

30 de abril de 2015 | Intercâmbio
RepositóriUM -



Plano de intervenção para a gestão de dados científicos na UMinho

Ricardo Saraiva

rsaraiva@s dum.uminho.pt

Tópicos da apresentação

1. Motivação e enquadramento: gestão de dados científicos na Universidade do Minho
2. Âmbito e caracterização do estudo desenvolvido
3. Resultados do estudo sobre curadoria e gestão de dados científicos
4. Resultados do estudo sobre o acervo de dados científicos existente
5. Conclusões e trabalho futuro



1

MOTIVAÇÃO & ENQUADRAMENTO



Motivação & enquadramento



Universidade do Minho
Serviços de Documentação

- A gestão de dados científicos tem assumindo **crecente relevância no contexto das instituições**.
- **reforçar** a sua capacidade de gestão e curadoria dos seus recursos para **apoiarem as necessidades de dados** dos investigadores.
- **Requisitos dos financiadores de ciência:**
 - **Comissão Europeia (CE)** criou recentemente, no âmbito do programa quadro Horizonte 2020, um **projeto-piloto de dados abertos**.
 - A **FCT** em Portugal adotou no contexto da sua política de Acesso

A Digital Curation Centre
Working level guide

DICC JISC

How to Develop Research Data Management Services
- a guide for HEIs
Sarah Jones, Graham Prior and Angus Whyte

The EU Framework Programme for Research and Innovation
HORIZON 2020

FCT Fundação para a Ciência e a Tecnologia
SECRETARIA DE INOVAÇÃO E CIÊNCIA

Política sobre a Disponibilização de Dados e outros Resultados de Projetos de I&D Financiados Pela FCT
(atualizada em 5 de maio de 2016)

- 1. Introdução**

Tem sido reconhecido em várias instâncias que a disponibilização aberta dos resultados da investigação realizada com financiamento público tem significativos benefícios sociais e económicos. Em sintonia com o que tem vindo crescendo e acordado por agências públicas de financiamento de I&D de outros países e com as recomendações da Comissão Europeia de 17 de Julho de 2012¹, a FCT – Fundação para a Ciência e a Tecnologia, III – adota a presente política sobre a disponibilização de resultados de investigação e desenvolvimento (I&D) que financia.
- 2. Orientações gerais**

A FCT encoraja os investigadores a disponibilizarem os dados resultantes dos projetos de I&D que financia em bases de dados de acesso aberto apropriadas, quando tal for possível. Entre outras, figuram, por exemplo, entre muitas outras, as bases de dados de sequências de nucleótidos tipo DDB/EMBL/GenBank e de estrutura de proteínas tipo RCSB-PDB/EMBL-EBI/PDB. Reconvém, contudo, que há resultados de I&D que, por natureza, não podem ser disponibilizados desta forma, a disponibilização por outra forma adequada fica ao critério dos investigadores.

A disponibilização dos resultados não pode, em qualquer circunstância, incluir informação que comprometa a privacidade ou a segurança de indivíduos ou entidades envolvidas na investigação, sujeitando-se aos condicionamentos legais vigentes. A FCT encoraja ainda os investigadores a promoverem a sua participação em iniciativas nacionais e internacionais, que procurem as formas mais adequadas de partilha de dados nas diferentes áreas do conhecimento.
- 3. Recomendações aos beneficiários de financiamento**

A FCT recomenda que os beneficiários de financiamento partilhem os dados primários e outros materiais produzidos em projetos financiados pela FCT com outros investigadores, e não mais do que outros investigadores e dentro de um prazo razoável, segundo as melhores práticas da disciplina científica em que se insere o projeto em questão.

Nas candidaturas a financiamento apresentadas à FCT, sugere-se que os candidatos 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, incluindo:

 - 1) descrições de tipos de dados, amostras, coleções, software, modelos, materiais curriculares e outros materiais que se prevê venham a ser produzidos no âmbito das atividades financiadas;

¹ Comunicação Tercera sobre acesso a ciência, informação, inovação e benefícios públicos investidos em inovação.
Recomendação do Conselho de 17 de Julho de 2012 sobre o acesso a dados científicos.

SDUM - QUAR 2014



Universidade do Minho
Serviços de Documentação

OBJ.3

Desenvolver um programa para a criação de serviços de gestão de dados científicos na UMinho

Ind. 3.1 Diagnóstico dos dados científicos na UMinho

Ind. 3.2

Conteúdos/Serviços de apoio e consultoria de gestão dados científicos

Ind. 3.3 Gestão dos dados produzidos/recolhidos nos SDUM

Plano de intervenção dos SDUM



Universidade do Minho
Serviços de Documentação

- Resulta de um trabalho de análise da equipa de projetos Open Access
- Define **10 eixos de intervenção**
- Enquadrados em **4 áreas estratégicas**
- **33 ações programadas para 2014 e 2015**



Serviços de Documentação da Universidade do Minho

PLANO DE INTERVENÇÃO

2014-2015

SERVIÇOS DE SUPORTE À GESTÃO
DOS DADOS CIENTÍFICOS NA UM



Março de 2014

Áreas estratégicas do plano de intervenção:



Universidade do Minho
Serviços de Documentação

Diagnóstico & política

Serviços de suporte & informação

Infraestruturas & normalização

Competências & recursos humanos

Poster ConfOA 2014

PLANO DE INTERVENÇÃO DOS SERVIÇOS DE DOCUMENTAÇÃO PARA SUPORTE À GESTÃO DOS DADOS CIENTÍFICOS NA UMINHO

Os Serviços de Documentação da Universidade do Minho (SDUM) estabeleceram para os anos de 2014 e 2015 um plano de intervenção para os serviços de suporte à Gestão dos Dados Científicos na UMinho. Esta nova área de intervenção inscrita nos objetivos estratégicos dos SDUM, visa contribuir para valorizar a investigação e o conhecimento produzidos na UM e demonstrar a importância do desenvolvimento de serviços que suportem a ação dos investigadores na gestão de dados científicos.

10 ÁREAS DE INTERVENÇÃO **4** ÁREAS ESTRATÉGICAS **33** AÇÕES DE IMPLEMENTAÇÃO

Diagnóstico & política
Análise da realidade da gestão dos dados científicos produzidos na UMinho.
Realizar observação de práticas de produção, planeamento e armazenamento de dados científicos e estado dos conteúdos institucionais dentro da UMinho.
Desenvolvimento da política institucional para a gestão de dados científicos.
Participação ativa na discussão, formulação e implementação da política, incluindo a disseminação de boas práticas e incentivo à adoção de políticas de dados abertos no ciclo de vida dos dados de investigação.

Serviços de suporte & informação
Serviços de apoio e consultoria na gestão de dados científicos.
Inclui planos de gestão de dados para projetos financiados e desenvolvimento de materiais informativos e guias formativos.
Serviços de curadoria para o ciclo de vida dos dados.
Apoiar os investigadores e grupos de investigação no ciclo de vida dos dados científicos (recolha, análise, armazenamento e publicação) promovendo sempre de armazenamento, pesquisa e acesso.
Promoção da referência e citação dos dados científicos.
Promover a contextualização dos dados, clarificando regras a utilizar, melhorando a utilização dos identificadores persistentes.

Infraestrutura & normalização
Serviço de repositório de dados científicos.
Oligar no sentido de criar um repositório de dados em aberto disponível ao suportar sempre de depósito de dados científicos.
Promoção da interoperabilidade de infraestruturas de dados científicos.
Localização de parcerias com investigadores, grupos de investigação, repositórios, arquivos e bases de dados para promover o acesso, utilização e partilha dos dados.
Serviços de metadados para dados de investigação.
Realizar o reconhecimento e identificação de normas e esquemas de metadados e eventual envolvimento no desenvolvimento de serviços de metadados para dados científicos.

Competências & recursos humanos
Especialização de um serviço interno dedicado para dados científicos e desenvolvimento de competências dos colaboradores SDUM.
Focar na formação interna da equipa dos SDUM tendo em vista a melhoria das competências e definição de serviços a prestar.
Melhorar a prática da gestão dos dados científicos e a disponibilização de dados abertos.
Inclui o estabelecimento de metodologias internas, por um planeamento da gestão de dados científicos e para a disponibilização de dados em acesso aberto de forma documentada.

Poster desenvolvido por: José Carvalho, Ricardo Soares, Filipa Rodrigues, Cláudia Soares e Rafael Travençolo
www.sdum.uminho.pt



Eixos de intervenção

- 1. Análise da realidade da gestão dos dados científicos produzidos na UMinho.**
- 2. Serviços de apoio e consultoria na gestão de dados científicos.**
- 3. Especificação de um serviço interno dedicado para dados científicos e desenvolvimento de competências dos colaboradores SDUM.**
- 4. Desenvolvimento da política institucional para a gestão de dados científicos.**
- 5. Promoção da interoperabilidade de infraestruturas de dados científicos.**
- 6. Serviços de curadoria para o ciclo de vida dos dados científicos.**
- 7. Serviços de metadados para dados de investigação.**
- 8. Promoção da referenciação e citação dos dados científicos.**

Inquérito sobre Dados Científicos



Universidade do Minho
Serviços de Documentação

Inquérito sobre Dados Científicos

Este inquérito por questionário pretende recolher informações com vista a **científicos** (*research data / scientific data*) que são gerados no âmbito da

Para o propósito deste estudo, considera-se *research data* todos os dados utilizados no decorrer da sua investigação na Universidade do Minho. Os dados podem ser dados primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por

O questionário demorará cerca de 10 minutos a ser respondido.

Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais

Existem 53 pessoas



Este inquérito por questionário pretende recolher informações com vista a proporcionar um melhor conhecimento sobre os dados científicos ('research data' / 'scientific data') que são gerados no âmbito da investigação produzida na Universidade do Minho. Para o propósito deste estudo, considera-se *research data* todos os dados que atualmente possui, que foram recolhidos e/ou utilizados no decorrer da sua investigação na Universidade do Minho. Os dados científicos podem ser dados primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por terceiros. O questionário demorará cerca de 10 minutos a ser respondido. Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais, por favor, contacte: openaccess@sdum.uminho.pt

Section A: A. Identificação

A1. Unidade Orgânica de Ensino e de Investigação (UOEI)

- Escola de Arquitectura (EA)
- Escola de Ciências (EC)
- Escola de Ciências da Saúde (ECS)
- Escola de Direito (ED)
- Escola de Economia e Gestão (EEG)
- Escola de Engenharia (EE)
- Escola de Psicologia (EP)
- Escola Superior de Enfermagem (ESE)
- Instituto de Ciências Sociais (ICS)
- Instituto de Educação (IE)
- Instituto de Letras e Ciências Humanas (ILCH)
- Outro

A2. Categoria Profissional

- Docente Universitário
- Investigador
- Técnico Superior
- Bolseiro(a) de Investigação



2

ÂMBITO & CARACTERIZAÇÃO



Âmbito e caracterização



Universidade do Minho
Serviços de Documentação

- Inquérito por questionário:
 - **Anónimo** e composto por **53 perguntas**
 - Disponibilizado **via web** usando o software **LimeSurvey**
 - Dirigido a investigadores, docentes e centros de investigação da UMinho
- Divulgado exclusivamente através de **e-mail enviado pelo Vice-Reitor** responsável pela área da investigação.

- Período de resposta : **14 de março e 17 de abril de 2014.**
226 respostas válidas
de um universo total de 353 respostas.

Estrutura do inquérito



Este inquérito por questionário pretende recolher informações com vista a proporcionar um melhor conhecimento sobre os dados científicos ('research data', 'scientific data') que são gerados no âmbito da investigação produzida na Universidade do Minho. Para o propósito deste estudo, considera-se *research data* todos os dados que atualmente possui, que foram recolhidos e/ou utilizados no decorrer da sua investigação na Universidade do Minho. Os dados científicos podem ser dados primários recolhidos por si ou pelo seu grupo de investigação ou dados secundários fornecidos por terceiros. O questionário demorará cerca de 10 minutos a ser respondido. Agradecemos desde já a sua colaboração! Para esclarecimentos adicionais, por favor, contacte: openaccess@sdum.uminho.pt

Section A: A. Identificação

A1. Unidade Orgânica de Ensino e de Investigação (UOEI)

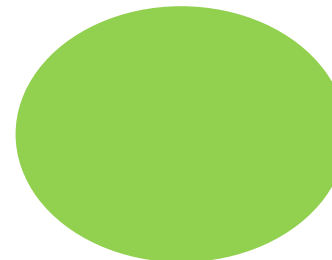
- Escola de Arquitectura (EA)
- Escola de Ciências (EC)
- Escola de Ciências da Saúde (ECS)
- Escola de Direito (ED)
- Escola de Economia e Gestão (EEG)
- Escola de Engenharia (EE)
- Escola de Psicologia (EP)
- Escola Superior de Enfermagem (ESE)
- Instituto de Ciências Sociais (ICS)
- Instituto de Educação (IE)
- Instituto de Letras e Ciências Humanas (ILCH)
- Outro

Outro

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

A2. Categoria Profissional

- Docente Universitário
- Investigador
- Técnico Superior
- Bolseiro(a) de Investigação



Caracterização dos participantes



Universidade do Minho
Serviços de Documentação

- **Respostas de 10 das 11 unidades** existentes na UMinho;
- **Metade dos participantes pertence à Escola de Engenharia (55,8%)** seguindo-se a Escola de Ciências com 13,7%;
- Categoria profissional, identificaram-se: **54,9% docentes universitários, 27,0% bolseiros e 20,4% investigadores;**
- Perfil da atividade: **52,2% são participantes em projetos, 35% coordenadores de projetos, 15,5% alunos de doutoramento;**
- Representatividade da **amostra**, regista-se a



3

CURADORIA E GESTÃO DE DADOS CIENTÍFICOS



3. Curadoria e gestão de dados científicos (1/9)



Universidade do Minho
Serviços de Documentação

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

No seu trabalho como investigador ou dos projetos em que participa **produz ou possui dados científicos?**

93% SIM

3. Curadoria e gestão de dados científicos (2/9)



	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	

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

75% Experimentais . 48% Estatísticos . 41% Imagens

3. Curadoria e gestão de dados científicos (3/9)



	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	

Qual é o volume global aproximado de dados acumulados por si ou pelo seu grupo de investigação?

25% não consegue estimar
24% entre 1-10GB
21% entre 10-100GB



3. Curadoria e gestão de dados científicos (4/9)

	<i>N</i>	<i>%</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	

Onde são guardados os dados?

80% disco externo

52% disco de PC

32% dropbox

3. Curadoria e gestão de dados científicos (5/9)



Universidade do Minho
Serviços de Documentação

	N	%
Não	169	80,5%
Sim	41	19,5%
	210	

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)?

80% NÃO

3. Curadoria e gestão de dados científicos (6/9)



Universidade do Minho
Serviços de Documentação

	N	%
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	

Qual foi o **principal motivo para o desenvolvimento de um plano de gestão de dados?**

44% para aceder, analisar e consultar outros dados

15% tamanho da equipa

15% volume de dados associados ao

3. Curadoria e gestão de dados científicos (7/9)



Universidade do Minho
Serviços de Documentação

	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	

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

57% acesso restrito ao grupo ou projeto

25% acesso aberto imediato ou após embargo

3. Curadoria e gestão de dados científicos (8/9)



	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	

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

79% Confidencialidade / propriedade intelectual

44% potencial reutilização

3. Curadoria e gestão de dados científicos (9/9)



Universidade do Minho
Serviços de Documentação

	N	%
Não	176	83,8%
Sim	34	16,2%
	210	

Tem conhecimento do projeto-piloto de dados abertos da Comissão Europeia no âmbito do programa quadro H2020?

84% NÃO

16% SIM



4

ACERVO DE DADOS CIENTÍFICOS EXISTENTE



4. Acervo de dados científicos existentes (1/4)



Universidade do Minho
Serviços de Documentação

	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	

Período de preservação estimado

27% 2-5 anos

27% 5-10 anos

22% indefinido

4. Acervo de dados científicos existentes (2/4)



	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

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

34% 1-10 GB

30% 10-100 GB

14% 100 GB - 1 TB



4. Acervo de dados científicos existentes (3/4)

	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	

Que tipo de dados deverão ser preservados?
69% dados brutos de instrumentos
69% documentos de texto
68% folhas de cálculo

4. Acervo de dados científicos existentes (4/4)



Universidade do Minho
Serviços de Documentação

	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	

que tipo de serviços gostaria que a UMinho fornecesse para garantir o acesso a esses dados em termos futuros?

73% sistema de backup de ficheiros

65% serviços de armazenamento e preservação

40% apoio técnico na elaboração de



5

CONCLUSÕES & TRABALHO FUTURO



Conclusões e trabalho futuro



Universidade do Minho
Serviços de Documentação

- Permitiu obter dados para **consolidar e ajustar o plano de intervenção** para serviços de suporte à Gestão dos Dados Científicos na UMinho.
- As mais de 200 respostas recolhidas permitem obter uma **melhor panorâmica do tipo de dados produzidos** no âmbito da investigação realizada na UMinho.
- A **grande diferenciação** entre as Unidades Orgânicas de Ensino e Investigação no nível de participação do estudo é **um indicador de diferentes graus de importância e de sensibilização**.
- Os resultados indiciam **abertura para a disponibilização em breve de serviços**

Conclusões e trabalho futuro



Universidade do Minho
Serviços de Documentação

- **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

Conclusões e trabalho futuro



Universidade do Minho
Serviços de Documentação

- Os serviços já em realização pelos SDUM (suporte da **ferramenta DMPonline para planos de gestão de dados; guias (in)formativos e sessões de formação interna...**) e os projetos a desenvolver proximamente na Universidade do Minho serão:
 - ✓ em consonância com as orientações da Reitoria,
 - ✓ em articulação com outros Serviços da UMinho,
 - ✓ em parceria com parceiros relevantes fora da UMinho.



Relatório e dados disponíveis no

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>

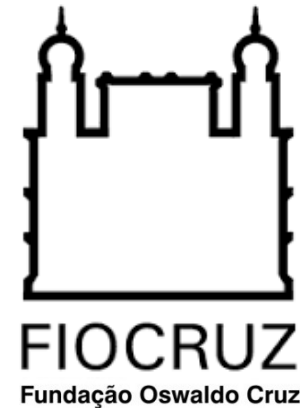


DADOS DO INQUÉRITO

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



30 de abril de 2015 | Intercâmbio
RepositóriUM -



Gestão de dados científicos

openaccess@s dum.uminho.pt

Universidade do Minho Serviços de Documentação



Tópicos da apresentação

1. Dados científicos: conceitos e tipologias
2. Gestão de dados científicos: relevância e desenvolvimentos
3. Ciclo de vida dos dados científicos e etapas na gestão
4. O papel das bibliotecas:
Recomendações da LIBER
5. Outras sugestões



1

DADOS CIENTÍFICOS: CONCEITOS E TIPOLOGIAS



O que são dados científicos...



Universidade do Minho
Serviços de Documentação

Data



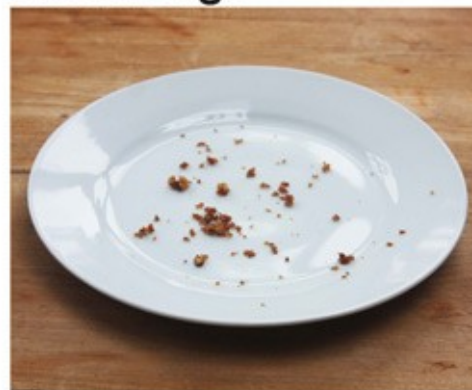
Information



Presentation



Knowledge

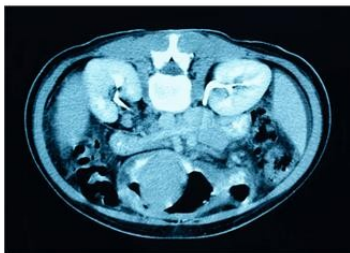


EpicGraphic.com

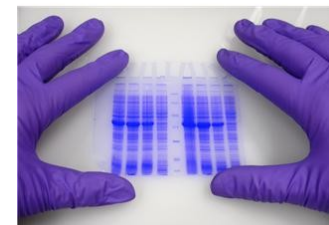
<http://epicgraphic.com/data-cake>



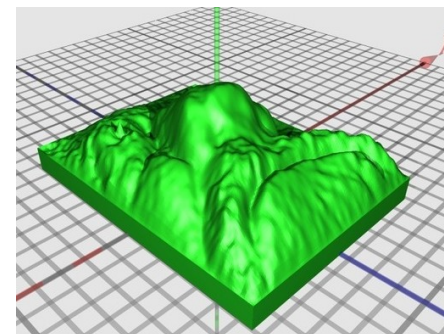
O que são dados científicos...



10/29/04	124.761606	129.27356	122.260995	209.662
11/1/04	119.977679	129.534731	124.739135	176.316
11/2/04	130.46875	135.839924	130.84732	168.289
11/3/04	135.895502	149.510531	140.795689	120.686
11/4/04	134.127052	140.495868	132.823819	206.138
11/5/04	129.851598	137.880438	124.888856	189.675
11/8/04	123.797241	131.84633	126.146789	202.496
11/9/04	118.435374	130.691651	112.877008	140.366
11/10/04	112.401212	121.561443	114.237637	125.298
11/11/04	112.388488	128.496503	113.302591	192.223
11/12/04	129.011813	138.880759	128.517198	108.701
11/15/04	127.077465	139.289941	129.528986	127.406
11/16/04	124.9785	135.363241	127.454638	129.669
11/17/04	124.294035	133.242253	124.704841	244.567
11/18/04	125.663717	135.159011	125.476984	169.271
11/19/04	123.704853	127.612613	124.25382	170.401
11/22/04	118.926697	122.818967	115.379664	134.970



Todo o tipo de registos produzidas no decurso da pesquisa.



Definição da OCDE



Universidade do Minho
Serviços de Documentação

“registos factuais usados como fontes primárias na investigação científica, e que são geralmente aceites na comunidade científica como necessários para validar os resultados de investigação”.





Dados científicos

- Podem assumir várias **formas** (texto, números, imagens fixas, imagens em movimento, etc.) e **dimensões**, desde registos de observações individuais ou ensaios de pequenos laboratórios que não ultrapassarão algumas centenas de kilobytes, até aos dados produzidos pelo *Large Hadron Collider (LHC)* do CERN, que pode gerar várias dezenas de petabytes por dia.

“Os repositórios de dados científicos: estado da arte” - [http://](http://hdl.handle.net/1822/10830)

hdl.handle.net/1822/10830



Dados científicos

- Analógicos ou digitais
- Livros de laboratório ou software
- Criados na forma digital ('born digital')
- Ou convertidos para a forma digital ('digitised')



Dados científicos

- Produzidos ou utilizados no contexto de pesquisa científica:
 - **recolhidos ou criados para efeito de processamento científico**, como os dados atmosféricos usados para previsão meteorológica, ou os dados recolhidos de sensores para monitorizar o estado de um edifício.
 - **obtidos como resultados do processamento automático de objetos** (por exemplo, uma coleção de imagens processada para obter os respetivos histogramas de cor, que constituem assim novos dados).
 - **dados que não são produzidos para investigação mas que acabam por ser objeto dela**, como as contribuições que os utilizadores de uma rede social fazem na forma de textos, fotografias ou outros objetos e que acabam por ser utilizados para estudos sociológicos.

Tipos de dados científicos



Universidade do Minho
Serviços de Documentação

- Experimentais
- Computacionais
- Observacionais
- Numéricos
- Estatísticos
- Textuais
- Imagens
- Áudio
- Multimédia
- ...

Ver MANTRA: http://datalib.edina.ac.uk/xerte/play.php?template_id=9#item0



Dados científicos

- Para se constituírem como **verdadeiramente úteis, os dados científicos devem possuir estrutura e organização.**
 - Os conjuntos de dados (“*datasets*”) são uma das unidades essenciais.
 - Os conjuntos de dados são coleções de informações ou factos relacionados entre si e registados num formato comum.
 - Por exemplo, os resultados de um estudo de opinião por entrevista numa investigação sociológica constituem um conjunto de dados, composto pelos registos individuais das entrevistas.

Os repositórios de dados científicos: estado da arte” - [http://](http://hdl.handle.net/1822/10830)

hdl.handle.net/1822/10830

Dados científicos

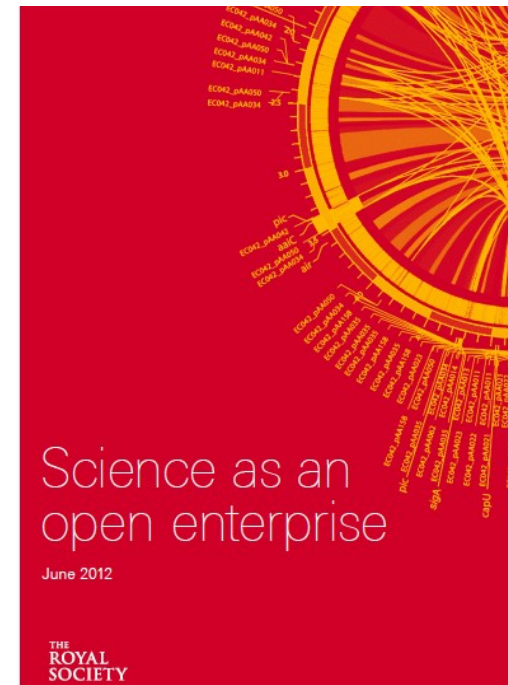


Universidade do Minho
Serviços de Documentação

Data terms

Data relationships	Definition
Data	Numbers, characters or images that designate an attribute of a phenomenon.
Information	Data become information when they are combined together in ways that have the potential to reveal patterns in the phenomenon.
Knowledge	Information yields knowledge when it supports non-trivial, true claims about a phenomenon.

<http://royalsociety.org/policy/projects/science-public-enterprise/report>



Dados científicos



Universidade do Minho
Serviços de Documentação

Data type	Definition
Big Data	Data that requires massive computing power to process.
Broad Data	Structured big data, so that it is freely available through the web to everyone, eg on websites like www.data.gov
Data	Qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (eg direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation.
Data-gap	When data becomes detached from the published conclusions
Data-intensive science	Science that involves large or even massive datasets
Data-led approach	Where hypotheses are constructed after identifying relationships in the dataset.
Data-led science	The use of massive datasets to find patterns as the basis of research.
Dataset	A collection of factual information held in electronic form where all or most of the information has been collected for the purpose of provision of a service by the authority or carrying out of any other function of the authority. Datasets contain factual information which is not the product of analysis or interpretation other than calculation, is not an official statistic, and is unaltered and un-adapted since recording.
Linked Data	Linked data is described by a unique identifier. It contains identifiers for other resources that would not otherwise be connected, in a way that can be automatically processed.
Metadata	Metadata "data about data", contains information about the data, such as how it was generated, who created it, its format, licensing terms, and standards it follows.
Open Data	Open data is data that meets the criteria of being available, assessable and intelligible.
Semantic Data	Data that are tagged with particular vocabularies to describe relationships between data.





2

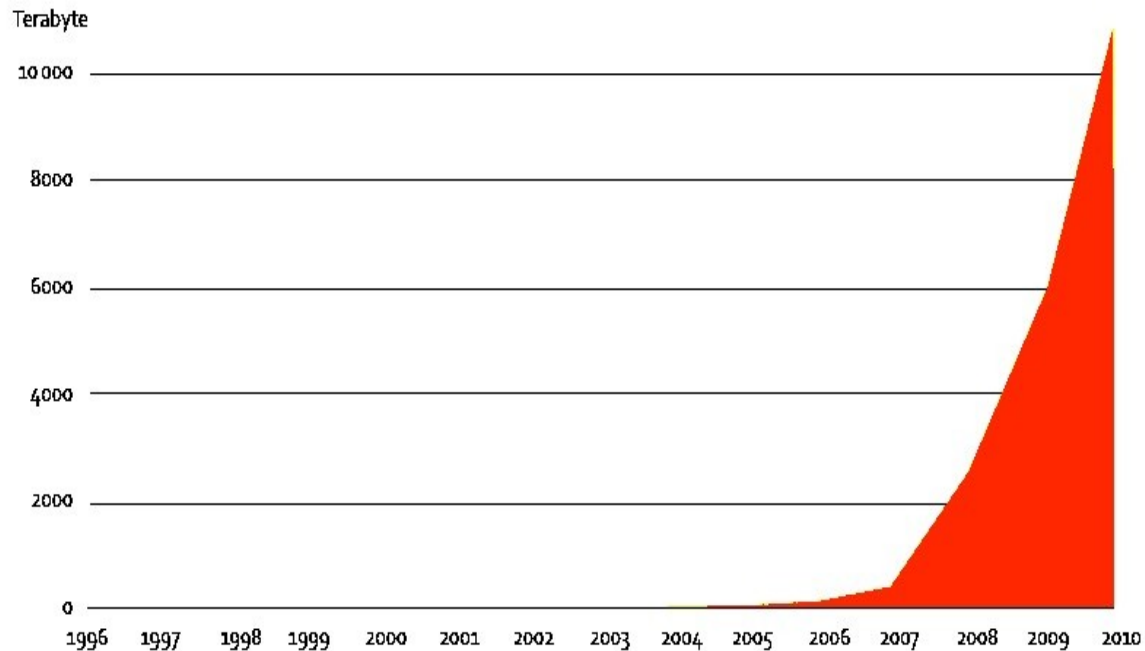
GESTÃO DE DADOS CIENTÍFICOS: RELEVÂNCIA E DESENVOLVIMENTOS





Relevância do tema

Maior consciência da **necessidade de gerir o acesso e a utilização dos dados** produzidos ou recolhidos no âmbito das atividades de investigação, garantindo a sua preservação



European Bioinformatics Institute (EBI)



Relevância do tema

- Cresce a **sensibilidade dos investigadores** para os efeitos da verdadeira explosão na produção de dados científicos:
 - crescimento global das atividades de investigação
 - consequência dos novos métodos e instrumentos de pesquisa e registo que originam cada vez maiores volumes de dados.
 - Onde o trabalho de investigação é baseado na análise de dados recolhidos de forma distribuída parece ter sido mais precoce e



Relevância do tema

- **Os organismos que produzem e financiam ciência** têm manifestado nestes últimos anos redobrada atenção à necessidade de assegurar meios e mecanismos para a gestão dos dados científicos.

Open Access no Horizonte 2020



Universidade do Minho
Serviços de Documentação



Depositar uma cópia legível por máquinas da versão publicada ou da versão final com revisão por pares aceite para publicação num repositório digital de publicações científicas.



Multi-beneficiary General
Model Grant Agreement

Version 1.0
11 December 2013

http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf



Dados Científicos

Abertos

Linhas de orientação:

Projeto-piloto que visa melhorar e maximizar o acesso e a reutilização dos dados de investigação gerados por projetos.

Nos projetos participantes no piloto é requerido o depósito de dados de investigação:

- Os dados, incluindo metadados associados, necessários para validar os resultados apresentados em publicações científicas, logo que possível;
- Outros dados, incluindo metadados associados, conforme especificado e dentro dos prazos estabelecidos no plano de gestão de dados.

www.openaire.eu/open-access/open-access-in-h2020/h2020-oa-data

Política da FCT (Portugal)



Universidade do Minho
Serviços de Documentação

FCT Fundação para a Ciência e a Tecnologia
agência de investigação científica

Disponibilização de Dados e outros Resultados de I&D

DADOS

Orientações gerais

Investigadores são **encorajados a disponibilizarem os dados (e outros *outputs*) resultantes** dos projetos de I&D financiados pela FCT em bases de dados de Acesso Aberto (Ex. bases de dados de sequências de nucleótidos tipo DDBJ/EMBL/GenBank)

Disponibilização fica ao critério dos investigadores

A partilha de dados têm enorme potencial para acelerar avanços no conhecimento, a sua aplicação e transformação e deve ser encorajada a médio prazo

Reforço da importância dos dados abertos (1/3)



Universidade do Minho
Serviços de Documentação

The screenshot shows the American Economic Association (AEA) website. The header includes the AEA logo and navigation links for Journals, Annual Meeting, EconLit, Jobs, Resources, and Members. The main content area displays the article title "Growth in a Time of Debt" from the May 2010 issue. It provides the article citation, DOI (10.1257/aer.100.2.573), and full-text access link. The authors listed are Carmen M. Reinhart (U MD) and Kenneth S. Rogoff (Harvard U). JEL Classifications include E23, E31, H63, O11, and O47. A sidebar on the right offers quick tools like social media sharing, print, export, and email alerts, along with an explore section for the journal's archive.

Kenneth Rogoff e Carmen Reinhart, economistas norte-americanos, publicaram em 2010 um estudo – **Crescimento num tempo de dívida** – que concluía, usando uma base de dados histórica de 20 países avançados, que, nas economias onde a dívida pública superava os 90% do PIB, a taxa de crescimento média era de apenas 0,1%, em contraponto com crescimentos entre 3% e 4% nos países com dívida abaixo de 90%.

<http://www.aeaweb.org/articles.php?doi=10.1257/aer.100.2.573>

Reforço da importância dos dados abertos (2/3)



Universidade do Minho
Serviços de Documentação



Carmen Reinhart



Kenneth Rogoff

Yet, they were proven wrong. Thomas Herndon, Michael Ash and Robert Pollin from the University of Massachusetts (UMass) **tried to replicate the results of Reinhart and Rogoff** and criticised them on the basis of three reasons:

- **Coding errors:** due to a spreadsheet error five countries were excluded completely from the sample resulting in significant error of the average real GDP growth and the debt/GDP ratio in several categories
- **Selective exclusion of available data and data gaps:** Reinhart and Rogoff exclude Australia (1946-1950), New Zealand (1946-1949) and Canada (1946-1950). This exclusion is alone responsible for a significant reduction of the estimated real GDP growth in the highest public debt/GDP category
- **Unconventional weighting of summary statistics:** the authors do not discuss their decision to weight equally by country rather than by country-year, which could be arbitrary and ignores the issue of serial correlation.

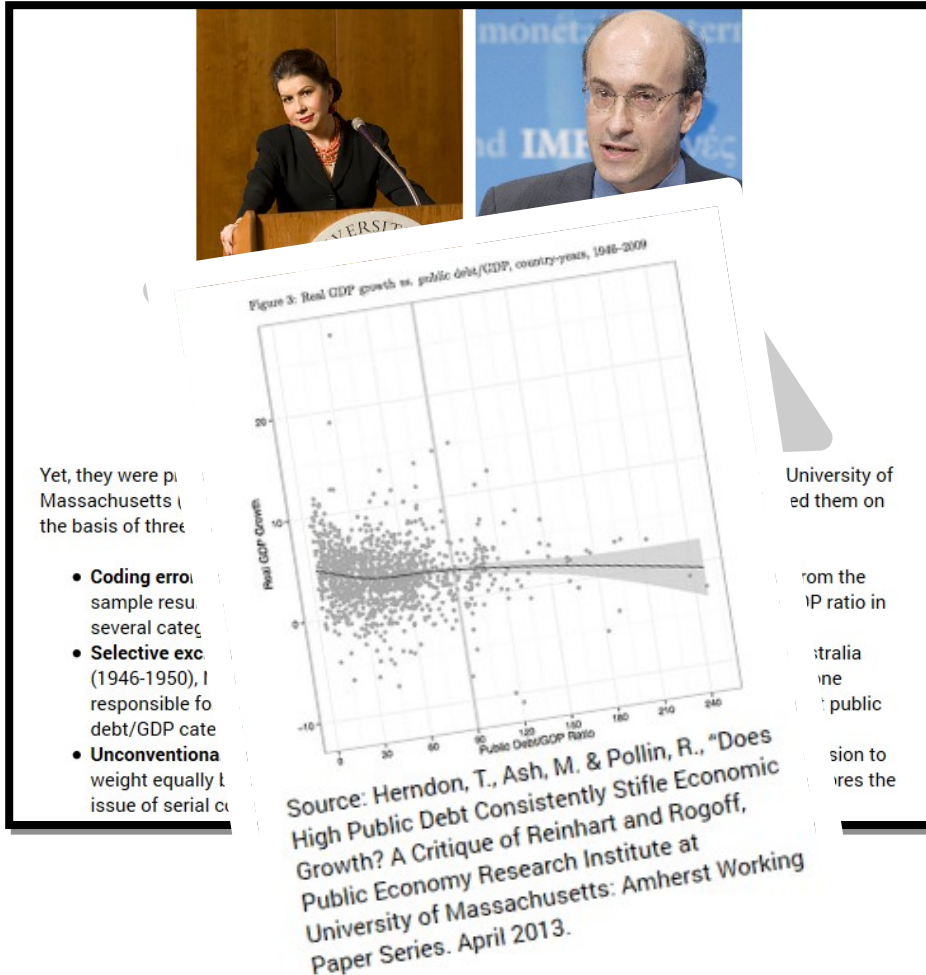
Outra pesquisa publicada por 3 economistas da UMass – Thomas Herndon, Michael Ash e Robert Pollin – **usou a mesma base de dados de Rogoff e Reinhart** (cedida por estes) e **chegou a conclusões diferentes**. Em vez de um crescimento médio de 0,1%, encontraram uma variação do PIB para os países com dívida acima de 90% de 2,2%. **Concluíram que a relação entre crescimento e dívida pública varia muito de acordo com a época e o país analisado.**

<http://blogs.lse.ac.uk/impactofsocialsciences/2013/04/24/reinhart-rogoff-revisited-why-we-need-open-data-in-economics>

Reforço da importância dos dados abertos (3/3)



Universidade do Minho
Serviços de Documentação



Detectados 3 erros nos dados de Rogoff e Reinhart:

- 1) Ficaram de fora da análise, três casos de países com dívidas elevadas e crescimento saudável – Austrália, Nova Zelândia e Canadá.
- 2) Cálculos feitos de uma forma considerada pouco convencional que baixou a taxa de crescimento média em períodos de dívida elevada.
- 3) Foi feito um erro na folha de cálculo do Excel que excluiu cinco países da análise.

<http://blogs.lse.ac.uk/impactofsocialsciences/2013/04/24/reinhart-rogoff-revisited-why-we-need-open-data-in-economics>

Dados ligados às publicações...



Universidade do Minho
Serviços de Documentação



The American Economic Review: Data Availability Policy

It is the policy of the *American Economic Review* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Authors of accepted papers that contain empirical work, simulations, or experimental work must provide to the *Review*, prior to publication, the data, programs, and other details of the computations sufficient to permit replication. These will be posted on the *AER* Web site. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.

As soon as possible after acceptance, authors are expected to send their data, programs, and sufficient details to permit replication, in electronic form, to the *AER* office. Please send the files via e-mail to aeraccept@aeapubs.org, indicating the manuscript number. Questions regarding any aspect of this policy should be forwarded to the Editor.

Our policies differ somewhat for econometric and simulation papers, and for experimental papers.

Aparentemente a *American Economic Review*, aparentemente, não terá conseguido impor a sua própria política de disponibilidade de dados...

Caminho para uma maior disponibilidade dos dados...

Não tem só a ver com a verificação desses dados:

Tal como o Relatório da Royal Society enfatiza:

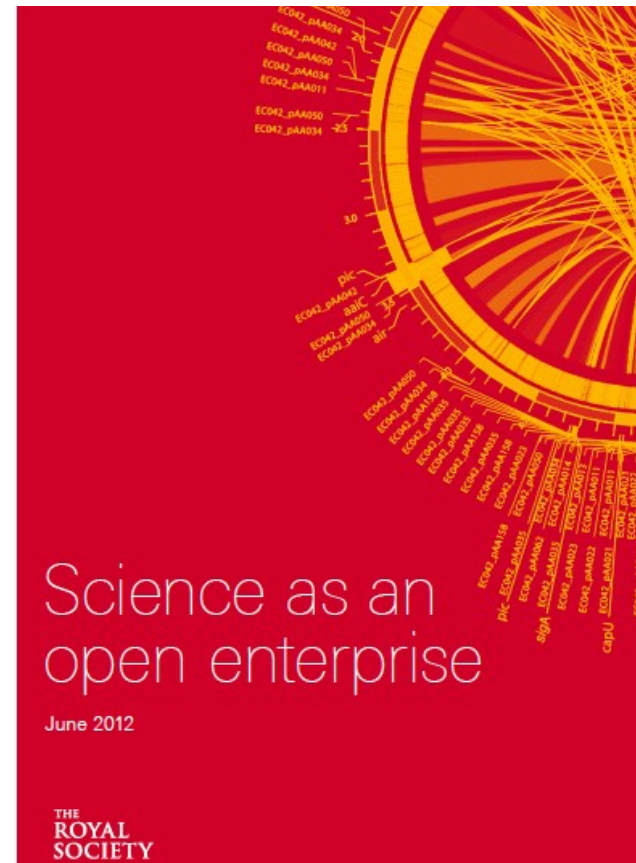
- Potencial de **reutilização** dos dados.

- Necessidade de

<http://royalsociety.org/policy/projects/science-public-enterprise/report>

partilha rápida dos

dados.



Relevância do tema



Universidade do Minho
Serviços de Documentação

Unlocking the full value of scientific data

Digital Agenda
1 0 0 1 1 0 0 1 0 1 0 1 1 1 0 1 1 1 0 0 0 0 1 0 0
for Europe

"European Digital"
Neelie Kroes

"Information and Communication Technologies (ICT) are the most recent transformational factors in science."

The Digital Agenda for Europe outlines policies and actions to maximise the benefit of the digital revolution for all. Supporting research and innovation is a key priority of the Agenda, essential if we want to establish a flourishing digital economy by 2020.

Scientific research is supported by its infrastructures: technical tools and instruments and socio-economic systems for organising and sharing knowledge. These have been in constant change for many centuries reflecting advances in technology and change in political systems. Key inventions like the microscope or the telescope resulted in huge

scientific progress by allowing the validation or rejection of theories; and the invention of book printing in the 15th century and the organisation of knowledge in research libraries allowed unprecedented access to knowledge.

Information and Communication Technologies (ICT) are the most recent transformational factors in science. They enable close and almost instantaneous collaboration between scientists all over the world and they provide access to unprecedented volumes of scientific information that can in turn be processed on powerful computational platforms. Many younger scientific disciplines would not even

RIDING THE WAVE How Europe can gain from the rising tide of scientific data

“To make progress in science, we need to be open and share... sharing data, and having the forum to openly use and build on what is shared, are essential to science. They fuel the progress and

Relevância do tema



Universidade do Minho
Serviços de Documentação



http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707

Relevância do tema



Universidade do Minho
Serviços de Documentação



»Modern science needs the free flow of knowledge... in an e-infrastructure that is open across national borders, disciplines and scientific communities«
Nelle Kroes (European Commission, 2012)

Supporting Open Science in Europe



Who benefits from OpenAIRE?

- EU researchers who access, deposit and link to research output
- National Open Access initiatives
- Repository managers
- Policy makers and funders who monitor funded work
- Publishers who wish to raise visibility of output
- Potential data providers who want to explore linking up their research

What is OpenAIRE?

- A Participatory European Open Access infrastructure to manage scientific publications and associated information via repository networks
- Harvests and indexes FP7 Open Access publications
- Harvests subsets of related data, and other contextual information, cross-linking them to demonstrate Enhanced Publications
- The OpenAIRE portal provides a suite of services
 - deposit and access
 - guidelines and a helpdesk
- OpenAIRE runs a series of workshops, and produces

Why is OpenAIRE important?

- By facilitating Open Science and Open Access, OpenAIRE allows scientists to access, reuse and enhance and research output
- OpenAIRE provides a cross-discipline support service for European Scientists
- Tools such as publication usage statistics
- OpenAIRE is based on
 - versatile technology and innovative research
 - European outreach effort which advocates Open Access

Who is OpenAIRE?

- OpenAIRE is an FP7 funded project, now in its second phase of funding until May 2014
- 41 project partners include 3 scientific communities: EB, DANS and BADC
- Collaboration with EuroCRIS, EUDAT, DataCite, COAR, LIBER, SPARC Europe
- Contact
Project Coordinator:
Mike Hatzopoulos, mike@dluoarg
- www.openaire.eu

Relevância do tema



Universidade do Minho
Serviços de Documentação



Repositórios Científicos ligam ao OpenAIRE

PUBLICAÇÕES

DADOS CIENTÍFICOS

INFORMAÇÃO DE FINANCIAMENTO



Relevância do tema



Universidade do Minho
Serviços de Documentação

Multiplicam-se as **iniciativas** de **Open Access**, **Open Data**.

– *Panton Principles for Open Data in Science*: <http://pantonprinciples.org>



Open Knowledge
Foundation



“Entendemos por dados abertos, em ciência, aqueles que estão livremente disponíveis na Internet pública permitindo a qualquer utilizador baixar, copiar, analisar, reprocessar, usá-los com um software ou utilizá-los para qualquer outra finalidade sem barreiras financeiras, legais, técnicas ou outras além do acesso à Internet.

Para este fim, os dados relacionados com ciência publicada devem ser explicitamente colocados em domínio público.”

Algumas leituras e recursos de referência



Universidade do Minho
Serviços de Documentação

- OCDE
 - OECD. *Declaration on Access to Research Data From Public Funding*, Paris, 2004.
http://www.oecd.org/document/15/0,3343,en_2649_34487_25998799_1_1_1_1,00.html
 - *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris, 2007. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Várias iniciativas dos Conselhos de Investigação do Reino Unido
 - RCUK. (2011). Common Principles on Data Policy - <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>
 - Jones, Sarah. (2011). Summary of UK research funders' expectations for the content of data management and sharing plans <http://www.dcc.ac.uk/>
- Iniciativas no Contexto da União Europeia produzidas pela Comissão Europeia no âmbito da Agenda Digital Europeia e ERA
 - CE (2012): Communication on a reinforced European Research Area partnership for excellence and growth
 - Communication Towards better access to scientific information
 - Recommendation on access to and preservation of scientific information
- EUA com iniciativas da National Science Foundation - <http://www.nsf.gov>

A importância da gestão de dados científicos



Universidade do Minho
Serviços de Documentação



<https://www.youtube.com/watch?v=N2zK3sAtr-4>



3

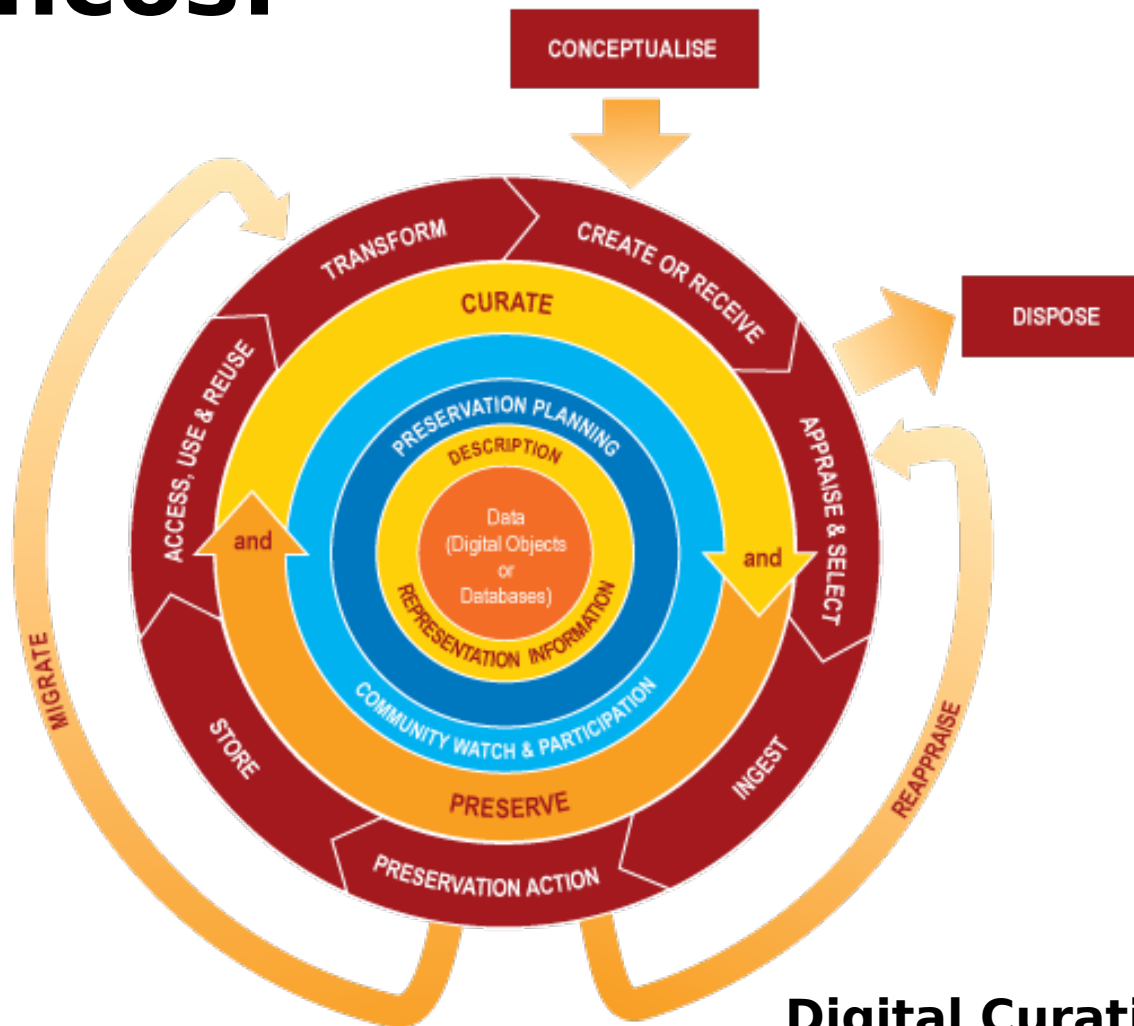
CICLO DE VIDA DOS DADOS CIENTÍFICOS E ETAPAS NA GESTÃO



O que está envolvido na gestão de dados científicos:



Universidade do Minho
Serviços de Documentação



Digital Curation Centre

O que está envolvido na gestão de dados científicos:



Universidade do Minho
Serviços de Documentação

- Planeamento e gestão de dados
- Criação de dados
- Documentar os dados
- Acesso e utilização dos dados
- Armazenamento e backups
- Partilha dos dados
- Preservação dos dados

Curadoria



Universidade do Minho
Serviços de Documentação



- Curadoria de dados é a atividade de gestão e utilização de dados desde o momento da sua criação para garantir o seu armazenamento, a sua pesquisa e respetiva reutilização.

“the active management and appraisal of data over the lifecycle of scholarly and scientific interest”

Digital Curation Centre

O ciclo de vida dos dados e a sua gestão:

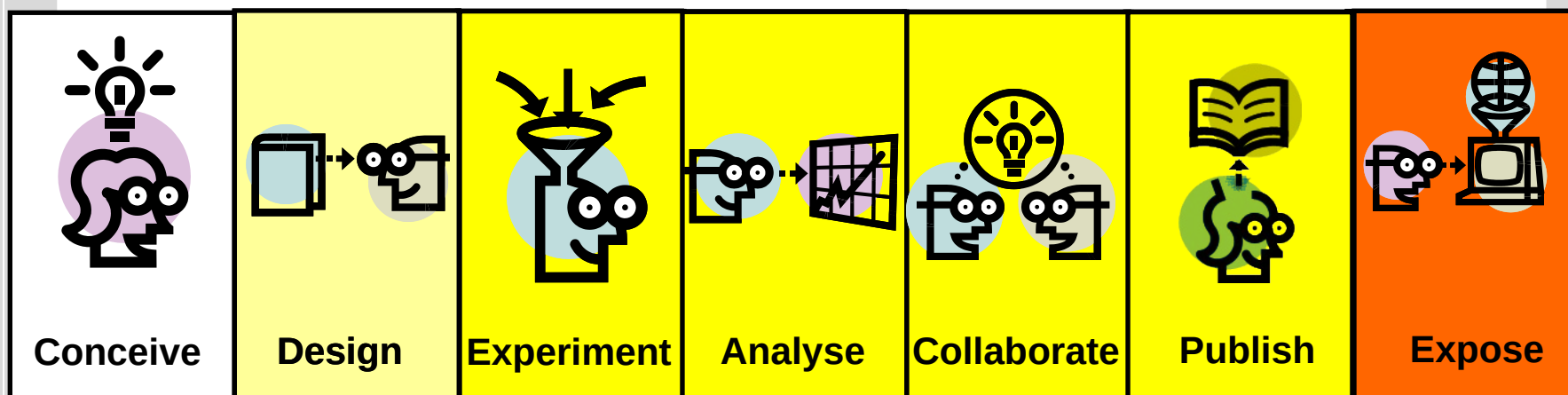


Universidade do Minho
Serviços de Documentação

Gestão e
Planeamento
de dados

Plataforma de
Gestão de dados científicos

Repositórios disciplinares
Ou institucionais e
Revistas científicas



+ info sobre ciclo dos dados



Universidade do Minho
Serviços de Documentação

<http://www.data-archive.ac.uk/create-manage/life-cycle>

Algumas leituras e recursos de referência



Universidade do Minho
Serviços de Documentação

- Knowledge-Exchange (2011), ["A Surfboard for Riding the Wave Towards a four country action programme on research data"](#)
- Ball, A and Duke, M (2011) How to Cite Datasets and Link to Publications. Digital Curation Centre http://www.dcc.ac.uk/webfm_send/525
- Green, T (2009), "We Need Publishing Standards for Datasets and Data Tables", OECD Publishing White Paper, OECD Publishing. doi: 10.1787/603233448430 <http://dx.doi.org/10.1787/603233448430>
- MaRDI-Gross project *DMP Planning for Big Science Projects* (2012). <http://arxiv.org/abs/1208.3754v1>
- Digital Curation Centre - <http://www.dcc.ac.uk>
- International Journal of Digital Curation - <http://www.ijdc.net/>
- Data Science Journal - <http://www.jstage.jst.go.jp/browse/dsj>
- Open Knowledge Foundation: <http://okfn.org/>
- OpenAIRE: <http://www.openaire.eu>



Outros recursos...



DCC: Managing Research Data.mov



TheDigitalCuration · 55 videos

✓ Subscrito ⚙

2 422

👍 12 👎 0

Digital Curation Center

<https://www.youtube.com/watch?v=2JBQS0qKOBU>



4

O PAPEL DAS BIBLIOTECAS: RECOMENDAÇÕES DA LIBER



Recomendações para as Bibliotecas: gestão de dados científicos



Universidade do Minho
Serviços de Documentação

- LIBER




Ten recommendations
for libraries to get
started with research
data management

Final report of the LIBER working group on E-Science / Research Data Management

Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

Recomendações para as
Bibliotecas:

gestão de dados científicos



Universidade do Minho
Serviços de Documentação

- LIBER

1. Apoio na gestão dos dados científicos

*Offer research data management support, including data management **plans for grant applications**, intellectual property **rights** advice and **information** materials. Assist faculty with data management plans and the integration of data management into the curriculum.*



Serviços da Data Library da Universidade de Edimburgo



THE UNIVERSITY of EDINBURGH

Schools & departments

Search



Contact

Home University home > Schools & departments > Information Services > Research Support > Research data support > **Data Library and consultancy**

Students

Researchers & Teachers

Support staff

Library

Computing

Learning Technology

Research Support

Help & Consultancy

About IS

Service info

Related links

Help

Research data support

Data Library and consultancy

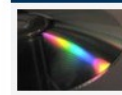
Overview of Data Library and consultancy service.



Finding data

Browse dataset holdings and find free data on the internet.

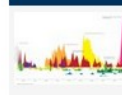
> Finding data



Accessing data

How to access data from national data services and the Data Library collection.

> Accessing data



Using data

User guides and online learning resources for using research data.

> Using data



Contact us

Get help by talking to a data librarian or sending an email query.

> Contact us

<http://>

www.ed.ac.uk/schools-departments/information-services/research-support/data-library/consultancy



Serviços, recursos e guias para a gestão de dados científicos na Universidade de Monash

The screenshot displays the Monash University Library website. At the top, there is a navigation bar with links for 'Skip to content', 'Monash home', 'Study', 'Research', 'Find people and places', 'Current students', 'Library', and 'my.monash'. A search bar is also present. Below the navigation bar, the 'MONASH University Library' logo is visible. The main content area features a menu with categories: 'Library home', 'About us', 'Libraries', 'Collections', 'Research and learning skills', 'Managing research data', 'Services and facilities', and 'Alumni and visitors'. Under 'Managing research data', there are sub-links: 'About research data management', 'Guidelines', 'Advice and planning', 'Skills development', and 'Resources and activities'. The main content area is divided into three sections: 1. 'Research data management' with an image of two people working on a laptop and a text box stating 'Manage research data effectively' and 'Produce higher quality data, increase your research impact, and protect your data from loss or misuse. Some of the benefits of data planning.' 2. 'Managing research data - a roadmap' with an image of a smartphone and a text box stating 'The Monash University Research Data Management Strategy and Strategic Plan 2012-2015 was publicly released in April 2012. It outlines future initiatives at Monash.' 3. A sidebar with a 'Strategy' section containing 'Guidelines', 'Advice and planning', and 'Skills development'. The main text in this section states: 'Monash University is leading efforts to improve the management of research data. Well-managed research data is more discoverable and available for re-use, and contributes to increased research impact, enhanced research practice (including collaboration) and improved education outcomes.'

[http://
www.monash.edu/library/researchdata](http://www.monash.edu/library/researchdata)

Importância do plano de gestão de dados científicos (GDC)



MANTRA - Ellie Bates - Importance of data management plan in a PhD work

<http://youtu.be/1stLmJUO81A>

Requisitos dos financiadores



Universidade do Minho
Serviços de Documentação

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	○
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	○
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

Papel da biblioteca - na prática...



Universidade do Minho
Serviços de Documentação



Preparing data management plans for NSF grant applications

[National Science Foundation](http://youtu.be/Lc82pxxRkMo)

<http://youtu.be/Lc82pxxRkMo>

Recomendações para as
Bibliotecas:

gestão de dados científicos

Universidade do Minho
Serviços de Documentação

- LIBER

2. Desenvolvimento de normas e serviços de metadados

Engage in the development of **metadata** and data **standards** and provide metadata **services** for research data.



DataCite

DataCite - International Data Citation

DataCite Metadata Schema for the Publication and Citation of Research Data

Version 3.0 July 2013

Documentation Revision January 2014

doi:10.5438/0008

http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.0.pdf

In this section

Briefing Papers

How-to Guides

Developing RDM Services

Curation Lifecycle Model

Curation Reference Manual

Policy and legal

Data Management Plans

Tools

Case studies

Repository audit and
assessment

Standards

Disciplinary Metadata

DIFFUSE

Publications and presentations

Roles

Curation journals

Informatics research

External resources

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

<http://www.dcc.ac.uk/resources/metadata-standards>



Documentation and metadata

Match the metadata

Drag the metadata field to its corresponding metadata category.

Descriptive	Administrative	Structural
location ✓	data processing ✓	variable list ✓
funder ✓	file formats ✓	database schema ✓
project start and end dates ✓	rights statement ✓	taxonomy ✓
subject ✓	Project identifier ✓	directory & file listing ✓
language ✓	methodology ✓	data dictionary ✓
data citation ✓	sources ✓	
title ✓	version ✓	
data creator ✓		

Check Answers



11 / 19



http://datalib.edina.ac.uk/xerte/play.php?template_id=4



- Um dos primeiros requisitos é que os conjuntos de dados sejam **acompanhados de informação que descreva** a sua origem (tempo ou espaço, métodos e instrumentos de recolha), âmbito, autoria, propriedade e condições de reutilização, ou seja, de metadados.
- Em paralelo com a **interoperabilidade** tecnológica, a existência de metadados adequados e normalizados é um requisito essencial para o acesso e reutilização dos dados científicos.

Documentação e metadados

- Repositórios de dados científicos e a normalização dos metadados.
- Difícil uniformizar a descrição entre domínios científicos.
- A descrição dos conjuntos de dados ainda se pode considerar muito pouco desenvolvida.
- Ao nível do conjunto têm sido adoptados modelos genéricos como o do Dublin Core, já muito utilizado nos repositórios institucionais.
- Projecto DataShare (Edimburgo, Oxford e Southampton)

Necessidade dos metadados



Universidade do Minho
Serviços de Documentação



Os metadados
podem não parecer
necessários quando
se trabalha sozinho
no seu próprio
computador, mas
são cruciais
quando são
partilhados online.

Dados
devidamente
documentados
permitem que
outros interpretem

Recomendações para as Bibliotecas: gestão de dados científicos



Universidade do Minho
Serviços de Documentação

- LIBER

3. Desenvolvimento de competências

Create **Data Librarian** posts and develop professional **staff skills** for data librarianship.

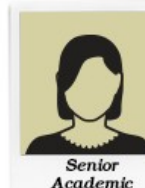
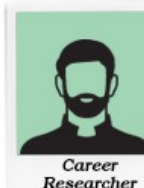
Mantra



Universidade do Minho
Serviços de Documentação



MANTRA is an online course designed for researchers or others planning to manage digital data as part of the research process.



HOME

ABOUT

ACKNOWLEDGEMENTS

DIY TRAINING KIT FOR LIBRARIANS

FEEDBACK

LEARNING UNITS



Research data explained



File formats & transformation



Data protection, rights & access



Data management plans



Documentation & metadata



Sharing, preservation & licensing



Organising data



Storage & security



Software practicals



<http://data.lib.edina.ac.uk/mantra/>

Recomendações para as
Bibliotecas:

gestão de dados científicos

Universidade do Minho
Serviços de Documentação

- LIBER

4. Políticas institucionais de gestão dos dados

Actively participate in institutional research **data policy development**, including resource plans. Encourage and adopt **open data** policies where appropriate in the research **data life cycle**.

Políticas institucionais de GDC



Universidade do Minho
Serviços de Documentação



Research data management policies - Jeff Haywood

http://youtu.be/V8IldfBAr_0

Recomendações para as
Bibliotecas:

gestão de dados científicos

Universidade do Minho
Serviços de Documentação

**- LIBER 5. Trabalhar em
articulação com os
parceiros para fomentar
infraestruturas**

Liaise and partner with researchers, research groups, data archives and data centers to foster an interoperable infrastructure for data access, discovery and data sharing.



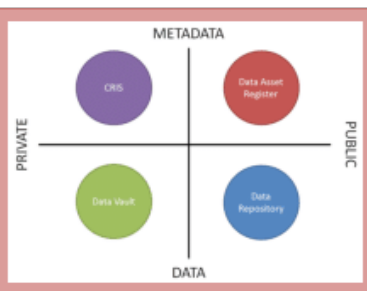
Data Curation Systems Development: Evolving thoughts

The University of Edinburgh

“The Four Quadrants of Research Data Curation Systems”

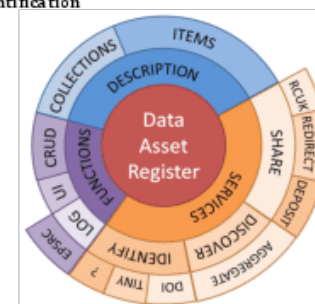
Classifying the different curation systems involved in recording, describing, storing, and sharing research data.

Issues: Interoperability, advocacy, requirements, similarities, differences, preservation, authority, trust.

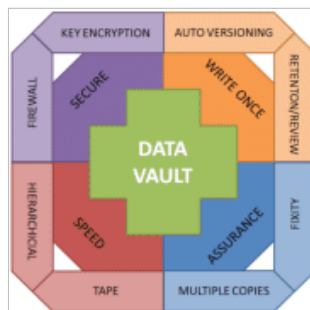


Data Asset Register

- Describe data items and collections (held internally or externally)
- Function: Create / Read / Update / Delete, Log actions, User Interface
- Services: Value added services to aid sharing, discovery, and identification

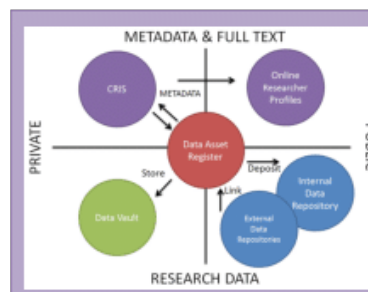


<http://dx.doi.org/10.6084/m9.figshare.902835>



Data Vault

- Provide secure storage
- Long term assurance and preservation requirements
- Consider speed issues: is high speed required (expensive)?
- Write once feature with automatic versioning to protect data



In practice...

How might this be delivered in practice?

Issues: Desire to minimize user interfaces, where is the master copy held, challenges of integration, incentivizing sharing, no duplication of effort by users, transferring data between active Data Store and Data Vault or DataShare (open repository).



Recomendações para as Bibliotecas: gestão de dados científicos

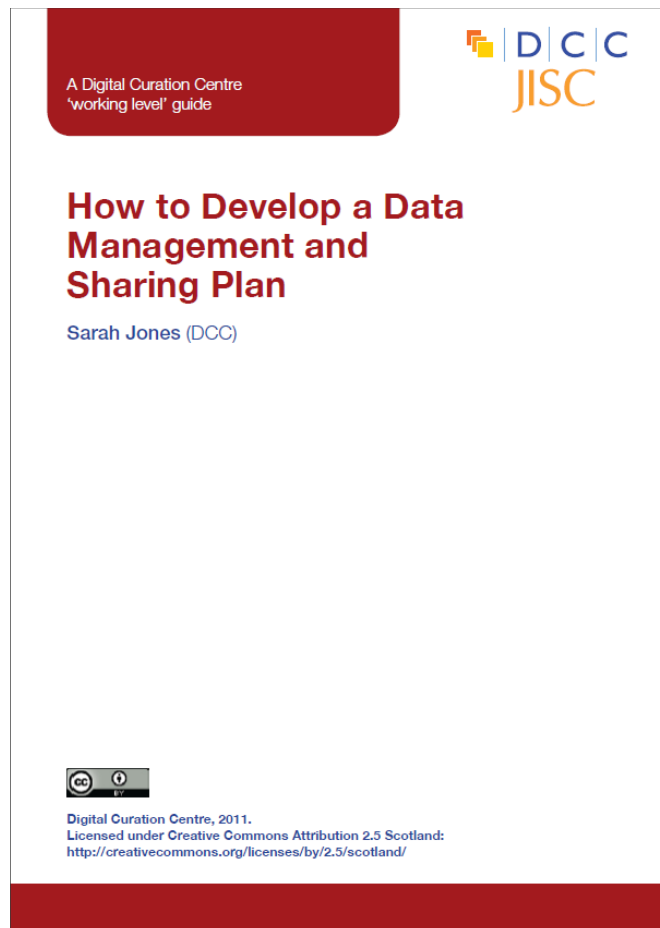


Universidade do Minho
Serviços de Documentação

- LIBER

6. Disponibilizar serviços de curadoria no ciclo de vida dos dados científicos

Support the lifecycle for research data by providing services for **storage**, **discovery** and **permanent access**.



DMP^{online}

The  | **D|C|C** Data Management Planning Tool

[https://
dmponline.dcc.ac.uk](https://dmponline.dcc.ac.uk)

www.dcc.ac.uk/resources/how-guides/develop-data-plan

Recomendações para as Bibliotecas: gestão de dados científicos



Universidade do Minho
Serviços de Documentação

- LIBER

7. Citação dos dados e links e identificadores permanentes

Promote research **data citation** by applying persistent identifiers to research data.



DataCite

Helping you to find,
access, and reuse data

DataCite Metadata Version 3.0 Available for Comment and Testing

Published by Sergio Ruiz on 3 June 2013 - 9:17pm

The DataCite Metadata Working Group is pleased to present a new version of the DataCite Schema for review, comment and testing. We will be giving greatest priority to feedback from DataCite clients and members, but we invite the scholarly and research community to contribute to this discussion as well.

With Version 3.0, we have worked to respond appropriately to a large number of feature requests while adhering to our goal of remaining domain agnostic. As well, we continue to strive for simplicity wherever possible.

Tags:

[datacite](#), [Metadata](#), [Schema](#), [identifiers](#), [DOI](#), [standards](#)

[Read more](#)

DataCite Summer Meeting - 19-20 September 2013 - Washington DC

Published by Sergio Ruiz on 31 May 2013 - 12:55pm

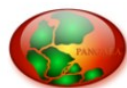
Why cite
data?

What is
DataCite?

What do
we do?

DOI resolver

Resolve a DOI string (e.g. 10.5284/1000164) by entering it below:



PANGAEA®
Data Publisher for Earth & Environmental Science

Not logged in (log in or sign up)

Always quote citation when using data!

[Show Map](#) [Google Earth](#)

Data Description

Citation: Ramirez-Llodra, Eva; Company Claret, Joan Baptista; Sardà, Francesc; De Mol, Ben (2012): Distribution of marine litter at bathyal and abyssal depths in the Mediterranean Sea. *Institut de Ciències del Mar*, Unpublished dataset #778322

Abstract: Marine litter was collected with an OTMS and Agassiz trawl from the Blanes margin and canyon (Catalan Sea, NW Mediterranean) between 900 and 1500 m depth (PROMETEO project) and from the Western, Central and Eastern Mediterranean basins at 1200, 2000 and 3000 m depth (BIOFUN project) in the framework of the HERMIONE programme. Litter was separated in 8 categories and weighed.

Related to: Ramirez-Llodra, Eva; Company Claret, Joan Baptista; Sardà, Francesc; De Mol, Ben: Distribution of marine litter at bathyal and abyssal depths in the Mediterranean Sea. *data submission PDI-1669, Progress in Oceanography*, submitted [↗](#)

Project(s): Hotspot Ecosystem Research and Mans Impact On European Seas (HERMIONE) [↗](#)

Event(s): BF1A01 [↗](#) * *Latitude Start:* 38.419500 * *Longitude Start:* 1.772667 * *Latitude End:* 38.408000 * *Longitude End:* 1.788667 * *Date/Time Start:* 2009-05-30T04:25:00 * *Date/Time End:* 2009-05-30T05:05:00 * *Elevation:* -1200.0 m * *Location:* Western Mediterranean [↗](#) * *Campaign:* BIOFUN_1 [↗](#) * *Basis:* Sarmiento de Gamboa [↗](#) * *Device:* Agassiz Trawl [↗](#)

BF1A02 [↗](#) * *Latitude Start:* 38.417500 * *Longitude Start:* 1.774333 * *Latitude End:* 38.393667 * *Longitude End:* 1.808000 * *Date/Time Start:* 2009-05-30T17:15:00 * *Date/Time End:* 2009-05-30T18:15:00 * *Elevation:* -1200.0 m * *Location:* Western Mediterranean [↗](#) * *Campaign:* BIOFUN_1 [↗](#) * *Basis:* Sarmiento de Gamboa [↗](#) * *Device:* Agassiz Trawl [↗](#)

BF1A03 [↗](#) * *Latitude Start:* 38.391167 * *Longitude Start:* 1.805333 * *Latitude End:* 38.414167 * *Longitude End:* 1.777000 * *Date/Time Start:* 2009-05-30T20:10:00 * *Date/Time End:* 2009-05-30T21:10:00 * *Elevation:* -1200.0 m * *Location:* Western Mediterranean [↗](#) * *Campaign:* BIOFUN_1 [↗](#) * *Basis:* Sarmiento de Gamboa [↗](#) * *Device:* Agassiz Trawl [↗](#)

Size: 0 datasets



[Download Data](#) (login required)

[Download ZIP](#)

Data Description

Citation:

Ramirez-Llodra, Eva; Company Claret, Joan Baptista; Sardà, Francesc; De Mol, Ben (2012): Distribution of marine litter at bathyal and abyssal depths in the Mediterranean Sea. *Institut de Ciències del Mar*, Unpublished dataset #778322

Recomendações para as Bibliotecas: gestão de dados científicos

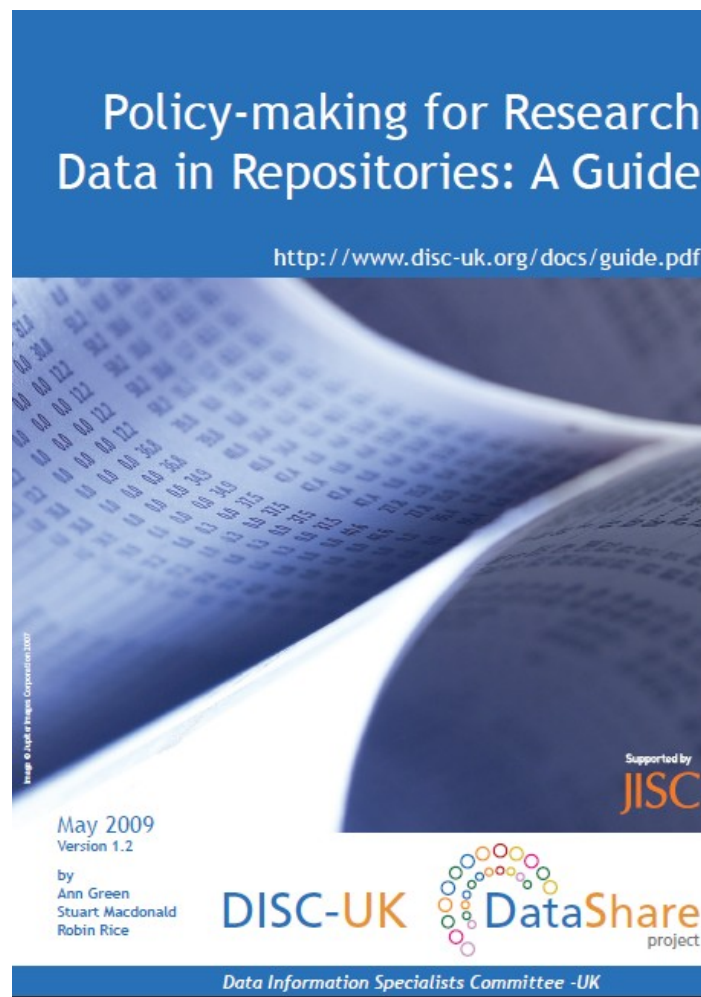


Universidade do Minho
Serviços de Documentação

- LIBER

8. Repositórios de dados

Provide an **institutional Data Catalogue** or **Data Repository**, depending on available **infrastructure**.



<http://www.disc-uk.org/docs/guide.pdf>

Projetos piloto em desenvolvimento



Universidade do Minho
Serviços de Documentação

Promove a **partilha** de **informação** entre a comunidade científica

06:18

NTE **Repositório de Dados Científicos da U.Porto**
from **Novas Tecnologias na Educação** 2 years ago NOT YET RATED

Vídeo informativo do projecto "Repositório de Dados Científicos da U.Porto".

<https://vimeo.com/33413787>

zenodo Research. Shared.

Search 617 records for:

Filter by types -

Recent Uploads

02 November 2013 **Journal article** **Open access**
Risk Perception Research Literature and Data Review
 Vrábliková, Katerina
 In this new discussion paper series, the Prague SECONOMICS team intends to allow the broader academic community taking part in an on-going discussion about risks and threats as well as trade-offs between them and security. [...]
 Uploaded by Belem Gallego on 23 May 2013.

03 March 2013 **Journal article** **Embargoed access**
Growth and galvanic replacement of silver nanocubes in organic media
 Polavarapu, Lakshminarayana ; Liz-Marzán, Luis M.
 Although metal nanoparticles with various shapes can be prepared in polar organic solvents, little has been advanced toward the shape-controlled synthesis in non-polar solvents. [...]
 Uploaded by Luis Liz-Marzan on 23 May 2013.

New to ZENODO?

- Research. Shared.** – all research outputs from across all fields of science are welcome!
- Citable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citable.
- Community Collections** – accept or reject uploads to your own community collections (e.g. workshops, EU projects or your complete own digital repository).
- Funding** – integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- Flexible licensing** – because not everything is under Creative Commons.
- Safe** – your research output is stored safely for the future in same cloud infrastructure as research data from CERN's Large Hadron Collider.
- DropBox integration** – upload files straight from your DropBox.

Read more about features and benefits.

www.zenodo.org

zenodo Research. Shared.

Home / Publications / Branch-specific plasticity enables self-organization of nonlinear computation in single neurons

27 July 2011 **Journal article** **Open access**

Branch-specific plasticity enables self-organization of nonlinear computation in single neurons

Legenstein, Robert; Maass, Wolfgang
 (show affiliations)

It has been conjectured that nonlinear processing in dendritic branches endows individual neurons with the capability to perform complex computational operations that are needed in order to solve for example the binding problem. However, it is not clear how single neurons could acquire such functionality in a self-organized manner, since most theoretical studies of synaptic plasticity and learning concentrate on neuron models without nonlinear dendritic properties. In the meantime, a complex picture of information processing with dendritic spikes and a variety of plasticity mechanisms in single neurons has emerged from experiments. In particular, new experimental data on dendritic branch strength potentiation in rat hippocampus have not yet been incorporated into such models. In this article, we investigate how experimentally observed plasticity mechanisms, such as depolarization-dependent STDP and branch-strength potentiation could be integrated to self-organize nonlinear neural computations with dendritic spikes. We provide a mathematical proof that in a simplified setup these plasticity mechanisms induce a competition between dendritic branches, a novel concept in the analysis of single neuron adaptivity. We show via computer simulations that such dendritic competition enables a single neuron to become member of several neuronal ensembles, and to acquire nonlinear computational capabilities, such as for example the capability to bind multiple input features. Hence our results suggest that nonlinear neural computation may self-organize in single neurons through the interaction of local synaptic and dendritic plasticity mechanisms.

1 / 16

Development/Functionality/Region

Branch-Specific Plasticity Enables Self-Organization of Nonlinear Computation in Single Neurons

Robert Legenstein and Wolfgang Maass
 Institute for Electrical Engineering, Graz University of Technology, 8010 Graz, Austria

It has been conjectured that nonlinear processing in dendritic branches endows individual neurons with the capability to perform complex computational operations that are needed in order to solve for example the binding problem. However, it is not clear how single neurons could acquire such functionality in a self-organized manner, since most theoretical studies of synaptic plasticity and learning concentrate on neuron models without nonlinear dendritic properties. In the meantime, a complex picture of information processing with

Files

Name	Date	Size	
LegensteinMaass_2011.pdf	07 Feb 2013	1.4 MB	<input type="button" value="Preview"/> <input type="button" value="Download"/>

Comments

Related content

Teemed by 2
 36 reads on Mendeley
 2 reads on CitULike

Publication date: 27 July 2011
DOI: 10.1523/JNEUROSCI.5684-10.2011
Report number(s): OpenAIRE-BRAIN_JNETS-2011-001
Published in: The Journal of Neuroscience: the official journal of the Society for Neuroscience: 30 (2011) no. 31, pp. 10878-10882
Grant(s): BRAIN-i-NETS - Novel Brain-Inspired Learning Paradigms for Large-Scale Neuronal Networks (243914)
Collection(s): Communities > European Commission Funded Research (OpenAIRE)
 Publications > Journal articles
 Open Access
Uploaded by: Robert Legenstein (on 07 February 2013)

New to ZENODO?

Read more about features and benefits.

Share

Cite as
 Legenstein, Robert et al (2011). Branch-specific plasticity enables self-organization of nonlinear computation in single neurons. The Journal of Neuroscience: the official journal of the Society for Neuroscience: 30 (2011) no. 31, pp. 10878-10882. 10.1523/JNEUROSCI.5684-10.2011
 Further citation formats: DOI Citation Formatter

Export
 BibTeX, DataCite, DC, EndNote, NLM, RefWorks, MARC, MARCXML



A destacar no Zenodo

- Permitir que todos possam partilhar e citar os dados.
- Recebe dados multidisciplinares de comunidade científica, pesquisador e instituição.
- Proporciona um interface para ligar dados à informação do projeto de financiamento.
- Qualquer dado ou coleção armazenada é passível de agregação via OAI-PMH.
- Fácil de usar.

Recomendações para as
Bibliotecas:

gestão de dados científicos


Universidade do Minho
Serviços de Documentação

- LIBER

**9. Envolve-se e pratique a
gestão de dados científicos**

Get involved in subject specific **data
management practice.**



 [Página principal](#) Percorrer por ▾ Ajudas ▾ português ▾

Percorrer por tipos de documento scientific data

Índice:

ou inserir as letras iniciais:

Ordenar por: ▾ Em ordem: ▾ Resultados/Página: ▾ Autores/Registo: ▾

Mostrar 1-18 de um total de 18 resultados.

Data	Título	Autor(es)	Tipo
2014	Análise dos resultados do inquérito sobre dados científicos produzidos na Universidade do Minho [Dados]	Rodrigues, Eloy ✳; Boavida, Clara ✳; Truta, Raquel ✳ et al.	other; scientific data
2013	A aplicação de um questionário sobre a qualidade do atendimento aos utilizadores das bibliotecas dos Serviços de Documentação da Universidade do Minho : dados de 2013	Universidade do Minho. Serviços de Documentação	other; scientific data
2014	A aplicação de um questionário sobre a qualidade do atendimento aos	Universidade do Minho. Serviços de	other;

<http://repositorium.sdum.uminho.pt/browse?type=type&order=ASC&rpp=20&value=scientific+data>

Dados Abertos das Instituição Gov.PT



Universidade do Minho
Serviços de Documentação



GOVERNO | CIDADÃOS | EMPRESAS |

PESQUISA |



English A- A+ Notícias Perguntas Frequentes Glossário LOGIN

dados.gov
BETA
A AGÊNCIA PARA A MODERNIZAÇÃO ADMINISTRATIVA

INÍCIO


CATÁLOGO de DADOS

CATÁLOGO de APLICAÇÕES

ESTATÍSTICAS

PROGRAMADORES

PARTICIPE



Nova série de dados disponibilizada. Contas Nacionais Trimestrais por Setor Institucional - 1º Trimestre 2012

1 2 3

Novos Dados

Novas Ideias

Moradas e coordenadas das Lojas do Cidadão
Agência para a Modernização Administrativa

Contratação Pública - Anúncios 2014
Instituto da Construção e do Imobiliário

Contratação Pública - Contratos 2014
Instituto da Construção e do Imobiliário

Ver catálogo de dados

TOP

Fornecedor
Agência para a Modernização Administrativa é o nosso fornecedor mais popular com 161626 downloads.

Ideia
Actualmente a ideia mais votada é "Estatística sobre a Economia do Mar" com 5 votos.

★ ESCOLHA DO EDITOR

Ideia
Estatística sobre a Economia do Mar

Dados

Aplicação
Receitas vs Despesas por Município

O dados.gov nas redes sociais



Categorias de dados

<http://dados.gov.pt>

Bibliotecas:

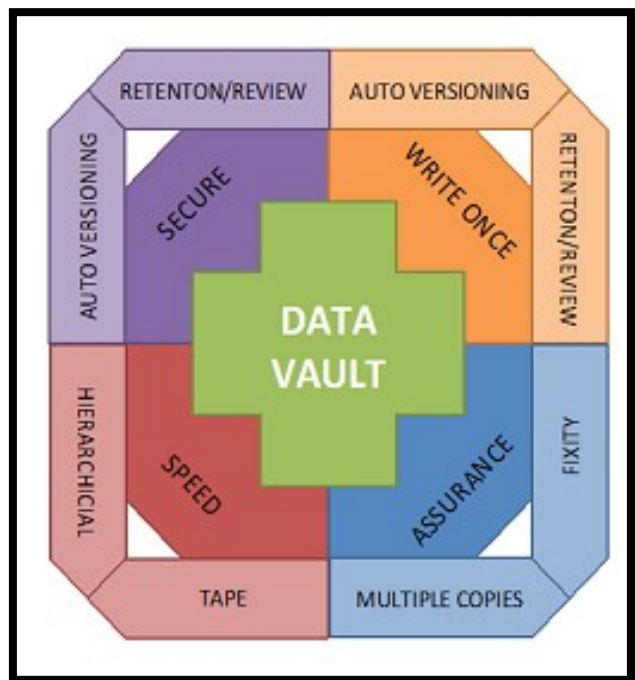
gestão de dados científicos

Universidade do Minho
Serviços de Documentação

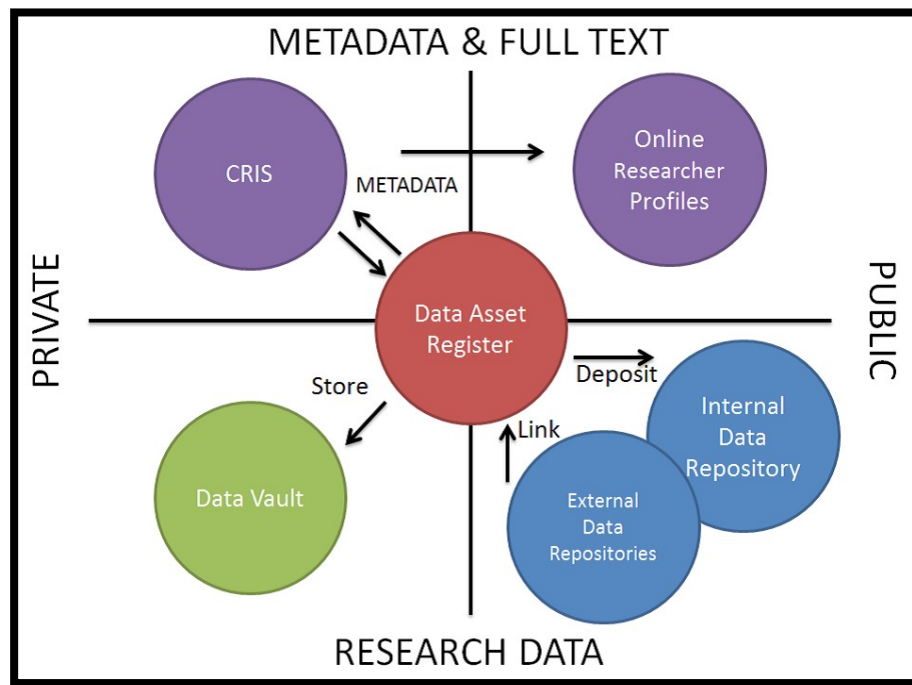
- LIBER

10. Oferecer ou mediar serviços de armazenamento

Offer or mediate secure **storage** for dynamic and static research data in co-operation with institutional IT units and/or seek exploitation of appropriate cloud services.



<http://datablog.is.ed.ac.uk/2013/12/20/thinking-about-a-data-vault>



<http://datablog.is.ed.ac.uk/2013/12/12/thinking-about-research-data-asset-registers>



5

OUTRAS SUGESTÕES



Sugestão de 7 regras para as Universidades (1/2)



Universidade do Minho
Serviços de Documentação

1. Compreender como a instituição lida com dados científicos.
 - <http://www.data-audit.eu>
2. Conceber um caso para Gestão de Dados Científicos e reunir apoios
 - <http://www.dcc.ac.uk/news/managing-research-data-video>
3. Definir o posicionamento da sua instituição sobre a Gestão dos Dados Científicos para instituir uma política e estratégia.
 - <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies>
 - <http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps>
4. Assegurar que os investigadores estão conscientes sobre os dados que existem disponíveis
 - <http://datalib.edina.ac.uk/mantra>

Sugestão de 7 regras para as Universidades (2/2)



Universidade do Minho
Serviços de Documentação

5. Providenciar serviços robustos de armazenamento de dados fáceis de utilizar.

- <http://research-data-toolkit.herts.ac.uk/2013/05/the-cost-of-a-bit-of-a-ddud>

6. Tornar mais fácil aos outros encontrarem e citarem dados de investigação.

- <http://www.datacite.org>
- <https://peerj.com/preprints/1v1>

7. Permanecer à frente dos seus pares.
“Seven Rules of successful research data management in universities”

(July 2013) <http://data.bris.ac.uk/files/2013/06/data-bris-benefits-report-V2.pdf>
<http://www.openaccess.nl/news/380-seven-rules-of-successful-research-data-management-in-universities>

Obrigado pela atenção!



Universidade do Minho
Serviços de Documentação

Thank
You

Mahalo

Kiitos

Tack

Grazie

Toda

Thanks

Obrigado

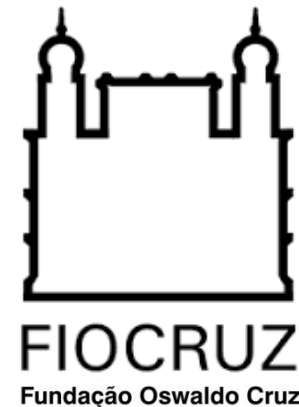
Takk

Gracias

Merci

Ricardo Saraiva
rsaraiva@s dum.uminho.pt

29 de abril de 2015 | Intercâmbio
RepositóriUM -



Perspetivas sobre a preservação digital

Ricardo Saraiva

rsaraiva@s dum.uminho.pt

Tópicos da apresentação

- Introdução/contextualização
- Preservação Digital
- Repositórios de Acesso Aberto e Preservação Digital
- Conclusões e Recomendações



Uma definição....

“...preservação digital o conjunto de processos responsáveis por garantir o acesso continuado à informação digital durante longos períodos de tempo, i.e. períodos de tempo superiores à esperança de vida do ambiente tecnológico necessário à interpretação e/ou reprodução dessa informação (Webb, 2003).”

Preservação Digital



Universidade do Minho
Serviços de Documentação

- **A preservação digital preocupa-se com a capacidade de manter a informação digital acessível, interpretável e autêntica**, mesmo na presença de uma plataforma tecnológica diferente daquela inicialmente utilizada no momento da sua criação.
- No mundo atual, onde cada vez mais organizações e pessoas dependem da informação digital que produzem, **torna-se premente a implementação de técnicas e de políticas concertadas que vão no sentido de garantir a continuidade e a acessibilidade a este tipo de informação** (Ex. registos de saúde, segurança social, etc.)
- Contudo, durante períodos de tempo alargados, é **comum haver evoluções tecnológicas ao nível do hardware, redes, arquiteturas de software, sistemas de gestão, esquemas e necessidades metadados, e até alterações à tutela responsável pela conservação destes registos.**

Exemplos obsolescência tecnológica



Universidade do Minho
Serviços de Documentação



Cassete vídeo
Betamax



Disquete de 3.5
polegadas



Cassete vídeo VHS



Leitor Disquete de 3.5
polegadas

Modelo de referência OAIS



Universidade do Minho
Serviços de Documentação

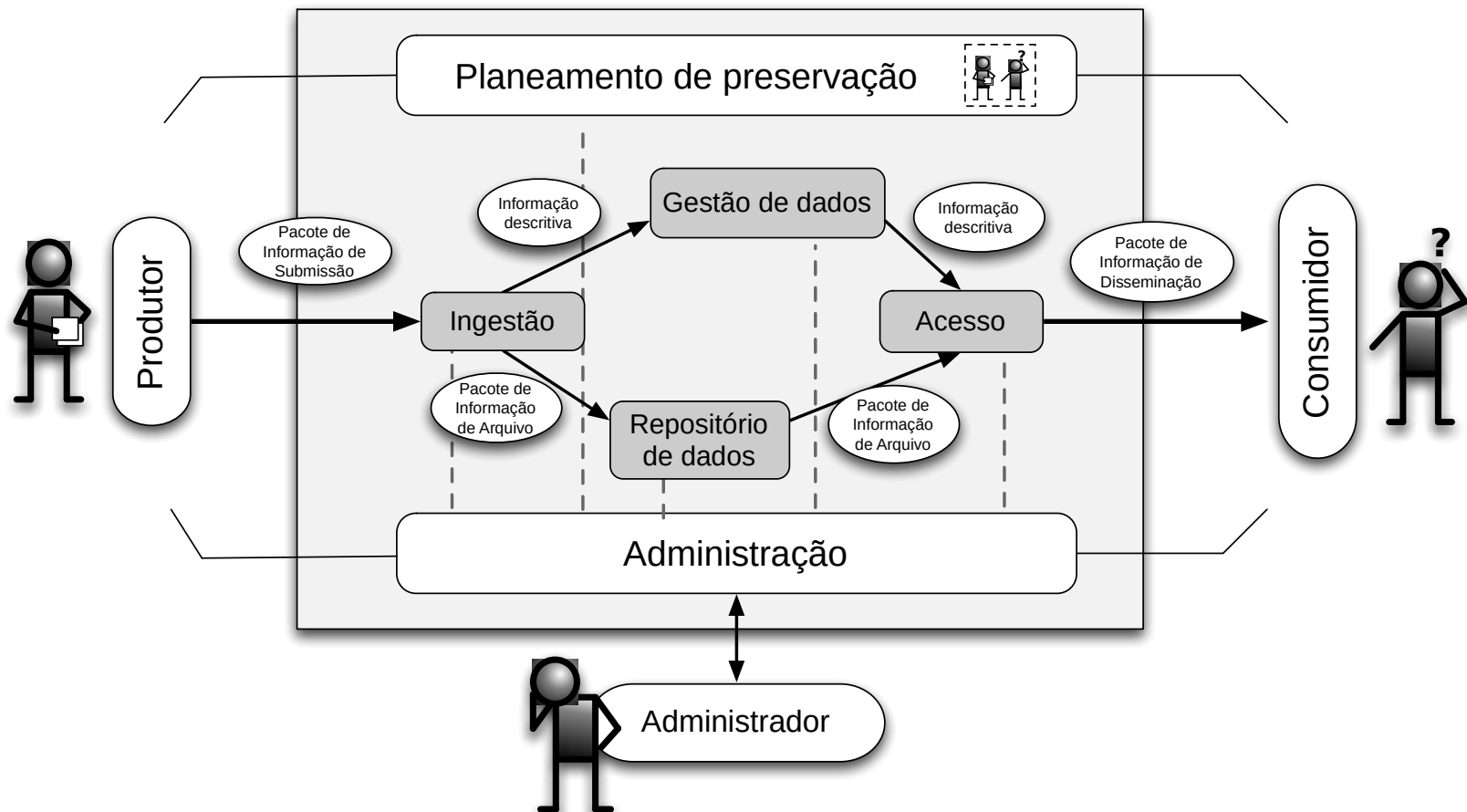
- Em 1990, o Consultative Committee for Space Data Systems (CCSDS) iniciou um esforço conjunto com a International Organization for Standardization (ISO) com o objetivo de desenvolver um conjunto de normas capazes de regular o armazenamento a longo-prazo de informação digital produzida no âmbito de missões espaciais.
- **Desta iniciativa nasceu o modelo de referência OAIS (Open Archival Information System), um modelo conceptual que visa identificar os componentes funcionais que deverão fazer parte de um sistema de informação dedicado à preservação digital, bem como as suas interfaces internas e externas e os objetos de informação trocados no seu interior.**

Modelo de referência OAIS



Universidade do Minho
Serviços de Documentação

Componentes funcionais identificados e os pacotes de informação trocados no interior do sistema.



Estratégias de preservação



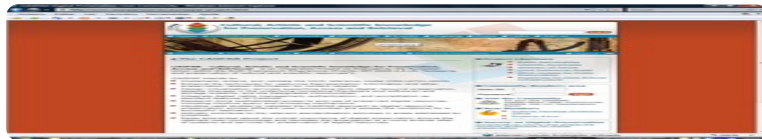
Universidade do Minho
Serviços de Documentação

- **Refrescamento** - consiste na transferência de informação de um suporte físico de armazenamento, para outro geralmente mais atual, antes que o primeiro se deteriore ou se torne irremediavelmente obsoleto (ex. disquete, disco rígido, CD-ROM, etc.);
- **Emulação** - baseia-se na utilização de software especial, designado vulgarmente por Emulador, que é capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software, inicialmente incompatível (ex. Videojogos).
- **Migração/conversão** - transferência periódica de material digital de uma dada configuração de hardware/software para uma outra, ou de uma geração de tecnologia para outra subsequente. (ex. conversão de formatos)
- **Encapsulamento** - consiste em conservar, juntamente com a informação digital, toda a metainformação necessária e suficiente que permita futuramente desenvolver conversores, visualizadores ou emuladores (ex. Arquivos da Web).

Projetos e arquiteturas de preservação para repositórios



Universidade do Minho
Serviços de Documentação



RepoMMan Project



The REMAP Project



Projeto CASPAR



Universidade do Minho
Serviços de Documentação

- O projeto CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) foi um projeto integrado, financiado pela UE, que abordou uma série de questões no domínio da preservação e de curadoria em diversas disciplinas. O CASPAR decorreu entre Abril de 2006 e Setembro de 2009, e nessa fase desenvolveu uma arquitetura e um fluxo de trabalho para preservação, conjuntamente com ferramentas integradas na própria arquitetura.



Projeto CRiB

- A CRiB (Conversion and Recommendation of Digital Object Formats) é uma arquitetura de serviço orientado (SOA), desenvolvida na Universidade do Minho, concebida para ajudar instituições com património cultural a implementar soluções de preservação baseadas em migração de formatos.
- O sistema possui serviços para identificar o formato de um objeto digital, monitorizar formatos quase obsoletos e compará-los com as soluções de migração existentes e em seguida produzir recomendações e estratégias de migração adequadas.
- Com base na arquitetura CRiB, a Direção Geral de Arquivos (DGARQ) de Portugal, já desenvolveu e criou o *Repository of Authentic Digital Objects* (RODA), um repositório multimédia, assente na plataforma Fedora Commons, com o qual a DGARQ espera incorporar documentos eletrónicos de instituições nacionais.

Projeto PANIC



Universidade do Minho
Serviços de Documentação

- O projeto PANIC (Preservation web services Architecture for New media, Interactive Collections and scientific data) foi conduzido pelo *Distributed Systems Technology Centre* (DTSC) e pela Universidade de Queensland, entre 2003 e 2006. O objetivo do projeto consistia na elaboração de uma arquitetura que pudesse ser adicionada a repositórios e que lhes permitisse suportar a preservação de objetos digitais.



Projeto Planets

- O Planets (Preservation and Long-term Access through NETworked Services) foi um projeto previsto para quatro anos e iniciado em Junho de 2006. Tratou-se de um projeto financiado pela União Europeia (EU) com vista a construir serviços práticos de curadoria e ferramentas para bens culturais e científicos.
- A arquitetura de preservação desenvolvida pelo projeto é conhecida como o *Planets Interoperability Framework*. Esta arquitetura foi desenhada e projetada para funcionar em paralelo com os *softwares* de repositórios existentes, ingerir material diretamente do repositório e tornar os resultados de preservação disponíveis para nova ingestão no repositório.



Projeto SCAPE

- O projeto SCAPE (SCALable Preservation Environments) é um projecto europeu em curso, que visa avançar o estado da arte no domínio da preservação digital. Reunindo especialistas de bibliotecas, laboratórios de investigação, universidades, empresas, entre outros. O objetivo é investigar e desenvolver sistemas de preservação digital escaláveis capazes de lidar com milhões de objetos em tempo útil.
- Nesta iniciativa, existe também uma participação do sector empresarial em atividades de investigação & desenvolvimento, estando representado por três empresas: a KEEP SOLUTIONS (Portugal), a Ex Libris Group e a Microsoft Research.

Exemplos de ferramentas



Universidade do Minho
Serviços de Documentação

Sistemas que visam suportar a preservação possuem necessidades de metadados muito específicas e que resultam de diversos projetos e iniciativas. Para além do OAIS, foram surgindo outras iniciativas que procuram fornecer uma orientação mais abrangente sobre os metadados a recolher....

- PREMIS : <http://www.loc.gov/standards/premis/>
- CAIRO: <http://cairo.paradigm.ac.uk/>
- AONS II: <http://www.apsr.edu.au/aons2/>



Conclusões

- A preservação digital, no âmbito dos repositórios de acesso aberto, **tem evoluído significativamente, como se comprova pelas múltiplas atividades, iniciativas e projetos**, que se têm vindo a conhecer nos últimos anos.
- Apesar da evolução, **o número de repositórios com políticas, estratégias e ações consolidadas ainda é residual.**
- Ainda **subsistem dúvidas, incertezas e lacunas quanto aos papéis e responsabilidades**, isto é, quem deverá ser responsável pela preservação e curadoria, a qualidade e interoperabilidade dos repositórios e sobre qual o enquadramento jurídico apropriado em termos de preservação digital.
- **Em Portugal, dos 45 repositórios** atualmente registados no portal RCAAP, **nenhum possuirá uma política de preservação formal.**



Recomendações

- Constituir, no âmbito da comunidade RCAAP, um grupo de interesse no domínio da preservação e curadoria digital, com iniciativas, atividades e canais de comunicação próprios, e no quadro do qual se poderiam concretizar a generalidade das recomendações e linhas de orientação apresentadas em seguida;
- Realizar um recenseamento e caracterização da situação existente nos repositórios de acesso aberto portugueses, no que diz respeito às políticas, procedimentos e estratégias de preservação digital;
- Avaliar, definir e concretizar um projeto piloto, no domínio da preservação digital, com a participação de vários repositórios portugueses, com o recurso a arquitetura(s) que possa(m) dotar os repositórios participantes de ferramentas abrangentes em termos de preservação digital;



Recomendações

- Acompanhar e, se possível e adequado, cooperar com iniciativas, serviços e projetos relevantes em curso na área da preservação digital, com o intuito de conhecer, utilizar e promover as boas práticas neste domínio. Exemplo possível cooperação entre o projeto RCAAP e/ou os repositórios individualmente com o RODA – Repositório de Objetos Digitais Autênticos;
- Desenvolver e/ou disseminar documentos de divulgação, formação e suporte, como Briefing papers, modelos de políticas e procedimentos, boas práticas e casos exemplares de preservação digital;
- Realizar ações de sensibilização, divulgação e formação destinadas a gestores de repositórios e responsáveis institucionais;



Recomendações

- Avaliar e identificar os recursos e os custos envolvidos nos processos de preservação digital, em especial os relativos aos recursos humanos, quer quanto ao esforço (tempo de trabalho) envolvido, quer quanto às competências requeridas, e eventuais necessidades de formação;
- Identificar eventuais conditionalismos éticos e legais, no que concerne a ações de preservação que possam incidir em conteúdos já depositados;
- Sensibilizar, incentivar e apoiar os autores e os gestores dos repositórios para a utilização de formatos de ficheiro que facilitem a preservação a longo prazo, aquando do depósito de documentos no(s) repositório(s).

Perspetivas e trabalho futuro



Universidade do Minho
Serviços de Documentação

- Desenvolvimento de um serviço de conversão tendo por base ferramentas open-source de conversão (com capacidade de expansão via plugins);
- Integração com processo de depósito do DSpace e implementação nos repositórios da rede RCAAP;
- Desenvolvimento de ferramenta de processamento de ficheiros já armazenados com integração com o DSpace e apresentação de relatórios (integrado com validador). (Incluir validação dos tipos de documento para preservação digital usando o formato **Flexible**

Ao nível do validador...



Universidade do Minho
Serviços de Documentação

Validador RCAAP

Funcionalidades

Sobre

Contacto

Ajuda

English

Verificar acesso ao
texto integral?

- Não
 Sim (processo mais demorado!)

Perfis de Validação

DRIVER - Valida o conteúdo do repositório de acordo com directrizes DRIVER ([Mais info.](#))

OpenAIRE - Valida o conteúdo do repositório de acordo com directrizes OpenAIRE ([Mais info.](#))

FCT - Valida o conteúdo do repositório de acordo com directrizes FCT ([Mais info.](#))

Teses e Dissertações - Valida o conteúdo do repositório de acordo com depósito legal de teses e dissertações ([Mais info.](#))

Ficheiros e Formatos - Verifica se os ficheiros do repositório são válidos e os seus formatos estão no conjunto de formatos preservation friendly ([Mais info.](#))

Acesso Embargado - Lista todos os documentos em acesso embargado ([Mais info.](#))

✓ Verifica se os ficheiros são válidos (usando o FITS) e os seus formatos (mimetype) estão no conjunto de formatos preservation friendly

Editar passo 1

Concluir



FEVEREIRO DE 12

ESTADO DA ARTE EM PRESERVAÇÃO DIGITAL



Disponível no RepositóriUM em:
<http://hdl.handle.net/1822/17049>



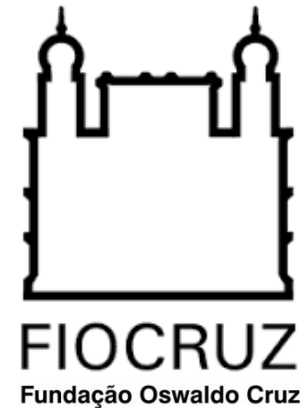
Muito Obrigado pela atenção!



Universidade do Minho
Serviços de Documentação



30 de abril de 2015 | Intercâmbio
RepositóriUM -



RepositóriUM

Sessão de perguntas e respostas

Ricardo Saraiva

rsaraiva@s dum.uminho.pt