

Enhanced metadata standards supported by the National Microbiome Data Collaborative

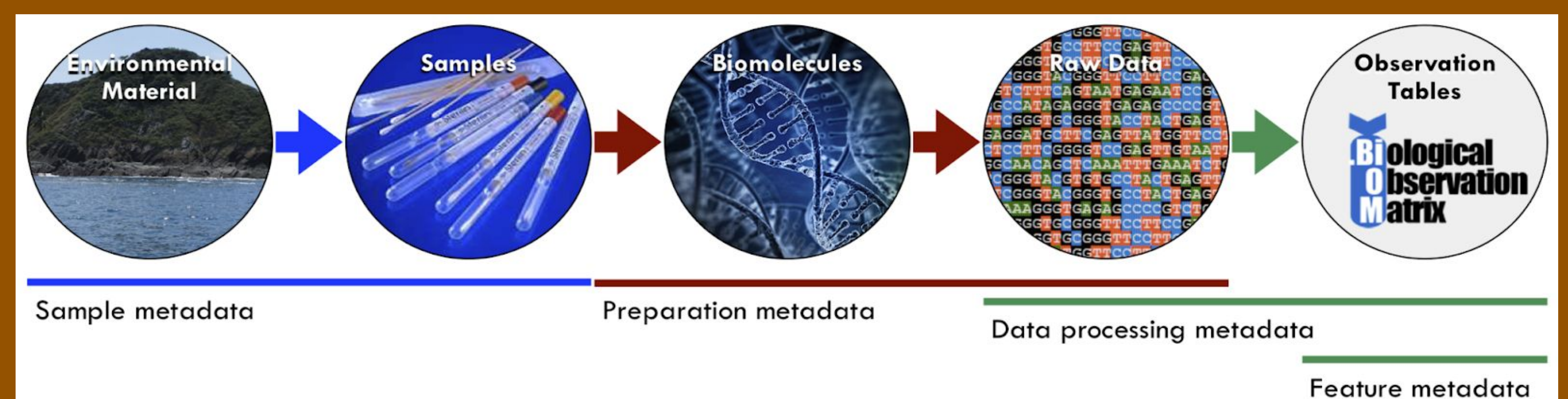
Chris Mungall^{1*} (cjmungall@lbl.gov), Faiza Ahmed², Anubhav³, Jeffrey Baumes², Jonathan Beezley², Mark Borkum³, Lisa Bramer³, Shane Canon¹, Patrick Chain⁴, Danielle Christianson¹, Yuri Corilo³, Karen Davenport⁴, Brandon Davis², Meghan Drake⁵, William D. Duncan¹, Kjersten Fagnan¹, Mark Flynn⁴, David Hays¹, Bin Hu⁴, Marcel Huntemann¹, Julia Kelliher⁴, Sonya Lebedeva¹, Po-E Li⁴, Mary Lipton³, Chien-Chi Lo⁴, Douglas Mans³, Stanton Martin⁵, Lee Ann McCue³, David Millard³, Kayd Miller¹, Nigel Mouncey¹, Paul Piehowski³, Elais Player Jackson⁴, Anastasiya Prymolenna³, Samuel Purvine³, TBK Reddy¹, Rachel Richardson³, Migun Shakya⁴, Montana Smith³, Jagadish Chandrabose Sundaramurthi¹, Deepak Unni¹, Pajau Vangay¹, Bruce Wilson⁵, Donny Winston⁶, Elisha Wood-Charlson¹, Yan Xu⁴, Emiley Eloe-Fadrosh¹

¹ Lawrence Berkeley National Laboratory, Berkeley, CA; ² Kitware, Clifton Park, NY; ³ Pacific Northwest National Laboratory, Richland, WA; ⁴ Los Alamos National Laboratory, Los Alamos, NM; ⁵ Oak Ridge National Laboratory, Oak Ridge, TN; ⁶ Polyneme LLC, New York, NY

Project Goals:

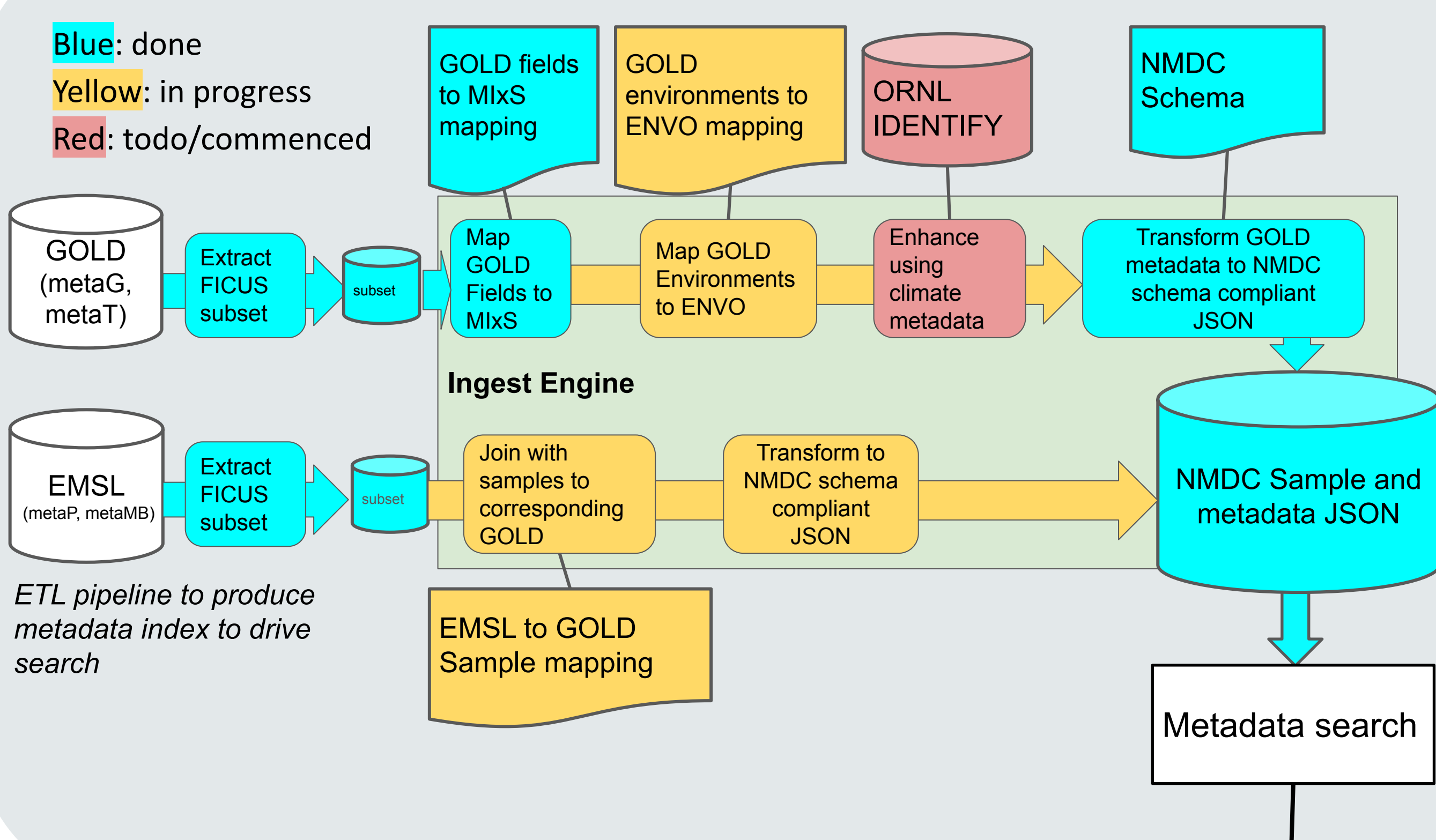
- Metadata plays a vital role in the analysis and interpretation of multi-omics microbiome data such as metagenome, metatranscriptome, metaproteome and metabolome.
- We have created a FAIR schema by integrating various resources including the Genomic Standards Consortium MixS environmental packages, Environment Ontology (EnvO) and Genomes Online Database (GOLD) to enhance the metadata of biosamples.
- We extended the W3C PROV standard for metadata about computational workflows and adopted standards and ontologies including GFF3, KEGG, PSI-MS for multi-omics and functional annotation.

Microbiome Metadata

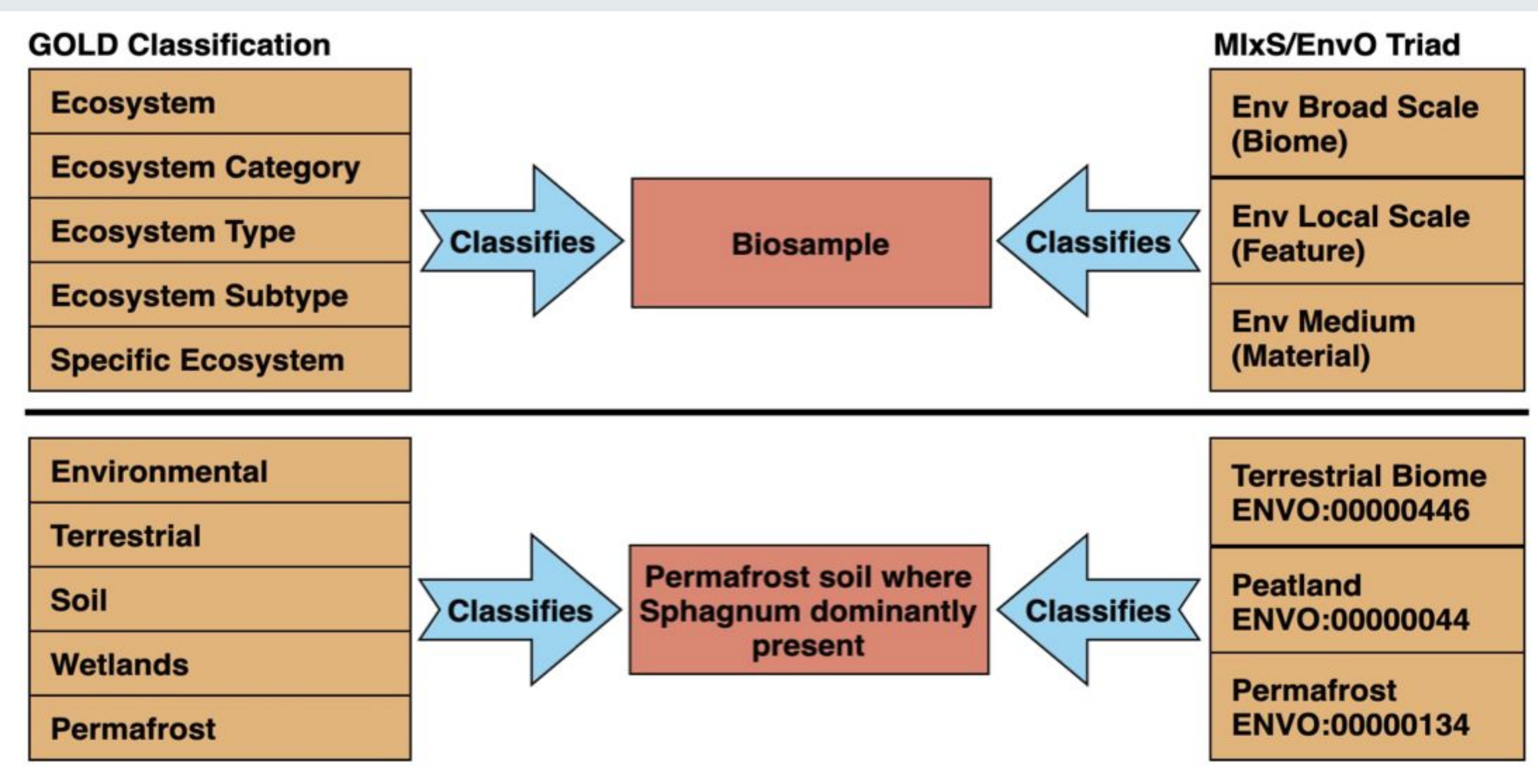


Microbiome metadata types: Information that contextualizes sample including its geographic location and collection date, sample preparation, data processing methods, and data products produced from a biological sample (Luke et al., 2020. Introduction to Metadata and Ontologies: Everything You Always Wanted to Know About Metadata and Ontologies (But Were Afraid to Ask). DOI: 10.25979/1607365.

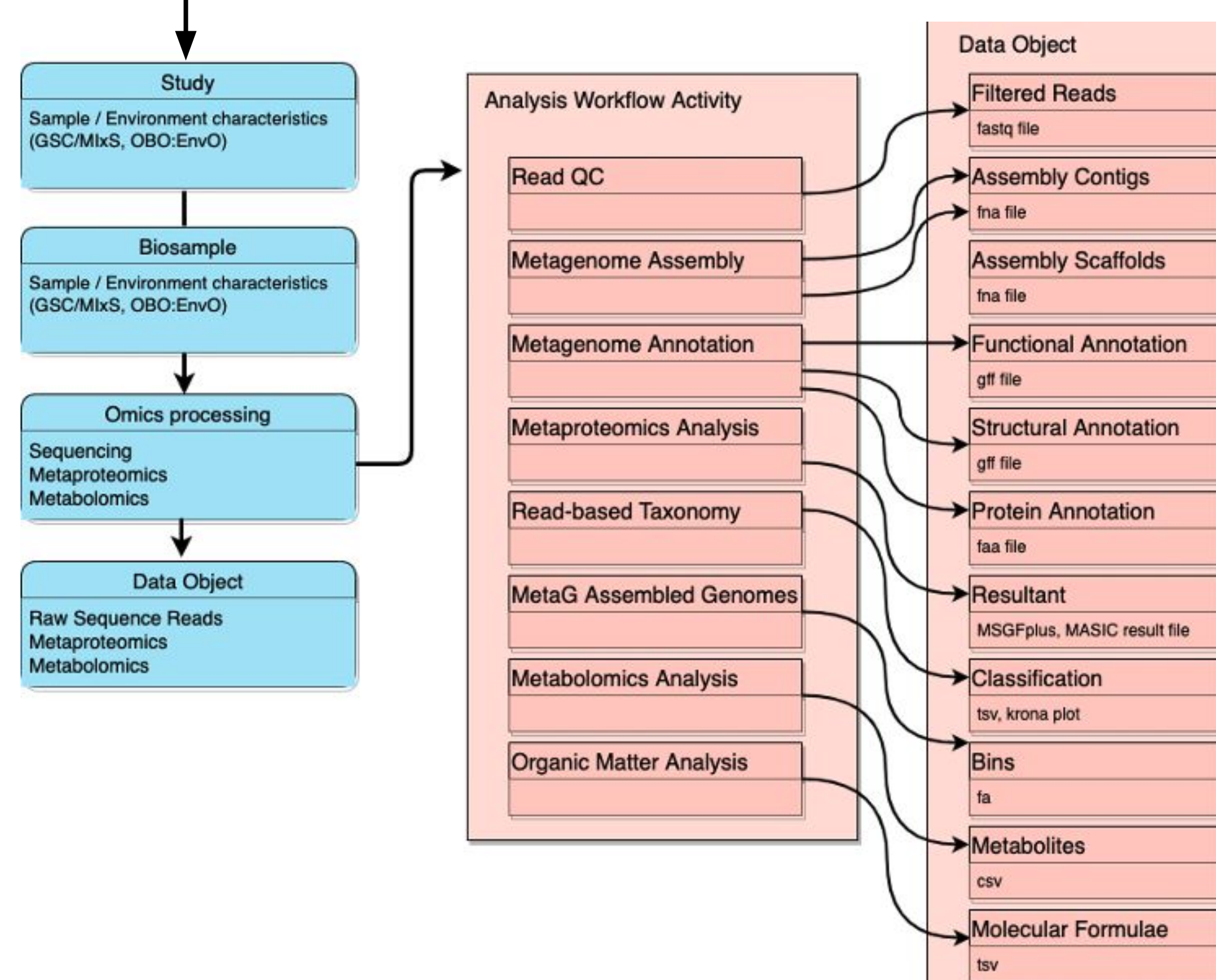
Enhanced Metadata Standards for Microbiome Research



Mapping of GOLD ecosystem paths with MixS/EnvO triad terms

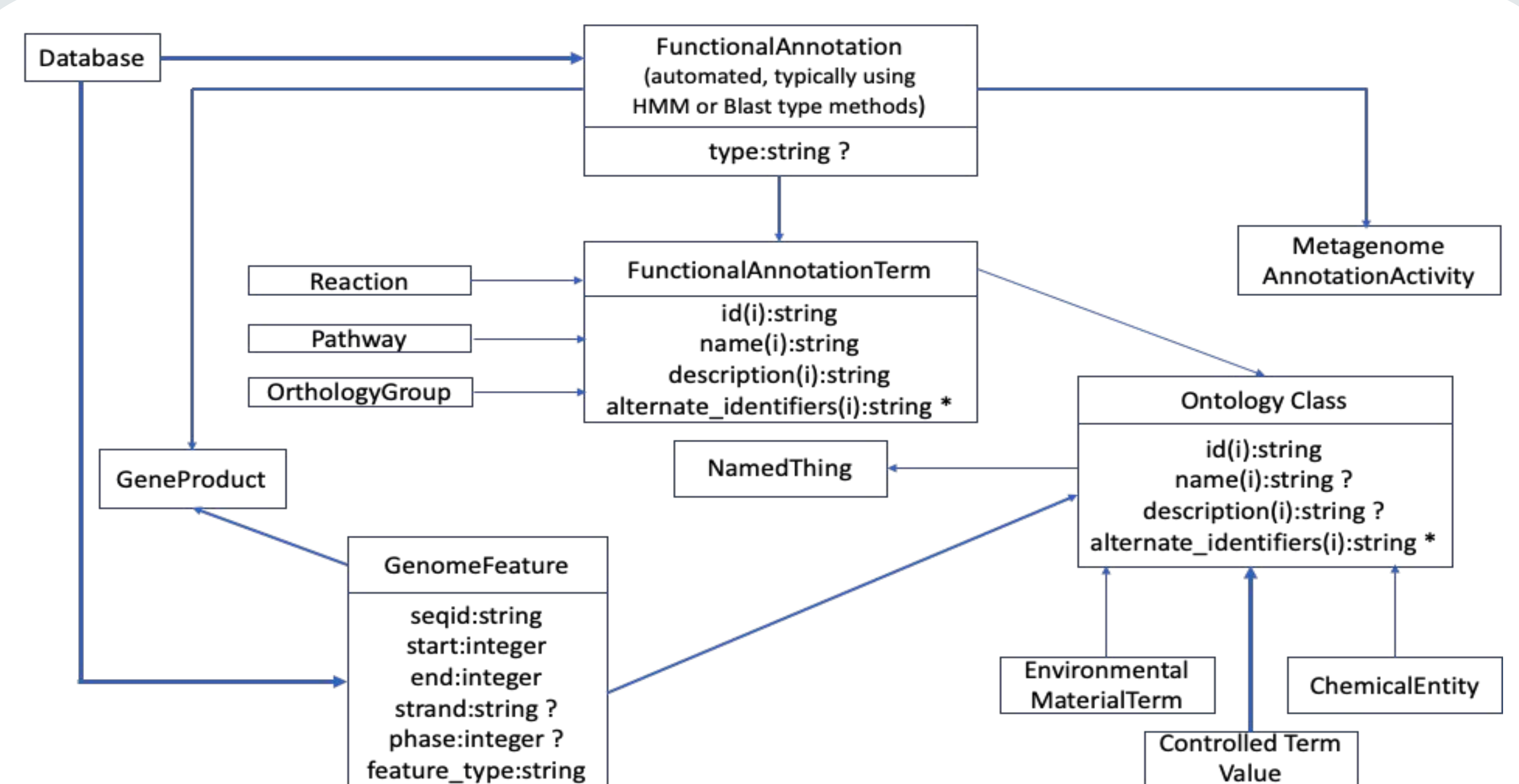


The environmental features of a biosample in GOLD is described by a 5-level ecosystem classification path that was mapped with MixS-adopted Environment Ontology triad terms in the NMDC schema for further metadata enhancement. EnvO triad for a given biosample was determined using various GOLD fields including the biosample name, description, habitat, sample collection site, identifier, ecosystem path and study description.



The metadata enhancement is achieved using various standards and ontologies including MixS and EnvO for biosamples, and W3C PROV, GFF3 and PSI-MS for multi-omics workflows.

Additional details: <https://github.com/microbiomedata/nmdc-metadata>



An illustration for metadata of workflows shown using NMDC schema classes of Genome Feature, Functional Annotation, Functional Annotation Terms and Ontology Class. The NMDC schema allows for the development of consistent interoperable linkages between metadata types.

Towards FAIR microbiome data

Integration of various standards and ontologies for enhancing the environmental context of biosamples ensures data are **findable** and **accessible**, and together with workflow metadata produces **interoperable** and **reusable** annotated data products.