**Blake’s Recommendation for Date Format**: ***YYYYMMDD***

Data Definitions:

**Metadata Folder:**

This folder contains data from Baylor. The prefix “P-XXXX” in the file names indicate different batch numbers (pool number).

***Columns*:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Metadata | SampleID\_samplesheet | String, TEXT | These are the labels of the bottle samples from WWTP sites | (need Mike’s input) |
| Site | String, TEXT | Names of the WWTP | Lubbock, Wichita and Brownsville don’t have site names yet |
| Collection Date | Date  Format: m/dd/yyyy | Date that the samples arrive at the lab | (need Mike’s input) |
| Flow Rate | Real Number, Float, Decimal |  |  |

**qPCR \_coverage Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| qPCR | Sample\_ID | String, TEXT |  |  |
| CMMR\_Barcode | String, TEXT |  |  |
| Target | String, TEXT |  |  |
| Ct | Real Number,  Double |  |  |
| copiesperml | Real Number, Float, Decimal |  |  |
| SampleName | String, TEXT |  |  |
| date\_of\_collection | Date  Format: m/dd/yyyy |  |  |

**Genome\_coverage Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Genome\_coverage | Sample\_ID | String, TEXT |  |  |
| accession | String, TEXT |  |  |
| Start\_base | Integer |  |  |
| End\_base | Integer |  |  |
| Mean\_depth | Real Number, Float, Decimal |  |  |

**Site\_coding Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Site\_coding | Name | String, TEXT |  |  |
| Code | String, TEXT |  |  |

**Taxonomical\_profiles Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| taxonomical\_profiles | accession | String, TEXT |  |  |
| reference\_length | Real Number,  Integer |  |  |
|  | covered\_base | Real Number,  Integer |  |  |
|  | reads\_aligned | Real Number,  Integer |  |  |
|  | mean\_coverage | Real Number,  double |  |  |
|  | RPKM | Real Number,  double |  |  |
|  | sequence\_name | String, TEXT |  |  |
|  | taxid | String, TEXT |  |  |
|  | kingdom | String, TEXT |  |  |
|  | phylum | String, TEXT |  |  |
|  | class\_var | String, TEXT |  |  |
|  | order\_var | String, TEXT |  |  |
|  | family | String, TEXT |  |  |
|  | genus | String, TEXT |  |  |
|  | species | String, TEXT |  |  |
|  | subspecies | String, TEXT |  |  |
|  | strain | String, TEXT |  |  |
|  | sample\_ID | String, TEXT |  |  |
|  | seq\_pool\_ID | String, TEXT |  |  |
|  | total\_filtered\_reads\_in\_sample | Real Number,  double |  |  |

**WWTP\_Location Folder:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Folder | Column Names | Data Type | Description | Comments |
| Wwtp\_location | State | String, TEXT |  |  |
| County | String, TEXT |  |  |
| City | String, TEXT |  |  |
| WWTP | String, TEXT |  |  |
| lat | Real Number, Float, Decimal |  |  |
|  | lon | Real Number, Float, Decimal |  |  |
|  | county\_centroid\_lat | Real Number, Float, Decimal |  |  |
|  | county\_centroid\_lon | Real Number, Float, Decimal |  |  |
|  | city\_centroid\_lat | Real Number, Float, Decimal |  |  |
|  | city\_centroid\_lon | Real Number, Float, Decimal |  |  |