



Estratégia Institucional para a gestão de dados de investigação na UMINHO: o papel dos SDUM

Pedro Príncipe^a, Antónia Correia^b, Paula Moura^c, Eloy Rodrigues^d

^a*Universidade do Minho, Portugal, pedroprincipe@sdum.uminho.pt*

^b*Universidade do Minho, Portugal, antoniacorreia@sdum.uminho.pt*

^c*Universidade do Minho, Portugal, paulamoura@sdum.uminho.pt*

^d*Universidade do Minho, Portugal, eloy@sdum.uminho.pt*

Resumo

A Universidade do Minho tem em curso o desenvolvimento de uma estratégia institucional para a gestão de dados de investigação. Os Serviços de Documentação (SDUM) participam ativamente na definição e realização das ações delineadas na estratégia para a Universidade, assumindo-se como serviço de suporte ao desenvolvimento de infraestruturas e sistemas para servir a gestão de dados ao longo do ciclo de vida da investigação, e para as atividades de capacitação dos gestores de dados e dos investigadores em matéria de gestão e partilha de dados de investigação.

Esta apresentação resume a atividade que os SDUM têm vindo a desenvolver no apoio à gestão de dados na UMinho e na operacionalização de serviços e sistemas para a infraestrutura de dados institucional, nomeadamente: i) no apoio à definição da estratégia da UMinho para a gestão de dados, ii) no trabalho de diagnóstico dos dados produzidos nas unidades de investigação, processo de gestão e identificação das necessidades de suporte à gestão de dados, iii) no desenvolvimento de ações de apoio informativo e programas integrados de formação, iv) na planificação e operacionalização de sistemas e criação de ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados.

Palavras-chave: Gestão de Dados, Dados de Investigação, Universidade do Minho, Repositório de dados.

Conteúdo da apresentação

Estratégia Institucional da Universidade do Minho para a gestão de dados de investigação

A Universidade do Minho tem em curso a definição da estratégia institucional para a gestão de dados de investigação que inclui a formulação de uma política institucional, a planificação de infraestrutura e sistemas para dados, a produção de serviços e ferramentas de apoio à gestão dos dados, o desenvolvimento de programas de capacitação e o reforço de serviços de apoio e consultoria no domínio de questões legais e de proteção de dados.

Os Serviços de Documentação assumem-se como serviço institucional de suporte à realização da estratégia em desenvolvimento, particularmente:

- no apoio à definição de uma política institucional da Universidade do Minho relativa à gestão e partilha dos dados de investigação, face à experiência que os SDUM detêm no domínio das políticas de acesso aberto;

- na criação e disponibilização serviços integrados e ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados, nomeadamente repositório de dados abertos e planos de gestão de dados;
- no desenvolvimento de ações e programas integrados que promovam e valorizem as competências dos investigadores em matéria de gestão e partilha de dados de investigação.

O plano de intervenção dos SDUM para os serviços de suporte à gestão de dados

Os SDUM desenvolveram em 2014 e 2015 um plano de intervenção para a gestão de dados que partiu da compreensão da necessidade de atuar de forma estruturada neste domínio emergente da atividade das bibliotecas e serviços de informação e documentação nas instituições de investigação e ensino superior.

O plano de intervenção tinha como primeiros objetivos, por um lado, o desenvolvimento de processos de diagnóstico para melhor conhecer a realidade da gestão dos dados produzidos na UMinho, e por outro, a sensibilização da gestão para a importância de definição de uma política institucional para a gestão de dados científicos. Estas duas premissas do plano de intervenção foram concretizadas através da realização de um estudo diagnóstico aos investigadores da UMinho em 2014 e de um programa de auditoria aos processos de curadoria de dados em 2017.

Diagnóstico à gestão de dados na Universidade do Minho

Os SDUM realizaram em 2014 um estudo¹, junto dos investigadores e centros de investigação, no sentido de se obter um melhor conhecimento sobre os dados gerados no âmbito da investigação realizada na UMinho e as práticas de gestão dos dados, avaliando o tipo de dados, identificar as práticas de gestão de dados de investigação e estimar a pertinência de serviços de suporte neste domínio. Este estudo baseou-se num inquérito por questionário dirigido a todos os investigadores e docentes e serviu de fundamentação à definição da estratégia institucional. Dos principais resultados obtidos concluiu-se que existia necessidade e abertura na comunidade para a implementação de serviços e ferramentas para apoio à gestão de dados, nomeadamente:

- 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, referenciação, acesso e reutilização de dados a longo prazo.

Mais recentemente e como forma de complementar o referido estudo, realizou-se um programa de diagnóstico e auditoria de dados de investigação junto de seis unidades de investigação. O programa de auditoria, baseado no projeto *Data Audit (Data Asset Framework - www.data-audit.eu)* desenvolvido

¹ Relatório completo do inquérito: <http://hdl.handle.net/1822/29948>; Dados do inquérito: <http://hdl.handle.net/1822/29949>.

pelo *Digital Curation Center* para as instituições de ensino superior no Reino Unido, foi realizado com base em dois questionários e uma entrevista de diagnóstico, tendo como público-alvo investigadores e curadores de dados e realizada em 3 etapas: 1) Caracterizar e analisar os processos de gestão associados aos dados de investigação produzidos e recolhidos pelas atividades da unidade de investigação (aplicação de questionário e entrevista); 2) Descrever de forma detalhada os dados de investigação produzidos e avaliar papéis e responsabilidade na gestão desses dados (aplicação de questionário), 3) Estabelecer orientações para a gestão dos dados de investigação e recomendações de melhoria (elaboração de relatório).

Serviços de suporte à gestão de dados

No âmbito do desenvolvimento de um programa integrado de serviços de suporte à gestão de dados de investigação, os SDUM prepararam recentemente dois importantes recursos para apoiar a elaboração de planos de gestão de dados e a capacitação na área da gestão e dados abertos.

Para o primeiro, relativo aos planos de gestão de dados, os SDUM prepararam com recurso à aplicação online – *DMPonline* <https://dmponline.dcc.ac.uk/> –, disponibilizada pelo *Digital Curation Center*, dois modelos para facilitar a elaboração de planos para projetos com financiamento pela Comissão Europeia (programa H2020) e pela FCT.

Adicionalmente, os SDUM prepararam um guia de apoio no sistema de guias dos SDUM (*LibGuides*) destinado a investigadores, particularmente estudantes de doutoramento, identificando os principais conteúdos de formação sobre: dados abertos, repositórios de dados, planos de gestão de dados, armazenamento e preservação de dados, licenças e reutilização de dados. Este guia é um importante recurso para desenvolver a atividade de apoio informativo à comunidade e como ponto de partida para a preparação de outros materiais a integrar no programa de formação a desenvolver.

Sistema de repositório de dados abertos

A disponibilização de um serviço de repositório de dados abertos integra-se na estratégia da UMinho de assegurar uma infraestrutura de sistemas e serviços à comunidade para a gestão de dados ao longo do ciclo de vida da investigação. Os SDUM estão a desenvolver um repositório de dados para registo e abertura dos dados, permitindo o depósito de dados abertos, embargados ou fechados, com atribuição de identificadores persistentes e preservação dos conjuntos de dados.

A conceção e criação do Repositório de Dados será desenvolvido com recurso a uma solução de software de código aberto, no quadro da construção de uma infraestrutura flexível e interoperável, tendo como modelo as melhores práticas em instituições de investigação no espaço europeu e mundial.

Os SDUM analisaram várias soluções de plataformas de software para repositório de dados, merecendo particular atenção os sistemas *DSpace*, *CKAN*, *Invenio* e *Dataverse*. A opção final recaiu pelo software *Dataverse*, uma plataforma desenvolvida e em utilização pela Universidade de Harvard, que se apresenta

como uma solução *open source* de aplicação web para partilha, preservação, citação, exploração e análise de dados de investigação. As 10 principais razões para a escolha da plataforma de software *Dataverse* foram:

- Versatilidade na organização de coleções (*dataverses*), necessária para abarcar de forma explícita todas as Unidades Operacionais e de Investigação da Universidade;
- Polivalência na adaptação dos esquemas de metadados disciplinares específicos para cada *dataverse* (coleção), de uma lista de 6 esquemas já configurados, sendo possível adicionar mais;
- Flexibilidade na definição de papéis e responsabilidades nos *workflows* de curadoria de dados e na administração das coleções, existindo oito diferentes papéis por defeito e sendo fácil criar novos papéis;
- Autonomia na operação dos processos de curadoria e na utilização do sistema, podendo o utilizador ser criador de coleções (*dataverses*), criar *datasets*, depositar ficheiros, fazer curadoria, gerindo o processo de publicação dos *datasets* e das próprias coleções;
- Escalabilidade e potencial de integração com outros sistemas, existindo serviços e ferramentas integradas para análise de dados, computação, *upload* de ficheiros e georreferenciação, entre outros;
- Multiplicidade de opções na disponibilização e acesso aos dados, desde dados abertos, dados restritos com metadados públicos, dados disponíveis após identificação dos utilizadores e assinatura de termos de uso (*guest book*), assegurando de forma ágil o princípio “*as open as possible, as closed as necessary*” difundido pela Comissão Europeia na sua política Open Science;
- Versatilidade nas opções de serviço de armazenamento de dados, sendo possível usar diferentes modelos ou sistemas associados ao repositório, em servidores próprios ou na *cloud*;
- Agregação de dados de fontes externas, configurável por OAI-PMH e de forma ágil em ambiente de *dashboard*, permitindo a pesquisa e identificação de dados de produtores da UMinho;
- Comunidade de instituições de investigação com instalações *dataverse* de grande relevância, desde logo os autores do sistema, a Harvard University, sendo os utilizadores do sistema muito ativos a contribuir para o desenvolvimento do software;
- Transparência nos desenvolvimentos futuros do software, com a apresentação pública e periódica do *roadmap* de atualizações da plataforma, via website do projeto *Dataverse*, e a organização de reuniões online mensais da comunidade de utilizadores.

Este serviço a disponibilizar a toda a comunidade em janeiro de 2019, é um dos componentes de uma infraestrutura de dados que visa, simultaneamente, dotar os investigadores da UMinho de sistemas e tecnologias que os capacitem para competir com os seus pares de instituições congéneres mundiais e facultar as ferramentas para cumprirem de forma ágil e capaz os princípios FAIR da gestão de dados e os requisitos de financiadores de ciência e inovação.

Conclusões

Esta apresentação resume a atividade que os SDUM têm vindo a desenvolver no apoio à gestão de dados na UMinho e na operacionalização de serviços e sistemas para a infraestrutura de dados institucional, nomeadamente: i) no apoio à definição da estratégia da UMinho para a gestão de dados, ii) no trabalho de diagnóstico dos dados produzidos nas unidades de investigação, processo de gestão e identificação das necessidades de suporte à gestão de dados, iii) no desenvolvimento de ações de apoio informativo e programas integrados de formação, iv) na planificação e operacionalização de sistemas e criação de ferramentas de apoio à gestão de dados que assegurem a interligação e utilização dos componentes da infraestrutura institucional de dados.