



Universidade do Minho

ESTRATÉGIA INSTITUCIONAL PARA A GESTÃO DOS DADOS DE INVESTIGAÇÃO: ESTUDO E RECOMENDAÇÕES

Relatório do grupo de trabalho constituído por:

Rui L. Reis, Vice-reitor (coordenador), Eloy Rodrigues (Serviços de Documentação), Eugénio Ferreira (Centro de Engenharia Biológica), Joana Palha (Instituto de Ciências da Vida e da Saúde), José Manuel Meijome (Centro de Física), Manuela Gomes (Grupo de Investigação 3B's), Miguel Gonçalves (Centro de Investigação em Psicologia), Moisés Martins (Centro de Estudos de Comunicação e Sociedade), Natália Monteiro (Núcleo de Investigação em Políticas Económicas), Paulo Novais (Centro Algoritmi), Pedro Príncipe (Serviços de Documentação).

Setembro de 2017

ÍNDICE

Sumário executivo	3
Introdução	4
1. ESTUDO E ANÁLISE	5
1.1. Gestão de Dados de Investigação na UMinho	5
1.1.1. Curadoria e gestão de dados científicos na UMinho	5
1.1.2. Acervo de dados científicos existentes produzidos na UMinho	9
1.1.3. Diagnóstico a comunidades piloto: práticas de armazenamento e depósito de dados e caracterização dos processos de gestão dos dados	11
1.2. Políticas para dados dos financiadores de Ciência e Inovação	16
1.2.1. Política Nacional de Ciência Aberta	17
1.2.2. Dados de Investigação Abertos no Horizonte 2020	17
1.2.3. Estratégias institucionais	18
1.3. Princípios FAIR para a gestão de dados de investigação	19
2. RECOMENDAÇÕES	20
2.1. Política de gestão de dados na UMinho	20
2.2. Infraestrutura para os dados de investigação	21
2.3. ferramentas de apoio à gestão dos dados	22
2.4. Capacitação	23
2.5. Questões legais, de proteção de dados e de propriedade intelectual	24
Conclusão	26
Quadro síntese das recomendações	26
Diagrama dos componentes da estratégia da UMinho para a Gestão de Dados de Investigação	27

SUMÁRIO EXECUTIVO

O presente relatório apresenta o estudo e as recomendações para o desenvolvimento de uma estratégia institucional na Universidade do Minho para a gestão de dados de investigação. O relatório foi preparado pelo grupo de trabalho nomeado pelo Senhor Reitor da UMinho a 12 de abril de 2017 no despacho nº RT-27/2017. O Grupo iniciou os seus trabalhos no mês de maio com o mandato de *“produzir um estudo e recomendações que sustentem a formulação de uma estratégia institucional relativa aos dados de investigação, incluindo a identificação de políticas, infraestruturas e serviços para a gestão dos dados de investigação na UMinho”*. A metodologia adotada passou pela realização de duas reuniões presenciais, correspondendo às duas fases do trabalho – inicialmente diagnóstico e posteriormente desenho de soluções e formulação de recomendações –, e pela redação deste documento, de forma assíncrona e não presencial.

O relatório reflete a metodologia adotada e encontra-se estruturado em duas partes: 1ª) estudo e análise e 2ª) recomendações. Para a primeira parte, o grupo de trabalho tomou como base o estudo realizado em 2014 pelos Serviços de Documentação da UMinho (SDUM) que fornece informação relevante sobre os dados gerados no âmbito da investigação realizada na UMinho e sobre as práticas de gestão desses dados. Esta parte do estudo foi complementada com a análise dos primeiros resultados do programa de diagnóstico iniciado pelos SDUM e aplicado a seis comunidades piloto no primeiro semestre de 2017 com o intuito de caracterizar os processos de gestão dos dados e avaliar papéis e responsabilidade associados.

Neste relatório analisam-se ainda as políticas para dados dos financiadores de ciência e inovação, com particular enfoque para os requisitos dos dados de investigação abertos no Horizonte 2020 e princípios FAIR para a gestão de dados. Reflete-se ainda sobre a Política Nacional de Ciência Aberta e o desenvolvimento de estratégias institucionais nas universidades.

A secção prospetiva do presente relatório apresenta recomendações estruturadas em cinco tópicos: 1) Política, 2) Infraestrutura e sistemas para dados, 3) Serviços e ferramentas de apoio à gestão dos dados, 4) Capacitação e 5) Questões legais, de proteção de dados e de propriedade intelectual. O relatório desenvolve os componentes e detalhes das seguintes recomendações:

1. Definição de uma política institucional da Universidade do Minho relativa à gestão e partilha dos dados de investigação.
2. Disponibilização de uma infraestrutura de sistemas e serviços integrados a disponibilizar à comunidade da UMinho para a gestão de dados ao longo do ciclo de vida da investigação.
3. Criação de ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados.
4. Desenvolvimento de ações e programas integrados que promovam e valorizem as competências dos investigadores em matéria de gestão e partilha de dados de investigação.
5. Disponibilização e reforço de serviços de apoio e consultoria no domínio da proteção de dados, licenciamento e reutilização de dados e propriedade intelectual.

INTRODUÇÃO

Um crescente número de organismos de financiamento de ciência e instituições de investigação tem vindo a adotar políticas relativas à gestão e partilha dos dados produzidos no âmbito de projetos de investigação. Estas políticas e mandatos fundamentam-se não apenas na preocupação de redução dos custos associados à recolha de dados evitando a sua duplicação, mas também nos aspetos relacionados com a reprodutibilidade da investigação, no quadro da agenda de *Responsible Research and Innovation* (RRI). Adicionalmente, muitos editores de revistas apresentam também aos autores requisitos para a identificação dos conjuntos de dados associados aos artigos propostos para publicação, procurando que estejam acessíveis para consulta em repositórios de dados.

A Comissão Europeia (CE), no quadro das políticas de Ciência Aberta, definiu para o programa Horizonte 2020 uma política de dados abertos que visa melhorar e maximizar o acesso e a reutilização dos dados de investigação gerados pelos projetos que financia. Os requisitos da CE¹ exigem aos investigadores o depósito dos dados de investigação, incluindo metadados associados, necessários para validar os resultados apresentados em publicações científicas, especificando ainda a necessidade de definição de planos para a gestão dos dados produzidos. Os requisitos para os dados abertos no H2020 foram de 2014 a 2016 enquadrados por um projeto-piloto abrangendo apenas algumas áreas científicas do programa, mas a partir de janeiro de 2017 os mesmos requisitos são aplicados por defeito a todos os projetos aprovados no programa-quadro.

Em Portugal, a Fundação para a Ciência e a Tecnologia adotou em 2014, no quadro mais amplo da política de Acesso Aberto, orientações que encorajam os investigadores a disponibilizarem os dados resultantes dos projetos de I&D em bases de dados de Acesso Aberto apropriadas, incentivando ainda os investigadores a promoverem e/ou participarem em iniciativas nacionais e internacionais que procurem as formas mais adequadas de partilha de dados nas diferentes áreas do conhecimento.

Verificada a importância crescente da gestão e abertura dos dados nos atuais processos de investigação e desenvolvimento, enquanto “produto” dotado de valor próprio, a nível internacional um número crescente de entidades dos sistemas científico e de ensino superior tem vindo a adotar políticas relativas aos dados de investigação. Apesar da sua diversidade, essas políticas tendem a incluir e valorizar duas dimensões básicas: os requisitos relativos à gestão de dados de investigação, sobretudo no que tange à elaboração de planos de gestão de dados; e os requisitos relacionados com a partilha e o acesso aberto aos dados produzidos, através de repositórios de dados.

¹ Guidelines on Data Management in Horizon 2020:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

A Universidade do Minho tem protagonizado, ao longo dos últimos 15 anos, um papel pioneiro nos domínios do acesso aberto e da ciência aberta, nomeadamente através do estabelecimento do RepositóriUM (2003) e da definição de políticas de auto-arquivo e acesso aberto das publicações dos seus membros (2005 e 2011). Importa, neste quadro, definir orientações que assegurem a continuidade e o aprofundamento da posição da Universidade nesse domínio.

Neste contexto é indispensável conceber estratégias de atuação institucional que maximizem o valor dos dados gerados nas atividades de investigação desenvolvidas na UMinho e, simultaneamente, facilitem o cumprimento das exigências dos organismos financiadores da ciência. Na conceção de uma estratégia integrada para garantir sistemas e serviços de suporte à gestão de dados científicos nas instituições de investigação e ensino superior desenvolvidos de forma coerente, devem seguir-se três etapas fundamentais: 1) compreender a sua posição atual; 2) definir onde se quer estar no futuro; 3) traçar um programa de atividades para fazer essa transição.

1. ESTUDO E ANÁLISE

1.1. GESTÃO DE DADOS DE INVESTIGAÇÃO NA UMINHO

Em 2014 os Serviços de Documentação da UMinho (SDUM) realizaram um estudo² junto dos investigadores e centros de investigação, no sentido de se obter um melhor conhecimento sobre os dados gerados no âmbito da investigação realizada na Universidade e as práticas de gestão dos dados, procurando avaliar o tipo de produção de dados, identificar as práticas de gestão de dados de investigação e estimar a pertinência de serviços de suporte neste domínio. Este estudo baseou-se num inquérito por questionário dirigido a todos os investigadores e docentes da Universidade do Minho, realizado em março e abril de 2014, tendo sido divulgado junto dos investigadores e centros de investigação exclusivamente através de mensagem de correio eletrónico enviada pelo Vice-Reitor responsável pela investigação. O questionário anónimo foi respondido via web com a recolha de 226 respostas completas.

1.1.1. CURADORIA E GESTÃO DE DADOS CIENTÍFICOS NA UMINHO

O estudo levado a cabo pelos SDUM em 2014 apresentou dados relevantes relativamente às atividades de curadoria e gestão de dados permitindo identificar o tipo e volume dos conjuntos de dados, as práticas de segurança e preservação dos dados e compreender o potencial de reutilização dos dados em novas investigações e o tipo de acesso associado.

Os resultados obtidos relativamente aos dados decorrentes da atividade científica dos investigadores e à forma como são produzidos, revelam uma grande preocupação com a

² Relatório completo do inquérito: <http://hdl.handle.net/1822/29948>; Dados do inquérito: <http://hdl.handle.net/1822/29949>.

necessidade de salvaguarda desses dados (81,0%), tendo, por contraponto, a maioria (80,5%) afirmado não possuir um plano de gestão de dados. De forma sumária e com relevância para o presente relatório destacam-se os seguintes resultados do referido estudo:

Tipos de dados de investigação produzidos:

A maioria dos conjuntos de dados são identificados como dados experimentais (74,8%), seguidos de dados estatísticos (47,6%), de imagens (41,4%) e dados numéricos (36,2%). Os dados multimédia e áudio apresentaram percentagens residuais e nos “Outros tipos de dados” foram referidos os dados cartográficos e os desenhos enquanto registo de escavações.

	N	%
Experimentais	157	74,8%
Estatísticos	100	47,6%
Imagens	87	41,4%
Numéricos	76	36,2%
Observacionais	71	33,8%
Computacionais	64	30,5%
Textuais	63	30,0%
Multimédia	17	8,1%
Áudio	15	7,1%
Outro	5	2,4%
	210	

Tabela 1 – Dados científicos produzidos no contexto da investigação

Volume global aproximado dos conjuntos de dados (datasets) acumulados:

24,8% dos participantes do questionário respondeu não saber exatamente qual o volume global dos *datasets* por si acumulados. 24,3% indicou que o volume dos seus dados de investigação se situa entre 1 – 10 GB, sendo que 21% respondeu que o volume se situa entre 10 – 100 GB. Os respondentes que assinalaram acumular mais de 1 TB representaram, no seu conjunto, 9,5%, sendo que destes mais de metade (5,2% do total) reúne entre 1 e 10 TB.

	N	%
Menor que 1 GB	18	8,6%
1 - 10 GB	51	24,3%
10 - 100 GB	44	21,0%
100 - 1 TB	25	11,9%
1 - 10 TB's	11	5,2%
10 - 100 TB's	4	1,9%
Mais de 100 TB's	5	2,4%
Não sei	52	24,8%
	210	

Tabela 2 – Volume global dos datasets acumulados

Regularidade da realização de cópias de segurança dos dados de investigação produzidos:

81% dos participantes indicam realizar com regularidade cópias de segurança dos dados científicos que produzem. Contudo, a percentagem daqueles que referiram não saber se são realizadas (12,9%) ou mesmo que não são realizadas cópias (6,2%) é ainda significativa.

	N	%
Sim	170	81,0%
Não	13	6,2%
Não sei	27	12,9%
	210	

Tabela 3 – Realização das cópias de segurança dos dados científicos

Frequência das cópias de segurança:

29,4% indicou ser realizado de forma *ad hoc* (ou seja, sem periodicidade definida), 22,4% que os *backups* são realizados mensalmente, 18,8% semanalmente e 11,8% diariamente. 11,8% dos participantes indicou ainda não saber qual a frequência dos *backups*.

	N	%
Ad hoc	50	29,4%
Mensalmente	38	22,4%
Semanalmente	32	18,8%
Diariamente	20	11,8%
Não sei	20	11,8%
Outro	9	5,3%
	170	

Tabela 4 – Frequência dos *backups*

Sistema de armazenamento dos dados científicos:

80% dos participantes indicaram que a preservação é realizada em disco externo, 52,4% em disco do computador pessoal, 32,4% na aplicação *Dropbox*, 23,5% num servidor local, 21,8% numa *USB/Flash drive* e 17,1% num disco de equipamento, sendo as restantes percentagens diluídas de forma residual por outros equipamentos. Na opção Outro foram apontados alguns serviços de armazenamento e sincronização de arquivos como por exemplo o *Google Drive*, o correio eletrónico e o suporte papel.

	N	%
Disco externo	136	80,0%
Disco de PC	89	52,4%
Dropbox	55	32,4%
Servidor local	40	23,5%
USB/ Flash drive	37	21,8%
Disco de equipamento	29	17,1%
CD/ DVD	14	8,2%
Outro	8	4,7%
Servidor internacional	3	1,8%
Servidor nacional	2	1,2%
Tape	2	1,2%
	170	

Tabela 5 – Local de armazenamento dos dados científicos

Existência de plano para a gestão de dados resultantes da sua investigação (por exemplo, política de preservação de dados, política de gestão de registos, estratégia de eliminação de dados):

A grande maioria (80,5%) dos respondentes indicou que não possui um plano de gestão de dados decorrentes da investigação realizada.

	N	%
Não	169	80,5%
Sim	41	19,5%
	210	

Tabela 6 – Plano de gestão de dados científicos

Estatuto de acesso usual dos dados

A maioria indicou que os dados obtidos no âmbito da investigação têm estatuto de acesso restrito ao grupo ou projeto (57,5%) ou mesmo apenas ao investigador (15,5%). Apenas 13,3% responderam que os seus dados ficam imediatamente disponíveis em acesso aberto e 10,3% referiram que os seus dados são de acesso aberto após um período de embargo.

	N	%
Acesso restrito ao grupo/ projeto de investigação	130	57,5%
Acesso restrito ao investigador	35	15,5%
São todos de acesso aberto imediato	30	13,3%
São de acesso aberto após período de embargo (por favor especifique a duração na caixa de texto)	23	10,2%
Outro	8	3,5%
	226	

Tabela 7 – Tipo de acesso dos dados científicos

Preocupações relativamente ao acesso aberto aos dados

As maiores preocupações quanto ao acesso aberto aos dados de investigação indicadas foram as relativas à confidencialidade/direitos de autor (79,0%), à potencial reutilização (44,3%) e a possível má interpretação dos dados (31,0%).

	N	%
Confidencialidade/ direitos de propriedade intelectual	166	79,0%
Potencial reutilização	93	44,3%
Possível má interpretação dos dados	65	31,0%
Tempo/ esforço despendido na sua obtenção	46	21,9%
Valor comercial dos dados	41	19,5%
Outro	10	4,8%
	210	

Tabela 8 – Preocupações relativamente ao acesso aberto

1.1.2. ACERVO DE DADOS CIENTÍFICOS EXISTENTES PRODUZIDOS NA UMINHO

O estudo permitiu recolher informação sobre os dados que os investigadores consideraram essenciais, o esforço associado à sua criação, os períodos de preservação estimados, a frequência de uso, a dimensão e formato dos conjuntos de dados, e ainda que tipo de serviços gostariam que a Universidade do Minho fornecesse. De forma sumária e com relevância para o presente relatório destacam-se os seguintes resultados:

Tipos de dados a preservar:

Em relação ao tipo de dados que deverão ser preservados verificou-se que mais de metade, 69,0%, referiu os dados brutos obtidos através de instrumentos e os documentos de texto, 67,6% assinalou a folha de cálculo, 59,0% imagens, digitalizações ou raios X e 53,3% os dados brutos gerados por programas.

	N	%
Dados brutos obtidos através de instrumentos	145	69,0%
Documento de texto (ex. Word, PDF, ...)	145	69,0%
Folha de cálculo (ex. Excel, ...)	142	67,6%
Imagens, digitalizações ou raios x	124	59,0%
Dados brutos gerados por programas	112	53,3%
Notas de laboratório	97	46,2%
Ficheiros SPSS	81	38,6%
Bases de dados (ex. MySQL, Oracle, ...)	80	38,1%
Vídeo	61	29,0%
Software	55	26,2%
Dados clínicos	53	25,2%
Áudio	42	20,0%
Outro formato proprietário	25	11,9%
Outro	5	2,4%
	210	

Tabela 9 – Tipos de dados preservados

Tipo de serviços a fornecer pela UMinho para garantir o acesso aos dados

Quanto ao tipo de serviços que gostariam de ver fornecidos pela UMinho para acesso aos dados em termos futuros, verificou-se que o serviço mais apontado foi o sistema de *backup* de ficheiros (72,9%), logo seguido por serviços de armazenamento e preservação (64,8%). O apoio técnico na elaboração de planos de gestão de dados teve 40,0% das respostas e a formação e consultoria na gestão de dados científicos 32,9%.

	N	%
Sistema de backup de ficheiros	153	72,9%
Serviços de armazenamento e preservação	136	64,8%
Apoio técnico na elaboração de planos de gestão de dados	84	40,0%
Formação e consultoria na gestão de dados científicos	69	32,9%
Serviços de partilha e acesso a dados	42	20,0%
Informação sobre repositórios temáticos de dados	33	15,7%
Identificação de normas e esquemas disciplinares de descrição de dados	14	6,7%
Outro	6	2,9%
	210	

Tabela 10 – Tipos de serviços a fornecer pela UMinho

1.1.3. DIAGNÓSTICO A COMUNIDADES PILOTO: PRÁTICAS DE ARMAZENAMENTO E DEPÓSITO DE DADOS E CARACTERIZAÇÃO DOS PROCESSOS DE GESTÃO DOS DADOS

Adicionalmente ao inquérito realizado em 2014 os SDUM propuseram-se prosseguir este trabalho de diagnóstico com o enfoque em comunidades piloto. O programa de diagnóstico foi preparado no ano de 2016 e está em aplicação no presente ano. Esta iniciativa de diagnóstico e auditoria de dados de investigação dos Serviços de Documentação da UMinho, baseada no projeto “Data Audit (Data Asset Framework)” desenvolvido pelo Digital Curation Center para as instituições de ensino superior no Reino Unido, visa o desenvolvimento iterativo de um diagnóstico dos dados de investigação produzidos na unidade de investigação e identificação das necessidades de suporte à gestão de dados. A metodologia deste programa baseia-se na identificação das comunidades piloto e aplicação de entrevistas e questionários, tendo como público-alvos investigadores e curadores de dados.

Os SDUM realizam este programa com as seguintes etapas:

1. Caracterizar e analisar os processos de gestão associados aos dados de investigação produzidos e recolhidos pelas atividades da unidade de investigação;
2. Descrever de forma detalhada os dados de investigação produzidos e avaliar papéis e responsabilidade na gestão desses dados;
3. Estabelecer orientações para a gestão dos dados de investigação e recomendações de melhoria.

Por forma a recolher mais dados para a elaboração do presente relatório e tendo em conta que apenas tinham sido recolhidos dados no primeiro semestre de 2017 de uma comunidade piloto (CIEd), os SDUM optaram por adaptar o programa e os respetivos questionários que compõem o diagnóstico por forma a recolher informações junto dos centros de investigação representados na comissão nomeada para elaboração deste relatório.

Deste modo, foram recolhidas de 26 de maio a 13 de junho respostas aos questionários que tinham como objetivo obter um breve retrato das práticas de armazenamento e depósito de dados em projetos das unidades de investigação representadas no grupo (questionário 2: <https://goo.gl/forms/aQ4hjyvEmdHdn9gv1>) e para [diagnóstico e caracterização dos processos de gestão de dados \(questionário 1\)](#). Às respostas recolhidas no período acima referido, devem ser acrescentadas as recolhidas resultado da primeira comunidade piloto (CIEd) no primeiro trimestre do ano.

Participaram na recolha de informação de diagnóstico, respondendo aos dois questionários, as seguintes escolas/departamentos ou centros/grupos de investigação:

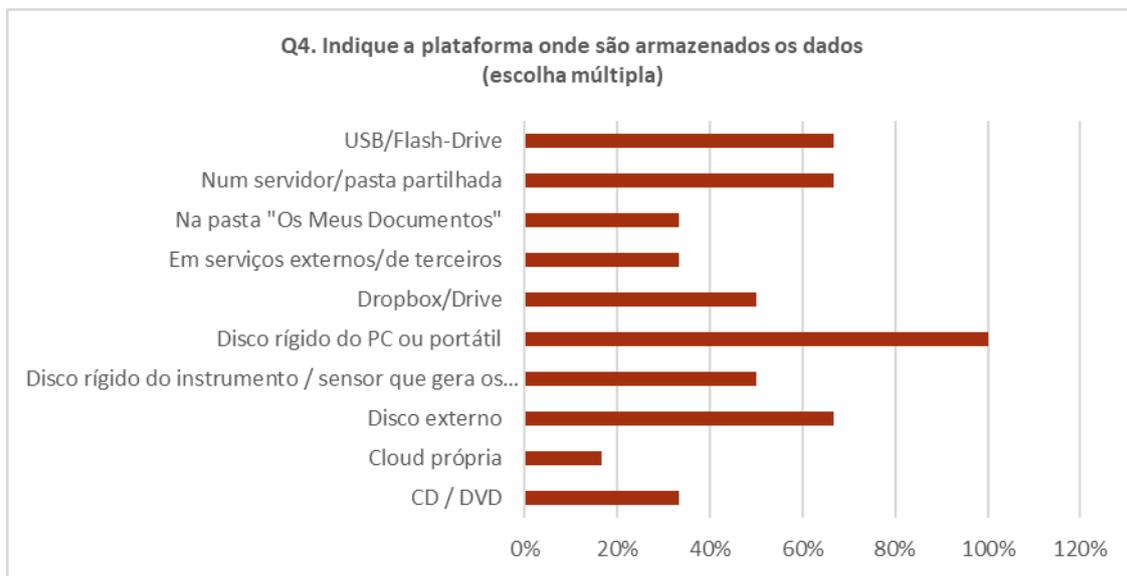
- Centro de Investigação em Educação (CIEd),
- Centro de Física e LIP Minho do Dep. de Física da Escola de Ciências,
- Escola de Economia e Gestão,
- Escola de Medicina e ICVS,

- Centro de Estudos de Comunicação e Sociedade (CECS),
- Centro de Engenharia Biológica da Escola de Engenharia.

As respostas recolhidas nos dois questionários revelam um alinhamento com os dados recolhidos no estudo de 2014 realizado pelos SDUM e descrito anteriormente, nomeadamente nas categorias ou de tipos de dados de investigação produzidos no centro/grupo de investigação e no volume aproximado dos dados usualmente detidos no contexto de um projeto da unidade de investigação.

Retrato das práticas de armazenamento e depósito de dados:

Plataformas de armazenamento dos dados:



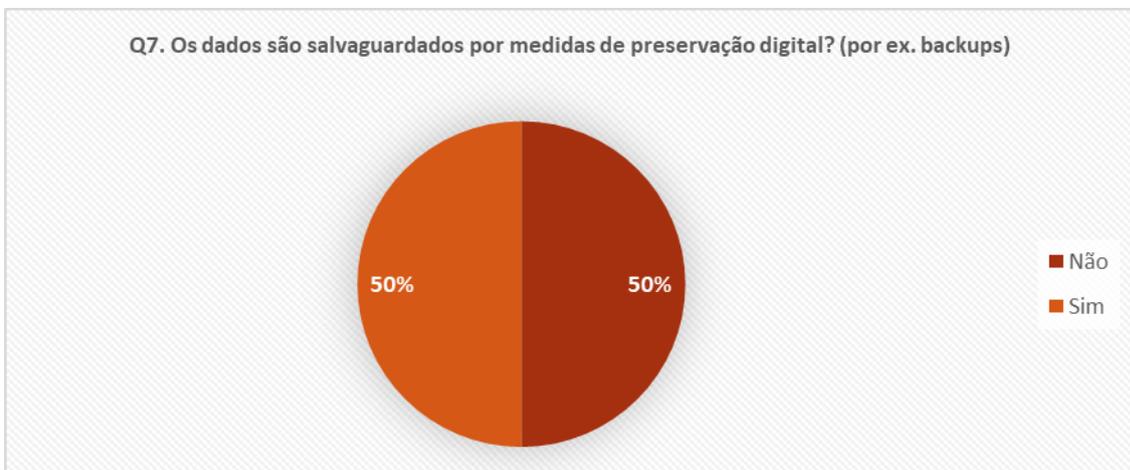
Capacidade disponível para armazenamento de dados:



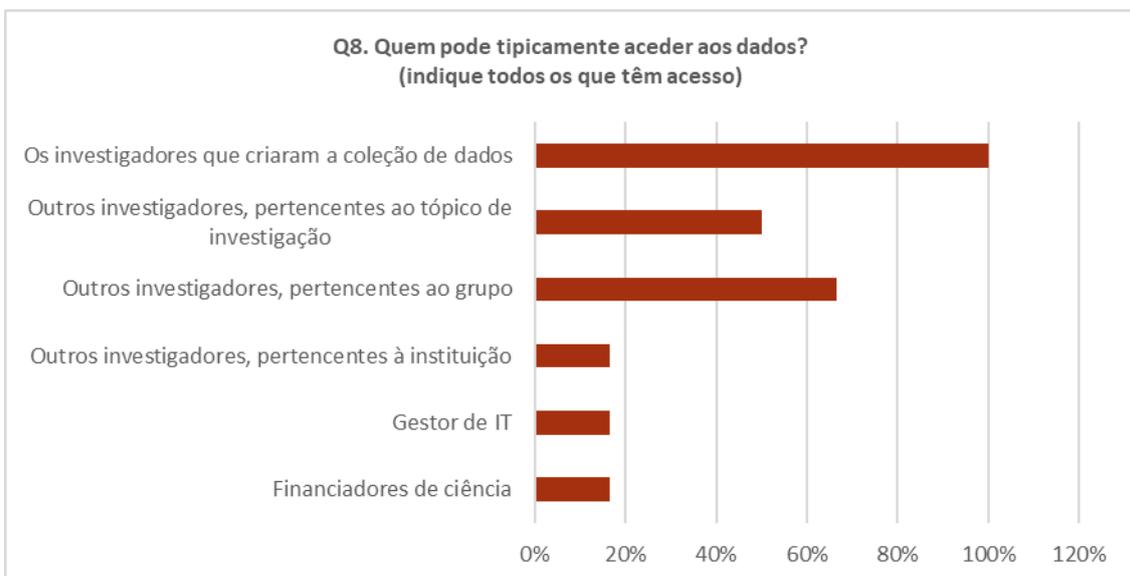
Aplicação de normas e diretrizes para garantir a qualidade da documentação dos dados:



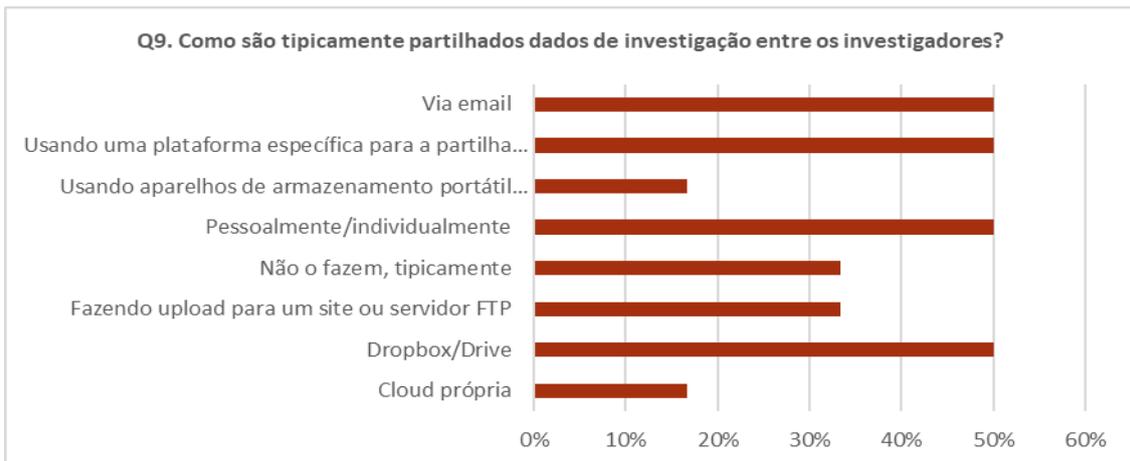
Salvaguardada dos dados por medidas de preservação digital:



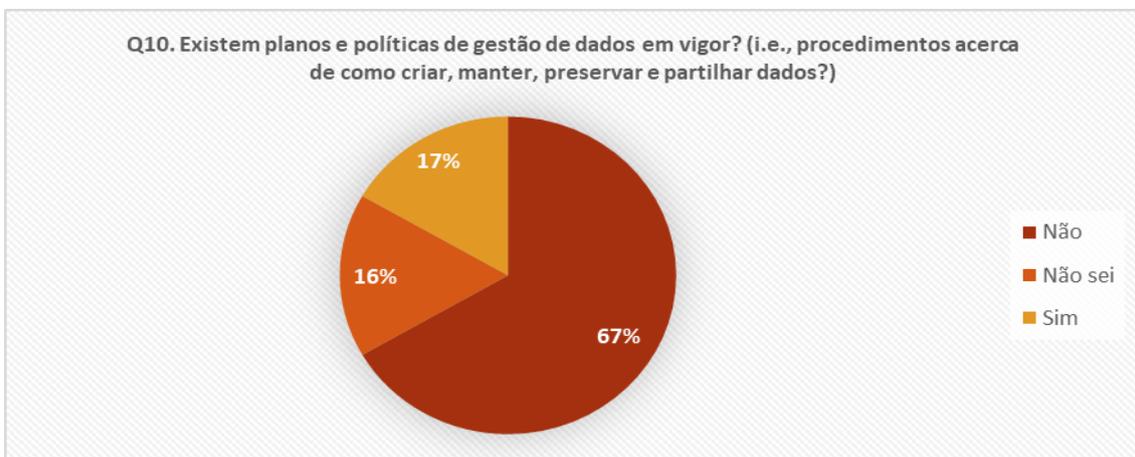
Políticas de acesso aos dados:



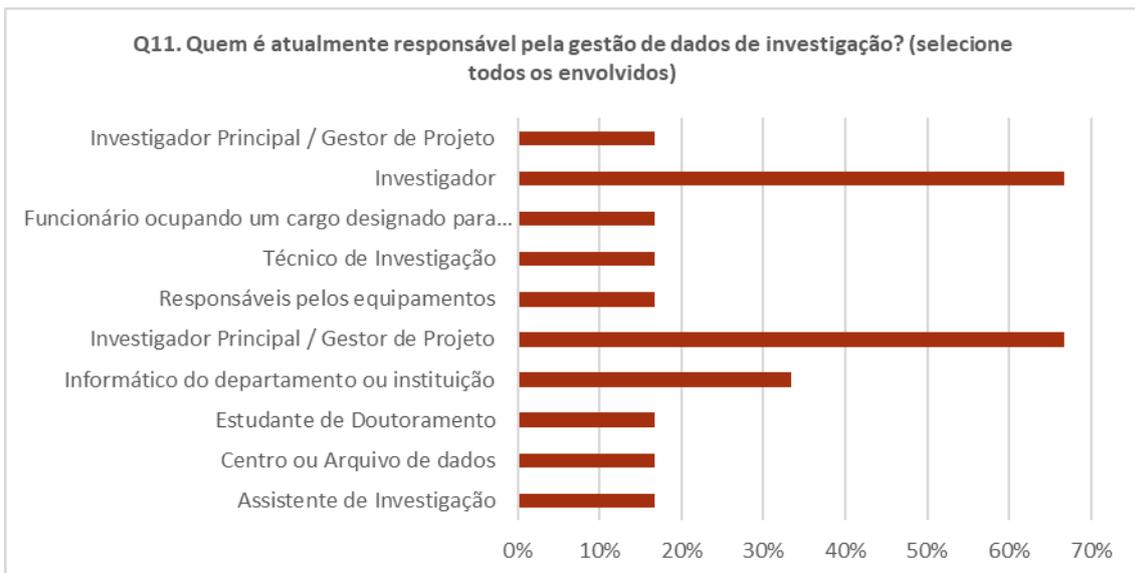
Partilha dos dados de investigação entre os investigadores:



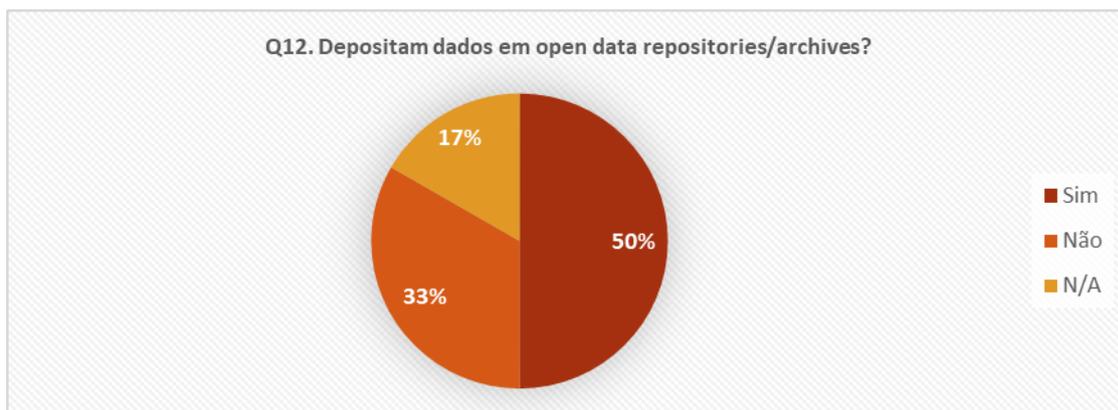
Planos e políticas de gestão de dados em vigor:



Responsabilidades na gestão de dados de investigação:

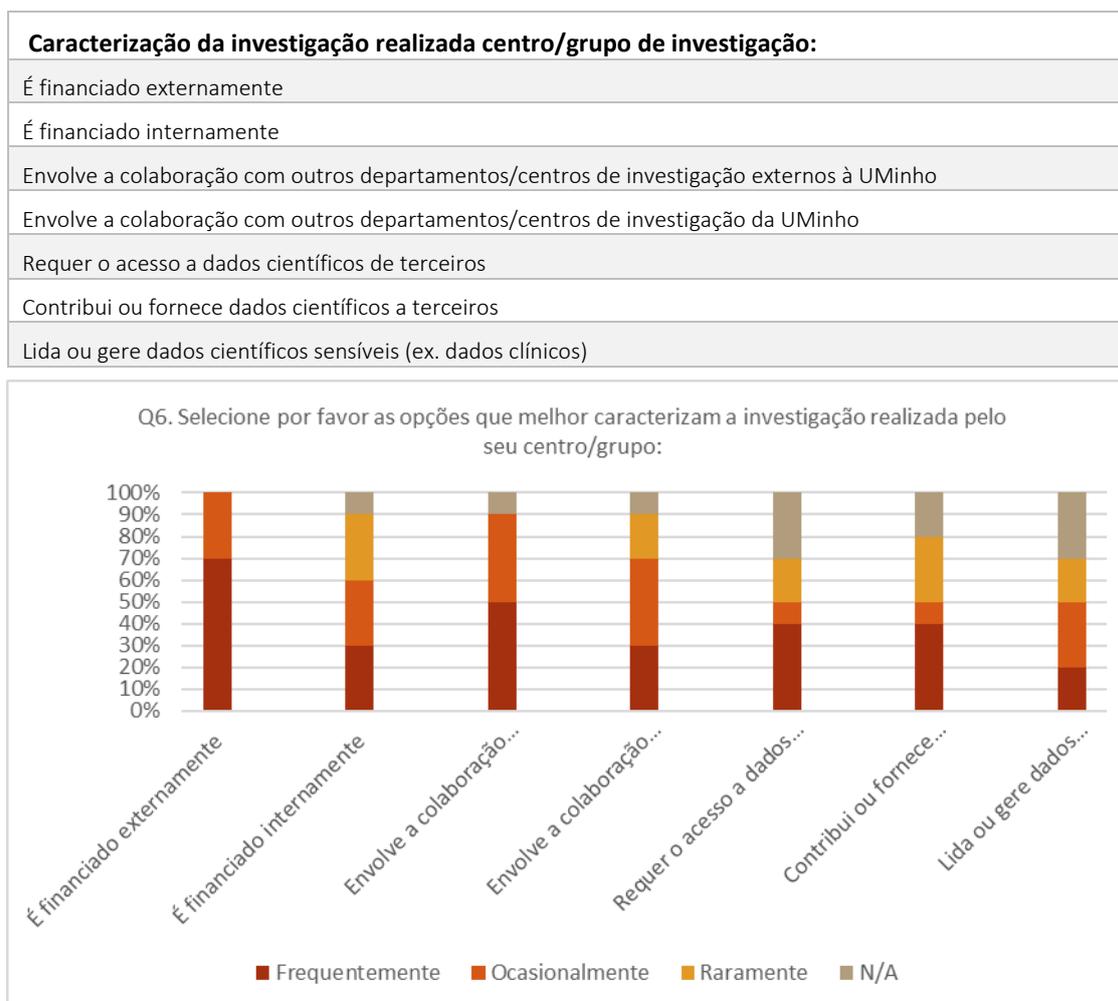


Depósito de dados em Repositórios ou Arquivos abertos:



Caracterização dos processos de gestão de dados:

Na informação recolhida no âmbito do diagnóstico e caracterização dos processos de gestão (questionário/entrevista adaptado do programa de diagnóstico de comunidades piloto a desenvolver pelos SDUM) destacam-se os seguintes resultados:



1.2. POLÍTICAS PARA DADOS DOS FINANCIADORES DE CIÊNCIA E INOVAÇÃO

A nível internacional, na última década, tem vindo a ser definidas e implementadas políticas relativas à gestão, acesso aberto e partilha dos dados de investigação por um crescente número de financiadores e instituições que realizam investigação.

Os financiadores têm sido os principais promotores de políticas de dados de investigação. Diversos financiadores nacionais na Europa (como os diversos *Research Councils* do Reino Unido, ou agências de financiamento da Alemanha e da Áustria), nos Estados Unidos (como a *National Science Foundation*) ou na Austrália (diversos *Research Council*), não apenas públicos, mas também privados (como o *Wellcome Trust* ou a *Bill e Melinda Gates Foundation*) definiram, na última década, políticas relativas aos dados de investigação que resultem do seu funcionamento. Apesar de várias especificidades e diferenças entre estas políticas, a generalidade requer a definição de planos de gestão de dados e estimula a partilha e o acesso aberto aos dados produzidos.

No contexto europeu, é particularmente relevante o estabelecimento de um projeto piloto de dados abertos pela Comissão Europeia no âmbito do programa Horizonte 2020. A política de dados abertos no H2020 aplicou-se apenas a cerca de 20% dos projetos iniciados entre 2014 e 2016, mas tornou-se universal (ainda que com opção de *opt-out*) a partir de 2017. A política do H2020, cuja orientação geral é “tão aberto quanto possível, tão fechado quanto necessário”, concretizou-se na cláusula 29.3 dos contratos de financiamento, que estabelece o requisito de depósito num repositório dos dados associados a publicações, ou de outros dados especificados no plano de gestão de dados, que passou a ser um entregável obrigatório dos projetos.

No que diz respeito às políticas relacionadas com a gestão e partilha dos dados científicos, a situação é ainda muito incipiente em Portugal. Não se conhecem quaisquer documentos de estratégia ou políticas relativas aos dados científicos de instituições que realizam investigação (quer de instituições individuais como as universidades, quer de órgãos coletivos como o Conselho de Reitores das Universidades Portuguesas).

A única política estabelecida em Portugal é a “Política sobre a disponibilização de dados e outros resultados de investigação científica financiada pela FCT”. A política da FCT recomenda aos investigadores que partilhem os “(...) dados primários e outros materiais produzidos em projetos financiados pela FCT com outros investigadores, a não mais do que custos incrementais e dentro de um prazo razoável, segundo as melhores práticas da disciplina científica em que se insere o projeto em questão”. A política sugere ainda que os candidatos a financiamento da FCT incluam “um plano de gestão de dados e outros resultados da investigação, contendo informação específica sobre o modo como promoverão a ampla disseminação e partilha dos resultados de I&D”.

1.2.1. POLÍTICA NACIONAL DE CIÊNCIA ABERTA

O Governo e o Ministério da Ciência, Tecnologia e Ensino Superior definiram como prioridade o compromisso da ciência com os princípios e práticas da Ciência Aberta, estando empenhados na elaboração e implementação de uma Política Nacional de Ciência Aberta assente no enunciado de que o Conhecimento é de Todos e para Todos.

O atual Governo perspetiva a Ciência Aberta não “como um fim em si mesmo, mas um meio que permite a translação do conhecimento científico para a comunidade científica, a sociedade e as empresas possibilitando desta forma ampliar o reconhecimento e o impacto social e económico da ciência. A Ciência Aberta é mais do que a disponibilização em acesso aberto de dados e publicações, representa a abertura do processo científico enquanto um todo, reforçando o conceito de responsabilidade social científica”³.

O MCTES publicou em fevereiro de 2016 um documento com os [Princípios Orientadores para a Ciência Aberta | Conhecimento para Todos](#) e em Março de 2016, o Conselho de Ministros aprovou uma [resolução](#) com as orientações para a Política Nacional de Ciência Aberta e mandata o MCTES para criar um [Grupo de Trabalho Interministerial \(GT-PNCA\)](#).

O GT-PNCA estruturou-se em quatro subgrupos (acesso aberto e dados abertos, infraestruturas e preservação, avaliação, responsabilidade social), uma comissão executiva (reunindo os coordenadores de cada subgrupo) e uma comissão consultiva. O grupo de trabalho produziu um primeiro documento de diagnóstico em outubro de 2016, e um segundo documento, contendo um significativo conjunto de recomendações em março de 2017. Este segundo documento encontra-se ainda a recolher o parecer e comentários da comissão consultiva.

1.2.2. DADOS DE INVESTIGAÇÃO ABERTOS NO HORIZONTE 2020

A Comissão Europeia (CE) criou em dezembro de 2013, no âmbito do programa quadro Horizonte 2020, um projeto-piloto de dados abertos que visa melhorar e maximizar o acesso e a reutilização dos dados de investigação gerados por projetos financiados. Decorrente da avaliação do piloto desenvolvido entre 2014 e 2016, a política de dados abertos da CE no H2020 passou, em janeiro de 2017, a aplicar-se a todas as áreas do programa. A CE requer aos beneficiários do financiamento H2020 o depósito dos dados de investigação, incluindo metadados associados, necessários para validar os resultados apresentados em publicações científicas, especificando ainda a definição de planos adequados para a gestão dos dados científicos produzidos.

A CE apresentou em abril de 2016 o seu projeto para disponibilizar serviços baseados na computação em nuvem e infraestruturas de dados de craveira mundial que permitam à comunidade científica, às empresas e às administrações públicas beneficiarem das vantagens oferecidas pela revolução do “Big Data”. A iniciativa europeia para a nuvem ([European Open](#)

³ <http://www.portugal.gov.pt/media/18506199/20160210-mctes-ciencia-aberta.pdf>

[Science Cloud](#)) facilitará o acesso e a reutilização de dados pelos investigadores e inovadores e reduzirá o custo do armazenamento de dados e das análises de elevado desempenho. O livre acesso aos dados da investigação pode contribuir para estimular a competitividade da Europa, beneficiando as empresas em fase de arranque, as PME e a inovação baseada nos dados, nomeadamente no campo da medicina e da saúde pública, podendo ainda incentivar o aparecimento de novas indústrias.

A iniciativa da União Europeia “European Open Science Cloud” (EOSC) tem como objetivo acelerar e apoiar a atual transição para uma Ciência Aberta e inovação aberta mais eficazes no mercado único digital. Deve permitir o acesso confiável a serviços, sistemas e a reutilização de dados científicos partilhados entre fronteiras disciplinares, sociais e geográficas. A EOSC é compreendida como um ambiente federado para abertura e reutilização de dados de investigação, baseado nos elementos existentes e emergentes nos Estados Membros da UE, com orientação e governação internacionais leves e um elevado grau de liberdade em relação à implementação prática.

A EOSC é, de facto, uma infraestrutura europeia, mas deve ser globalmente interoperável e acessível. Inclui os conhecimentos humanos necessários, os recursos, as normas, as boas práticas e as infraestruturas técnicas inerentes. Um aspeto importante da EOSC é a gestão sistemática de dados e a administração a longo prazo de dados de investigação e serviços na Europa e em todo o mundo. No entanto, a gestão de dados não é um fim em si mesmo, sendo o desígnio da EOSC a fronteira da ciência e da inovação na Europa⁴.

1.2.3. ESTRATÉGIAS INSTITUCIONAIS

Como anteriormente referido, um número crescente de instituições que realizam investigação, em particular universidades, têm vindo a estabelecer estratégias, políticas, infraestruturas e serviços no domínio dos dados de investigação. Para além das iniciativas individuais, têm sido definidas orientações coletivas para a promoção da gestão e partilha dos dados de investigação.

Por exemplo, em 2013 a LERU (League of European Research Universities) produziu o [LERU Roadmap for Research Data](#), oferecendo orientação às universidades europeias para enfrentarem os desafios e oportunidades relacionadas com os dados de investigação.

Outro exemplo relevante é o da Alemanha onde o Conselho de Reitores (HRK) tem vindo a discutir sistematicamente o tema da gestão dos dados de investigação e a produzir documentos orientadores para as universidades⁵.

⁴ [Realising the European Open Science Cloud: First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud](#)

⁵ Management of research data – a key strategic challenge for university management, recommendation by the 16th General Meeting of the German Rectors' Conference (HRK) - 13/05/2014
http://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_EN.pdf

Finalmente, a European University Association (EUA), estabeleceu em 2015 um [Expert Group on Science 2.0/Open Science](#), que produziu o [EUA Roadmap on Open Access to Research Publications](#) em 2016, prevendo-se que produza um documento semelhante sobre os dados de investigação proximamente.

1.3. PRINCÍPIOS FAIR PARA A GESTÃO DE DADOS DE INVESTIGAÇÃO

Recentemente⁶ foram formalmente divulgados os princípios FAIR (Findable, Accessible, Interoperable, and Re-usable) propostos para guiar a gestão dos dados de investigação e que resultam do contributo de um conjunto diversificado de parceiros no domínio dos dados científicos vindos das Universidades, da Indústria, das Agências de financiamento e dos Editores. O texto dos princípios⁷ preconiza quatro princípios basilares que devem guiar a gestão e administração dos dados - pesquisáveis, acessíveis, interoperáveis e reutilizáveis, e apresenta como premissa a necessidade urgente de melhorar a infraestrutura de apoio à reutilização de dados de investigação.

Numa abordagem prática os princípios FAIR preconizam as seguintes características:

- **Findable:** Registrar identificadores persistentes (PIDs), providenciar metadados, registar os dados num recurso pesquisável ou repositório.
- **Accessible:** Recuperável pelo PID usando o protocolo standard, os metadados devem permanecer acessíveis mesmo que os dados não.
- **Interoperable:** Usar linguagens formais e amplamente aplicáveis, usar vocabulários padrão, referências qualificadas.
- **Reusable:** Metadados com qualidade, licença e proveniência com clareza, utilização de padrões da comunidade disciplinar.

Adicionalmente, desde 2016, no contexto do lançamento pela Comissão Europeia da iniciativa “European Open Science Cloud” (EOSC), um grupo de especialistas de diferentes Estados-Membro (‘early mover’ EU member states) está a preparar a iniciativa GO FAIR que se apresenta como uma proposta para implementação prática da EOSC baseada nos princípios FAIR.

How university management can guide the development of research data management. , recommendation by the 19th General Meeting of the German Rectors' Conference (HRK) 10/11/2015
www.hrk.de/fileadmin/migrated/content/uploads/Empfehlung_Forschungsdatenmanagement_10112015_EN_02.pdf

⁶ <https://www.force11.org/group/fairgroup/fairprinciples>

⁷ <https://www.nature.com/articles/sdata201618>

2. RECOMENDAÇÕES

Em função do estudo realizado do contexto nacional e internacional e tendo por base a análise do diagnóstico desenvolvido sobre processos e práticas de gestão de dados na Universidade do Minho, apresentam-se de seguida as recomendações da estratégia institucional integrada da UMinho para a Gestão dos Dados de Investigação. As recomendações encontram-se estruturadas em cinco tópicos: política, infraestrutura e sistemas, serviços e ferramentas de apoio, capacitação e questões legais de proteção de dados e de propriedade intelectual.

2.1. POLÍTICA DE GESTÃO DE DADOS NA UMINHO

A definição de uma política institucional é a base para uma estratégia e a ação coordenada em matéria de gestão de dados na UMinho. São dois os principais objetivos que a estratégia e as políticas da UMinho devem prosseguir. Por um lado, beneficiar a ciência e a investigação, estimulando processos de gestão, curadoria, partilha e reutilização dos dados que garantam a integridade da investigação, promovam melhorias da sua eficiência e potenciem os seus impactos sociais e económicos.

Por outro lado, pretende-se garantir que a Universidade do Minho, e os que nela realizam investigação, dispõe de ferramentas, serviços e procedimentos que alinham com as melhores práticas internacionais neste domínio e facilitem o cumprimento das políticas relativas aos dados de investigação das agências de financiamento. Trata-se, portanto, de garantir a competitividade da UMinho num domínio que é cada vez mais importante para a avaliação e a obtenção de financiamentos.

Recomenda-se a definição de uma política institucional da Universidade do Minho relativa à gestão e partilha dos dados de investigação, com os seguintes elementos e características:

- a) Enquadra a sua existência na estratégia da UMinho para a investigação, nas tendências internacionais de política científica, na experiência da UMinho no domínio do acesso aberto e da ciência aberta;
- b) Inclui uma definição do que se entende por dados de investigação que são abrangidos pela política;
- c) Alinha com o princípio geral do H2020 - tão aberto quanto possível, tão fechado quanto necessário;
- d) Identifica os principais serviços, ferramentas e infraestruturas de suporte à gestão dos dados de investigação que a UMinho disponibiliza e os seus responsáveis;
- e) Reconhece e acolhe o Regulamento Geral sobre a Proteção de Dados da União Europeia;
- f) Enquadra e respeita o Código de Ética da Universidade do Minho.

2.2. INFRAESTRUTURA PARA OS DADOS DE INVESTIGAÇÃO

A implementação de uma infraestrutura institucional de suporte a políticas e práticas para a gestão de dados tem como premissa a sua integração e/ou interoperabilidade com o ecossistema mais amplo de serviços e plataformas disponíveis a nível nacional - existentes ou em desenvolvimento -, internacional e no contexto temático e disciplinar, tendo em conta os sistemas e recursos ao dispor nas comunidades científicas. Apesar da produção de dados de investigação se realizar naturalmente em contextos disciplinares, as atividades ocorrem em contexto institucional e beneficiam da existência de soluções institucionais, particularmente para a designada cauda longa dos dados científicos - os pouco volumosos conjuntos de dados produzidos e recolhidos pelos centros e grupos de investigação nos vários projetos científicos que desenvolvem.

Disponibilizar uma infraestrutura de dados que responda às diferentes necessidades de curadoria de dados ao longo do ciclo de vida da investigação é o desafio que se coloca à UMinho para conseguir, simultaneamente, dotar os seus investigadores de sistemas e tecnologias que os capacitem para competir com os seus pares de instituições congéneres mundiais e facultar as ferramentas para cumprirem de forma ágil e capaz os princípios FAIR da gestão de dados e os requisitos de financiadores de ciência e inovação.

No desenvolvimento dos diferentes componentes da infraestrutura para gestão de dados na UMinho, deve garantir-se a interligação com infraestruturas nacionais e europeias de investigação e outras de âmbito disciplinar e internacional que promovem a Ciência Aberta, assegurando os padrões de interoperabilidade de sistemas e guiar-se pelas melhores práticas ou diretrizes estabelecidas no âmbito de organismos e projetos como a *Research Data Alliance*, *OpenAIRE* ou *Datacite*. Neste desenvolvimento assume-se ainda como premissa a complexidade e diversidade dos tipos de dados, compreendendo que os resultados da investigação não se resumem a publicações e dados, existindo uma variedade de artefactos de investigação e produtos associados ao processo científicos (ex. software, modelos, métodos, testes) que carecem da necessária integração nos componentes da infraestrutura institucional.

O primeiro e mais significativo desafio em função do diagnóstico levado a cabo em 2014 e mais recentemente pelos SDUM, aponta para a necessidade de proporcionar a cada investigador um volume de armazenamento mínimo compatível com a sua atividade científica na instituição, que pode ser variável em função do tipo e do volume de dados produzidos em cada área disciplinar, proporcionando ainda garantias de segurança, estabilidade e redundância dos dados face a ameaças externas em constante expansão e mutação.

Recomenda-se o desenvolvimento a disponibilização de uma infraestrutura de sistemas e serviços integrados a disponibilizar à comunidade de investigadores da UMinho para a gestão de dados ao longo do ciclo de vida da investigação, incluindo os seguintes componentes:

- a. **Data box** - sistema de armazenamento ativo de dados a disponibilizar universalmente à comunidade UMinho (por indivíduo ou projeto) com funcionamento na nuvem e integrado com ferramentas de documentação dos dados.
- b. **Open Data repository** (Data RepositóriUM) - sistema de repositório de dados para registo e abertura dos dados, permitindo o depósito de dados abertos, embargados ou fechados, com atribuição de identificadores persistentes (DOI) e preservação dos conjuntos de dados.
- c. **Software repository** - sistema de registo, documentação e partilha de software aberto associado às atividades de investigação, assegurando o seu licenciamento, a identificação de autoria e a integração com plataformas de dados e publicações.
- d. **Catálogo de dados** - sistema para identificação e registo de todos os dados publicados como resultado da investigação de projetos ou investigadores da UMinho que agrega os dados depositados no Data Repository e permite o registo de outros dados depositados em repositórios disciplinares ou de editores e data centres temáticos.
- e. **Data vault** - sistema para preservação de dados integrado com o serviço de repositório e de armazenamento, permitindo a seleção e identificação dos conjuntos de dados a preservar e assegurar a integridade a longo prazo.

2.3. FERRAMENTAS DE APOIO À GESTÃO DOS DADOS

Os Serviços institucionais para apoio à gestão de dados visam sobretudo atribuir contexto aos sistemas disponíveis para o ciclo de vida de investigação, interligar os diferentes componentes da infraestrutura, disponibilizar ferramentas e informação de suporte que qualifique a comunidade científica na melhoria contínua das práticas de gestão e dados.

Com o desenvolvimento efetivo de serviços de apoio e consultoria na gestão de dados será possível cumprir de forma mais ágil aos requisitos de gestão e partilha de dados impostos pelos organismos financiadores de ciência e inovação e corresponder adequadamente aos estímulos que se preveem com a dinamização da iniciativa European Open Science Cloud.

Os serviços de apoio e consultoria visam igualmente melhorar a prática da gestão dos dados e a disponibilização de dados abertos, contribuindo para a o estabelecimento de metodologias de planeamento e para a disponibilização de dados em acesso aberto devidamente documentados.

O apoio de pessoal técnico habilitado para ocorrer às diferentes necessidades de gestão de dados ao longo do ciclo de vida de investigação, apoiando o processo de documentação dos dados antes mesmo do depósito no repositório é de grande relevância, pelo que a aposta na formação de curadores de dados será um eixo de intervenção essencial dos serviços de apoio.

Recomenda-se a criação de serviços e ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados, nomeadamente:

- a. **Planos de Gestão de Dados online** - disponibilização de ferramenta online⁸ para a criação de Planos de Gestão de Dados com inclusão de guias de apoio, permitindo o cumprimento dos requisitos dos financiadores e a melhoria gradual das práticas de armazenamento e documentação dos dados.
- b. **Guia de Metadados e documentação para dados** - informação sobre serviços de metadados e documentação para dados de investigação que promova o conhecimento e identificação de normas e esquemas de metadados disciplinares e a sua coerente e progressiva integração nos sistemas de armazenamento e repositório.
- c. **Seletor de licenças para dados** - disponibilização de ferramenta online⁹ para seleção das licenças adequadas à proteção do conjunto de dados e abertura dos dados à reutilização.
- d. **Registo DOI para conjuntos de dados** - Integração da funcionalidade de atribuição de *Digital Object Identifier* (Identificador de Objeto Digital) no Open Data RepositórioUM para exponenciar a visibilidade dos dados depositador e a sua adequada citação.
- e. **Serviço de apoio e consultoria na gestão de dados** – operacionalização de um serviço de coordenação dos sistemas e ferramentas a disponibilizar aos investigadores e articulação dos vários componentes, incluindo iniciativas de capacitação, que compõem a estratégia de gestão de dados em desenvolvimento na UMinho.

2.4. CAPACITAÇÃO

Promover e valorizar as competências dos investigadores em matéria de gestão e partilha de dados de investigação no contexto da Ciência Aberta é uma necessidade e um desafio prioritário para a comunidade de investigadores da UMinho.

Neste sentido, devem desenvolver-se um conjunto de ações e programas integrados de apoio e formação de investigadores que abarquem as necessidades de gestão de dados nas diferentes fases do ciclo de vida, desde o planeamento, com particular enfoque nos planos de gestão de dados, até ao licenciamento dos dados com o depósito e partilha em repositórios abertos, passando pelos processos de armazenamento e documentação dos dados, sem esquecer a promoção da citação e reutilização de dados.

As iniciativas de âmbito formativo devem promover a utilização das infraestruturas disponíveis e a criar e integrar os serviços de apoio e consultoria em desenvolvimento na UMinho pelos SDUM e outras unidades. Deve também, procurar-se integrar gradualmente este tipo de capacitação no âmbito das unidades curriculares de Projeto de Tese, Metodologias de Investigação e outras similares.

⁸ Versão institucional da DMPonline – ferramenta open source do Digital Curation Center do Reino Unido (<https://dmponline.dcc.ac.uk/>)

⁹ Versão institucional da License Selector – ferramenta open source desenvolvida no projeto europeu EUDAT (<https://www.eudat.eu/services/userdoc/license-selector>)

O investimento necessário para a dinamização de atividade de capacitação deve incluir igualmente a integração em iniciativas de âmbito nacional ou internacional com a inclusão ou a dinamização de programas na UMinho, nomeadamente de cariz internacional e disciplinar relacionados com a Ciência Aberta e a análise e processamento de dados.

Recomenda-se o desenvolvimento de programas integrados que promovam e valorizem as competências dos investigadores em matéria de gestão e partilha de dados de investigação, nomeadamente:

- a. Incluir **ações de formação** sobre gestão de dados no catálogo de cursos de literacia de informação oferecidos pelos **SDUM**;
- b. Criar materiais informativos e **guias formativos** dos SDUM sobre os requisitos dos financiadores, particularmente da Comissão Europeia e da FCT, e ainda sobre as políticas e práticas de editores relativamente à identificação dos dados.
- c. Oferecer materiais e cursos de **formação a distância** sobre Gestão de Dados e Ciência Aberta, designadamente com adaptação e tradução de recursos online disponibilizados no portal FOSTER Open Science e materiais do OpenAIRE.
- d. Promover sessões de esclarecimento e debate nos SDUM e nas Escolas que promovam boas práticas de “FAIR data”¹⁰ na gestão e partilha de dados, promovendo boas-práticas e iniciativas emergentes (**sessões *Data Champions* ou *FAIR data talks***).
- e. Realizar **curros práticos** de gestão de dados com enfoque na análise e processamento de dados, adaptando localmente os workshops das iniciativas “datacarpentry.org” e “softwarecarpentry” e promovendo escolas de verão de Data Science.
- f. Incluir conteúdos sobre boas práticas para a gestão de dados de investigação - nomeadamente planos de gestão de dados, referênciação e citação de dados, licenciamento e reutilização de dados e proteção de dados pessoais- na **formação pós-graduada da UMinho**, nomeadamente nas Unidades Curriculares de Projeto de Tese e Metodologia de Investigação.

2.5. QUESTÕES LEGAIS, DE PROTEÇÃO DE DADOS E DE PROPRIEDADE INTELECTUAL

A proteção de dados é um princípio hoje consagrado na legislação nacional e internacional. Há uma série de orientações da OCDE e das Nações Unidas, sendo que a Convenção Europeia para a Proteção dos Direitos Humanos e das Liberdades Fundamentais incorpora o direito humano à proteção de dados. Na União Europeia, a Carta dos Direitos Fundamentais garante a proteção dos dados pessoais. A legislação de proteção de dados é amplamente harmonizada pela Diretiva 95/46 /CE, com a indicação que as regras da diretiva serão substituídas e harmonizadas pelo Regulamento Geral de Proteção de Dados que entrará em vigor no próximo ano.

¹⁰ Aplicação dos Princípios FAIR data (Findable, Accessible, Interoperable, Reusable)

O novo Regulamento Geral de Proteção de Dados, a aplicar a partir de 25 de maio de 2018, em todos os Estados Membros da EU, abrangerá instituições públicas e privadas e terá implicações organizacionais e orçamentais significativas. Deve por isso acautelar-se o quanto antes as implicações da aplicação dos requisitos deste novo regulamento, especificamente no âmbito dos dados de investigação e dos procedimentos regulamentares a respeitar.

Se os dados pessoais estão em causa, as regras de proteção de dados devem ser consideradas. Muitas vezes, quando se abordam as questões legais relacionadas com gestão de dados discutem-se os temas da privacidade e dos dados pessoais. Estas questões são centrais quando são recolhidos dados de investigações de pessoas ou ensaios clínicos, pelo que nestes casos os investigadores devem assegurar a observância do Código de Conduta Ética da Universidade do Minho que se constitui como a linha de orientação ética que enquadra as atividades de investigação. Em muitos desses casos, os dados não podem ser partilhados e somente alguns dos dados podem, depois de procedimentos rigorosos de anonimização, ser partilhados em repositórios abertos.

Outra área relevante e complexa é a do licenciamento dos conjuntos de dados e bases de dados, considerando a grande diversidade, na natureza e formatos, do que se considera dados de investigação. Para garantir a utilização adequada do licenciamento em cada situação, é necessário compreender quais os direitos que podem estar envolvidos e conhecer as diferentes opções de licenças.

Os princípios da política de dados abertos da Comissão Europeia no âmbito do programa Horizonte 2020 definida como *“as open as possible, as closed as necessary”*, deve enquadrar a ação de promoção da gestão de dados de investigação na UMinho.

Recomenda-se a disponibilização e reforço de serviços de apoio e consultoria no domínio da proteção de dados, direitos de autor, licenciamento e reutilização de dados e propriedade intelectual, visando nomeadamente:

- a) Adaptação da informação relevante para a ferramenta online de seleção das licenças adequadas à abertura e depósito dos conjuntos de dados produzidos na UMinho.
- b) Criação de guias informativos sobre os princípios basilares da proteção de dados, direitos de autor no contexto das bases de dados, licenciamento e reutilização de dados e outras questões associadas à propriedade intelectual.
- c) Disponibilização de um programa de capacitação integrado na formação pós-graduada vocacionado a trabalhar as questões legais associadas aos dados, com particular enfoque na proteção de dados pessoais e licenciamento para abertura e reutilização de dados.
- d) Desenvolvimento das ações necessárias para acautelar as implicações institucionais no domínio da atividade de investigação do novo Regulamento Geral de Proteção de Dados a aplicar a partir de 25 de maio de 2018.

CONCLUSÃO

Tendo em conta a dimensão e a qualidade da sua comunidade de docentes e investigadores, a sua forte inserção nas redes e comunidades de investigação em múltiplas áreas científicas, a sua significativa experiência nos domínios do acesso aberto e da ciência aberta, bem como sua participação, ou mesmo liderança, em vários dos mais importantes projetos, iniciativas e infraestruturas nacionais e europeias nestes domínios, a Universidade do Minho está simultaneamente “obrigada” e apta a responder aos vários desafios relacionados com a gestão dos dados de investigação.

Tal como concluiu e propõe este Grupo de Trabalho, a resposta a esses desafios passa pela definição de uma estratégia que inclua três elementos ou dimensões fundamentais – política institucional, infraestruturas e serviços, capacitação e promoção das competências da comunidade UMinho – e que seja capaz de articular esses componentes com as iniciativas, organizações e infraestruturas nacionais e internacionais (especialmente europeias) já existentes ou em desenvolvimento.

QUADRO SÍNTESE DAS RECOMENDAÇÕES

Definição de uma política institucional da Universidade do Minho relativa à gestão e partilha dos dados de investigação.	<ul style="list-style-type: none">• Alinhamento com as tendências internacionais de política científica, o princípio geral do H2020 e com a experiência da UMinho no domínio do acesso aberto e da ciência aberta.• Identificação dos principais serviços, ferramentas e infraestruturas de suporte à gestão dos dados de investigação.• Aplica o Regulamento Geral sobre a Proteção de Dados da União Europeia e respeita o Código de Ética da Universidade do Minho.• Promove uma cultura de curadoria dos dados ao longo do ciclo de vida da investigação e promove a abertura e partilha.
Disponibilização de uma infraestrutura de sistemas e serviços integrados a disponibilizar à comunidade da UMinho para a gestão de dados ao longo do ciclo de vida da investigação.	<ul style="list-style-type: none">• Data box - sistema de armazenamento ativo de dados.• Open Data RepositóriUM - sistema de repositório de dados para depósito e abertura dos dados.• Software repository - sistema de registo, documentação e partilha de software aberto associado às atividades de investigação na UMinho.• Catálogo de dados - sistema para identificação, registo e agregação dos dados publicados como resultado da investigação de projetos ou investigadores da UMinho.• Data vault - sistema para preservação de dados integrado com o serviço de repositório e de armazenamento.
Criação de ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados.	<ul style="list-style-type: none">• Disponibilização de ferramenta online para a criação de Planos de Gestão de Dados.• Informação sobre serviços de metadados para dados de investigação.• Disponibilização de ferramenta para seleção das licenças para proteção e abertura dos dados.• Atribuição e registo de Digital Object Identifiers (DOI).• Serviço de apoio e consultoria na gestão de dados.

<p>Desenvolvimento de ações e programas integrados que promovam e valorizem as competências dos investigadores em matéria de gestão e partilha de dados de investigação.</p>	<ul style="list-style-type: none"> • Incluir ações de formação sobre gestão de dados no catálogo de cursos de literacia de informação dos SDUM. • Oferecer materiais e cursos de formação a distância em matéria de dados abertos e ciência aberta. • Promover sessões de esclarecimento e debate nos SDUM e nas Escolas que promovam boas práticas de “FAIR data” na gestão e partilha de dados (sessões <i>Data Champions</i> ou <i>FAIR data talks</i>). • Realizar cursos práticos de formação sobre gestão de dados com enfoque em <i>Data Science</i>. • Incluir conteúdos sobre boas práticas para a gestão de dados de investigação na formação pós-graduada da UMinho, nomeadamente nas Unidades Curriculares de Projeto de Tese e Metodologia de Investigação
<p>Disponibilização e reforço de serviços de apoio e consultoria no domínio da proteção de dados, licenciamento e reutilização de dados e propriedade intelectual.</p>	<ul style="list-style-type: none"> • Adaptação da informação relevante para a ferramenta online de seleção das licenças adequadas à abertura e depósito dos dados. • Criação de guias informativos sobre proteção de dados, direitos de autor, licenciamento de dados e propriedade intelectual. • Disponibilização de um programa de capacitação integrado na formação pós-graduada vocacionado a questões legais associadas aos dados. • Desenvolvimento da aplicação do novo Regulamento Geral de Proteção de Dados na UMinho.

DIAGRAMA DOS COMPONENTES DA ESTRATÉGIA DA UMINHO PARA A GESTÃO DE DADOS DE INVESTIGAÇÃO

