

RECUPERAÇÃO E CONSOLIDAÇÃO DO ACERVO DE DADOS GEOQUÍMICOS DA CPRM: PANORAMA ATUAL

Mota, C. E.¹; Larizzatti, J. H.¹; Moreira, A. P. C.¹; Porto, C. G.¹; Silveira, F. V.²; Cunha, F.¹;

¹CPRM/Serviço Geológico do Brasil, Rio de Janeiro-RJ; ²CPRM/Serviço Geológico do Brasil, Brasília-DF

RESUMO: A CPRM possui um amplo e valioso banco de dados de geoquímica, resultado de mais de 40 anos de levantamentos. O acervo compreende um volume aproximado de 450.000 amostras de diversos materiais geológicos, analisadas por diversos métodos analíticos. Atualmente, este acervo encontra-se em processo de reorganização e validação, para ser disponibilizado à sociedade em breve, via GEOBANK. O objetivo é apresentar um panorama do estado atual do acervo geoquímico e quais são as perspectivas futuras. O motivo é que na base de dados ocorrem inconsistências, inclusive casos de localização inadequada. Além disso, novas informações geoquímicas são continuamente produzidas, inclusive reanálise de alíquotas existentes no acervo, e que precisam estar integradas com os dados de legado. Para a análise dos dados foram utilizados softwares livres e de código aberto, como o Quantum GIS (Ferramenta SIG) e o RapidMiner (*ETL e Business Intelligence*), além de *scripts* Python e bibliotecas como Django, Pandas e Matplotlib. A unificação destes dados foi implementada no banco de dados PostgreSQL 9.3/PostGIS 2.0.7. Em linhas gerais, o acervo geoquímico pode ser subdividido em dois conjuntos de dados: O acervo histórico, do início da década de 1970 até o final da década de 2000, extraídas do GEOBANK; e Levantamentos geoquímicos do final da década de 2010, de posse da DIGEOQ. A partir de observações e pesquisa de documentos internos, um modelo conceitual foi elaborado que permitisse a integração entre as fontes distintas. Uma das diretrizes do modelo é a preservação das coordenadas originais e as transformações para objetos *Geometry*, projetados em SIRGAS2000. O resgate das coordenadas e parâmetros cartográficos foi feito preditivamente, através de consultas às bases cartográficas da época. Os dados geográficos recentes, na sua maioria, são coletados via GPS, com exceção de informações enviadas em planilhas ou fichas. Os erros locais devem-se as seguintes razões: (1) Erros humanos por digitação incorreta, principalmente em época pré-GIS; (2) Erros computacionais, pela definição incorreta dos algoritmos de conversão de coordenadas. Atualmente, duas atividades estão em execução: A consolidação e a consistência dos dados. O trabalho de consolidação representa exatamente a unificação de todos os dados produzidos, inclusive os sabidamente incorretos, para a análise de padrões e requisitos e validar os procedimentos descritos nos manuais da CPRM. O processo de consistência corresponde a verificação dos metadados, pontos (inclusive sistemas de coordenadas) e de dados analíticos. Esta etapa implica em consulta aos mapas e relatórios originais para reajustar as amostras que porventura estejam deslocadas, incluir ou excluir amostras faltantes, ajustar metadados e boletins analíticos. Paulatinamente, os dados (novos e históricos) corrigidos e validados pela DIGEOQ serão disponibilizados ao público via GEOBANK, nas modalidades de consulta a metadados e download de resultados analíticos, em formato CSV, compactados em arquivos ZIP. Por fim, a CPRM, através de ações como a descrita neste trabalho, reforça o compromisso em dar credibilidade técnica aos dados disponibilizados para benefício da sociedade.

PALAVRAS-CHAVE: GEOQUÍMICA, ACERVO, BANCO DE DADOS