialmente está no vocabulário controlado e isto afeta a indexação, chamada de codificação pelos patologistas e médicos codificadores do HUSJ.

Tabela 4.1: Análise crítica das observações sobre uso de terminologias e pesquisa.

Observação	Análise crítica / Impacto
1. Relatórios codificados pelo SNOMED (morfologia, topografia, procedimento e modificadores).	Garante a descrição, mas depende da versão usada; se não atualizada, compromete interoperabilidade futura.
2. Listagem traduzida dos termos mais utilizados, baseada no SNOMED antigo.	Facilita pesquisa local, mas gera risco de desvio em relação ao padrão internacional, dificultando integração com SNOMED CT.
3. Os termos dos relatórios são pesquisáveis no SAPIJ. É possível pesquisar por código ou descrição e retirar um excel com todos os relatórios que contém aquele termo, mas é necessário filtrar e rever os resultados por falta de fiabilidade na codificação.	Processo moroso, com esforço humano elevado e risco de inconsistência; modelo deve prever automação da triagem.
4. Códigos SNOMED não visíveis ao patologista, apenas termos descritores.	Facilita usabilidade.
5. SNOMED considerado mais completo para classificação dos casos.	Reforça pertinência de adoção no modelo, mas exige alinhamento à versão mais recente.
6. SNOMED adaptado à realidade do serviço, traduzido de inglês para português pelo secretariado e com listas de termos de uso frequente.	Garante aplicabilidade local, mas dificulta a interoperabilidade e padronização; modelo deve prever equivalências terminológicas.
7. Terminologia consistente ao longo das décadas.	Consistência interna é positiva, mas pode dificultar a adaptação à nova versão de termos controlados; modelo deve prever atualização periódica.
8. SNOMED usado mais para pesquisa científica que para clínica.	Modelo deve promover equilíbrio entre pesquisa e apoio ao diagnóstico.

Tabela 4.2: Análise crítica das observações sobre marcas de produção.

Observação	Análise crítica / Impacto
9. Relatório “E” como primeiro nível de relatório complementar (até “Z”).	Deve-se processar os documentos como compostos.
10. Apenas relatórios extemporâneos são manuscritos.	Reduz complexidade, mas exige OCR/HTR para integração; modelo deve prever captura digital desses casos.
11. Protocolos estruturados para estômago, mama e ginecologia (narrativo vs estruturado).	Os protocolos de mama e ginecologia viram relatórios estruturados com base no protocolo, já os relatórios de estômago viram texto narrativo, mas com base na estrutura do protocolo. Ou seja, o texto é narrativo, mas segue sempre uma sequência do que foi preenchido no protocolo.
12. Quando existe um pedido de levantamento de informação, procede a pesquisa pelo órgão > encontra o código > pesquisa pelo código > encontra termos relacionados ao órgão de referência.	Reforça valor do controle terminológico.
13. Relatórios ainda não têm estrutura única, mesmo com protocolos.	Ausência de uniformidade compromete a atribuição de metainformação automatizada; modelo deve propor estrutura flexível, mas consistente.

Tabela 4.3: Análise crítica das observações sobre circuito de produção e uso.

Observação	Análise crítica / Impacto
14. São enviados ao IPATIMUP aqueles muito específicos, como para decisão terapêutica (oncologia), apuração de diagnóstico ou esclarecimento de dúvidas. Geralmente não é feito um novo relatório, apenas é anexado o que foi enviado ao processo. Não configura um novo relatório, é apenas anexo.	Modelo deve integrar relatórios externos com preservação de contexto.
15. Atualmente está em teste um sistema de reconhecimento de voz e transcrição para o processo de macro e microscopia (INVOX medical). Vem se mostrando promissor e serve como um pré relatório.	Potencial para acelerar produção e enriquecer metainformação, mas precisa de validação terminológica.
16. Não há um sistema de OCR utilizado atualmente, o que estava no VCintegrator já não funciona mais. Não existem testes feitos com mineração de texto (NLP).	Limitação para automação.

As observações apresentadas nas três tabelas evidenciam que o modelo de enriquecimento da metainformação precisa lidar com múltiplos desafios: (i) a heterogeneidade terminológica decorrente do uso de versões antigas e adaptadas do SNOMED, (ii) a coexistência de relatórios

estruturados e manuscritos, que exige estratégias híbridas de normalização e leitura OCR, (iii) a necessidade de interoperabilidade com outros serviços hospitalares e instituições externas, e (iv) a limitação tecnológica atual, marcada pela ausência de OCR funcional.

Após a observação inicial e posteriores visitas ao Serviço, foi possível perceber também que ao contrário do que ilustra a figura 2.3, do ciclo de vida de um objeto de informação, o Serviço de Anatomia Patológica gere os seus relatórios também com alguma diferença. Existem 3 tópicos que chamam a atenção neste quesito:

1. Existe uma informatização manual dos relatórios por parte do secretariado quando são realizados manualmente pelo patologista;
2. Não existe fase de validação quando isto ocorre, o que implica em um relatório transcrito por uma segunda pessoa que não o patologista;
3. Não há um procedimento ou regras quanto a disseminação para outros sistemas de informação, desde que foi implementado o novo sistema de gestão do Serviço, o que causa falhas no processo de integração com o sistema de registo oncológico (RON) e com o JOne.

A compreensão sobre a organização do Serviço e dos seus patologistas foi enriquecida pela primeira das entrevistas realizada com um patologista, assistente hospitalar graduado sénior, com 28 anos de trabalho no HUSJ e especialista na área histológica, onde percebeu-se que o Serviço está dividido em setores, um dos quais é o setor da citologia, o outro é da histologia e o das autópsias como setores principais. Outros setores também constituem o Serviço, aqueles que dizem respeito a exames complementares de diagnóstico. O capítulo seguinte descreve no que consiste estas áreas, contextualizando como o arquivo da Anatomia Patológica está organizado atualmente.

4.2 O arquivo de anatomia patológica do HUSJ

O arquivo do HUSJ é atualmente responsável pelo tratamento e disponibilização de parte dos documentos, relatórios citológicos e histológicos da anatomia patológica, selecionados para análise e tratamento nesta pesquisa. Os documentos estão atualmente organizados por períodos cronológicos, identificando-se distintos formatos de produção dos mesmos. É possível visualizar na tabela 4.4 como essa organização se apresenta atualmente e as características dos documentos em causa, tendo presente amostras dos relatórios.

O relatório citológico ou histológico são documentos que compreendem diferentes metodologias e objetivos diagnósticos. Um exame citológico visa uma análise de células isoladas ou pequenos aglomerados celulares obtidos de fluidos, secreções ou aspirações, sem a preservação da arquitetura tecidual. É com base na citologia, que se investiga as características morfológicas das células. Alguns exemplos de exames como estes são o papanicolau, utilizado para rastrear lesões precursoras do carcinoma cervical, a aspiração com agulha fina (FNA) de nódulos tireodíanos, mamários ou linfonodais e o exame citológico de fluidos cavitários, como derrames pleurais ou peritoneais, para detectar células neoplásicas.

O relatório citológico por regra compreende a identificação e os dados clínicos, a descrição macroscópica, a descrição microscópica das células observadas e a conclusão diagnóstica, que pode indicar a ausência de malignidade, a presença de lesões de baixo ou alto grau ou a suspeita de neoplasia, e também pode incorporar um comentário com orientação complementar.

Já o relatório histológico, por outro lado, surge da análise de fragmentos ou constituintes inteiros do tecido, processados em blocos de parafina, seccionados e corados para reter a arquitetura tecidual. Os exemplos incluem biópsias de pele para diagnosticar carcinoma basocelular, biópsias gástricas para avaliar gastrite e *helicobacter pylori*, amostras cirúrgicas de mastectomia para estadiamento do câncer de mama ou ressecções intestinais para avaliar doenças inflamatórias e tumores.

O relatório histológico geralmente apresenta dados clínicos e de identificação, descrição macroscópica da amostra, descrição microscópica detalhando padrões, tipos de células, grau de diferenciação, margens cirúrgicas e invasão linfovascular, além do diagnóstico conclusivo, que delinea o tipo histológico, o grau e a extensão da lesão. Quando aplicável, também engloba o estadiamento da neoplasia, como o sistema TNM (Tumor primário, Linfonodo e Metástase) e observações adicionais, como a indicação de investigações imunohistoquímicas.

Em resumo, a citologia opera com células isoladas e é vantajosa para a triagem e o diagnóstico minimamente invasivos, e é acelerada, enquanto a histologia examina os tecidos preservando a arquitetura, facilitando o diagnóstico e o estadiamento definitivos, embora necessite de mais tempo de processamento. Ambos são complementares na prática do patologista.

O tamanho do arquivo de anatomia patológica acumulado durante todos esses anos é significativo, com aproximadamente 120000 relatórios. Devido ao alto custo financeiro e de tempo para tratamento em massa destes relatórios, optou-se por implementar um processo de digitalização e acesso por demanda, ou seja, atualmente a forma de acesso a esses documentos passa pela solicitação prévia por parte dos clínicos. A partir do pedido, o Serviço de Arquivo digitaliza, classifica o documento, insere no RCD e dessa forma o acesso é disponibilizado aos clínicos do Serviço. A nível de satisfação, todo esse processo faz com que os clínicos manifestem dificuldades relativamente ao pedido para ter o acesso, esse entrave para recuperação da informação desejada desperta insatisfação nos utilizadores em questão.

Tabela 4.4: Natureza e dimensão do arquivo de anatomia patológica.

Amostra	Período temporal	Tipologia	Natureza	Dimensão
Amostra nº1	1962 a jun. 1976	Histológico	Relatórios datilografados (formato A5, frente e verso com informação) com numeração manuscrita, carimbo e nº; inclui existência de relatório complementares (agrafado ao 1º, tendo ambos os relatórios o mesmo nº atribuído).	120 582
Amostra nº2 e nº3	Jul. 1976 a 1996	Histológico	Relatórios datilografados (formato A5, frente e verso com informação) com numeração manuscrita por ano (inicia sempre nº 1/ano); inclui existência de relatório complementares (agrafado ao 1º, tendo ambos os relatórios o mesmo nº atribuído).	180 539
Amostra nº4 e nº5	1997 a 2001	Histológico	Relatório (formato A4, inclui notas nos versos) inclui autocolante com nº e código de barras. Nota: parte do ano 1997 tem relatórios dobrados.	60 812
Amostra nº6	2002 a 2011	Histológico	Relatório com vinheta de identificação com nº e código de barras; últimos anos inclui requisição e relatório no verso.	196 018
Amostra nº6	2012 a 2020	Histológico	Impressão da requisição JONE e relatório de exame disponível no VCIntegrator com notas manuscritas.	260 758
Amostra nº7 e nº8	1977 a 1996	Citológico	Relatórios datilografados (formato A5, frente e verso com informação) com numeração manuscrita, nº sequencial por ano carimbo.	124 390
Amostra nº9	1997 a 2002	Citológico	Relatórios (formato A4, frente e verso com informação) com numeração manuscrita por ano; inclui relatórios complementares e registros adicionais agraphados.	51 830
Amostra nº10	2003 a 2020	Citológico	Relatório (formato A4; inclui notas nos versos, com aumento do nº de versos em branco); inclui autocolante com nº e código de barras; inclui relatórios complementares agraphados; anos mais recentes inclui requisições no verso.	191 310

É de se notar que o arquivo não se encontra em total custódia do Serviço de Arquivo. Os relatórios datados entre 2003 a 2018 se encontram no Arquivo, portanto aqueles anteriores a 2003 estão no Serviço de Anatomia Patológica, fator que adiciona mais entraves na recuperação do relatório em algumas situações.

O diagrama 4.2 relata uma atividade relacionada ao pedido de documento realizado por um clínico da cirurgia geral que ao atender uma doente na urgência teve conhecimento que a doente já havia realizado uma cirurgia no HUSJ, relativamente à qual existam resultados de Anatomia Patológica. O diagrama deixa claro que o processo passa por diversas fases, no caso do respectivo relatório já estar no RCD, indexado de forma a ser recuperado no durante a própria consulta/urgência, o processo de prestação de cuidados seria potencialmente mais ágil e efetivo.

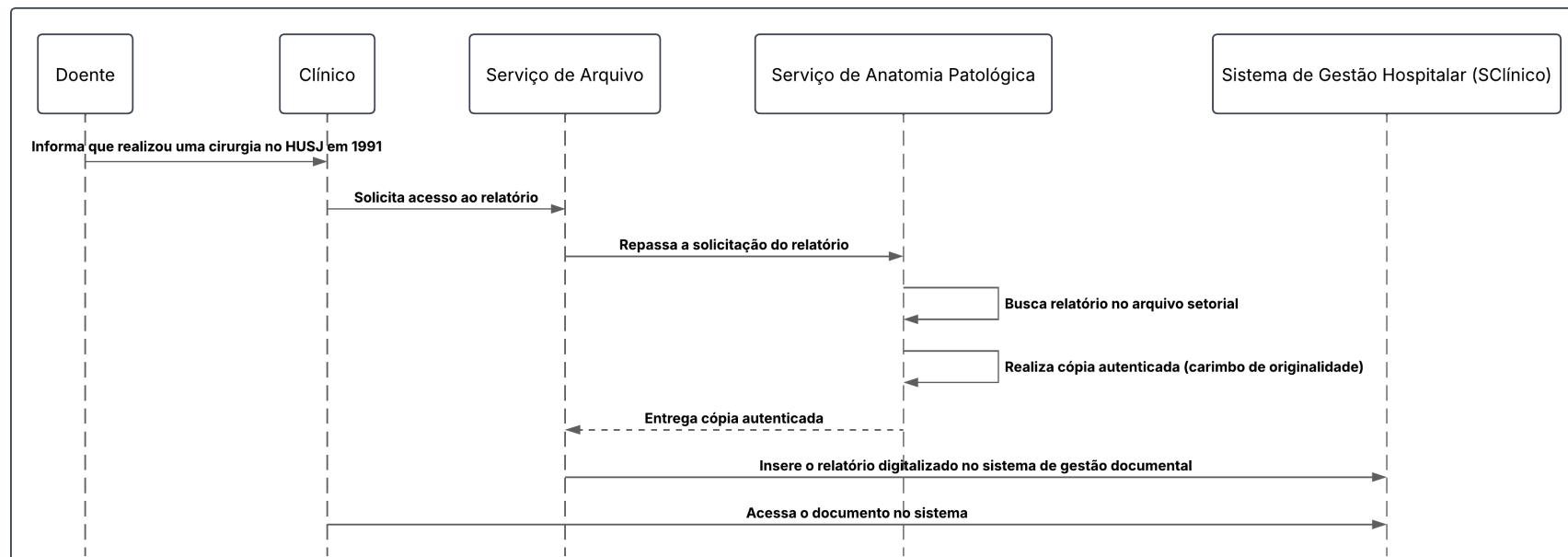


Figura 4.2: Diagrama de sequência para o *use case* "Pedido de documento retrospectivo de anatomia em contexto de urgência".

O diagrama também ilustra a diversidade das necessidades informacionais dos clínicos em relação aos relatórios, sendo de referir que essas necessidades vão para além do Serviço de Anatomia. O clínico precisa contatar o arquivo ou pedir para que alguém o faça, para verificar a existência do relatório e solicitar o acesso, aguardando que possa ficar disponível.

Quando este relatório não está em posse do Serviço de Arquivo, como é o caso daqueles dados antes de 2003, o técnico do Arquivo tem ainda que contatar o Serviço de Anatomia Patológica para verificar sua existência no arquivo setorial. Para além da necessidade de despendido demasiado tempo dos profissionais da informação, o tempo para que o documento seja disponibilizado para acesso dos clínicos e a própria indisponibilidade imediata do documento no RCD é o prejuízo decorrente de todo este processo atual.

Como forma de dar início à compreensão dos processos atuais realizados no Serviço de Arquivo, a prática teve como início a observação participativa no processo de preparação e digitalização dos relatórios. A partir dessa observação foi possível compreender este processo em questão, que está esquematizado no diagrama 4.3.

O processo ilustrado pelo diagrama delinea o circuito organizacional que converte a documentação clínica em papel para digital de forma confiável, com metainformação descritiva associada, recuperação e reutilização. É possível perceber que as tarefas despendidas são realizadas a partir de recursos internos por cada entidade descrita, a comunidade HUSJ, a equipa de preparação e digitalização e a coordenadora do RCD.

A lógica do processo em si começa com a categorização e priorização, seja de retrospectivo (processos clínicos) ou ativo (documentos soltos). A equipa de preparação organiza então a documentação a ser tratada de acordo com a prioridade.

A equipa de preparação padroniza a documentação recebida em lotes, verifica a integridade do processo clínico ou do documento e gera registos descritivos, tirando partido da informação em resumo clínico do utente, inserindo-os no SPD (Suporte de Preparação e Digitalização, que consiste em um sistema utilizado no Arquivo de pré-processamento dos documentos antes de serem inseridos no RCD). O resumo clínico consiste em um conjunto de informações sobre episódios de consultas, internamentos, ambulatorios e outras tipologias no SONHO, que são importadas para o RCD para preenchimento da metainformação relativa ao relatório em questão. A equipa de digitalização executa a captura de imagens via *PaperStream*, organiza e monitora os lotes de digitalização e conduz o controle de qualidade. Caso ocorra alguma falha neste fluxo de preparação, a coordenadora do RCD supervisiona as exceções, regista erros relatados pelas equipas e examina as causas relativas aos erros encontrados.

Tratando-se de um **retrospectivo**, verifica-se o processo clínico do doente para certificar-se se está completo ou não. Quando incompleto, a solicitação de documentação ausente é iniciada (Tarefa 2). No caso de ser um processo completo, aqueles que sejam anteriores à informatização, ou seja, documentos produzidos em ambulatório anteriores a 1992 ou documentos de internamento anteriores a 1996, requerem a inserção manual de metainformação descritiva no SPD devido a falta de dados (Tarefa 3). Mais a frente esse tópico será melhor detalhado em conjunto das figuras 4.10 e 4.11.

No caso de serem processos posteriores à informatização, após 1992 para ambulatório e após 1996 para internamentos, basta que sejam registados no SPD a partir do número do episódio, pois a informação é recuperada do resumo clínico no SONHO, e associados à caixa onde se encontram (Tarefa 4*). Essas tarefas controlam o registo da unidade arquivística e sua vinculação física (caixa), um requisito para a rastreabilidade.

Na verificação de duplicação, o que se realiza é perguntar se o processo de registo já existe, garantindo a historicidade do objeto digital (Tarefa 5). Caso negativo, é iniciada a preparação da versão 0 do objeto (Tarefa 6). Em ambos os cenários, a conclusão do registo de preparação ocorre para que o documento siga para a digitalização (Tarefa 7).

Na digitalização, o documento é digitalizado como retrospectivo no RCD (Tarefa 8) e posteriormente volta a equipa de preparação para realização do controlo de qualidade (Tarefa 9), que avalia a conformidade (integridade, legibilidade e correspondência entre o objeto digital e a sua versão em papel). Caso esteja tudo de acordo, avança para a verificação da ingestão no RCD (Tarefa 10), ou seja, a confirmação de que o objeto digital e a metainformação associada a ele foram ingeridos de forma correta no RCD e comunicação ao requerente de que o documento já se encontra disponível para consulta. Caso não esteja tudo de acordo na tarefa 9, o documento volta para o início da preparação no SPD e todo o processo é refeito.

Caso haja algum problema na ingestão, tarefa 10, a tarefa de analisar e corrigir é ativada (Tarefa 12), envolvendo um ciclo de correção até ser corrigido. A persistência da falha pode ser escalada a coordenação do RCD (Tarefa 13), que compreende o erro, abordando as causas procedimentais ou tecnológicas com os responsáveis pelo sistema.

Em caso do documento ser classificado como um **ativo**, são preparados lotes de documentos recebidos no Arquivo (Tarefa 14), são registados no SPD (Tarefa 15*) e passam para a equipa de digitalização em uma folha de preparação, para assinatura do técnico responsável de cada fase e falhas que tenham ocorrido em cada fase, são posteriormente digitalizados em *PaperStream* (Tarefa 16) e passam pelo controlo de qualidade (Tarefa 17).

Vale ressaltar que nas tarefas 4* e 15*, efetua-se a associação da metainformação a partir da recolha de elementos no próprio documento por parte do técnico. O tipo de episódio, que pode ser consulta, internamento, ambulatório, bloco e outros, o número de episódio, que por norma se encontra no documento, mas pode ser que não esteja descrito e o técnico tenha que buscar no SONHO, pelo nome completo e data de nascimento. Neste caso, o técnico busca no resumo clínico do utente pela data de admissão ou a data que mais se aproxima do documento para atribuir um número de episódio no SPD, pois sem este número é impossível inserir o documento no RCD, uma vez que o RCD busca os dados do utente através deste número. Portanto, o número de episódio e data do episódio dependem da informação estar descrita no documento ou informatizada no SONHO. Outros campos necessários, mas que não dependem do SONHO, são a tipologia documental, a tipologia de descrição, a secção, data do documento e número de folhas. Caso esteja tudo em conformidade, o processo se encerra com a finalização do lote. Caso tenha algum problema no processo de digitalização, o lote completo ou apenas o documento que teve problema, volta a ser digitalizado.

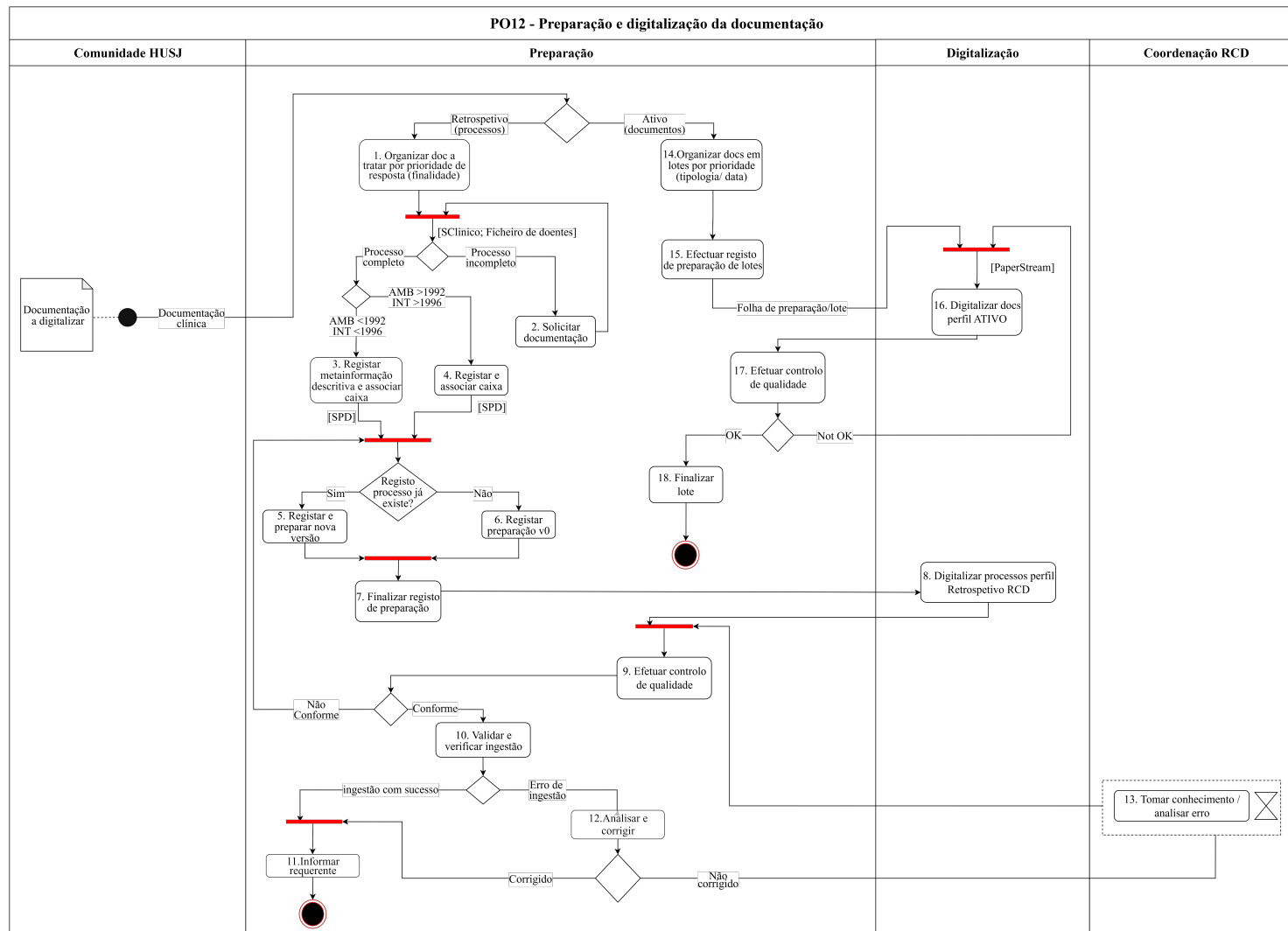


Figura 4.3: Processo de preparação e digitalização - SPD.

Para proceder a preparação de um ativo no SPD, no menu preparação, como mostra a figura 4.3, é necessário realizar o correto preenchimento dos campos requisitados, sendo esses o tipo de episódio, número do episódio, tipologia, tipo de descrição, secção, descrição e número da caixa.

Ativos - Preparação

Tipo de Episódio: Consulta *

Nº de Episódio: XXXXXXX *

Data do Episódio: 12-05-1999 *

Tipologia Documental: Resultado de MCDT *

Tipologia Descrição: Exame histológico - ver... *

Secção: MCDT *

Descrição da Tipologia: Relatório H02/XXXX

Data do Documento: 15-03-1999 *

Serviço: Arquivo *

Destino: Arquivo *

Nº de Folhas: 3 *

Nº de Caixa: APAT1/2025 *

* Campos de preenchimento obrigatório

Adicionar Imprimir Etiqueta Mira

Nenhum registro encontrado.

Figura 4.4: Ambiente de preparação de ativos - SPD.

A partir de uma amostra de 10 documentos de 1996 (amostra 4), referida na tabela 4.4, foi possível iniciar atividades práticas para perceber o processo de preparação, a estrutura dos relatórios em questão e iniciar uma análise sobre a metainformação contida nos documentos. Em ambiente de teste no SPD, como ilustra a figura 4.4, deu-se início ao processo de teste com relatórios histológicos, onde os tipos de episódios eram de consulta ou internamento, classificados como : (a) tipologia de resultados de MCDT (Meio Complementar de Diagnóstico e Terapêutica), (b) tipo de descrição exame histológico, (c) secção MCDT, (d) número de caixa APAT1/2025, (e) número de folhas e a descrição da tipologia, ou seja, das características físicas, número do código do exame e outros detalhes a depender das características do relatório em papel.

Para fins de teste, em relação a esse campo, de descrição da tipologia, foi inserida informação quanto ao estado, como a presença de relatórios ora datilografados com notas manuais, ora apenas com notas manuais e desenhos referentes a peça, como mostra a figura 4.5, número do código (exemplo do lote analisado: H96/0000), alguns com mais de uma folha, sendo as outras referentes a exames complementares. As figuras 4.6, 4.8 e 4.9 ilustram relatórios de 1988 e 2002 que fazem parte de outras amostras também tratadas pelo mesmo processo de análise e preparação de documento.

ANATOMIA PATOLÓGICA

EXAME N.º

João

NI: Data: 20/11/2002-8:16 NO:

CIRURGIA B

Data Nasc: - 61 Anos Fem. Tel.

RESULTADO A ENVIAR PARA: CONSULTA ☐ INTERNAMENTO ☒ SERVIÇO DE: cirurgia B

PRODUTO / ÓRGÃO / LOCAL: Tumor da mama direita + envolvimento axilar TÉCNICA DE COLHEITA: cirurgia

ENVIADO: SEM FIXAÇÃO ☐ (Colheita efectuada às ____ h ____ m) FIXADO ☐ EM:

DIAGNÓSTICO CLÍNICO / INFORMAÇÕES CLÍNICAS: Não de mama
Bupir - carcinoma ductal invasor

PROBLEMAS A ESCLARECER: carcinoma em histologia

DATA 21/11/02 O MÉDICO: VINHETA 32592

Relatório de Anatomia Patológica:

Baga de cor. conserv. de 103 de 14,5x9,5x 5cm. Ao corte observava-se neoplasia dura, esbranquiçada, limites mal definidos, de 0,8x0,4x cm. Amarelo. Situa-se a: 1,1cm do laranja; a 1,3cm do verde; 1,4 do azul; 1,5cm do amarelo; 1,3cm do amarelo.

B2-B2-2f-neopl. (dist. ao laranja) 12x1,2x0,99
B3-B4-2f- " " " azul
B5-B5-2f- logo a seguir à neopl. (B5-carbonizável-CDIS)
B7-7f-a 1,5cm da neopl. (CDIS?)
B8-B9-4f-par. rest.
B10-B12-11f-17ggs linf

Data: 22/11/2002

B13-4f-2ggs linf. B15-B16-3f-1gg linf. Total: 16e-33f

B14-4f-2ggs linf.

H.S.J. - Mod. 904 - 80 Grs. - Tip. Mota & Ferreira, Lda. CÓD. 531427

Desenhos referentes à peça:

Figura 4.5: Exemplo de relatório de 2002 com desenhos referentes à peça.

Fonte: Arquivo do Serviço de Anatomia Patológica HUSJ.

 ANATOMIA PATOLÓGICA EXAME HISTOLÓGICO N.º [redacted]

Hospital de S. João
PORTO

[redacted] C. TROCHA [redacted] 10 2 67 [redacted] *Lu. Abast.*

☐ Internamento ☒ Consulta

Diagnóstico clínico [redacted] ável: [redacted]

Natureza da peça: *ulc*

Colheita por: *Cirurgia* Fixação em: *Formol*

3 / *11* / 1988 O Médico, [redacted]

Resultado para [redacted] / [redacted] / 198

Data do início da lesão *Quente e*

Sede *terço posterior do dorso do nariz*

Evolução *lenta*

Aspecto macroscópico *"ulcer rodem"*

1261-745
4.11.88
FC. J. O. B.

Pequeno cuboide de 2,5x1,8 cm,
centrado por lesão recente, ma-
Região exacta onde foi colhido o fragmento milonado, de
superfície ulcerada, em 1,7 cm

Neoplasia: primitiva? secundária (recidiva? metástase?) *de novo*

Se há outro diagnóstico histológico anterior, indicar o resultado *Imun-*

Outras análises: *NT par.*

Tratamento já feito *Os brônquios laterais de*
acord. com a peça não parecem em
volúmen pelo processo neoplásico

A anamnese oferece outros factos dignos de atenção, tais como cancro, sífilis, alcoolismo, tabagismo, irritações crónicas, vícios de nutrição, etc.?

Observações: *Ho. ante - lesão tem*
amp. pouco ovóide e arred. de
em branco - acm. deidade e tem
lesão expansiva. O plano de
acord. com a peça, e, no plano
profundo, m. de
próximo de lesão, parecendo mais a intercepção.

/ / 198 O Chefe de Serviço [redacted]

H. S. J. - Mod. Citograf 904 - 1000 bl. c/ 4x50

Figura 4.6: Exemplo de relatório de 1988 - frente.

Fonte: Arquivo do Serviço de Anatomia Patológica HUSJ.

RELATÓRIO

Retalho cutâneo de 2,5 x 1,8cm, centrado por lesão saliente de 1,7cm de maior dimensão com superfície ligeiramente mamilonada, que ao corte exhibe crescimento expansivo, consistência carnuda e cor branca-acinzentada.

O exame histológico mostra neoplasia epitelial constituída por células de tipo basalóide, identificando-se diversos padrões de crescimento: sólido, adenóide, trabecular e queratótico. A epiderme superficial está adelgada, não se identificando "pontos de contacto" entre a neoplasia e a epiderme. No estroma da neoplasia identificam-se alguns melanófagos.

As margens laterais de secção cirúrgica não estão envolvidas pelo processo neoplásico. No plano profundo a neoplasia infiltra a camada muscular. A secção cirúrgica não interessou, neste plano, o processo neoplásico embora esteja muito próximo do mesmo.

O DIRECTOR DO SERVIÇO, O MÉDICO,

Conclusão: Carcinoma basocelular (com origem no folículo piloso?).

PORTO 88 / 11 / 10

Figura 4.7: Exemplo de relatório de 1988 - verso.

Fonte: Arquivo do Serviço de Anatomia Patológica HUSJ.

simkash
pod



Hospital de S. João
PORTO

ANATOMIA PATOLÓGICA

EXAME N.º

Masculino

Fil.: *
SEGURANÇA SOCIAL



RESULTADO A ENVIAR PARA: CONSULTA ☒ INTERNAMENTO ☐ SERVIÇO DE: CASAS

PRODUTO / ÓRGÃO / LOCAL: Fígado Fígado TÉCNICA DE COLHEITA: Porto T. J. J. J. J.

ENVIADO: SEM FIXAÇÃO ☐ (Colheita efectuada às ____ h ____ m) FIXADO ☒ EM: Fígado

DIAGNÓSTICO CLÍNICO / INFORMAÇÕES CLÍNICAS: Dt de etiologia por esclerose.
D. J. J. J.

PROBLEMAS A ESCLARECER: _____

DATA 20/11/02 O MÉDICO _____

VINHETA
35566

Relatório de Anatomia Patológica:

Morfologia : M-01000 ° Análise : H02

No exame histológico do produto de biopsia hepática observa-se um fragmento de fígado de pequenas dimensões, com preservação da arquitectura trabecular. Observa-se expansão dos espaços porta por escasso infiltrado inflamatório de carácter inespecífico. Na transição epitélio-mesenquimatosa não se observam lesões necro-inflamatórias. Ausência de lesões dos ductos biliares. Em localização intra-lobular identificam-se lesões necro-inflamatórias dispersas de intensidade ligeira. Ausência de esteatose ou colestase. Siderose em células do revestimento sinusoidal. Não se identificam megamitocôndrias, glóbulos hialinos PAS-D positivos ou corpos hialinos de Mallory.

A pesquisa de cobre e proteína fixadora do cobre por métodos histoquímicos apropriado (Shikata e Rodamina) foi negativa.

CONCLUSÃO: Fígado de arquitectura trabecular com alterações inflamatórias portais e intra-lobulares discretas e de carácter inespecífico, a valorizar no contexto clínico e analítico (ver descrição detalhada feita anteriormente).



Data: 2 / 12 / 2002 O Médico: _____

H. S. J. - Mod. 904 - 400 bl. - (50x4) - 80g - PAPERLUMINO

Figura 4.8: Exemplo de relatório de 2002.

Fonte: Arquivo do Serviço de Anatomia Patológica HUSJ.

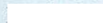
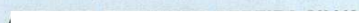
HOSPITAL DE S. JOÃO - SERVIÇO DE ANATOMIA PATOLÓGICA	
Relatório Complementar 	
	
CONSULTA DE GASTROENTEROLOGIA	
Nº D.: 	
Nome: 	
D. N.: 	
Morada: 	
Cód. P.: 	
Data : 20-11-2002	
RELATÓRIO ANATOMOPATOLÓGICO	
Morfologia : M-01000	
No exame histológico do produto de biopsia hepática observa-se um fragmento de fígado de pequenas dimensões, com preservação da arquitetura trabecular. Observa-se expansão dos espaços porta por escasso infiltrado inflamatório de carácter inespecífico. Na transição epitélio-mesenquimatosa não se observam lesões necro-inflamatórias. Ausência de lesões dos ductos biliares. Em localização intra-lobular identificam-se lesões necro-inflamatórias dispersas de intensidade ligeira. Ausência de esteatose ou colestase. Siderose em células do revestimento sinusoidal. Não se identificam megamitocôndrias, glóbulos hialinos PAS-D positivos ou corpos hialinos de Mallory. A pesquisa de cobre e proteína fixadora do cobre por métodos histoquímicos apropriado (Shikata e Rodamina) foi negativa. CONCLUSÃO: Fígado de arquitectura trabecular com alterações inflamatórias portais e intra-lobulares discretas e de carácter inespecífico, a valorizar no contexto clínico e analítico (ver descrição detalhada feita anteriormente). RELATÓRIO COMPLEMENTAR (29/7/03) Realizou-se o estudo imunocitoquímico (I-/03) com anticorpos "para" o AgHBs e AgHBc. Não se observou imunorreactividade com nenhum dos anticorpos utilizados. CONCLUSÃO: Os resultados deste estudo não permitem demonstrar a existência de sinais de infecção pelo VHB.	
Data: 29-07-2003	O Médico: 

Figura 4.9: Exemplo de relatório complementar de 2002.
Fonte: Arquivo do Serviço de Anatomia Patológica HUSJ.

No caso de existir um documento composto, ou seja, um conjunto de documentos simples que corresponde a um mesmo procedimento, constituído por um relatório final e relatórios complementares ou de outras instituições, o processo consiste apenas informar na descrição os detalhes sobre as folhas que acompanham o relatório principal, sendo que a data mais recente em qualquer uma das folhas é a data colocada como final no registo no SPD. Ou seja, os relatórios complementares são todos processados junto ao inicial como um único documento e com um único código no SPD. Foi levantada a possibilidade de processar esses relatórios complementares como um documento simples, para recuperação do relatório específico e não de todos aqueles relacionados a ele. Posteriormente, essa questão foi feita em visita ao Serviço de Anatomia Patológica e verificou-se que o relatório complementar deve ser necessariamente processado como um documento composto junto ao inicial para que se tenha a visão global de todos os exames realizados para um mesmo caso clínico.

Ao fim da fase de preparação, cada relatório principal recebe um código QR com uma numeração e assim são rastreados em todas as etapas, tal como pode ser visto na figura 4.8. Após esta fase, acompanhados por um formulário de registo para preenchimento sobre cada fase de processamento, como ilustra a figura E.1 em apêndice, os documentos seguem para a parte de digitalização, onde é lido primeiramente o código QR da "mira", uma vez por lote, para que a digitalizadora tenha um padrão de cor, para que as versões digitalizadas sejam o mais fiéis à realidade e posteriormente seja possível avaliar a qualidade das digitalizações. Dessa forma, os relatórios passam pela digitalizadora em lotes e os códigos são automaticamente lidos. Aqueles que não são lidos, são inseridos manualmente e existe um rastreio por parte do técnico da quantidade de falhas de leitura do código para revisão contínua do processo, sendo essa informação anotada no formulário de registo que acompanha o lote.

A etapa seguinte é a de controle de qualidade, onde é revisado o documento digitalizado em todos os pormenores para garantir a qualidade e fiabilidade à versão original. Caso não esteja correspondente ao desejável, o documento é marcado como não conforme (NOT OK), a digitalização do documento em causa é repetida e substituído pelo que estava não conforme.

Após as etapas descritas, o documento é inserido no repositório (RCD) e fica disponível para consulta, acompanhado de 56 metadados relacionados ao mesmo, descritos em detalhe na tabela D.1 em apêndice. Dentre estes metadados, são compreendidos aqueles relacionados ao número do doente, número antigo do doente, nome do doente, número do Sistema Nacional de Saúde (SNS), data de nascimento, sexo, número do episódio, especialidade clínica de admissão, especialidade clínica de alta e outros, mas ainda não se encontra um para termos indexadores.

Como pode ser verificado na tabela 4.6, os quatro últimos relatórios foram detetados como sem metainformação, o que significa que os episódios não foram encontrados no SONHO. Esse fator impede que sejam inseridos no SPD e consequentemente no RCD sem que sejam introduzida a metainformação previamente de forma manual. Alguns dados essenciais, por exemplo, em um documento de internamento, são o nome e número do médico de admissão e alta, que sem estes a autenticidade do documento fica comprometida e é impossível inserir o documento no RCD, uma vez que o repositório visa a preservação da informação.

Em casos de internamentos, como mostra a figura 4.10, é necessário levantar os microfilmes e no SONHO para coleta do número do doente, o número do episódio, nome do doente, número sequencial, data da admissão, o nome e número de ordem do médico da admissão e da alta, Número do SNS do doente, data da alta e especialidade. A metainformação requisitada nos documentos de ambulatório são menos exigentes, como mostra a figura 4.11 o que facilita que sejam processados, diferentemente dos documentos de internamento. Em caso de documentos de tipologia de consulta anteriores à informatização, não é possível inserir metainformação, pois estes não tem número de episódio atribuído, passam a ter a partir da informatização. Para todo este processo de tratamento de doentes não informatizados, o percurso necessário demanda em média 30 minutos por relatório, caso já se tenha todos os materiais de apoio necessários para inserir os dados do utente e a metainformação.

Devido ao nível de detalhe, é algo moroso e algumas vezes impossível de se fazer devido a falta de informações necessárias. Esse é um dos fatores que fazem do processo de tratamento do arquivo de anatomia patológica, principalmente com relatórios anteriores a 1992. Ou seja, existem dois cenários para o tratamento destes relatórios:

- Cenário A: O doente não é informatizado, é necessário recuperar a identificação do doente no SONHO, constatar se faleceu depois de 1999, sendo possível recuperar os dados a partir do RNU ou se faleceu antes de 1999. Nestes casos, a recuperação é feita com base no documento ou exige pesquisa nos ficheiros dos doentes. Para este cenário, só devem ser digitalizados relatórios à pedido ou os que sejam de reutilização recorrente pelo Serviço de Anatomia Patológica.
- Cenário B: O doente já está informatizado, a sua identificação é recuperada para o SONHO com base nos dados existentes no RNU ou nos registos de processo/ficha do doente. Para relatórios de internamentos entre 1992 e 1996, o número do episódio é produzido em SPD para associação ao documento. Nos casos de urgências e consultas, que não possuem número de episódio atribuído pelo HUSJ durante estes anos, é necessário criar um número sequencial pelo SPD para que seja possível introduzir um episódio.

Estes cenários implicam (i) ter a consciência da limitada capacidade de cruzar os dados de um mesmo doente, quando não está informatizado no HUSJ; (ii) alimentar o RCD com identificações de doentes limitadas aos documentos, o que gera riscos de erro e completude; (iii) e a informatização prévia do ficheiro manualmente dos doentes e fichas de internamento para suporte ao trabalho de criação de identificações e metainformação de episódios.

  **SUPORTE PREPARAÇÃO e DIGITALIZAÇÃO** **UAT**  Registo Processo  Pesquisa  Preparação Processo  Gestão  Relatórios  Gestão de Protocolos  Pedidos ao Arquivo  Ativos  Sair

+ Inserir Metainformação

Tipo de Processo:	Internamento				
Nº Doente:	18011749				
Nº Episódio:	18013610				
Nome Doente:	TESTE DOIS				
Nº Sequencial:	1313954	Nº SNS:		Nº Antigo Doente:	
Data Nascimento:	01-01-1959	Sexo:	☒ F ☐ M		
Data Admissão:	18-04-2018	Data Alta:	28/04/2018		
Nº OM Admissão:	22222	Med. Admissão:	MARIA MARIA	Especialidade Admissão:	SERV. INTERN. TESTE
Nº OM Alta:	33333	Med. Alta:	JOAO JOAO	Especialidade Alta:	SERV. INTERN. TESTE

Nenhum registo encontrado.

Figura 4.10: Campos de preenchimento de metainformação para processos de internamento - SPD.

+ Inserir Metainformação

Tipo de Processo:

Nº Doente:

Nome Doente:

Nº Sequencial:

Data Nascimento:

Nº SNS:

Sexo:

Nº Antigo Doente:

Nenhum registo encontrado.

Figura 4.11: Campos de preenchimento de metainformação para processos de ambulatório - SPD.

4.3 Caracterização dos relatórios citológicos e histológicos

Os relatórios citológicos e histológicos foram elaborados em diferentes formatos ao longo dos anos e classificados também de maneiras diferentes. Inicialmente, levando-se em consideração o ano de 1962, o Serviço de Anatomia Patológica elaborava seus relatórios em formato A5, com requisição do clínico, alguma estrutura para preenchimento do exame macroscópico, com notas manuscritas e o relatório, propriamente dito, datilografado no verso em texto livre. A estrutura da descrição seguia padrões aconselhados pela OMS, mas dependiam do trabalho manual do patologista para detalhar todos os aspetos necessários quanto ao tamanho da peça, coloração e demais requisitos específicos a depender da especialidade.

Com o passar do tempo, como pode-se constatar na tabela 4.4 e nos exemplos de relatórios nas figuras 4.6 e 4.8, houve uma evolução do formato da folha, na descrição do relatório, na estrutura e na codificação. Os primeiros relatórios informatizados no Serviço de Anatomia Patológica, em 2004, eram chamados de relatórios Tipo, o objetivo era logo diminuir trabalho dos patologistas, portanto, foram feitos muitos relatórios tipo baseados no que era habitual a maior parte dos especialistas dizerem naquela situação. Estes relatórios consistiam em uma versão pré formatada do texto referente ao tipo de exame e o patologista apenas precisava ajustar alguns elementos, por exemplo, acrescentava as medidas e mudava algumas características, como a cor, a dimensão, o aspeto das coisas e tudo o resto era mais ou menos.

Por exemplo, a informação é basicamente a mesma quando existe uma apendicite aguda, os apêndices com apendicite aguda são quase todos iguais, o que varia é o tamanho, se tem perfuração, se não tem perfuração. Pode ser uma coisa crónica ou pode ser uma coisa aguda. Quando a peça e a requisição são recebidas no serviço, os patologistas já podiam fazer essa escolha do relatório desejado e faziam a análise em cima do relatório pré formatado.

O diagrama 4.12, produzido a partir da observação e em conversa com patologistas e técnicos, ilustra o processo de produção dos relatórios histológicos no Serviço de Anatomia Patológica, onde diversos agentes contribuem para a sua realização. Neste processo é importante destacar alguns aspetos na atribuição do código SNOMED para a morfologia e topografia, enviada posteriormente para o serviço de codificação em caso de cancro. Em entrevista com uma médica codificadora, percebeu-se que a etapa de revisão da codificação SNOMED realizada pelo patologista e, por diversas vezes, a consulta ao próprio relatório se mostra necessário devido aos erros cometidos na atribuição do código SNOMED.

A médica relata a recorrência de um mesmo código que é atribuído diversas vezes quando são codificados vários relatórios de uma vez por um mesmo patologista. Ou seja, existe um padrão de codificação por lotes de relatórios produzidos, o que coloca em causa a fiabilidade da codificação por parte dos patologistas e requer uma revisão por relatórios por parte do Serviço de codificação antes que seja atribuído um código ICD.

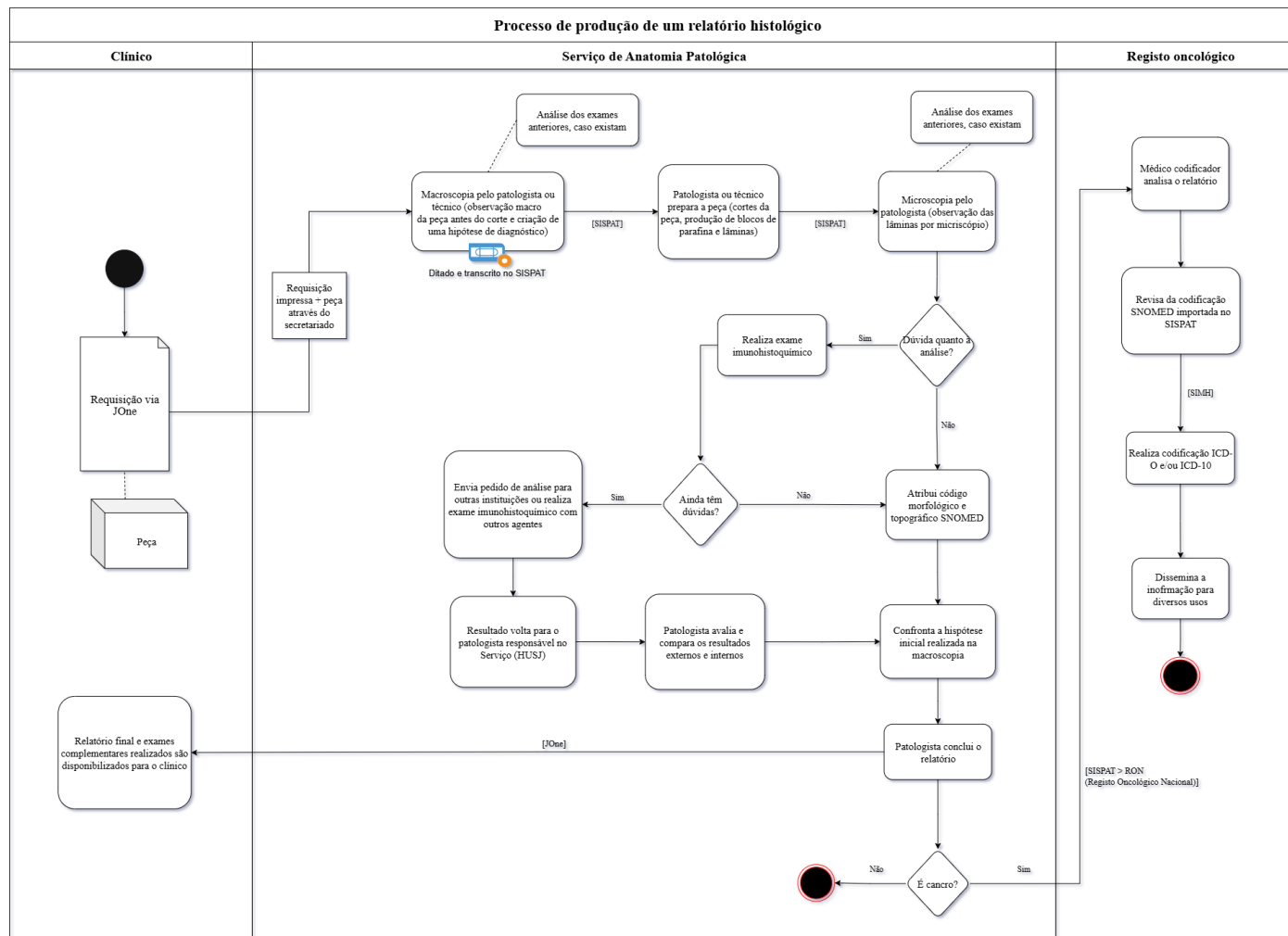


Figura 4.12: Processo de produção de relatórios histológicos no Serviço de Anatomia Patológica.

Em continuidade do processo observado para preparação dos relatórios, como descrito na secção anterior, cada um foi preparado no ambiente de teste do SPD, digitalizado, revisado pelo controlo de qualidade e por fim, inseridos no RCD. Todo o processo de análise documental foi descrito nas tabelas 4.5, 4.6 e 4.7, seleccionadas para exemplificar a análise aprofundada por cada relatório de cada amostra, pois as amostras 3, 4 e 5 apresentavam mais detalhes que enriqueciam a compreensão da estrutura dos relatórios à data que foram produzidos.

Como forma de adequar e sistematizar a análise documental dos relatórios citológicos e histológicos, e análise os vocabulários controlados, os dados levantados serão registados em um quadro que contenha:

- Data da análise
- Número do código do relatório
- Nº do registo no SPD
- Notas descritivas do relatório
- Notas reflexivas da investigadora
- Termos identificados no relatório

Tabela 4.5: Análise crítica descritiva dos relatórios - Amostra 3.

Data	Número do código	Nº de registo	Notas descritivas do relatório	Termos encontrados
13/03/2025	30**/94	290	Relatório em formato A5, com vinheta, estruturado, com requisição manuscrita, com notas manuscritas, relatório final datilografado no verso em texto livre, com assinatura manuscrita. Com marcas de agrafador, marcas na vertical no verso, e dobras.	
	30**/94	289	Relatório em formato A5, com vinheta, estruturado, com requisição duplicada por papel químico e notas datilografadas no verso em texto livre. Sem autocolante do código, o mesmo aparece manuscrito no canto superior direito. Com marcas e amassados.	Parênquima placentário; corioamnionite; funiculite; vilite linfo-histiocítica; edema do estroma vilositário; basal sub-trofoblástica; classificação de blanc; necrópsico.
	30**/94	291	Relatório em formato A5, estruturado, com requisição e notas duplicadas por papel químico, com notas manuscritas, relatório final datilografado no verso em texto livre, correções manuais do datilografado, com assinatura manuscrita, com data do documento corrigida manualmente. Código escrito manualmente, Com marcas de agrafador, marcas na vertical no verso, e amassados.	
	30**/94	292	Relatório estruturado, com requisição manuscrita, com notas manuscritas, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3003/94. O relatório contém rasgos por agrafador.	
	30**/94	293	Relatório estruturado, com requisição e notas duplicada por papel químico, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3004/94.0 relatório contém rasgos por agrafador.	Alterações no endométrio do tipo secretor e Arias-Stella; vilosidade coriais; eritoblastos; saco gestacional.

Data	Número do código	Nº de registo	Notas descritivas do relatório	Termos encontrados
	30**/94	294	Relatório estruturado, com requisição e notas duplicada por papel químico, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3005/94. O relatório contém rasgos por agrafador.	
	30**/94	295	Relatório estruturado, com requisição e notas duplicada por papel químico, com notas manuscritas, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3006/94. O relatório contém rasgos por agrafador	
	30**/94	296	Relatório estruturado, com requisição e notas duplicada por papel químico, com notas manuscritas e rasuras, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3007/94. Com marcas na vertical no verso, e amassados. Folha mais fina e esverdeada.	
	30**/94	297	Relatório estruturado, com requisição e notas duplicadas por papel químico, com notas manuscritas, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3009/94. Com marcas na vertical no verso, e amassados. Folha mais fina e esverdeada.	Retalhos de mucosa entérica; linfoplasmocitário de córion; plimorfonucleares eosinófilos; gastroenterite eosinófica.
	30**/94	298	Relatório estruturado, com requisição e notas duplicadas por papel químico, com notas manuscritas, relatório final datilografado no verso, com assinatura manuscrita e com código manuscrito 3010/94.	Retalhos de pleura; pleurite crónica exsudativa (em organização).

Tabela 4.6: Análise crítica descritiva dos relatórios - Amostra 4.

Data	Número do código	Nº de registo	Notas descritivas do relatório	Notas reflexivas	Termos encontrados
19/03/2025	H96/75**	299	Relatório em formato A5, estruturado, com vinheta, com requisição duplicada por papel químico, com relatório datilografado no verso, com correções manuscritas, com assinatura manuscrita. O relatório apresenta cortes nas bordas e restauração com fita cola transparente. O código aparece com autocolante em todos.	Verificar se a fita atrapalha o processo de digitalização e leitura por OCR	apêndice ileo-cecal; sangue; pus; apendicite aguda; necrozada
	H96/75**	300	Relatório em formato A5, estruturado, sem vinheta, com dados do doente, requisição e notas duplicadas por papel químico, com relatório datilografado no verso, com assinatura manuscrita. O relatório apresenta cortes nas bordas, mancha na vertical no verso e restauração com fita cola transparente.		pigmentação hemossiderínica; histiocitoma fibroso; bainhas tendinosas; exérese
	H96/75**	301	Relatório em formato A5, estruturado, com vinheta, com requisição manuscrita, com notas manuscritas, com relatório datilografado no verso, com correções datilografadas nas notas datilografadas, com assinatura manuscrita. O relatório apresenta cortes nas bordas e restauração com fita cola transparente.	As correções manuscritas e datilografadas podem, e certamente vão, atrapalhar a leitura dos documentos por OCR	placenta; cordão umbilical; vilosidades coriárias; necrose fibrinóide; hematoma retro-placental; cadáver; maceração; morte intrauterina; petequias no timo; pericárdio; pleura; IG (idade gestacional); anóxia aguda terminal
	H96/75**	302	Relatório em formato A5, estruturado, com vinheta, com requisição e notas duplicadas por papel químico, com relatório datilografado no verso, com assinatura manuscrita e restauração com fita cola transparente no verso.		fibrina subcoriária; perivilositário; microenfartes; parênquima placentário

Data	Número do código	Nº de registo	Notas descritivas do relatório	Notas reflexivas	Termos encontrados
	H96/75**	303	Relatório em formato A5, estruturado, com vinheta, com requisição duplicada por papel químico, com notas manuscritas, com relatório datilografado no verso, com assinatura manuscrita. O relatório apresenta rasgos, dobras e manchas.		histiocitose sinusal; peri-ganglionar
	H96/75**	304	Relatório em formato A5, estruturado, com vinheta, com requisição duplicada por papel químico, com notas manuscritas, com relatório datilografado no verso, com correções manuscritas e datilografadas, com assinatura manuscrita. O relatório apresenta rasgos, dobras e manchas.	As correções manuscritas e datilografadas podem, e certamente vão, atrapalhar a leitura dos documentos por OCR	colecistite crónica litiasítica
	H96/75**	-	Não tem metadados, correções manuscritas nas notas datilografadas		vesícula biliar; colecistite; litiasítica; divertículos de Pocktansky-Aschoff; incrustações calcárias
	H96/75**	-	Não tem metadados, correções manuscritas e datilografadas nas notas datilografadas	20/03: Não é possível encontrar o episódio no SONHOI, ou seja, não é possível preparar no SPD sem inserir manualmente os metadados	endométrio; cistos de Naboth; esclerotoftes; leiomiomatose; leiomioma; cervicite crónica cística
	H96/75**	-	Não tem metadados, correções manuscritas e datilografadas nas notas datilografadas		tecido adiposo; ulceração; metaplasia
	H96/75**	-	Não tem metadados		condiloma invertido

Tabela 4.7: Análise crítica descritiva dos relatórios - Amostra 5.

Data	Número do código	Nº de registo	Notas descritivas do relatório	Notas reflexivas	Termos encontrados
21/03/2025	H97/99**	305	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com relatório complementar datilografado no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.	Esses relatórios já contam com morfologia (código SNOMED CT) e estão todos dobrados para ficar no formato A5.	D5-80430; cirrose hepática; toxicidade alcoólica; megamitocôndrias; corpos hialinos de mallory; esteatose; colestase; ground-glass.
	H97/99**	306	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.	F1, F2, F3... significa frasco 1, frasco 2...	M-74001; mucosa gástrica do antro; adenomatosa; displasia leve; pólio adenomatoso séssil.
	H97/99**	307	Relatório em formato A5, com estrutura para preenchimento, com requisição e notas duplicadas por papel químico, com notas datilografadas em texto livre no verso, com assinatura manuscrita. Com marcas e furos.	Não conta com código de morfologia	gastrite crônica; helicobacteripylori.
	H97/99**	308	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.		M-43000; mucosa duodenal; glândulas de Brunner.
	H97/99**	309	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.		M-00100; pólio hiperplásico; mucosa rectal.
	H97/99**	310	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.		M-72042; pólios hiperplásicos.

Data	Número do código	Nº de registro	Notas descritivas do relatório	Notas reflexivas	Termos encontrados
	H97/99**	311	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras e manchas.		M-43000; gastrite crônica; atrofia glândular
	H97/99**	312	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.		M-43000; recto colite crônica
	H97/99**	313	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas no verso em texto livre, com assinatura manuscrita. Com marcas de dobras.		M-40000; infiltrado polimorfo no córion; rectocolite.
	H97/99**	314	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com notas manuscritas em texto livre, com assinatura manuscrita. Com marcas de dobras e furos.		M-67050; medula óssea; hematopoiética; edema; célula estromais; fibroblastos; degenerescência.
	H97/99**	315	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com assinatura manuscrita, com notas manuscritas em texto livre no verso. Com marcas de dobras e manchas.		M-44000; parênquima pulmonar; necrose; caseoso; células gigantes de tipo Langhans; bacilos álcool-ácido; inflamação granulomatosa; etiologia tuberculosa.
	H97/99**	316	Relatório em formato A4, sem estrutura para preenchimento, com requisição duplicada por papel químico, com notas datilografadas em texto livre, com assinatura manuscrita. Com marcas de dobras e manchas.		M-53202; hiperplasia das criptas; infiltrado inflamatório.

O número do episódio consiste em um número inicial, seguido por uma barra, uma letra (**A** para internamentos e **C** para consultas) e o número do ano, por exemplo 0000/A62. Os relatórios a partir de 1976 seguem uma estrutura semelhante, mas o número do episódio modifica-se para o número do exame, seguido de barra e ano (0000/76). A partir de 1997, o formato do documento modifica-se por completo, sendo em folha A4, com requisição, incluindo por vezes alguma estrutura para preenchimento ou não, com notas manuscritas e datilografadas em texto livre e com notas também no verso. A sua classificação já apresenta uma sequência diferente, sendo C para citológico e H para histológico e ano, seguido por barra e número do exame (C97/0000).

Com o objetivo de traçar um padrão de descrição nos relatórios, para análise de progressão dos termos utilizados na área, a análise documental permitiu identificar termos relacionados a doenças, procedimentos, morfologias e topografias. Algumas peculiaridades dos manuscritos foram identificados com a ajuda dos técnicos do Arquivo, por exemplo, a letra q com traço em cima é para abreviação do termo "que" e uma palavra seguida de um "corte" ou traço na diagonal é abreviação de "mente" (analítica/ = analiticamente).

A análise detalhada das amostras evidencia que a heterogeneidade dos relatórios citológicos e histológicos, tanto em termos de formato físico (A5, A4, datilografados, manuscritos, informatizados) como de estrutura descritiva, representa simultaneamente uma riqueza e um limitante. Se, por um lado, esta diversidade permite compreender a evolução histórica da prática descritiva e possibilita comparações terminológicas, por outro, cria dificuldades substanciais para processos

automatizados de digitalização e indexação. Em particular, as rasuras, notas manuscritas e correções interferem diretamente na eficácia de ferramentas de OCR, o que implica a necessidade de pré-processamento e validação específicas no modelo de enriquecimento proposto.

Outro aspecto crítico observado é a coexistência de relatórios estruturados e narrativos, bem como o uso irregular de terminologias padronizadas. Enquanto alguns relatórios já incorporam códigos SNOMED, outros mantêm apenas descrições livres, o que compromete a consistência na automatização. Para o modelo de enriquecimento da metainformação, isto significa que será necessário desenvolver mecanismos de equivalência terminológica que permitam mapear descrições livres para vocabulários controlados, de forma a garantir interoperabilidade e recuperar informação com maior precisão.

Em síntese, a caracterização dos relatórios demonstra que a evolução documental do Serviço de Anatomia Patológica gerou um arquivo complexo, marcado por camadas históricas de formatos e terminologias. Este cenário reforça a necessidade de um modelo de enriquecimento que não apenas normalize e atualize vocabulários, mas que também contemple terminologias locais e estratégias de integração entre documentos de diferentes épocas, garantindo tanto a preservação do valor histórico como a usabilidade futura em contexto clínico e científico.

Capítulo 5

Análise da atividade informacional dos utilizadores e arquivo

Este capítulo dedica-se à análise das atividades informacionais no Serviço de Anatomia Patológica, no que se refere à produção e uso do arquivo de relatórios histológicos e citológicos, em articulação com as questões de investigação formuladas e com o desenho metodológico previamente apresentado. A partir de entrevistas com patologistas, internos e assistentes técnicos, procura-se compreender como diferentes utilizadores acedem, pesquisam e utilizam a informação arquivística no contexto do serviço. O trabalho foca-se nas necessidades informacionais, nas práticas de pesquisa e nas dificuldades encontradas, permitindo observar o nível de eficácia da resposta do arquivo às exigências assistenciais, de apoio clínico e de investigação. Os resultados aqui discutidos não visam a generalização estatística, mas sim a construção de uma leitura interpretativa crítica e situada do comportamento de pesquisa, de acordo com a lógica qualitativa e exploratória do estudo.

5.1 Comportamento de pesquisa dos patologistas e investigadores

Como forma de enquadrar o comportamento de pesquisa dos patologistas e investigadores, conforme descrito no desenho da investigação 3.1, foram realizadas 11 entrevistas, sendo 9 com patologistas, 1 com assistente técnica e 1 com técnico de informática. A população constituiu-se por 20 especialistas em anatomia patológica e 11 internos em formação específica, portanto 31 pessoas. A amostragem intencional e por quotas (Pajo, 2023), não probabilística, foi utilizada com a intenção de selecionar uma amostra que seja capaz de proporcionar uma variedade de pessoas com mais e menos idade e em diferentes níveis de experiência e formação académica.

Como descrito no quadro 5.1, a amostra de 9 a 12 participantes foi assim definida para que fosse possível rastrear também o comportamento de pesquisa de diferentes faixas etárias e de diferentes percursos profissionais. Considerando a abordagem qualitativa da investigação, voltada para entender contextos, práticas e comportamento de pesquisa, não para generalização estatística, o critério principal é a saturação teórica, e não o tamanho proporcional da amostra.

Tabela 5.1: Caracterização da amostra de entrevistados.

População	Total	Amostra	Justificativa
Especialistas	20	6–8	Captar variações de experiência e atuação
Internos em formação específica	11	3–4	Compreender perspectivas em formação e dificuldades

As entrevistas decorreram durante o mês de junho, devido à realidade de agenda dos patologistas, sendo de refletir que esta investigação acompanhou a mudança da Direção do Serviço de Anatomia Patológica, o que dificultou o início das entrevistas, devido à necessidade de articulação com a Direção do Serviço.

As entrevistas tiveram duração média de 40 a 50 minutos, com gravação de áudio, divididas em 3 blocos principais, sendo estes (i) o contexto organizacional no Serviço, (ii) organização e acesso ao arquivo e (iii) comportamento de pesquisa. Todos os participantes foram voluntários, o que contribuiu para um diálogo proveitoso e enriquecedor para a investigação.

O tratamento e análise dos dados obtidos a partir das entrevistas teve início com o processo de transcrição e correção das transcrições no Transkriptor, realizado a partir do upload dos ficheiros, transcrição automática, nomeação dos arquivos por número de sequência das entrevistas, correção de gramática e limpeza de diálogos paralelos sem pertinência para a pesquisa.

Depois desta fase, deu-se início à análise no NVIVO da seguinte forma:

1. Criação de codebook, tabela C.1, a partir do guião de entrevista e dos temas principais abordados;
2. Upload de entrevistas;
3. Codificação das entrevistas;
4. Seleção de todas as falas referentes ao contexto organizacional em todas as entrevistas;
5. Seleção de falas por entrevistas de acordo com o tema abordado;
6. Produção de tabelas e gráficos a partir dos ficheiros criados com o grupo de falas por tema.

A partir das entrevistas realizadas foi possível identificar diferentes tópicos pertinentes para compreender o comportamento e as necessidades informacionais, no que se refere os relatórios de anatomia patológica. Os temas foram divididos com a intenção de satisfazer não apenas a questão do comportamento de pesquisa e recuperação da informação, mas também tópicos sobre a estrutura dos relatórios, terminologias utilizadas, organização e acesso ao arquivo, utilização do arquivo em suporte de papel e digital e outros contributos para a totalidade da investigação. O apêndice C ilustra o *codebook* e a árvore de itens codificados.

Para dar início à análise, foram identificados os cargos e o tempo de serviço de cada um dos participantes. Como descrito no capítulo 3, foram selecionados de forma intencional e por quotas, ou seja, não probabilística, o que possibilitou as entrevistas de perfis em diferentes momentos da carreira e de idade diferentes.

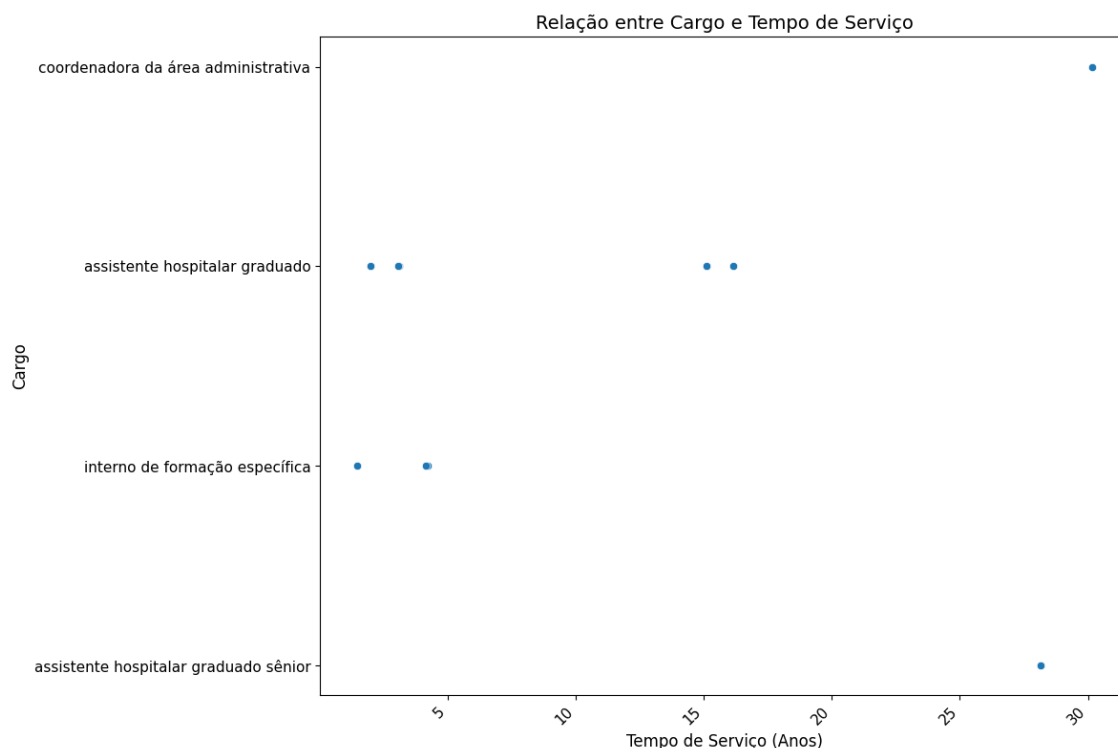


Figura 5.1: Perfil dos participantes.

Dos patologistas entrevistados, incluem-se 2 perfis principais, sendo especialistas ou internos em formação específica. Os médicos internos em formação específica participantes (3) são todos menores de 30 anos, ao oposto que os médicos assistentes hospitalares graduados (4) são todos maiores de 30 anos, bem como o assistente graduado sênior (1) e a coordenadora da área administrativa (1), têm mais de 50 anos.

A idade dos participantes é um ponto relevante, pois permite explorar melhor a perspectiva quanto a uma mudança no comportamento de produção dos relatórios no suporte de papel para o suporte digital ao longo dos anos. Visto que apenas 2 participantes, com mais idade, confirmaram que fazem uso de notas manuscritas como complemento aos relatórios ou até realizam todo o relatório manualmente, mesmo com sistema disponível, pode-se perceber que ainda existe um hábito de descrição manual.

Quando o assunto é realizar notas manuscritas ou utilizá-las como forma de complemento ao relatório, o comportamento dos dois grupos entrevistados diferem-se em alguns aspetos, abordados na tabela 5.2 e 5.3, relativas a importância e finalidade das anotações manuscritas e a prática propriamente dita de realizar anotações manuscritas ou digitais.

Tabela 5.2: Importância e finalidade das anotações manuscritas.

Categoria	Relevância e finalidade das anotações manuscritas
Patologistas (uso interno da Anatomia Patológica)	- Orientação e guia pessoal durante o exame. - Compreensão do raciocínio e conclusões anteriores. - Dúvidas para discussão com outros especialistas. - Organização de fragmentos macroscópicos. - Para identificação de casos especiais.
Clínicos (não patologistas)	- As notas não interessam, basta o relatório final.

As anotações manuscritas têm grande importância e finalidade no contexto da investigação, pois permitem revisitar o caso a partir da perspectiva do patologista que analisou a peça, possibilitando compreender a evolução do processo de análise das lâminas no âmbito do Serviço e entre diferentes especialistas. Além disso, oferecem pistas que não estão necessariamente registadas no relatório final, funcionando como complemento valioso para a interpretação e o estudo posterior. No entanto, apresentam alguns desafios, como a ausência de datação das notas e sua autoria, o uso de códigos ou formas de descrição pessoais e a consequente necessidade de interpretação especulativa. Para superar essas limitações, sugere-se implementações futuras, como a utilização de recursos de gravação em áudio que sejam automaticamente transformados em texto datado, garantindo maior precisão e acessibilidade das informações.

Tabela 5.3: Prática de anotações manuscritas x digitais.

Categoria	Notas manuscritas x notas digitais
Patologistas recém formados	- Notas complementares ao relatório são de preferência pelo digital (são inseridas em um campo de outros ou notas no próprio sistema). - Notas manuscritas são pessoais para lembretes, não complementam os relatórios finais. - Evitam ao máximo escrever manualmente.
Patologistas seniores	- Preferência pela produção de relatórios manuscritos, principalmente para análise microscópica. - Relatórios e notas manuscritas são transcritas para o SIS-PAT, mas também são arquivadas.

As anotações complementares não são anexadas ao relatório, mas cumprem funções específicas em situações particulares. O registo manual é utilizado, por exemplo, quando há falhas no sistema informático de produção de relatórios, na elaboração de requisições clínicas como os pedidos de soros, durante a fase da macroscopia para organização de frascos e outros materiais, bem como no contexto da autópsia fetal.

Ainda sobre este quesito, de produção de relatórios, os patologistas disseram que existem protocolos *standard* que guiam a produção destes relatórios, principalmente quanto o teor de descrição que os patologistas devem empregar. Alguns dos protocolos citados são CAP (*College of American Pathologists*), AJCC (*American Joint Committee on Cancer*) e o UICC (*Union for International Cancer Control*).

O protocolo da CAP é mais utilizado para neoplasias, ou seja para patologia maligna, onde existe um conjunto já bem estabelecido de parâmetros que devem ser mencionados no relatório. Esses protocolos da CAP são extremamente úteis e os formulários que usam no serviço foram construídos com base em várias fontes, nomeadamente nesses protocolos da CAP e UICC.

Os médicos internos em formação específica quando são responsáveis pela produção de um relatório, corrigem com o patologista especialista designado para orientação e dizem que conforme tomam conhecimento do estilo de descrição do patologista em questão e o que eles valorizam, os internos começam a descrever mais à maneira do patologista que está a orientá-los. Este é um fator demasiado importante que também chama muito a atenção para a forma como indexam/codificam os relatórios. Esse tópico será melhor abordado na secção 5.2.

A patologia oncológica é mais estabelecida de acordo com a CAP. A patologia inflamatória não está tão bem estabelecido como deve ser descrita, mas também depende de cada uma das áreas. Para os vocabulários controlados esse aspeto também é árduo, pois os processos inflamatórios não são necessariamente morfologia ou doença e no SNOMED não está bem sistematizado quanto a esses processos.

Em entrevista com os patologistas, desde os seniores até os mais novos, foi possível concretizar que no momento de pesquisa por termos, ou seja, quando se tem alguma dúvida quanto ao termo em si e sua forma controlada, os patologistas nunca utilizam o SNOMED browser, apesar de ser o vocabulário utilizado no Serviço.

A partir da figura 5.2 é possível perceber que a fonte de pesquisa para levantamento terminológico compreende principalmente a Google e o website da OMS (Organização Mundial da Saúde). Vale ressaltar que menção ao ChatGPT nas entrevistas na verdade se refere a não utilização do mesmo, ou seja, é mencionado como uma ferramenta que não é utilizada.

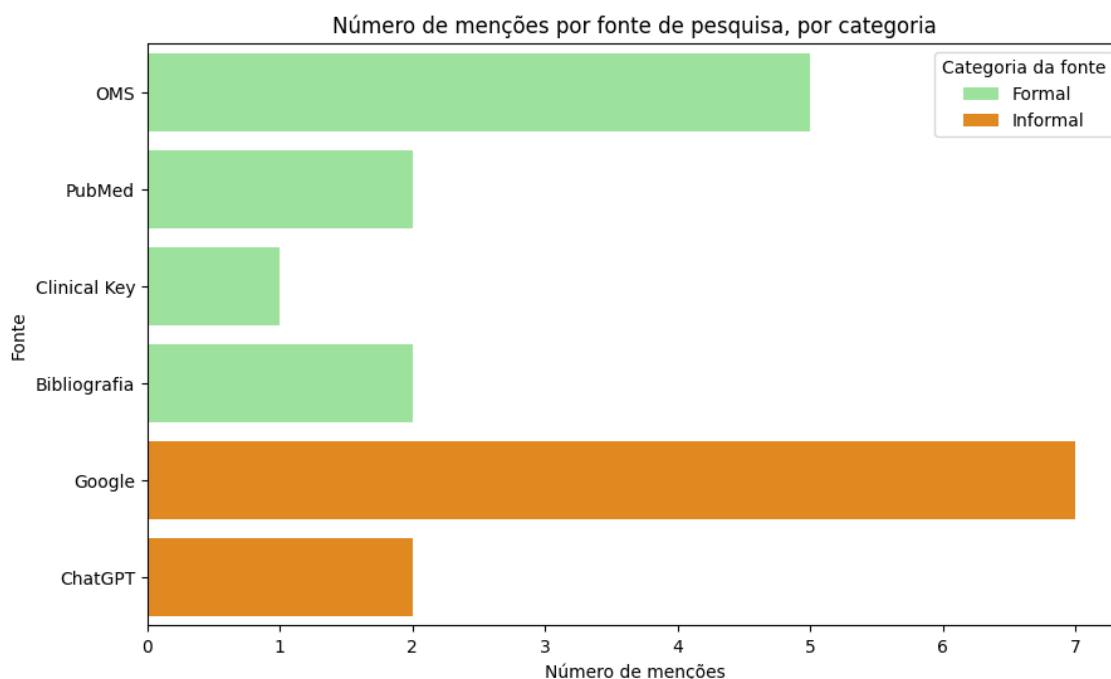


Figura 5.2: Fontes de pesquisa referidas nas entrevistas.

Como forma de diferenciar o processo de pesquisa por termos, com objetivo final de recuperar o termo controlado ou mais recente, e o processo de levantamento de relatório por meio dos termos, os diagramas 5.3 e 5.4 mostram detalhadamente como estes processos ocorrem no âmbito do Serviço de Anatomia Patológica.

O primeiro diagrama, figura 5.3, procura traduzir o processo de busca de um termo controlado, ou a forma autorizada de um termo, que se inicia com a seleção de uma ferramenta de pesquisa. É importante pontuar que, apesar da existência do SNOMED CT *browser* e do ICD *browser*, todos os patologistas entrevistados afirmaram que não fazem uso destas ferramentas para pesquisa terminológica. Eles relatam que quando precisam pesquisar para encontrar a versão autorizada de um termo, não são instruídos nem durante a formação na faculdade e nem no Serviço a buscar pelo SNOMED *browser*, sendo este o motivo principal da não utilização. Os outros meios citados parecem a eles mais acessíveis e convencionais que o acesso a vocabulários controlados como o SNOMED *browser*.

No processo de busca por termos autorizados, o patologista atua de forma exploratória, utilizando fontes externas e não necessariamente controladas. Embora haja recursos como o SNOMED CT *browser*, eles não estão incorporados aos sistemas utilizados no cotidiano do serviço, o que compromete a uniformidade da codificação.

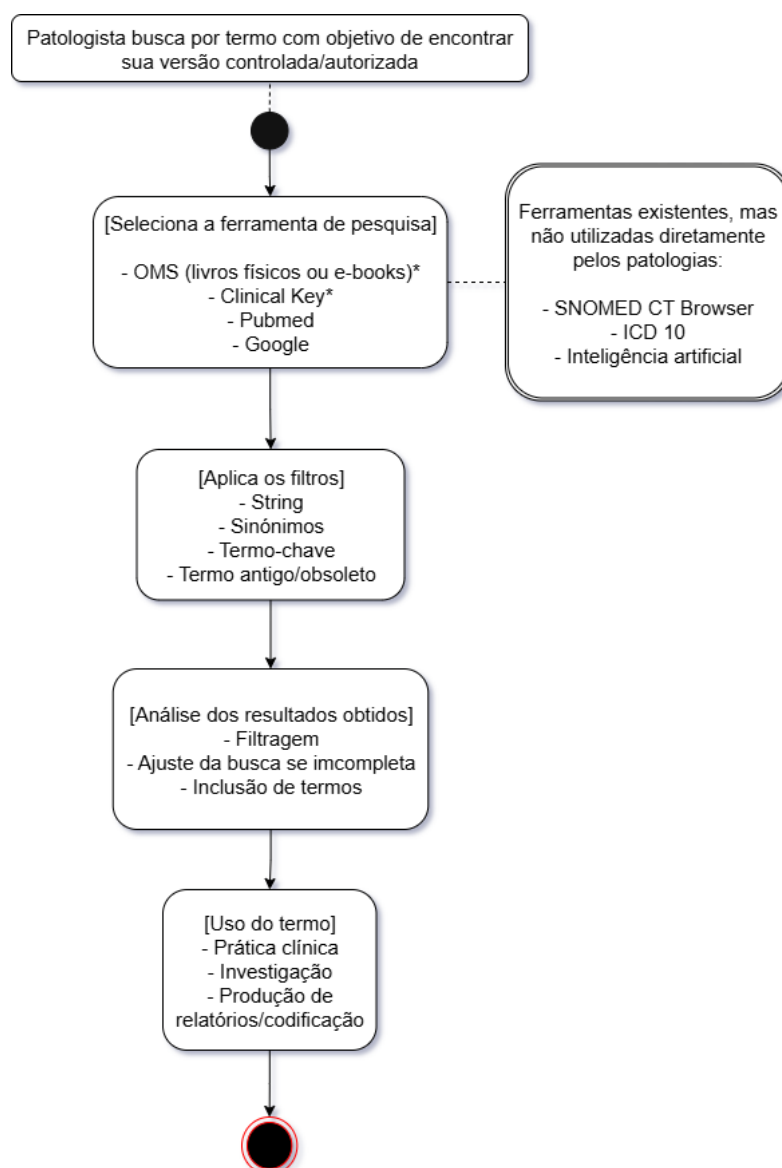


Figura 5.3: Diagrama de processo de pesquisa para encontrar o termo controlado.

O website da OMS¹ é utilizado para aceder aos livros de anatomia patológica disponibilizados a partir de uma assinatura paga. O *ClinicalKey*² é uma ferramenta da empresa *Elsevier* que consiste em uma base de dados de livros, artigos, terminologias e outras fontes capazes de dar respostas com base em evidências científicas, caracterizar patologias, medicamentos e outros temas, também em regime pago para acesso, amplamente utilizado pelos patologistas mais jovens e internos em formação específica.

A *Pubmed*, como citado no capítulo 2, subcapítulo 2.2, é uma base de dados científico-técnica gratuita que utiliza o MeSH como indexador, e fornece artigos e livros da área biomédica e de

¹ <https://publications.iarc.who.int/Book-And-Report-Series/Who-Classification-Of-Tumours>

² <https://www.elsevier.com/products/clinicalkey>

ciências da saúde. A ferramenta *Google* que funciona como ferramenta de pesquisa utilizada amplamente por todos os patologistas entrevistados para consulta de terminologia.

O segundo passo deste processo ocorre pela aplicação de filtros na pesquisa, ou seja, a seleção de termos específicos para encontrar o que se pretende, sinónimos ou até termos obsoletos para chegar ao termo atual. A terceira etapa é a análise crítica dos resultados, onde ocorre a filtragem do que foi encontrado, ajustes de pesquisa e/ou inclusão de termos para uma nova busca. Por fim, o patologista utiliza o termo para a prática clínica, para investigação ou na produção ou codificação dos relatórios.

Tanto neste como no processo seguinte, a dependência de ferramentas não integradas não beneficia o patologista no uso direto dos vocabulários controlados, como o SNOMED CT, o que gera perda de padronização terminológica. Em ambos os fluxos, também existe a preocupação com termos obsoletos, mas não existe um mecanismo automático capaz de atualizar o patologista em caso de uso de um termo em desuso. O processo então é manual e depende da experiência e tempo disponível do patologista.

O segundo diagrama, representado pela figura 5.4, exemplifica a forma como os patologistas recuperam casos já registados em sistemas de informação, como o SISPAT ou SAPJ, *Pubmed* ou *Google*. Um exemplo real de como isto ocorre é quando um patologista necessita recuperar informação sobre casos existentes em uma patologia clínica pouco recorrente no Serviço e precisa rastrear, seja no próprio Serviço ou em todo o mundo, relatórios que descrevam como foi analisado. Destaca-se que, quando as entrevistas ocorreram, o Serviço de Anatomia Patológica havia migrado de sistema, SAPJ para SISPAT, há 2 semanas. Ou seja, os patologistas não haviam recebido formação para utilizar o novo sistema, por isso, não conseguiram dar detalhes de como ocorre a pesquisa por terminologia ou levantamento de relatórios por termos no SISPAT.

Primeiro, há uma decisão por parte do patologista, se faz a busca de forma independente ou se solicita que o secretariado realize a mediação da pesquisa. Apenas 3 dos 9 patologistas entrevistados afirmam recorrer à mediação por escolha, seja devido ao melhor desempenho do SAPJ em máquinas específicas no secretariado ou para poupar-lhes tempo enquanto trabalham na produção de relatórios.

Quando não existe a necessidade de mediação, seja porque preferem ter controlo do processo de pesquisa e filtragem ou por acharem que em casos de terminologias mais complexas, que o secretariado não tem o olhar médico para selecionar o que é necessário, o patologista seleciona a ferramenta de pesquisa, aplica os filtros, que podem ser termos relacionados a morfologia, topografia, doenças ou termos obsoletos, por exemplo, escolhe a forma como vai pesquisar, seja por texto livre ou termo controlado (codificador), no caso de pesquisa pelo SISPAT ou SAPJ, e por fim, filtra os resultados para obter apenas aquilo que responde às suas questões.

Quando existe a necessidade de mediação, o patologista tem de explicar o que se pretende com a pesquisa, para dar contexto ao secretariado, o assistente técnico seleciona os termos que traduzem a dúvida e realiza a pesquisa no SISPAT ou SAPJ por termos de morfologia, topografia ou outros, o sistema gera uma listagem em excel com todos os relatórios que contém os termos

pesquisados associados, filtra as inconsistências geradas pelos erros de codificação, o técnico envia para o patologista, que também filtra para validar o que foi selecionado.

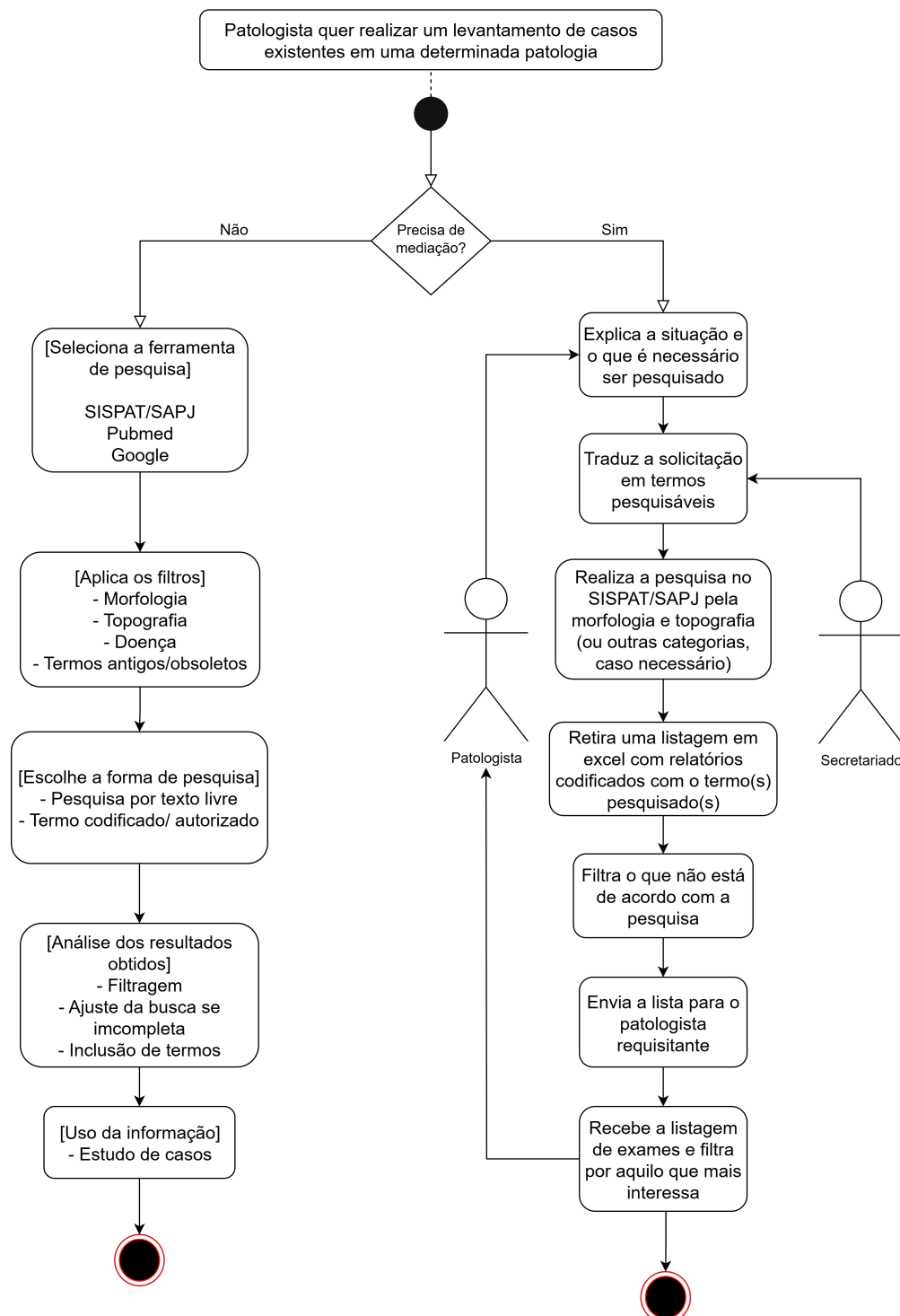


Figura 5.4: Diagrama de processo de pesquisa por termos para levantamento de casos.

Ou seja, com a mediação há um fluxo mais colaborativo e burocrático, que envolve as diferentes atuações, que revela ser um meio de suporte, mas que nem sempre funciona, pois os patologistas confirmam que o olhar que eles têm para a seleção primária de casos, a partir do primeiro resultado da pesquisa, ou mesmo na pesquisa, muitas vezes faz toda a diferença para o que se pretende.

Ambos os diagramas revelam lacunas no processo de padronização terminológica, que será melhor abordado na subsecção seguinte, o que afeta a descrição, indexação e consequentemente a recuperação dos relatórios.

No processo de levantamento de casos e relatórios, existe uma maior dependência dos sistemas do próprio serviço, mas a busca é permeada por etapas de tradução de termos e filtragem manual. A necessidade de mediação mostra que o vocabulário utilizado muitas vezes não é intuitivo nem diretamente acessível para os patologistas, mesmo sendo os próprios a codificar os relatórios. Esse aspecto confirma que o processo de codificação dos relatórios ainda não é o ideal, o que resulta em erros nesta fase de atribuição de códigos e recuperação da informação.

Desta forma os fluxos demonstram a integração insuficiente entre vocabulários controlados e sistemas internos, o que gera redundância, inconsistência e retrabalho. Além de reforçar que o enriquecimento da metainformação surge como necessidade para unificar terminologias, facilitar a atualização de termos obsoletos, reduzir a dependência de mediação e promover a interoperabilidade entre a prática clínica, investigação e arquivo.

5.2 Sistematização de termos descritores em anatomia patológica

Com base nas entrevistas realizadas e na observação dos patologistas, percebeu-se que a padronização terminológica em anatomia patológica se fundamenta sobretudo em sistemas de classificação internacionais, com destaque para o SNOMED CT e o ICD-O no âmbito das doenças oncológicas, bem como o ICD-10/11 e bibliografia especializada no caso das patologias inflamatórias. Esses referenciais são vistos como essenciais para assegurar que os relatórios se integrem numa linguagem médica universal e inteligível em diferentes contextos clínicos e geográficos. Apesar disso, a tradução dos termos do inglês para o português apresenta desafios frequentes.

Segundo a coordenadora do Serviço de Anatomia Patológica, o SNOMED anterior ao SNOMED RT (Read Terminology), sendo provável que seja SNOMED International Version 3.5, publicada entre 1998 e 2001, era utilizado com várias adaptações para se adequar à realidade e rotina. Ou seja, a versão original foi traduzida conforme a necessidade dos termos utilizados na rotina, além de integrar termos e códigos de classes como Doenças na classe Morfologia. Atualmente, desde maio de 2025, foi implementado um novo sistema, SISPAT, que utiliza o SNOMED CT, versão mais recente.

Muitos profissionais relatam a necessidade de realizar adaptações quando não existe correspondência direta, ou mesmo de manter o termo original em inglês para evitar ambiguidades. Essa prática reflete problemas referentes a adaptação à nomenclatura científica internacional e, por outro

lado, a clareza necessária na comunicação em língua portuguesa, considerando que os relatórios podem circular para além do meio médico.

O domínio da terminologia não se limita à formação inicial, mas é consolidado sobretudo durante a prática profissional. Diversos entrevistados apontam que a aprendizagem ocorre no internato, em interação com colegas mais experientes, e na adaptação contínua às exigências do serviço. Assim, a normalização terminológica depende não apenas de normas internacionais, mas também de processos de socialização profissional e da transmissão de saber tácito no contexto de trabalho.

Outro ponto recorrente é o carácter evolutivo das classificações. A nomenclatura oncológica, por exemplo, sofreu alterações ao longo do tempo, incorporando descrições que em alguma altura eram consideradas irrelevantes, mas que se tornaram fundamentais para o diagnóstico. Em paralelo, surgem categorias provisórias que agrupam entidades mal caracterizadas e das quais derivam novas classificações. Ainda assim, reconhece-se que a atualização dos vocabulários não acompanham totalmente o ritmo do avanço científico, levando a lacunas e a classificações pouco precisas em algumas áreas, como na patologia linfática.

Na prática quotidiana, as dificuldades de codificação são evidentes. Relatórios podem ser codificados de forma genérica, sem atualização posterior quando o diagnóstico é refinado, comprometendo a qualidade da informação. Por essa razão, em pesquisas retrospectivas para levantamento de relatórios, os patologistas muitas vezes recorrem ao texto livre para recuperar termos que não estão devidamente representados nos códigos. Isso demonstra que, embora indispensável para fins estatísticos de recuperação da informação e de interoperabilidade, para o patologista a codificação não substitui a riqueza descritiva do que está no próprio relatório.

Por fim, embora exista consenso quanto à importância de utilizar vocabulários controlados e de manter o maior rigor possível, observa-se que não é a maioria dos patologistas que aderem esta ideia. Os internos em formação, patologistas e técnicos confirmaram em entrevista o baixo rigor no processo de codificação, levando muitas vezes à atribuição de um código mais genérico a diferentes especificidades dentro de uma mesma patologia identificada, diminuindo a granularidade da descrição temática. A falta de atualização sistemática em relação ao surgimento de novos termos e a utilização incorreta dos mesmos são apontadas como fragilidades. Assim, para além da existência de sistemas de classificação, a consolidação de uma terminologia padronizada em anatomia patológica requer investimentos contínuos em formação e atualização profissional.

As reflexões apresentadas a partir das entrevistas permitem constatar que a padronização terminológica em anatomia patológica depende, simultaneamente, de sistemas de classificação internacionais, de práticas locais de adaptação e de processos contínuos de formação profissional.

Para ilustrar de modo concreto como essa dinâmica se manifesta a nível documental, apresenta-se, na tabela 5.4, um quadro comparativo de termos identificados em relatórios do Arquivo do Serviço de Anatomia Patológica do Hospital de São João (SAPJ) e sua correspondência atual nos sistemas SNOMED CT.

Tabela 5.4: Sistematização de termos de morfologias - SNOMED CT x SNOMED SAPJ.

Ano do relatório	Termos identificados - Relatórios	Termos atuais - SNOMED CT	Termo obsoleto - SNOMED CT ou SNOMED SAPJ
1962	Fibromiomatose difusa	125569002 Diffuse fibromatosis (morphologic abnormality)	M-88901 Leiomiomatose/fibromiomatose
1962	Pólipos mucosos	55310007 Mucous polyp of cervix (disorder)	M-76840 Polipo mucoso
1962	Salpingite folicular	88157006 Salpingitis (disorder)	D7-71234 Salpingite Follicular
1962	Formações císticas	16217341000119101 Cystic nodule (morphologic abnormality)	363641000 Cystic lesion - nondiagnostic (morphologic abnormality); M-33400 Cisto
1977	Endometriose	129103003 Endometriosis (disorder)	D7-72000 Endometriose
1977	Bacilos de Doderlein	45707005 Lactobacillus acidophilus (organism)	L-20106 Bacilos de Doderlein
1977	Numerosos polinucleares neutrófilos	1197483002 Functional disorder of polymorphonuclear neutrophil (disorder)	Sem termo correspondente
1988	Incrustações calcárias	18115005 Pathologic calcification, calcified structure (morphologic abnormality)	M-55400 Calcificação/deposição de cálcio
1988	Ovário esclero-quístico	781067001 Polycystic ovary (disorder)	69878008 Polycystic ovaries (disorder)
1994	Arias-Stella	67217006 Arias-Stella effect (morphologic abnormality)	M-79430 Arias-Stella
1996	Apendicite aguda	85189001 Acute appendicitis (disorder)	Sem termo correspondente
1996	Colecistite	76581006 Cholecystitis (disorder)	D5-85102 Colecistite sem cálculos
1996	Divertículos de Rokitansky-Aschoff	22149007 Aschoff-Rokitansky sinuses (disorder)	Sem termo correspondente
1997	Carcinoma epidermoide	1162767002 Squamous cell carcinoma (morphologic abnormality)	M-80723 Carcinoma epidermoide não queratinizante, SOE
1997	Adenocarcinoma	1187332001 Adenocarcinoma (morphologic abnormality)	M-81403 Adenocarcinoma
1997	Cirrose hepática	19943007 Cirrhosis of liver (disorder)	D5-80600 Cirrose hepática, SOE
1997	Corpos hialinos de Mallory	20894005 Mallory bodies (morphologic abnormality)	Sem termo correspondente
1997	Cistos de Naboth	767356009 Mucous retention cyst of cervix uteri (disorder)	24565001 Nabothian follicles on cervix (disorder); D7-71550 Cisto de Naboth
1997	Glândulas de Brunner	235700007 Hyperplasia of Brunner glands of duodenum (disorder)	413215003 Brunner's gland hyperplasia (disorder)
	Síndrome de Von Hippel-Lindau	Sem termo correspondente	Sem termo correspondente

Esta sistematização evidencia tanto a permanência de certas categorias diagnósticas como a obsolescência ou substituição de outras, revelando, na prática, o caráter evolutivo e negociado da terminologia em anatomia patológica. Os termos mais atuais retirados do SNOMED CT *browser* estão em inglês, pois não existe tradução para o português. Para além dos termos selecionados neste quadro, mais de 110 termos foram analisados para comparação da sua versão controlada mais recente e a utilizada pelo Serviço de Anatomia Patológica.

A existência de várias versões do SNOMED ao longo do tempo causam inconsistência. Com a mudança constante, o próprio SNOMED estabelece uma dificuldade de encontrar um termo que signifique exatamente o que os patologistas querem dizer. A sistematização apresentada no quadro 5.4 confirma a natureza dinâmica da terminologia em anatomia patológica. Nota-se que alguns termos historicamente utilizados nos relatórios mantêm correspondência direta no SNOMED CT atual, como ocorre com “endometriose” ou “adenocarcinoma”, sinalizando a estabilidade e a consolidação desses termos ao longo do tempo.

Por outro lado, observa-se a ocorrência de termos que foram substituídos ou considerados obsoletos, como “fibromiomatose difusa” ou “pólipos mucosos”. Nesses casos, a nomenclatura mais recente apresenta descrições mais precisas, alinhadas com avanços científicos e clínicos, ao passo que os termos anteriores permanecem apenas como registos históricos. Isso demonstra a necessidade constante de atualização e de vigilância terminológica, já que a persistência do uso de termos desatualizados pode comprometer a clareza da comunicação diagnóstica.

Um terceiro aspeto a destacar são os termos para os quais não existe correspondência direta na classificação atual. Casos como “síndrome de Von Hippel–Lindau” ou “corpos hialinos de Mallory” revelam limitações do sistema de codificação, que nem sempre abarca categorias menos frequentes ou descrições de processos inflamatórios específicas. Nessas situações, evidencia-se a relevância do texto livre nos relatórios de anatomia patológica como recurso indispensável para assegurar a precisão descritiva.

Assim, a análise do quadro permite compreender que a terminologia em anatomia patológica se constrói entre permanência, atualização e lacunas de termos. A coexistência desses três eixos – estabilidade, mudança e ausência de correspondência – reforça a ideia de que a padronização terminológica não é um processo acabado, mas antes um campo em constante negociação entre a prática clínica, a evolução científica e a codificação normativa.

O que se pretende posteriormente, uma vez que esta investigação percebeu um teor mais profundo devido a variável da forma de codificação a depender do patologista que a utiliza, é selecionar códigos SNOMED e trabalhar apenas em alguns deles, criando uma tabela com o períodos temporais, os termos equivalentes em cada período (coletados a partir do texto livre) e o que estava a ser descrito no relatório, para perceber o significado autorizado e o significado atribuído ao termos pelo patologista.

Em síntese, a análise conjunta das entrevistas com os patologistas e da análise documental entre termos contidos nos relatórios e o vocabulário atual evidencia que a padronização terminológica em anatomia patológica resulta de um processo híbrido: simultaneamente baseado em sistemas internacionais de referência, mas condicionado pelas práticas locais de adaptação e pela

evolução contínua do conhecimento científico. A permanência de certos termos, a substituição de outros e as lacunas de correspondência confirmam que a nomenclatura não é estática, mas sim um campo em constante negociação, no qual se articulam necessidades clínicas, rigor científico e exigências de interoperabilidade. Isto mostra a relevância de investir na capacidade de recuperar todo o texto e não apenas termos específicos dependentes de um vocabulário controlado.

5.3 Testes com ferramenta OCR nos relatórios

5.3.1 Transkribus

O Transkribus, descrito na seção 2.3, é uma plataforma que conta com inteligência artificial, especificamente algoritmos avançados de *machine learning*, para reconhecimento de texto, análise de imagens e identificação estrutural de documentos históricos manuscritos. Desenvolvido por pesquisadores e arquivistas, o software oferece reconhecimento de texto com alta precisão, suporte multilíngue e ferramentas para treinar modelos específicos de caligrafia.

A transcrição dos relatórios em papel, retirados das amostras selecionadas para esta investigação, para texto digital, realizados com esta ferramenta, conta também com uso de *machine learning* para treinamento dos modelos. Nesta seção será descrito todo o processo de leitura dos relatórios através do software, a preparação dos dados, leitura por OCR/HTR, correção e treinamento do modelo.

No âmbito deste projeto, houve uma participação no workshop "Transkribus: Transcrição Automática de Manuscritos", que ocorreu no dia 07 de junho de 2025, na Biblioteca e Arquivo Municipal de Felgueiras, com a participação da equipa responsável pelo projeto TraPrInq, 2025. Neste workshop, em conversa com os investigadores responsáveis pelo projeto, foi dito que o Transkribus está muito focado em projetos de transcrição particulares, como o TraPrInq, 2025, otimizando a ferramenta para cada transcrição em específico. Isto quer dizer que para gerar um modelo suficientemente bom para o caso específico do arquivo de anatomia patológica do HUSJ seriam necessários aproximadamente de 18 meses em testes e correções.

Como a participação ocorreu após os testes, todo o esforço de aprender a utilizar a ferramenta e as limitações para o contexto do HUSJ foram independentes. Entretanto, este intercâmbio de ideias na sessão do workshop permitiu reconhecer de forma mais segura e realista, o que validou a ideia da ferramenta demandar muito para que seja suficientemente funcional.

Previamente, na fase de observação participativa, todas as amostras foram preparadas no SPD, digitalizadas (em TIFF), analisadas pelo controle de qualidade com apoio do Serviço de Arquivo. Posteriormente as que seriam utilizadas para teste no Transkribus, amostras 1 (1962), 2 (1988), 6 (2002) e 7 (1977), descritas na tabela 4.4, foram anonimizadas com as ferramentas PAINT e GIMP, ou seja, as amostras 1 e 6 foram transformadas em PDF para análise do software com diferentes suportes de arquivo. As amostras citadas foram intencionalmente selecionadas, pois cada uma tem estruturas diferentes, qualidade de leitura diferentes e dessa forma o modelo poderia

ser testado e analisado para os diferentes formatos de relatórios existentes nas amostras analisadas na investigação.

A anonimização considera a omissão dos seguintes dados: nome, data de nascimento, naturalidade, residência, número do exame e número de exame retrospectivo citado no corpo do texto do relatórios, boletim clínico, assinatura do clínico e assinatura do chefe de serviço.

Com uso de uma conta gratuita, as primeiras tentativas de leitura por OCR na plataforma, com uso dos diferentes modelos disponíveis para a língua portuguesa, não obtiveram sucesso. Os primeiros testes foram realizados da seguinte forma: (i) inserir os relatórios em pastas por amostra e ano (AM-1-1962), (ii) selecionar 1 relatório por vez (frente e verso), (iii) selecionar a língua do modelo (português), (iv) selecionar o modelo desejado, listados na figura 5.5 e (v) dar início à leitura.

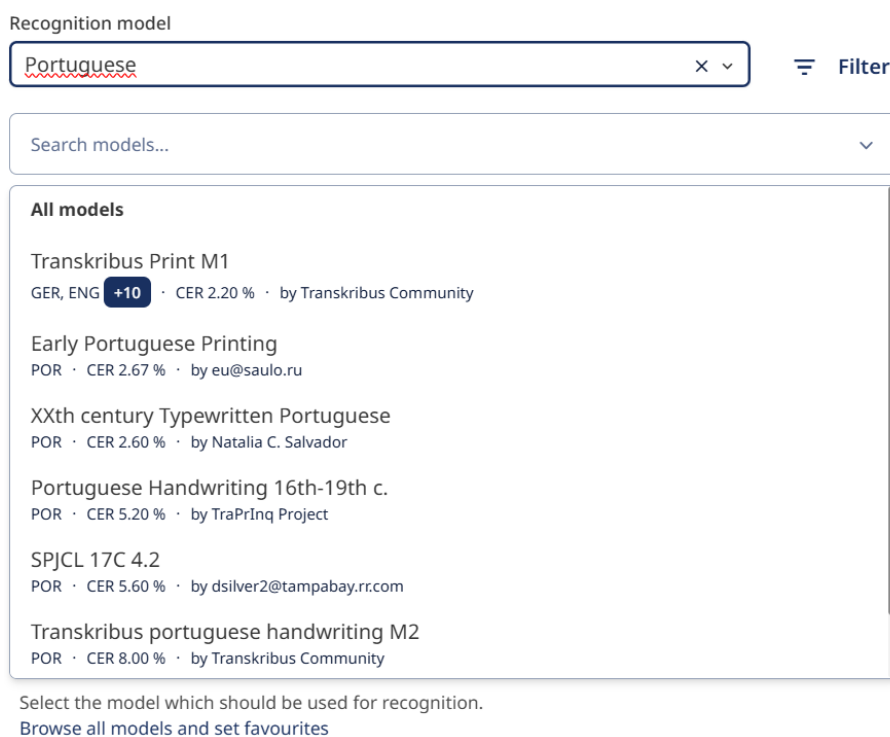


Figura 5.5: Modelos em português na plataforma Transkribus.

Como pode ser verificado na figura 5.5, cada modelo tem uma percentagem de erro de leitura de caracteres (*Character Error Rate* - CER). Com uma conta paga, que varia o custo de 900 créditos por ano entre 99€ para contas pessoais e 399€ para contas escolares, os processos de análise dos caracteres foram mais detalhados e melhor sucedidos, apesar de ainda contar com muitos erros ortográficos e principalmente falta de reconhecimento dos manuscritos. O custo do serviço pago para organizações depende de um contacto direto com a plataforma.

Foram inseridos e transcritos 40 relatórios, em sua maioria manuscritos, pois a intenção era mesmo testar a ferramenta para estes que são mais desafiadores para a transcrição automática. A amostra 6, de 2002, que apesar de alguns relatórios contarem com imagens desenhadas e notas

manuscritas, tem todo o relatório final digitado, tiveram mais sucesso na transcrição destas partes digitadas, como mostra a figura 5.6. Esta figura apresenta um relatório já corrigido manualmente, mas o seu resultado inicial se assemelhava ao resultado apresentado pela figura 5.7. Percebe-se que a transcrição sem correção apresenta bons resultados, apesar de não estarem estruturados corretamente como no relatório. A leitura de cada relatório demorou em média 30 segundos, mas a correção manual demanda de 2 a 3 minutos para ajustes estruturais e de gramática.

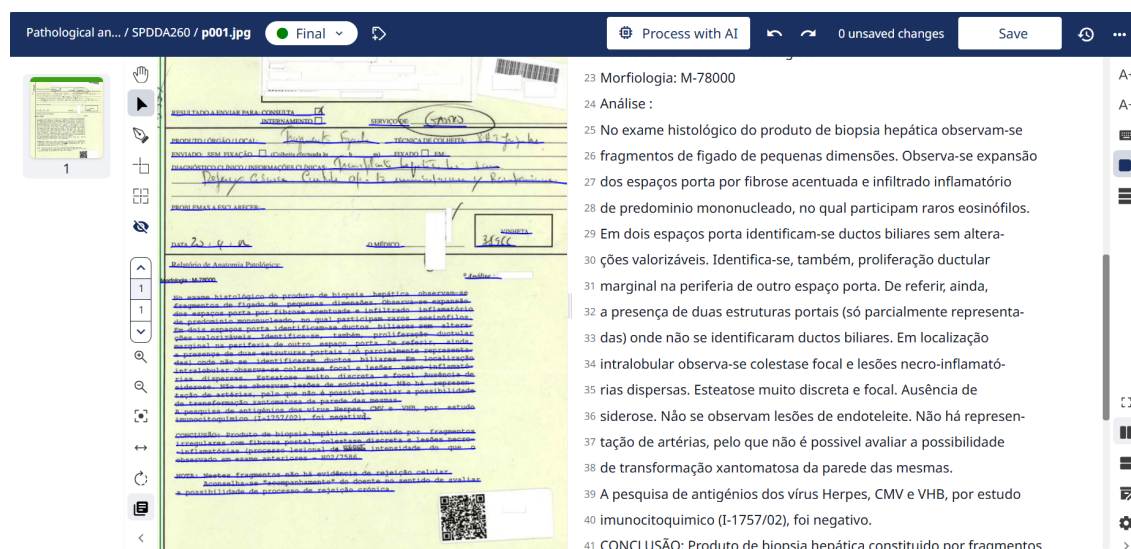


Figura 5.6: Teste com relatório digitado com correção manual - Transkribus.

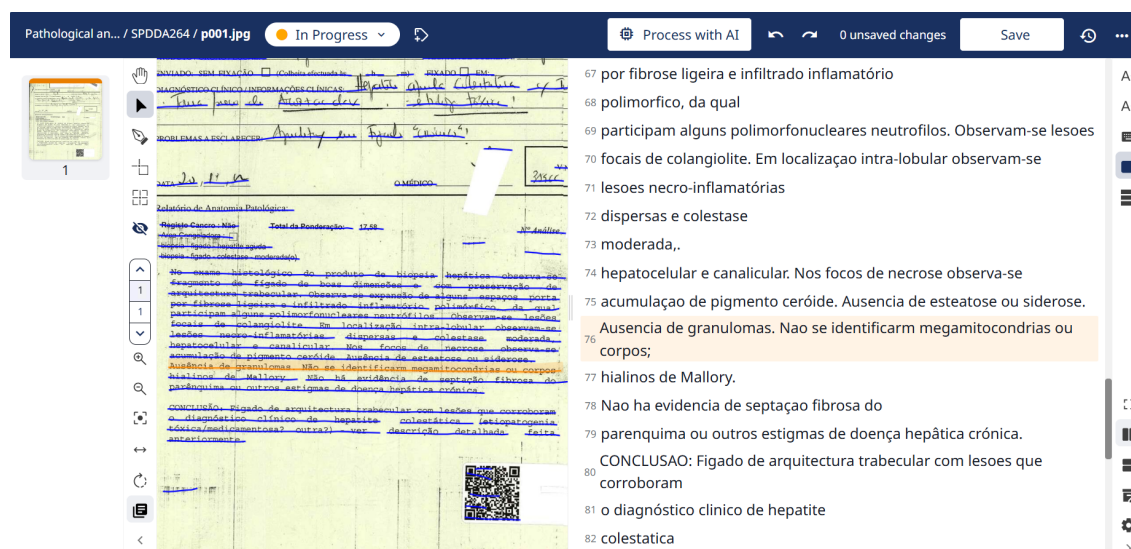


Figura 5.7: Teste com relatório digitado sem correção - Transkribus.

Assim como nos relatórios de 1988, a leitura se mostra muito eficiente para a parte do texto corrido, mas pela falta de reconhecimento adequado da estrutura, os textos dos campos de preenchimento não são transcritos e dispostos de forma adequada nos resultados. A figura 5.8 ilustra esta afirmação.

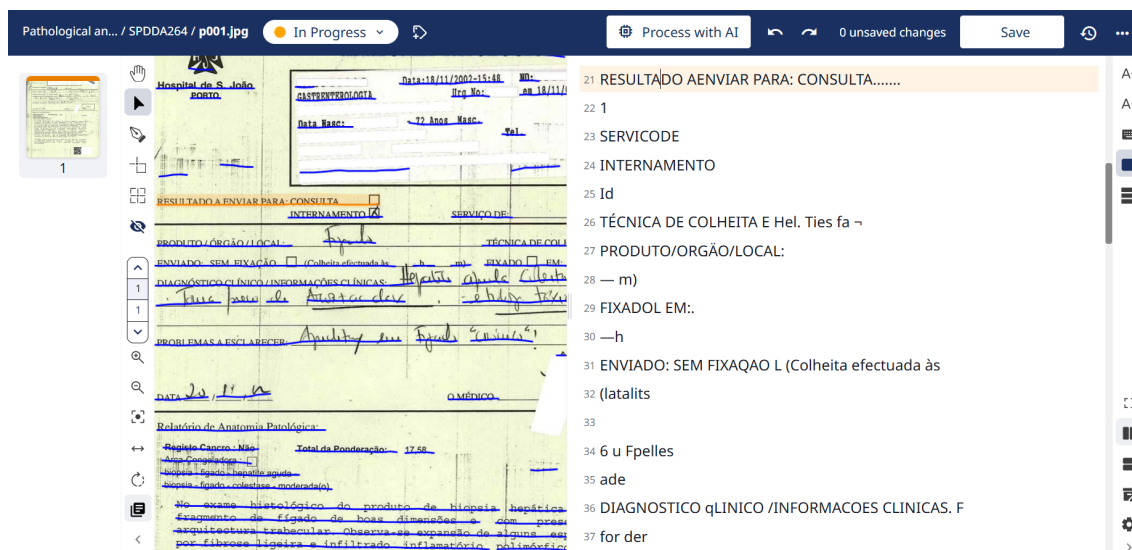


Figura 5.8: Transcrição de campos para preenchimento no relatório - Transkribus.

Para os relatórios datilografados o resultado é o mesmo, como pode-se constatar nas figuras 5.9 e 5.10, texto transcrito com qualidade, apesar de necessitar algumas correções e sem estrutura. Para estes, a dificuldade maior nas correções manuais foi a leitura, uma vez que haviam rasuras e os próprios termos muito específicos da área, que por vezes estavam ilegíveis, tornavam a sua leitura mais difícil.

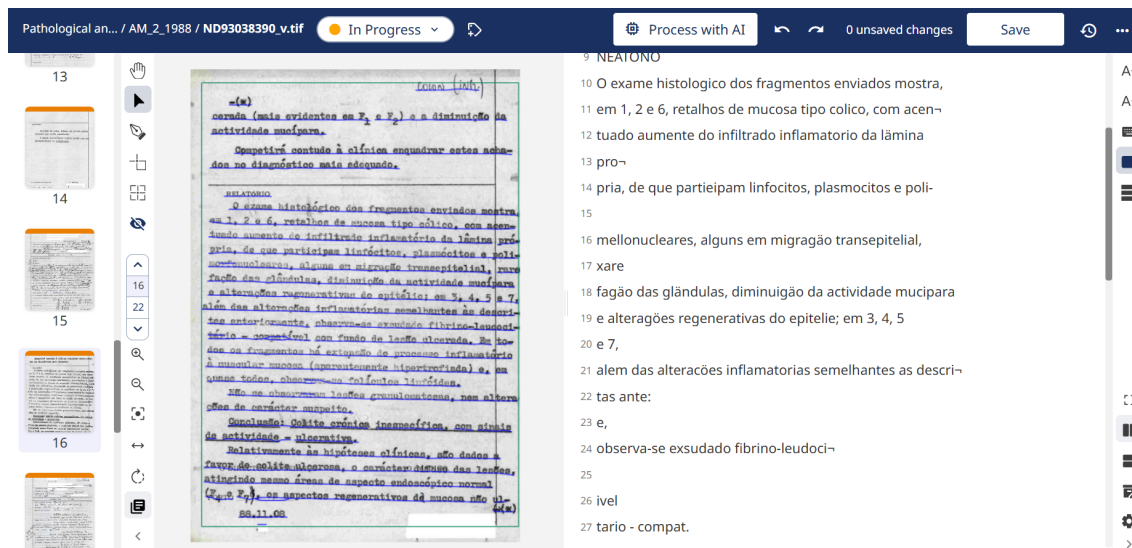


Figura 5.9: Teste com relatório datilografado - Transkribus.

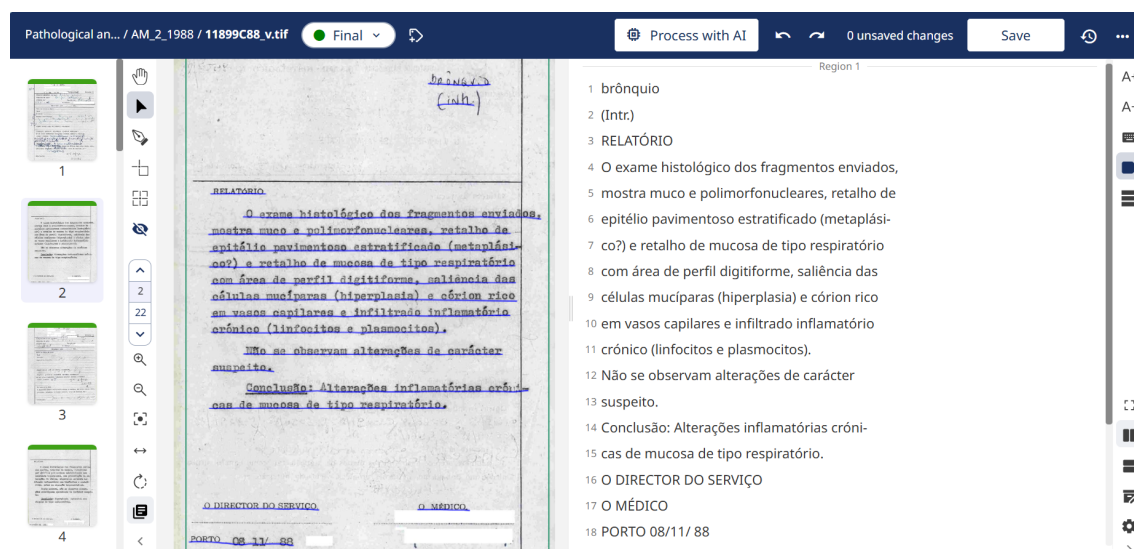


Figura 5.10: Teste com relatório pós correção - Transkribus.

Já o teste com os manuscritos mostrou que a ferramenta necessita de adaptação a longo prazo com diversas fases de treinamento de um modelo específico para estes relatórios, pois os modelos já existentes não são capazes de ler suficientemente bem o que está em manuscrito nos relatórios. A duração da transcrição destes relatórios também durou em média 30 segundos, mas devido ao alto índice de erros ou de falta de transcrição, para a correção o tempo foi de 5 minutos por relatório.

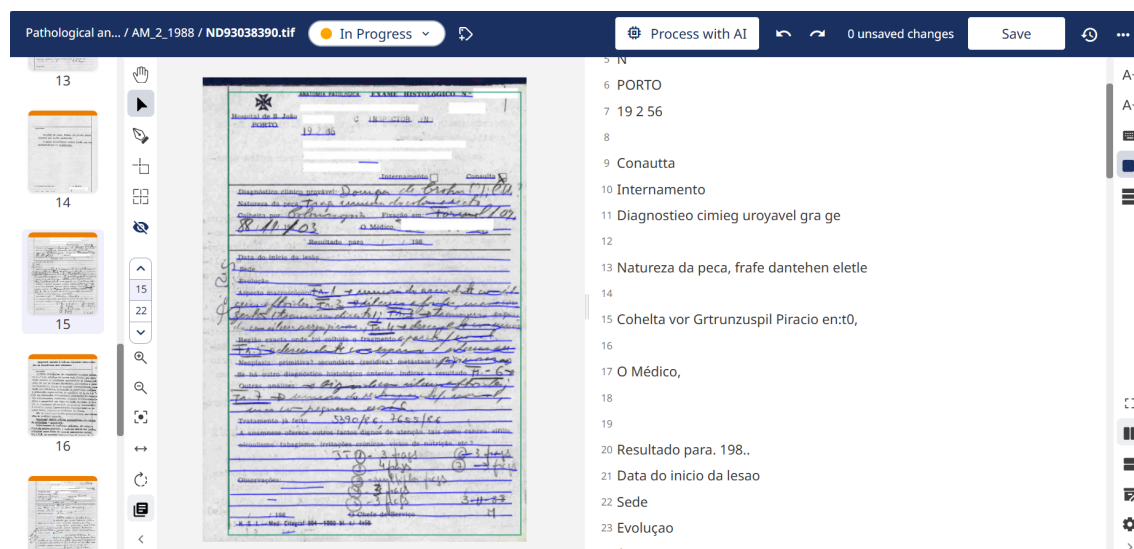


Figura 5.11: Teste com relatório manuscrito - Transkribus.

Em síntese, os testes realizados com o Transkribus evidenciam tanto o potencial quanto as limitações da ferramenta para o contexto do arquivo de anatomia patológica do HUSJ. Enquanto os relatórios datilografados e digitados apresentaram resultados satisfatórios após correções pontuais, a transcrição de manuscritos revelou-se inviável sem um processo prolongado de treinamento

de modelos específicos. A experiência reforça a utilidade do software como recurso auxiliar para digitalização e análise documental, mas também confirma que, no caso particular desta investigação, o investimento de tempo e recursos necessários para atingir resultados consistentes ultrapassa a escala deste projeto, exigindo estratégias complementares para assegurar a preservação e a acessibilidade dos relatórios históricos.

5.3.2 Landing AI

O Landing AI é uma empresa que desenvolve diferentes produtos baseados em inteligência artificial, voltados para a extração de dados em documentos, detecção de elementos em imagens e criação de modelos especializados. O produto *Agentic Document Extraction* é apresentado como uma solução que vai além do OCR e dos LLMs (*Large Language Models*), sendo capaz, segundo a própria empresa, de compreender estruturas complexas sem as modificar e mantendo as relações entre elementos e dados identificados. Além disso, afirma-se que a ferramenta consegue extrair com exatidão informações de tabelas, gráficos e imagens.

Para além disto, por ser uma plataforma de código fechado e que necessariamente se paga para utilizar, existem problemas no que se refere a segurança dos dados dos utentes, uma vez que não é possível que todos os relatórios sejam anonimizados para este processo e tampouco faz sentido fazê-lo em um cenário real, pois o acesso aos dados do doente que detém o relatório é essencial para a prestação de cuidados e até para a investigação.

Foi possível realizar testes a partir da versão experimental gratuita do *Agentic Document Extraction*. O processo é simples: basta inserir o ficheiro e a ferramenta processa automaticamente texto digitado, datilografado, manuscrito e até imagens. A figura 5.12 mostra o teste inicial de extração de texto, a figura 5.13 ilustra a análise de elementos gráficos, enquanto a figura 5.14 evidencia a extração estruturada de dados. Portanto, ao se observar com atenção os resultados, nota-se que a ferramenta por muitas vezes cria termos que se aproximem ao original, em detrimento da informação correta, para que seja preenchida toda a transcrição.

Esse comportamento é particularmente problemático no âmbito dos relatórios de anatomia patológica, pois pode levar a conclusões incorretas e comprometer tanto a integridade do registo clínico como a utilização em investigação científica. Nesse sentido, há mais perigo em aceitar uma informação inventada pela IA do que em ter de recorrer à transcrição manual.

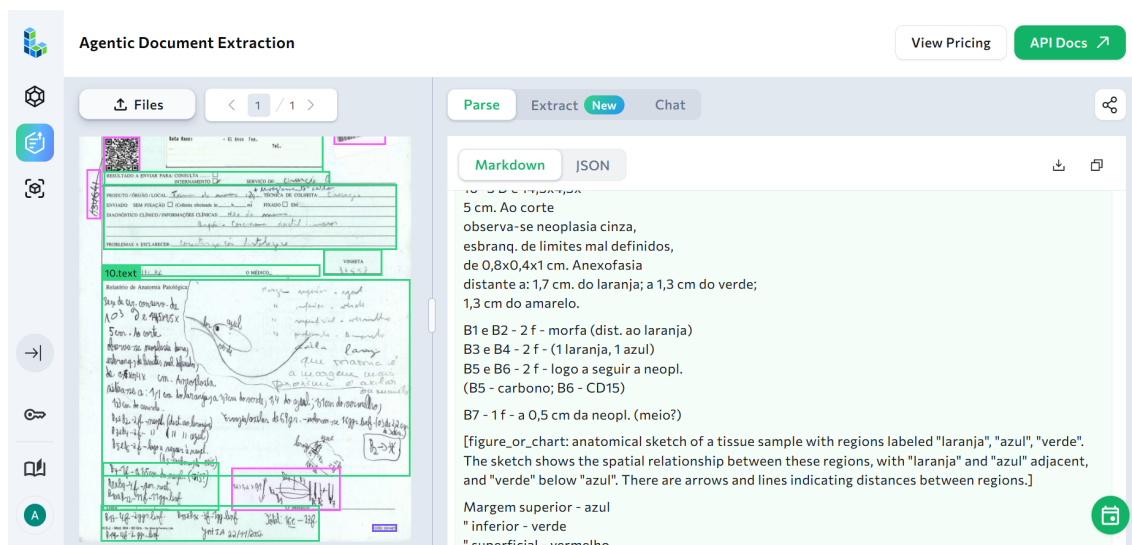


Figura 5.12: Teste inicial Landing AI

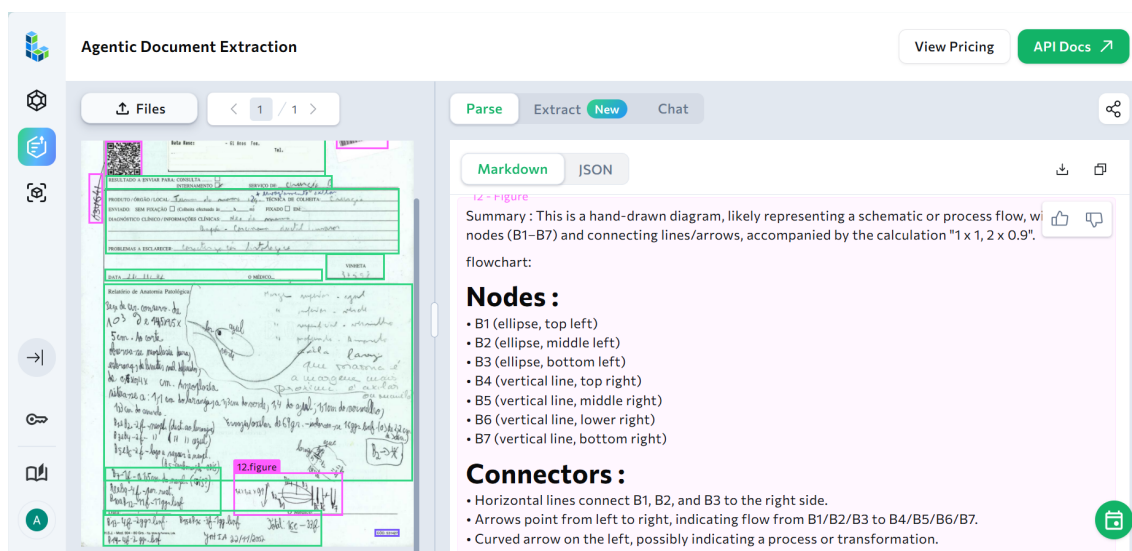


Figura 5.13: Análise de imagem no Landing AI

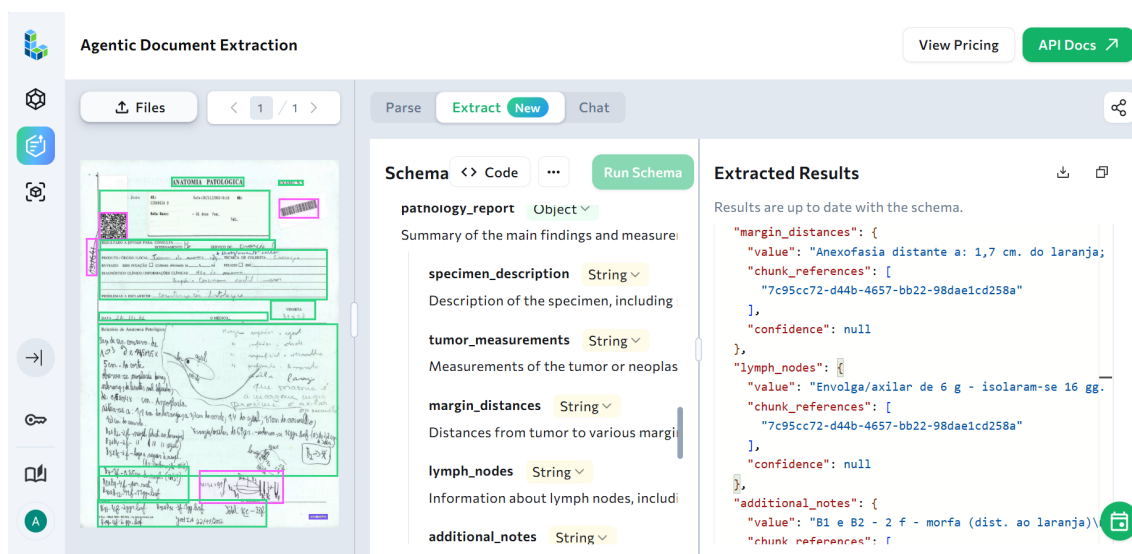


Figura 5.14: Extração de dados no Landing AI

É uma ferramenta com grande potencial, mas os entraves quanto a segurança dos dados é o principal fator que a impossibilita de seguir como a escolhida para este processo. Assim, embora se destaque pelo desempenho técnico, a sua aplicabilidade real permanece restrita, reforçando a necessidade de alternativas que combinem precisão, segurança e adaptabilidade às especificidades deste tipo de acervo documental.

5.3.3 RODA

O sistema RODA (*Repository of Authentic Digital Objects*), construído com base no OAIS (*Open Archival Information System*), constitui a estrutura de repositório digital em uso no arquivo, no RCD, orientado para a preservação e acesso de longo prazo a documentos arquivísticos. Uma das funcionalidades do sistema é a possibilidade de incorporação de mecanismos de OCR, que permitem estruturar e tornar pesquisável a informação contida em relatórios digitalizados. Para tal, o RODA recorre a diversos padrões de estruturação de metainformação, como a EAD (*Encoded Archival Description*), PREMIS (*PREservation Metadata: Implementation Strategies*) e METS (*Metadata Encoding and Transmission Standard*).

O OCR utilizado atualmente no RODA é o Tesseract, um dos projetos *open source* mais consolidados nesta área. Atualmente mantido pela *Google*, o Tesseract utiliza redes neurais convolucionais (CNNs) para a extração de caracteres, oferecendo rapidez, estabilidade e a vantagem de poder ser executado de forma local, garantindo assim a privacidade e soberania sobre os dados sensíveis. Entretanto, o Tesseract apresenta limitações no que tange o reconhecimento de manuscritos, sendo mais eficaz em textos datilografados ou impressos. No contexto do arquivo de anatomia patológica, esta limitação relativa a não leitura de manuscritos, impacta diretamente em aproximadamente 43 anos de relatórios que eram produzidos de forma manuscrita e datilografada.

Foram feitos testes com o OCR do RODA a partir dos documentos que foram previamente processados e inseridos no RCD, o resultado foi pouco satisfatório em relação ao potencial de

estruturação do texto transcrito, uma vez que a ferramenta entrega o texto sem formatação alguma. Ou seja, o *layout* é desconsiderado e apenas o conteúdo do relatório é apresentado. Além disto, caso o documento esteja com uma qualidade de imagem razoável já é o suficiente para que as falhas sejam maiores na transcrição.

Como alternativa, ferramentas comerciais como Nanonets OCR apresentam funcionalidades de extração mais avançadas, incluindo modelos treinados para lidar com diferentes tipos de caligrafia. Contudo, estas soluções implicam custos de subscrição, dependência de processamento em nuvem e potenciais riscos quanto à confidencialidades da informação em saúde, o que restringe a sua aplicabilidade direta em ambientes hospitalares.

Diante deste cenário, podem ser considerados dois caminhos complementares:

- Cenário A: utilização do OCR ativo no RODA para digitalização e pesquisa em texto livre em relatórios datilografados e digitados, além da criação de um campo para indexação dos relatórios a partir do código SNOMED inserido no próprio relatório. Esta abordagem, embora limitada quanto aos manuscritos, apresenta vantagens de privacidade e baixo custo de implementação, mas exigiria investimentos adicionais em hardware dedicado para garantir desempenho adequado, face ao volume documental.
- Cenário B: desenvolvimento de um modelo específico, treinado com base em diferentes formatos de relatórios produzidos pelo serviço de anatomia patológica. Esta opção poderia incorporar técnicas de *machine learning* orientadas para o reconhecimento de manuscritos e integrar o SNOMED CT como mecanismo de indexação automática, ampliando as possibilidades de recuperação por termos controlados e, naturalmente, em texto livre. Dessa forma, abriria-se um caminho para o enriquecimento da metainformação descritiva, um dos eixos centrais desta investigação, de forma automática e contribuindo para a reutilização dos dados.

Como forma de sintetizar os pontos principais encontrados nos testes das ferramentas, a tabela 5.5 compara as vantagens, limitações e adequação de cada uma à realidade do HUSJ.

Tabela 5.5: Comparação entre ferramentas de OCR aplicadas ao arquivo do HUSJ

Ferramenta	Vantagens	Limitações	Adequação ao HUSJ
Transkribus	- Especializada em documentos históricos e manuscritos. - Permite treinar modelos com caligrafias específicas. - Interface amigável para correção colaborativa.	- Necessita de treino extensivo para alcançar bons resultados. - Dependência de técnicos exclusivos para esta atividade. - Custos crescentes para grandes volumes.	Elevado potencial para relatórios manuscritos, desde 1962 a 2004, desde que haja investimento em treino de modelos próprios.
Landing AI	- Plataforma flexível de IA. - Permite personalização de modelos para OCR em contextos específicos. - Suporte a diferentes tipos de documentos.	- Exige infraestrutura de treino e manutenção de modelos. - Problemas de confidencialidade e privacidade em um cenário real. - Não é uma solução dedicada exclusivamente a OCR.	Adequada em projetos de investigação e desenvolvimento, mas complexa para adoção direta em ambiente hospitalar.
RODA (Tesseract)	- Open source e integrado em repositórios digitais. - Baseado em Tesseract: rápido, estável e executável localmente. - Garante privacidade e soberania sobre os dados.	- Baixa eficácia em manuscritos. - Requer hardware dedicado para grandes volumes. - Funções básicas de OCR comparadas a outras soluções.	Adequado para relatórios digitados e datilografados; insuficiente para manuscritos históricos sem apoio de modelos adicionais.

A comparação entre as três ferramentas evidencia que não existe uma solução única capaz de responder integralmente às necessidades do arquivo. Enquanto o Transkribus se mostra promissor para a transcrição dos relatórios manuscritos, o Landing AI destaca-se pela flexibilidade e potencial de desenvolvimento de modelos personalizados, embora com uma elevada complexidade técnica. Já o RODA, integrado na infraestrutura do HUSJ, oferece mais estabilidade, privacidade e adequação para os relatórios digitados, mas permanece limitado no tratamento dos relatórios manuscritos.

Esta análise destaca a importância de uma estratégia na seleção de tecnologias, guiada pela especificidades documentais e pelos recursos disponíveis no hospital. Neste sentido, torna-se necessário avançar para uma proposta de política de tratamento técnico que desenhe uma padronização terminológica, a digitalização, o OCR e o enriquecimento da metainformação descritiva.

5.4 Proposta de política de tratamento técnico

Durante e após a investigação foi possível estabelecer pontos cruciais para o desenvolvimento de uma política de tratamento técnico para o arquivo de anatomia patológica do HUSJ. A abordagem pós-custodial ajudou a desenvolver estes aspetos para processo dos documentos a partir da visão holística, do contexto de produção dos relatórios, dos processos técnicos envolvidos e a análise retrospectiva de todo o processo de gestão deste arquivo. Portanto, com intenção de melhorar o

fluxo de tratamento destes documentos, a recuperação da informação, a interoperabilidade e acessibilidade, foram definidas orientações técnicas a considerar numa futura política de tratamento, em alinhamento com o objetivo do trabalho.

Quanto aos objetivos e princípios, é primordial garantir a preservação de longo prazo, a rastreabilidade e o acesso qualificado aos relatórios citológicos e histológicos, tanto para apoio à decisão clínica como à investigação científica. Por isso, implementar um modelo sistémico, centrado na informação enquanto um fenómeno social, dinâmico e contextual, valorizando a participação ativa das comunidades de utilizadores, patologistas, clínicos e investigadores, na construção do processo descritivo e de indexação se mostra essencial para o desenvolvimento de uma política efetiva.

Quanto às linhas orientadoras fundamentais a trabalhar no desenvolvimento de uma estratégia piloto, foram divididas em 5 tópicos, sendo eles (i) Estruturação e digitalização documental; (ii) Enriquecimento da metainformação descritiva; (iii) Automação e qualidade descritiva; (iv) Padronização, formação e governança; (v) Segurança, ética e conformidade. Pode-se sugerir que em relação a:

1. Estruturação e digitalização documental:

- Todos os relatórios anteriores e posteriores à informatização devem ser integrados no RCD, com tratamento prioritário por critérios de demanda e relevância científica e clínica. Seria interessante uma seleção prévia para processamento de relatórios relevantes para os registos clínicos atuais, com ajuda dos patologistas, e gradualmente tratar os relatórios dos mais recentes para os mais antigos. Neste aspeto, relativo aos relatórios anteriores à 1992, em que o utente não esteja informatizado no SONHO, num primeiro momento só devem ser digitalizados a pedido ou os que tiverem sido referidos como reutilizados pela Anatomia Patológica para doentes ativos. Para aqueles relatórios posteriores a 1992, em suporte de papel, deve-se tratar dos mais recentes aos mais antigos, por um critério de seleção do que se define como relatório pertinente ou não.
- O processo de preparação documental deve ser padronizado, sendo a identificação, avaliação da integridade, classificação por tipologia (citológico/histológico), rastreio por QR code e verificação rigorosa da metainformação antes e após digitalização.
- A política deve determinar que, na digitalização, os documentos com ausência de metainformação devem ser objeto de investigação complementar ou anotação manual para garantir sua futura identificação. Todas as versões (provisória, final, complementares) devem ser agrupadas sob um único registo, respeitando a historicidade documental.

2. Enriquecimento da metainformação descritiva:

- Inserir elementos de metadados no RCD para termos indexadores/codificadores do documento, como forma de materializar o vocabulário para a visualização e navegação do utilizador. A ideia é que seja possível clicar no termo para encontrar documentos indexados pelo mesmo termo.

- Estabelecer o SNOMED CT como vocabulário de referência, promovendo a migração progressiva dos códigos antigos e adaptações locais para a versão mais recente, mapeando termos históricos aos atuais com uma tabela de correspondência. Além de buscar a atualização contínua, com registo de versões, para acompanhar a evolução científica e terminológica.
- Prever campos para registo de conteúdo de texto livre, anotando termos não contemplados no vocabulário controlado, complementando a estrutura descritiva. Ou seja, não se pretende abandonar por completo os termos não controlados utilizados na rotina do Serviço de Anatomia Patológica, mas sim incluir estes termos de forma adicional ao vocabulário controlado para buscar não só a padronização, mas a utilização plena do arquivo pela comunidade designada.

3. Automação e qualidade descritiva:

- Adequar uma ferramenta de OCR, com prioridade para soluções com capacidade de customização, sugerindo-se a aplicação do Transkribus aos manuscritos e o Teressact a documentos digitais e, futuramente, modelos treinados localmente com *machine learning*/PNL para reconhecimento automático de termos clínicos, morfologias e topografias. O modelo criado será alimentado por uma ontologia, também desenvolvida localmente, que sistematize o termo controlado e seu significado, a partir do SNOMED CT, e sua correspondência cronológica na evolução terminológica em anatomia patológica.
- Definir protocolo para correção/revisão humana dos resultados do OCR e para atribuição automatizada de termos, reduzindo a subjetividade, garantindo precisão, *recall* e padronização terminológica.
- Desenvolver projetos-piloto regulares para aferição da qualidade do OCR, mineração de texto e resultados de pesquisa no RCD, com retorno ativo da comunidade de utilizadores e indicadores de desempenho, como a precisão na extração automática de termos, tempo de processamento e índice de recuperação da informação.

4. Padronização, formação e governança:

- Implementar ciclos de formação e atualização para patologistas, técnicos e secretariado em normalização terminológica, boas práticas descritivas, uso do SNOMED CT, fluxo de preparação e digitalização e ferramentas de automação.
- Criar grupo de trabalho responsável pela curadoria terminológica, alinhamento de versões do SNOMED CT, triagem de ambiguidades e supervisão ética do uso de IA.
- Reavaliar a política periodicamente com base em resultados dos processos técnicos estabelecidos, avanços tecnológicos e contexto científicos, garantindo o processo de revisão contínua.

5. Segurança, ética e conformidade:

- Garantir a anonimização dos dados onde aplicável, controle do acesso por permissões, logs de consulta e auditoria, com total conformidade ao RGPD e às considerações da Núcleo de Saúde Digital Global e Relações Internacionais e Unidade de Cibersegurança, 2025 e do Regulamento Europeu da IA na saúde.

- Prever estratégias de preservação digital e interoperabilidade do repositório com outros sistemas hospitalares, permitindo acesso transparente, seguro e rastreável aos registos arquivísticos, e facilitando a pesquisa.

No que se refere ao enriquecimento de metainformação, que é o foco do presente trabalho, a proposta de inserir elementos de metadados no RCD para termos indexadores/codificadores do documento, como forma de materializar o vocabulário para a visualização e navegação do utilizador, sugere-se desenvolver este campo com base no *standard CDA (Clinical Document Architecture)* (2006), que foi utilizado como inspiração para definição dos elementos de metadados do RCD. Concretamente, propõe-se utilizar os campos da tabela 5.6. Os termos seriam alimentados por um tesouro pré-definido com termos controlados SNOMED CT e termos em linguagem natural coletados dos textos livres transcritos, como sugerido nas orientações.

Tabela 5.6: Proposta de elementos adicionais no RCD para metainformação descritiva em relatórios citológicos e histológicos.

Campo	Vocabulário	Exemplos (SNOMED CT + linguagem natural)	Observações
Diagnóstico Principal	SNOMED CT	1187332001 Adenocarcinoma (morphologic abnormality) / "adenocarcinoma invasivo"	Código controlado e termo natural.
Diagnósticos Secundários / Manifestação (MFST)	SNOMED CT	94381002 Metastatic malignant neoplasm to liver (disorder) / "nódulo metastático no fígado"	Permite ligar diagnóstico a manifestações.
Topografia (TOP)	SNOMED CT	3341006 Right lung structure (body structure) / "lesão no lobo superior direito"	Localização anatómica.
Causa (CAUS)	SNOMED CT + linguagem natural	240532009 Human papillomavirus infection (disorder) / "associado a HPV"	Relaciona causa e achado.
Razão / Justificativa do Exame (RSON)	SNOMED CT + Linguagem natural	/ 44991000119100 Abnormal uterine bleeding (disorder) / "sangramento uterino anormal"	Extraído do resumo clínico ou requisição.
Observações microscópicas	SNOMED CT + Linguagem natural	/ 4916000 Pleomorphism (morphologic abnormality) / "células com hipercromasia e pleomorfismo"	Detalhes morfológicos.
Procedimentos relacionados	SNOMED CT + Linguagem natural	86259008 Biopsy of liver (procedure) (86273004) / "biópsia core do fígado"	Associa o procedimento ao relatório.
Estado / Qualidade da amostra	Linguagem natural	"material insuficiente", "fixação inadequada"	Essencial para reuso científico.
Termos livres	Linguagem natural	"necrose", "lesão benigna", "reação inflamatória"	Apoio a indexação controlada.

Com estas orientações propõe-se garantir que o arquivo de anatomia do HUSJ seja um ativo estratégico para a decisão clínica e investigação, promovendo a máxima usabilidade da informação por meio do enriquecimento de metainformação descritiva, automação ética e padronização

terminológica, alinhando-se às melhores praticas da ciência da informação, arquivística hospitalar e transformação digital.

Capítulo 6

Conclusões e trabalho futuro

Este capítulo apresenta as conclusões da investigação e indica perspectivas de trabalho futuro. Partindo da análise do arquivo de anatomia patológica do HUSJ, destacam-se os principais contributos do estudo, que evidenciam tanto as dificuldades estruturais, técnicas e terminológicas na recuperação da informação, apontando cenários relativos ao potencial do enriquecimento da metainformação descritiva para promover maior acessibilidade, interoperabilidade e reutilização dos registos. Neste contexto, são assinaladas as limitações encontradas e referidas propostas de continuidade, de forma a consolidar e aprofundar os resultados alcançados. Na presente conclusão, importa referir contributos, limitações e trabalhos futuros, apresentando-se uma reflexão quanto ao trabalho desenvolvido.

6.1 Contributos

A investigação desenvolvida teve como ponto de partida a constatação de que o arquivo de anatomia patológica do Hospital Universitário de São João representa um repositório singular de informação clínica e científica, acumulada ao longo de mais de seis décadas, mas cuja utilização tem sido dificultada por fatores estruturais, técnicos e terminológicos.

A falta de conexão entre os suportes, a coexistência de documentos manuscritos, datilografados e nado-digitais, a ausência de padronização terminológica e falta de uma política clara de tratamento técnico configuram-se como entraves significativos à recuperação da informação. Neste contexto, a questão central desta dissertação consistiu em compreender de que forma o enriquecimento da metainformação descritiva poderia apoiar a definição de uma política de tratamento do arquivo, tornando-o mais acessível, interoperável e reutilizável.

É importante referir que, apesar de definido como fase final, na tabela 3.1, o teste de utilizador com os patologistas entrevistados não foi realizado, devido ao atraso das entrevistas, à dificuldade de seleção de uma ferramenta OCR adequada para a realidade do Arquivo e a falhas dificuldades na sua implementação no RCD apra realização de testes. A intenção dar continuidade à investigação para que seja possível realizar esta fase e aplicar, de facto, a política de tratamento e o fluxo de enriquecimento de metainformação proposta.

Os contributos deste estudo podem ser analisados em diferentes dimensões. Em primeiro lugar, em relação aos aspetos científico e académico, a dissertação reforça a relevância da Ciência da Informação aplicada ao domínio da saúde, mais especificamente à arquivística no âmbito hospitalar. A proposta de um modelo de enriquecimento da metainformação, apoiado em normas de descrição arquivística e em vocabulários controlados, demonstra que é possível combinar práticas tradicionais de descrição com ferramentas tecnológicas atuais. Além disto, ao sistematizar a evolução terminológica em anatomia patológica e ao discutir a aplicabilidade do SNOMED CT para o contexto do arquivo, a investigação contribui para o debate da integração de terminologia internacionalmente utilizada e com a realidade da indexação/codificação local acumulada ao longo dos anos.

Do mesmo modo, as entrevistas conduzidas com patologistas e investigadores, posteriormente analisadas e codificadas no NVIVO, permitiram compreender as necessidades informacionais dos utilizadores e os obstáculos enfrentados no quotidiano da recuperação da informação, produção e indexação/codificação de relatórios e terminologias na área da anatomia patológica. Esse levantamento empírico constituiu a base para a definição de campos de metainformação descritiva e para a proposta de uma política de tratamento técnico adaptada ao contexto do HUSJ.

Em relação ao plano técnico e metodológico, a dissertação demonstrou a viabilidade da aplicação de ferramentas de reconhecimento ótico de caracteres (OCR) ao tratamento de relatórios históricos. Os testes realizados com Transkribus, Landing AI e RODA revelaram que, apesar de limitações inerentes à qualidade dos documentos originais e limitações pontuais das próprias ferramentas em relação à realidade do hospital, é possível digitalizar e extrair texto de forma eficiente, desde que o processo seja acompanhado de protocolos de validação humana.

Em quarto lugar, em relação a parte prática e institucional, a investigação oferece um contributo aplicável, funcionando como uma espécie de guia para a implementação de uma política de tratamento técnico do arquivo. A política delineada integra três eixos centrais: (i) a sistematização de descritores terminológicos a partir de vocabulários controlados; (ii) a utilização de tecnologias de digitalização e extração de texto para acelerar o tratamento massivo dos relatórios; (iii) e a definição de diretrizes de preservação, acesso e interoperabilidade com o Repositório Clínico Digital. Ao propor este modelo, a investigação mostra que é possível transformar o arquivo em um ativo estratégico para a comunidade do hospital, útil não apenas para a prática clínica, mas também para a investigação científica e para a preservação da memória institucional.

Por fim, a nível social e estratégico, a dissertação destaca que a disponibilização estruturada dos relatórios de anatomia patológica representa um ganho para múltiplos atores. Para os clínicos, uma vez que permite acesso mais rápido e preciso à informação essencial ao diagnóstico e à continuidade de cuidados. Para os investigadores, tendo presente que abre novas possibilidades de reutilização de dados históricos em estudos epidemiológicos, de recorrência de doenças ou de evolução terapêutica. Para a instituição, considerando que confere ao arquivo um estatuto patrimonial, transformando-o em memória de décadas de prática médica e científica e da própria comunidade portuense e de todas ligadas ao HUSJ.

Em síntese, conclui-se que o enriquecimento da metainformação descritiva constitui um instrumento central para a modernização do arquivo de anatomia patológica do HUSJ. O modelo proposto não elimina, por si só, todas as dificuldades técnicas e organizacionais identificadas, mas oferece um caminho claro e fundamental para superar as fragilidades existentes. Ao propor a articulação entre terminologia controlada, tecnologia de OCR, análise empírica das necessidades informacionais e diretrizes institucionais, a investigação demonstra que é possível avançar de um arquivo fragmentado para um sistema estruturado, sustentável e alinhado com os princípios da Ciência da Informação.

6.2 Trabalho futuro

Embora os resultados obtidos sejam significativos, abrem-se diversas linhas de trabalho futuro que poderão aprofundar e expandir os contributos desta investigação.

Um primeiro caminho sugere um desenvolvimento de técnicas de processamento de linguagem natural e mineração de texto aplicadas ao texto digitalizado. A automatização da extração de entidades clínicas, como diagnósticos, topografias, morfologias, permitirá enriquecer de forma significativa o processo de atribuição de termos indexadores aos documentos e reduzir a dependência de processos manuais.

Um segundo eixo de continuidade do trabalho, diz respeito à análise cronológica da evolução terminológica em anatomia patológica. Levando-se em consideração o significado adquirido a partir do perfil do patologista que o atribuiu ao relatório, utilizando-se a informação contextual e pessoal dos patologistas do serviço, seria possível montar uma ontologia por períodos que seja capaz de explicitar o real significado de um termo utilizado. Pretende-se verificar o motivo da utilização do termo e seu padrão de uso pelo patologista. Desta forma, é possível criar uma ferramenta com apoio da IA capaz de atribuir a um termo específico, um padrão descritivo em texto livre nos relatórios que o caracterize.

Sugere-se também que futuros trabalhos avaliem empiricamente a implementação da política de tratamento técnico proposta nesta dissertação. Indicadores como a precisão da recuperação da informação, o tempo médio de resposta a pedidos clínicos, o grau de interoperabilidade com outros sistemas hospitalares e a satisfação dos utilizadores poderão fornecer dados objetivos sobre a eficácia das medidas recomendadas.

Em suma, os trabalhos futuros deverão explorar tanto as dimensões técnicas, relacionadas com a evolução da inteligência artificial e das ontologias terminológicas, como as dimensões organizacionais e éticas, garantindo que os arquivos clínicos evoluam para ativos da comunidade e para disseminação do conhecimento, ao serviço da prática clínica e da investigação científica.

Referências

- ALMEIDA, Jobson Louis Santos De; PERUCCHI, Valmira; FREIRE, Gustavo Henrique Araújo, 2019. A pesquisa-ação como estratégia metodológica na Ciência da Informação. *Perspectivas em Gestão & Conhecimento*. Vol. 9, n.º 3. URL DOI: [10.21714/2236-417X2019v9n3p130](https://doi.org/10.21714/2236-417X2019v9n3p130).
- ALVARENGA, Lúcia, 2003. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*. Vol. 8, n.º 15, pp. 18–40. URL DOI: [10.5007/1518-2924.2003v8n15p18](https://doi.org/10.5007/1518-2924.2003v8n15p18).
- BERTALANFFY, Ludwig von, 1968. *General System Theory: Foundations, development, applications*. George Braziller.
- BRADBURY, Hilary (ed.), 2015. *The SAGE Handbook of Action Research*. Third edition. London: SAGE Publications Ltd. ISBN 978-1-4462-9454-3.
- BRADLEY, Chris; HENDERSON, Deborah; EARLEY, Susan; DATA ADMINISTRATION MANAGEMENT ASSOCIATION (eds.), 2024. *DAMA-DMBOK: Data Management Body of Knowledge*. Second edition, revised. Basking Ridge, New Jersey: Technics Publications. ISBN 978-1-63462-236-3.
- BUCKLAND, Michael K., 1991. Information as thing. *Journal of the American Society for Information Science*. Vol. 42, n.º 5, pp. 351–360. URL DOI: [10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-ASI5>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3).
- BURGER, Gerard; ABU-HANNA, Ameen; DE KEIZER, Nicolette; CORNET, Ronald, 2016. Natural language processing in pathology: a scoping review. *Journal of Clinical Pathology*. Vol. 69, n.º 11, pp. 949–955. URL DOI: [10.1136/jclinpath-2016-203872](https://doi.org/10.1136/jclinpath-2016-203872).
- CAPURRO, Rafael; HJØRLAND, Birger, 2003. The concept of information. *Annual Review of Information Science and Technology*. Vol. 37, n.º 1, pp. 343–411. URL DOI: [10.1002/aris.1440370109](https://doi.org/10.1002/aris.1440370109).
- CARRAJÓ, Lino; PENAS, Ángel; MELCÓN, Rubén; GONZÁLEZ, Fco Javier; COUTO, Eduardo, 2008. From Documents on Paper to Electronic Medical Records. *Studies in health technology and informatics*. Vol. 136, pp. 395–400.
- CASTRO, Onofre Ferreira; TEIXEIRA JÚNIOR, Mário; TAVARES, Paulo, 1977. Implantação de Processamento Eletrônico de Dados Na Seção de Anatomia Patológica Do Instituto Nacional de Câncer (INCa). *Revista Brasileira de Cancerologia*, pp. 31–36.
- CHEN, Richard J.; LU, Ming Y.; CHEN, Tiffany Y.; WILLIAMSON, Drew F. K.; MAHMOOD, Faisal, 2021. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*. Vol. 5, n.º 6, pp. 493–497. URL DOI: [10.1038/s41551-021-00751-8](https://doi.org/10.1038/s41551-021-00751-8).
- CRESWELL, John W.; POT, Cheryl N., 2018. *Qualitative Inquiry & Research Design: Choosing among Five Approaches*. Fourth edition, international student edition. Los Angeles London New Dehli Singapore Washington DC Melbourne: SAGE. ISBN 978-1-5063-6117-8.

- CUNHA, Murilo Bastos Da; CAVALCANTI, Cordélia Robalinho De Oliveira, 2008. *Dicionário de Biblioteconomia e Arquivologia*. Brasília: Briquet de Lemos. ISBN 978-85-85637-35-4.
- CURRÁS, Emilia, 1995. *Tesaurus: linguagens terminológicas*. Brasília: IBICT. ISBN 85-7013-052-X.
- DOLIN, R. H.; ALSCHULER, L.; BOYER, S.; BEEBE, C.; BEHLEN, F. M.; BIRON, P. V.; SHABO (SHVO), A., 2006. HL7 Clinical Document Architecture, Release 2. *Journal of the American Medical Informatics Association* [online]. Vol. 13, n.º 1, pp. 30–39 [acedido em 2025-09-12]. URL DOI: [10.1197/jamia.M1888](https://doi.org/10.1197/jamia.M1888).
- DURANTI, Luciana, 1994. Registros documentais contemporâneos como provas de ação. *Estudos Históricos*. Vol. 7, n.º 13, pp. 49–64. ISSN 2178-1494.
- EHSANI, Sepehr; KIEHL, Tim-Rasmus; BERNSTEIN, Andrea; GENTILI, Fred; ASA, Sylvia L; CROUL, Sidney E, 2008. Creation of a retrospective searchable neuropathologic database from print archives at Toronto's University Health Network. *Laboratory Investigation*. Vol. 88, n.º 1, pp. 89–93. URL DOI: [10.1038/labinvest.3700694](https://doi.org/10.1038/labinvest.3700694).
- FONSECA, Maria Odila Kahl, 2005. *Arquivologia e ciência da informação*. 1. ed. Rio de Janeiro: FGV. ISBN 978-85-225-0503-6.
- FUJITA, Mariângela Spotti Lopes; ALVES, Roberta Caroline Vesu; ALMEIDA, Carlos Cândido De (eds.), 2020. *Modelos de Leitura Documentária Para Indexação: Abordagens Teóricas Interdisciplinares e Aplicações Em Diferentes Tipos de Documentos*. Marília, SP: Editora UNESP. ISBN 978-65-86546-07-1.
- GHOSH, Sanjib Kumar, 2017. Giovanni Battista Morgagni (1682–1771): Father of Pathologic Anatomy and Pioneer of Modern Medicine. *Anatomical Science International*. Vol. 92, n.º 3, pp. 305–312. URL DOI: [10.1007/s12565-016-0373-7](https://doi.org/10.1007/s12565-016-0373-7).
- GILLILAND, Anne J., 2008. Setting the Stage. Em: *Introduction to Metadata*. 2nd ed. Ed. por BACA, Murtha. Los Angeles: Getty Research Institute. ISBN 978-0-89236-896-9.
- GONÇALVES, Fernanda; DAVID, Gabriel, 2011. Are We on the Right Track to Paperless Hospitals? Em: *Colóquio Internacional MEDINFOR II – A Medicina na Era da Informação*.
- GOUVEIA, Luís Borges, 2025. Inteligência Artificial e memória profunda: transformando o acesso e a interpretação dos arquivos. *Cadernos BAD*. N.º 1–2.
- HAYNES, David, 2018. *Metadata for Information Management and Retrieval: Understanding Metadata and Its Use*. 2nd edition. London: Facet Publishing. ISBN 978-1-85604-824-8.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2013. *ISO 25964-2:2013, Information and Documentation — Thesauri and Interoperability with Other Vocabularies*. Disponível em: <https://www.iso.org/obp/ui/en/#iso:std:iso:25964:-2:ed-1:v1:en>.
- KARIM, Kelvin, 2001. Assessing the strengths and weaknesses of action research. *Nursing Standard*. Vol. 15, n.º 26, pp. 33–35. URL DOI: [10.7748/ns2001.03.15.26.33.c2994](https://doi.org/10.7748/ns2001.03.15.26.33.c2994).
- KOERICH, Magda Santos; BACKES, Dirce Stein; SOUSA, Francisca Georgina Macêdo De; ERDMANN, Alacoque Lorenzini; ALBURQUERQUE, Gelson Luiz, 2009. Pesquisa-Ação: Ferramenta Metodológica Para a Pesquisa Qualitativa. *Revista Eletrônica de Enfermagem*. Vol. 11, n.º 3. URL DOI: [10.5216/ree.v11.47234](https://doi.org/10.5216/ree.v11.47234).
- LAKATOS, Eva Maria; MARCONI, Marina De Andrade, 2017. *Fundamentos de metodologia científica*. 8 ed. São Paulo: Editora Atlas. ISBN 978-85-97-01076-3.

- LANCASTER, F. W., 2004. *Indexação e resumos: teoria e prática*. 2ª ed. Brasília: Briquet de Lemos/Livros. ISBN 978-85-85637-24-8.
- LE COADIC, Yves-François, 1996. *A ciência da informação*. Briquet de Lemos. ISBN 85-85637-08-0.
- MÉNDEZ, Eva; HOOLAND, Seth Van, 2014. Metadata typology and metadata uses. Em: SICILIA, Miguel-Angel. *Handbook of Metadata, Semantics and Ontologies*. WORLD SCIENTIFIC, pp. 9–39. ISBN 978-981-283-630-4. URL DOI: [10.1142/9789812836304_0002](https://doi.org/10.1142/9789812836304_0002).
- NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010. *ANSI/NISO Z39.19-2005 (R2010), Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. NISO. URL DOI: [10.3789/ansi.niso.z39.19-2005R2010](https://doi.org/10.3789/ansi.niso.z39.19-2005R2010).
- NATIONAL LIBRARY OF MEDICINE, 2025a. *Overview of SNOMED CT* [online]. Estados Unidos [acesso em 2025-01-17]. URL: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html.
- NATIONAL LIBRARY OF MEDICINE, 2025b. *Welcome to Medical Subject Headings* [online]. Estados Unidos [acesso em 2025-01-16]. URL: <https://www.nlm.nih.gov/mesh/meshhome.html>.
- NGUYEN, Anthony; O'DWYER, John; VU, Thanh; WEBB, Penelope M; JOHNATTY, Sharon E; SPURDLE, Amanda B, 2020. Generating high-quality data abstractions from scanned clinical records: text-mining-assisted extraction of endometrial carcinoma pathology features as proof of principle. *BMJ Open*. Vol. 10, n.º 6, e037740. URL DOI: [10.1136/bmjopen-2020-037740](https://doi.org/10.1136/bmjopen-2020-037740).
- NÚCLEO DE SAÚDE DIGITAL GLOBAL E RELAÇÕES INTERNACIONAIS E UNIDADE DE CIBERSEGURANÇA, 2025. *Inteligência Artificial na Saúde em Portugal: Regulamentação, Impactos e Perspetivas de Futuro*.
- PAJO, Bora, 2023. *Introduction to Research Methods: A Hands-on Approach*. 2nd Revised edition. Thousand Oaks: SAGE Publications Inc.
- PARLAMENTO EUROPEU E CONSELHO DA UNIÃO EUROPEIA, 2025. *Regulamento (UE) 2025/327 do Parlamento Europeu e do Conselho, de 11 de fevereiro de 2025, relativo ao Espaço Europeu de Dados de Saúde e que altera a Diretiva 2011/24/UE e o Regulamento (UE) 2024/2847, (Texto relevante para efeitos do EEE)*. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32025R0327>. Jornal Oficial da União Europeia, L 45, p. 1–34.
- RAMDHANI, Tosan Wiar; BUDI, Indra; PURWANDARI, Betty, 2021. Optical Character Recognition Engines Performance Comparison in Information Extraction. *International Journal of Advanced Computer Science and Applications*. Vol. 12, n.º 8. URL DOI: [10.14569/IJACSA.2021.0120814](https://doi.org/10.14569/IJACSA.2021.0120814).
- READ-COOP, 2022. *Transkribus Lite 2.3. – New Features: Smart Search, Tag Manager, and Layout Relations* [online]. Transkribus blog [acesso em 2025-02-02]. URL: <https://blog.transkribus.org/en/transkribus-lite-2-3>.
- RIBEIRO, Fernanda, 2008. Medicina e Ciência da Informação: uma abordagem integradora e interdisciplinar. Em: *Colóquio Internacional - A Medicina na era da Informação*, pp. 111–125.
- SANTOS, Fernando Pinto; NOGUEIRA, Mafalda, 2023. *Trabalho final de mestrado : antes, durante e depois*. [Coimbra]: Almedina.

- SHORTLIFFE, Edward H.; BARNETT, G. Octo, 2014. Biomedical Data: Their Acquisition, Storage, and Use. Em: SHORTLIFFE, Edward H.; CIMINO, James J. (eds.). *Biomedical Informatics*. London: Springer London, pp. 39–66. URL DOI: [10.1007/978-1-4471-4474-8_2](https://doi.org/10.1007/978-1-4471-4474-8_2).
- SILVA, Armando Malheiro; RIBEIRO, Fernanda, 2000. A Avaliação Em Arquivística: Reformulação Teórico-Prática de Uma Operação Metodológica. *Páginas a&b: Arquivos E Bibliotecas*. N.º 5, pp. 57–113. ISSN 0873-5670.
- SILVA, Armando Malheiro; RIBEIRO, Fernanda; RAMOS, Julio; REAL, Manuel Luís, 1998. *Arquivística: teoria e prática de uma ciência da informação*. Porto: Edições Afrontamento. Biblioteca das ciências do homem, n.º 2. ISBN 978-972-36-0483-2.
- SMITH, Matthew A.; BARNES, E. Leon; CHIOSEA, Simion I., 2011. Pathology Archive. *American Journal of Clinical Pathology*. Vol. 135, n.º 5, pp. 753–759. URL DOI: [10.1309/AJCP3CVA2NAVUUVU](https://doi.org/10.1309/AJCP3CVA2NAVUUVU).
- SOUZA, Amanda Damasceno De, 2021. *O discurso na prática clínica e as terminologias de padronização: investigando a conexão*. Minas Gerais. Disponível em: <http://hdl.handle.net/1843/38044>. Tese de doutoramento. Universidade Federal de Minas Gerais.
- STRAUS, Sharon E.; GLASZIOU, Paul; RICHARDSON, W. Scott; HAYNES, R. Brian, 2019. *Evidence-Based Medicine: How to Practice and Teach EBM*. Fifth edition. Edinburgh ; London ; New York: Elsevier. ISBN 978-0-7020-6296-4.
- TRAPRINQ, 2025. *O projeto* [online]. [acedido em 2025-01-30]. URL: <https://traprinq.mozellosite.com/o-projeto/>.
- TRAQUEIA, Ana; PACHECO, Emilce; TAVEIRA, Eugénia, 2021. Reflexão crítica sobre métodos e técnicas de recolha de dados: investigação-ação. Em: *Reflexões em torno de Metodologias de Investigação: métodos*. Aveiro: UA Editora.
- UNITED NATIONS REGIONAL INFORMATION CENTRE, 2025. *Objetivos de Desenvolvimento Sustentável* [online]. [acedido em 2025-08-09]. URL: <https://unric.org/pt/objetivos-de-desenvolvimento-sustentavel/>.
- VARMA, Murali; COLLINS, Laura C; CHETTY, Runjan; KARAMCHANDANI, Dipti M; TALIA, Karen; DORMER, John; VYAS, Monika; CONN, Brendan; GUZMÁN-AROCHO, Yai-leen D; JONES, Adam V; PRING, Miranda; MCCLUGGAGE, W Glenn, 2024. Macroscopic Examination of Pathology Specimens: A Critical Reappraisal. *Journal of Clinical Pathology*. Vol. 77, n.º 3, pp. 164–168. URL DOI: [10.1136/jcp-2023-209045](https://doi.org/10.1136/jcp-2023-209045).
- WORLD HEALTH ORGANIZATION, 2025. *Importance of ICD* [online]. [acedido em 2025-03-26]. URL: <https://www.who.int/standards/classifications/frequently-asked-questions/importance-of-icd>.

Apêndice A

Guião de entrevista

—A entrevista a ser conduzida faz parte de uma investigação sobre o arquivo de anatomia patológica do Hospital Universitário São João, tendo como foco o enriquecimento de metainformação dos relatórios citológicos e histológicos. A investigação foi aceite pela Comissão de Ética do HUSJ, com número CES 459/24. **Esta fase tem como objetivo caracterizar as práticas de estruturação dos relatórios histológicos e citológicos, no presente e no passado, bem como compreender o comportamento de pesquisa dos clínicos do Serviço de Anatomia Patológica, com intuito de contribuir para um eficiente modelo de enriquecimento de metainformação descritiva do arquivo de anatomia patológica.** Esta entrevista será gravada por áudio, sem coleta de imagem, sempre com o consentimento prévio dos participantes. As gravações servirão apenas para o processo de recolha de dados para a investigação a realizar. Caso pretendam, os entrevistados poderão ter acesso à respetiva gravação efetuada e terão ainda acesso aos resultados da presente investigação, atingidos com a sua colaboração.

—[NOME DO PARTICIPANTE], concorda com a gravação da entrevista nestes termos?

—O termo de consentimento informado, livre e esclarecido para participação em investigação foi assinado pelo participante.

Título: Caracterização das práticas de estruturação dos relatórios de anatomia patológica e diagnóstico do comportamento de pesquisa dos clínicos e investigadores

Data:

Local:

Horário de início:

Entrevistadora:

Participante:

Perguntas de entrevista

Contexto organizacional

1. **Que função desempenha no Serviço de Anatomia Patológica?**
2. **Há quanto tempo está a trabalhar no HUSJ e, especificamente, neste cargo?**

3. **Em relação ao percurso profissional e académico, quais fatores salientaria como relevantes para a elaboração do relatórios de anatomia patológica, no que se refere a estrutura e terminologia?** *Por exemplo, experiência académica, profissional, experiências no estrangeiro ou pessoal*

Organização e acesso ao arquivo

4. **Costuma aceder a relatórios citológicos e histológicos anteriores dos doentes para prestação de cuidados ou investigação?** *O clínico apenas acessa os relatórios digitais ou também utiliza relatórios arquivados?*

- (a) *Caso sim, como o faz?*
- (b) *Caso sim, busca para ter mesmo acesso ao relatório final ou o interessa também as versões anotadas em arquivo?*
- (c) *Caso não, tem conhecimento que estes relatórios arquivados possuem notas relativas à análise? Acha que essas notas tenham relevância para a prestação de cuidados ou para investigação?*
- (d) *Em ambos os casos, qual a sua perceção em relação ao acesso dos relatórios anteriores à informatização? Por exemplo, constrangimentos para acesso aos documentos em formato de papel.*

5. **Qual a sua opinião em relação à eficácia do software que utiliza atualmente para gerar relatórios no Serviço de Anatomia Patológica?** *Do ponto de vista das funcionalidades de produção dos relatórios e para a recuperação da informação*

- (a) *O clínico realiza notas manuscritas no relatório ou na fase de macroscopia/microscopia como forma de complementar a utilização do software?*
- (b) *Estas notas são inseridas no sistema posteriormente ou só ficam acessíveis no arquivo?*

6. **O clínico já utilizou o Repositório Clínico Digital (RCD)?**

- (a) *Caso sim, quanto a estes relatórios mais antigos, anteriores à informatização, poderia apontar dificuldades e vantagens à disponibilização destes para prestação de cuidados e investigação científica no RCD? *Acompanha amostras anonimizadas.*
- (b) *!Caso não tenha sido claramente respondido! Quais as vantagens e desvantagens que percebe na organização e acesso dos relatórios, citológicos e histológicos, no modelo atual já informatizado (documentos digitais)?*
- (c) *Caso não, peço que imagine um cenário em que o clínico necessite aceder aos relatórios anteriores de um doente e estes estão em formato de papel. Quais as vantagens e desvantagens que o clínico identifica na disponibilização destes relatórios em formato digital ao invés de formato em papel? Ou seja, enxerga apenas coisas boas ao*

inserimos esses relatórios em um repositório ou identifica também dificuldades neste processo?

7. **No atual modelo, de que forma o processo de pesquisa poderia ser mais efetivo? Por exemplo, conseguir pesquisar pela morfologia e ter a opção de filtrar pela topografia.**

Comportamento de pesquisa

8. **Enquanto clínico ou investigador, quais as classificações ou vocabulários controlados da área médica utiliza com maior frequência para produção do relatório? Por exemplo, SNOMED CT, MeSH, DeCS, ICD ou outro? Ver cartões na figura A.1.**

- (a) *Utiliza apenas no âmbito do HUSJ ou em outras áreas, como de ensino, estudo, investigação ou prestação de cuidados?*
- (b) *No âmbito do HUSJ, utilizam alguma classificação que não o SNOMED para realizar os relatórios ou tomar notas?*
- (c) *Do ponto de vista da evolução na produção de relatórios da anatomia, como descreve a utilização das classificações na área da saúde? Por exemplo, entende que houve uma evolução na utilização no HUSJ, se tornou mais ou menos rigorosa?*
- (d) *Procede a pesquisa por termos da área da saúde, especificamente da anatomia patológica, de outra forma que não os vocabulários controlados? Por exemplo, ferramentas de Inteligência Artificial.*

9. **De que forma o clínico procede às pesquisas por termos descritores ou palavras de pesquisa da anatomia patológica no âmbito do HUSJ e/ou em outros locais de trabalho ou pesquisa?**

- (a) *Em relação ao processo de pesquisa dos termos, qual a percepção do clínico quanto aos pontos facilitadores e dificuldades para o acesso?*
- (b) *Caso a mediação da pesquisa seja realizada pelo Secretariado, como é para si o processo e qual a sua percepção sobre ele? Benefícios e dificuldades.*



Figura A.1: Cards dos vocabulários controlados.

Apêndice B

Termo de consentimento

CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM INVESTIGAÇÃO

De acordo com a Convenção de Oviedo¹

Por favor, leia com atenção a seguinte informação. Se achar que algo está incorreto ou não está claro, não hesite em solicitar mais informações. Se concorda com a proposta que lhe foi feita, queira assinar este documento.

Título da investigação: Tratamento e disponibilização de arquivos de anatomia patológica: modelo de enriquecimento de metainformação.

Enquadramento: A presente entrevista enquadra-se no âmbito da dissertação de mestrado da estudante Ana Carolina Ramalho, intitulado “Tratamento e disponibilização de arquivos de anatomia patológica: modelo de enriquecimento de metainformação”. O mestrado surge no contexto do curso Mestrado em Ciência da Informação, da Universidade do Porto, sob orientação do Professor Doutor Gabriel David, coorientação da Professora Dr^a. Fernanda Gonçalves e tutoria da Dr^a Joana Gomes.

Explicação da entrevista e conservação dos dados: Esta fase tem como objetivo caracterizar as práticas de estruturação dos relatórios histológicos e citológicos, no presente e no passado, bem como compreender o comportamento de pesquisa dos clínicos do Serviço de Anatomia Patológica, com intuito de contribuir para um eficiente modelo de enriquecimento de metainformação descritiva do arquivo de anatomia patológica.

As entrevistas terão por base um guião semi-estruturado, que possibilita guiar a entrevista, mas com flexibilidade para inserir informação complementar às perguntas pré definidas. A entrevista pretende perceber as necessidades informacionais e comportamento do utilizador, com base na rotina clínica no Serviço de Anatomia Patológica, para ser possível, através das suas respostas, retirar as devidas conclusões para esta investigação. A entrevista não possui um limite de tempo mínimo ou máximo para a sua realização. No entanto, estima-se que o mesmo não se estenderá para além de uma hora.

¹<https://www.ministeriopublico.pt/instrumento/convencao-para-proteccao-dos-direitos-do-homem-e-da-dignidade-do-ser-humano-face-22>

As reuniões serão gravadas por áudio, sem coleta de imagem, sempre com a informação e consentimento prévio dos participantes. As gravações servirão apenas para o processo de recolha de dados para a investigação em curso. Caso pretendam, os entrevistados poderão ter acesso a respetiva gravação efetuada e terão ainda acesso aos resultados da presente investigação, atingidos com a sua colaboração.

O prazo de conservação dos dados, ou seja, destas gravações de áudio e suas respetivas transcrições, tem o prazo máximo de 6 meses, com o propósito de análise qualitativa e quantitativa dos dados e validação da análise. Com o intuito único de envio da gravação e notificação acerca da destruição das gravações e destruição da respetiva transcrição, será coletado o email do participante. Em nenhum momento do processo de mestrado, ou para lá deste, será desvendada a identificação do participante, nem a gravação da sua tarefa no estudo. Os dados provenientes da entrevista serão devidamente anonimizados e, posteriormente, utilizados para publicação de eventuais artigos científicos e apresentação e publicação da dissertação de mestrado.

Condições e financiamento: O estudo não é financiado por qualquer instituição. A participação neste estudo é totalmente voluntária e não envolve qualquer tipo de compensação financeira ou benefício material. Desta forma, não existem quaisquer custos ou perdas associadas à sua participação. A sua decisão de participar ou não participar no estudo não terá impacto negativo na sua relação com a investigadora ou com qualquer entidade envolvida.

Confidencialidade e anonimato: O presente estudo garante a confidencialidade e o uso exclusivo dos dados recolhidos, nunca tornando pública a identificação dos participantes. Está assegurado que os contactos provenientes da entrevista serão realizados em ambiente de contexto académico.

Em caso de dúvidas relativas à investigação, acompanhamento do tratamento dos dados ou para retirar o consentimento, o participante poderá contactar o Elo de Ligação da ULS São João responsável pelo acompanhamento da investigação através do email: mfernanda.goncalves@ulssjoao.min-saude.pt

Caso tenha dúvidas sobre a forma como os seus dados pessoais são tratados, para saber mais sobre os seus direitos no âmbito da proteção de dados, pode entrar em contacto com o Encarregado de Proteção de Dados da ULS SÃO JOÃO através do endereço de email: epd@ulssjoao.min-saude.pt

Caso considere que os seus dados não estão a ser objeto de tratamento legítimo, pode, a todo o momento, apresentar uma reclamação junto da autoridade competente, a Comissão Nacional de Proteção de Dados (www.cnpd.pt).

Eu, Ana Carolina Ramalho Sampaio, estudante do Mestrado em Ciência da Informação, da Universidade do Porto, com o endereço eletrónico “up202301906@up.pt”, agradeço pela disponibilidade e prontidão para a participação neste estudo.

DATA:

ASSINATURA:

Declaro ter lido e compreendido este documento, bem como as informações que me foram fornecidas pela pessoa que acima assina. Foi-me garantida a possibilidade de, em qualquer altura, recusar participar neste estudo sem qualquer tipo de consequências. Desta forma, aceito participar neste estudo e permito a utilização dos dados que de forma voluntária forneço, confiando em que apenas serão utilizados para esta investigação e nas garantias de confidencialidade e anonimato que me são dadas pela investigadora.

NOME:

EMAIL:

DATA:

ASSINATURA:

DOCUMENTO É COMPOSTO POR 2 PÁGINAS E É FEITO EM DUPLICADO: UMA VIA PARA A INVESTIGADORA, OUTRA PARA A PESSOA QUE CONSENTE.

Apêndice C

Ferramentas para análise das entrevistas no NVIVO

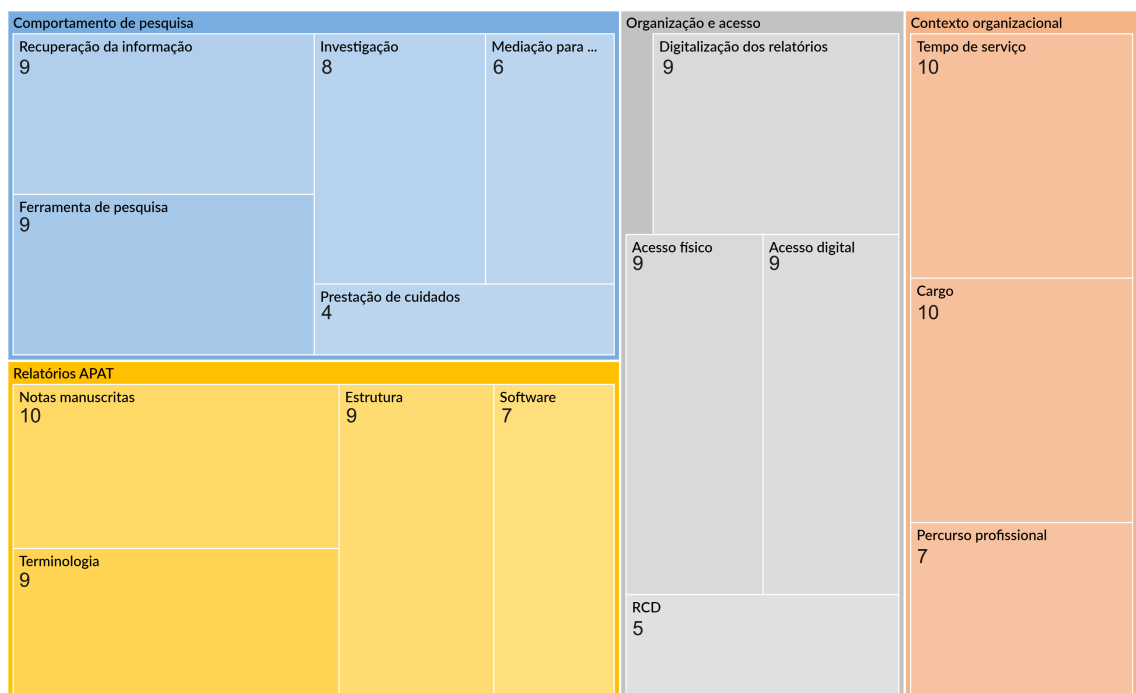


Figura C.1: Áreas por número de entrevistas (itens) codificadas.

A figura C.1 representa por tamanho da área, o número de itens codificados, e por cores, sendo as mais escuras com mais referências codificadas. Os campos maiores, nomeadamente comportamento de pesquisa, organização e acesso, contexto organizacional e relatórios da anatomia patológica, ficaram com maior destaque pelo número de itens e referências agregados, ou seja, que foram codificados nos pelo código-filho. Os códigos com maior número de referências, com base na tabela C.1, são terminologia (44), ferramenta de pesquisa (29), notas manuscritas (29) e estrutura (20), o que demonstra os principais temas desenvolvidos nas entrevistas.

Tabela C.1: Codebook NVIVO.

Nome	Descrição	Arquivos	Referências
Comportamento de pesquisa	Informações relacionadas a forma de pesquisa por termos, vocabulários controlados, recuperação da informação, mediação por secretariado e etc	9	29
Ferramenta de pesquisa	Especificamente por termos descritores	9	29
Investigação	Pesquisas com objetivo de investigação	8	12
Mediação para pesquisa	Perceção sobre a mediação para levantamento de termos e relatórios por meio do secretariado	6	8
Prestação de cuidados	Pesquisas com objetivo de prestação de cuidados	4	7
Recuperação da informação	Formas de pesquisa atuais e expectativas para uma recuperação efetiva	9	18
Contexto organizacional			
Cargo	Função desempenhada no Serviço de Anatomia Patológica	10	10
Percurso profissional	Área de trabalho e outros fatores relevantes para a estruturação dos relatórios e terminologia utilizada	7	8
Tempo de serviço	Ano de início no cargo indicado	10	10
Organização e acesso	Informações relacionadas à organização e acesso ao arquivo/relatórios de APAT antigamente e atualmente	1	2
Acesso digital	Perspectivas sobre o acesso aos relatórios em formato digital	9	13
Acesso físico	Facilitadores e dificuldades em acessar os relatórios em formato de papel	9	18
Digitalização dos relatórios	Perspectivas quanto à digitalização e disponibilização dos relatórios antigos no repositório	9	18
RCD	Conhecimento ou não do Repositório Clínico Digital	5	6
Relatórios APAT	Informações relacionadas à estruturação e elaboração dos relatórios de anatomia patológica	9	20
Estrutura	Informações relacionadas à forma como era estruturado o relatório antigamente e como é hoje em dia	9	20
Notas manuscritas	Falas relativas à realização ou não de notas manuscritas	10	29
Software	Sistema para criação e pesquisa de relatórios	7	13
Terminologia	Terminologia utilizada nos relatórios, como fazem a seleção de termos, se utilizam linguagem natural ou controlada	9	44

Apêndice D

Modelo de metadados atual do RCD

Tabela D.1: Esquema de metadados do RCD (56 elementos).

Nº	Nome Elemento	Exemplo de Valor	Caraterização	Descrição	Tipos de processo	Preenchimento	Obs
1	NumSequencial	969190	máx. 8 dígitos	Nº sequencial atribuído no sistema SONHO.	INT	Obrigatório	Dado enviado pelo CHSJ
2	NumDoente	19012345	7-8 dígitos	Nº interno, irrepetível, equivale ao processo do doente.	INT	Obrigatório	Dado enviado pelo CHSJ
3	NumAntigoDoente	74046541	até 8 alfanuméricos	Nº do processo antes da informatização (1992).	INT	Opcional	Dado enviado pelo CHSJ
4	NomeDoente	João Silva	até 150 car.	Nome completo do doente.	INT	Obrigatório	Leitura CB + validação com NumEpisodio
5	NumSNS	211012156	9 dígitos	Nº irrepetível no SNS.	INT	Opcional	Dado enviado pelo CHSJ
6	DataNascimento	05/05/2000	DD/MM/AAAA	Data de nascimento.	INT	Obrigatório	Dado enviado pelo CHSJ
7	Sexo	F	M/F/H	Sexo ou género.	INT	Obrigatório	—
8	TipoSubprocesso	INT	INT/AMB	Tipo de subprocesso.	INT	Obrigatório	—
9	NumEpisodio	19000150	7-8 dígitos	Nº de episódio hospitalar.	INT	Obrigatório	Associa dados descritivos
10	EspecialidadeAdmissao	Medicina Intensiva	Tabela controlada	Especialidade de admissão.	INT	Obrigatório	—
11	EspecialidadeAlta	Pediatria	Tabela controlada	Especialidade de alta.	INT	Obrigatório	—
12	DataAdmissao	19/06/2015	DD/MM/AAAA	Data de admissão.	INT	Obrigatório	—
13	DataAlta	27/07/2015	DD/MM/AAAA	Data de alta.	INT	Obrigatório	—
14	MedicoAdmissao	António Lopes	até 100 car.	Médico que registou admissão.	INT	Obrigatório	—
15	NumOMadmissao	58584	até 5 dígitos	Nº da Ordem do médico admissão.	INT	Obrigatório	—

Nº	Nome Elemento	Exemplo de Valor	Caraterização	Descrição	Tipos de processo	Preenchimento	Obs
16	MedicoAlta	António Lopes	até 100 car.	Médico que registou alta.	INT	Obrigatório	—
17	NumOMAlta	58584	até 5 dígitos	Nº da Ordem do médico alta.	INT	Obrigatório	—
18	Seccoes	ALT	Lista controlada	Classificação por secções.	INT	Obrigatório	ALT, CIR, MED...
19	TipologiaDocumental	Eletrocardiograma	Tabela controlada	Classificação tipológica.	INT	Opcional	Não usado retros- petivo
20	DataDocumento	13/04/2016	DD/MM/AAAA	Data de produção.	INT	Opcional	—
21	HoraDocumento	17:30	HH:MM	Hora de produção.	INT	Opcional	—
22	NumVolume	1	1-2 dígitos	Nº de volume.	INT	Obrigatório	—
23	NumFolha	25	até 3 dígitos	Nº de folha.	INT	Obrigatório	—
24	TipoSuporte	Papel	Lista controlada	Tipo de suporte original.	INT	Obrigatório	Papel, digital, mi- crofilme
25	EstadoConservacao	Bom	Lista controlada	Estado físico.	INT	Opcional	Bom, Regular, Mau
26	Assinatura	Sim	Booleano	Indica assinatura manuscrita.	INT	Opcional	—
27	Carimbo	Sim	Booleano	Indica carimbo.	INT	Opcional	—
28	Observacoes	Texto livre	até 255 car.	Notas adicionais.	INT	Opcional	—
29	NomeMae	Maria Silva	até 150 car.	Nome da mãe.	INT	Opcional	—
30	NomePai	José Silva	até 150 car.	Nome do pai.	INT	Opcional	—
31	EstadoCivil	Casado	Lista controlada	Estado civil.	INT	Opcional	Solteiro, Casado...
32	Profissao	Professor	até 100 car.	Profissão.	INT	Opcional	—
33	Naturalidade	Porto	até 100 car.	Local de nascimento.	INT	Opcional	—
34	Nacionalidade	Portuguesa	até 100 car.	Nacionalidade.	INT	Opcional	—
35	Morada	Rua X, nº Y	até 255 car.	Endereço completo.	INT	Opcional	—

Nº	Nome Elemento	Exemplo de Valor	Caraterização	Descrição	Tipos de processo	Preenchimento	Obs
36	CodPostal	4100-000	NNNN-NNN	Código postal.	INT	Opcional	—
37	Localidade	Porto	até 100 car.	Localidade.	INT	Opcional	—
38	Distrito	Porto	até 100 car.	Distrito.	INT	Opcional	—
39	Telefone	220000000	até 15 car.	Telefone.	INT	Opcional	—
40	Email	exemplo@pt	até 150 car.	Email.	INT	Opcional	—
41	NomeResponsavel	Ana Silva	até 150 car.	Nome do responsável.	INT	Opcional	—
42	GrauParentesco	Mãe	até 50 car.	Parentesco do responsável.	INT	Opcional	—
43	ContactoResp.	220000111	até 15 car.	Telefone do responsável.	INT	Opcional	—
44	DataCriacaoProc.	01/01/1990	DD/MM/AAAA	Data de criação.	INT	Obrigatório	—
45	DataEncerram.Proc.	31/12/2020	DD/MM/AAAA	Data de encerramento.	INT	Opcional	—
46	MotivoEncerram.	Óbito	Lista controlada	Motivo do encerramento.	INT	Opcional	Óbito, Transferência...
47	DataDigitalizacao	15/03/2025	DD/MM/AAAA	Data de digitalização.	INT	Obrigatório	—
48	OperadorDigitaliz.	Carlos Sousa	até 100 car.	Nome do operador.	INT	Obrigatório	—
49	SoftwareOCR	ABBY	até 100 car.	Software OCR.	INT	Opcional	—
50	QualidadeImagem	300dpi	até 50 car.	Resolução.	INT	Obrigatório	—
51	FormatoImagem	TIFF	até 10 car.	Formato de ficheiro.	INT	Obrigatório	TIFF, JPEG, PDF
52	TamanhoFicheiro	2MB	até 10 car.	Tamanho digital.	INT	Obrigatório	—
53	HashFicheiro	abc123	até 64 car.	Hash de integridade.	INT	Obrigatório	—
54	Backup	Sim	Booleano	Indica cópia de segurança.	INT	Obrigatório	—
55	LocalArmazenam.	Servidor A	até 100 car.	Local de arquivo.	INT	Obrigatório	—
56	PoliticaRetencao	20 anos	até 50 car.	Tempo de retenção.	INT	Obrigatório	Segundo legislação

Apêndice E

Formulário de registo de lotes - SPD

Centro de Gestão da Informação | **Serviço de Arquivo** N.º



Registo de lotes de documentos ativos

Registo

1. Registo de entrada

N.º Caixa [Cota]		Primeiro registo [N.º SPD]	Último registo [N.º SPD]	Prioridade* *Seguir IT - envio doc depósitos			
Quantidade de documentos				1	2	3	4

Observações do operador de registo de entrada
[Exemplo: registo nº XXX anulado.]

Ass. Func.: _____ N.º Mec. |_____|_____|_____|_____|_____|_____|_____|_____| Data ____/____/____ Hora: ____:____

Digitalização

2. Digitalização dos documentos ativos

Confirma a informação recebida? Sim ☐ Não ☐

Caso seja "Não" voltar ao menu anterior "1.Registo".

Observações do operador de digitalização

Ass. Func.: _____ N.º Mec. |_____|_____|_____|_____|_____|_____|_____|_____| Data ____/____/____ Hora: ____:____

Controlo de qualidade

3. Controlo de qualidade

Confirma a informação recebida? Sim ☐ Não ☐

Caso seja "Não" voltar ao menu anterior "2.Digitalização".

Observações do operador do controlo de qualidade

Ass. Func.: _____ N.º Mec. |_____|_____|_____|_____|_____|_____|_____|_____| Data ____/____/____ Hora: ____:____

Ingestão

4. Verificação da Ingestão no RCD

Confirma que a informação foi ingerida no RCD? Sim ☐ Não ☐

Caso seja "Não" voltar ao menu anterior "3.Controlo de qualidade".

Observações do operador de verificação no RCD

Ass. Func.: _____ N.º Mec. |_____|_____|_____|_____|_____|_____|_____|_____| Data ____/____/____ Hora: ____:____

Elaborado por:

Verificado por:

Aprovado por:

Página 1 de 1

CENTRO HOSPITALAR DE SÃO JOÃO, E.P.E. (Sede) Alameda Professor Hernâni Monteiro 4200-319 Porto T+351 225512100 F+351 225025766 E geral@chj.min-saude.pt W http://portal-chj.min-saude.pt

N.º 1 ARQ-H001-0

Figura E.1: Formulário de registo de lotes de documentos ativos.