Re.Data
Rede para a Gestão de Dados de Investigação

BOAS PRÁTICAS & CASOS DE USO

Gestão de dados na genómica e transcriptómica de abelhas (Apis mellifera)

O grupo de investigação desenvolve projetos que envolvem a sequenciação em larga escala do genoma (DNA) e do transcriptoma da abelha. Entre os principais projetos encontram-se MEDIBEES e Better-B, que abrangem milhares de indivíduos, e o projeto BEE3Pomics, dedicado ao estudo do transcriptoma. A equipa responsável pela análise de dados é constituída por cinco elementos e trabalha com volumes muito elevados de informação, frequentemente na ordem dos terabytes.

A gestão de dados constitui um desafio central. Os dados brutos fornecidos pelas empresas de sequenciação são armazenados no cluster do IPB, com mecanismos que garantem duplicação e integridade. Estes dados são organizados em pastas com permissões de leitura para todos os membros da equipa, sem possibilidade de modificação, assegurando que o raw data permanece inalterado.

O processamento e a análise são efetuados em contas individuais, sendo os resultados finais duplicados e organizados em diretórios partilhados. Para fins de publicação, os dados genéticos são depositados em repositórios disciplinares, como o NCBI Sequence Read Archive (SRA). As análises integram igualmente genomas externos disponíveis no SRA. Sempre que adequado, informação adicional é disponibilizada em materiais suplementares.

CIMO Centro de Investigação de Montanha

ipb INSTITUTO POLITÉCNICO DE BRAGANÇA

Gestão de dados na genómica e transcriptómica de abelhas (Apis mellifera)

Investigadora Entrevistada

Dora Henriques

Etapa do Ciclo de Vida dos Dados

- Planeamento
- Partilha
- Reutilização

Estrutura da narrativa

- Descrição
- Três lições aprendidas
- Três desafios futuros
- Cinco questões sobre GDI

VER RECURSO COMPLETO

Consórcio
Universidade do Minho · Universidade de Coimbra · iscte · ipb · NOVA
Apoio
FCCN serviços digitais fct · fct Fundação para a Ciência e a Tecnologia
Financiamento
PRR Plano de Recuperação e Resiliência · República Portuguesa · Financiado pela União Europeia NextGenerationEU

> *A partilha de dados contribui para acelerar o progresso científico e aumentar a transparência. No campo da genómica, os mesmos dados podem ser utilizados para responder a questões diversas, maximizando o retorno do investimento público.*
>
> Dora Henriques

3 LIÇÕES APRENDIDAS

O planeamento é essencial

A definição de um plano de gestão de dados antes do início do projeto é fundamental para evitar redundâncias, clarificar responsabilidades e garantir eficiência ao longo de todo o ciclo de vida dos dados. O planeamento deve incluir a definição de acessos, a escolha das plataformas tecnológicas, a estrutura de organização da informação e os procedimentos de backup e preservação.

Os repositórios institucionais são estratégicos

A utilização de plataformas institucionais, como o dados.ip.pt, revela-se crucial para uma gestão eficaz e sustentável dos dados de investigação. Estas ferramentas facilitam o controlo de versões, promovem a colaboração entre equipas e permitem a atribuição de identificadores persistentes (DOI), contribuindo para uma maior transparência e organização.

Os princípios FAIR são indispensáveis

A adoção dos princípios FAIR (Findable, Accessible, Interoperable, Reusable) aumenta significativamente a transparência, a reprodutibilidade e o impacto da investigação. Dados bem documentados, acessíveis e interoperáveis facilitam a sua reutilização por outros investigadores, ampliam o alcance dos resultados e reforçam a credibilidade científica. Integrar estes princípios desde as fases iniciais do projeto permite uma gestão mais robusta e alinhada com as melhores práticas internacionais.

3 DESAFIOS FUTUROS

INTEGRAÇÃO TOTAL DA EQUIPA

É fundamental assegurar que todos os membros do grupo utilizam o repositório institucional de forma consistente e integrada. A adoção desigual destas plataformas reduz significativamente a sua utilidade e compromete a eficiência da gestão de dados.

PLANOS DE GESTÃO DE DADOS ROBUSTOS

A elaboração de planos de gestão de dados que sejam simultaneamente aceites pelas entidades financiadoras e úteis para a equipa de investigação continua a ser um desafio. Estes planos devem ser suficientemente estruturados para garantir conformidade e qualidade, mas também flexíveis para acompanhar a natureza dinâmica dos projetos.

FERRAMENTAS PARA MONITORIZAÇÃO GLOBAL

A implementação de sistemas institucionais de acompanhamento — semelhantes a plataformas como o GitLab — é essencial para monitorizar prazos, entregas, DOIs e deliverables, assegurando uma gestão integrada e reduzindo o risco de perda de informação.