



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

RNA

BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS

Funded by



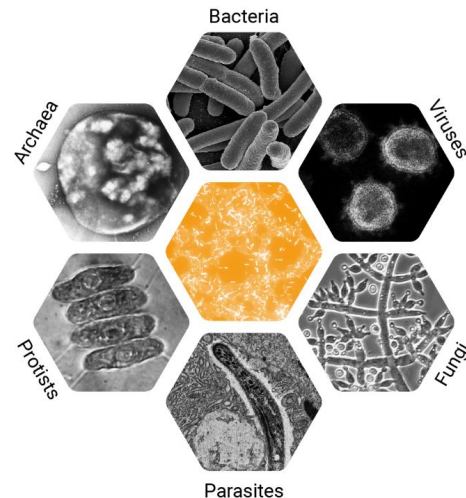
Deutsche
Forschungsgemeinschaft
German Research Foundation



VirJenDB

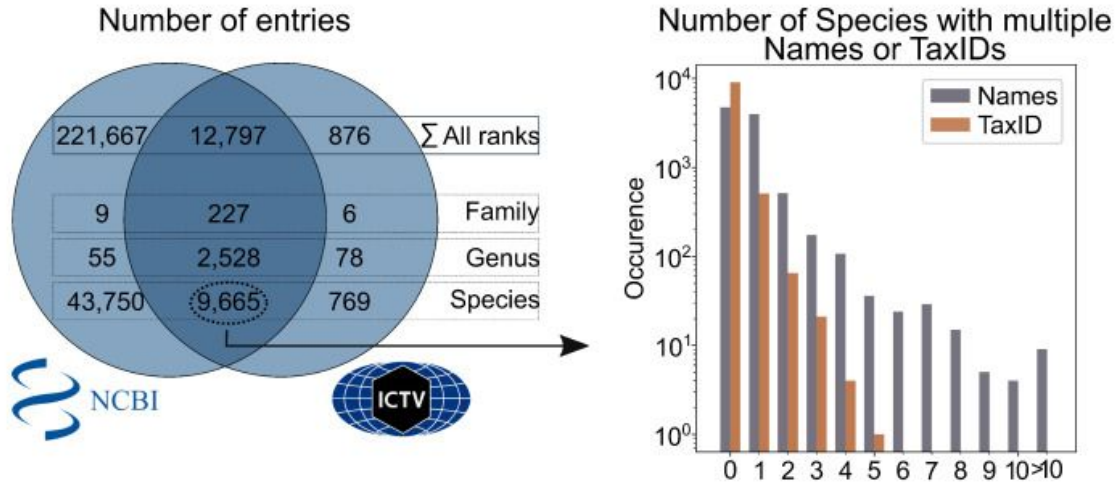
40th TBI Winterseminar
Bled, Slovenia

Shahram Saghaei
10 February 2025



<https://nfdi4microbiota.de/>

Background: (virus) sequencing data volume keeps increasing



- Currently >24 active virus genome databases (Ritsch *et al* 2023 Viruses)
- >18K virus genomes in NCBI RefSeq
 - Represent only a fraction of available virus sequences
 - IMG/VR contains >15M UViGs
 - VirusDIP reports >30M sequences
- Need for efficient and organized storage of sequencing data and downstream datasets (alignments, predicted structures)
- Also a need for FAIR and Open and well-curated options for data re-use

Aims



Develop user-friendly database to lower barriers for virus -omics data analysis

Focus global efforts to curate and correct virus metadata



Support virologists to deposit their virus sequences with rich metadata

FAIR and Open solutions: integration of tools and API development

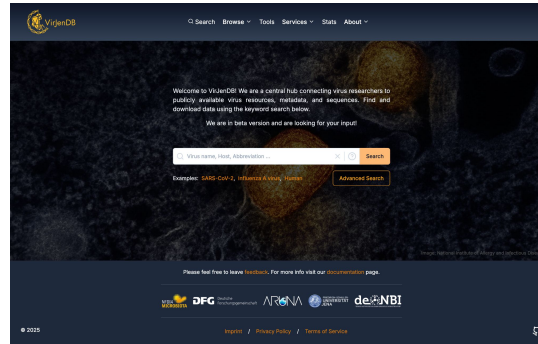


FAIR Principles: Wilkinson et al 2016 *Sci Data*
Open Science: <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

Database Implementation

Repositories

Knowledge Resources



VirJenDB: A Central Hub for Virus Research

virjendb.org
96 Metadata Fields
>13 Million Records

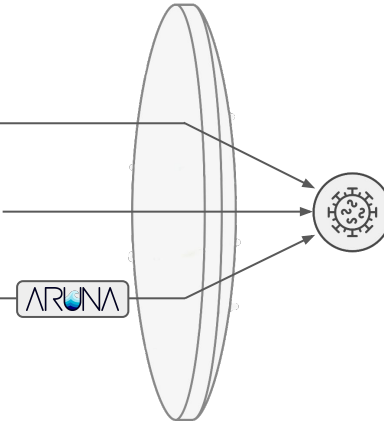
Reference Resources



Metadata



Sequences



Browse



Keyword Search



Submission



Metadata Schema Expansion

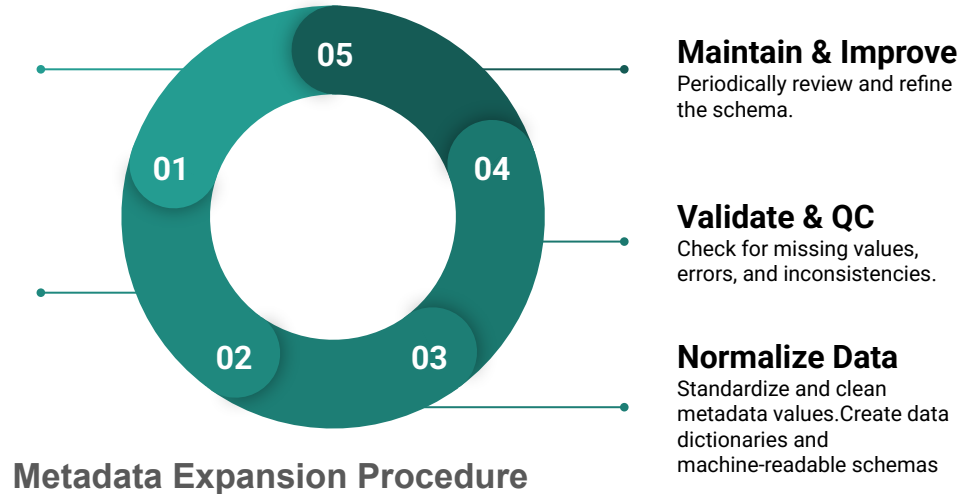
- Improved Data Standardization & Interoperability
- Better Querying & Filtering Capabilities
- Enhanced Data Compatibility & Cross-Referencing
- Supports Future-Proofing & Scalability
- Avoids Information Loss

Assess Standards

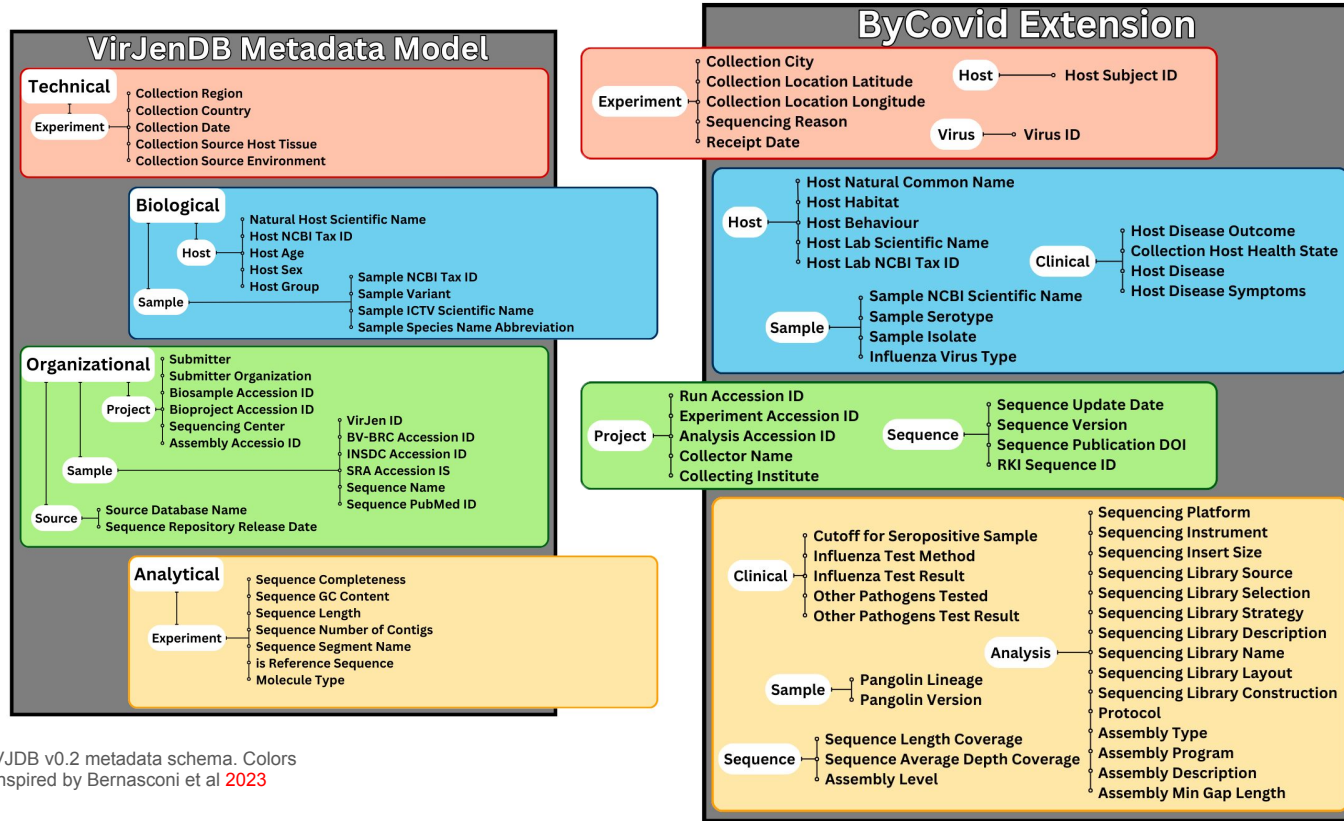
Compare existing metadata with established databases (INSDCI, GSC etc.)

Define Schema

Set field types, naming conventions, and controlled vocabularies.



Metadata Schema Expansion



VJDB v0.2 metadata schema. Colors inspired by Bernasconi et al 2023



MlxS standards (MIGS-VI)

ROBERT KOCH INSTITUT



NFDI 4 BIODIVERSITY



- Expand the metadata schemas by 62 metadata field
- Pipeline for quality checking and correction of metadata

Help Metadata Enrichment by Easier Data Deposition

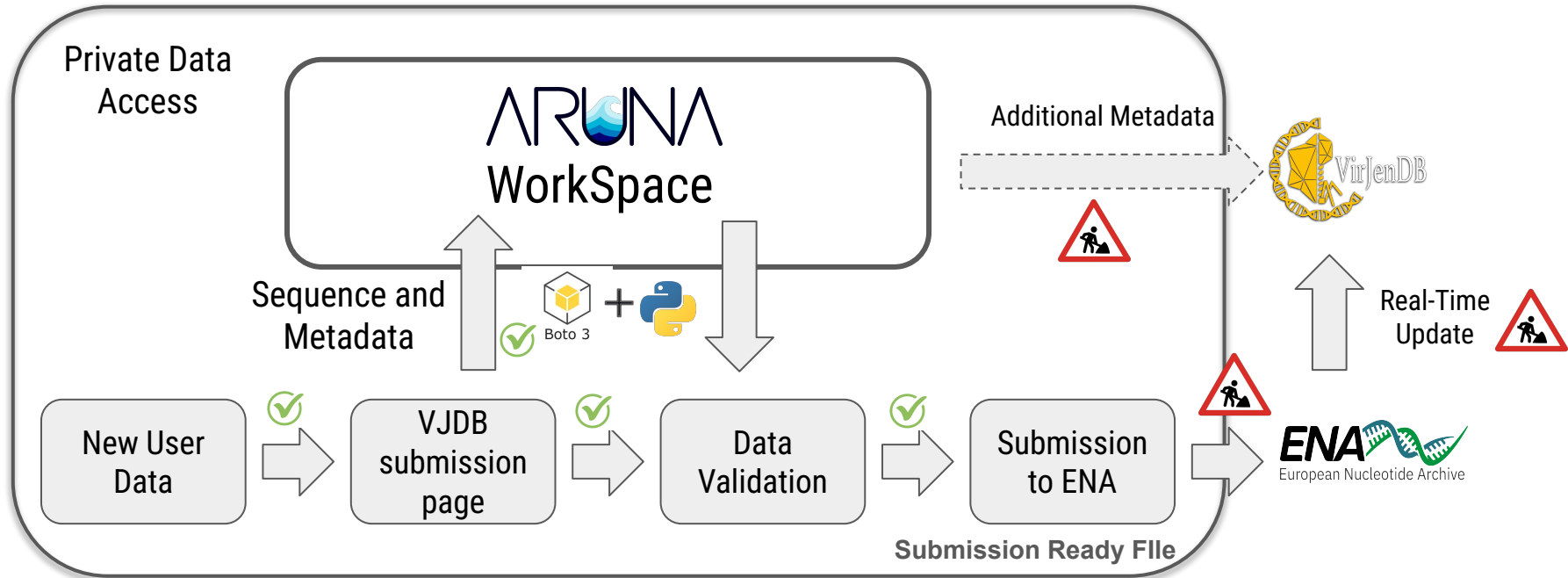
- Reduces Errors
 - Suggestion and autocomplete
 - Controlled Vocabularies
- Improves Data Quality
 - Wizard assistant
- Speeds Up Submission
 - Batch Submission
- Enhances Accessibility
- Supports FAIR Data Principles
 - Machine-Readable

The screenshot shows the VirjeDB submission wizard interface. The header includes the VirjeDB logo and navigation links: Search, Browse, Tools, Services, Stats, and About. The wizard is divided into two tabs: 'Wizard' (active) and 'Sample'. The 'Wizard' tab contains four steps:

- What do you want to do?** with options: 'Submission from scratch' (selected) and 'Submission from save file'.
- What do you want to submit?** with checkboxes: 'Study/Project' (checked), 'Sample' (checked), 'Raw Reads/Run/Experiment' (unchecked), and 'Genome Assembly' (unchecked).
- Do you already have a STUDY/PROJECT in ENA and want to submit to it?** with 'X' and checkmark buttons.
- Do you want to upload a influenza virus sequence?** with 'X' and checkmark buttons.

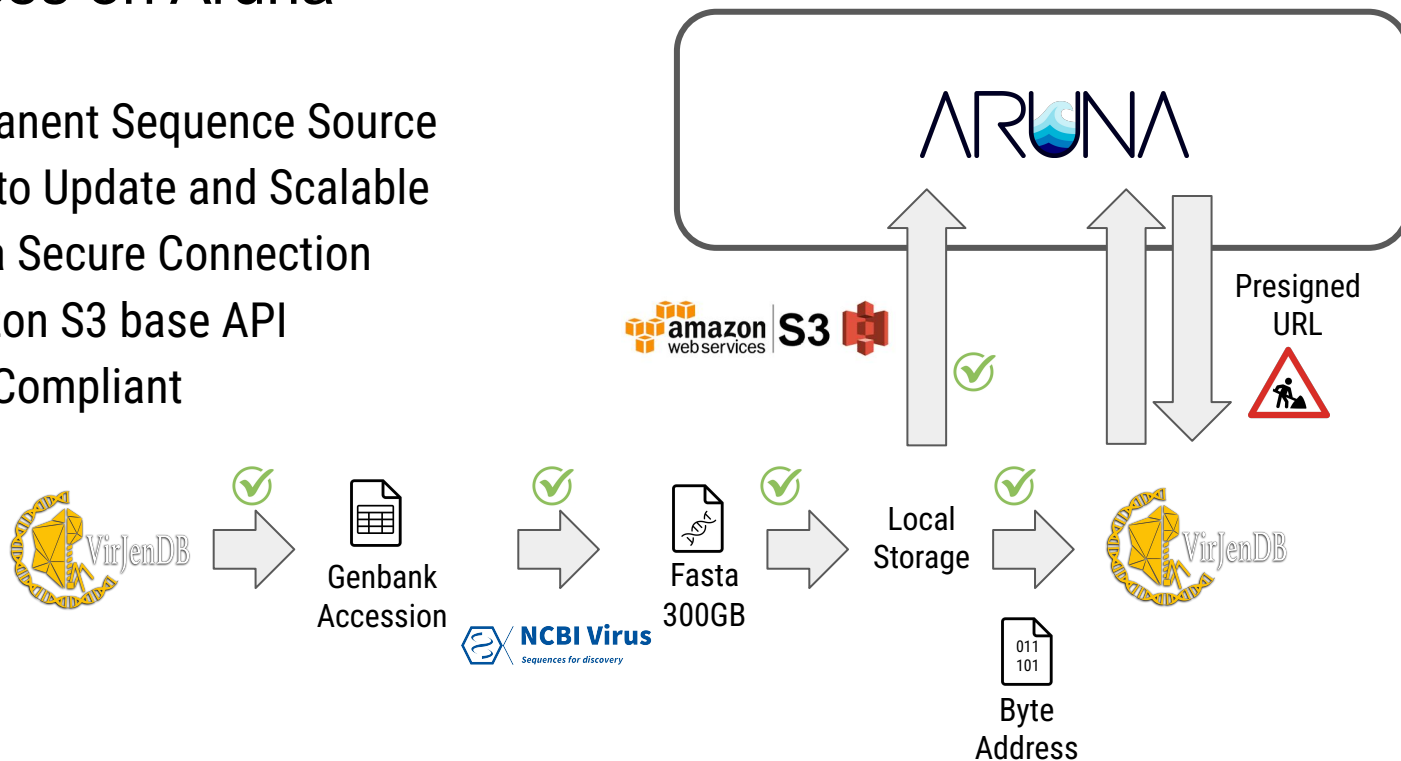
A yellow sidebar on the left contains icons for a warning (!), download, and search. The footer includes a feedback link and logos for NFDI4 INFRASTRUCTURE, DFG Deutsche Forschungsgemeinschaft, ARINA, UNIVERSITÄT JENA, and de NBI.

Pilot Submission Tool (BY-COVID)



Sequences on Aruna

- Permanent Sequence Source
- Easy to Update and Scalable
- Aruna Secure Connection
- Amazon S3 base API
- FAIR Compliant

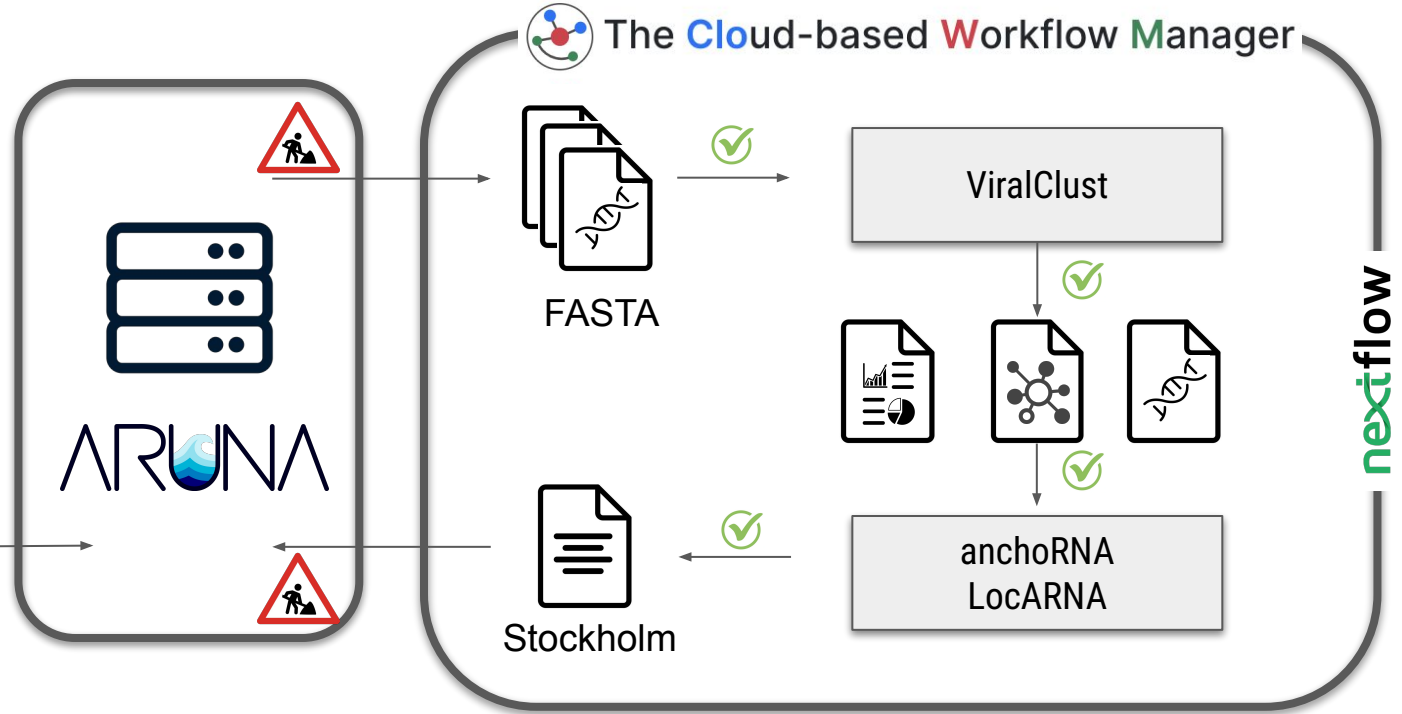


<https://arunastorage.github.io/documentation/v2.0.x/>

CloWM Access Workflow Diagram



Sandra Triebel
Kevin Lamkiewicz



<https://github.com/rmajena/viralclust>



Developer Team

Dr. Frank Förster

Jannis Hochmuth

Sebastian Beyvers

Lukas Fabian Brehm

Maria Hansen



Sarah Krautwurst

Sandra Triebel

Hamed Ziraksaz



Prof. Dr. Alexander Sczyrba

Dr. Sebastian Jünemann

Jan Krüger

Shahram Saghaei
PhD Candidate

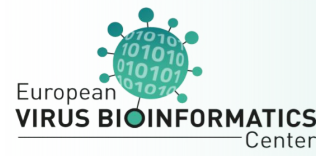
Kilian Ossetek
PhD Candidate

Dr. Noriko Cassman
Junior Group Leader

Prof. Dr. Manja Marz
Principal Investigator



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA





FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Thanks for your attention!



We value your input!

Visit: <https://virjendb.org>

Contact:
virjendb@uni-jena.de

