Goldschmidt 2026 Abstract
https://doi.org/10.46427/gold2026.36061

GeoReM 2.0 – First tentative steps toward automated curation and checking for data completeness

AXEL D. RENNO[1], MARIE KATRINE TRAUN[2], DIETER GARBE-SCHÖNBERG[3], LEANDER KALLAS[2], ADRIAN STURM[4], MATTHIAS WILLBOLD[2] AND GERHARD WÖRNER[2]

[1]Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology
[2]University of Göttingen
[3]Kiel University
[4]Göttingen State and University Library

The GeoReM (Geological and Environmental Reference Material) database was initiated in 2005 by Klaus Peter Jochum and developed with colleagues at the Max Planck Institute for Chemistry in Mainz, Germany. Today, GeoReM comprises 453,666 data values extracted from 13,946 publications, covering 3,954 geological, microanalytical, environmental, and biological reference materials (RM) —reflecting a major curatorial effort. After Jochum's passing, responsibility for GeoReM was transferred to the International Association of Geoanalysts (IAG) in August 2025. Infrastructure support is now provided by the Digital Geochemical Data Infrastructure (DIGIS) team at the University of Göttingen.

Since the early 2020s, several developments have created new challenges for GeoReM. These include a rise in publishing activity, often accompanied by insufficiently reviewed articles in so-called predatory journals; a substantial growth in data publication volumes in line with FAIR principles; and increasing data output driven by advances and automation in geoanalytical methods.

For GeoReM 2.0, data curation must adapt to these changes. Previously, curation relied on experience-based searches within a regularly updated list of trusted journals, supplemented by voluntary submission of published datasets by authors. To start adapting to the increasing data volumes, we evaluated data completeness using a case selection of the IGG-B1 to B8 boron isotope reference materials [1, 2]. An extended citation map was applied to identify whether relevant data from original or subsequent citations had been incorporated into GeoReM. The results are presented alongside a statistical assessment of how newly integrated datasets refine knowledge of these eight reference materials. The method also enables efficient identification of missing data for other reference materials.

In addition, the complete citation data of GeoReM are analyzed within a knowledge graph that visualizes reference material usage and related reference material clusters in the geoanalytical community as a network mind map. The integration of bibliographic data in the RIS format (Reference Information System) also used by journals into GeoReM 2.0 enhances the traditional "combing" approach by enabling systematic citation mapping with automated triggers for newly published studies.

[1] Tonarini et al. (2003) https://doi.org/10.1111/j.1751-908X.2003.tb00710.x
[2] Gonfiantini et al. (2003) https://doi.org/10.1111/j.1751-908X.2003.tb00711.x