

foundingGIDE: メタデータ調和によるバイオイメージングデータリソースの国際連携

京田 耕司¹、山縣 友紀^{2,3}、糸賀 裕弥¹、藤澤 絵美¹、大浪 修一^{1,2} (¹RIKEN BDR, ²RIKEN R-IH, ³RIKEN-BRC)

概要

foundingGIDEは、日欧豪の国際共同プロジェクトであり、バイオイメージング分野におけるデータ共有、再利用を担っているデータリソース間の相互運用を可能にする基盤の確立を目的としている。

背景

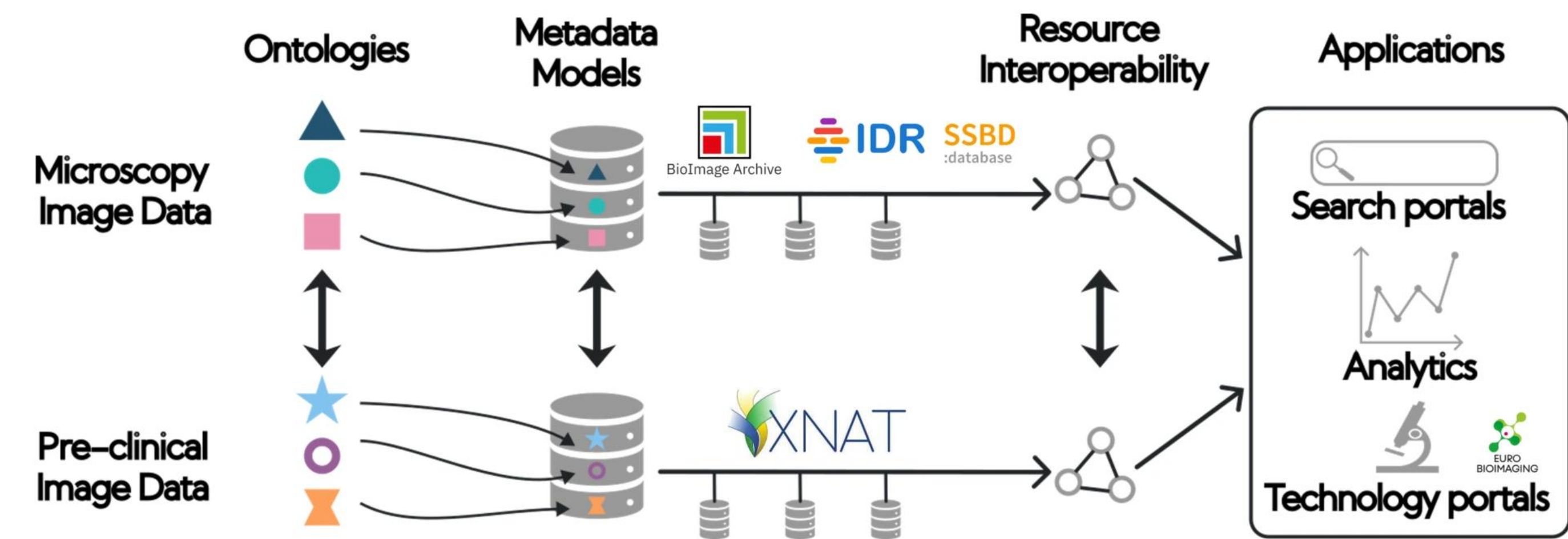
- バイオイメージング分野において多様なリポジトリが構築されているが、各データリソースで使用するメタデータモデルや語彙、オントロジーが統一されておらず、横断的な検索や解析が困難である。

方法

- 主要なバイオイメージングデータのデータリソースのメタデータ項目と統制語彙・オントロジーの比較解析を行い、メタデータ調和の要件を定義することにより、データリソース間の相互運用を可能にする。

参加機関と対象リソース

- 参加: 理研 (WP6リード) , EMBL-EBI, GerBI, UNSYD等
- 対象: バイオイメージ (SSBD, Image Data Resource, BioImage Archive) 、プレクリニカルイメージ



メタデータ調和のための要件

- メタデータの比較解析の結果とユースケースを考慮して、バイオイメーシングデータのデータリソースのメタデータを調和するための要件を定義した。

ユースケース

- データ引用（著者）
- データ検索（幅広い分野のユーザーを想定）
- データの比較と統合
- データ解析
 - AIを基盤としたセグメンテーション・クラス分類手法等の開発
 - AIを利用した大規模データに対する自動アノテーション手法の開発

メタデータ調和のための要件

Component	Description	Format	Access Query
Study Description	Description of the study	Text	Free text
Authors	Information about the authors	First name and Last name; Affiliation, ORCID	Author name, ORCID, Credit, Organization name, ROR
Organization	Information about the organization	Name URL, ROR	Name, ROR
Publication	Information on the related publication	Text (Title), Publication year, DOI (Journal)	DOI
License	Information about the dataset's license	Creative Commons with the version and URL	CC version, URL
Release Date	The release date of the data	ISO 8601 format (including time/time zone)	Release date, range search
Imaging Method	The method used to obtain image data	FBBI, EDAM Bioimaging Ontology term and ID	FBBI or EDAM Bioimaging Ontology ID, term
Cell Line	Information about the cell line	Cell Line Ontology term and ID	CLO ID, term
Organism	Information about the organism studied	NCBI Taxonomy term and ID, BioSample (geographic location)	NCBI Taxonomy ID, term
Gene	Information about related genes	Ensembl Gene or NCBI Gene name or ID	Gene ID, name
Compound	Information about the compound	ChEBI or PubChem term and ID	Compound ID, term
Antibody	Information about the antibody used	FBBI or ChEBI term and ID, RRID	Antibody ID, term
Channel – Content	Content related to the channel	FBBI term and ID	Channel content ID, term
Channel – Biological Entity	Biological entities related to the channel	Experimental Factor Ontology term and UniProt ID	Biological entity ID, term
Instrument	Information about the instrument used	Compliant with 4DN-BINA-OME-QUAREP (NBO-Q) standards, PIDInst	Instrument name, ID
Dimension	Information about data dimensions	Compliant with OME-Zarr axes info. (T, C, Z, Y, X)	Dimension specification
Pixel/Voxel Size/Time resolution	Size of the pixels or voxels, Time interval	Pixel/Voxel size and time interval stored with units	Pixel/voxel size, Time interval
Study Unique ID	Unique ID for the study	Accession ID, DOI	Accession ID, DOI
Dataset Unique ID	Unique ID for the dataset	Accession ID	Accession ID
Pathology/Disease (Biological Entity)	Pathology/Disease related to the biological entity	SNOMED-CT, DOID, ICD-11 or MONDO	Pathology/Disease ID, term
Phenotype (Analysis Data)	Phenotypic data related to the analysis	Cell Morphology Phenotype Ontology, MPO, or hPO	Phenotype ID, term
Organ	Information about anatomical entities	UBERON, Radlex Ontology	UBERON ID, term
Analyzed Data	Information about software, workflow, annotation data	GitHub URL, (release), DOI (Zenodo, Figshare), RRID, Dataset ID	Name, DOI

段階的な調和

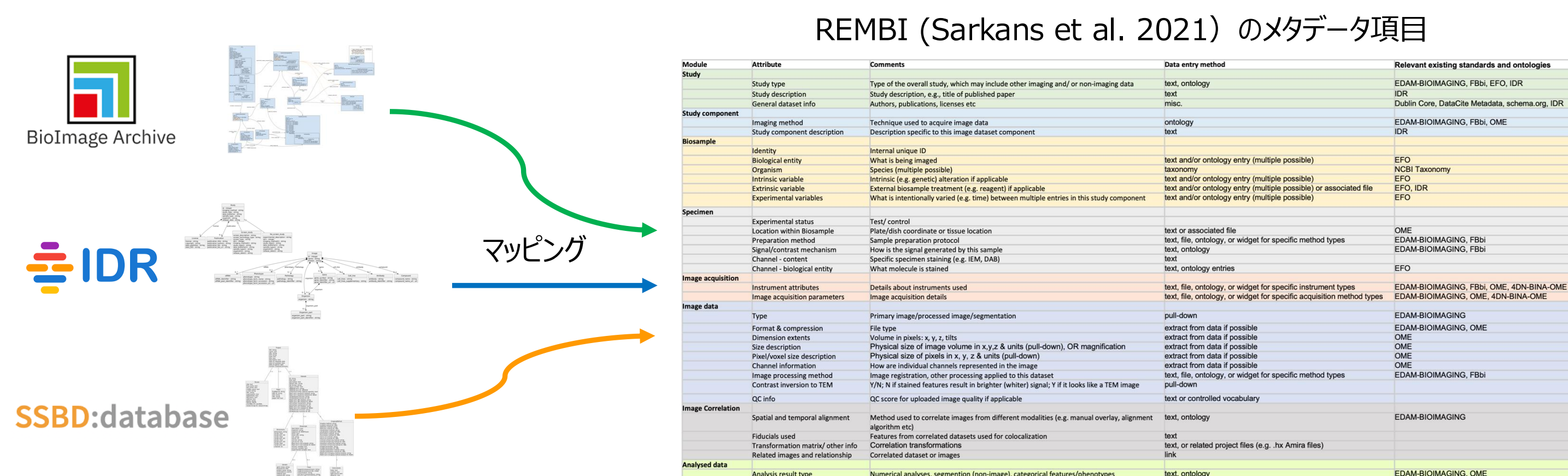
- フェーズ1: 広く共有されている項目の標準化
- フェーズ2: 部分的に重複する語彙間のマッピング
- フェーズ3: 各リソース固有の項目への対応

メタデータの比較解析

主要なバイオイメーjingデータリソース (BioImage Archive, Image Data Resource, SSBD) のメタデータ項目、語彙・オントロジーの比較を行う。

方法

- 各データリソースのメタデータモデルからメタデータ項目を抽出し、REMBIの項目に当てはめて項目の共通点と差異を明らかにする。
- さらに、メタデータモデルで使用されている統制語彙・オントロジーの比較を行い、その共通点と差異を明らかにする。



メタデータ項目の比較

[illegible]

統制語彙・オントロジーの比較

Category	Ontology/Controlled vocabulary	Acronym	RIA	IDA	SNAD
Imaging Method	Biological Imaging Methods Ontology	FBBI	✓	✓	✓
Taxonomy	NCBI Organismal Classification	NCBITaxon	✓	✓	✓
Cell Line	Cell Line Ontology	CLO			
Cell type	Cell Ontology	CL			✓
Gene	Ensembl	ENSG		✓	✓
	Enrtrr Gene	NCBI_Gene		✓	✓
Protein	Universal Protein Resource	UniProt	✓	✓	✓
Chemical Compound		PubChem			✓
	Chemical Entities of Biological Interest Ontology	ChEBI			✓
Experimental Condition	Experimental Factor Ontology	EFO	✓	✓	✓
	Ontology for Biomedical Investigation	OBi			✓
Unit	Units of Measurement Ontology	UO			✓
Disease	Systematized Nomenclature of Medicine-Clinical Terms	SNOMED CT			✓
Cellular Phenotype	Cellular Microscopy Phenotype Ontology	CMPO	✓		✓
Biological Process	Uberon-Multi-species Anatomy Ontology	UBERON			✓
Biological Process	Gene Ontology	GO			✓
Cellular Component					✓
Molecular Function					✓
Controlled Vocabulary	Medical Subject Headings	MeSH			✓

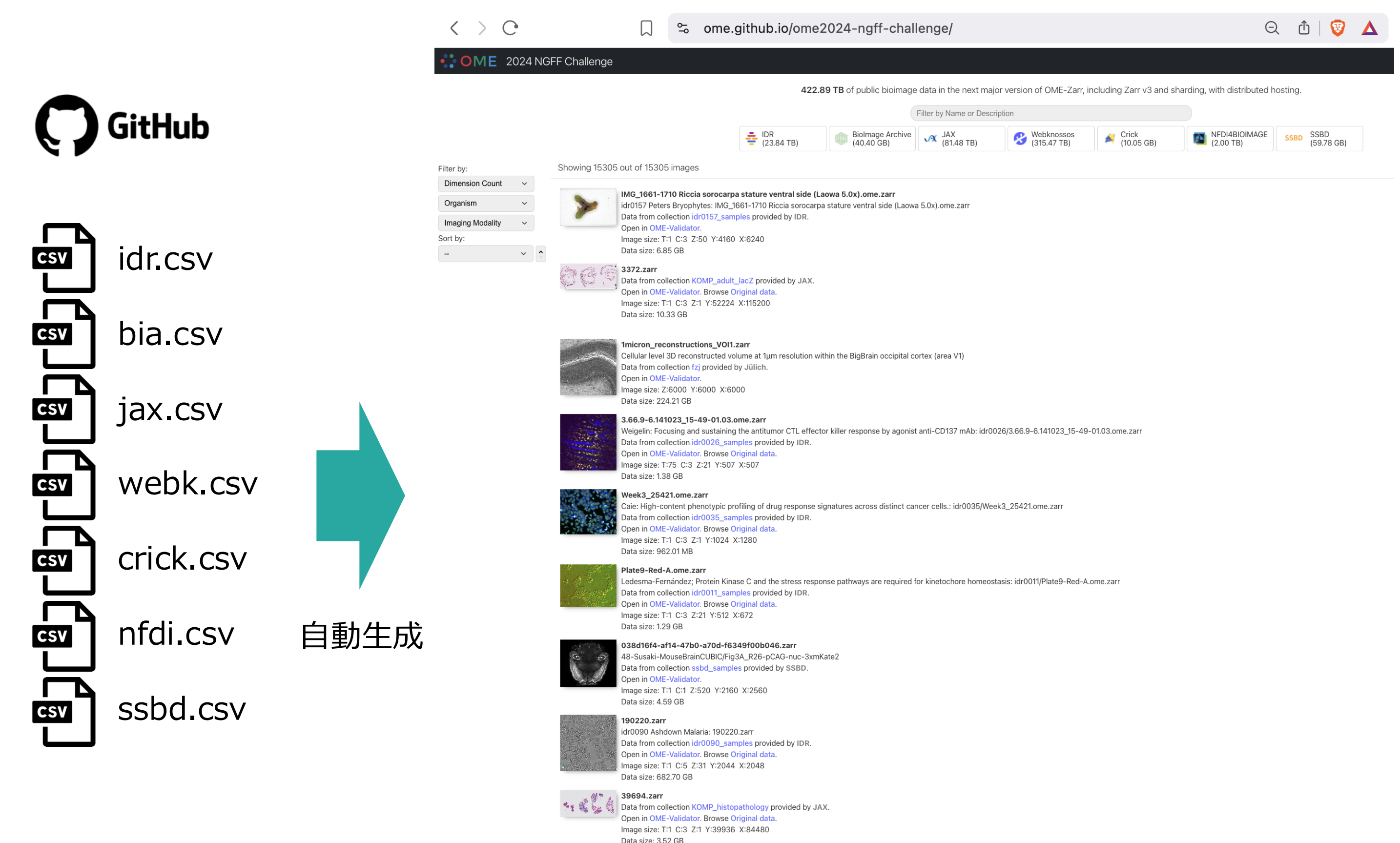
- 共通項目が存在した。
- 同じ項目に対して異なるオントロジーが使用されている場合があった。

リソース間相互検索

各データリソースが要件に基づくメタデータを提供することにより、複数のデータリソースを横断する検索機能を実現する。

現在の進捗

- データリソースでの対応
 - IDRでテスト検索サイトを立ち上げた。
- ポータルサイトの作成
 - OME2024 NGFF Challengeにて検索サイトを立ち上げた。



今後の展開

- データの発見性、アクセス性、再利用性のより一層の向上
- AI/ML対応データ基盤の支援に向けた取り組み
- 各種画像解析ツール（Fiji, napari等）との連携強化
- LLM等による自動アノテーション