

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMUNICAÇÃO E INFORMAÇÃO

DENISE RAMIRES MACHADO

DADOS DE PESQUISA EM REPOSITÓRIO INSTITUCIONAL:
o caso do Edinburgh DataShare

Porto Alegre
2015

DENISE RAMIRES MACHADO

DADOS DE PESQUISA EM REPOSITÓRIO INSTITUCIONAL:
o caso do Edinburgh DataShare

Dissertação apresentada para o Programa de Pós-Graduação em Comunicação e Informação, Linha de Pesquisa Informação, Redes Sociais e Tecnologias, da Faculdade de Biblioteconomia e Comunicação, da Universidade Federal do Rio Grande do Sul.

Orientadora: Prof^ª. Dra. Ida Regina Chittó Stumpf

Porto Alegre
2015

CIP - Catalogação na Publicação

Machado, Denise Ramires

Dados de pesquisa em repositório institucional : o caso do Edinburgh DataShare / Denise Ramires Machado. -- 2015.

111 f.

Orientadora: Ida Regina C. Stumpf.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Comunicação e Informação, Porto Alegre, BR-RS, 2015.

1. Repositórios de dados de pesquisa. 2. Dados de pesquisa. 3. Comunicação científica. I. Stumpf, Ida Regina C., orient. II. Título.

Denise Ramires Machado

DADOS DE PESQUISA EM REPOSITÓRIO INSTITUCIONAL:
o caso do Edinburgh DataShare

Dissertação apresentada para o Programa de Pós-Graduação em Comunicação e Informação, Linha de Pesquisa Informação, Redes Sociais e Tecnologias, da Faculdade de Biblioteconomia e Comunicação, da Universidade Federal do Rio Grande do Sul.

Data: 12 de maio de 2015.

BANCA EXAMINADORA:

Prof. Dr. Rafael Port da Rocha
Universidade Federal do Rio Grande do Sul – FABICO

Prof^ª. Dra. Sonia Elisa Caregnato
Universidade Federal do Rio Grande do Sul – PPGCOM

Prof^ª. Dra. Ana Maria Mielniczuk de Moura
Universidade Federal do Rio Grande do Sul– PPGCOM

Prof^ª. Dra. Samile Andrea de Souza Vanz (Suplente)
Universidade Federal do Rio Grande do Sul – PPGCOM

AGRADECIMENTOS

Gostaria de agradecer em primeiro lugar a Deus, pois sem Ele nada é possível!

Agradeço também à minha orientadora, Profa. Dra. Ida Stumpf, por toda a colaboração e parceria nesse processo.

À minha família, em especial ao meu marido Guilherme, ao meu pai Romeu, irmãos, irmãs, sobrinha e sobrinho e às minhas companheirinhas-cachorro Tiffany e Mina que tanto alegraram minha vida nesse período.

Às Dra. Márcia e Vladia e à Bérlys (terapeuta holística).

Ao pessoal do Espaço Vida Saudável Osvaldo Aranha: as anfitriãs Rosario e Linda e as amigas de lá: Marlene, Miss e Carla.

Às chefes e colegas de trabalho da Comissão de Automação. Com um carinho especial destaque: Zita, Zuleika e Beatriz.

Às amigas de longa data: Bruna e Marinez, que sempre estiveram prontas para conversar sobre tudo, mesmo que fosse pela milésima vez sobre repositórios de dados de pesquisa!

So instead of sitting in the library and waiting for the books to come to you, you proactively go to where people are doing the research. [...] So data services, data management, and data curation are proactive services.
(Professor Jian Qin - Syracuse University)

RESUMO

Este estudo analisou as relações entre diretrizes e práticas do repositório institucional universitário de dados de pesquisa Edinburgh DataShare, no contexto do gerenciamento de dados de pesquisa, baseando-se nas seguintes categorias: responsabilidade, conteúdo, aspectos legais, padrões, preservação digital, política de acesso e uso e sustentabilidade e financiamento. A metodologia foi um estudo de caso qualitativo com levantamento de documentos e dados na Internet e observação direta do repositório. Utilizou o software NVivo para registro e análise dos dados. Nas características do Edinburgh DataShare quanto à responsabilidade, há a presença de profissionais de tecnologia da informação (TI) e profissionais da informação (com destaque para bibliotecários) na equipe. Sobre o conteúdo, foram examinados os metadados de 161 itens recuperados em 09 de outubro de 2014. Como o repositório inicialmente foi povoado através de um projeto piloto, a maioria dos itens (103 itens) foi criada por um mesmo pesquisador, Bert Remijnsen, da área de assunto de Linguística e Idioma Inglês (Linguistics and English Language). Com relação aos aspectos legais, a adoção de licenças abertas é uma opção que o Edinburgh DataShare oferece, com a licença ODC-BY, mas também há a opção de não escolher essa licença e indicar licenças diferentes ou outras informações relativas a copyright no metadado dc.rights. Sobre os padrões, o software utilizado é o DSpace, que permite a interoperabilidade com outros sistemas internos e externos à Universidade. O uso de um perfil de aplicação do padrão Dublin Core qualificado, específico para conjunto de dados de pesquisa, facilita a recuperação da informação e a interoperabilidade com outros sistemas, por usar um padrão reconhecido mundialmente. O fluxo de depósito documentado e disponível na Internet e a inserção no fluxo de gerenciamento de dados de pesquisa da Universidade trazem segurança e estabilidade para os serviços do repositório. Há uma política de preservação digital do repositório que norteia as ações de preservação. Somente o Handle era utilizado como identificador permanente até o início de novembro de 2014, e a partir desse momento, passou a ser incluído também o DOI. Sobre o acesso e uso, o depósito é efetuado por pessoas vinculadas à Universidade, de dados de pesquisa da Universidade, e o acesso aos metadados e à maioria dos materiais é livre para todos, sem necessidade de identificação. A recuperação da informação está em desenvolvimento e em fevereiro de 2015 uma nova forma de pesquisa foi disponibilizada, ampliando as possibilidades de acesso, compartilhamento e uso dos conjuntos de dados de pesquisa, potencializando o alcance dos objetivos do Edinburgh DataShare. Com relação à sustentabilidade e ao financiamento, poucas informações foram recuperadas, porém ficou evidente, por seu início ter sido por um projeto financiado dentro do Jisc Repositories and Preservation Programme e por ter sido incluído em uma estrutura da Universidade, que é um projeto que exige um grande investimento e que necessita de apoio institucional para assegurar sua continuidade. O Edinburgh DataShare é uma parte essencial do gerenciamento de dados de pesquisa da Universidade, mas não é o único mecanismo de curadoria digital utilizado pela Universidade. O Edinburgh DataShare está cumprindo a tarefa de complementar o ciclo da comunicação científica e proporcionar as condições de criação das chamadas publicações ampliadas ao oferecer os serviços que permitem que os pesquisadores vinculem seus dados de pesquisa às suas publicações através de identificadores permanentes. Apresentando as relações entre as diretrizes e as práticas do Edinburgh DataShare, no contexto do gerenciamento de dados de pesquisa, foi possível perceber uma relação de construção contínua das diretrizes e das práticas. Essa forma de construção traz como consequência algumas diferenças observadas entre as diretrizes e as práticas, visto que nem sempre elas estão no mesmo momento de maturidade.

Palavras-chave: Repositórios institucionais de dados de pesquisa. Dados de pesquisa. Comunicação científica. Edinburgh DataShare.

ABSTRACT

This study examined the relationship between policies and practices of the university research data institutional repository Edinburgh DataShare in the context of the management of research data, based on the following categories: responsibility, content, legal aspects, standards, digital preservation, policy access and use, and sustainability and financing. The methodology was a qualitative case study of survey documents and data on the Internet and direct observation of the repository. It used NVivo software for recording and data analysis. The characteristics of Edinburgh DataShare for accountability, there is the presence of information technology (IT) professionals and information professionals (especially librarians) in the team. On content, metadata 161 items recovered on 09 October 2014 were examined since the repository was initially populated by a pilot project, most of the items (103 items) was created by the same investigator, Bert Remijsen, of subject area of Linguistics and English Language (Linguistics and Language Inglês). Regarding the legal aspects, the adoption of open licenses is an option that Edinburgh DataShare offers, with the ODC-BY license, but there is also the option of not choosing the license and indicate different licenses or other information concerning copyright metadata in dc .rights. About the standards, the software used is DSpace, which enables interoperability with other internal and external systems to the University. The use of a standard application profile Qualified Dublin Core, for specific set of research data, facilitates the retrieval of information and interoperability with other systems, by using a standard recognized worldwide. The deposit flow documented and available on the Internet and the inclusion in the University's research data management workflow bring security and stability to the services repository. There is a digital preservation policy repository that guides the preservation actions. Only Handle was used as a permanent identifier to the beginning of November 2014, and from that moment, became also included the DOI. On access and use, the deposit is made by people linked to the University, research data from the University, and access to metadata and to most materials is free for all, without identification. Information retrieval is under development; in February 2015 a new form of research was available, expanding the possibilities of access, sharing and use of sets of research data, increasing the scope of the objectives of Edinburgh DataShare. With regard to sustainability and financing, little information was retrieved, but it was evident, by the beginning was a project funded by JISC Repositories and within the Preservation Programme and have been included in a structure of the University, which is a project that requires a big investment and requires institutional support to ensure its continuity. Edinburgh DataShare is an essential part of the University research data management, but is not the only digital curation mechanism used by the University. Edinburgh DataShare is fulfilling additional task the cycle of scientific communication and provide the conditions for the creation of so-called extended publications by providing services that enable researchers to bind your search data to their publications through permanent identifiers. Introducing the relationship between the guidelines and the Edinburgh DataShare practices in the context of research data management, it was possible to see a continuous relationship building guidelines and practices. This form of construction brings as a consequence some differences between the guidelines and practices, as they are not always at the same time of maturity.

Keywords: Research data institutional repositories. Research data. Scientific communication. Edinburgh DataShare.

LISTA DE FIGURAS

Figura 1 – Paradigmas da ciência	19
Figura 2 – Modelo de Brody e Harnad: acesso e impacto da pesquisa maximizados por meio do autoarquivamento	22
Figura 3 – Modelo do ciclo de vida da curadoria digital	27
Figura 4 – Modelo de Sales e Sayão: modelo tradicional alterado pelo reuso de dados científicos	30
Figura 5 – Áreas estratégicas da implementação da política de gerenciamento de dados de pesquisa da University of Edinburgh	65
Figura 6 – Opções de busca e pesquisa por listas no Edinburgh DataShare	71
Figura 7 – Opção de recuperação da informação Discover no Edinburgh DataShare ..	72
Figura 8 – Continuum do compartilhamento de dados de pesquisa	80
Figura 9 – Artigo da revista Classical and Quantum Gravity	83
Figura 10 – Artigo depositado no arXiv.org (Versão 2)	84
Figura 11 – Item do Edinburgh DataShare	85

LISTA DE GRÁFICOS

Gráfico 1 – Quantidade de tipos de dados de pesquisa por item	50
Gráfico 2 – Tipos de dados de pesquisa	51
Gráfico 3 – Extensões dos arquivos dos itens	52
Gráfico 4 – Quantidade de assuntos (metadado dc.subject) por item	54
Gráfico 5 – Tipos de licenças dos itens	59
Gráfico 6 – Informações do metadado dc.rights nos itens que têm somente o Depositor Agreement	60

LISTA DE QUADROS

Quadro 1 – Repositórios de dados de pesquisa recuperados no OpenDOAR	35
Quadro 2 – Repositórios de dados de pesquisa recuperados no Databib	37
Quadro 3 – Repositórios de dados de pesquisa recuperados no re3data.org	38
Quadro 4 – Faculdades e Escolas	40
Quadro 5 – Critérios e operacionalização	44
Quadro 6 – Padrão de metadados do Edinburgh DataShare	62

LISTA DE TABELAS

Tabela 1 – Tipos de dados de pesquisa dos itens	50
Tabela 2 – Dez criadores que têm mais itens no repositório	53
Tabela 3 – Assuntos (metadado dc.subject) dos conjuntos de dados do Edinburgh DataShare	55
Tabela 4 – Frequência de assuntos (metadado dc.subject.classification) dos conjuntos de dados do Edinburgh DataShare	57

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Justificativa	15
2 OBJETIVOS	17
2.1 Objetivo geral	17
2.2 Objetivos específicos	17
3 MARCOS TEÓRICOS	18
3.1 Comunicação científica	18
3.2 Repositórios institucionais: conceituação, objetivos, políticas e diretrizes	20
3.3 Dados de pesquisa: definição e sua importância para a comunicação científica	23
3.4 Curadoria digital	26
3.5 Repositórios institucionais de dados de pesquisa	29
3.6 Políticas de Informação para repositórios: aspectos a serem observados nas diretrizes e nas práticas	31
4 PROCEDIMENTOS METODOLÓGICOS	33
4.1 Procedimentos de escolha do objeto de estudo	33
4.2 Contexto do estudo	40
4.3 Objeto de estudo	42
4.4 Definições dos termos	42
4.5 Categorias de análise	43
4.6 Fontes de coleta de dados	45
4.7 Instrumentos de coleta de dados	45
4.8 Procedimentos de coleta, análise e tratamento dos dados	46
4.9 Limitações da pesquisa	47
5 ANÁLISE E DISCUSSÃO DOS DADOS	48
5.1 Características do Repositório Edinburgh DataShare	48
5.2 Edinburgh Datashare e o gerenciamento de dados de pesquisa	76
5.3 Edinburgh Datashare e o novo modelo de comunicação científica	82

6 CONSIDERAÇÕES FINAIS	86
REFERÊNCIAS	91
APÊNDICE A - NVIVO	97
APÊNDICE B - PLANILHA EXCEL	102
APÊNDICE C - ROTEIRO DE OBSERVAÇÃO.....	103
ANEXO A - POLÍTICA DE GERENCIAMENTO DE DADOS DE PESQUISA	107
ANEXO B - POLÍTICA DE PRESERVAÇÃO	108
ANEXO C - POLÍTICA DE SUBMISSÃO	109
ANEXO D - TABELA DE FORMATOS DE ARQUIVO UK DATA ARCHIVE	110

1 INTRODUÇÃO

O fazer científico costuma ser reconhecido por seus produtos finais: publicações acadêmicas, como artigos em periódicos e livros e capítulos de livros. Os trabalhos apresentados em eventos também trazem reconhecimento, apesar de serem uma etapa na construção das publicações finais. Nas últimas décadas, com essas publicações sendo disponibilizadas na Internet e tendo sua visibilidade aumentada, as universidades passaram a utilizar repositórios institucionais para armazenar esses produtos, e as redes sociais passaram a ser meios utilizados por diversas pessoas para compartilhar os *links* dessas publicações.

Mas há outra parte do fazer científico que não é tão estruturado quanto os produtos que costumam trazer reconhecimento aos pesquisadores: os dados de pesquisa, também conhecidos como dados científicos. Enquanto artigos, livros e anais de evento costumam ter estruturas mais ou menos consolidadas, construídas há muitos anos, os dados de pesquisa podem variar enormemente, incluindo fatores referentes à metodologia adotada para coleta, formas de registro, se são avaliados ou não por pares, entre outros. O compartilhamento dos dados é relevante para a verificação da pesquisa, para construir novas pesquisas com custos menores.

Com o aumento na intensidade da produção de dados de pesquisa característica da e-Science – definida por Gray (2009) como “onde a tecnologia da informação encontra a ciência”, o mundo vive hoje sob o que se conhece como um “dilúvio de dados”, mais especificamente de dados digitais de pesquisa, os quais podem ser tanto originalmente digitais (nascidos digitais) quanto digitalizados a partir dos formatos analógicos, ampliando cada vez mais a quantidade e a variedade dos dados disponíveis digitalmente.

Instituições pelo mundo inteiro têm estudado soluções para o gerenciamento dos dados de pesquisa, pois essa tem se tornado uma questão cada vez mais urgente para os governos, os pesquisadores e o público, pois está diretamente relacionada com a sustentabilidade da pesquisa financiada com recursos públicos, e com a responsabilidade social do pesquisador.

A Ciência da Informação (CI) nos últimos anos tem se ocupado de estudar o gerenciamento dos dados de pesquisa. Sales e Sayão (2012) notam que essa não é uma questão totalmente nova para a CI, mas com o desenvolvimento das mídias eletrônicas e sua influência na comunicação científica, bem como questões referentes à obsolescência tecnológica, a necessidade desse gerenciamento assumiu maior relevância.

Devido à variedade dos dados de pesquisa e à necessidade de estabelecer um denominador comum para embasar estudos que permitam seu gerenciamento, é necessário estabelecer definições para eles. Rodrigues *et al.* (2010, p. 48), no relatório sobre o estado da arte de repositórios de dados científicos em Portugal, estabelecem: “Dados que são produzidos no contexto de investigação científica ou que de alguma forma são usados em investigação”. A Organização para a Cooperação Econômica e Desenvolvimento (OCDE), apresenta da seguinte forma o que são os dados de pesquisa:

[...] “dados de pesquisa” são definidos como registros factuais (números, registros textuais, imagens e sons) utilizados como fontes primárias para a pesquisa científica, e que são geralmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. (OCDE, 2007, p.13, tradução nossa)

Esses dados de pesquisa são recursos de altas possibilidades, pois representam a base a partir da qual os cientistas constroem os seus conhecimentos científicos, sendo necessários não somente para validação, mas também para o desenvolvimento de novos saberes. Entretanto, deve-se ter em conta que, como Procter, Halfpenny e Voss (2012) alertam, esses dados não terão valor até que sejam gerenciados de forma a assegurar sua recuperação, acessibilidade e reuso.

As universidades são agentes-chave nesse ambiente, pois estão diretamente envolvidas na produção de conhecimentos científicos. No mundo inteiro elas têm implantado repositórios institucionais para preservar, divulgar e aumentar a visibilidade de suas produções científicas. Segundo a ideia de Rodrigues e outros (2010) esses repositórios poderiam ser uma opção para resolver a situação relacionada ao gerenciamento de dados de pesquisa, pois existe um número crescente desse tipo de repositório, com um forte respaldo institucional.

O uso dos repositórios institucionais para a disponibilização e o compartilhamento dos dados de pesquisa pode ajudar as universidades a avançar em seus desafios, por ser um tipo de política tecnológica que inova e permite o desenvolvimento e o fortalecimento de redes de pesquisa e de projetos interdisciplinares (FRANCO; AFONSO; BORDIGNON, 2012).

A possibilidade de encontrar dados de diferentes áreas do conhecimento em um único repositório institucional e multidisciplinar pode favorecer a descoberta de oportunidades de colaboração científica dentro da própria instituição. O reuso dos dados de pesquisa torna-se então uma possibilidade concreta, pois sua disponibilização no repositório institucional amplia a visibilidade desses dados tanto interna quanto externamente. Nesse caso, colaborações internacionais também são facilitadas, e o tempo e recursos despendidos nas pesquisas diminuem.

Destaca-se na criação dos repositórios a perspectiva de aumentar a visibilidade e a contribuição das universidades para a ciência, através do aumento do impacto e a disseminação das pesquisas, garantindo o acesso e a preservação da memória da produção institucional. Como os dados de pesquisa constituem parte importante da produção intelectual das instituições e do processo de construção e comunicação do conhecimento científico (quando compartilhados), chega-se ao tema de pesquisa deste trabalho:

A articulação entre diretrizes e práticas de repositórios institucionais de dados de pesquisa (de universidades) no processo de gerenciamento de dados de pesquisa, no contexto da comunicação científica, através do estudo de um repositório desse tipo já em atividade.

1.1 Justificativa

No Brasil ainda existem poucos estudos relacionados a e-Science e ao próprio gerenciamento de dados de pesquisa. Costa e Cunha (2014) destacam que, de acordo com sua pesquisa em bases de dados bibliográficas, a contribuição da Biblioteconomia e da Ciência da Informação para essa área ainda está começando.

Esta pesquisa tem a intenção de contribuir para a exploração desse território ao se inserir no contexto dos repositórios institucionais de dados de pesquisa, buscando encontrar as articulações entre as diretrizes e as práticas dos repositórios no processo de gerenciamento de dados de pesquisa e as possíveis implicações para a comunicação científica.

Com o incremento das atividades de e-Science e o aumento da quantidade de dados de pesquisa em formato digital, torna-se relevante analisar as contribuições dos repositórios institucionais na curadoria digital dos dados de pesquisa, visto que esses repositórios têm por uma de suas finalidades a preservação e o aumento da visibilidade da produção intelectual de sua instituição.

Sabe-se que a ciência tem um papel relevante em muitas atividades da humanidade, mas para determinado conhecimento ser considerado científico ele necessita ser comunicado, avaliado e aceito pelos pares (demais pesquisadores). Além da comunicação dos trabalhos escritos, verifica-se a tendência de se compartilhar também os dados coletados pela pesquisa, para que outros pesquisadores possam reutilizá-los, possibilitando novas visões sobre esses dados já coletados ou ampliando esses dados através da combinação com outros conjuntos de dados. As maneiras de proporcionar o compartilhamento dos dados de pesquisa têm sido

diferentes. Nos últimos anos, tem crescido o uso de repositórios institucionais para a preservação e compartilhamento dos dados de pesquisa.

Repositórios institucionais têm sido objetos de diversas pesquisas, desde as que abordam os aspectos mais básicos de seu funcionamento e as questões políticas e sociais envolvidas, como as relacionadas ao acesso aberto, até as ligadas à preservação e ao compartilhamento da produção intelectual das instituições. Analisar contribuições desses repositórios na curadoria digital dos dados de pesquisa e as possibilidades que eles oferecem para a comunicação científica faz-se relevante por trazer uma reflexão sobre as práticas existentes e possibilitar mudanças ou novas visões e alternativas.

Desta forma, torna-se importante analisar os repositórios institucionais de dados de pesquisa, a fim de verificar quais estratégias estão sendo utilizadas por esses repositórios para permitir a divulgação, o acesso e o reuso desses dados, bem como vincular esses achados ao processo de gerenciamento de dados de pesquisa.

Tendo em vista as justificativas apresentadas, elaborou-se a seguinte questão de pesquisa:

Como as diretrizes e as práticas do repositório institucional de dados de pesquisa de uma Universidade se articulam em seu processo de gerenciamento de dados de pesquisa?

2 OBJETIVOS

Nas subseções a seguir estão delineados os objetivos gerais e específicos desta pesquisa.

2.1 Objetivo geral

Analisar as características (diretrizes e práticas) do Edinburgh DataShare, um repositório institucional universitário de dados de pesquisa, no contexto do gerenciamento de dados de pesquisa da University of Edinburgh.

2.2 Objetivos específicos

- a) Delinear as características do Edinburgh DataShare quanto a:
 - responsabilidade,
 - conteúdo,
 - aspectos legais,
 - padrões,
 - preservação digital,
 - política de acesso e uso e,
 - sustentabilidade e financiamento;
- b) Relacionar as diretrizes e as práticas (observadas a partir das características) com o processo de gerenciamento de dados de pesquisa da Universidade;
- c) Verificar como as diretrizes e as práticas se relacionam com a comunicação científica.

3 MARCOS TEÓRICOS

Os principais assuntos que servem de marcos teóricos relacionados à temática da pesquisa são: comunicação científica, repositórios institucionais, dados de pesquisa e sua importância na comunicação científica, curadoria digital e repositórios institucionais de dados de pesquisa.

3.1 Comunicação científica

Para que os resultados de uma pesquisa científica possam ser considerados parte da ciência, é necessário que eles sejam comunicados, debatidos, analisados, avaliados e validados pelos outros cientistas de um determinado campo científico.

Meadows (1999, p. 161) coloca que “a realização de pesquisas e a comunicação de seus resultados são atividades inseparáveis”. O mesmo autor comenta que no início de um projeto de pesquisa, a maioria da comunicação é informal, e com o avanço da pesquisa os meios vão tornando-se cada vez mais formais. Após a conclusão do projeto, normalmente os resultados são submetidos à avaliação pelos pares e à publicação.

Conforme Bourdieu (2008), o campo científico é um espaço de luta, no qual os pesquisadores concorrem dentro de posições e privilégios obtidos em lutas anteriores. O mesmo autor destaca que os motivos da luta são os seguintes:

O que está em luta são os monopólios da *autoridade científica* (capacidade técnica e poder social) e da *competência científica* (capacidade de falar e agir legitimamente, isto é, de maneira autorizada e com autoridade) que são socialmente outorgadas a um agente determinado. (BOURDIEU, 2008, p. 112, grifos do autor).

As representações da ciência, portanto, estão ligadas a determinados modos de ver e entender a ciência, bem como às posições que os agentes ocupam no campo científico. Ter autoridade e competência científica significa conseguir legitimação social no campo científico para pronunciar-se de maneira autorizada nesse espaço.

Apesar de não utilizar o conceito de campo científico, mas sim o de comunidade científica, Targino (2000) destaca a necessidade da comunicação científica para a atividade científica:

A comunicação científica é indispensável à atividade científica, pois permite somar os esforços individuais dos membros das comunidades científicas. Eles trocam continuamente informações com seus pares, emitindo-as para seus sucessores e/ou adquirindo-as de seus predecessores. (TARGINO, 2000, p. 10).

Os estudos da ciência têm como finalidade conhecer e entender as situações envolvidas no fazer científico a fim de fornecer explicações sobre o funcionamento da ciência.

Chalmers (1994) procura elaborar uma defesa restrita da ciência, apropriando-se de algumas críticas, porém tentando relativizar menos a ciência, buscando ver como é fabricado (construído) o conhecimento científico. Também destaca que a busca pelas metas da ciência relaciona-se com outras metas externas à ciência.

Essas metas abrangem necessidades econômicas, sociais e as possibilidades tecnológicas de cada espaço e momento. Em um mundo globalizado, as diferenças locais continuam existindo, porém torna-se possível ver que fatores mais amplos estão afetando a ciência, embora cada campo tenha suas especificidades.

Um fator muito importante para a ciência atual é o que Gray (2009) chama de quarto paradigma da ciência, a e-Science, também conhecida como ciência orientada aos dados.

Figura 1 – Paradigmas da ciência



Fonte: Medeiros e Caregnato (2012, p. 314), traduzido de Gray (2009, p. xviii)

A e-Science é um paradigma da ciência digital, que lida com dados originalmente digitais ou digitalizados. Cada vez mais se alimenta do aumento da capacidade de armazenamento e processamento dos computadores.

Medeiros e Caregnato (2012) notam que a literatura registra diferentes grafias para denominar a e-Science, bem como existem outras formas que são equivalentes, como *Cyberinfrastructure*, *cyberscience*, *eInfrastructure* e *eResearch*.

As mudanças relacionadas a e-Science e ao acesso aberto podem ser consideradas como consequências do contexto socioeconômico mundial, junto com o contexto tecnológico, e têm modificado a comunicação científica.

Quanto ao acesso aberto, tem-se que as bibliotecas e os pesquisadores estavam com dificuldade de manter as assinaturas dos periódicos científicos, cujo custo estava crescendo acima da possibilidade de pagamento. Alguns pesquisadores e instituições passaram a questionar essa situação, pois muitas das pesquisas eram financiadas com recursos públicos, e depois para acessar seus resultados, publicados nos periódicos, era necessário que mais recursos públicos fossem despendidos nas assinaturas. Assim, a ideia do acesso aberto (*open access*) às produções científicas começou a crescer e a modificar o fluxo das informações científicas.

Os repositórios institucionais tornaram-se uma das principais formas de incentivar o acesso aberto às produções científicas. Na seção seguinte são apresentadas as características dos repositórios institucionais, bem como seus objetivos e políticas.

3.2 Repositórios institucionais: conceituação, objetivos, políticas e diretrizes

Nos últimos anos, com a evolução dos computadores e as facilidades comunicacionais oferecidas pela disseminação da internet, várias mudanças vêm ocorrendo na comunicação científica. De acordo com Rosa e Gomes:

O desenvolvimento das TIC trouxe consigo as condições tecnológicas para o desenvolvimento de Repositórios Institucionais (RI) on-line. Criados a partir de 2002, os RI integram o novo fluxo estabelecido para a comunicação científica que tem no acesso livre o seu fundamento. (ROSA; GOMES, 2010, p. 24)

Conforme Leite (2009), os repositórios institucionais são os repositórios que tratam a informação científica ou acadêmica. Esse autor também considera que os “[...] repositórios institucionais constituem serviços de informação de valor agregado com forte potencial para o

atendimento dos imperativos de novos modos de produção, gestão e comunicação do conhecimento científico." (LEITE, 2009, p. 95). A partir dessa visão, o autor situa os repositórios institucionais como tendo um papel de “reformulação e melhoria do sistema de comunicação científica”.

Para definir de modo didático, Leite (2009) divide as abordagens para a gestão do repositório em abordagem rígida e em abordagem flexível. A abordagem rígida vê os repositórios institucionais como tendo a função de fortalecer a comunicação científica formal, ou seja, tem como foco o sistema de publicações científicas. Já a abordagem flexível amplia a abordagem rígida, pois prevê que, além das publicações avaliadas pelos pares, sejam disponibilizados conteúdos de natureza acadêmico-científica da instituição como, por exemplo, objetos de aprendizagem e dados científicos.

Apesar dessa classificação, o mais importante, segundo Rosa e Gomes (2010, p. 28) é que: "Seja o sistema de comunicação científica formal ou informal, o que importa, em termos sociais e de desenvolvimento científico, é o desenvolvimento da pesquisa e a transferência dos resultados”.

Um ponto de vista dos pesquisadores sobre os repositórios institucionais pode ser extraído da tese de Pavão:

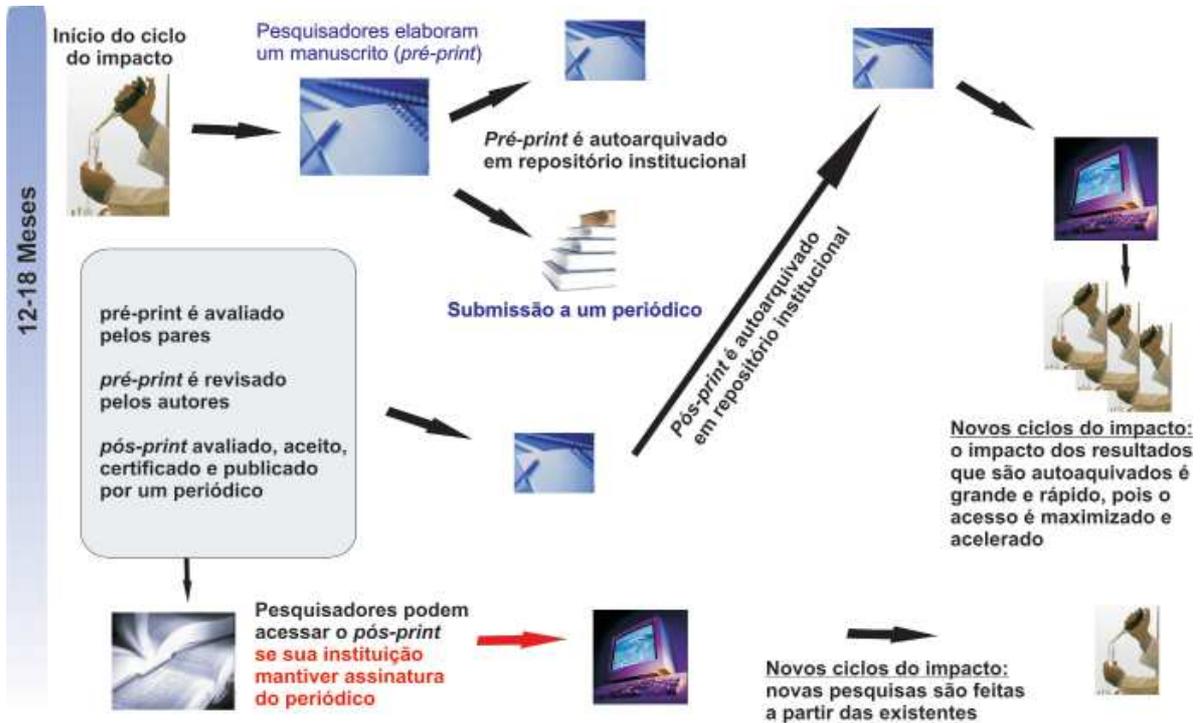
Os repositórios institucionais, como percebidos pelos entrevistados, contribuem para a melhoria do sistema de comunicação da ciência, e a visibilidade proporcionada por eles influenciará positivamente no avanço científico. Podem fornecer um componente importante para a reformulação do sistema de comunicação científica, expandem o acesso às pesquisas e podem reduzir o monopólio dos periódicos. (PAVÃO, 2010, p. 117)

A visibilidade trazida pelos repositórios é citada por diversos autores (LEITE, 2009; PAVÃO *et al.* 2012; ROSA; GOMES, 2010), como auxiliar no aumento do impacto dos resultados da pesquisa e da própria instituição, além do pesquisador.

Brody e Harnad apresentaram em 2004 um modelo da comunicação científica com o impacto da pesquisa maximizado pela disponibilização em acesso aberto através do auto-arquivamento. Esse modelo não é detalhado com relação aos tipos de documento, pois seu enfoque é na dinâmica da comunicação científica relacionada com o tempo. Mais exatamente, Brody e Harnad procuraram apresentar como o acesso e o impacto da pesquisa podem ser aumentados pelo auto-arquivamento das produções científicas, e o exemplo utilizado foi o de artigos de periódicos, por sua grande importância para a comunicação científica de diversas

áreas, mas nada impede que esse modelo seja aplicado a outros tipos de documento, como capítulos de livro. Leite (2009) fez uma tradução desse modelo:

Figura 2 – Modelo de Brody e Harnad: acesso e impacto da pesquisa maximizados por meio do autoarquivamento



Fonte: Brody e Harnad (2004)¹, *apud* Leite (2009, p. 15)

O papel dos repositórios institucionais na comunicação científica é ressaltado nesse modelo: eles são responsáveis pelo armazenamento e disseminação em acesso aberto de pré-prints e pós-prints dos trabalhos publicados em periódicos científicos. Os repositórios cumprem essa função de complementar e maximizar o impacto das pesquisas, facilitando o acesso aos resultados de pesquisa.

Leite (2009) também destaca a importância dos repositórios institucionais nos processos de organização, recuperação e disseminação da informação:

Repositórios institucionais proporcionam as condições necessárias à adequada gestão da informação científica. Além de permitir a organização e recuperação da informação, como em sistemas de informação em ciência e tecnologia convencionais, enfatizam os processos de disseminação da informação. (LEITE, 2009, p. 97)

¹ BRODY, Tim; HARNAD, Stevan. **The Research-Impact Cycle**. 2004. Disponível em: <www.eprints.org/jan2005/ppts/harnad.ppt>. Acesso em: 15 ago. 2013.

Enfim, a tecnologia vinculada às necessidades surgidas no próprio campo científico veio a contribuir com mudanças nas práticas de comunicação científica. O modelo de Brody e Harnad apresenta que o tempo para determinado trabalho produzir impacto é mais rápido do que o da comunicação científica alterada apenas pela internet, por oferecer o benefício do acesso aberto aos pré-prints e pós-prints.

É importante ressaltar a seguinte declaração de Kuramoto (2009, p. 203) sobre os repositórios institucionais no contexto mundial: “Para os países, a construção de uma rede de RI integrando a sua produção científica nacional, significa dar maior transparência aos investimentos em ciência, assim como maior governabilidade.”.

3.3 Dados de pesquisa: definição e sua importância para a comunicação científica

As seguintes definições de conjuntos de dados, dados científicos e curadoria de dados foram retiradas do relatório “Os repositórios de dados científicos: estado da arte”, do Projeto Repositório Científico de Acesso Aberto de Portugal (RODRIGUES *et al.*, 2010):

- a) conjunto de dados (dataset): um conjunto de dados, ou conjunto de dados científicos, reúne itens de natureza semelhante que foram colecionados para servirem de base a investigação;
- b) curadoria de dados (data curation): a curadoria de dados designa o conjunto de ações que garantem que um conjunto de dados é genuíno, permitindo o seu uso por outros que não os seus produtores;
- c) dados científicos (research data): dados que são produzidos no contexto de investigação científica ou que de alguma forma são usados em investigação.

Retomando a definição da Organização para a Cooperação Econômica e Desenvolvimento (OCDE), tem-se que os:

[...] “dados de pesquisa” são definidos como registros factuais (números, registros textuais, imagens e sons) utilizados como fontes primárias para a pesquisa científica, e que são geralmente aceitos na comunidade científica como necessários para validar os resultados da pesquisa. (OCDE, 2007, p.13, tradução nossa)

O que há em comum entre as definições de dados de pesquisa é que eles são gerados ou trabalhados no âmbito de uma pesquisa, porém eles podem ser dos mais variados tipos. Conforme relatório da Research Information Network (2008), os tipos de dados de pesquisa

quanto a seus propósitos e processos são: experimentos científicos, modelos ou simulações, observações, dados derivados e dados canônicos ou de referência.

São destacadas a seguir as definições de cada um desses tipos de dados de pesquisa:

- experimentos científicos, que em princípio podem ser reproduzidos, apesar de que pode ser difícil reproduzi-los na prática, ou ainda podem não serem eficazes em termos de custos para serem reproduzidos;
- modelos ou simulações, para os quais pode ser mais importante preservar o modelo e os metadados associados que os dados computacionais provenientes do modelo;
- observações - das astronômicas às zoológicas - de fenômenos específicos em um lugar ou tempo específicos, no qual dos dados normalmente constituem um registro único e insubstituível;
- dados derivados, resultantes do processamento ou combinação de dados primários ou outros dados (onde se deve ter cuidado sobre os direitos dos proprietários dos dados primários)
- dados canônicos ou de referência relacionados, por exemplo, a sequências de genes, estruturas químicas ou textos literários. (RESEARCH INFORMATION NETWORK, 2008, p. 5, tradução nossa)

Sobre os dados derivados, Cragin *et al.* (2010 p. 4029, tradução nossa) sustentam: “[...] conjuntos de dados compostos também são comuns; este tipo de conjunto de dados é derivado de muitas fontes de dados, integradas e analisadas para produzir um conjunto de dados novo e complexo.”

A partir dessa variedade de tipos de dados de pesquisa, há a compreensão que o tratamento a ser dado a esses tipos de materiais exige muita atenção e estudo contínuo, pois cada um desses tipos pode se apresentar de muitas maneiras, em diferentes formatos de arquivo. Borgman (2012, p. 1059) destaca que “[...] os dados de pesquisa têm tantas formas, são manejados de tantas maneiras, usando diferentes abordagens e frequentemente são difíceis de interpretar quando são removidos de seu contexto inicial”.

O objetivo maior atualmente quanto aos dados de pesquisa refere-se à necessidade de aproveitar melhor esses recursos, através de seu reuso por outros pesquisadores, que não tenham necessariamente participado da coleta/criação dos dados originais. É uma questão de economia de tempo e recursos financeiros, bem como é um melhor aproveitamento desse produto da atividade científica.

O compartilhamento e o reuso dos dados de pesquisa existem há tempos, mas normalmente ocorrem através do fornecimento direto dos dados de um pesquisador a outro, após contato pessoal. O uso dos repositórios para a disponibilização dos dados de pesquisa permite ampliar o escopo de distribuição, visibilidade e possibilidades de reuso desses dados.

Essa nova situação causa dúvidas nos pesquisadores, principalmente com relação a problemas na interpretação dos dados por outras pessoas e o mau uso dos dados, tanto por

engano comum ou por má fé, bem como questões relacionadas ao reconhecimento pela criação dos dados. Cragin e outros (2010) reconhecem que ainda há muito o que aprender sobre as condições para compartilhamento e os efeitos causados pelos incidentes de "mau uso" dos dados, bem como a necessidade de que sejam realizadas pesquisas que criem uma taxonomia dos tipos de mau uso, a fim de facilitar a tomada de decisão sobre as políticas institucionais de tratamento dos dados e serviços para comunicação científica.

Borgman (2012) destaca que os dados têm mais chance de serem compartilhados quando as políticas beneficiam os produtores dos dados. A autora também apresenta quatro razões para o compartilhamento dos dados de pesquisa:

- a) reproduzir ou verificar a pesquisa;
- b) tornar disponível para o público os resultados de pesquisas financiadas com verbas públicas;
- c) proporcionar a outros pesquisadores fazer novos questionamentos sobre os dados;
- d) avançar a pesquisa e a inovação.

Atualmente, no exterior, a pressão para compartilhar os dados vem tanto das agências de financiamento, quanto das entidades representativas, das editoras dos periódicos, educadores, do público em geral e dos próprios pesquisadores (BORGMAN, 2012). No Brasil essa questão ainda é incipiente, os trabalhos de Luis Sayão e Luana Sales, da Comissão Nacional de Energia Nuclear têm trazido essas questões relacionadas com os dados de pesquisa e a curadoria digital, a qual consiste em um tratamento complexo desses dados visando sua preservação a longo prazo, com qualidade na recuperação, identificação e uso dos conjuntos de dados de pesquisa.

Ainda é possível discutir se os dados de pesquisa constituem ou não uma publicação, tendo em vista que estão surgindo periódicos especializados na publicação de dados de pesquisa (como exemplos, tem-se o "Dataset Papers in Science", publicado pela editora Hindawi, e o "Geoscience Data Journal", publicado pela Wiley). Em alguns momentos, os dados de pesquisa inclusive podem ser agregados aos e-prints (artigos, trabalhos de evento, teses, dissertações etc.) para os quais foram o material primário de pesquisa, assim constituindo publicações ampliadas (SALES, SAYÃO, SOUZA, 2013).

Apesar da discussão em torno da questão da publicação, não é possível discordar do fato de que a organização de conjuntos de dados de pesquisa é uma produção dos pesquisadores, tendo em vista que a definição de determinada observação, artefato ou registro

como dado de pesquisa ocorre através do olhar de um pesquisador (BORGMAN, 2012), e é de fato uma transformação, um trabalho intelectual inerente ao trabalho da pesquisa.

3.4 Curadoria digital

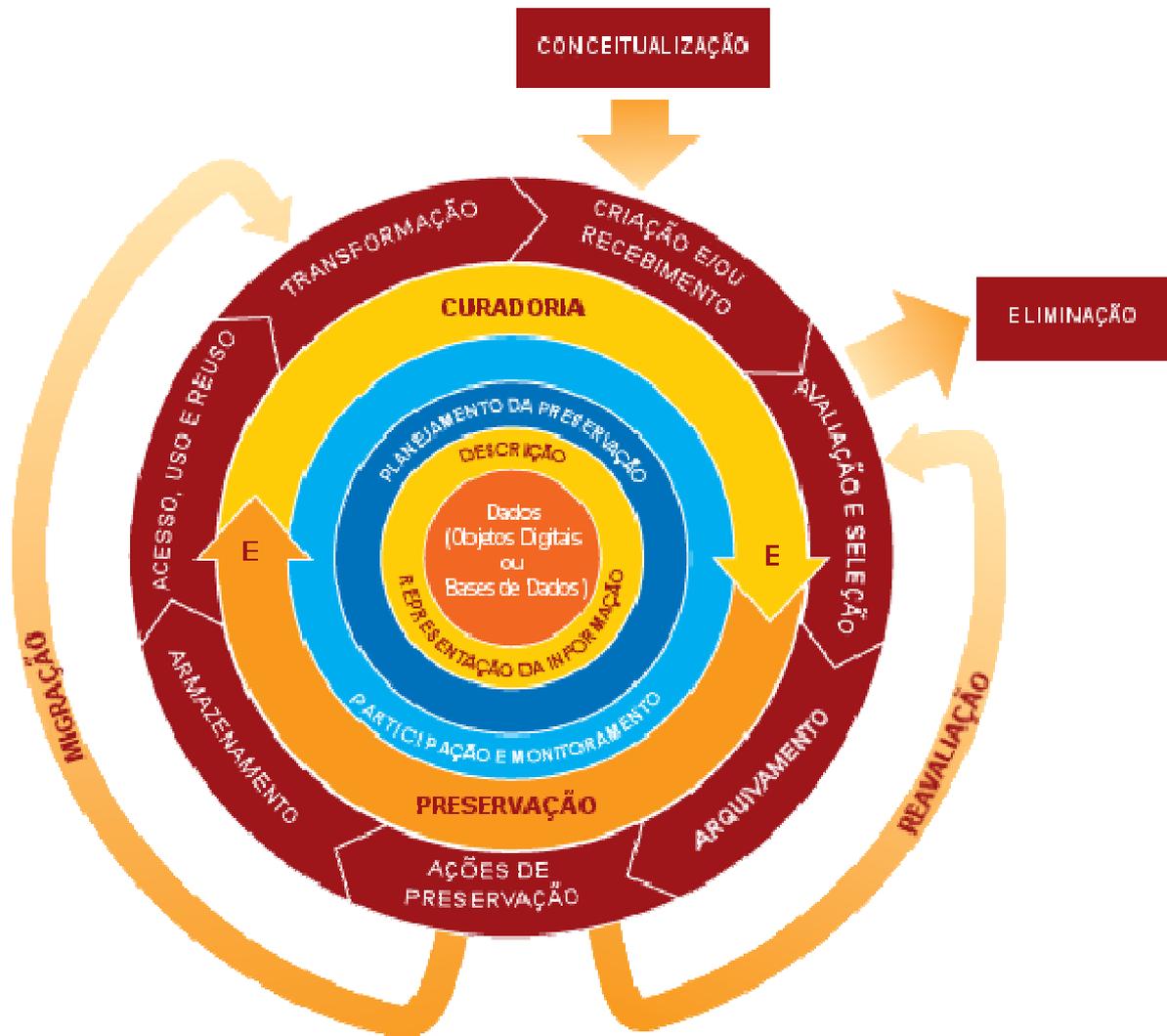
A curadoria digital ultrapassa o âmbito dos repositórios: compreende todo um amplo espectro de ações referentes ao tratamento dos dados de pesquisa, desde sua concepção, dentro do planejamento da pesquisa, passando pela coleta, tratamento e armazenamento desses dados nos repositórios, o compartilhamento e o reuso desses dados, bem como seu eventual descarte. Aspectos éticos, legais e técnicos estão relacionados com a curadoria digital.

Por esse motivo, é importante ter uma visão global desse processo. O modelo do ciclo de vida da curadoria digital do Digital Curation Centre (DCC) traz essa visão. Ele admite que esse modelo é um ideal, que provavelmente na vida real as atividades não acontecem exatamente dessa forma, mas esse ideal ajuda no planejamento e na correção dos problemas que são encontrados no decorrer do processo de curadoria digital.

Higgins (2011) comenta que a curadoria digital surgiu como uma nova disciplina em um processo interativo de discussão e estabelecimento de metas. Conforme a autora, no Reino Unido a ênfase estratégica modificou-se gradualmente de uma preservação passiva para uma curadoria pró-ativa. O Digital Curation Centre está comprometido em promover sua abordagem baseada no ciclo de vida do gerenciamento de materiais digitais a fim de que a curadoria e a preservação a longo termo tenham sucesso (HIGGINS, 2008).

O modelo do ciclo de vida da curadoria digital do DCC é apresentado em forma de gráfico no qual podem ser vistas diferentes ações, que podem ser utilizadas no planejamento das atividades em nível organizacional. Ele também identifica ações para todo o ciclo de vida, ações seqüenciais e ações ocasionais, que devem acontecer somente em casos específicos.

Figura 3 – Modelo do ciclo de vida da curadoria digital



Fonte: Adaptado de Digital Curation Centre <<http://www.dcc.ac.uk>>, pela autora, utilizando a tradução de Sayão e Sales (2012).

Abaixo são apresentadas ações envolvidas no modelo, segundo Sayão e Sales (2012):

Ações para todo o ciclo de vida:

- **Descrição e a representação da informação** - é efetivada pela atribuição de metadados administrativos, técnicos, estruturais e de representação de acordo com os padrões apropriados; visa assegurar a descrição adequada e o controle de longo prazo; compreende também a coleta e a atribuição de informações de representação necessárias para o entendimento do dado e para a sua apresentação (ou renderização).
- **Planejamento da Preservação** – é necessária a definição de um plano de preservação cujo espectro englobe todo o ciclo de vida da curadoria do material digital, incluindo gestão, administração, políticas, e tecnologias.
- **Participação e monitoramento** – enfatiza a necessidade de atenção para as atividades que se desenrolam no âmbito das comunidades envolvidas com o problema de curadoria, bem como a necessidade de participação no desenvolvimento de padrões, de ferramentas e de software adequados ao problema e que possam também serem compartilhados;

- **Curadoria e preservação** - estar continuamente alerta e empreender as ações administrativas e gerenciais planejadas para a curadoria e preservação por todo o ciclo de vida da curadoria. (SAYÃO e SALES, 2012 p. 185-186).

Ações sequenciais:

- **Conceitualização** – conceber e planejar a criação do dado, incluindo os métodos de captura e as opções de armazenamento; questões tais como propriedade intelectual, embargos e restrições, financiamento, responsabilidades, objetivos específicos da pesquisa, ferramentas de captura e calibração devem ser registradas.
- **Criação e/ou Recebimento** – compreende a criação do dado incluindo o elenco de metadados necessários à sua gestão e compreensão, ou seja, metadados administrativos, descritivos, estruturais e técnicos (os metadados de preservação também podem ser incluídos no momento da criação do dado). Nem sempre os dados são arquivados por quem os gerou, dessa forma, esse estágio inclui também a recepção dos dados segundo políticas bem documentadas, sejam dos seus criadores, de outros arquivos, de repositórios ou centro de dados; quando necessário, assinalar metadados apropriados para a curadoria e a preservação dos dados recebidos.
- **Avaliação e seleção** – avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação por longo prazo; manter-se aderente tanto às boas práticas quanto às políticas pertinentes e também às exigências legais.
- **Arquivamento** – transferir o dado para um arquivo, repositório, centro de dados ou outro custodiante apropriado.
- **Ações de preservação** – promover ações para assegurar a preservação de longo prazo e a retenção do dado de natureza oficial; as ações de preservação devem assegurar que o dado permaneça autêntico, confiável e capaz de ser usado enquanto mantém sua integridade; essas ações de preservação incluem: a limpeza do dado e a sua validação, a adição de metadados de preservação e de informação de representação e a garantia de estruturas de dados ou formatos de arquivos aceitáveis.
- **Armazenamento** – armazenar o dado de forma segura mantendo a aderência aos padrões relevantes.
- **Acesso, uso e reuso** – garantir que o dado possa ser cotidianamente acessado tanto pela sua comunidade-alvo, quanto pelos demais usuários interessados no reuso do dado; isto pode ser realizado por meio de publicação disponível para as comunidades interessadas; porém, controle de acesso e procedimentos de autenticação podem ser aplicados.
- **Transformação** – compreende a criação de novos dados a partir do original, por exemplo, pelo processo de migração para diferentes formatos ou pela criação de subconjuntos - realizada por meio de seleção ou formulação de consultas – derivando novos resultados que podem ser publicados. (SAYÃO e SALES, 2012 p. 186).

Ações ocasionais:

- **Eliminação** – eliminar os dados que não foram selecionados para curadoria e preservação de longo prazo de acordo com políticas documentadas, diretrizes ou exigências legais.
- **Reavaliação** – retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção para curadoria.
- **Migração** – migrar os dados para um formato diferente; isto pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado contra a obsolescência de hardware e de software. (SAYÃO e SALES, 2012, p. 186-187).

Todas essas ações são importantes por permitirem uma visão geral do processo de curadoria. Apesar de muitas vezes o profissional da informação só encontrar com o pesquisador a partir da fase de arquivamento, é muito mais proveitoso se os pesquisadores e

os profissionais da informação comecem a trabalhar juntos no início do ciclo de pesquisa. Isso permite que se identifiquem formas de trabalho e curadoria dos dados que possam auxiliar a identificar os metadados pertinentes e outras questões que facilitem o depósito dos dados antes do fim do projeto de pesquisa, melhorando a qualidade dos dados a serem depositados e reduzindo os custos na fase de arquivamento (CRAGIN *et al.*, 2010).

3.5 Repositórios institucionais de dados de pesquisa

É interessante destacar que a curadoria de dados exige um trabalho em parceria entre os pesquisadores e os prestadores de serviço de curadoria, muitas vezes o responsável pelo repositório. Para tanto, é necessário que a equipe esteja devidamente capacitada. Leite (2009), ao comentar sobre equipes de repositórios institucionais em geral, destaca que é importante:

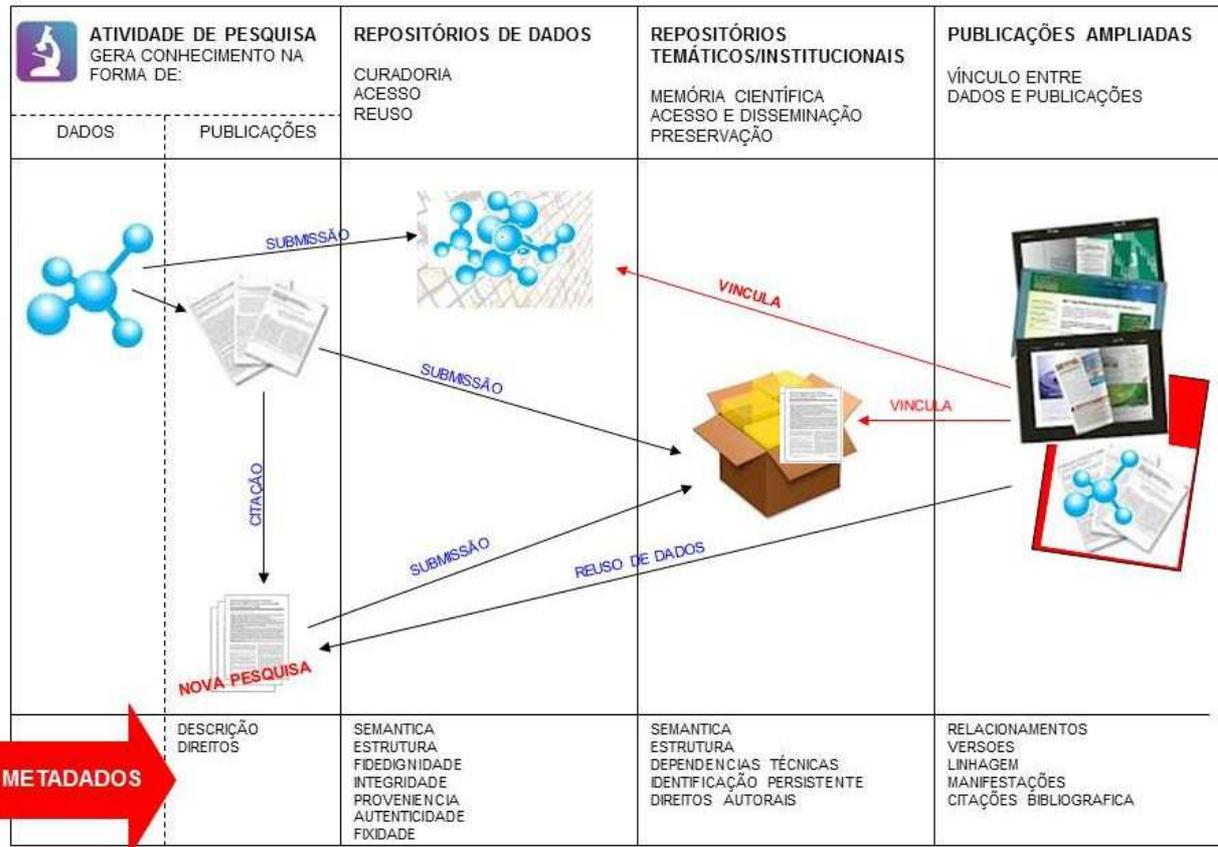
[...] além da internalização dos conceitos fundamentais, os profissionais estejam aptos a compreender o contexto da comunicação científica e sua lógica, atores, componentes, processos e as forças que a governam, assim como as propriedades da informação e do conhecimento científico, o papel das tecnologias de informação e o funcionamento das comunidades científicas e suas diferenças disciplinares na produção e comunicação do conhecimento. (LEITE, 2009, p. 41)

Esse fator continua sendo relevante também para o tratamento dos dados científicos, ainda mais porque as diferenças disciplinares ficam mais evidentes nos dados do que já transparece nas publicações tradicionais. O relatório da RCAAP (RODRIGUES *et al.*, 2010, p. 35) ressalta que “o uso de normas adequadas à curadoria dos dados científicos, garantindo a sua preservação e as condições para a sua reutilização é uma tarefa que requer conhecimentos técnicos e disciplinares”. Percebe-se que um dos principais desafios da utilização de repositórios institucionais na curadoria de dados científicos é a integração entre a dimensão multidisciplinar da instituição e as necessidades específicas de cada disciplina para os seus dados científicos.

Além disso, há a constatação de que os repositórios institucionais serão subutilizados como recursos para os dados científicos se eles não suportarem as práticas de dados já existentes e garantirem proteção contra o uso incorreto dos dados. (CRAGIN *et al.*, 2010)

Sales e Sayão (2012) notam que a curadoria dos dados de pesquisa tem trazido mudanças também para a comunicação científica, e propõem o seguinte modelo para uma possível alteração no ciclo da informação científica:

Figura 4 – Modelo de Sales e Sayão: modelo tradicional alterado pelo reuso de dados científicos



Fonte: Sales e Sayão (2012, p. 131)

Este modelo destaca as possibilidades de vinculação entre os dados de pesquisa e o trabalho final (publicação) relacionada a esses dados, criando o que se conhece como publicação ampliada. Os repositórios, tanto institucionais quanto de dados de pesquisa têm um papel preponderante nesse modelo, pois permitem o acesso, o reuso e garantem a integridade e a preservação da informação científica.

É relevante ressaltar o reuso dos dados de pesquisa e a citação aos trabalhos como componente do processo de comunicação científica. Outro diferencial do modelo de Sales e Sayão é a coluna inferior que descreve condições dos metadados como componentes importantes nesse ciclo.

Os repositórios institucionais podem facilitar o gerenciamento dos dados de pesquisa. Conforme Cragin e outros (2010, p. 4036), “dentro das universidades os repositórios também fazem parte de uma estrutura de serviços e de tecnologia maior, que pode contribuir ao desenvolvimento de um serviço de curadoria digital”.

Hernández-Pérez e García-Moreno (2013) destacam a importância dos repositórios de dados abertos de pesquisa para a transparência da pesquisa e também para uma economia de

tempo e recursos, bem como para uma retribuição à sociedade dos investimentos na ciência, mas que os problemas relacionados ao compartilhamento dos dados dessa forma são relacionados à conscientização dos pesquisadores, ao planejamento da instituição e à formação dos pesquisadores e dos profissionais da informação para trabalharem com o ciclo de vida dos dados de pesquisa.

Arano e outros (2011) destacam que os repositórios institucionais são uma excelente oportunidade para disponibilizar os dados de pesquisa, proporcionando um tratamento padronizado e manuais de boas práticas para esses materiais.

3.6 Políticas de Informação para repositórios: aspectos a serem observados nas diretrizes e nas práticas

As políticas de informação têm um papel fundamental na escolha dos materiais que constituem um repositório. Definem diretrizes que norteiam as boas práticas, as quais trarão maior consistência e continuidade ao repositório. Tomaél e Silva (2007) organizaram as políticas de informação para um repositório institucional em sete categorias, contemplando aspectos que devem ser levados em conta no planejamento e no desenvolvimento de um desenvolvimento de um repositório institucional. São elas:

- a) *Responsabilidade*: Profissionais responsáveis pela gestão do repositório. Como a instituição se estrutura para organizar o repositório;
- b) *Conteúdo*: Conteúdos permitidos no repositório, os tipos de formato de arquivo aceitos e características dos criadores e dos assuntos dos itens;
- c) *Aspectos legais*: Licenças utilizadas para a disponibilização dos dados de pesquisa;
- d) *Padrões*: Padrões tecnológicos, de metadados, de fluxos de trabalho e de submissão, de interface e de interoperabilidade;
- e) *Preservação digital*: Existência de plano de preservação digital e aspectos contemplados nesse plano, e uso de identificadores digitais;
- f) *Política de acesso e uso*: Permissões de acesso e uso aos dados de pesquisa depositados. Suporte ao pesquisador que submete e ao que usa os dados de pesquisa do repositório;
- g) *Sustentabilidade e financiamento*: Estratégias de financiamento e sustentabilidade do repositório, principalmente visando os aspectos de equipe e infraestrutura.

Essas categorias permitem visualizar, de um modo geral, as principais áreas para as quais as políticas devem ser definidas, lembrando sempre que a realidade da instituição deve ser priorizada.

Ao adicionar a complexidade do tratamento dos dados de pesquisa aos repositórios institucionais, abre-se todo um campo de trabalho, porém as categorias básicas ajudam a ter um ponto de partida estruturado para analisar e organizar essa informação. Assim, as sete categorias de Tomaél e Silva (2007) são a base para a análise deste estudo das diretrizes e práticas de um repositório institucional de dados de pesquisa e suas políticas de informação no processo de gerenciamento desses dados.

4 PROCEDIMENTOS METODOLÓGICOS

Este trabalho caracteriza-se como um estudo de caso de um repositório institucional de dados de pesquisa.

Estudos de caso são escolhidos como método de pesquisa quando há uma necessidade de entender fenômenos sociais complexos, de acordo com Yin (2010). No caso desta pesquisa isso se aplica, pois repositórios não são apenas objetos tecnológicos, mas sim um fenômeno resultante de diferentes atividades humanas complexas para sua efetivação. Além disso, esta pesquisa trabalha com a questão **“Como as diretrizes e as práticas do repositório institucional de dados de pesquisa de uma Universidade se articulam em seu processo de gerenciamento de dados de pesquisa?”**, que busca apreender um fenômeno atual, e a pesquisadora não tem controle sobre o repositório que está sendo estudado. Yin (2010) afirma que a escolha do estudo de caso como método de pesquisa é vantajosa quando a questão de pesquisa é do tipo “como” ou “por que” e está sendo feita sobre eventos contemporâneos e algo que o pesquisador tem pouco ou nenhum controle.

Para analisar repositórios de dados de pesquisa, é necessário obter uma amostra dos mesmos e, neste caso, por ser um objeto de estudo recente, é possível e até desejável selecionar os repositórios intencionalmente, dentre um conjunto de repositórios, como Fragoso, Recuero e Amaral (2011) afirmam, a respeito da escolha do tipo de amostragem.

As escolhas efetuadas para esta proposta seguem a ideia de que o método consiste em “[...] um conjunto de decisões e opções particulares que são feitas ao longo de um processo de investigação.” (LOPES, 2010, p. 94). Dessa forma, as escolhas estão especificadas visando facilitar a visualização do caminho percorrido na elaboração do estudo, mas destaca-se que a dinâmica do processo de pesquisa compreende idas e voltas, interações, combinações e recombinações lógicas das etapas deste caminho (LOPES, 2010).

4.1 Procedimentos de escolha do objeto de estudo

Para a seleção do objeto de estudo (repositório institucional de universidades que contem dados de pesquisa), foi pensada a utilização de diretórios de repositórios.

Muitos estudos utilizam bases de dados e diretórios on-line para selecionar e identificar seus objetos de pesquisa. Esta pesquisa utiliza três deles: OpenDOAR, Databib e re3data.

Oliver e Swain (2006) concluíram que investimentos no OpenDOAR e em diretórios semelhantes trariam benefícios para a ciência, pois tais ferramentas poderiam facilitar a observação e análise do desenvolvimento das atividades científicas. Luzi e outros (2012) realizaram uma pesquisa sobre os dados de pesquisa em acesso aberto, utilizando o diretório OpenDOAR.

Hernández-Pérez e García-Moreno (2013), em seu trabalho sobre dados abertos e repositórios de dados citaram o Databib e o re3data.org como os catálogos de repositórios mais importantes neste momento.

Como Pampel e outros (2013) destacam, os repositórios de dados de pesquisa institucionais de universidades costumam ser multidisciplinares, por abordarem diferentes áreas do conhecimento. Repositórios disciplinares costumam ter ferramentas mais específicas para sua área do conhecimento, e por isso sua abordagem de gerenciamento dos dados costuma ser diferente dos repositórios multidisciplinares.

O repositório foi escolhido através dos seguintes procedimentos:

- a) Busca nos diretórios OpenDOAR, Databib e re3data.org pelos seguintes critérios: repositórios institucionais de universidades e multidisciplinares e que contenham dados de pesquisa (research datasets);
- b) Visita aos repositórios encontrados na busca e seleção daqueles que contêm documentação sobre dados de pesquisa e nos quais se podem identificar facilmente a presença de dados científicos.

A seguir, apresentam-se as características básicas de cada um dos diretórios pesquisados, bem como os resultados das buscas neles realizadas.

OpenDOAR: lançado em 2005, foi resultado de um trabalho conjunto entre a University of Nottingham (Reino Unido) e a Lund University (Suécia), como parte do projeto SHERPA. Atualmente está sob a responsabilidade da University of Nottingham, e é um diretório que contém dados sobre repositórios de acesso aberto do mundo inteiro, inclusive repositórios que contém conjuntos de dados de pesquisa (*datasets*). A equipe do OpenDOAR visita, analisa e atribui metadados a esses repositórios, a fim de obter categorizações e análises que permitem um amplo uso dessas informações (OPENDOAR, 2011).

Em 21 de janeiro de 2014, o OpenDOAR tinha 2563 repositórios cadastrados.

O diretório OpenDOAR foi pesquisado usando os seguintes termos:

- a) No campo de Pesquisa Geral: university
- b) No campo Subject Área: multidisciplinary
- c) No campo Content type: datasets
- d) No campo Repository type: institutional

O resultado foi de 42 repositórios compatíveis com os critérios de busca, conforme Quadro 1. Os repositórios listados no OpenDOAR foram visitados, e se destacaram cinco que tinham mais informações sobre o repositório e sobre o gerenciamento dos dados de pesquisa: DSpace @ Cambridge, Edinburgh DataShare, Open Research Exeter (ORE), RUCore (Rutgers University Community Repository) e ScholarSphere.

Quadro 1 – Repositórios de dados de pesquisa recuperados no OpenDOAR

Diretório	Instituição Responsável
1. Academic Commons	Columbia University Libraries Columbia University (Estados Unidos)
2. Adelaide Research & Scholarship (AR&S)	University of Adelaide (Austrália)
3. Chiba University's Repository for Access To Outcomes from Research (CURATOR)	Chiba University (Japão)
4. Claremont Colleges Digital Library (CCDL)	Claremont University Consortium (Estados Unidos)
5. Dépôt Institutionnel Numérique (PAPYRUS)	Université de Montreal (Canadá)
6. Digital Case	Kelvin Smith Library Case Western Reserve University (Case) (Estados Unidos)
7. Digital Collections Repository	Alber B. Alkek Library Texas State University - San Marcos (TxState) (Estados Unidos)
8. DigiTool@Florida Atlantic University	Florida Atlantic University (FAU) (Estados Unidos)
9. DiscoverArchive	Jean and Alexander Heard Library Vanderbilt University (Estados Unidos)
10. DSpace @ Cambridge	Cambridge University Library and Computing Service University of Cambridge (Reino Unido)
11. DSpace at University of Calgary	University of Calgary (Canadá)
12. Edinburgh DataShare	Data Library University of Edinburgh (Reino Unido)
13. Edinburgh Research Archive (ERA)	Edinburgh University Library University of Edinburgh (Reino Unido)

14. Education and Research Archive (ERA)	University of Alberta (Canadá)
15. Institutional Repository of Tsinghua University	Tsinghua University (China)
16. ISU ReD: Research and eData	Illinois State University (Estados Unidos)
17. Kanazawa University Repository for Academic Resources (KURA)	Kanazawa University (Japão)
18. KU ScholarWorks	Information Services University of Kansas (KU) (Estados Unidos)
19. Leicester Research Archive (LRA)	University of Leicester Library University of Leicester (Reino Unido)
20. Libra	University of Virginia (Estados Unidos)
21. Loyola eCommons	Loyola University Chicago (Estados Unidos)
22. Minds @ UNiversity of Wisconsin	University of Wisconsin Systems (UW) (Estados Unidos)
23. Monash University ARROW Repository	Monash University Library Monash University (Austrália)
24. Obihiro University of Agriculture and Veterinary Medicine Academic Repository	Obihiro University of Agriculture and Veterinary Medicine (Japão)
25. Odum Institute Data Archive	Odum Institute at University of North Carolina (Estados Unidos)
26. Open Research Exeter (O R E)	University of Exeter University of Exeter (Reino Unido)
27. ORBIT (Online Research Database In Technology)	DTU (Technical University of Denmark) (Dinamarca)
28. QSpace at Queen's University	Queen's University Library Queen's University (Canadá)
29. Repositorio Institucional del ITESO (ReI)	Iteso, A.C. Instituto Tecnológico de Estudios Superiores de Occidente (ITESO) Jesuit University of Guadalajara (México)
30. RUCore (Rutgers University Community Repository)	Rutgers University Libraries Rutgers University (Estados Unidos)
31. ScholarSphere	Information Technology Services Pennsylvania State University (Penn State) (Estados Unidos)
32. SUCRA (Saitama University Cyber Repository of Academic Resources)	Saitama University (Japão)
33. Sussex Research Online	University Library University of Sussex (Reino Unido)
34. University of Delaware Library Institutional Repository	University of Delaware Library University of Delaware (UD) (Estados Unidos)
35. University of Hull Institutional Repository (Hydra)	University of Hull (Reino Unido)
36. University of Melbourne Digital Repository (UMER)	Academic Registrar University of Melbourne (Austrália)
37. UNSWorks	Library University of New South Wales (Austrália)
38. UPF Digital Repository	Universitat Pompeu Fabra (UPF) (Espanha)

39. UPSpace (UPSpace at the University of Pretoria)	University of Pretoria - Department of Library Services University of Pretoria (UP) (África do Sul)
40. UQ eSpace (University of Queensland eSpace)	University of Queensland (UQ) (Austrália)
41. UR Research (University of Rochester Digital Repository)	Rush Rhees Library University of Rochester (Estados Unidos)
42. USPERR (University of the South Pacific Electronic Research Repository)	University of the South Pacific (USP) (Fiji)

Fonte: OpenDOAR <<http://www.opendoar.org/>>, em 21 jan. 2014.

Databib: ferramenta criada para ajudar as pessoas a identificar e localizar repositórios online de dados de pesquisa. É mais específico que o OpenDOAR, pois seu foco é em repositórios de dados de pesquisa, e outra diferença é que ele aceita registrar também repositórios que não sejam de acesso aberto. A descrição dos repositórios é feita tanto por bibliógrafos quanto pelos usuários do diretório Databib. O diretório é hospedado pelas Bibliotecas da Purdue University, dos Estados Unidos, mas seus conselhos consultivo e editorial têm integrantes vinculados a diversas universidades do mundo (DATABIB, 2012, [2013]).

Em 21 de janeiro de 2014, o diretório Databib tinha 602 repositórios cadastrados.

O diretório Databib foi pesquisado usando os seguintes termos:

No campo All fields (Pesquisa Geral): multidisciplinary university

O resultado foi de oito repositórios compatíveis com os critérios de busca, conforme Quadro 2.

Quadro 2 – Repositórios de dados de pesquisa recuperados no Databib

Diretório	Instituição Responsável
1. 3TU.Datacentrum	3TU.Federation of Technical Universities of the Netherlands (Holanda)
2. DataBox	The Royal Library (Dinamarca)
3. DataStaR	Albert R. Mann Library, Cornell University (Estados Unidos)
4. Edinburgh DataShare	University of Edinburgh (Reino Unido)
5. Griffith University Research Data Repository	Griffith University Research Data Repository (Austrália)
6. Purdue University Research Repository	Purdue University (Estados Unidos)
7. Scholarly Database (SDB)	Cyberinfrastructure for Network Science Center, SLIS Indiana University (Estados Unidos)
8. ScholarSphere	Penn State University Libraries and Information Technology Services (Estados Unidos)

Fonte: Databib <<http://databib.org/>>, em 21 jan. 2014.

re3data.org: teve início em 2012, e os parceiros do projeto são o departamento de Library and Information Services (LIS) do GFZ German Research Centre for Geosciences, a Berlin School of Library and Information Science na Humboldt-Universität zu Berlin e a KIT Library do Karlsruhe Institute of Technology (KIT). O objetivo do re3data.org é criar um registro mundial dos repositórios de dados de pesquisa, oferecendo orientação aos pesquisadores na área de dados de pesquisa, pensando neles tanto como produtores quanto usuários de dados (PAMPEL *et al.*, 2013).

Em 21 de janeiro de 2014, o diretório re3data.org tinha 632 repositórios cadastrados.

O diretório re3data.org foi pesquisado usando os seguintes termos:

No campo de Pesquisa Geral: multidisciplinary university

O resultado foi de 20 repositórios compatíveis com os critérios de busca, conforme Quadro 3.

Ao verificar mais detalhadamente as informações apresentadas sobre esses repositórios, quatro (J. Craig Venter Institute, Longitudinal Aging Study Amsterdam, Survey of Health, Ageing and Retirement in Europe e THREDDS Data Server) não eram multidisciplinares ou de universidades. Os outros 16 encaixavam-se nos critérios de um modo geral.

Quadro 3 – Repositórios de dados de pesquisa recuperados no re3data.org

Diretório	Instituição Responsável
1. DataBox	Copenhagen University Library Service (Dinamarca)
2. EDINA	EDINA - University of Edinburgh, Data Library (Reino Unido)
3. Edinburgh DataShare	The University of Edinburgh, Informations Systems, Data Library (Reino Unido)
4. HeidICON	Universität Heidelberg, Unviertsitätsbibliothek Additional name Heidelberg University, Library (Alemanha)
5. IQSS Dataverse network	Harvard University, Institute for Quantitative Social Sciences (Estados Unidos)
6. J. Craig Venter Institute	J. Craig Venter Institute National Institute of Allergy and Infectious Disease (Estados Unidos)
7. Khazar University Institutional Repository	Khazar University, Library and Information Center (Azerbaijão)
8. Longitudinal Aging Study Amsterdam	VU University Medical Centre, Department of Sociology, Longitudinal Aging Study Amsterdam Ministry of Health Welfare and Sports, Directorate of Long-Term Care (Holanda)
9. Merritt	California Digital Library University of California Curation Center (Estados Unidos)

10. myExperiment	University of Southampton, Electronics and Computr Science, Intelligence, Agents, Multimedia Group University of Manchester, School of Computer Science University of Oxford Etc. (total de oito instituições) (Reino Unido)
11. Open Data LMU	Ludwig-Maximilians-Universität München, Universitätsbibliothek (Alemanha)
12. Phaidra	Universität Wien University of Vienna (Áustria)
13. PUB data	Universität Bielefeld (Alemanha)
14. Scholars Portal Dataverse	OCUL Ontario Council of University Librairies (Canadá) Harvard University, The Institute for Quantitative Social Science (Reino Unido)
15. ScholarSphere	Pennsylvania State University, Information Technology Services Pennsylvania State University, Libraries (Estados Unidos)
16. Stanford Digital Repository	Stanford University, Stanford University Libraries (Estados Unidos)
17. Survey of Health, Ageing and Retirement in Europe	Max Planck Institute for Social Law and Social Policy, Munich Center for the Economics of Aging (Alemanha) Survey of Health, Ageing and Retirement in Europe, Research Data Center (Holanda) Etc...(total de sete instituições)
18. THREDDS Data Server	University Corporation for Atmospheric Research, Unidata National Science Foundation (Estados Unidos)
19. UCD Digital Library	University College Dublin, Library Irish Research Council for the Humanities and Social Sciences Irish Nurses and Midwives Organisation National Digital Learning Resources (Irlanda)
20. USU-IR	University of Sumatera Utara Perpustakaan Universitas Sumatera Utara (Indonésia)

Fonte: re3data.org <<http://www.re3data.org/>>, em 21 jan. 2014.

Ao utilizar os diretórios, percebeu-se que, além de suas características gerais e de seus objetivos, os aspectos relativos à recuperação da informação e à documentação disponível fizeram diferença na recuperação. O diretório OpenDOAR permite uma busca e recuperação mais precisa, mas os outros dois diretórios têm como vantagem serem específicos sobre repositórios de dados de pesquisa.

Analisando as listas obtidas, constatou-se que os repositórios Edinburgh DataShare e ScholarSphere encontravam-se dentro dos parâmetros e apareciam nas três listas recuperadas. No final, escolheu-se o repositório Edinburgh DataShare para este estudo de caso, por ser um repositório separado do repositório institucional de publicações acadêmicas de sua instituição, enquanto o repositório ScholarSphere abrigava tanto os dados de pesquisa, quanto as publicações de sua universidade.

4.2 Contexto do estudo

A University of Edinburgh foi fundada em 1583, e é uma das universidades que constantemente está entre as 50 melhores do mundo de acordo com o QS World University Rankings. Na edição 2013/2014 desse ranking, ocupa a 17ª posição. Localizada na cidade de Edinburgo, na Escócia, Reino Unido, tem tradição de pesquisas inovadoras e de impacto internacional em diferentes áreas do conhecimento (UNIVERSITY OF EDINBURGH, c2014).

A Universidade é constituída por três faculdades: The College of Humanities & Social Science, The College of Medicine & Veterinary Medicine e The College of Science & Engineering, cada uma delas com suas escolas, totalizando 22 escolas, conforme quadro abaixo:

Quadro 4 – Faculdades e Escolas

Faculdades	The College of Humanities & Social Science	The College of Medicine & Veterinary Medicine	The College of Science & Engineering
Escolas	<ul style="list-style-type: none"> - The University of Edinburgh Business School - The School of Divinity - The School of Economics - Edinburgh College of Art - The School of Health in Social Science - The School of History, Classics & Archaeology - Edinburgh Law School - The School of Literatures, Languages & Cultures - Moray House School of Education - The School of Philosophy, Psychology & Language Science - The School of Social & Political Science 	<ul style="list-style-type: none"> - The School of Biomedical Sciences - The School of Clinical Sciences - The School of Molecular, Genetic & Population Health Sciences - The Royal (Dick) School of Veterinary Studies 	<ul style="list-style-type: none"> - The School of Biological Sciences - The School of Chemistry - The School of Engineering - The School of GeoSciences - The School of Informatics - The School of Mathematics - The School of Physics & Astronomy

Fonte: University of Edinburgh (c2014, p. 25).

É uma Universidade que tem grande interesse na geração do conhecimento e no cuidado com a pesquisa, pois temos que sua missão é “[...] a criação, disseminação e curadoria do conhecimento”² (UNIVERSITY OF EDINBURGH, 2011). Assim sendo, torna-se claro o porquê de a Universidade ter percebido a necessidade de gerenciar os seus dados de pesquisa. Há uma política de gerenciamento de dados de pesquisa que foi adotada desde 16 de maio de 2011, quando foi aprovada na Corte da Universidade. Reconhece-se que a política ainda levará anos até que seja totalmente implantada, mas é um marco importante para determinar qual a visão e a orientação da Universidade quanto a seus dados de pesquisa (UNIVERSITY OF EDINBURGH, 2015).

Destacam-se os seguintes aspectos da política: todas as novas propostas a partir da data de adoção da política devem ter um plano de gerenciamento dos dados de pesquisa, a Universidade proverá suporte e treinamento aos pesquisadores, bem como mecanismos para prover o acesso a esses dados, respeitando direitos, mas incentivando o acesso aberto aos dados. Muito importante: os dados devem ficar em algum serviço de dados ou repositório, seja temático ou da Universidade (UNIVERSITY OF EDINBURGH, 2015).

Atualmente a University of Edinburgh tem dois repositórios institucionais:

- a) O Edinburgh Research Archive (ERA), que arquiva documentos escritos por autores vinculados à Universidade, que não são controlados por editoras comerciais. Inclui teses e dissertações, relatórios de projetos, documentos informativos e outros materiais esgotados. É coordenado pelo Scholarly Communications Team, um grupo de trabalho dentro do Research & Learning Services, da Library & University Collections, que faz parte do setor de Information Services. (UNIVERSITY OF EDINBURGH, 2014).
- b) O Edinburgh DataShare, que é um repositório de dados de pesquisa. Ele permite que os pesquisadores dessa universidade depositem, compartilhem e licenciem seus dados de pesquisa. É coordenado pelo Data Library, que faz parte da EDINA and Data Library, que também faz parte do setor de Information Services. (UNIVERSITY OF EDINBURGH, 2014).

² The University’s mission is the creation, dissemination and curation of knowledge.

4.3 Objeto de estudo

O objeto deste estudo de caso é o repositório de dados de pesquisa Edinburgh DataShare³, o repositório de dados de pesquisa da University of Edinburgh. Ele permite que os pesquisadores dessa universidade depositem, compartilhem e licenciem seus dados de pesquisa. Ele foi criado como parte do projeto DISC-UK DataShare, entre 2007 e 2009, financiado pelo JISC-UK (anteriormente conhecido como Joint Information Systems Committee) (UNIVERSITY OF EDINBURGH, 2012).

4.4 Definições dos termos

Abaixo estão listadas as definições dos termos mais utilizados no contexto deste estudo:

- a) diretriz: "conjunto de instruções a serem seguidas para que sejam alcançados os objetivos de uma organização" (CUNHA; CAVALCANTI, 2008, p. 128);
- b) prática: "ação que o homem exerce sobre as coisas, aplicação de um conhecimento em uma ação concreta, efetiva" (JAPIASSÚ; MARCONDES, 2008, p. 223);
- c) política de informação: "plano para a provisão e acesso à informação no âmbito de uma [...] organização [...]" (CUNHA; CAVALCANTI, 2008, p. 285);
- d) dados de pesquisa: dados coletados, observados ou criados no contexto de uma pesquisa científica para análise ou validação dos resultados dessa pesquisa (OCDE, 2007; RODRIGUES et al. 2010; UNIVERSITY OF EDINBURGH, 2013);
- e) conjunto de dados: um conjunto de arquivos que contém tanto os dados de pesquisa - normalmente numéricos ou codificados - e documentação suficiente para fazer os dados reusáveis. (UNIVERSITY OF EDINBURGH, 2013)⁴;
- f) documentação: quaisquer arquivos digitais como tabelas de codificação, relatórios técnicos ou metodológicos ou guias do usuário, os quais expliquem a produção dos dados de pesquisa, sua proveniência, processamento ou interpretação (UNIVERSITY OF EDINBURGH, 2013)⁵;

³ Edinburgh DataShare: <http://datashare.is.ed.ac.uk/>

⁴ **Dataset:** A set of files containing both research data - usually numeric or encoded - and documentation sufficient to make the data re-usable.

⁵ **Documentation:** Any digital files such as a codebook, technical or methodology report or user guide, which explain the research data's production, provenance, processing or interpretation.

- g) comunidade: “um termo do DSpace usado para agrupar coleções em um repositório. Pode ser uma escola ou outra unidade da Universidade” (UNIVERSITY OF EDINBURGH, 2013)⁶;
- h) coleção: “um grupo de conjuntos de dados ou itens. Pode receber o nome da unidade de pesquisa responsável pela criação dos dados ou receber um nome próprio”. (UNIVERSITY OF EDINBURGH, 2013)⁷;
- i) interoperabilidade técnica: “[...] padrões e protocolos de comunicação, transporte, armazenamento e codificação de informações, tais como Z39.50, OAI-PMH, ISO-ILL, XML [...]” (SAYÃO, 2007, p. 39);
- j) interoperabilidade semântica: “[...] adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados e tesouros; (SAYÃO, 2007, p. 39);
- k) interoperabilidade política/humana: “considera as implicações para a organização, equipe e usuários de tornar as informações mais amplamente disponíveis. (SAYÃO, 2007, p. 39).

4.5 Categorias de análise

As principais categorias de análise das diretrizes e práticas de informação para repositórios institucionais foram baseadas no trabalho de Tomaél e Silva (2007), sendo elas já apresentadas em mais detalhes na seção 3.6: responsabilidade, conteúdo, aspectos legais, padrões, preservação digital, política de acesso e uso e sustentabilidade e financiamento.

A operacionalização de cada uma das categorias foi baseada, além do trabalho de Tomaél e Silva (2007), em informações presentes nos trabalhos dos seguintes autores: Torino (2010), e Green, Macdonald e Rice (2009). Afinal, como França (2001, p. 43) afirma: “[...] não existe uma única forma e um só caminho para o conhecimento”. Muitas vezes torna-se necessário combinar modelos e questões advindas de diferentes autores, para tentar chegar a um modelo que ajude a representar e abordar os diversos problemas de pesquisa existentes.

⁶ **Community:** A DSpace term used to group collections in a repository. Can be a school or other unit of the University.

⁷ **Collection:** A group of datasets, or items. Can be named after the research unit responsible for creating the data or given its own name.

A seguir apresentamos o Quadro 5 com os critérios e a operacionalização de cada uma das categorias para o momento da coleta dos dados.

Quadro 5 – Critérios e operacionalização

Critérios	Operacionalização
Responsabilidade	Verificar como a instituição se estrutura administrativamente para gerir o repositório. Verificar quais são os profissionais responsáveis pela gestão do repositório.
Conteúdo	Determinar que tipos de documento são armazenados no repositório. Determinar quais os formatos de arquivo aceitos. Verificar características relacionadas aos criadores dos itens Verificar características dos assuntos dos itens
Aspectos legais	Determinar as licenças sob as quais os dados de pesquisa são disponibilizados.
Padrões	Verificar os padrões utilizados, sejam eles tecnológicos (hardware e software), de metadados, de fluxos de trabalho (na gestão do repositório e fluxo de publicação), de interface (usabilidade e acessibilidade, níveis), e de interoperabilidade (política, semântica e técnica).
Preservação	Verificar a existência de plano de preservação digital, destacando seus principais aspectos, caso exista. Verificar o uso de sistema de identificadores permanentes. Verificar qual sistema é utilizado.
Política de acesso e uso	Verificar as políticas de submissão. Destacar os materiais de suporte à submissão. Verificar a recuperação da informação (tipos de busca disponíveis). Verificar o suporte à recuperação da informação. Verificar quem pode acessar os metadados, e quem pode acessar os dados de pesquisa. Verificar as funcionalidades de visualização e análise dos dados. Verificar o suporte à citação dos dados de pesquisa.
Sustentabilidade e financiamento	Verificar que fatores relativos ao financiamento do repositório são destacados. Verificar como se obtêm os valores necessários para custear a equipe de trabalho e a infraestrutura.

Fonte: Adaptação da autora sobre dados dos trabalhos de Tomaél e Silva (2007), Torino (2010), e Green, Macdonald e Rice (2009).

4.6 Fontes de coleta de dados

Para a etapa de seleção do objeto de estudo, as fontes de coleta de dados foram os diretórios de repositórios OpenDOAR, Databib e Re3data.

Para o estudo propriamente dito, as fontes de coleta de dados foram os documentos sobre dados de pesquisa do repositório Edinburgh DataShare e da University of Edinburgh, como manuais, tutoriais, páginas de ajuda, entre outros, documentos que discorriam sobre esse repositório de dados de pesquisa, como apresentações e postagens em blogs, mesmo que não tenham sido produzidos pela instituição. A outra fonte de dados foi o próprio repositório, na medida em que foi visitado para observação dos aspectos destacados nesta pesquisa.

4.7 Instrumentos de coleta, organização, tratamento e análise e dos dados

Como instrumento principal de coleta, organização, tratamento e análise de dados foram utilizadas bases de dados organizadas no Software NVivo, organizadas em pastas e os nós foram as categorias de análise: responsabilidade, conteúdo, aspectos legais, padrões, preservação digital, política de acesso e uso, e sustentabilidade e financiamento. Os nós secundários contemplaram a operacionalização das questões. As fontes de dados foram inseridos nessas bases de dados, através do plugin Ncapture (para páginas web) e carregamento dos arquivos PDF. O Apêndice A traz imagens desse software, representando as etapas de coleta e organização dos dados.

A escolha do software NVivo⁸ foi devido ao fato que ele permite o tratamento de dados qualitativos, bem como de dados advindos de diferentes tipos de fontes, como textos, imagens, vídeos e áudio, por exemplo.

Outro instrumento utilizado foi um roteiro de observação (disponível no Apêndice C) para a etapa de explorar o repositório a fim de verificar as práticas de tratamento dos dados de pesquisa nesses repositórios. Para registro visual das observações e complementar as anotações no roteiro, as páginas observadas também foram capturadas no NVivo, através da ferramenta NCapture.

⁸ NVivo é um software que suporta métodos de pesquisa qualitativos e mistos. Ele permite reunir, organizar e analisar conteúdos de entrevistas, discussões em grupo, pesquisas, áudio, mídia social e páginas web. Tem ferramentas de busca, consulta e visualização, que possibilitam análises diferenciadas. Página web: <http://www.qsrinternational.com/other-languages_portuguese.aspx>. A autora adquiriu pessoalmente a licença através da empresa SoftTop, revendedora autorizada no Brasil pela QSR International, detentora do software.

Além desses instrumentos, notou-se a necessidade de criar outro instrumento para estudar os metadados dos conjuntos de dados de pesquisa recuperados, a fim de sistematizar essas informações. Criou-se uma planilha (disponível no Apêndice B), com as colunas Handle (para identificar o item do repositório), N. Tipo (quantos tipos foram atribuídos ao item), Tipo (tipos atribuídos ao item), N. Criadores (quantos criadores foram atribuídos ao item), Criadores (criadores atribuídos ao item), N. Assuntos (quantos assuntos foram atribuídos ao item), Assuntos (assuntos atribuídos ao item), N. Arquivos (quantos assuntos formam o item), Extensões dos arquivos (extensões dos arquivos que formam o item), Coleção (coleção à qual o item pertence), Date accessioned (data de inclusão do item no repositório), Data available (data de disponibilização do item), Licença (licenças disponíveis ao clicar nos *links* de licenças associadas ao item), dc.rights (conteúdo do metadado dc.rights).

4.8 Procedimentos de coleta, organização, tratamento e análise e dos dados

Os procedimentos de coleta, organização, tratamento e análise dos dados estão descritos a seguir:

- a) Pesquisa documental de documentos produzidos pelo Edinburgh DataShare e pela University of Edinburgh, como manuais, tutoriais e instruções (estes dados foram coletados em 28 de julho de 2014) e documentos sobre esse repositório e relacionados à curadoria dos dados da University of Edinburgh, como postagens em blogs e apresentações em eventos (estes materiais foram sendo coletados durante o ano, quando eram encontrados);
- b) A fim de facilitar a análise e a identificação dos tópicos, os documentos foram inseridos em uma base de dados no software NVivo, através do plugin Ncapture (para páginas web) e por carregamento dos arquivos PDF. Foram criados nós para as principais categorias. Quando trechos de texto ou imagem relacionados a essas categorias foram identificados nos documentos, foram selecionados e relacionados aos nós. Foram criados nós secundários para a operacionalização das categorias, ligados aos nós primários;
- c) Exploração do repositório em busca de visualizar os conjuntos de dados de pesquisa (datasets) e os tópicos a serem analisados, descritos no roteiro de observação, o qual foi preenchido para registro visual das observações. Para complementar as anotações no roteiro, as páginas observadas também foram capturadas no NVivo,

através da ferramenta NCapture (a observação formal ocorreu de 09 a 13 de outubro de 2014, nela foram capturados os dados que servem como uma fotografia da situação que é analisada nesta pesquisa, e o roteiro de observação foi preenchido, no entanto, decidiu-se acompanhar informalmente o repositório em momentos posteriores, a fim de complementar a observação formal);

- d) As informações dos metadados de cada item foram inseridas na planilha, conforme descrito na seção anterior;
- e) As informações do repositório foram analisadas tópico a tópico, incluindo as informações da observação e as informações obtidas na pesquisa documental, registrada no NVivo;
- f) Foram relacionadas as informações obtidas sobre as diretrizes e as práticas as informações sobre o processo de gerenciamento de dados de pesquisa da University of Edinburgh e com o modelo do ciclo de vida da curadoria digital do Digital Curation Centre (HIGGINS, 2008; 2012);
- g) Foram realizadas considerações sobre as informações obtidas nessa pesquisa e o modelo de comunicação científica proposto por Sales e Sayão (2012).

4.9 Limitações da pesquisa

Os repositórios institucionais de dados de pesquisa ainda são um assunto relativamente novo, e esta abordagem de pesquisa procurou obter uma visão das relações entre as características de um repositório específico e o processo de gerenciamento de dados de pesquisa, bem como com a comunicação científica em um modo mais amplo. Mas ao escolher um caso específico, tem-se como limitação da pesquisa de que não se podem fazer generalizações numéricas, mas sim observar articulações e movimentos.

5 ANÁLISE E DISCUSSÃO DOS DADOS

Nas seções a seguir serão apresentadas as análises relacionadas a cada objetivo, ou seja, delinear as características do repositório Edinburgh DataShare (responsabilidade, conteúdo, aspectos legais, padrões, preservação digital, política de acesso e uso, e sustentabilidade e financiamento), depois relacionar as diretrizes e as práticas com o processo de gerenciamento de dados de pesquisa, usando o modelo do ciclo de vida da curadoria digital do DCC e finalmente verificar como as diretrizes e as práticas se relacionam com a comunicação científica, usando o modelo de Sales e Sayão (2012).

5.1 Características do Repositório Edinburgh DataShare

Nesta seção são explorados os resultados obtidos para cada uma das características do repositório Edinburgh DataShare, utilizando a classificação de Tomaél e Silva (2007), com aspectos observados baseados também em Torino (2010), e Green, Macdonald e Rice (2009).

5.1.1 Responsabilidade

Ao procurar dados sobre a responsabilidade do Edinburgh DataShare, encontrou-se a informação de que o repositório está sob a responsabilidade da Data Library. A Data Library é parte da divisão EDINA and Data Library, que por sua vez faz parte do setor Information Services da University of Edinburgh.

O setor Information Services une diversos serviços ligados à área de tratamento da informação, mostrando quão próximos estão os setores ligados à ciência da informação/Biblioteconomia e à tecnologia da informação na visão estrutural da Universidade.

São sete os profissionais diretamente ligados à Data Library: Robin Rice (Data Librarian) Stuart Macdonald (Associate Data Librarian) Claire Bradford (Information Officer) George Hamilton (Software Engineer) Rocío von Jungfeld (Data Library Assistant) Laine Ruus (Assistant Data Librarian) Pauline Ward (Data Library Assistant).

Têm-se, assim, profissionais ligados às áreas de biblioteconomia/ciência da informação e tecnologia da informação como diretamente envolvidos com o repositório. Os outros profissionais da divisão EDINA and Data Library também têm funções semelhantes, e

outras complementares, como webdesign e helpdesk. Há colaboração também com as outras divisões do setor Information Services, principalmente com a divisão Library and University Collections, que é responsável pelo repositório de publicações de acesso aberto, o Edinburgh Research Archive (ERA). Os pesquisadores são responsáveis pelas coleções nas quais depositam seus itens.

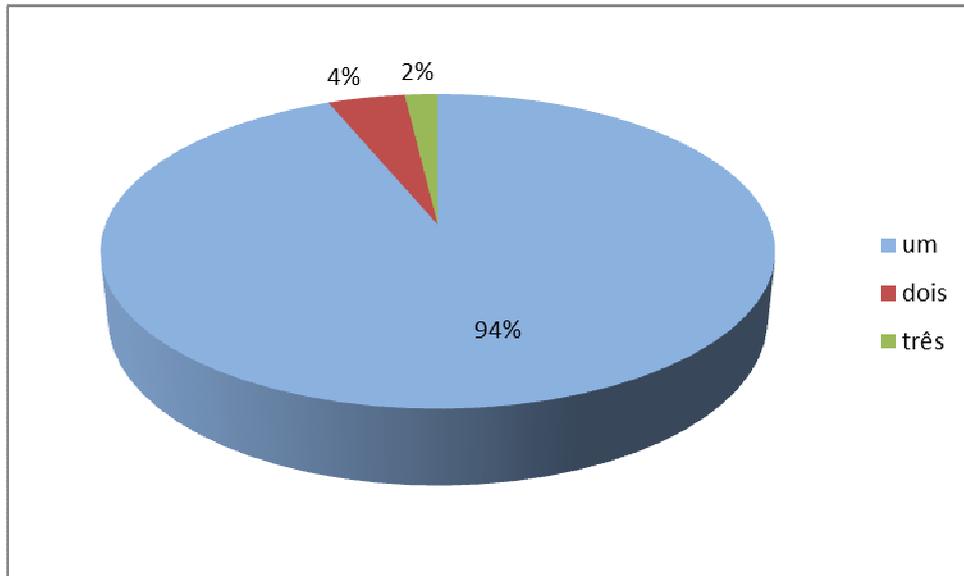
Tomaél e Silva (2007) consideram que a principal função dos bibliotecários, em instituições acadêmicas e de pesquisa, é organizar e manter o conteúdo digital gerado por autores das suas instituições. Os dados de pesquisa podem também ser considerados bens digitais (nascidos digitais ou digitalizados) das instituições, e neste caso temos bibliotecários envolvidos no projeto do repositório institucional de dados de pesquisa.

Destaca-se também que, para Leite (2009), é importante que a equipe do repositório seja devidamente capacitada e comprometida com o projeto. Ele considera que uma equipe multidisciplinar, com bibliotecário, analista de sistemas e profissional de comunicação/marketing atende aos principais requisitos de recursos humanos de um repositório institucional, com no mínimo tendo os dois primeiros profissionais como base. Percebe-se que, no caso da University of Edinburgh, o critério mínimo é preenchido, pois há na Data Library profissionais da informação, bibliotecários especializados em dados, engenheiro de software, e na EDINA há outros profissionais que complementam, inclusive com um profissional de webdesign, área relacionada com a tecnologia da informação e com a comunicação.

5.1.2 Conteúdo

Por ser um repositório institucional de dados de pesquisa, o Edinburgh DataShare conta com dados de pesquisa como seu tipo documental. Mas os dados de pesquisa podem ser tão variados quanto as pesquisas que podem ser conduzidas. Dados de pesquisa não são organizados de uma única maneira.

Ao buscar os dados de pesquisa existentes no repositório em 09 de outubro de 2014, obteve-se um total de 161 itens, sendo que 94% deles indicavam somente um tipo (metadado dc.type), conforme pode ser visto no Gráfico 1.

Gráfico 1 – Quantidade de tipos de dados de pesquisa por item

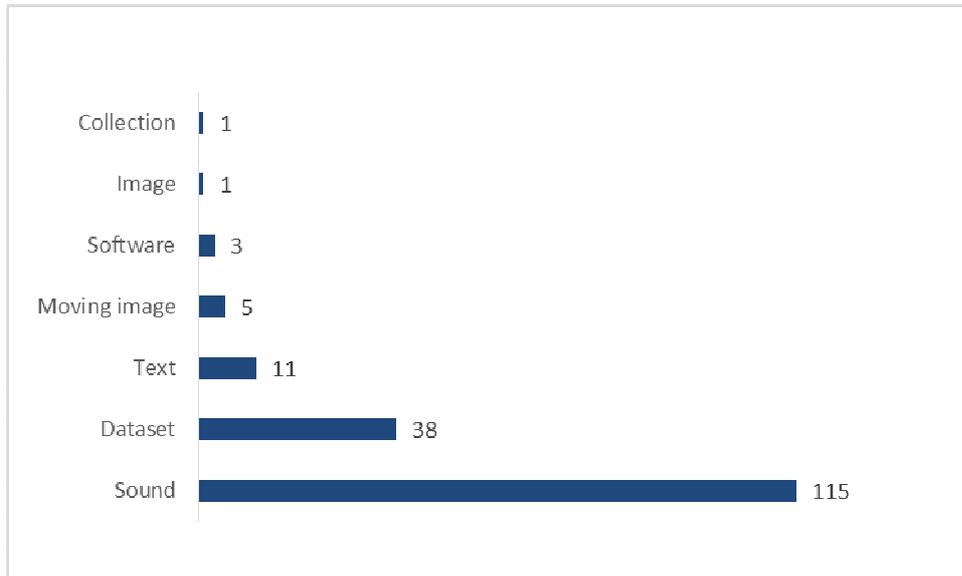
Fonte: Dados da pesquisa.

A Tabela 1 apresenta os tipos de dados de pesquisa conforme estão apresentados nos itens, ou seja, os itens que têm mais de um metadado dc.type aparecem agrupados, conforme os valores desse metadado. O Gráfico 2 foi gerado separando todos os tipos, por isso a soma do percentual é maior do que 100%. Os tipos de dados de pesquisa mais frequentes no repositório são os áudios, seguidos pelos conjuntos de dados, conforme pode ser visualizado no Gráfico 2. Esses tipos de recurso são previstos na lista do Dublin Core Metadata Initiative (DCMI USAGE BOARD, 2010).

Tabela 1 – Tipos de dados de pesquisa dos itens

Tipos	Frequência	Percentual (%)
Sound	106	65,84
Dataset	34	21,12
Text; Sound	6	3,73
Moving image	5	3,11
Dataset; Sound; Text	3	1,86
Software	2	1,24
Text	2	1,24
Collection	1	0,62
Dataset; Software	1	0,62
Image	1	0,62
Total Geral	161	100%

Fonte: Dados da pesquisa.

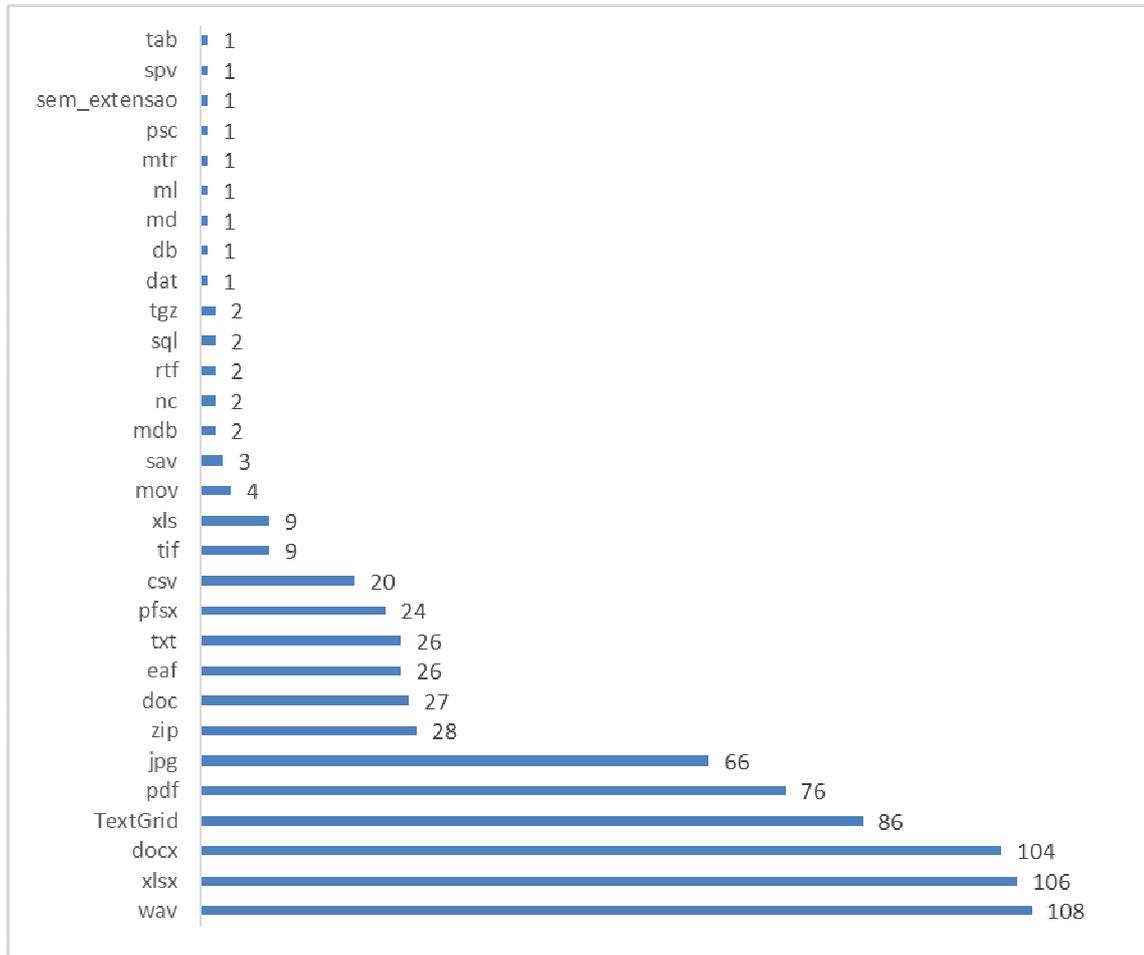
Gráfico 2 – Tipos de dados de pesquisa

Fonte: Dados da pesquisa.

Há muitos itens que contêm arquivos de áudio (wav), textuais (docx, doc, txt, pdf), imagens (jpg, tif) e planilhas (xlsx, xls). Os formatos dos arquivos foram identificados através da lista dos arquivos que estão no final de cada um dos itens no formato completo. É possível ver, comparando os Gráficos 2 e 3, que somente um item foi caracterizado com o tipo de recurso imagem (indicando que a imagem seria o tipo predominante no recurso), mas que há 66 itens que contêm ao menos um arquivo jpg e nove itens que contêm arquivos tif. Outra situação a ser destacada é a presença dos arquivos comprimidos, com as extensões zip e tgz, que dificulta a identificação dos formatos, pois é necessária a etapa da extração dos arquivos. Neste trabalho os arquivos comprimidos não foram abertos para identificação dos formatos dos arquivos que estavam dentro deles. Formatos proprietários também podem trazer problemas para a preservação, ao contrário dos arquivos em formatos abertos.

Há relações entre o que é descrito no metadado dc.type e os formatos dos arquivos disponíveis: há 115 itens com tipo=áudio e 108 itens com arquivos no formato=wav. Os outros sete itens com tipo=áudio e que não contêm arquivos wav contêm arquivos comprimidos (zip). Há a presença de planilhas xlsx em 107 itens, porém poucos itens contêm tipo=conjunto de dados. A maioria dos itens que contêm essas planilhas são tipo=áudio, e os arquivos das planilhas são metadados mais detalhados dos arquivos de áudio.

Gráfico 3 – Extensões dos arquivos dos itens



Fonte: Dados da pesquisa.

Há um total de 29 diferentes formatos de extensões, e um arquivo sem a indicação de extensão (é um arquivo do tipo texto, conforme identificado pelo repositório). O repositório recomenda aos depositantes que observem as diretrizes de formatos de arquivos do UK Data Archive (Anexo D).

Green, Macdonald e Rice (2009) ressaltam a ligação que a escolha dos formatos a serem aceitos no repositório tem com a preservação digital dos mesmos. Já há uma variedade de formatos de arquivos disponíveis, e poucas unidades da Universidade ainda estão participando desse projeto. Será importante que as políticas continuem se desenvolvendo nesse caminho, para assegurar o acesso a esses dados de pesquisa ao longo do tempo.

O Edinburgh DataShare começou como um projeto piloto, com algumas poucas unidades servindo como base para o início do repositório. Assim, pode-se perceber que há uma tendência de os dados de pesquisa nele contidos não representarem ainda todas as áreas do conhecimento, mas prioritariamente as áreas específicas contempladas no projeto piloto.

A Tabela 2 apresenta a lista dos 10 criadores que têm mais itens no repositório.

Tabela 2 – Dez criadores que têm mais itens no repositório

Criador	Quantidade de itens	Percentual (%)
Remijsen, Bert	103	63,98
Gwado Ayoker, Otto	62	38,51
Penn de Ngong, John	38	23,60
Ladd, D. Robert	38	23,60
Impey, Angela	38	23,60
Ayuel Ring, Peter Malek	38	23,60
Meyerhoff, Miriam	38	23,60
Deng Yak, Simon Yak	38	23,60
Reid, Tatiana	38	23,60
Ajuet Deng, Elizabeth Achol	38	23,60

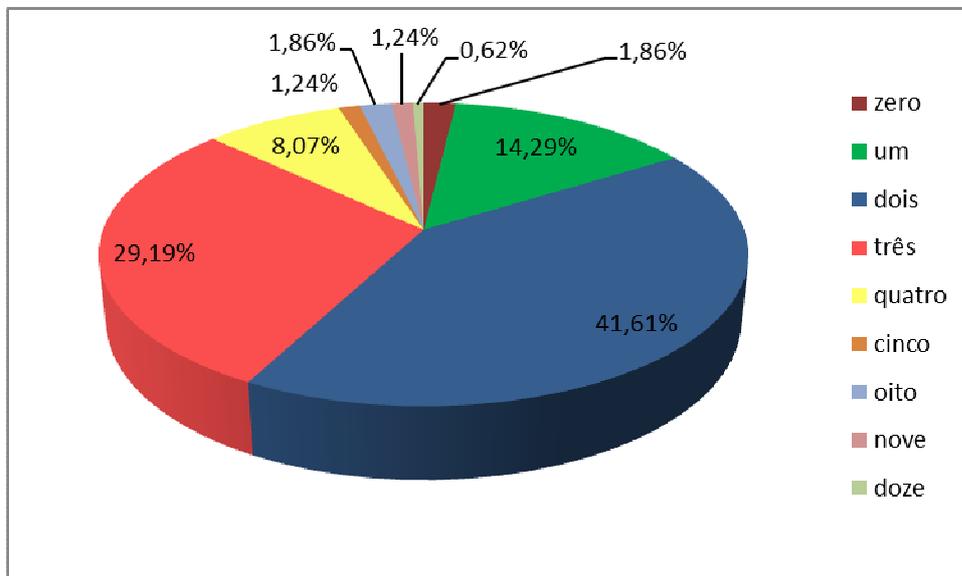
Fonte: Dados da pesquisa.

Ao analisar a Tabela 2, fica claro que quase 64% dos itens do repositório foram criados pelo pesquisador Bert Remijsen⁹, da área de assunto de Linguística e Idioma Inglês (Linguistics and English Language), e que faz parte da School of Philosophy, Psychology and Language Sciences, da University of Edinburgh. O que não aparece nessa tabela é que todos os outros criadores que estão nessa lista dos “10 mais” criaram todos seus itens em colaboração com Bert Remijsen. Os 62 itens criados com Otto Gwado Ayoker pertencem à coleção Collection of Shilluk narratives and songs, e os 38 itens criados com as outras pessoas, à coleção A Collection of Dinka Songs. Apenas três itens desse pesquisador não foram criados em colaboração. Assim, quaisquer dados numéricos relacionados aos itens do Edinburgh DataShare devem ser pensados com relação a essa situação peculiar, que provavelmente se diluirá no passar do tempo, na medida em que mais itens sejam depositados no repositório.

⁹ Página web: <http://www.lel.ed.ac.uk/~bert/index.html>

Sobre os assuntos dos itens do repositório: a maioria dos itens foi descrito com dois ou três assuntos (metadado dc.subject). Poucos itens foram descritos com muitos assuntos, dentre eles destaca-se o item “General Catalogue” (<http://datashare.is.ed.ac.uk/handle/10283/380>) que foi descrito com 12 assuntos e tem também o maior número de criadores (55). Existe também outro metadado relacionado com assunto, o metadado Subject Classification (dc.subject.classification), que é um pouco mais controlado, como explicado logo adiante, porém nenhum dos metadados de assunto é obrigatório.

Gráfico 4 – Quantidade de assuntos (metadado dc.subject) por item



Fonte: Dados da pesquisa.

A Tabela 3 apresenta as informações do metadado dc.subject de cada item. É perceptível a não padronização com relação ao uso de maiúsculas. É possível identificar rapidamente os assuntos incluídos nas duas maiores coleções do pesquisador Bert Remijsen.

A não obrigatoriedade de um metadado de assunto pode ser problemática, visto que em um repositório de dados de pesquisa a busca por texto completo nem sempre é possível, devido à variedade dos formatos de arquivo utilizados, ao contrário dos repositórios de publicações científicas, que normalmente adotam padrões como arquivos PDF e indexam o texto completo para busca. Assim, a inclusão de assuntos amplia as possibilidades de recuperação da informação.

**Tabela 3 – Assuntos (metadado dc.subject) dos conjuntos de dados do
Edinburgh DataShare**

Assuntos	Total
Shilluk, oral culture	62
songs, Dinka, African literature	38
electronic music, performance practice, computer music, electroacoustic	4
Temperature, Lake, ATSR	3
(sem assunto)	3
Lake, Temperature, ATSR	2
Harvard sentences, Spanish, phonetically balanced	2
water, seawater, refractive index, emissivity, infrared, ATSR, AATSR	1
visual linguistic treebank dataset images vdr visual dependency representation	1
tone, tonal alignment, fundamental frequency, autosegmental theory, Shilluk	1
Telephone communication, Information and communication technology, Telephone usage survey	1
Technology, ICT skills, teaching and learning, digital literacy	1
stress hyperglycaemia, hyperglycaemia, hyperglycemia, diabetes, risk prediction, routine data, emergency admissions, risk calculator	1
speech in noise, Lombard, Hurricane Challenge, Harvard sentences, Modified Rhyme Test	1
Sea Surface Temperature, ATSR, Climate, Sea Surface Temperature Anomaly	1
Scholarly Communication, Institutional Repository, Information Environment, Open Access	1
sample survey, academic history, academic results, attitudes to schooling, study habits, reasons for leaving school, work, further education, recreation	1
Saints	1
research, teaching, learning, collaboration, tools, administration, Web2 tools	1
research, computing, support service requirements	1
Reproduction	1
referring expression generation, visual salience, visual clutter	1
QUANGO, non-governmental organisation	1
Phonetics	1
performance practice, electronic music, electroacoustic, computer music, score	1
overcrowding, occupational classification, Census Areas Through Time, car ownership, unemployment, area deprivation, social class, Scotland, Carstairs index	1
openurl, router, EDINA	1
ocaml celtic studies medieval	1
Numerical analysis, Special functions, Confluent hypergeometric function, Gauss hypergeometric function	1
Nilo-Saharan, Dinka, tone, three-level vowel length, phonetics, tone, tonal alignment	1
Magic, Sorcery, Necromancy, Witchcraft	1
librarianship, digital library	1
Information Environment, Scholarly Communication, Open Access, Institutional Repository	1
Information Environment, Scholarly Communication, Institutional Repository, Open Access	1

Imposing Memory Task, mindreading, metarepresentation, evolutionary psychology	1
Historical recordings, Children	1
Gold open access, journal impact factor	1
Epigenetics Allele Expression Histone Modification SNP	1
English, Harvard sentences, repeated speech, read speech, audio recordings	1
End credit sequence, music, television serial, drama serial	1
elections, elections results	1
Dinka, number marking, noun morphology, verb morphology	1
Diapix, non-native, conversational, spontaneous, Spanish, English	1
Dialects, Songs, Interviews, Intonation, Phonetic practice, Readings, Lectures, Poems, Synthetic Speech, Ear Training Tapes, Xhosa, Accents	1
conserved eukaryote signaling, protein kinase phylogeny, <i>Ostreococcus tauri</i> , TOR signaling, MAPK cascade, circadian clock, phosphorylation, model kinome	1
Clinical trials, stroke, RCTs, randomized controlled trials	1
clinical trials, randomised controlled trials, RCTs,stroke,	1
clear speech, infant-direct speech, computer directed speech, non-native-directed speech, Lombard speech	1
channel heads, drainage density lidar	1
Cardinal vowel, Phonological and phonetic contexts, Spoken sentences	1
Ambel, oral culture	1
Ambel, language description	1
Aesop's Fables, Historical recordings, Phonetics	1
ACT	1
Total	161

Fonte: Dados da pesquisa.

Como o objetivo dessa seção é uma caracterização mais abrangente do repositório, e já está demonstrado o viés que existe com relação às coleções da área de linguística, em virtude de quase 64% dos itens serem criados por um pesquisador dessa área, não será feita uma análise detalhada das ligações entre os assuntos.

Na análise do outro metadado de assunto (dc.subject.classification) percebeu-se que havia diferenças entre os metadados dos registros incluídos em épocas diferentes, oito iniciavam com a expressão “Research Subject Categories”. Para preparar a Tabela 4, foi considerada apenas a primeira parte do assunto, antes das subdivisões e a expressão “Research Subject Categories” foi excluída, para evidenciar as categorias mais gerais de assunto representadas no repositório.

Tabela 4 – Frequência de assuntos (metadado dc.subject.classification) dos conjuntos de dados do Edinburgh DataShare

Classificação de assunto (metadado dc.subject.classification)	Frequência	Percentual (%)
Eastern Asiatic African American and Australasian Languages Literature and related subjects	66	40,99
Musicology	38	23,60
(sem assunto)	11	6,83
Physical Sciences	8	4,97
Mathematical and Computer Sciences	7	4,35
Linguistics Classics and related subjects	6	3,73
Creative Arts and Design	5	3,11
Education	5	3,11
Biological Sciences	3	1,86
Medicine and Dentistry	3	1,86
Social studies	3	1,86
Historical and Philosophical studies	2	1,24
Mass Communications and Documentation	2	1,24
Bioinformatics	1	0,62
Geomorphology	1	0,62
Total	161	100

Fonte: Dados da pesquisa.

Tendo em vista que o metadado dc.subject.classification é controlado, os assuntos deveriam ser codificados usando o Joint Academic Coding System (JISC)¹⁰. Além de ser um campo mais controlado, outra característica que distingue o metadado dc.subject.classification do metadado dc.subject é que ele não é repetível, ou seja, o item só pode ter uma ocorrência desse metadado. Ele serve mais para identificar a área de assunto predominante do conjunto de dados de pesquisa. É realmente preocupante que esse metadado não seja obrigatório, pois neste caso não foram apenas três itens sem o metadado, como no caso do dc.subject, mas sim 11 itens, e não somente itens mais antigos, mas também recentes.

É interessante destacar que, provavelmente por boa parte dos registros serem provenientes do projeto piloto, talvez a funcionalidade de escolha na lista ainda não estivesse totalmente implantada no momento do depósito desses materiais, por isso os problemas observados no preenchimento de um metadado controlado. Talvez fosse relevante a padronização (correção) desse metadado nos registros que já estão no repositório, a inclusão

¹⁰ Esse sistema de classificação é mantido pelo Universities and Colleges Admissions Service (UCAS) e pela Higher Education Statistics Agency (HESA) do Reino Unido. É um sistema que serve para classificar os cursos das universidades por assunto, mas esse sistema também costuma ser utilizado por outras instituições para organizar seus recursos.

nos 11 itens que não o tem e a transformação desse metadado em obrigatório, pois além de facilitar a recuperação da informação para o usuário final dos dados de pesquisa, ele pode ser utilizado inclusive para análises bibliométricas e/ou gerenciais do repositório, já que um mesmo metadado pode contemplar mais de uma utilidade, conforme Qin, Ball e Greenberg (2012).

5.1.3 Aspectos legais

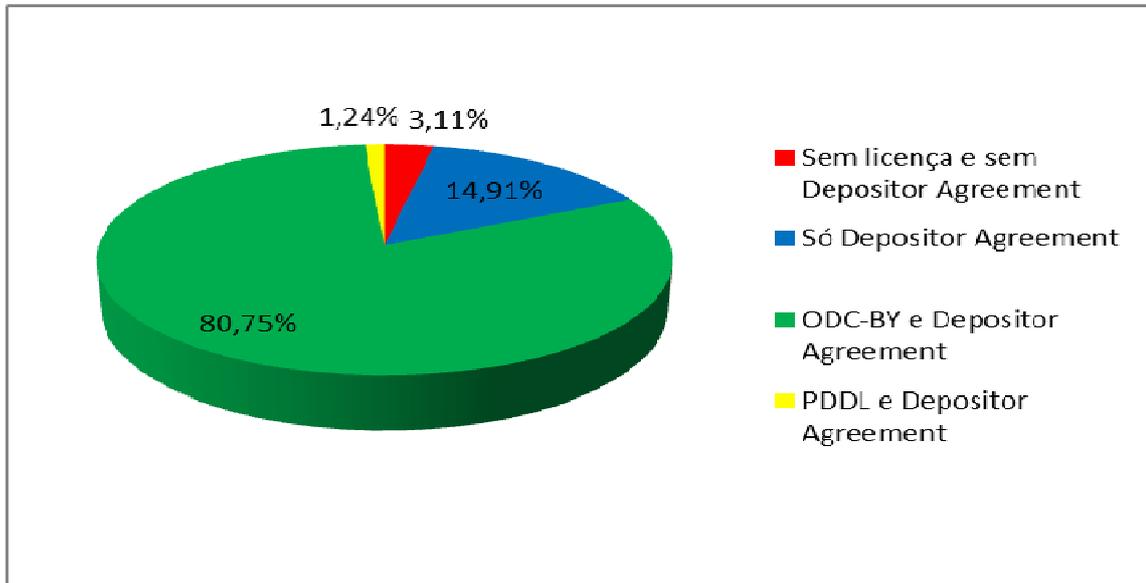
Os repositórios institucionais de dados de pesquisa, como os demais repositórios institucionais, também precisam estabelecer parâmetros relacionados aos aspectos legais dos materiais neles depositados, procurando não infringir copyrights e assegurar a propriedade intelectual dos criadores dos dados de pesquisa em primeiro lugar.

Leite (2009) destaca que há dois tipos de licenças que os repositórios institucionais normalmente trabalham: as licenças de depósito, as quais são um acordo entre o autor e a instituição e garantem ao repositório o direito de preservar e distribuir o trabalho depositado e as licenças de uso, as quais são um acordo entre o autor e o usuário final e que informam o os usos que podem ser feitos do material recuperado no repositório.

De acordo com as diretrizes sobre licenças do Edinburgh DataShare, há duas opções para os depositantes:

- a) Selecionar a licença de atribuição Open Data Commons (ODC-BY), a qual permite que o usuário final possa utilizar os conjuntos de dados, desde que informe o criador dos dados. Essa licença é semelhante à licença de atribuição Creative Commons (CC-BY), mas é específica para bases de dados/conjuntos de dados;
- b) ou selecionar nenhuma licença (e aceitar somente o acordo do depositante - Depositor Agreement) e indicar no metadado de direitos suas próprias condições de licenciamento/uso dos seus dados.

Gráfico 5 – Tipos de licenças dos itens



Fonte: Dados da pesquisa.

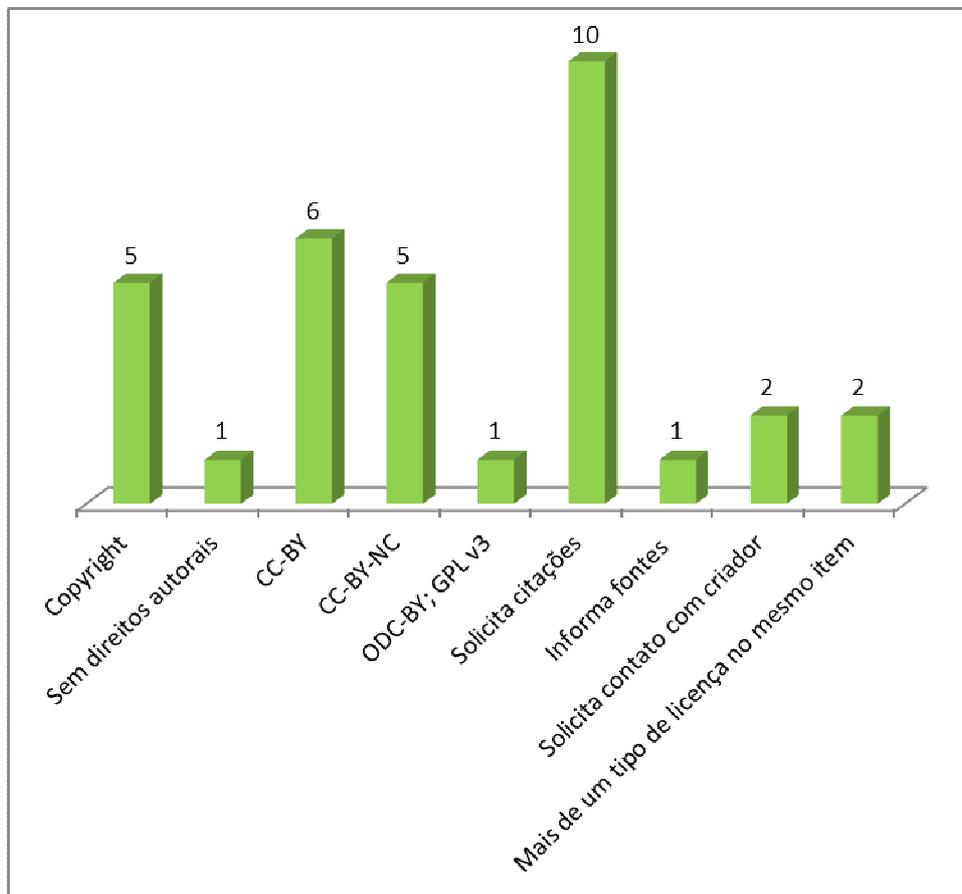
Na prática, as seguintes situações foram observadas, a partir da identificação dos arquivos de licença relacionados no final de cada um dos itens no formato completo: 130 itens (80,75%) têm como licença a ODC-BY. A maioria desses itens (100 itens) pertence às coleções “A Collection of Dinka Songs” e “Collection of Shilluk narratives and songs” da área de lingüística, do prof. Bert Remijsen.

Dois itens (1,24%) têm a licença de domínio público da Open Data Commons - Public Domain Dedication and Licence (PDDL). Essa licença não faz nenhuma exigência para o usuário dos dados, é a mais aberta de todas, porém não faz parte das diretrizes. Não foi possível identificar algum padrão que fosse exclusivo dos itens que contêm essa licença, ambos são do tipo dataset (conjunto de dados de pesquisa), porém há outros itens com esse tipo que não usam essa licença. Como essa licença não faz parte das diretrizes, pode ter sido utilizada nos primeiros momentos do projeto piloto, já que os handles desses itens não estão entre os números mais altos do repositório, um deles é dos primeiros itens (<http://hdl.handle.net/10283/47> e <http://hdl.handle.net/10283/123>).

Dos cinco itens (3,11%) que não têm nenhum tipo de licença vinculada, nem o Depositor Agreement, somente um tem informação de copyright no metadado dc.rights. Em comum entre esses registros que não têm nenhuma informação de licença nem nos metadados, tem-se que eles são os primeiros handles do repositório (<http://hdl.handle.net/10283/7>, <http://hdl.handle.net/10283/8>, <http://hdl.handle.net/10283/9>, <http://hdl.handle.net/10283/10> e <http://hdl.handle.net/10283/46>).

Dos 24 itens (14,91%) que têm somente o Depositor Agreement, ou seja, escolheram não licenciar seus dados através de um arquivo de licença do repositório, 23 têm informação de licença no metadado dc.rights. Esses são os mais interessantes para se analisar, visto que as informações disponíveis no campo “dc.rights” têm muito a dizer, pois apresentam dados mais detalhados sobre os direitos/licenças desses itens depositados, indicando necessidades mais específicas do que as licenças disponibilizadas no formulário de submissão do repositório (atualmente é só uma, ODC-BY).

Gráfico 6 – Informações do metadado dc.rights nos itens que têm somente o Depositor Agreement



Fonte: Dados da pesquisa.

Um dos itens declara claramente que não há direitos de propriedade intelectual, e também é do tipo dataset. Dez itens informam a necessidade de citação de uma obra ou dos dados, quando do uso dos dados de pesquisa. Cinco itens incluem informação de copyright. Dois itens informam a necessidade de contatar o criador dos dados ou detentor de copyright antes de determinados usos ou do fim do embargo.

Seis atribuem uma licença CC-BY, sendo que um deles atribui essa licença a algumas partes do item, outras partes do item têm outros copyrights descritos no metadado. Um atribui uma licença ODC-BY para os dados e uma licença GPL v3 para o código fonte do software, todos pertencentes ao mesmo item. Cinco registros atribuem licenças CC-BY-NC. Apenas um registro informa apenas que os dados foram obtidos livremente em determinados sites e os sites nos quais os dados foram obtidos.

Como o repositório disponibiliza automaticamente o metadado dc.identifier.citation, traz uma facilidade ao usuário final na tarefa de citar os dados conforme as solicitações dos criadores no metadado dc.rights e nas licenças com atribuição (BY).

Tomaél e Silva (2007, p. [7], grifo das autoras) ressaltam que “um repositório institucional deve ser apenas o custodiante da informação que armazena e dissemina, e não seu proprietário. Portanto, um repositório institucional não pode digitalizar e dar acesso a informações protegidas por *copyright*.”. No caso dos dados de pesquisa do Edinburgh DataShare que informam terem copyright, cinco são da própria University of Edinburgh, e um do criador, Thomas Backdahl, que decidiu compartilhar seus arquivos para verificação dos resultados de seu trabalho. Neste caso, o repositório continua sendo um custodiante e não proprietário dos dados, pois os criadores aceitaram o Depositor Agreement e inseriram suas próprias condições de uso como julgaram necessárias.

5.1.4 Padrões

Os padrões são fundamentais em um repositório institucional. Nesta seção são apresentados os resultados encontrados referentes aos padrões tecnológicos, de metadados, de fluxo de trabalho, de interface e de interoperabilidade.

Padrões tecnológicos (hardware e software): Não foram encontradas diretrizes separadas das práticas neste caso. Sobre o hardware também não foram encontradas informações. O sistema é online, e a plataforma de software utilizada é o DSpace, um dos sistemas de código aberto para repositórios mais em uso no mundo atualmente. É o mesmo sistema utilizado pelo repositório de publicações da Universidade. Afirmam que o serviço é projetado para ser altamente resiliente. Ele roda em Solaris¹¹, nos Kings Buildings¹²

¹¹ Sistema operacional baseado em UNIX originalmente desenvolvido pela Sun e atualmente desenvolvido e comercializado pela Oracle. Pagina web: <<http://www.oracle.com/br/products/servers-storage/solaris/index.html>>.

utilizando espaço de armazenamento da Universidade e é automaticamente replicado em um backup na Appleton Tower¹². É importante que o hardware e o software sejam adequados à proposta de um repositório institucional, conforme Tomaél e Silva (2007), e apesar e não se ter obtido informações sobre o hardware, o software está de acordo com o que é utilizado atualmente em repositórios institucionais.

Padrões de metadados: Foi possível verificar tanto nas diretrizes quanto nas práticas o uso de um perfil de aplicação do padrão de metadados Dublin Core qualificado. Há pequenas diferenças observadas entre as diretrizes e as práticas, aparentemente decorrentes da própria natureza de construção de um projeto como esse repositório. Percebe-se que em itens mais antigos (por exemplo, nos registros <http://hdl.handle.net/10283/7>, <http://hdl.handle.net/10283/8>, <http://hdl.handle.net/10283/9>, <http://hdl.handle.net/10283/10>) existem metadados que estão fora do padrão descrito nas diretrizes, e que não se repetem em registros mais recentes.

Quadro 6 – Padrão de metadados do Edinburgh DataShare

Diretrizes (<i>Perfil de aplicação para datasets</i>)	Práticas (<i>Metadados encontrados nos itens</i>)
dc.contributor	dc.contributor
<i>dc.contributor.advisor</i> *	dc.contributor.advisor (registros antigos)
<i>dc.contributor.author</i> *	dc.contributor.author (registros antigos)
dc.contributor.other	dc.contributor.other
dc.coverage.spatial	dc.coverage.spatial
dc.coverage.temporal	dc.coverage.temporal
dc.creator	dc.creator
<i>dc.description</i> *	dc.description (registros antigos)
dc.description.abstract	dc.description.abstract
<i>dc.description.sponsorship</i> *	dc.description.sponsorship (registros antigos)
dc.date.accessioned	dc.date.accessioned (incluído automaticamente)
dc.date.available	dc.date.available
dc.date.issued	dc.date.issued (incluído automaticamente)
dc.description.tableofcontents	dc.description.tableofcontents
dc.identifier.citation	dc.identifier.citation (incluído automaticamente)
dc.identifier.uri	dc.identifier.uri (incluído automaticamente)
dc.language.iso	dc.language.iso
dc.publisher	dc.publisher
<i>dc.relation.hasversion</i> *	dc.relation.hasversion
dc.relation.isversionof	dc.relation.isversionof
dc.relation.isreferencedby	dc.relation.isreferencedby
dc.relation.replaces	dc.relation.replaces
dc.relations.isreplacedby	dc.relation.isreplacedby

¹² Prédios da University of Edinburgh.

dc.rights	dc.rights
dc.source	dc.source
dc.subject	dc.subject
dc.subject.classification	dc.subject.classification
dc.title	dc.title
dc.title.alternative	dc.title.alternative
dc.type	dc.type
dc.date.embargo	dc.date.embargo

Fonte: Dados da pesquisa.

* Metadados que aparecem apenas nos registros e no formato Dublin Core qualificado, não no perfil de aplicação para datasets

Rodrigues e outros (2010) já comentavam a dificuldade de se uniformizar a descrição entre áreas científicas diferentes, que é o que ocorre em repositórios institucionais de universidades que decidem abrigar dados de pesquisa de diversas áreas do conhecimento. Alguns conjuntos de dados de pesquisa disponíveis no Edinburgh DataShare contêm outros metadados em arquivos que fazem parte dos seus itens, como é possível observar nos itens das coleções “A Collection of Dinka Songs” e “Collection of Shilluk narratives and songs”, que contêm arquivos de planilha com metadados acerca dos arquivos de áudio, ou seja, o repositório tem seu padrão de metadados, para a administração e a recuperação de conteúdos, mas é importante que os criadores dos dados também mantenham seus dados organizados conforme as práticas de suas comunidades científicas.

Padrões de fluxos de trabalho: Não foi possível observar diretamente as práticas do fluxo de submissão de materiais. Neste caso, recorreu-se aos materiais disponibilizados pelo repositório com suas orientações e diretrizes de submissão e notícias da Universidade relacionadas a esse fluxo.

Os materiais são bem ilustrativos sobre os passos que o Edinburgh DataShare apresenta para o depositante incluir seu item no repositório. Trazem as instruções passo a passo sobre o preenchimento dos campos de dados (metadados) e a inclusão dos arquivos, bem como sobre a inclusão de uma licença aberta (caso seja essa a escolha do depositante). Segue um fluxo padrão do software DSpace, com as adaptações para os metadados escolhidos para os dados de pesquisa.

Outra forma de fluxo de trabalho que estava sendo preparada pela equipe do Edinburgh DataShare era o depósito através do laboratório eletrônico RSpace electronic notebook, que estava sendo implantado na Universidade. Após selecionar registros do RSpace para serem depositados, seria necessário clicar no botão DataShare. Assim, abriria uma janela

para o depositante fornecer algumas informações sobre o depósito. Depois de confirmar alguns detalhes, o depósito seria efetuado diretamente no Edinburgh DataShare.

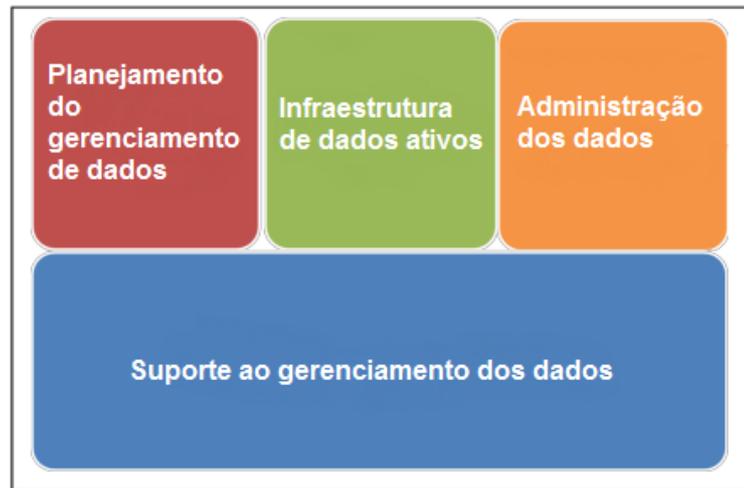
O outro meio é o uso do protocolo SWORD para depósito em lote de arquivos grandes e/ou em grande número. São utilizados *checksums*, ou seja, códigos para conferir a integridade dos arquivos.

Não foram encontradas informações relativas ao fluxo administrativo do repositório, as mais próximas são as informadas na seção Responsabilidade deste trabalho.

É relevante destacar que o repositório Edinburgh DataShare faz parte das estratégias da Universidade para o gerenciamento de seus dados de pesquisa. Esse gerenciamento abrange ações mais amplas do que o próprio repositório, e a Universidade tem uma política de gerenciamento de dados de pesquisa aprovada desde 16 de maio de 2011 (Anexo A) que contempla dez pontos chave e está sendo implantada desde então. O mapa da implementação dessa política contempla quatro áreas estratégicas (RICE *et al.*, 2013; UNIVERSITY OF EDINBURGH, 2014b):

- a) Planejamento do gerenciamento de dados (*Data Management Planning*): são os serviços para as atividades de planejamento que ocorrem normalmente antes de os dados de pesquisa serem coletados ou criados. Relaciona-se com as cláusulas 3 e 4 da Política;
- b) Infraestrutura de dados ativos (*Active Data Infrastructure*): são as ferramentas para guardar os dados ativos, ou seja, que estão sendo utilizados em atividades atuais de pesquisa, fornecendo facilidades para ajudar o pesquisador no trabalho com os dados. Relaciona-se com as cláusulas 5 e 8 da Política;
- c) Administração dos dados (*Data Stewardship*): consiste em ferramentas e serviços que ajudam na descrição, depósito e contínuo gerenciamento de dados de pesquisas finalizadas. As atividades do Edinburgh DataShare fazem parte desta área estratégica. Relaciona-se com as cláusulas 6, 7, 9 e 10 da Política;
- d) Suporte ao gerenciamento dos dados (*Data Management Support*): abrange a consultoria e os serviços de suporte em geral. Relaciona-se com as cláusulas 1, 2 e 4 da Política.

Figura 5 – Áreas estratégicas da implementação da política de gerenciamento de dados de pesquisa da University of Edinburgh



Fonte: Adaptado e traduzido de University of Edinburgh (2014b, p.2) pela autora.

Essa inserção do Edinburgh DataShare no contexto da University of Edinburgh do gerenciamento dos dados de pesquisa e o seu papel como parte de um projeto piloto no Reino Unido sobre repositórios institucionais de dados de pesquisa (que disponibilizou os resultados desse projeto livremente na Internet) está de acordo com o que Leite (2009) afirma sobre os repositórios institucionais em geral, que devem ter funções tanto no contexto da instituição quanto no sistema de comunicação científica global.

Padrões de interface: Não foram encontradas diretrizes separadas das práticas neste caso. Há diferentes níveis de interface, com diferentes permissões para administradores, depositantes e visitantes, conforme as autorizações dos usuários, o que é um dos tópicos abordados por Tomaél e Silva (2007). O mais próximo de indicação de acessibilidade foram as indicações de compatibilidade com os padrões do WC3 de validade do XHTML 1.0 e do CSS. A observação da usabilidade foi com relação aos depositantes-pilotos, que resultaram em melhorias no processo de depósito.

Neste caso provavelmente será importante a condução de estudos futuros de uso do repositório e usabilidade para o usuário final. Como depois do período de observação formal deste estudo foram disponibilizadas para visualização no repositório as estatísticas para cada item, coleção e comunidade, esse é um novo dado que abre novas possibilidades de estudo de acesso aos metadados e downloads dos itens, além dos estudos altmétricos, que poderão ser efetuados em outro momento, para verificar o uso. A usabilidade poderá ser estudada tanto

pessoalmente ou através de formulários digitais, cabendo aos administradores do repositório escolherem o que melhor satisfaz suas necessidades.

Padrões de interoperabilidade: A interoperabilidade é, segundo Leite (2009), um dos fatores que caracterizam um repositório institucional. Com relação à interoperabilidade política, foi possível ver tanto nas diretrizes quanto na prática um comprometimento com as políticas do Reino Unido para o tratamento dos dados de pesquisa, sendo a University of Edinburgh uma das universidades parceiras na construção dessas políticas nacionais. A divulgação de suas políticas de forma ampla e bem documentada na Internet pode ser considerada também uma expressão da interoperabilidade política, seguindo a definição de Sayão (2007), que a relaciona a tornar as informações amplamente disponíveis.

Sobre a interoperabilidade técnica, o repositório permite coleta dos metadados (harvesting) através do protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), e é interoperável com sistemas internos e externos à University of Edinburgh.

Sobre interoperabilidade semântica, a adoção de um esquema de metadados está descrita nas diretrizes, e pôde ser verificada nas práticas, pois os metadados estão em Dublin Core qualificado, mais exatamente foi criado um perfil de aplicação específico para a descrição de conjuntos de dados.

A adesão a esses padrões de interoperabilidade garante uma confiabilidade maior no repositório, pois demonstra uma preocupação com a inserção em um cenário científico maior do que a instituição na qual foi criado.

5.1.5 Preservação digital

A preservação digital é um tópico que causa algumas controvérsias com relação a repositórios em geral. Harnad (2006), para tentar esclarecer a polêmica sobre preservação digital nos repositórios, explicou que existe a preservação com p minúsculo e a Preservação com P maiúsculo, sendo que a primeira é realizada pelos repositórios como um meio para atingir seu fim, que é a disponibilização imediata de conteúdo, e a segunda é um fim em si mesma.

Márdero Arellano e Leite (2009) continuam a explicação sobre a preservação digital em repositórios institucionais:

"As ferramentas para repositórios institucionais não são originalmente destinadas a projetos de preservação digital. A motivação primordial não diz respeito à garantia da longevidade dos conteúdos digitais, mesmo que elas acompanhem o desenvolvimento nas áreas de preservação em longo prazo de conteúdos digitais." (MÁRDERO ARELLANO; LEITE, 2009, p. 9)

Assim, esse é o ponto de vista a partir do qual foram observadas as menções à preservação digital encontradas nesta pesquisa. Nesta seção são apresentados os resultados referentes ao plano de preservação e aos identificadores permanentes.

Plano de preservação: Não foi encontrado um plano de preservação, somente uma política de preservação do repositório. A política (Anexo B) prevê que os itens serão retidos indefinidamente e que, em caso do repositório ser fechado, a base de dados será transferida a outro arquivo apropriado. O repositório tentará garantir que os itens continuarão legíveis e acessíveis migrando os arquivos a novos formatos, mantendo o bitstream original junto com os formatos atualizados. Os itens passarão por backups regulares. Itens não são deletados, mas sim removidos da visão pública. Os itens poderão ser retirados da visualização pública por questões de violação de copyright, plágio, requisitos legais, segurança nacional e pesquisa falsificada, por exemplo, mas continuará a existir uma página com um aviso para prevenir os *links* quebrados. Itens recebem um *checksum* (código para conferir a integridade do arquivo) para facilitar a detecção de alterações. Erratas e correções podem ser incluídas no registro original, ou pode ser depositada uma nova versão atualizada. Nesse caso, deve haver *links* entre a versão antiga e a nova, e esta deve estar claramente identificada. A versão mais antiga pode ser retirada da visão pública.

Não foi possível observar na prática a maioria dessas questões. Apenas alguns itens tinham versões mais recentes, e nesses casos foi possível observar os *links* entre as versões, que a versão mais nova estava bem identificada e que a versão mais antiga não havia sido retirada da visão pública, porém não eram versões mais antigas e mais novas por erro, e sim por continuidade. Como exemplos podemos citar os itens ARC-Lake v1.1 – Global (<http://hdl.handle.net/10283/131>), ARC-Lake v2.0 – Global (<http://hdl.handle.net/10283/383>). Nos metadados do item ARC-Lake v1.1 – Global, há o metadado `dc.relation.isreplacedby`,

indicando o handle do ARC-Lake v2.0 – Global, e no item ARC-Lake v2.0 – Global há o metadado dc.relation.replaces indicando o handle do ARC-Lake v1.1 – Global.

Identificadores permanentes: Nas diretrizes é indicado o uso dos identificadores permanentes. No período da observação formal, o repositório utilizava somente o Handle system¹³. A partir do dia 05 de novembro de 2014, foram incluídos Digital Object Identifiers (DOIs)¹⁴ para os itens que já estavam disponíveis no repositório. Além de importante para a preservação digital, o uso de identificadores permanentes é indispensável para a interoperabilidade (SAYÃO, 2007).

Assim, percebe-se que, como os resultados das categorias anteriores, os resultados relacionados à categoria preservação digital mostra que há um processo em nítido andamento no Edinburgh DataShare, que em alguns tópicos está mais desenvolvido e em outros ainda em construção, porém isto não é nenhum demérito, e sim uma característica evidenciada.

5.1.6 Política de acesso e uso

Nesta seção são apresentados os dados referentes à atividade fim do repositório, suas políticas e práticas de acesso ou uso, observadas a partir dos seguintes aspectos: políticas de submissão, materiais de suporte à submissão, recuperação da informação, acesso aos metadados e aos dados de pesquisa, funcionalidades de visualização e análise dos dados e suporte à citação de dados de pesquisa.

Políticas de submissão: somente pessoas com vínculo com a University of Edinburgh podem submeter itens ao Edinburgh DataShare. Podem ser membros atuais ou aposentados, colaboradores ou visitantes que desenvolvam projetos na Universidade ou estudantes, desde que autorizados por seus orientadores. A política estabelece que a responsabilidade do conteúdo do item seja dos autores/depositantes, e que eles têm o direito de estabelecer a data a

¹³ “O Handle System (<http://www.handle.net>) é um sistema distribuído de computadores concebido para assinalar, armazenar, administrar e resolver identificadores ou nomes persistentes de objetos digitais conhecidos como handles” (SAYÃO, 2007, p. 30).

¹⁴ “O Digital Object Identifier (<http://www.doi.org>) é uma aplicação específica do Handle System voltada para a identificação persistente de recursos digitais sobre os quais possam ser atribuídos direitos de propriedade intelectual, bem como para o intercâmbio de informações sobre essas propriedades em um ambiente de rede.” (SAYÃO, 2007, p. 31).

partir de qual o item pode ser acessado (data de embargo). Os administradores do repositório reservam-se o direito de vetar itens que não estejam de acordo com: adequação à regra dos depositantes, relevância para o escopo do repositório, validade do leiaute e do formato e exclusão de spam, bem como de excluir itens se constatada violação de copyright.

Este é um caso em que se percebe que a política (Anexo C) é um tipo de diretriz, e dois quesitos puderam ser observados na prática. O que foi possível verificar neste caso foi a situação da data de embargo, que funcionou corretamente, pois os itens que continham data de embargo (<http://hdl.handle.net/10283/625> - embargo até 31/10/2014; e <http://hdl.handle.net/10283/163> - embargo até 06/11/2018) só permitiam a visualização dos metadados, não o download dos arquivos.

O outro ponto que foi possível observar foi a situação dos depositantes, através do conteúdo do metadado dc.contributor, o qual indicava o nome do depositante. Foram 28 depositantes diferentes. Desses, três são/eram estudantes, nove profissionais ligados aos serviços de informação ou infraestrutura de dados e 16 professores/pesquisadores, tendo vínculo com a University of Edinburgh no período do depósito.

Essa situação está de acordo com o que Green, Macdonald e Rice (2009) e Tomaél e Silva (2007) salientam sobre as políticas dos repositórios, que devem determinar quem pode depositar documentos no repositório.

Materiais de suporte à submissão: Neste caso não foram encontradas diretrizes separadas da prática. Existe uma documentação on-line que serve de suporte aos depositantes: Lista de conferência para o depósito (*Checklist for deposit*); Como depositar (*How to deposit*); Guia de usuário do depositante - Guia do usuário do DataShare (*Depositor's user guide - DataShare User Guide*) – tem ilustrações das telas; Políticas e definições no nível de serviço (*Policies and service level definition*); e Nossas definições (*Ours definitions*). Essa documentação abrange desde as definições básicas do que são dados de pesquisa, as políticas do repositório, e também o passo a passo da solicitação da criação de comunidades e coleções, como fazer o depósito e o preenchimento dos campos do formulário.

A equipe da Data Library está disponível para atendimentos por e-mail conforme a demanda dos depositantes.

Além desses materiais, é disponibilizado pela University of Edinburgh também o curso on-line MANTRA¹⁵, que abrange todo o processo de gerenciamento de dados de pesquisa, não somente o repositório. Esse curso consiste em módulos de treinamento on-line que abordam tópicos como: Dados de pesquisa explicados; Planos de gerenciamento de dados; Organizando dados; Formatos de arquivo e transformação; Documentação, metadados e citação; Armazenamento e segurança; Proteção, direitos e acesso a dados; Compartilhamento, preservação e licenciamento; Tutoriais de manejo de dados; e foram desenvolvidos pela Data Library para apresentar melhores práticas no gerenciamento de dados de pesquisa em diferentes áreas disciplinares. Está disponível através de uma licença CC-BY, que permite sua ampla distribuição, desde que a atribuição à EDINA and Data Library da University of Edinburgh seja informada.

Recuperação da informação: Sobre as diretrizes - há a informação de que o repositório oferece uma seleção de campos de metadados úteis para a descoberta dos conjuntos de dados de pesquisa, através do Google e outras ferramentas de busca, para maximizar a visibilidade e o impacto. Há também a informação de que o repositório oferece as facilidades e busca (*search*) e pesquisa por listas (*browse*). Não há informações sobre suporte à recuperação de informações no repositório.

No período da observação formal, estavam disponíveis as funcionalidades de busca (*search*) e pesquisa por listas (*browse*). As opções de *search*: Search Edinburgh DataShare; Search this Collection. O repositório permite a adição de filtros ao resultado de uma busca simples.

¹⁵MANTRA: <http://datalib.edina.ac.uk/mantra/>

Figura 6 – Opções de busca e pesquisa por listas no Edinburgh DataShare

The image shows two sections of the Edinburgh DataShare interface. The top section, titled "Search Edinburgh Data Share", contains a search input field, a "Go" button, and two radio button options: "Search Edinburgh DataShare" (which is selected) and "This Collection". The bottom section, titled "Browse", lists two categories: "Edinburgh DataShare" and "This Collection". Under "Edinburgh DataShare", the options are "Research Communities", "Date Issued", "Data Creators", "Titles", and "Subjects". Under "This Collection", the options are "Date Issued", "Data Creators", "Titles", and "Subjects".

Fonte: Edinburgh DataShare. <<http://datashare.is.ed.ac.uk/>>, em 09 out. 2014.

No dia 23 de fevereiro de 2015, as opções de pesquisa por lista foram alteradas para incluírem somente as comunidades de pesquisa (Research Communities) e os títulos (Titles), e foi incluída na página inicial do repositório uma opção chamada Discover (descoberta), que é semelhante à pesquisa por lista, mas que apresenta algumas opções visíveis antes de fazer a busca.

Figura 7 – Opção de recuperação da informação Discover no Edinburgh DataShare

Discover	
Date Issued	2015 (23) 2014 (110) 2013 (24)
Data Creator	Remijsen, Bert (126) Gwado Ayoker, Otto (85) Ajuet Deng, Elizabeth Achol (38) ... View More
Subject Classification	African Language studies (87) Musicology (38) Music (22) ... View More
Subject Keywords	oral culture (86) Shilluk (86) Dinka (40) ... View More
Funder	Volkswagen Foundation (86) AHRC - Arts and Humanities Research Council (50) University of Edinburgh (36) ... View More
Type	sound (138) Dataset (41) moving image (21) ... View More
Spatial Coverage	SS (126) South Sudan (119) UK (46) ... View More

Fonte: Edinburgh DataShare. <<http://datashare.is.ed.ac.uk/>>, em 01 mar. 2015.

Algo que é muito instigante neste tipo de estudo é que, apesar de estudar uma “fotografia” de um momento, há a possibilidade de acompanhar as mudanças, e percebe-se mais uma vez a dinamicidade dos objetos de estudo que estão disponíveis on-line. O contínuo aprimoramento é uma marca visível neste objeto de estudo.

Acesso aos metadados e aos dados de pesquisa: As diretrizes informam que qualquer um pode acessar os metadados sem custo, e que podem ser reutilizados em qualquer meio sem solicitação de permissão para propósitos que não visem lucro e não os revenda comercialmente desde que informando o identificador OAI ou um *link* para o registro original dos metadados. Com relação à prática, foi possível acessar os metadados dos itens tanto pela interface de pesquisa quanto pelo protocolo OAI-PMH em <http://datashare.is.ed.ac.uk/dspace-oai/request>, inclusive dos itens que estavam no período de embargo.

A política relativa aos dados de pesquisa destaca que alguns itens têm diferentes permissões/direitos e que há alguns itens com acesso controlado. Na prática, itens que não tinham embargo puderam ser baixados facilmente, existindo inclusive um botão que permite fazer o download de todos os arquivos de um item ao mesmo tempo.

O acesso aos metadados está informado nas diretrizes de acordo com o que autores recomendam para repositórios institucionais –em geral ou específicos de dados de pesquisa (GREEN; MACDONALD; RICE, 2009; TOMAÉL; SILVA, 2007).

Funcionalidades de visualização e análise dos dados: Não foram encontradas funcionalidades de visualização e análise dos dados disponíveis no repositório. Ainda está nos planos da equipe do Edinburgh DataShare realizar estudos para implantar funcionalidades como essas, bem como a visualização de multimídias no repositório.

Um exemplo de funcionalidade de visualização e análise dos dados em um repositório institucional de dados de pesquisa, que ainda está em fase de testes, é da Universidade do Porto, a qual criou uma extensão para o DSpace para quando o usuário está identificado e no item há dados em forma de tabelas, cada tabela tem seus próprios metadados e ela pode ser navegada através de uma interface especial, que possibilita busca e filtro dentro dos dados e a exportação de subconjuntos dos dados sem a necessidade de fazer download (RIBEIRO *et al.*, 2012).

Suporte à citação de dados de pesquisa: Todos os 161 itens têm o metadado de citação, o qual é inserido automaticamente. Nas diretrizes, há a informação de que esse metadado é incluído para ajudar o usuário a cumprir com os requisitos de atribuição (BY) do repositório. Além disso, os registros são coletados (*harvested*) pelo Data Citation Index, da Thomson Reuters.

Abaixo está o modelo que serve para a criação automática do metadado de citação (dc.identifier.citation). Desde a implantação do DOI (após a coleta dos dados desta pesquisa), ele passou a fazer parte desse metadado, sendo incluído no final.

Criador dos dados: Sobrenome, Nome/Iniciais; Sobrenome, Nome/Iniciais. (Date Accessioned – apenas o ANO). Título, Período de tempo [Tipo de item]. Publicador dos dados / Financiador. Nota: Se nenhum nome pessoal é informado, substitua pelo Publicador dos dados¹⁶.

Esse suporte à citação é muito importante para auxiliar o usuário final a cumprir as exigências dos criadores quanto à atribuição dos dados aos criadores. Ao criar o metadado de citação automaticamente também ajuda aos depositantes, pois também diminui a possibilidade de erros na entrada de dados e facilita o processo de depósito ao não fazer com que eles entrem o mesmo dado duas vezes.

5.1.7 Sustentabilidade e Financiamento

Nos documentos utilizados como base para a pesquisa, o apoio financeiro do JISC para a criação do projeto piloto do Edinburgh DataShare, dentro do Jisc Repositories and Preservation Programme é bastante destacado, sendo um resultado do projeto DataShare do Data Information Specialists Committee-UK (DISC-UK), de 2007 a 2009, que explorava formas para os pesquisadores de algumas universidades (Edinburgh, Oxford e Southampton) compartilharem seus dados de pesquisa na Internet.

Como a maioria dos documentos utilizados como base para esta pesquisa tinham como público-alvo os pesquisadores possíveis depositantes de dados de pesquisa no Edinburgh DataShare, a informação mais presente é de que o repositório é free-at-point-of-use, ou seja, o pesquisador da Universidade pode depositar, compartilhar e licenciar seus dados de pesquisa para descoberta on-line e reuso por outros.

¹⁶ Data Creator(s): Family Name, Given Name/Initials; Family Name, Given Name/Initials. (Date Accessioned - YEAR ONLY). Title, Time Period [Item Type]. Data Publisher / Funder. Note: If no personal name is provided, substitute with Data Publisher.

Assim, as informações sobre as estratégias de financiamento ficaram limitadas, bem como quais fatores são considerados importantes para custear a equipe de trabalho e a infraestrutura. Apesar disso, é possível inferir dos dados coletados um compromisso da University of Edinburgh com a comunidade científica, e com o Edinburgh DataShare, que tornou-se um serviço chave em sua iniciativa de gerenciamento de dados de pesquisa (RICE, 2013/2014), principalmente com relação à política de preservação e o acordo do depositante (Depositor agreement) que, apesar de não ser um plano efetivo demonstra o interesse da instituição de dar continuidade ao trabalho com os dados de pesquisa depositados.

Tomaél e Silva (2007, p. [10]) destacam que: “É vital que a instituição reconheça o repositório institucional como um sério compromisso de longo prazo, para a comunidade universitária, bem como para o espaço acadêmico em geral e para o grande público.”.

5.2 Edinburgh DataShare e o gerenciamento de dados de pesquisa

O gerenciamento dos dados de pesquisa gerados por uma instituição deve estar previsto por políticas dessa instituição. Sayão e Sales (2014) ressaltam que a curadoria digital, questões relativas ao arquivamento persistente, preservação digital e modelos de informação para a preservação dos dados científicos são questões-chave devido à relevância dos dados para a ciência atual. Sobre esse tema, Higgins (2012), destaca:

O gerenciamento dos dados é facilitado por uma série de políticas documentadas, estratégias e procedimentos, e pode ser de fato inviável sem eles. Políticas no nível organizacional ou no nível do financiador das pesquisas sustentam as atividades necessárias, assegurando que elas estão inseridas em um fluxo de trabalho da instituição de pesquisa ou da universidade. (HIGGINS, 2012, p. 20)

Ou seja, sem políticas bem documentadas, disseminadas dentro da instituição e com estratégias e procedimentos estabelecidos é pequena a chance de um projeto de gerenciamento de dados de pesquisa funcionar.

Como apresentado nos resultados, na categoria Padrões, na parte sobre o fluxo de trabalho, o Edinburgh DataShare faz parte de uma política institucional da University of Edinburgh para o tratamento de seus dados de pesquisa. A Política de gerenciamento de dados de pesquisa da Universidade apresenta dez itens que estão sendo implantados através de quatro áreas estratégicas: planejamento do gerenciamento de dados (Data Management Planning), infraestrutura de dados ativos (Active Data Infrastructure), administração dos dados (Data Stewardship) e suporte ao gerenciamento dos dados (Data Management Support). A estratégia à qual o Edinburgh DataShare está relacionada é a Administração dos dados, e refere-se ao papel de armazenar e permitir o acesso aos dados após o término da pesquisa.

Estando bem inserido no contexto institucional do gerenciamento dos dados de pesquisa da University of Edinburgh, o Edinburgh DataShare pode oferecer contribuições nesse processo, estando continuamente se renovando e adquirindo novas características para suprir as necessidades de sua comunidade.

Analisando o Edinburgh DataShare dentro do processo de gerenciamento dos dados de pesquisa da University of Edinburgh usando o ponto de vista do Modelo do ciclo de vida da curadoria digital do Digital Curation Centre (HIGGINS, 2008), apresentado com a tradução de Sayão e Sales (2012) na seção 3.4, tem-se os seguintes resultados:

5.2.1 Ações sequenciais

- a) *Conceitualização*: Essa fase não está ligada diretamente ao repositório. Está conectada ao suporte que a Universidade fornece aos seus pesquisadores no processo de planejamento do gerenciamento dos dados de pesquisa, com o curso MANTRA e com a área estratégica de planejamento;
- b) *Criação e/ou recebimento*: Abrange a criação dos dados e dos metadados relacionados. Tem relação com a área estratégica de infraestrutura de dados ativa, pois é para os dados que estão em uso;
- c) *Avaliação e seleção*: Funciona como o processo de desenvolvimento de coleções nas bibliotecas, segundo Higgins (2012), sendo necessário ter atenção às políticas de informação e às exigências legais. Nesta ação há conexões tanto com o repositório institucional de dados de pesquisa quanto com outras ferramentas internas de controle institucional dos dados de pesquisa. (Data Vault – Cofre de dados);
- d) *Arquivamento*: Esta é a ação na qual o repositório entra definitivamente, como curador dos dados, mas sua principal função aparece mais adiante. Nesta ação há a transferência dos dados finais para o repositório, sendo relacionada com a terceira área estratégica, a administração dos dados;
- e) *Ações de preservação*: Aqui já está conectado ao repositório e às outras iniciativas da Universidade. Relaciona-se às características de preservação, mas também aos metadados e ao conteúdo, principalmente quando se refere aos formatos de arquivo;
- f) *Armazenamento*: Relaciona-se ao hardware e software utilizado, bem como aos padrões, interoperabilidade, URLs permanentes, backups e formas seguras de armazenamento;

- g) *Acesso, uso e reuso*: Essa é a ação que reflete a principal função do repositório, sua finalidade. Tem relação com o controle de acesso, de depósito, mas também às facilidades para compartilhar e promover o uso dos dados de pesquisa disponibilizados;
- h) *Transformação*: Criação de novos dados a partir do original, criando subconjuntos, migração para diferentes formatos, não foi possível observar ainda.

5.2.2 Ações para todo o ciclo de vida

- a) *Descrição e representação da informação*: Essa ação está relacionada a todo o ciclo de vida, pois deve ser considerada desde o planejamento do gerenciamento dos dados. No caso do Edinburgh DataShare, em vários itens há arquivos com metadados específicos, além dos metadados em Dublin Core qualificado do repositório. Por ser um repositório institucional que abrange diversas áreas do conhecimento, não há metadados que contemplem todas as especificidades de cada área;
- b) *Planejamento da Preservação*: Envolve mais do que o repositório, por enquanto o repositório tem uma política de preservação, mas a Universidade está trabalhando em estratégias mais abrangentes com relação ao gerenciamento de dados de pesquisa;
- c) *Participação e monitoramento*: O desenvolvimento dessa ação é demonstrado pela participação da University of Edinburgh em políticas do Reino Unido na questão da curadoria digital, e pelo envolvimento de representantes dos pesquisadores e das equipes de suporte (Information Services) no projeto interno de gerenciamento de dados de pesquisa da Universidade.

5.2.3 Ações ocasionais

- a) *Eliminação*: Não foi possível observar, talvez somente se ocorrer algum problema de copyright, conforme a política de preservação do repositório;

- b) *Reavaliação*: Também não foi possível observar. Provavelmente necessitaria de outra estratégia para observação, como uma entrevista para saber se um dado foi reavaliado.
- c) *Migração*: Como as outras ações ocasionais, talvez seja necessário mais tempo para poder observar a ocorrência dessa ação.

Enfim, o Edinburgh DataShare não está isolado das demais iniciativas da instituição, mas é um componente chave em torno do qual outras ideias foram se agregando. Essas atividades culminaram na elaboração de uma política institucional de gerenciamento de dados de pesquisa e outros serviços que ajudam os pesquisadores a efetuar a curadoria de seus dados de pesquisa visando sua preservação, compartilhamento e reuso.

Higgins (2012) descreve os três estilos de gerenciamento de dados de pesquisa que podem ser seguidos por uma organização:

1. O gerenciamento dos dados é realizado pelo grupo de pesquisa durante todo o ciclo da curadoria e os dados permanecem sob sua curadoria quando a pesquisa é finalizada;
2. Um profissional da informação mantém uma visão geral das atividades de gerenciamento dos dados dentro da instituição, bem como recomendações de boas práticas, enquanto os dados são gerenciados e mantidos pelo grupo de pesquisa;
3. O gerenciamento dos dados é realizado pelo grupo de pesquisa enquanto a pesquisa está em andamento, e depois os dados são transferidos para um curador como a biblioteca da instituição, um repositório temático ou uma biblioteca de dados (data library).

Assim, a University of Edinburgh parece estar se encaminhando para um tipo de gerenciamento semelhante ao terceiro estilo, sem contudo deixar de recomendar boas práticas e fornecer suporte aos pesquisadores e estudantes sobre o gerenciamento de seus dados de pesquisa, através de páginas web com guias sobre gerenciamento de dados de pesquisa¹⁷, o curso on-line MANTRA, workshops, cursos e treinamentos presenciais sobre dados de pesquisa e atendimento através da web, e-mail, e telefone.

¹⁷Página web: <http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management>

Figura 8 – Continuum do compartilhamento de dados de pesquisa



Fonte: Adaptado e traduzido de Rice (2007) pela autora.

Através da Figura 8, Rice (2007) apresenta o *continuum* do compartilhamento dos dados de pesquisa, que faz um resumo dos tipos de gerenciamento e distribuição de dados de pesquisa, desde a forma mais típica, em HDs pessoais até uma forma mais elaborada, com computação de alta performance, com dados em rede, sendo compartilhados inclusive máquina-à-máquina.

Os repositórios institucionais ficam entre a disponibilização dos metadados com itens em embargo e o acesso aberto com o download imediato dos arquivos. A meta do projeto DataShare (do qual o Edinburgh DataShare era um dos participantes) era subir um pouco mais nesse *continuum* e formar uma rede de repositórios, com metadados de qualidade assegurada, guias disponíveis para os depositantes, anonimização/consentimento para compartilhamento obtido dos sujeitos, documentação minuciosa sobre criação dos dados e metodologia inclusa, IDs permanentes, formatos validados e adequados para distribuição, comprometimento com a preservação baseada em migração.

Com relação à meta do projeto DataShare, há ainda alguns aspectos que poderiam ser melhorados com relação aos metadados, principalmente com os de assunto, que não são obrigatórios e nem os de classificação estão padronizados. Contudo, os outros pontos estão muito bem contemplados:

- a) Há documentação detalhada disponível para os depositantes, como os guias e a Universidade também disponibiliza o curso on-line MANTRA que traz explicações sobre como gerenciar dados de pesquisa;
- b) Ainda estão sendo feitos estudos sobre a anonimização/consentimento, principalmente na área de psicologia, mas não estão sendo disponibilizados no repositório itens nessa situação;
- c) Os pesquisadores que estão depositando dados têm depositado arquivos com documentação específica sobre os dados de pesquisa;
- d) Há IDs permanentes no repositório (Handles e DOIs);
- e) Há recomendações sobre os formatos de arquivo ;
- f) A política de preservação e o *Depositor Agreement* preveem a migração dos arquivos.

Enfim, o Edinburgh DataShare demonstra estar se encaminhando para um papel mais forte dentro da instituição e um repositório institucional de dados de pesquisa mais consolidado no contexto do compartilhamento de dados de pesquisa.

5.3 Edinburgh Datashare e o novo modelo de comunicação científica

Os dados de pesquisa são um dos produtos da pesquisa científica, junto com os trabalhos publicados. Fazem parte do ciclo da comunicação científica, porém até pouco tempo atrás não tinham uma visibilidade fora dos grupos de pesquisa nos quais eram criados, ficando restritos às três primeiras camadas do *continuum* do compartilhamento de dados de Rice (2007).

Com a inserção dos dados nos repositórios de dados de pesquisa, abriram-se novas oportunidades para a comunicação científica, pois os dados passaram a estar disponíveis com identificadores permanentes, custodiados por entidades que se responsabilizam por mantê-los e realizar seu gerenciamento e curadoria, facilitando sua citação e a criação de conexões entre os trabalhos publicados e os dados disponibilizados nos repositórios.

O modelo de comunicação científica de Sales e Sayão (2012), inserido na seção 3.5, que apresenta o modelo tradicional de comunicação científica, alterado pelo reuso dos dados científicos, contempla essas conexões entre dados e publicações, ressaltando o que é conhecido como publicação ampliada.

A função do repositório de dados, segundo esse modelo, está em promover a curadoria, o acesso e o reuso. O Edinburgh DataShare promove o acesso e o reuso dos dados e parte da curadoria, pois outras funções da curadoria são promovidas pela Universidade, conforme visto nas seções anteriores. De acordo com o modelo, os metadados que o repositório de dados deve prover abrangem as condições de semântica, estrutura, fidedignidade, integridade, proveniência, autenticidade e fixidade (permanência) fornecendo dados prontos para o reuso. Seguindo os dados apresentados nas seções anteriores, os metadados do Edinburgh DataShare abrangem essas condições.

A inclusão de dados em repositórios de dados de pesquisa e a submissão dos manuscritos aos periódicos fazem parte desse modelo. Após a aprovação, os autores sugerem a submissão do “[...] *pos-print* em um Repositório 2 (Institucional ou Temático), que interoperando com o Repositório 1 gerará um terceiro tipo de documento chamado ‘documento ampliado’, ou seja, exatamente a soma do documento avaliado com os dados que o geraram.” (SALES; SAYÃO, 2012, p. 132). Assim se faz a publicação ampliada e se permite o reuso e a verificação da pesquisa científica, através do acesso imediato aos dados de pesquisa, no momento da leitura de um trabalho científico.

Apresenta-se, a seguir, um exemplo encontrado no Edinburgh DataShare de um conjunto de dados, Symmetry operators Mathematica notebook, do criador Thomas Backdahl (<http://hdl.handle.net/10283/541>). Esse item contém arquivos que servem para verificação dos resultados de um artigo escrito por ele e outros dois autores, conforme informações inseridas no metadado dc.rights.

Ao olhar a página da revista na qual o artigo foi publicado (Figura 9), ela tem os dados de pesquisa como dados suplementares, e uma nota de ver também - *See also* - ligando o artigo aos dados disponíveis no repositório.

Figura 9 – Artigo da revista Classical and Quantum Gravity

The screenshot shows the IOPscience website interface. At the top, there is a search bar with the text 'Pesquisar' and a navigation menu with 'Journals', 'Books', and 'Login'. Below the navigation bar, the journal title 'Classical and Quantum Gravity' is displayed. The article information includes the volume and issue: 'Classical and Quantum Gravity > Volume 31 > Number 13', the authors 'Lars Andersson et al 2014', and the DOI '10.1088/0264-9381/31/13/135015'. The article title is 'Second order symmetry operators', marked as 'IOPSELECT'. The authors listed are 'Lars Andersson¹, Thomas Bäckdahl² and Pieter Blue²'. There are options to 'Tag this article', 'PDF (551 KB)', and 'View article'. Below the article information, there are tabs for 'Abstract', 'References', 'Supplementary Data', and 'Metrics'. The 'Supplementary Data' tab is active, showing a download icon and the text 'Supplementary data. (1 MB Mathematica notebook)'. A red box highlights the text 'See also: hdl.handle.net/10283/541'.

Fonte: <http://iopscience.iop.org/0264-9381/31/13/135015/media>

Verificou-se que o artigo estava depositado em um repositório temático da área de Física, o arXiv.org, e a versão dois do artigo ficou exatamente caracterizada como a publicação ampliada do modelo de Sales e Sayão (2012), pois é um artigo depositado em um repositório temático depois da aceitação do artigo pela revista, e tem o vínculo com os dados depositados no repositório institucional de dados de pesquisa (Figura 10).

Figura 10 – Artigo depositado no arXiv.org (Versão 2)

The screenshot shows the arXiv.org interface for the article 'Second order symmetry operators'. The browser address bar shows 'arxiv.org/abs/1402.6252'. The page header includes the Cornell University Library logo and the text 'arXiv.org > gr-qc > arXiv:1402.6252'. The article title is 'Second order symmetry operators' by Lars Andersson, Thomas Bäckdahl, and Pieter Blue. The submission date is 'Submitted on 25 Feb 2014 (v1), last revised 19 Jun 2014 (this version, v2)'. The abstract describes systematic calculations in spinor language for the conformal wave equation, Dirac-Weyl equation, and Maxwell equation on a curved four-dimensional Lorentzian manifold. The comments section includes a red box around the text 'Mathematica notebook available from this http URL'. The subjects are 'General Relativity and Quantum Cosmology (gr-qc), Analysis of PDEs (math.AP)'. The MSC classes are '58J70, 35Q61, 83C60, 83C50, 83-04'. The journal reference is 'Class. Quantum Grav. 31 135015 (2014)'. The DOI is '10.1088/0264-9381/31/13/135015'. The cite as information is 'arXiv:1402.6252 [gr-qc] (or arXiv:1402.6252v2 [gr-qc] for this version)'. The submission history shows it was submitted by Thomas Bäckdahl on 17.42.06 GMT (32kb) from hdl.handle.net/10283/541.4. The source is 'Fonte: http://arxiv.org/abs/1402.6252'.

Ao clicar no *link*, abre o item Symmetry operators Mathematica notebook, do repositório Edinburgh DataShare (Figura 11), confirmando a vinculação e a situação de publicação ampliada. Assim, pode-se inferir que o Edinburgh DataShare tem condições de servir como um repositório de dados de pesquisa adequado ao novo modelo de comunicação científica ao prover identificadores permanentes para fazer as ligações entre os dados de pesquisa e as publicações, e gerar as publicações ampliadas.

Dos 161 itens desta pesquisa, 16 têm o metadado `dc.relation.isreferencedby`, e dois desses itens têm duas ocorrências dele que apontam para documentos diferentes.

Além do documento e do item que foi utilizado como modelo, apenas dois outros documentos (vinculados a dois itens diferentes) apresentam a URL direta do item do Edinburgh DataShare. Um terceiro documento cita a existência dos dados no Edinburgh DataShare mas o *link* é para o repositório como um todo, não para o item relacionado.

Quatro dos documentos têm os dados de pesquisa vinculados como material suplementar, porém não têm *links* para o Edinburgh DataShare.

Figura 11 – Item do Edinburgh DataShare

Symmetry operators Mathematica notebook

 Download Dataset 

 [Mostrar registro completo](#) 

Título:	Symmetry operators Mathematica notebook
Autor:	Backdahl, Thomas
Data:	2014-04-17
Citation:	Backdahl, Thomas. (2014). Symmetry operators Mathematica notebook, [Software]. http://dx.doi.org/10.7488/ds/25 .
Dataset Description (abstract):	This a collection of files intended for verification of the results in the paper Andersson, L, Bäckdahl, T, Blue, P, "Second order symmetry operators", (2014) arXiv:1402.6252 [gr-qc] to appear in Class. Quantum Grav.

Arquivos	Tamanho	Formato	Visualização
SymOpMaster.zip	2.546Mb	application/zip	 Download 

Os arquivos de licença a seguir estão associados a este item:

 Licença original

Este item aparece na(s) seguinte(s) coleção(s)

Fonte: <http://hdl.handle.net/10283/541>

Enfim, ainda é necessário um trabalho de suporte junto aos pesquisadores para fortalecer o uso das URLs permanentes dos itens do Edinburgh DataShare nas publicações depositadas em repositórios, e até nos periódicos e anais de eventos, visando a estruturação delas como publicações ampliadas.

6 CONSIDERAÇÕES FINAIS

Este estudo analisou as relações entre diretrizes e práticas do repositório institucional universitário de dados de pesquisa Edinburgh DataShare, no contexto do gerenciamento de dados de pesquisa, baseando-se nas seguintes categorias: responsabilidade, conteúdo, aspectos legais, padrões, preservação digital, política de acesso e uso e sustentabilidade e financiamento (TOMAÉL; SILVA, 2007). Consideramos que os objetivos específicos foram alcançados e, através destes, o objetivo geral.

Abordando as características do Edinburgh DataShare quanto à responsabilidade, há a presença de profissionais de tecnologia da informação (TI) e profissionais da informação (com destaque para bibliotecários) na equipe do Edinburgh DataShare. Essa parceria é importante para ambos profissionais e para o repositório, pois permite a complementação de habilidades informacionais diferentes. Cada profissional tem seu espaço de ação/atividades e também seu ponto de vista sobre o usuário final, mas na área de intersecção, no qual se comunicam e interagem é onde ocorre o grande trabalho para o desenvolvimento do repositório.

Sobre o conteúdo, foram examinados os metadados de 161 itens recuperados em 09 de outubro de 2014. Como o repositório inicialmente foi povoado através de um projeto piloto, a maioria dos itens (103 itens) foi criada por um mesmo pesquisador, Bert Remijnsen, da área de assunto de Linguística e Idioma Inglês (Linguistics and English Language), e as análises numéricas do conjunto de itens revelam esse viés. Isso implica que análises estatísticas simples dos itens deste repositório, neste momento, revelariam mais sobre os conjuntos de dados desse pesquisador do que sobre os conjuntos de dados de pesquisa da University of Edinburgh em geral. Ao analisar quantitativamente esses itens, esse fator deve ser indicado.

Sobre os assuntos, há um metadado de assunto livre, o `dc.subject`, e um metadado de assunto controlado, o `dc.subject.classification`, porém nenhum dos dois é obrigatório. Esse fato diminui as possibilidades de recuperação da informação, ainda mais em se tratando de dados de pesquisa, que muitas vezes estão em formatos que não permitem uma busca em texto completo, como ocorre com as publicações formais. Isso pode ocasionar a não recuperação de itens do repositório, o que é um problema, pois a finalidade de um repositório institucional de dados de pesquisa é permitir o acesso, o compartilhamento e o reuso dos dados. Se os conjuntos de dados de pesquisa não forem recuperados, não serão acessados, compartilhados, nem reusados. Então se torna extremamente relevante ampliar ao máximo as possibilidades de recuperação da informação, com qualidade.

Com relação aos aspectos legais, a adoção de licenças abertas é uma opção que o Edinburgh DataShare oferece, com a licença ODC-BY, mas também há a opção de não escolher essa licença e indicar licenças diferentes ou outras informações relativas a copyright no metadado dc.rights. Isso implica em uma liberdade maior na escolha das licenças, porém pode ser uma dificuldade para pesquisadores que não estão acostumados com todas essas opções. Por isso, é importante uma documentação de apoio ao depósito, um formulário com dicas para os depositantes e cursos sobre acesso aberto e dados de pesquisa.

Sobre os padrões, nos documentos que foram utilizados como base do estudo foram recuperadas informações sobre o software, porém não sobre o hardware. O DSpace é a plataforma utilizada para o repositório, e permite a interoperabilidade com outros sistemas internos e externos à Universidade. O uso de um perfil de aplicação do padrão Dublin Core qualificado, específico para conjunto de dados de pesquisa, facilita a recuperação da informação e a interoperabilidade com outros sistemas, por usar um padrão reconhecido mundialmente. Ter um fluxo de depósito documentado e disponível na Internet e estar inserido no fluxo de gerenciamento de dados de pesquisa da Universidade traz segurança e estabilidade para os serviços oferecidos aos pesquisadores.

Conforme nossa observação, concluída em fevereiro de 2015, ainda não há um plano de preservação digital para o repositório. O que existe atualmente é uma política, e ela norteia as ações de preservação que provavelmente resultarão em um plano, visto que alterações no Edinburgh DataShare puderam ser vistas no decorrer deste estudo. O uso de identificadores permanentes até o início de novembro de 2014 era somente do Handle, e a partir disso, passou a ser utilizado também o DOI.

Sobre o acesso e uso, há consistência com a prática de outros repositórios institucionais, ou seja, o depósito é efetuado por pessoas vinculadas à Universidade, de materiais da Universidade, e o acesso aos metadados e à maioria dos materiais é livre para todos, sem necessidade de identificação. A recuperação da informação está evoluindo, pois em fevereiro de 2015 foi incluída uma nova funcionalidade de busca (descoberta) na página inicial, com mais opções de pesquisa para os usuários. Ao oferecer novas formas de recuperação, as possibilidades de acesso, compartilhamento e uso dos conjuntos de dados de pesquisa existentes no repositório são ampliadas, potencializando o alcance dos objetivos do Edinburgh DataShare.

Com relação à sustentabilidade e ao financiamento, poucas informações foram recuperadas, talvez porque os documentos utilizados como base da pesquisa tinham mais foco no depósito e no gerenciamento dos dados de pesquisa, não em como obter financiamento

para a criação de repositórios institucionais de dados de pesquisa. Mas ficou claro, por seu início ter sido por um projeto financiado dentro do Jisc Repositories and Preservation Programme e por ter sido incluído em uma estrutura da Universidade, que é um projeto que exige um grande investimento e que necessita de apoio institucional para assegurar sua manutenção e continuidade.

Pensando no Edinburgh DataShare em relação ao ciclo de vida do gerenciamento de dados de pesquisa, percebe-se que ele é uma parte essencial do gerenciamento de dados de pesquisa da Universidade, mas é importante lembrar: ele não é o único mecanismo existente, e há outros que devem ser utilizados para que a curadoria digital da Universidade tenha sucesso, como os laboratórios de dados ativos, o cofre de dados, e outros ainda em desenvolvimento.

Conforme o modelo de Sales e Sayão (2012), os repositórios de dados de pesquisa têm o papel de complementar o ciclo da comunicação científica, trazendo um elemento que já fazia parte da comunicação científica, porém não era tão disponível quanto antes, ou seja, os dados de pesquisa. Além disso, servem para proporcionar as condições de criação das chamadas publicações ampliadas, com a combinação das publicações tradicionais e dos dados de pesquisa que as geraram, e que permitem tanto verificar os resultados apresentados nas publicações quanto ser base de novas pesquisas, como Sales e Sayão (2012) trazem no seu modelo. O Edinburgh DataShare está cumprindo essa tarefa ao proporcionar os serviços que permitem que os pesquisadores vinculem seus dados de pesquisa às suas publicações através de identificadores permanentes.

Apresentando as relações entre as diretrizes e as práticas do Edinburgh DataShare, no contexto do gerenciamento de dados de pesquisa, foi possível perceber uma relação de construção contínua das diretrizes e das práticas. A política de gerenciamento de dados de pesquisa da University of Edinburgh foi estabelecida após o projeto piloto do repositório de dados de pesquisa. Essa forma de construção traz como consequência algumas diferenças observadas entre as diretrizes e as práticas, visto que nem sempre elas estão no mesmo momento de maturidade. Há diretrizes que são estabelecidas inicialmente, como o padrão de metadados e o fluxo de depósito, mas que sofrem alterações com o tempo, para se adequar às necessidades observadas nas práticas, e há reflexos disso nas práticas como, por exemplo, as inconsistências nos metadados de assunto controlado (dc.subject.classification), que teve formas diferentes de ser preenchido na prática. Há práticas que ainda estão se estabelecendo, como as alterações na recuperação da informação, e as diretrizes a esse respeito também ainda estão sendo construídas. A inserção do Edinburgh DataShare como parte da estratégia da

University of Edinburgh em seu gerenciamento dos dados de pesquisa é de importância para seu posicionamento e continuidade, e já está começando a demonstrar a participação no ciclo de comunicação científica, conforme o modelo verificado neste trabalho.

Outro aspecto metodológico a ser destacado, foi a escolha de realizar uma pesquisa qualitativa com levantamento de documentos e dados na Internet e observação direta do repositório em vez de utilizar questionários e entrevistas. Essa estratégia serviu para explorar as informações já disponíveis na rede e verificar o quanto se poderia obter de informações sem ter de contatar diretamente as pessoas envolvidas no projeto do repositório. Assim, com uma observação formal e o acompanhamento informal foi possível verificar o quão dinâmico é o repositório, pois novas funcionalidades foram implantadas no decorrer do período deste estudo e outras estão em planejamento. Este estudo não termina aqui, o Edinburgh DataShare continuará a ser observado, pois é importante atualizar-se sobre suas novidades. Provavelmente, em outro momento, possam ser realizados questionários ou entrevistas com questões mais direcionadas aos pontos sobre os quais não foi possível recuperar informações na Internet. Além disso, o tipo de estudo que foi realizado traz a possibilidade de fazer sugestões à equipe do repositório.

Acerca da metodologia de estudo, foi válido e importante usar o software NVivo, pois aprender a utilização de uma ferramenta que tem diversas possibilidades para o apoio da pesquisa, é algo relevante na formação acadêmica. Em outros projetos, talvez seja produtivo utilizar o NVivo já na fase de levantamento bibliográfico, pois suas funcionalidades permitem uma melhor organização das informações, que podem ser integradas posteriormente nos momentos de análise.

Enfim, julga-se acertada a escolha do repositório Edinburgh DataShare, tanto pela sua importância no Reino Unido quanto no cenário mundial da curadoria digital. Além disso, ele destaca-se pelo pioneirismo nessa área, pela constante evolução e por seus gestores terem decidido dar um tratamento destacado aos dados de pesquisa sem, contudo, deixar de compartilhar ideias com os colegas do repositório Edinburgh Research Archive, o repositório de publicações científicas de acesso aberto. Outra ação acertada é a inclusão do repositório dentro do contexto institucional do gerenciamento de dados de pesquisa, que o torna oficial e com suporte da instituição. O estudo do Edinburgh DataShare em geral foi empolgante, e foi muito bom ter percebido a dinâmica do repositório nas observações e os traços da constituição do repositório nos metadados dos itens. Foi relevante verificar que documentos disponíveis na Internet podem trazer tantas riquezas de dados ao serem analisados qualitativamente.

Como sugestões ao Edinburgh DataShare, ficariam: o ajuste/correção dos metadados que estão fora do padrão estabelecido, pois ainda são poucos itens na base, e fica mais fácil efetuar correções enquanto não há um grande número de registros em bases de dados; e a inclusão obrigatória do metadado `dc.subject.classification`, que poderá ter diversas utilidades na recuperação da informação, dentre elas, ser utilizado com fins administrativos, de ver as áreas de assunto que estão efetuando depósitos no repositório.

Este trabalho é um ponto de partida que pode servir para basear futuros estudos e escolhas de modos de trabalhar com os dados de pesquisa. Caregnato e Rocha (2013, p. [6]) falam da criação do Centro de Documentação e Acervo Digital de Pesquisa (CEDAP/UFRGS), como um “[...] centro de infra-estrutura para o compartilhamento e curadoria de dados de pesquisa da instituição, digitais natos e decorrentes da digitalização”. A Universidade Federal do Rio Grande do Sul está dando passos muito importantes para esse caminho, e reflexões sobre as práticas e teorias existentes a ajudarão a avançar mais rápido, assim como o reuso dos dados de pesquisa ajudará os pesquisadores.

O compartilhamento dos dados na ciência é fundamental à verificação dos resultados das pesquisas e também para que outros pesquisadores possam reutilizá-los, fazendo novas interpretações desses dados ou ampliando-os ao combiná-los com outros conjuntos de dados. Os repositórios de dados de pesquisa ajudam ao avanço no *continuum* do compartilhamento de dados de pesquisa (RICE, 2007), ao permitir uma estrutura mais organizada, estável e acessível para os dados de pesquisa.

Para as universidades, ter um repositório institucional de dados de pesquisa traz diversos benefícios: a estrutura para compartilhamento dos dados de pesquisa de seus pesquisadores é o principal, mas a organização de seus dados de pesquisa, uma motivação para pensar no gerenciamento dos dados de pesquisa da universidade, e até o aumento da visibilidade da universidade são possibilidades fortes.

Enquanto acadêmica e profissional da informação, estudar um repositório institucional de dados de pesquisa foi desafiador, porém a vontade de entender como um repositório pioneiro de uma universidade tradicional lidava com seus dados de pesquisa foi maior, principalmente por ser bibliotecária e trabalhar no Sistema de Bibliotecas da Universidade Federal do Rio Grande do Sul. A importância de se compartilhar os dados de pesquisa tanto para reduzir o tempo de pesquisa, economizar recursos ou permitir novas combinações de dados tornou-se tão evidente para mim que certamente compartilharei também essas informações o quanto for possível.

REFERÊNCIAS

- ARANO, Silvia *et al.* La comunidad “Recursos y datos primarios” de la Universitat Pompeu Fabra: los repositorios institucionales como infraestructuras científicas: estudio de caso. **Revista Española de Documentación Científica**, Madrid, v. 34, n. 3, p. 385-407, jul.-sep. 2011. Disponível em: <<http://redc.revistas.csic.es/index.php/redc/article/view/704/780>>. Acesso em: 15 dez. 2013.
- BORGMAN, Christine L. The conundrum of sharing research data. **Journal of the American Society for Information Science and Technology**, Maryland, v. 63, n. 6, p. 1059–1078, June 2012. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/pdf>>. Acesso em: 21 nov. 2013.
- BOURDIEU, Pierre. O campo científico. In: ORTIZ, Renato (Org.). **A sociologia de Pierre Bourdieu**. São Paulo: Olho d'água, 2008, p. 112-143.
- CAREGNATO, Sônia Elisa; ROCHA, Rafael Port da. Compartilhamento e curadoria de dados de pesquisa: uma proposta para a ciência produzida na Universidade Federal do Rio Grande do Sul. Conferência sobre Tecnologia, Cultura e Memória, 2013, Recife. **Anais...** Recife: UFPE, 2013. Disponível em: <http://www.liber.ufpe.br/ctcm2013/anais/files/6a.CCDP_PCUFRGS.pdf>. Acesso em: 30 abr. 2014.
- CHALMERS, Alan. **A fabricação da ciência**. São Paulo: UNESP, c1994.
- COSTA, Maira Murrieta; CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.19, n.3, p.189-206, jul./set. 2014. Disponível em: <<http://www.scielo.br/pdf/pci/v19n3/a10v19n3.pdf>>. Acesso em: 04 fev. 2015.
- CRAGIN, Melissa H. *et al.* Data sharing, small science and institutional repositories. **Philosophical Transactions of the Royal Society A**, London, v. 368, n. 1926, p. 4023–4038, Sept. 2010. Disponível em: <<http://rsta.royalsocietypublishing.org/content/368/1926/4023.full.pdf+html>>. Acesso em: 17 ago. 2012.
- CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. Diretriz. In: _____. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos/Livros, 2008. P. 128.
- DATA Asset Framework: implementation guide. October, 2009. Disponível em: <http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf>. Acesso em 25 set. 2012.

DATABIB. **About Databib**. [2013]. Disponível em: <<http://databib.org/about.php>>. Acesso em: 06 ago. 2013.

_____. **Databib: Guidelines for Bibliographers**. Draft April 3, 2012. Disponível em: <http://databib.org/Databib_Rubric_Draft.pdf>. Acesso em: 06 ago. 2013.

DCMI Usage Board. **DCMI Type Vocabulary**. 2010-10-11. Disponível em: <<http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/>>. Acesso em: 15 dez. 2014.

FRAGOSO, Suely; RECUERO, Raquel; AMARAL, Adriana. **Métodos de pesquisa para internet**. Porto Alegre: Sulina, 2011.

FRANÇA, Vera Veiga. O objeto da comunicação/a comunicação como objeto. In: HOHFELDT, Antônio; MARTINO, Luiz C.; FRANÇA, Vera Veiga (org.). **Teorias da comunicação: conceitos, escolas e tendências**. Petrópolis, RJ: Vozes, 2001, p. 39-60.

FRANCO, Maria Estela Dal Pai; AFONSO, Mariangela Rosa; BORDIGNON, Luciane Spanhol. Gestão universitária: qualidade, investigação científica e inovação educacional. **Revista Gestão Universitária na América Latina (GUAL)**, Florianópolis, v. 5, n. 1, p. 83-103, jan./abr. 2012. Disponível em: <https://periodicos.ufsc.br/index.php/gual/article/view/1983-4535.2012v5n1p83/21938>. Acesso em 17 jan. 2014.

GRAY, J. Jim Gray on science: a transformed scientific method. In: HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm: data-intensive scientific discovery**. Washington: Microsoft Research, 2009. Disponível em: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf>. Acesso em: 10 nov. 2014.

GREEN, Ann; MACDONALD, Stuart; RICE, Robin. **Policy-Making for Research Data in Repositories: a guide**. S.l.: DISC-UK, 2009. Disponível em: <<http://www.disc-uk.org/docs/guide.pdf>>. Acesso em: 02 jan. 2014.

HARNAD, Stevan. preservation vs. Preservation [Blog Post]. In: **Open Access Archivangelism** [Blog]. March 3. 2006. Disponível em: <<http://openaccess.eprints.org/index.php/?archives/67-preservation-vs.-Preservation.html>>. Acesso em: 27 ago. 2014.

HERNÁNDEZ-PÉREZ, Tony; GARCÍA-MORENO, María-Antonia. Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios. **El Profesional de la Información**, Barcelona, v. 22, n. 3, p. 259-263, 2013. Disponível em: <http://eprints.rclis.org/19524/1/postprint_Hernandez_Garcia.pdf>. Acesso em: 21 nov. 2013.

HIGGINS, Sarah. Digital Curation: The Emergence of a New Discipline. **International Journal of Digital Curation**, Edinburgh, v. 6, n. 2, p. 78-88, 2011. Disponível em: <<http://dx.doi.org/10.2218/ijdc.v6i2.191>>. Acesso em: 15 dez. 2013.

HIGGINS, Sarah. The DCC Curation Lifecycle Model. **International Journal of Digital Curation**, Edinburgh, v. 3, n. 1, p. 134-140, 2008. Disponível em: <<http://dx.doi.org/10.2218/ijdc.v3i1.48>>. Acesso em: 10 set. 2013.

HIGGINS, Sarah. The Lifecycle of Data Management. In: PRYOR, Graham (Org.). **Managing research data**. Londres: Facet Publishing, 2012. Cap. 2, p. 17-46.

JAPIASSÚ, Hilton; MARCONDES, Danilo. Prática/prático. In: _____. **Dicionário básico de filosofia**. Rio de Janeiro: Zahar, 2008. Disponível em: <http://www.portaldapesquisa.com.br/databases/sites?action=bookpage&publisher=zahar&db=zahar180&book_id=978-85-7110-095-4&chapter=002&page=0223>. Acesso em 17 set. 2014.

KURAMOTO, Hélio. Repositórios institucionais: políticas e mandatos. In: SAYÃO, Luis *et al.* (Org.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: UFBA, 2009, p. 203-217. Disponível em: <<http://www.repositorio.ufba.br/ri/handle/ufba/473>>. Acesso em: 29 maio 2013.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília, DF: IBICT, 2009. Disponível em: <<http://kuramoto.files.wordpress.com/2009/11/repositorios-institucionais-f-leite.pdf>>. Acesso em: 13 set. 2012.

LOPES, Maria Immacolata Vassallo de. **Pesquisa em comunicação**. 10. ed. São Paulo, SP: Loyola, 2010.

LUZI, Daniela *et al.* Enhancing diffusion of scientific contents: Open data in Repositories. **Grey Journal** (TGJ), Amsterdam, v. 8, n.2, 2012. Disponível em: <http://greynet.org/images/GL13-S4P_Luzi_et_al.pdf>. Acesso em: 14 ago. 2013.

MEADOWS, A. J. **A comunicação científica**. Brasília, DF: Briquet de Lemos Livros, 1999.

MEDEIROS, Jackson da Silva; CAREGNATO, Sônia Elisa. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, 2012. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/view/488/373>>. Acesso em: 30 abr. 2014.

OCDE. **OECD Principles and Guidelines for Access to Research Data from Public Funding**. Paris: OCDE, 2007. Disponível em: <<http://www.oecd.org/sti/sci-tech/38500813.pdf>>. Acesso em: 08 ago. 2013.

OLIVER, Kathleen; SWAIN, Robert. Directories of institutional repositories: research results & recommendations. In: IFLA GENERAL CONFERENCE AND COUNCIL. 72nd, 2006 ago. 2006, Seul, Coréia. **Programme and proceedings**. IFLA, 2006. Disponível em: <http://archive.ifla.org/IV/ifla72/papers/151-Oliver_Swain-en.pdf>. Acesso em: 06 ago. 2013.

OPENDOAR. **About OpenDOAR**: Directory of Open Access Repositories. 15 Feb. 2011. Disponível em: <<http://www.opendoar.org/about.html>>. Acesso em: 06 ago. 2013.

PAMPEL, Heinz *et al.* **Making Research Data Repositories Visible**: The re3data.org Registry. PLOS ONE, v. 8, n. 11, nov. 2013. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0078080>>. Acesso em: 21 jan. 2014.

PAVÃO, Caterina Groppo *et al.* Contribución del acceso abierto a la visibilidad de la literatura científica en una institución de educación superior. **e-colabora**. Colômbia, v. 2, n. 3, p. 48-66, 2012. Disponível em: <<http://hdl.handle.net/10183/37188>>. Acesso em: 18 set. 2012.

PAVÃO, Caterina Groppo. **Contribuição dos repositórios institucionais à comunicação científica**: um estudo na Universidade Federal do Rio Grande do Sul. 2010. 151 f. Dissertação (Mestrado)-Universidade Federal do Rio Grande do Sul. Faculdade de Biblioteconomia e Comunicação. Programa de Pós-Graduação em Comunicação e Informação, Porto Alegre, 2010. Disponível em: <<http://hdl.handle.net/10183/20932>>. Acesso em: 13 set. 2012

PROCTER, Rob; HALFPENNY, Peter; VOSS, Alex. Research data management: opportunities and challenges for HEIs. In: PRYOR, Graham (Org.). **Managing research data**. Londres: Facet Publishing, 2012. Cap. 7, p. 135-150.

QIN, Jian; BALL, Alex; GREENBERG, Jane. Functional and Architectural Requirements for Metadata: supporting discovery and management of scientific data. In: International Conference on Dublin Core and Metadata Applications, 12th, 2012, Kuching, Sarawak, Malaysia. **Proceedings**. Singapore: DCMI, 2012. Disponível em: <http://opus.bath.ac.uk/34849/1/107_396_1_PB.pdf>. Acesso em: 12 nov. 2014.

RESEARCH INFORMATION NETWORK. **Stewardship of digital research data - principles and guidelines**. London: RIN, 2008. Disponível em: <<http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf>>. Acesso em: 02 jan. 2014.

RIBEIRO, Cristina *et al.* Repositório de dados na U.PORTO: um fluxo de curadoria suportado numa extensão ao DSpace. In: Conferência Luso-Brasileira sobre Acesso Aberto, 3., 2012, Lisboa. **Apresentações**. Lisboa: Universidade Nova de Lisboa, 2012. Disponível em: <<http://hdl.handle.net/10216/65088>>. Acesso em: 26 set. 2014.

RICE, Robin. **Data Sharing Continuum**, [Gráfico]. DISC-UK, September, 2007. Disponível em: <http://www.disc-uk.org/docs/data_sharing_continuum.pdf>. Acesso em: 12 nov. 2014.

RICE, Robin. Edinburgh DataShare: Reflections from a Data Repository Manager. **ASIS&T Bulletin**, Maryland, v. 40, n.2, p. 39-40, Dec.2013/Jan.2014. Disponível em: <https://www.asis.org/Bulletin/Dec-13/DecJan14_RDAP_Rice.pdf>. Acesso em: 24 out. 2014.

RICE, Robin et al. Implementing the Research Data Management Policy: University of Edinburgh Roadmap. **International Journal of Digital Curation**, Edinburgh. Vol. 6, No. 2, 2011, p. 78-88. Disponível em: <<http://dx.doi.org/10.2218/ijdc.v8i2.283>>. Acesso em: 28 jul. 2014.

RODRIGUES, Eloy *et al.* **Os repositórios de dados científicos: estado da arte.** Versão 1.0. [Portugal]: Projeto RCAAP, julho 2010. Disponível em: <<http://hdl.handle.net/1822/10830>>. Acesso em: 21 ago. 2012.

ROSA, Flávia; GOMES, Maria João. Comunicação científica: das restrições ao acesso livre. In: GOMES, Maria João; ROSA, Flávia (Orgs.). **Repositórios institucionais: democratizando o acesso ao conhecimento.** Salvador: EDUFBA, 2010. p 11-34. Disponível em: <<https://repositorio.ufba.br/ri/handle/ri/616>>. Acesso em: 13 set. 2012.

SALES, Luana Farias; SAYÃO, Luís Fernando. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. esp. 2, p. 118-135, 2012. Disponível em: <<http://dx.doi.org/10.5007/1518-2924.2012v17nesp2p118>>. Acesso em: 12 jun. 2013.

SALES, Luana Farias; SAYÃO, Luís Fernando; SOUZA, Rosali Fernandez. Publicações ampliadas: um novo modelo de publicação acadêmica para o ambiente de e-science. In: Encontro Nacional de Pesquisa em Ciência da Informação, 14., 2013, Florianópolis. **Anais**. Florianópolis: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação, 2013. Disponível em: <<http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/402/198>>. Acesso em: 27 ago. 2014.

SAYÃO, Luís Fernando. Padrões para bibliotecas digitais abertas e interoperáveis. **Encontros Bibli**, Florianópolis, n. esp., 1º sem. 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12nesp1p18/436>>. Acesso em: 16 dez. 2014.

SAYÃO, Luis Fernando; SALES, Luana Farias. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. **Informação & Informação**, Londrina, v. 22, n. 3, 2012. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12224/8586>>. Acesso em: 15 dez. 2013.

SAYÃO, Luis Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS: Revista Eletrônica de Comunicação, Informação, Inovação e Saúde**, Rio de Janeiro, v. 8, n. 2, p. 76-92, jun. 2014. Disponível em: <<http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611/1252>>. Acesso em: 02 set. 2014.

TOMAÉL, Maria Inês; SILVA, Terezinha Elisabeth da. Repositórios institucionais: diretrizes para políticas de informação. In: Encontro Nacional de Pesquisa em Ciência da Informação - ENANCIB, 8., 2007, Salvador. **Anais ...**. Salvador: UFBA, 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT5--142.pdf>>. Acesso em: 17 dez. 2014

TORINO, Lígia Patrícia. **Organização da produção científica em repositórios institucionais**: um parâmetro para a UTFPR. 2010. 150 f. Dissertação (Mestrado profissionalizante)-- Programa de Pós-Graduação em Gestão da Informação, Universidade Estadual de Londrina, Londrina, 2010.

UNIVERSITY OF EDINBURGH. **About Information Services**. Aug 12, 2014. Disponível em: <<http://www.ed.ac.uk/schools-departments/information-services/about>>. Acesso em: 09 fev. 2015.

UNIVERSITY OF EDINBURGH. **Data repository**: DataShare: Our definitions. Sep. 4, 2013. Disponível em: <<http://www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/definitions>>. Acesso em: 03 jul. 2014.

UNIVERSITY OF EDINBURGH. **Edinburgh DataShare [Lealeft]**. Edinburgh: University of Edinburgh, 2012. Disponível em: <http://www.ed.ac.uk/polopoly_fs/1.125129!/fileManager/datashare_leaflet.pdf>. Acesso em: 27 fev. 2014.

UNIVERSITY OF EDINBURGH. **Mission**. Apr 5, 2011. Disponível em: <<http://www.ed.ac.uk/about/mission-governance/mission>>. Acesso em: 09 fev. 2015.

UNIVERSITY OF EDINBURGH. **Research Data Management Policy**. Feb 5, 2015. Disponível em: <<http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy>>. Acesso em: 09 fev. 2015.

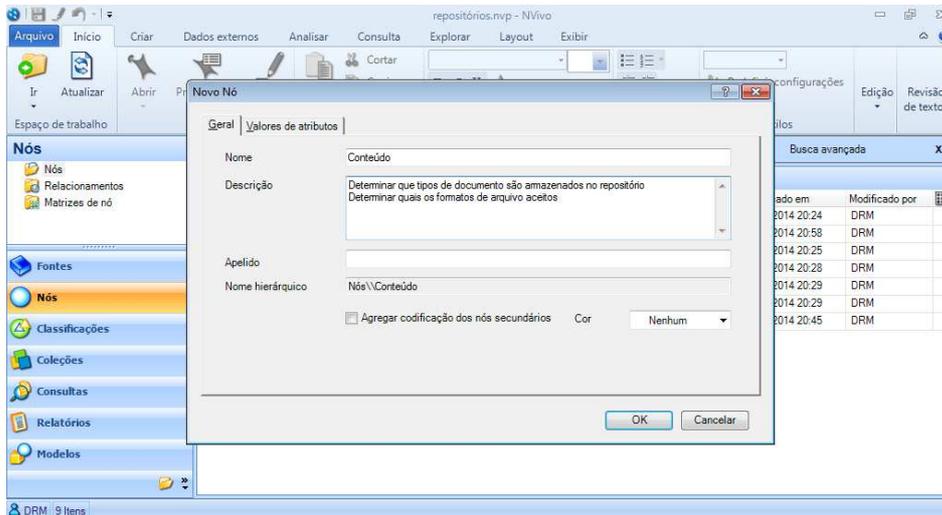
UNIVERSITY OF EDINBURGH. **Undergraduate prospectus**: 2015 entry. Edinburgh: University of Edinburgh, c2014. Disponível em: <http://www.ed.ac.uk/polopoly_fs/1.134097!/fileManager/undergraduate-prospectus-2015-entry.pdf>. Acesso em: 09 fev. 2015.

UNIVERSITY OF EDINBURGH. Information Services RDM Policy Implementation Committee. **Research Data Management (RDM) Roadmap**: August 2012 – May 2015. Version 1.2. January 2014b. Disponível em: <http://www.ed.ac.uk/polopoly_fs/1.116873!/fileManager/UoE-RDM-Roadmap-140106.pdf>. Acesso em: 28 jul. 2014.

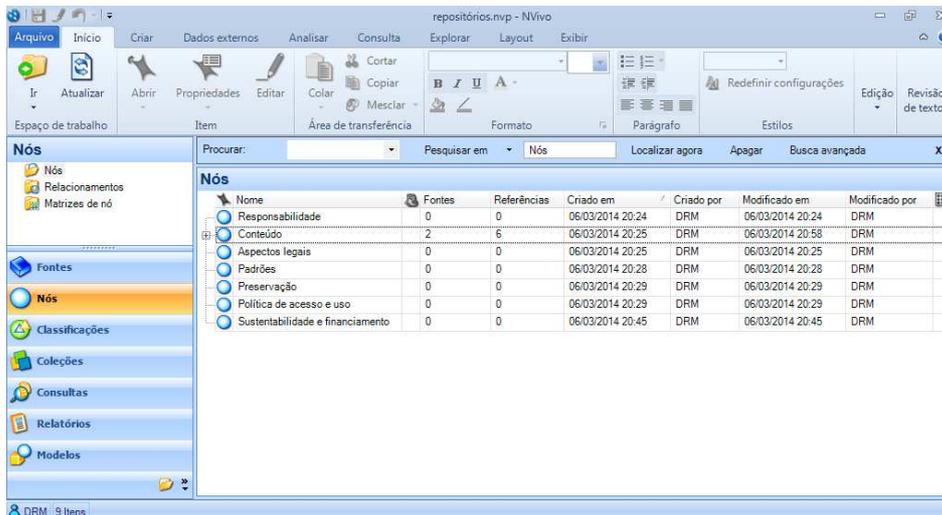
YIN, Robert K. **Estudo de caso**: planejamento e métodos. 4. ed. Porto Alegre: Bookman, 2010.

APÊNDICE A - NVIVO

Nesta primeira etapa, os nós são criados a partir das categorias. No campo Descrição de cada nó foram colocadas as operacionalizações das categorias.

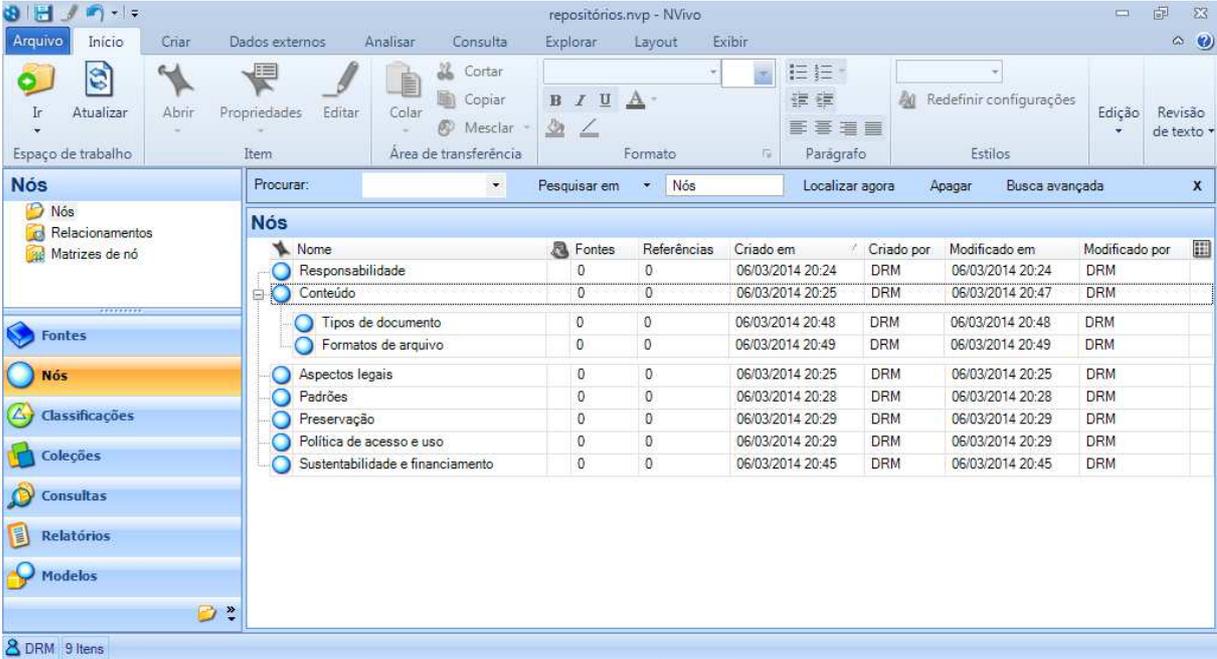


Esta imagem apresenta as categorias gerais em nós no programa NVivo.



Nesta etapa foram criados os sub-nós, baseados nas operacionalizações.

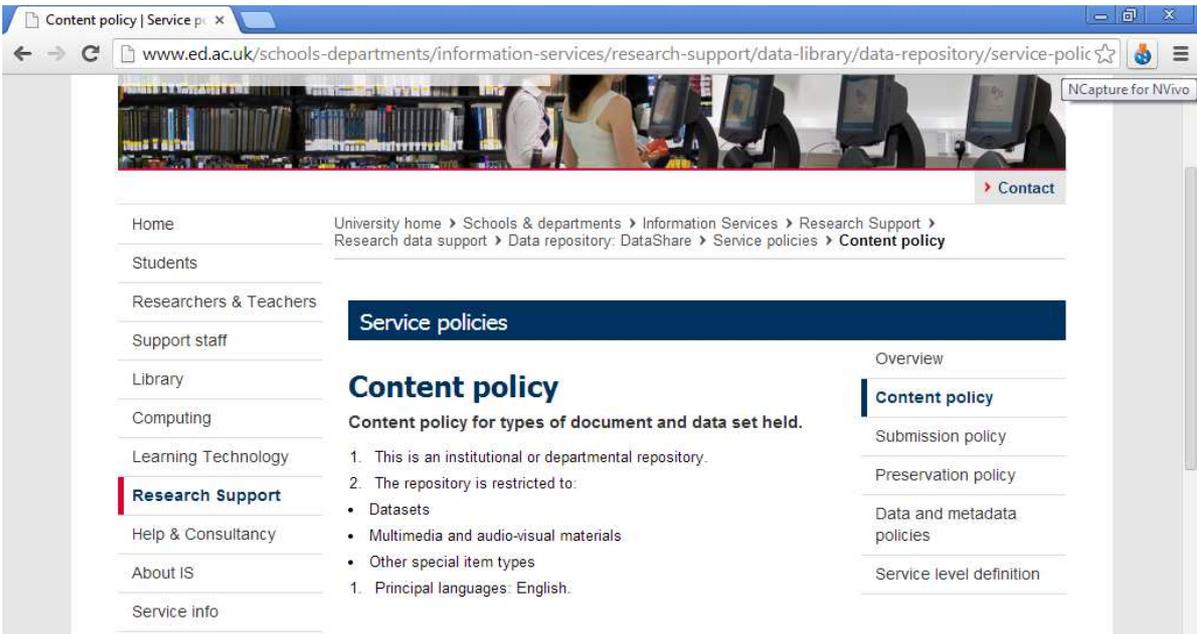
Como exemplo, estão os sub-nós da categoria conteúdo: tipo de documento e formato de arquivo.



The screenshot shows the NVivo software interface with a tree view of nodes. The 'Conteúdo' node is selected and expanded, showing the following sub-nodes:

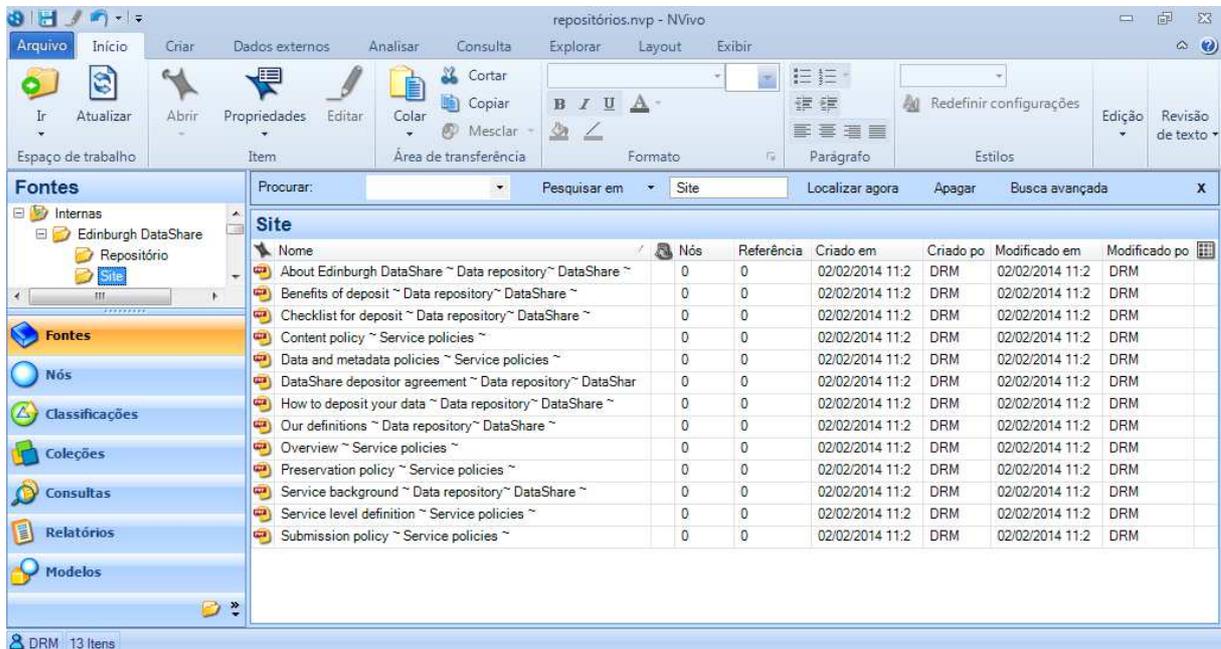
Nome	Fontes	Referências	Criado em	Criado por	Modificado em	Modificado por
Responsabilidade	0	0	06/03/2014 20:24	DRM	06/03/2014 20:24	DRM
Conteúdo	0	0	06/03/2014 20:25	DRM	06/03/2014 20:47	DRM
Tipos de documento	0	0	06/03/2014 20:48	DRM	06/03/2014 20:48	DRM
Formatos de arquivo	0	0	06/03/2014 20:49	DRM	06/03/2014 20:49	DRM
Aspectos legais	0	0	06/03/2014 20:25	DRM	06/03/2014 20:25	DRM
Padrões	0	0	06/03/2014 20:28	DRM	06/03/2014 20:28	DRM
Preservação	0	0	06/03/2014 20:29	DRM	06/03/2014 20:29	DRM
Política de acesso e uso	0	0	06/03/2014 20:29	DRM	06/03/2014 20:29	DRM
Sustentabilidade e financiamento	0	0	06/03/2014 20:45	DRM	06/03/2014 20:45	DRM

A imagem a seguir apresenta o uso do plug-in NCapture, usado para capturar as páginas web para o NVivo, em forma de arquivos PDF.

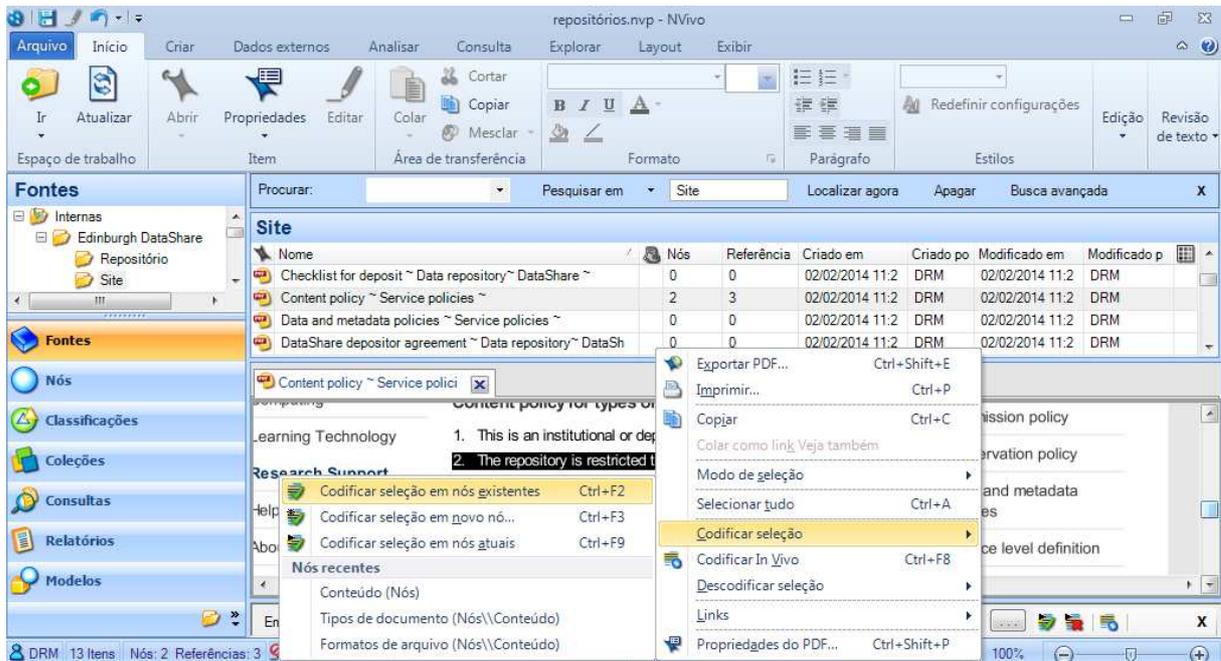


The screenshot shows a web browser window displaying a website page. The URL is www.ed.ac.uk/schools-departments/information-services/research-support/data-library/data-repository/service-policy. The page content includes a navigation menu on the left, a main heading 'Service policies', and a sub-heading 'Content policy'. A small 'NCapture for NVivo' window is visible in the top right corner of the browser window.

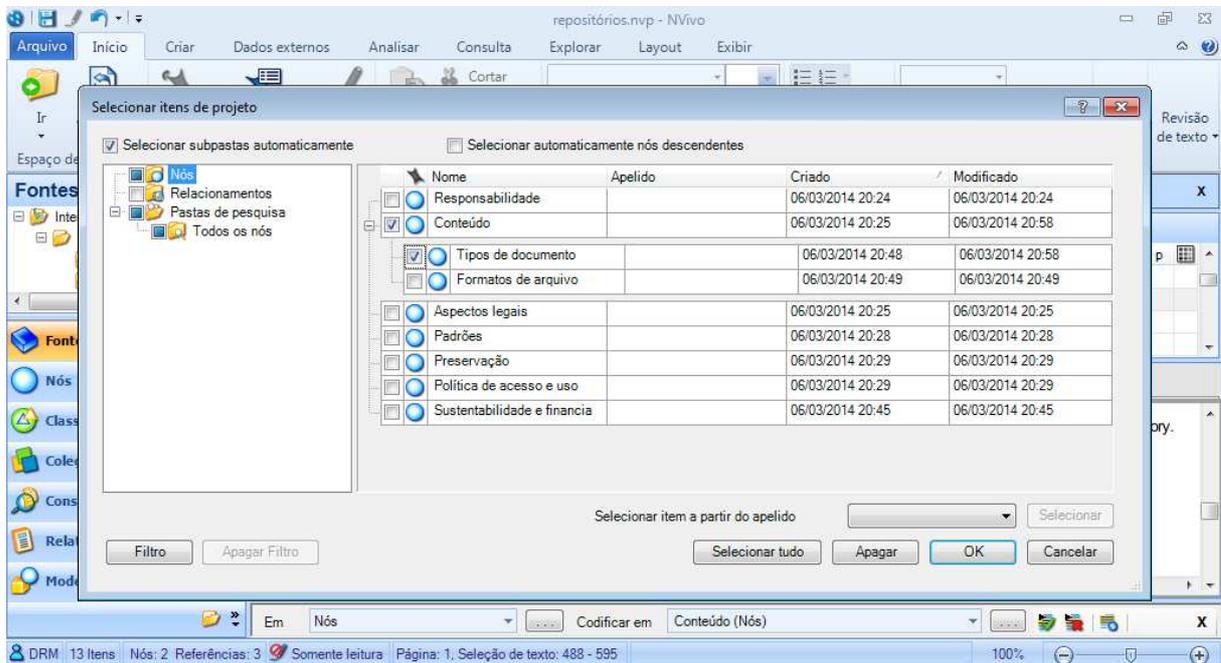
A imagem abaixo apresenta a lista de fontes (do site do Edinburgh DataShare), que foram incluídas no NVivo através do plugin NCapture.



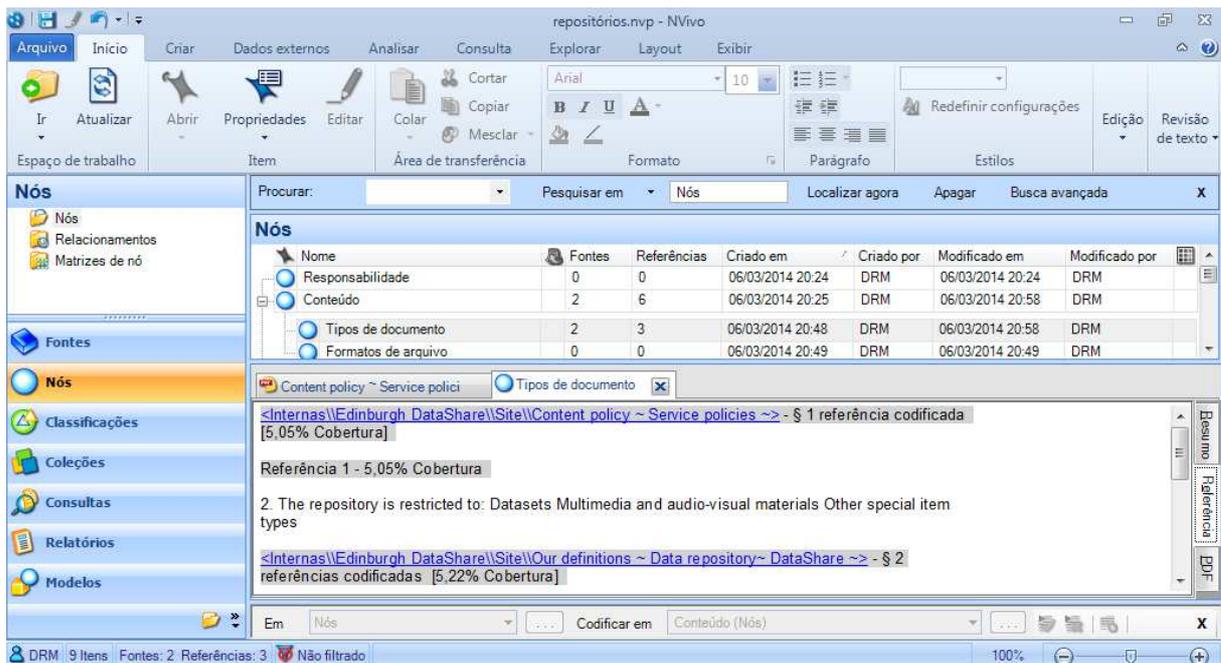
A figura abaixo apresenta a seleção de um texto para a atribuição de nós e sub-nós



A imagem abaixo apresenta a atribuição de um nó e sub-nós ao texto.



A imagem a seguir apresenta a visualização de todos os trechos de material relacionados a um determinado nó, o que serve para facilitar a análise.



A imagem abaixo apresenta a lista de fontes (do site do Edinburgh DataShare), que foram incluídas através do NCapture, com a lista de nós codificados.

The screenshot shows the NVivo interface with the following data in the 'Site' table:

Nome	Nós	Referência	Criado em	Criado po	Modificado em	Modificado po
About Edinburgh DataShare ~ Data repository~ DataShare ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Benefits of deposit ~ Data repository~ DataShare ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Checklist for deposit ~ Data repository~ DataShare ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Content policy ~ Service policies ~	2	3	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Data and metadata policies ~ Service policies ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
DataShare depositor agreement ~ Data repository~ DataShar	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
How to deposit your data ~ Data repository~ DataShare ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Our definitions ~ Data repository~ DataShare ~	2	6	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Overview ~ Service policies ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Preservation policy ~ Service policies ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Service background ~ Data repository~ DataShare ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Service level definition ~ Service policies ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM
Submission policy ~ Service policies ~	0	0	02/02/2014 11:2	DRM	02/02/2014 11:2	DRM

At the bottom of the interface, it shows 'DRM 13 Itens'.

APÊNDICE B - PLANILHA EXCEL

Microsoft Excel interface showing a spreadsheet titled "datashare_ - Microsoft Excel uso não comercial". The spreadsheet contains data with columns labeled A through N. The data is organized into rows, with the first row (row 1) serving as a header for the data below.

Handle	N. Tipo	Tipo	N. Criadores	Criadores	N. Assuntos	Assuntos	N. Arquivos	Extensões dos arquivos	Coleção	Date accessioned	Data available	Licença	dc.rights
571	1	Dataset	0		0		9	pdf; xls; zip	Scottish yo	22/07/2014	22/07/2014	Licença origina	Authors who publish using this dataset
163	1	Dataset	4	Morris, Pa	1	ACT	4	sav; csv; pdf; docx	PHD by Res	04/09/2012	06/11/2018	Licença origina	While this dataset is under embargo pl
343	2	Text; So	1	Mayo, Cat	1	clear spee	11	txt; zip; wav	LISTA (The i	24/09/2013	08/11/2013	Licença origina	License: Creative Commons Attributor
182	1	Sound	9	Remijsen,	3	songs; Dir	16	xlsx; docx; pdf; wav; eaf; j	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
184	1	Sound	9	Remijsen,	3	songs; Dir	5	xlsx; docx; pdf; wav; doc	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
185	1	Sound	9	Remijsen,	3	songs; Dir	9	xlsx; docx; pdf; wav; eaf; j	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
186	1	Sound	9	Remijsen,	3	songs; Dir	12	xlsx; docx; pdf; wav; eaf; j	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
187	1	Sound	9	Remijsen,	3	songs; Dir	89	xlsx; docx; pdf; wav; eaf; j	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
614	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; pdf;	Collection c	19/08/2014	19/08/2014	ODC-BY; Licenç	---
564	2	Text; So	1	Arnold, La	2	Ambel; or	5	wav; eaf; pdf; txt; jpg	The docum	08/07/2014	08/07/2014	ODC-BY; Licenç	---
566	1	Sound	1	Arnold, La	2	Ambel; lai	6	wav; eaf; pdf; txt; jpg	The docum	15/07/2014	15/07/2014	ODC-BY; Licenç	---
615	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; pdf;	Collection c	19/08/2014	19/08/2014	ODC-BY; Licenç	---
548	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; jgg;	Collection c	19/05/2014	19/05/2014	ODC-BY; Licenç	---
549	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; jgg;	Collection c	19/05/2014	19/05/2014	ODC-BY; Licenç	---
550	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; jgg;	Collection c	19/05/2014	19/05/2014	ODC-BY; Licenç	---
188	1	Sound	9	Remijsen,	3	songs; Dir	13	xlsx; docx; pdf; eaf; pfsx;	A Collector	24/09/2012	08/11/2013	ODC-BY; Licenç	---
132	1	Dataset	2	MacCallur	3	Temperat	5	zip; pdf	ATSR Repro	08/12/2011	08/11/2013	ODC-BY; Licenç	---
131	1	Dataset	2	MacCallur	3	Temperat	9	zip; pdf	ATSR Repro	08/12/2011	08/11/2013	ODC-BY; Licenç	---
130	1	Dataset	2	MacCallur	3	Temperat	11	zip; pdf	ATSR Repro	08/12/2011	08/11/2013	ODC-BY; Licenç	---
383	1	Dataset	2	MacCallur	3	Lake; Tem	8	zip; pdf	ATSR Repro	25/10/2013	08/11/2013	ODC-BY; Licenç	---
384	1	Dataset	2	MacCallur	3	Lake; Tem	9	zip; pdf	ATSR Repro	25/10/2013	08/11/2013	ODC-BY; Licenç	---
392	1	Sound	0		1	Phonetics	33	doc; txt; xls; csv; zip	The Univer:	11/11/2013	11/11/2013	ODC-BY; Licenç	---
557	1	Sound	2	Gwado Ay	2	Shilluk; or	6	wav; TextGrid; docx; jgg;	Collection c	04/06/2014	04/06/2014	ODC-BY; Licenç	---
536	1	Dataset	3	MacCallur	4	Sea Surfac	3	nc; pdf	ATSR Repro	17/03/2014	17/03/2014	ODC-BY; Licenç	---

APÊNDICE C - ROTEIRO DE OBSERVAÇÃO

1 INFORMAÇÕES GERAIS

Nome do Repositório: _____

URL: _____

Instituição Responsável: _____

Data de implantação do repositório: _____

Total de conjuntos de dados de pesquisa disponíveis: _____

Data da análise: _____

Comentários:

2 RESPONSABILIDADE (ver se é possível observar)

2.1 Setores responsáveis:

Tecnologia da Informação (TI)

Biblioteca

Outros. Quais? _____

Nomes dos setores responsáveis: _____

Comentários:

2.2 Profissionais responsáveis:

Profissionais de TI

Bibliotecários

Pesquisadores

Especialistas em áreas temáticas

Outros. Quais? _____

Comentários:

3 CONTEÚDO

3.1 Tipos de documento

somente dados de pesquisa

dados de pesquisa com documentação explicativa

outros tipos de documento

Comentários:

3.2 Áreas de assunto encontradas: _____

Comentários:

3.3 Formatos de arquivo encontrados:

- () abertos
() proprietários

Listar formatos de arquivo encontrados: _____

Comentários:

4 ASPECTOS LEGAIS

4.1 Quantidade de licenças adotadas:

- () uma mesma licença para todos os dados de pesquisa depositados
() licenças diferentes para cada conjunto de dados de pesquisa

Comentários:

4.2 Licenças adotadas: _____

Comentários:

5 PADRÕES

5.1 Plataforma tecnológica

- Hardware: _____
- Software: _____

Comentários:

5.2 Metadados

Quais padrões? _____

Quais campos? _____

Como é o formulário de auto-arquivamento? _____ (ver se é possível observar)

Comentários:

5.3 Fluxos de trabalho (ver se é possível observar)

Gestão do repositório: _____

Publicação: _____

Comentários:

5.4 Interface

Há estudos de usabilidade? Resultados: _____

Recursos de acessibilidade encontrados: _____

Há níveis diferentes de interface? Quais?

- () submissão
- () administração
- () visitantes

Comentários:

5.5 Interoperabilidade

- () Interoperabilidade política: _____
- () Interoperabilidade semântica: _____
- () Interoperabilidade técnica: _____

Comentários:

6 PRESERVAÇÃO

Há plano de preservação? _____

Destacar os aspectos relevantes do plano de preservação: _____

Comentários:

Há o uso de identificadores permanentes? Quais? _____

Comentários:

7 POLÍTICA DE ACESSO E USO

7.1 Submissão (ver se é possível observar, pois se não permitir depósito de pessoas de fora da instituição, não será possível observar em andamento, apenas teremos dados indiretos, através do metadado dc.description.provenance, caso exista)

Quem pode submeter? _____

Como ocorre a submissão? _____

Terceiros podem submeter itens, com a autorização do autor? _____

Como funciona a questão da atualização e versões dos conjuntos de dados de pesquisa?

Há mediação/avaliação da submissão? Quem realiza?

Há explicações/tutoriais sobre como realizar a submissão?

Há explicações disponíveis sobre o gerenciamento dos dados de pesquisa e sua preparação para disponibilização no repositório? Quais são? _____

Comentários:

7.2. Recuperação da informação

Quais são os tipos de busca disponíveis no repositório:

- pesquisa simples por palavras
- pesquisa avançada (permite o uso de operadores booleanos entre os campos)
- pesquisa nos metadados
- pesquisa em texto completo
- pesquisa por tipo de dados de pesquisa
- pesquisa ou filtro por formato de arquivo
- por autor ou colaborador
- por data de publicação
- por título
- por assunto

Comentários:

7.3 Acesso e reuso

É possível acessar os metadados sem cadastro ou login da instituição?

É possível acessar os dados de pesquisa sem cadastro ou login da instituição?

Há facilidades para visualização e análise dos dados sem a necessidade de fazer o download dos dados? _____

É possível fazer o download dos dados? _____

Há a possibilidade de compartilhamento dos dados de pesquisa através do repositório? Como ocorre? _____

Há explicações ou tutoriais sobre como citar os dados de pesquisa utilizados? Quais? _____

Comentários:

8 SUSTENTABILIDADE E FINANCIAMENTO (ver se é possível observar no repositório)

8.1 Como é tratada a questão da sustentabilidade e do financiamento do repositório?

Comentários:

8.2 Como são custeadas a equipe de trabalho e a infraestrutura? _____

Comentários:

ANEXO A - POLÍTICA DE GERENCIAMENTO DE DADOS DE PESQUISA

Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.
9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

ANEXO B - POLÍTICA DE PRESERVAÇÃO

Preservation policy

1. Items will be retained indefinitely.
2. The repository will try to ensure continued readability and accessibility:
Items will be migrated to new file formats where necessary. It may not be possible to guarantee the readability of some unusual file formats.
3. The repository regularly backs up its files according to current best practice.
4. The original bit stream is retained for all items, in addition to any upgraded formats.
5. Items may be removed at the request of the author/copyright holder.
6. Acceptable reasons for withdrawal include:
Proven copyright violation or plagiarism
Legal requirements and proven violations
National Security
Falsified research
7. Withdrawn items are not deleted per se, but are removed from public view.
8. Withdrawn items' identifiers/URLs are retained indefinitely.
9. URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories, with:
A link to a replacement version, where available
A note explaining the reasons for withdrawal
10. Errata and corrigenda lists may be included with the original record if required.
11. If necessary, an updated version may be deposited:
The earlier version may be withdrawn from public view. There will be links between earlier and later versions, with the most recent version clearly identified.
12. Items are allocated a checksum to facilitate the detection of alterations.
13. In the event of the repository being closed down, the database will be transferred to another appropriate archive.

ANEXO C - POLÍTICA DE SUBMISSÃO

Submission policy

Concerning depositors, quality and copyright.

1. Items may only be deposited by accredited members of the institution, or their delegated agents.
2. The administrator only vets items for the eligibility of authors/depositors, relevance to the scope of the repository, valid layout & format, and the exclusion of spam.
3. The validity and authenticity of the content of submissions is the sole responsibility of the depositor.
4. Items can be deposited at any time, but will not be made publicly visible until any publishers' or funders' embargo period has expired.
5. Any copyright violations are entirely the responsibility of the authors/depositors.
6. If the repository receives proof of copyright violation, the relevant item will be removed immediately.

ANEXO D - TABELA DE FORMATOS DE ARQUIVO UK DATA ARCHIVE

FILE FORMATS TABLE

The UK Data Archive works with different data formats for different purposes. There are optimal data formats that are used for long-term preservation of data.

This table contains guidance on file formats accepted by the UK Data Archive for deposited data. We welcome queries from researchers about appropriate file formats for working and preservation, particularly early in the research process.

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
<p>Quantitative tabular data with extensive metadata</p> <p>a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data</p>	<p>SPSS portable format (.por)</p> <p>delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) containing metadata information</p> <p>some structured text or mark-up file containing metadata information, e.g. DDI XML file</p>	<p>proprietary formats of statistical packages e.g. SPSS (.sav), Stata (.dta)</p> <p>MS Access (.mdb/.accdb)</p>
<p>Quantitative tabular data with minimal metadata</p> <p>a matrix of data with or without column headings or variable names, but no other metadata or labelling</p>	<p>comma-separated values (CSV) file (.csv)</p> <p>tab-delimited file (.tab)</p> <p>including delimited text of given character set with SQL data definition statements where appropriate</p>	<p>delimited text of given character set - only characters not present in the data should be used as delimiters (.txt)</p> <p>widely-used formats, e.g. MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf) and OpenDocument Spreadsheet (.ods)</p>
<p>Geospatial data</p> <p>vector and raster data</p>	<p>ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn)</p> <p>geo-referenced TIFF (.tif, .tiff)</p> <p>CAD data (.dwg)</p> <p>tabular GIS attribute data</p>	<p>ESRI Geodatabase format (.mdb)</p> <p>MapInfo Interchange Format (.mif) for vector data</p> <p>Keyhole Mark-up Language (KML) (.kml)</p> <p>Adobe Illustrator (.ai), CAD data (.dxf or .svg)</p> <p>binary formats of GIS and CAD packages</p>

Type of data	Acceptable formats for sharing, reuse and preservation	Other acceptable formats for data preservation
Qualitative data textual	eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) Rich Text Format (.rtf) plain text data, ASCII (.txt)	Hypertext Mark-up Language (HTML) (.html) widely-used proprietary formats, e.g. MS Word (.doc/.docx) some proprietary/software-specific formats, e.g. NUD*IST, NVivo and ATLAS.ti
Digital image data	TIFF version 6 uncompressed (.tif)	JPEG (.jpeg, .jpg) but only if created in this format TIFF (other versions) (.tif, .tiff) Adobe Portable Document Format (PDF/A, PDF) (.pdf) standard applicable RAW image format (.raw) Photoshop files (.psd)
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) but only if created in this format Audio Interchange File Format (AIFF) (.aif) Waveform Audio Format (WAV) (.wav)
Digital video data	MPEG-4 (.mp4) motion JPEG 2000 (.mj2)	
Documentation and scripts	Rich Text Format (.rtf) PDF/A or PDF (.pdf) HTML (.htm) OpenDocument Text (.odt)	plain text (.txt) some widely-used proprietary formats, e.g. MS Word (.doc/.docx) or MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0

Fonte: <http://www.data-archive.ac.uk/create-manage/format/formats-table>