Minority Report: A Execução do PGD em Análise

Daniel Agostinho[a], João Pereira[b], Alexandre Sayal[c], Bruno Direito[d]

[a]*Centro de Imagem Biomédica e Investigação Translacional (CIBIT), Centro de informática e Sistemas da Universidade de Coimbra (CISUC), Universidade de Coimbra, danielagostinho@dei.uc.pt*
[b]*Centro de Imagem Biomédica e Investigação Translacional (CIBIT), Universidade de Coimbra, jpereira@uc.pt*
[c] *Centro de Imagem Biomédica e Investigação Translacional (CIBIT), Universidade de Coimbra, alexandre.campos@uc.pt*
[d]*Centro de informática e Sistemas da Universidade de Coimbra (CISUC), bruno.direito@uc.pt*

Resumo

A crescente exigência de transparência na gestão de dados de investigação impulsiona a necessidade de mecanismos de acompanhamento eficazes. Este trabalho apresenta um *pipeline* de monitorização automatizado da execução do Plano de Gestão de Dados (PGD), concebido para fornecer relatórios quantitativos sobre a conformidade e o progresso da curadoria de dados. Tendo como base um PGD acionável por máquina (*machine-actionable DMP*), o conhecimento do *standard* de dados do projeto e o acesso direto à pasta de armazenamento do *dataset*, esta solução visa transformar o PGD de um documento estático num instrumento de gestão dinâmico e verificável.
O *pipeline* opera em três fases fundamentais. Primeiro, realiza a leitura e interpretação do PGD acionável por máquina para extrair não só as regras e requisitos para organização (nomeação de ficheiros, metadados, formatos, políticas de acesso), mas também os dados quantitativos previstos, como o número de itens ou o volume de armazenamento (este passo necessita da articulação dos requisitos com a construção do template de PGD). Na segunda fase, o sistema acede à pasta onde o *dataset* está a ser armazenado (após a devida curadoria) para inspecionar a estrutura de pastas, os nomes dos ficheiros, os metadados associados e os formatos utilizados. A terceira e última fase consiste na comparação entre as diretrizes estipuladas no PGD e o estado real do *dataset*, gerando relatórios de execução que quantificam a conformidade ao plano. Este

*todos os autores contribuiram de igual forma para o trabalho.

plano é facilmente atualizado, quer através da atualização do PGD - o processo pode ser automaticamente accionado após consulta de novas versões.
A implementação deste *pipeline* oferece benefícios significativos para **o investigador responsável**, proporcionando um acompanhamento contínuo e objetivo da execução do projeto, com elementos quantitativos que facilitam a gestão da equipa, valorizando o PGD e a necessidade da sua atualização e para **a organização**, oferecendo uma visão agregada e concreta dos requisitos de armazenamento, curadoria e recursos computacionais por projeto e no seu total.
Esta visão global permite uma gestão mais eficiente dos recursos partilhados, bem como a capacidade de estimar com maior precisão as necessidades para futuros projetos, promovendo uma cultura de excelência na gestão de dados de investigação.
Palavras-chave: maDMP, data center, gestão de dados

Designação do projeto/infraestrutura/iniciativa

LASI

Público-alvo

gestores de repositórios e data centers, gestores de ciência, curadores de dados, especialistas de informática, investigadores

Ligações web úteis

https://argos.openaire.eu/
https://bids.neuroimaging.io/