Preparing Yale's material cycles data for global distribution:
An overview of the Yale STAF Database

Yale
Center for Industrial Ecology

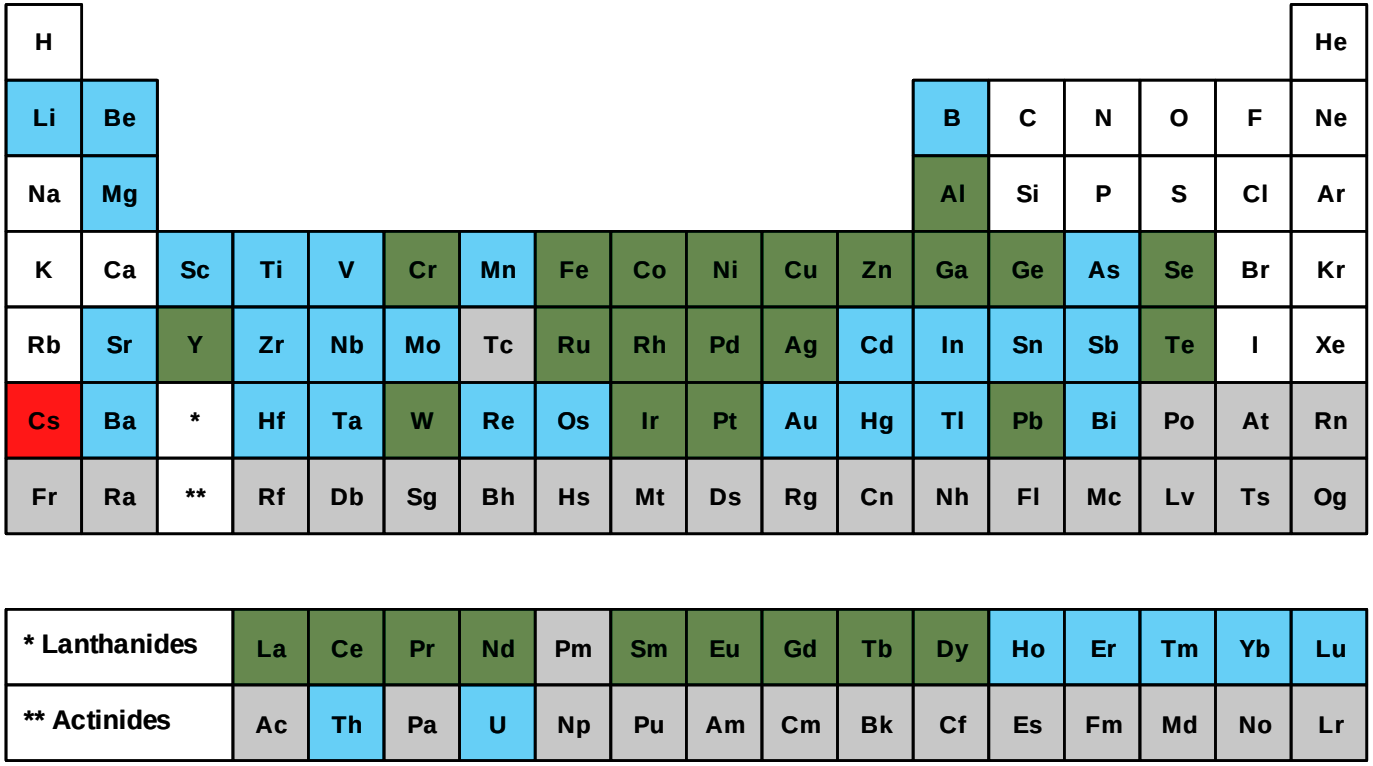
Dr. Rupert J. Myers
Postdoctoral Associate

An abundance of data…

Graedel research group (1997 -)

17+ years of material cycles (2000 -)

7+ years of criticality assessments (2010 -)



35 Non-primordial radioactive elements

20 No data

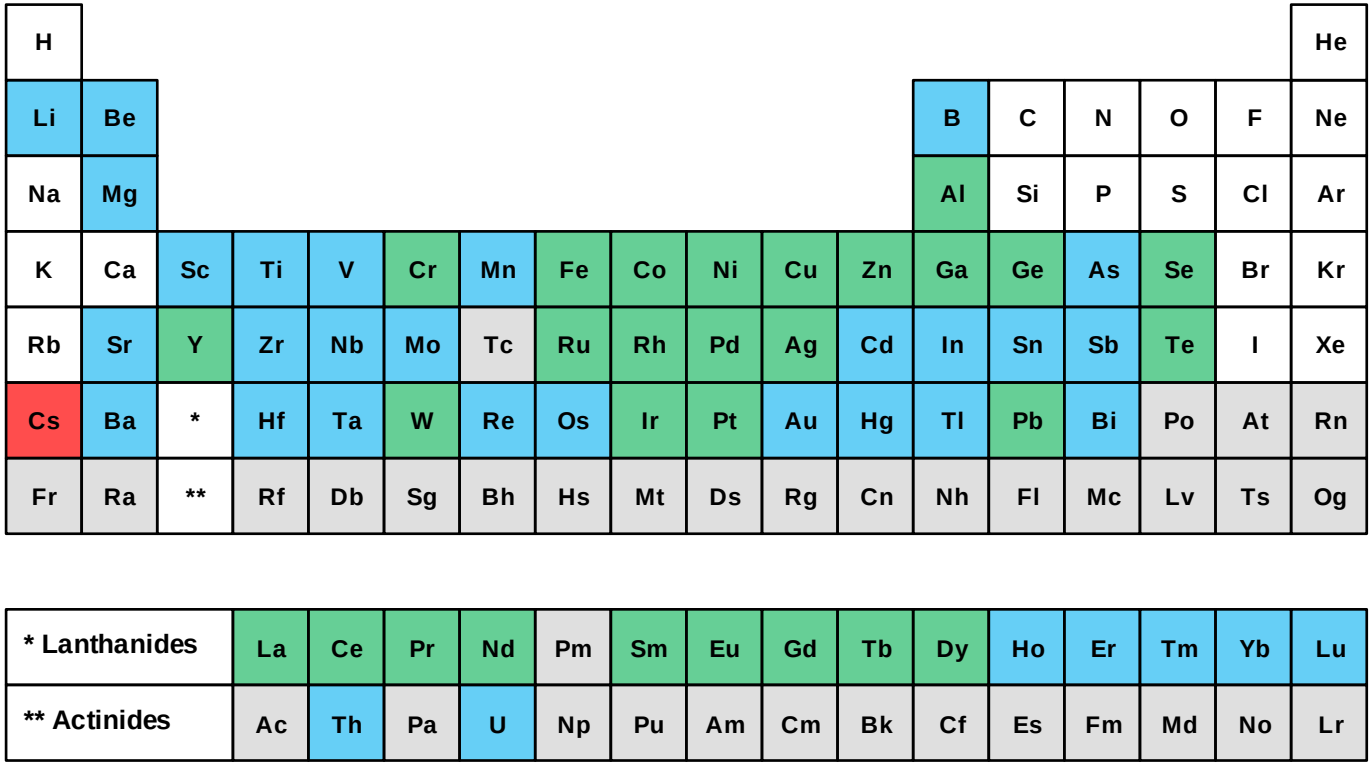
1 Material cycle data

33 Criticality data

29 Criticality and Material cycle data

Yale STAF Database

Unifying ~20 years of material cycle and criticality data generated by Tom Graedel's research group (Yale) into a database, including:



35 Non-primordial radioactive elements

20 No data

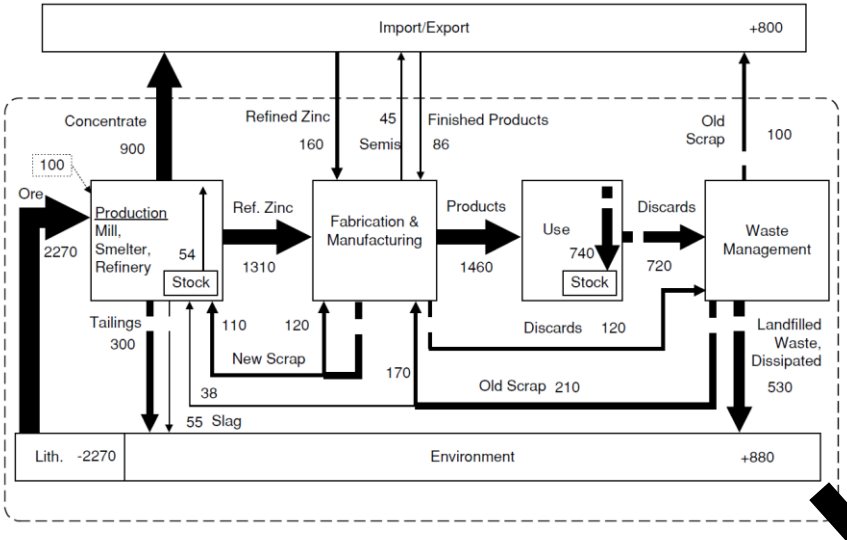
1 Material cycle data

33 Criticality data

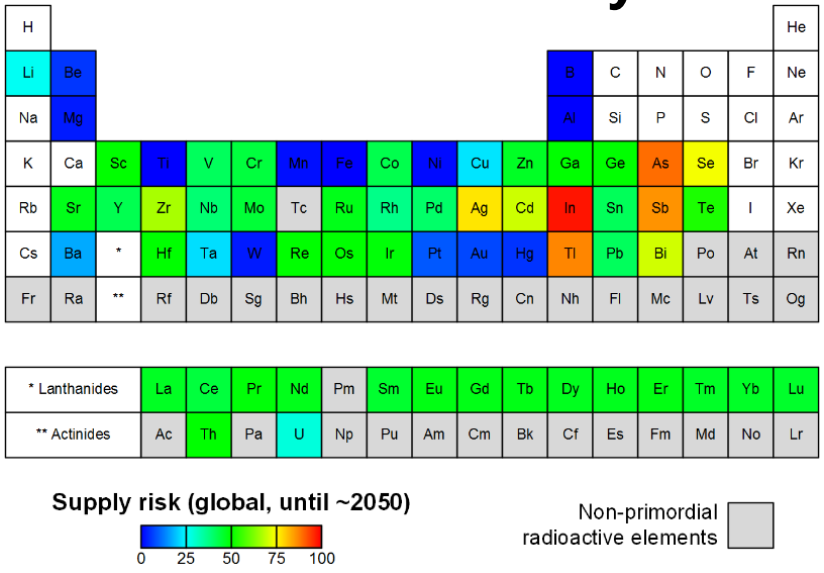
29 Criticality and Material cycle data

National Science Foundation Grant #1636509, T.E. Graedel (PI), B.K. Reck (Co-PI), R.J. Myers (Co-PI), Preparing the Yale metal life cycles database for global distribution, Yale University, USD$233,785 (2016-2018)

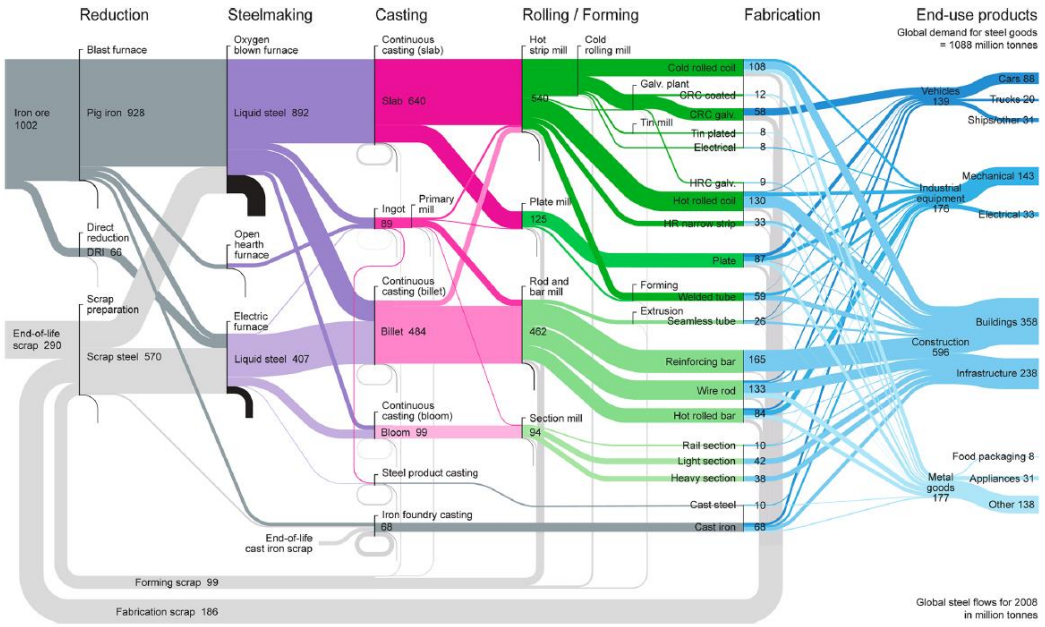
Material cycles



Material criticality



Unified data structure for material cycles



Cullen et al., *Environ. Sci. Technol.*, 2012, 46, 13048-13055

Messy for complex data analysis

No standardized visual labeling system

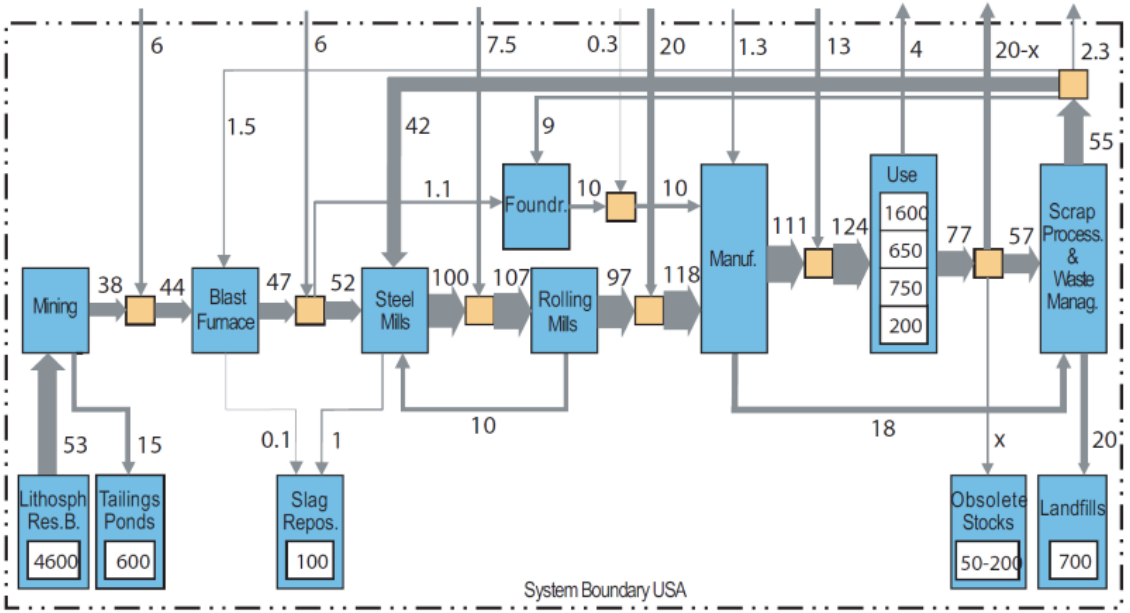
Additional logic needed to translate visualization into computational data structure

Individual datasets

Typically inconsistently formatted

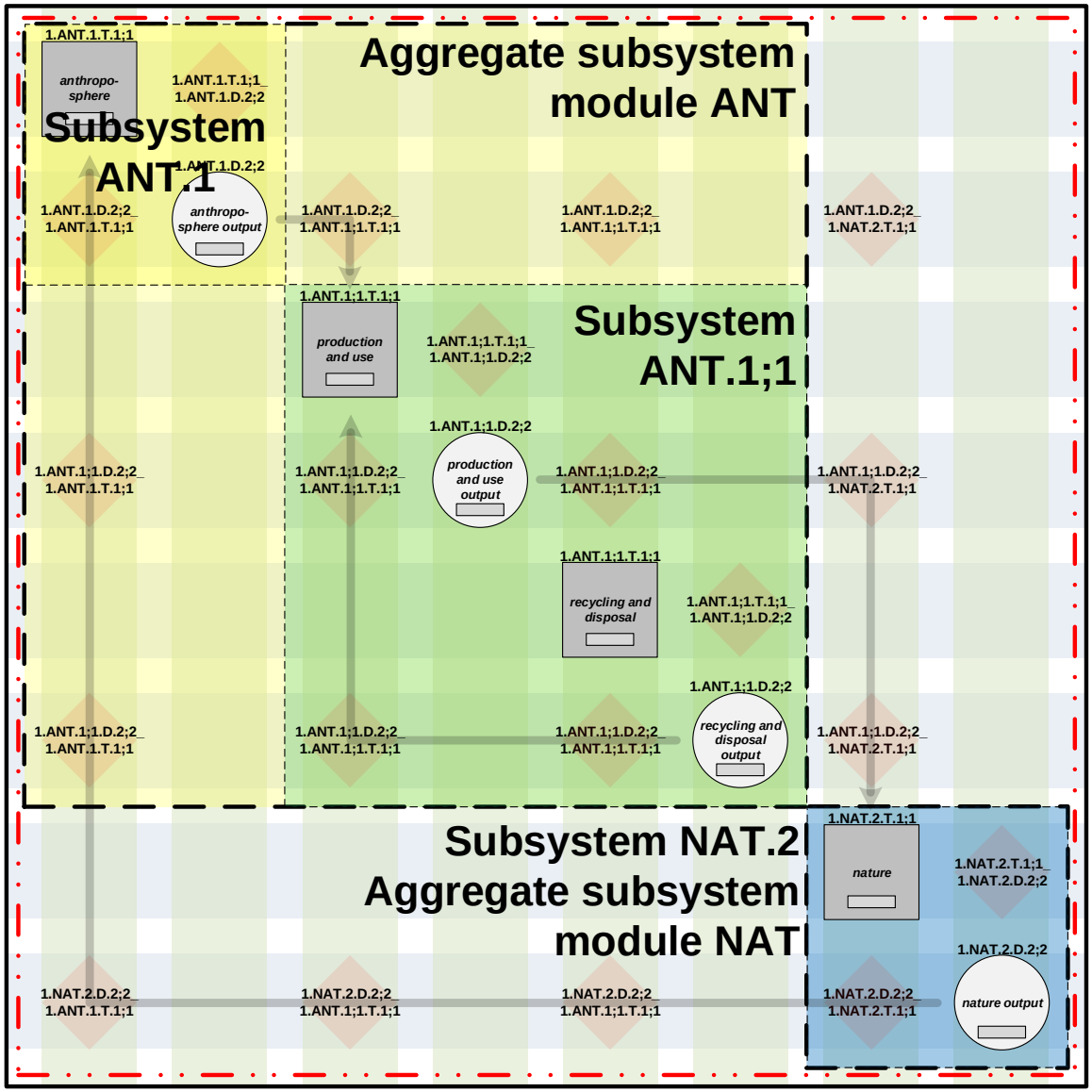
Often incompletely documented

Rarely describe multiple data types (kg, $, …)



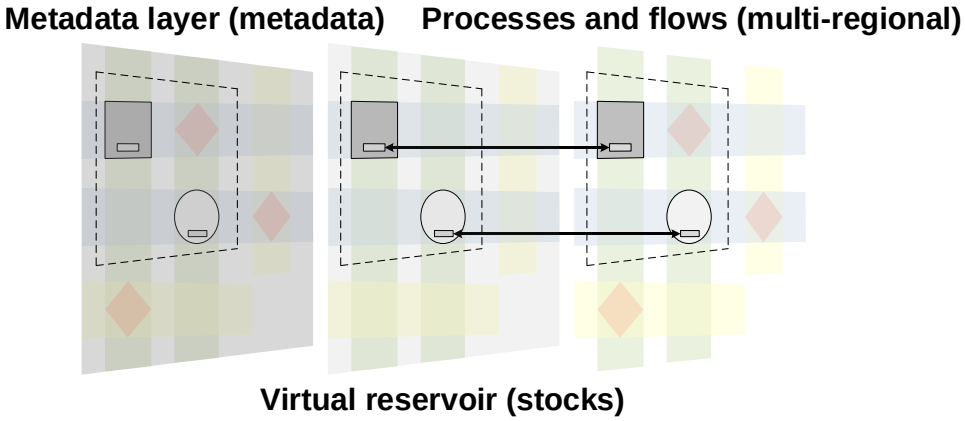
Müller et al., *Proc. Natl. Acad. Sci. USA*, 2006, 103(44), 16111-16116

Unified Material Information System (UMIS)



Myers et al., *J. Ind. Ecol.*, under review;
1 invited presentation/1 presentation/1 panel at the ISIE-ISSST 2017 Conference (25-29 June)

UMIS: key components

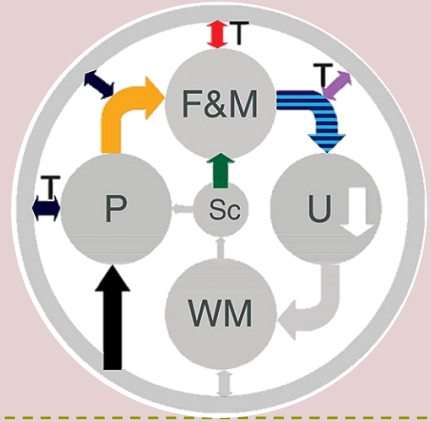


Metadata

Material reservoir (virtual) - stocks

Spatial boundaries - trade

(Life) cycle stages – processes & flows

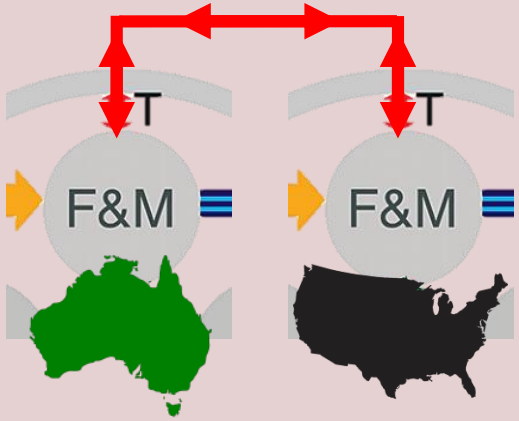


F&M = fabrication & manufacturing

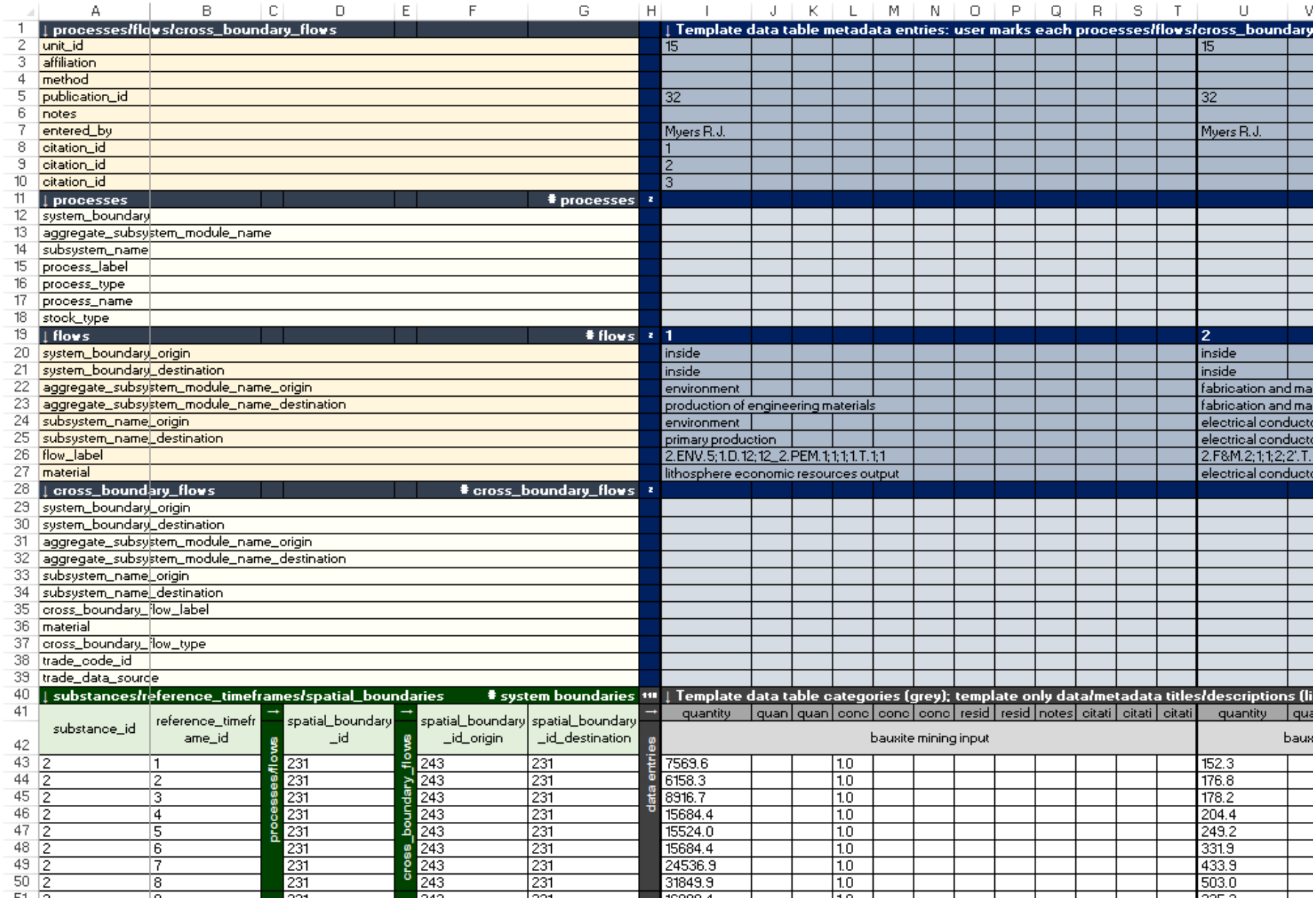
U = use

…

Material hierarchy for each process/flow



Data entry, ‘templates’, publications → CSV



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	↓ processes/flows/cross_boundary_flows								↓ Template data table metadata entries: user marks each processes/flows/cross_boundary													
2	unit_id								15												15	
3	affiliation																					
4	method																					
5	publication_id								32												32	
6	notes																					
7	entered_by								Myers R.J.												Myers R.J.	
8	citation_id								1													
9	citation_id								2													
10	citation_id								3													
11	↓ processes						# processes															
12	system_boundary																					
13	aggregate_subsystem_module_name																					
14	subsystem_name																					
15	process_label																					
16	process_type																					
17	process_name																					
18	stock_type																					
19	↓ flows						# flows		1												2	
20	system_boundary_origin								inside												inside	
21	system_boundary_destination								inside												inside	
22	aggregate_subsystem_module_name_origin								environment												fabrication and ma	
23	aggregate_subsystem_module_name_destination								production of engineering materials												fabrication and ma	
24	subsystem_name_origin								environment												electrical conducto	
25	subsystem_name_destination								primary production												electrical conducto	
26	flow_label								2.ENV.5;1.D.12;12_2.PEM.1;1;1;1.T.1;1												2.F&M.2;1;1;2;2'.T.	
27	material								lithosphere economic resources output												electrical conducto	
28	↓ cross_boundary_flows						# cross_boundary_flows															
29	system_boundary_origin																					
30	system_boundary_destination																					
31	aggregate_subsystem_module_name_origin																					
32	aggregate_subsystem_module_name_destination																					
33	subsystem_name_origin																					
34	subsystem_name_destination																					
35	cross_boundary_flow_label																					
36	material																					
37	cross_boundary_flow_type																					
38	trade_code_id																					
39	trade_data_source																					
40	↓ substances/reference_timeframes/spatial_boundaries						# system boundaries		↓ Template data table categories (grey); template only data/metadata titles/descriptions (li													
41	substance_id	reference_timefr ame_id	processes/flows	spatial_boundary _id	cross_boundary_flows	spatial_boundary _id_origin	spatial_boundary _id_destination	data entries	quantity	quan	quan	conc	conc	conc	resid	resid	notes	citati	citati	citati	quantity	qua
42									bauxite mining input												bauxi	
43	2	1		231		243	231		7569.6			1.0									152.3	
44	2	2		231		243	231		6158.3			1.0									176.8	
45	2	3		231		243	231		8916.7			1.0									178.2	
46	2	4		231		243	231		15684.4			1.0									204.4	
47	2	5		231		243	231		15524.0			1.0									249.2	
48	2	6		231		243	231		15684.4			1.0									331.9	
49	2	7		231		243	231		24536.9			1.0									433.9	
50	2	8		231		243	231		31849.9			1.0									503.0	

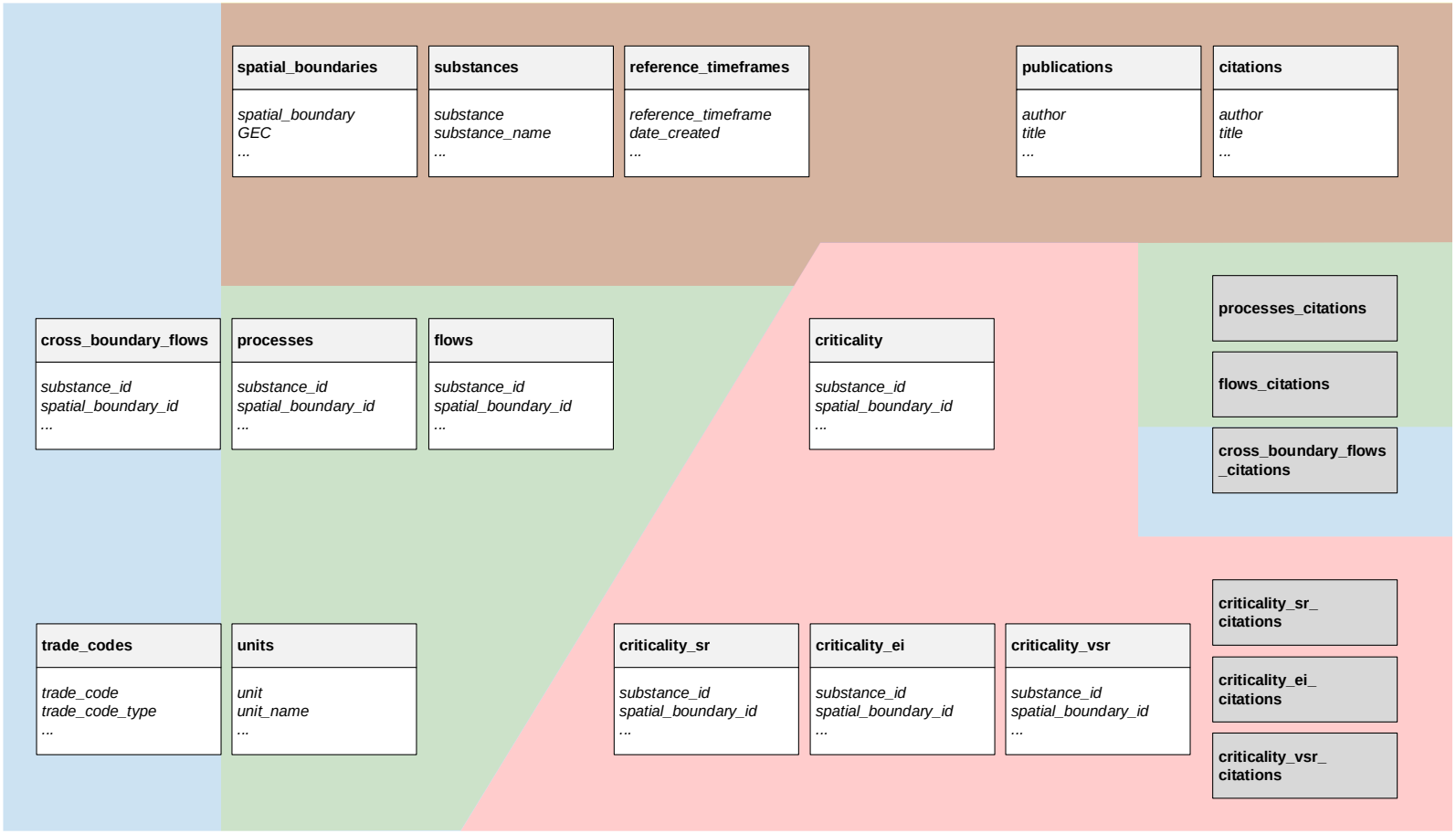
Data table: flows

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	substance_id	spatial_boundary_id	reference_timeframe_id	system_boundary_origin	system_boundary_destination	aggregate_subsystem_module_origin	aggregate_subsystem_module_destination	subsystem_origin	subsystem_destination	flow_label	material	quantity	quantity_uncertainty	quantity_reliability	concentr
2	13	4	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
3	13	10	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	168	\N	\N	
4	13	13	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	946	\N	\N	
5	13	14	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
6	13	246	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
7	13	30	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	36	\N	\N	
8	13	39	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	724	\N	\N	
9	13	44	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	5118	\N	\N	
10	13	45	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	708	\N	\N	
11	13	247	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	1194	\N	\N	
12	13	49	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	3	\N	\N	
13	13	62	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	38	\N	\N	
14	13	67	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
15	13	76	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
16	13	77	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
17	13	83	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
18	13	86	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
19	13	101	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	39	\N	\N	
20	13	102	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	1165	\N	\N	
21	13	103	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	149	\N	\N	
22	13	107	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
23	13	108	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
24	13	110	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	1	\N	\N	
25	13	114	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
26	13	129	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
27	13	138	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	414	\N	\N	
28	13	144	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	8	\N	\N	
29	13	150	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
30	13	152	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
31	13	159	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	0	\N	\N	
32	13	166	101	outside	inside	environment	production o	\N	\N	RM0601_F	unclassifi	622	\N	\N	

Yale STAF Database structure

Material cycles
(multiple spatial boundaries)

Material cycles & criticality
(single & multiple spatial boundaries)



Material cycles
(single & multiple spatial boundaries)

Criticality

MySQL®

Summary

Developing a database to archive Yale's material cycle and criticality data (63 elements, >100,000 data entries)

To do so, we have developed a general and comprehensive data structure for materials cycles data, the Unified Materials Information System (UMIS)

Published data are transformed into the UMIS structure using 'templates', which are then archived into the database (MySQL), and will eventually be freely available (USGS)

National Science Foundation
Grant #1636509

Yale
Center for Industrial Ecology

Yale

Dr. Rupert J. Myers
rupert.myers@yale.edu
Dr. Tomer Fishman
Dr. Barbara K. Reck
Prof. Thomas E. Graedel

USGS

Mr. Michael S. Baker
Dr. Nedal T. Nassar