DM Vitals – Use Case Scenario – Prof 123

Field: Molecular Biology

File formats:

* BAM – disciplinary standard open source
* VCF – curated annotated version of the file
* XLS

File sizes and volume:

* Has large files up to 100 GB 🡪100-1,000 files

Data ownership:

* For his projects he uses his own samples and owns the data
* For consortium projects unsure about data ownership
* Is not aware of campus data policies
* Uses an electronic notebook

Data backup and organization:

* Stores data on campus IT systems and office computer
* Uses as structured directory
* Uses a file naming standard
* Projects use a basic data management plan

Collaboration

* Uses campus tools for collaborating internally
* Files are too large to email
* Only shares “small files” externally
* Uses versioning software

Documentation

* Documents algorithms and processes
* Does not use a metadata standard

Preservation

* Has lost files recently
* Do not have old or obsolete formats
* Currently keeps all forms of the data from raw to analyzed