FaceBase Support for NIH's Data Management and Sharing Policy

FaceBase 2025 Bootcamp for Users and Contributors

Planning for Data Management – Why bother?

Comply with new (2023) NIH Data Management and Sharing Plan requirements

Ensure data is reusable

- Standard data formats
- Rich metadata
- Standard ontologies

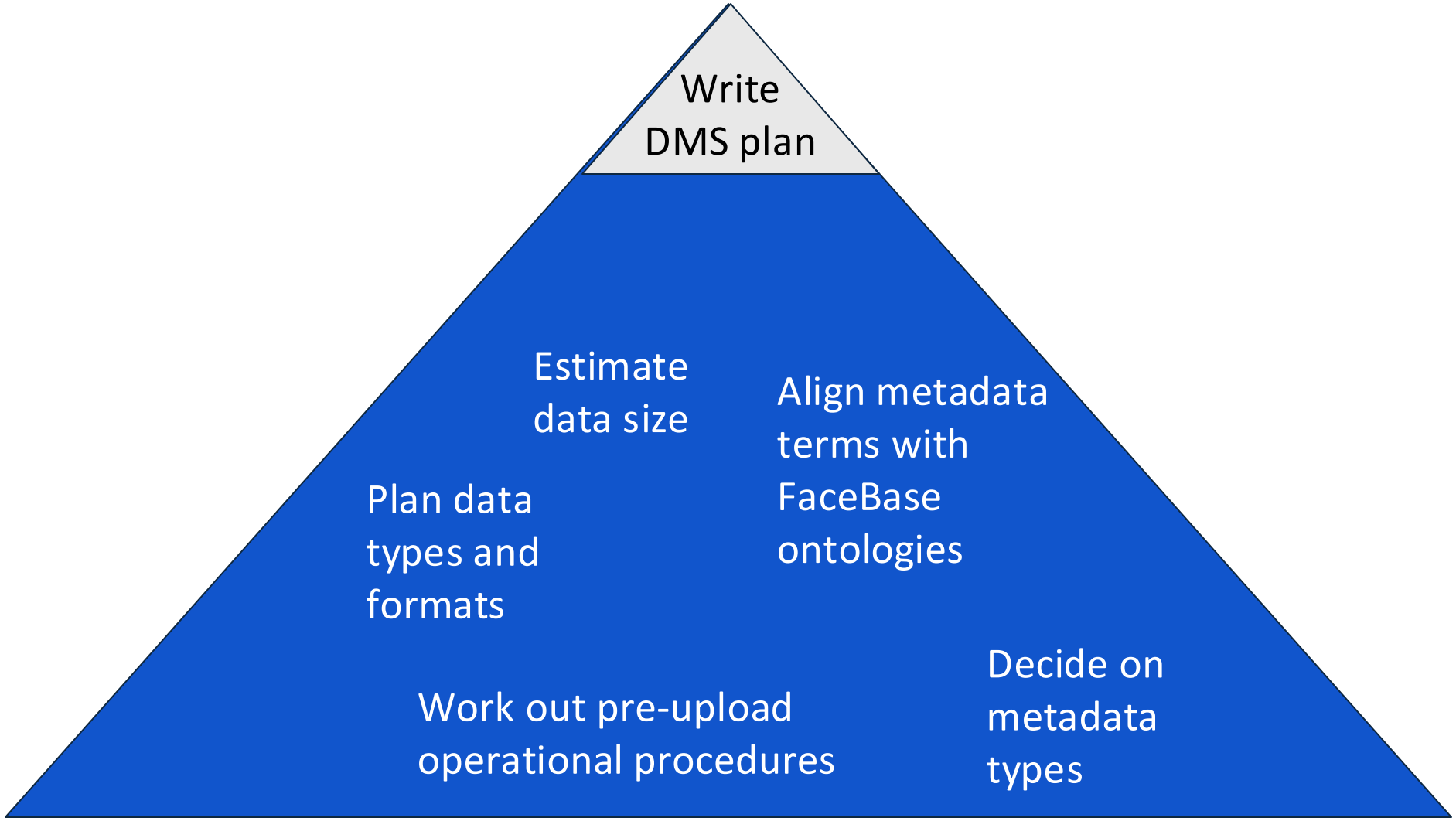
Understand and budget for the required level of effort

Elements of a DMS Plan

Data Type	Identifying data to be preserved and shared
Related tools, software & code	Tools and software needed to access and manipulate data. If needed specify how they can be accessed
Standards	Standards to be applied to scientific data and metadata
Data preservation, access & timelines	Repository to be used, persistent unique identifier, and when/how long data will be available
Access, distribution, reuse consideration	Description of factors for data access, distribution, or reuse
Oversight of data management and sharing	Plan compliance will be monitored/ managed and by whom

See NOT-OD-21-014 for more information (slide courtesy of Alicia Chou)

Planning for Data Management and Sharing

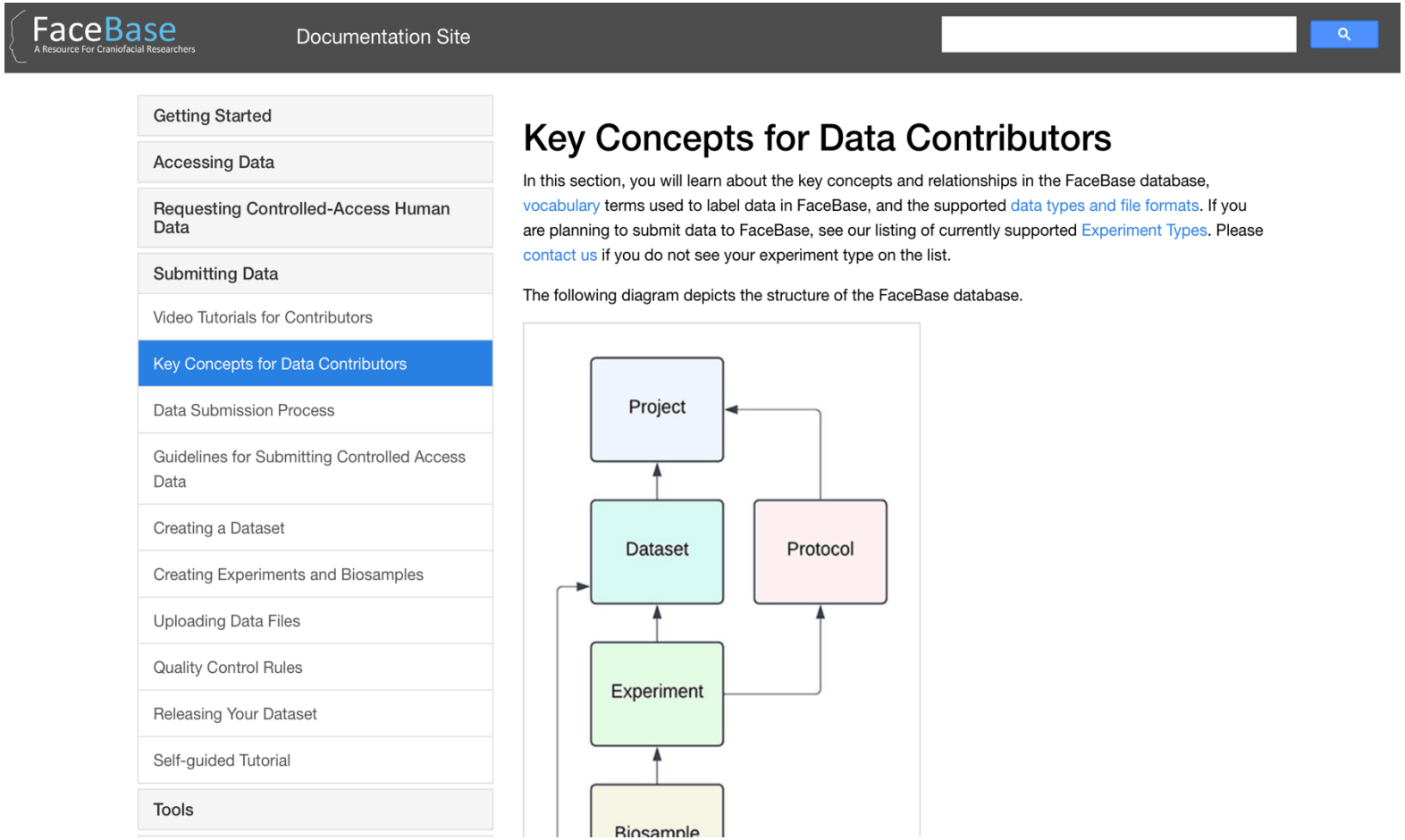


Data Management Considerations

- When, and how often, will you upload your data?
  - We encourage uploads early and often
  - But you'll need to have collected some data before it's approved for FaceBase
- How will you manage your data before it's uploaded?
  - How will you organize your data and metadata before you upload?
  - Do you have space for it?
  - Do you have security and backup policies in place?
- Will you need to do any preprocessing?
  - Does this involve tools you'll develop yourself?
  - If so, will you make it available on an open-source repository like github?

FaceBase-Specific Considerations

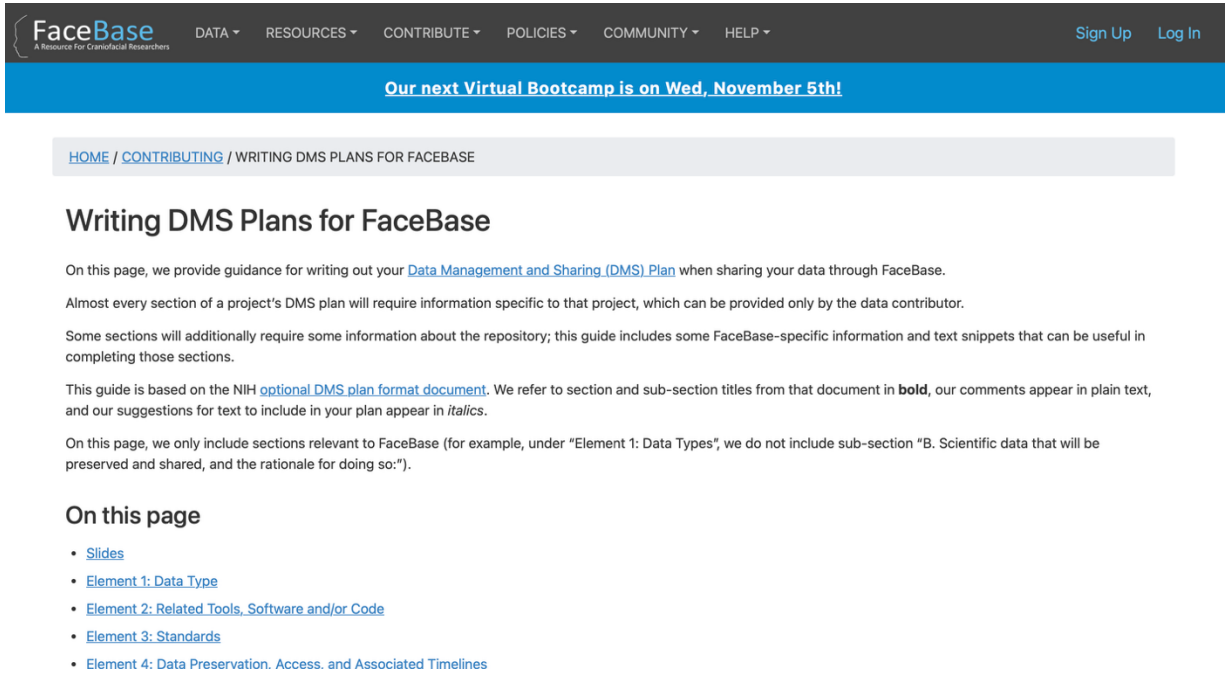
- The FaceBase repository currently supports certain:
  - Data Types
  - File Formats
  - Metadata elements
    - Expressed in specific ontologies
- If you have data/metadata that doesn't quite fit this model, talk to us!



https://docs.facebase.org/docs/Data-Submission-Key-Concepts/

Data Planning

Choose a data repository – is FaceBase the right fit?	FaceBase mission FaceBase data priorities
Consider some data issues	Data types and formats Metadata types and ontologies Pre-upload data management
Write data management plan	Some parts only you can answer FaceBase provides some boilerplate text •Complete answers to some questions •Snippets that can be used in others



https://www.facebase.org/contributing/dms

Data Types (Not an Exhaustive List)

See “Key Concepts for Data Contributors” for a full list of currently supported data and experiment types and species.

- Some currently supported data types:
  - Sequencing data (and derived processed and track data)
  - Imaging data (2D or 3D)
  - Surface/Mesh data
- Species: human, mouse, zebrafish, chick, xenopus
- Some experiment types:
  - Tomography/MRI
  - Gene Expression
  - Epigenetics
  - Microscopy

Data Formats

- Some Supported Data Formats:
  - Sequencing Data:
    - Raw: fastq
    - Processed: fastqc, count, tpm, fpkm, bam, bai, and measures in tsv
    - Track Data: BED (.bed), bigBed (.bb), and bigWig (.bw)
  - Imaging Data: TIFF, OME-TIFF, NIfTI
  - Surface Model / Mesh Data: Wavefront OBJ
- Considerations for new formats:
  - Is the format “open” (via standards or de facto openness)?
  - Are free or widely used tools available?

Metadata

- FaceBase has some minimum metadata requirements
  - Protocols for each experiment (we recommend the Nature Protocol Exchange format)
  - Species, developmental stage, and anatomy for each biosample
  - Additional requirements depending on data type
- FaceBase supports many more optional metadata elements
  - The more metadata you provide, the more discoverable and reproduceable your data will be
  - If you want to provide more than we currently collect, we'll probably accommodate those too.
- How will you express the metadata?
  - FaceBase uses standard ontologies for different metadata types

Some Ontologies Used by FaceBase

- Anatomy: UBERON
- Chromatin modifier: ZFIN, NGI, HGNC, Ensemble, MGI
- Data type: OBI, SMOMEDCT, CHMO
- Experiment type: MMO, ERO, CHMO, SCTID, OBI, STATO
- Gene: NCBI
- Phenotype: chmo, cmmo, fma, MP, HP, DOID
- Sex: UBERON
- Species: NCBI Taxon
- Strain: MGI
- Syndrome: MONDO
- Transcription factor: MGI, ZFIN, Gene_ORFName, Ensembl, HGNC

Adding New Ontologies

- We add new ontologies when necessary.
- Considerations:
  - Is the ontology standardized and widely used?
  - Does the DOC community generally agree that it's a good fit?
  - Does it overlap with currently-supported FaceBase ontologies?

Writing the DMS Plan

Useful resources:

- NIH “Writing a Data Management and Sharing Plan”
  - Includes format and sample plans
  - Includes links to additional requirements for specific institutes, programs, and offices
- “Writing a DMS Plan for FaceBase”
  - Based on NIH DMS plan format

DMS Plan - Element 1: Data Type

Question	Contributor activity	FaceBase provides
Types and amount of data to be generated	Provides entire answer	
Types and amount of data to be shared (and rationale)	Provides bulk of answer	Boilerplate text snippets about supported data types.
Metadata, other relevant data, and associated documentation	Writes answer using FaceBase-provided text snippets based on their data and metadata.	Boilerplate text snippets about required and optional metadata.

DMS Plan - Element 2: Related Tools, Software and/or Code:

Question	Contributor activity	FaceBase provides
Specialized tools, software, and/or code needed to access or manipulate shared scientific data, and how they can be accessed.	Provides bulk of answer	Boilerplate text about FaceBase tools for visualizing and annotating various types of data, which the contributor can include if relevant.

DMS Plan - Element 3: Standards

Question	Contributor activity	FaceBase provides
State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and how these data standards will be applied	Constructs answer from text snippets for all relevant metadata types.	Text snippets for the standards that FaceBase supports.

DMS Plan - Element 4: Data Preservation, Access, and Associated Timelines

Question	Contributor activity	FaceBase provides
Repository where data and metadata will be archived		FaceBase: www.facebase.org
How scientific data will be findable and identifiable		Text about persistent record IDs generated by FaceBase and optional DataCite DOIs.
When and how long the data will be made available.	"When" - either after curation or after publication.	"How long" text

DMS Plan - Element 5: Access, Distribution, or Reuse Considerations

Question	Contributor activity	FaceBase provides
Factors affecting subsequent access, distribution, or reuse of scientific data	Provides entire answer	
Whether access to scientific data will be controlled	Provides bulk of answer	Some text about sharing of public and protected human data
Protections for privacy, rights, and confidentiality of human research participants	Provides entire answer	

DMS Plan - Element 6: Oversight of Data Management and Sharing

Question	Contributor activity	FaceBase provides
Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).	Provides text specific to their project	Partial (FaceBase-specific) text.

Further Resources

- NIH “Writing a Data Management & Sharing Plan”: https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan
- FaceBase “Key Concepts for Contributors”: https://docs.facebase.org/docs/Data-Submission-Key-Concepts/
- “Writing a DMS Plan for FaceBase”: https://www.facebase.org/contributing/dms/
- FaceBase monthly office hours
- help@facebase.org
- Contact us at any point in this process, but please fill out our “Request to Submit Your Data to FaceBase” form once you’re funded!