

Supplemental Files in Electronic Theses and Dissertations: Implications for Policy and Practice

Sarah L. Shreeves - sshreeve@illinois.edu - @sshreeves
University of Illinois at Urbana-Champaign



IDCC 2013 – Amsterdam, Netherlands – 14-16 January 2013

ETDs and Supplemental Files at UIUC

- ETDs and supplemental files (in all formats) accepted since 2010.
- Must be approved by the advisor / committee.
- Must be described in an appendix.
- No standard for description in Grad College policy.

3538 ETDs deposited between 2010-05 and 2012-08

- 59% Doctoral Dissertations
- 41% Masters Theses
- 63% Immediately Open Access
- 20% Illinois only for 2 years
- 17% Closed for 2 years

Of these **77 (2%)** ETDs deposited with at least one supplemental file

- 47% Doctoral dissertations
- 53% Masters Theses
- 73% Immediately Open Access
- 19% Illinois only for 2 years
- 8% Closed for two years

- Currently investigating what policies should be in place, if any, for format, metadata, and other requirements for supplemental files.

Examples of Supplemental Files and Documentation in Appendix

```
FieldweedIdentify.m uses the all of the weed images obtained from laboratory
experiment to construct the SVM model. And such model is used to classify
the weed image captured in the field.
% Those weed image include:
% 1. barleygrass
% 2. crabgrass
% 3. Lambsquarters
% 4. Velvetleaf
% 5. waterhemp
% notice: 1. Corn is not include

% Author: chufan Lin
% Date: Oct 17th, 2009
clear all
close all
clc

nocorn = xlsread('C:\ClassifiedImages\FieldImages\wocorn.xls');
label1 = nocorn(:,16);
inst1 = nocorn(:,9:13);

% [bestcv1,bestc1,bestg1] = cvselect(label1,inst1);
% returned cvs and best (c,g):
% For nocorn data: best (c,g) = (4096, 0.0078), cv = 85.34%;
% Accuracy = 91.53%;
load model1.mat

time = [];
scrSz = get(0,'screenSize');
plant = 1;

date = 1;
while date <= 0;
    date = input('type in the date to be processed: ');
    if date == 0
        disp('procedure done, exiting...')
        close all
        break
    end
    dir_old = ['C:\ClassifiedImages\FieldImages\image 0' num2str(date) '\'];
    dir_new = ['C:\ClassifiedImages\FieldImages\images\image 0' num2str(date) '\'];

    if exist(num2str(dir_new),'file') == 0
        mkdir(num2str(dir_new)); % make a new directory
    end
end
```

Script replicated in Appendix

	A	B	C	D	E	F	G	H	I	
1	APPENDIX D: Equine Populations in Braunschweig-Lüneburg, 1535-1770									
37										
38	Amt Ilten	1535/6, H	1564, H				1585, Pf	1585, H		
39	Ahlten (ohne Heuflinge und Öffnen /		52 /				129	61		
40	Anderten (ohne Heuflinge)	48	49 /				91	52		
41	Bällem	33	36 /				113	46		
42	Dolgen		33 /				70	42		
43	Döhren									
44	Evern									
45	Gretenberg									
46	Harber									
47	Heymer (ohne der Heußl. Wiese)									
48	Höverde									
49	Ilten (ohne der Heußl. Wiese)									
50	Laatzten									
51	Lehete									
52	Klein Lopke (Amt Ilten)									
53	Retmar (Amt Ilten), Herrenleute									
54	Retmar (Amt Ilten), Junkerleute	8	8 /				[18]	[8]		
55	Selnde (Amt Ilten)	47	47 /				149	50		
56	Wulfel	26	30 /				56	33		
57										
58	Amt Medingen									
59	Addenstorf									
60	Alliam (Aljam)									
61	Almstorf									
62										

Appendix D: Equine Populations in Braunschweig-Lüneburg, 1535-1770
Given the particular utility of this Appendix as a digital document that can be easily modified to allow for speedy comparative work, it does not represent well in a fixed medium. Please see the supplemental file Appendix D deposit.

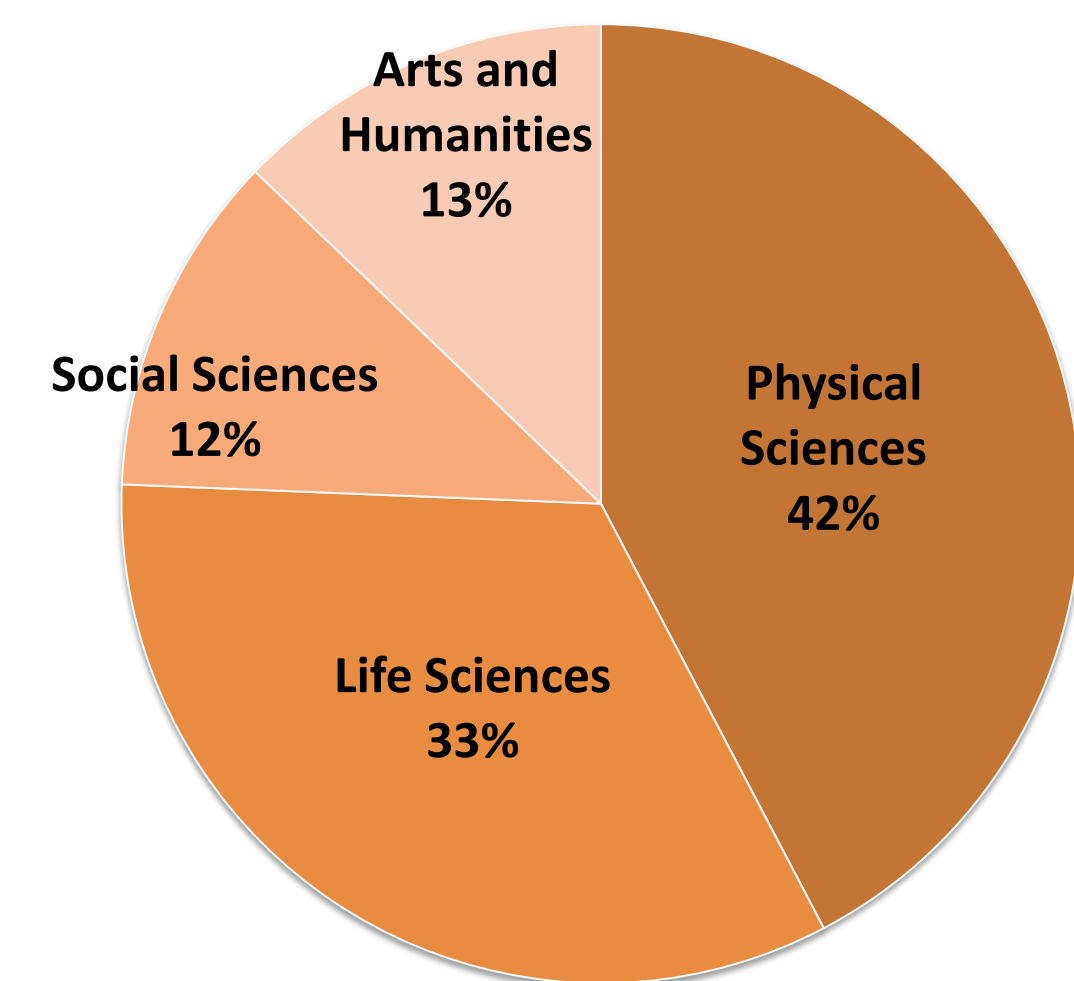


Appendix B: Video #1 of Trap-Jaw Ant Escape Jump
A standard speed video example of *Odontomachus brunneus* jumping out of an experimental antlion pit can be found in a supplemental file named *larabee_appendix_B.mov*

Supplement 2. Transcripts expression ratios significantly expressed in at least one treatment relative to ambient undamaged conditions 3 days and 2 weeks after the beginning of the experiment, (P < 0.01) p values are for adjusted all comparisons are to ambient undamaged plants

Affy Identifier	p value	3 days						2 weeks																				
		CO2	O2	O3	O4	O5	O6	CO2	O2	O3	O4	O5	O6															
Gma.1.1.AT.at	0.0001	1.405198822	4.46851E-05	0.62629805	0.415	0.88	3.0027E-06	1.67	2E-08	2.3	0.876	1.04	0.86	0.95	0.906	1	1	0.97	0.99	0.96	1	1	0.99	0.98	1.01	0.99996	1	
Gma.1.1.S1.at	0.0029	2.2952452	0.75092765	0.68631809	0.85	1.31	0.00131684	24.9	2E-04	55	0.915	14.6	0.4	0.3	0.9	0.64	2.89	0.63	2.65	0.75	1.4	1	1.24	0.29	3.91	0.99156	4.52	
Gma.10013.1.S1.a	0.0022	0.83369999	0.706104274	0.552084391	0.011	0.79	0.55395116	0.95	0.006	0	0.335	0.9	0.02	0.74	0.556	1	0.99	1.01	0.97	0.97	0.81	1	0.94	0.94	0.93	0.99995	0.99	
Gma.1002.1.A1.a	0.0089	1.29544973	0.83652142	0.93412311	0.523	1.4	0.04446007	0.68	0.351	0.9	0.014	0.63	0.15	0.69	0.296	1.2	0.98	1.04	0.46	1.24	0.96	1	0.78	0.24	1.05	0.99911	1.1	
Gma.1002.2.S1.a	0.0044	1.38526299	0.99179213	0.977387911	0.093	1.35	0.00297314	0.67	0.725	0.9	0.02	0.66	0.82	0.87	0.171	1.2	0.99	0.98	0.58	1.19	0.99	1	0.75	1.21	0.95	0.99995	1.09	
Gma.1003.1.S1.a	0.0081	1.16642754	0.55208215	0.64042623	0.014	1.24	0.13964365	1.11	0.002	1.3	0.071	1.16	0.48	0.307	0.9	0.98	1.04	0.88	0.97	0.65	1	0.7	0.9	0.96	1.01	0.99995	0.95	
Gma.1003.8.1.S1.a	0.1332	1.224852178	0.6682638	0.899703394	0.722	1.09	0.08852894	0.68	0.388	0.8	0.014	0.61	0.08	0.61	0.115	1.3	0.99	0.99	0.48	1.25	0.07	0.7	1	0.95	0.22	0.73	0.99995	1.05
Gma.1004.9.1.A1.a	0.1142	0.87678972	0.87215037	0.95724495	0.182	0.84	0.06874208	1.27	0.418	1.1	0.138	1.21	0.97	0.99	0.259	0.9	0.98	0.96	0.41	0.86	0.21	1.1	1	1.01	0.63	1.08	0.99156	0.86
Gma.1005.1.S1.a	0.0427	1.17098491	0.89526029	0.89286133	0.285	1.15	0.04028865	1.19	0.003	1.4																		
Gma.1005.0.1.S1.a	0.0022	0.76557906	0.74047611	0.93542774	0.011	0.72	0.09342437	0.85	8E-04	0.6																		
Gma.1005.3.1.A1.a	8E-05	1.24376236	0.85898069	1.02917235	0.604	1.28	0.217676697	0.94	0.008	1.2																		
Gma.1006.2.1.S1.a	0.2192	1.11945062	0.928224074	0.963852939	0.631	1.08	0.14743481	1.27	0.006	1.4																		
Gma.1006.9.1.A1.a	0.0072	0.92287823	0.233022678	0.884661623	0.084	0.73	0.96157633	1.01	0.084	0.8																		
Gma.1007.1.S1.a	0.5273	0.965312201	0.030386297	0.63737812	0.031	0.8	0.00652828	0.76	0.001	0.7																		
Gma.1007.2.S1.a	0.5952	0.96027722	0.04616685	0.33860687	0.045	0.8	0.001265023	0.76	0.003	0.7																		
Gma.1007.6.1.S1.a	1E-05	0.48133579	0.405646863	1.22413037	0.03	0.61	0.60133447	0.88	2E-04	0.4																		
Gma.1007.8.2.S1.a	3E-05	0.779117609	0.056632091	0.872627674	2E-05	0.68	0.154423172	0.91	9E-05	0.7																		
Gma.1008.0.1.S1.a	0.5491	0.93028038	0.490130991	1.15611729	0.77	1.09	0.93668076	0.97	0.622	0.9																		
Gma.1008.6.1.S1.a	0.0021	1.16078292	0.99903696	0.993817244	0.062	1.15	0.78142624	1.02	0.017	1.2																		
Gma.1008.7.1.S1.a	0.0177	0.96889141	0.96204044	1.02947057	0.32	0.89	0.08584193	1.22	0.087	1.2																		
Gma.1009.1.S1.a	0.0127	0.80888106	0.324899626	0.5877736	0.008	0.24	0.23570962	1.58	0.157	0.7																		
Gma.1009.3.1.A1.a	0.054	0.95543322	0.08395955	0.950383741	0.007	0.91	0.642659931	0.98	0.007	0.9																		
Gma.1009.5.1.S1.a	0.1187	0.9118552	0.85295721	0.97188657	0.244	0.85	0.513787878	1.08	0.66	0.9																		
Gma.1009.9.1.S1.a	0.0031	0.5682031	0.39109013	1.3014927	0.372	0.74	0.69198922	0.86	0.011	0.5																		
Gma.1010.1.S1.a	0.0342	0.276470165	0.03689195	0.25046462	0.853	0.86	0.012912931	1.27	8E-04	1.8																		
Gma.1010.9.1.S1.a	2E-05	0.76028637	0.9384312	1.024899684	0.007	0.78	0.607647829	0.95	5E-04	0.7																		
Gma.1011.1.A1.at	3E-06	0.582482326	0.45849322	0.873548629	5E-05	0.51	0.600120842	0.82	1E-04	0.5	0.228	0.8	0.45	0.163	0.8	0.95	1.1	0.93	0.92	0.34	1.1	1	0.94	0.28	1.24	0.99995	0.87	

Supplemental File(s) by Discipline

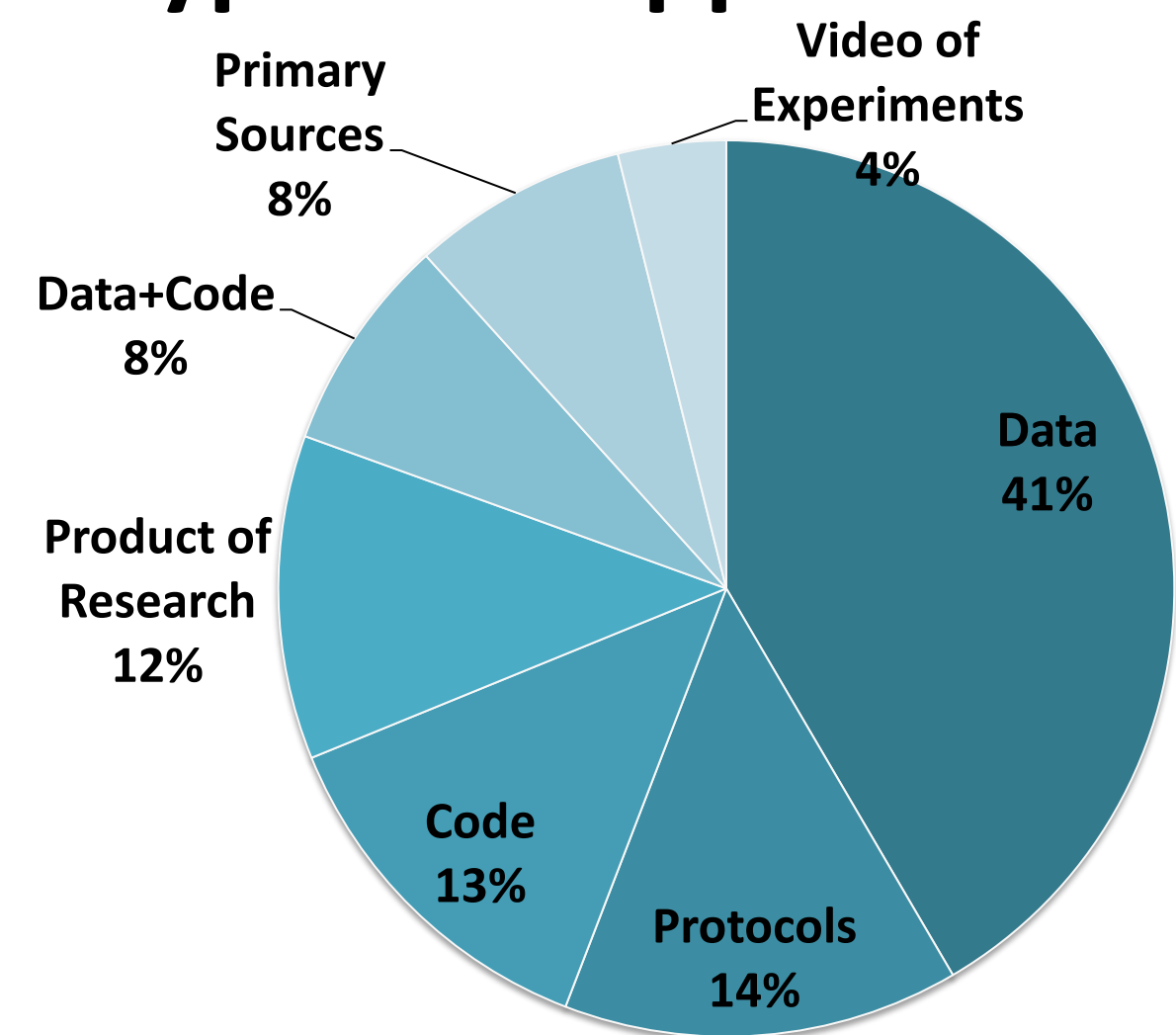


Numbers of Supplemental Files

- 29% - 1 file
- 36% - 2-5 files
- 16% - 6-20 files
- 19% - 21 or more files

One thesis had over **2000** supplemental files (data+code) included in a **zip** file

Types of Supplemental Files



Formats of Supplemental Files



Preliminary Findings

Wide variety of formats. Code is a particular concern. Type of supplementary file also varies. Data and data+code make up about half of the files. Basic description of what a supplementary file is generally included in the appendix, but contained within the text of the ETD itself. Definite need for more structured description of supplemental files that can assist both in use and long term access and preservation. More investigation needed into when supplemental files are included and when they are not.