

# MATRR

Monkey Alcohol Tissue Research Resource

Data Upload

# Notify MATRR

## Submit Research Updates

*Please provide any information available regarding your successful investigation. Only one field is required.*

<b>PMID</b> <i>PubMed ID number of publication</i>	<input type="text"/>
<b>Research Update</b> <i>File to Upload</i>	<input type="button" value="Choose File"/> VeryDetailedReport.docx
<b>Comments</b> <i>Please briefly describe the progress made thus far.</i>	<input type="text" value="Very important research"/>
<b>Submitted Grants</b> <i>Please describe any grant submissions resulting from the MATRR tissues. Include title, funding agency, expected review date, etc.</i>	<input type="text" value="Very fancy grants"/>
<b>Data Available</b>	<input checked="" type="checkbox"/> Data is available for upload to MATRR. Please contact me to arrange this integration into the MATRR.

Checkbox for "In Progress" and "Complete" updates

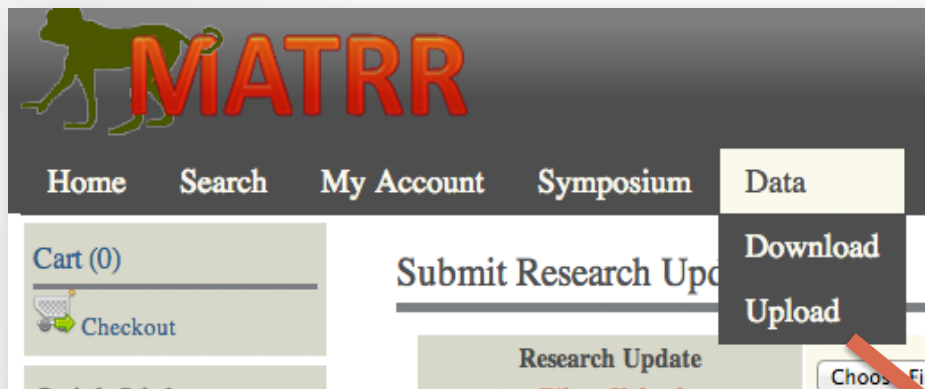
# Notify MATRR

Alternative methods of communication are also encouraged



# Upload Data

The MATRR admin will add 'Data' to your menu which gives you permission to upload files directly to the MATRR web server

A screenshot of the 'MATRR Data Upload' form. The form has a title 'MATRR Data Upload' and a subtitle 'You can upload your MATRR data here.' Below the subtitle, there is a 'Data:' label followed by a 'Choose File' button and the text 'No file chosen'. At the bottom of the form, there is an 'Upload' button. A red arrow points from the 'Upload' option in the 'Data' dropdown menu to the 'MATRR Data Upload' form.


# Upload Data

## What to submit to MATRR

- Format
  - Microsoft Excel, OpenOffice
  - Tab or Comma delimited files are my favorite
- All valid data
  - (In)significant data is still data
  - Include control's data
  - Only finalized data
- Instructions describing your data
  - Chances are we have no idea what your acronyms mean
  - We definitely don't know what the units are for the acronyms
  - If it's a calculated data type, please tell us how it's calculated
    - $\text{Concentration} = .04$
    - ethanol in mL, mass in kg
    - $\text{g/kg} = \text{Concentration} * \text{ethanol} / \text{mass}$

# Upload Data

## Please DO NOT Send

- Machine-unreadable data formats
  - ++Super Important++
  -  Colors, Sentences, Special Flags\*+, nested spreadsheets
- Images are hard to work with
  - I can't extract data from graphs, images, or PDFs
- Raw Data
  - On a case-by-case basis
  - Can be overwhelming and provide little articulation with existing data
- Viruses, malware, worms, key-loggers, backdoors....
  - Nobody wants to deal with a sick computer

# Dates

- Excel is a jerk.
- Localized date formats are inconsistent and easily misunderstood
  - 12/06/02 is not a date, it's 1:
    - Dec 6<sup>th</sup> 2002 in one country (US, usually)
    - Jun 12<sup>th</sup> 2012 others (France, Brazil, few others)
- Published in 1988, ISO 8601 exists for a reason
  - YYYY-MM-DD, with zero-padded months and days
  - Bonus: ISO dates alphabetically sort chronologically
- Excel is still a jerk
  - Prepend dates with a single quote ('YYYY-MM-DD) to override Excel's auto-magical date formatting.

# Columns and Rows

- Columns are variable types
- Rows are observations

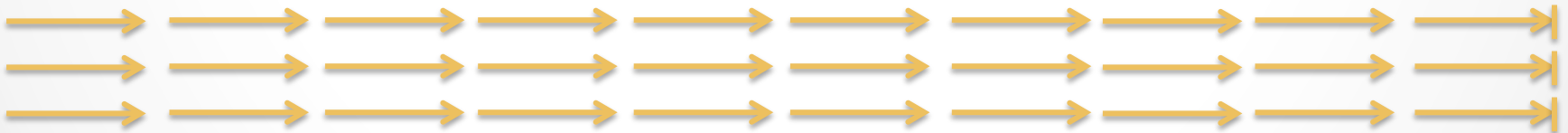
	A1CF	A2M	A4GALT	A4GNT	AAAS	AACS	AADAC	AADACL2	AADACL3
26077	0.01469	0.59021	4.67471	0	9.51025	14.29235	0	0	0.04726
26078	0.00423	0.66487	5.74714	0	9.69338	16.97913	0	0	0.02042
26082	0	0.85862	5.00389	0	7.40104	11.20429	0	0	0

This file has 16 rows and 16,045 columns. And that's fine.

- Computers read spreadsheets left-to-right until the end of a row



- Then a computer performs the \_exact\_ same actions on the next row



- Until the end of the file.



# Consistent Formatting

- Columns are always in the same order

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Animal	BLANK MATR	Collect Date	Run Date	Exp. Period	Exp Day	Sess Start	Time	weight	EtOH consum	Dose consum	22 Hour Dose	Diff sample t	BEC (mg%)
2	1050		5/1/09		22 hour			5:50:14	6.8142857	477	2.8	3.6		28
3	1051		5/1/09		22 hour			5:55:36	5.4181818	149	1.1	1.1		26
4	1055		5/1/09		22 hour			5:58:27	6.35	127	0.8	0.8		23
5	1058		5/1/09		22 hour			5:53:03	9.8	98	0.4	0.6		18
6	1059		5/1/09		22 hour			6:01:30	7.32	183	1	1.3		14
7	1050		5/6/09		22 hour			5:58:46	6.9	276	1.6	2.8		13



	A	B	C	D	E	F	G
1	Cohort	MATRR ID	Lifetime EtOH Intake (g/kg)	Lifetime EtOH Intake (mL)	Average BEC, Induction (mg pct)	Average BEC, Open Access (mg pct)	Average BEC, Lifetime (mg pct)
2	INIA Rhesus 4	10048	602.1496649	166930	57.74222222	27.10630435	35.72265625
3	INIA Rhesus 4	10049	1169.874458	295370.9	56.30722222	35.93688889	41.75698413
4	INIA Rhesus 4	10050					
5	INIA Rhesus 4	10051	1177.850023	296559.5	53.75888889	52.055	52.53421875
6	INIA Rhesus 4	10052	291.5542312	55867.3	24.79722222	2.368478261	8.6765625

- New/added variable types are appended to the right,
  - No matter what the new data is
  - This saves a lot of time/effort

# Colors and Fonts and etc.

- Colors and fonts are COMPLETELY ignored
  - Loadable data formats (CSV and tab-delimited text) don't even support these features
  - SOLUTION: Add more columns
- Special flags aren't helpful
  - " 101.15\* " isn't a number and will cause errors when loading data
  - SOLUTION: Add more columns
  - Don't merge or "ditto"-mark cells
- Consistency, consistency, consistency
  - "22hr" and "22 hour" aren't the same thing to a computer
    - Either is fine, but please choose one
- DRY principle (Don't repeat yourself)
  - "Every piece of knowledge must have a single, unambiguous, authoritative representation within a system"
  - I only need columns/features unique to the observation

- There should only be 1 header row per file
  - Differently-formatted data should be saved to a different file
- All rows should have the same number of cells

	A	B	C	D	E	F	G	H	I	J	K
1	<b>Control</b>										
2	NECROPSY DATE AND TIME	MONKEY ID #	BAL	Lifetime Intake (g/kg)	Lifetime Intake (ml)	File Name	Frequency (Hz)	IEI (ms)	Amplitude	Rise (ms)	Decay (ms)
3	4/22/09 AM	26212	0	0	0	09422006	1	2	3	4	5
4											
5											
6											
7	<b>Ethanol</b>										
8	NECROPSY DATE AND TIME	MONKEY ID #	BAL	Lifetime Intake (g/kg)	Lifetime Intake (ml)	File Name	Frequency (Hz)	IEI (ms)	Amplitude	Rise (ms)	Decay (ms)
9	4/13/09 AM	25086	150	2336.7	476966.5	09413004	1	2	3	4	5
10	4/13/09 AM	25086	150	2336.7	476966.5	09413008	14	15	16	17	18
11	4/15/09 AM	25091	139	2559.3							
12			92	2506							
13	4/15/09 PM	25090	92	2506							
14			92	2506	463043	09415023	66	67	68	69	70
15			266	4882.5	934244.4	09416003	79	80	81	82	83
16			266	4882.5	934244.4	09416009	92	93	94	95	96
17	4/16/09 AM	25093	266	4882.5	934244.4	09416012	105	106	107	108	109
18			266	4882.5	934244.4	09416018	118	119	120	121	122
19			112	2186	505701.7	09416027	131	132	133	134	135

Doesn't matter to a computer, all the columns are the same.

All duplicate headers, unnecessary and ill-advised

Don't merge cells! Leave them blank, or duplicate data

- Loadable file formats don't support most features
- Any notes/definitions should be in a separate file.

Shading isn't a readable feature

	A	B	C	D	E	F	G	H	I	J	K	L	M
						25882	26016	26077	26082	26089	26104	26148	26168
						0	0	0.0625	0	0.0625	0	0.007812	0
						0.001953	0	0	0	0	0	0	0.000488
4	25790	0	0.001556	1	0.000244	0.000122	0	0.000122	0	0	0.015625	0.015747	0
5	25811	0	0	0.000244	1	0.0625	0	0.0625	0	0.000488	0	0.066406	0
6	25882	0	0.001953	0.000122	0.0625	1	0	0.03125	0	0	0	0.03125	0.000488
7	26016	0	0	0	0	0	1	0	0	0	0	0	0.033203
8	26077	0.0625	0	0.000122	0.0625	0.03125	0	1	0	0.125	0	0.03125	0
9	26082	0	0	0	0	0	0	0	1	0	0	0	0
10	26089	0.0625	0	0	0.000488	0	0	0.125	0	1	0	0	0
11	26104	0	0	0.015625	0	0	0	0	0	1	0	0	0
12	26148	0.007812	0	0.015747	0.066406	0.03125	0	0.03125	0	0	0	1	0
13	26168	0	0.000488	0	0	0.000488	0.033203	0	0	0	0	0	1
14													
15													
16													
17													
18													
19													
20	<b>Relation</b>		<b>Kinship</b>										
21	Self/Monozygotic Twins		1										
22	Parent-Offspring		0.25										
23	Full-Siblings		0.25										
24	Half-Siblings		0.125										
25	Grandparent-Grandchild		0.125										
26	Uncle-nephew		0.125										
27	First cousins		0.0625										
28	Second cousins		0.015625										
29													

IDs in blue are the control animals

Kinship coefficients in red are at the first cousin level or closer

Kinship coefficients in purple are at the second cousin level or closer, up to first cousin.

Notes and nested spreadsheets

# Summary

## Do:

- Add as many columns as you could possibly want
- Format data consistently
- Name/label data consistently
- Describe your data, in detail
  - Separately, NOT in the data
- Use a standardized date format
- Send us ALL your final data

## Don't:

- Insert new columns between existing columns
- Rearrange columns
- Color/flag/mark data cells
- Merge cells together
- Merge spreadsheets together
- Send us images or graphs
- Send us unfinished data