

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

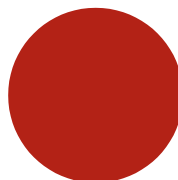
Gestão e Partilha de Dados Marinhos e Ambientais. Estudo de Caso Aplicado ao CIIMAR Watch.

Adriana Luís M. Garcia

M
2023

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

FACULDADE DE LETRAS DA UNIVERSIDADE DO PORTO

Gestão e Partilha de Dados Marinhos e Ambientais. Estudo de Caso aplicado ao CIIMAR Watch

Adriana Luís Martins Garcia



Mestrado em Ciência da Informação

Orientador: Prof. Doutor João Aguiar Castro

Segundo orientador: Doutora Maria Paola Tomasino

24 de outubro de 2025

Gestão e Partilha de Dados Marinhos e Ambientais. Estudo de Caso aplicado ao CIIMAR Watch

Adriana Luís Martins Garcia

Mestrado em Ciência da Informação

Aprovado em provas públicas pelo Júri:

Presidente: Prof. Doutora Mariana Curado Malta

Arguente: Prof. Doutora Ana Alice Baptista

Vogal: Prof. Doutor João Daniel Aguiar de Castro

24 de outubro de 2025

Resumo

A gestão de dados de investigação (GDI) tem vindo a afirmar-se como um elemento central na investigação científica, permitindo assegurar a integridade, acessibilidade e reutilização dos dados ao longo do seu ciclo de vida. Alinhada com os princípios da ciência aberta e com normas internacionais, a GDI contribui para reforçar a transparência, a colaboração e a inovação científica. No domínio marinho, onde se produzem dados de grande volume, diversidade e complexidade, a gestão eficaz assume particular relevância. A heterogeneidade de metodologias e formatos, associada à necessidade de integrar dimensões físicas, químicas e biológicas, torna a GDI indispensável para garantir a interoperabilidade e a utilização sustentável dos dados no apoio à investigação e à tomada de decisão ambiental.

Esta dissertação analisa a gestão e partilha de dados no contexto do programa CIIMAR Watch, uma iniciativa multidisciplinar dedicada à monitorização de parâmetros biológicos, químicos e físicos na costa norte de Portugal. O estudo centra-se na identificação de práticas, desafios e necessidades associadas à gestão de dados de investigação, procurando compreender como estas se articulam num contexto de crescente produção e diversidade de informação.

Com base numa abordagem qualitativa e exploratória, foram conduzidas entrevistas semiestruturadas a investigadores de diferentes áreas científicas integrados no programa. A análise permitiu identificar padrões, desafios e necessidades específicas ao longo do ciclo de vida dos dados, abrangendo aspetos como o registo, armazenamento, documentação, metadados, partilha e preservação.

Os resultados evidenciam a coexistência de práticas diversas, condicionadas pelas especificidades disciplinares e pela ausência de normas transversais, mas também um reconhecimento generalizado da importância de melhorar a gestão de dados. A partir das perspetivas dos investigadores, foram delineados requisitos e recomendações alinhados com os princípios FAIR e com normas internacionais relevantes, assim como um modelo de metadados, que poderão servir de base para a implementação futura de infraestruturas de apoio e serviços de *data stewardship* no CIIMAR.

O contributo desta dissertação reside, assim, na caracterização das práticas e necessidades de gestão de dados num contexto multidisciplinar de monitorização marinha, apontando caminhos para o fortalecimento da ciência aberta e para a sustentabilidade dos dados de investigação a longo prazo.

Desta forma, este trabalho abre perspetivas para desenvolvimentos futuros, entre os quais se destacam: adoção de cadernos de laboratório eletrónicos que facilitem o registo e rastreabilidade dos dados; a criação de um manual de boas práticas e protocolos adaptados ao contexto multidisciplinar do CIIMAR; a promoção de ações de formação dirigidas aos investigadores; e a experimentação da proposta de modelo de metadados delineada nesta dissertação. Estas iniciativas poderão contribuir para consolidar uma cultura de gestão de dados mais sistemática, colaborativa e alinhada com os princípios da ciência aberta.

Abstract

Research data management (RDM) has become a central element in scientific research, ensuring the integrity, accessibility and reuse of data throughout its life cycle. Aligned with the principles of open science and international standards, RDM contributes to strengthening transparency, collaboration and scientific innovation. In the domain of oceans, where data is produced in large volumes and is highly diverse and complex, effective management is particularly important. The heterogeneity of methodologies and formats, combined with the need to integrate physical, chemical and biological dimensions, makes RDM indispensable for ensuring the interoperability and sustainable use of data in support of research and environmental decision-making.

This dissertation analyses data management and sharing in the context of the CIIMAR Watch programme, a multidisciplinary initiative dedicated to monitoring biological, chemical and physical parameters on the northern coast of Portugal. This investigation focuses on identifying practices, challenges and needs associated with research data management, aiming to understand how these relate to each other in a context of increasing production and diversity of information.

Based on a qualitative and exploratory approach, semi-structured interviews were conducted with researchers from different scientific domains involved in the programme. The analysis identified patterns, challenges, and specific needs throughout the data lifecycle, covering aspects such as registration, storage, documentation, metadata, sharing, and preservation.

The results show the coexistence of diverse practices, conditioned by specific disciplinary characteristics and the absence of cross-cutting standards, but also a widespread recognition of the importance of improving data management. Based on the researchers' perspectives, requirements and recommendations were outlined in line with FAIR principles and relevant international standards, as well as a metadata model, which could serve as a basis for the future implementation of support infrastructures and data stewardship services at CIIMAR.

The contribution of this dissertation therefore lies in characterising data management practices and needs in a multidisciplinary marine monitoring context, highlighting ways to strengthen open science and ensure the long-term sustainability of research data.

Hence, this work opens up perspectives for future developments, among which the following stand out: adoption of electronic laboratory notebooks that facilitate data recording and traceability; creation of a manual of good practices and protocols adapted to the multidisciplinary context of CIIMAR; promotion of training actions aimed at researchers; and experimentation with the proposed metadata model outlined in this dissertation. These initiatives could contribute to consolidating a more systematic and collaborative data management culture, aligned with the principles of open science.

Objetivos de Desenvolvimento Sustentável (ODS)

A dissertação contribui para os Objetivos de Desenvolvimento Sustentável (ODS) ao reforçar a gestão e a partilha de dados marinhos e ambientais, promovendo práticas alinhadas com a ciência aberta e os princípios FAIR. Os contributos centram-se sobretudo nos ODS 14, ODS 13 e ODS 17, resumidos na seguinte tabela:

ODS	Meta	Contributo da dissertação	Indicadores de desempenho e métricas
13	13.1	Melhoria da qualidade e da interoperabilidade de dados ambientais (ex. temperatura, salinidade, pH) através da normalização das práticas de registo e metadados.	Conjuntos de dados documentados com o perfil mínimo; reutilização em estudos sobre alterações climáticas.
	13.3	Identificação de necessidades de formação e sensibilização para a gestão de dados e ciência aberta, promovendo a adoção progressiva de práticas FAIR.	Inclusão de recomendações de capacitação e boas práticas em futuras campanhas de monitorização.
14	14.2	Proposta de um perfil mínimo de metadados que assegura a padronização e a acessibilidade de dados marinhos e ambientais, potenciando a sua reutilização em conservação e gestão costeira.	Integração do perfil mínimo nas recomendações internas; potencial depósito dos conjuntos de dados em repositórios marinhos (ex. OBIS, EMODnet).
17	17.6	Promoção da interoperabilidade e da ciência aberta através da proposta de práticas de documentação e metadados alinhadas com normas internacionais (ex. ISO 19115, Darwin Core).	Grau de compatibilidade do perfil proposto com padrões internacionais; potencial integração em infraestruturas marinhas (ex. EMODnet).

Agradecimentos

Em primeiro lugar, agradeço profundamente ao meu orientador, Professor Doutor João Aguiar Castro, e à minha coorientadora, Doutora Maria Paola Tomasino. A ambos, obrigada pela paciência, pela orientação atenta e pela confiança em mim, mesmo nos momentos em que me sentia perdida. Ao Professor João, pela disponibilidade e incentivo constante, e à Maria, por me ter sempre acolhido com tanta atenção e generosidade.

Aos meus professores, pela dedicação e pelo conhecimento partilhado que tanto contribuíram para a minha formação.

Aos investigadores que aceitaram colaborar neste trabalho, pela disponibilidade e contributo essencial para os resultados alcançados.

Aos meus amigos do mestrado, verdadeiros companheiros desta jornada e que levo além dela, por todas as partilhas e momentos felizes, que tornaram o caminho mais leve.

Aos meus amigos de sempre que, espalhados pelo mundo, se fizeram presentes e acreditaram em mim mesmo quando eu duvidava. Tenho sorte de vos ter na minha vida.

Ao Pedro e à sua memória, pela amizade, pela presença que continua a acompanhar-me, e por ter feito parte do tempo em que o Porto se tornou casa.

Aos meus pais, pelo que fazem por mim.

E de forma muito especial, à minha irmã, Ana, e ao meu cunhado, Renan, por me abrirem a porta de casa, mas sobretudo por me darem um lar, no verdadeiro sentido da palavra. A vossa amizade, apoio e alegria foram um pilar essencial para que eu pudesse terminar esta dissertação com serenidade.

Adriana Luís Garcia

“A força não está em ser-se firme e forte mas em ser-se flexível. A árvore, que é flexível, aguenta-se no meio da ventania.”

J. Krishnamurti

Sumário

1	Introdução	1
1.1	Contexto	2
1.2	Motivação e Objetivos	3
1.3	Abordagem Metodológica	4
1.4	Estrutura da Dissertação	5
2	Gestão de Dados de Investigação	6
2.1	Dados de Investigação	7
2.2	Princípios FAIR e Ciência Aberta	8
2.3	Políticas de dados e iniciativas como a <i>Research Data Alliance</i>	10
2.4	Curadoria e ciclo de vida dos dados	11
2.5	Metadados e documentação	12
2.6	Os investigadores e a partilha de dados	13
3	Gestão de Dados de Investigação Marinhos e Ambientais	16
3.1	Especificidade dos Dados Marinhos e Ambientais	16
3.2	Iniciativas e Serviços Relevantes/Infraestruturas e Projetos Relevantes	17
3.3	Vocabulários Controlados e Normas e Ontologias	18
3.4	Repositórios de Dados	20
3.5	Tendências na Investigação sobre Gestão de Dados Marinhos	22
3.5.1	Revisão das Principais Contribuições Científicas	23
3.5.2	Resultados Bibliométricos	25
4	Enquadramento do CIIMAR Watch	28
4.1	O CIIMAR e o programa CIIMAR Watch	28
4.2	Estrutura, objetivos e funcionamento do programa	29
4.3	Multidisciplinaridade e diversidade de perfis dos investigadores	30
5	Metodologia e Desenho da Investigação	31
5.1	Abordagem Geral	31
5.2	Entrevistas Semiestruturadas: Instrumento e Processo	31
5.3	Amostra e Perfil dos Participantes	32
5.3.1	Investigadores por Área Científica	32
5.3.2	Investigadores por Ecossistema	33
5.4	Técnica de Análise dos Dados	33

6	Práticas e Desafios na Gestão e Partilha de Dados no CIIMAR Watch	37
6.1	Resultados	37
6.1.1	Tipologias de dados gerados	37
6.1.2	Anotação e Registo de Dados	38
6.1.3	Armazenamento e Organização dos Dados	41
6.1.4	Documentação e Metadados	44
6.1.5	Reutilização e Partilha de Dados	45
6.1.6	Práticas, Desafios e Requisitos	48
6.2	Discussão dos Resultados	49
7	Recomendações para a Gestão de Dados no CIIMAR Watch	53
7.1	Recomendações Gerais	53
7.2	Interoperabilidade, Perfis de Aplicação e Ontologias	55
7.3	Proposta de um Conjunto Mínimo de Metadados para o CIIMAR Watch	56
8	Conclusão e Trabalho Futuro	59
	Referências	62
A	Guião das Entrevistas	69
B	Consentimento Informado	71
C	Modelo XML do Minimal Metadata Set para o CIIMAR Watch	73

Lista de Figuras

2.1	Modelo de Curadoria do Ciclo de Vida dos Dados do DCC	12
3.1	Principais temas na literatura (SCOPUS, 2025)	25
3.2	Produção científica anual (SCOPUS, 2025)	26
3.3	Distribuição geográfica da produção científica (SCOPUS, 2025)	27
3.4	Autores mais relevantes na produção científica sobre gestão de dados marinhos (SCOPUS, 2025)	27
4.1	Estações de Amostragem do CIIMAR Watch	29
5.1	Exemplo de anotação no CATMA	36

Lista de Tabelas

3.1	Exemplos de repositórios de dados marinhos relevantes e respetivos padrões de metadados	21
5.1	Resumo dos participantes por área, função e ecossistema	33
5.2	Grelha temática de categorias utilizadas na análise	35
6.1	Síntese das tipologias, métodos, formatos e práticas de registo dos dados no CIIMAR Watch	40
6.2	Síntese das práticas de armazenamento e organização dos dados no CIIMAR Watch	43
6.3	Síntese dos principais resultados e respetiva interpretação crítica.	52
7.1	Conjunto mínimo de metadados proposto para o CIIMAR Watch.	58

Lista de Acrónimos

BODC	British Oceanographic Data Centre
CATMA	Computer Assisted Text Markup and Analysis
CDI	Common Data Index
CF	Climate and Forecast Standard Names
CIIMAR	Centro Interdisciplinar de Investigação Marinha e Ambiental
CMEMS	Copernicus Marine Environment Monitoring Service
CSW	Catalogue Service for the Web
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
DwC	Darwin Core
ELN	Electronic Lab Notebook
EML	Ecological Metadata Language
EMODnet	European Marine Observation and Data Network
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable and Reusable
FASTA	FASTA format
GBIF	Global Biodiversity Information Facility
GDI	Gestão de Dados de Investigação
GEO BON	Group on Earth Observations Biodiversity Observation Network
INSPIRE	Infrastructure for Spatial Information in Europe
IOC	Intergovernmental Oceanographic Commission
ISO	International Organization for Standardization
LSID	Life Science Identifier
MBON	Marine Biodiversity Observation Network
NetCDF	Network Common Data Form
NERC	Natural Environment Research Council
NVS	NERC Vocabulary Server
OBIS	Ocean Biodiversity Information System
OBPS	Ocean Best Practices System
ODV	Ocean Data View
OGC	Open Geospatial Consortium
ONC	Ocean Networks Canada
QARTOD	Quality Assurance of Real-Time Oceanographic Data
RDA	Research Data Alliance
re3data	Registry of Research Data Repositories
SeaDataNet	Pan-European Sea Data Network
SDC	SeaDataCloud
UNESCO	United Nations Educational, Scientific and Cultural Organization
VLIZ	Flanders Marine Institute
WFS	Web Feature Service
WMS	Web Map Service
WoRMS	World Register of Marine Species

Capítulo 1

Introdução

Na investigação científica, e em particular nas ciências naturais e ambientais, onde os dados são complexos e produzidos em larga escala, é essencial torná-los acessíveis e compreensíveis, permitindo a sua reutilização. A gestão de dados de investigação (GDI) é o processo que abrange todo o ciclo de vida dos dados, desde o planeamento à reutilização e preservação, tornando-se um dos pilares da ciência aberta e colaborativa, cada vez mais incentivada por financiadores e entidades governamentais. Alinhada com normas e políticas globais de boas práticas, como os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), a GDI facilita o acesso responsável e a preservação de dados, assegurando a sua reutilização e continuidade para investigações futuras. Embora amplamente debatida por profissionais da informação e investigadores interessados na prática de gestão de dados, esta deve ser uma prioridade transversal a todos os produtores e utilizadores de dados, com vista à adoção de práticas adequadas de gestão (Connie, Clare et al., 2019).

Em áreas como a biologia marinha, este cenário torna-se particularmente relevante, onde os sistemas de observação, em constante evolução, produzem volumes crescentes de dados diversificados (Tanhua et al., 2019), exigindo práticas consistentes de gestão para garantir a integridade e a qualidade da investigação, para além de diminuir o impacto ambiental associado à investigação científica no oceano. A monitorização marinha e ambiental gera conjuntos de dados complexos e em atualização contínua, sendo essencial uma gestão eficaz não apenas para garantir a reprodutibilidade dos estudos, mas também para potenciar a partilha e integração dos dados e apoiar investigações sustentáveis a longo prazo. Neste contexto, a adoção de boas práticas de gestão de dados é fundamental para assegurar o acesso, a interoperabilidade e, em última instância, a reutilização responsável dos dados científicos, promovendo a colaboração entre projetos e investigadores.

O Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), da Universidade do Porto, é um instituto de referência na investigação científica relacionada com os recursos marinhos e a sustentabilidade ambiental. Este centro desenvolve projetos multidisciplinares que produzem dados fundamentais para compreender e preservar os ecossistemas marinhos, contribuindo para o avanço da ciência e para a tomada de decisões informadas em diversas áreas, como a ecologia, a biologia marinha e a oceanografia.

Entre as suas iniciativas destaca-se o programa CIIMAR Watch, um programa dedicado à monitorização de dados biológicos, químicos e físicos da costa norte de Portugal, que adota uma abordagem integrada ao envolver investigadores de várias disciplinas científicas para reunir e consolidar dados multidisciplinares. A diversidade dos dados gerados e das áreas científicas envolvidas exige um método unificado e eficaz de gestão e partilha gerados pelos investigadores no contexto do programa.

Neste contexto, a identificação das necessidades e requisitos dos investigadores envolvidos no programa revela-se essencial para assegurar que as práticas de gestão e partilha de dados a adotar estejam adequadamente alinhadas com as especificidades do CIIMAR Watch e com as exigências metodológicas das diferentes áreas representadas. Esta análise poderá, simultaneamente, fornecer uma base informada para o desenvolvimento de orientações mais abrangente, potenciando a sua aplicabilidade a outros grupos de investigação no CIIMAR.

1.1 Contexto

A gestão e partilha dos dados de investigação é cada vez mais importante para a promoção de uma ciência aberta e transparente, assim como para a colaboração científica. Em áreas como a monitorização marinha, onde a integração de diversas disciplinas é não apenas uma prática habitual, mas uma necessidade essencial para lidar com a complexidade dos ecossistemas, a gestão de dados de investigação torna-se um aspeto estruturante para garantir a qualidade, integridade e reutilização dos dados gerados.

O CIIMAR Watch, desenvolvido pelo Centro Interdisciplinar de Investigação Marinha e Ambiental, dedica-se à monitorização de dados multidisciplinares – biológicos, químicos e físicos – da costa norte de Portugal, com estações localizadas no Porto e em Viana do Castelo. Os dados gerados resultam dos contributos de investigadores de diversas áreas marinhas, incluindo oceanografia, climatologia, química ambiental, sensorização remota, toxicologia ambiental, ecologia microbiana, ecologia bentónica e pelágica, ecologia de mamíferos marinhos e biodiversidade, promovendo uma estratégia colaborativa e holística da monitorização marinha e ambiental.

Os principais objetivos deste programa consistem em disponibilizar em acesso aberto os dados recolhidos, promovendo boas práticas na sua gestão e partilha, com ênfase na acessibilidade, reutilização e preservação a longo prazo. Embora o CIIMAR tenha recentemente criado um Gabinete de Ciência Aberta, trata-se de um serviço ainda em fase inicial de implementação. Neste contexto, é atribuída especial atenção ao desenvolvimento de um sistema sólido de *data stewardship* que apoie os investigadores na gestão eficaz dos seus dados e na adoção dos princípios FAIR. Assim, este programa serve de base para explorar estratégias e práticas que poderão ser adotadas de forma mais ampla.

Além disso, o programa visa alinhar as suas práticas com as de outras iniciativas europeias, promovendo uma abordagem mais uniforme e colaborativa na monitorização ambiental. Esta harmonização não se limita a aspetos processuais, mas procura sobretudo potenciar a partilha, o uso

e reutilização de dados provenientes de diversas fontes, facilitando a colaboração e contribuindo para o avanço da ciência aberta à escala global.

1.2 Motivação e Objetivos

A integração de diferentes áreas do conhecimento é essencial na monitorização marinha e ambiental, dada a complexidade dos ecossistemas em estudo. Seguindo o CIIMAR Watch uma abordagem multidisciplinar, o programa gera uma grande quantidade de dados, tornando urgente a adoção de práticas de gestão que favoreçam a acessibilidade e a reutilização dos mesmos. Apesar do enorme potencial para impulsionar a ciência marinha, a gestão eficaz desses dados enfrenta desafios que devem ser abordados para otimizar os processos de registo, organização e partilha. A diversidade de práticas adotadas pelos investigadores, a ausência de uma abordagem uniforme e a falta de utilização sistemática de metadados e repositórios adequados comprometem a eficiência da gestão e partilha de dados. Além disso, a aplicação dos princípios FAIR no contexto do CIIMAR Watch ainda se encontra numa fase inicial, o que representa ao mesmo tempo um desafio e uma oportunidade para reforçar a acessibilidade e a reutilização dos dados a longo prazo.

Questões de investigação

Face ao exposto, este estudo tem como objetivo analisar as práticas de gestão de dados dos investigadores do CIIMAR Watch, identificar os principais desafios e requisitos necessários à melhoria desses processos e esboçar recomendações alinhadas com os princípios FAIR e com as necessidades do programa.

1. Quais são as práticas de gestão de dados atualmente adotadas pelos investigadores do programa CIIMAR Watch?
2. Quais são os principais desafios, necessidades e requisitos identificados pelos investigadores para melhorar as práticas de gestão e partilha de dados?

Objetivo geral

Analisar as práticas de gestão e partilha de dados de investigação no contexto do programa CIIMAR Watch, identificando desafios, necessidades e requisitos para a adoção de práticas alinhadas com os princípios FAIR e adaptadas à realidade multidisciplinar do programa.

Objetivos específicos

1. Caracterizar as práticas de gestão de dados adotadas pelos investigadores do CIIMAR Watch, abrangendo as diferentes etapas do ciclo de vida dos dados (registo, armazenamento, documentação, metadados, partilha e reutilização).
2. Identificar os principais desafios, necessidades e requisitos associados à gestão e partilha de dados, considerando as perspetivas dos investigadores e as orientações presentes na literatura.

3. Definir requisitos técnicos e organizacionais para a criação de metadados e infraestruturas adequadas à realidade do programa, assegurando a interoperabilidade e a acessibilidade dos dados a longo prazo.
4. Comparar as práticas observadas com boas práticas internacionais de gestão de dados em ciências marinhas e ambientais, avaliando o seu potencial de adaptação ao contexto do CIIMAR Watch.
5. Propor recomendações para a melhoria das práticas de gestão de dados e para o desenvolvimento de um serviço de *data stewardship* no CIIMAR, alinhadas com os princípios FAIR e com as necessidades institucionais identificadas.

1.3 Abordagem Metodológica

Esta dissertação baseia-se numa abordagem qualitativa e exploratória, adequada à natureza interpretativa do objeto de estudo e ao propósito de compreender práticas, perceções e desafios associados à gestão e partilha de dados em contextos multidisciplinares. Esta abordagem visa aprofundar a compreensão das experiências e perspetivas dos investigadores do programa CIIMAR Watch, privilegiando a análise contextual e descritiva das suas práticas de gestão e partilha de dados.

A recolha de dados foi realizada através de entrevistas semiestruturadas a investigadores integrados no programa CIIMAR Watch, permitindo uma exploração flexível de temas centrais relativos ao ciclo de vida dos dados. As entrevistas foram transcritas e analisadas com recurso ao software **CATMA**, segundo uma lógica de codificação temática. Este processo permitiu identificar práticas e desafios emergentes, alinhando-se com o objetivo de caracterizar e compreender as práticas de gestão de dados no contexto do programa.

Complementarmente, ao longo do processo de investigação foram desenvolvidas atividades que contribuíram para o reforço do enquadramento teórico e para a consolidação das competências metodológicas associadas à análise do caso de estudo. Parte dos resultados preliminares relativos ao estado da arte em gestão de dados marinhos foi apresentada no *Fórum de Gestão de Dados de Investigação* (Viseu, 2024), em formato de póster, o que permitiu validar a pertinência do estudo e promover o diálogo com outros investigadores e profissionais da área. O póster encontra-se disponível em acesso aberto na plataforma Zenodo.¹

Adicionalmente, foi frequentado o curso online "*Ocean Data Management*"² promovido pelo Flanders Marine Institute (VLIZ), que proporcionou uma visão abrangente sobre práticas, ferramentas e infraestruturas relevantes para a gestão de dados marinhos e ambientais. Embora estas atividades não integrem diretamente a metodologia de recolha e análise de dados, contribuíram para o enriquecimento do enquadramento teórico e para a identificação de boas práticas e *standards* relevantes para a análise do caso de estudo.

¹<https://zenodo.org/records/14764362>

²<https://oceanexpert.org/event/4056>

1.4 Estrutura da Dissertação

Em adição a esta introdução, esta dissertação contém mais 7 capítulos.

O Capítulo 2 apresenta uma revisão de literatura sobre a gestão de dados de investigação, abordando conceitos-chave, os princípios FAIR, a ciência aberta, o ciclo de vida dos dados, as políticas e os principais desafios enfrentados pelos investigadores.

O Capítulo 3 analisa a gestão de dados em ciências marinhas e ambientais, explorando as especificidades deste domínio, projetos e infraestruturas internacionais relevantes, vocabulários e normas utilizadas, repositórios disponíveis e boas práticas e desafios associados.

O Capítulo 4 caracteriza o estudo de caso, descrevendo o CIIMAR e o programa CIIMAR Watch, os seus objetivos e funcionamento, e a diversidade de perfis dos investigadores envolvidos.

O Capítulo 5 detalha a metodologia, desde a construção do guião de entrevistas semiestruturadas até à codificação e análise qualitativa das transcrições, explicando as etapas de preparação, recolha e análise de dados com recurso a ferramentas digitais.

O Capítulo 6 apresenta os resultados do estudo e a respetiva análise, estruturados em diferentes dimensões da gestão e partilha de dados no âmbito do CIIMAR Watch. Inicia-se com a caracterização das tipologias de dados gerados, seguindo-se a análise das práticas de anotação e registo, e de armazenamento e organização. Posteriormente, são abordadas as práticas de documentação e metadados, bem como os aspetos relativos à reutilização e partilha de dados, incluindo uma subsecção dedicada às experiências, barreiras e expectativas quanto a repositórios de dados. O capítulo prossegue com a síntese das práticas, desafios e requisitos identificados e encerra com a discussão dos resultados.

O Capítulo 7 reúne as recomendações derivadas dos resultados, propondo orientações para a melhoria da gestão e partilha de dados no contexto do programa CIIMAR Watch, incluindo a apresentação de um modelo preliminar de metadados adaptado à realidade do programa.

Por fim, o Capítulo 8 encerra a dissertação com uma reflexão crítica sobre os resultados, destacando os contributos do estudo, as suas limitações e perspetivas para trabalho futuro.

Capítulo 2

Gestão de Dados de Investigação

A gestão de dados de investigação tornou-se uma área central no âmbito da Ciência da Informação, cuja origem remonta a atividades especializadas em bibliotecas e centros de documentação. Com a evolução tecnológica e a crescente ênfase na ciência aberta e na reprodutibilidade, a gestão e a partilha de dados passaram a ser reconhecidas como práticas essenciais, tanto para investigadores como para profissionais da informação (Bawden et al., 2022).

Neste contexto, conceitos como *data stewardship* e *data curation*, assumem particular relevância e devem ser compreendidos de forma complementar. A curadoria digital (*digital curation*) consiste na gestão ativa dos dados ao longo do seu ciclo de vida, garantindo acessibilidade, preservação, autenticidade e integridade; envolve atividades como organizar, descrever, limpar, enriquecer e preservar dados para suportar a partilha e reutilização. Whyte et al., 2011, sublinham que a GDI implica "a organização dos dados, desde a sua entrada no ciclo de investigação até à disseminação e arquivamento de resultados valiosos", enquanto Borgman, 2012 enfatiza que, sendo dispendiosos de produzir, os dados só realizam o seu valor quando podem ser efetivamente reutilizados.

Por sua vez o *data stewardship* foca-se na gestão responsável de coleções de dados partilhados e na sua governação, incorporando a noção de "cuidado a longo prazo" de ativos digitais valiosos no futuro (Thanos, 2017; Wilkinson et al., 2016). Em termos práticos, a publicação e a (re)utilização de dados resultam da articulação entre curadoria e stewardship: as atividades de preservação e curadoria garantem que conjuntos de dados valiosos e onerosos se mantêm seguros e úteis ao longo do tempo (Cox; Pinfield, 2014), ao passo que a governação e os mecanismos de stewardship sustentam a persistência, a confiança e reutilização dos dados em ecossistemas partilhados (Thanos, 2017).

A literatura é consistente ao apontar que a curadoria de dados reduz ameaças ao seu valor de longo prazo e mitiga o risco de obsolescência digital, aumentando o valor dos dados existentes ao torná-los disponíveis para investigação de elevada qualidade (Whyte et al., 2011; Borgman, 2012). Neste contexto, os curadores de dados desempenham um papel fundamental, uma vez que a documentação contextual, a normalização e o enriquecimento da informação potenciam a descoberta, a integração e a reutilização dos dados (Corti et al., 2014).

Deste modo, a GDI pode ser entendida como o processo que engloba criar, encontrar, organizar, documentar, armazenar, partilhar e preservar dados, enfrentando desafios como a descoberta, o contexto, a preservação a longo prazo, a motivação para a partilha e a facilitação da reutilização (Cox; Verbaan, 2018). O seu valor como componente chave da integridade e da reprodutibilidade tem sido reiterado por financiadores, governos e instituições, que reforçam a adoção de práticas alinhadas com princípios FAIR e com políticas institucionais.

Em paralelo, o envolvimento dos investigadores é determinante para consolidar boas práticas. Contudo, existem barreiras - perceção, de custos de tempo, resistência à partilha, assimetrias disciplinares - que explicam a distância entre atitudes favoráveis e práticas efetivas (Surkis et al., 2015). Estas dinâmicas sublinham a importância da colaboração entre profissionais da informação e equipas científicas, articulando data stewardship e curadoria no quotidiano dos projetos.

Tendo delimitado o enquadramento conceptual da GDI, analisam-se de seguida as noções de dados de investigação, as suas tipologias e implicações para gestão, documentação e reutilização.

2.1 Dados de Investigação

Os dados de investigação são dados gerados e utilizados no âmbito da investigação científica, constituindo recursos fundamentais para a formulação de conclusões, a tomada de decisões e a promoção da inovação e do progresso científico.

A Organização para a Cooperação e o Desenvolvimento Económico (OCDE), na sua publicação sobre *princípios e guidelines para o acesso a dados de investigação com financiamento público*¹, caracteriza os dados de investigação como registos empíricos - numéricos, textuais e multimédia -, que servem como fontes fundamentais na investigação científica. Por outro lado, Briney (2015) define dados como "qualquer coisa sobre a qual se efetue uma análise". A literatura evidencia a existência de múltiplas definições de dados, cuja aplicabilidade varia consoante a área científica e o contexto de investigação.

Para além de definirem o objeto central da investigação, os dados possuem um valor intrínseco e funcional: são a base do conhecimento científico, permitem validar e reproduzir resultados, aumentam a visibilidade através de citação quando partilhados, e abrem novas oportunidades de investigação.

Os dados de investigação podem assumir diferentes naturezas, dependendo da sua origem. Dados observacionais são obtidos através de observações diretas, num determinado espaço e tempo, não podendo ser recolhidos novamente; dados experimentais são obtidos através da realização de experiências controladas, podendo ser reproduzíveis. Dados computacionais resultam da execução de modelos ou simulações, gerando dados sintéticos. A sua reprodutibilidade depende, contudo, da disponibilidade do software, hardware e dados utilizados na simulação original (Borgman, 2012; Thanos, 2017).

¹<https://doi.org/10.1787/9789264034020-en-fr>

Os dados podem ser classificados como brutos ou processados, segundo o glossário CODATA². Os dados brutos, como mencionado por Willis et al., 2012, representam os dados originais, exatamente como foram recolhidos durante o processo de investigação, e que não sofreram qualquer manipulação ou processamento. Por outro lado, os dados processados já foram alvo de manipulação, que pode incluir limpeza, estruturação ou análise, tornando os dados mais compreensíveis e utilizáveis.

Em alguns casos, a preservação de dados brutos é crucial para garantir a reprodutibilidade e a transparência da investigação (Castro, 2020). Esta distinção é particularmente relevante para a gestão de dados de investigação, dado que os diferentes estágios de processamento implicam requisitos distintos de documentação, armazenamento e partilha.

O volume crescente de dados coloca pressão direta sobre os serviços de GDI, exigindo soluções para armazenamento, arquivamento, preservação a longo prazo, cópias de segurança e gestão de custos e infraestruturas (Sheikh et al., 2023; Cox; Verbaan, 2018). Entre os principais obstáculos identificados contam-se problemas de armazenamento, questões de propriedade intelectual, limitações organizacionais, défices de pessoal qualificado e constrangimentos financeiros para soluções de armazenamento escalável (Berman, 2017; Burgi et al., 2017; Chiware et al., 2018; Cox; Kennan et al., 2019; Tang et al., 2019; Tripathi et al., 2017; Muhammad Ashiq et al., 2020).

O armazenamento em nuvem, embora cada vez mais utilizado, apresenta igualmente limites práticos, como custos adicionais e preocupações relacionadas com a segurança e a localização dos dados, o que exige políticas claras de proteção e a manutenção de cópias offline (Corti et al., 2014). Por sua vez, os dados processados tendem a ser mais fáceis de gerir e partilhar, mas requerem documentação rigorosa dos métodos aplicados, registo da proveniência e anotação normalizada para garantir que permaneçam compreensíveis e reutilizáveis por terceiros (Cox; Verbaan, 2018; Corti et al., 2014; Higman et al., 2019; Birkbeck et al., 2022).

A colaboração interdisciplinar resulta frequentemente em conjuntos de dados complexos e heterogêneos, dada a diversidade de normas, requisitos e expectativas de cada disciplina. Para que estes dados possam ser partilhados e reutilizados de forma eficaz, é essencial que sejam tornados FAIR - localizáveis, acessíveis, interoperáveis e reutilizáveis (Wilkinson et al., 2016). Este processo exige um empenho específico em cada área científica para definir as melhores práticas de formatação e descrição dos dados, assegurando que possam ser integrados e utilizados de forma eficaz em diversos contextos. Importa, contudo, distinguir entre a reutilização por humanos — que requer documentação clara e contextual — e a reutilização por máquinas, que depende da existência de metadados normalizados e semanticamente estruturados que permitam a interpretação automática dos dados.

2.2 Princípios FAIR e Ciência Aberta

O conceito de dados FAIR (*Findable, Accessible, Interoperable, and Reusable*) surgiu num contexto de rápido crescimento do volume de dados científicos e da necessidade de maximizar o

²<http://www.codata.org/>

seu impacto e valor. A necessidade de preservar e tornar os dados científicos abertos e acessíveis é debatida desde, pelo menos, os anos 90 (Ray, 2014).

Antes da formalização dos princípios FAIR, várias iniciativas e políticas internacionais já sublinhavam a importância da abertura e da gestão adequada dos dados de investigação. Em 2007, os Princípios e Diretrizes da OCDE para o "Acesso a Dados de Investigação com Financiamento Público" destacavam os benefícios científicos, económicos e sociais de garantir o acesso aberto a dados financiados por recursos públicos. Mais tarde, em 2012, o relatório da Royal Society, *Science as an Open Enterprise*³, reforçou que, para além de serem abertos, os dados deveriam ser acessíveis, interoperáveis e reutilizáveis. Esta ideia foi também reiterada na Declaração dos Ministros da Ciência do G8⁴, em 2013, onde se defendia que os dados científicos fossem localizáveis e utilizados de forma fácil e integrada, respeitando as normas de qualidade. Já em 2016, no âmbito do programa *Horizon 2020*, a Comissão Europeia introduziu orientações específicas para incentivar práticas FAIR, promovendo a gestão eficaz e a partilha responsável dos dados de investigação.

Em 2014, numa conferência do *Lorentz Center*⁵, em Leiden (Países Baixos), foram concebidos os Princípios FAIR, que vieram a ser publicados em 2016 pelo grupo *FORCE11* (Directorate-General for Research and Innovation (European Commission), 2018), destacando a importância de tornar os dados de investigação localizáveis, acessíveis, interoperáveis e reutilizáveis.

O primeiro princípio, **Findable**, foca-se na necessidade dos dados e metadados serem facilmente encontrados por humanos e máquinas. Isso implica que os dados devem ser identificados com identificadores únicos e persistentes, descritos com metadados ricos, e registados ou indexados em recursos pesquisáveis de forma a serem encontrados por potenciais utilizadores.

O princípio **Accessible** refere-se à disponibilidade dos dados, garantindo que possam ser recuperados de forma segura através de protocolos claramente definidos, com metadados acessíveis, mesmo que nem todos os dados estejam abertos.

Já o princípio **Interoperable** destaca a necessidade dos dados de diferentes fontes serem integrados e trabalhados em conjunto com o mínimo de esforço. Para isso, é essencial recorrer a linguagens formais, acessíveis, partilhadas e amplamente reconhecidas para a representação do conhecimento.

Por fim, o princípio **Reusable** enfatiza que os dados devem ser suficientemente bem descritos e documentados para que possam ser reutilizados em novos contextos. Além disso, devem estar acompanhados de licenças claras que definam as condições de reutilização, informações de proveniência detalhadas e conformidade com as normas da comunidade (Wilkinson et al., 2016).

Estes princípios, apesar de conectados, são independentes, ou seja, cada princípio pode ser aplicado individualmente ou em combinação com outros, de forma incremental. Os princípios FAIR são um guia que ajuda os investigadores e gestores de dados a avaliar se as suas ações estão efetivamente a contribuir para a abertura e acesso aos dados, sendo que este é um processo

³<https://royalsociety.org/topics-policy/projects/science-public-enterprise/report>

⁴<https://www.gov.uk/government/news/g8-science-ministers-statement>

⁵<https://www.lorentzcenter.nl/>

contínuo e não um estado absoluto (Directorate-General for Research and Innovation (European Commission), 2018). A conformidade dos dados com os princípios FAIR pode variar, dependendo das circunstâncias e da implementação progressiva de práticas recomendadas, com o objetivo de melhorar continuamente a sua gestão e impacto.

2.3 Políticas de dados e iniciativas como a *Research Data Alliance*

As políticas de gestão de dados de investigação e iniciativas internacionais como a *Research Data Alliance* (RDA)⁶ constituem elementos fundamentais para a consolidação de uma cultura de partilha e reutilização de dados na comunidade científica (Andrikopoulou et al., 2022, Cox; Pinfield, 2014). A implementação destes elementos reflete a crescente valorização da transparência e da ciência aberta, assumindo os dados financiados com recursos públicos como um bem público.

As políticas de dados de investigação, enquanto declarações formais de princípios, têm como objetivo orientar práticas institucionais relacionadas com a recolha, organização, armazenamento, partilha e preservação de dados (Ray, 2014; Corti et al., 2014; Karimova et al., 2021). Uma política bem estruturada reduz mal-entendidos, clarifica funções e responsabilidades (Green et al., 2009; Cox; Pinfield, 2014), e promove a organização dos dados ao longo de todo o ciclo de vida da investigação, desde a recolha inicial até à disseminação e preservação dos resultados (Read et al., 2015).

O desenvolvimento destas políticas tem sido fortemente impulsionado por organismos de financiamento e por editores científicos, que exigem cada vez mais que os dados sejam disponibilizados em conformidade com princípios de acessibilidade, interoperabilidade e reutilização. No Reino Unido, por exemplo, todos os Conselhos de Investigação adotaram em 2011 os Common Principles on Data Policy (Cox; Verbaan, 2018; Corti et al., 2014), destacando a centralidade dos dados para o avanço do conhecimento científico. Nos Estados Unidos, instituições como a National Science Foundation (NSF) e os National Institutes of Health (NIH) assumem igualmente um papel determinante ao exigir que projetos financiados contemplem planos claros de gestão de dados (Ray, 2014; Read et al., 2015; Sheikh et al., 2023). Um marco relevante foi ainda o memorando do Gabinete de Política Científica e Tecnológica da Casa Branca (OSTP), em 2013, que estabeleceu a obrigatoriedade de Data Management Plans (DMPs) para todas as agências federais (Ray, 2014).

Apesar do seu impacto positivo, a implementação destas políticas continua a enfrentar desafios significativos. A diversidade de orientações entre financiadores, a falta de consistência conceptual (Birkbeck et al., 2022; Williams et al., 2017) e a ausência de mecanismos de verificação ou financiamento adequado (Cox; Verbaan, 2018; Ray, 2014; Sheikh et al., 2023) podem gerar resistência entre os investigadores, que por vezes percecionam os DMPs como um fardo administrativo em vez de uma ferramenta estratégica (Birkbeck et al., 2022; Corti et al., 2014; Connie, Clare et al., 2019). É, por isso, essencial que estes planos sejam integrados como parte das estratégias institucionais e acompanhados de apoio formativo e técnico (Murtaza Ashiq et al., 2020; Birkbeck et al.,

⁶<https://www.rd-alliance.org/>

2022; Karimova et al., 2021), assegurando que a documentação dos métodos e processos torne os dados compreensíveis e utilizáveis (Ray, 2014).

Murtaza Ashiq et al., 2020 fala da necessidade de "treino para investigadores e profissionais de bibliotecas" e de "colaboração entre *stakeholders* e departamentos de serviços universitários". Birkbeck et al., 2022 sugere que a resistência pode ser resolvida com "educação que construa a sua capacidade e confiança" e que é um "desafio institucional capacitar os investigadores em RDM". Por sua vez, Karimova et al., 2021 destaca que "é essencial ter pessoal especializado para auxiliar os investigadores nestas tarefas" e que "os serviços de consulta e sessões de formação" são iniciativas comuns de apoio.

Neste panorama, a Research Data Alliance (RDA) desempenha um papel central ao fornecer uma infraestrutura social e técnica que apoia a partilha e a reutilização abertas de dados. Criada em 2013, com o apoio da Comissão Europeia, da NSF e do Governo Australiano, a RDA funciona como um fórum neutro que reúne voluntários de diferentes setores para debater e implementar soluções em gestão de dados. Através dos seus Working Groups (WGs) e Interest Groups (IGs), a organização promove o desenvolvimento de recomendações práticas, abordando desde questões de interoperabilidade até problemas específicos de partilha de dados.

As contribuições da RDA estendem-se à promoção da aplicação dos princípios FAIR, à criação de redes nacionais para a sua disseminação local, e ao incentivo de boas práticas de GDI, incluindo a elaboração de DMPs, a certificação de repositórios e a definição de recomendações para a governação de dados, como práticas de citação adaptadas a diferentes domínios científicos. Ao fomentar o contacto entre especialistas e ao desenvolver soluções interdisciplinares, a RDA tem contribuído de forma decisiva para moldar o panorama da gestão de dados, promovendo uma cultura global orientada para a abertura, a qualidade e a reutilização.

2.4 Curadoria e ciclo de vida dos dados

A curadoria de dados é essencial para assegurar que os dados de investigação são devidamente geridos, descritos, preservados e disponibilizados, garantindo a sua reutilização (Corti et al., 2014). Este processo não é responsabilidade exclusiva dos investigadores, exigindo a colaboração com curadores de dados, que atuam como ponte entre a investigação e serviços de apoio como bibliotecas, equipas de IT ou unidades responsáveis por questões jurídicas, éticas e de privacidade (Connie, Clare et al., 2019). Para além de promoverem boas práticas de descrição, organização e preservação, estes profissionais desempenham um papel central na elaboração de planos de gestão de dados, assegurando o uso de modelos de metadados adequados e favorecendo a interoperabilidade e reutilização (Rodrigues et al., 2019).

Os dados, criados no decurso da investigação, possuem também o seu próprio ciclo de vida (Cox; Verbaan, 2018). Entre os modelos desenvolvidos para o estruturar, destaca-se o Modelo de Curadoria do Ciclo de Vida do Digital Curation Centre (DCC)⁷, representado na Figura 2.1.

⁷Disponível em: <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>

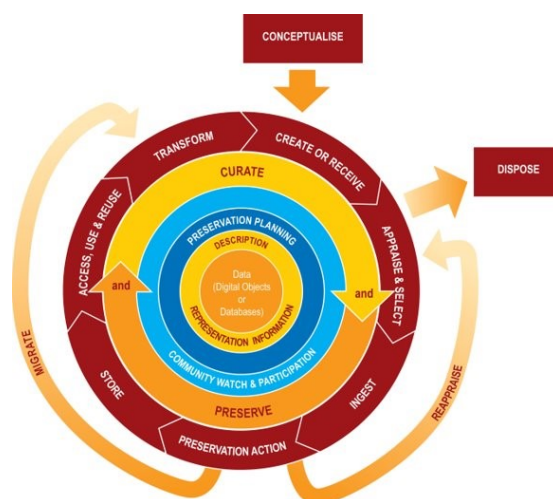


Figura 2.1: Modelo de Curadoria do Ciclo de Vida dos Dados do DCC

Este modelo identifica três categorias de ações. As ações completas são contínuas e acompanham todo o ciclo de vida dos dados, incluindo a descrição e representação da informação, o planeamento da preservação, a monitorização das necessidades da comunidade científica e a implementação das medidas de curadoria. As ações sequenciais ocorrem em momentos específicos, desde a conceptualização e criação dos dados até à sua avaliação e seleção, ingestão em repositórios, armazenamento, aplicação de medidas de preservação e disponibilização para acesso, reutilização e eventual transformação. Já as ações ocasionais são aplicadas apenas quando necessário, como no descarte de dados que não serão preservados, na reavaliação de conjuntos que falham nos critérios de validação ou na migração para formatos mais adequados à preservação a longo prazo.

Embora mais direcionado para o trabalho dos curadores de dados, o modelo do DCC oferece uma visão abrangente que pode apoiar investigadores e instituições na planificação da gestão dos seus dados, assegurando que todas as fases do ciclo de vida são devidamente consideradas.

2.5 Metadados e documentação

A documentação e os metadados constituem elementos centrais na gestão de dados de investigação, pois permitem contextualizar, descrever e tornar os dados compreensíveis e reutilizáveis ao longo do tempo. Embora frequentemente utilizados como sinónimos, ambos desempenham funções distintas: a documentação refere-se à descrição mais ampla de um projeto ou conjunto de dados, incluindo informação sobre os objetivos, metodologias, instrumentos de recolha, locais de amostragem ou condições experimentais (Corti et al., 2014); os metadados, por sua vez, dizem respeito a descrições estruturadas diretamente associadas a cada ficheiro ou registo de dados, definidas segundo esquemas ou padrões partilhados pela comunidade científica (Qin et al., 2012; Corti et al., 2014).

Os metadados asseguram que os dados sejam interpretáveis e interoperáveis, permitindo que possam ser descobertos, acedidos e integrados em diferentes sistemas (Ray, 2014). No domínio marinho e ambiental, padrões como o ISO 19115 (informação geográfica), o Darwin Core (biodiversidade) (Wieczorek et al., 2012; Ribeiro et al., 2018), ou o CF Conventions (dados climáticos e oceanográficos) são amplamente utilizados, fornecendo uma estrutura consistente para descrever variáveis, unidades de medida, coordenadas espaciais e temporais (Corti et al., 2014). A adoção destes padrões promove a interoperabilidade entre repositórios e aumenta a possibilidade de integração de dados em infraestruturas internacionais (Genova et al., 2017), como o SeaDataNet⁸ ou o Ocean Biodiversity Information System (OBIS)⁹.

Apesar da sua relevância, a criação de metadados enfrenta diversos desafios. Entre os mais referidos encontram-se o tempo e esforço exigidos na sua elaboração (Jahnke et al., 2012; Ray, 2014), a dificuldade em produzir descrições detalhadas para dados puramente numéricos (Amorim et al., 2017), e a falta de formação específica entre os investigadores (Ray, 2014; Tenopir; Dalton et al., 2015). Em muitos casos, os metadados acabam por ser produzidos de forma genérica ou incompleta, comprometendo a utilidade futura dos dados e a sua conformidade com os princípios FAIR (Ray, 2014; Tenopir; Allard et al., 2011).

A documentação, embora menos formalizada, desempenha igualmente um papel crítico. Notas de campo, folhas de campanha, protocolos laboratoriais ou relatórios técnicos fornecem contexto essencial sobre a recolha e o tratamento dos dados (Corti et al., 2014; Cox; Verbaan, 2018). Contudo, quando estes documentos não são devidamente preservados ou associados aos dados principais, a compreensão do conjunto pode ser limitada, dificultando a sua reutilização (Higman et al., 2019).

Neste sentido, tanto a documentação como os metadados devem ser encarados como dimensões complementares: a documentação garante a contextualização e compreensão global dos dados, enquanto os metadados asseguram a sua normalização e interoperabilidade técnica. A conjugação de ambas é indispensável para garantir que os dados possam ser corretamente interpretados e reutilizados por investigadores de diferentes áreas disciplinares, alinhando-se com os objetivos da ciência aberta e com os princípios FAIR.

2.6 Os investigadores e a partilha de dados

A partilha dos dados por parte dos investigadores é um tema complexo e influenciado por diversos fatores, incluindo as suas práticas, perceções e o contexto institucional em que trabalham (Cox; Verbaan, 2018). Os investigadores podem ter diferentes perspetivas sobre a partilha de dados, e essas perspetivas são moldadas tanto por fatores individuais como organizacionais.

Embora a partilha de dados seja fundamental para a validação científica, muitos investigadores ainda hesitam em disponibilizar os seus dados. Apesar de muitos investigadores demonstrarem interesse em partilhar dados de outros, poucos são os que realmente partilham os seus próprios

⁸<https://www.seadatanet.org/>

⁹<https://obis.org/>

dados, o que reflete uma clara discrepância entre o interesse e a prática (Tenopir; Allard et al., 2011).

Ao longo do tempo, a percepção dos investigadores sobre a partilha de dados tem vindo a melhorar. Embora as preocupações com a reutilização dos dados de terceiros ainda sejam preva-lentes, existe uma tendência crescente para considerar a partilha como algo valioso. A mudança nas atitudes dos investigadores reflete assim uma maior valorização da partilha de dados, sendo que uma grande parte já reconhece a sua importância para a continuidade da pesquisa e para a validação dos resultados. Este fenómeno pode ser observado em estudos que indicam uma maior disposição para disponibilizar dados, especialmente quando existem incentivos como o crédito por citação e a visibilidade das suas publicações (Tenopir; Dalton et al., 2015).

Neste sentido, diversos estudos sublinham as principais **motivações e benefícios** percebidos pelos investigadores. O aumento da visibilidade e da taxa de citação está frequentemente associado ao acesso detalhado aos dados de investigação, sendo que muitos cientistas consideram essencial receber reconhecimento formal pela partilha dos seus dados (Tenopir; Dalton et al., 2015). Um estudo de Tenopir; Allard et al. (2011) indica que cerca de 80% dos cientistas se mostram dispostos a partilhar dados com a comunidade de investigação e educação. A oportunidade de colaborar em projetos que reutilizam os seus dados é outra motivação importante, particularmente nas ciências da saúde e medicina (Tenopir; Allard et al., 2011; Tenopir; Dalton et al., 2015). A partilha é ainda entendida como condição para a reprodutibilidade e transparência da investigação, permitindo verificar descobertas científicas (Cox; Verbaan, 2018). Para além disso, os dados permanecem um recurso valioso mesmo após o término de um projeto, constituindo novas oportunidades de investigação e evitando duplicação de esforços (Corti et al., 2014). O contexto disciplinar também influencia: um estudo de Kim et al. (2012) revelou que cientistas de biologia, química e ecologia viam a partilha como crítica, enquanto investigadores de informática, engenharia e matemática tendiam a considerá-la menos relevante. Pressões de financiadores e revistas, expectativas dos pares, existência de repositórios e normas, e o desejo de evidenciar a qualidade dos dados foram apontados como motivadores adicionais (Kim et al., 2012; Sayogo et al., 2013).

Por outro lado, existem **barreiras e desafios** concretos que condicionam a prática da parti-lha. A falta de tempo e de financiamento é uma das razões mais frequentemente referidas para não disponibilizar dados de forma estruturada (Tenopir; Allard et al., 2011). A estas somam-se receios de má interpretação ou uso indevido, bem como preocupações relacionadas com a prop-riedade intelectual. Quando os dados são sensíveis, como acontece em muitas ciências sociais, a sua preservação exige procedimentos rigorosos de anonimização e consentimento (Castro, 2020). A falta de competências e de práticas estabelecidas de descrição constitui outro entrave: a cria-ção de metadados é percecionada como complexa e demorada, e a documentação inconsistente fragiliza a reutilização. Estudos revelam frustração entre investigadores quando dados não são adequadamente anotados, sobretudo após a saída de colegas de projeto (Wiley et al., 2018; Curty, 2016). Resistências culturais também persistem, como a percepção de perda de controlo quando os dados são armazenados fora do computador pessoal (Connie, Clare et al., 2019) ou a ideia de que “ninguém estará interessado nos dados brutos, apenas nas tabelas e figuras geradas” (Castro,

2020). Em alguns casos, investigadores receiam ainda que revistas de topo não aceitem artigos se os dados tiverem já sido explorados em outros estudos (Castro, 2020).

A heterogeneidade disciplinar contribui para a diversidade de práticas: enquanto algumas áreas adotam políticas e culturas favoráveis à partilha, noutras subsistem barreiras estruturais (Kim et al., 2012; Sayogo et al., 2013). Esta diversidade repercute-se também nas **práticas atuais**, onde prevalece o armazenamento informal. Um estudo reporta que mais de 90% dos investigadores armazenam os seus dados em computadores pessoais, sendo esta a prática mais comum (Tenopir; Dalton et al., 2015). O depósito formal em repositórios é ainda limitado: apenas cerca de 12% dos dados financiados por bolsas de investigação são arquivados de forma sistemática, enquanto 45% circulam informalmente e 44% permanecem restritos às equipas de investigação (Pienta et al., 2010). A documentação ad-hoc, através de cadernos de laboratório ou folhas de cálculo, é também frequente, dificultando a organização e a reutilização posterior (Castro, 2020).

Em suma, a partilha de dados é percecionada pelos investigadores como simultaneamente necessária e desafiante. As motivações e benefícios coexistem com barreiras e resistências, e as diferenças disciplinares e institucionais moldam fortemente os comportamentos. A promoção de políticas claras, incentivos adequados e serviços de apoio poderá reduzir estas barreiras e consolidar práticas mais consistentes e sustentáveis de gestão e partilha de dados.

Capítulo 3

Gestão de Dados de Investigação Marinhos e Ambientais

A gestão dos dados provenientes da investigação marinha desempenha um papel crucial para compreender a complexa interação entre os aspetos físicos, químicos e biológicos do oceano, de forma a obter uma visão integrada dos seus ecossistemas. Através da monitorização e análise das condições dos oceanos, é possível prever impactos ambientais e tomar decisões informadas sobre a conservação e utilização sustentável dos recursos marinhos, assim como elaborar políticas ambientais sustentadas por dados científicos de confiança (G. M. R. Manzella et al., 2022). Neste sentido, têm sido desenvolvidas diversas iniciativas e projetos internacionais de forma a garantir que os dados marinhos sejam recolhidos, geridos e partilhados de formas eficaz e acessível.

3.1 Especificidade dos Dados Marinhos e Ambientais

Os dados marinhos e ambientais possuem especificidades que os tornam particularmente complexos na sua recolha, gestão e utilização. São dados intrinsecamente diversos e heterogêneos, abrangendo medições físicas, químicas, biológicas, geológicas e até antropogénicas (G. M. Manzella et al., 2022). Esta diversidade decorre da multiplicidade de metodologias e objetivos de recolha (D. M. Schaap et al., 2022), bem como da introdução constante de novos sensores e tecnologias, que ampliam o volume e a variedade dos dados produzidos (Beja et al., 2022). Consequentemente, surgem desafios acrescidos para a sua normalização, interoperabilidade e reutilização (De Pooter et al., 2017).

O ambiente de recolha acrescenta dificuldades próprias: sensores subaquáticos operam sob condições adversas, como pressão elevada, correntes fortes e bioincrustação em águas costeiras, fatores que podem comprometer a qualidade e continuidade das medições. Sem metadados que registem estas condições e intervenções, a fiabilidade dos dados pode ser afetada (G. M. Manzella et al., 2022).

A investigação marinha é, além disso, marcada pela sua natureza multidisciplinar. Diferentes áreas — física, química, biologia ou geologia — desenvolveram práticas e padrões próprios,

frequentemente resultando em sistemas de gestão de dados isolados e difíceis de integrar (S. Simoncelli et al., 2022). Por isso, a colaboração e a interoperabilidade são condições fundamentais para a valorização destes dados, ainda que a diversidade de abordagens continue a representar um desafio significativo.

Neste sentido, têm sido desenvolvidas diversas iniciativas, infraestruturas e normas internacionais que procuram responder a estas especificidades, promovendo a harmonização dos dados marinhos e ambientais e facilitando a sua partilha e reutilização em contextos multidisciplinares.

3.2 Iniciativas e Serviços Relevantes/Infraestruturas e Projetos Relevantes

A gestão eficaz dos dados marinhos exige não apenas boas práticas locais, mas também o enquadramento em redes e infraestruturas que permitam a interoperabilidade, partilha e reutilização dos dados à escala regional e global. Diversas iniciativas internacionais têm desempenhado um papel crucial na normalização, organização e disseminação dos dados marinhos, promovendo uma ciência mais aberta, colaborativa e sustentada.

OceanBestPractices Systems (OBPS) O OBPS¹, iniciativa apoiada pela Comissão Oceanográfica Intergovernamental (IOC) da UNESCO, funciona como uma plataforma global que agrega, organiza e divulga boas práticas e metodologias normalizadas no domínio das ciências marinhas. O seu principal objetivo é facilitar o acesso a práticas validadas pela comunidade científica, promovendo a harmonização de processos de recolha, gestão e disseminação de dados oceânicos.

EMODnet (European Marine Observation and Data Network) A EMODnet² é uma rede pan-europeia de organizações financiadas pela União Europeia que disponibiliza dados marinhos abertos, interoperáveis e de alta qualidade, organizados por áreas temáticas. A rede promove o acesso gratuito e integrado a dados provenientes de diferentes países e instituições, contribuindo para a investigação, o ordenamento do espaço marítimo, a economia azul e a definição de políticas ambientais.

SeaDataNet Complementando o EMODnet, o SeaDataNet³ é uma infraestrutura pan-europeia dedicada à gestão de dados oceanográficos. Integra institutos e centros de investigação marinha de vários países, promovendo a normalização dos metadados, o uso de vocabulários controlados, e o desenvolvimento de ferramentas de interoperabilidade para acesso e partilha de dados. O SeaDataNet disponibiliza serviços como o *Common Data Index* (CDI) e o *SeaDataCloud*, facilitando o intercâmbio de dados físicos, químicos, biológicos e geológicos.

¹<https://www.oceanbestpractices.org/>

²<https://emodnet.ec.europa.eu/en>

³<https://www.seadatanet.org/>

Marine Biodiversity Observation Network (MBON) O MBON é uma rede internacional que se visa monitorizar e partilhar dados sobre biodiversidade marinha. Está integrado no sistema de observação global GEO BON e procura articular esforços locais e regionais para produzir indicadores úteis para a ciência e a política. Os dados gerados são frequentemente publicados no OBIS (*Ocean Biodiversity Information System*)⁴, uma plataforma global de dados sobre biodiversidade marinha, que permite o acesso aberto e normalizado a milhões de registos de espécies em ambientes marinhos.

Estas infraestruturas têm desempenhado um papel fundamental na construção de ecossistemas de dados colaborativos, na disseminação de boas práticas e no reforço da literacia em gestão de dados marinhos. A sua existência demonstra que o desenvolvimento de políticas eficazes de gestão e partilha de dados depende não apenas de iniciativas locais, mas também da integração em redes de cooperação científica mais amplas.

3.3 Vocabulários Controlados e Normas e Ontologias

A utilização de vocabulários controlados e de normas específicas constitui um dos pilares fundamentais da gestão de dados de investigação, garantindo a consistência, a interoperabilidade e a reutilização a longo prazo. Os vocabulários controlados correspondem a listas de termos cuidadosamente selecionados, com significados inequívocos e definidos pela comunidade científica, assegurando que unidades de informação como parâmetros de medição, plataformas de recolha ou nomes taxonómicos são descritas de forma normalizada (G. M. R. Manzella et al., 2022). Esta normalização evita ambiguidades, facilita a pesquisa e recuperação de dados por humanos e máquinas e assegura a interpretação válida dos valores em diferentes contextos disciplinares.

No domínio marinho e ambiental, destacam-se alguns exemplos relevantes. O NERC Vocabulary Server (NVS)⁵, desenvolvido pelo British Oceanographic Data Centre (BODC), organiza thesaurus hierárquicos aplicados a múltiplos domínios oceanográficos, sendo amplamente utilizado pela infraestrutura SeaDataNet para a marcação de metadados e dados. Estes vocabulários abrangem parâmetros e grupos de parâmetros, unidades de medida, tipos de sensores, plataformas e fabricantes, permitindo uma descrição granular e semanticamente enriquecida. O World Register of Marine Species (WoRMS)⁶ constitui uma referência taxonómica global, fornecendo nomes validados de espécies marinhas e ligando-os a identificadores persistentes (LSIDs). Este padrão é utilizado, por exemplo, pelo Ocean Biodiversity Information System (OBIS) como espinha dorsal taxonómica. De forma complementar, o Marine Regions disponibiliza topónimos e áreas marinhas georreferenciadas, enquanto o Climate and Forecast (CF) Standard Names garantem a descrição inequívoca de variáveis ambientais em modelos e redes de observação. Finalmente, o Darwin

⁴<https://obis.org/>

⁵<https://vocab.nerc.ac.uk/>

⁶<https://www.marinespecies.org/>

Core (DwC)⁷, amplamente adotado por infraestruturas como o OBIS ou o Global Biodiversity Information Facility (GBIF)⁸, fornece um esquema unificado para a descrição de dados biológicos, facilitando a integração de múltiplos conjuntos de dados.

As normas, por sua vez, estabelecem especificações e requisitos técnicos que asseguram a qualidade, a fiabilidade e a harmonização dos dados produzidos. Entre as mais relevantes para a gestão de dados marinhos destaca-se a família ISO 19100, em particular a ISO 19115, que define estruturas de metadados para conjuntos de dados geoespaciais e é amplamente utilizada em infraestruturas como a SeaDataNet e o EMODnet. Outras normas relevantes incluem a ISO 8601, que especifica a representação padronizada de datas e horas, e a ISO/IEC 25012, que propõe um modelo de qualidade de dados baseado em características como precisão, consistência e credibilidade.

A Diretiva INSPIRE da União Europeia é outro marco importante, ao estabelecer uma infraestrutura comum de informação espacial que exige a harmonização de dados geográficos e ambientais produzidos pelos Estados-Membros. Os metadados do Common Data Index (CDI) da SeaDataNet, por exemplo, cumprem integralmente os requisitos do INSPIRE, assegurando a compatibilidade com serviços europeus de monitorização ambiental.

Para além das normas ISO e INSPIRE, destacam-se ainda outras normas técnicas com forte aplicação na área marinha. O formato NetCDF (Network Common Data Form)⁹, em particular na sua versão 4 com conformidade CF, é amplamente reconhecido para a partilha de dados ambientais e oceanográficos, sendo recomendado pela comunidade internacional para a distribuição de dados científicos de larga escala. O QARTOD (Quality Assurance of Real-Time Oceanographic Data) define diretrizes de controlo de qualidade para dados oceanográficos em tempo real, aplicadas por diversas redes de monitorização como a Ocean Networks Canada. No mesmo sentido, os serviços abertos definidos pelo Open Geospatial Consortium (OGC) — como WMS (Web Map Service)¹⁰, WFS (Web Feature Service)¹¹ ou CSW (Catalogue Service for the Web)¹² — são essenciais para a interoperabilidade, permitindo a partilha e visualização de dados geoespaciais em formatos normalizados.

Algumas normas combinam a natureza de vocabulário e de norma. O Darwin Core, além de vocabulário controlado, é também considerado uma norma que sustenta a integração global de dados de biodiversidade, garantindo a sua utilização em plataformas como o OBIS e o GBIF (Beja et al., 2022).

Apesar da diversidade de iniciativas, a adoção consistente destes vocabulários e normas continua a enfrentar desafios. A heterogeneidade dos dados, a complexidade técnica dos padrões e a falta de formação específica entre investigadores contribuem para práticas desiguais e, por vezes, incompletas. Para além disso, a implementação de normas e vocabulários exige coordenação

⁷<https://dwc.tdwg.org/>

⁸<https://www.gbif.org/>

⁹<https://www.unidata.ucar.edu/software/netcdf>

¹⁰<https://www.ogc.org/standards/wms/>

¹¹<https://www.ogc.org/standards/wfs/>

¹²<https://www.ogc.org/standards/cat/>

internacional e financiamento contínuo, sem os quais é difícil assegurar a sustentabilidade destas práticas (G. M. R. Manzella et al., 2022). Projetos e infraestruturas como a SeaDataNet, o EMOD-net, o OBIS ou a ONC trabalham ativamente na promoção da harmonização, desenvolvendo ferramentas de validação, serviços de conversão e mecanismos de apoio comunitário. A conjugação destes esforços é essencial para garantir a qualidade, a interoperabilidade e a reutilização dos dados marinhos, assegurando a sua integração em redes globais de observação e contribuindo para a ciência aberta.

Para além dos vocabulários e normas, importa destacar também o papel das ontologias, que vão além das classificações hierárquicas ao permitirem descrever relações semânticas complexas entre conceitos (S. Simoncelli et al., 2022). Enquanto uma taxonomia ou vocabulário assegura consistência terminológica, uma ontologia estabelece conexões ricas entre entidades, processos e atributos, tornando a informação legível tanto para humanos como para máquinas.

3.4 Repositórios de Dados

Os repositórios de dados são sistemas essenciais na gestão de dados marinhos e ambientais, desempenhando um papel fundamental na preservação, disseminação e reutilização da produção científica. Ao garantir o armazenamento seguro e o acesso estruturado a conjuntos de dados, estas infraestruturas permitem que os dados sejam partilhados com outros investigadores, aumentem a transparência dos processos científicos e potenciem novas análises, contribuindo para uma ciência mais aberta e colaborativa. Para além disso, os repositórios facilitam o cumprimento de requisitos institucionais e de financiadores, que cada vez mais exigem o depósito e a partilha dos dados de investigação.

Entre os diretórios mais relevantes que catalogam e descrevem repositórios científicos destaca-se o *Registry of Research Data Repositories* (re3data)¹³, um serviço global que reúne informações sobre centenas de repositórios de dados, abrangendo diversas disciplinas científicas. O re3data disponibiliza metadados descritivos sobre cada repositório, incluindo área temática, condições de acesso, políticas de partilha, utilização de identificadores persistentes, formatos suportados e certificações de confiança. Por esse motivo, é amplamente referenciado em políticas de gestão de dados de investigação, como as diretrizes de acesso aberto do programa *Horizon 2020* da Comissão Europeia.

Com o objetivo de identificar opções viáveis para o depósito e a partilha de dados marinhos e ambientais, foram analisados alguns repositórios listados no re3data, considerando a sua adequação ao contexto do CIIMAR Watch. Esta análise teve em conta critérios como o tipo de dados aceites, os requisitos de metadados, as políticas de acesso e a possibilidade de reutilização dos dados depositados. A identificação de repositórios relevantes visa apoiar os investigadores na adoção de práticas alinhadas com os princípios FAIR e facilitar a integração dos dados do programa em redes internacionais de investigação.

¹³<https://www.re3data.org/>

Para ilustrar a diversidade de repositórios relevantes para os dados marinhos e ambientais, foram selecionados sete exemplos a partir da plataforma re3data, considerando a sua relevância para diferentes áreas científicas, o tipo de dados suportados, os standards de metadados adotados e os mecanismos de acesso técnico disponíveis. O SeaDataNet e o Copernicus Marine Environmental Monitoring Service assumem especial importância no domínio da oceanografia europeia, oferecendo suporte a formatos normalizados como NetCDF e APIs de acesso. O WoRMS (World Register of Marine Species) e o OBIS (Ocean Biodiversity Information System) representam repositórios especializados em biodiversidade marinha, ambos amplamente reconhecidos internacionalmente e com adoção de padrões como Darwin Core e EML (Ecological Metadata Language)¹⁴. O GBIF (Global Biodiversity Information Facility) complementa esta dimensão ao integrar dados de biodiversidade em escala global, promovendo a interoperabilidade entre diferentes domínios científicos. Já o PANGAEA¹⁵ destaca-se como um repositório interdisciplinar de ciências da Terra e do mar, suportando padrões como ISO 19115 e Dublin Core. Finalmente, o ENA (European Nucleotide Archive) representa uma infraestrutura especializada em biologia molecular, assegurando a curadoria de sequências genômicas através de standards como INSDC e MIxS.

Tabela 3.1: Exemplos de repositórios de dados marinhos relevantes e respectivos padrões de metadados

Repositório	Área Científica	Metadados
PANGAEA	Ciências da Terra e do Mar	ISO 19115, Dublin Core
ENA (European Nucleotide Archive)	Biologia Molecular	INSDC, MIxS
WoRMS (World Register of Marine Species)	Biodiversidade	Darwin Core, Dublin Core, EML
OBIS (World Register of Marine Species)	Biodiversidade / Oceanografia	Darwin Core, EML, FGDC, ISO 19115
SeaDataNet	Oceanografia	ISO 19115
GBIF (Global Biodiversity Information Facility)	Biodiversidade Global	Darwin Core
Copernicus Marine EMS	Oceanografia	CF Conventions, NetCDF

A Tabela 3.1 sintetiza um conjunto de repositórios relevantes para os dados marinhos e ambientais, identificados através do re3data. Estes repositórios diferem quanto às áreas científicas cobertas, aos standards de metadados que adotam e aos mecanismos de acesso disponíveis, constituindo referências importantes para projetos como o CIIMAR Watch, que ambicionam disponibilizar dados em acesso aberto.

Os padrões de metadados asseguram a consistência e a interoperabilidade dos dados a nível internacional. Por exemplo, o ISO 19115 é amplamente utilizado em contextos oceanográficos

¹⁴<https://eml.ecoinformatics.org/>

¹⁵<https://www.pangaea.de/>

para descrever dados geoespaciais, enquanto o Darwin Core constitui a referência para a biodiversidade, permitindo a descrição normalizada de espécies, habitats e ocorrências. A utilização de tais padrões facilita a integração dos dados em infraestruturas internacionais e promove a sua reutilização por diferentes comunidades científicas.

Entre os repositórios internacionais, destacam-se o OBIS e o WoRMS, centrais para a harmonização taxonómica e a disponibilização global de dados de biodiversidade marinha, bem como o GBIF, que amplia o acesso à biodiversidade numa escala global. No domínio da oceanografia, o SeaDataNet e o Copernicus Marine EMS asseguram a interoperabilidade de dados físico-químicos e climáticos em redes europeias. Já o ENA (European Nucleotide Archive) responde a necessidades específicas da biologia molecular, através de padrões como o INSDC (International Nucleotide Sequence Database Collaboration) e o MIXS (Minimum Information about Any Sequence), enquanto o PANGAEA constitui uma infraestrutura interdisciplinar para dados das ciências da Terra e do mar.

Neste último caso, importa salientar que o CIIMAR conta com um investigador associado como editor do PANGAEA, o que facilita a submissão, curadoria e validação de conjuntos de dados produzidos no âmbito de projetos do centro. Esta ligação institucional reforça o potencial do PANGAEA como canal privilegiado para a partilha de dados do CIIMAR Watch, integrando-os numa infraestrutura global de referência.

3.5 Tendências na Investigação sobre Gestão de Dados Marinhos

A revisão de literatura sobre gestão de dados marinhos e ambientais partiu de uma pesquisa sistemática realizada na base de dados SCOPUS, com o objetivo de identificar os principais trabalhos publicados sobre gestão de dados no contexto da biologia marinha e oceanografia. A expressão utilizada combinou os termos "Data Management", "Marine Biology", "Ocean Data" e "Marine Data", aplicando filtros para excluir publicações editoriais ou informais, e incluir estudos com palavras-chave relacionadas com aquisição de dados, princípios FAIR, e metadados.

Com esta pesquisa foram recuperadas 297 publicações que evidenciam algumas tendências claras:

- A relevância da interoperabilidade e dos padrões internacionais de metadados (ex.: Darwin Core, ISO 19115, CF Conventions).
- A importância crescente da aplicação dos princípios FAIR como guia para a gestão e partilha de dados.
- A valorização de repositórios disciplinares (SeaDataNet, EMODnet, OBIS) enquanto infraestruturas fundamentais para a curadoria e a disponibilização de dados marinhos.
- A emergência de temas ligados a *big data*, integração multiescalar e uso de ontologias para descrever relações complexas.

Com base nesta representação foi feita a seleção de artigos que, pela sua relevância conceptual e empírica, permitiram aprofundar o estado da arte sobre gestão de dados marinhos e ambientais.

Desta pesquisa foi extraído um subconjunto de 12 publicações diretamente relevantes para a gestão de dados em biologia marinha e oceanografia. A seleção privilegiou trabalhos que documentam soluções técnicas (infraestruturas, normas, semântica), modelos organizacionais (qualidade e governação) e orientações para tornar os dados FAIR. O objetivo desta síntese é caracterizar o estado da arte com base nesses artigos, evidenciando tendências e lacunas que enquadram o presente estudo.

3.5.1 Revisão das Principais Contribuições Científicas

A literatura recente converge na adoção dos princípios FAIR e na construção de infraestruturas interoperáveis como pilares da gestão de dados marinhos. A crescente quantidade e heterogeneidade de dados marinhos, provenientes de diversas plataformas e sensores, realça a necessidade urgente destas abordagens. Em termos estratégicos, os princípios FAIR articulados com *Marine Spatial/Data Infrastructures* (MSDI) mostram-se eficazes para integrar geodados multidisciplinares, aumentar a descoberta e suportar reutilização em larga escala (Foglini et al., 2023).

A gestão de dados de imagem marinha tem recebido atenção particular, dado o seu rápido crescimento e diversidade de formatos. Schoening et al., 2022 sublinham a importância de estruturar metadados em *Image FAIR Digital Objects* (iFDOs), de forma a viabilizar a citação, a partilha e a reutilização de grandes volumes visuais. Esta preocupação com fundações sólidas – identificadores persistentes, vocabulários controlados e infraestruturas interoperáveis – revela-se indispensável para a aplicação efetiva dos princípios FAIR no quotidiano dos investigadores.

No plano operacional, diversas experiências de e-infraestruturas abertas e baseadas em normas têm sido desenvolvidas. O observatório OBSEA¹⁶ por exemplo, demonstrou uma e-infraestrutura modular e robusta que gere dados marinhos multiparamétricos em tempo real, seguindo os princípios FAIR e utilizando ferramentas de código aberto e padrões aceites pela comunidade (Martínez et al., 2023). A automação de fluxos de trabalho, que inclui a aquisição, controlo de qualidade e publicação de dados, é crucial para reduzir a fricção e aumentar a eficiência em toda a cadeia de valor dos dados em tais infraestruturas (Tanhua et al., 2019). De forma complementar, a adoção de *Image FAIR Digital Objects* (iFDOs) em infraestruturas abertas é proposta para facilitar a gestão e reutilização de imagens marinhas, reduzindo a fragmentação e a sobrecarga de gestão (Schoening et al., 2022). O observatório Koster Seafloor, por sua vez, exemplifica uma abordagem modular de código aberto para a análise de grandes volumes de vídeo subaquático, demonstrando a importância de ferramentas abertas para a gestão de imagens (Anton et al., 2021).

Partescano et al. (2021), mostram como os *workflows* do EMODnet Chemistry conseguem integrar múltiplas fontes através de controlos de qualidade e vocabulários normalizados, resultando em produtos consistentes e reutilizáveis.

¹⁶<https://www.obsea.es/>

Em paralelo, a componente semântica emerge como vetor determinante para ultrapassar a heterogeneidade disciplinar (Ren et al., 2021; Tanhua et al., 2019). Nesse sentido, o sistema DOLP (Data Ontology and List-Based Publishing), que integra uma estrutura otimizada de consulta de serviços (QSQL), antecipou a publicação e descoberta mais eficientes de serviços de dados. Este sistema utiliza modelos semânticos unificados, como a OEDO (Ocean Environmental Data Ontology), e listas pré-inferidas de serviços semanticamente relacionados para reduzir o tempo de consulta e melhorar a taxa de acesso (Ren et al., 2021). Evoluções mais recentes propõem ontologias de domínio, como a OEDO, para representar, publicar e reunir serviços de dados ambientais oceânicos, promovendo raciocínio semântico e metadados estruturados para acelerar a descoberta e integrar fontes diversas (Ren et al., 2021). Em conjunto, estes contributos apontam para a necessidade de combinar perfis de aplicação pragmáticos com camadas ontológicas que suportem “interoperabilidade semântica”.

A gestão da qualidade é tratada na literatura como uma capacidade organizacional. Leadbetter et al. (2020) descrevem o modelo de *Data Management Quality Management Systems* promovido no âmbito do IOC-IODE, enfatizando que a estabilização de processos e responsabilidades é essencial para a melhoria cumulativa da qualidade dos dados. Esta visão converge com recomendações para políticas FAIR institucionais, licenças legíveis por máquina e proveniência explícita ao longo do ciclo de vida dos dados.

A gestão de dados marinhos enfrenta desafios complexos que são reconhecidos há já algum tempo. Por exemplo, Bechini et al. (2013) já destacavam os problemas inerentes à recolha de dados *in situ* heterogêneos durante cruzeiros oceanográficos, onde a diversidade de formatos gerados por cada instrumento exige práticas sistemáticas de gestão de dados e sistemas de informação estruturados. No domínio mais específico das observações biológicas, a harmonização e reutilização de dados de longo prazo é crucial para *frameworks* de Variáveis Essenciais (EVs). Zilioli et al. (2021) exploram abordagens práticas para tornar os dados históricos de biodiversidade marinha compatíveis com as Variáveis Essenciais de Biodiversidade (EBVs) e Variáveis Essenciais Oceânicas Biológicas e de Ecossistemas (BioEco EOVs), sublinhando a necessidade de processamento de dados de múltiplas fontes para gerar índices consistentes que suportem objetivos de conservação.

Com uma visão de futuro, a construção de um ecossistema digital de dados oceânicos eficaz é um objetivo primordial. Simone Simoncelli et al. (2021) enfatizam a necessidade de um quadro colaborativo que envolva produtores, gestores e utilizadores de dados para concretizar a visão da Década dos Oceanos de um "oceano transparente e acessível". Os autores descrevem um ecossistema digital composto por elementos múltiplos, interoperáveis e escaláveis, desenhado para facilitar o acesso e a integração de dados, e para impulsionar o desenvolvimento de novas aplicações e serviços com capacidades preditivas e suporte à tomada de decisões em tempo real.

De forma transversal, Tanhua et al. (2019) destacam que a adoção de princípios FAIR requer não apenas infraestruturas técnicas, mas também mudanças culturais e políticas, incluindo a capacitação de investigadores em início de carreira e a definição de políticas institucionais de partilha. D. M. A. Schaap et al. (2021) reforçam este ponto, ao apresentarem uma visão integrada sobre

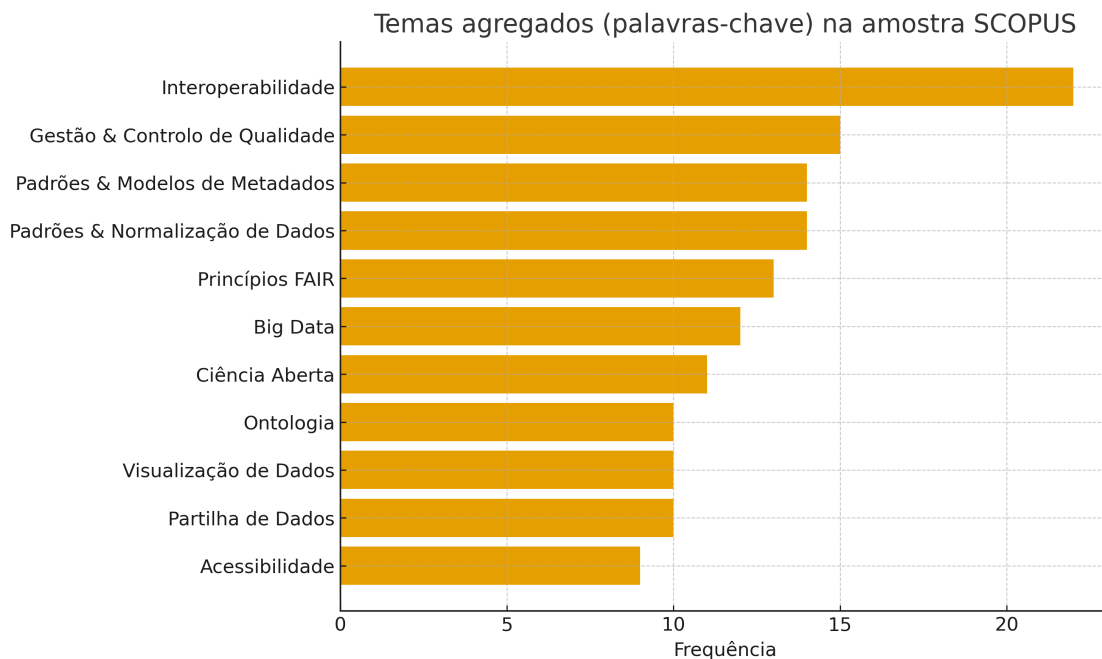


Figura 3.1: Principais temas na literatura (SCOPUS, 2025)

normas, infraestruturas e a ligação entre FAIRness e *big data* em oceanografia.

3.5.2 Resultados Bibliométricos

Para além da análise qualitativa dos 12 artigos selecionados, foi realizada uma análise quantitativa à totalidade da amostra de 297 publicações recuperadas da SCOPUS. Os gráficos seguintes sintetizam as principais tendências em termos de evolução temporal, distribuição por áreas temáticas e palavras-chave mais recorrentes. Estes resultados oferecem um enquadramento mais amplo, permitindo contextualizar o estado da arte identificado.

Temas Principais

A análise das palavras-chave (Figura 3.1) destaca a interoperabilidade como o eixo dominante, o que confirma a centralidade das normas e da integração multidisciplinar na literatura. Em segundo plano surgem temas ligados à qualidade e normalização, refletindo a necessidade de dados confiáveis e comparáveis. A presença de termos como FAIR, ciência aberta e partilha de dados demonstra que as dimensões políticas e culturais da gestão de dados ganharam relevância para além das questões técnicas. Já a emergência de termos como *big data* e visualização de dados mostra como o aumento do volume e da complexidade da informação está a reconfigurar as práticas de análise e de disseminação científica.

Produção Científica Anual (SCOPUS, 2025)

A evolução temporal (Figura 3.2) mostra uma curva inicial muito lenta até meados dos anos 1990, seguida de um crescimento contínuo e, sobretudo, um salto significativo a partir de 2005. Este aumento coincide com a consolidação das primeiras grandes infraestruturas europeias (ex.:

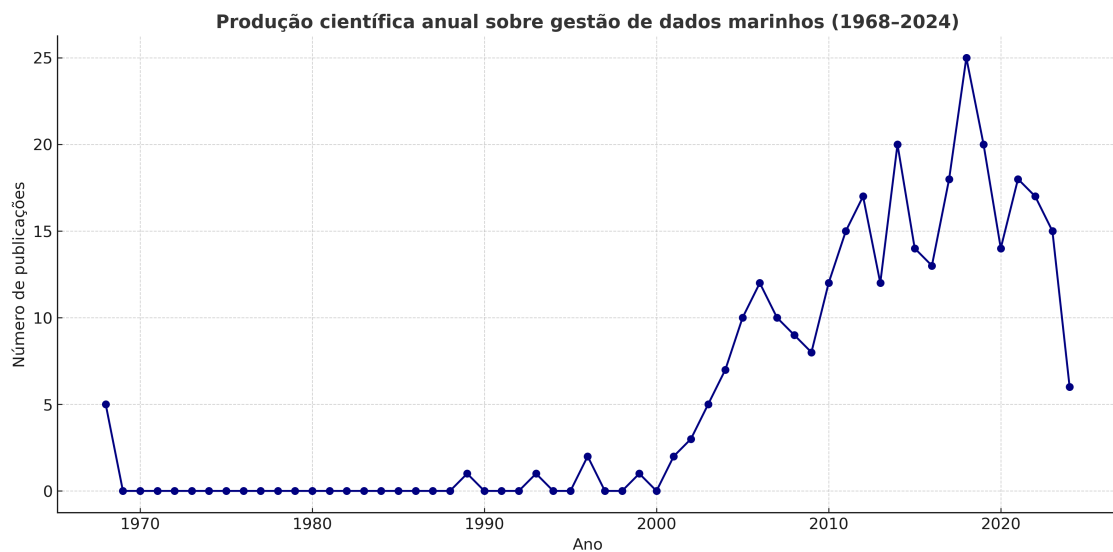


Figura 3.2: Produção científica anual (SCOPUS, 2025)

SeaDataNet) e com a intensificação do debate sobre dados FAIR. O pico em 2015 marca a maturidade destas discussões, enquanto os anos recentes confirmam uma estabilização em patamares elevados. Isto sugere que a gestão de dados marinhos deixou de ser um nicho técnico, tendo-se vindo a tornar uma área estruturante da investigação oceânica.

Produção por País

A distribuição geográfica das publicações (Figura 3.3) revela a forte centralidade dos Estados Unidos e da China, que juntos concentram mais de 20% da produção global. Estes países têm liderado a criação de infraestruturas digitais de dados marinhos e políticas de ciência aberta. A Itália surge logo a seguir, refletindo o envolvimento ativo em redes como a EMODnet e a SeaDataNet. Portugal aparece apenas com uma publicação, revelando ainda uma presença residual neste domínio, o que demonstra a necessidade de desenvolver mais investigação nesta área.

Autores Relevantes

A análise dos autores mais relevantes (Figura 3.4) evidencia um núcleo restrito de investigadores que têm contribuído de forma sistemática para o campo da gestão de dados marinhos. Destacam-se Collins, Fiolek, Giorgetti e Novellino, todos com seis artigos registados, embora com diferentes valores fracionados — o que significa que, apesar de estarem presentes em vários trabalhos, o seu peso relativo enquanto autores principais varia. Este aspeto é particularmente relevante, pois permite diferenciar entre quem coordena efetivamente linhas de investigação neste domínio e quem participa em colaborações de grande escala.

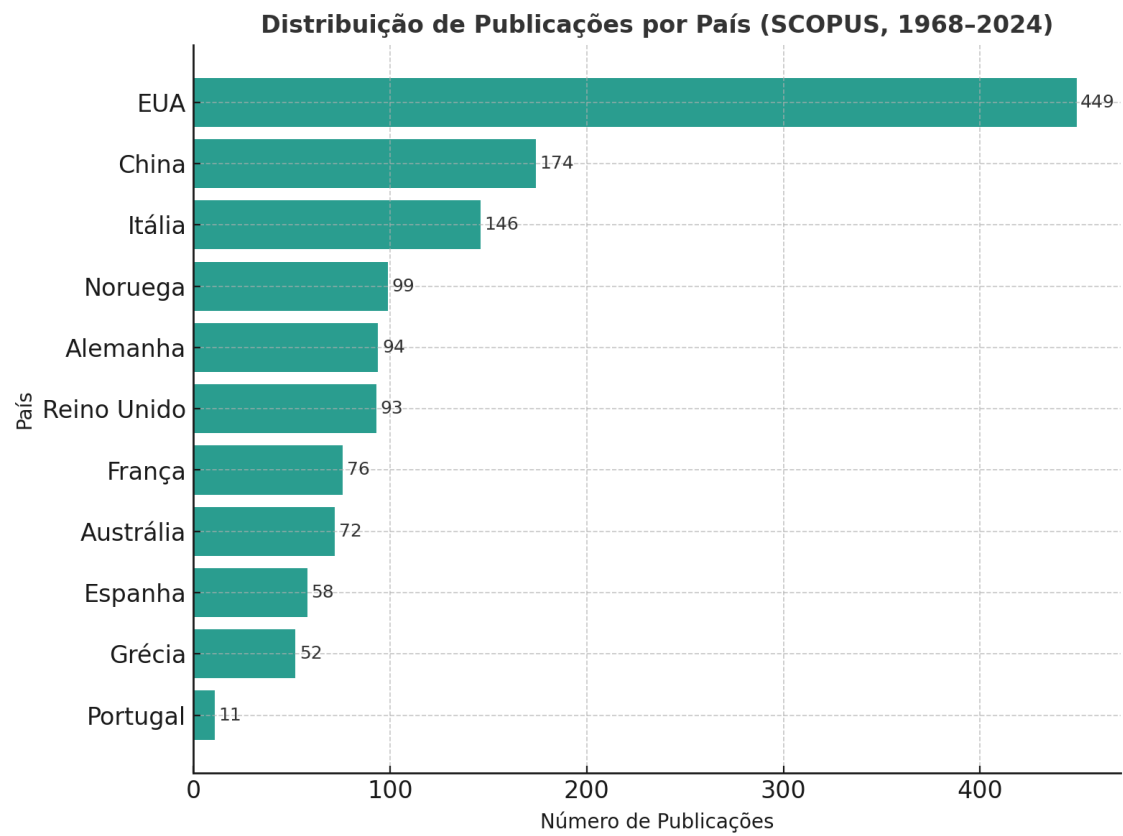


Figura 3.3: Distribuição geográfica da produção científica (SCOPUS, 2025)

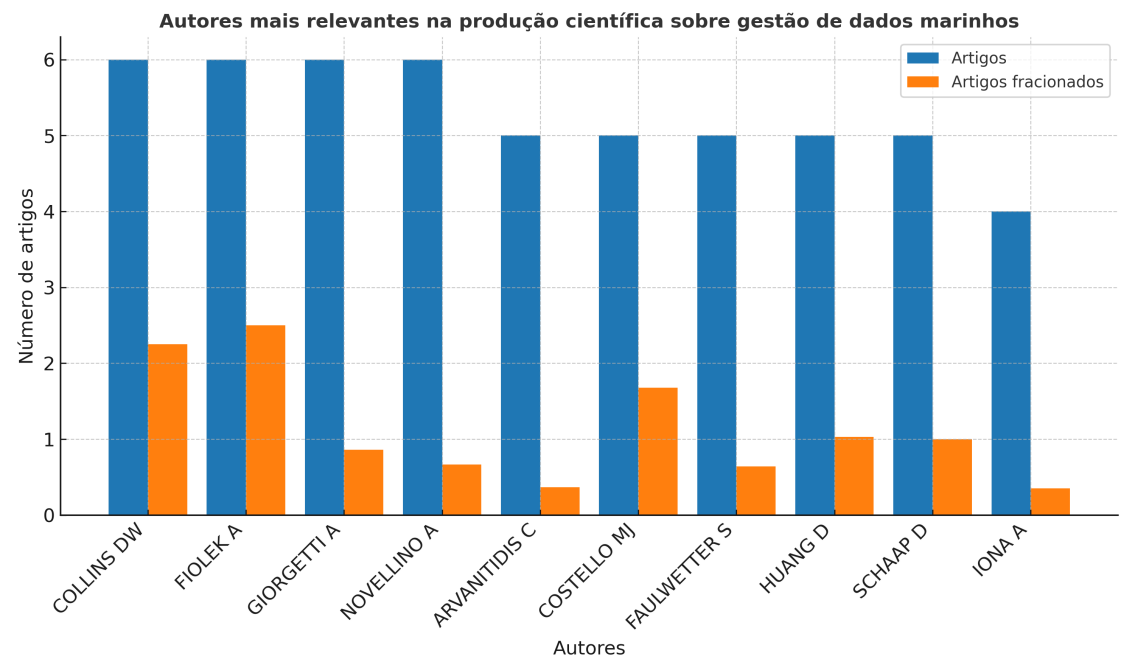


Figura 3.4: Autores mais relevantes na produção científica sobre gestão de dados marinhos (SCOPUS, 2025)

Capítulo 4

Enquadramento do CIIMAR Watch

4.1 O CIIMAR e o programa CIIMAR Watch

O Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) é uma instituição de investigação da Universidade do Porto, cujo foco principal se centra no estudo do oceano e dos ecossistemas costeiros, promovendo o desenvolvimento sustentável através da investigação científica, inovação tecnológica e literacia do oceano. A sua atuação estende-se por áreas como biodiversidade e funcionamento dos ecossistemas, alterações globais, biotecnologia azul e valorização de recursos marinhos.

Inserido neste contexto, o programa CIIMAR Watch surgiu como uma iniciativa de monitorização ambiental da costa norte de Portugal com início em 2023, reunindo uma equipa multidisciplinar de investigadores do CIIMAR com o objetivo de recolher, tratar e disponibilizar dados marinhos e ambientais em acesso aberto. O programa representa um esforço coletivo de recolha sistemática de dados físico-químicos, biológicos e ecológicos, promovendo a colaboração entre equipas de diferentes áreas científicas.

O CIIMAR Watch realiza campanhas regulares de monitorização em vários pontos da costa, abrangendo habitats costeiros, intertidais, subtidais e estuarinos - nomeadamente os estuários do Douro e do Lima. As estações de amostragem estão organizadas em duas *Super Stations* (VIANA WATCH e PORTO WATCH), com uma frequência de amostragem sazonal e capacidade para recolha de dados em tempo real, e um conjunto de estações adicionais com frequência bi-anual.

As variáveis monitorizadas variam consoante o tipo de estação e incluem, por exemplo: velocidade superficial da água, parâmetros físico-químicos (como temperatura, salinidade, oxigénio dissolvido, pH), nutrientes, composição de comunidades biológicas (plâncton, macroinvertebrados, macrófitas, peixes), contaminantes, microplásticos e dados genéticos. A diversidade de dados recolhidos reflete a natureza abrangente do programa e a sua preocupação com o estado ecológico e químico dos ecossistemas costeiros e estuarinos.

O trabalho de campo é realizado com uma periodicidade bem definida: as *super stations* são amostradas sazonalmente (quatro vezes por ano), enquanto as restantes estações seguem uma

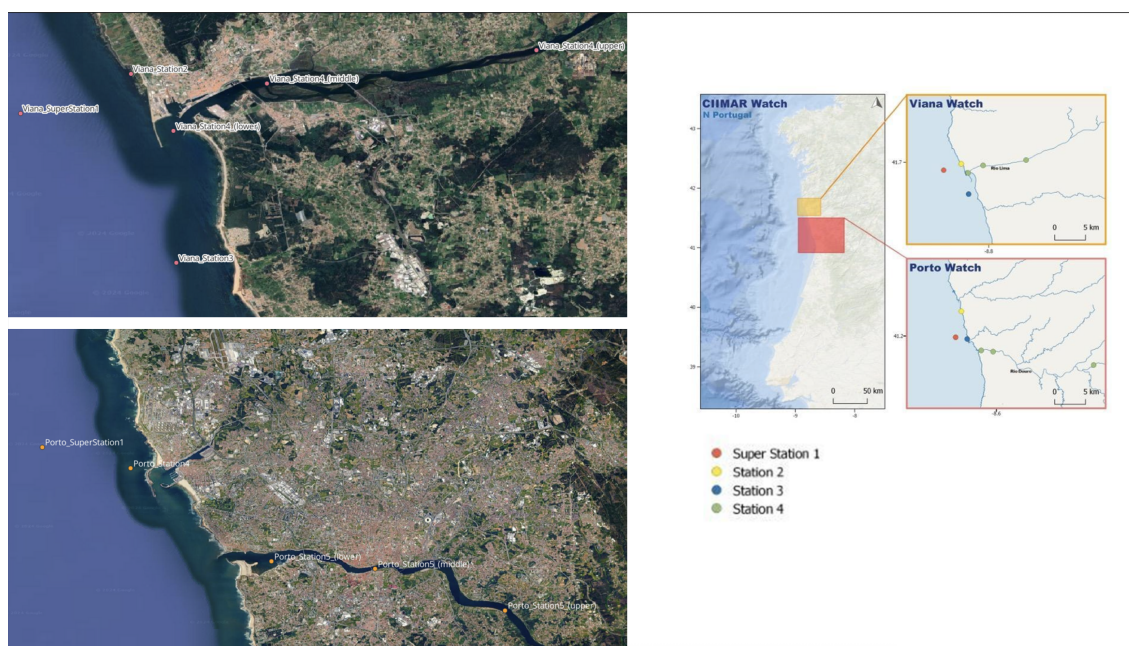


Figura 4.1: Estações de Amostragem do CIIMAR Watch

frequência bi-anual, normalmente no verão e inverno. A logística das campanhas envolve a articulação entre várias equipas e depende de fatores como marés, condições do mar e disponibilidade de embarcações e equipamentos.

O programa é coordenado por um grupo central de investigadores e técnicos, incluindo responsáveis pelas diferentes áreas de monitorização, pela gestão de dados e pela operação logística. Este modelo de governação e colaboração tem permitido ao CIIMAR Watch consolidar-se como um programa estruturante no CIIMAR, com potencial para se tornar uma referência nacional em monitorização ambiental costeira.

4.2 Estrutura, objetivos e funcionamento do programa

O programa CIIMAR Watch assenta numa estrutura colaborativa, que integra investigadores e técnicos de diferentes grupos do CIIMAR, organizados em torno das áreas de especialidade envolvidas na recolha e análise dos dados. Esta estrutura é apoiada por um modelo de gestão que assegura a coordenação logística, científica e técnica das campanhas, bem como a consolidação e partilha dos dados.

O principal objetivo do programa é garantir uma recolha sistemática, contínua e multidisciplinar de dados marinhos e ambientais ao longo da costa norte de Portugal, promovendo a sua disponibilização em acesso aberto para toda a comunidade científica e sociedade em geral. Pretende-se, assim, reforçar o conhecimento sobre o estado ecológico e químico das zonas costeiras e estuárias, apoiar políticas públicas de conservação e uso sustentável dos recursos marinhos, e fomentar a investigação interdisciplinar.

Para além da componente científica, o programa assume uma missão de literacia do oceano e sensibilização ambiental, promovendo o envolvimento de estudantes, investigadores em início de carreira e técnicos em formação. O trabalho de campo é, frequentemente, realizado coma participação ativa de equipas diversas, o que contribui para a formação prática e o desenvolvimento de competências em ciência marinha.

As campanhas de monitorização seguem um calendário regular e uma metodologia padronizada, ainda que flexível o suficiente para acomodar as especificidades de cada equipa. A amostragem obedece a protocolos definidos internamente, mas é adaptada consoante as variáveis em estudo e as condições do terreno. A organização logística das campanhas, a distribuição de tarefas e a posterior gestão dos dados recolhidos são coordenadas por um grupo central de investigadores e técnicos, em articulação com os coordenadores de cada equipa temática.

Os dados recolhidos são tratados e armazenados numa infraestrutura comum, alojada no Google Drive institucional, que funciona como repositório interno. A inserção e organização dos dados nesta plataforma são acompanhadas por um técnico e gestora de dados responsável pelas práticas de ciência aberta no CIIMAR. Embora ainda não exista um repositório institucional próprio para o programa, esta estrutura digital partilhada tem permitido centralizar a informação e assegurar uma base de trabalho comum a todas as equipas.

4.3 Multidisciplinaridade e diversidade de perfis dos investigadores

Um dos aspetos mais distintivos do CIIMAR Watch é a composição multidisciplinar das suas equipas. O programa integra investigadores de diversas áreas científicas - desde biologia marinha, ecotoxicologia e microbiologia, até à química analítica, engenharia ambiental, oceanografia e geociências - refletindo a complexidade dos sistemas costeiros e a necessidade de abordagens complementares para a sua monitorização.

A coexistência de perspetivas disciplinares distintas contribui para uma maior riqueza na interpretação dos dados, mas também levanta desafios ao nível da gestão e documentação da informação. Cada área tem as suas práticas próprias de registo, formatos de dados preferenciais e normas de validação, o que acentua a necessidade de mecanismos comuns que promovam a interoperabilidade e a coerência entre equipas.

Apesar destas diferenças, a maioria dos investigadores manifesta uma postura colaborativa, reconhecendo o valor da partilha de dados e do trabalho interdisciplinar. Esta abertura apresenta uma oportunidade para desenvolver modelos de gestão de dados mais completos e adaptáveis, que considerem não apenas requisitos técnicos, mas também as especificidades humanas e metodológicas das equipas.

Capítulo 5

Metodologia e Desenho da Investigação

5.1 Abordagem Geral

A presente investigação adota uma abordagem qualitativa, de natureza exploratória, com o objetivo de compreender as práticas de gestão e partilha de dados no contexto do CIIMAR Watch. Esta opção metodológica, assente num estudo de caso (Pickard, 2013), foi considerada a mais adequada dada a complexidade, heterogeneidade e natureza ainda pouco estudada das dinâmicas associadas à gestão de dados em programas de monitorização multidisciplinares.

Esta abordagem possibilitou compreender as perspetivas e experiências dos investigadores envolvidos no programa, explorando as suas práticas, perceções e necessidades concretas associadas à gestão dos dados recolhidos e gerados.

A estratégia adotada baseou-se na realização de entrevistas semiestruturadas, que permitiram aprofundar o conhecimento sobre processos, contextos e racionalidades dificilmente captáveis através de métodos quantitativos ou instrumentos estruturados.

5.2 Entrevistas Semiestruturadas: Instrumento e Processo

A recolha de dados empíricos foi realizada através de entrevistas semiestruturadas, conduzidas junto de investigadores e técnicos envolvidos no programa CIIMAR Watch, provenientes de diferentes áreas científicas. Esta técnica revelou-se adequada por permitir explorar em profundidade as experiências, práticas e perspetivas dos participantes, mantendo simultaneamente uma estrutura flexível que possibilitou adaptar o guião às especificidades de cada perfil.

As entrevistas foram concebidas em alinhamento com os objetivos e as questões de investigação definidos no Capítulo 1, assegurando a coerência entre o desenho metodológico e o propósito do estudo — compreender as práticas de gestão e partilha de dados no contexto do CIIMAR Watch e identificar os principais desafios, necessidades e requisitos para a sua melhoria.

As entrevistas tiveram como pilar um guião desenvolvido com base no *Data Curation Profiles Toolkit*¹, adaptado à realidade do programa em estudo. O guião foi enviado previamente a cada

¹<https://docs.lib.purdue.edu/dcptoolkit/>

participante, permitindo-lhes refletir sobre as questões abordadas e preparar exemplos concretos das suas práticas. A estrutura do guião procurou cobrir as principais etapas do ciclo de vida dos dados, desde a recolha à preservação e partilha, de modo a captar não só os procedimentos adotados, mas também as dificuldades, perceções e necessidades sentidas pelos investigadores. O guião completo encontra-se em anexo A.

As entrevistas decorreram entre março e junho de 2025, num formato misto: algumas foram realizadas presencialmente nas instalações do CIIMAR, enquanto outras ocorreram por videoconferência. Todas as entrevistas foram gravadas mediante consentimento informado dos participantes B e, posteriormente, transcritas manualmente de forma resumida, com foco nos conteúdos mais relevantes para a análise. As transcrições foram validadas pelos participantes, garantindo que refletiam fielmente as suas respostas e intenções.

A seleção dos participantes seguiu uma amostragem intencional, com o objetivo de abranger a diversidade de áreas científicas, funções e níveis de envolvimento no programa. Foram entrevistados nove participantes, incluindo alunos de doutoramento, investigadores e um técnico de investigação com funções de apoio logístico ao programa. Esta diversidade assegurou uma visão abrangente e multidisciplinar sobre as práticas e desafios associados à gestão e partilha de dados no CIIMAR Watch.

5.3 Amostra e Perfil dos Participantes

Esta secção apresenta os perfis dos investigadores participantes no estudo, organizados de acordo com a sua área científica, o(s) ecossistema(s) com que trabalham e o tipo de dados produzidos. O objetivo é oferecer uma contextualização que permita compreender a diversidade de práticas de gestão e partilha de dados no âmbito do CIIMAR Watch.

A amostra desta investigação é composta por dez participantes com diferentes funções, especializações e níveis de envolvimento no programa CIIMAR Watch. A sua seleção foi feita com base em critérios de relevância e diversidade, assegurando uma representação equilibrada entre áreas científicas, papéis funcionais e contacto direto com os dados. Esta diversidade é essencial para compreender a multiplicidade de práticas e necessidades à gestão de dados num programa multidisciplinar.

5.3.1 Investigadores por Área Científica

Os participantes distribuem-se por várias áreas científicas, refletindo a diversidade disciplinar do CIIMAR Watch:

- Ecologia (estuarina, microbiana, cetáceos e biogeoquímica)
- Ecotoxicologia
- Química Ambiental
- Oceanografia Física

Esta diversidade disciplinar permite evidenciar como os métodos de trabalho, os tipos de dados produzidos e as exigências de documentação variam significativamente entre áreas científicas. Muitos destes dados são particularmente sensíveis ao tempo, ao local e ao método de recolha, o que exige abordagens diferenciadas de registo, armazenamento, anotação e partilha. Estas diferenças constituem, precisamente, uma das questões centrais analisadas nesta dissertação.

5.3.2 Investigadores por Ecossistema

O CIIMAR Watch realiza atividades de monitorização em quatro ecossistemas principais: estuarino, intertidal, costeiro e subtidal. Alguns participantes atuam em ecossistemas específicos, enquanto outros intervêm de forma transversal em vários contextos.

Esta distribuição reforça o caráter multidimensional da investigação e sublinha a necessidade de práticas de gestão de dados adaptadas a diferentes condições ambientais, metodologias de amostragem e tipos de observações.

A *tabela 5.1* apresenta um resumo dos participantes da investigação, organizados por nome, função no programa, área científica ou técnica principal, e ecossistemas em que atuam.

Tabela 5.1: Resumo dos participantes por área, função e ecossistema

Nome	Função	Área Científica / Técnica	Ecossistema
Óscar Babé	Técnico de Investigação	Monitorização / Mergulho	Intertidal
Vânia Freitas	Investigadora Doutorada	Ecologia Estuarina	Estuarino
Catarina Magalhães	Coordenadora do Programa e Investigadora Sénior	Ecologia Microbiana	Costeiro / Subtidal
Marisa Almeida	Investigadora Principal	Química Ambiental	Estuarino
Martina Illari	Investigadora Auxiliar	Ecologia	Estuarino
Isabel Iglesias	Investigadora Sénior	Oceanografia Física	Estuarino
Raul Valente	Investigador Doutorado	Ecologia de Cetáceos	Costeiro / Subtidal
Luís Afonso	Doutorando	Ecologia Biogeoquímica	Costeiro / Subtidal
Luís Vieira	Investigador	Ecologia e Ecotoxicologia	Costeiro

5.4 Técnica de Análise dos Dados

A análise dos dados seguiu uma abordagem qualitativa, com foco na identificação de padrões, práticas correntes e desafios enfrentados pelos investigadores no contexto da gestão de dados do CIIMAR Watch. Esta abordagem permitiu uma leitura interpretativa das experiências relatadas, integrando diferentes dimensões das práticas de gestão ao longo do ciclo de vida dos dados.

As entrevistas foram transcritas manualmente com base nas gravações autorizadas. Optou-se por transcrições resumidas, centradas nos conteúdos mais relevantes para os objetivos da investigação, em vez de uma transcrição integral.

A análise foi realizada com o apoio do software CATMA (Computer Assisted Text Markup and Analysis)², que possibilitou a codificação e anotação sistemática dos dados qualitativos. A partir da marcação textual com categorias temáticas, foi possível segmentar, relacionar e reorganizar as respostas dos participantes em torno de tópicos-chave que refletem as diferentes dimensões da gestão e partilha de dados.

O processo de análise seguiu as orientações metodológicas de Miles et al. (2014), recorrendo a uma lógica iterativa e indutiva. As categorias foram desenvolvidas de forma emergente, com base nos dados recolhidos, sendo posteriormente refinadas ao longo das etapas de análise.

O procedimento decorreu em várias fases:

1. Transcrição e anotação inicial das entrevistas, com familiarização com os dados;
2. Codificação exploratória, com criação de etiquetas emergentes à medida que surgiam padrões e temas recorrentes nas respostas;
3. Revisão e reorganização das categorias, com fusão, divisão e eliminação de redundâncias;
4. Revisão das transcrições e anotações, de acordo com o procedimento estabelecido no ponto anterior, para assegurar a consistência da codificação;
5. Interpretação dos dados, com base nas categorias finais, relacionando-as com os objetivos e questões de investigação.

Ao longo da análise, foi utilizada uma tabela temática inicialmente baseada nos objetivos e questões de investigação, mas ajustada progressivamente com base nos dados recolhidos. O processo de codificação decorreu de forma iterativa: após uma leitura aberta e exploratória das entrevistas, foram criadas categorias preliminares, posteriormente revistas, fundidas ou subdivididas de acordo com a recorrência e relevância dos temas emergentes. As categorias finais resultaram deste processo de codificação sistemática e incluíram temas como o registo e a anotação dos dados, formatos, documentação, armazenamento, organização, reutilização, partilha, utilização de repositórios, metadados e preocupações expressas pelos investigadores.

As categorias temáticas apresentadas na tabela 5.2 foram definidas de forma a assegurar a correspondência direta com as questões de investigação e os objetivos do estudo. Assim, as categorias que descrevem práticas de registo, anotação, armazenamento, documentação, metadados e partilha respondem à primeira questão de investigação, centrada nas práticas atualmente utilizadas pelos investigadores do programa CIIMAR Watch.

Por sua vez, as categorias associadas a preocupações, requisitos, reutilização e preservação dos dados refletem a segunda questão de investigação, relacionada com os principais desafios e requisitos para a implementação de infraestruturas e práticas de gestão de dados mais eficazes.

²<https://catma.de/>

Esta correspondência assegura a coerência entre o processo de codificação e os objetivos analíticos, permitindo uma leitura estruturada e comparável dos resultados face às questões que orientam a dissertação.

Tabela 5.2: Grelha temática de categorias utilizadas na análise

Categoria	Descrição
Práticas de Anotação de Dados	Observações, comentários ou notas complementares feitas no momento da recolha ou tratamento dos dados.
Práticas de Armazenamento	Locais e formatos utilizados para guardar os dados (ex: discos externos, cloud, servidores).
Documentação	Descrição contextual dos dados, incluindo objetivos, metodologias e instrumentos utilizados.
Formatos	Tipos de ficheiros em que os dados são guardados (ex: .csv, .xlsx, .txt).
Metadados	Informação sobre a utilização de metadados.
Práticas de Organização de Dados	Métodos e estruturas utilizadas para organizar os dados em pastas, ficheiros ou sistemas.
Partilha de Dados	Práticas associadas à disponibilização de dados a terceiros, dentro ou fora da equipa.
Preocupações dos Investigadores	Dúvidas, receios ou limitações apontadas pelos investigadores sobre a gestão e partilha dos dados.
Técnicas de Preservação	Mecanismos e preocupações com a preservação de longo prazo dos dados.
Registo dos Dados	Métodos utilizados para registar dados em campo ou laboratório.
Utilização de Repositórios	Referência a repositórios de dados utilizados.
Requisitos dos Investigadores	Necessidades apontadas para melhorar a gestão e partilha de dados.
Reutilização de Dados	Práticas que permitem a reutilização dos dados.

O processo metodológico estabeleceu uma relação direta entre os objetivos de investigação, o guião das entrevistas e as categorias de análise. Os objetivos gerais orientaram a construção do guião de entrevista, garantindo que as questões abordadas refletissem os principais eixos de análise pretendidos. Subsequentemente, as categorias utilizadas no processo de anotação e codificação foram derivadas desses mesmos objetivos e dimensões temáticas, assegurando uma coerência metodológica entre recolha e análise dos dados.

A título ilustrativo, a Figura 5.1 apresenta um exemplo de anotação temática realizada no CATMA, com destaque para categorias como partilha de dados, condições de repositório e preocupações dos investigadores.

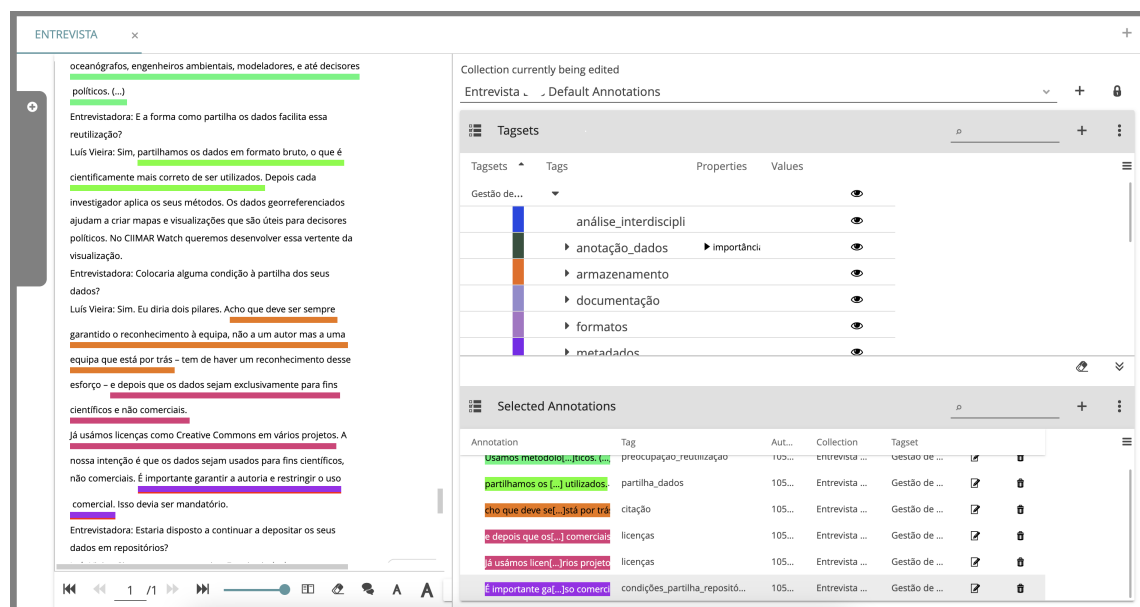


Figura 5.1: Exemplo de anotação no CATMA

Capítulo 6

Práticas e Desafios na Gestão e Partilha de Dados no CIIMAR Watch

Este capítulo apresenta os principais resultados da análise qualitativa realizada com base nas entrevistas conduzidas com investigadores e técnicos envolvidos no programa CIIMAR Watch. O objetivo é descrever as práticas, desafios e necessidades associadas à gestão, organização, documentação, partilha e reutilização dos dados de investigação no contexto deste programa multidisciplinar, procurando explicitar os obstáculos e requisitos identificados pelos participantes.

As categorias de análise emergiram do processo de codificação indutiva descrito no capítulo 5 sendo posteriormente sistematizadas na *tabela 5.2*. Estas categorias refletem as diferentes dimensões da gestão de dados identificadas nos testemunhos dos participantes e constituem a base para a estruturação dos resultados apresentados nas secções seguintes.

6.1 Resultados

6.1.1 Tipologias de dados gerados

A diversidade de áreas científicas envolvidas no CIIMAR Watch traduz-se numa ampla variedade de dados recolhidos, que refletem diferentes dimensões do ambiente marinho e costeiro. Esta tipologia abrange dados físicos, químicos e biológicos, obtidos através de metodologias de amostragem e instrumentação diversas, adaptadas aos objetivos de cada equipa.

Nos **dados físicos**, os investigadores referem a utilização de sondas multiparamétricas, CTDs (Conductivity, Temperature, Depth) e sensores *in situ*, que permitem medir variáveis como temperatura, salinidade, oxigénio dissolvido, pH, condutividade e turbidez. Foram também mencionados parâmetros como velocidade e nível de maré, sobretudo nos contextos estuarinos. Estes dados são recolhidos em campanhas de campo e exportados a partir dos equipamentos em formatos de texto simples ou compatíveis com *Excel*, sendo posteriormente tratados em programas como *R*¹ ou *Matlab*².

¹<https://www.r-project.org/>

²<https://www.mathworks.com/products/matlab.html>

Relativamente aos **dados químicos**, são recolhidas amostras de água e sedimentos para análise laboratorial de nutrientes, metais pesados e compostos orgânicos. Estes dados são registados em folhas de *Excel*, onde são calculadas e organizadas as concentrações obtidas, muitas vezes através de matrizes criadas individualmente pelos técnicos e alunos responsáveis. Segundo os testemunhos, não existe ainda um modelo uniforme para o tratamento e organização destes dados, sendo a estrutura dependente do projeto e da equipa envolvida.

No caso dos **dados biológicos**, as práticas incluem a recolha de comunidades fitoplancónicas e zooplancónicas, macroinvertebrados bentónicos, macroalgas, peixes e outros organismos, através de redes de plâncton, arrastos, armadilhas ou mergulho científico. Em campanhas subaquáticas, são também produzidos registos fotográficos (por exemplo, quadrados de 50×50 cm) para identificação de espécies e caracterização de habitats. Foram ainda mencionados dados de sequenciação genética (NGS – Next-Generation Sequencing), utilizados em estudos de ecologia microbiana, que geram ficheiros em formato *FASTA* de diferentes dimensões, associados a dados de metagenómica e metabarcoding. Estes conjuntos de dados são frequentemente articulados com variáveis ambientais, permitindo análises integradas entre dados físicos, químicos e biológicos.

Verificou-se ainda a coexistência de **dados brutos** (*raw data*) e **dados processados**, refletindo diferentes fases do ciclo de vida dos dados. Os dados brutos incluem ficheiros provenientes de equipamentos com formatos proprietários — como ficheiros de sondas ou de sequenciação (*.hex*, *.dat*, *.FASTA*). Os dados processados são frequentemente convertidos e organizados em formatos tabulares, como *.csv*, *.xlsx* ou *.txt*, ou acompanhados por documentação complementar em *.docx*. A diversidade destes formatos evidencia as múltiplas origens e naturezas dos dados recolhidos, refletindo a heterogeneidade das metodologias e a natureza multidisciplinar do programa CIIMAR Watch. Em complemento, são também produzidos **registos visuais e documentação digitalizada**, como fotografias em formatos *.jpg*, *.png* e *.tiff*, utilizadas para a identificação de espécies ou caracterização de habitats, bem como digitalizações de fichas laboratoriais e de campo em *.pdf*. Estes elementos funcionam como suporte adicional à análise e validação dos dados, garantindo a preservação de informação contextual relevante.

6.1.2 Anotação e Registo de Dados

A análise das entrevistas evidencia diversidade significativa nas práticas de **registo e anotação** entre as equipas do CIIMAR Watch, em função do contexto de trabalho e da área científica. Em termos operacionais, o registo ocorre de duas formas, muitas vezes *complementares*: (i) **registo manual** em pranchetas subaquáticas e folhas de campo (incluindo fichas específicas por campanha ou atividade de laboratório); e (ii) **registo digital** diretamente em folhas de cálculo (*Excel*) ou em ficheiros gerados pelos próprios equipamentos durante a recolha.

Nos cenários em que o registo é manual, os investigadores descrevem a passagem subsequente dessa informação para um suporte digital, preservando os originais (ou as digitalizações das fichas) e estruturando os dados segundo grelhas usadas internamente pelas equipas. Quando o registo é digital desde a origem, a recolha segue *logsheets* partilhadas ou modelos internos, nos quais

se introduzem identificadores de amostra, data, local e outros campos mínimos definidos para a atividade.

Os dados recolhidos por sensores ou equipamentos são frequentemente armazenados em cartões de memória e exportados de acordo com os critérios de organização de cada equipa. Quando recolhidos manualmente, são digitalizados ou introduzidos em folhas de cálculo, e quando recolhidos automaticamente, exportados a partir dos equipamentos em formatos adequados ao tratamento posterior.

A **anotação** surge como prática complementar ao registo: consiste em acrescentar *observações contextuais* e notas explicativas que clarificam condições de recolha e decisões metodológicas. Entre os exemplos recorrentes contam-se apontamentos sobre o *estado do tempo* e as condições do mar, observações sobre *visibilidade* ou *corrente* em mergulho, marcação de *incidentes* (avarias, imprevistos, atrasos), bem como referências a *ajustes de procedimento* realizados no terreno. Estas anotações são introduzidas em **colunas de comentários** nas folhas de cálculo, em **campos de observações** das *logsheets* ou em **notas de campo e de laboratório** associadas ao registo principal.

Vários participantes referem ainda o uso de **registos visuais** — como fotografias das áreas de amostragem, imagens das fichas de campo e registos do local — como apoio ao registo e à anotação. Estes elementos funcionam como *comprovativo* e *memória contextual*, permitindo validar posteriormente os valores registados e esclarecer dúvidas surgidas na fase de organização ou tratamento.

Os participantes distinguiram ainda entre **dados brutos** (*raw data*) — tal como saem dos equipamentos ou são recolhidos em campo — e **dados processados ou trabalhados**, sujeitos a organização, filtragem ou análise para integração em estudos ou relatórios científicos. Como um investigador referiu, “*os equipamentos fornecem dados nos seus próprios formatos, o que varia bastante*”, o que ilustra a diversidade e complexidade dos ficheiros brutos. Já os dados processados correspondem às análises realizadas posteriormente; como referiu uma das participantes que trabalha com sequenciação, “*as análises de downstreams, como curvas de diversidade ou relações entre parâmetros ambientais e biológicos, são feitas normalmente no R*”, evidenciando a transição para formatos mais estruturados e analiticamente tratados.

A Tabela 6.1 sintetiza as principais tipologias de dados gerados no âmbito do CIIMAR Watch, detalhando as suas características, métodos, formatos e práticas de registo, bem como os suportes utilizados para a sua organização. A informação apresentada reflete a diversidade metodológica e a coexistência de abordagens manuais e digitais entre as diferentes áreas científicas do programa.

Tabela 6.1: Síntese das tipologias, métodos, formatos e práticas de registo dos dados no CIIMAR Watch

Tipo de Dados	Características Principais	Métodos, Formatos e Práticas de Registo	Suportes Utilizados
Físicos	Variáveis ambientais como temperatura, salinidade, oxigénio dissolvido, pH, condutividade e turbidez; incluem também velocidade e nível de maré.	Recolha com sondas multiparamétricas, CTDs e sensores <i>in situ</i> . Exportação de ficheiros .txt, .csv, .dat e .hex. Registo automático pelos equipamentos, complementado com anotações sobre condições ambientais (vento, maré, temperatura do ar) e falhas de sensores.	Ficheiros de texto; folhas de campo; <i>logsheets</i> partilhadas.
Químicos	Dados laboratoriais de nutrientes, metais pesados e compostos orgânicos, obtidos a partir de amostras de água e sedimentos.	Processamento analítico em laboratório, com cálculos registados em folhas de <i>Excel</i> e ficheiros .csv ou .xlsx. Anotações sobre amostras, replicações, calibrações e incidentes laboratoriais.	Folhas de cálculo; protocolos e notas complementares (.docx, .pdf); digitalização de fichas.
Biológicos	Dados de comunidades fitoplactónicas e zooplactónicas, macroinvertebrados bentónicos, macroalgas, peixes e outros organismos; incluem também dados genómicos e registos visuais.	Recolha por redes de plâncton, arrastos, armadilhas e mergulho científico. Dados genómicos exportados em formato .FASTA; registos fotográficos em .jpg, .png e .tiff. Registo manual em pranchetas e folhas de campo, com anotações sobre visibilidade, corrente, fauna e substrato, incidentes e ajustes de procedimento.	Pranchetas; folhas de campo; colunas de comentários em folhas de cálculo; registos fotográficos e <i>logsheets</i> .

6.1.3 Armazenamento e Organização dos Dados

As entrevistas realizadas aos investigadores que integram o programa CIIMAR Watch permitem observar uma grande diversidade nas práticas de armazenamento e organização de dados. Embora o programa disponibilize uma estrutura partilhada no Google Drive, onde cada equipa deposita os dados recolhidos e processados, as práticas associadas ao armazenamento, organização e manutenção dos ficheiros variam entre os investigadores. Esta diversidade resulta, em parte, das diferentes áreas científicas envolvidas, das metodologias adotadas e da coexistência entre dados produzidos no âmbito do programa e dados provenientes de outros projetos.

O Google Drive do programa funciona como repositório central, com pastas estruturadas por campanhas e coordenado por um técnico responsável que acompanha o tratamento e a inserção de dados na plataforma. O projeto conta ainda com o apoio da gestora de dados do CIIMAR, responsável pela ciência aberta, que colabora com os investigadores na submissão dos dados recolhidos. Para além desta *cloud*, é comum o recurso a discos externos, computadores pessoais e servidores locais, sobretudo para guardar dados brutos, *backups* ou ficheiros de grandes dimensões.

Em vários casos, as equipas organizam as pastas do Google Drive de acordo com a estrutura das campanhas de amostragem ou por tipologia de dados (físicos, químicos e biológicos). Alguns investigadores referem que criam subpastas específicas para o tratamento e validação, de modo a manter a separação entre versões preliminares e dados consolidados. Outros mencionam que esta estrutura é replicada localmente, em discos externos ou computadores pessoais, para facilitar o trabalho *offline* e a partilha seletiva com colaboradores. Contudo, nem todas as equipas seguem a mesma lógica de organização, e muitas utilizam sistemas próprios de pastas, com convenções de nomeação implícitas e dependentes da experiência individual.

Alguns investigadores descrevem também uma separação informal entre dados “ativos” (em uso corrente) e dados “arquivados” (de campanhas anteriores ou já analisados). Em certos casos, esta distinção traduz-se em pastas designadas como “trabalho” e “arquivo”, mas sem uma política definida para migração de ficheiros. Outros investigadores mantêm apenas as versões mais recentes na *cloud*, guardando versões antigas em dispositivos locais ou servidores do CIIMAR, o que gera múltiplos locais de armazenamento para o mesmo conjunto de dados.

Apesar desta variabilidade, há equipas que demonstram níveis elevados de organização, com estruturas de diretórios padronizadas, convenções internas para nomeação de ficheiros e registos de alterações documentados em folhas de cálculo partilhadas. Alguns participantes referiram o uso de logs ou listas de controlo para acompanhar atualizações, identificando quem realizou alterações e quando. Outros reconheceram que essa prática é feita de forma mais informal, por exemplo, através da criação manual de ficheiros duplicados com versões numeradas (“versao_final”, “versao_corrigida”, etc.), o que leva à coexistência de múltiplas versões do mesmo *dataset* sem rastreamento automatizado.

As práticas de preservação também variam: a maioria das equipas digitaliza fichas de campo e armazena as versões digitalizadas no Google Drive ou em discos externos, mantendo, quando possível, os originais físicos em arquivo. Estas digitalizações, geralmente em formato .pdf ou

.jpg, são consideradas uma forma segura de garantir a durabilidade da informação. Ainda assim, alguns participantes salientam a dificuldade de conservar documentos em papel e a inexistência de mecanismos automáticos de cópias de segurança, o que leva à dependência de práticas individuais para prevenir perdas.

Outros, no entanto, manifestaram preocupações em relação à limitação de espaço nas plataformas utilizadas, sobretudo no caso de ficheiros de grandes dimensões, como imagens subaquáticas ou dados de sequenciação, que muitas vezes não podem ser totalmente carregados para o Google Drive. Alguns investigadores referem que, por esse motivo, mantêm cópias locais ou guardam parte dos dados em discos externos, o que implica uma gestão manual dos ficheiros e a necessidade de criar múltiplas versões para diferentes finalidades. Em vários casos, estas versões são armazenadas em pastas distintas, com nomes semelhantes ou repetidos, sem um controlo automatizado de alterações, o que pode levar à coexistência de diferentes ficheiros para o mesmo conjunto de dados. Também foi referida a inexistência de um sistema de cópias de segurança automático, levando à dependência de práticas individuais para garantir a preservação dos dados e aumentando o risco de perda ou duplicação de informação.

Estas preocupações foram expressas de forma recorrente pelos participantes, conforme ilustram os excertos seguintes:

“Preocupa-me a possibilidade de os dados originais serem alterados ou apagados, porque muitas pessoas têm acesso aos ficheiros.”

“O meu maior medo é perder os dados. As folhas, às vezes, digitalizam-se e depois desaparecem, porque são reutilizadas — por exemplo, nas campanhas de mergulho. É difícil guardar papel físico. Mas uma versão digitalizada é mais fácil de armazenar. Ainda assim, é bom manter os originais sempre que possível.”

Verificou-se, assim, que a gestão e o armazenamento dos dados combinam soluções institucionais e práticas individuais, resultando numa organização heterogénea e adaptada às necessidades de cada equipa.

A Tabela 6.2 sintetiza as principais práticas de armazenamento e organização identificadas entre os investigadores do CIIMAR Watch, destacando a coexistência de soluções institucionais e individuais, bem como as limitações mais frequentemente mencionadas em relação à gestão e preservação dos dados.

Tabela 6.2: Síntese das práticas de armazenamento e organização dos dados no CIIMAR Watch

Componente	Práticas observadas	Limitações e preocupações relatadas	Suportes utilizados
Infraestrutura partilhada (cloud)	Google Drive organizado por campanhas, coordenado por um técnico responsável; utilizado para deposição de dados processados e documentação associada.	Limitação de espaço; múltiplos acessos simultâneos; ausência de controlo de versões e de rastreamento automático de alterações.	Google Drive (pastas partilhadas por campanhas).
Armazenamento local e backup	Utilização de discos externos, computadores pessoais e servidores locais para guardar dados brutos, cópias de segurança e ficheiros de grandes dimensões.	Dispersão de dados entre dispositivos; risco de perda, duplicação ou desatualização de versões; ausência de rotina de sincronização.	Discos externos; computadores pessoais; servidores do CIIMAR.
Estrutura e organização de pastas	Pastas organizadas por campanhas, tipo de dados ou anos de recolha; algumas equipas seguem convenções internas de nomeação e registo de alterações.	Estruturas e critérios não uniformes entre equipas; convenções de nomeação implícitas ou pessoais.	Pastas partilhadas; diretórios locais; logs de alterações internos.
Separação entre dados ativos e arquivados	Dados correntes mantidos em diretórios de trabalho; dados de campanhas anteriores arquivados noutras pastas ou dispositivos.	Distinção informal e nem sempre refletida na estrutura das pastas; falta de critérios definidos para migração de dados.	Google Drive; discos externos; servidores locais.
Preservação e segurança	Digitalização de fichas de campo e armazenamento das versões digitalizadas; manutenção pontual de originais em papel.	Dificuldade em conservar documentos físicos; risco de eliminação accidental; preocupação com integridade dos ficheiros digitais.	Ficheiros digitalizados (.pdf, .jpg); originais físicos.

6.1.4 Documentação e Metadados

A análise revelou que os investigadores produzem documentação de apoio aos dados, sobretudo através de relatórios, notas de campo, *logsheets* ou ficheiros complementares (como os *README*), que asseguram um enquadramento mínimo para a compreensão dos *datasets*. Estes documentos incluem habitualmente informação sobre o local e data de amostragem, parâmetros medidos, responsáveis pela recolha, instrumentos utilizados e eventuais observações contextuais. Apesar disso, não existe uma prática sistemática ou normalizada, sendo comum que a documentação varie de acordo com a experiência individual ou as exigências específicas de cada projeto.

Um dos participantes destacou a importância de registos claros durante as campanhas, referindo: *“Tenho README associados com os dados, das campanhas (...) Nas campanhas é necessário ficar bem registado, por isso ficam bem documentados.”* Outro investigador descreveu a mesma preocupação, mas motivada por requisitos específicos de projeto: *“Nesse caso fiz um ficheiro bastante bem documentado, com o metafile, e tentei pôr o mais claro possível.”* Estes testemunhos ilustram a diversidade de práticas, desde anotações informais até formas mais estruturadas de documentação, elaboradas em ficheiros *.txt*, *.docx* ou *.pdf*, frequentemente armazenados junto dos conjuntos de dados.

No que respeita aos metadados, as perspetivas são igualmente diversas. Para alguns investigadores, a sua organização é feita de forma interna e adaptada ao funcionamento da equipa: *“São estruturas internas. Os metadados são organizados pelos gestores de dados (dos projetos).”* Outros reconhecem que, na prática, os metadados são arquivados de forma automática, sem grande intervenção manual: *“Sim, os metadados são arquivados da forma como foram gerados pelas sondas.”*

Em termos de padronização, observou-se que apenas uma parte dos participantes recorre a estruturas formais de metadados, normalmente quando os dados são partilhados em projetos externos ou em plataformas internacionais. Nestes casos, foram mencionados modelos baseados em *templates* fornecidos por projetos como o EMODnet ou o SeaDataNet, ou em esquemas simplificados inspirados em normas como o ISO 19115 ou o Darwin Core, embora adaptados internamente. A maioria, contudo, recorre a descrições livres, sem correspondência a esquemas ou vocabulários controlados.

Nem sempre a criação e submissão de metadados é percebida como uma tarefa simples. Como descreveu um investigador, *“depositar dados com metadados não é fácil — é preciso seguir categorias específicas, usar ontologias específicas, e às vezes processar os dados no R. Requer bastante conhecimento técnico.”* Por outro lado, alguns investigadores relataram esforços ativos de organização e normalização interna, sobretudo em contextos específicos como as campanhas de amostragem de água: *“Temos uma organização interna na equipa para os metadados, especialmente para as amostragens de água, onde associamos vários metadados aos resultados. Mas não seguimos nenhum modelo formal ou standard internacional — usamos uma estrutura interna própria ou as que nos são dadas nos projetos específicos.”*

De forma transversal, foi ainda reforçada a importância de garantir que todos os conjuntos

de dados sejam acompanhados por descrições estruturadas, sendo que, como afirmou um dos participantes, *“esses dados têm de estar acompanhados por metadados bem padronizados.”* Em conjunto, estes testemunhos evidenciam uma consciência clara da relevância dos metadados, ainda que coexistam limitações práticas e soluções *ad hoc*.

Importa referir que, em vários testemunhos, emergiu também uma certa confusão conceptual entre documentação e metadados. Enquanto a documentação é entendida como um conjunto de descrições contextuais ou relatórios que acompanham os dados, os metadados deveriam constituir registos estruturados e normalizados. No entanto, alguns participantes tendem a utilizar os dois termos para referência à mesma prática, refletindo a ausência de distinção formal no contexto do programa e a dependência de práticas individuais para a descrição e contextualização dos dados.

6.1.5 Reutilização e Partilha de Dados

A análise das entrevistas evidencia uma diversidade de práticas no que diz respeito à partilha e reutilização de dados. Embora exista um reconhecimento generalizado da importância de disponibilizar dados em acesso aberto, as estratégias utilizadas pelos investigadores revelam-se heterogêneas e, por vezes, pouco sistemáticas.

No que respeita à partilha, alguns participantes referem a utilização de repositórios institucionais ou generalistas, ainda que de forma pontual. Um participante afirmou que utiliza o *Zenodo* e faz uso das suas licenças, indicando uma preocupação em garantir condições de reutilização explícitas. Contudo, a maioria admite que a partilha em repositórios de dados ainda não é uma prática regular:

“Depositar não tanto, utilizo mais para ir buscar. Já descarreguei dados do EMOD-net, para os modelos numéricos preciso de forçamentos que vamos buscar a essas bases de dados. Para depositar, para já ainda não.”

“Sempre que posso, tento publicar os dados, mas não faço isso de forma sistemática.”

Na ausência de submissão formal, a partilha tende a ocorrer de forma mais direta e localizada, sobretudo através de plataformas colaborativas internas: *“Quando partilho é mais direto com um colega ou assim.”*

Alguns investigadores relacionaram diretamente a partilha com a origem do financiamento público, sublinhando que os dados devem ser entendidos como um bem coletivo. Como afirmou um participante:

“Os dados não são nossos, são financiados pela FCT, mas os dados não são nossos, por isso é muito importante a partilha. Se há um embargo no início por estudos pendentes ou muito inovadores, também percebo isso e é muito importante, mas uma vez ultrapassado, têm de ser partilhados. Principalmente atualmente, com as leis da Europa.”

A **reutilização de dados** é reconhecida como uma possibilidade relevante, mas que depende fortemente da forma como estes são documentados e disponibilizados. Como salientado por um investigador: *“Os dados em bruto precisam sempre de uma explicação associada, caso contrário podem parecer caóticos. Por isso, durante o processamento, organizamo-los de forma clara e compreensível, por exemplo em Excel e com colunas que facilitam a análise.”* Esta preocupação é reiterada por outro participante, que refere: *“Procuramos apresentar os dados de forma clara, especialmente os valores finais. A raw data é mais complexa porque precisa de ser processada, mas deixamos os ficheiros disponíveis (...)”*

Foram igualmente apontadas preocupações quanto à gestão e utilização dos dados. Um participante salientou que *“a principal preocupação é que a nossa própria equipa consiga encontrar e compreender os dados”*, sublinhando a dificuldade em assegurar a organização e a legibilidade interna antes mesmo de considerar a partilha externa.

Apesar destas fragilidades, existem exemplos de boas práticas, como a publicação de *raw data* em bases de dados internacionais: *“Atualmente, publicamos sobretudo a raw data, para garantir que os resultados sejam reproduzíveis. Os métodos usados precisam de ser aplicados à raw data para reproduzir os mesmos resultados.”*

Repositórios de Dados: Experiências, Barreiras e Expectativas —

As entrevistas revelam que a utilização de repositórios de dados entre os investigadores do CIIMAR Watch é ainda pouco homogênea, oscilando entre experiências pontuais de depósito, práticas informais de partilha e a ausência total de submissão formal. Um dos participantes evidencia a adoção de repositórios generalistas e a preocupação com a aplicação de licenças abertas.

Noutro testemunho, é referida a utilização regular de plataformas internacionais consideradas de referência:

“(...) utilizamos as plataformas que achamos ser as melhores da Europa, as mais usadas, que têm mais arquivos e capacidade para receber, são resilientes. Aquelas que são também, de alguma forma, sugeridas nos projetos europeus. (...) Nós publicamos muitos dados nas bases de dados sem os publicar antes [em artigo]. Eu acho que é melhor correr-se o risco de alguém publicar aquela nossa questão científica, porque eles podem usar os dados para outras questões científicas.”

Contudo, nem todos os investigadores utilizam repositórios de dados de forma sistemática. Alguns admitem utilizar repositórios para consultar dados, mas não com o propósito de depositar os seus próprios conjuntos. Em vários casos, a partilha limita-se a contextos colaborativos próximos, ficando os dados fora de circuitos institucionais ou públicos, como admite um participante: *“Quando partilho é mais direto com um colega ou assim.”* No entanto, há alguma consciência da importância da partilha de dados além do próprio contexto de investigação, como refere outro participante: *“Até há pouco tempo, os dados que eu tinha estavam processados e eram apresentados nos artigos publicados, portanto os dados brutos eram guardados por mim e não eram disponibilizados para mais nada. Recentemente ganhei consciência disso e ando a organizar coisas antigas de forma a publicá-las no Zenodo.”*

A análise das entrevistas permite identificar dois caminhos principais. Por um lado, os repositórios generalistas, como o *Zenodo*, são valorizados pela sua facilidade de utilização, integração com publicações e visibilidade imediata, mas a sua utilização surge sobretudo em iniciativas individuais. Por outro, os repositórios disciplinares ou específicos — referidos genericamente como “*bases de dados internacionais*” — são reconhecidos como espaços adequados para a curadoria e partilha de dados marinhos, ainda que a sua adoção não seja sistemática.

As entrevistas revelam que a utilização de repositórios de dados entre os investigadores do CIIMAR Watch é ainda pouco homogênea, oscilando entre experiências pontuais de depósito, práticas informais de partilha e a ausência total de submissão formal. Um dos participantes evidencia a adoção de repositórios generalistas e a preocupação com a aplicação de licenças abertas, mencionando explicitamente o uso do *Zenodo* para garantir condições claras de reutilização dos dados.

Enquanto alguns investigadores referem já ter depositado dados em plataformas abertas, a maioria admite que ainda não o faz de forma sistemática, utilizando os repositórios sobretudo para descarregar informação relevante para os seus próprios estudos. Como referido por um dos participantes, “*utilizo mais para ir buscar do que para depositar*”, demonstrando que, embora exista familiaridade com o conceito de repositório, a prática de partilha de dados próprios permanece incipiente.

Entre aqueles que já partilham dados, a motivação principal associa-se à transparência e ao reconhecimento de que os dados resultam de financiamento público, e por isso devem ser considerados um bem coletivo. Esta perceção é expressa de forma clara no testemunho de um dos participantes, que afirmou: “*Os dados não são nossos, são financiados pela FCT (...), é muito importante a partilha.*”

Nos restantes casos, a partilha tende a ocorrer de forma direta e informal, sobretudo entre colaboradores próximos, fora de plataformas institucionais ou públicas: “*Quando partilho é mais direto com um colega ou assim.*” Esta tendência reforça a ideia de que, no contexto atual do programa, a partilha é ainda fortemente dependente de redes pessoais de confiança e da iniciativa individual de cada investigador.

Entre as barreiras mais frequentemente mencionadas estão a falta de tempo para preparar os dados para depósito, a ausência de orientações claras sobre onde e como o fazer, e a perceção de que os conjuntos de dados necessitam de tratamento ou validação antes da publicação. Estas dificuldades ajudam a explicar a adoção desigual de práticas de partilha estruturada e a preferência por soluções mais simples, como o envio direto de ficheiros entre colegas.

Apesar disso, alguns participantes revelam uma evolução nas suas perceções e práticas, relatando esforços recentes de organização e retrospectiva de dados antigos com vista à sua publicação em plataformas abertas. Um dos testemunhos ilustra bem essa mudança de postura: “*Até há pouco tempo, os dados que eu tinha estavam processados e eram apresentados nos artigos publicados (...). Recentemente ganhei consciência disso e ando a organizar coisas antigas de forma a publicá-las no Zenodo.*”

A análise global demonstra, assim, que a utilização de repositórios generalistas como o *Zenodo* surge associada a iniciativas individuais e a uma crescente consciência sobre a importância da disponibilização aberta. Já os repositórios disciplinares ou especializados são referidos apenas de forma genérica, designados como “*bases de dados internacionais*”, e associados sobretudo a equipas com maior experiência em projetos europeus ou colaboração externa. Embora a adoção ainda não seja sistemática, observa-se uma tendência gradual de abertura e de valorização da partilha como componente essencial da investigação desenvolvida no âmbito do CIIMAR Watch.

6.1.6 Práticas, Desafios e Requisitos

As entrevistas realizadas no âmbito do CIIMAR Watch revelam uma grande diversidade nas práticas de gestão de dados, bem como nos desafios e requisitos identificados pelos investigadores.

No que respeita às **práticas**, observam-se estratégias variadas de registo e armazenamento. Alguns investigadores recorrem a soluções simples de cópia de segurança, enquanto outros descrevem esquemas mais complexos, combinando discos externos com plataformas institucionais. Um participante, por exemplo, referiu: “*Normalmente tenho um ou mais discos externos onde faço cópias. Para além disso, partilho com o responsável daquele ecossistema — ele também guarda os dados noutros discos. E também estão no Google Drive, para estarem sempre acessíveis.*” Também são frequentes cópias em suportes portáteis, como ilustrado pela afirmação: “*Cópias em pen drives, especialmente em (discos) SSDs — tentamos evitar discos mecânicos por causa do risco de falha.*” Esta multiplicidade de estratégias demonstra a preocupação com a redundância e segurança, mas evidencia igualmente a ausência de uma política comum de gestão.

Entre os principais **desafios**, destacam-se a acessibilidade e a preservação a longo prazo. Um investigador observou: “*Por exemplo, se eu quisesse encontrar dados de temperatura de um verão de 2024, encontrava rapidamente, mas alguém de fora ia demorar mais.*” Outro acrescentou que a fiabilidade depende fortemente da validação metodológica: “*Na minha área, a fiabilidade dos dados depende muito da forma como foram obtidos e da validação dos métodos.*” A limitação de espaço de armazenamento é outra dificuldade recorrente: “*Há limitações de espaço nas drives — especialmente com os ficheiros de raw data, que são bastante pesados.*” Em síntese, os testemunhos revelam que os investigadores enfrentam constrangimentos que vão da organização interna à infraestrutura técnica, comprometendo a gestão sustentável dos dados e a sua reutilização futura.

No plano dos **requisitos**, emergem várias propostas que apontam para soluções mais estruturadas e sustentáveis. Um participante sugeriu que “*faria sentido cada investigador ter um espaço onde armazenasse os dados até à standardização dos formatos*”, enquanto outro sublinhou a importância da usabilidade, defendendo “*um repositório em que os dados sejam facilmente acessíveis e intercomunicáveis, com metadados bem ligados aos dados e com uma interface fácil de utilizar.*” Também a adoção de identificadores persistentes é considerada crucial — “*Identificadores persistentes — DOIs — são essenciais*” — assim como a padronização dos metadados: “*Esses dados têm de estar acompanhados por metadados bem padronizados.*” A necessidade de reconhecimento académico é igualmente valorizada: “*Fundamental garantir o reconhecimento*

dos autores, com citação adequada nos trabalhos que utilizem os dados compartilhados.” Estas observações revelam uma clara preocupação em alinhar as práticas locais com normas internacionais, garantindo simultaneamente a valorização dos investigadores e o reconhecimento do seu contributo.

Por fim, a **colaboração** surge como uma necessidade transversal. Como sintetizou um dos participantes: “A comunicação é muito importante, (...) conversamos muito porque são estudos multidisciplinares e acabas por precisar dos dados todos.” Este testemunho confirma que a gestão e a necessidade de partilha dos dados no CIIMAR Watch não são apenas questões técnicas, mas também processos sociais e colaborativos, essenciais para a produção de conhecimento integrado e multidisciplinar.

6.2 Discussão dos Resultados

A análise das entrevistas demonstra que a gestão de dados no CIIMAR Watch é marcada por uma forte heterogeneidade, reflexo da natureza multidisciplinar do programa. Essa diversidade traduz-se em práticas que oscilam entre soluções informais e boas práticas mais estruturadas, mas em todos os casos os testemunhos revelam desafios comuns e necessidades partilhadas. Esta pluralidade de abordagens reflete não apenas a diversidade disciplinar, mas também diferentes níveis de maturidade em termos de literacia de dados, o que reforça a necessidade de um enquadramento institucional capaz de promover coerência sem inibir a autonomia científica. Deste modo, a heterogeneidade observada deve ser entendida como fragilidade, mas como um ponto de partida para a construção de práticas partilhadas.

A coexistência de dados brutos e processados, bem como a valorização de dados auxiliares e contextuais, confirma que a gestão de dados marinhos exige uma atenção particular à documentação e rastreabilidade. A dificuldade em acompanhar as transformações aplicadas aos dados, relatada pelos investigadores, reflete um risco já assinalado por Borgman (2012) e Whyte et al. (2011): sem informação clara sobre proveniência e processamento, a transparência e a reprodutibilidade ficam comprometidas. Este resultado sugere uma lacuna estrutural que não se resolve apenas com boas práticas individuais, mas requer políticas internas de curadoria e normalização, à semelhança do que defendem Whyte et al. (2011) e Cox; Verbaan (2018). Assim, torna-se evidente que a reprodutibilidade dos dados depende tanto da infraestrutura técnica como da cultura organizacional.

A variabilidade nas práticas de anotação — desde registos informais em cadernos até descrições detalhadas em ficheiros *README* — mostra a ausência de uma cultura comum no programa. Os testemunhos de que “sem o contexto anotado os dados podem tornar-se incompreensíveis ou difíceis de reutilizar” alinham-se com os alertas de Corti et al. (2014) e de Higman et al. (2019), que identificam a documentação contextual como fator decisivo para a qualidade dos dados. O reconhecimento, por parte dos investigadores, de que o “contexto anotado” é essencial revela uma consciência prática da importância dos metadados, ainda que sem um enquadramento formal. Este

tipo de consciência empírica pode constituir um ponto de partida valioso para a criação de um serviço de *data stewardship* interno, sugerindo uma maturidade incipiente mas promissora no que respeita à gestão de dados.

A coexistência de múltiplas soluções — desde o Google Drive partilhado até discos externos e *pen drives* — revela uma dispersão que fragiliza a preservação e a rastreabilidade. O receio de perda ou alteração dos dados (“preocupa-me a possibilidade de os dados originais serem alterados ou apagados”) é consistente com estudos que apontam a insegurança dos sistemas de armazenamento como uma das principais preocupações dos investigadores (Sheikh et al., 2023; Cox; Kennan et al., 2019). Por outro lado, a experiência positiva de investigadores que já implementam convenções internas mostra que boas práticas locais podem servir de base para a criação de normas institucionais. Este equilíbrio entre práticas informais e estratégias consolidadas evidencia que a normalização não precisa de ser imposta de forma rígida, mas pode evoluir a partir de práticas espontâneas e eficazes.

A sobreposição entre os conceitos de documentação e metadados, identificada nos testemunhos, é também observada na literatura (Borgman, 2012; Connie, Clare et al., 2019). A dificuldade em distinguir ambos reflete-se em práticas desiguais, onde os metadados são por vezes produzidos apenas quando exigidos em projetos ou repositórios. No entanto, a consciência partilhada de que “esses dados têm de estar acompanhados por metadados bem padronizados” revela uma abertura para a adoção de modelos normalizados, em linha com normas como o ISO 19115 ou o Darwin Core, fundamentais para a interoperabilidade internacional (G. M. R. Manzella et al., 2022). Deste modo, embora o conhecimento terminológico ainda seja limitado, observa-se uma clara disposição para integrar práticas mais estruturadas.

Alguns investigadores expressaram uma postura favorável à partilha de dados, sobretudo porque esta é frequentemente obrigatória em projetos financiados. Esta obrigatoriedade, mais do que um impulso ético, revela uma adesão normativa à ciência aberta. Contudo, a conformidade pode funcionar como catalisador de mudança cultural: à medida que os investigadores integram estas práticas nos seus fluxos de trabalho, a partilha tende a ser progressivamente interiorizada como parte do processo científico. Enquanto Tenopir; Allard et al. (2011) e Sayogo et al. (2013) identificam resistências estruturais à partilha, o caso do CIIMAR Watch demonstra que a imposição de requisitos institucionais pode funcionar como motor de transformação, aproximando comportamentos individuais das normas FAIR. Este resultado reforça a ideia de que a mudança cultural na gestão de dados é gradual e depende de incentivos institucionais claros.

Esta visão está em consonância com as orientações internacionais e europeias. Em 2007, a OCDE defendia que os dados de investigação financiados com recursos públicos deveriam ser disponibilizados em acesso aberto, reconhecendo os benefícios científicos, económicos e sociais dessa prática (OECD, 2007). No contexto europeu, o programa Horizon 2020 e, mais recentemente, o Horizon Europe, tornaram obrigatória a elaboração de planos de gestão de dados e incentivaram a partilha em conformidade com os princípios FAIR (Directorate-General for Research and Innovation (European Commission), 2018). Também em Portugal, a FCT (Fundação

para a Ciência e Tecnologia)³ tem vindo a alinhar-se com estas diretrizes, promovendo políticas de ciência aberta que colocam a partilha de dados como pilar estratégico. Deste modo, o contexto institucional português insere-se num movimento internacional mais amplo, que procura consolidar a cultura da transparência e da interoperabilidade.

Ao admitir a legitimidade de embargos temporários — por exemplo, em casos de estudos inovadores ou ainda em curso —, os investigadores demonstram uma compreensão equilibrada das exigências de abertura. Esta posição reflete o que a literatura tem identificado como uma condição comum para facilitar a adesão à ciência aberta: permitir períodos transitórios de embargo que salvaguardem os interesses científicos imediatos, sem comprometer a disponibilização futura dos dados (Tenopir; Dalton et al., 2015; Sayogo et al., 2013). No contexto do CIIMAR Watch, a ausência de uma infraestrutura de dados não é sinónimo de desorganização, mas sim de um estágio intermédio de maturidade organizacional. A existência de práticas dispersas, mas conscientes, sugere que o próximo passo deve ser a institucionalização de um serviço de gestão de dados que capitalize o conhecimento tácito já existente. Assim, observa-se um cenário de transição, em que a vontade individual e a pressão institucional convergem para a consolidação de uma estratégia de dados mais coesa.

A utilização pontual de repositórios, como o Zenodo, e a confiança em plataformas disciplinares europeias revelam duas tendências complementares: a valorização de soluções generalistas pela sua facilidade de utilização e a preferência por infraestruturas disciplinares pelo seu alinhamento com projetos internacionais. O testemunho de que “é melhor correr-se o risco de alguém publicar aquela nossa questão científica, porque eles podem usar os dados para outras questões” ilustra um compromisso com a ciência aberta que ultrapassa o receio de competição. Esta atitude revela uma compreensão mais madura da partilha, em que o valor coletivo do conhecimento supera o benefício individual. Esta visão encontra eco em iniciativas como o OBIS ou o SeaData-Net, que defendem a abertura dos dados como forma de maximizar o seu valor científico (G. M. R. Manzella et al., 2022).

De forma transversal, os resultados confirmam tendências já assinaladas pela literatura internacional. A fragmentação das práticas e a falta de tempo e recursos são obstáculos recorrentes (Cox; Pinfield, 2014; Sheikh et al., 2023), enquanto a necessidade de repositórios intuitivos e de reconhecimento académico (“fundamental garantir o reconhecimento dos autores”) refletem preocupações comuns a várias disciplinas (Tenopir, 2015). Deste modo, a ênfase na colaboração — “acabamos por precisar dos dados todos, por isso é muito importante comunicar” — reforça a ideia de que a gestão de dados deve ser entendida não apenas como questão técnica, mas também social, dependente de comunicação, confiança e responsabilidade partilhada entre equipas (Connie, Clare et al., 2019).

De forma a consolidar a análise apresentada, a Tabela 6.3 sintetiza as principais evidências empíricas e a respetiva interpretação crítica. Esta sistematização permite observar, de forma integrada, as correspondências entre as categorias de análise, os resultados obtidos e os significados que emergem da discussão. A leitura conjunta destes elementos evidencia a coerência entre as

³<https://www.fct.pt/>

práticas observadas no contexto do CIIMAR Watch e as tendências apontadas pela literatura internacional sobre gestão de dados de investigação.

Tabela 6.3: Síntese dos principais resultados e respetiva interpretação crítica.

Tema / Categoria	Principais Resultados	Interpretação Crítica / Discussão
Registo e anotação	Registos informais e ausência de padronização.	Evidencia a necessidade de políticas internas e de formação em literacia de dados; a consciência empírica do valor do contexto pode servir de base à criação de boas práticas comuns.
Documentação e metadados	Confusão frequente entre documentação e metadados.	Indica um estágio inicial de maturidade; existe abertura para normalização e adoção de normas como o ISO 19115 ou o Darwin Core.
Armazenamento e preservação	Utilização de múltiplas soluções dispersas (Google Drive, discos externos, pen drives).	Fragilidade na rastreabilidade e risco de perda de informação; boas práticas locais podem ser institucionalizadas para reforçar a preservação e a coerência.
Partilha e acesso	Partilha de dados ocorre sobretudo por obrigatoriedade de financiadores.	Reflete uma adesão normativa à ciência aberta que, embora não espontânea, pode gerar mudança cultural e consolidar práticas FAIR.
Barreiras e desafios	Falta de tempo, recursos e apoio institucional; receio de competição científica.	Obstáculos de natureza cultural e organizacional, mais do que técnica; relevância de políticas de incentivo e reconhecimento académico.

Capítulo 7

Recomendações para a Gestão de Dados no CIIMAR Watch

Com base nos resultados alcançados na Secção 6.1 e na literatura analisada, foi elaborado um conjunto de recomendações que visam orientar a consolidação de práticas de gestão de dados no CIIMAR Watch. Estas recomendações procuram responder diretamente às fragilidades identificadas nas entrevistas — nomeadamente a ausência de normas consistentes, as dificuldades de anotação, a dispersão dos locais de armazenamento e as barreiras à partilha e reutilização —, alinhando-se com as boas práticas internacionais em gestão de dados de investigação.

As propostas aqui apresentadas articulam-se em dois níveis complementares: (i) recomendações gerais, de carácter transversal e aplicáveis a todo o programa; e (ii) recomendações específicas, centradas na interoperabilidade e na definição de um conjunto mínimo de metadados adaptado ao contexto multidisciplinar do CIIMAR Watch. Desta forma, as medidas propostas respondem diretamente às questões de investigação que orientaram o estudo: compreender as práticas de gestão de dados atualmente adotadas pelos investigadores e identificar os desafios, necessidades e requisitos que poderão orientar a criação de infraestruturas adequadas de gestão e partilha.

Estas recomendações têm como objetivo apoiar a criação de um ecossistema institucional de gestão de dados sustentado por princípios FAIR e práticas de curadoria contínua. Pretende-se, assim, contribuir para a transição de um modelo informal e fragmentado — identificado nas entrevistas como reflexo da diversidade disciplinar e da ausência de mecanismos de uniformização — para uma estrutura institucionalizada, baseada em políticas claras, suporte técnico especializado e incentivos à partilha responsável.

7.1 Recomendações Gerais

As recomendações gerais incidem sobre aspetos organizacionais e estruturais essenciais ao fortalecimento de uma cultura institucional de gestão de dados. Propõem-se as seguintes linhas de ação:

- **Integração do ciclo de vida dos dados na gestão institucional:** reconhecer que os dados percorrem várias fases — da recolha à partilha e reutilização — e que todas devem ser planeadas, documentadas e acompanhadas de forma sistemática. Recomenda-se a inclusão da gestão de dados nos procedimentos padrão do programa, articulando-a com as etapas de planeamento e execução científica.
- **Implementação de um serviço de *data stewardship*:** criar uma estrutura permanente de apoio técnico e organizacional dedicada à gestão de dados. Este serviço deverá oferecer acompanhamento na elaboração de planos de gestão de dados (DMPs), apoiar a adoção de padrões de metadados, promover a curadoria dos datasets e assegurar a articulação com repositórios institucionais e temáticos.
- **Normalização e documentação de procedimentos:** desenvolver manuais e guias internos que estabeleçam convenções comuns para nomeação de ficheiros, descrição de variáveis, formatos e armazenamento. A uniformização destes aspetos é essencial para garantir rastreabilidade e consistência entre equipas e campanhas.
- **Preservação e *backups*:** definir uma política de redundância (regra 3-2-1) que assegure cópias locais e externas, bem como a verificação periódica da integridade dos dados. A política deve incluir responsabilidades definidas e cronogramas de atualização.
- **Partilha responsável e transparente:** adotar políticas que promovam o depósito de dados em repositórios abertos após períodos de embargo justificados, acompanhadas de licenças claras de utilização (ex.: CC-BY) e de identificadores persistentes (ex.: DOI). Recomenda-se que cada dataset partilhado inclua documentação mínima e instruções de citação.
- **Promoção da reutilização e colaboração:** criar mecanismos de reconhecimento dos autores dos dados (citação, coautoria em casos relevantes, métricas de utilização) e incentivar a reutilização responsável. Recomenda-se a capacitação contínua dos investigadores, com formações regulares sobre boas práticas de anotação, metadados, repositórios e princípios FAIR.

Estas recomendações constituem a base para a consolidação de práticas institucionais no CIIMAR Watch. A criação de um serviço de *data stewardship*, em particular, deve ser entendida como elemento central desta estratégia, articulando os restantes eixos de ação e promovendo a integração entre investigadores, dados e infraestruturas digitais.

Para além destas medidas estruturantes, a consolidação de uma política de gestão de dados no CIIMAR Watch requer um conjunto de ações complementares de natureza prática e formativa. Entre as prioridades identificadas destacam-se a criação de um serviço de *data stewardship* no CIIMAR, com funções de apoio técnico e coordenação institucional, a adoção progressiva de cadernos de laboratório eletrónicos (ELNs) que assegurem registo estruturado e rastreabilidade, e a elaboração de um manual interno de boas práticas desenvolvido de forma colaborativa entre as equipas do programa.

Estas medidas devem ser acompanhadas de iniciativas de capacitação contínua, dirigidas a investigadores em diferentes fases da carreira, com enfoque na literacia em ciência aberta, princípios FAIR e práticas de curadoria de dados. Recomenda-se ainda a realização de testes-piloto do modelo de metadados proposto, avaliando a sua aplicabilidade em diferentes tipologias de dados e a sua integração em repositórios internacionais. Conjuntamente, estas ações constituem um plano operativo para a institucionalização da gestão de dados no CIIMAR Watch e para a consolidação de uma cultura de partilha responsável e sustentável.

7.2 Interoperabilidade, Perfis de Aplicação e Ontologias

Os resultados do Capítulo 6 evidenciaram fragilidades significativas na gestão de dados do CIIMAR Watch, em particular a ausência de convenções comuns de anotação, a variabilidade nos métodos de documentação e a dispersão dos locais de armazenamento. Estas lacunas comprometem a clareza, a rastreabilidade e a reutilização dos dados, demonstrando a necessidade de adotar uma abordagem mais estruturada e partilhada.

A centralidade da interoperabilidade decorre não apenas de exigências técnicas, mas também das políticas de ciência aberta promovidas pela FCT e pela União Europeia. Como sublinhado por um dos investigadores, “os dados não são nossos, são financiados pela FCT (...), por isso é muito importante a partilha”. Esta visão é consistente com as recomendações internacionais que defendem o acesso aberto a dados financiados publicamente, assegurando transparência, reprodutibilidade e equidade no acesso ao conhecimento científico.

No contexto europeu, iniciativas como a EMODnet, SeaDataNet e Blue-Cloud¹ têm vindo a reforçar a interoperabilidade entre infraestruturas marinhas, incorporando tecnologias de computação em nuvem, *big data analytics* e integração com a European Open Science Cloud (EOSC)². A participação de programas como o CIIMAR Watch em redes desta natureza permitiria aumentar a visibilidade e o impacto dos dados produzidos, garantindo simultaneamente a sua compatibilidade técnica e semântica com sistemas internacionais.

A interoperabilidade pode ser desenvolvida em três níveis complementares:

1. **Interoperabilidade técnica**, que assegura o uso de formatos normalizados e protocolos de partilha compatíveis (ex.: OGC, NetCDF, ISO 19115);
2. **Interoperabilidade semântica**, baseada na adoção de vocabulários controlados (ex.: WoRMS, CF Conventions, NERC Vocabulary Server) que garantem coerência terminológica entre domínios científicos;
3. **Interoperabilidade organizacional**, que requer políticas institucionais claras, mecanismos de coordenação entre equipas e integração com plataformas externas.

A construção de um perfil de metadados para o programa pode ser entendida em três níveis: ao nível do **conceito**, identificam-se os elementos essenciais que precisam de ser registados; ao

¹<https://blue-cloud.org/>

²<https://eosc.eu/eosc-about/>

nível da **descrição**, esses elementos são operacionalizados em campos específicos, alinhados com padrões internacionais; e, num nível mais avançado, estes elementos podem evoluir para uma **ontologia**, capaz de formalizar não apenas os atributos, mas também as relações entre dados, amostras, locais e métodos. Esta evolução permitiria alcançar a chamada *interoperabilidade semântica*, ampliando as possibilidades de integração e reutilização em redes internacionais.

Este enquadramento está em consonância com o **Singapore Framework for Dublin Core Application Profiles** (Nilsson et al., 2008), que estrutura a definição de perfis em três camadas — *foundation standards*, *domain standards* e *application profiles*. A aplicação deste modelo ao CIIMAR Watch permitirá adaptar normas amplamente utilizadas, como o Darwin Core e o ISO 19115, às necessidades específicas do programa, promovendo a consistência interna e a compatibilidade externa.

Para além da normalização, recomenda-se o alinhamento estratégico do CIIMAR Watch com as principais infraestruturas europeias de dados marinhos e com as iniciativas emergentes da European Open Science Cloud. Estas plataformas podem servir de base para o desenvolvimento de protótipos aplicados ao programa, potenciando a colaboração interdisciplinar e preparando o CIIMAR Watch para futuras iniciativas de grande escala.

Neste contexto, a operacionalização da interoperabilidade no CIIMAR Watch exige a definição de um modelo de descrição de dados que, sem perder simplicidade, assegure consistência, rastreabilidade e alinhamento com normas reconhecidas internacionalmente. A criação de um *Minimal Metadata Set* representa o primeiro passo nesse processo: um núcleo de elementos essenciais que permite uniformizar a documentação e facilitar a integração dos dados em repositórios externos e infraestruturas colaborativas. Este conjunto constitui, assim, uma solução pragmática e adaptável, capaz de traduzir os princípios FAIR em práticas concretas de registo e curadoria no contexto multidisciplinar do programa.

7.3 Proposta de um Conjunto Mínimo de Metadados para o CIIMAR Watch

Os resultados desta investigação evidenciam a necessidade de um instrumento que assegure a documentação mínima, mas essencial, dos dados produzidos no CIIMAR Watch. As entrevistas revelaram práticas de anotação pouco sistemáticas, ausência de orientações partilhadas e dificuldade em garantir que os dados permanecem compreensíveis e reutilizáveis por terceiros. Estas limitações comprometem a interoperabilidade e a aplicação efetiva dos princípios FAIR, reforçando a pertinência de definir um conjunto mínimo de metadados aplicável a todo o programa.

A proposta de um *Minimal Metadata Set* surge, assim, como uma primeira iteração de um **perfil de aplicação leve**, concebido para harmonizar práticas de documentação no contexto do CIIMAR Watch. Este conjunto constitui um núcleo comum que pode ser progressivamente expandido, servindo de base para o desenvolvimento de um perfil de aplicação completo, alinhado

com normas internacionais como o Darwin Core, a ISO 19115 e as CF Conventions. O objetivo é promover uma adoção gradual, garantindo que a estrutura proposta se adapta à realidade multidisciplinar do programa e exige um esforço mínimo adicional dos investigadores.

A definição deste conjunto assenta em três fundamentos principais:

1. **Resultados empíricos desta investigação:** a necessidade de maior clareza na anotação, registo e organização dos dados foi reiteradamente sublinhada pelos investigadores, destacando a importância de informações contextuais como local, data, método e condições de recolha.
2. **Normas internacionais de referência:** elementos fundamentais de padrões como o Darwin Core (ex.: identificador, taxonomia, local, data), ISO 19115 (ex.: coordenadas, formatos, responsável) e CF Conventions (ex.: variáveis ambientais) foram considerados como base de adaptação.
3. **Princípios FAIR e boas práticas de RDM:** assegurar que todos os dados do CIIMAR Watch possam ser encontrados, compreendidos e reutilizados em contextos internacionais, reforçando a interoperabilidade com repositórios externos e plataformas europeias.

O *Minimal Metadata Set* organiza-se em quatro blocos de informação — Identificação, Contexto e Proveniência, Dados Essenciais de Recolha e Estrutura e Formato —, sintetizados na Tabela 7.1.

Este conjunto de campos representa um ponto de partida comum para todos os investigadores do CIIMAR Watch, constituindo a base de uma política de documentação coerente e sustentável. A sua aplicação permitirá uniformizar práticas, reforçar a interoperabilidade e facilitar a integração futura dos dados em repositórios disciplinares internacionais, como o SeaDataNet ou o OBIS.

Deve, contudo, ser entendido como um **modelo em evolução**, suscetível de atualização após teste e à medida que forem definidos requisitos mais específicos ou ampliados os tipos de dados recolhidos.

Tabela 7.1: Conjunto mínimo de metadados proposto para o CIIMAR Watch.

Bloco	Campos e descrição
Identificação do Dataset	Título: nome descritivo e consistente com o conteúdo do dataset; Responsável: pessoa ou equipa responsável pela recolha e curadoria; Data de criação: data de produção ou compilação do dataset; Identificador persistente: quando aplicável, atribuição de DOI ou outro identificador único que assegure rastreabilidade e citação.
Contexto e Proveniência	Descrição breve: resumo do conteúdo e finalidade do dataset; Objetivo da recolha: enquadramento científico e finalidade dos dados; Projeto associado: referência ao projeto ou campanha no âmbito do qual os dados foram gerados; Contactos: informação de contacto do responsável, equipa ou instituição detentora dos dados.
Dados Essenciais de Recolha	Local: coordenadas geográficas (latitude/longitude) e/ou topónimo; Data/hora: registo normalizado segundo o padrão ISO 8601; Método/protocolo de amostragem: descrição do procedimento adotado, incluindo referências a guias ou normas internas; Equipamento utilizado: identificação dos instrumentos ou sensores usados na recolha dos dados.
Estrutura e Formato	Tipo de dados: classificação segundo a natureza da variável (física, química, biológica, etc.); Formato de ficheiro: especificação do formato digital (ex.: CSV, NetCDF, XLSX); Variáveis principais: nome e descrição das variáveis observadas ou medidas; Unidades de medida: de acordo com o Sistema Internacional (SI) ou convenções da área; Verões e alterações: indicação de revisões, atualizações ou correções relevantes.

Capítulo 8

Conclusão e Trabalho Futuro

A presente dissertação teve como propósito analisar as práticas de gestão e partilha de dados de investigação no âmbito do programa CIIMAR Watch, procurando compreender como investigadores de diferentes áreas científicas lidam com aspetos fundamentais do ciclo de vida dos dados: registo, anotação, armazenamento, documentação, metadados, partilha e reutilização.

Partindo das questões de investigação formuladas no Capítulo 1 — (i) identificar as práticas de gestão de dados atualmente adotadas pelos investigadores do programa e (ii) reconhecer os principais desafios, necessidades e requisitos para a melhoria das práticas de gestão e partilha, considerando as orientações descritas na literatura e as especificidades do CIIMAR Watch —, foi possível construir um retrato aprofundado da realidade do programa e das suas necessidades emergentes.

Os resultados evidenciaram uma grande heterogeneidade disciplinar, refletida em práticas distintas de registo e anotação, frequentemente dependentes da experiência pessoal ou da tradição da área científica de origem. Esta diversidade confirma que as práticas de gestão de dados no CIIMAR Watch são marcadas por uma forte descentralização e ausência de orientações comuns, o que conduz a soluções informais e locais, muitas vezes eficazes a curto prazo, mas pouco sustentáveis no longo prazo. A coexistência de métodos formais e informais — desde cadernos de campo e folhas de campanha até ficheiros digitais dispersos — demonstra que o programa possui um capital científico rico, mas carece de uma estrutura institucional que permita transformar esse conhecimento individual em práticas partilhadas e sistematizadas.

No que respeita ao armazenamento e organização, constatou-se a predominância de soluções híbridas, combinando o Google Drive institucional com discos externos, computadores pessoais e dispositivos portáteis. Embora esta multiplicidade de suportes assegure alguma redundância, gera igualmente riscos para a preservação e integridade dos dados, bem como dificuldades na rastreabilidade e na reprodutibilidade. A ausência de políticas formais de *backup* e controlo de versões foi identificada pelos próprios investigadores como uma vulnerabilidade estrutural. Esta constatação reforça a importância de políticas institucionais que incorporem práticas de preservação planeada, definindo responsabilidades e calendários de atualização.

Relativamente à documentação e aos metadados, verificou-se uma sobreposição conceptual

que dificulta a distinção entre ambos. Muitos investigadores produzem documentação contextual (ex.: relatórios, ficheiros README, notas de campo), mas a criação de metadados estruturados e normalizados ainda não é prática generalizada. Apesar disso, existe uma consciência crescente de que a padronização é essencial para assegurar a interoperabilidade e a reutilização dos dados, especialmente em contextos colaborativos. A adoção de normas como o Darwin Core, o ISO 19115 ou as CF Conventions é reconhecida como necessária, ainda que o seu uso dependa de apoio técnico e de formação específica.

As entrevistas revelaram também que os investigadores enfrentam desafios de natureza sobretudo organizacional e cultural — e não apenas técnica. A falta de tempo, de formação em gestão de dados e de apoio institucional adequado limita a adoção de boas práticas, enquanto o receio de uso indevido e a ausência de mecanismos de reconhecimento reduzem a predisposição para partilhar dados. Estas barreiras refletem a necessidade de um enquadramento mais sólido, capaz de equilibrar a proteção dos investigadores com a abertura dos dados. Os resultados indicam ainda que a partilha tende a ocorrer de forma localizada e informal (ex.: entre colegas de confiança), sendo pontual a submissão em repositórios generalistas (como o Zenodo) ou disciplinares. Contudo, alguns investigadores demonstraram já uma clara predisposição para disponibilizar dados em acesso aberto, mesmo antes da publicação de artigos, sinalizando uma mudança cultural gradual em direção a práticas mais abertas e colaborativas.

Neste contexto, as recomendações propostas no Capítulo 7 respondem diretamente aos desafios e requisitos identificados. A criação de um serviço de *data stewardship* no CIIMAR surge como uma prioridade estruturante, permitindo oferecer apoio técnico, formativo e organizacional aos investigadores e garantindo a coerência das práticas entre equipas e campanhas. A par disso, a adoção de cadernos de laboratório eletrónicos (ELNs), a elaboração de um manual de boas práticas e a realização de ações de formação contínua representam passos concretos para uniformizar procedimentos, aumentar a literacia em ciência aberta e consolidar uma cultura de curadoria de dados. O desenvolvimento e teste do *Minimal Metadata Set* proposto nesta dissertação constitui igualmente um marco essencial para operacionalizar a interoperabilidade e traduzir os princípios FAIR em práticas aplicáveis ao contexto multidisciplinar do programa.

Em síntese, a primeira questão de investigação permitiu confirmar que as práticas de gestão de dados no CIIMAR Watch são diversas, pouco uniformizadas e fortemente dependentes das dinâmicas individuais e disciplinares. A segunda questão revelou que os principais desafios e necessidades se situam ao nível da ausência de políticas institucionais e necessidade de formação contínua. As respostas a ambas convergem na identificação de um caminho de transformação: a transição de uma gestão fragmentada e informal para uma estrutura institucional de apoio, capaz de assegurar a qualidade, preservação e partilha responsável dos dados de investigação.

O contributo central desta dissertação consiste, assim, em mapear e sistematizar práticas, desafios e requisitos no contexto de um programa multidisciplinar de monitorização marinha. Para além de caracterizar empiricamente as experiências dos investigadores, foi possível:

- identificar barreiras comuns à gestão e partilha de dados, de natureza técnica, organizacional e cultural;

- propor recomendações alinhadas com normas internacionais e princípios FAIR;
- desenhar um conjunto mínimo de metadados adaptado ao CIIMAR Watch, concebido como instrumento prático de uniformização e ponto de partida para a interoperabilidade com infraestruturas internacionais.

Reconhecem-se, no entanto, algumas limitações do estudo. A amostra de nove participantes, embora representativa da diversidade disciplinar do programa, não cobre a totalidade das áreas científicas envolvidas. A inexistência de observação direta das práticas de gestão em contexto de campo ou laboratório deveu-se a constrangimentos logísticos, relacionados com a disponibilidade dos investigadores e a frequência das campanhas. Este tipo de observação poderia, contudo, ter contribuído para enriquecer a análise, permitindo triangulação com os testemunhos recolhidos e uma compreensão mais detalhada das práticas efetivamente adotadas no momento da recolha, organização e anotação dos dados.

Estas limitações abrem caminho para trabalho futuro. Entre as perspetivas mais relevantes, destacam-se:

- a implementação de um serviço de *data stewardship* no CIIMAR, capaz de oferecer apoio técnico, formativo e organizacional aos investigadores;
- a adoção de cadernos de laboratório eletrónicos (ELNs) que assegurem registo estruturado, rastreabilidade e preservação de dados desde a fase inicial da investigação;
- a elaboração de um manual de boas práticas, desenvolvido de forma colaborativa, adaptado às especificidades multidisciplinares do programa;
- a realização de ações de formação contínua, direcionadas a investigadores em diferentes estágios de carreira, promovendo competências em ciência aberta, FAIR e gestão de dados;
- a experimentação prática do modelo de metadados delineado nesta dissertação, avaliando a sua aplicabilidade em diferentes tipologias de dados e a sua integração em repositórios internacionais (ex.: OBIS, SeaDataNet, PANGAEA).

Em suma, este trabalho representa um passo estruturante na consolidação de uma cultura de gestão de dados sistemática, colaborativa e alinhada com a ciência aberta no CIIMAR Watch. As recomendações e propostas aqui apresentadas pretendem servir de base prática para o desenvolvimento de soluções futuras, adaptáveis às necessidades do programa. Ao abrir caminho para práticas mais consistentes, esta dissertação contribui não apenas para o fortalecimento da investigação realizada no CIIMAR, mas também para a inserção do centro em redes internacionais de partilha e valorização de dados científicos.

Referências

- AMORIM, Ricardo Carvalho; CASTRO, João Aguiar; ROCHA DA SILVA, João; RIBEIRO, Cristina, 2017. A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Universal Access in the Information Society* **online** vol. 16, n.º 4, pp. 851–862 [acedido em 2025-01-03]. ISSN 1615-5289, ISSN 1615-5297. URL DOI: [10.1007/s10209-016-0475-y](https://doi.org/10.1007/s10209-016-0475-y) (ver p. 13).
- ANDRIKOPOULOU, Angeliki; ROWLEY, Jennifer; WALTON, Geoff, 2022. Research Data Management (RDM) and the Evolving Identity of Academic Libraries and Librarians: A Literature Review. *New Review of Academic Librarianship* **online** vol. 28, n.º 4, pp. 349–365 [acedido em 2025-08-24]. ISSN 1361-4533. URL DOI: [10.1080/13614533.2021.1964549](https://doi.org/10.1080/13614533.2021.1964549). Publisher: Routledge _eprint: <https://doi.org/10.1080/13614533.2021.1964549> (ver p. 10).
- ANTON, Victor; GERMISHUYS, Jannes; BERGSTRÖM, Per; LINDEGARTH, Mats; OBST, Matthias, 2021. An open-source approach to analysing large amounts of subsea movie data for marine ecological research: the Koster Seafloor Observatory. *Biodiversity Data Journal*. Vol. 9, e60548. URL DOI: [10.3897/BDJ.9.e60548](https://doi.org/10.3897/BDJ.9.e60548) (ver p. 23).
- ASHIQ, Muhammad; REHMAN, Shafiq Ur; MUJTABA, Ghulam, 2020. Future challenges and emerging role of academic libraries in Pakistan: a phenomenology approach. *Information Development*, pp. 1–16. URL DOI: [10.1177/0266666919897410](https://doi.org/10.1177/0266666919897410) (ver p. 8).
- ASHIQ, Murtaza; USMANI, Muhammad Haroon; NAEEM, Muhammad, 2020. A systematic literature review on research data management practices and services. *Global Knowledge, Memory and Communication* **online** vol. 71, n.º 8/9, pp. 649–671 [acedido em 2025-01-05]. ISSN 2514-9342. URL DOI: [10.1108/GKMC-07-2020-0103](https://doi.org/10.1108/GKMC-07-2020-0103). Publisher: Emerald Publishing Limited (ver pp. 10, 11).
- BAWDEN, David; ROBINSON, Lyn, 2022. *Introduction to Information Science*. 2ª ed. Facet Publishing (ver p. 6).
- BECHINI, Alessio; VETRANO, Anna, 2013. Management and storage of in situ oceanographic data: An ECM-based approach. *Information Systems*. Vol. 38, n.º 4, pp. 485–496. URL DOI: [10.1016/j.is.2012.10.004](https://doi.org/10.1016/j.is.2012.10.004) (ver p. 24).
- BEJA, Joana; VANDEPITTE, Leen; BENSON, Abigail; VAN DE PUTTE, Anton; LEAR, Dan; DE POOTER, Daphnis; MONCOIFFÉ, Gwenaëlle; NICHOLLS, John; WAMBIJI, Nina; MILOSLAVICH, Patricia; GEROVASILEIOU, Vasilis, 2022. Data services in ocean science with a focus on the biology. Em: MANZELLA, Giuseppe M.R.; NOVELLINO, Antonio (eds.). *Ocean Science Data Collection, Management, Networking and Services*. Amsterdam: Elsevier, pp. 21–44 (ver pp. 16, 19).

- BERMAN, Elizabeth A., 2017. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM: integrated findings to develop research data services. *Journal of eScience Librarianship*. Vol. 6, n.º 1, e1104. URL DOI: [10.7191/jeslib.2017.1104](https://doi.org/10.7191/jeslib.2017.1104) (ver p. 8).
- BIRKBECK, Gail; NAGLE, Tadhg; SAMMON, David, 2022. Challenges in research data management practices: a literature analysis. *Journal of Decision Systems*. Vol. 31, n.º sup1, pp. 153–167. URL DOI: [10.1080/12460125.2022.2074653](https://doi.org/10.1080/12460125.2022.2074653) (ver pp. 8, 10, 11).
- BORGMAN, Christine L., 2012. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* **online** vol. 63, n.º 6, pp. 1059–1078 [acedido em 2025-01-21]. ISSN 1532-2890. URL DOI: [10.1002/asi.22634](https://doi.org/10.1002/asi.22634) (ver pp. 6, 7, 49, 50).
- BRINEY, Kristin, 2015. *Data Management for Researchers: Organize, maintain and share your data for research success*. Pelagic Publishing (ver p. 7).
- BURGI, P.; COLLEAGUES, 2017. Challenges in Research Data Management: A Swiss Perspective. *Data Science Journal*. Vol. 16, n.º 1, pp. 1–10. URL DOI: [10.5334/dsj-2017-012](https://doi.org/10.5334/dsj-2017-012) (ver p. 8).
- CASTRO, João Aguiar, 2020. *Engaging researchers in research data management: creating metadata models for multi-domain dataset description*. **urlalso**: <https://hdl.handle.net/10216/129396>. Tese de doutoramento (ver pp. 8, 14, 15).
- CHIWARE, Elisha RT; BECKER, D., 2018. Research Data Management Services in Southern Africa: A Readiness Survey of Academic and Research Libraries. *African Journal of Library, Archives and Information Science*. Vol. 28, n.º 1, pp. 1–16 (ver p. 8).
- CONNIE, CLARE; CRUZ, Maria; PAPADOPOULOU, Elli; SAVAGE, James; TEPEREK, Marta; WANG, Yan; WITKOWSKA, Iza; YEOMANS, Joanne, 2019. *Engaging Researchers with Data Management* (ver pp. 1, 10, 11, 14, 50, 51).
- CORTI, Louise; EYNDEN, Veerle Van den; BISHOP, Libby; WOOLLARD, Matthew, 2014. *Managing and Sharing Research Data: a guide to good practice*. SAGE publications, Inc. (ver pp. 6, 8, 10, 11, 12, 13, 14, 49).
- COX, Andrew M.; KENNAN, Mary Anne; LYON, Liz; PINFIELD, Stephen; SBAFFI, Laura, 2019. Maturing research data services and the transformation of academic libraries. *Journal of Documentation*. Vol. 75, n.º 6, pp. 1432–1462. URL DOI: [10.1108/JD-12-2018-0211](https://doi.org/10.1108/JD-12-2018-0211) (ver pp. 8, 50).
- COX, Andrew M.; PINFIELD, Stephen, 2014. Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science* **online** vol. 46, n.º 4, pp. 299–316 [acedido em 2025-01-18]. ISSN 0961-0006. URL DOI: [10.1177/0961000613492542](https://doi.org/10.1177/0961000613492542) (ver pp. 6, 10, 51).
- COX, Andrew M.; VERBAAN, Eddy, 2018. *Exploring Research Data Management*. Facet Publishing (ver pp. 7, 8, 10, 11, 13, 14, 49).
- CURTY, Renata G., 2016. Factors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study. *International Journal of Digital Curation*. Vol. 11, n.º 1, pp. 96–117. URL DOI: [10.2218/ijdc.v11i1.401](https://doi.org/10.2218/ijdc.v11i1.401) (ver p. 14).

- DE POOTER, David; APPELTANS, Ward; BAILLY, Nicolas; BRISTOL, Stacey; DENEUDT, Klaas; ELIEZER, Marla; FUJIOKA, Eileen; GIORGETTI, Alessandra; GOLDSTEIN, Phil; LEWIS, Mark; LIPIZER, Marina; MACKAY, Keith; MARIN, Martine; MONCOIFFE, Gwennaelle; NIKOLOPOULOU, Stella; PROVOOST, Pieter; RAUCH, Sven; ROUBICEK, Alexander; TORRES, Carlos; VAN DE PUTTE, Anton; VANDEPITTE, Leen; VANHOORNE, Bart; VINCI, Michela; WAMBIJI, Nyawira; WATTS, David; SALAS, Eduardo K.; HERNANDEZ, Francisco, 2017. Toward a new data standard for combined marine biological and environmental datasets – expanding OBIS beyond species occurrences. *Biodiversity Data Journal*. Vol. 5, e10989. URL DOI: [10.3897/BDJ.5.e10989](https://doi.org/10.3897/BDJ.5.e10989) (ver p. 16).
- DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION (EUROPEAN COMMISSION), 2018. *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data* **online** [acedido em 2025-01-21]. Rel. téc. Publications Office of the European Union. URL: <https://data.europa.eu/doi/10.2777/1524> (ver pp. 9, 10, 50).
- FOGLINI, Federica; GRANDE, Valentina, 2023. A Marine Spatial Data Infrastructure to manage multidisciplinary, inhomogeneous and fragmented geodata in a FAIR perspective ... the Adriatic Sea experience. *Oceanologia* **online** vol. 65, n.º 1, pp. 260–277 [acedido em 2025-01-06]. ISSN 0078-3234. URL DOI: [10.1016/j.oceano.2022.11.002](https://doi.org/10.1016/j.oceano.2022.11.002) (ver p. 23).
- GENOVA, Françoise; ARVISET, Christophe; ALMAS, Bridget M.; BARTOLO, Laura; BROEDER, Daan; LAW, Emily; MCMAHON, Brian, 2017. Building a Disciplinary, World-Wide Data Infrastructure. *Data Science Journal* **online** vol. 16, p. 16 [acedido em 2025-01-21]. ISSN 1683-1470. URL DOI: [10.5334/dsj-2017-016](https://doi.org/10.5334/dsj-2017-016). arXiv:1703.06450 [astro-ph] (ver p. 13).
- GREEN, A.; MACDONALD, S.; RICE, R., 2009. *Policy-making for Research Data in Repositories: A Guide*. 2009-05. Rel. téc. DISC-UK Datashare Project. **urlalso**: <http://www.disc-uk.org/docs/guide.pdf> (ver p. 10).
- HIGMAN, Rosie; BANGERT, Daniel; JONES, Sarah, 2019. Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights* **online** vol. 32, n.º 1 [acedido em 2025-01-18]. ISSN 2048-7754. URL DOI: [10.1629/uksg.468](https://doi.org/10.1629/uksg.468) (ver pp. 8, 13, 49).
- JAHNKE, L.; ASHER, A., 2012. *The problem of data*. Washington, DC. Rel. téc. Council on Library e Information Resources. **urlalso**: <http://www.clir.org/pubs/reports/pub154/pub154.pdf> (ver p. 13).
- KARIMOVA, Yulia; RIBEIRO, Cristina; DAVID, Gabriel, 2021. Institutional Support for Data Management Plans: Five Case Studies. Em: GAROUFALLOU, Emmanouel; OVALLE-PERANDONES, María-Antonia (eds.). *Metadata and Semantic Research*. Cham: Springer International Publishing, pp. 308–319. ISBN 978-3-030-71903-6. URL DOI: [10.1007/978-3-030-71903-6_29](https://doi.org/10.1007/978-3-030-71903-6_29) (ver pp. 10, 11).
- KIM, Youngseek; STANTON, Jeffrey M., 2012. Institutional and individual influences on scientists' data sharing practices. *Journal of Computational Science Education*. Vol. 3, n.º 1, pp. 47–56 (ver pp. 14, 15).
- LEADBETTER, Adam; CARR, Ramona; FLYNN, Sarah; MEANEY, Will; MORAN, Siobhan; BOGAN, Yvonne; BROPHY, Laura; LYONS, Kieran; STOKES, David; THOMAS, Rob, 2020. Implementation of a Data Management Quality Management Framework at the Marine Institute, Ireland. *Earth Science Informatics* **online** vol. 13, n.º 2, pp. 509–521 [acedido em 2025-01-06]. ISSN 1865-0481. URL DOI: [10.1007/s12145-019-00432-w](https://doi.org/10.1007/s12145-019-00432-w) (ver p. 24).

- MANZELLA, Giuseppe M. R.; NOVELLINO, Antonio (eds.), 2022. *Ocean science data: collection, management, networking and services*. Amsterdam: Elsevier. ISBN 978-0-12-823427-3 (ver pp. 16, 18, 20, 50, 51).
- MANZELLA, Giuseppe M.R.; DE STROBEL, Federico; PINARDI, Nadia; EMERY, William, 2022. A narrative of historical, methodological, and technological observations in marine science. Em: MANZELLA, Giuseppe M.R.; NOVELLINO, Antonio (eds.). *Ocean Science Data Collection, Management, Networking and Services*. Amsterdam: Elsevier, pp. 1–20 (ver p. 16).
- MARTÍNEZ, Enoc; GARCIA-BENADÍ, Albert; TOMA, Daniel M.; CARANDELL, Matias; NOGUERAS, Marc; RÍO, Joaquín del, 2023. An e-Infrastructure for FAIR Data Management of Underwater Observatories. Em: *OCEANS 2023 - Limerick online*pp. 1–6 [acedido em 2024-10-15]. URL DOI: [10.1109/OCEANS_Limerick52467.2023.10244451](https://doi.org/10.1109/OCEANS_Limerick52467.2023.10244451) (ver p. 23).
- MILES, Matthew B.; HUBERMAN, A. Michael; SALDAÑA, Johnny, 2014. *Qualitative Data Analysis: A Methods Sourcebook*. 3rd. Thousand Oaks, California: SAGE Publications, Inc. ISBN 978-1-4522-5787-7 (ver p. 34).
- NILSSON, Mikael; BAKER, Thomas; JOHNSTON, Pete, 2008. The Singapore Framework for Dublin Core Application Profiles. *International Conference on Dublin Core and Metadata Applications*, pp. 1–8. **urlalso**: <http://dcpapers.dublincore.org/pubs/article/view/925> (ver p. 56).
- OECD, 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding online*2007-04. [acedido em 2025-01-06]. Rel. téc. URL: https://www.oecd.org/en/publications/oecd-principles-and-guidelines-for-access-to-research-data-from-public-funding_9789264034020-en-fr.html (ver p. 50).
- PARTESCANO, Elena; JACK, Maria Eugenia Molina; VINCI, Matteo; COCIANCICH, Alexia; ALTENBURGER, Alessandro; GIORGETTI, Alessandra; GALGANI, Francois, 2021. Data quality and FAIR principles applied to marine litter data in Europe. *Marine Pollution Bulletin online*vol. 173, p. 112965 [acedido em 2025-01-06]. ISSN 0025-326X. URL DOI: [10.1016/j.marpolbul.2021.112965](https://doi.org/10.1016/j.marpolbul.2021.112965) (ver p. 23).
- PICKARD, Alison Jane, 2013. *Research methods in information*. 2.^a. Facet Publishing (ver p. 31).
- PIENTA, A.; ALTER, G.; LYLE, J., 2010. The enduring value of social science research: The use and reuse of primary research data. Em: *The Organisation, Economics and Policy of Scientific Research Workshop*. Torino, Italy. **urlalso**: <http://hdl.handle.net/2027.42/78307>. Paper presented (ver p. 15).
- QIN, J.; BALL, A.; GREENBERG, J., 2012. Functional and architectural requirements for meta-data: supporting discovery and management of scientific data. Em: *Proceedings of the International Conference on Dublin Core and Metadata Applications*, pp. 62–71 (ver p. 12).
- RAY, Joyce M., 2014. *Research data management: Practical Strategies for Information Professionals*. Purdue University Press. Charleston Insights in Library, Archival, and Information Sciences (ver pp. 9, 10, 11, 13).
- READ, K. B.; SURKIS, A.; LARSON, C.; MCCRILLIS, A.; GRAFF, A.; NICHOLSON, J.; XU, J., 2015. Starting the data conversation: informing data services at an academic health sciences library. *Journal of the Medical Library Association*. Vol. 103, n.º 3, pp. 131–135. URL DOI: [10.3163/1536-5050.103.3.005](https://doi.org/10.3163/1536-5050.103.3.005) (ver p. 10).

- REN, Xiao-Li; LI, Ming; CHEN, Yuxin; YU, Qiang; WANG, Xiaojie, 2021. Improving Ocean Data Services with Semantics and Quick Index. *Journal of Computer Science and Technology*. Vol. 36, n.º 6, pp. 1400–1413. URL DOI: [10.1007/s11390-021-1463-0](https://doi.org/10.1007/s11390-021-1463-0) (ver p. 24).
- RIBEIRO, Cristina; SILVA, João Rocha da; CASTRO, João Aguiar; AMORIM, Ricardo Carvalho; LOPES, João Correia; DAVID, Gabriel, 2018. Research Data Management Tools and Workflows: Experimental Work at the University of Porto. *IASSIST Quarterly online* vol. 42, n.º 2, pp. 1–17 [acedido em 2025-01-20]. ISSN 07391137. URL DOI: [10.29173/iq925](https://doi.org/10.29173/iq925) (ver p. 13).
- RODRIGUES, Joana; CASTRO, João Aguiar; SILVA, João Rocha da; RIBEIRO, Cristina, 2019. Hands-On Data Publishing with Researchers: Five Experiments with Metadata in Multiple Domains. Em: MANGHI, Paolo; CANDELA, Leonardo; SILVELLO, Gianmaria (eds.). *Digital Libraries: Supporting Open Science*. Cham: Springer International Publishing, pp. 274–288. ISBN 978-3-030-11226-4. URL DOI: [10.1007/978-3-030-11226-4_22](https://doi.org/10.1007/978-3-030-11226-4_22) (ver p. 11).
- SAYOGO, Djoko S.; PARDO, Theresa A., 2013. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*. Vol. 30, n.º 1, pp. 19–31. URL DOI: [10.1016/j.giq.2012.06.011](https://doi.org/10.1016/j.giq.2012.06.011) (ver pp. 14, 15, 50, 51).
- SCHAAP, Dick M. A.; NOVELLINO, Antonio; FICHAUT, Michèle; BULLA, Giuseppe M. R. et al., 2021. Data management infrastructures and their practices in Europe. Em: *Ocean Science Data: Collection, Management, Networking and Services*. Elsevier. URL DOI: [10.1016/B978-0-12-823427-3.00007-4](https://doi.org/10.1016/B978-0-12-823427-3.00007-4) (ver p. 24).
- SCHAAP, Dick M.A.; NOVELLINO, Antonio; FICHAUT, Michele; MANZELLA, Giuseppe M.R., 2022. Data management infrastructures and their practices in Europe. Em: MANZELLA, Giuseppe M.R.; NOVELLINO, Antonio (eds.). *Ocean Science Data Collection, Management, Networking and Services*. Amsterdam: Elsevier, pp. 45–68 (ver p. 16).
- SCHOENING, Timm; DURDEN, Jennifer M.; FABER, Claas; FELDEN, Janine; HEGER, Karl; HOVING, Henk-Jan T.; KIKO, Rainer; KÖSER, Kevin; KRÄMMER, Christopher; KWASNITSCHKA, Tom; MÖLLER, Klas Ove; NAKATH, David; NASS, Andrea; NATTKEMPER, Tim W.; PURSER, Autun; ZUROWIETZ, Martin, 2022. Making marine image data FAIR. *Scientific Data*. Vol. 9, p. 414. URL DOI: [10.1038/s41597-022-01491-3](https://doi.org/10.1038/s41597-022-01491-3) (ver p. 23).
- SHEIKH, Arslan; MALIK, Amara; ADNAN, Rubina, 2023. Evolution of research data management in academic libraries: A review of the literature. *Information Development*. Vol. 41, n.º 2, pp. 305–319. ISSN 0266-6669. URL DOI: [10.1177/02666669231157405](https://doi.org/10.1177/02666669231157405). Publisher: SAGE Publications Ltd (ver pp. 8, 10, 50, 51).
- SIMONCELLI, S.; MANZELLA, G.M.R.; STORTO, A.; PISANO, A.; LIPIZER, M.; BARTH, A.; MYROSHNYCHENKO, V.; BOYER, T.; TROUPIN, C.; COATANOAN, C.; PITITTO, A.; SCHLITZER, R.; SCHAAP, D.; DIGGS, S., 2022. A collaborative framework among data producers, managers, and users. Em: MANZELLA, Giuseppe M.R.; NOVELLINO, Antonio (eds.). *Ocean Science Data Collection, Management, Networking and Services*. Amsterdam: Elsevier, pp. 69–94 (ver pp. 17, 20).
- SIMONCELLI, Simone; MANZELLA, Giuseppe M. R.; STORTO, Andrea; PISANO, Antonello; LIPIZER, Michela et al., 2021. A collaborative framework among data producers, managers, and users. Em: *Ocean Science Data: Collection, Management, Networking and Services*. Elsevier. URL DOI: [10.1016/B978-0-12-823427-3.00001-3](https://doi.org/10.1016/B978-0-12-823427-3.00001-3) (ver p. 24).

- SURKIS, Alisa; READ, Kevin, 2015. Research data management. *Journal of the Medical Library Association : JMLA* **online**vol. 103, n.º 3, pp. 154–156 [accedido em 2025-01-05]. ISSN 1536-5050. URL DOI: [10.3163/1536-5050.103.3.011](https://doi.org/10.3163/1536-5050.103.3.011) (ver p. 7).
- TANG, Rong; HU, Zhichao, 2019. Providing research data management (RDM) services in libraries: preparedness, roles, challenges, and training for RDM practice. *Data and Information Management*. Vol. 3, n.º 2, pp. 84–101. URL DOI: [10.2478/dim-2018-0019](https://doi.org/10.2478/dim-2018-0019) (ver p. 8).
- TANHUA, Toste; POULIQUEN, Sylvie; HAUSMAN, Jessica; O'BRIEN, Kevin; BRICHER, Pip; BRUIN, Taco de; BUCK, Justin J. H.; BURGER, Eugene F.; CARVAL, Thierry; CASEY, Kenneth S.; DIGGS, Steve; GIORGETTI, Alessandra; GLAVES, Helen; HARSCOAT, Valerie; KINKADE, Danie; MUELBERT, Jose H.; NOVELLINO, Antonio; PFEIL, Benjamin; PULSIFER, Peter L.; VAN DE PUTTE, Anton; ROBINSON, Erin; SCHAAP, Dick; SMIRNOV, Alexander; SMITH, Neville; SNOWDEN, Derrick; SPEARS, Tobias; STALL, Shelley; TACOMA, Marten; THUISSE, Peter; TRONSTAD, Stein; VANDENBERGHE, Thomas; WENGREN, Micah; WYBORN, Lesley; ZHAO, Zhiming, 2019. Ocean FAIR Data Services. *Frontiers in Marine Science* **online**vol. 6 [accedido em 2024-12-02]. ISSN 2296-7745. URL DOI: [10.3389/fmars.2019.00440](https://doi.org/10.3389/fmars.2019.00440) (ver pp. 1, 23, 24).
- TENOPIR, Carol, 2015. Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship* **online**[accedido em 2025-01-21]. URL: [10.7191/jeslib.2015.1085](https://doi.org/10.7191/jeslib.2015.1085) (ver p. 51).
- TENOPIR, Carol; ALLARD, Suzie; DOUGLASS, Kimberly; AYDINOGLU, Arsev Umur; WU, Lei; READ, Eleanor; MANOFF, Maribeth; FRAME, Mike, 2011. Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE* **online**vol. 6, n.º 6, e21101 [accedido em 2025-01-18]. ISSN 1932-6203. URL DOI: [10.1371/journal.pone.0021101](https://doi.org/10.1371/journal.pone.0021101) (ver pp. 13, 14, 50).
- TENOPIR, Carol; DALTON, Elizabeth D.; ALLARD, Suzie; FRAME, Mike; PJESIVAC, Ivanka; BIRCH, Ben; POLLOCK, Danielle; DORSETT, Kristina, 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* **online**vol. 10, n.º 8, e0134826 [accedido em 2025-01-21]. ISSN 1932-6203. URL DOI: [10.1371/journal.pone.0134826](https://doi.org/10.1371/journal.pone.0134826) (ver pp. 13, 14, 15, 51).
- THANOS, Costantino, 2017. Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies. *Publications* **online**vol. 5, n.º 1, p. 2 [accedido em 2025-01-21]. ISSN 2304-6775. URL DOI: [10.3390/publications5010002](https://doi.org/10.3390/publications5010002) (ver pp. 6, 7).
- TRIPATHI, Manorama; CHAND, Mahesh; SONKAR, Satish K.; JEEVAN, V.K. J., 2017. A brief assessment of researchers' perceptions towards research data in India. *IFLA Journal*. Vol. 43, n.º 1, pp. 22–39. URL DOI: [10.1177/0340035216686983](https://doi.org/10.1177/0340035216686983) (ver p. 8).
- WHYTE, Angus; TEDDS, Jane, 2011. *Making the Case for Research Data Management*. DCC Briefing Papers. Digital Curation Centre. **urlalso**: <http://www.dcc.ac.uk/resources/briefing-papers>. Accessed: 2025-09-09 (ver pp. 6, 49).
- WIECZOREK, J.; BLOOM, S.; GURALNICK, R.; BLUM, S.; DÖRING, M.; GIOVANNI, R.; ROBERTSON, T.; VIEGLAIS, D., 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE*. Vol. 7. URL DOI: [10.1371/journal.pone.0029715](https://doi.org/10.1371/journal.pone.0029715) (ver p. 13).
- WILEY, Christie A.; KERBY, Erin E., 2018. Managing Research Data: Graduate Student and Postdoctoral Researcher Perspectives. *Issues in Science and Technology Librarianship*. N.º 89 (ver p. 14).

- WILKINSON, Mark D.; DUMONTIER, Michel; AALBERSBERG, IJsbrand Jan; APPLETON, Gabrielle; AXTON, Myles; BAAK, Arie; BLOMBERG, Niklas; BOITEN, Jan-Willem; DA SILVA SANTOS, Luiz Bonino; BOURNE, Philip E.; BOUWMAN, Jildau; BROOKES, Anthony J.; CLARK, Tim; CROSAS, Mercè; DILLO, Ingrid; DUMON, Olivier; EDMUNDS, Scott; EVELO, Chris T.; FINKERS, Richard; GONZALEZ-BELTRAN, Alejandra; GRAY, Alasdair J.G.; GROTH, Paul; GOBLE, Carole; GRETHE, Jeffrey S.; HERINGA, Jaap; 'T HOEN, Peter A.C.; HOOFT, Rob; KUHN, Tobias; KOK, Ruben; KOK, Joost; LUSHER, Scott J.; MARTONE, Maryann E.; MONS, Albert; PACKER, Abel L.; PERSSON, Bengt; ROCCA-SERRA, Philippe; ROOS, Marco; VAN SCHAIK, Rene; SANSONE, Susanna-Assunta; SCHULTES, Erik; SENGSTAG, Thierry; SLATER, Ted; STRAWN, George; SWERTZ, Morris A.; THOMPSON, Mark; VAN DER LEI, Johan; VAN MULLIGEN, Erik; VELTEROP, Jan; WAAGMEESTER, Andra; WITTENBURG, Peter; WOLSTENCROFT, Katherine; ZHAO, Jun; MONS, Barend, 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **online** vol. 3, n.º 1, p. 160018 [acedido em 2025-01-03]. ISSN 2052-4463. URL DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18) (ver pp. 6, 8, 9).
- WILLIAMS, M.; BAGWELL, J.; NAHM ZOZUS, M., 2017. Data management plans: The missing perspective. *Journal of Biomedical Informatics*. Vol. 71, pp. 130–142. URL DOI: [10.1016/j.jbi.2017.05.004](https://doi.org/10.1016/j.jbi.2017.05.004) (ver p. 10).
- WILLIS, Craig; GREENBERG, Jane; WHITE, Hollie, 2012. Analysis and synthesis of metadata goals for scientific data. *Journal of the American Society for Information Science and Technology* **online** vol. 63, n.º 8, pp. 1505–1520 [acedido em 2025-01-21]. ISSN 1532-2890. URL DOI: [10.1002/asi.22683](https://doi.org/10.1002/asi.22683) (ver p. 8).
- ZILIOLI, Martina; BERGAMI, Caterina; CARRARA, Paola; FUGAZZA, Cristiano; OGGIONI, Alessandro et al., 2021. Enabling the Reuse of Long-Term Marine Biological Observations in Essential Variables Frameworks Through a Practical Approach. *Frontiers in Marine Science*. Vol. 8, p. 645997. URL DOI: [10.3389/fmars.2021.645997](https://doi.org/10.3389/fmars.2021.645997) (ver p. 24).

Anexo A

Guião das Entrevistas

Contexto da Investigação

1. Pode descrever brevemente o seu trabalho de investigação?
2. Que tipo de dados recolhe? No presente, está a trabalhar com algum conjunto de dados?
3. Com que frequência são recolhidos e em que formatos costumam estar os dados?
4. Qual o contexto de recolha de dados, e quais as principais metodologias utilizadas para a sua recolha?

Organização dos dados

1. Onde armazena os dados que gera e como os organiza?
2. Como avalia as suas práticas de organização de dados?
3. Que tipo de dificuldades sente na organização ou localização dos dados?
4. Quando precisa de aceder aos dados que já guardou, utiliza alguma estratégia para os localizar facilmente?
5. Já teve alguma formação em gestão de dados?

Descrição dos dados

1. Que práticas utiliza para garantir que os dados estão corretamente documentados e fáceis de encontrar?
2. Usa alguma ferramenta para apoiar as anotações dos dados? Se sim, qual?
3. Os dados que gera são normalmente acompanhados de metadados?
4. Considera a anotação de dados uma prática relevante? De que forma?
5. Tem preocupação em garantir que os dados que produz sejam facilmente utilizados por outros investigadores?

Partilha e Acesso

1. Como é feita a partilha dos dados entre os membros do programa/projetos nos quais está envolvido?
2. Está familiarizado com o conceito de repositório de dados? O que entende por?
3. Já explorou os dados depositados de algum repositório?
4. Considera que os dados que gera podem ser facilmente reutilizados por outros investigadores de diferentes domínios científicos?
5. Colocaria alguma condição à partilha dos seus dados? Em caso afirmativo, quais seriam essas condições?
6. Estaria disposto a submeter os seus dados num repositório de dados?
7. Alguma vez depositou os seus dados num repositório de dados? Se sim, que tipos de dados e onde?
8. Quais seriam os requisitos para um repositório de dados ideal para a sua área de investigação?

Anexo B

Consentimento Informado



CONSENTIMENTO INFORMADO, LIVRE E ESCLARECIDO PARA PARTICIPAÇÃO EM INVESTIGAÇÃO de acordo com a Declaração de Helsínquia ¹ e a Convenção de Oviedo ²

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 do estudo: Gestão e Partilha de Dados Marinhos e Ambientais. Estudo de Caso aplicado ao CIIMAR Watch.

Enquadramento: A presente participação enquadra-se no âmbito da dissertação de mestrado da estudante Adriana Luís Martins Garcia, intitulada “Gestão e Partilha de Dados Marinhos e Ambientais. Estudo de Caso aplicado ao CIIMAR Watch”. O Mestrado surge no contexto do curso Mestrado em Ciência da Informação, da Universidade do Porto, sob orientação do Professor Doutor João Daniel Aguiar de Castro.

Explicação do estudo: Este estudo investiga a gestão e partilha de dados no âmbito do programa CIIMAR Watch, procurando compreender como os investigadores lidam com a organização, armazenamento, documentação e acesso aos dados gerados no contexto da sua investigação. O objetivo é identificar desafios e necessidades, contribuindo para o desenvolvimento de melhores práticas e infraestruturas de gestão de dados para o programa.

A participação no estudo envolve a realização de uma entrevista semiestruturada, que será gravada em áudio para posterior transcrição e análise. Os participantes serão identificados, pois o estudo pretende contextualizar as respostas no âmbito das suas áreas de investigação.

As gravações serão eliminadas no prazo de 365 dias. As transcrições poderão ser utilizadas na dissertação e em potenciais publicações científicas, sendo preservadas de forma a respeitar os princípios éticos da investigação.

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 os investigadores ou com qualquer entidade envolvida.

Confidencialidade e identificação: O presente estudo garante a confidencialidade e o uso exclusivo dos dados recolhidos para fins académicos. Em consonância com os princípios da transparência, os participantes serão identificados pelo nome, cargo e área de estudo, informações que serão utilizadas única e exclusivamente para a análise e divulgação dos resultados no âmbito deste estudo. As entrevistas serão gravadas para efeitos de registo e análise, garantindo-se que o seu conteúdo será tratado com sigilo e utilizado apenas no contexto académico. Os participantes declaram, por este meio, o seu consentimento para a gravação das entrevistas, bem como para a utilização dos dados nos termos acima descritos.

Eu, Adriana Luís Martins Garcia, estudante do Mestrado em Ciência da Informação, da Universidade do Porto, com o endereço eletrónico “garcia.adrianaluis@gmail.com”, agradeço a (nome do participante), pela disponibilidade e prontidão para a participação neste estudo.

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 pelo investigador.

Nome:

Assinatura:

Data: / /

**ESTE DOCUMENTO É COMPOSTO POR 1 PÁGINA E É FEITO EM DUPLICADO:
UMA VIA PARA O/A INVESTIGADOR/A, OUTRA PARA A PESSOA QUE CONSENTE.**

¹ http://epidemiologia.med.up.pt/pdfs/Helsinki_2013.pdf

² <http://dre.pt/pdf1sdip/2001/01/002A00/00140036.pdf>

Anexo C

Modelo XML do Minimal Metadata Set para o CIIMAR Watch

```
1 <metadata>
2   <!-- ===== IDENTIFICACAO ===== -->
3   <title></title>
4   <creator></creator>
5   <date></date>
6   <identifier type="DOI"></identifier>
7
8   <!-- ===== CONTEXTO ===== -->
9   <context>
10    <description></description>
11    <project>CIIMAR Watch</project>
12    <institute>CIIMAR      Centro Interdisciplinar de Investiga o Marinha e
        Ambiental</institute>
13    <contact>
14      <name></name>
15      <email></email>
16    </contact>
17    <funding></funding>
18    <license></license>
19    <accessRights></accessRights>
20  </context>
21
22  <!-- ===== RECOLHA / AQUISICAO ===== -->
23  <collection>
24    <superStation></superStation>
25    <location>
26      <latitude></latitude>
27      <longitude></longitude>
28      <placeName></placeName>
29      <depth unit="m"></depth>
30    </location>
```

```

31         <datetime></datetime>
32         <temporalCoverage>
33             <start></start>
34             <end></end>
35             <samplingInterval unit=""></samplingInterval>
36         </temporalCoverage>
37         <methodology></methodology>
38         <equipment></equipment>
39         <responsibleTeam></responsibleTeam>
40     </collection>
41
42     <!-- ===== ESTRUTURA DOS DADOS ===== -->
43     <dataStructure>
44         <dataType></dataType>
45         <discipline></discipline>
46         <fileFormat></fileFormat>
47         <encoding></encoding>
48         <variables>
49             <variable name="" unit="">
50                 <description></description>
51                 <standardName></standardName>
52                 <missingValue></missingValue>
53             </variable>
54         </variables>
55         <qualityControl></qualityControl>
56         <version></version>
57     </dataStructure>
58
59     <!-- ===== RELACOES E PROVENIENCIA ===== -->
60     <relations>
61         <relatedDataset></relatedDataset>
62         <relatedPublication></relatedPublication>
63     </relations>
64     <provenance>
65         <processingStep></processingStep>
66         <software></software>
67     </provenance>
68
69     <!-- ===== CITACAO ===== -->
70     <citation>
71         <recommended></recommended>
72         <acknowledgement></acknowledgement>
73     </citation>
74 </metadata>

```

Nota: Este modelo constitui uma estrutura inicial que poderá evoluir para integração futura com esquemas normalizados como Darwin Core e ISO 19115.