

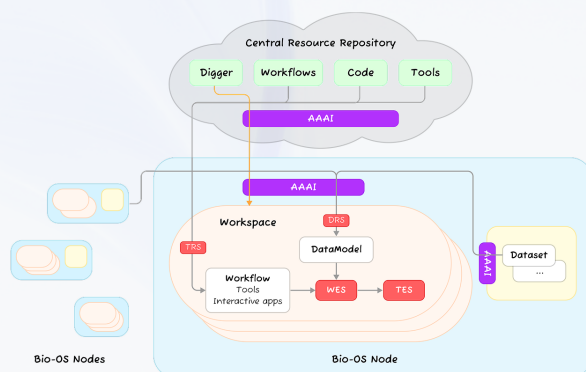
Bio-OS: Real-World Implementation of GA4GH Cloud Standards in China



Authors: Jilong Liu, Jianwen Zhou, Zhaoqiang Li, Qingyu Xiao, Kaixin Yang, Ziru Chen, Lu Yao, Hengshen Liu, Yibo Miao, Junyu Luo, Yiping Chen, Ruifeng jing, Yixue Li
Orgnization: Guangzhou Laboratory, Guangzhou 510005, Guangdong Province, China | Contact: liu_jilong@gzlab.ac.cn

The Challenge: Interoperability and Development of Biomedical Cloud Platforms

While cloud computing has revolutionized data processing and collaboration in life sciences, significant challenges persist in developing tailored cloud platforms for biomedical research and ensuring seamless interoperability. The Global Alliance for Genomics and Health (GA4GH) Cloud application programming interfaces (APIs) provide a critical framework to address these issues. To facilitate large-scale, cross-institutional biomedical data utilization in China while enhancing global connectivity, we developed Bio-medical big data operating system (Bio-OS) based on GA4GH Cloud standards.



Bio-OS is designed for large-scale, cross-institutional biomedical data utilization based on GA4GH Cloud standards. It not only strictly complies with the four key service specifications — TRS, DRS, WES, and TES — but also conducts functional expansion and optimization in response to the actual needs of biomedical research.

Core Features: Standards Implementation and Novel Extensions



Workspace RO-Crate integrates comprehensive research components including project introductions, datasets, analytical workflows, interactive analysis instances, and operational logs, thereby facilitating the standardized export and streamlined distribution of research workspaces.

The workspace paradigm organizes research projects encompassing introduction, datasets, workflows, interactive analysis instances and logs. We developed Workspace RO-Crate as a standardized format for workspace export and distribution, complemented by Digger (<https://network.miracle.ac.cn/digger>) repository for artifact sharing.

Our Solution: Bio-OS – A GA4GH-Compliant Cloud-Native Framework

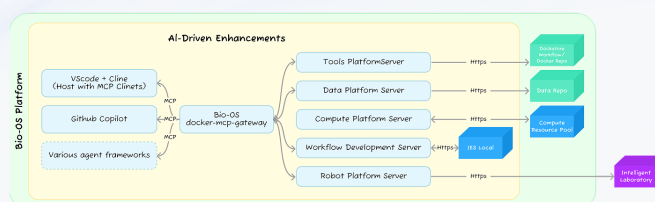


Bio-OS has established nodes across North, East, and South China, enabling standardized connectivity with international institutions.

Bio-OS is an open-source, cloud-native framework engineered to accelerate the development of interoperable biomedical cloud platforms. We established a centralized registry service to interconnect Bio-OS instances, forming a collaborative resource network. Bio-OS implements GA4GH standards including Workflow Execution Service (WES), Task Execution Service (TES), Data Repository Service (DRS), Tool Registry Service (TRS), and Authentication and Authorization Infrastructure (AAI) Passport, while introducing novel features: Interactive Execution Service (IES) for tertiary analysis support, workspace-centric data management, and AI-driven enhancements.

Impact: AI Integration and Multi-Institutional Validation

Enhanced machine and Large Language Model operability through Workspace RO-Crate enabled development of Bio-OS Copilot, **transforming Bio-OS into an operable environment for AI agents and enabling the AI4LifeScience paradigm.** This leverages a VS Code extension for natural language-based workspace manipulation. These innovations promote adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles, enhancing reproducibility and collaboration. Multi-institutional deployments validate Bio-OS's efficacy in supporting data-intensive biomedical applications and establishing robust cloud platforms for life sciences research.



Bio-OS exposes a suite of core functionalities—including data processing, workflow development, and task execution—as MCP (Model Context Protocol) Servers. Leveraging VS Code integrated with LLM via an extension, users can invoke these MCP Servers to operate Bio-OS.