Interactive Serology Plate Inspector: A Guide for Data Uploading

Background and Objective

The Interactive Serology Plate Inspector (I-SPI) developed at Immunoodle, is designed to streamline quality control and quality assurance for multiplex bead assays starting from raw assay data. Raw bead array data, including assays from Luminex, require careful preparation for the data to be processed into a tidy way. The purpose of this document is to provide a guide to ensure data is processed into the correct way and show how inconsistencies are revealed when using I-SPI.

The Interactive Serology Plate Inspector is connected to a PostgreSQL database which stores data in a format in Data Science coined by Hadley Wickham known as tidy data (Wickham). Most datasets are not tidy, including those from raw assay outputs. In the I-SPI we developed interactive methods to parse the datasets uploaded into tidy data which can be stored in relational databases. Despite this, data needs to be handled with care ensuring the data parsers handle the data in the correct way.

Anatomy of Raw Assay Data

Raw assay data from multiplex bead assays consists of two primary components, the plate metadata and the raw assay MFI measurements and bead count values for antigens by specimen type on the plate. The plate metadata includes a file name consisting of a file path with key information including the study accession, experiment accession, and plate number. A shortened Plate ID is also in the metadata as well as the acquisition date (in the form of day-month-year, time), reader serial number, RP1 volts, and RP1 target. Additional information is not currently processed in I-SPI.

In the main dataset with MFI values, the specimen type (in the type column) must be a string from one of the following options: X (samples), S (standard curve samples), C (controls), and B (blanks).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	**File Name: C:\Users\BIO-PLEX\GRP_IMI_ERASME\TEST\Maternal Immunity\IgG subclasses\Bulk subclasses - May-June 2025\IgG2\IgG2 plates 1-5 (30-07-25)\TESTSTUDY IgG2 Plate 9 06082025.rbx**														
2	**Acquisition Date: 06-Aug-2025, 02:04 PM**														
3	**Reader Serial Number: LX10000000000**														
4	**RP1 PMT (Volts): 508.18**														
5	**RP1 Target: 3652**														
6	**Plate ID: TESTSTUDY_IgG2_plate_9_06082025**														
7															
8	**Well**	**Type**	**Description**	**PT (75)**	**FHA (27)**	**PRN (30)**	**TT (53)**	**DT (78)**	**ACT (42)**	**Pentamer (22)**	**IPV1 (19)**	**IPV2 (64)**	**IPV3 (36)**	**% Agg Beads**	**Sampling Errors**
9	A1	X1	151_TT_wP_prevacc_5000	42.0 (87)	60.0 (50)	211.5 (88)	115.0 (95)	74.5 (126)	91.0 (117)	74.0 (96)	20.0 (108)	20.0 (110)	24.5 (104)	8	
10	B1	X13	152_TdaP_wP_prevacc_5000	723.5 (50)	79.5 (52)	1988.0 (69)	1337.0 (77)	666.0 (69)	1004.0 (88)	550.0 (72)	44.0 (63)	35.0 (79)	121.0 (77)	9	
11	C1	X25	167_TdaP_aP_prevacc_5000	50.0 (61)	243.5 (50)	131.0 (56)	1189.0 (72)	65.0 (83)	88.0 (98)	47.0 (69)	16.0 (102)	82.0 (95)	97.0 (74)	8	
12	D1	X37	174_TdaP_wP_prevacc_5000	56.5 (50)	228.0 (55)	383.0 (62)	143.5 (68)	155.0 (66)	165.5 (70)	137.0 (67)	33.0 (84)	41.0 (86)	51.5 (62)	11	
13	E1	X49	169_TT_wP_prevacc_5000	43.0 (63)	187.0 (50)	175.5 (64)	702.0 (93)	37.0 (87)	75.0 (80)	33.5 (82)	16.0 (75)	16.0 (100)	23.0 (96)	10	

Figure 1: An example of a raw file from a multiplex bead assay including the plate metadata bolded in the first 6 rows and the beginning of the raw assay measurements with a header beginning on row 8. The first 3 columns of the main dataset are consistent across all plates and are Well, Type, and Description. The next columns are the antigens on the plate with the bead region associated with the antigen uniquely identifying the headset being read by the

detector with the antigen. The last 2 columns which follow the antigen columns are the % Agg Beads and the Sampling Errors.

Preparing Assay Data for Smooth Parsing

The primary dataset for the plate from the raw assay data contains a column named "description". This column is not tidy since it contains specific information that corresponds to the type of column and can contain a combination of distinct information. When more than one piece of information is present in the description it is delaminated with an underscore.

Type X contains the patient id, study group, timepoint and serum dilution.

Type S contains the standard curve source followed by the dilution.

Type C contains the source of the control followed by the dilution.

Type B contains the source of the blanks followed by the dilution.

For the controls (Type C), examples include P2 and P5 where P is short for Pooled. The numbers are not dilutions but rather different pooled samples. This can be dependent on the lab protocols as well. A string that does not contain an underscore is suggested so data does not separate in unexpected ways during parsing and assigning source to the controls.

The patient identifier is a numeric integer as well as the dilutions in each of the specimen types outlined above. The study group, timepoint, and all the sources are characters and not solely numbers. I-SPI is designed to handle serum dilutions ranging from 1 to 100000.

To ensure a bead count is processed for the antigens the reported bead count should be in parenthesis following the name of the antigen. I-SPI will store this information to the PostgreSQL database.

When the Interactive Serology Plate Inspector parses data it separates the description field into multiple columns, so each variable is in its assigned column. If there are additional underscores in the name of a field such as the source which takes one component this separates out the source into two columns which increases difficulty in source assignment. It is recommended that unneeded underscores are avoided.

The following are examples of descriptions of sources that will pass the plate validation step in the application: SD_20, NIBSC_120, NIBSC140_20. In the last case, NIBSC140 will be identified as the source and the standard curve serum dilution will be identified as 20.

Examples of descriptions that will cause downstream issues in source assignment include NIBSCO_140_20. In this case, there is an additional underscore resulting in three components instead of two when the value is split via the underscore. When assigning the source, it may be

NIBSCO instead of the intended NIBSCO_140. A similar result occurs if there is a slash in the source such as in NIBSCO/140_20.

Handling Missing Information

In the raw spreadsheet for the plate, it is important for information to be filled into the description column. Blanks can still be processed if assignment of the source and dilutions are assigned to columns with data. The dilution for the blanks should be set to 1. If the dilutions are not 1 the records can be modified in the data parser after assignment of the columns corresponding to the data and before uploading to the PostgreSQL database.

In addition to the dilution of the blanks, for the cleanest data uploading, ensure that the description contains the relevant information including the source, such as PBS, and that it follows the outline provided in the previous section.

When samples are missing and blanks are used in wells of the plates instead of samples, they should be treated as blanks. Figure 2 shows an example where blanks are used in sample wells before correcting for importing. To do this change the X_n to a B_n , where *n* is an integer, in the type column and use the description format for Blanks outlined in "Preparing Assay Data for Smooth Parsing" (Figure 3).

Well	Type	Description	PT (75)	FHA (27)	PRN (30)	TT (53)	DT (78)	ACT (42)	Pentamer (22)	IPV1 (19)	IPV2 (64)	IPV3 (36)	% Agg Beads	Sampling Errors
A12	X12	Blank	12.0 (51)	608.0 (71)	24.0 (53)	6.0 (50)	17.0 (55)	23.0 (88)	10.0 (76)	11.0 (64)	7.0 (56)	5.5 (82)	12	
B12	X24	Blank	10.0 (68)	433.5 (50)	26.0 (63)	5.0 (55)	19.5 (80)	28.0 (109)	7.0 (79)	9.0 (66)	8.0 (84)	5.0 (81)	12	
C12	X36	Blank	13.0 (64)	361.0 (50)	42.0 (73)	6.0 (75)	20.0 (92)	29.0 (94)	8.0 (66)	10.5 (84)	9.0 (70)	7.0 (70)	12	

Figure 2: An example that will not upload. The blanks are used in sample wells but are of type X. Description is also Blank and does not correspond to a valid description for a type.

Well	Type	Description	PT (75)	FHA (27)	PRN (30)	TT (53)	DT (78)	ACT (42)	Pentamer (22)	IPV1 (19)	IPV2 (64)	IPV3 (36)	% Agg Beads	Sampling Errors
A12	B12	PBS_1	12.0 (51)	608.0 (71)	24.0 (53)	6.0 (50)	17.0 (55)	23.0 (88)	10.0 (76)	11.0 (64)	7.0 (56)	5.5 (82)	12	
B12	B24	PBS_1	10.0 (68)	433.5 (50)	26.0 (63)	5.0 (55)	19.5 (80)	28.0 (109)	7.0 (79)	9.0 (66)	8.0 (84)	5.0 (81)	12	
C12	B36	PBS_1	13.0 (64)	361.0 (50)	42.0 (73)	6.0 (75)	20.0 (92)	29.0 (94)	8.0 (66)	10.5 (84)	9.0 (70)	7.0 (70)	12	

Figure 3: An example of correcting data in Figure 2 for successful data upload. The Blanks are assigned B followed by a unique number in the type column and the description have the source of the blank (PBS) followed by the dilution of 1.

Ensure that to the right of the antigen's MFI value there the bead count in the form of an integer for the antigen in parenthesis. If this is missing a bead count would be missing for the antigen in that well.

Assigning and Saving Data

When data is parsed and assigned if multiple columns are selected, which can be the case for assigning timepoints and study group, underscores are used to separate that information. When saving components the strings created in the assignment process must have a length less than or equal to the maximum length that is indicated for that variable.

Plate Validator

To validate a plate before completing the upload I-SPI, the file can be validated using the rules outlined. The following is pseudocode for the validator.

Validate the plate metadata:

> For each line in the plate metadata create key-value pairs where the key is the text up to the first colon and the value is the text after the colon.
>
> For the file name, verify the file name is a valid file path and contains the study name, experiment name and plate number.
>
> Verifies that the acquisition date is in the form day number-3-character month abbreviation- 4-digit year number, time with AM or PM. The comma and hyphen are included. Example: 01-Oct-2025, 4:27 PM
>
> The RP1 PMT (Volts) and the RP1 Target keys must have a numeric value. Decimal and integer values are allowed.
>
> The Plate ID must be a character string.

If one of these verification steps does not pass, information is provided in a table to identify where in the plate metadata corrections need to be applied.

Validate the Primary Dataset:

> For each record, verify that the bead count is an integer that follows the MFI value in a set of parentheses. Typically, there is a space between the MFI reading and the beginning of the parenthesis, but this is not required.
>
> Verify that the data in the type column begins with an X, S, C, or B followed by an integer.
>
> Based on the type validate the description based on the structural requirements for that type. Specifically verify that:
>
> > Type X contains the patient id as an integer, study group, timepoint and serum dilution as an integer separated by underscores.

Type S contains the standard curve source (character string) followed by the dilution (integer) all separated by underscores.

Type C contains the source of the control (character string) followed by the dilution (integer) all separated by underscores.

Type B contains the source of the blanks (character string) followed by the dilution (integer) separated by an underscore.

For each of the above criteria that are not satisfied when a plate is selected, a message will be included in a table that describes where in the raw file corrections need to be applied.

When all previous checks are satisfied:

If the keyword Blank is in the description, do not allow uploading. Provide a choice of a set of keywords to place in the description: empty_well indicates the well is skipped in uploading. use_as_blank indicates it will be used as a blank and assigned to type B.

Proceed to plate uploading once all checks are validated.

Plate Uploading Parser

After the plate is validated I-SPI allows the plate metadata and then the plate data to be parsed and uploaded to the database. The first step is uploading the plate metadata and following that the different specimen types present on the plate (X, S, C, and B). When uploading the plate metadata, I-SPI automatically parses the plate metadata to automate uploading by extracting each of the components in the metadata. If missing or not in the correct form the user can manually edit the table to fill in the correct information. The sample dilution factor must lie between 1 and 100000 (inclusive) and the plate variable must be a string in the form plate_n where n ranges between 1 and 99 and optionally is followed by a lowercase letter (a-z). The lowercase letter is important when a plate is rerun, and the plate has been already uploaded as plates must be unique in an experiment. The following are examples of the table that will upload and will not upload.

variable	value
file_name	C:\Users\BIO-PLEX\GRP_IMI_ERASME\TEST\Maternal Immunity\IgG subclasses\Bulk subclasses - May-June 2025\IgG2\IgG2 plates 1-5 (30-07-25)\TEST IgG2 Plate 9 06082025.rbx
acquisition_date	06-Aug-2025, 02:04 PM
reader_serial_number	LX10022082422
rp1_pmt_volts	508.18
rp1_target	3652
plateid	TEST_IgG2_plate_9_06082025
plate_id	C:\Users\BIO-PLEX\GRP_IMI_ERASME\TEST\MaternalImmunity\IgGsubclasses\Bulksubclasses-May-June2025\IgG2\IgG2plates1-5(30-07-25)\TESTIgG2Plate906082025.rbx
plate	
sample_dilution_factor	
study_accession	validator
experiment_accession	IgG2
auth0_user	seamus.owen.stein@dartmouth.edu
workspace_id	51

Figure 4: An example of parsed metadata that will not upload to I-SPI. This metadata will not upload since the sample dilution factor variable has a blank value and the plate has a blank value. Blank values are highlighted in red. The user can either correct this in the raw file ensuring plate and the number is in lowercase and include a dilution factor in the file name or the easier way is to enter the appropriate plate value and the correct sample dilution factor. The study accession and experiment accession is read in from what the user has selected as the study and experiment from the UI in the application.

variable	value
file_name	C:\Users\BIO-PLEX\GRP_IMI_ERASME\TEST\Maternal Immunity\IgG subclasses\Bulk subclasses - May-June 2025\IgG2\IgG2 plates 1-5 (30-07-25)\TEST IgG2 Plate 9 06082025.rbx
acquisition_date	06-Aug-2025, 02:04 PM
reader_serial_number	LX10022082422
rp1_pmt_volts	508.18
rp1_target	3652
plateid	TEST_IgG2_plate_9_06082025
plate_id	C:\Users\BIO-PLEX\GRP_IMI_ERASME\TEST\MaternalImmunity\IgGsubclasses\Bulksubclasses-May-June2025\IgG2\IgG2plates1-5(30-07-25)\TESTIgG2Plate906082025.rbx
plate	plate_9
sample_dilution_factor	5000
study_accession	validator
experiment_accession	IgG2
auth0_user	seamus.owen.stein@dartmouth.edu
workspace_id	51

Figure 5: An example of parsed metadata that will upload to I-SPI. All values associated with the variables are correctly parsed and filled in. A user can correct editable fields that are not greyed out, such as the plate and sample dilution factor by typing the desired value into the field before uploading the plate metadata.

Data Inconsistencies Revealed by I-SPI

When there are inconsistencies in the data such as variations in the names of antigens and timepoints, they are revealed and can be seen places including study-level overview figures. When antigens have slightly different names including underscores or a variant number from another, they are treated as different though the intention may be that the antigens are the same. The same concept is applicable for timepoints.

Discussion

This document serves as a companion to the Interactive Serology Plate Inspector and outlines the importance of careful preparation and loading raw assay data from multiplex immunoassays into I-SPI and provides guidance so when uploaded data is stored in a consistent way following the practice of tidy data.

References

Wickham, H. . (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1–23. https://doi.org/10.18637/jss.v059.i10