|  |
| --- |
| Form 3: Data Collection  |
| Purpose and scope of the data collected | *Provide a short description of the purpose and the scope of the data collection exercise* |
| Date | *Insert dates when the data were collected* |
| Person that completed this form | *Provide the name and details of the person who was responsible for completing this form* |
| SOPs used | *Indicate the SOPs used for the data collection (including the version numbers)* |

**Individuals involved**

*List here the key individuals involved in data collection, planning, coordination, external and internal quality control, etc.:*

|  |  |  |  |
| --- | --- | --- | --- |
| **Name**  | **Contact**  | **Institution**  | **Role for data collection**  |
| *Name* | *Email address and/or phone number*  | *Institution name*  | Coordinator |
| *Name* | *Email address and/or phone number*  | *Institution name*  | Trainer |
| *Name* | *Email address and/or phone number*  | *Institution name*  | Sample interpretation |
| *Name* | *Email address and/or phone number*  | *Institution name*  | Sample interpretation |
| *Name* | *Email address and/or phone number*  | *Institution name*  | Etc. |

**Sample unit allocation to interpreters**

*Provide an overview of sample unit allocation to interpreters*

|  |  |  |  |
| --- | --- | --- | --- |
| Number of sample units | Interpreter name  | File name  | File archive location  |
| *X sample* units | *Interpreter 1*  | *collection\_data\_date[year/month/day]\_version number.csv*  | *Link to cloud storage or folder path to repository*  |

**Cross validation**

*Describe how many duplicate samples were assessed and how these were chosen. Explain for multiple assessments how the reference interpretation is decided, particularly in the case of a tiebreaker.*

*For each interpreter, provide the following confusion matrix:*

|  |  |  |  |
| --- | --- | --- | --- |
|   | *Class 1 (reference)*  | *Class 2 (reference)*  | *Class k (reference)*  |
| *Class 1 (interpreter)*  | *Counts of sample points*  | *Counts of sample points*  | *Counts of sample points*  |
| *Class 2 (interpreter)*  | *Counts of sample points*  | *Counts of sample points*  | *Counts of sample points*  |
| *Class k (interpreter)*  | *Counts of sample points*  | *Counts of sample points*  | *Counts of sample points*  |

*Provide the following overview of overall agreement for each interpreter*

|  |  |
| --- | --- |
| Interpreter  | Overall agreement  |
| *Interpreter 1*  | *Sum of counts in all of the diagonal cells/ Sum of all counts*  |
| *Interpreter 2* | *Sum of counts in all of the diagonal cells/ Sum of all counts*  |
| *Interpreter n* | *Sum of counts in all of the diagonal cells/ Sum of all counts*  |

*Provide the following overview of per-class agreements:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|   | All interpreters agreeing  | One interpreter disagreeing  | Two interpreters disagreeing  | Etc.  |
| *Class 1 (reference)*  | *Percentage*  | *Percentage*  | *Percentage*  | *Percentage*  |
| *Class 2 (reference)*  | *Percentage*  | *Percentage*  | *Percentage*  | *Percentage*  |
| *Class k (reference)*  | *Percentage*  | *Percentage*  | *Percentage*  | *Percentage*  |
| Total  | *Percentage*  | *Percentage*  | *Percentage*  | *Percentage*  |

*Analyze the per-class agreement amongst interpreters*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | All interpreters agreeing | One interpreter disagreeing | Two interpreters disagreeing | *Etc.* |
| *Class 1 (reference)* | *Percentage* | *Percentage* | *Percentage* | *Percentage* |
| *Class 2 (reference)* | *Percentage* | *Percentage* | *Percentage* | *Percentage* |
| *Class 3 (reference)* | *Percentage* | *Percentage* | *Percentage* | *Percentage* |
| Total | *Percentage* | *Percentage* | *Percentage* | *Percentage* |