



**Universidade do Minho**  
Serviços de Documentação

# **Análise dos Resultados do Inquérito sobre Dados Científicos produzidos na Universidade do Minho**

*Relatório do Inquérito por Questionário*



# **ANÁLISE DOS RESULTADOS DO INQUÉRITO SOBRE DADOS CIENTÍFICOS PRODUZIDOS NA UNIVERSIDADE DO MINHO**

RELATÓRIO DESENVOLVIDO PELOS

**SERVIÇOS DE DOCUMENTAÇÃO DA UNIVERSIDADE DO MINHO**

**GABINETE DE PROJETOS OPEN ACCESS**

[HTTP://OPENACCESS.SDUM.UMINHO.PT](http://openaccess.sdum.uminho.pt)

JULHO 2014

LICENÇA CREATIVE COMMONS

*Este trabalho foi licenciado com uma Licença Creative Commons – Atribuição 4.0 Internacional.*

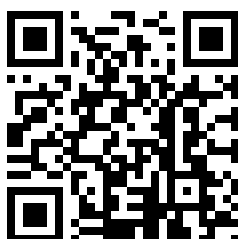


CITAÇÃO

Rodrigues, E., Boavida, C., Truta, R., Príncipe, P., Carvalho J. & Saraiva, R. (2014). Análise dos resultados do inquérito sobre dados científicos produzidos na Universidade do Minho. Universidade do Minho, Braga. Disponível no RepositóriUM em: <http://hdl.handle.net/1822/29948>

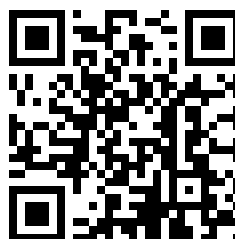
RELATÓRIO DO INQUÉRITO

<http://hdl.handle.net/1822/29948>



DADOS DO INQUÉRITO

<http://hdl.handle.net/1822/29949>



## Índice

<b>SUMÁRIO EXECUTIVO .....</b>	<b>3</b>
<b>1 INTRODUÇÃO .....</b>	<b>6</b>
1.1 Enquadramento e objetivos .....	6
1.2 Metodologia .....	8
1.3 Siglas .....	9
<b>2 CARACTERIZAÇÃO DOS PARTICIPANTES .....</b>	<b>10</b>
2.1 Unidade orgânica de ensino e de investigação (UOEI) .....	10
2.2 Categoria profissional .....	10
2.3 Perfil de atividade / Responsabilidade .....	11
2.4 Domínio científico .....	11
2.4.1 Área científica .....	12
2.5 Sexo .....	14
<b>3 CURADORIA E GESTÃO DE DADOS CIENTÍFICOS .....</b>	<b>15</b>
3.1 No âmbito do seu trabalho como investigador ou dos projetos em que participa produz ou possui dados científicos? .....	15
3.2 Desenvolve a sua investigação como parte de um grupo ou individualmente? .....	15
3.3 Que tipos de dados científicos são produzidos na investigação que realiza/em que participa? .....	16
3.4 Em valores aproximados qual é o volume global dos datasets acumulados por si ou pelo seu grupo de investigação? .....	16
3.5 São realizadas com regularidade cópias de segurança dos dados científicos produzidos por si ou pelo seu grupo de investigação? .....	17
3.6 Quem é o responsável pela realização das cópias de segurança dos dados científicos produzidos por si ou pelo seu grupo de investigação? .....	17
3.7 Com que frequência são feitos os backups? .....	18
3.8 Que tipo de dados científicos preserva usualmente? .....	18
3.9 Onde são guardados? .....	19
3.10 Atualmente, possui algum 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)? .....	19
3.11 Qual foi o principal motivo para o desenvolvimento da sua estratégia? .....	20
3.12 Por favor, indique o motivo pelo qual não possui um plano para a gestão de dados resultantes da sua investigação .....	20
3.13 Após a 'conclusão' de uma investigação, com que frequência são reutilizados os dados produzidos por si ou pelo seu grupo de investigação? .....	21
3.14 No que concerne ao tipo de acesso, qual é o estatuto usual dos seus dados? .....	21
3.15 Atualmente, permite que outros possam aceder aos dados produzidos por si ou pelo seu grupo de investigação? .....	22
3.16 Quem pode aceder aos dados produzidos por si ou pelo seu grupo de investigação? .....	22
3.17 Que tipo de questões o preocupa relativamente ao seu acesso aberto? .....	23
3.18 Já alguma vez foi solicitado para colocar os seus dados científicos disponíveis em acesso aberto, fora do âmbito de uma publicação (por exemplo, exigido pelo financiador do projeto)? ...	23
3.18.1 Comentários à pergunta anterior: .....	24
3.19 Tem conhecimento do projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa quadro Horizonte 2020? .....	24
<b>4 ACERVO DE DADOS CIENTÍFICOS EXISTENTE .....</b>	<b>25</b>

4.1	Que tipos de dados deverão ser preservados? .....	25
4.2	Esforço associado à criação de dados? .....	26
4.3	Período de preservação estimado .....	27
4.4	Frequência de uso .....	28
4.5	Dimensão 'final' estimada dos dados (tendo por exemplo um projeto tipo).....	29
4.6	Algum dos seus 'datasets' está armazenado num formato proprietário? .....	30
4.6.1	Por favor, identifique o(s) formato(s) .....	31
4.7	Por favor, indique que tipo de serviços gostaria que a Universidade do Minho fornecesse para garantir o acesso a esses dados em termos futuros?.....	31
4.8	Por favor, indique se estaria disposto a participar num projeto no âmbito da curadoria de dados científicos promovido pela Universidade do Minho e/ou numa reunião de 'follow-up' após este inquérito?.....	33
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>34</b>
	<b>ÍNDICE DE TABELAS.....</b>	<b>37</b>
	<b>ÍNDICE DE GRÁFICOS.....</b>	<b>39</b>
	<b>ANEXO 1: QUESTIONÁRIO .....</b>	<b>40</b>

## Sumário executivo

O presente estudo foi desenvolvido pelos Serviços de Documentação da Universidade do Minho, no âmbito do plano de intervenção 2014-2015 para os serviços de suporte à gestão dos dados científicos na UMinho que estabelece como primeira ação a análise da realidade da gestão dos dados científicos produzidos na UMinho.

Este relatório apresenta os resultados do inquérito por questionário realizado junto dos investigadores e centros de investigação da Universidade do Minho, que teve como objetivo proporcionar um melhor conhecimento sobre os dados científicos que são gerados no âmbito da investigação produzida na Universidade do Minho, procurando identificar o tipo de produção de dados e as práticas de gestão associadas e averiguar a pertinência de serviços de suporte neste domínio.

O questionário que consubstancia o presente estudo foi composto por 53 perguntas e estruturado em três grupos de questões. O primeiro grupo caracterizou o perfil do investigador participante no estudo. O segundo grupo caracterizou o nível do conhecimento, opinião e atitude face à curadoria e gestão de dados científicos. Por fim, o terceiro grupo caracterizou o acervo de dados científicos que cada investigador ou centro de investigação detém.

O questionário foi respondido exclusivamente via web com recurso à plataforma *LimeSurvey*, entre 14 de março e 17 de abril de 2014, tendo sido recolhidas 226 respostas completas.

A caracterização dos participantes do estudo revelou que mais de metade dos participantes no inquérito pertence à Escola de Engenharia (EE) (55,8%) seguindo-se a Escola de Ciências (EC) com 13,7% de respostas. A maioria dos participantes no inquérito é docente universitário (54,9%), representando os bolseiros de investigação 27,0% das respostas. A maioria dos participantes teve participação em projetos (52,2%) e 35% indicou assumir a coordenação de projetos. As respostas distribuíram-se por todos os domínios científicos sendo o mais representativo o da engenharia e tecnologia com 50,4% das respostas. A distribuição por género foi de 51,3% participantes do sexo masculino e 48,7% do sexo feminino.

A esmagadora maioria (92,9%) dos participantes afirmou possuir dados científicos resultantes das investigações que realizam na Universidade do Minho. Aproximadamente metade dos

participantes que possuem dados científicos (41,9%) realizou a sua investigação integrado numa equipa de trabalho e com os dados geridos em grupo.

O tipo de dados mais produzidos foram os dados experimentais (74,8%), seguido de dados estatísticos (47,6%), de imagens (41,4%) e de dados numéricos (36,2%). Cerca de um terço dos participantes que respondeu possuir dados científicos não tem conhecimento exato do volume aproximado de dados acumulados por si ou pelo seu grupo de investigação (24,8%) e cerca de um terço afirmou que o volume dos seus dados científicos não ultrapassa os 10 GB (8,6% menos de 1GB e 24,3% entre 1 e 10 GB).

A grande maioria dos que responderam ao inquérito realizou com regularidade cópias de segurança dos dados científicos produzidos (81,0%), sendo o responsável pela execução dessas cópias de segurança o participante de projeto (49,4%). Para 29,4% dos participantes essas cópias foram feitas *ad hoc*, ou seja sem periodicidade definida, ou mensalmente (22,4%) e preservam todo e qualquer tipo de dados científicos (86,5%), na sua maioria, usando discos externos (80,0%) ou discos de PC (52,4%).

Quando questionados sobre a existência de um plano para a gestão de dados resultantes da sua investigação, 80,5% dos participantes do inquérito indicou não possuir um plano, sendo o motivo mais referido para explicar esse facto a inexistência de políticas de gestão de dados no seio do seu grupo de investigação (31,4%). Aqueles que afirmaram possuir um plano para a gestão de dados (19,5%) apontaram como principal razão para o seu desenvolvimento o facto de ser um requisito da investigação para aceder, analisar e consultar dados de outros (43,9%).

Após a conclusão de uma investigação, 60,9% dos participantes referiu reutilizar com muita frequência em outros projetos de investigação os dados produzidos por si ou pelo seu grupo de investigação, nomeadamente em dissertações de mestrado (59,1%) e em teses de doutoramento (56,2%). As respostas indiciam que a maioria dos dados científicos obtidos no âmbito da investigação científica realizada permanece em acesso restrito ao grupo ou projeto (57,5%). A maioria dos participantes referiu não permitir que outros pudessem aceder aos dados produzidos por si ou pelo seu grupo de investigação (56,2%). Os restantes 43,8% afirmaram-se disponíveis para partilhar os dados entre o grupo de investigação (59,8%), entre alunos/colegas do departamento (39,1%), entre alunos/colegas da instituição (37,0%) e como suporte às publicações (31,5%).

As questões que mais preocupam os inquiridos que possuem dados científicos foram a confidencialidade e direitos de propriedade intelectual (79,0%), sendo que 92,4% nunca foi instado a colocar, fora do âmbito de uma publicação, os seus dados científicos em acesso aberto. Os que afirmaram ter colocado, fora do âmbito de uma publicação, os seus dados científicos disponíveis em acesso aberto referiram que disponibilizaram os dados científicos em redes sociais, como por exemplo *ResearchGate* e *LinkedIn*, especificando que tal foi realizado por solicitação de outros investigadores, por recomendação da Comissão Europeia ou da própria universidade e para publicação em revistas científicas de acesso aberto. Relativamente ao projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa quadro Horizonte 2020, 83,8% afirmou não ter conhecimento deste projeto.

Sobre o acervo de dados existente na Universidade do Minho, 69,0% respondeu que devem ser preservados os dados brutos obtidos através de instrumentos e documentos de texto (ex. Word, PDF), 67,6% a folha de cálculo e 59,0% apontou as imagens, digitalizações ou raios X. Sobre a previsão do esforço associado à criação de dados científicos, 33,8% indicou ser de meses, 28,1% de anos e 20,0% de semanas. Relativamente ao período de preservação de dados científicos, 26,7% dos participantes responderam que este período deve corresponder entre 2 a 5 anos ou a 5 a 10 anos. Já em relação à frequência do uso dos dados científicos, 30,0% responderam que fazem um uso semanal, 17,6% usam os dados diariamente e 17,1% mensalmente. Tendo por exemplo um projeto tipo, a dimensão final estimada dos dados produzidos foi de 1 a 10 GB para 34,3% dos participantes do estudo e de 10 a 100 GB para 30,0%.

Quanto ao tipo de serviços que os respondentes gostariam de ver facultados pela Universidade do Minho relativamente ao acesso dos dados em termos futuros, 72,9% apontaram um sistema de *backup* de ficheiros como serviço preferencial, 64,8% preferiram serviços de armazenamento e preservação. A destacar ainda que 40,0% apontou o apoio técnico na elaboração de planos de gestão de dados como um serviço necessário e nesta linha 32,9% indicaram a formação e consultoria na gestão de dados científicos. Por último, 19,0% dos participantes manifestaram o seu interesse em participar num projeto promovido pela Universidade do Minho no âmbito da curadoria de dados científicos.

## 1 Introdução

### 1.1 Enquadramento e objetivos

Este relatório apresenta os resultados do estudo efetuado pelos Serviços de Documentação da Universidade do Minho (SDUM) junto dos investigadores e centros de investigação, no sentido de se obter um melhor conhecimento sobre os dados científicos gerados no âmbito da investigação realizada na Universidade do Minho e as práticas associadas para gestão dos dados. O estudo resulta do plano de intervenção que os SDUM estabeleceram para os serviços de suporte à Gestão dos Dados Científicos na UMinho, enquadrando-se no eixo um deste plano, que inclui diversas ações visando a análise da realidade da gestão dos dados científicos produzidos na UMinho.

O plano de intervenção estratégica para suporte à gestão dos dados científicos na UMinho estabelece dez eixos de intervenção que incluem um total de 33 ações. A atuação dos SDUM neste domínio está planeada a dois anos e passará por: diagnóstico e observação prática junto de comunidades piloto, criação de serviços de apoio e consultoria, desenvolvimento de competências dos colaboradores SDUM, formulação e implementação da política institucional, disseminação de boas práticas e incentivo à adoção de políticas de dados abertos, promoção da interoperabilidade de infraestruturas de dados científicos, apoio à conceção de serviços de curadoria para o ciclo de vida dos dados científicos que inclui o repositório de dados, promoção da referenciação e citação dos dados científicos e melhorar a prática da gestão dos dados abertos nos Serviços de Documentação.

Entende-se por dados científicos (*research data/scientific data*) todos os dados que foram recolhidos e/ou utilizados pelos investigadores no decorrer das suas investigações na Universidade do Minho. As questões relacionadas com os dados científicos, nomeadamente as da preservação, acessibilidade, integridade e utilização, associadas à curadoria e gestão de dados científicos assumem atualmente particular relevância no contexto das instituições de investigação científica e dos organismos financiadores da ciência. Muitas instituições de investigação reconhecem hoje a importância de desenvolver serviços de gestão de dados científicos, alicerçados numa política institucional e delineados numa estratégia integrada suportada em sistemas de armazenamento, preservação e acesso e serviços de apoio e planeamento. Paralelamente, um número significativo de agentes financiadores de ciência tem



manifestado redobrada atenção à necessidade de assegurar meios e mecanismos para preservação e reutilização dos dados científicos, requerendo aos investigadores beneficiários o planeamento adequado da gestão dos dados científicos e às instituições de investigação a definição de políticas e serviços que suportem a curadoria dos dados científicos no decorrer do ciclo de vida da investigação.

A Comissão Europeia (CE) criou recentemente, no âmbito do programa quadro Horizonte 2020<sup>1</sup>, um projeto-piloto de dados abertos que visa melhorar e maximizar o acesso e a reutilização dos dados de investigação gerados por projetos financiados. A CE requer aos participantes deste piloto 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.

Na mesma linha, a Fundação para a Ciência e a Tecnologia<sup>2</sup> adotou, a 5 de maio de 2014, no quadro mais amplo da política de Acesso Aberto, orientações sobre a disponibilização de dados e outros resultados de projetos financiados pela FCT, 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.

Neste contexto é indispensável conceber estratégias de atuação institucional que respondam às exigências dos organismos financiadores da ciência e que maximizem o valor dos dados gerados nas atividades de investigação. 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<sup>3</sup>.

---

<sup>1</sup> *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020* disponível em: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

<sup>2</sup> Política sobre a disponibilização de dados e outros resultados de projetos de I&D financiados pela FCT disponível em: [https://www.fct.pt/documentos/PoliticaAcessoAberto\\_Dados.pdf](https://www.fct.pt/documentos/PoliticaAcessoAberto_Dados.pdf)

<sup>3</sup> Jones, S., Pryor, G. & Whyte, A. (2013). *How to Develop Research Data Management Services - a guide for HEIs. DCC How-to Guides*. Edinburgh: Digital Curation Centre <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services>

O relatório está estruturado em cinco partes, sendo a primeira constituída por esta breve introdução, na qual se descreve o estudo realizado e o contextualiza face à crescente visibilidade dos temas referentes à curadoria, gestão de dados científicos e à importância do estabelecimento de estratégias de ação institucionais. Na segunda parte, dedicada à “Caraterização dos participantes”, faz-se um enquadramento por unidades orgânicas de ensino, categoria profissional, perfil de atividade, domínio científico, área científica e sexo dos participantes. Na terceira parte, designada de “Curadoria e gestão de dados científicos”, apresentam-se os resultados obtidos relativamente aos dados decorrentes da atividade científica dos participantes, e como são produzidos, quanto aos diferentes tipos e volumes, às práticas de segurança, preservação, gestão e reutilização dos dados, e ainda o tipo de acesso permitido aos dados gerados. Por fim, na quarta parte designada por “Acervo de dados científicos existentes”, reúne-se a informação recolhida sobre os dados que os participantes consideram essenciais para o seu trabalho individual ou do grupo de investigação, o esforço associado à criação de dados, período de preservação estimado, frequência de uso, dimensão e formato dos conjuntos de dados, e ainda o tipo de serviços que gostariam que a Universidade do Minho fornecesse para garantir o acesso a esses dados em termos futuros. Finalmente, na quinta e última parte apresentam-se algumas conclusões do estudo e identificam-se os eixos de atuação futura para os serviços de suporte à gestão de dados científicos na UMinho.

## 1.2 Metodologia

O inquérito por questionário foi 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, composto por quarenta e uma perguntas, foi respondido via web com recurso à plataforma *LimeSurvey*.

O período de resposta decorreu entre os dias 14 de março e 17 de abril de 2014 e resultou na obtenção de 226 respostas válidas de um universo de 353 respostas, sendo que 127 foram consideradas incompletas, não fazendo parte do tratamento final dos resultados.

Os dados recolhidos com recurso ao sistema *LimeSurvey* foram armazenados numa tabela única e examinados com recurso ao *software* de tratamento estatístico SPSS (*Statistical Package for the Social Sciences*) versão 22. Nos testes de hipótese efetuados, nomeadamente o teste de independência do Qui-quadrado, utilizaram-se níveis de significância de 5%. As relações

estatisticamente significativas estão devidamente assinaladas. Procedeu-se à exclusão de categorias pouco representativas sempre que os testes de hipótese efetuados assim o exigiram.

### 1.3 Siglas

CE – Comissão Europeia

EA – Escola de Arquitetura

EC – Escola de Ciências

ECS – Escola de Ciências da Saúde

ED – Escola de Direito

EEG – Escola de Economia e Gestão

EE – Escola de Engenharia

EPsi – Escola de Psicologia

ESE – Escola Superior de Enfermagem

ETI – Equivalente a Tempo Integral

FCT – Fundação para a Ciência e a Tecnologia

ICS – Instituto de Ciências Sociais

I&D – Investigação & Desenvolvimento

IE – Instituto de Educação

ILCH – Instituto de Letras e Ciências Humanas

SDUM – Serviços de Documentação da Universidade do Minho

UOEI - Unidade Orgânica de Ensino e Investigação

UMinho – Universidade do Minho

## 2 Caracterização dos participantes

Foram definidas cinco variáveis para obter a caracterização dos participantes neste inquérito por questionário: unidade orgânica de ensino e de investigação, categoria profissional, perfil de atividade/responsabilidade, domínio científico (subdividido em área científica) e sexo. No sentido de aferir a representatividade da amostra constituída pelos participantes no inquérito foi feito um paralelo entre os dados deste estudo e os dados disponíveis no Relatório de Atividades 2012 da Universidade do Minho<sup>4</sup>, representando esta amostra 22% do total de efetivos de pessoal docente (ETIs).

### 2.1 Unidade orgânica de ensino e de investigação (UOEI)

Mais de metade dos participantes do estudo indicaram pertencer à EE (55,8%). Os restantes 44,2% distribuíram-se pela EC (13,7%), ECS (7,5%), sendo a restante percentagem distribuída de forma quase residual pelas restantes UOEI.

	<i>N</i>	%
Escola de Arquitectura (EA)	0	0,0%
Escola de Ciências (EC)	31	13,7%
Escola de Ciências da Saúde (ECS)	17	7,5%
Escola de Direito (ED)	2	0,9%
Escola de Economia e Gestão (EEG)	15	6,6%
Escola de Engenharia (EE)	126	55,8%
Escola de Psicologia (EPsi)	7	3,1%
Escola Superior de Enfermagem (ESE)	2	0,9%
Instituto de Ciências Sociais (ICS)	16	7,1%
Instituto de Educação (IE)	13	5,8%
Instituto de Letras e Ciências Humanas (ILCH)	4	1,8%
	226	

Tabela 1 – Número de participantes por UOEI

### 2.2 Categoria profissional

Em relação à categoria profissional (ver Tabela 2), verificou-se que as categorias mais representativas foram os docentes universitários (54,9%), bolsiros de investigação (27,0%) e investigadores (20,4%).

<sup>4</sup> Relatório de Atividades 2012 disponível em: <http://www.uminho.pt/uminho/informacao-institucional/relatorios>, consultado a 21 de maio de 2014.

	<i>N</i>	%
Docente Universitário	124	54,9%
Investigador	46	20,4%
Técnico Superior	2	0,9%
Bolseiro(a) de Investigação	61	27,0%
Outra	7	3,1%
	226	

Tabela 2 – Número de participantes por categoria profissional

A categoria ‘Outra’ obteve 7 respostas, sendo a maioria das respostas referentes a diferentes tipos de bolseiros de investigação.

### 2.3 Perfil de atividade / Responsabilidade

A maioria dos participantes do estudo teve como atividade/responsabilidade a participação em projetos (52,2%), 35,0% a coordenação de projetos e 15,5% indicaram ser alunos de doutoramento.

	<i>N</i>	%
Responsável de centro de investigação	11	4,9%
Coordenador de projeto	79	35,0%
Participante de projeto	118	52,2%
Aluno de doutoramento	35	15,5%
Outro	28	12,4%
	226	

Tabela 3 - Número de participantes por perfil de atividade / responsabilidade

De realçar que a opção Outro perfil de atividade/responsabilidade abrangeu 12,4% das respostas. Da análise das respostas abertas resultou que a maioria ( $N=17$ ) frequenta pós-doutoramento.

### 2.4 Domínio científico

Já em relação ao domínio científico dos participantes deste estudo (ver Tabela 4), a área da engenharia e tecnologia é maioritária (50,4%), seguindo-se a área das ciências sociais (19,9%) e a das ciências médicas (12,8%).

	<i>N</i>	%
Ciências Naturais	27	11,9%
Engenharia e Tecnologia	114	50,4%
Ciências Médicas	29	12,8%
Ciências Agrárias	1	0,4%
Ciências Sociais	45	19,9%
Humanidades	10	4,4%
	226	

Tabela 4 - Número de participantes por domínio científico

### 2.4.1 Área científica

Cada um dos domínios científicos incluiu diferentes áreas científicas. Nas ciências naturais obteve-se um maior número de respostas nas áreas das ciências biológicas (44,4%), seguida das ciências físicas (22,2%).

	<i>N</i>	%
Matemáticas	5	18,5%
Ciências da Computação e da Informação	0	0,0%
Ciências Físicas	6	22,2%
Ciências Químicas	3	11,1%
Ciências da Terra e do Ambiente	0	0,0%
Ciências Biológicas	12	44,4%
Outras Ciências Naturais	1	3,7%
	27	

Tabela 5 - Número de participantes por área científica: ciências naturais

Analisando as respostas da engenharia e tecnologia, 17,5% são da engenharia eletrotécnica, eletrónica e informática, 16,7% da engenharia dos materiais e 11,4% da biotecnologia industrial.

	<i>N</i>	%
Engenharia Civil	11	9,6%
Engenharia Eletrotécnica, Eletrónica e Informática	20	17,5%
Engenharia Mecânica	2	1,8%
Engenharia Química	7	6,1%
Engenharia dos Materiais	19	16,7%
Engenharia Médica	8	7,0%
Engenharia do Ambiente	2	1,8%
Biotecnologia Industrial	13	11,4%
Biotecnologia Ambiental	6	5,3%
Nanotecnologia	4	3,5%
Outras Engenharias e Tecnologias	22	19,3%
	114	

Tabela 6 - Número de participantes por área científica: engenharia e tecnologia

Das ciências médicas, a área científica das ciências da saúde obteve a maior percentagem de respostas (51,7%).

	<i>N</i>	%
Medicina Básica	0	0,0%
Medicina Clínica	4	13,8%
Ciências da Saúde	15	51,7%
Biotecnologia Médica	6	20,7%
Outras Ciências Médicas	4	13,8%
	29	

Tabela 7 - Número de participantes por área científica: ciências médicas

Obteve-se apenas 1 resposta nas ciências agrárias.

	<i>N</i>	%
Agricultura, Sicultura e Pescas	0	0,0%
Ciência Animal e dos Laticínios	0	0,0%
Ciências Veterinária	0	0,0%
Biotecnologia Agrária	0	0,0%
Outras Ciências Agrárias	1	100,0%
	1	

Tabela 8 - Número de participantes por área científica: ciências agrárias

Nas ciências sociais, as respostas distribuíram-se pela economia e gestão (24,4%), psicologia (22,2%) e ciências da educação (15,6%).

	<i>N</i>	%
Psicologia	10	22,2%
Economia e Gestão	11	24,4%
Ciências da Educação	7	15,6%
Sociologia	2	4,4%
Direito	2	4,4%
Ciências Políticas	4	8,9%
Geografia Económica e Social	0	0,0%
Ciências da Comunicação	4	8,9%
Outras Ciências Sociais	5	11,1%
	45	

Tabela 9 - Número de participantes por área científica: ciências sociais

Nas humanidades (ver Tabela 10) verificou-se uma distribuição equitativa na história e arqueologia e línguas e literaturas, ambas com 40,0% de respostas.

	<i>N</i>	%
História e Arqueologia	4	40,0%
Línguas e Literaturas	4	40,0%
Filosofia, Ética e Religião	0	0,0%
Artes	0	0,0%
Outras Humanidades	2	20,0%
	10	

Tabela 10 - Número de participantes por área científica: humanidades

## 2.5 Sexo

Por fim, em relação à caracterização dos participantes, esta última questão mostrou-nos que houve uma certa equidade nas respostas, apesar de a maioria ter sido dada pelo sexo masculino com 51,3%, em detrimento das respostas femininas com 48,7%.

	<i>N</i>	%
Masculino	116	51,3%
Feminino	110	48,7%
	226	

Tabela 11 – Número de participantes por sexo



### 3 Curadoria e gestão de dados científicos

Por curadoria e/ou gestão de dados científicos entende-se o conjunto de ações que garantem que um conjunto de dados é genuíno, permitindo o seu uso por outros que não os seus produtores. As ações de curadoria e de gestão de dados têm alguma interseção, sendo as de gestão mais independentes do conteúdo e do uso.

Neste terceiro grupo, composto por vinte questões, pretendeu-se compreender se os participantes deste estudo possuem dados científicos decorrentes das suas produções científicas; os diferentes tipos e volumes; se são feitas cópias de segurança aos dados, por quem e com que frequência; o tipo de dados científicos que habitualmente são preservados; se os participantes têm algum plano para a gestão dos dados que resultam da investigação; se esses mesmos dados são reutilizados em novas investigações e qual o tipo de acesso desses mesmos dados.

#### 3.1 No âmbito do seu trabalho como investigador ou dos projetos em que participa produz ou possui dados científicos?

Em larga escala, 92,9%, dos participantes deste estudo indicaram produzir ou possuir dados científicos, resultantes das investigações que realizam.

	<i>N</i>	%
Sim	210	92,9%
Não	16	7,1%
	226	

Tabela 12 – Dados científicos produzidos no âmbito do trabalho como investigador

#### 3.2 Desenvolve a sua investigação como parte de um grupo ou individualmente?

Esta questão teve como objetivo aferir o tipo de envolvimento e método de trabalho dos investigadores participantes no estudo, tendo-se apurado que 41,9% está integrado numa equipa e com os dados geridos pelo grupo, 26,7% realizaram parte da investigação de forma individual e outra parte em equipa. Como parte de uma equipa mas sendo cada membro responsável pela curadoria dos seus próprios dados responderam 25,7%. A indicação de investigação exclusivamente individual foi residual entre os participantes (ver Tabela 13).

	<i>N</i>	%
Como parte de uma equipa e com os dados geridos pela mesma	88	41,9%
Como parte de uma equipa, mas cada elemento da equipa cuida dos seus próprios dados	54	25,7%
Parte da investigação realizada de forma independente e outra parte em equipa	56	26,7%
Apenas individualmente	11	5,2%
Outro	1	0,5%
	210	

Tabela 13 – Investigação desenvolvida individualmente ou em grupo

### 3.3 Que tipos de dados científicos são produzidos na investigação que realiza/em que participa?

Das respostas obtidas, é possível verificar que a maioria dos dados que foram produzidos no âmbito da investigação foram dados experimentais (74,8%), seguidos de dados estatísticos (47,6%), de imagens (41,4%) e dados numéricos (36,2%).

	<i>N</i>	%
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 14 – Dados científicos produzidos no contexto da investigação

Os dados multimédia e áudio apresentaram percentagens residuais e nos 2,4% que indicam Outro tipo de dados científicos as respostas mais relevantes foram os dados cartográficos e os desenhos enquanto registo de escavações.

### 3.4 Em valores aproximados qual é o volume global dos *datasets* acumulados por si ou pelo seu grupo de investigação?

A esta questão do estudo 24,8% dos participantes responderam que não sabiam qual o volume global dos *datasets*. Já 24,3% responderam que o volume dos seus dados científicos 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 (ver Tabela 15).

	<i>N</i>	%
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 15 – Volume global dos *datasets* acumulados

### 3.5 São realizadas com regularidade cópias de segurança dos dados científicos produzidos por si ou pelo seu grupo de investigação?

Os resultados desta questão indicaram que 81,0% dos participantes realizam 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.

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

Tabela 16 – Realização das cópias de segurança dos dados científicos

### 3.6 Quem é o responsável pela realização das cópias de segurança dos dados científicos produzidos por si ou pelo seu grupo de investigação?

Quase metade dos participantes (49,4%) respondeu que as cópias de segurança são executadas por participantes de projetos enquanto 35,3% indicou que a responsabilidade é dos coordenadores de projeto. Os alunos de doutoramento foram apontados por 22,9% dos participantes como responsáveis pelas cópias de segurança.

	<i>N</i>	%
Participante de projeto	84	49,4%
Coordenador de projeto	60	35,3%
Outro	43	25,3%
Aluno de doutoramento	39	22,9%
Não sei	24	14,1%
Responsável de centro de investigação	5	2,9%
	170	

Tabela 17 – Responsável pela realização das cópias de segurança dos dados científicos

Importa realçar que a opção Outro reuniu 25,3% das respostas. Da análise das respostas abertas desse grupo apurou-se que 62,8% (15,8% do total de participantes do questionário) indicou que as cópias de segurança dos dados científicos são realizadas pelo próprio.

	<i>N</i>	%
Próprio	27	62,8%
Bolseiro de Investigação	2	4,7%
Cada investigador do grupo	6	14,0%
Responsável TI	6	14,0%
Ninguém	1	2,3%
Secretariado	1	2,3%
	43	

Tabela 18 – Responsável pela realização das cópias de segurança dos dados científicos, análise da opção Outro

### 3.7 Com que frequência são feitos os *backups*?

Quanto à frequência da execução de cópias de segurança 29,4% dos inquiridos respondeu que é uma realização *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*.

	<i>N</i>	%
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 19 – Frequência dos *backups*

Já na opção de resposta Outro (5,3%), a análise das respostas abertas indicou que a maioria (*N*=4) realiza *backups* semestralmente.

### 3.8 Que tipo de dados científicos preserva usualmente?

A esmagadora maioria dos participantes (86,5%) indicou que todos os dados científicos são preservados, enquanto 6,5% respondeu que apenas são preservados os dados críticos para o projeto ou os dados críticos para a publicação (ver Tabela 20).

	<i>N</i>	%
Tudo	147	86,5%
Dados críticos para o projeto	11	6,5%
Dados críticos para publicação	11	6,5%
Não sei	1	0,6%
	170	

Tabela 20 – Tipos de dados científicos preservados

### 3.9 Onde são guardados?

Decorrente da questão anterior, 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.

	<i>N</i>	%
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 21 – Local de armazenamento dos dados científicos

Já a opção *Outro*, referente ao local onde são guardados os dados científicos, representou 4,7% das respostas. Foram apontados alguns serviços de armazenamento e sincronização de arquivos como por exemplo o *SurDoc* e o *Google Drive*, o correio eletrónico e o suporte papel.

### 3.10 Atualmente, possui algum 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 (ver Tabela 22).

	<i>N</i>	%
Não	169	80,5%
Sim	41	19,5%
	210	

Tabela 22 – Plano de gestão de dados científicos

### 3.11 Qual foi o principal motivo para o desenvolvimento da sua estratégia?

Dos que afirmaram possuir um plano de gestão de dados, 43,9% responderam que é um requisito da investigação para aceder, analisar e consultar dados de outros. Nesta aferição dos motivos da criação do plano, o tamanho da equipa, assim como o volume de dados associados ao projeto foram apontados por 14,6% dos respondentes, sendo os restantes motivos, como a exigência do financiador, complexidade dos dados e ausência de política de dados, referidos por uma percentagem mais reduzida de participantes.

	<i>N</i>	%
Requisito da investigação para aceder, analisar e consultar dados de outros	18	43,9%
Tamanho da equipa do projeto (ou seja, vários criadores e utilizadores de dados)	6	14,6%
Volume de dados associados ao projeto	6	14,6%
Exigência do financiador do projeto	3	7,3%
Complexidade de dados associados com o projeto (por exemplo, vários formatos)	3	7,3%
Outro	3	7,3%
Ausência de política de gestão de dados no grupo de investigação	2	4,9%
	41	

Tabela 23 – Motivos para o desenvolvimento do plano para a gestão de dados

Na resposta Outro motivo para o desenvolvimento da estratégia, foram apontados o sistema de qualidade e o rigor científico.

### 3.12 Por favor, indique o motivo pelo qual não possui um plano para a gestão de dados resultantes da sua investigação

Nesta questão, pretendeu-se saber os motivos que levam os participantes a não terem um plano de gestão de dados científicos aquando da realização da investigação. Quase um terço (31,4%) respondeu que isso se deve à inexistência de políticas de gestão de dados no seio do grupo de investigação, 17,8% afirmou que isso não é necessário/adequado, enquanto igual percentagem (17,8%) referiu não saber por que não existe um plano de gestão de dados. O tempo e o esforço necessários (13,6%), o facto de não ser uma exigência do financiador do projeto (6,5%), a falta

de apoio e/ou orientação (6,5%) e a falta de formação/conhecimentos (5,3%), foram os outros motivos assinalados para a não existência de planos de gestão de dados.

	<i>N</i>	%
Ausência de política de gestão de dados no grupo de investigação	53	31,4%
Não é necessária / adequada à área ou grupo de investigação	30	17,8%
Não sei	30	17,8%
Tempo e esforço necessários	23	13,6%
Não é exigida pelo financiador do projeto	11	6,5%
A falta de apoio local / orientação	11	6,5%
A falta de formação / conhecimentos no grupo de investigação	9	5,3%
Outro	2	1,2%
	169	

Tabela 24 – Motivos para não possuir um plano para a gestão de dados

### 3.13 Após a 'conclusão' de uma investigação, com que frequência são reutilizados os dados produzidos por si ou pelo seu grupo de investigação?

Os inquiridos afirmaram reutilizar com frequência/muita frequência os dados produzidos por si ou pelo seu grupo de investigação em outros projetos de investigação (60,9%), em dissertações de mestrado (59,1%) e em teses de doutoramento (56,2%).

	Muito frequente		Com frequência		Não sei		Pouco frequente		Nunca	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Outros projetos de investigação	33	15,7%	95	45,2%	27	12,9%	49	23,3%	6	2,9%
Atividades de ensino	19	9,0%	82	39,0%	37	17,6%	57	27,1%	15	7,1%
Dissertações de mestrado	26	12,4%	98	46,7%	35	16,7%	32	15,2%	19	9,0%
Teses de doutoramento	34	16,2%	84	40,0%	39	18,6%	31	14,8%	22	10,5%
Não são reutilizados	15	7,1%	15	7,1%	45	21,4%	76	36,2%	59	28,1%
	210		210		210		210		210	

Tabela 25 – Frequência de reutilização dos dados produzidos

### 3.14 No que concerne ao tipo de acesso, qual é o estatuto usual dos seus dados?

Como é possível verificar na Tabela 26, a maioria dos participantes indicou que os dados científicos obtidos no âmbito da investigação científica têm estatuto de acesso restrito ao grupo ou projeto (57,5%) ou mesmo apenas ao investigador (15,5%). Apenas 13,3% dos participantes responderam que os seus dados ficam imediatamente disponíveis em acesso aberto imediato e 10,3% referiram que os seus dados são de acesso aberto após um período de embargo, que varia entre 1 e 5 anos (mas com maior frequência de 1 a 2 anos) dependendo do tipo de contrato de financiamento e das políticas editoriais das revistas científicas.

	<i>N</i>	%
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 26 – Tipo de acesso dos dados científicos

Em relação à opção Outro tipo de acesso, a maioria das respostas foram inconclusivas. No entanto, as que foram consideradas relacionaram-se com as políticas editoriais das revistas científicas.

### 3.15 Atualmente, permite que outros possam aceder aos dados produzidos por si ou pelo seu grupo de investigação?

Avaliando a questão se outras pessoas podem ou não aceder aos dados gerados no âmbito da investigação, a maioria dos participantes afirmou que não (56,2%), contra 43,8% que responderam que permitem o acesso a terceiros.

	<i>N</i>	%
Não	118	56,2%
Sim	92	43,8%
	210	

Tabela 27 – Permissão de acesso aos dados científicos

### 3.16 Quem pode aceder aos dados produzidos por si ou pelo seu grupo de investigação?

De entre os que responderam afirmativamente à questão anterior (3.15) a maioria respondeu que o acesso é permitido entre o grupo de investigação (59,8%), 39,1% permite o acesso a alunos/colegas do departamento, 37,0% a alunos/colegas da instituição e 31,5% permite que sejam um suporte às publicações (ver Tabela 28).



	<i>N</i>	%
Grupo de investigação	55	59,8%
Alunos / Colegas do departamento	36	39,1%
Alunos / Colegas da instituição	34	37,0%
Como suporte às publicações	29	31,5%
Público em geral	19	20,7%
Outras instituições	17	18,5%
Outro	5	5,4%
	92	

Tabela 28 – Quem pode aceder aos dados científicos produzidos

Apenas 20,7% respondeu dar acesso aos dados pelo público em geral (ou seja, em bom rigor, acesso aberto).

### 3.17 Que tipo de questões o preocupa relativamente ao seu acesso aberto?

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

	<i>N</i>	%
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 29 – Preocupações relativamente ao acesso aberto

Dos 4,8% dos participantes que responderam Outro, a maioria ( $N=8$ ), afirmou não ter qualquer tipo de preocupação relativamente ao acesso aberto dos dados científicos.

### 3.18 Já alguma vez foi solicitado para colocar os seus dados científicos disponíveis em acesso aberto, fora do âmbito de uma publicação (por exemplo, exigido pelo financiador do projeto)?

Questionados sobre se alguma vez lhes foi solicitado a disponibilização dos dados em acesso aberto fora do âmbito de uma publicação, a esmagadora maioria (92,4%) respondeu que não (ver Tabela 30).

	<i>N</i>	%
Não	194	92,4%
Sim	16	7,6%
	210	

Tabela 30 – Colocação dos dados científicos em acesso aberto

### 3.18.1 Comentários à pergunta anterior:

Dos 7,6% dos inquiridos que afirmaram já lhes ter sido solicitada a disponibilização dos seus dados científicos em acesso aberto, fora do âmbito de uma publicação, contabilizaram-se seis comentários, entre os quais, a disponibilização dos dados científicos em redes sociais, como por exemplo o *ResearchGate* e *LinkedIn* ( $N=2$ ), por solicitação de outros investigadores ( $N=2$ ), por recomendação da Comissão Europeia e da própria universidade ( $N=1$ ) e para publicação em revistas científicas de acesso aberto ( $N=1$ ).

### 3.19 Tem conhecimento do projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa quadro Horizonte 2020?

Questionados sobre o conhecimento do projeto-piloto de dados abertos da Comissão Europeia, a esmagadora maioria (83,8%) respondeu de forma negativa.

	<i>N</i>	%
Não	176	83,8%
Sim	34	16,2%
	210	

Tabela 31 – Conhecimento sobre o projeto-piloto de dados abertos da CE

## 4 Acervo de dados científicos existente

Nesta secção do questionário, composta por 10 perguntas, pretendeu-se recolher informações sobre os dados que os participantes consideram primordiais para o seu próprio trabalho ou grupo de investigação, o esforço associado à criação de dados, período de preservação estimado, frequência de uso, dimensão e formato dos conjuntos de dados, e ainda o tipo de serviços que gostariam que a Universidade do Minho facultasse para contribuir para uma melhor gestão dos dados científicos.

### 4.1 Que tipos de dados deverão ser preservados?

Em relação ao tipo de dados que deverão ser preservados verificou-se que mais de metade dos inquiridos, 69,0%, respondeu que os dados brutos obtidos através de instrumentos e os documentos de texto (ex. Word, PDF) deverão ser preservados, 67,6% referiu a folha de cálculo, 59,0% apontou as imagens, digitalizações ou raios X e 53,3% indicou os dados brutos gerados por programas.

	<i>N</i>	%
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 32 – Tipos de dados preservados

Observaram-se variações significativas entre os participantes por UOEL na resposta a esta questão. Enquanto entre os participantes da ECS os dados brutos obtidos através de instrumentos; imagens, digitalizações e raios x; notas de laboratório; ficheiros SPSS e dados clínicos foram os mais referidos, entre os respondentes da EE verificou-se que os dados brutos obtidos através de instrumentos; folha de cálculo (ex. Excel...); imagens, digitalizações e raios x;

dados brutos gerados por programas; notas de laboratório e *software* foram as opções preferenciais. As respostas dos participantes da EA e da ED foram excluídas da análise pelo seu número ser pouco representativo.

	EC	ECS	EEG	EE	Epsi	ESE	ICS	IE	ILCH
Dados brutos obtidos através de instrumentos	11,8%	12,7%	11,1%	13,0%	14,3%	11,1%	10,6%	11,3%	3,2%
Documento de texto (ex. Word, PDF, ...)	13,2%	9,5%	15,6%	12,1%	11,4%	0,0%	15,2%	15,1%	12,9%
Folha de cálculo (ex. Excel, ...)	11,1%	10,3%	8,9%	13,4%	8,6%	0,0%	10,6%	11,3%	9,7%
Imagens, digitalizações ou raios x	14,6%	12,7%	2,2%	11,0%	5,7%	11,1%	9,1%	3,8%	12,9%
Dados brutos gerados por programas	9,0%	7,9%	8,9%	11,0%	2,9%	0,0%	7,6%	5,7%	3,2%
Notas de laboratório	11,1%	10,3%	2,2%	9,5%	5,7%	11,1%	0,0%	5,7%	3,2%
Ficheiros SPSS	4,2%	9,5%	8,9%	5,6%	20,0%	22,2%	9,1%	13,2%	6,5%
Bases de dados (ex. MySQL, Oracle, ...)	5,6%	6,3%	11,1%	5,6%	2,9%	11,1%	16,7%	11,3%	12,9%
Vídeo	4,9%	6,3%	4,4%	4,0%	5,7%	11,1%	10,6%	5,7%	12,9%
Software	3,5%	0,8%	8,9%	5,9%	0,0%	11,1%	0,0%	3,8%	3,2%
Dados clínicos	4,9%	9,5%	0,0%	4,4%	8,6%	0,0%	0,0%	3,8%	3,2%
Áudio	2,8%	2,4%	8,9%	1,8%	14,3%	11,1%	9,1%	5,7%	12,9%
Outro formato proprietário	2,8%	1,6%	6,7%	2,4%	0,0%	0,0%	0,0%	1,9%	3,2%
Outro	0,7%	0,0%	2,2%	0,1%	0,0%	0,0%	1,5%	1,9%	0,0%

Tabela 33 – Tipos de dados preservados por UOEI

#### 4.2 Esforço associado à criação de dados?

O esforço associado à criação de dados científicos referido pelos participantes do estudo, medido a partir de períodos de tempo, foi de meses para 33,8%, de anos para 28,1% e de semanas para 20,0%.

	N	%
Meses	71	33,8%
Anos	59	28,1%
Semanas	42	20,0%
Dias	24	11,4%
Horas	10	4,8%
Outro	2	1,0%
Sem resposta	2	1,0%
	210	

Tabela 34 – Esforço associado à criação de dados

Também quanto ao esforço associado à criação de dados se verificaram variações estatisticamente significativas entre os participantes das diversas UOEI. Verificou-se que para a maioria dos participantes da EEG, EE, IE e do ILCH o esforço associado à criação de dados é de meses. Para a maioria dos participantes da ECS e do ICS o esforço associado à criação de dados é de anos. Nas restantes UOEI as respostas distribuem-se de forma mais dispersa pelos

diferentes períodos. Na análise do esforço associado à criação de dados por UOEI as respostas dos participantes da EA, ED e da ESE foram excluídas por estarem pouco representadas.

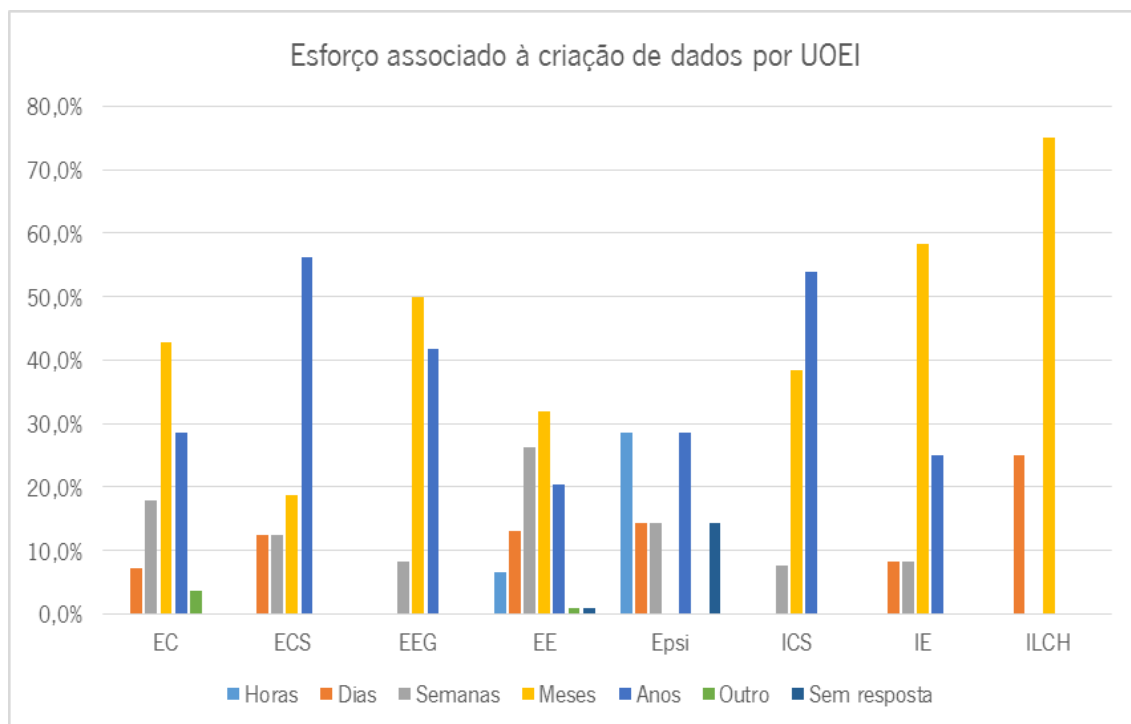


Gráfico 1 – Esforço associado à criação de dados por UOEI

### 4.3 Período de preservação estimado

Relativamente ao período de preservação de dados científicos, 26,7% responderam que este período deve ser de 2 a 5 anos ou de 5 a 10 anos.

	N	%
2 - 5 anos	56	26,7%
5 - 10 anos	56	26,7%
Indefinido	47	22,4%
10 - 20 anos	18	8,6%
Não sei	16	7,6%
1 - 2 anos	10	4,8%
20 - 100 anos	5	2,4%
Menos de 1 ano	1	0,5%
Sem resposta	1	0,5%
Outro	0	0,0%
	210	

Tabela 35 – Período de preservação estimado

Na análise do período de preservação estimado dos dados científicos por UOEI foram excluídas as respostas de participantes da EA, ED e da ESE por estarem pouco representadas. Um período

de preservação de 2 a 5 anos foi o mais indicado pelos participantes do IE e da EE, enquanto um período de 5 a 10 anos foi o mais indicado pelos participantes da ECS e o período de 10 a 20 anos o mais apontado entre os participantes da EPsi. No caso dos participantes da EC e do ILCH o maior número de respostas apontou para um período de preservação indefinido.

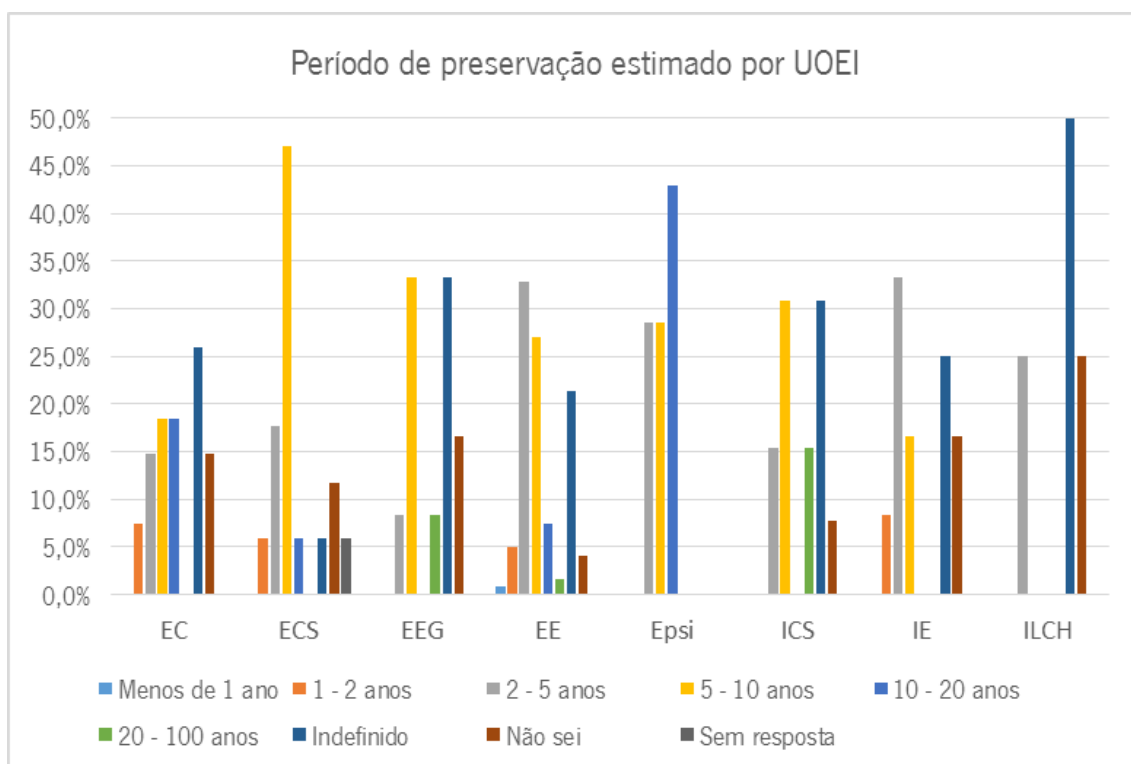


Gráfico 2 – Período de preservação estimado por UOEI

#### 4.4 Frequência de uso

Já em relação à frequência do uso dos dados científicos (ver Tabela 36), 30,0% dos participantes responderam que fazem um uso semanal, 17,6% usam os dados diariamente, 17,1% mensalmente, 11,4% trimestralmente e 11,0% fazem um uso pontual dos dados.

A opção Outro obteve 6,7% das respostas, metade das quais referindo ser muito variável a frequência de uso dos dados científicos.

	<i>N</i>	%
Semanal	63	30,0%
Diária	37	17,6%
Mensal	36	17,1%
Trimestral	24	11,4%
Pontual	23	11,0%
Outro	14	6,7%
Anual	10	4,8%
Sem resposta	3	1,4%
	210	

Tabela 36 – Frequência de uso

Na análise da frequência de uso por UOEI as respostas dos participantes da EA, ED e da ESE foram excluídas por estarem pouco representadas. O uso semanal do acervo de dados científicos foi o mais indicado pelos participantes da EC, ECS, EE, Epsi e do ICS. A frequência de uso trimestral foi a que recolheu maior número de respostas entre os participantes da EEG e do IE. As respostas dos participantes do ILCH dividem-se entre uma frequência de uso trimestral, diária e pontual.

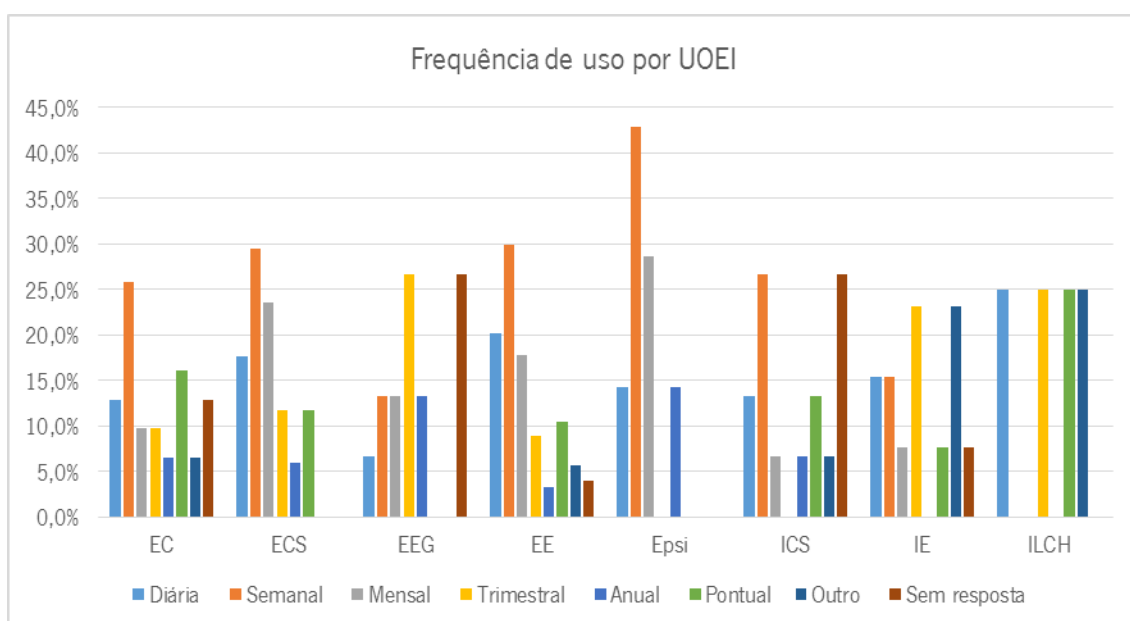


Gráfico 3 – Frequência de uso por UOEI

#### 4.5 Dimensão 'final' estimada dos dados (tendo por exemplo um projeto tipo)

A dimensão final estimada dos dados produzidos (ver Tabela 37) foi de 1 a 10 GB para 34,3% dos participantes e de 10 a 100 GB para 30,0%. Sublinhe-se ainda que o número de respostas

indicando uma dimensão inferior a 1GB (11,0%) foi maior do que as que indicaram mais de 1 TB (6,7%).

	N	%
1 - 10 GB	72	34,3%
10 - 100 GB	63	30,0%
100 GB - 1 TB	30	14,3%
Menos de 1 GB	23	11,0%
1 - 10 TB's	10	4,8%
Sem resposta	8	3,8%
Múltiplos TB's	4	1,9%
	210	

Tabela 37 – Dimensão estimada dos dados

Da análise seguinte, sobre a variação da dimensão estimada dos dados por UOEI, foram excluídas as respostas dos participantes da EA, ED e da ESE por estarem pouco representadas. A dimensão dos dados indicada pelo maior número de respondentes da maioria das UOEI (EC, EEG, EE, EPsi e IE) foi de 1 e 10 GB, enquanto os participantes da ECS e do ICS estimaram uma dimensão dos dados entre os 10 e os 100 GB, e dos do ILCH distribuíram-se por quatro dimensões diferentes.

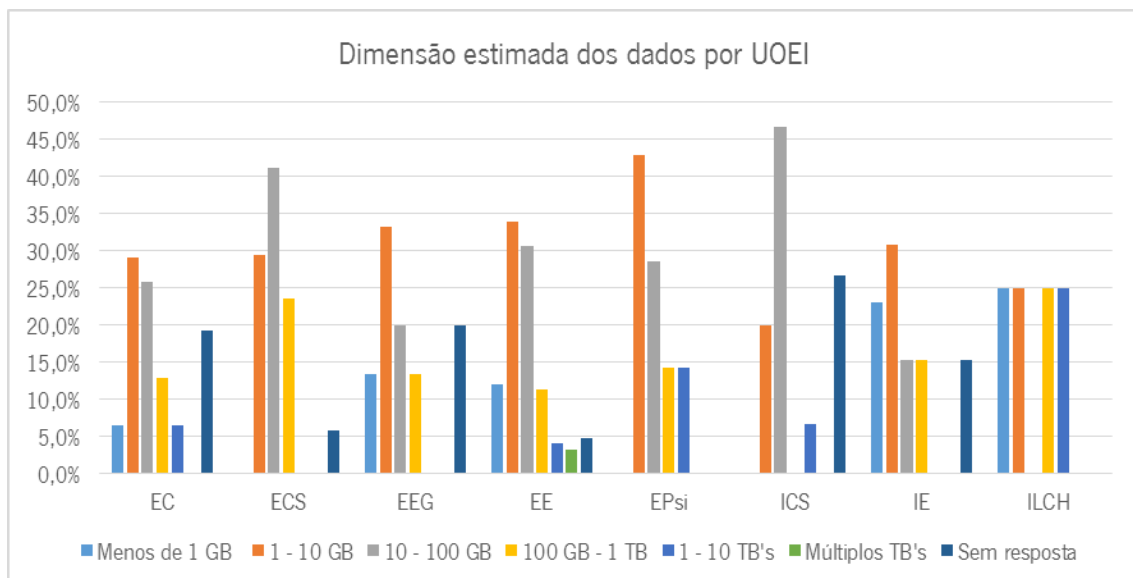


Gráfico 4 – Dimensão estimada dos dados por UOEI

#### 4.6 Algum dos seus 'datasets' está armazenado num formato proprietário?

Relativamente a esta questão (ver Tabela 38), 90,5% dos participantes responderam não possuir *datasets* armazenados num formato proprietário.



	<i>N</i>	%
Não	190	90,5%
Sim	12	5,7%
Sem resposta	8	3,8%
	210	

Tabela 38 – *Datasets* armazenados num formato proprietário

#### 4.6.1 Por favor, identifique o(s) formato(s)

Dos 5,7% dos inquiridos que identificaram os formatos proprietários utilizados para armazenar os seus *datasets* mencionaram formatos associados a sistemas como: Oracle *database*, Matlab, SPSS, Stata, MySQL, alguns formatos associados a *software* de equipamentos específicos como FT-IR Perkin Elmer e *standards* internacionais como Health Level-7.

Da análise destas últimas duas perguntas, pode deduzir-se que existiu uma deficiente interpretação do conceito de formato proprietário, eventualmente entendido por alguns dos participantes como um formato criado pelos próprios, e não propriamente baseado na aplicação ou *software* em que foi gerado (*open source* ou proprietário) como implícito na formulação da pergunta.

#### 4.7 Por favor, indique que tipo de serviços gostaria que a Universidade do Minho fornecesse para garantir o acesso a esses dados em termos futuros?

Quanto ao tipo de serviços que os respondentes gostariam de ver fornecidos pela Universidade do Minho, relativamente ao acesso dos 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 recebeu 40,0% das respostas e a formação e consultoria na gestão de dados científicos 32,9%.

	<i>N</i>	%
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 39 – Tipos de serviços fornecidos pela UMinho

Também nesta questão se registaram variações estatisticamente significativas nas respostas por UOEI. O serviço mais indicado pelos participantes da ECS e da EE foi o sistema de *backup* de ficheiros, enquanto o mais selecionado pelos participantes da EC foram os serviços de armazenamento e preservação. Os participantes da EPsi elegeram o apoio técnico na elaboração de planos de gestão de dados, e os participantes da EEG manifestaram preferência pela formação e consultoria na gestão de dados científicos. Nas outras UOEI as respostas maioritárias distribuíram-se por mais de uma opção. Desta análise foram excluídas as respostas dos participantes da EA e da ED por estarem pouco representadas.

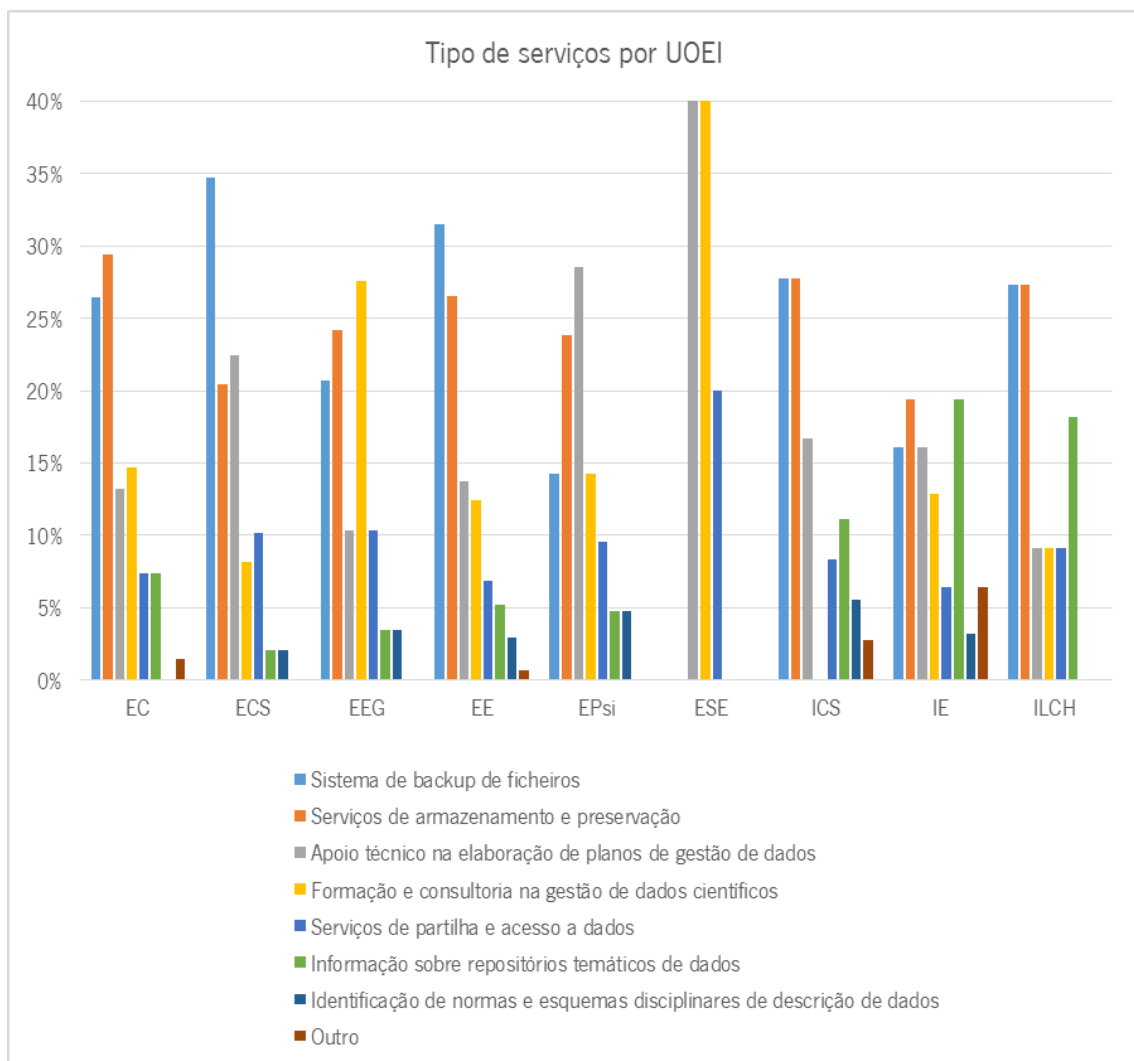


Gráfico 5 – Tipo de serviços fornecidos pela UMinho por UOEI

4.8 Por favor, indique se estaria disposto a participar num projeto no âmbito da curadoria de dados científicos promovido pela Universidade do Minho e/ou numa reunião de 'follow-up' após este inquérito?

81,0% dos participantes manifestou não estar interessado em participar num projeto promovido pela Universidade do Minho no que diz respeito à curadoria de dados científicos. Ao contrário, 19,0% (40 investigadores) demonstraram essa disponibilidade, pelo que se considera existirem condições para continuar e aprofundar o diagnóstico dos dados científicos na UMinho iniciado com este estudo.

	<i>N</i>	%
Não	170	81,0%
Sim	40	19,0%
	210	

Tabela 40 – Participação em projeto UMinho no âmbito da curadoria de dados científicos

## 5 Conclusões

O desenvolvimento do presente estudo permitiu obter dados para consolidar e ajustar o plano de intervenção para os serviços de suporte à Gestão dos Dados Científicos na UMinho delineado para os anos de 2014 e 2015. Os resultados obtidos contribuem para melhorar a definição de algumas das ações programadas no plano de intervenção, possibilitam estabelecer níveis de prioridade mais ajustados à realidade da Universidade e permitem fundamentar a atuação dos Serviços de Documentação neste domínio, particularmente junto dos responsáveis institucionais e dos outros parceiros relevantes.

Em termos de relevância e representatividade da amostra, registe-se a participação de aproximadamente 20% do pessoal docente e investigador que se pretendia inquirir. Apesar do nível de participação no estudo ser bastante diferenciado entre as UOEl (o que em si mesmo pode ser um indicador de diferentes níveis de importância e de sensibilização), as mais de 200 respostas recolhidas permitem obter uma panorâmica do tipo de dados produzidos no âmbito da investigação realizada na UMinho, aferir o nível de sensibilidade para as questões em estudo, antever as necessidades da comunidade e projetar serviços na instituição.

O estudo demonstrou que a esmagadora maioria dos participantes (92,9%) produz dados científicos e que maioritariamente são dados experimentais e estatísticos. Relativamente ao volume aproximado dos dados acumulados existiu uma grande variação nas respostas, mas o facto de cerca de um terço dos participantes ter situado esse volume até 10 GB, e de menos de 10% ter referido volumes iguais ou superiores a 1 TB, permite perspetivar o tipo de necessidades primárias relativamente a serviços de *backup* e armazenamento.

A grande maioria (81,0%) afirmou realizar regularmente cópias de segurança dos dados produzidos, de forma não sistemática e maioritariamente com recurso a discos externos ou discos de PC, mas apontou como serviços prioritários a disponibilizar a necessidade de um sistema de *backup* de ficheiros e de serviços para armazenamento e preservação. Estes resultados reforçam a pertinência de avaliar de forma transversal na Universidade a necessidade de serviços centralizados, que permitam maior segurança nas rotinas de armazenamento e promovam melhores práticas de preservação dos dados científicos dos docentes e investigadores da UMinho.

Os resultados evidenciam a ausência (81,5% das respostas) de planos para a gestão de dados gerados nas atividades de investigação, principalmente devido à inexistência de políticas no seio dos grupos de investigação. No entanto, é bastante significativa a reutilização dos dados científicos produzidos em outros projetos (60,9%), dissertações de mestrado (59,1%) ou teses de doutoramento (56,2%), facto que reforça a conveniência de um planeamento adequado à gestão dos dados. Nesta linha, e em função do número significativo de respostas favorável à disponibilização de serviços de apoio na elaboração de planos de gestão de dados ou de formação e consultoria na gestão de dados científicos, fica reforçada a relevância de ações específicas já planeadas e a desenvolver proximamente pelos Serviços de Documentação neste domínio.

O significativo desconhecimento (83,8%) revelado sobre o projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa-quadro Horizonte 2020, não sendo um facto surpreendente, dado ser um projeto bastante recente, deve traduzir-se no desenvolvimento de ações de informação a breve prazo, já que os princípios estabelecidos pela Comissão Europeia constam também de políticas de outros agentes financiadores da ciência, nomeadamente da Fundação para a Ciência e a Tecnologia em Portugal.

Ficou também evidenciada nas respostas recolhidas a existência de uma grande restrição na possibilidade de acesso aos dados por outros que não sejam do grupo ou projeto de investigação (57,5%). No entanto, os resultados revelam que nunca foi solicitado a estes docentes e investigadores a colocação dos seus dados científicos disponíveis em acesso aberto, fora do âmbito de uma publicação. As maiores preocupações quanto ao acesso aberto aos dados científicos indicados pelos participantes foram as relativas à confidencialidade/direitos de autor.

Face aos resultados obtidos com o questionário pode concluir-se que existe necessidade, espaço e abertura na comunidade da UMinho para o desenvolvimento, implementação e disponibilização de serviços e ferramentas associadas à gestão de dados científicos, em particular:

- Serviços de armazenamento e *backup* de dados;
- Serviços de apoio e consultoria, com particular enfoque nas ferramentas de suporte ao desenvolvimento de planos de gestão de dados;

- Serviços de repositório de dados científicos, possibilitando o armazenamento, referênciação, acesso e reutilização de dados a longo prazo.

No seguimento do seu plano de intervenção (cuja adequação e relevância foi também confirmada pelos resultados do inquérito) os Serviços de Documentação da Universidade do Minho, em consonância com as orientações da Reitoria, e em articulação com outros Serviços e parceiros relevantes no interior e no exterior da Universidade, irão desenvolver e começar a disponibilizar alguns destes serviços nos próximos meses.

## Índice de tabelas

TABELA 1 – NÚMERO DE PARTICIPANTES POR UOEI.....	10
TABELA 2 – NÚMERO DE PARTICIPANTES POR CATEGORIA PROFISSIONAL .....	11
TABELA 3 - NÚMERO DE PARTICIPANTES POR PERFIL DE ATIVIDADE / RESPONSABILIDADE .....	11
TABELA 4 - NÚMERO DE PARTICIPANTES POR DOMÍNIO CIENTÍFICO .....	12
TABELA 5 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: CIÊNCIAS NATURAIS.....	12
TABELA 6 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: ENGENHARIA E TECNOLOGIA .....	12
TABELA 7 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: CIÊNCIAS MÉDICAS .....	13
TABELA 8 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: CIÊNCIAS AGRÁRIAS .....	13
TABELA 9 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: CIÊNCIAS SOCIAIS .....	13
TABELA 10 - NÚMERO DE PARTICIPANTES POR ÁREA CIENTÍFICA: HUMANIDADES.....	14
TABELA 11 – NÚMERO DE PARTICIPANTES POR SEXO .....	14
TABELA 12 – DADOS CIENTÍFICOS PRODUZIDOS NO ÂMBITO DO TRABALHO COMO INVESTIGADOR.....	15
TABELA 13 – INVESTIGAÇÃO DESENVOLVIDA INDIVIDUALMENTE OU EM GRUPO.....	16
TABELA 14 – DADOS CIENTÍFICOS PRODUZIDOS NO CONTEXTO DA INVESTIGAÇÃO.....	16
TABELA 15 – VOLUME GLOBAL DOS <i>DATASETS</i> ACUMULADOS.....	17
TABELA 16 – REALIZAÇÃO DAS CÓPIAS DE SEGURANÇA DOS DADOS CIENTÍFICOS.....	17
TABELA 17 – RESPONSÁVEL PELA REALIZAÇÃO DAS CÓPIAS DE SEGURANÇA DOS DADOS CIENTÍFICOS.....	17
TABELA 18 – RESPONSÁVEL PELA REALIZAÇÃO DAS CÓPIAS DE SEGURANÇA DOS DADOS CIENTÍFICOS, ANÁLISE DA OPÇÃO OUTRO .....	18
TABELA 19 – FREQUÊNCIA DOS <i>BACKUPS</i> .....	18
TABELA 20 – TIPOS DE DADOS CIENTÍFICOS PRESERVADOS .....	19
TABELA 21 – LOCAL DE ARMAZENAMENTO DOS DADOS CIENTÍFICOS.....	19
TABELA 22 – PLANO DE GESTÃO DE DADOS CIENTÍFICOS .....	20
TABELA 23 – MOTIVOS PARA O DESENVOLVIMENTO DO PLANO PARA A GESTÃO DE DADOS.....	20
TABELA 24 – MOTIVOS PARA NÃO POSSUIR UM PLANO PARA A GESTÃO DE DADOS .....	21
TABELA 25 – FREQUÊNCIA DE REUTILIZAÇÃO DOS DADOS PRODUZIDOS.....	21
TABELA 26 – TIPO DE ACESSO DOS DADOS CIENTÍFICOS .....	22
TABELA 27 – PERMISSÃO DE ACESSO AOS DADOS CIENTÍFICOS .....	22
TABELA 28 – QUEM PODE ACEDER AOS DADOS CIENTÍFICOS PRODUZIDOS.....	23
TABELA 29 – PREOCUPAÇÕES RELATIVAMENTE AO ACESSO ABERTO .....	23
TABELA 30 – COLOCAÇÃO DOS DADOS CIENTÍFICOS EM ACESSO ABERTO .....	24
TABELA 31 – CONHECIMENTO SOBRE O PROJETO-PILOTO DE DADOS ABERTOS DA CE.....	24
TABELA 32 – TIPOS DE DADOS PRESERVADOS.....	25
TABELA 33 – TIPOS DE DADOS PRESERVADOS POR UOEI .....	26
TABELA 34 – ESFORÇO ASSOCIADO À CRIAÇÃO DE DADOS .....	26

TABELA 35 – PERÍODO DE PRESERVAÇÃO ESTIMADO .....	27
TABELA 36 – FREQUÊNCIA DE USO .....	29
TABELA 37 – DIMENSÃO ESTIMADA DOS DADOS .....	30
TABELA 38 – <i>DATASETS</i> ARMAZENADOS NUM FORMATO PROPRIETÁRIO.....	31
TABELA 39 – TIPOS DE SERVIÇOS FORNECIDOS PELA UMINHO.....	31
TABELA 40 – PARTICIPAÇÃO EM PROJETO UMINHO NO ÂMBITO DA CURADORIA DE DADOS CIENTÍFICOS .....	33



## Índice de gráficos

GRÁFICO 1 – ESFORÇO ASSOCIADO À CRIAÇÃO DE DADOS POR UOEI .....	27
GRÁFICO 2 – PERÍODO DE PRESERVAÇÃO ESTIMADO POR UOEI .....	28
GRÁFICO 3 – FREQUÊNCIA DE USO POR UOEI.....	29
GRÁFICO 4 – DIMENSÃO ESTIMADA DOS DADOS POR UOEI .....	30
GRÁFICO 5 – TIPO DE SERVIÇOS FORNECIDOS PELA UMINHO POR UOEI .....	32

## Anexo 1: Questionário





Outro

Outro

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

### A3. Perfil de Atividade / Responsabilidade

Responsável de centro de investigação

Coordenador de projeto

Participante de projeto

Aluno de doutoramento

Outro

Outro

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

### A4. Domínio Científico

Ciências Naturais

Engenharia e Tecnologia

Ciências Médicas

Ciências Agrárias

Ciências Sociais

Humanidades

### A5. Área Científica: Ciências Naturais

Matemáticas

Ciências da Computação e da Informação

Ciências Físicas

Ciências Químicas

Ciências da Terra e do Ambiente

Ciências Biológicas

Outras Ciências Naturais



**A6. Área Científica: Engenharia e Tecnologia**

- Engenharia Civil
- Engenharia Eletrotécnica, Eletrónica e Informática
- Engenharia Mecânica
- Engenharia Química
- Engenharia dos Materiais
- Engenharia Médica
- Engenharia do Ambiente
- Biotechnology Industrial
- Biotechnology Ambiental
- Nanotechnology
- Outras Engenharias e Tecnologias

**A7. Área Científica: Ciências Médicas**

- Medicina Básica
- Medicina Clínica
- Ciências da Saúde
- Biotechnology Médica
- Outras Ciências Médicas

**A8. Área Científica: Ciências Agrárias**

- Agricultura, Sivicultura e Pescas
- Ciência Animal e dos Laticínios
- Ciências Veterinárias
- Biotechnology Agrária
- Outras Ciências Agrárias



### A9. Área Científica: Ciências Sociais

- Psicologia
- Economia e Gestão
- Ciências da Educação
- Sociologia
- Direito
- Ciências Políticas
- Geografia Económica e Social
- Ciências da Comunicação
- Outras Ciências Sociais

### A10. Área Científica: Humanidades

- História e Arqueologia
- Línguas e Literaturas
- Filosofia, Ética e Religião
- Artes
- Outras Humanidades

### A11. Sexo

- Feminino
- Masculino

## Section B: B. Curadoria e gestão de dados científicos

Por **curadoria (ou gestão) de dados científicos** entende-se o conjunto de ações que garantem que um conjunto de dados é genuíno, permitindo o seu uso por outros que não os seus produtores. As ações de curadoria e de gestão de dados têm alguma interseção, sendo as de gestão mais independentes do conteúdo e do uso. Nesta seção, deve considerar ainda o termo **dataset** como um conjunto de dados científicos resultantes da sua atividade de investigação, podendo incluir dados numéricos produzidos por experiências computacionais, resultados obtidos com equipamento experimental, imagens ou áudio criados a partir de entrevistas ou dados recolhidos como parte de um projeto ou até mesmo dados recolhidos a partir de inquéritos relacionados com um projeto.

### B1. No âmbito do seu trabalho como investigador ou dos projetos em que participa produz ou possui dados científicos?

*Atenção: Caso responda "não" o questionário será dado por terminado.*

- Sim
- Não





















**C6. Algum dos seus 'datasets' está armazenado num formato proprietário?**

Sim

Não

**C7. Por favor, identifique o(s) formato(s)**

**C8. Por favor, indique que tipo de serviços gostaria que a Universidade do Minho fornecesse para garantir o acesso a esses dados em termos futuros?**

Serviços de armazenamento e preservação

Sistema de backup de ficheiros

Apoio técnico na elaboração de planos de gestão de dados

Formação e consultoria na gestão de dados científicos

Serviços de partilha e acesso a dados

Informação sobre repositórios temáticos de dados

Identificação de normas e esquemas disciplinares de descrição de dados

Outro

Outro

**C9. Por favor, indique se estaria disposto a participar num projeto no âmbito da curadoria de dados científicos promovido pela Universidade do Minho e/ou numa reunião de 'follow-up' após este inquérito?**

Sim

Não

**C10. Por favor, indique-nos o seu nome e correio eletrónico para eventual contacto**

Nome

Correio eletrónico



**Inquérito submetido com sucesso. Muito obrigado pela sua colaboração! Para esclarecimentos adicionais, por favor, contacte: [openaccess@sdum.uminho.pt](mailto:openaccess@sdum.uminho.pt) A equipa dos SDUM**