



Partilha de *Datasets* de Imagens: Desafios e Boas Práticas em Saúde e Agricultura

Rebeca B. Kalbermatter^a, João Mendes^b

^aCeDRI, Instituto Politécnico de Bragança, kalbermatter@ipb.pt

^bCeDRI, Instituto Politécnico de Bragança, joao.cmendes@ipb.pt

Resumo

A investigação científica contemporânea depende cada vez mais de *datasets* de qualidade, especialmente em áreas onde a análise de imagens desempenha um papel central. Tanto na saúde como na agricultura, a criação e partilha de *datasets* constituem um pilar essencial para a reproduzibilidade, a comparação de resultados e o avanço do conhecimento.

Na área da saúde, a abundância de imagens médicas contrasta com as fortes restrições éticas e legais relativas à sua utilização. Questões de anonimização, proteção da privacidade e normalização tornam complexa a criação de repositórios abertos. Ainda assim, projetos internacionais têm mostrado como a aplicação de padrões robustos e a adoção de licenças claras permitem equilibrar segurança e acesso, fomentando investigação de alto impacto.

Já na agricultura, o problema é de natureza distinta: existem pouquíssimos *datasets* abertos e estruturados, sobretudo em domínios especializados como a olivicultura. A ausência de dados anotados limita o desenvolvimento de algoritmos de visão computacional capazes de responder a problemas concretos, desde a deteção de pragas até a avaliação da maturação da cultura. A experiência acumulada na criação de *datasets* agrícolas mostra que a definição de esquemas de metadados, o pré-processamento de imagens, a padronização de formatos e a disponibilização em repositórios abertos são passos críticos para ultrapassar esta lacuna.

Esta apresentação conjuga assim as duas perspetivas, saúde e agricultura, para destacar semelhanças e diferenças nos desafios enfrentados, apontando lições que podem ser mutuamente aproveitadas. Irão ainda ser discutidos um conjunto de boas práticas comuns, desde *guidelines* de anotação até estratégias de versionamento e licenciamento, bem como um *checklist* prático para investigadores que pretendam disponibilizar *datasets* de imagens reutilizáveis.

Conclui-se que a comunidade GDI pode beneficiar de uma maior articulação entre áreas distintas, partilhando experiências que contribuam para uma cultura sólida de dados FAIR (*Findable, Accessible, Interoperable, Reusable*).

Palavras-chave: *datasets, imagens, partilha de dados, saúde, agricultura*

Designação do projeto/infraestrutura/iniciativa

Infraestrutura de **Datasets** de Imagens em Saúde e Agricultura.

Público-alvo

Investigadores, cientistas de dados, gestores de ciência, bibliotecários, curadores de dados, técnicos de informação.

Ligações web úteis

<https://doi.org/10.34620/dadosipb/ZXAKE6>

<https://doi.org/10.34620/dadosipb/TVYE8K>