

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

**GESTÃO DE DADOS DE
INVESTIGAÇÃO NO DOMÍNIO
DA OCEANOGRAFIA
BIOLÓGICA: criação e avaliação
de um perfil de aplicação
baseado em ontologia**

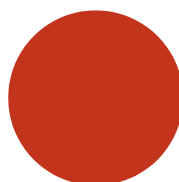
Rúbia Tatiana Gattelli

M

2015

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



Rúbia Tatiana Gattelli

Gestão de dados de investigação no domínio da
Oceanografia Biológica: criação e avaliação de um perfil
de aplicação baseado em ontologia

Dissertação realizada no âmbito do Mestrado em Ciência da Informação,
orientada pela Professora Doutora Maria Cristina de Carvalho Alves Ribeiro

Faculdade de Engenharia e Faculdade de Letras
Universidade do Porto

Julho de 2015

Gestão de dados de investigação no domínio da Oceanografia Biológica: criação e avaliação de um perfil de aplicação baseado em ontologia

Rúbia Tatiana Gattelli

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada pela Professora Doutora Maria Cristina de Carvalho Alves Ribeiro

Membros do Júri

Professor Doutor António Manuel Lucas Soares

Faculdade de Engenharia - Universidade do Porto

Professora Doutora Ana Alice Rodrigues Pereira Baptista

Faculdade de Engenharia - Universidade do Minho

Professora Doutora Maria Cristina de Carvalho de Alves Ribeiro

Faculdade de Engenharia - Universidade do Porto

“Research cannot flourish if data are not preserved and made accessible.”
(Nature, v. 461. n. 7261, Set. 2009)

Agradecimentos

Durante estes dois anos de mestrado foram vários os que contribuíram para esta caminhada, sem os quais eu não a teria concluído com tanto êxito e satisfação.

Em primeiro lugar agradeço à Orientadora Profa. Doutora Cristina Ribeiro, que imediatamente me acolheu como sua orientada, me inseriu na equipe do InfoLab e me introduziu no mundo da gestão de dados de investigação.

Aos colegas do InfoLab, pelo apoio e orientação nos momentos de dúvidas, pelo companheirismo e bom humor, pelos almoços e “tea time” nas tardes de inverno. Especialmente agradeço ao João Castro pela abertura e disponibilidade em me ajudar sempre que precisei, e ao João Rocha, pela ingestão do perfil de aplicação no Dendro.

À Universidade Federal do Rio Grande – FURG, na qual trabalho e a qual me concedeu a oportunidade de estar em Portugal para cursar o Mestrado em Ciência da Informação.

Aos investigadores do Instituto de Oceanografia/ FURG Prof. Dr. Leonir André Colling, Dr. Fabio Lameiro Rodrigues e Dr. Marcos Alaniz Rodrigues, pela inestimável colaboração na elaboração do perfil de aplicação para o domínio da Oceanografia Biológica. Aos investigadores do CIIMAR/UP Dra. Martina di Iulio Ilarri e Dr. Allan T. de Souza, pela valiosa participação na Campanha Dendro e avaliação do perfil de aplicação.

Um agradecimento especial para minha querida mãe, que desde muito cedo me ensinou o valor do conhecimento e da educação nas nossas vidas.

Às minhas grandes amigas e companheiras de mestrado Ângela Santos e Elisângela Moura, que tornaram estes dois anos muito mais fáceis, alegres e gostosos de viver. Gurias, obrigada pela amizade sincera e a troca de experiências incrível.

Ao meu marido e companheiro de todas as horas Marcos, pelo apoio incondicional, por acreditar em mim e me estimular a perseguir meus objetivos. Por ter me acompanhado nesta aventura portuguesa, ter me consolado nas horas difíceis e dado lindas palavras de incentivo para eu seguir adiante. Obrigada por me fazer rir sempre e ser meu sol particular.

Deixo um agradecimento especial a todos os familiares e amigos que, apesar da distância, sempre tiveram uma palavra de carinho e incentivo. A todos os professores e colegas do MCI, com quem aprendi muito nesta caminhada, e a todos os que, de uma forma ou outra contribuíram para o sucesso desta etapa.

Resumo

“Dados de investigação” tem sido um tema frequente nas agendas de vários grupos envolvidos com investigação, sejam universidades, agências financiadoras, editores de literatura científica, governos ou a própria sociedade. Por este motivo, a questão da gestão de dados de investigação, está no cerne da preocupação das comunidades científicas. Com o objetivo de apoiar investigadores na tarefa de gerir seus conjuntos de dados, este estudo propõe a elaboração de uma ferramenta de curadoria digital feita sob medida para descrição de dados do domínio da Oceanografia Biológica. Trata-se do desenho de um modelo de metadados, que combina descritores genéricos e específicos do domínio, chamado de perfil de aplicação e representado em uma ontologia. Esta é incorporada a uma plataforma colaborativa e multidisciplinar de gestão de dados (Dendro) concebida na Faculdade de Engenharia da Universidade do Porto, e fica disponível para que os investigadores descrevam seus dados ali depositados e possam avaliar o desempenho da ferramenta e da plataforma. O armazenamento e descrição de dados na plataforma Dendro visa a preparação para depósito dos mesmos em repositórios. Este trabalho conta com a colaboração de investigadores da área da Oceanografia Biológica para elaboração e avaliação do perfil de aplicação, e com a colaboração da equipe do projeto Dendro na integração do perfil na plataforma de gestão de dados. Ele pretende contribuir para o cenário científico ampliando a discussão sobre as boas práticas de gestão de dados de investigação e envolvendo investigadores de forma a conscientizá-los sobre a importância da preservação e partilha dos mesmos.

Palavras-chave: dados de investigação, gestão de dados de investigação, curadoria digital, metadados, perfil de aplicação, ontologia, Oceanografia Biológica.

Abstract

Research data has been a frequent theme on the agenda of several groups dealing with research, such as universities, funding agencies, science literature editors, government or society itself. For this reason, the research data management issue is at the center of the concerns of scientific communities. With the aim of supporting researchers on the management of their data, this study presents a digital curation tool tailored to the description of data from the Biological Oceanography domain. This tool is a metadata model that combines generic and domain-specific descriptors, to compose a so-called application profile, and it is formalized as an ontology. The ontology is then incorporated in a collaborative and multidisciplinary data management platform (Dendro), developed at the Faculty of Engineering of the University of Porto, and it becomes available for the researchers to deposit and describe their data. It is then possible to evaluate the performance of the tool and of the platform itself. The storage and description of data on Dendro paves the way for data deposit in external repositories. This work has the collaboration of Biological Oceanography researchers for the design and evaluation of the application profile, and the collaboration of the Dendro project team for the integration of the profile into the data management platform. It aims to contribute to the scientific scenario by broadening the discussion on best practices in research data management and by engaging researchers to make them aware of the importance of preserving and sharing data.

Keywords: research data, research data management, digital curation, metadata, application profile, ontology, Biological Oceanography

Lista de Abreviaturas e Siglas

ANDS – Australian National Data Services

CIIMAR – Centro Interdisciplinar de Investigação Marinha e Ambiental

DC – Dublin Core

DCC – Digital Curation Centre

DCMI – Dublin Core Metadata Initiative

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária

EML – Ecological Metadata Language

FEUP – Faculdade de Engenharia da Universidade do Porto

FURG – Universidade Federal do Rio Grande

ICMAN – Instituto de Ciencias Marinas de Andalucia

IEEE – Institute of Electrical and Electronics Engineers

INFOLAB – Information Systems Research Group

IO – Instituto de Oceanografia

JISC – Joint Information Systems Committee

MANTRA – Research Data Management Training

NASA – National Aeronautics and Space Administration

NISO – National Information Standards Organization

NOCS – Ocean and Earth Science, National Oceanography Centre Southampton

OECD – Organisation for Economic Co-Operation and Development

SEORG – Sociedade de Estudos Oceanográficos do Rio Grande

SNBU – Seminário Nacional de Bibliotecas Universitárias

Lista de Figuras

- Figura 1** – Ciclo de vida da investigação científica
- Figura 2** – Ciclo de vida dos dados de investigação
- Figura 3** – Ciclo de vida da curadoria digital
- Figura 4** – Conjunto de 15 elementos básicos Dublin Core
- Figura 5** – Excerto de um perfil de aplicação em Mecânica das Fraturas
- Figura 6** – Excerto de um perfil de aplicação em Mecânica das Fraturas – elementos Dublin Core
- Figura 7** – Mapa de conceitos “Atividades de Investigação em Oceanografia Biológica
- Figura 8** – Excerto de processo de seleção de conceitos
- Figura 9** – Excerto de seleção de conceitos conforme objetivos estabelecidos
- Figura 10** – Excerto de processo de seleção de descritores
- Figura 11** – Classe BiologicalOceanographyObservation
- Figura 12** – Propriedades da ontologia Biological Oceanography
- Figura 13** – Subpropriedades na ontologia Biological Oceanography
- Figura 14** – *Annotation Properties e Domains (Intersection)* na ontologia Biological Oceanography
- Figura 15** – Lista de conjuntos de descritores no Dendro
- Figura 16** – Ontologia “Biological Oceanography” na lista de descritores do Dendro
- Figura 17** – Apresentação dos descritores na ontologia “Biological Oceanography” da Plataforma Dendro
- Figura 18** – Exemplo de descrição sem recomendação de descritores
- Figura 19** – Exemplo de descrição com recomendação de descritores

Lista de Tabelas

Tabela 1 – Padrões de metadados

Tabela 2 – Transposição de conceitos para descritores (Partes 1 a 7)

Tabela 3 – Perfil de aplicação Biological Oceanography

Tabela 4 – Ontologias

Tabela 5 – Termos de ontologias para representação de descritores

Sumário

Resumo.....	5
Abstract.....	6
Lista de Abreviaturas e Siglas.....	7
Lista de Figuras.....	8
Lista de Tabelas.....	9
Introdução.....	13
1 Dados de investigação: gestão e curadoria digital no contexto da e-Science.....	20
1.1 Contexto e definição de dados de investigação.....	20
1.2 Ciclo de vida e gestão de dados de investigação.....	26
1.3 Curadoria de dados de investigação.....	31
1.4 Metadados em curadoria digital: perfis de aplicação e ontologias.....	35
1.4.1 Metadados.....	36
1.4.2 Perfis de aplicação.....	40
1.4.3 Ontologias.....	43
2 A investigação em Oceanografia Biológica em contexto organizacional.....	46
2.1 Universidade Federal do Rio Grande (FURG) – uma Universidade voltada para o ecossistema costeiro.....	46
2.1.2 O Instituto de Oceanografia da FURG.....	51
2.1.3 O Núcleo da Oceanografia Biológica.....	53
2.2 Instituições internacionais que realizam investigação em Oceanografia.....	55
3 Metodologia.....	63
3.1 Procedimentos.....	63
3.1.1 Construção de um perfil de aplicação baseado em ontologia.....	63
3.1.2 Promoção e avaliação do depósito de conjuntos de dados de investigação na plataforma Dendro.....	66
3.2 Participantes.....	67
3.3 Recolha de dados.....	67
3.4 Materiais.....	68
3.4.1 Dendro.....	68
3.4.2 CMapTool.....	69
3.4.3 Protégé.....	69
4 Desenho e avaliação de um perfil de aplicação para atividades de investigação na Oceanografia Biológica.....	70
4.1 A produção de dados no domínio da Oceanografia Biológica.....	71

4.2 Passo-a-passo do desenho do perfil de aplicação	74
4.2.1 Levantamento de requisitos do domínio junto aos investigadores	74
4.2.2 Mapa de conceitos e escolha de descritores	77
4.2.3 Esquemas de metadados e normas de descrição de dados	81
4.3 Formalização da ontologia e sua ingestão na plataforma Dendro	93
4.4 Depósito de dados na plataforma e avaliação do perfil de aplicação	103
Conclusões e perspectivas futuras	112
Referências	118
Anexos	123

Introdução

Enquadramento e motivação

Desde a última década tem havido um crescimento da importância de questões globais ligadas à ciência. Temas relacionados com gestão, acesso e reutilização de dados provenientes de atividades de investigação científica, sobretudo se estes forem resultantes de projetos de investigação financiados com recursos públicos, têm atraído atenção tanto de comunidades científicas quanto da esfera política internacional.

Começa a se firmar em cenário global um novo paradigma científico fortemente ligado à tecnologia da informação e orientado por e para os dados de investigação, o qual se convencionou chamar de “e-Science”. O termo, segundo Costa e Cunha (2014) foi cunhado por John Taylor, “Director General of Research Councils, Office of Science and Technology” do “National e-Science Centre” no Reino Unido. Taylor o definiu como uma ciência em grande escala executada através de colaborações globais possibilitadas pela Internet, as quais irão requisitar acesso a grandes coleções de dados, recursos computacionais em grande escala e visualização de alto desempenho para os cientistas.

Dentre as características da “e-Science” ressaltadas pelos autores da área, das quais podem ser citadas o trabalho colaborativo, a multidisciplinaridade e o uso de aparato tecnológico, os dados provenientes das atividades de investigação (referidos na literatura em língua portuguesa como “dados de investigação” – termo adotado neste trabalho, ou “dados científicos”, e na literatura em língua inglesa como “research data”) desempenham um papel central. Os dados de investigação passam a ganhar status de importância e maior visibilidade neste novo paradigma científico, posto que cresce o entendimento de que são a base fundamental da ciência e da investigação científica. Os dados de investigação são o insumo a partir do qual a informação e o conhecimento são derivados, e têm o propósito de produzir e validar resultados de investigação.

De acordo com Corti *et. al.* (2014) o período a partir dos anos 2000 tem visto uma série de empreendimentos em direção ao compartilhamento de dados, bem como ao desenvolvimento de capacidades humana e material para fazê-lo. São vistas várias

iniciativas que visam assegurar a qualidade, sustentabilidade e acessibilidade aos dados de investigação. Da mesma forma, também tem crescido a demanda pela transparência e acesso aos dados, estas advindas não somente dos ambientes acadêmicos, onde em grande parte os dados são produzidos, mas de diferentes setores, como financiadores de investigação, editores de literatura científica, governos e a sociedade em geral.

Igualmente há que se referir a preocupação crescente das comunidades científicas em lidar com o aumento progressivo dos dados, impulsionado pelo incremento do aparato tecnológico utilizado para gerar ou recolher dados durante suas ações de investigação. Este fenômeno, constatado e reportado na literatura como “dilúvio de dados”, (do inglês “data deluge”) fez emergir a atenção em relação às políticas e práticas de gestão de dados de investigação.

A gestão de dados de investigação compreende todas as práticas, manipulações, melhoramentos e processos que possam assegurar a qualidade dos dados, bem como, garantir que estejam organizados, documentados, preservados, sustentáveis, acessíveis e reutilizáveis (CORTI *et al.*, 2014). Portanto, questões relacionadas ao armazenamento, utilização e preservação dos dados constituem problemas concretos para as comunidades de investigadores.

Na tentativa de oferecer soluções para tais problemas uma área relacionada com a curadoria digital de dados científicos tem se afirmado. A curadoria digital, de acordo com o Digital Curation Centre, centro de investigação em curadoria de informação digital no Reino Unido, envolve “a manutenção, preservação e adição de valor aos dados de investigação digitais ao longo do seu ciclo de vida”.

Dentre as atividades a que se dedica a curadoria de dados de investigação, destaca-se aqui a criação de metadados normalizados, essenciais para descrição, acesso e reutilização dos dados científicos. A apropriada descrição de conjuntos de dados é importante para que estes sejam entendidos e, assim, reutilizados. Uma descrição clara requer um conhecimento profundo dos dados e do seu processo de criação, fatores que são diferentes em cada domínio de conhecimento. A utilização de metadados normalizados é importante para garantir interoperabilidade, no entanto, estes podem não contemplar todas as características e especificidades das diferentes áreas de domínio. Logo, diferentes áreas de investigação demandam a criação de diferentes conjuntos de metadados, para que os dados possam ser descritos apropriadamente no que diz respeito às suas peculiaridades.

À definição de um conjunto de metadados específico para uma necessidade específica de descrição em determinado domínio chamamos de perfil de aplicação.

Um perfil de aplicação cumpre sua função em uma plataforma eletrônica onde se queira armazenar e descrever determinado recurso, e uma boa forma para representá-lo dentro de uma plataforma web é fazê-lo através do uso de uma ontologia. Uma ontologia é a ferramenta adequada para representar um modelo de metadados criado sob medida para diferentes domínios, uma vez que apresenta capacidade semântica para representar tais especificidades, pode evoluir facilmente e ainda ser integrada por sistemas de gestão de dados.

Levando em consideração o cenário da produção e gestão de dados de investigação por comunidades científicas, este trabalho depara-se com a seguinte questão de investigação: “Como apoiar um pequeno grupo de investigadores brasileiros do domínio da Oceanografia Biológica na gestão de seus dados de investigação?” Partindo desta questão, chega-se à hipótese de que um esquema de metadados baseado em ontologia, elaborado sob medida para descrição de dados do domínio específico e abrigado em uma plataforma de gestão de dados, fornece o apoio necessário aos investigadores na gestão de dados de investigação.

A proposta deste trabalho é, portanto, fornecer um modelo de metadados, chamado perfil de aplicação, para descrição dos dados de investigação da área de Oceanografia Biológica. Para a construção do perfil de aplicação é necessário mapear algumas especificidades do domínio e de suas atividades de investigação, de modo a selecionar e/ou criar metadados adequados. Portanto, a colaboração de investigadores, que são os reais conhecedores do domínio, é requisito fundamental deste processo.

Os metadados do perfil de aplicação, também chamados de descritores, são provenientes de padrões ou normas existentes para descrição de dados e, em caso de haver uma necessidade específica de descrição cujos padrões preexistentes não possam sanar, outros metadados podem ser criados. Desta maneira, perfis de aplicação são elaborados para atividades específicas de investigação e têm o objetivo de resolver necessidades específicas de descrição.

O conjunto de descritores criado para as atividades de investigação do domínio da Oceanografia Biológica é formalizado neste trabalho em uma ontologia, onde suas propriedades serão a representação formal destes descritores. O tipo de ontologia criada

para este propósito é uma ontologia ligeira, que possui uma estrutura mais simples, apresenta um número de classes limitado e não estabelece muitas relações entre conceitos. Esta torna-se uma alternativa viável para apoiar modelos de dados de um sistema de gestão uma vez que é facilmente manejável por curadores (CASTRO, SILVA e RIBEIRO, 2014).

Após a construção da ontologia esta é incorporada no Dendro, uma plataforma colaborativa de gestão de dados para pequenos grupos de investigação, desenvolvida na Faculdade de Engenharia da Universidade do Porto. A plataforma Dendro foi desenhada para dar apoio aos investigadores a começarem a descrição de dados desde o momento em que estes são criados; seu modelo de dados encoraja os curadores a modelar ontologias ligeiras que satisfaçam a necessidade de descrição dos dados em cada área de domínio específica (CASTRO, SILVA e RIBEIRO, 2014).

O perfil de aplicação criado e formalizado em ontologia passa a compor uma lista de vocabulários com descritores específicos de várias áreas do conhecimento, contribuindo assim para a multidisciplinaridade da plataforma Dendro. O objetivo é que investigadores de vários domínios possam fazer o armazenamento e descrição de seus conjuntos de dados. Estes, desta maneira, ficam preparados para serem partilhados com outros repositórios ou sistemas de gestão de dados, caso seja do interesse do investigador.

Como parte do projeto Dendro, foi realizada uma campanha de experiências de interação dos investigadores com a plataforma Dendro, como forma de promovê-la como ferramenta de gestão de dados e também para aprimorar suas funcionalidades. Este estudo colaborou com a campanha de avaliação, participando das interações dos investigadores da área de Oceanografia Biológica com a plataforma e colaborando em sessões em outros domínios.

Para participar da campanha foram convidados investigadores da Universidade do Porto daquelas áreas em que foram criados os conjuntos de descritores. Na experiência estes procederam ao depósito de um conjunto de dados na plataforma e à sua descrição utilizando os vocabulários disponíveis. A experiência foi avaliada e serve de subsídio para melhoramentos na plataforma. Como parte desta experiência de avaliação foi testado um sistema de recomendação de descritores com vista a facilitar a tarefa de anotação dos dados.

Concomitante com estas experiências os investigadores da Oceanografia Biológica foram convidados a realizar uma avaliação sobre o perfil de aplicação “Biological Oceanography”. Depois de sua interação com o Dendro e experiência de descrição de dados de investigação foram solicitados a responder um inquérito para avaliar a usabilidade dos descritores do perfil de aplicação e sua compatibilidade com a investigação no domínio.

Objetivos e resultados esperados

O objetivo principal deste trabalho é apoiar investigadores do domínio da Oceanografia Biológica na gestão de seus dados de investigação, partindo da colaboração com um grupo desta área. O apoio à gestão de dados implica questões como depósito, preservação e partilha de dados, motivo pelo qual o objetivo central do estudo está imediatamente suportado por outros dois grandes objetivos, que são:

- elaborar um modelo de metadados, chamado perfil de aplicação, baseado em ontologia, a ser usado em plataforma de gestão de dados de investigação para descrever o domínio; e
- avaliar o modelo de metadados criado através de uma experiência de interação de investigadores com a plataforma de gestão de dados, na qual são convidados a depositar e descrever seus conjuntos de dados de investigação.

Estes três objetivos estão intimamente ligados no trabalho. Acredita-se que a elaboração de um perfil de aplicação para fornecer descritores específicos do domínio é um incentivo aos investigadores para que descrevam seus dados, uma vez que este procura contemplar as características principais da área. Desta forma, o investigador estará familiarizado com os termos apresentados e não terá resistência ou dificuldade na descrição dos seus conjuntos de dados. O perfil de aplicação é representado por uma ontologia a ser incorporada em uma plataforma de gestão de dados de investigação. Neste trabalho a ontologia foi integrada na plataforma Dendro que está em desenvolvimento na Faculdade de Engenharia da Universidade do Porto.

Após a formalização em ontologia e ingestão do perfil de aplicação na plataforma Dendro, um grupo de investigadores é convidado a depositar e descrever seus conjuntos de dados com o intuito de avaliar o desempenho, não somente do perfil de aplicação, mas da própria plataforma.

Estrutura da dissertação

Este trabalho encontra-se estruturado em quatro capítulos abaixo detalhados.

O capítulo 1, intitulado “Dados de investigação: gestão e curadoria digital no contexto da *e-Science*” trata da revisão de literatura onde, inicialmente é apresentado o contexto onde emergiram as questões relativas à gestão de dados de investigação, contexto este cujas características o definiram como *e-Science*. Também são apresentados conceitos relativos aos dados de investigação, tipologia, ciclo de vida e gestão de dados. A seguir discorre-se sobre curadoria de dados de investigação, a definição de curadoria digital com o seu ciclo de vida e são apresentadas algumas ferramentas de curadoria digital, sendo estas os metadados, os perfis de aplicação e as ontologias.

O segundo capítulo da dissertação, intitulado “A investigação em Oceanografia Biológica em contexto organizacional” procura traçar um panorama da investigação em Oceanografia Biológica, área de domínio escolhida para realização deste trabalho, no contexto da Universidade Federal do Rio Grande (FURG) e em outras Universidades internacionais. Inicialmente traça um histórico da FURG e sua ligação com o meio ambiente, explicitando sua vocação institucional voltada para os ecossistemas costeiros. Neste ambiente surgiu o primeiro curso de graduação de Oceanologia do país e mais tarde, os cursos de pós-graduação em Oceanografia, traduzindo-se em um ambiente propício à investigação em Oceanografia Biológica. Em seguida contextualiza as investigações neste domínio em universidades do exterior, como forma de traçar semelhanças e diferenças de nível organizacional com a FURG.

O terceiro capítulo trata da metodologia e apresenta de forma mais resumida a abordagem metodológica utilizada para chegar ao objetivo final do estudo. Está subdividido em “Procedimentos”, onde estão descritos os passos para a construção de um perfil de aplicação baseado em ontologia para o domínio da Oceanografia Biológica;

“Participantes”, onde são citados os colaboradores do estudo, nomeadamente os investigadores da área; “Recolha de dados”, onde estão descritos as técnicas e instrumentos de recolha de dados utilizados; e “Materiais”, onde são citados os softwares utilizados para a elaboração do trabalho.

O quarto capítulo se dedica a documentar todo o processo realizado para elaboração e avaliação do perfil de aplicação “Biological Oceanography”. Fala sobre atividades de investigação e produção de dados no domínio para então descrever o passo-a-passo realizado para desenhar o perfil de aplicação. Em seguida ilustra a formalização do perfil em ontologia, que é incorporada na plataforma Dendro, para a gestão colaborativa de dados de investigação. Por fim, são descritas algumas experiências de teste da plataforma Dendro e a concomitante avaliação do perfil de aplicação elaborado para descrever os dados do domínio da Oceanografia Biológica.

1 Dados de investigação: gestão e curadoria digital no contexto da e-Science

1.1 Contexto e definição de dados de investigação

Muito se tem debatido a respeito do modo como os avanços das tecnologias da informação e comunicação têm afetado continuamente a sociedade. O seu impacto nos setores económico, cultural, organizacional e espaço-temporal é discutido por vários estudiosos na tentativa de definir as novas relações emergentes destes contextos.

Werthein cita a existência de um novo paradigma técnico-económico que passa a usar a informação como insumo principal para transformações técnicas, organizacionais e administrativas. (WERTHEIN, 2000, p. 71). De fato, este novo paradigma é o da tecnologia da informação, que expressa a essência da presente transformação tecnológica em suas relações com a economia e a sociedade. Castells é outro autor que defende que o processo de transformação social está associado ao paradigma tecnológico, baseado nas tecnologias da informação e da comunicação, e define este tipo de sociedade como “sociedade em rede”, onde a tecnologia é condição necessária, mas não suficiente, para sua organização (CASTELLS, 2006, p. 17).

Considerando o poder transformador do paradigma tecnológico é seguro afirmar que ambientes académicos e de investigação científica não ficaram imunes à sua influência. A adesão ao uso de novos equipamentos tecnológicos, de hardware e software, e o uso da Internet aumentaram e melhoraram as tarefas ligadas ao fazer científico e, conseqüentemente, proporcionaram o aumento da produção de dados de investigação e de informação científica. Além disso, também se estabeleceram novas relações entre os cientistas e a comunicação científica, cujas práticas também tem sido transformadas pela tecnologia da informação e comunicação, especialmente a Internet. Segundo Sawant (2012, p. 21-22) a maneira como académicos e investigadores comunicam seus resultados de investigação se modificou com o advento da comunicação eletrônica, a qual permite disseminação e acesso a publicações e outros documentos produzidos no âmbito académico, em escala global, além de afetar as publicações permitindo novos modelos de publicação.

Para a autora a Internet possibilitou comunicação entre acadêmicos em três importantes vias, para além dos meios formais de comunicação: arquivos de acesso aberto, dos quais são exemplo os repositórios institucionais; publicações de acesso aberto, representados pelos periódicos de acesso aberto; e outros meios de comunicação académica através dos canais disponíveis na chamada web 2.0, tais como blogs, wikis, redes sociais e outras plataformas. Este novo modelo de comunicação entre cientistas, emergente do paradigma tecnológico, evidenciou, portanto, não somente novas possibilidades de publicação, com publicação de conteúdos diferenciados em plataforma web, mas, sobretudo, uma nova postura de partilha de informação e de trabalho colaborativo entre investigadores.

A facilidade e a democratização do acesso à informação, bem como a de publicação, transformaram a rotina dos centros de investigação e o conhecimento se tornou disponível a esferas onde antes o acesso era restrito, como, por exemplo, a investigadores iniciantes e centros de investigação de países em desenvolvimento, onde o salto tecnológico tem sido sentido nos últimos anos.

O impacto do avanço tecnológico mudou as características da atividade científica e fez surgir expressões como “e-Science” e “e-Research”. Com elas, os centros destinados às comunidades científicas, estas formadas por investigadores e profissionais da Ciência da Informação, se empenham em debater e definir estes tópicos, como é o caso da “Association of Research Libraries”¹, nos Estados Unidos da América.

Segundo um estudo realizado por Amanda L. Whitmire, uma investigadora que se tornou bibliotecária na Universidade do Estado de Oregon, nos Estados Unidos, os termos “e-Science” e “e-Research” são usados de forma desconexa e confusa, muitas vezes para se referir à mesma coisa. A questão é cultural, enquanto a comunidade de cientistas conhece e utiliza o termo “e-Science”, a comunidade da Biblioteconomia e Ciência da Informação, adota o jargão “e-Research”. Tal confusão, na opinião de Whitmire, é problemática, pois ao mesmo tempo que as Bibliotecas promovem esforços para prover serviços aos investigadores, arriscam alienar do processo seus potenciais parceiros usando uma linguagem que eles não entendem (WHITMIRE, 2013, p. 68).

¹ Mais informações em: <http://www.arl.org/index.php>

Do ponto de vista da comunidade de bibliotecas a “Association of Research Libraries” define “e-Research” como sendo:

formas colaborativas, em rede, em grande escala e intensivamente baseada em computação, de investigação e educação em todas as disciplinas, incluindo todas as ciências naturais e físicas, disciplinas tecnológicas e aplicadas relacionadas, biomedicina, ciências sociais e humanidades digitais (ARL, 2014, tradução nossa).²

Partilhando aspetos comuns, o termo “e-Science” é definido por John Taylor, Diretor Geral dos Conselhos de Investigação, Escritório de Ciência e Tecnologia do Reino Unido, como:

ciência em grande escala que será crescentemente executada através de colaborações globais distribuídas possibilitadas pela Internet. Tipicamente, uma característica de tal empreendimento científico colaborativo é que irão requisitar acesso a grandes coleções de dados, recursos computacionais em grande escala e visualização de alto desempenho para os cientistas (TAYLOR, 2014, tradução nossa).³

Taylor ainda acrescenta que e-Science trata de colaboração global em áreas fundamentais da ciência e da geração de infraestrutura que permite que isso ocorra.

Considerando as duas definições, nota-se que ambas mencionam características semelhantes no que se refere a utilização intensiva da tecnologia computacional nas atividades de investigação, elevando a ciência para uma escala maior e tornando a relação entre cientistas mais global e colaborativa. De qualquer forma, apesar das semelhanças, o termo “e-Science” é mais amplamente disseminado, aparece em maior número na literatura científica e, portanto, pode ser melhor definido.

² “computationally intensive, large-scale, networked and collaborative forms of research and scholarship across all disciplines, including all of the natural and physical sciences, related applied and technological disciplines, biomedicine, social science and the digital humanities”. Disponível em: <http://www.arl.org/focus-areas/e-research#.VKcZvyusWSo>.

³ “large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists”. Disponível em: <http://www.nesc.ac.uk/nesc/define.html>.

Em adição à definição de Taylor, Bohle (2013), bibliotecária e arquivista americana cuja experiência inclui trabalho em bibliotecas de instituições de investigação científica como a NASA, explica o que significa “e-Science” para ela:

e-science é a aplicação da tecnologia de computador para a realização de investigação científica moderna, incluindo a preparação, a experimentação, a recolha de dados, divulgação de resultados, e armazenamento de longo prazo e acessibilidade de todos os materiais gerados através do processo científico. Estes podem incluir a modelagem e análise de dados, cadernos de laboratório eletrônicos ou digitalizados, conjuntos de dados brutos e trabalhados, produção de manuscritos e versões preliminares, pré-publicações, e publicações impressas e/ou eletrônicas (BOHLE, 2013, tradução nossa).⁴

Na sua definição, Bhole dá ênfase às características práticas desta nova maneira de fazer ciência, tornando-a mais facilmente compreensível. Definição igualmente simples pode ser consultada na homepage da “IEEE International Conference on eScience”, que, em sua versão reduzida, afirma que “e-Science promove inovação na investigação colaborativa e baseada no uso intensivo de recursos computacionais ou de dados, em todas as disciplinas e durante todo o ciclo de vida da investigação” (tradução nossa)⁵.

“e-Science”, portanto, pode ser definida como uma metodologia de investigação (GORE, 2011, p. 168; PENNOCK, 2006; WHITMIRE, 2013, p. 69), que utiliza largamente os recursos tecnológicos disponíveis para empreender as atividades científicas, as quais, por sua vez, tiveram suas características modificadas pela tecnologia. Inclusive, duas das principais características que definem a “e-Science” são o trabalho em rede e os dados que são produzidos nas investigações (GORE, 2011, p. 168).

⁴ “E-science is the application of computer technology to the undertaking of modern scientific investigation, including the preparation, experimentation, data collection, results dissemination, and long-term storage and accessibility of all materials generated through the scientific process. These may include data modeling and analysis, electronic/digitized laboratory notebooks, raw and fitted data sets, manuscript production and draft versions, preprints, and print and/or electronic publications”. Disponível em: http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/

⁵ “eScience promotes innovation in collaborative, computationally- or data-intensive research across all disciplines, throughout the research lifecycle”. Disponível em: <https://escience-conference.org/>

A produção intensa e crescente de dados de investigação, impulsionada pela adoção de ferramentas, técnicas e práticas proporcionadas pela revolução digital, trouxe para estes grande visibilidade. De acordo com Sharma (2014, p. 138) hoje, a inundação de dados de investigação, chamada de dilúvio de dados, está a influenciar o processo de tomada de decisão no âmbito científico.

Para Corti *et al.* (2014) os dados de investigação são a pedra angular do conhecimento científico, da aprendizagem e inovação, eles explicam e desenvolvem a humanidade e o mundo à nossa volta. Os autores advogam em favor do acesso aberto aos dados produzidos ou recolhidos nas atividades de investigação, pois estes permitem que resultados e descobertas de investigação possam ser verificados quando necessário.

A definição de Corti *et al.* vem ao encontro da definição da OCDE (2007, p. 13) que afirma que dados de investigação são “registos factuais usados como fontes primárias na investigação científica, e que são geralmente aceites na comunidade científica como necessários para validar os resultados de investigação”⁶.

A “Australian National Data Service” (ANDS) declara que é um desafio prover uma definição oficial para dados de investigação, uma vez que qualquer definição depende do contexto na qual foi questionada, no entanto, oferece algumas definições que recolheu de políticas de gestão de dados de universidades australianas. A Universidade de Tecnologia de Queensland, por exemplo, define os dados de investigação “sob a forma de fatos, observações, imagens, resultados de programas de computador, gravações, medidas ou experiências em que um argumento, teoria, teste ou hipótese, ou outro resultado de investigação é baseado”⁷ (tradução nossa). Os dados podem estar em qualquer formato ou mídia, e podem ser brutos ou processados. Para a Universidade de Melbourne dados são “fatos, observações ou experiências em que um argumento, teoria ou o teste se baseia. Os dados podem ser numéricos, descritivos ou

⁶ “factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings”. OECD Principles and guidelines for access to research data from public funding. Disponível em: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

⁷ “Research data means data in the form of facts, observations, images, computer program results, recordings, measurements or experiences on which an argument, theory, test or hypothesis, or another research output is based”. **ANDS Guide**: What is research data? Disponível em: <http://ands.org.au/guides/what-is-research-data.html>

visuais. Os dados podem ser crus ou analisados, experimentais ou observacionais”⁸ (tradução nossa). Ela não faz distinção entre dados físicos e digitais e tampouco registros derivados da investigação.

De fato, os dados de investigação podem ser de diferentes tipologias e usados de formas variadas por instituições e investigadores. Portanto, são passíveis de definições diferenciadas e abertas, e seu significado irá depender do tipo de contexto onde é gerado e aplicado e do propósito de uso. Assume-se, no entanto, que hoje a maioria dos conjuntos de dados recolhidos e gerados nas atividades de investigação já estejam no formato digital, produzidos neste formato ou, ao menos, digitalizados.

Para a MANTRA, entidade ligada à Universidade de Edimburgo que oferece cursos online gratuitos na área de gestão de dados digitais, os dados de investigação podem ser vistos como o mais baixo nível de abstração do qual informação e conhecimento são derivados. Podem ser primários ou secundários, de acordo com o responsável pela sua recolha, sendo que os dados primários são recolhidos pelo investigador que conduz a pesquisa; e os dados secundários são recolhidos por alguém que não é utilizador. A utilização dos dados secundários pode impor maiores dificuldades ao investigador, uma vez que ele não possui o controle e conhecimentos sobre sua origem e organização.

Quanto à origem a MANTRA fornece a seguinte classificação para os dados de investigação:

Observacional: capturados em tempo real, são usualmente únicos e insubstituíveis. Exemplo: imagens do cérebro.

Experimental: advindos de resultados de experiências e equipamentos de laboratório.

Simulação: gerados a partir de modelos de teste, onde o modelo e metadados podem ser mais importantes que os dados de saída do modelo. Exemplo: modelos económicos ou climáticos.

⁸ “Data are facts, observations or experiences on which an argument, theory or test is based. Data may be numerical, descriptive or visual. Data may be raw or analysed, experimental or observational”. **ANDS Guide:** What is research data? Disponível em: <http://ands.org.au/guides/what-is-research-data.html>

Derivados ou compilados: resultantes do processamento ou combinação de dados brutos. Exemplo: bases de dados compiladas, dados de censos, resultados de análise automática de textos (“text mining”).

Referência ou Canônicos: a conglomeração ou coleção de conjuntos de dados menores, muito provavelmente publicados e curados. Exemplo: bancos de dados de genes ou bases de dados cristalográficas.

Em relação à tipologia os formatos de arquivos dados podem ser: textual, numérico, multimídia, modelos, softwares, específicos de disciplina e específicos de instrumentos.

São objetos de dados de investigação: documentos de texto (ex.: Word) e planilhas (ex.: Excel); notas de campo e laboratório, diários; questionários, transcrições; material de áudio e vídeo; fotografias, filmes; respostas de questionários; slides, artefatos, espécies, amostras; arquivos com dados estatísticos; metodologias, workflows, etc.

Tais definições são amplas o suficiente para considerar a importância dos dados produzidos nas mais variadas disciplinas, não somente como resultados de investigação, mas como insumo para novos empreendimentos científicos.

1.2 Ciclo de vida e gestão de dados de investigação

Se, por um lado, cresce a noção da importância dos dados de investigação para a atividade científica, por outro, surge a preocupação das comunidades científicas em lidar com volumes cada vez maiores de dados. Com o aumento do volume de dados, a necessidade de encontrar formas de gestão e curadoria dos dados, com vistas à sua partilha e reutilização também cresce (HODSON, 2009).

Em relação ao ciclo de vida dos dados CORTI *et al* (2014) afirmam que a maioria dos dados frequentemente têm muito mais tempo de vida que o projeto de investigação que os cria. Investigadores podem continuar a trabalhar com os dados mesmo depois que o financiamento de um projeto tenha encerrado, através de projetos de acompanhamento que podem analisar os dados ou juntar a estes uma quantidade extra de dados. Ainda, os dados podem ser reutilizados e usados para diferentes fins por

outros investigadores. Para que isso seja possível, os dados devem ser bem geridos durante o curso de vida do projeto de investigação.

O JISC⁹ propõe um modelo (figura 1) que estabelece o ciclo de vida de um projeto de investigação onde se pode visualizar na etapa “processo de investigação” a geração dos dados, que ocorre através de simulação, experimentação ou observação. Neste caso o ciclo de vida dos dados é inerente ao do projeto de investigação e questões como a gestão, análise e partilha de dados são partes importantes deste ciclo, como se pode observar no modelo. Toda a etapa “processo de investigação” ocorre em ambientes de investigação virtuais que fornecem as ferramentas necessárias para a gestão do processo de investigação e permitem a colaboração entre investigadores, instituições ou mesmo países que partilham do mesmo objeto de investigação.

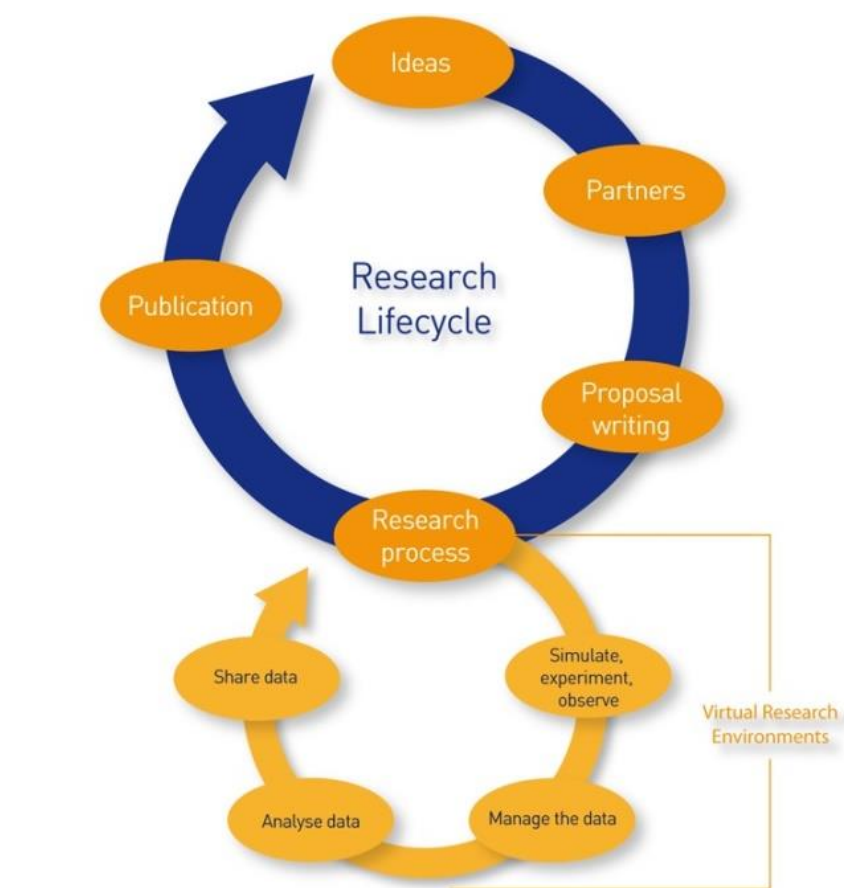


Figura 1 – Ciclo de vida da investigação científica (Joint Information Systems Committee, 2014)

⁹ **How Jisc is helping researchers.** Disponível em: <http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx#ideas>

O ciclo de vida dos dados de investigação está intimamente ligado à sua gestão. Gestão de dados refere-se a todos os aspectos da criação, hospedagem, fornecimento, manutenção e arquivamento e preservação de dados. É uma das áreas essenciais da conduta responsável de investigação (MANTRA, 2014). Por gestão de dados se entende todas as práticas, manipulações, melhoramentos e processos que asseguram que os dados de investigação sejam de alta qualidade, bem como, estejam organizados, documentados, preservados, sustentáveis, acessíveis e reutilizáveis (CORTI *et al.*, 2014).

Tais aspetos podem ser visualizados no diagrama do ciclo de vida dos dados de investigação cedido pelo DCC para o MANTRA¹⁰:



Figura 2 – Ciclo de vida dos dados de investigação (MANTRA, 2014)

Cada um dos quadros do diagrama corresponde a uma atividade envolvida na gestão de dados de investigação, por exemplo:

Create – a criação de dados envolve a escolha cuidadosa do tipo e formato dos dados e a forma como eles são criados. Nesta etapa é também aconselhável a elaboração de um plano de gestão de dados.

¹⁰ **Data management plans:** Research data lifecycle. Disponível em: <http://datalib.edina.ac.uk/mantra/datamanagementplans/#>

Document – documentar dados significa fornecer informação sobre eles aos seus utilizadores (e ao próprio criador dos dados) para que estes sejam entendidos.

Use – se relaciona ao acesso e uso dos dados, e suscita questões a respeito de sua organização. Pode envolver questões de direito e níveis de acesso aos dados.

Store – armazenamento e backup dos dados durante o projeto.

Share – a partilha dos dados envolve torná-los publicamente disponíveis, quando possível, ao fim de um projeto.

Preserve – se refere à maneira como os dados serão preservados depois do fim do projeto.

A gestão dos dados de investigação preocupa-se com todo o ciclo de vida dos dados, dentro do ciclo de vida do projeto de investigação. É desejável que investigadores e instituições que acolhem atividades de investigação desenvolvam habilidades para bem gerir seus conjuntos de dados, uma vez que as boas práticas de gestão garantem que os dados estejam apropriadamente preservados e tornados acessíveis a longo prazo, de tal forma que eles poderão estar aptos a serem reutilizados em investigações futuras. Também é recomendável que se faça um plano formal de gestão de dados de investigação, apropriadamente registado em documento, que contenha todos os aspetos de como serão tratados os dados provenientes de um projeto, durante o seu ciclo de vida e depois de seu encerramento.

Para além das comunidades científicas tradicionais, representadas em sua maioria pelas instituições de ensino, centros e grupos de investigação, a questão da gestão dos dados é tema que também está no topo das agendas de outros grupos que, de alguma forma, estão envolvidos com investigação, como as agências financiadoras, alguns editores de literatura científica, governos, e a própria sociedade interessada em descobertas e inovações científicas. As razões que movem estes grupos podem ser diferenciadas das que movem os primeiros, porém não menos importantes, sobretudo se os dados em questão são fruto de projetos financiados com recursos públicos. Neste contexto, os dados podem ser considerados como bens públicos, que devem estar acessíveis para a sociedade em geral.

Dados também podem ser solicitados por editores para verificação de resultados de investigação publicados, ou por agências financiadoras, para comprovação de fim de um projeto ou outra prestação de contas. Para os governos, a boa gestão de dados de investigação significa transparência nas suas relações com a comunidade, e economia, uma vez que dados bem geridos podem evitar novos investimentos de recolha para investigações futuras.

Dada a importância da gestão de dados de investigação no atual contexto de e-Science, considera-se desejável que políticas públicas sejam desenvolvidas e implantadas para orientar e padronizar a conduta dos investigadores e outros responsáveis em relação aos dados produzidos no âmbito da investigação científica. Para Hodson (2009) são necessárias políticas nos níveis internacional, nacional e institucional, as quais devem estar enraizadas nas ações que os próprios investigadores terão de tomar. Tais políticas são fundamentais para criar uma cultura de boas práticas de gestão de dados.

Em 2009, a Revista Nature lança uma nota sobre a negligência em relação aos dados de investigação e sua partilha, apesar de sua importância para o sucesso dos projetos de investigação. Justifica tal comportamento em razão de uma lacuna de orientações técnicas, institucionais e culturais necessárias para suportar o acesso aberto aos dados e advoga pela responsabilidade partilhada em fazê-lo. De fato, esta ainda pode ser considerada a realidade dos dias atuais. Afirma que as agências financiadoras e os investigadores devem garantir suporte de hardware e software necessários para armazenar os dados. Ainda alega que as instituições devem assumir a responsabilidade pela preservação de dados digitais e torná-los acessíveis em longo prazo (através de sua disponibilização em biblioteca ou repositórios digitais), sendo as próprias bibliotecas universitárias fortes candidatos a assumir este papel. Por fim, sustenta que universidades e disciplinas individuais precisam empreender um vigoroso programa de educação e sensibilização sobre os dados. Considera uma disciplina sobre gestão de dados tão importante quanto uma disciplina de estatística, e a primeira deveria fazer parte dos currículos dos cursos de pós-graduação tal qual a última (NATURE EDITOR, 2009, p. 145).

1.3 Curadoria de dados de investigação

Juntamente com a preocupação com a gestão dos dados, uma nova área de estudos tem se afirmado, que é a da curadoria digital. A maioria das pessoas já está familiarizada com curadores de museus ou curadores de coleções diversas (pinturas ou insetos, por exemplo), agora existem os curadores que zelam por coleções de objetos digitais, neste caso, os dados de investigação que devem estar em formato digital. O sentido de curadoria para objetos digitais é o mesmo das coleções já conhecidas, e tem a ver com cuidado para preservação ao longo do tempo, prevendo aspectos de descrição e contextualização. A curadoria digital de dados de investigação garante que estes sejam preservados, estejam disponíveis e possam ser acessados e reutilizados pelo público interessado.

A nova maneira de fazer ciência imposta pela e-Science, colaborativa, interdisciplinar e, sobretudo, baseada na tecnologia e no acesso e partilha de coleções de dados de investigação, também trouxe desafios à curadoria digital. As mudanças tecnológicas e a flexibilidade na utilização das ferramentas de tecnologia da informação colocam os dados em risco e fazem crescer problemas de estratégia e de questões práticas relacionadas à criação, gestão e cuidado dos dados de investigação (LORD *et al.*, 2004).

A fragilidade inerente às mídias digitais e a obsolescência tecnológica que colocam os dados em risco, são problemas que podem ser sanados com as atividades de curadoria digital. Para Abbott (2008) todas as atividades envolvidas na gestão de dados, boas práticas de digitalização e documentação, e garantia de disponibilidade e adequação para descoberta e reuso dos dados no futuro, fazem parte da curadoria digital. Tais atividades podem envolver tecnologias para conversão dos dados em diferentes formatos digitais, descrição de conjuntos de dados (metadados), armazenamento e acesso aos dados de investigação (plataformas digitais).

Curadoria digital, de acordo com o Digital Curation Centre, centro de investigação em curadoria de informação digital no Reino Unido, envolve “a manutenção, preservação e adição de valor aos dados de investigação digitais ao longo

do seu ciclo de vida”¹¹. Acrescenta ainda que a curadoria digital reduz a duplicação de esforços na criação de dados e aumenta o valor de longo prazo dos dados existentes, tornando-os disponíveis para futuras investigações. Além disso, a gestão ativa dos dados reduz ameaças ao valor de investigação de longo prazo e ainda diminui o risco de obsolescência digital.

Para o DCC, a curadoria digital e a preservação de dados são processos que devem atuar em sinergia e requerem atenção e investimentos de tempo e recursos adequados. O ciclo de vida da curadoria digital, proposto pelo DCC, que compreende passos desde a concepção ou recolha dos dados, também prevê ações de preservação. Curar e preservar são atividades-chave que se encontram no centro do modelo¹² que pode ser visualizado na Figura 3.

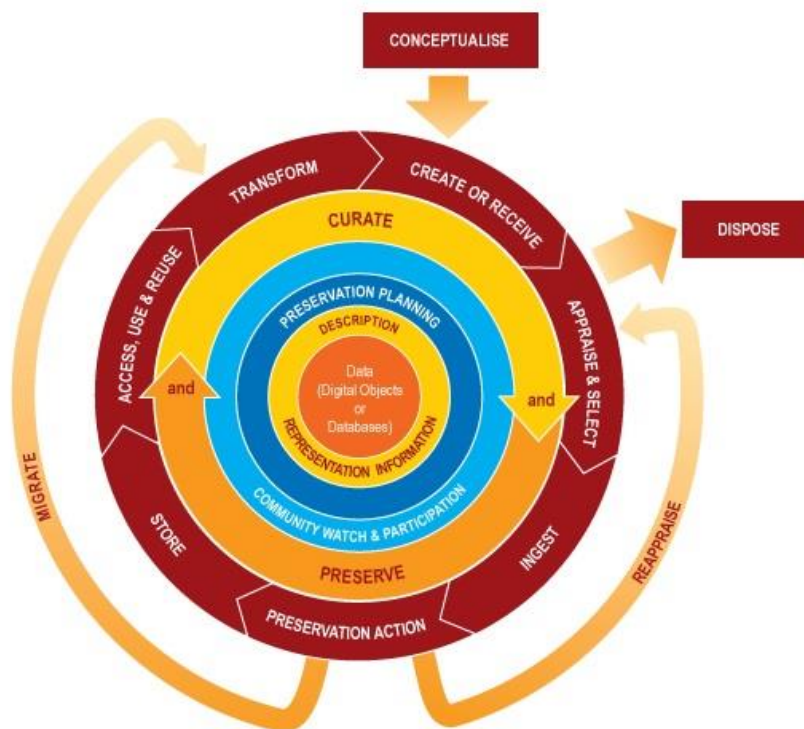


Figura 3 – Ciclo de vida da curadoria digital (Digital Curation Centre, 2014.)

¹¹ “Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle”. Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>

¹² **DCC Curation Lifecycle Model**. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

No relatório JISC sobre curadoria de dados da e-Science, Lord e Macdonald (2004) também colocam as atividades de curadoria digital e de preservação em paralelo ao diferenciar uma da outra, e ainda acrescentam a estas duas a atividade de arquivamento, apesar desta última compartilhar características com a preservação. Assim, estas seriam as três atividades fundamentais para a curadoria na e-Science.

Curadoria para os autores é definida como a atividade de gestão e promoção do uso de dados desde a sua criação, para assegurar que estejam adequados para finalidades contemporâneas e disponíveis para reuso. Esta atividade pode significar o enriquecimento ou atualização dos dados para que permaneçam utilizáveis (LORD e MACDONALD, 2004). Em adição, Lord *et al.* (2004) afirmam que níveis mais elevados de curadoria também poderão envolver a manutenção de ligações web e outros materiais publicados.

O arquivamento é uma atividade de curadoria que garante que os dados sejam adequadamente selecionados, armazenados, possam ser acedidos e que a sua integridade física e lógica esteja mantida ao longo do tempo, incluindo a segurança e autenticidade (LORD e MACDONALD, 2004). No modelo do ciclo de vida da curadoria digital (figura 3) os passos de avaliar e selecionar (*appraise and select*), ingestão dos dados em ambientes seguros (*ingest*), armazenamento (*store*), e acesso, utilização e reutilização (*access, use and reuse*) estão relacionados à atividade de arquivamento.

A preservação é uma atividade de arquivamento em que os itens específicos de dados são mantidos ao longo do tempo para que eles ainda possam ser acessados e compreendidos mesmo depois de mudanças sucessivas e obsolescência de tecnologias (LORD e MACDONALD, 2004). Desta forma, todos os passos do modelo do ciclo de vida da curadoria digital (Figura 3) acima citados, que estão relacionados com o arquivamento, somados às ações de preservação (*preservation action*) fazem parte desta atividade.

Portanto, estratégias de curadoria, preservação e armazenamento são concomitantes na gestão de dados de investigação, e essenciais para que estes estejam seguros e inteligíveis, e assim possam ser partilhados e reutilizados pelas comunidades de investigadores. Iniciativas em direção à curadoria de dados podem ser vistas nas comunidades científicas através de repositórios temáticos de dados, que armazenam dados de apenas uma área do conhecimento; repositórios institucionais de dados, que procuram englobar todos os domínios trabalhados em determinada universidade, por ser

um serviço institucional; e centros de dados, que fornecem as orientações, conhecimento e ferramentas necessárias para a curadoria profissional de dados.

No entanto, esta ainda é uma realidade para poucos. Países como Estados Unidos da América, Austrália e parte do Reino Unido são exemplos mais frequentemente encontrados na literatura, pois estão mais avançados em relação a políticas públicas, estratégias e práticas, bem como infraestruturas para a gestão de dados de investigação, enquanto existem outros países onde a mera discussão do tema ainda é bastante incipiente. No Brasil, por exemplo, a temática ainda é pouco explorada, segundo informações de Cunha (2014), apresentadas no SNBU - Seminário Nacional de Bibliotecas Universitárias¹³. O autor não menciona nenhuma iniciativa concretizada, além de um possível envolvimento da EMBRAPA Informática Agropecuária¹⁴, instituição pública brasileira líder em pesquisa na área da agropecuária, e alguns poucos eventos realizados entre os anos de 2012 e 2014.

O mesmo se pode dizer em relação a grandes e pequenos grupos de investigação. Enquanto centros de investigação de áreas mais consolidadas possuem os recursos e infraestruturas necessárias para empreender atividades de curadoria, pequenos grupos de investigação que constituem a chamada “pequena ciência”, normalmente oriundos de instituições universitárias, não possuem nem um nem outro para cuidar dos conjuntos de dados para além do termo dos seus projetos de investigação (RODRIGUES *et al.*, 2010).

No caso destes pequenos grupos de investigadores que trabalham dentro das universidades, pode-se notar que existe uma demanda por curadoria digital de seus dados de investigação, que, no entanto, resultam em soluções “caseiras”. Cada grupo faz o armazenamento de seus conjuntos de dados da forma que acha mais adequada, seja nos seus laboratórios, computadores pessoais ou nas nuvens, e não se verifica nenhuma prática de adição de informação extra sobre os dados. Diante disso, acabam por depender de iniciativas advindas da Universidade, como o caso dos repositórios institucionais.

¹³ SNBU 2014. Disponível em: <https://www.bu.ufmg.br/snbu2014/>

¹⁴ EMBRAPA Informática Agropecuária. Disponível em: <https://www.embrapa.br/informatica-agropecuaria>

O desafio dos repositórios institucionais como solução de curadoria de dados de investigação para estes pequenos grupos ou mesmo investigadores individuais é precisamente a “necessidade de conjugar a dimensão institucional (muito ampla e multidisciplinar no caso das universidades) com a dimensão disciplinar (com os seus requisitos específicos)” (RODRIGUES *et al.*, 2010). Os repositórios institucionais não possuem capacidade semântica para representar as especificidades dos diferentes conjuntos de dados produzidos nos diferentes domínios. Os padrões de metadados usados nestes repositórios precisam ser genéricos o suficiente para contemplar todo o tipo de descrição, o que negligencia aspetos que podem ser considerados importantes na descrição de dados de domínios específicos.

1.4 Metadados em curadoria digital: perfis de aplicação e ontologias

A definição de curadoria digital do DCC afirma que uma das atividades de curadoria é a “adição de valor” aos dados de investigação ao longo de seu ciclo de vida. Adicionar valor aos dados significa adicionar informação extra sobre eles e o contexto onde foram criados, para garantir que estejam bem documentados. Alguma desta informação pode ser: como e porque foram criados; como foram preparados ou digitalizados; o que significam, qual seu conteúdo e estrutura; ou quaisquer alterações ou codificações que os dados tenham sofrido durante seu ciclo de vida.

A boa documentação, isto é, boa descrição dos dados, é crítica para o entendimento dos dados em curto, médio e mais longo prazo, e é vital para o sucesso da preservação dos dados a longo prazo (CORTI *et al.*, 2014). Também é fundamental para as práticas de partilha e reutilização, pois é necessário que os dados sejam entendidos por outros investigadores, ou mesmo pelos próprios investigadores que os criaram depois de passado algum tempo.

Outro termo encontrado na literatura sobre curadoria digital usado para designar a documentação é “anotação”. Abbott (2008) define anotação como o processo de adicionar notas sobre ou para alguma coisa. A autora afirma que a anotação desempenha um papel importante no trabalho académico, uma vez que explicações, descrições e interpretações podem informar futuras pesquisas sobre os dados anotados. Para ela, os metadados são tipo o de anotação descritiva crucial para recuperação da informação.

É aconselhável que se comece a fazer a descrição dos dados de investigação logo no início do projeto quando os mesmos são gerados, e que se continue através do seu ciclo de vida, pois isso torna a tarefa mais eficiente e os investigadores não correm o risco de perder alguma informação importante. A organização dos dados pode ser fundamental para o sucesso do projeto de investigação.

1.4.1 Metadados

Dentre as atividades a que se dedica a curadoria de dados de investigação, destaca-se aqui a utilização de metadados normalizados, essenciais para descrição, acesso e reutilização dos dados de investigação. Eles são estruturas organizadas que orientam a forma como os dados serão descritos no seu armazenamento em uma plataforma digital, e ainda fornecem pontos de acesso para que a informação seja posteriormente recuperada.

Para Corti *et al.* (2014) os metadados são considerados um subconjunto específico de documentação de dados, aos quais fornecem informação estruturada e padronizada que explica a proposta, origem, referências de tempo, localização geográfica, e cria autoria, condições de acesso e termos de uso da coleção de dados.

De acordo com a NISO (National Information Standards Organization, EUA)¹⁵ (2004) o termo metadados é usado de forma diferente em diferentes comunidades. Alguns o usam para se referir à informação entendível por máquina, enquanto outros usam o termo apenas para os registos que descrevem recursos eletrônicos. Para a NISO, além de descrever, explicar e localizar um recurso, os metadados tornam a recuperação, uso ou gerenciamento de um recurso de informação mais fácil.

Na definição do DCMI (Dublin Core Metadata Initiative)¹⁶ a palavra "metadados" significa "dados sobre dados", definição comum encontrada na literatura em razão do prefixo "meta". Ainda, para o DCMI metadados estão ligados à descrição de recursos e articulam um contexto para objetos de interesse - "recursos" como

¹⁵ National Information Standards Organization. **Understanding metadata**. Disponível em: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>

¹⁶ Dublin Core Metadata Initiative. **Metadata basics**. Disponível em: <http://dublincore.org/metadata-basics/>

arquivos MP3, livros da biblioteca, ou imagens de satélite - sob a forma de "descrições de recursos".

No guia da ANDS¹⁷ o termo metadados refere-se à informação usada para descrever itens físicos e itens digitais (ficheiros, documentos, imagens, conjuntos de dados, etc.). Um catálogo da biblioteca, por exemplo, é composto de metadados que descrevem os livros, revistas e outros recursos detidos pela biblioteca. Ainda, os metadados podem descrever itens, quando a descrição é de um único objeto, e grupos de itens, quando a descrição é feita ao nível de coleção. Acrescenta que os metadados agregam valor aos recursos, o que para os dados de investigação é ainda mais importante, uma vez que estes sem o contexto fornecido pelos metadados não seriam entendidos.

Quanto à tipologia os elementos de metadados podem ser: **descritivos** (descrevem um recurso para fins como a descoberta e identificação, pode incluir elementos como título, resumo, autor e palavras-chave); **estruturais** (indicam como objetos compostos são colocados juntos, por exemplo, como as páginas são ordenadas para formar capítulos; ou, no caso de dados de investigação, como os diferentes componentes de um conjunto de dados associados se relacionam entre si); e **administrativos** (fornecem informações para ajudar a gerenciar um recurso, como quando e como ele foi criado, tipo de arquivo e outras informações técnicas, e quem pode acessá-lo; dois subconjuntos de metadados administrativos são os metadados de gestão de direitos, que tratam de direitos de propriedade intelectual, e os metadados de preservação, que contêm a informação necessária para arquivar e preservar recurso) (NISO, 2004, MANTRA, 2014).

Uma razão importante para a criação de metadados descritivos é facilitar a descoberta de informação relevante. Além da descoberta de recursos, os metadados podem ajudar a organizar os recursos eletrônicos, facilitar a interoperabilidade e integração de recursos de legado, fornecer identificação digital e suporte de arquivamento e preservação (NISO, 2004).

Existem muitos padrões ou esquemas de metadados prontos que são conhecidos, um exemplo a destacar é o Dublin Core (DC)¹⁸, que foi criado para descrever recursos

¹⁷ ANDS Guides. **Metadata**. Disponível em: <http://ands.org.au/guides/metadata-awareness.html>

¹⁸ Mais informações em: <http://dublincore.org/>

digitais na web. As principais características do DC são sua simplicidade de uso, que permite que mesmo pessoas sem experiência em indexação possam usá-lo para descrever recursos, e seu escopo internacional, que garante a interoperabilidade de registros, já que é largamente conhecido e utilizado em todo o mundo. O conjunto de elementos de metadados Dublin Core é um vocabulário formado por quinze propriedades para uso na descrição de recursos. São elementos bastante amplos e genéricos, portanto, servem para descrever toda a gama de recursos disponíveis na web. Como se pode visualizar na Figura 4, os elementos básicos do DC se assemelham aos dos catálogos online utilizados para descrever os recursos bibliográficos nas bibliotecas. No entanto, o DC é amplamente usado fora desta comunidade.

Elemento	Descrição
Title	O nome do recurso.
Creator	A principal entidade responsável por criar o recurso. Ex.: pode ser uma pessoa, organização ou serviço.
Subject	Os assuntos do recurso. Ex.: palavras-chave, termos extraídos de vocabulários controlados, etc.
Description	Descrição do recurso. Ex.: um resumo, uma tabela de conteúdo, uma representação gráfica ou uma descrição de texto livre do recurso.
Publisher	A entidade responsável por tornar o recurso disponível.
Contributor	Uma entidade responsável por fazer contribuições para o recurso. Ex.: pode ser uma pessoa, organização ou serviço.
Date	Um ponto ou período de tempo, associado a um evento do ciclo de vida do recurso.
Type	A natureza ou gênero do recurso. Ex.: texto, som, imagem, software.
Format	O formato de arquivo, suporte físico ou dimensões do recurso. Ex.: dimensões de tamanho e duração (jpeg, pdf, duração de 30 min., tamanho 1 GB, etc).
Identifier	Uma referência não ambígua para o recurso em um determinado contexto. Ex.: ISSN, ISBN, DOI.
Source	Um recurso relacionado a partir do qual o recurso descrito é derivado. Ex.: uma URL.
Language	A língua em que o recurso está registrado.
Relation	Um recurso relacionado. Ex.: se a obra faz parte de uma coleção, se é a mesma obra em outro suporte, se é uma adaptação (filme) etc.
Coverage	Tópico espacial ou temporal do recurso, ou a jurisdição em que o recurso é relevante. Ex.: nome de um local, coordenadas geográficas, datas, períodos de tempo, etc.
Rights	Informação sobre os direitos do recurso. Ex.: propriedade intelectual, etc.

Fonte: Adaptado de Dublin Core Metadata Initiative - <http://dublincore.org/documents/dces/>

Figura 4 – Conjunto de 15 elementos básicos Dublin Core (Dublin Core Metadata Initiative, 2014)

O conjunto dos quinze elementos básicos o DC faz parte de um conjunto maior de vocabulários e especificações técnicas mantidos pelo Dublin Core Metadata Initiative (DCMI). Citam-se aqui os elementos de metadados qualificados, que são um refinamento dos primeiros, a serem utilizados quando estes são insuficientes para descrever algum recurso. O DC permite ainda que comunidades adicionem ou qualifiquem elementos de acordo com suas necessidades de descrição. Em relação à descrição de conjuntos de dados de investigação a comunidade Dublin Core se beneficia de um fórum de discussão, onde indivíduos e organizações podem trocar informação e conhecimento sobre o tema. Denominado “Science & Metadata Community”¹⁹, o fórum está focado nos desafios de metadados específicos para curadoria de dados de investigação.

Em relação à gestão de dados de investigação, o editorial da revista Nature de 2009 relaciona o software de gerenciamento de metadados como uma importante faceta deste processo. Se refere aos metadados como as ferramentas que agilizam o processo de anotação de dados com uma descrição do que os dados querem dizer, qual o instrumento que os recolheu, que algoritmos têm sido usados para processá-los e assim por diante - a informação que é essencial para que outros cientistas possam reutilizar os dados de forma eficaz (NATURE Editor, 2009, p. 145).

Exemplo de um projeto que envolve a gestão de metadados é o da Biblioteca Central do “National Oceanic Atmospheric Administration – NOAA”²⁰, que em colaboração com o Gabinete do “Ocean Exploration and Research” (OER), visa atender aos requisitos para o uso de metadados, arquivamento, preservação, gerenciamento e fornecimento de acesso online para vídeos digitais e de imagens estáticas de suas expedições oceanográficas.

O projeto, nomeado “Video Data Management System” - VDMS foi desenvolvido para criar e gerenciar as informações de metadados para fornecer acesso on-line e off-line para vídeo e imagens obtidas durante expedições oceanográficas. As diretrizes de metadados desenvolvidas durante o projeto ajudaram cientistas e gestores de dados a criar metadados descritivos e técnicos para os seus dados de vídeo e imagem. Cientistas, bibliotecários e arquivistas, em seguida, usaram essas informações para criar

¹⁹ Mais informações em: <http://dublincore.org/groups/scidata/>

²⁰ National Oceanic Atmospheric Administration – <http://www.noaa.gov/>

registos de metadados MARC 21²¹, MARC XML²² e The Federal Geographic Data Committee (FGDC)²³.

1.4.2 Perfis de aplicação

Uma descrição clara requer um conhecimento profundo dos dados e do seu processo de criação, fatores que são diferentes em cada domínio de conhecimento. A criação de metadados normalizados é importante para garantir interoperabilidade, no entanto não contempla todas as características e especificidades das diferentes áreas de domínio. Logo, diferentes áreas de investigação demandam a criação de diferentes conjuntos de metadados, para que os dados possam ser descritos apropriadamente no que diz respeito às suas peculiaridades.

A necessidade de descrever conjuntos de dados de diferentes domínios tem levado à criação de perfis de aplicação (CASTRO, RIBEIRO e SILVA, 2013). Para Heery e Patel (2000) os perfis de aplicação (do inglês *application profile*) são um tipo de esquema de metadados que consistem de elementos de dados retirados de um ou mais esquemas existentes, combinados entre si e otimizados para uma aplicação local em particular. É o que as autoras chamam de abordagem “mix and match”.

A partir de 2000, a comunidade Dublin Core colocou também o seu foco nos perfis de aplicação - a ideia de que registos de metadados usariam o Dublin Core juntamente com outros vocabulários especializados para atender aos requisitos específicos de aplicação. A DCMI, entendendo que contextos específicos requerem descritores específicos, também fornece orientações para a criação de perfis de aplicação²⁴. O documento explica os principais componentes de um perfil de aplicação Dublin Core e guia o utilizador através do processo de desenvolvimento de um perfil.

²¹ Mais informações em Library of Congress – MARC 21 Format of Bibliographic Data: <http://www.loc.gov/marc/bibliographic/>

²² Mais informações em Library of Congress – MARCXML MARC 21 XML Schema: <http://www.loc.gov/standards/marcxml/>

²³ Mais informações em: <https://www.fgdc.gov/metadata/geospatial-metadata-standards>

²⁴ Guidelines for Dublin Core Application Profiles. Disponível em: <http://dublincore.org/documents/profile-guidelines/>

Heery e Patel (2000) citam as características de um perfil de aplicação, das quais foram destacadas três:

Um perfil de aplicação pode recorrer a um ou mais nomes de esquemas existentes - pode usar elementos de um ou mais conjuntos diferentes de elementos, mas o perfil de aplicação não pode criar novos elementos não definidos nos nomes de esquemas existentes.

Um perfil de aplicação não introduz novos elementos de dados - todos os elementos de um perfil de aplicação são retirados de outros lugares, a partir de esquemas de nomes distintos. Se um implementador pretende criar elementos "novos" que não existem em outros lugares, então, eles devem criar o seu próprio nome de esquema, e assumir a responsabilidade de declarar e manter esse esquema.

Um perfil de aplicação pode refinar as definições padrão - O perfil de aplicação pode refinar as definições dentro do esquema, mas só pode fazer a definição semanticamente mais estreita ou mais específica. Isto leva em conta situações onde determinadas implementações usam domínio específico, ou recurso específico de linguagem.

Castro, Ribeiro e Silva (2013) afirmam que a quantidade de esforço que os investigadores precisam investir para descrever seus dados pode limitar a sua disposição em partilhá-los. Daí que, em vez de fazer os investigadores seguirem esquemas estritos de metadados, propõem que eles devem ter acesso a um perfil de aplicação feito sob medida para o seu próprio domínio.

Os autores acima citados conduziram na Faculdade de Engenharia da Universidade do Porto (FEUP) um experimento onde criaram um perfil de aplicação para ajudar investigadores da área de Engenharia Mecânica na de gestão dos seus dados gerados nas atividades de investigação no domínio da mecânica das fraturas. O perfil de aplicação combinou elementos Dublin Core, Ecological Metadata Language²⁵ e elementos específicos do domínio.

²⁵ Mais informações em: <http://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language>

TABLE 2- Fracture Mechanics application profile

Descriptor	Description	Example
eml:methods	Procedures that are used in the creation or the subsequent processing of a dataset	Free text
eml: instrumentation	Description of the instruments used in the data collection or quality control and quality assurance	INSTROM-1125
fm:specimen	Type of specimen used in the experiment	Pinus pinaster (Wood)
fm:specimenLength	Specimen geometric length	L= 400 mm
fm:specimenWidth	Specimen geometric width	B=20mm
fm:specimenHeight	Specimen geometric height	2h = 20 mm
fm:specimenInitialCrackLength	The crack in the double cantilever beam specimen prior to the fracture test	ao= 150mm

Figura 5 – Excerto de um perfil de aplicação em Mecânica das Fraturas (CASTRO, J. A.; RIBEIRO, C. e SILVA, J. R. da. Designing an Application Profile using qualified Dublin Core: a case study with fracture mechanics datasets.)

A Figura 5 exibe elementos de prefixo “eml”, que foram extraídos do esquema de metadados EML (Ecological Metadata Language), e “fm”, que foram criados especificamente para o domínio e denominado “Fracture Mechanics”. A Figura 6 exibe os elementos de prefixo “dc” selecionados do esquema de metadados Dublin Core.

TABLE 1 – Dublin Core terms used in the Fracture Mechanics application profile

DC		Qualified DC
dc:title (required)	dc:date (required)	dcterms:references (if available)
dc:subject (required)	dc:identifier (if available)	dcterms:isReferencedBy(if available)
dc:description (required)		dcterms:format:medium (recommended)

Figura 6 – Excerto de um perfil de aplicação em Mecânica das Fraturas – elementos Dublin Core (CASTRO, J. A.; RIBEIRO, C. e SILVA, J. R. da. Designing an Application Profile using qualified Dublin Core: a case study with fracture mechanics datasets.)

Os autores ressaltam a importância de usar elementos de metadados existentes para manter-se certo grau de interoperabilidade nos registros dos conjuntos de dados, o que permite que as comunidades de investigadores possam trabalhar de acordo com os

parâmetros da e-Science, ou seja, de forma colaborativa e compartilhando material digital de forma eficiente.

O objetivo de criar um perfil de aplicação é, sobretudo, apresentar uma ferramenta de curadoria digital para que os investigadores possam adicionar valor aos seus dados de investigação da maneira mais expedita. Os elementos de metadados são criados, selecionados e combinados para se ajustar às necessidades de descrição de dados de investigação de domínios específicos, de forma a garantir precisão e abrangência necessárias. Além disso, o perfil de aplicação não oferece dificuldades para a anotação dos dados, pois os investigadores estão familiarizados com a linguagem que os descritores do perfil apresentam.

1.4.3 Ontologias

Mesmo um perfil de aplicação pode apresentar alguma dificuldade de compreensão e adoção em plataformas para armazenamento e gestão de dados de investigação. De acordo com Castro, Silva e Ribeiro (2014), as ontologias têm sido apresentadas como uma possível solução para a descrição de dados, pois satisfazem os requisitos desejáveis de metadados e são capazes de representar as especificidades semânticas de cada domínio de investigação. Portanto, uma ontologia é uma boa solução para representar um perfil de aplicação dentro de uma plataforma web.

As ontologias podem ser de diversos tipos e são utilizadas em diversas áreas do conhecimento para organizar informação (ALMEIDA e BAX, 2003), sua adoção tem aumentado muito na medida em que cresce o volume de informação digital na web. Considerando o atual dilúvio de dados de investigação produzidos nas mais variadas áreas do conhecimento se pode inferir que as ontologias são ferramentas apropriadas para representar e organizar tal informação, absorvendo o vocabulário específico de cada área. Segundo Almeida (2014) o termo ontologia tem sido usado desde os anos de 1960 para se referir a uma estrutura de conceitos representados por um vocabulário lógico. Nos anos de 1990, o termo manteve sua presença no contexto do conjunto de tecnologias que se convencionou rotular de “Web Semântica”.

A Web Semântica é uma iniciativa de Tim Berners-Lee, criador da Web, para tentar melhorar (ou mesmo otimizar) a representação e a exploração da informação na

web (PICKLER, 2007). Uma explicação sobre o que faz a Web Semântica pode ser encontrada no sítio eletrônico do World Wide Web Consortium (W3C)²⁶, onde se lê:

Em adição a clássica “Web dos documentos” a W3C está ajudando a construir uma pilha de tecnologias para apoiar a “Web de dados”, os tipos de dados que se pode encontrar em bancos de dados. O objetivo final da Web de dados é possibilitar que os computadores sejam capazes de fazer o trabalho mais útil e desenvolver sistemas que podem apoiar interações confiáveis através da rede. O termo “Web Semântica” refere-se a visão da W3C da Web de dados ligados. As tecnologias da Web Semântica permitem às pessoas criar armazéns de dados na Web, construir vocabulários, e escrever regras para manipulação de dados. Dados ligados são capacitados por tecnologias como RDF, SPARQL, OWL, e SKOS (tradução nossa).

O termo ontologia, inicialmente empregado em Filosofia, foi incorporado pelas Ciências da Informação e Computação. “Em Ciência da Computação, ontologias são aplicadas à modelagem, tanto em sistemas baseados em bancos de dados quanto em sistemas de representação do conhecimento” (ALMEIDA, 2014). Na área de Organização e Representação do Conhecimento uma ontologia é um esquema legível por computador que possui a propriedade de representar um mundo específico, é usada para facilitar o compartilhamento e reutilização da informação na web, ao mesmo tempo em que confere um contexto semântico à web.

Segundo Guarino (1998) apud Almeida e Bax (2003) uma “ontologia se refere a um artefato constituído por um vocabulário usado para descrever uma certa realidade, mais um conjunto de fatos explícitos e aceitos que dizem respeito ao sentido pretendido para as palavras do vocabulário”. No contexto da descrição de dados a ontologia corresponde a uma linguagem formal que representa determinada realidade, ou seja, o perfil de aplicação com vocabulário específico aceito na sua comunidade de utilizadores.

²⁶ In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS. Disponível em: <http://www.w3.org/standards/semanticweb/>

Silva *et al.* (2014) apresentam um experimento conduzido no uso de ontologias para a descrição de conjuntos de dados do domínio da biodiversidade. O grupo adotou recomendação europeia da INSPIRE como primeira ferramenta de metadados para descrição dos dados e modelou com base nesta a ontologia que se chamou BIOME. A ontologia combinou conceitos da INSPIRE²⁷ com elementos definidos em outras ontologias, como o Dublin Core.

Para os autores as linguagens ontológicas são tecnologias de representação do conhecimento de uso geral e podem ser adotadas para captar a natureza de registos de metadados. Elas transmitem não só as regras sintáticas que impõem a correção de um registo de metadados, mas também o significado de cada descritor utilizado em um registo de uma maneira processável por máquina. As ontologias são essenciais para a descrição de recursos na Web Semântica (SILVA *et al.*, 2014).

Investigadores da área de curadoria digital da FEUP também empreendem outras iniciativas no uso de ontologias para descrição de conjuntos de dados de investigação. Castro, Silva e Ribeiro (2014) modularam ontologias para descrição de conjuntos de dados nos domínios de mecânica das fraturas e química analítica. Atualmente, estão desenvolvendo o Dendro²⁸, uma plataforma para gestão de dados de investigação baseada em ontologia, que tem o objetivo de ser multidisciplinar, uma vez que abarca ontologias de diversos domínios.

²⁷ Mais informações em: <http://inspire.ec.europa.eu/index.cfm/pageid/48>

²⁸ Mais informações em: <http://dendro.fe.up.pt/#>

2 A investigação em Oceanografia Biológica em contexto organizacional

2.1 Universidade Federal do Rio Grande (FURG) – uma Universidade voltada para o ecossistema costeiro

A Universidade Federal do Rio Grande (FURG)²⁹ é uma instituição pública de ensino superior, situada na região sul do Brasil, com sede na cidade de Rio Grande, estado do Rio Grande do Sul. Nasceu na segunda metade do século XX quando cresceu a preocupação com a evasão dos estudantes que iam para outras cidades a fim de completar seus estudos, uma vez que em Rio Grande não havia escolas de nível superior. Estes estudantes raramente voltavam para dar sua contribuição para o crescimento e desenvolvimento local. Então, na década de 1950, a partir da iniciativa de intelectuais dispostos a mudar esta realidade, instituiu-se a Escola de Engenharia, cuja área apresentava grande demanda de profissionais na época.

A seguir à Escola de Engenharia surgiram outras faculdades, nomeadamente a Faculdade de Ciências Políticas e Econômicas e a Faculdade de Direito, esta mantida financeiramente pela Mitra Diocesana da cidade vizinha, Pelotas, porém com funcionamento em Rio Grande. Em função do vínculo cultural estabelecido entre as duas cidades também se criou a Faculdade Católica de Filosofia de Rio Grande, que oferecia os cursos de Filosofia e Pedagogia. Assim se sucedeu a criação de outros cursos de nível superior na cidade que, em 1968, já podia oferecer um bom número deles.

O surgimento oficial da Universidade se deu em 1969, quando a filosofia educacional do país admitia, a título precário, o funcionamento de escolas isoladas no sistema de ensino superior, e a Reforma Universitária preconizava a aglutinação de unidades independentes, menores, em complexos estruturais maiores, organizados em função de objetivos comuns. Assim, foi assinado o Decreto-Lei n.º 774, pelo então Presidente da República Arthur da Costa e Silva, autorizando o funcionamento da Universidade do Rio Grande - URG. Surgia, então, a 20 de agosto de 1969, no cenário cultural brasileiro, esta Universidade que aqui se apresenta.

²⁹ Universidade Federal do Rio Grande - <http://furg.br/>

Em 1987 a FURG passa à condição de Fundação Pública, com seu funcionamento custeado precipuamente por recursos da União Federal. Hoje a Universidade tem grande importância no cenário educacional nacional, sobretudo no que se refere ao oferecimento de formação em cursos ligados ao meio ambiente marinho e costeiro. Inclusive, o ano de 1987 também marca a definição, pelo Conselho Universitário, da Filosofia e Política para a Universidade do Rio Grande, mediante a qual esta assume como vocação institucional o Ecossistema Costeiro, que orientará as atividades de ensino, pesquisa e extensão.

A FURG é uma instituição que se insere de maneira particularmente privilegiada num ecossistema costeiro, do qual formula sua orientação filosófica vocacionada para as características históricas, culturais e sociais próprias de sua posição ambiental regional. Voltar-se ao ecossistema costeiro implica enunciar alternativas reais de desenvolvimento harmônico entre sociedade e natureza, especialmente na região costeira em que está inserida. (Projeto Pedagógico Institucional, 2011-2022).

Os ambientes costeiros e oceânicos adjacentes são áreas especiais no planeta. Eles se formam e se caracterizam através da interação de processos entre a atmosfera, o continente e a região oceânica. Essa confluência de processos reflete-se numa concentração de energias naturais que tornam essas áreas altamente dinâmicas, produtivas e com características únicas. Tais áreas, por sua capacidade produtiva e pela facilidade de transporte e acesso pelo meio hídrico, têm sido historicamente ocupadas pelo homem de forma destacada (Projeto Pedagógico Institucional, 2011-2022).

Essa ocupação humana numa área de elevada produtividade e oferta de recursos ambientais produziu seu desenvolvimento e estabeleceu suas peculiaridades socioeconômicas, históricas e culturais. Todos esses elementos estabeleceram o que hoje reconhecemos como os ecossistemas costeiros e oceânicos. Portanto, para um completo conhecimento e o desejado uso sustentável desses ecossistemas com bem-estar social, é imperativo seu estudo e a formação de profissionais que nele atuem através de um enfoque multidisciplinar e integrado. Nesse sentido, seu completo conhecimento torna-se indissociável dos aportes necessários das Ciências Biológicas, Agrárias, Exatas e da Terra, Saúde, Humanas, Sociais Aplicadas, Engenharias e Linguística, Letras e Artes (Projeto Pedagógico Institucional, 2011-2022).

A visão da Universidade também está ligada à sua vocação institucional. Sendo assim, pretende consolidar sua imagem, nacional e internacionalmente, como referência

em educação, desenvolvimento tecnológico e estudo dos ecossistemas costeiros e oceânicos. Sua missão está igualmente comprometida com o meio ambiente, sendo “promover o avanço do conhecimento e a educação plena com excelência, formando profissionais capazes de contribuir para o desenvolvimento humano e a melhoria da qualidade socioambiental”.

Abaixo são citadas algumas datas importantes na história da Universidade que se relacionam com sua vocação ambiental, fortemente voltada aos ecossistemas costeiros:

1970 - É criado o curso de Oceanologia, pioneiro no Brasil em nível de graduação, formando profissionais nas linhas biológicas e geológicas.

1975 - No dia 25 de julho é reconhecido o curso de Oceanologia, pelo Decreto nº 76.028. No mesmo ano, buscando desenvolver a pesquisa marinha, é implantado o Projeto Atlântico, cujo principal objetivo era o desenvolvimento da pesquisa tecnológica na região oceânica sul-brasileira. Na época, muito contribuiu para os estudos realizados a doação que a FURG recebeu da Fundação Cidade do Rio Grande: o Museu Oceanográfico.

1978 - São inaugurados, no Campus Carreiros, os primeiros prédios da Base Oceanográfica Atlântica. É onde se localizam a maioria dos laboratórios de pesquisa em Oceanografia.

1979 - É criado o Curso de Mestrado em Oceanografia Biológica, primeiro curso de pós-graduação *stricto sensu* da FURG.

1982 - A Universidade do Rio Grande é convidada a participar do projeto Antártico Brasileiro – PROANTAR, mediante convênio celebrado entre a FURG e a Comissão Interministerial para os Recursos do Mar – CIRM –, considerando a política governamental brasileira de exploração do continente antártico e a localização geográfica e vocacional eminentemente marítima da Universidade.

1983 - É inaugurado um pavilhão que representa a primeira fase de construção da Estação de Apoio Antártico – ESANTAR – pelos Ministros da Marinha, Maximiano Eduardo da Silva Fonseca; da Educação e Cultura, Hester de Figueiredo Ferraz, e pelo Reitor da FURG, Prof. Fernando Lopes Pedone.

1990 - Passa a ter desempenho operacional a Estação Marinha de Aquicultura, no Balneário Cassino, cujo projeto tem por objetivo principal a reprodução em cativeiro de camarão, peixe-rei e tainha.

1992 - É implementado o Doutorado em Oceanografia Biológica, primeiro curso de doutoramento da FURG.

1995 – Instalação do Programa Internacional Train-Sea Coast

1999 - Criação do Programa de Pós-Graduação em Aquicultura

2004 - É aprovado pelo Conselho de Ensino, Pesquisa e Extensão e começa a funcionar o Programa de Doutorado em Oceanografia Física, Química e Geológica.

2006 - Começam a funcionar os Programas de Pós-Graduação Doutorados em Educação Ambiental e em Ciências Fisiológicas - Fisiologia Animal Comparada e Mestrado em Biologia de Ambientes Aquáticos e Continentais. Também marca o início do Projeto Amazônia Azul: A Experiência Embarcada

2007 - Início do Doutorado do Programa de Pós-Graduação em Aquicultura

2009 - Criação do Programa de Pós-Graduação em Gerenciamento Costeiro

Atualmente, em termos de estrutura a FURG tem sua sede no Campus Carreiros da cidade de Rio Grande, onde funciona a Reitoria e a maior parte da sua Administração, bem como grande parte de todos os cursos oferecidos pela Universidade, e se ramifica por mais três cidades no Estado do Rio Grande do Sul, possuindo campi instalados e em funcionamento em Santa Vitória do Palmar, São Lourenço do Sul e Santo Antônio da Patrulha. Na cidade de Rio Grande possui ainda o Campus Saúde, onde funciona o Hospital Universitário, hospital escola que atende a comunidade através do Sistema Único de Saúde; o Complexo de Museus e Centros Associados, formado pelo Museu Oceanográfico “Eliézer de Carvalho Rios”, Museu Antártico, Eco-Museu da Ilha da Pólvora, Museu Náutico, Centro de Recuperação de Animais Marinhos (CRAM) e Centro de Convívio dos Meninos do Mar (CCMar); e a Estação Marinha de Aquicultura, situada no Balneário Cassino.

Com atenção à sua macroestrutura organizacional a FURG, estrutura-se nos termos do Estatuto da Universidade, em Gabinete do Reitor, Pró-Reitorias, Secretaria Executiva dos Conselhos, Assessorias Jurídica, de Relações Internacionais, de

Comunicação Social e de Projetos Estratégicos, Ouvidoria, Auditoria Interna, Comissões Permanentes e Órgãos Vinculados. São sete as Pró-Reitorias da Universidade, nomeadamente: Pró-Reitoria de Graduação (PROGRAD); Pró-Reitoria de Pesquisa e Pós-Graduação (PROPESP); Pró-Reitoria de Extensão e Cultura (PROEXC); Pró-Reitoria de Planejamento e Administração (PROPLAD); Pró-Reitoria de Assuntos Estudantis (PRAE); Pró-Reitoria de Infra-estrutura (PROINFRA); e Pró-Reitoria de Gestão e Desenvolvimento de Pessoas (PROGEP). Para além desta macroestrutura, são inúmeros os órgãos suplementares que estão vinculados à Reitoria e suas Pró-Reitorias.

Para o desenvolvimento das atividades de ensino, pesquisa e extensão, a FURG estrutura-se em Unidades Acadêmicas, entes perfeitamente definidos, com funções próprias e organizações semelhantes, instituídas como órgãos abertos a toda a Instituição. São treze ao total: Centro de Ciência Computacionais (C3), Escola de Enfermagem (EEnf), Escola de Engenharia (EE), Faculdade de Direito (FADIR), Faculdade de Medicina (FAMED), Instituto de Ciências Biológicas (ICB), Instituto de Ciências Econômicas, Administrativas e Contábeis (ICEAC), Instituto de Ciências Humanas e Informação (ICHI), Instituto de Educação (IE), Instituto de Letras e Artes (ILA), Instituto de Matemática, Estatística e Física (IMEF), Instituto de Oceanografia (IO), e Escola de Química e Alimentos (EQA).

Com relação à população universitária, o Boletim Estatístico 2013, produzido e editado pela Universidade, aponta que no ano de 2012 a FURG apresentava uma população total de 14.350 pessoas, das quais 2.235 eram estudantes de pós-graduação (especialização, mestrado e doutorado), 9.522 eram estudantes de graduação, 759 eram docentes (efetivos, substitutos e visitantes), e 1.098 eram do corpo Técnico Administrativo em Educação. A FURG ainda mantém, em conjunto com a Prefeitura Municipal, uma escola de Ensino Fundamental para atender basicamente as crianças da comunidade do entorno da Universidade, cuja população foi estimada em 736 alunos.

Ainda no que se refere a dados estatísticos, segundo informações oficiais do Boletim Estatístico 2013, que tem como base o ano anterior, em 2012 a FURG possuía 53 cursos de Graduação (nas modalidades presencial e à distância), 15 cursos de Especialização (nas modalidades presencial e à distância), 21 cursos de Mestrado e 11 cursos de Doutorado.

2.1.2 O Instituto de Oceanografia da FURG

O estudo oceanográfico sistemático teve início no Brasil com a criação do Instituto Paulista de Oceanografia (1946), entidade posteriormente absorvida pela Universidade de São Paulo, passando a denominar-se Instituto Oceanográfico. Desde então, foram criados diversos núcleos de pesquisas oceanográficas no país, sendo hoje raras as universidades brasileiras localizadas na região litorânea que não desenvolvam algum tipo de trabalho voltado para o mar.

A criação dos cursos de graduação foi decisiva para o desenvolvimento da Oceanografia no Brasil. O curso de Oceanologia da FURG, criado em 1970, o primeiro do país, deu início ao primeiro ciclo de criação de cursos da modalidade, que se estendeu até o começo da década de 90.

A gênese do curso de Oceanologia da FURG está na Sociedade de Estudos Oceanográficos do Rio Grande – SEORG, criada em 20 de março de 1953 por um grupo de entusiastas, entre os quais estavam Eliezer de Carvalho Rios, Boaventura Barcellos, Nicolas Vilhar e Cícero Vassão, que já há algum tempo pensavam em implantar na região um centro para estudos ligados ao oceano. Com o apoio da Prefeitura Municipal, que cedeu o prédio localizado no interior da Praça Tamandaré, onde permaneceu de 1953 a 1972, o Museu Oceanográfico passou a desenvolver pesquisas de laboratório e implantou uma exposição com seu acervo para visitação, contribuindo para despertar na população local o interesse pelas ciências do mar.

A estrutura e prática já utilizadas para a pesquisa oceanográfica desde os anos 1950 foi o embrião do curso de Oceanologia, que viria a ser criado pela Universidade de Rio Grande em 27 de agosto de 1970. Na falta de referências curriculares no Brasil, o Prof. Pery Riet Correa, da Faculdade de Medicina, enviou correspondências para centros universitários de diversos países, entre os quais Estados Unidos da América, França e Rússia, obtendo retorno de algumas universidades, inclusive com o oferecimento de apoio de pesquisadores para a organização do novo curso e a formação de recursos humanos.

No presente, o Instituto de Oceanografia (IO)³⁰ da FURG reúne todos os docentes que atuam na área de Oceanografia na Universidade, possui três Programas de Pós-Graduação (Oceanografia Biológica, Oceanografia Física, Química e Geológica e Aquicultura), um Curso de Especialização (Ecologia Aquática Costeira) e um Curso de Graduação (Oceanologia). Em 2010 teve início o curso de Tecnólogo em Gestão Ambiental, oferecido em Rio Grande e também na cidade de São Lourenço do Sul. Tem como missão promover o ensino, a pesquisa e a extensão em ciências do mar, de forma a produzir, organizar e disseminar o conhecimento na sociedade e contribuir para preservar o meio ambiente.

O Instituto de Oceanografia, em termos de estrutura organizacional, está dividido em núcleos acadêmicos, os quais são: Núcleo de Oceanografia Biológica, Núcleo de Oceanografia Física, Núcleo de Oceanografia Química, Núcleo de Oceanografia Geológica, Núcleo de Gerenciamento Costeiro, Núcleo de Recursos Renováveis, e Núcleo de Aquicultura e Biotecnologia Marinha.

Em termos de infraestrutura o IO conta com os laboratórios nos Núcleos Acadêmicos, situados na Base Oceanográfica do Campus Carreiros, com a Estação Marinha de Aquicultura Prof. Marcos A. Marchiori (EMA), situada no Balneário Cassino, e a Estação Saco do Justino, conta ainda com instituições do complexo de Museus da Universidade, sendo estas o Museu Oceanográfico Prof. Eliézer de Carvalho Rios, Eco-Museu da Ilha da Pólvora, Centro de Recuperação de Animais Marinhos e Museu Náutico. Também tem o suporte da Biblioteca Setorial em Oceanografia, com acervo especializado na área.

Ainda em relação à infraestrutura, para além de veículos tracionados utilizados nas saídas de campo, a frota do Instituto conta com o Navio Oceanográfico “Atlântico Sul”, a Lancha Oceanográfica “Larus” e o Bote “Prof. Moralles”, embarcações imprescindíveis para o desenvolvimento das atividades de investigação dos Núcleos Acadêmicos e a formação dos oceanólogos. Ademais, conta com o apoio de navios da Marinha do Brasil, os quais podem ser partilhados para estudos costeiros e oceânicos.

No final de 1976 a Universidade lançou a revista Atlântica, o primeiro periódico científico editado pela instituição. Ao longo de seus quase 30 anos, a revista “Atlântica” granjeou progressivamente reputação como um periódico científico de qualidade, que

³⁰ Universidade Federal do Rio Grande. Instituto de Oceanografia - <http://www.io.furg.br/>

pode ser encontrado nas bibliotecas das principais instituições de todo o mundo. Além desta publicação, o IO também edita os “Cadernos de Ecologia Aquática”.

Atualmente o IO participa em algumas redes de pesquisa em âmbito nacional, tais como o Programa de Pesquisa Ecológica de Longa Duração (PELD - <http://www.peld.furg.br/>), Instituto Nacional de Ciência e Tecnologia da Criosfera (INCT da Criosfera - <http://inct.cnpq.br/web/inct-criosfera>), Instituto Nacional de Ciência e Tecnologia do Mar (INCT do Mar – Centro de Oceanografia Integrada - <http://inct.cnpq.br/web/inct-mar-coi>), e Sub-rede Zonas Costeiras – Rede Clima e INCT para Mudanças Climáticas (<http://www.mudancasclimaticas.zonascosteiras.furg.br/index.php?lang=pt>).

2.1.3 O Núcleo da Oceanografia Biológica

Ao Núcleo da Oceanografia Biológica do Instituto de Oceanografia está ligado o Programa de Pós-Graduação em Oceanografia Biológica³¹, que oferece os cursos de mestrado e doutorado nas seguintes linhas de pesquisa: Oceanografia Pesqueira; Biotecnologia Marinha e Oceanografia Aplicada; Poluição, Modelagem e Gerenciamento Ambiental; e Sistemática, Biologia e Ecologia de Organismos Marinhos e Estuarinos.

Em termos de estrutura organizacional estão ligados ao Núcleo de Oceanografia Biológica os seguintes laboratórios de ensino e investigação: Laboratório de Ecologia Vegetal Costeira; Laboratório de Crustáceos Decápodes; Laboratório de Ecologia do Ictioplâncton; Laboratório de Ecologia de Invertebrados Bentônicos; Laboratório de Fitoplâncton e Microorganismos Marinhos; Laboratório de Ictiologia; Laboratório de Mamíferos Marinhos e Tartarugas Marinhas; Laboratório de Elasmobrânquios e Aves Marinhas; e Laboratório de Zôoplancton.

Os principais objetivos do Programa de Pós-graduação em Oceanografia Biológica são os de proporcionar conhecimentos aprofundados no campo da

³¹ Instituto de Oceanografia. Programa de Pós-Graduação em Oceanografia Biológica - <http://www.ocbio.furg.br/>

oceanografia biológica e preparar recursos humanos para a investigação científica, o magistério superior e o mercado de trabalho não acadêmico. Especificamente, se almeja: a) ampliar e difundir o conhecimento da oceanografia biológica dos ambientes marinhos e costeiros do Brasil e suas inter-relações com o oceano global; b) fornecer recursos humanos especializados e de alto nível para o crescente mercado nacional de profissionais ligados as ciências do mar auxiliando, também, os tomadores de decisão em temas ligados ao mar; c) formar profissionais de capacidade acadêmico-científicas de alto nível, capacitando e qualificando as ciências do mar no país e, desta maneira, propiciar a inserção do Brasil, com qualidade, nos fóruns internacionais da área; d) promover a integração científica e aplicada entre os profissionais ligados às ciências do mar da FURG, do Brasil e da comunidade internacional e; e) contribuir para o desenvolvimento e integração regional, nacional e internacional buscando parceiros em outras instituições de ensino para a Integração das Ciências do Mar.

As pesquisas, no âmbito do programa de pós-graduação, estão voltadas ao estudo das relações ecológicas entre os componentes bióticos e abióticos dos ambientes costeiro e marinho, à oceanografia pesqueira, estudos de poluição marinha, modelagem ecológica e gerenciamento ambiental, bem como oceanografia aplicada e biotecnologia marinha. Os estudos desenvolvem-se numa ampla variedade de ambientes como, por exemplo, estuários, águas costeiras e mar profundo do sudoeste do Oceano Atlântico, abrangendo toda Zona Econômica Exclusiva brasileira, ilhas oceânicas tropicais e região Antártica.

Devido à determinante geográfica da cidade de Rio Grande, grande parte dos estudos do programa de pós-graduação em oceanografia biológica se desenvolve no estuário da Lagoa dos Patos, que margeia a cidade. O estuário possui importância ecológica, econômica e social para as comunidades que vivem em seus arredores, onde são encontrados muitos pescadores artesanais, algumas indústrias de pescados e um potente polo industrial. Também, serve como corredor de escoamento fluvial da produção interna do estado e do país através do Super Porto de Rio Grande (SCHOWCHOW e ZAMBONI, 2007, p. 15).

Odum (1997) coloca os estuários a par das florestas tropicais e dos recifes de coral, como ecossistemas naturalmente produtivos, produtividade que decorre das características físicas (abióticas) dos mesmos. Portanto, são recursos a serem estudados

em toda a sua riqueza a fim de assegurar a harmoniosa interação com o homem, bem como sua preservação.

2.2 Instituições internacionais que realizam investigação em Oceanografia

A investigação em Oceanografia no IO-FURG se desenvolve prementemente no âmbito dos programas de pós-graduação que este oferece, mais especificamente nos laboratórios ligados organicamente à estrutura dos Núcleos Acadêmicos. Portanto, se pode ter uma ideia dos temas investigados no domínio da Oceanografia Biológica ao se vislumbrar a estrutura organizacional dos laboratórios ligados ao seu núcleo. Todavia, há que se considerar que as estruturas organizacionais das instituições são concebidas de maneiras diferentes, consoante a uma série de variáveis que definem cada uma delas. Veja a seguir exemplos de instituições fora do Brasil que realizam investigação na área da Oceanografia Biológica, que apresentam estruturas organizacionais diferenciadas.

A “North Carolina State University³²”, Universidade americana que dispõe de programa de intercâmbio com universidades brasileiras, possui dentro do seu “College of Sciences” o departamento “Marine, Earth And Atmospheric Sciences³³”, o qual oferece programas de investigação nas áreas das Ciências Atmosféricas, Ciências Marinhas, e Ciências da Terra. No âmbito das Ciências Marinhas o departamento oferece programas de especialização e investigação nas quatro fundamentais disciplinas da Oceanografia (Física, Química, Biológica e Geológica), assim como na emergente área da meteorologia marinha, com vistas a explorar os processos e dinâmicas dos oceanos. Esta estrutura é praticamente idêntica ao do Instituto de Oceanografia da FURG no que se refere à Pós-Graduação. No entanto, quanto aos grupos de investigação das Ciências Marinhas estes são separados por temáticas, onde os limites que separam os quatro tipos de oceanografia não são claramente definidos, o que difere da estrutura do IO-FURG, onde os laboratórios estão separados pelos Núcleos

³² North Carolina State University - <https://www.ncsu.edu/>

³³ North Carolina State University – Department of Marine, Earth And Atmospheric Sciences - <http://www.meas.ncsu.edu/>

Acadêmicos. Alguns dos grupos de investigação em Ciências Marinhas do departamento “Marine, Earth And Atmospheric Sciences” são:

- Benthic Studies
- Coastal Ocean Initiatives
- Estuarine & Coastal Processes
- Global Biogeochemical Cycles of Carbon and Silica
- Marine Ecology and Conservation Program
- Marine Geophysics and Underwater Acoustics

Ainda em território americano, convém mencionar relevantes instituições de ensino e investigação que são referência na área das Ciências do Mar. Uma destas instituições que é reconhecida como um centro de investigação de ponta é o “Scripps Institution of Oceanography³⁴”, da Universidade da Califórnia em San Diego³⁵. O programa doutoral do Scripps está organizado em três programas acadêmicos separados, cuja estrutura se assemelha ao do Instituto de Oceanografia da FURG, sendo eles “Climate-Ocean-Atmosphere Program”; “Geosciences of the Earth, Oceans, and Planets”; e o “Ocean Biosciences Program”. O programa “Ocean Biosciences” apresenta duas linhas de formação, uma em “Oceanografia Biológica” e outra “Biologia Marinha”. A linha de investigação em Oceanografia Biológica incide sobre as interações de populações de organismos marinhos umas com as outras e com o seu ambiente físico e químico, que é também a incidência de investigação da Oceanografia Biológica na FURG. O programa Ocean Biosciences apresenta vinte e cinco (25) grupos de investigação, enquanto o Programa de Pós-Graduação em Oceanografia Biológica na FURG apresenta onze (11), o que reafirma a grandeza da investigação do primeiro.

Além do “Scripps Institution of Oceanography”, outra instituição americana referenciada pelas universidades brasileiras como centro de significativa investigação na área das Ciências do Mar é a “Woods Hole Oceanographic Institution³⁶”, localizada em Boston, estado de Massachusetts. A Woods Hole Oceanographic Institution é uma instituição privada e sem fins lucrativos em investigação oceanográfica e no estudo e

³⁴ Scripps Institution of Oceanography - <https://scripps.ucsd.edu/>

³⁵ University of California, San Diego - <https://ucsd.edu/>

³⁶ Woods Hole Oceanographic Institution - <http://www.whoi.edu/>

exploração do oceano. As áreas de investigação deste Instituto de Oceanografia (abaixo listadas) estão organizadas de forma que não compartilham muitas semelhanças com a estrutura apresentada pelo Instituto de Oceanografia da FURG. No que se refere à investigação em Oceanografia Biológica podemos encontrar, por exemplo, dentro da área de investigação “Coastal Science”, os laboratórios “Benthic Ecology and Nearshore Oceanographic Laboratory” e “Fish Ecology Laboratory” que poderiam ser comparados aos laboratórios “Ecologia de Invertebrados Bentônicos” e “Ictiologia”, que promovem investigação nos mesmos tópicos.

Algumas das áreas de investigação do Woods Hole Oceanographic Institution:

- Climate & Ocean
- Coastal Science
- Ocean Chemistry
- Ocean Circulation
- Ocean Life
- Pollar Systems
- Pollution

Ainda na América do Norte, porém no Canadá, o “Department of Earth, Ocean & Atmospheric Sciences” – EOAs³⁷ da “University of British Columbia³⁸”, Campus de Vancouver, oferece um programa de pós-graduação em Oceanografia, onde os candidatos devem optar por uma das seguintes áreas de especialização: Biological Oceanography; Marine Chemistry and Geochemistry; Physical Oceanography and Atmospheric Sciences. Esta estrutura organizacional assemelha-se à da pós-graduação oferecida pelo IO-FURG, no entanto, no que tange à investigação o EOAs é um departamento integrado e não há divisões internas formais. No entanto, a investigação pode ser agrupada de diversas maneiras: por temas, por infraestrutura, e por grupo ou projeto de investigação.

Em continente europeu, outro exemplo que se pode tomar é do Institute of Oceanology “Fridtjof Nansen³⁹”, situado na Bulgária. As atividades de investigação do

³⁷ Department of Earth, Ocean & Atmospheric Sciences – EOAs - <http://www.eos.ubc.ca/research/>

³⁸ University of British Columbia - <https://www.ubc.ca/>

³⁹ Institute of Oceanology “Fridtjof Nansen” - http://www.io-bas.bg/index_en.html

Instituto se concentram nos domínios marinhos da física, química, geologia e arqueologia, biologia e ecologia, na dinâmica costeira e nas tecnologias oceânicas. Ele realiza monitoramentos da parte búlgara do Mar Negro, desenvolve atividades especializadas e de consultoria e prepara cientistas marinhos. Quanto à estrutura organizacional está dividido em departamentos científicos, os quais são: “Marine Chemistry”; “Marine Biology and Ecology”; “Coastal Zone Dynamics”; “Marine Geology and Archeology”; e “Ocean Technologies”. Embora esta divisão estrutural não apresente equivalência com a do IO-FURG, pode-se perceber semelhanças entre as atividades desenvolvidas pelo departamento “Marine Biology and Ecology”, as quais estão listadas abaixo, e as desenvolvidas no Núcleo da Oceanografia Biológica do Instituto de Oceanografia da FURG:

- Estudo da biodiversidade taxonômica e funcional do Mar Negro e as interações de cadeia alimentar;
- Investigação da resposta da biota à força externa - pressão antrópica e impacto climático mundial;
- Desenvolvimento de diretrizes para a monitorização e análise laboratorial da flora e fauna do Mar Negro e padrões de classificação de qualidade ecológica das elementos biológicos de qualidade sensu Diretiva Quadro da Água 2000 / CE / 60 e sua aplicação na Bulgária;
- Realização de estudos de qualidade ecológica e avaliação de estoque de peixes comercialmente importantes ao longo da costa búlgara;
- Formulação das bases científicas para o desenvolvimento sustentável do ecossistema do Mar Negro e dos seus recursos biológicos.

O exemplo seguinte é de um importante centro de investigação em Oceanografia na Europa chamado “Ocean and Earth Science, National Oceanography Centre Southampton⁴⁰”, também conhecido como NOCS. Este faz parte da “University of

⁴⁰ Ocean and Earth Science, National Oceanography Centre Southampton - <http://www.southampton.ac.uk/oes/research/index.page>

Southampton⁴¹”, no Reino Unido, e é uma notável instituição para a investigação e educação em ciências do mar e da terra. A investigação no NOCS se estende da costa para o oceano profundo e procura abordar as questões científicas mais prementes do nosso tempo. O centro está dividido em seis grupos temáticos de investigação, que em conjunto com outros fatores, fazem do NOCS um local de formação em investigação interdisciplinar. São eles:

- Geochemistry
- Geology and Geophysics
- Marine Biology and Ecology
- Palaeoceanography and Palaeoclimate
- Marine Biogeochemistry
- Physical Oceanography

Na tentativa de fazer uma comparação com o Instituto de Oceanografia da FURG, poder-se-ia dizer que o grupo de investigação em Ecologia e Biologia Marinha seria equivalente ao Núcleo da Oceanografia Biológica.

Também no Reino Unido a “University of Liverpool⁴²”, através do “Department of Earth, Ocean and Ecological Sciences⁴³” da “School of Environmental Sciences” promove investigação e educação na área da Oceanografia. A investigação neste departamento está dividida em “Earth Sciences”, “Oceans and Climate” e “Ecology and Marine Biology”. Na divisão “Ecology and Marine Biology”, que seria a mais semelhante a Oceanografia Biológica, as principais áreas de investigação são:

- Dinâmica temporal e espacial nos ecossistemas
- Compreender a contribuição funcional da biodiversidade nos ecossistemas

⁴¹ University of Southampton - <http://www.southampton.ac.uk/>

⁴² University of Liverpool - <http://www.liv.ac.uk/>

⁴³ University of Liverpool - Department of Earth, Ocean and Ecological Sciences - <http://www.liv.ac.uk/environmental-sciences/>

- Gestão baseada nos ecossistemas
- Ecologia, fisiologia e comportamento de aves marinhas
- Modelagem de mudanças de composição em recifes de coral
- Ecologia e manejo do fogo em charnecas e savanas
- Impacto e controle de espécies invasoras
- Ecologia da restauração
- Medição de impactos ecológicos de longo prazo

Na Espanha, o “Instituto de Ciencias Marinas de Andalucía” – ICMAN⁴⁴ possui dois departamentos aos quais estão ligados os grupos de investigação. Ao “Departamento de Biología Marina y Acuicultura” estão ligados o “Grupo de Investigación Acuicultura” o qual investiga na linha de fisiologia e cultivo de larvas de peixes; o “Grupo de Investigación Biología Celular y Molecular en Organismos Marinos” que investiga a linha histofisiologia e histopatologia; e o “Grupo de Investigación Ecología de Esteros y Zonas Intermareales”, cuja linha de investigação se concentra nas Comunidades aquáticas de estuários e outros habitats costeiros. Já ao “Departamento de Ecología y Gestión Costera” estão ligados os grupos “Oceanografía de Ecosistemas”; “Ecotoxicología Marina”; e “Ecología de la macrofauna acuática”. Embora com uma estrutura organizacional bastante diferenciada, se pode notar uma paridade acentuada entre as áreas de investigação do ICMAN e as do IO-FURG, sobretudo no Núcleo de Oceanografía Biológica.

Os últimos exemplos de investigação na área de investigação marinha são advindos da Oceania. O primeiro se refere à “University of Otago⁴⁵”, da Nova Zelândia, cujo “Department of Marine Science⁴⁶”, criado em 1992, apresenta uma longa história de ensino e investigação marinha em Otago. O departamento oferece cursos de graduação e pós-graduação com foco em ciência marinha pura e aplicada, oceanografia e aquicultura.

⁴⁴ Instituto de Ciencias Marinas de Andalucía – ICMAN - <http://www.icman.csic.es/>

⁴⁵ University of Otago - <http://www.otago.ac.nz/>

⁴⁶ University of Otago - Department of Marine Science - <http://www.otago.ac.nz/marinescience/index.html>

A investigação em Ciências Marinhas está dividida por vários temas, sendo estes:

- Antarctic Science
- Biology and Conservation of Marine Mammals
- Changing Southern Seas
- Managing and Conserving Marine Resources
- Marine Discovery
- Marine Ecosystems
- Ocean Physics
- Shelf and Deep Sea Systems

Para finalizar a prospecção das estruturas organizacionais de instituições que realizam investigação no domínio da oceanografia, cita-se o exemplo do “UWA Oceans Institute⁴⁷”, da universidade australiana “University of Western Australia⁴⁸”. O Instituto tem o objetivo de capitalizar os pontos fortes de investigação existentes na UWA - em áreas como a oceanografia, ecologia, engenharia, gestão de recursos e governança - e utilizá-los para oferecer soluções de oceano para grandes desafios da humanidade. A investigação marinha no UWA Oceans Institute cobre as seguintes áreas:

- Blue Water Oceanography
- Coastal Processes
- Marine Ecology
- Ocean Engineering
- Marine Biological Resources, Management, Governance and Conservation
- Impacts of Climate Change in the Marine Environment
- Marine Observation, Monitoring and Risk Management

⁴⁷ University of Western Australia - UWA Oceans Institute - <http://www.oceans.uwa.edu.au/>

⁴⁸ University of Western Australia - <http://www.uwa.edu.au/>

Diante da exposição acima, pode-se concluir que instituições de diferentes países, apresentam normalmente estruturas organizacionais diferentes, porém, muitas vezes podem executar investigação nos mesmos tópicos dentro de uma área de conhecimento.

3 Metodologia

Este estudo enquadra-se no paradigma das “Design Science Research”, que, segundo Hevner *et al.*, 2004, tendo suas raízes na engenharia e nas ciências do artificial, compreendem a criação de artefatos para a resolução de problemas. Neste caso, o artefato criado é uma ferramenta de curadoria digital para apoiar investigadores na gestão de conjuntos de dados de investigação.

A seguir é descrita a metodologia adotada para a construção de tal ferramenta, cujos passos são abordados em duas etapas. A primeira compreende a elaboração de um perfil de aplicação baseado em ontologia para descrição de conjuntos de dados de investigação para o domínio da Oceanografia Biológica, o que pretende solucionar a questão da descrição de conjuntos de dados. A segunda se refere à promoção e avaliação do depósito de conjuntos de dados na plataforma Dendro, o que pretende solucionar a questão do armazenamento seguro dos dados, com vistas à preservação e partilha dos mesmos.

3.1 Procedimentos

3.1.1 Construção de um perfil de aplicação baseado em ontologia

Um perfil de aplicação compreende a criação de metadados para descrição da informação, neste caso, conjuntos de dados de investigação. Consiste na combinação de elementos de metadados extraídos de um ou mais esquemas existentes com outros elementos criados sob medida para o domínio, e otimizados para uma aplicação particular (HEERY e PATEL, 2000). Um perfil de aplicação é, no fundo, um modelo de metadados criado sob medida para uma necessidade específica de descrição.

O perfil de aplicação criado para este estudo é adequado a uma experiência específica de investigação no domínio da Oceanografia Biológica, o qual foi chamado “Biological Oceanography”. Para empreender na sua criação foi utilizada a abordagem metodológica que vem sendo executada no Information Systems Laboratory, da Faculdade de Engenharia da Universidade do Porto. Também foram consultados os

“Guidelines for Dublin Core Application Profiles⁴⁹”, que embasaram e complementaram a abordagem metodológica em questão, cujos passos estão abaixo descritos.

Para elaborar um perfil de aplicação com metadados que descrevessem apropriadamente os dados de investigação provenientes do domínio da Oceanografia Biológica foi necessário adquirir um algum conhecimento sobre a área. Obter informação sobre questões práticas, como procedimentos e instrumentos utilizados nas atividades de investigação, métodos de recolha, tratamento e armazenagem de dados, além de familiarização com a linguagem técnica do domínio, foram importantes pré-requisitos para se conseguir tal percepção.

Para isso foram entrevistados investigadores da área, que responderam a perguntas genéricas sobre o modo como realizam a gestão de seus dados de investigação e o que pensam sobre preservação e partilha de dados. Também responderam a perguntas mais específicas relativas ao workflow dos seus conjuntos de dados e procedimentos específicos de investigação. A entrevista é considerada um importante instrumento para iniciar a abordagem de levantamento de requisitos para o perfil de aplicação, uma vez que permite ao criador do perfil obter um entendimento do domínio a partir da visão do investigador e, em contrapartida, permite ao investigador obter uma compreensão sobre gestão de dados mais alargada a partir do ponto de vista do criador do perfil de aplicação.

Para esta etapa do estudo também se precedeu à análise de conteúdo de publicações científicas dos investigadores (dois ou três artigos publicados, conforme a disponibilidade do investigador), com o objetivo de captar conceitos-chave para o entendimento da investigação na área. Para compreensão da linguagem técnica por vezes se recorreu a dicionários, wikis ou vocabulários controlados, e para entendimento de procedimentos específicos, eventualmente foi necessário recorrer ao próprio investigador.

As duas tarefas acima descritas forneceram embasamento para familiarização com termos específicos do domínio e entendimento da área e das experiências de investigação, e culminaram na elaboração de um mapa com os principais conceitos levantados.

⁴⁹ Dublin Core Metadata Initiative. **Guidelines for Dublin Core Application Profiles** - (<http://dublincore.org/documents/profile-guidelines/#DCMI-MT>),

A elaboração do mapa conceitual é, portanto, o terceiro passo para estruturar e formalizar o conhecimento do fluxo de trabalho e produção de dados de investigação na área. Para este domínio foram elaborados inicialmente três mapas de conceitos, um para cada uma das subáreas dos investigadores entrevistados. Estes foram enviados por correio eletrónico aos investigadores para serem analisados e discutidos e, assim serem feitas as modificações necessárias. O passo seguinte foi reunir os três mapas em apenas um que contemplasse as três subáreas. Este foi novamente enviado aos investigadores para ser validado.

Deste mapa foram extraídos os conceitos mais importantes a serem adotados como potenciais descritores (metadados) no perfil de aplicação e na formalização da ontologia. No entanto, sentiu-se a necessidade de acrescentar descritores que não constavam no mapa de conceitos, mas que seriam importantes para a descrição de um conjunto de dados. Decidiu-se seguir a recomendação Dublin Core Application Profiles e estabelecer objetivos para o perfil de aplicação. A partir daí os descritores foram selecionados a partir dos objetivos.

Ainda, anteriormente a elaboração do perfil de aplicação foi necessário proceder ao estudo de esquemas de metadados e normas existentes para descrição de dados concernentes ao domínio da Oceanografia Biológica e áreas afins, tais como Ecologia e Biologia, para que fossem escolhidos os elementos apropriados à sua formação. Também foi estudado um padrão de metadados genérico, para o que foi escolhido o Dublin Core, por ser amplamente conhecido e usado para descrição de recursos na web.

Em seguida foi elaborado o perfil de aplicação para descrição de dados do domínio da Oceanografia Biológica, através da escolha de elementos de metadados dos esquemas estudados, combinados com elementos novos, criados a partir dos conceitos representados no mapa. Para cada um dos descritores selecionados ou criados foi atribuída uma definição e um exemplo de aplicação.

Depois de elaborado o perfil de aplicação, foram pesquisadas ontologias cujos vocabulários pudessem representar os descritores. O objetivo era importar e aproveitar vocabulários de outras ontologias para a formalização do perfil de aplicação. A ontologia construída constituiu-se apenas de algumas classes e propriedades, e por apresentar este formato simplificado é chamada de “ontologia ligeira” (“lightweight ontology”) (Castro, Silva e Ribeiro, 2014). Nesta ontologia, os descritores do perfil de aplicação são representados por propriedades.

Após a sua formalização a ontologia foi incorporada na plataforma colaborativa de gestão de dados de investigação Dendro, onde cumpre sua função como modelo de metadados para descrição de dados de investigação.

3.1.2 Promoção e avaliação do depósito de conjuntos de dados de investigação na plataforma Dendro

Convidar investigadores a armazenar e descrever seus dados de investigação é uma das principais atividades neste estudo, a qual está ligada à plataforma Dendro, uma plataforma colaborativa e multidisciplinar para gestão de dados concebida na FEUP. Depois de pronto o perfil de aplicação é formalizado em ontologia e esta incorporada na plataforma Dendro para cumprir sua função na descrição de dados de investigação. A plataforma já conta com vocabulários genéricos de descrição e com vocabulários específicos de outras áreas do conhecimento.

Para sensibilizar os investigadores quanto à importância das boas práticas de gestão de dados, sobretudo no que se refere às questões de depósito e descrição de dados visando sua preservação e partilha, está sendo realizada uma campanha de interações de investigadores da Universidade do Porto com a plataforma Dendro. Nesta campanha investigadores das áreas que possuem ontologias no Dendro foram convidados a depositar e descrever seus conjuntos de dados e outros produtos gerados a partir destes, como artigos ou relatórios de investigação. Estas experiências visam avaliar o desempenho das funcionalidades da plataforma, bem como a aplicabilidade dos seus vocabulários para descrição de dados.

No contexto da Campanha Dendro, que também contou com a participação de investigadores da área de Ciências do Mar, foi elaborado um breve inquérito (que pode ser visto no anexo I) para avaliar especificamente o perfil de aplicação “Biological Oceanography”. Esse foi composto por questões genéricas sobre validade e utilidade dos descritores do perfil, e os investigadores desta área que participaram da experiência o responderam após sua interação com a plataforma.

3.2 Participantes

Para o levantamento dos requisitos necessários ao desenho do perfil de aplicação foram convidados a participar do estudo três (3) investigadores da área da Oceanografia Biológica, sendo que cada um pertence a uma subárea específica dentro do domínio. Os investigadores em questão desenvolvem sua investigação junto aos laboratórios de Crustáceos Decápodes, de Ictiologia, e de Ecologia de Invertebrados Bentônicos, na Universidade Federal do Rio Grande, no sul do Brasil.

Para a fase posterior, de utilização e avaliação do perfil de aplicação já formalizado em ontologia e incorporado na plataforma Dendro, foram convidados dois (2) investigadores da mesma área, pertencentes à Universidade do Porto. Para avaliar o perfil de aplicação foi dada preferência a investigadores que não tivessem colaborado na sua construção, de modo que a avaliação não ficasse tendenciosa. Ainda, uma avaliação feita por investigadores do mesmo domínio, porém de uma instituição e de um país diferentes, podem demonstrar a abrangência dos descritores criados.

3.3 Recolha de dados

A recolha de dados do estudo também foi realizada em duas etapas, uma necessária para a construção do perfil de aplicação, e a outra com o objetivo de avaliar a ferramenta já pronta e incorporada à plataforma Dendro.

Para a fase anterior à elaboração do perfil de aplicação a recolha de dados foi feita através de:

- Entrevista semiestruturada com investigadores: cada um dos investigadores foi entrevistado individualmente, com o propósito de fornecer informação a respeito dos procedimentos de investigação e gestão de dados na sua área. As entrevistas foram conduzidas com base em um guião e gravadas, para o áudio ser posteriormente analisado.
- Análise de conteúdo de publicações: foram analisados os conteúdos de artigos científicos publicados pelos investigadores participantes do estudo. O objetivo era captar conceitos que fossem importantes para o entendimento do domínio e embasar a posterior elaboração do mapa de conceitos.

- Estudo de esquemas de metadados: foram examinados um esquema de metadados genérico (Dublin Core) e quatro esquemas de metadados específicos do domínio e áreas afins (Darwin Core, Ecological Metadata Language, OBIS Data Schema e Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile) para a escolha de elementos de descritores a compor o perfil de aplicação.

Para a fase posterior, após a formalização da ontologia e sua incorporação na plataforma de gestão de dados, a recolha de dados foi realizada através de um inquérito, no qual os investigadores registam suas impressões acerca do uso da ferramenta para descrição de conjuntos de dados na plataforma Dendro.

Foram, portanto, utilizados dois instrumentos de recolha de dados neste estudo. Na etapa pré-perfil de aplicação, foi utilizado o “Guião de Entrevista Curadoria de Dados”, adaptado de instrumento disponibilizado pelo Data Curation Profile Toolkit⁵⁰. Na etapa pós-perfil de aplicação, foi utilizado o “Inquérito para avaliação do perfil de aplicação Biological Oceanography”, respondido pelos investigadores após sua utilização na plataforma Dendro. Os dois guiões encontram-se nos anexos F e I.

3.4 Materiais

3.4.1 Dendro

O Dendro⁵¹ é uma plataforma baseada na web para depósito e troca de dados de investigação, projetada para ajudar os investigadores a armazenar os seus dados (à maneira da aplicação Dropbox), e para descrevê-los de forma colaborativa. Seu principal objetivo é ajudar os investigadores a descreverem seus dados na medida em eles são produzidos, tornando mais fácil a sua partilha com um repositório externo (institucional, por exemplo). Desta forma os dados podem ser citados por outros, bem como pelos seus próprios autores em publicações.

⁵⁰ Data Curation Profiles Project. Data Curation Profile Toolkit - <http://datacurationprofiles.org/>

⁵¹ Dendro - <http://dendro.fe.up.pt/>

A plataforma Dendro foi desenhada com uma interface amigável para utilizadores sem conhecimento em gestão de dados. Utiliza conceitos de ontologias específicas de domínio, que se configuram em metadados criados sob medida para descrever dados destes domínios. Portanto, permite aos utilizadores construir uma base de conhecimento usando ontologias em segundo plano, que lhes dá prioridade em focar na escolha dos descritores semanticamente compatíveis com seus domínios sem se preocupar com questões de design e implementação que surgem do uso de ontologias. (Silva, *et al.* 2014).

3.4.2 CMapTool

Para a elaboração dos mapas de conceitos extraídos da análise de conteúdo dos artigos científicos foi utilizado o software CMap Tool⁵². O software, disponível para descarga gratuita, é resultado de pesquisa realizada na “Florida Institution of Human & Máquina Cognition⁵³” e capacita os usuários a construir, navegar, compartilhar e criticar modelos de conhecimento representados através de mapas conceituais.

3.4.3 Protégé

Para a construção da ontologia foi utilizado o software Protégé⁵⁴, uma ferramenta disponibilizada gratuitamente pela Universidade de Stanford⁵⁵. Protégé é uma plataforma gratuita, de código aberto que fornece um conjunto de ferramentas para a construção de modelos de domínio e aplicações baseadas em conhecimento com ontologias.

⁵² Mais informações em: CMap Tool - <http://cmap.ihmc.us/>

⁵³ Florida Institution of Human & Máquina Cognition - <http://www.ihmc.us/>

⁵⁴ Mais informações em: Protégé - (<http://protege.stanford.edu/>)

⁵⁵ Stanford University - <https://www.stanford.edu/>

4 Desenho e avaliação de um perfil de aplicação para atividades de investigação na Oceanografia Biológica

O perfil de aplicação construído no âmbito deste estudo se enquadra na metodologia que está a ser aplicada no “Information Systems Research Group⁵⁶”, do Departamento de Engenharia Informática da FEUP, conhecido como InfoLab. Investigadores deste laboratório têm trabalhado na elaboração de perfis de aplicação baseados em ontologias com o objetivo de estes serem incorporados à plataforma Dendro, a qual é destinada a suportar a gestão colaborativa de dados de investigação multi-domínio. Estes perfis são desenhados para situações e necessidades específicas com a colaboração de especialistas dos domínios. Retratam, portanto, atividades específicas de investigação, não tendo a pretensão de abranger o domínio em sua totalidade, antes pelo contrário, têm o objetivo de captar e desenhar os requisitos de gestão de dados em grupos de investigação específicos.

Desenhar um perfil de aplicação que seja compreensivo de um domínio é uma tarefa bastante complexa, que demanda uma carga pesada de recursos (pessoas, orçamento, tempo). A tarefa de criar um esquema de metadados sob medida exige que se tenha, antes de tudo, conhecimento sobre o domínio para o qual se quer criar o perfil e suas particularidades. Ainda que seja uma área de conhecimento coesa, a criação de metadados está diretamente ligada à prática investigativa da área, pelo que muitas de suas características podem não ser captadas à exaustão devido a uma série de variáveis contextuais, sejam institucionais, sociais, culturais ou econômicas, que podem afetar as atividades de investigação dentro de uma mesma área de conhecimento.

Por este motivo, para este trabalho foi desenhado um perfil de aplicação para o domínio da Oceanografia Biológica que retrata especificamente as atividades de investigação de três grupos distintos oriundos da Universidade Federal do Rio Grande, no Sul do Brasil, sendo estes: Laboratório de Crustáceos Decápodes, Laboratório de Ictiologia, e Laboratório de Ecologia de Invertebrados Bentônicos. De cada um dos grupos participou do estudo um investigador, do que se pode depreender que o desenho do perfil de aplicação levou em consideração as atividades específicas destes investigadores. Embora tenha se tentado fazer uma generalização para o restante do

⁵⁶ Information Systems Research Group - <http://infofab.fe.up.pt/>

grupo ao qual o investigador pertence, admite-se que alguns aspectos podem ter ficado de fora.

No subitem a seguir é feita uma caracterização das atividades de investigação nas subáreas da Oceanografia Biológica estudadas no âmbito deste trabalho, com foco central na produção de dados de investigação gerados a partir destas atividades. São mencionados também alguns aspectos em relação à organização e armazenamento dos dados, além de questões de preservação e partilha. As informações que constam neste subitem são derivadas das entrevistas realizadas com os investigadores da Universidade Federal do Rio Grande.

4.1 A produção de dados no domínio da Oceanografia Biológica

A produção de dados de investigação se dá no contexto de três subáreas do domínio da Oceanografia Biológica, mais concretamente nas áreas de ecologia de peixes, de crustáceos decápodes e de organismos invertebrados bentônicos. As três subáreas objetivam investigar as interações ecológicas entre organismos marinhos e estuarinos e seus parâmetros ambientais.

Os dados gerados a partir de suas atividades de investigação são sobretudo observacionais, ou seja, são captados em tempo real. As atividades dividem-se em duas etapas distintas, que foram designadas como atividades de campo e atividades de laboratório. As atividades de campo dizem respeito aos eventos de coleta de material biológico e ao registo dos parâmetros ambientais. As atividades realizadas em laboratório dizem respeito aos tipos de interações necessárias realizadas com o material biológico coletado em campo, o que pode envolver experimentos, dando origem também a dados experimentais.

O material biológico capturado nos eventos de coleta consiste nos organismos estudados pelas áreas (peixes, crustáceos e bentos) e em sedimento (substrato do fundo de corpos d'água, neste caso o estuário da Lagoa dos Patos). Os parâmetros ambientais podem ser registados durante os eventos de coleta ou ser independente destes, de acordo com uma periodicidade previamente estipulada. Eles são chamados pelos investigadores de dados abióticos e os principais são temperatura da água, salinidade, transparência e profundidade da coluna d'água. Para estas duas atividades de campo são aplicados

métodos e utilizados instrumentos e ferramentas específicos, que levam em consideração o tipo de material que se quer coletar ou o tipo de variável ambiental que se pretende registrar.

Tanto para os eventos de coleta de material biológico quanto para os registros de parâmetros ambientais são definidos e registados alguns dados espaciais e alguns dados temporais, os quais, como os próprios nomes sugerem, são elementos importantes para contextualizar estas atividades territorial e temporalmente. Os dados espaciais referem-se a: nome do local onde será realizado o evento, nomes dos pontos específicos de coleta daquele local, suas coordenadas geográficas, sendo também definida a quantidade de pontos de coleta e a extensão do perfil. Os dados temporais referem-se a: data do evento, período de abrangência dos eventos, a periodicidade em que se realizam os eventos e a estação do ano em que se realizam.

Após os eventos de coleta, o material biológico capturado é levado para o laboratório onde é processado. Primeiramente é feita uma separação de todo o material coletado, a qual é chamada de triagem. Na triagem dos sedimentos estes são separados e calculados os elementos que os compõem, tais como argila, areia, silte e matéria orgânica. Na triagem dos organismos, os indivíduos são separados por espécies e então são feitas uma série de contagens (as quantidades de espécies, quantidade de indivíduos por espécie e quantidade total de indivíduos capturados) e tomadas uma série de medidas (biometrias dos indivíduos, que são diferentes para cada espécie, mas que consistem basicamente em medidas de largura, comprimento e peso). Ainda podem ser verificadas outras variáveis, como o sexo, estágio de vida e o estágio de muda (para crustáceos) dos indivíduos. Todos os dados derivados destas interações com os organismos são designados pelos investigadores como dados bióticos. Da mesma forma que as atividades de campo, as atividades de laboratório também demandam métodos e instrumental específicos.

Em relação ao registo e organização dos dados normalmente estes são primeiramente anotados em planilha em papel, tanto os dados capturados em campo como os capturados em laboratório. Existe, então, uma fase posterior onde as planilhas em papel são transpostas para ficheiros Excel. Esta é considerada uma fase delicada onde se deve ter muita atenção para não se cometerem erros de digitação dos valores. As planilhas eletrônicas são muito simples, contêm nos cabeçalhos das colunas os nomes das medidas que são tomadas, e referências em relação a datas e locais onde são

realizadas as coletas. Estas informações aparecem de maneira abreviada. Em alguns casos é feita uma legenda para as siglas, em outros não, por se entender que os investigadores que trabalham no mesmo laboratório já possuem conhecimento para interpretar as planilhas (quando entram estudantes ou investigadores novos as planilhas são explicadas por um membro antigo).

Os ficheiros Excel são salvos em pastas com nomes que possam identificar seu conteúdo e são armazenados, normalmente nos computadores dos laboratórios. Às vezes são feitas cópias de segurança ou os dados são armazenados em drives na nuvem, mas isso depende da iniciativa individual de cada investigador.

Do ponto de vista da investigação os investigadores apontam quatro fases para o ciclo de vida dos dados: 1) coleta dos dados brutos (bióticos e abióticos); 2) transposição dos dados para planilhas Excel e armazenamento; 3) análise estatística dos dados (uso de softwares específicos); 4) elaboração de produtos finais (teses, artigos publicados, etc.).

O cuidado que os investigadores têm em relação ao armazenamento e preservação é em relação aos dados brutos, este foi o tipo de dados que todos referiram em entrevista ser o mais importante a preservar. Os investigadores têm noção de que estes dados preservados podem dar origem a novos estudos. Dados que já sofreram algum tipo de análise são dados a conhecer apenas nas dissertações, teses, relatórios ou artigos publicados.

Questões de gestão e partilha de dados não são institucionalizadas, portanto são praticadas com pouco critério ou padronização. Também não há exigência por parte das agências que financiam os projetos de investigação que os investigadores apresentem planos de gestão de dados de investigação ou pratiquem formalmente a sua partilha através de depósito em repositórios.

A partilha é uma questão subjetiva ao discernimento do investigador ou do líder do seu grupo de investigação, logo, pode ou não ser praticada. Normalmente apenas é realizada quando os dados foram coletados em conjunto com investigadores de outras instituições. Ultimamente reconhecem que tem havido uma maior cooperação entre laboratórios da mesma instituição, quando possível, na tentativa de reduzir custos.

A partilha ocorre através da permuta dos ficheiros digitais que podem conter dados abióticos e bióticos. Em muitos casos também é praticada a partilha dos animais

capturados, em vez de compartilhar os dados brutos da triagem, pois estes podem não ser suficientes ou condizentes com os objetivos de investigação de um terceiro. Desta forma, este pode extrair os dados que considerar necessários e gerar outro banco de dados. Os dados bióticos tratados são eventualmente disponibilizados para análises conjuntas com autores da mesma área de outras instituições. Estas partilhas por via de regra estão condicionadas à parceria nas publicações de monografias, dissertações, teses e artigos científicos.

4.2 Passo-a-passo do desenho do perfil de aplicação

Nos tópicos seguintes estão documentados os passos realizados para a concretização de um conjunto de metadados para descrição de dados de investigação no domínio da Oceanografia Biológica.

4.2.1 Levantamento de requisitos do domínio junto aos investigadores

O primeiro passo a ser dado em direção à elaboração de um perfil de aplicação consistente com as necessidades de descrição de um domínio deve ser o envolvimento dos investigadores da área. Para esta experiência contou-se com a colaboração de três investigadores do Instituto de Oceanografia da Universidade Federal do Rio Grande, onde foi realizada uma apresentação sobre o trabalho a ser realizado para elaboração do perfil de aplicação. Além disso, a apresentação abordou o tema da gestão de dados de investigação com o objetivo de sensibilizar os investigadores em relação às boas práticas de gestão.

Além dos investigadores que colaboraram com o estudo outros demonstraram interesse e participaram da apresentação. A temática da gestão de dados gerou entre os investigadores presentes alguma discussão em relação à partilha de dados de investigação no contexto brasileiro, falou-se sobre a cultura da não partilha de dados e de como esta visão precisa mudar, uma vez que estes são produtos de projetos essencialmente financiados com recursos públicos. Percebeu-se uma carência dos grupos de investigação no Brasil por políticas públicas que venham a regular estas

questões. Hoje não há nenhuma exigência por parte das entidades financiadoras de projetos pela gestão e transparência dos dados gerados pelas atividades de investigação.

A fase seguinte compreendeu o levantamento de requisitos necessários para elaboração do perfil de aplicação, na qual o contato direto com os investigadores/colaboradores do estudo foi importante. Para este levantamento de requisitos, que se constituem de elementos para o entendimento do domínio, foi programada uma entrevista com cada um dos colaboradores. Após a entrevista lhes foi solicitado o envio de pelo menos dois artigos seus já publicados e, se possível de um conjunto de dados. Caso o envio do conjunto de dados não fosse possível foi pedido que este fosse apenas mostrado, para se obter uma ideia mais concreta de como estes se apresentam.

Para a entrevista foi enviado previamente aos investigadores o “Guião de Entrevista Curadoria de Dados” (Anexo F), para que tivessem a oportunidade de se preparar para a mesma. Este guião é usualmente utilizado pelo InfoLab/FEUP e foi traduzido e adaptado por um de seus membros de um instrumento disponibilizado pelo Data Curation Profile Toolkit. O instrumento, que pode ser visto em anexo, aborda questões sobre o fluxo dos dados produzidos durante o ciclo de investigação e questões sobre atividades de gestão, relativamente ao armazenamento e partilha de dados.

A entrevista é um instrumento de recolha de dados adequado para uma primeira abordagem do estudo. Para além de ser orientada por um guião, ela oferece oportunidade de entrevistador e entrevistado conversarem abertamente e sanarem dúvidas de ambos os lados. Permite ao entrevistador obter um entendimento mais acurado do domínio a partir da visão do entrevistado e permite ao entrevistado obter uma compreensão mais alargada sobre gestão de dados a partir do ponto de vista do entrevistador.

Após a entrevista já foi possível obter um entendimento dos conjuntos de dados gerados a partir das atividades de investigação da área. Convém ressaltar que a entrevista é um procedimento recomendado pelo Data Curation Profiles Project⁵⁷ adotado pelo InfoLab. Outras atividades complementares podem ser realizadas para levantar requisitos do domínio, como participar dos procedimentos de investigação, eventos de coleta, experimentos de laboratório, assistir defesas de teses e dissertações,

⁵⁷ Mais informações em: Data Curation Profiles Project - <http://datacurationprofiles.org/about>

conversar com outros investigadores da área, etc. Também pode se querer fazer uma entrevista com outros tipos de questões envolvidas, portanto, pode-se adaptar o mesmo guião conforme objetivos e necessidades próprios.

Ao lado do entendimento dos conjuntos de dados, a entrevista ainda possibilitou uma melhor compreensão da leitura dos artigos publicados pelos investigadores. Apesar disso, nesta fase ainda foi encontrada alguma dificuldade em compreender situações específicas, procedimentos envolvidos ou linguagem técnica. Para sanar estas dúvidas se recorreu a pesquisas na Internet, consulta a dicionários e vocabulários da área, e aos próprios investigadores, os quais foram a melhor fonte de informação. Por este motivo a disponibilidade deles é fundamental para o processo.

Durante a leitura de artigos se procedeu à análise de conteúdo dos mesmos, com identificação e seleção de conceitos considerados importantes. Análise de conteúdo é uma tarefa bastante subjetiva, no entanto pode ser orientada pela Norma Portuguesa NP 3715, de 1989. A norma sugere quais as partes mais importantes de um documento escrito que devem ser analisadas, caso não haja disponibilidade para uma análise do texto integral. Ainda orienta na identificação de conceitos mais representativos do conteúdo do documento, que devem ser agrupados em grelhas segundo critérios que indica.

Portanto, com base na norma, foram estabelecidos critérios gerais e dentro destes alocados os conceitos mais representativos do conteúdo do documento. Abaixo estão os critérios. Em anexo pode ser visto um exemplo completo da análise de conteúdo realizada (Anexo B), juntamente com o artigo analisado (Anexo A).

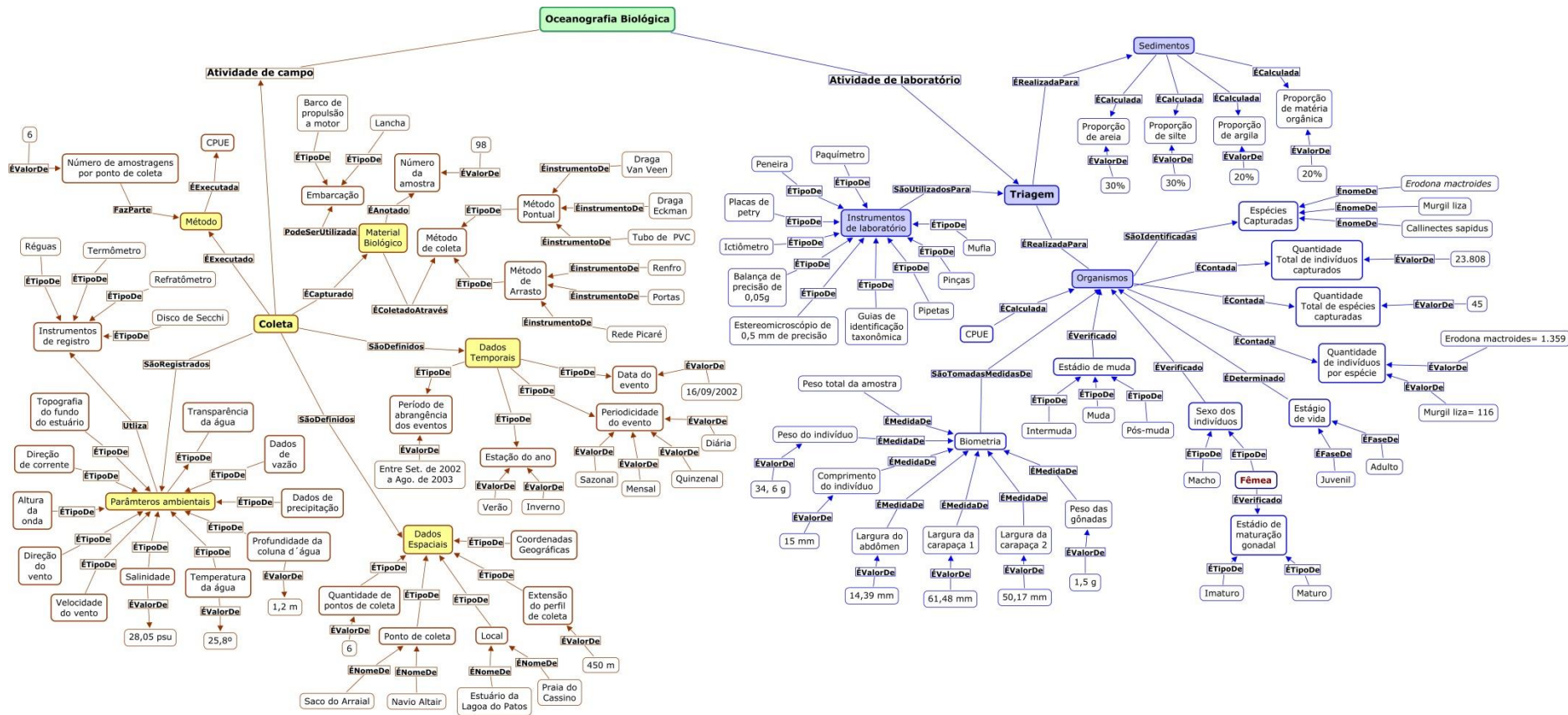
- Identificação de título e autores
- Assuntos
- Variáveis de espaço
- Variáveis de tempo
- Metodologia e instrumentos
- Material coletado/triagem
- Parâmetros ambientais

4.2.2 Mapa de conceitos e escolha de descritores

O levantamento de requisitos (entrevista e análise de conteúdo dos artigos) levou à construção de um mapa de conceitos, onde foi possível modelar o conhecimento da área ao estabelecer relações entre os conceitos. Estes, fundamentalmente extraídos da análise de conteúdo das publicações, formaram um desenho das atividades práticas de investigação e variáveis intrínsecas para a produção de dados.

Por se tratar de três subáreas do mesmo domínio, que apesar de compartilharem características em comum ainda mantêm suas particularidades, foram construídos inicialmente três mapas (que podem ser conferidos nos Anexos C, D e E). Depois de validados com cada um dos investigadores das áreas, estes foram posteriormente unidos em apenas um mapa mais completo do domínio, que se encontra a seguir:

FIGURA 7 – Mapa de Conceitos “Atividades de Investigação em Oceanografia Biológica



A tarefa de modelar conceitos no mapa pode dispende algum tempo e ocasionar várias modificações do mapa original. Isto se deve ao fato de que, no decorrer desta ação, ocorre um aprimoramento sobre a visão de organização e de relações conceituais do conhecimento que se está a estruturar. Ademais, solicita-se sempre a intervenção dos reais conhecedores do domínio (investigadores), o que pode resultar em modificações adicionais. No entanto, tal tarefa é relevante para o processo de aprendizagem, o que culmina em um mapa de conceitos consistente e, por consequência, na escolha mais adequada de potenciais descritores para conceber o perfil de aplicação.

A etapa subsequente é da escolha dos conceitos do mapa que poderão se tornar os descritores do perfil de aplicação. Nesta fase foi importante dar aos investigadores noção do conceito de descritor (metadados) e qual sua função na gestão de dados de investigação. Comparações com exemplos concretos de anotação, que sejam conhecidos dos colaboradores, são úteis. Neste caso foi usado o exemplo de uma descrição de um livro constante no catálogo da biblioteca da Universidade.

Ainda assim, percebeu-se que os investigadores confundiam os dados com os metadados pois, inicialmente, escolheram muitos conceitos do mapa que se constituíam nos próprios dados de investigação, como por exemplo, todos os parâmetros ambientais (que constituem os dados abióticos), ou todas as medidas de biometria (que constituem os dados bióticos). Foi necessário dotá-los de um nível de abstração um pouco maior no que se refere à organização e representação da informação, noção que a tarefa de descrição exige. Em seguida pode ser conferido um excerto do processo de escolha de conceitos pelos investigadores. As partes com realce amarelo e as escritas em vermelho faz parte da interação do criador do perfil de aplicação neste processo.

- **Nas Atividades de Campo:**
 - Número da Amostra;
 - Número de Amostras por Ponto de Coleta (poderia ser descrito nos procedimentos de amostragem, campo que poderia incluir também outros detalhes, como CPUE, por ex.);
 - Parâmetros Ambientais (todos) (realmente necessário para descrever o conjunto de dados? Eles já constam no dataset e podem ser recuperados de outra forma em uma pesquisa);
 - Instrumentos de Coleta de Indivíduos;
 - Período da Coleta;
 - Data da Coleta;
 - Periodicidade da Coleta;
 - Local.
- **Nas Atividades de Laboratório:**
 - Biometria dos Organismos (São informações que constam no dataset, teria que abranger um número muito grande de medidas para todas as espécies);
 - Identificação de Espécies;
 - Total de Indivíduos Capturados;
 - Total de Espécies Capturadas;
 - Quantidade de Indivíduos por Espécie;
 - Estágio de Maturação Gonadal (Também constam no dataset, não seria muito detalhamento?).

Figura 8 – Excerto de processo de seleção de conceitos

Para delimitar a escolha dos conceitos e tornar o processo simplificado foi necessário delinear objetivos para o perfil de aplicação. Estes foram estabelecidos de acordo com demonstrações de dúvidas dos investigadores durante o processo, tais como a quantidade de conceitos que se poderia escolher. Para contornar estas situações os seguintes objetivos foram delimitados:

- Escolher descritores que cubram o domínio de forma abrangente, porém o mais completa possível. Caso o número de descritores fique um pouco extenso não é problema, pois no momento da descrição pode-se escolher usar apenas os que se quer. O importante é que estejam disponíveis em caso de necessidade.
- Definir alguns aspectos de ordem administrativa, relativos ao projeto, tais como descrição ou resumo do projeto, responsável/chefe pelo laboratório, órgão de financiamento.
- Definir elementos que identifiquem o conjunto de dados, tais como título, autoria, assunto, natureza ou tipologia dos dados.
- Identificar material coletado.
- Quantificar espécies e indivíduos coletados.
- Localizar os eventos de coleta de dados no tempo (data, periodicidade de coleta) e no espaço (local, coordenadas geográficas).
- Especificar métodos e instrumentos para amostragens.
- Especificar destinação de materiais coletados.

Da mesma maneira como surgiu a dúvida em relação ao número de descritores, os investigadores também sentiram falta de descritores cujos conceitos não estavam expressos no mapa, descritores estes que seriam úteis numa situação real de descrição. Houve a necessidade de introduzir conceitos que descrevessem aspectos administrativos do projeto do qual derivam os dados, assim como aspetos que caracterizem o próprio conjunto de dados. Esta necessidade foi expressa em forma de objetivo e a partir daí foram selecionados alguns conceitos que os pudessem representar.

A partir daí os conceitos do mapa e os outros que surgiram foram classificados e posicionados conforme os objetivos estabelecidos. Abaixo está um exemplo de como a atividade se desenhou:

CONCEITOS CONFORME OBJETIVOS ESTABELECIDOS

- Definir aspetos de ordem administrativa, relativos ao projeto, tais como descrição ou resumo do projeto, responsável/chefe pelo laboratório, órgão de financiamento.
 - Nome do Laboratório /e/ou/ Nome da Instituição (Lab. Crustáceos Decápodos – FURG)
 - Órgão de financiamento
 - Descrição /ou/ resumo do projeto
 - Data de vigência do projeto

- Definir elementos que identifiquem o conjunto de dados, tais como título, autoria, assunto, natureza ou tipologia dos dados.
 - Autor (responsável pelo projeto que origina a coleta dos dados)
 - Contato do autor (e-mail)
 - Equipe (nomes das pessoas que colaboraram na coleta de dados)
 - Título
 - Assunto (palavras-chave que identifiquem o assunto dos dados)
 - Tipologia dos dados (qualitativo e quantitativo; texto; imagens; valores numéricos, etc.)

- Identificar material coletado.
 - Tipo de material coletado (sedimentos, amostra de água, organismos vivos, organismo inteiro, partes anatômicas, etc.)
 - Identificação das espécies (nome científico – táxon mais básico + autor + ano de identificação)

Figura 9 – Excerto de seleção de conceitos conforme objetivos estabelecidos

4.2.3 Esquemas de metadados e normas de descrição de dados

Durante a etapa de escolha dos conceitos teve início o processo de pesquisa e estudo de normas de descrição de dados e esquemas de metadados existentes no domínio da Oceanografia Biológica ou em domínios relacionados. O conhecimento destas normas também foi útil para orientar os investigadores na seleção de descritores, uma vez que estas poderiam abordar aspetos que não estavam contemplados até o momento, mas que, todavia, poderiam ser cruciais numa situação de descrição.

Depois da seleção definitiva de todos os conceitos iniciou-se a fase de transposição de conceitos para os descritores que formariam o perfil de aplicação. A maioria dos conceitos encontrou compatibilidade semântica nos descritores constantes nos esquemas estudados. Os que não puderam ser encontrados foram criados.

Os descritores atribuídos ao perfil de aplicação foram classificados em duas categorias: descritores gerais e específicos do domínio. Os descritores gerais possuem uma ampla abrangência e se aplicam para descrição de recursos dos mais variados suportes e de áreas de conhecimento diversificadas. São exemplos deste tipo os metadados que representam aspectos descritivos do projeto de investigação, do próprio conjunto de dados, ou de documentos gerados a partir dos dados (artigos, relatórios, etc.), tais como, “title”, “creator”, “date” ou “subject”.

Os descritores específicos são os que buscam mapear o fluxo das atividades de investigação específicas do domínio. Também utilizam vocabulário próprio da área de investigação. Existem casos em que os descritores podem ser generalizados para outros domínios, como é o caso dos descritores “methods” e “instrumentation”, porém a maioria só faz sentido dentro de uma mesma área de investigação ou área interdisciplinar, como é o caso dos descritores “scientific name” e “life stage”, que se inserem no contexto das ciências biológicas.

Portanto, foram estudados esquemas e normas que contemplassem as duas categorias de descritores, de forma a elaborar um perfil de aplicação completo. Tais esquemas foram localizados no sítio eletrónico do Digital Curation Centre⁵⁸, onde há uma secção que lista padrões para descrição de dados, bem como casos de usos, em vários domínios. Esta secção organiza os padrões por disciplinas ou em lista geral.

⁵⁸ Digital Curation Centre. Resources. Metadata Standards. Disciplinary Metadata - (<http://www.dcc.ac.uk/resources/metadata-standards>)

De acordo com esta classificação por disciplinas foram escolhidos quatro padrões nas categorias “Biology” (Darwin Core, EML e OBIS) e “Earth Science” (FGDC/CSDGM) e um padrão da categoria “General Research Data” (Dublin Core). Estes estão listados na Tabela 1:

Tabela 1 – Padrões de metadados

Standard	Description
Darwin Core	<p>A body of standards, including a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing reference definitions, examples, and commentaries.</p> <p>See: http://rs.tdwg.org/dwc/terms/index.htm</p>
Dublin Core	<p>A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards.</p> <p>See: http://dublincore.org/documents/dcmi-terms/</p>
Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile	<p>A profile of the FGDC/CSDGM metadata standard, intended to support the collection and processing of biological data.</p> <p>See: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf</p>
EML - Ecological Metadata Language	<p>Ecological Metadata Language (EML) is a metadata specification particularly developed for the ecology discipline. It is based on prior work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications).</p> <p>See: https://knb.ecoinformatics.org/#external//emlparser/docs/eml-2.1.1/index.html</p>
OBIS Data Schema	<p>The OBIS schema is a list of data fields with names, descriptions, and format notes. It is an extension to the Darwin Core Version 2 standard. The OBIS Schema also contains some additional fields for holding information that the Darwin Core does not handle.</p> <p>See: http://www.iobis.org/node/304</p>

FONTES: Digital Curation Centre (<http://www.dcc.ac.uk/resources/metadata-standards/list>) e Ocean Biogeography Information System (<http://iobis.org/node/304>)

Após análise dos padrões de metadados foi estabelecida a seguinte ordem de preferência de uso:

- 1) Dublin Core: primeiro na ordem de preferência por ser um padrão amplamente conhecido e utilizado em várias aplicações na web, portanto o seu uso garante interoperabilidade com outras plataformas e repositórios; o Dublin Core representa os metadados de ordem geral para descrição de recursos no perfil de aplicação e, além do mais, possui uma ontologia já carregada na plataforma Dendro;
- 2) EML: norma bastante consistente e abrangente no domínio da Ecologia, portanto aplicável ao domínio da Oceanografia Biológica, que estuda as interações ecológicas entre ambientes e organismos marinhos e estuarinos; também já havia sido utilizada em trabalhos prévios realizados pelo InfoLab;
- 3) Darwin Core: apresenta um vocabulário bastante compatível com o domínio, vindo a complementar a norma EML, além disso, também fornece uma ontologia para representação dos conceitos;
- 4) OBIS: apesar de ser voltada para oceanografia esta norma é uma extensão do Darwin Core e sua aplicabilidade se restringe ao portal OBIS – Ocean Biogeographic Information System, um repositório de dados;
- 5) CSDGM Part 1: Biological Data Profile: o vocabulário desta norma se apresentou mais restrito do que o das normas anteriormente citadas, motivo pelo qual foi preterido a estas;
- 6) BIOLOGICAL OCEANOGRAPHY: por fim, nos casos em que nenhuma das normas estudadas conseguiu representar o conceito que se queria expressar foi criado o vocabulário “Biological Oceanography”, prefixo “bioen”.

Ao todo foram selecionados trinta e um (31) conceitos, entre gerais e específicos, para compor o perfil de aplicação. Para cada um dos conceitos foram analisadas as normas apresentadas na tabela acima a fim de se encontrar compatibilidade semântica entre conceito e descritor. Os descritores que correspondiam semanticamente aos conceitos selecionados eram escolhidos para compor o perfil conforme a ordem de preferência apresentada. Para os conceitos que não encontraram conformidade semântica os descritores foram criados sob medida.

Na Tabela 2 é possível visualizar todos os conceitos selecionados, os descritores correspondentes, os respectivos padrões de metadados e uma descrição de cada descritor

de acordo com seu padrão: A tabela foi dividida em sete (7) partes nas quais os descritores estão agrupados conforme os objetivos do perfil de aplicação.

A Parte 1 da tabela apresenta quatro (4) descritores selecionados de acordo com o objetivo de definir alguns aspectos de ordens administrativa ou descritiva relativos ao projeto de investigação ao qual os dados estão ligados, como agência financiadora do projeto, ou instituição que o abriga.

Tabela 2 – Parte 1 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
OBJETIVO: Definir aspectos de ordem administrativa relativos ao projeto			
1) Nome do Laboratório /e/ou/ Nome da Instituição	Organization Name	EML	The responsible party field contains the full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. Ex.: National Center for Ecological Analysis and Synthesis
2) Órgão de financiamento	Funding	EML	The funding field is used to provide information about funding sources for the project such as: grant and contract numbers; names and addresses of funding sources. Other funding-related information may also be included.
3) Data de vigência do projeto	Date	Dublin Core	A point or period of time associated with an event in the lifecycle of the resource.
4) Descrição /ou/ resumo do projeto	Description	Dublin Core	An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

A Parte 2 apresenta seis (6) descritores selecionados para atingir o objetivo de definir elementos que identifiquem e descrevam o conjunto de dados e quem o criou, com vista à recuperação desta informação. Tanto a Parte 1 como a parte 2 são compostas por descritores de ordem mais genérica, aplicáveis a descrição de recursos variados. Grande parte dos descritores são oriundos do padrão Dublin Core.

Tabela 2 – Parte 2 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
----------	---------------------	---------	-----------

OBJETIVO: Definir elementos que identifiquem o conjunto de dados			
5) Autor (responsável pelo projeto que origina a coleta dos dados)	Creator	Dublin Core	An entity primarily responsible for making the resource.
6) Contato do autor (e-mail)	Electronic Mail Address	EML	The electronic mail address is the email address for the party. It is intended to be an Internet SMTP email address, which should consist of a username followed by the @ symbol, followed by the email server domain name address. Other address types are allowable. Example(s): my-email@mydomain.edu
7) Equipe que contribuiu com coleta de dados	Contributor	Dublin Core	An entity responsible for making contributions to the resource. Examples of a Contributor include a person, an organization, or a service.
8) Título	Title	Dublin Core	A name given to the resource.
9) Assunto	Subject	Dublin Core	The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes.
10) Tipologia dos dados	Type	Dublin Core	The nature or genre of the resource. E.g.: Text, sound, image.

A terceira parte da Tabela 2 exibe cinco (5) descritores selecionados de acordo com o objetivo de identificar material e organismos capturados em eventos de coleta. A partir desta da Parte 3, os descritores apresentados podem ser classificados como específicos do domínio.

Tabela 2 – Parte 3 – Transposição de conceitos para descritores

CONCEITO	DESCRITOR SUGERIDO	ESQUEMA	DESCRIÇÃO
OBJETIVO: Identificar material coletado			
11) Tipo de material coletado	Specimen	EML	A word or phrase describing the type of specimen collected. Example(s): herbarium specimens,

			blood samples, photographs, individuals, or batches
12) Identificação das espécies	Scientific Name	Darwin Core	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. Examples: "Coleoptera" (order), "Vespertilionidae" (family), "Manis" (genus), "Ctenomys sociabilis" (genus + specificEpithet), "Ambystoma tigrinum diaboli" (genus + specificEpithet + infraspecificEpithet), "Roptrocerus typographi (Györfi, 1952)"
13) Nome comum	Common Name	EML	Specification of applicable common names. These common names may be general descriptions of a group of organisms if appropriate. Example(s): insects, vertebrate, grasses, waterfowl, vascular plants, red maple.
14) Estágio de vida das espécies coletadas	Life Stage	Darwin Core	The age class or life stage of the biological individual(s) at the time the sampling event. Recommended best practice is to use a controlled vocabulary. Examples: "egg", "eft", "juvenile", "adult", "2 adults 4 juveniles".
15) Sexos dos indivíduos	Sex	Darwin Core	The sex of the biological individual(s) collected. Examples: "female", "hermaphrodite", "8 males, 4 females".

A Parte 4 apresenta quatro (4) descritores selecionados de acordo com o objetivo de quantificar o material coletado, apresenta descritores de contagem e mensuração das espécies e indivíduos capturados.

Tabela 2 – Parte 4 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
OBJETIVO: Quantificar espécies e indivíduos coletados			
16) Total de Indivíduos Capturados	Individual Count	Darwin Core	The number of individuals represented present at the time of the sampling event.

17) Total de Espécies Capturadas	Species Count	Biological Oceanography	The total number of species caught in a sample event, expressed by arabic numbers. E.g.: 5, 17.
18) Quantidade de Indivíduos por Espécie	Individual Per Species	Biological Oceanography	The quantity of individuals caught per species in a sample event. E.g.: Callinectes sapidus = 5, Murgil liza = 17.
19) Peso total da amostra	Observed Weight	OBIS	The total biomass found in a collection/record event. Expressed as kg.

A Parte 5 apresenta seis (6) descritores selecionados com o objetivo de contextualizar os eventos de coleta de dados no tempo e no espaço.

Tabela 2 – Parte 5 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
OBJETIVO: Localizar os eventos de coleta de dados no tempo e no espaço			
20) Local	Coverage	Dublin Core	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates.
21) Descrição do local de coleta	Geographic Description	EML	geographicDescription is a short text description of a dataset's geographic areal domain. A text description is especially important to provide a geographic setting when the extent of the data set cannot be well described by the "boundingCoordinates", or in the case of data which are not specifically geospatial.
22) Data da coleta	Single Date Time	EML	The singleDateTime field is intended to describe a single date and time for an event. There is a choice between two options: a calendar date with a time, or a geologic age.
23) Período de abrangência das coletas	Begin Date	EML	A single time stamp signifying the beginning of some time period. There is a choice between two options: a calendar date with a time, or a geologic age.
24) Período de abrangência das coletas	End Date	EML	A single time stamp signifying the end of some time period. There is a choice between two options: a calendar date

			with a time, or a geologic age.
25) Periodicidade das coletas	Sampling Periodicity	Biological Oceanography	This field expresses the time interval between sample events. For e.g.: daily, fortnightly, monthly, seasonal...

A Parte 6 apresenta cinco (5) descritores selecionados com o objetivo descrever os métodos e os instrumentos utilizados tanto para os eventos de coleta em campo, como para as atividades de laboratório.

Tabela 2 – Parte 6 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
OBJETIVO: Especificar métodos e instrumentos para amostragens			
26) CPUE	Sampling Effort	Darwin Core	The amount of effort expended during an Event. Examples: "40 trap-nights", "10 observer-hours; 10 km by foot; 30 km by car".
27) Método de coleta	Sampling Description	EML	The samplingDescription field allows for a text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.
28) Método de triagem	Methods	EML	The methods field documents scientific methods used in the collection of this dataset. It includes information on items such as tools, instrument calibration and software.
29) Instrumentos de coleta / Instrumentos de laboratório	Instrumentation	EML	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance. The description should include vendor, model number, optional equipment, etc.
30) Rótulo da amostra	Sample Identification	Biological Oceanography	An identifier created at collection time to identify the specimen collected. This field allows numbers, letters, simple words or codes, e.g. a numerical sequence or a composite code of letters and numbers.

Por fim, a Parte 7 da Tabela 2 mostra um (1) descritor selecionado de acordo com o objetivo de descrever a destinação final dada a uma amostra depois de analisada.

Tabela 2 – Parte 7 – Transposição de conceitos para descritores

CONCEITO	DESCRIPTOR SUGERIDO	ESQUEMA	DESCRIÇÃO
----------	---------------------	---------	-----------

OBJETIVO: Especificar destinação de materiais coletados			
31) Destinação da amostra	Sample Destination	Biological Oceanography	Describes the final destination of a sample after used in the research analysis. For e.g.: discarded, destroyed, donated to another research project, deposited in the lab collection, etc.

Existiram casos em que alguns conceitos encontraram compatibilidade semântica com mais de um descritor, pelo que foi preciso avaliar junto a um dos investigadores qual a melhor alternativa. Nestes casos a escolha do descritor pela ordem de preferência de padrões foi flexibilizada por julgar que o descritor de outra norma se apresentou mais adequado. No exemplo abaixo, o descritor “organization name” da norma EML foi escolhido no lugar do descritor “contributor”, do padrão Dublin Core, embora este fosse o primeiro na lista de descritores preferidos.

DESCRITOR	METADADO SUGERIDO	ESQUEMA	DESCRIÇÃO DO METADADO
1) Nome do Laboratório /e/ou/ Nome da Instituição	<u>Institution Code</u>	OBIS	A "standard" code identifier that identifies the institution to which the collection belongs, if there is one. Use the code that is "standard" in your discipline, if there is one (no global registry exists for assigning institutional codes). If not, use a short version of the name of the institution. (e.g. "NMNH" for Smithsonian National Museum of Natural History or "Duke" for Duke University).
	<u>Contributor</u>	Dublin Core	An entity responsible for making contributions to the resource. Examples of a Contributor include a person, an organization, or a service.
	<u>Organization Name</u>	EML	The responsible party field contains the full name of the organization that is associated with the resource. This field is intended to describe which institution or overall organization is associated with the resource being described. Ex.: <u>National Center for Ecological Analysis and Synthesis</u>

Figura 10 – Excerto de processo de seleção de descritores

Desta forma, todos os descritores foram avaliados e validados junto aos investigadores para então ser montado o perfil de aplicação “Biological Oceanography”, que pode ser visto na tabela abaixo. Os descritores são apresentados com seus respectivos prefixos, que são as siglas das normas que os originaram. O prefixo “biocn” é do próprio perfil de aplicação “Biological Oceanography”, e foi atribuído para os descritores criados para o perfil. Cada descritor apresenta também uma descrição e um exemplo de aplicação.

Tabela 3 – Perfil de aplicação Biological Oceanography

APPLICATION PROFILE BIOLOGICAL OCEANOGRAPHY		
DESCRIPTOR	DESCRIPTION	EXEMPLE
eml:beginDate	A single time stamp signifying the beginning of some time period. There is a choice between two options: a calendar date with a time, or a geologic age.	2010-09-12
eml:commonName	Specification of applicable common names. These common names may be general descriptions of a group of organisms if appropriate.	Blue crab
dc:contributor	An entity responsible for making contributions to the resource. Examples of a Contributor include a person, an organization, or a service.	Interdisciplinary Centre of Marine and Environmental Research
dc:coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates.	Lagoa dos Patos
dc:creator	An entity primarily responsible for making the resource.	Silva, João da
dc:date	A point or period of time associated with an event in the lifecycle of the resource.	2013-09-16 a 2015-07-30
dc:description	An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.	Abstract of the research project
eml:eletronicMailAdress	The electronic mail address is the email address for the party. It is intended to be an Internet SMTP email address, which should consist of a username followed by the @ symbol, followed by the email server domain name address. Other address types are allowable.	email@mydomain.edu
eml:endDate	A single time stamp signifying the end of some time period. There is a choice between two options: a calendar date with a time, or a geologic age.	2011-08-25
eml:funding	The funding field is used to provide information about funding sources for the project such as: grant and contract numbers; names and addresses of funding sources. Other funding-related information may also be included.	FCT/Portugal (http://www.fct.pt/)
eml:geographicDescription	geographicDescription is a short text description of a dataset's geographic areal domain.	Fieldwork was performed along a transect with a length of about 450 m in the Saco do Arraial inlet, in

		a shallow plain in front of the eastern margin of Pombas Island (32°01'S, 52°07'W) in the Patos Lagoon estuarine area.
dwc:individualCount	The number of individuals represented present at the time of the sampling event.	85
biocn:individualsPerSpecies	The quantity of individuals caught per species in a sample event.	Callinectes sapidus = 5, Murgiliza = 17.
eml:instrumentation	The Instrumentation field allows the description of any instruments used in the data collection or quality control and quality assurance. The description should include vendor, model number, optional equipment, etc.	Beach seine net (9 m long; 1.5 m high) with a 13 mm stretch mesh in the wings and a 5 mm stretch mesh in the centre 3 m section.
dwc:lifeStage	The age class or life stage of the biological individual(s) at the time the sampling event. Recommended best practice is to use a controlled vocabulary.	2 adults, 4 juveniles
eml:methods	The methods field documents scientific methods used in the collection of this dataset. It includes information on items such as tools, instrument calibration and software.	In the laboratory, E. mactroides individuals were sorted from the sediment under stereomicroscope and then preserved in a 70% ethanol solution.
obis:observedWeight	The total biomass found in a collection/record event. Expressed as kg.	15 kg
eml:organizationName	The responsible party field contains the full name of the organization that is associated with the resource.	University of Porto/ InfoLab
biocn:sampleDestination	Describes the final destination of a sample after used in the research analysis.	Discarded.
biocn:sampleIdentification	An identifier created at collection time to identify the specimen collected. This field allows numbers, letters, simple words or codes, e.g. a numerical sequence or a composite code of letters and numbers.	PL86S (combines the letters "PL" which refer to a location Patos Lagoon, the number "86" representing the sample number, and the letter "S", to specify the season, in this case Summer).
eml:samplingDescription	Allows for a text-based/human readable description of the sampling procedures used in the research project. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article.	Six stations 90 m apart from each other were sampled monthly from September 2002 to August 2003. Three biological samples were taken from each station, with a 10 cm diameter corer, pushed 20 cm into the bottom.
dwc:samplingEffort	The amount of effort expended during	40 trap-nights

	an Event.	
biocn:samplingPeriodicity	This field expresses the time interval between sample events.	Monthly
dwc:scientificName	The full scientific name, with authorship and date information if known. When forming part of an identification, this should be the name in lowest level taxonomic rank that can be determined.	<i>Callinectes</i> <i>sapidus</i> Rathbun 1896
dwc:sex	The sex of the biological individual(s) collected.	8 males, 4 females
eml:singleDateTime	The singleDateTime field is intended to describe a single date and time for an event. There is a choice between two options: a calendar date with a time, or a geologic age.	2015-03-22 10AM
biocn:speciesCount	The total number of species caught in a sample event, expressed by arabic numbers.	17
eml:specimen	A word or phrase describing the type of specimen collected.	Sediment
dc:subject	The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes.	Reproductive biology
dc:title	A name given to the resource.	
dc:type	The nature or genre of the resource.	Numerical values

4.3 Formalização da ontologia e sua ingestão na plataforma Dendro

A etapa posterior à elaboração do perfil de aplicação trata da sua formalização em ontologia, para que desta forma possa ser incorporada à plataforma de gestão de dados e nela os descritores possam atuar como parte do modelo de metadados para a descrição de dados de investigação e outros recursos.

A ontologia dentro da plataforma funciona como um vocabulário usado para descrever um domínio, onde as propriedades são a representação dos descritores do perfil de aplicação. A ontologia é a ferramenta apropriada para fazer tal representação uma vez que apresenta a flexibilidade semântica que a tarefa exige. É importante que a linguagem da ontologia seja a mesma do perfil de aplicação construído em colaboração com os investigadores, para não haver barreira no entendimento dos descritores.

Uma das ideias centrais de representar o perfil de aplicação através de uma ontologia é poder aproveitar o vocabulário de outras ontologias para um objetivo particular. Portanto, para a construção da ontologia “Biological Oceanography” se procedeu a uma pesquisa de ontologias existentes que pudessem ter uma representação dos descritores do perfil de aplicação. Tal processo foi semelhante ao que foi feito com o estudo dos padrões de metadados para compor o perfil de aplicação, o objetivo é encontrar compatibilidade semântica entre os descritores do perfil e os termos das ontologias. Para este caso, não foi preciso analisar as relações entre os conceitos das ontologias, uma vez que esta “lightweight ontology” só se ocupa de classes e propriedades.

Foram analisadas algumas ontologias recomendadas pelos colegas do InfoLab e outras recomendadas pelo World Wide Web Consortium (W3C)⁵⁹, como podem ser visualizadas na Tabela 4:

Tabela 4 – Ontologias

Ontologia	URL	Descrição
Dublin Core	http://dublincore.org/documents/dcmi-terms/	This is a light weight RDFS vocabulary for describing generic metadata. DC is a moderately small ontology divided into 2 vocabularies: DC elements and DC terms. DC elements contain 15 properties. DC terms contain 22 classes and 55 properties.
Oboé	https://semtools.coinformatics.org/oboe	OBOE is a suite of OWL-DL ontologies for modeling and representing scientific observations. The OBOE model is designed as a generic data model with a number of constructs for defining observational data. Key features of OBOE include its ability to represent a wide range of measurement types, a mechanism for specifying measurement context, and the ability to associate the type of entity (e.g., sample, organism, etc.) being measured. OBOE is being used and developed within the Semtools project for describing a wide variety of ecological data stored within the Knowledge Network for Biodiversity (KNB) as well as extensions for ontology-based data annotation and discovery within the MetaCat software infrastructure.
CERIF	http://www.eurocris.org/ontologies/cerif/1.3#Funding	The <i>Common European Research Information Format (CERIF) Ontology Specification</i> provides basic concepts and properties for describing research information as semantic data. This document contains a friendly

⁵⁹ World Wide Web Consortium (W3C) - (http://www.w3.org/wiki/Good_Ontologies)

		description of the <i>Common European Research Information Format (CERIF) Ontology</i> developed by EuroCRIS.
Friend of a friend (FOAF)	http://xmlns.com/foaf/spec/	FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked. FOAF integrates three kinds of network: <i>social networks</i> of human collaboration, friendship and association; <i>representational networks</i> that describe a simplified view of a cartoon universe in factual terms, and <i>information networks</i> that use Web-based linking to share independently published descriptions of this inter-connected world. FOAF does not compete with socially-oriented Web sites; rather it provides an approach in which different sites can tell different parts of the larger story, and by which users can retain some control over their information in a non-proprietary format.
TGWD Ontologies The TaxonConcept Ontology The Darwin-SW Ontology	https://code.google.com/p/tdwg-rdf/wiki/BiodiversityOntologies#4 The Darwin-SW Ontology	Summary of some ontologies/data models which provide object properties that can connect instances of biodiversity-related classes
Marine TLO	http://www.ics.forth.gr/isl/MarineTLO/	MarineTLO is a top level ontology, generic enough to provide consistent abstractions or specifications of concepts included in all data models or ontologies of marine data sources and provide the necessary properties to make this distributed knowledge base a coherent source of facts relating observational data with the respective spatiotemporal context and categorical (systematic) domain knowledge.

Das ontologias acima, a “Dublin Core” e a “Friend of a Friend” já estão incorporadas na plataforma Dendro, portanto poderia haver compatibilidade de descritores. As ontologias “OBOE”, “CERIF” e “Marine TLO”, apesar de serem altamente recomendadas não apresentaram compatibilidade de vocabulário para com os descritores do perfil de aplicação, uma vez que as propriedades das ontologias necessitam ser a representação semântica fiel dos descritores, portanto não foram utilizadas para o vocabulário “Biological Oceanography”. A ontologia “Darwin Core” cuja norma já havia sido utilizada para compor o perfil de aplicação apresentou compatibilidade.

Após esta análise foi procedida uma outra para verificar quais descritores poderiam ser aproveitados das ontologias já existentes na plataforma Dendro. Verificou-se que muitos descritores do perfil de aplicação já estavam contemplados em outras

ontologias do Dendro, como se pode ver na tabela abaixo. Apenas os descritores inexistentes precisaram ser formalizados na ontologia “Biological Oceanography”.

A tabela abaixo mostra os descritores do perfil de aplicação, o padrão de metadados ao qual eles pertencem (se existente ou criado para o domínio), e em seguida o termo correspondente em ontologia e o nome da respectiva ontologia a que pertence:

Tabela 5 – Termos de ontologias para representação de descritores

DESCRITOR	PADRÃO	TERMO	ONTOLOGIA
Begin Date	EML	beginDate	Biological Oceanography
Common Name	EML	commonName	Biological Oceanography
Contributor	Dublin Core	contributor	Dublin Core
Coverage	Dublin Core	coverage	Dublin Core
Creator	Dublin Core	creator	Dublin Core
Date	Dublin Core	date	Dublin Core
Description	Dublin Core	description	Dublin Core
Electronic Mail Address	EML	personalMailbox	FOAF
End Date	EML	endDate	Biological Oceanography
Funding	EML	fundedBy	FOAF
Geographic Description	EML	geographicDescription	Biological Oceanography
Individual Count	Darwin Core	individualCount	Darwin Core
Individuals Per Species	BIOCN	individualsPerSpecies	Biological Oceanography
Instrumentation	EML	instrumentation	Dendro Research (InfoLab)
Life Stage	Darwin Core	lifeStage	Darwin Core
Methods	EML	method	Dendro Research (InfoLab)
Observed Weight	OBIS	observedWeight	Biological Oceanography
Organization Name	EML	organizationName	Biodiversity Evolution Studies (InfoLab)
Sample Destination	BIOCN	sampleDestination	Biological Oceanography
Sample Identification	BIOCN	sampleIdentification	Biological Oceanography
Sampling Description	EML	samplingDescription	Biological Oceanography
Sampling Effort	Darwin Core	samplingEffort	Darwin Core
Sampling Periodicity	BIOCN	samplingPeriodicity	Biological Oceanography
Scientific Name	Darwin Core	scientificName	Darwin Core

Sex	Darwin Core	sex	Darwin Core
Single Date Time	EML	singleDateTime	Biological Oceanography
Species Count	BIOCN	speciesCount	Biological Oceanography
Specimen	EML	specimen	Double Cantilever Beam (InfoLab)
Subject	Dublin Core	subject	Dublin Core
Title	Dublin Core	title	Dublin Core
Type	Dublin Core	type	Dublin Core

A partir deste mapeamento os descritores ainda não contemplados nas ontologias da plataforma Dendro foram formalizados no software Protégé. Foi utilizada como base para sua criação a ontologia “Dendro Reasearch”, onde foi criada uma classe chamada “Observation”, por se tratar de um tipo de investigação de caráter sobretudo observacional. Dentro desta foi criada uma subclasse com nome do domínio “BiologicalOceanography”. As classes podem ser visualizadas na Figura 11:

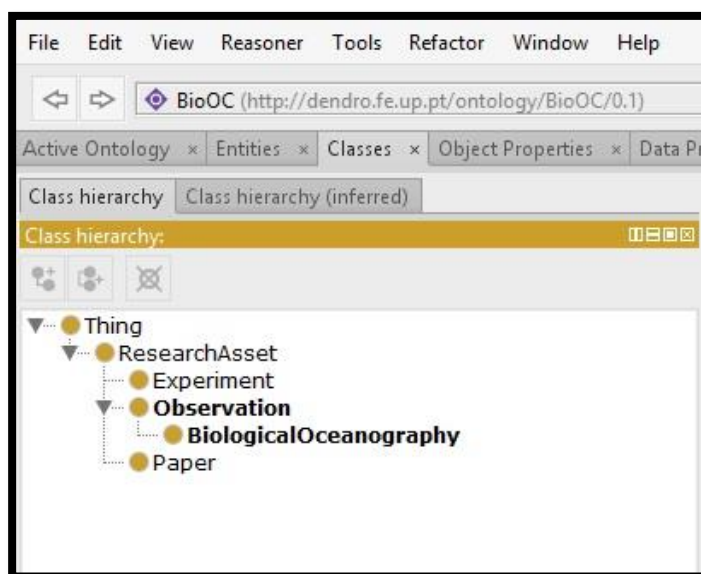


Figura 11 – Classe BiologicalOceanographyObservation

A seguir foram criadas as *data properties* com os nomes dos descritores do perfil de aplicação, como se pode ver na figura a seguir. As propriedades “Measurement”, “Method” e “Sample” já existiam na ontologia Dendro Research e foram aproveitadas. As propriedades restantes, que se encontram em negrito, foram criadas.



Figura 12 – Propriedades da ontologia Biological Oceanography

Algumas propriedades foram consideradas subpropriedades de “Measurement” e “Method” e colocadas como tal. Por exemplo, os descritores “Individual Count”, que é o número total de indivíduos coletados, “Individuals per Species”, que é a quantidade de indivíduos coletados por espécie capturada, “Observed Weight”, que se refere ao total da amostra coletada em quilos, e “Species Count”, quantidade total de espécies capturadas num evento de coleta, foram considerados unidades de contagem e medidas, portanto posicionados na propriedade “Measurement”.

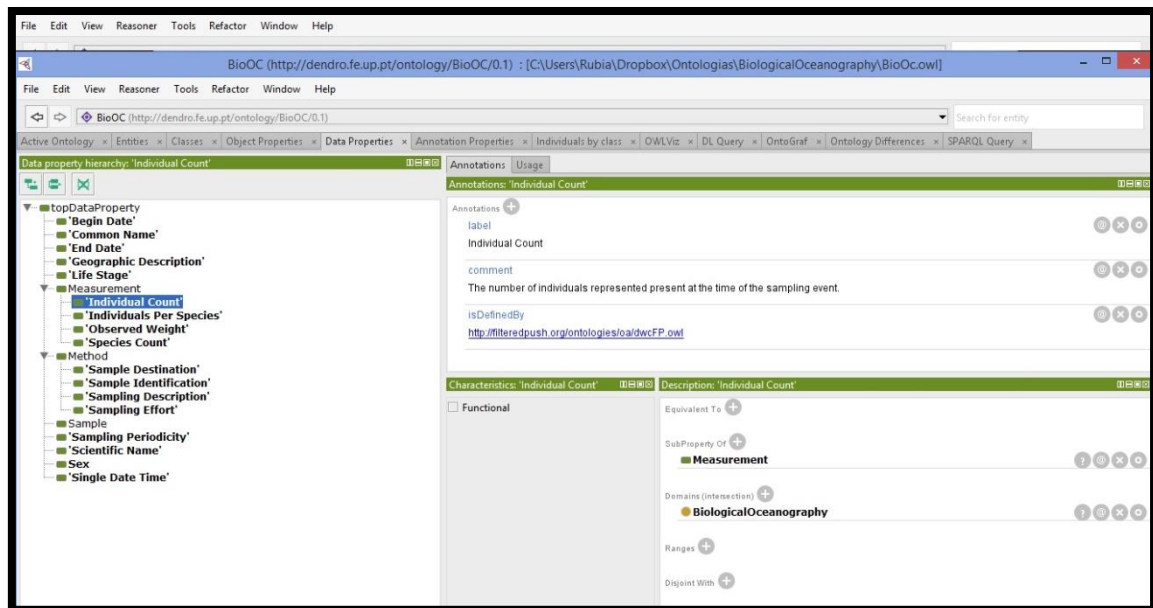


Figura 13 –Subpropriedades na ontologia Biological Oceanography

Para cada uma das propriedades criadas foi inserida informação adicional para sua caracterização. Por exemplo, foram inseridas a *annotation property* “label”, para estabelecer o formato como o descritor deve aparecer; “comment”, onde constou uma descrição do conceito de cada descritor e um exemplo; e “isDefinedBy”, que foi usada apenas nos casos de um descritor ter sido criado por alguma norma específica. Neste caso, foi colocada a URL da norma que definiu o descritor, como no exemplo abaixo, em que foi colocada a URL da norma EML que definiu o descritor “Begin Date”. Ainda foi feita a relação das propriedades com as respectivas classes, através de *Domains (intersection)*, a maioria das propriedades foi ligada à subclasse de “Observation” “BiologicalOceanography”, ver Figura 14.

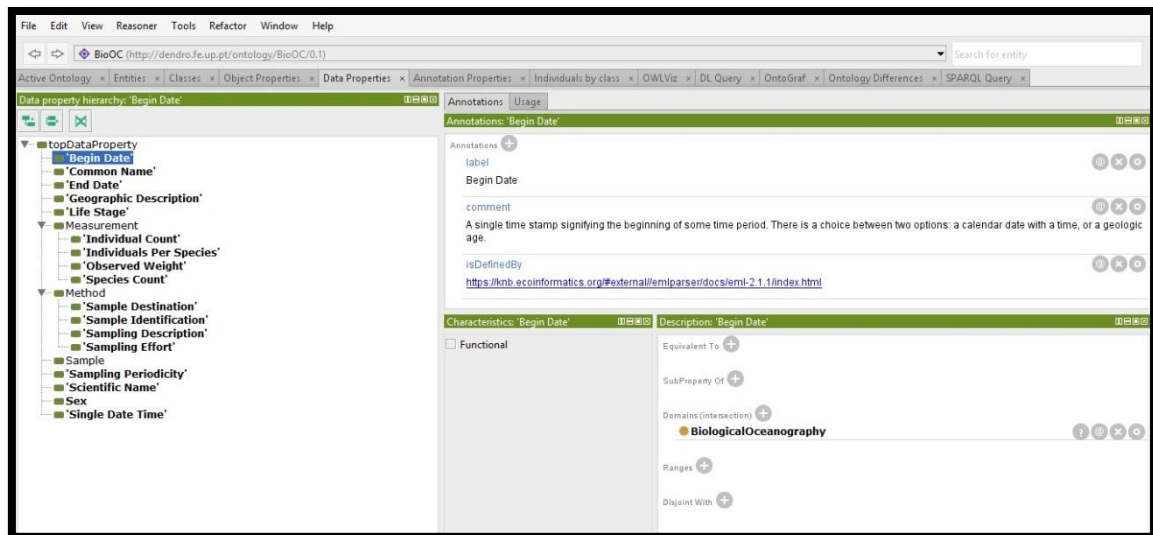


Figura 14 – Annotation Properties e Domains (Intersection) na ontologia Biological Oceanography

Depois de pronta a ontologia foi introduzida na plataforma de gestão de dados Dendro, a qual apresenta uma série de perfis de aplicação formalizados em ontologia. Alguns destes perfis são constituídos de descritores genéricos, que se aplicam a várias situações de descrição independentemente do domínio a ser descrito, como é o caso, por exemplo, dos vocabulários Dublin Core, Friend of a Friend e Dendro Research (este último com descritores genéricos aplicáveis às ciências). O restante dos perfis já são constituídos de descritores para anotação de dados de investigação e outros recursos de domínios específicos.

Na Figura 15 se pode visualizar a interface de descrição da plataforma Dendro. À esquerda estão os recursos da plataforma para que os investigadores criem pastas para seus projetos e depositem ali os ficheiros (conjuntos de dados e outros recursos que pretendam armazenar). Ambos, pastas e ficheiros, são passíveis de descrição. À direita encontra-se a lista de conjuntos de descritores, sob o cabeçalho “Descriptors”. Os descritores destes vocabulários podem ser selecionados manualmente ou, quando ativado o modo “Auto”, são recomendados aos investigadores. Ao centro da figura está a área de descrição, onde aparecem alguns descritores já selecionados e preenchidos para descrever um ficheiro.

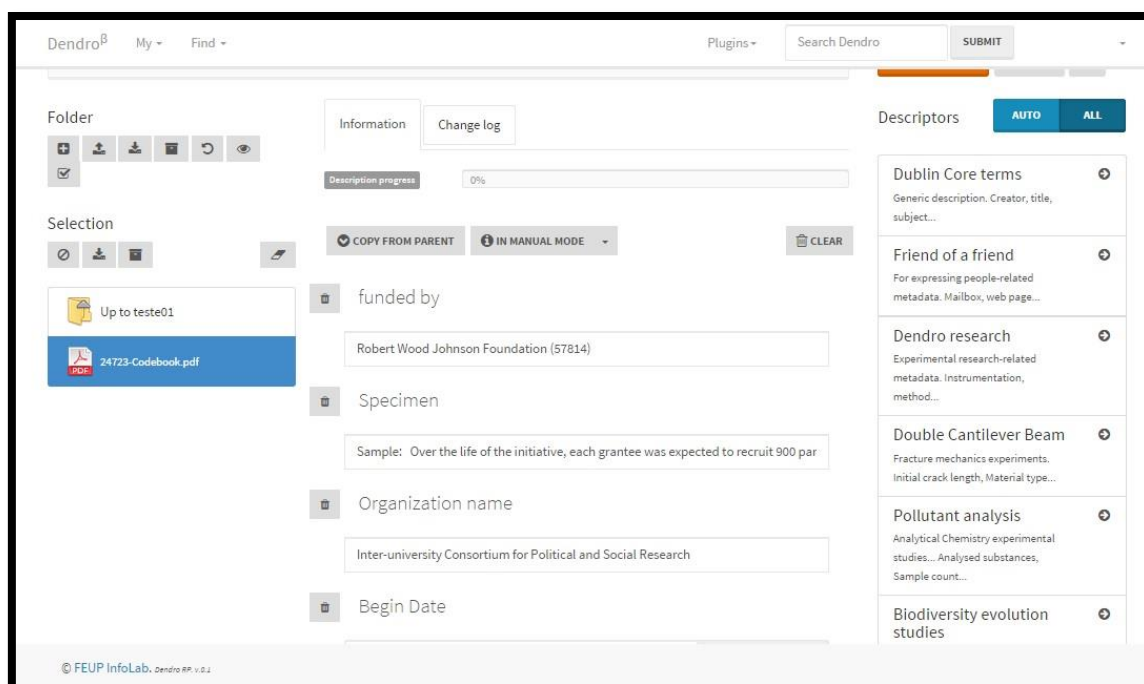


Figura 15 – Lista de conjuntos de descritores no Dendro

O vocabulário “Biological Oceanography” aparece nesta lista da direita e contribui para a oferta multidisciplinar de descritores da plataforma Dendro, apresentando descritores específicos do domínio da Oceanografia Biológica. A seguir, na Figura 16, se pode visualizar o vocabulário, que apresenta uma breve descrição de seu conteúdo. Para a visualização dos descritores é necessário carregar sobre o nome da ontologia ou sobre a seta, para então abrir o menu dos descritores.

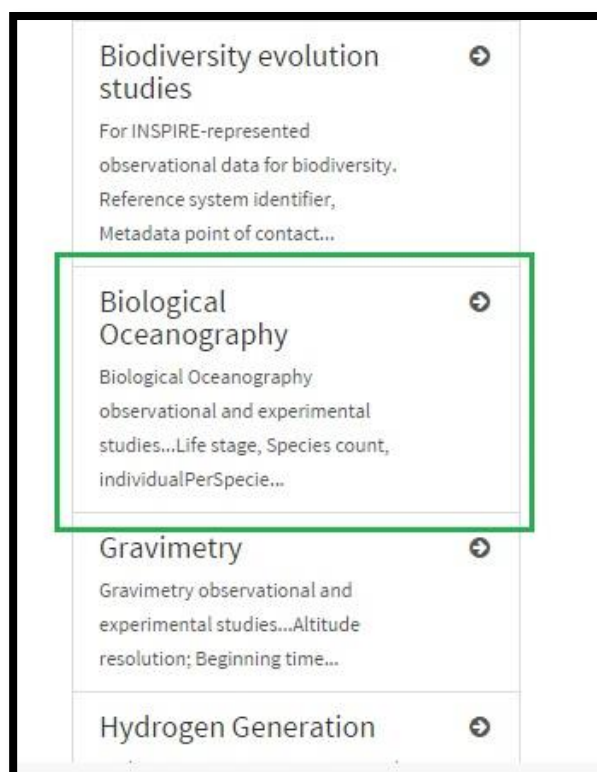


Figura 16 – Ontologia “Biological Oceanography” na lista de descritores do Dendro

Vê-se abaixo um exemplo de como se apresentam os descritores dentro da ontologia na plataforma. Estes ficam dispostos por ordem alfabética e apresentam a descrição que lhes foi associada na *annotation property* da ontologia. Deve-se ressaltar que o perfil de aplicação completo, criado para descrição de dados no domínio da Oceanografia Biológica, não pode ser visualizado apenas sob o vocabulário “Biological Oceanography”. Como muitos descritores já existiam previamente em outros vocabulários da plataforma Dendro, não foram todos os descritores do perfil formalizados na ontologia “Biological Oceanography”. Para se ter acesso ao perfil completo deve-se percorrer os seguintes vocabulários da plataforma Dendro:

- Biodiversity Evolution Studies (1 descritor)
- Biological Oceanography (17 descritores)
- Dendro Research (2 descritores)
- Double Cantilever Beam (1 descritor)
- Dublin Core (8 descritores)
- Friend of a Friend (2 descritores)

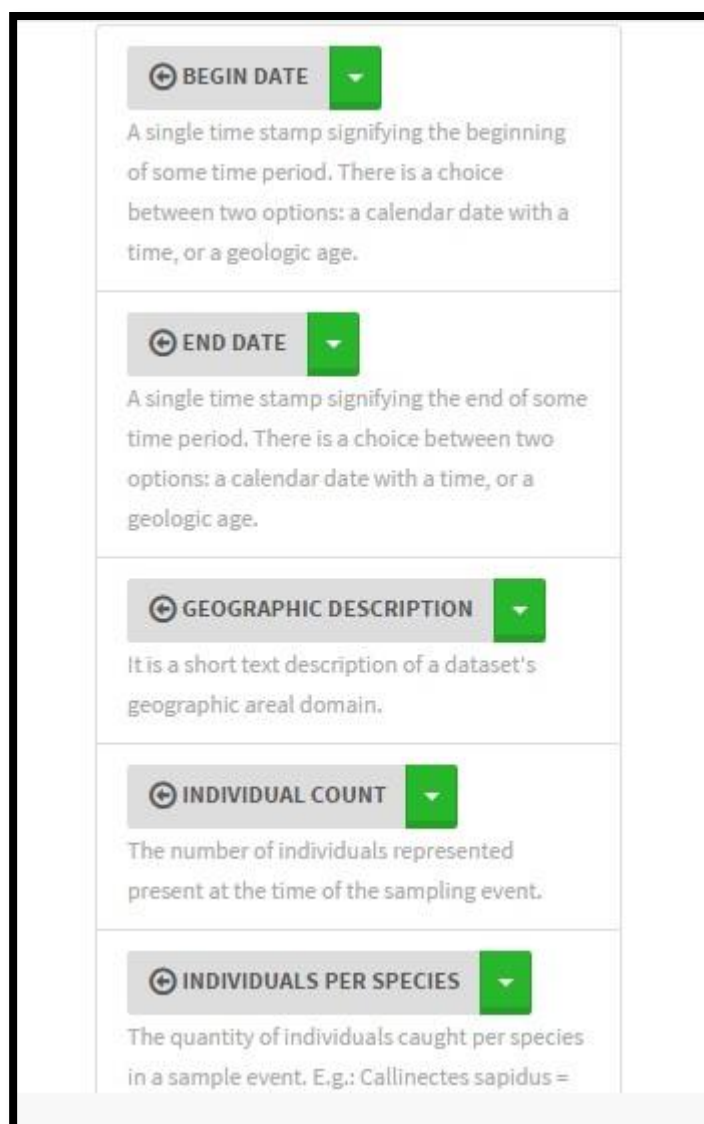


Figura 17 – Apresentação dos descritores na ontologia “Biological Oceanography” da plataforma Dendro

4.4 Depósito de dados na plataforma e avaliação do perfil de aplicação

Para além de construir uma ferramenta de curadoria digital com intuito de apoiar a gestão de dados de investigação também é importante avaliar sua eficácia num contexto de ordem prática. Ou seja, testar a aplicabilidade da ferramenta dentro de uma plataforma para a tarefa de anotação de dados de investigação.

De forma a promover a plataforma Dendro como uma ferramenta adequada à gestão de dados de investigação, sobretudo no que se refere à questão da descrição de dados, bem como para avaliar o seu desempenho nesta tarefa, uma série de experiências vem sendo realizada com investigadores da Universidade do Porto. Participam da campanha de avaliação da plataforma investigadores das áreas em que foram construídas ontologias para descrição de dados. Estes foram convidados a depositar, pelo menos, um conjunto de dados e um artigo publicado e proceder à sua descrição.

Para as experiências no Dendro optou-se por convidar investigadores que não colaboraram com a elaboração dos conjuntos de descritores, para não enviesar a avaliação. Além do mais, sendo os vocabulários construídos com base em experiências específicas de investigação, a opinião de investigadores diferentes pode fornecer subsídios para apreciação da qualidade e abrangência dos descritores.

A campanha de avaliação se realiza em três sessões, sendo que na primeira um investigador de cada área faz o depósito e descrição de seus recursos numa primeira versão da plataforma, que não usa um sistema de recomendação de descritores. Na segunda sessão outro investigador da mesma área procede ao depósito e descrição de seus recursos com o sistema de recomendação de descritores ativado na plataforma. A ideia é que a plataforma aprenda com a interação de seus utilizadores e passe a recomendar descritores para facilitar a tarefa de anotação. Uma terceira sessão será realizada com um terceiro investigador de cada área que fará uma avaliação das descrições feitas anteriormente.

Foi elaborado um guião (Anexo H) para as duas primeiras sessões para garantir que as mesmas sejam conduzidas de maneira uniforme. Como os investigadores convidados a participar da experiência não possuem conhecimento da plataforma nem tampouco da atividade de descrição, as sessões se iniciam com uma explicação do funcionamento do Dendro, suportado pelo Guião de Utilização do Dendro (Anexo G), e uma explicação da tarefa de descrição e do conceito de descritor. Antes do investigador ser de fato convidado a iniciar sua interação com a plataforma lhe é demonstrado um exemplo de descrição.

Para cada sessão foram designados dois elementos do InfoLab que, em adição às atividades de apresentação e demonstração acima explicadas, tinham a incumbência de cronometrar uma atividade de descrição, gravar o áudio da experiência e anotar qualquer dúvida ou aspecto que chamasse a atenção.

A colaboração deste estudo não se restringiu à elaboração do perfil de aplicação para o domínio da Oceanografia Biológica, mas se estendeu à participação da campanha de avaliação da plataforma Dendro com investigadores deste domínio. Participaram da campanha investigadores área de Ciências do Mar e Ambiente do Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)⁶⁰ da Universidade do Porto.

Apesar do objetivo geral da campanha de interações com o Dendro se centrar em testar a plataforma e seu sistema de recomendação de descritores, a experiência foi desenhada de forma a servir também os propósitos de validação das ontologias desenvolvidas. No nosso caso, a experiência serviu para inferir sobre a atividade de anotação de dados por parte dos investigadores. Também se aproveitou a ocasião para solicitar aos investigadores que respondessem a um inquérito para avaliar os descritores do perfil de aplicação “Biological Oceanography”.

O inquérito, que pode ser conferido no Anexo I, foi composto por questões sobre validade e usabilidade dos descritores do perfil. Foi encaminhado aos e-mails dos investigadores após sua interação com a plataforma Dendro, juntamente com o perfil de aplicação completo do domínio. A aplicação do inquérito se restringiu às sessões 1 e 2, uma vez que a sessão 3 está agendada para o mês de Julho de 2015, fora do âmbito deste estudo.

Apresenta-se a seguir um breve relato de cada uma das sessões da experiência, apresentando as respostas do investigador em relação ao perfil de aplicação.

Sessão 1- 08/06/2015, 17h

A experiência seguiu as orientações elaboradas para guiá-la e teve início com uma explicação da plataforma através do guião Dendro. De seguida a plataforma foi aberta e um exemplo de descrição concreto foi demonstrado. Foi explicado o conceito de descritor e demonstradas as ontologias genéricas e específicas disponíveis para anotação. A investigadora, convidada a explorar estes vocabulários, questiona se é ela quem terá que escolher entre os descritores dos vocabulários disponíveis para anotar os próprios dados. Na sessão 1 são os próprios investigadores que devem explorar os

⁶⁰ Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR) - <http://www.ciimar.up.pt/>

vocabulários disponíveis e selecionar os descritores de forma manual, pois nesta etapa o sistema de recomendação de descritores não está ativa. Foi explicado como deve ser feita a descrição, como adicionar descritores para a área de edição, salvar descritores como favoritos, de sua descrição e do projeto em que está a anotar, e gravar. Na adição de descritores favoritos ao projeto estes aparecem para colaboradores que adicionam e descrevem recursos no mesmo projeto. De seguida, foi solicitada a fazer a primeira interação com a plataforma, ao criar uma pasta com ficheiros associados. Foi solicitada a descrição da pasta e dos ficheiros. Nesta primeira descrição a investigadora contou com o nosso apoio.

A investigadora, a princípio, teve dificuldades em distinguir a descrição das pastas da descrição dos ficheiros, e demonstrou dificuldades em selecionar os descritores apropriados para a descrição de ambos. Percebeu-se que, no fundo, a tarefa de anotação se apresentou complexa, o que foi agravado pela falta de conhecimento dos vocabulários (ontologias) e respectivos descritores. Ficou ainda em dúvida na escolha dos descritores em função de seus ficheiros serem todos do mesmo assunto - pensou que os descritores seriam os mesmos para todos os ficheiros (embora o conteúdo fosse diverso – dados de coleta, artigo publicado).

Teve inicialmente um pouco de dificuldade na interação com a plataforma, como por exemplo, a navegação entre pastas e ficheiros, navegação entre e dentro das ontologias, como selecionar os descritores e carregá-los para a área de edição, como apagar descritores. Dificuldade que foi atenuada após uma maior interação com a plataforma. Percebeu-se que a plataforma pode ter uma interface mais amigável em alguns quesitos.

A investigadora não adicionou muitos descritores como favoritos durante sua tarefa de descrição, fazendo-o quando alertada para a importância e praticidade de fazê-lo. Também não demonstrou interesse em percorrer descritores das ontologias específicas de outras áreas, fixando-se nas ontologias genéricas e na ontologia de sua área.

Ao final da experiência foi solicitado um feedback da investigadora em relação às dificuldades sentidas (de linguagem, de uso da plataforma). Ela mencionou a dificuldade de navegação nas ontologias, achou que não era muito prática a forma de escolha dos descritores e sugeriu outro tipo de agrupamento dos descritores. Achou a plataforma em si relativamente fácil de usar e intuitiva, depois da demonstração.

Também considerou o Dendro uma ferramenta importante para gestão de dados. A experiência toda durou cerca de uma hora.

Em relação ao inquérito que avalia o perfil de aplicação, as respostas foram bastante positivas, demonstrando que os descritores do perfil são úteis e compatíveis com o domínio da Oceanografia Biológica. Quando questionada se usou os descritores deste vocabulário nas descrições nas interações com o Dendro ou se os usaria, a investigadora respondeu de forma positiva para as duas opções. Também afirmou que considera os descritores úteis e que os usaria para descrever seus dados se os fosse depositar em alguma plataforma.

Relativamente à compatibilidade semântica dos descritores com a linguagem utilizada no domínio, a investigadora declarou que considera o vocabulário criado semanticamente adequado à investigação no domínio. Também considera que os descritores são compatíveis com o domínio no que concerne à sua capacidade de descrever suas atividades de investigação, como por exemplo, os eventos de coleta.

Quanto à abrangência e completude do perfil de aplicação a investigadora considerou o perfil bastante completo e abrangente em relação aos aspectos de investigação do domínio da Oceanografia Biológica, não modificaria ou suprimiria qualquer um dos descritores. Porém, como sugestão, acrescentaria algo relacionado ao tamanho dos indivíduos coletados, por exemplo, “Mean size of the individual collected”.

Apresenta-se abaixo um exemplo de descrição que a investigadora fez na plataforma Dendro, selecionando manualmente os descritores dos conjuntos de metadados disponíveis na plataforma. Pode-se perceber que todos os descritores do início da anotação foram selecionados do vocabulário “Biological Oceanography”.

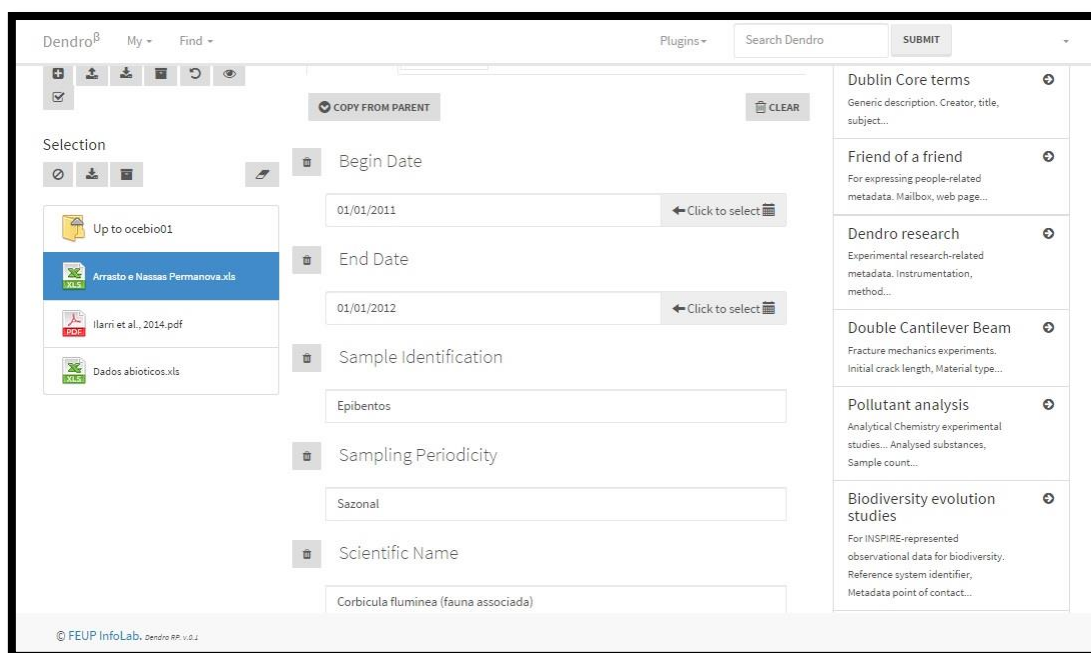


Figura 18 – Exemplo de descrição sem recomendação de descritores

Sessão 2- 22/06/2015, 16h

A experiência de interação do investigador com a plataforma de gestão de dados fez parte da segunda etapa da campanha de avaliação, nomeada de “sessão 2”, onde o investigador utiliza uma versão do Dendro habilitada com um sistema de recomendação de descritores. Ou seja, a plataforma indica para o investigador metadados compatíveis com sua área de atuação, que podem ser úteis para a atividade de anotação de seus dados.

Para dar início à experiência foi feita uma explanação sobre a plataforma Dendro, seus principais objetivos e funcionalidades com relação à descrição, preservação e partilha de dados de investigação. Também foi dada ao investigador uma noção de descrição de dados. Então, foi solicitado a este que criasse uma pasta no mesmo projeto anteriormente utilizado pelo investigador que participou da “sessão 1” da campanha de experiências com o Dendro. No caso, os dois investigadores trabalham colaborativamente no mesmo projeto. Depois de criada a pasta o investigador adicionou a ela um conjunto de dados.

A seguir foi brevemente explicado como se dá a atribuição de metadados nos níveis de anotação de pastas e ficheiros, para então mostrar os conjuntos de descritores genéricos e específicos de domínio disponíveis na plataforma. Foram apresentados um pouco mais detalhadamente os três vocabulários que contém descritores genéricos e também a interface do Dendro com modo de seleção manual de descritores, a qual o investigador foi convidado a explorar. Então, foi feita uma demonstração da plataforma com a recomendação de descritores ativa. Foi explicado como adicionar os descritores para a área de edição e as funcionalidades atribuídas a eles, como adição de favoritos, por exemplo.

O investigador inicia a tarefa de descrição tendo escolhido descrever apenas o ficheiro, não a pasta. De início, pensa que deve seleccionar e preencher todos os descritores lhe são recomendados, até ser alertado que não é necessário. Logo, passa a percorrer a lista de descritores recomendados e seleccionar somente aqueles que considera necessários para a anotação de seus dados.

O investigador percorre três páginas da recomendação e questiona se é necessário percorrer todas, pois já não encontra mais utilidade nos descritores que estão a ser recomendados. De facto, utiliza principalmente os descritores recomendados na primeira página. Percebe-se que, dos descritores recomendados, muitos dos que foram seleccionados pelo investigador pertencem ao vocabulário específico de sua área, “Biological Oceanography”.

Como os descritores recomendados já não satisfaziam sua necessidade de descrição foi aconselhado a percorrer todos os vocabulários no modo de seleção manual, para o caso de haver algum conjunto de descritores mais específico para o seu domínio.

Ao final da descrição é requerido a adicionar como favoritos aqueles descritores que considera importante e que possam ser reutilizados mais vezes. Neste momento se depara com um descritor do vocabulário “Biological Oceanography” referente à unidade temporal que considera mais apropriado para sua descrição do que o descritor que havia previamente escolhido, faz então uma edição de sua anotação, acrescentando este novo metadado e eliminando o anterior.

Para finalizar a experiência, como modo de complementar sua descrição, é convidado a fazer a descrição da pasta de forma resumida, utilizando apenas alguns

descritores que a caracterizem de forma a saber o que tem de conteúdo. Para isto pode, em vez de percorrer todos os vocabulários ou ativar o sistema de recomendação, fazer a pesquisa do descritor que quer utilizar. Apresenta um pouco de dúvida em relação à descrição do conteúdo da pasta, esta sanada pelo apoio dos curadores.

Não foram detectados quaisquer problemas de sistema. Ao final averiguou-se que as descrições ficaram todas guardadas. A experiência toda levou cerca de 30 minutos, metade do tempo que levou a experiência da sessão 1, o que se atribui ao sistema de recomendação de descritores, que escusa o investigador de ter que percorrer muitas listas de ontologias para encontrar os descritores pretendidos. O investigador relatou que não teve problemas em utilizar a plataforma e teve curiosidade em saber como funciona o sistema de recomendação.

O investigador que participou da segunda sessão de avaliação da plataforma Dendro foi também convidado a responder ao inquérito que avalia o perfil de aplicação “Biological Oceanography”. Assim como as respostas anteriores estas demonstraram a utilidade e compatibilidade dos descritores do perfil para com o domínio da Oceanografia Biológica. O investigador respondeu que usou os descritores deste vocabulário nas suas interações com o Dendro e os usaria em uma futura atividade de descrição. Também afirmou que considera os descritores úteis para a atividade de descrição de dados de investigação e produtos gerados a partir destes.

Quanto à conformidade semântica dos descritores com a linguagem utilizada no domínio e sua capacidade de descrever as atividades de investigação o investigador declarou que considera os descritores do vocabulário adequado em ambos os casos. Considerou o perfil de aplicação completo e abrangente, portanto não modificaria ou suprimiria qualquer um dos descritores. Apenas sugeriu a inclusão de opções pré-definidas em um descritor chamado “subárea”, por exemplo, Ecologia, Genética, Imunologia, Aquacultura e etc... A escolha da subárea agilizaria a tarefa de descrição.

Abaixo segue exemplo de descrição que o investigador fez na plataforma Dendro utilizando o sistema de recomendação de descritores. Pode-se perceber que a recomendação indicou descritores do perfil “Biological Oceanography”, os quais foram selecionados pelo investigador.

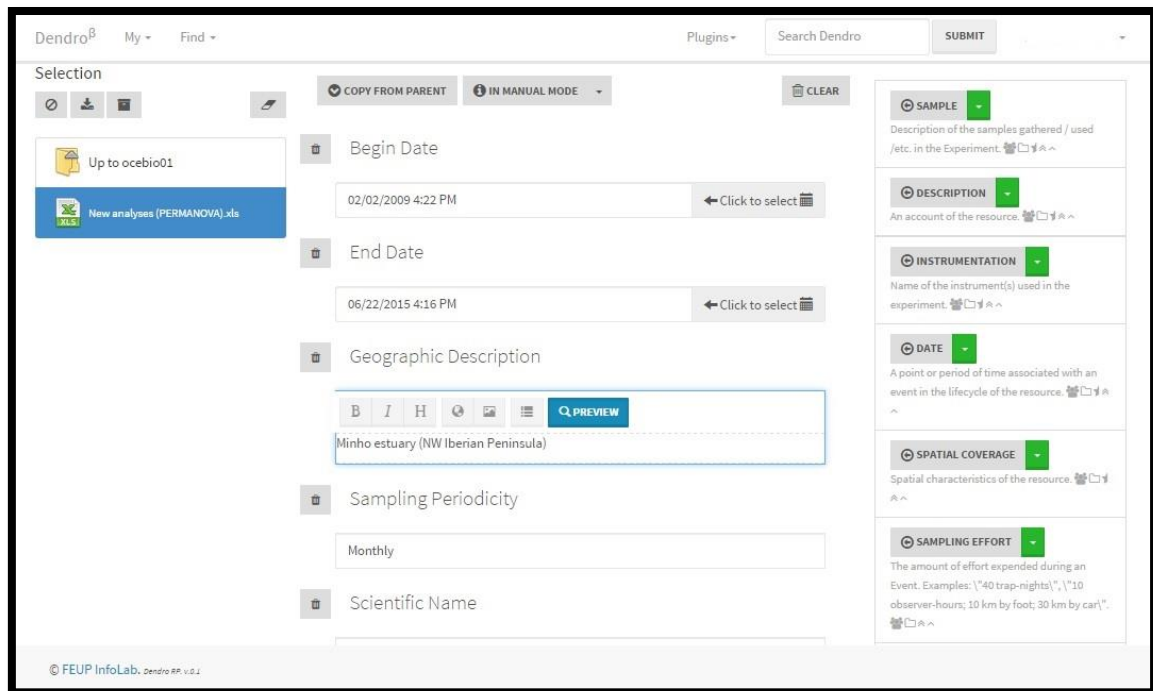


Figura 19 – Exemplo de descrição com recomendação de descritores

A qualidade das descrições feitas pelos investigadores não foi avaliada nas duas etapas de experimentação do Dendro, no entanto, notou-se um empenho de sua parte para descrever os dados da maneira mais correta possível. As perguntas dos investigadores acerca de padrões para descrição, como a língua ou unidades de medida a serem usadas, deixaram este fato evidente. Em nenhum dos casos houve uma descrição fictícia.

Em ambas as sessões foram usados descritores que compõem o perfil de aplicação “Biological Oceanography”, os quais estão distribuídos nas diversas ontologias da plataforma Dendro, e não apenas os descritores da ontologia “Biological Oceanography”.

Conclusões e perspectivas futuras

Os dados de investigação já têm sua importância consolidada no atual contexto científico, porém, percebe-se que a questão da gestão de dados de investigação ainda é bastante incipiente. Embora existam organizações com iniciativas proeminentes no desenvolvimento e fornecimento de ferramentas de curadoria digital, políticas de gestão, programas de treinamento e educação, a prática da gestão de dados como uma tarefa planeada que segue determinados padrões ainda não é comum entre grupos de investigação.

Tal realidade retrata o panorama de grupos de investigação da Universidade Federal do Rio Grande, situada no sul do Brasil, instituição na qual se realizou a primeira fase deste trabalho, e também de grupos de investigação da Universidade do Porto, onde foi realizada a segunda parte do trabalho. Nenhum destes grupos possuía planos de gestão de dados, da mesma forma estratégias de preservação e partilha se apresentaram praticamente inexistentes, feitas apenas em casos particulares, sem obedecer a critérios específicos. Tampouco haviam tido contato com ferramentas de gestão, que não as ferramentas tradicionais para armazenamento de dados.

Este cenário é compreensível se pensarmos na quantidade de trabalho que os investigadores já possuem para gerar e analisar dados, produzir, validar e comunicar resultados (a tarefa de gerir a literatura científica nem sempre é fácil). Para além disso, ainda necessitam se envolver com questões burocráticas dos projetos e das agências de financiamento. Para não citar quando o investigador tem sob sua responsabilidade a tarefa de gestão de pessoas ou bens materiais. Isto, por um lado, justifica a falta de iniciativa dos investigadores para as boas práticas de gestão, e por outro, prova que as iniciativas devem advir de gestores e curadores de dados.

Neste sentido este estudo pretendeu fornecer apoio à gestão de dados de investigação a pequenos grupos de investigadores, buscando abranger várias fases do ciclo de vida dos dados, logo depois de sua criação. Particularmente, se preocupou com questões de documentação, armazenamento e possibilidade de partilha dos mesmos.

Para resolver a questão da documentação de dados, que é essencial para que estes sejam inteligíveis, partilhados entre colaboradores e preservados ao longo do tempo, foi desenhado um perfil de aplicação para atividades de investigação no domínio

da Oceanografia Biológica, cujos descritores foram formalizados em ontologia. Esta ontologia foi introduzida numa plataforma de gestão de dados de forma a atuar como modelo de metadados a descrever este domínio. Na plataforma, portanto, conjuntos de dados podem ser depositados, descritos, e se for do interesse dos investigadores, partilhados com parceiros ou copiados para repositórios externos, o que vem a colaborar na solução de questões de preservação e partilha. Desta forma, os objetivos principais deste estudo foram atingidos.

Um perfil de aplicação pode ser desenhado por curadores com a colaboração de investigadores do domínio, ou sem esta colaboração, desde que o curador tenha algum conhecimento sobre ele. No entanto, a experiência de elaboração do perfil com a colaboração dos investigadores mostrou-se valiosa, pelo que se recomenda. É importante que eles sejam vistos como atores do processo, e não apenas como utilizadores da ferramenta.

O envolvimento dos investigadores é fundamental para um entendimento consistente do domínio, o que evita erros que podem afetar a linguagem do perfil de aplicação. Por este motivo a disponibilidade de contato com investigadores e a sua compreensão sobre a importância do processo é importante. Entender e utilizar a mesma linguagem do investigador no perfil de aplicação é outro fator a ser destacado, uma vez que esta não pode ser uma barreira para a descrição de dados mas, pelo contrário, deve ser facilitadora do processo de descrição.

Em relação à tarefa de anotação dos dados esta tem que ser o mais simples possível para o investigador, uma vez que suas atividades de investigação já lhe demandam um esforço considerável. A tarefa adicional de descrever dados pode ser um peso extra que o investigador tem que carregar. Somado a isso, o desconhecimento e inexperiência em relação à atividade de anotar pode ser um impedimento para que a realizem. Portanto, todos os recursos para facilitar a tarefa de descrição devem estar disponíveis ao utilizador, desde a criação de metadados compreensíveis até o emprego de tecnologia inteligente em repositórios ou plataformas de gestão de dados.

Num cenário ideal existiria um gestor de dados junto a cada grupo de investigação para fazer o tratamento desta informação tão logo ela fosse gerada. Esta tarefa poderia ficar a cargo de profissionais da informação, que têm conhecimento e experiência em representação e organização do conhecimento, assim como em gestão da informação. No entanto, este cenário está muito longe de se tornar realidade. O que

seria verdadeiramente ideal hoje, é que estes profissionais tomassem para si a responsabilidade de promover as boas práticas de gestão de dados junto aos investigadores e fornecessem a eles as ferramentas necessárias para praticá-las. A própria Revista Nature, em editorial de 2009, sugere que as bibliotecas das universidades tomem para si a responsabilidade da preservação e acessibilidade aos dados de investigação, através da criação de repositórios equivalentes a bibliotecas digitais. Junto destes há que se considerar a adição de um componente educacional.

Por outro lado, agências de financiamento também têm que reconhecer que faz parte de suas atribuições fornecer suporte para que investigadores preservem e partilhem seus dados. Para citar um exemplo, pouco se sabe sobre iniciativas tomadas por agências financiadoras de investigação científica no Brasil em direção à implantação de práticas uniformizadas de gestão de dados, ao desenvolvimento de políticas ou fornecimento programas de treinamento. Não basta fornecer o apoio financeiro necessário à produção dos dados, é preciso também assegurar que estes se mantenham relevantes de modo que possam ser reutilizados ao longo do tempo. Assim, o ciclo de investimentos nas atividades de investigação é otimizado, evitando repetições nos processos de recolha de dados.

Idealmente, ainda, o mesmo editorial da Nature acima mencionado sugere que universidades e cursos de pós-graduação disponibilizem em seus currículos uma disciplina de gestão da informação, para ensinar os alunos a gerir seus projetos e dados de investigação. De fato, é imperativo sensibilizar os investigadores para a elaboração de planos de gestão de dados, para a importância de documentar, preservar e partilhar dados de investigação. Este é um compromisso que deve ser partilhado pelas várias instâncias aqui nomeadas, porém deve haver um esforço dos próprios investigadores para que seja instalada a cultura do depósito, descrição e partilha de dados.

Neste momento, em muitos grupos de investigação, ainda prevalece uma visão de que os dados, mesmo que gerados com recursos públicos, pertencem a eles, portanto, é sua a decisão do que fazer com os dados. São iniciativas, como a que está a decorrer na FEUP, que contribuem para uma mudança de paradigma. A disponibilização aos investigadores de uma plataforma onde possam, de forma segura e fácil, depositar e documentar seus dados, pode ser o primeiro passo para que os mesmos sejam partilhados com repositórios externos. A campanha de interações dos investigadores

com a plataforma Dendro a decorrer está a ser positiva neste sentido, mostrando abertura e disponibilidade dos investigadores que dela participaram.

A experiência de elaboração de um perfil de aplicação para integrar a lista de conjuntos de descritores da plataforma Dendro foi muito satisfatória, uma vez que sua utilização para descrição de dados foi efetiva. Embora tenha sido elaborado para atividades de investigação específicas do domínio da Oceanografia Biológica no Brasil, na campanha de interação com o Dendro os descritores do perfil foram selecionados por investigadores de área correlata na Universidade do Porto, o que demonstra sua utilidade em situações práticas de descrição. Estes investigadores ainda avaliaram o perfil como completo e abrangente. Durante a experimentação do Dendro notou-se ainda que alguns descritores foram usados por investigadores de outras áreas, o que reforça a contribuição do perfil de aplicação para com uma plataforma que funciona de forma colaborativa e multidisciplinar.

A abordagem metodológica utilizada para elaborar o perfil de aplicação “Biological Oceanography” é flexível e pode ser repetida para construção de perfis em outras áreas do conhecimento. As experiências do InfoLab demonstraram que isso é factível, e que as descrições podem ser feitas em qualquer domínio, que trabalhe com qualquer tipo de dados. Os mapas de conceito se mostraram muito úteis para a modelagem do conhecimento nos domínios, tanto para o curador quanto para os investigadores. E a ontologia é uma ferramenta eficaz para formalização dos descritores e seu funcionamento dentro da plataforma de gestão de dados.

Como perspectiva de trabalho futuro é importante utilizar as experiências realizadas na plataforma Dendro para refinar o perfil de aplicação “Biological Oceanography”. Ainda, para um perfil mais completo se deve continuar o levantamento de requisitos em outras subáreas da Oceanografia Biológica, de modo a se conseguir uma maior amplitude na descrição do domínio.

Em relação ao restante dos descritores da plataforma também seria interessante haver uma parametrização para verificação de equivalência semântica entre os mesmos. Por exemplo, dentro de um mesmo vocabulário se tem o descritor “coverage” para se referir à cobertura espacial ou temporal de um recurso, e também os descritores mais específicos “spatial coverage” e “temporal coverage”, que se referem aos mesmos tópicos. Ainda há os descritores “surname” e “last name”, que servem para o mesmo propósito de descrição. Um controle semântico de descritores poderia reduzir a lista dos

mesmos, tornando-as mais consistentes e padronizadas e reduzindo o tempo de seleção de descritores quando houvesse a necessidade de percorrer os vocabulários.

Quanto às descrições feitas pelos investigadores, neste momento estas são feitas em linguagem natural e, embora os descritores apresentem exemplos de descrição, obedecem apenas a critérios subjetivos. Também não existe nenhuma regra quanto ao uso de repetição de metadados, por vezes estes são repetidos para conter informação diferente, por outras toda a informação é colocada dentro de um mesmo campo. Portanto, poderiam ser elaboradas algumas normas para padronizar a utilização da plataforma, assim como se poderia ponderar a utilização de thesauri ou vocabulários controlados. Tais aspectos poderiam ter um impacto positivo na recuperação da informação.

Por fim, na FURG deve se ampliar a discussão sobre a gestão de dados de investigação para outros domínios, de modo a conscientizar os investigadores da importância da preservação e partilha destes. Este poderia ser o embrião de um empreendimento para a implantação e uso de uma plataforma ou repositório de dados, tal como já existe o Repositório Institucional.

Referências

ABBOTT, Daisy. **Annotation**. DCC Briefing Papers: Introduction to Curation. Digital Curation Centre: Edinburgh, 2008. Disponível em:

<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/annotation>>

Acesso em: 29 Dez. 2014.

ABBOTT, Daisy. **What is digital curation?** DCC Briefing Papers: Introduction to Curation. Digital Curation Centre: Edinburgh, 2008. Disponível em:

<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>> Acesso em: 29 Dez. 2014.

ALMEIDA, Maurício Barcellos. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.19, n.3, p. 242-258, Jul./Set. 2014. Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1736>> Acesso em: 18 Nov. 2014.

ALMEIDA, Maurício B.; BAX, Marcello P. Visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003.

ARL. **Association Of Research Libraries**: empowering the research library community. Disponível em: <http://www.arl.org/focus-areas/e-research#.VKbwXiusWSo> > Acesso em: 09 Dez. 2014.

AUSTRALIAN NATIONAL DATA SERVICES. **ANDS Guide**: Metadata - awareness level. Disponível em: <http://ands.org.au/guides/metadata-awareness.html>> Acesso em: 13 Nov. 2014.

AUSTRALIAN NATIONAL DATA SERVICES. **ANDS Guide**: What is research data? Disponível em: <http://ands.org.au/guides/what-is-research-data.html>> Acesso em: 13 Nov. 2014.

BOHLE, Shannon. **What is e-science and how should it be managed?** Nature.com, Spektrum der Wissenschaft (Scientific American). 2013. Disponível em:

http://www.scilogs.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed/> Acesso em: 28 Dez. 2014.

CASTELLS, Manuel. A sociedade em rede: do conhecimento à política. In: CASTELLS, Manuel; CARDOSO, Gustavo (orgs.). **A sociedade em rede**: do conhecimento à acção política. Lisboa: Imprensa Nacional – Casa da Moeda, 2006. p. 17-30. (Coleção Debates: Presidência da República). Conferência promovida pelo Presidente da República, que decorreu a 4 e 5 de Março de 2005, no Centro Cultural de Belém.

CASTRO, João Aguiar; RIBEIRO, Cristina e SILVA, João Rocha da. **Designing an Application Profile using qualified Dublin Core**: a case study with fracture mechanics datasets. Proc. Int'l Conf. on Dublin Core and Metadata Applications 2013,

Lisboa, 2 a 6 de Setembro de 2013. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3685> Acesso em: 30 Set. 2014.

CASTRO, João Aguiar; SILVA, João Rocha da e RIBEIRO, Cristina. **Creating lightweight ontologies for dataset description: practical applications in a cross-domain research data management workflow**. Joint Conference on Digital Libraries, Londres, 8 a 12 de Setembro de 2014.

CORTI, Louise; VAN DEN EYNDEN, Veerle; BISHOP, Libby; WOOLLARD, Matthew. **Managing and sharing research data: a guide to good practice**. Los Angeles: SAGE, 2014. ISBN: 978-1-4462-6726-4.

COSTA, Maíra Murrieta e CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, vol. 19, n. 3, Jul./Set. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362014000300010&lng=pt&nrm=iso&tlng=en> Acesso em: 26 Mar. 2015.

CUNHA, Murilo Bastos da. **A Biblioteca Universitária e a gestão de dados científicos**. 49 slides. Disponível em: <https://www.bu.ufmg.br/snbu2014/wp-content/uploads/2014/12/PDF_Murilo-Bastos.pdf> Acesso em: 03 Dez. 2014.

NATURE EDITOR. Data's shameful neglect. *Nature*, v. 461, n. 7261, 10 Set. 2009, p. 145. Disponível em: <<http://www.nature.com/nature/journal/v461/n7261/full/461145a.html>> Acesso em: 07 Jun. 2015.

DENDRO. Disponível em: <<http://dendro.fe.up.pt/>> Acesso em: 15 Nov. 2014.

DUBLIN CORE METADATA INIATIVE. **Metadata basics**. Disponível em: <<http://dublincore.org/metadata-basics/>> Acesso em: 10 Jan. 2015.

GORE, Sally. eScience and data management resources on the web. **Medical Reference Services Quarterly**, 2011, v. 30, n. 2, p. 167-177.

FIOLEK, A. *et. al.* **Mapping, cross-walking, converting and exchanging Oceanographic metadata information in Video Data Management System**. In: Conference OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges. Disponível em: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5422072&tag=1 Acesso em: 12 Jan. 2015.

IEEE International Conference on eScience: a website for the conferences series. Disponível em: <<https://escience-conference.org/>> Acesso em: 28 Dez. 2014.

INSTITUTE OF OCEANOLOGY “FRIDTJOF NANSEN”. Disponível em: <http://www.io-bas.bg/index_en.html> Acesso em: 12 Abr. 2015.

INSTITUTO DE CIENCIAS MARINAS DE ANDALUCIA – ICMAN. Disponível em: <<http://www.icman.csic.es/>> Acesso em: 12 Abr. 2015.

INSTITUTO PORTUGUÊS DE QUALIDADE. Norma Portuguesa NP3715 – 1989.

HEERY, Rachel e PATEL, Manjula. Application profiles: mixing and matching metadata schemas. *Ariadne*, v. 25, Set. 2000. Disponível em: <http://www.ariadne.ac.uk/issue25/app-profiles> Acesso em: 22 Out. 2014.

HEVNER, Alan R. 2004. Design Science in Information Systems Research. *MIS Quarterly*, v. 28, n. 1, Mar. 2004, p. 75-105.

HODSON, Simon. **Meeting the research data challenge**: managing research data programme 2009-11. JISC briefing paper, 2009. Disponível em: <<http://webarchive.nationalarchives.gov.uk/20140702233839/http://jisc.ac.uk/publications/briefingpapers/2009/bpresearchdatachallenge.aspx#downloads>> Acesso em: 05 Dez. 2014.

LORD, Philip *et al.* From data deluge to data curation. In: Proceedings 3th UK e-Science All Hands Meeting, 2004. Disponível em: <http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf> Acesso em: 28 Out. 2014.

LORD, Philip and MACDONALD, Alison. eScience data curation. JISC: 2004. Disponível em: <<http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/publications/esciencedatacuration.pdf>> Acesso em: 19 Nov. 2014.

MANTRA – Research Data Management Training. Research data explaining. Edimburgo: University of Edinburgh, 2014. Disponível em: <http://datalib.edina.ac.uk/mantra/researchdataexplained/> > Acesso em: 16 Out. 2014.

NATIONAL INFORMATION STANDARDS ORGANIZATION. Understanding metadata. Bethesda, USA: NISO Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>> Acesso em: 17 Nov. 2014.

NORTH CAROLINA STATE UNIVERSITY. Department of Marine, Earth And Atmospheric Sciences. Disponível em: < <http://www.meas.ncsu.edu/>> Acesso em: 12 Abr. 2015.

OCEAN AND EARTH SCIENCE, NATIONAL OCEANOGRAPHY CENTRE SOUTHAMPTON – NOCS. Disponível em: <<http://www.southampton.ac.uk/oes/research/index.page>> Acesso em: 20 Mai. 2015.

ODUM, Eugene P. Fundamentos de Ecologia. – 5. ed. – Lisboa: Fundação Calouste Gulbekian, 1997.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD Principles and guidelines for access to research data from public funding**. Paris: OECD, 2007. Disponível em: <<http://www.oecd.org/dataoecd/9/61/38500813.pdf>> Acesso em: 10 Dez. 2014.

PENNOCK, Maureen. **Curating eScience data**. DCC Briefing Papers: Introduction to Curation. Digital Curation Centre: Edinburgh, 2006. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/curating-e-science-data>> Acesso em: 29 Dez. 2014.

PICKLER, Maria Elisa Velentim. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.12, n.1, p. 65-83, Jan./Abr. 2007. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/251>> Acesso em: 24 Set. 2014.

RODRIGUES, Eloy *et al.* **Os Repositórios de Dados Científicos**: estado da arte. Projecto RCAAP D24 – Relatório, 2010. Disponível em: <<http://hdl.handle.net/1822/10830>> Acesso em: 24 Out. 2014.

SAWANT, Sarika. Transformation of the scholarly communication cycle. **Library Hi Tech News**, New Jersey, v. 29, n. 10, p. 21-24, 2012. Disponível em: <<http://www.emeraldinsight.com/doi/full/10.1108/07419051211294482>> Acesso em: 01 Dez. 2014.

SCHOWCHOW, Rosane Quaresma; ZANBONI, Ademilson J. O Estuário da Lagoa dos Patos: um exemplo para o ensino de ecologia no nível médio. **Cadernos de Ecologia Aquática**, Rio Grande, v. 2, n. 2, p. 13-27. Ago./Dez 2007. Disponível em: <http://www.cadernos.ecologia.furg.br/images/artigos/04_Rosane.pdf> Acesso em: 30 Mar. 2015.

SCRIPPS INSTITUTION OF OCEANOGRAPHY. Disponível em: <<https://scripps.ucsd.edu/>> Acesso em: 12 Abr. 2015.

SHARMA, Sugam *et al.* A brief review on leading big data models. **Data Science Journal**, v. 13, n. 4, p. 138-157, Dez. 2014. Disponível em: <https://www.jstage.jst.go.jp/article/dsj/13/0/13_14-041/_article> Acesso em: 07 Jan. 2015.

SILVA, João Rocha da *et al.* **Beyond INSPIRE**: an ontology for biodiversity metadata records. 10th International Workshop on Ontology Content, 20 e 30 Out./ 2014, Amantea, Itália. Disponível em: <<http://dendro.fe.up.pt/#publications>> Acesso em: 11 Dez 2014.

SILVA, João Rocha da *et al.* **Dendro**: Collaborative Research Data Management Built on Linked Open Data. In: The Semantic Web: ESWC 2014 Satellite Events. Série Lecture Notes in Computer Science. Creta: Springer, 2014. Disponível em: http://link.springer.com/chapter/10.1007/978-3-319-11955-7_71 Acesso em: 01 de Outubro de 2014.

TAYLOR, John. Definig eScience. National eScience Centre. Disponível em: <<http://www.nesc.ac.uk/nesc/define.html>> Acesso em: 09 Dez. 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE. Disponível em: <<http://www.furg.br/>> Acesso em: 10 Dez. 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE. Instituto de Oceanografia. Disponível em: <<http://www.io.furg.br/>> Acesso em: 10 Dez. 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE. Instituto de Oceanografia. Programa de Pós-Graduação em Oceanografia Biológica. Disponível em: <<http://www.ocbio.furg.br/>> Acesso em: 10 Dez. 2014.

UNIVERSIDADE FEDERAL DO RIO GRANDE. Projeto Pedagógico Institucional. Disponível em: <<http://www4.furg.br/paginaFURG/arquivos/menu/000000292.pdf>> Acesso em: 12 Mar. 2015.

UNIVERSITY OF BRITISH COLUMBIA, Department of Earth, Ocean & Atmospheric Sciences. Disponível em: <<http://www.eos.ubc.ca/research/>> Acesso em: 20 Abr. 2015.

UNIVERSITY OF LIVERPOOL. Department of Earth, Ocean and Ecological Sciences. Disponível em: <<http://www.liv.ac.uk/environmental-sciences/>> Acesso em: 20 Mai. 2015.

UNIVERSITY OF OTAGO. Department of Marine Science. Disponível em: <<http://www.otago.ac.nz/marinescience/index.html>> Acesso em: 20 Mai. 2015.

UNIVERSITY OF WESTERN AUSTRALIA. Oceans Institute. Disponível em: <<http://www.oceans.uwa.edu.au/>> Acesso em: 13 Abr. 2015.

W3C. **Web Semântica**. Disponível em: <<http://www.w3.org/standards/semanticweb/>> Acesso em: 11 Jan. 2015.

W3C **Wiki. Good Ontologies**. Disponível em: <http://www.w3.org/wiki/Good_Ontologies> Acesso em: 04 Mai. 2015

WERTHEIN, Jorge. A sociedade da informação e seus desafios. **Ciência da Informação**, Brasília, v.29, n. 2, p. 71-77, Maio./Ago. 2000. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/254/1705>> Acesso em: 27 Dez. 2013.

WHITMIRE, Amanda L.. (2013). Thoughts on “eResearch: a scientist’s perspective. **Journal of eScience Librarianship**, v. 2, n. 2. p. 68-72. Disponível em: <<http://escholarship.umassmed.edu/jeslib/vol2/iss2/3/>> Acesso em: 09 Dez. 2014.

WOODS HOLE OCEANOGRAPHIC INSTITUTION. Disponível em: <<http://www.whoi.edu/>> Acesso em: 12 Abr. 2015.

Anexos

ANEXO A – Artigo “Seasonal variability on the structure of sublittoral macrozoobenthic association in the Patos Lagoon estuary, southern Brazil (2007) / Colling, Bemvenuti e Gandra”

Seasonal variability on the structure of sublittoral macrozoobenthic association in the Patos Lagoon estuary, southern Brazil

Leonir A. Colling, Carlos E. Bemvenuti & Michel S. Gandra

Laboratório de Ecologia de Invertebrados Bentônicos, Departamento de Oceanografia, FURG. Av. Itália Km 8, Caixa Postal 474, 96203-000 Rio Grande, RS, Brasil. (andre_colling@yahoo.com.br; docbemve@furg.br)

ABSTRACT. The aim of this study is to analyze and relate the spatial-temporal variability of macrozoobenthic assemblages to bottom characteristics and salinity fluctuations, in an estuarine shallow water region of Patos Lagoon. Monthly samples, between September 2002 and August 2003, were taken on six sampling stations (distant 90 m). Three biological samples with a 10 cm diameter corer, one sample for sediment analysis, fortnightly bottom topography measurements, and daily data of temperature and salinity were taken from each station. Two biotic and environmental conditions were identified: the first corresponding to spring and summer months, with low macrozoobenthos densities, low values of salinity, small variations in bottom topographic level and weak hydrodynamic activity. A second situation occurred in the months of fall and winter, which showed increased salinity, hydrodynamics and macrobenthos organisms. These results which contrast with previous studies carried out in the area, were attributed to failure in macrozoobenthos recruitments during summer period, especially of the bivalve *Erodona mactroides* Bosc, 1802 and the tanaid *Kalliapseudes schubartii* Mañe-Garzón, 1949. This results showed that recruitments of dominant species were influenced by salinity and hydrodynamic conditions.

KEYWORDS. Seasonal variability, benthos, estuary, abiotic variables, sublittoral habitats.

RESUMO. Variabilidade sazonal na estrutura da associação de macroinvertebrados bentônicos em uma enseada estuarina da Lagoa dos Patos, sul do Brasil. Objetiva-se analisar e relacionar a variabilidade espaço-temporal de uma associação macrozoobentônica com as características do substrato e variações da salinidade, numa enseada da região estuarina da Lagoa dos Patos. Amostragens mensais entre setembro de 2002 e agosto de 2003 foram realizadas em seis pontos de coleta distantes 90 m entre si. Em cada ponto foram tomadas três amostras biológicas com tubo extrator de 10 cm de diâmetro, uma amostra para análise do sedimento, medidas quinzenais da topografia do fundo e dados diários de temperatura e salinidade. Foram identificadas duas situações ambientais e bióticas bem definidas: uma correspondente aos meses de primavera e verão com baixas densidades do macrozoobentos, baixos valores de salinidade e pouca variação no nível do substrato, refletindo neste caso uma menor atividade hidrodinâmica. A outra situação ocorreu nos meses de outono e inverno, que mostrou uma situação oposta em relação às variáveis bióticas e abióticas. Esses resultados, que contrariam trabalhos anteriores efetuados na região, foram atribuídos a falhas nos recrutamentos do macrozoobentos durante o período de verão, especialmente do bivalve *Erodona mactroides* Bosc, 1802 e do tanaidáceo *Kalliapseudes schubartii* Mañe-Garzón, 1949. Verificou-se que os recrutamentos das espécies dominantes foram influenciados pelas condições de salinidade e pela hidrodinâmica.

PALAVRAS-CHAVE. Variabilidade sazonal, bentos, estuário, variáveis abióticas, habitats sublitorais.

The benthic macroinvertebrate assemblages are structured by many species of molluscs, polychaetes and crustaceans, among other organisms with more than 1 mm size, which show direct relation to the bottom conditions, resulting in an uniformity in lifestyle, despite their distinct phylogenetic origins (DAY *et al.*, 1989). These organisms have a key role in the estuary feeding web, acting as a link between the detritus deposited on the bottom and the higher trophic levels in the system, so contributing to an important resource for larger consumers, such as birds, decapods crustaceans and fishes (BEMVENUTI, 1997c). The structure of these assemblages includes attributes as specific composition, distribution, abundance, biomass, trophic relations and diversity of the organisms (LEVINTON, 1995; PEREIRA & SOARES-GOMES, 2002).

The analysis of the structure of benthic macroinvertebrates assemblages, from their attributes, has been useful in diagnostic studies and environmental monitoring (WARWICK, 1986). However, problems may arise from the interpretation of the data acquired during these studies, concerning the distinction whether they are related to pollution or the result of natural environment variability (CLARKE & WARWICK, 1994). In this latter case,

previous studies are important in order to accomplish sufficient temporal survey of the macrofauna associations and the natural variations on the environment parameters, along the water column and substrate (WEISBERG *et al.*, 1997).

Spatial and temporal variability have been the object of studies in many estuaries (YSEBAERT *et al.*, 2003; HOLLAND, 1985; NETTO & LANA, 1994; QUIJÓN & JARAMILLO, 1993; BILES *et al.*, 2003). In the Patos Lagoon estuarine area, different studies have been conducted about benthic communities in soft bottoms, such as the macrobenthic assemblage characterization (CAPITOLI *et al.*, 1978), temporal variability in the infralittoral and channel area (BEMVENUTI *et al.*, 1978), biological relationships (BEMVENUTI, 1987; 1988; 1994) and comparisons to other estuarine environments of southern Brazil (ROSA-FILHO & BEMVENUTI, 1998). However, there is a lack of studies on the effects of dynamic processes in the water column and substrate and the consequences on distribution and abundance of benthic macrofauna in the Patos Lagoon estuarine area. The present work aimed to analyze the spatial-temporal distribution of benthic macrofauna, in relation to the bottom characteristics and salinity variations along an annual period in that lagoon.

MATERIAL AND METHODS

Fieldwork was performed along a transect with a length of about 450 m in the Saco do Arraial inlet, in a shallow plain in front of the eastern margin of Pombas Island (32°01'S, 52°07'W) in the Patos Lagoon estuarine area (Fig. 1). Six stations 90 m apart from each other were sampled monthly from September 2002 to August 2003. Three biological samples were taken from each station, with a 10 cm diameter corer, pushed 20 cm into the bottom. In each station, a stratified sample (0-5 and 5-10 cm) of the sediment was also collected with same corer (10 cm diameter), and granulometric data were obtained through sieving and pipette analysis (SUGUIO, 1973). Biological samples were sieved in the field, using a 0.3 mm mesh size, fixed with formaldehyde 4% and stained with Bengal Rose. The macroinvertebrates were separated from the sediment matter to the lower possible taxon and preserved in ethanol 70%, with the aid of a stereomicroscope in the laboratory. Daily water temperature and salinity data were obtained from PELD - Programas Ecológicos de Longa Duração, Site 8 - CNPq, FURG - databank.

Aiming to evaluate the hydrodynamic effects on substrate erosion and/or accretion, and its influences upon the benthic community, bottom topography measurements were done fortnightly. Levelled references were fixed on every station (6) along the 450 m profile, where four bottom level measurements were also fortnightly taken.

The dominant species collected were measured with the help of a stereomicroscope (0.5 mm precision). *Erodona mactroides* Bosc, 1802 specimens until 1 mm length and *Kalliapseudes schubartii* Mañe-Garzón, 1949 until 3 mm length were classified as recruits.

Analysis of variance (One-way ANOVA, $\alpha = 0.05$) and Tukey's contrast test was applied to evaluate the temporal variability of fine sediments (silt + clay).

The Shannon-Wiener diversity index (H'), which

integrates the number of species and its abundance in the association, and Pielou's evenness index (CLARK, 1997), were computed and submitted to analysis of variance (Two-way ANOVA, $\alpha = 0.05$) (local x time), like the statistical differences of macrofauna densities too. The biological data was transformed ($\log X + 1$) (UNDERWOOD, 1997) and tested for normality (Kolmogorov-Smirnoff test) and homogeneity of variances (Cochran test and standard-deviation mean plots) prior to their use in statistical tests (UNDERWOOD, 1997). The contrast test of Tukey was applied whenever significant results occurred (MAGURRAN, 1998).

A multi dimensional scaling ordination technique (MDS) was applied employing the Bray-Curtis similarity index on $\log (X + 1)$ transformed species data. A similarity analysis (ANOSIM; $\alpha = 0.05$) was performed to verify possible differences between sample stations. Simper analysis was used to verify the species contribution to similarity between groups of samples (confirmed by ANOSIM) (CLARKE & WARWICK, 1994).

RESULTS

The salinity data showed low values during the first 4 months, previous to sampling period (Fig. 2). This pattern was also observed after the beginning of macrozoobenthos sampling (September, 2002) and remained until late summer (February, 2003) (Fig. 2). In this period, the salinity data showed low values with fortnightly means not higher than 5, except in September, 2002. An increase in salinity was verified from the second fortnight of February, 2003, when it reached up to 28, as well as fortnightly means above 5 (Fig. 2). Monthly means of temperature rised gradually from the beginning of spring to the end of summer (14.3 to 24.8 °C), and fell during the months of autumn and winter (23.1 to 16.6 °C) (Fig. 2).

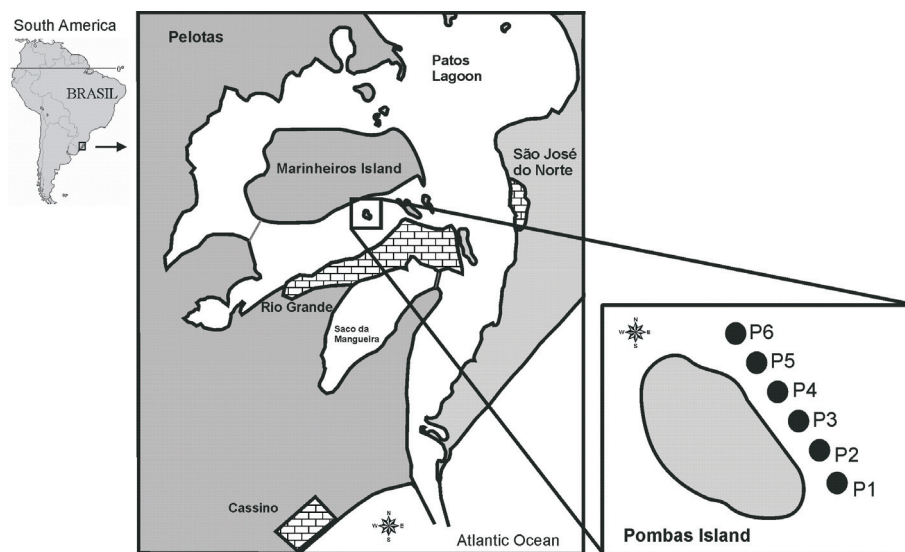


Fig. 1. Patos Lagoon estuarine region, southern Brazil, indicating the study area. P1 to P6, biological samples places. Hachured regions are inhabited zones.

Topographical measurements showed homogeneous values between spring and winter. The higher variability was verified in the level measurements in February, 2003. This trend was marked in stations #1, #2 and #3 (Figs. 3, 4) than in stations #5 and #6 (Fig. 5).

Grain size analysis showed that sediment of stations #1 (4.3%) and #2 (5.9%) had smaller percentages of mud (silt and clay, <0.064 mm), in relation to stations #3 (8.3%), #4 (11.2%), #5 (9.6%) and #6 (10.5%). Temporal variation of mud percentage showed significant differences for the winter period ($p < 0.05$), when the mud content was smaller than in the other periods.

A total of 23,808 organisms were collected, corresponding to 17 species (Tab. I). The dominance was of the bivalve *E. mactroides* (49%), the polychaetes *Nephtys fluviatilis* Monro, 1937 (12.2 %) and *Heteromastus similis* Southern, 1921 (12.8 %), the tanaid *Kalliapseudes schubartii* (15.7 %), and the isopod *Munna peterseni* Pires-Vanin, 1985 (3.7 %), that made-up 93 % of the total macrofauna. The values of diversity (H') and evenness (J') did not show significant variations ($p > 0.05$) between sampling stations, as well as between months. Benthic macrofauna showed significant differences of

density ($p < 0.05$) among seasons, when it registered very low values of density in the spring (2,885 ind.m⁻²) and in the summer (3,782 ind.m⁻²), in contrast to higher values registered in the autumn (27,028 ind.m⁻²) and in the winter (22,416 ind.m⁻²) (Fig. 6).

Most of the macrobenthic species increase abundance since autumn (Fig. 6), among them arise the dominant species *E. mactroides* and *K. schubartii* (Tab. IV), that showed strong recruitments in this period.

The Ordination Analysis (MDS) (Fig. 7) showed two major groups, composed by spring and summer samples, and other group containing autumn and winter samples. This results was confirmed by ANOSIM ($R = 0.9$; $p < 0.05$).

The juveniles of *E. mactroides* showed significant differences of density along the six stations, in February and March ($p < 0.05$), due to the small number of specimens found in stations #5 and #6. Recruitments of *E. mactroides* were always more intense in places with higher hydrodynamics, situated in front of the southern extremity of Pombas Island (stations #1 and #2), then the places with lower hydrodynamics, in an sheltered area due island protection (stations #5 and #6).

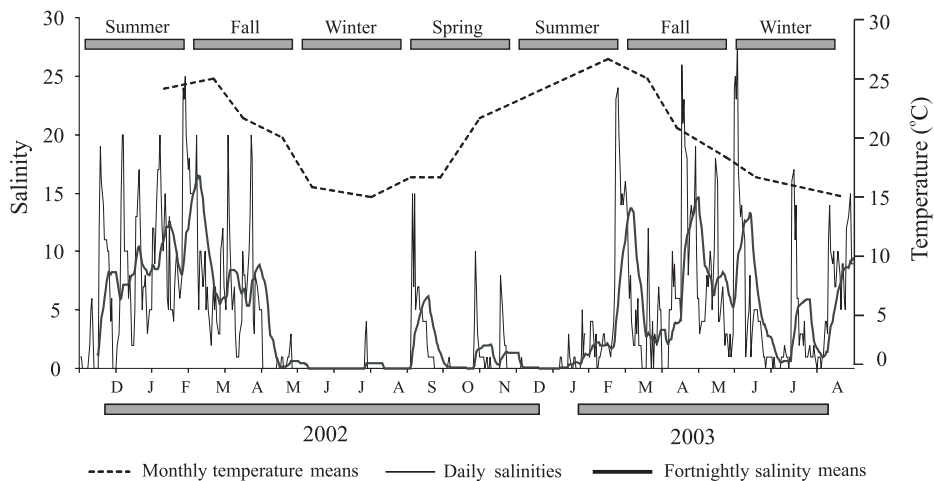


Fig. 2. Abiotic parameters (monthly temperature means, daily salinities and fortnightly salinity means) during study period on Patos Lagoon estuary, southern Brazil. December of 2001 to August of 2002 correspond to previous data, registered before the sampling period.

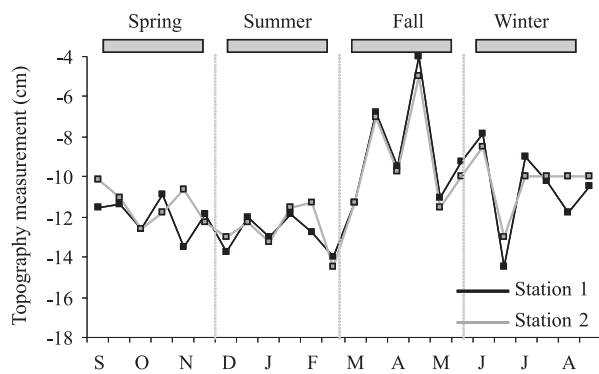


Fig. 3. Fortnightly variability (cm) on bottom topography measurements in stations 1 and 2, Patos Lagoon estuary, southern Brazil.

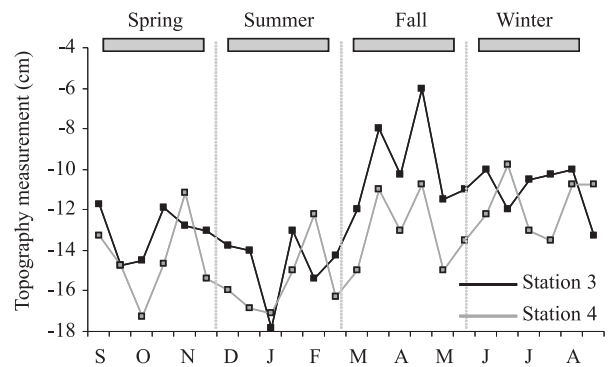


Fig. 4. Fortnightly variability (cm) on bottom topography measurements in stations 3 and 4, Patos Lagoon estuary, southern Brazil.

Tab. I. Values of mean density (mean) and standard deviation (Sd) (ind.m⁻²) from samples of macrofauna groups in the Patos Lagoon estuary, southern Brazil.

	September		Spring October		November		December		Summer January		February	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
	POLYCHAETA											
<i>Heteromastus similis</i>	1755	1019	906	388	580	196	856	454	488	371	1338	739
<i>Laonereis acuta</i>	134	207	78	89	142	144	85	116	14	41	78	89
<i>Nephtys fluviatilis</i>	913	377	425	218	672	246	616	297	856	528	856	425
CRUSTACEA												
<i>Munna peterseni</i>	234	539	57	146	85	151	7	30	42	76	28	54
<i>Diastylis sympterigiae</i>	92	150	42	76	149	236	92	150	163	209	21	66
<i>Kalliapseudes schubartii</i>	410	390	248	271	623	632	205	273	156	331	1019	780
<i>Kupellonura sp.</i>	99	183	7	30	85	218	21	90	21	49	28	93
<i>Mellita mangrovi</i>	0	0	0	0	0	0	0	0	0	0	0	0
<i>Misidopsis tortoneri</i>	7	30	0	0	7	30	0	0	0	0	0	0
<i>Pseudosphaeroma sp.</i>	0	0	14	41	28	120	21	66	7	30	0	0
<i>Sinelobus stanfordi</i>	290	332	64	126	71	133	7	30	21	49	7	30
<i>Cumacea n.d.</i>	0	0	0	0	0	0	0	0	14	41	0	0
MOLLUSCA												
<i>Erodona mactroides</i>	0	0	0	0	7	30	0	0	0	0	4140	3857
<i>Heleobia australis</i>	347	504	57	90	0	0	21	66	226	308	28	70
<i>Tagelus plebeius</i>	0	0	0	0	7	30	14	41	0	0	0	0
NEMERTINEA												
	0	0	0	0	0	0	0	0	7	30	0	0
HIRUDINEA												
	21	49	0	0	0	0	0	0	0	0	0	0

	March		Fall April		May		June		Winter July		August	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
	POLYCHAETA											
<i>Heteromastus similis</i>	4968	2356	5088	2215	3241	1279	2123	871	1762	573	2442	1098
<i>Laonereis acuta</i>	85	124	439	281	219	204	163	144	57	100	64	100
<i>Nephtys fluviatilis</i>	1897	763	4402	1357	3680	1038	4076	2738	2583	1456	3730	1825
CRUSTACEA												
<i>Munna peterseni</i>	120	334	446	548	3022	3834	1479	1951	665	979	1224	1148
<i>Diastylis sympterigiae</i>	28	93	35	73	156	292	170	441	21	66	255	393
<i>Kalliapseudes schubartii</i>	2887	3485	5909	5437	7219	5507	4338	3647	3836	3965	2739	2606
<i>Kupellonura sp.</i>	14	41	28	54	14	41	14	41	0	0	21	49
<i>Mellita mangrovi</i>	0	0	0	0	7	30	0	0	0	0	0	0
<i>Misidopsis tortoneri</i>	0	0	0	0	0	0	0	0	0	0	0	0
<i>Pseudosphaeroma sp.</i>	21	66	28	70	7	30	35	73	7	30	7	30
<i>Sinelobus stanfordi</i>	14	41	156	298	1182	1160	1408	1225	835	656	948	716
<i>Cumacea n.d.</i>	0	0	0	0	0	0	0	0	0	0	0	0
MOLLUSCA												
<i>Erodona mactroides</i>	11366	7466	12385	7114	11132	4642	13652	7445	8004	5651	9561	5664
<i>Heleobia australis</i>	219	156	177	219	488	358	488	336	219	209	318	325
<i>Tagelus plebeius</i>	0	0	0	0	0	0	0	0	0	0	0	0
NEMERTINEA												
	0	0	0	0	0	0	0	0	0	0	0	0
HIRUDINEA												
	0	0	0	0	0	0	0	0	0	0	0	0

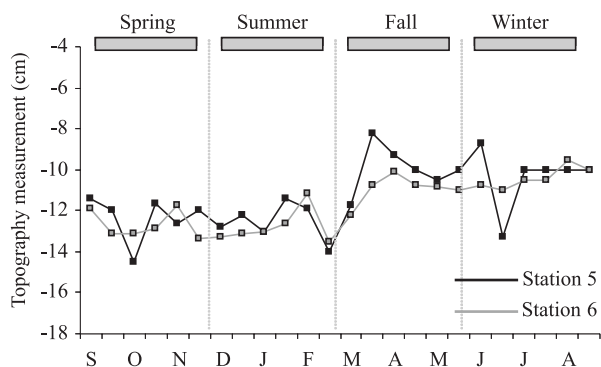
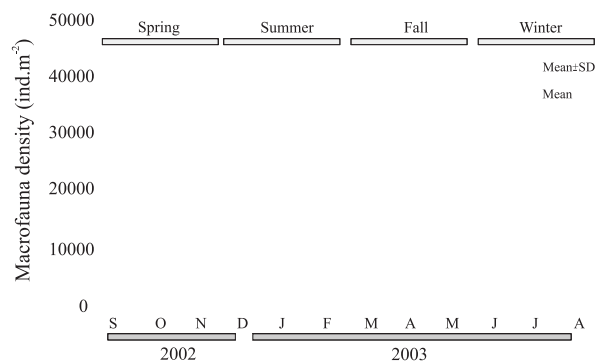


Fig. 5. Fortnightly variability (cm) on bottom topography measurements in stations 5 and 6, Patos Lagoon estuary, southern Brazil.

Fig. 6. Mean density (ind.m⁻²) of macrofauna for the study period in the Patos Lagoon estuary, southern Brazil.

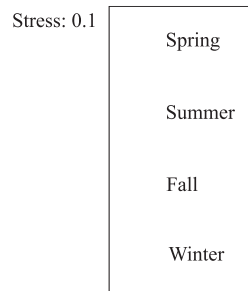


Fig. 7. Result of Multi-dimensional scaling ordination analysis (MDS) in the Patos Lagoon estuary, southern Brazil.

DISCUSSION

The results showed two well defined environmental and biotic situations: one corresponding to spring-summer period, with low macrozoobenthos densities, low values of salinity and little variation in bottom level, due to low hydrodynamics, and the another corresponding to the autumn and winter period, when an opposite pattern occurred.

The low density values recorded during the summer months in the studied shallow area contrast to those obtained in previous works conducted in Patos Lagoon. BEMVENUTI (1987) found that macrozoobenthos densities recorded in summer, above 17,000 ind.m⁻², were significantly higher than those found in the winter months (below 9,000 ind.m⁻²). The salinity conditions occurred in that work shows lowest salinity values in spring, with an increase in the summer – fall periods. Seasonal fluctuations in the estuarine area showed that reduction in the salinity average, during winter and spring may affect the abundance of benthic macrofauna assemblages, in contrast to summer and fall periods, with high salinity values, that increase macrozoobenthos densities, mainly due to recruitment of the dominant species (BEMVENUTI, 1987; BEMVENUTI & NETTO, 1998). Those high densities of macrozoobenthos in the summer were strongly influenced by the expressive recruitment of *K. schubartii* and coincided with the higher salinity and temperature values (BEMVENUTI, 1987).

It is well known that salinity influences the composition and the number of species in estuaries (LITTLE, 2000). Low salinity conditions impose severe conditions to estuarine organisms, which will therefore demand energy at high expenses, in response to osmoregulation (DAY *et al.*, 1989). The maintenance of this process could cause either mortality or inhibit their activity, limiting reproductive capacity and so its impact on future recruitment. The lower values of macrozoobenthos density coincide with low values of salinity registered between September, 2002 and February, 2003, when the estuarine area was under El-Niño influence, causing peaks of fresh water discharge, which made exceed mean values (GARCIA, 1997; GARCIA *et al.*, 2003).

Kalliapseudes schubartii is an estuarine species which can afford high levels of investment in reproduction, so responsible for the intense recruitments in summer months in the shallow water areas of Patos

Lagoon estuarine region (BEMVENUTI, 1997a, b). This species decrease its activity and seems under stress in low salinity conditions (G. Fillmann, pers. comm. Ecotoxicology Laboratory, Oceanography Department, FURG). The low salinity condition recorded during the spring-summer period, in the present study may have influenced the failures in the species recruitment.

The low *E. mactroides* recruitments intensity also influenced the low densities recorded in the present work, during the summer. These bivalve adult stocks are in northern portion of the estuarine area (BEMVENUTI *et al.*, 1978) as well along the pre-limnic and limnic area at the extreme northern section of Patos Lagoon (BEMVENUTI & NETTO, 1998). The species reproduction occurs in these areas and larvae drift along in the ebb tides colonizing the southern estuarine area, between late spring and the end of summer (BEMVENUTI *et al.*, 1978). This reproductive pattern in which larvae arrival depends on water transportation makes the recruitment of *E. mactroides* to be unpredictable in time and space (BEMVENUTI, 1997b). This pattern was also observed by the authors, during long term studies beginning in 1996 (PELD – Project, Site 8 – Patos Lagoon Estuary, unpublished data).

The fact that recruitment of *E. mactroides* had not occurred in the summer, only in the autumn in the present work, as shown in the works carried out by JORCIN (1996), influenced upon the macrozoobenthos densities, which recorded lower values than those registered in the study carried out by BEMVENUTI (1987). The coincidence between higher levels of recruitment in time (autumn) as well in space (stations #1 and #2), under greater hydrodynamics, reinforce the influence of this variable upon the distribution and abundance of the species in the southern area of Patos Lagoon estuary. In subtropical and temperate zones, summer periods are characterized by the settlement of new individuals into the coastal benthic system, due to the enhancement of the reproductive process and recruitment of macrozoobenthos. Therefore, environmental patterns such as salinity changes and hydrodynamic forces, that cause recruitment variability, add a relevant aspect on the dynamics of the benthic associations in temperate estuarine regions.

Acknowledgements. To the two anonymous referees who increased the quality of the manuscript. To the financial support from FAPERGS and CNPq (PELD – Programas Ecológicos de Longa Duração). Thanks also to Nilton Abreu, for help in the Laboratório de Ecologia de Invertebrados Bentônicos – Departamento de Oceanografia, FURG.

REFERENCES

- BEMVENUTI, C. E. 1987. Predation effects on a benthic community in estuarine soft sediments. *Atlântica* 1:5-32.
- _____. 1988. Impacto da predação sobre *Heteromastus similis* Southern, 1921 e *Nephtys fluviatilis* Monro, 1937 (Annelida, Polychaeta), em fundos moles estuarinos. *Atlântica* 10(1):8-102.
- _____. 1994. O poliqueta *Nephtys fluviatilis* Monro (1937) como predador da infauna na comunidade de fundos moles. *Atlântica* 16:87-98.
- _____. 1997a. Benthic invertebrates. In: SEELIGER, U.; ODEBRECHT, C. & CASTELLO, J. eds. **Subtropical convergence marine ecosystem. The coast and the sea in the warm temperate southwestern Atlantic**. Heidelberg, Springer Verlag. p.43-46.
- _____. 1997b. Unvegetated intertidal flats and subtidal bottoms. In:

- SEELIGER, U.; ODEBRECHT, C. & CASTELLO, J. eds. **Subtropical convergence marine ecosystem. The coast and the sea in the warm temperate southwestern Atlantic.** Heidelberg, Springer Verlag. p.78-82.
- _____. 1997c. Trophic structure. *In*: SEELIGER, U.; ODEBRECHT, C. & CASTELLO, J. eds. **Subtropical convergence marine ecosystem. The coast and the sea in the warm temperate southwestern Atlantic.** Heidelberg, Springer Verlag. p.70-73.
- BEMVENUTI, C. E.; CAPITOLI, R. R. & GIANUCA, N. M. 1978. Estudos de ecologia bentônica na região estuarial da Lagoa dos Patos. II - Distribuição quantitativa do macrobentos infralitoral. *Atlântica* **3**:23-32.
- BEMVENUTI, C. E. & NETTO, S. 1998. Distribution and seasonal patterns of the sublittoral benthic macrofauna of Patos Lagoon (South Brazil). *Revista Brasileira de Biologia* **58**(2):211-221.
- BILES, C. B.; SOLAN, M.; ISAKSSON, I.; PATERSON, D.; EMES, E.; RAFFAELLI, D. G. 2003. Flow modifies the effect of biodiversity on ecosystem functioning: an *in situ* study of estuarine sediments. *Journal of Experimental Marine Biology and Ecology* **285-286**:165-177.
- CAPITOLI, R. R.; BEMVENUTI, C. E. & GIANUCA, N. M. 1978. Estudos de ecologia bentônica na região estuarial da Lagoa dos Patos. I - As comunidades bentônicas. *Atlântica* **3**:5-22.
- CLARK, R. B. 1997. **Marine pollution.** Oxford, Clarendon. 161p.
- CLARKE, K. R. & WARWICK, R. M. 1994. **Changes in marine communities: an approach to statistical analysis and interpretation.** Plymouth, Natural Environmental Research Council. 234p.
- DAY, J. W.; HALL, C. A. S.; KEMP, W. M. & YÁÑEZ-ARANCIBIA, A. 1989. The estuarine bottom and benthic subsystem. *In*: DAY, J. W. ed. **Estuarine ecology.** New York, John Wiley & Sons, p. 338-376.
- GARCIA, C. A. E. 1997. Hydrographic characteristics. *In*: SEELIGER, U.; ODEBRECHT, C. & CASTELLO, J. eds. **Subtropical convergence marine ecosystem: the coast and the sea in the warm temperate southwestern Atlantic.** Berlin, Berlin Springer-Verlag. p.18-20.
- GARCIA, A. M.; VIEIRA, J. P. & WINEMILLER, K. O. 2003. Effects of 1997-1998 El Niño on the dynamics of the shallow-water fish assemblage of the Patos Lagoon Estuary (Brazil). *Estuarine, coastal and shelf Science* **57**:489-500.
- HOLLAND, A. F. 1985. Long-term variation of macrobenthos in an mesohaline region of Chesapeake Bay. *Estuaries* **8**:93-113.
- JORCIN, A. 1996. Distribución, abundancia y biomassa de *Erodona mactroides* (Mollusca, Bivalvia, DAUDIN, 1801), em la Laguna de Rocha (Dpto. De Rocha, Uruguay). *Revista Brasileira de Biologia* **56**(1):155-162.
- LEVINTON, J. S. 1995. **Marine biology: function, biodiversity, Ecology.** New York, Oxford University Press. 420p.
- LITTLE, C. 2000. **The biology of soft shores and estuaries.** New York, Oxford University. 252p.
- MAGURRAN, A. E. 1998. **Ecological diversity and its measurement.** London, Chapman & Hall. 242p.
- NETTO, S. A. & LANA, P. C. 1994. Effects of sediment disturbance on the structure of benthic fauna in a subtropical tidal creek of southeastern Brazil. *Marine Ecology Progress Series* **106**:239-247.
- PEREIRA, R. C. & SOARES-GOMES, A. eds. 2002. **Biologia marinha.** Rio de Janeiro, Interciência. 382p.
- QUIJÓN, P. & JARAMILLO, E. 1993. Temporal variability in the intertidal macroinfauna in the Queule River Estuary, South-Central Chile. *Estuarine, Coastal and Shelf Science* **43**(5):655-667.
- ROSA-FILHO, J. S. & C. E. BEMVENUTI. 1998. Caracterización de las comunidades macrobentónicas de fondos blandos en regiones estuarinas de Rio Grande do Sul (Brasil). *Thalassas* **14**:43-56.
- SUGUIO, K. 1973. **Introdução à sedimentologia.** São Paulo, EDUSP. 317p.
- UNDERWOOD, A. J. 1997. **Experiments in ecology – their logical design and interpretation using analysis of variance.** Cambridge, Cambridge University. 504p.
- WARWICK, R. M. 1986. A new method for detecting pollution effects on marine macrobenthic communities. *Marine Biology* **92**:557-562.
- WEISBERG, S. B.; RANASINGUE, J. A.; DAUER, D. M.; SCAFFNER, L. C.; DIAZ, R. J. & FITTHSEN, J. B. 1997. An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries* **20**(1):149-158.
- YSEBAERT, T.; HERMAN, P. M. J.; MEIRE, P.; CRAEYMEERSCH, J.; VEERBEK, H.; HEIP, C. H. R. 2003. Large-scale spatial patterns in estuaries: estuarine macrobenthic communities in the Schelde estuary, NW Europe. *Estuarine, coastal and shelf Science* **57**:335-355.

ANEXO B –Análise de conteúdo do artigo “Seasonal variability on the structure of sublittoral macrozoobenthic association in the Patos Lagoon estuary, southern Brazil (2007) / Colling, Bemvenuti e Gandra”

1 – Seasonal variability on the structure of sublittoral macrozoobenthic association in the Patos Lagoon estuary, southern Brazil (2007)

ASSUNTOS

Variabilidade espaço-temporal - Variabilidade sazonal
Região estuarina - Águas rasas
Processo de dinâmica da coluna d'água - Processo de dinâmica do substrato
Comunidade macrozoobentônica - Distribuição de macrofauna bentônica
Abundância de macrofauna bentônica

VARIÁVEIS DE ESPAÇO

Local de coleta – Estuário da Lagoa dos Patos
Ponto de coleta – entrada do Saco do Arraial, em frente à Ilha das Pombas
Quantidade de pontos de coleta – 6 (distantes 90 m entre si)
Extensão do perfil de coleta – 450 m

VARIÁVEIS DE TEMPO

Período de coletas – Setembro de 2002 a Agosto de 2003
Periodicidade da coleta – mensal (para sedimentologia e amostras biológicas); quinzenal para análise da topografia do fundo; diária (para salinidade e temperatura)

METODOLOGIA E INSTRUMENTOS

Método de coleta – Amostragem, com uso de um tubo extrator de 10 cm de diâmetro empurrado 20 cm para o fundo do substrato
Quantidade de amostragem por ponto – 3
Instrumento de coleta – tubo extrator de 10 cm de diâmetro, disco de Secchi (salinidade), termômetro (temperatura)
Outros “instrumentos” de campo – peneira com malha de 3mm para separar a amostra biológica do sedimento, formaldeído 4%, Bengal Rose, etanol 70%
Instrumento de laboratório – estereomicroscópio (0,5 mm de precisão) para medir os indivíduos capturados, pipeta para análise de granulometria

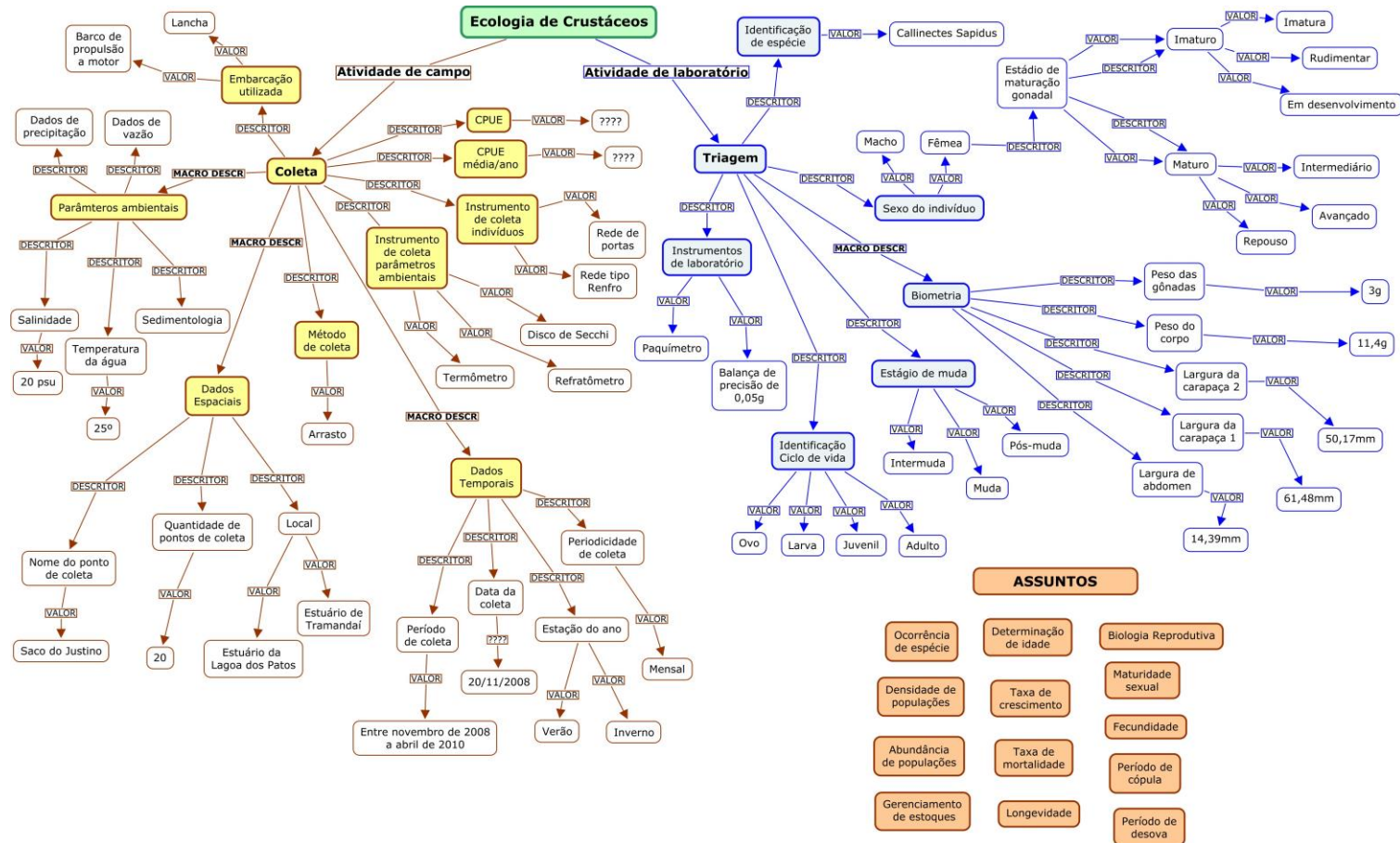
MATERIAL COLETADO/TRIAGEM

Identificação de espécies capturadas – *Erodona mactroides* Bosc, 1802
Quantidade total de organismos capturados – 23.808
Quantidade total de espécies capturadas – 17
Quantidade total de organismos por espécie – *Erodona mactroides*: 1359
Ciclo de vida – juvenis, adultos
Estação do ano – primavera, verão, outono e inverno
Biometria – comprimento do corpo
Sedimentologia – lama (silte e argila): 4,3%

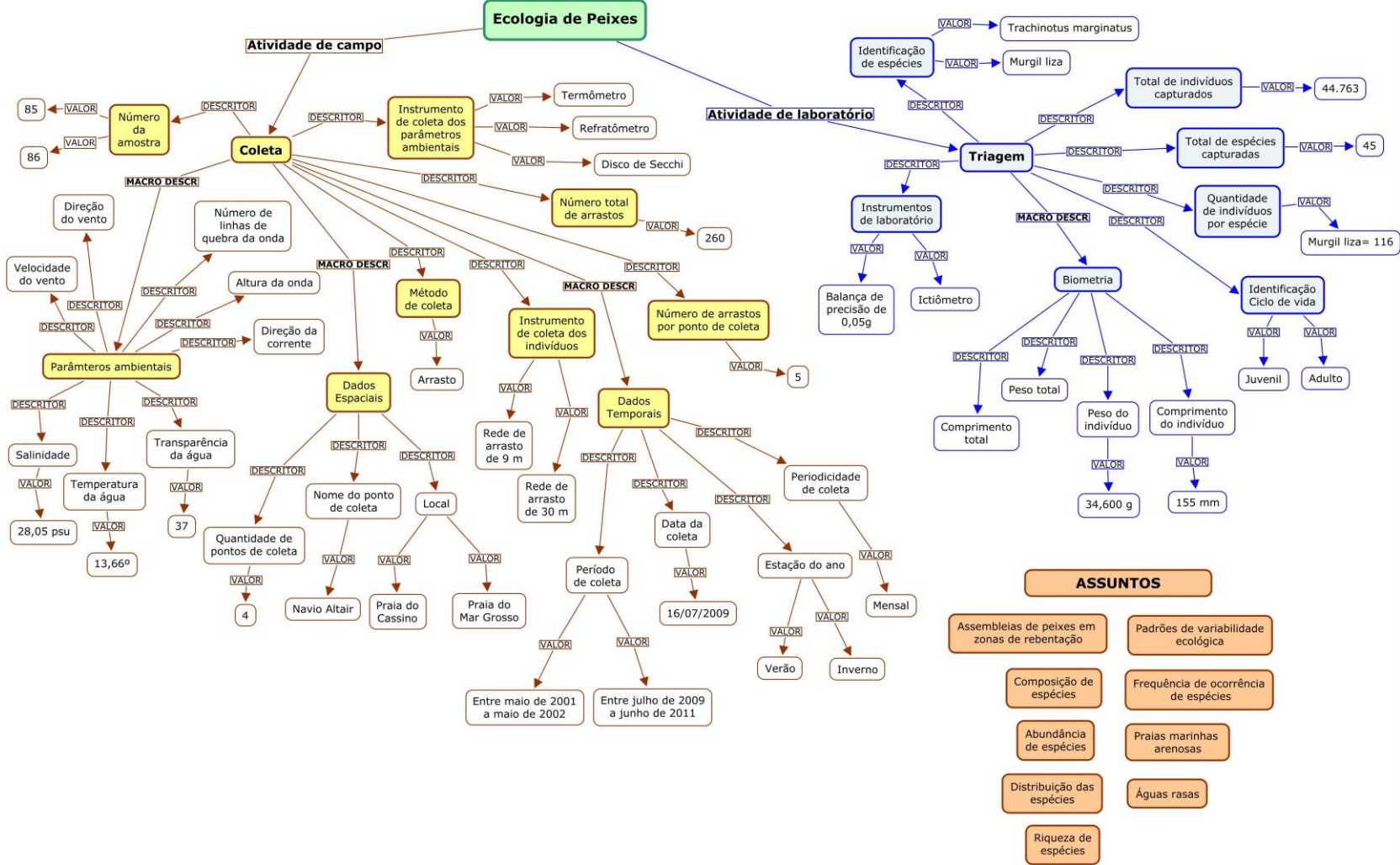
PARÂMETROS AMBIENTAIS

Temperatura da água - Salinidade da água - Topografia de fundo

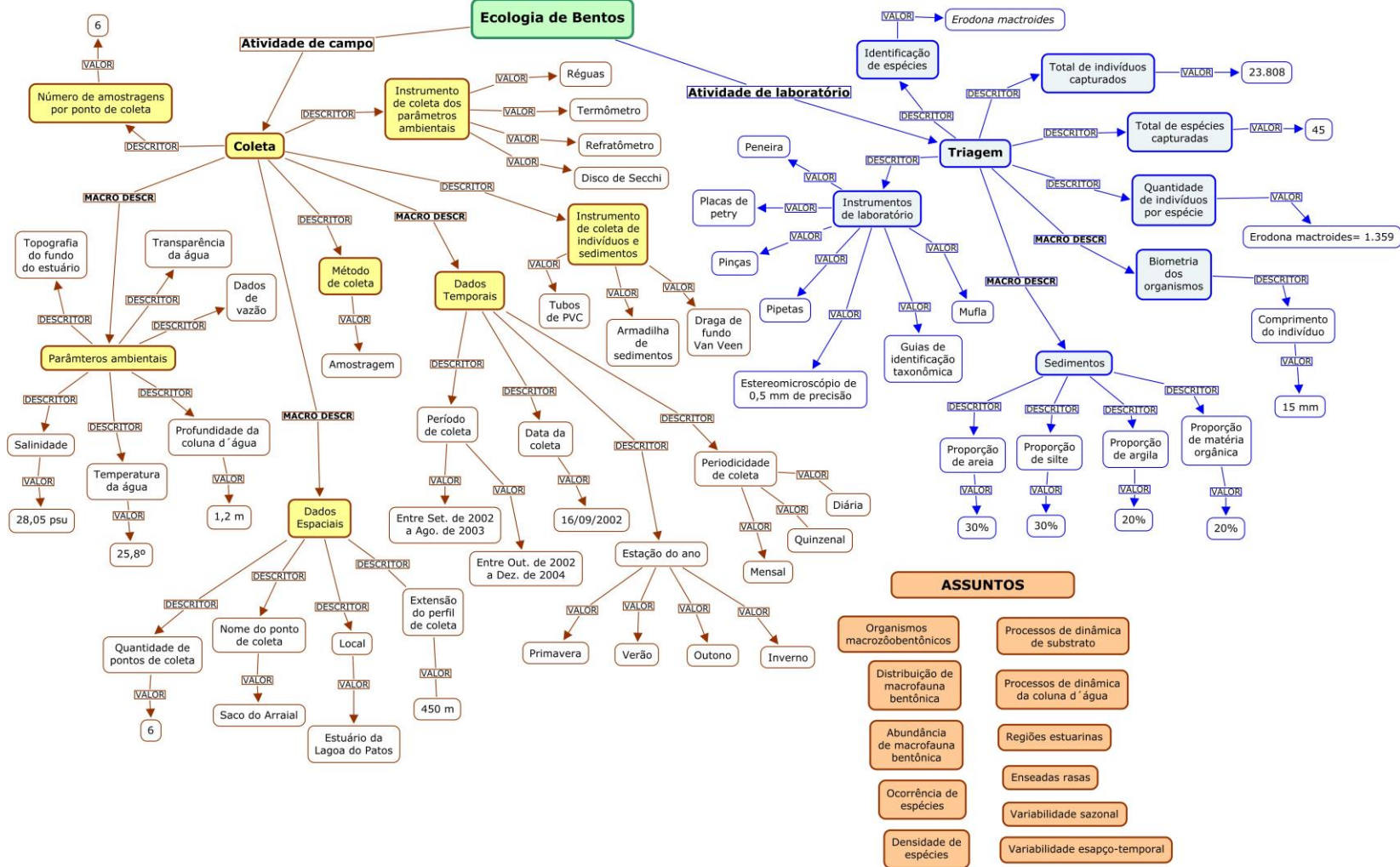
ANEXO C – Mapas de conceitos “Ecologia de Crustáceos”



ANEXO D – Mapas de conceitos “Ecologia de Peixes”



ANEXO E – Mapas de conceitos “Ecologia de Bentos”



ANEXO F – Guião de Entrevista Curadoria de Dados

Guião de Entrevista Curadoria de Dados

Investigador -
Domínio -
Data -

Início da Entrevista

Esta entrevista tem como objetivo recolher informação sobre o fluxo dos dados produzidos durante o ciclo de investigação. Entre outros aspectos, pretendo saber quais as actividades desenvolvidas para gerir os dados, assim como as necessidades de partilha e preservação dos dados.

Para efeitos de análise dos dados recolhidos, será útil proceder à gravação áudio desta entrevista. Posso contar com a sua colaboração e autorização para a gravação áudio?

-Deve (o entrevistado) ter em mente dados de investigação concretos, por exemplo aqueles com que esta a trabalhar neste momento.

Antes de iniciar a entrevista, permitir que o entrevistado possa esclarecer alguma dúvida que tenha.

- Tem alguma questão inicial?

Nota: Documento elaborado com base no Data Curation Profile Toolkit
<http://datacurationprofiles.org/>

Questão de fundo

Antes de mais, gostaria de saber mais sobre o projecto de investigação associado aos dados que vão ser abordados ao longo da entrevista. Peço-lhe que faça uma breve descrição sobre a investigação em que esta envolvido(a).

1 – O DataSet

1.1 Pode fazer uma descrição sobre os dados de investigação com os quais está a trabalhar, especificando o valor dos mesmos?

Referir características dos dados: (observação, experimental, simulação), o tipo de ficheiros, tamanho, quantidade, etc...

2 – Ciclo de vida dos dados / Preservação

2.1. Pode detalhar o processo de investigação do ponto de vista dos dados? O que quero saber em particular é o que acontece aos dados durante a investigação, e como os caracteriza em cada fase do processo de investigação.

Por exemplo, descrevendo o processo de investigação numa série de passos e actividades em torno dos dados. Alguns passos podem incluir dados em bruto, o processamento dos dados, a análise e o produto final da investigação.

2.1.1. O que acontece aos dados em cada fase?

2.1.2. Porquê? Com que objetivo?

2.1.3. Como acontece (que instrumentos são usados em particular)?

2.1.4. Quais as pessoas envolvidas?

2.2 Quais são os dados, de acordo com as fases do ciclo de vida dos dados, mais importantes a preservar?

- a gerir e manter ao longo do tempo.

Notas a debater

Prioridades de preservação.

Partes adicionais ou componentes dos dados que devam ser preservados, para além dos dados passíveis de ser partilhados.

O valor que os dados retêm ao longo do tempo, tanto para os investigadores como para os outros.

3 – Partilha

3.1. Para cada uma das fases discutidas anteriormente, pode indicar com quem podem ser partilhados e com quem?

	Ninguém	Colaboradores	Investigadores da mesma instituição	Investigadores da mesma área	Investigadores de outra área	Todos
Fase Inicial						
2ª Fase						
3ª Fase						
4ª Fase						
Outras						

3.2. Atualmente partilha os dados com alguém?

Se sim

3.2.1. Com quem?

3.2.2. Como é feita a partilha?

3.2.3. Existe alguma condição para essa partilha?
(*ex. confidencialidade*).

3.3. Que outras pessoas podem ter interesse em aceder aos dados?

3.4. Que valor podem os outros obter a partir da utilização destes dados?

4 – Preparação dos dados a submeter num repositório (no caso de existir esse interesse)

4.1. Que atividades devem ser consideradas para que os dados possam ser submetidos num repositório, ou transferidos para um serviço de curadoria de dados?

4.2. Em qual das fases, previamente discutidas, os dados são passíveis de ser submetidos?

4.3. Na possibilidade de um serviço que possa proteger o acesso de terceiros aos dados, durante um período de embargo, acha que os dados podem ser sujeitos a essa condição?

Embargo implica a restrição de acesso durante um período de tempo pré-definido antes de serem tornados públicos.

4.4 Em que circunstâncias?

Notas a debater

Revisão dos dados para identificar erros e garantir a qualidade.

Editar os dados para modificar ou remover qualquer variável ou informação que possa ser usada para identificar sujeitos.

Garantir conformidade com protocolos das identidades financiadoras.

Obter autorização de outros envolvidos e rever documentação e materiais que suportam o uso e curadoria dos dados.

5 – Organização e descrição dos dados

5.1. Pode explicar as actividade que desenvolve(m) para organizar os dados?
- materiais de apoio, anotações, cabeçalhos numa tabela, etc...

5.2. São utilizados normas para descrever os dados? Quais?

5.3. A descrição dos dados é uma atividade prioritária?

5.4. A descrição dos dados, se ocorre, providencia informação suficiente para que outras pessoas possam interpretar os dados?

5.5. Existe a necessidade de providenciar informação adicional, de contexto, para que outros possam interpretar e utilizar os dados?

6 - Propriedade intelectual

6.1. Quem detém propriedade sobre os dados?

6.2. Existe alguma entidade financiadora do trabalho de investigação? – Qual?

Se sim, essa entidade exige...

6.2.1. *a elaboração de um plano de gestão dos dados?*

6.2.2. *a partilha, publicação ou submissão dos dados num repositório?*

6.2.3. *a preservação dos dados após a investigação?*

6.3. Quem são os outros stakeholders destes dados?

Stakeholders podem ser definidos como: grupos, organizações, indivíduos ou outros que podem ser vistos como tendo feito um investimento nos dados, ou que devam ser consultados em questões relacionadas com os dados.

7 - Ferramentas

7.1. Que ferramentas estão envolvidas na captura e na utilização dos dados?

7.2. Em que medida essas ferramentas são utilizadas?

7.3. As ferramentas são necessárias para que outros possam reutilizar os dados?

7.4. Existem alternativas a estas ferramentas?

8- Relação entre os dados e publicações

8.1. As entidades que publicam os seus resultados, aceitam ou exigem informação adicional como os dados para publicação?

Se sim

8.1.1 *Como prepara os dados para serem submetidos a um dessas entidades?*

Considera importante um serviço que...

8.2.1. *relacione os seus dados com as publicações ou outros produtos da investigação?*

8.2.2. *permita que os dados possam ser integrados com outros dados?*

9 - Impacto

Considera importante obter informação sobre....

9.1.1. *estatísticas do uso dos dados (valores quantitativos)?*

9.1.2. *as pessoas que usaram os seus dados (valores qualitativos)?*

9.2. Que tipo informação gostaria de obter sobre os utilizadores, as pessoas que acederam ou usaram os dados gostaria de obter? Em que sentido?

9.3. Para além de estatísticas de utilização, existem outras medidas que gostaria de ver associadas aos dados?

Pedir informação detalhada se o investigador apresentar alguma sugestão.

FIM

ANEXO G – Guião de utilização do Dendro

How Dendro works

Produce data

Dendro manages all the files in your research environment. PDF, Word, audio, video, etc.

Describe your data and share them with your team

Simple interface
Collaboration
Full changes tracking

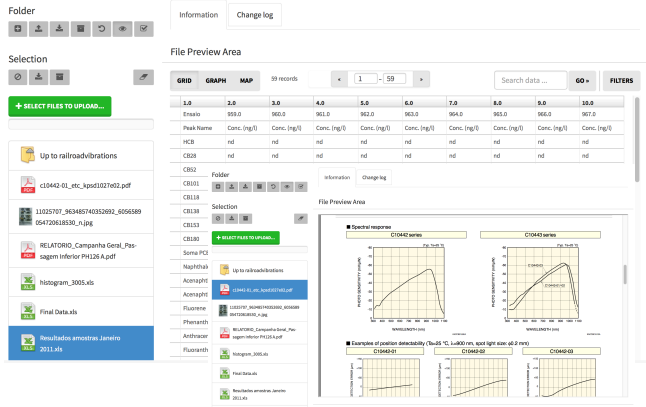
Export it*

Share only the data that you want to share

Cite

Foster collaboration by showcasing your data

*Fully integrated with known repository solutions



File previews

- PDF, Excel, Documents, etc can be directly previewed in the browser, when the "Edit mode" is disabled.

Full Change log

- Track the changes by project collaborators to the metadata in projects and files

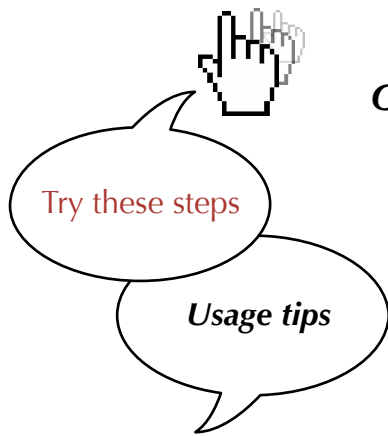
Revision Number	Time	Author	Summary
Current 2	Last Tuesday at 5:58 PM	Paulo Sousa	Added Subject with value "User interface testing".



Collaborative research data management platform

Quick start guide

For inquiries or assistance, please contact
João Rocha da Silva
joaorosilva@gmail.com
<http://dendro.fe.up.pt>



Options

- List your projects
- Find other users

Metadata viewer / editor

- See information about the selected file or folder
- Edit information

Search box

- Search the entire platform for interesting datasets

Tools

- Switch between edit and read-only modes. Read-only shows file previews, edit mode shows the information editor
- Share the data to an external repository
- Download information, in text-based formats

The screenshot shows the Dendro platform interface. At the top, there's a navigation bar with 'Dendro^β', 'My', 'Find', 'Plugins', 'Search Dendro', 'SUBMIT', and 'João Rocha'. Below this, a breadcrumb trail shows 'railroadvibrations > Base data gathering - August Run'. The main area is divided into three sections: a 'Folder' view on the left, a 'Metadata editor' in the center, and a 'Search' results panel on the right. The 'Folder' view shows a list of files and folders, with one file selected. The 'Metadata editor' shows fields for 'Description', 'Title', and 'Creator', with a 'PREVIEW' button. The 'Search' panel shows a list of search results with 'Dublin Core terms' selected. A hand icon with the number '1' points to the folder view, '2' to the 'EDIT MODE ON' button, '3' to the 'Dublin Core terms' search result, and '4' to the 'Select descriptor group' dropdown menu.

File and folder operations

- Create new folders
- Upload / Download files
- Backup the current folder
- Restore a past backup
- Show / Hide deleted files
- Operate on multiple files at once

File explorer

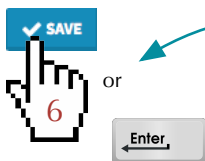
- Select a file to preview its contents or see its description
- Double-click a folder to open it
- Go up in the hierarchy by selecting the option at the top
- Select files for the File and Folder operations

Little time savers

- Replicate information from the parent folder
- Remove single descriptors

Save your changes

- Click save or press enter when editing a descriptor to save the current annotations



The screenshot shows the 'Audience Education Level' descriptor editor. It has a dropdown menu set to 'Audience Education Level' and a text input field containing 'MSc, PhD students in information science'. A hand icon with the number '5' points to the text input field.

Add a descriptor to fill in

- Click the green arrow on the right of the button to add multiple copies of the same descriptor (e.g. multiple authors)

Select descriptor group

The screenshot shows the 'Descriptors' panel. It has a dropdown menu set to 'Dublin Core terms'. Below the menu, there are several descriptors with their descriptions: 'Audience', 'Audience Education Level', 'Bibliographic Citation', 'Contributor', 'Coverage', and 'Creator'. A hand icon with the number '4' points to the 'Audience Education Level' descriptor.

ANEXO H – Processo de experiência com investigadores do Dendro

Experiências Dendro

Localização da máquina da experiência	Credenciais do user administrador

Projecto de demonstração

Deverá ser criado um projecto de demonstração do Dendro. O projecto de demonstração deve conter um dataset bem anotado e tem como objectivo demonstrar a cada investigador qual será o resultado final das experiências.

Registo de users

Os utilizadores deverão ser registados antecipadamente, e ter logins e passwords fornecidos por nós, em papel e na hora.

Criação de projetos

Os projectos deverão ser criados antecipadamente pelo utilizador administrador. Depois de terem sido criados projectos para cada área científica / grupo de investigação, serão adicionados os respectivos colaboradores a cada um dos projetos. Nenhum dos investigadores poderá administrar, para já, os seus projetos. Essa administração fica a nosso cargo.

Funcionalidades de recomendação de descritores

Na primeira fase de experiências não haverá nenhuma funcionalidade de recomendação ativa. Os investigadores terão que seleccionar sempre a ontologia da qual os descritores deverão ser retirados e depois acrescentar manualmente os descritores que pretendem preencher.

Guião para entrevista

A verificar antes da entrevista:

- [Username e Password](#) (testar) - falar com Ricardo se não funcionarem ou criar outros se for urgente
- Projeto vazio para o teste
- [Dendro de produção](#)

Guião

1. Dar noção do conceito de descritor (se necessário)
2. Apresentar os objetivos da avaliação
3. Apresentar a plataforma com base no projecto exemplo (a criar)
 - a. Conceito de projetos e navegação por pastas;
 - b. Layout da interface e módulos;
 - c. Breve apresentação e navegação das ontologias;
 - d. Seleção e preenchimento de descritores
4. Perguntar se há alguma dúvida antes de começar
5. Deixar o investigador à vontade durante a interação

Decisões

-> Fazer uma primeira demonstração de anotação ao utilizador, usando um projecto de uma área que não a dele. Neste exemplo, deverá ser criada uma pasta no projecto, deverá ser enviado um ficheiro e adicionados dois descritores. O processo de navegação pelos descritores (selecciona ontologia-> Selecciona descritor-> voltar atrás) deve também ser mostrado. Chamar a atenção para o facto do utilizador ter que clicar no botão de gravar para gravar as alterações, mas não precisa de o fazer sempre que preenche um descritor (pode adicionar vários descritores, preenchê-los e só gravar no final).

-> Deverão estar **presentes dois elementos** : um será o elemento principal, que fará as demonstrações, conversará com o investigador e esclarecerá as dúvidas que possam surgir no investigador durante a experiência; o segundo utilizador irá **cronometrar o trabalho** e registar os instantes de tempo nos quais ocorrer algum problema, dúvida ou pergunta do investigador.

Muito importante é registar onde o investigador fica “parado” sem saber o que fazer. Outros exemplos: investigador selecciona e apaga descritor; investigador desabafa em voz alta dúvidas; investigador demonstra com algum tipo de expressão que está com dúvidas (coça a cabeça por exemplo)

Também deverá registar quando começa e acaba uma tarefa (começo: começar a descrever; fim: gravar os descritores com sucesso).

-> Ajudas: No início da experiência apresentar o guião

(<https://www.dropbox.com/s/krt4839ua7jkytb/Gui%C3%B5es%20Dendro.pdf?dl=0>).

Explicar o fluxo mostrado pelas “mãozinhas” e explicar as principais áreas do interface do Dendro; no fundo, ler o folheto com o investigador. Durante a

experiência, responder a todas as perguntas do investigador sem receios. Não é necessário ocultar nada se o investigador tomar a iniciativa de perguntar alguma coisa.

-> Será necessário **gravar o áudio das experiências**. Levar um telemóvel para gravar audio e outro para cronometrar as experiências.

-> **Em caso de ocorrerem bugs**: Registrar o instante de tempo da ocorrência. Dizer que a solução ainda está a ser desenvolvida; tentar criar outra pasta ao lado daquela onde o erro ocorreu (mesmo nível) e recomeçar o processo de anotação.

-> **Inquérito no final da experiência**: Devemos enviar um email a agradecer muito a participação e a enviar o link para o inquérito a preencher. Pedir para preencher “na próxima semana, se possível”. LINK: <http://goo.gl/forms/5vEyfqdoWy>

ANEXO I – Inquérito para avaliação do perfil de aplicação Biological Oceanography”

Inquérito para avaliação do perfil de aplicação “Biological Oceanography”

Considerando o vocabulário “Biological Oceanography” que se encontra em anexo, contendo descritores gerais e descritores específicos da área de Oceanografia Biológica e áreas interdisciplinares, e considerando sua interação com a plataforma de gestão de dados “Dendro”, responda:

1. Usou ou usaria os descritores deste vocabulário para descrever seus dados de investigação? (Se escolheu descritores deste vocabulário quando fez depósito e descrição de conteúdos na plataforma Dendro)
2. Considera os descritores compatíveis com o domínio da Oceanografia Biológica? (Se os descritores refletem as atividades de investigação da área. Por exemplo, os descritores “sampling description” ou “sampling periodicity” conseguem descrever adequadamente atividades práticas de investigação na área, como os eventos de coleta?)
3. A linguagem dos descritores é semanticamente adequada à investigação no domínio da Oceanografia Biológica? (Se o vocabulário consegue captar a linguagem técnica usada na área da Oceanografia Biológica, por exemplo, os descritores “scientific name” e “life stage” pertencem à linguagem utilizada no domínio?)
4. Considera os descritores úteis para descrição de dados (ou produtos gerados a partir destes, como relatórios, artigos, etc.) no domínio da Oceanografia Biológica? (Se os descritores, tanto os gerais quanto os específicos, são úteis para descrição de conteúdo neste domínio, ou não, não usaria nenhum destes descritores para descrever conteúdos neste domínio)
5. Considera o perfil de aplicação completo? Acrescentaria, suprimiria ou modificaria descritores? (Considera que os descritores conseguem abranger vários aspectos da investigação na área? Por exemplo, os descritores gerais tem o objetivo de descrever questões mais formais dos projetos de investigação, artigos publicados ou dos próprios dados; os descritores específicos tem o objetivo de descrever aspectos mais práticos da investigação, dos eventos de coleta ou atividades de laboratório)
6. Alguma sugestão para melhorar o vocabulário “Biological Oceanography”?