Getting to Know You! (GtKY!)

A Pentaho Data Integration (PDI) transformation that ingests the raw JSON data for
the GtKY application API in order to profile, validate, distill, and model the API's
data source objects.

========================= TRANFORMATION INFO =========================

Author: Michael Hoovler
Email: hoovlermichael@gmail.com
Engine: Pentaho PDI (Kettle), v7.1.0.0
File Type: Kettle Transformation (.KTR)
File Format: XML

Source Type: Online API
Source Location: https://www.willowtreeapps.com/api/v1.0/profiles
Source Format: JSON

API Data Out1   'output all peeps'
API Data Out2   'output all matts'

========================= VALIDATION CONDITIONS =========================

BOTH of these conditions must be met in order to pass the validation requirements:

     (1) IFF the final count of rows processed by the 'output all peeps' step equals
          that of the 'unique_peeps' step

     (2) IFF the final number of rows processed by the 'output all matts' step matches
          that of the 'unique_matts' step

========================= DESCRIPTION & PURPOSE =========================

This transformation contains multiple sub-sections, and may be visually divided by
both horizontal grid placement, as well as placement on the vertical grid, in
addition to the evident color codes and step clusters.

Taken from left to right - considering the full vertical scale - the flow generally
follows a sequential pattern of processing, with a few notable distinctions I've
highlighted with differently-colored labels.

As viewed from top to bottom, with the full horizontal span in tow, there are two
distinct flows - the JSON input for which is given a green label. The upper-most
flow uses the given source URL as the JSON input datasource. The lower section is a
duplicate of the upper, but uses a local copy of the URL's JSON as the source. It
differs elsewhere inasmuch as this was the flow used to profile the data and to
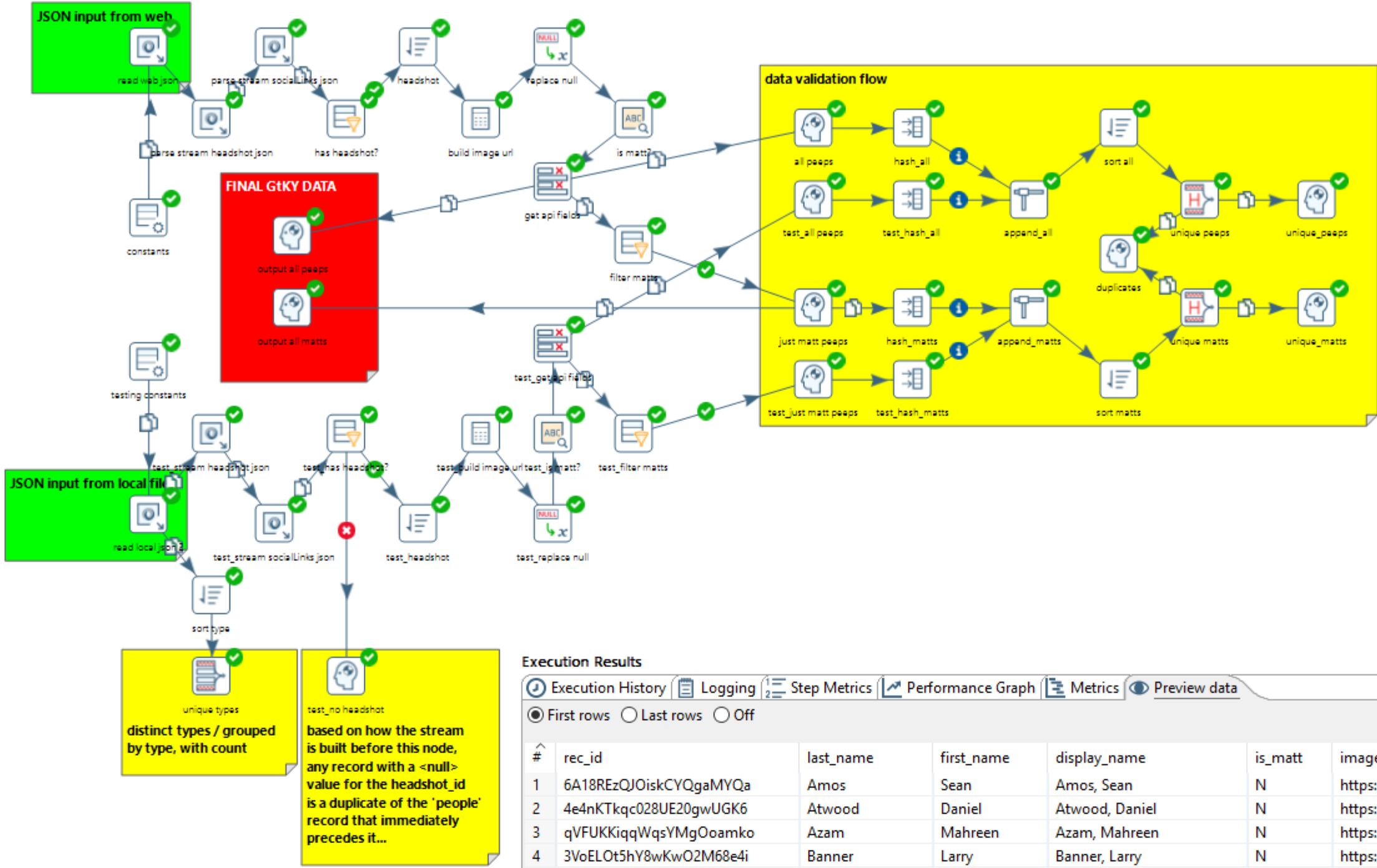provide end-points for the data stream absent from the upper.

The most obvious feature of the transformation that doesn't seem to fit the general
pattern is the 'data validation flow' section, which merges the outputs from both
upper and lower streams into a more homogeneous combination of the two. The purpose
of this section is evident by the 'unique_peeps' and 'unique_matts' nodes, the final
number of rows for which serve as validation for the 'FINAL GtNY DATA' portion.

Lastly, there is the 'FINAL GtNY DATA' itself, which simply groups the two API endpoints.
These two steps serve as the logical data sources for the GtKY application API.

JSON input from web

read web json

parse stream socialLinks json

has headshot?

replace null

is matt?

data validation flow

all peeps   hash_all   sort_all

test_all peeps   test_hash_all   append_all   unique_peeps   unique_peeps

just matt peeps   hash_matts   append_matts   unique_matts   unique_matts

test_just matt peeps   test_hash_matts   sort_matts

parse stream headshot json

build image url

FINAL GtKY DATA

output all peeps

output all matts

constants

get api fields

filter matts

testing constants

test_get api fields

test_replace null

JSON input from local file

read local json

test_stream headshot json

testing has headshot?

test_build image url test_is matt?   test_filter matts

test_headshot

test_replace null

sort type

distinct types / grouped
by type, with count

unique types

test_no headshot

test_no headshot

based on how the stream
is built before this node,
any record with a <null>
value for the headshot_id
is a duplicate of the 'people'
record that immediately
precedes it...

Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	testing constants	0	0	1	0	0	0	0	0	Finished	0.0s	250	-
2	constants	0	0	1	0	0	0	0	0	Finished	0.0s	500	-
3	read local json 2	0	0	650	325	0	0	0	0	Finished	0.2s	40,625	-
4	read web json	0	1	325	325	0	0	0	0	Finished	0.2s	1,658	-
5	test_stream headshot json	0	325	325	325	0	0	0	0	Finished	0.0s	13,000	20
6	parse stream headshot json	0	325	325	325	0	0	0	0	Finished	0.2s	1,617	-
7	parse stream socialLinks json	0	325	325	325	0	0	0	0	Finished	0.2s	1,585	-
8	has headshot?	0	325	153	325	0	0	0	0	Finished	0.0s	1,555	-
9	test_stream socialLinks json	0	325	325	325	0	0	0	0	Finished	0.0s	11,207	-
10	test_has headshot?	0	325	325	325	0	0	0	0	Finished	0.0s	9,559	-
11	headshot	0	153	153	0	0	0	0	0	Finished	0.0s	718	-
12	build image url	0	153	153	0	0	0	0	0	Finished	0.0s	705	-
13	replace null	0	153	153	0	0	0	0	0	Finished	0.0s	692	-
14	test_headshot	0	153	153	0	0	0	0	0	Finished	0.0s	4,135	-
15	test_build image url	0	153	153	0	0	0	0	0	Finished	0.0s	3,825	-
16	test_no headshot	0	172	172	0	0	0	0	0	Finished	0.0s	4,778	-
17	test_replace null	0	153	153	0	0	0	0	0	Finished	0.0s	3,400	-
18	test_is matt?	0	153	153	0	0	0	0	0	Finished	0.0s	3,188	-
19	test_get api fields	0	153	306	0	0	0	0	0	Finished	0.1s	5,885	-
20	test_all peeps	0	153	153	0	0	0	0	0	Finished	0.1s	2,684	-
21	test_filter matts	0	153	10	0	0	0	0	0	Finished	0.1s	2,550	-
22	is matt?	0	153	153	0	0	0	0	0	Finished	0.2s	683	-
23	get api fields	0	153	459	0	0	0	0	0	Finished	0.2s	2,013	-
24	all peeps	0	153	153	0	0	0	0	0	Finished	0.2s	657	-
25	filter matts	0	153	10	0	0	0	0	0	Finished	0.2s	654	-
26	just matt peeps	0	10	10	0	0	0	0	0	Finished	0.2s	84	-
27	test_just matt peeps	0	10	10	0	0	0	0	0	Finished	0.1s	161	-
28	test_hash_all	0	153	153	0	0	0	0	0	Finished	0.1s	2,508	-
29	test_hash_matts	0	10	10	0	0	0	0	0	Finished	0.1s	149	-
30	sort type	0	325	325	0	0	0	0	0	Finished	0.0s	14,130	-
31	unique types	0	325	5	0	0	0	0	0	Finished	0.2s	12,037	-
32	hash_all	0	153	153	0	0	0	0	0	Finished	0.2s	646	-
33	hash_matts	0	10	10	0	0	0	0	0	Finished	0.2s	41	-
34	append_matts	0	20	20	0	0	0	0	0	Finished	0.5s	37	-
35	append_all	0	306	306	0	0	0	0	0	Finished	0.5s	568	-
36	sort matts	0	20	20	0	0	0	0	0	Finished	0.5s	36	-
37	unique matts	0	20	10	0	0	10	0	0	Finished	0.6s	54	-
38	unique_matts	0	10	10	0	0	0	0	0	Finished	0.6s	18	-
39	sort all	0	306	306	0	0	0	0	0	Finished	0.6s	564	-
40	unique peeps	0	306	306	0	0	153	0	0	Finished	0.6s	841	-
41	duplicates	0	163	163	0	0	0	0	0	Finished	0.6s	293	-
42	output all peeps	0	153	153	0	0	0	0	0	Finished	0.2s	659	-
43	unique_peeps	0	153	153	0	0	0	0	0	Finished	0.6s	278	-
44	output all matts	0	10	10	0	0	0	0	0	Finished	0.2s	41	-

Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

● First rows ○ Last rows ○ Off

#	rec_id	last_name	first_name	display_name	is_matt	image_url
1	6A18REzQJOiskCYQgaMYQa	Amos	Sean	Amos, Sean	N	https://images.ctfassets.net/3cttzl4i3k1h/5oTPOfrAkwUS8y2w6lwsuO/1b0d814ee9b435e159573c4b701eab28/headshot_sean_amos.jpg
2	4e4nKTkqc028UE20gwUGK6	Atwood	Daniel	Atwood, Daniel	N	https://images.ctfassets.net/3cttzl4i3k1h/4xObs4rq6AaskQ6AcCcAKl/6b3ca5266c17dd96af018aa8870de52/3170c3b.jpg
3	qVFUKKiqqWqsYMgOoamko	Azam	Mahreen	Azam, Mahreen	N	https://images.ctfassets.net/3cttzl4i3k1h/5J4GjKsDUk6amwQASC6eSg/5c042758ba2876cbbabdda1cacedf93e/headshot_mahreen_azam.jpg
4	3VoELOt5hY8wKwO2M68e4i	Banner	Larry	Banner, Larry	N	https://images.ctfassets.net/3cttzl4i3k1h/62w63e7xteQGgKWqyaAuAE/0f004ec0debc8a52c56bb01563f8844/Larry.jpg
5	52sUSQZiQUqo0ege4wUmy4	Baranowski	Matthew	Baranowski, Matthew	Y	https://images.ctfassets.net/3cttzl4i3k1h/avLdCFJYHKqWOeagmWq6u/11b0d92856fbbba84c9da2c94f23c2c0/headshot_matt_baranowski2.jpg
6	10RbMxoaP0A6amcAoGeSC2	Batman	Elizabeth	Batman, Elizabeth	N	https://images.ctfassets.net/3cttzl4i3k1h/552Fh0Hfu8AWsSEYmWkEwe/d009da1a01e640d39aa50b6300e7ea32/headshot_lizzie_batman.jpg
7	4lh5w7atvOgYsEggioUiCU	Batten	Angela	Batten, Angela	N	https://images.ctfassets.net/3cttzl4i3k1h/iRC8dDilRqYSAgkSo4keS/55a8612209d30f18054c7450be9c3c40/headshot_angela_batton.jpg
8	4k4ksxINHqy8y02Gs22CkY	Bowles	Ashby	Bowles, Ashby	N	https://images.ctfassets.net/3cttzl4i3k1h/2l5ixOSn0AsyWaMe0mEGMs/09ef8a0e1f4838b3e5c561f88f640de7/headshot_ashby_bowles.jpg
9	2JK2zW1eG42C00wYeoC68y	Brameyer	Derek	Brameyer, Derek	N	https://images.ctfassets.net/3cttzl4i3k1h/3JSJ4RrJlKyqgUuiwG6Ywq/dfbda212a041978f5abbc9efb39572be/IMG_6700.jpg
10	2eKUsaLa0UgwaCGU0uAS6W	Brear	David	Brear, David	N	https://images.ctfassets.net/3cttzl4i3k1h/4cjP5JnhFYeaYkSOoos0IW/bb0145505280163888102dec31168d3f/headshot_david_brear.jpg
11	6btNXZhuo0uQSo4icYqGyO	Brown	Preston	Brown, Preston	N	https://images.ctfassets.net/3cttzl4i3k1h/3lciiXPNFelaWUA4GUMEO4/13db5b49dc2a746d03dbecbb05501f84/headshot_preston_brown.jpg
12	6CrM4BdoJ2yUmeaOAYsCA2	Brown	Rebekkah	Brown, Rebekkah	N	NA
13	7zHJnICoY86ek8oISwYUsA	Bryant	Barry	Bryant, Barry	N	https://images.ctfassets.net/3cttzl4i3k1h/5ZUiD3uOByWWuaSQsayAQ6/c630e7f851d5adb1876c118dc4811aed/featured-image-TEST1.png
14	6qi0txvBkWYK8AimoASS2w	Bryant-Ryback	Christine	Bryant-Ryback, Christine	N	https://images.ctfassets.net/3cttzl4i3k1h/4bSE4jIUXCQGQowcWauW6S/2b89c3e592265cca936d8ee4df0940ff/headshot_christine_braynt_ryback.jpg
15	5BEVM7wUY82AGqUluowmwc	Cady	Jarrod	Cady, Jarrod	N	https://images.ctfassets.net/3cttzl4i3k1h/63MK8b4lVyGmma2asSKqeG/7fd1b44e861b6a3683191ed367a520da/jarrod.png
16	GronavkyOq6WaWwy6ggcy	Carter	Andrew	Carter, Andrew	N	https://images.ctfassets.net/3cttzl4i3k1h/1YVnmELNi4CgUmlWUWgQE/ddd718224b5076b6d966ac732f94692b/headshot_andrew_carter.jpg
17	WMZBgojs4K2yoWGcasYci	Carter	Brandon	Carter, Brandon	N	https://images.ctfassets.net/3cttzl4i3k1h/7hsmsmmYIGYooGwAQYAWuS/11fdb68353caee3d2a73d4b8b4ed66/headshot_brandon_carter.jpg
18	2MJZzvffY4UwCAOUGE4q8q	Casey	Roger	Casey, Roger	N	NA
19	2UUFBWewvYsyYIsC08ugWK	Casino	Tatyana	Casino, Tatyana	N	https://images.ctfassets.net/3cttzl4i3k1h/6RONTEtFlICiM4a0ckskiw/2c9be82630792914bbc77bdd887d4c9/headshot_tatyana_casino.jpg
20	5o8FknMJ4kC0eouSgEcKUm	Centofante	Peter	Centofante, Peter	N	https://images.ctfassets.net/3cttzl4i3k1h/aUQmOos9OwOliA2qAS8Yq/34c2d097fa5dd441e7f19faf4fa5f807/IMG_1182.jpg
21	4KUjxVUAEmSGaAEqOSQeW	Chandler	Charlie	Chandler, Charlie	N	https://images.ctfassets.net/3cttzl4i3k1h/5YCG4GQso04o0wmoW2y8C8/99cd8570b00d4175cb3d5788eb9268a1/headshot_charlie_chandler.jpg
22	5oV9G8eKisW0M0GUCo4eke	Cherry	Joseph	Cherry, Joseph	N	https://images.ctfassets.net/3cttzl4i3k1h/1yUCBofluco4muYYslOwms/e6896398aac4889e8247059cfb58db1f/Joseph_Cherry.jpg
23	4qFhHJBYHY6WcgicAOK86m	Christian	Brittany	Christian, Brittany	N	https://images.ctfassets.net/3cttzl4i3k1h/4G8r6GkJoQGi0ckgu8OaUG/b3ba537338717e0ce309679ee1ece4ba/Brittany_Headshot.jpg
24	5Nr4SiYjM4umkQa2c88am	Cober	Erika	Cober, Erika	N	https://images.ctfassets.net/3cttzl4i3k1h/3nX4Cqo4C4iowKl0Y4okEA/173d9d8b40c8897889c0a95528fcbaaf/headshot_erika_cober.jpg
25	29Sn6GoIcE2G2uKgkqmcgg	Compton	Evan	Compton, Evan	N	https://images.ctfassets.net/3cttzl4i3k1h/2fYSUbmwlii2Yqi4i2Qewl/82e376360f1a22310e8a01a124e281a2/headshot_evan_compton.jpg
26	5zeP4yE3LiUk8iKCyyoEaQ	Conner	Kevin	Conner, Kevin	N	https://images.ctfassets.net/3cttzl4i3k1h/2jsQT4AK28SY6AcwQAYsmk/3427e7287bc1f6d308f78e7c03465b84/headshot_kevin_conner.jpg
27	19LrPt51IYuKq6qec8go68	Connor	Amanda	Connor, Amanda	N	https://images.ctfassets.net/3cttzl4i3k1h/Edx7eNECFEIY2OGg8WoAE/2805a3837ce2f7a28a8829e605c0d0bc/Connor_Photo.jpg
28	406rycFMu40YkMEQ8iW6s6	Cook	Abby	Cook, Abby	N	https://images.ctfassets.net/3cttzl4i3k1h/5PO8DVEsg0Wyo0iGcOq0Ui/6a21db753b91315ecba9852ab22723d4/abby.png
29	2Py3OXbu4MSKMYAEii2mwY	Dawson	Matthew	Dawson, Matthew	Y	https://images.ctfassets.net/3cttzl4i3k1h/77Md1XrN9m00USt82eyMuGY/91e97bf89afb939fbb723abec51898b7/headshot_matt_dawson2.jpg
30	6uLkL9vd60elmOoEGs44IS	de Wolfe	Charlotte	de Wolfe, Charlotte	N	https://images.ctfassets.net/3cttzl4i3k1h/5O6L3459egwGa0miey2CK4/93febc56ba2e1c636b35e020df994207/Image_uploaded_from_iOS_1_.jpg
31	XXRvvkg6KA80VqWaowa2	DeGroof	Amanda	DeGroof, Amanda	N	https://images.ctfassets.net/3cttzl4i3k1h/1LfMjWhfQEGS8YCA240UCa/6069799f091ea3c0fc8e9d98a9efc224/Amanda_DeGroof.jpg
32	ZXJst3C2o6CGSeOkaMw4W	Dengel	Tobias	Dengel, Tobias	N	https://images.ctfassets.net/3cttzl4i3k1h/5TtGDrZNAcm6oW6lGCQkaS/1afd3665e9b0d3a06f433dd7bebf737f/td.jpg
33	5p16RUZqGkUIOCCYYiKUWe	Dinon	Christopher	Dinon, Christopher	N	https://images.ctfassets.net/3cttzl4i3k1h/39k4Tey4wEe4GiIAmEiy0/96105e8c12a30fc4f0260f2e703fbb3e/Dinon.jpg
34	4YKWqsxefuuOUkQc8oiWaQ	Doyle	Frank	Doyle, Frank	N	https://images.ctfassets.net/3cttzl4i3k1h/4Qcr7oYHCgC4M2WU4klcao/9360ac59c805db9aa8bfe0eaccf1828e/headshot_frank_doyle.jpg
35	63zMjzatos6gkWEYkMAmsM	Dunn	Jordan	Dunn, Jordan	N	https://images.ctfassets.net/3cttzl4i3k1h/45Cl0BcKXCMK8laOul6OeW/fcf2e2790ac4d3aadff047efe34567ab/headshot_jordan_dunn.jpg
36	35M55sRl9ekgyK6qaMOcK0	Edmonds	Morgan	Edmonds, Morgan	N	https://images.ctfassets.net/3cttzl4i3k1h/4WgB8vctLWYlque8sUyO2m/58c10cef7a819c50bd717d72ca3b1a54/20170118-DSC_1496.png
37	7zNwpL4C3KqYqW88UWmokg	Ellis	Will	Ellis, Will	N	https://images.ctfassets.net/3cttzl4i3k1h/2JbFih4rzyK8E0elUmUCeo/b28d41104e6558ec3e12e1845f1ead59/headshot_will_ellis.jpg
38	1syHN53aOYlQYWgyQw0cos	Erwin	Taylor	Erwin, Taylor	N	https://images.ctfassets.net/3cttzl4i3k1h/N8Qwm1FGmWOcwll068gC8/e2e8a7c72479867bdbec38931595cb27/Taylor_Erwin.jpg
39	6xn7sUFT0WlccUUs6Wsu4m	Evans	Mike	Evans, Mike	N	https://images.ctfassets.net/3cttzl4i3k1h/3GSk7fOiEzukqGOwa4kqG8C/bf3afb42728fd5c7ea0369c3c05a0211/headshot_mike_evans.jpg
40	2OUgVqDQ8EyuoOaqMgYk4C	Fairchild	Charles	Fairchild, Charles	N	https://images.ctfassets.net/3cttzl4i3k1h/475K7gHErYCOuomwm4Ekli/4e40d27b84376edc08356d79f9e36d43/charlie.jpg
41	4aSoSNPp5e6yAmglqAGoly	Fishwick	James	Fishwick, James	N	https://images.ctfassets.net/3cttzl4i3k1h/EDy0YUkOVgIgmqsScAy0e/63a8741f72c3e473750a2534003c89fc/Fishwick.jpg
42	6fTnNvl8aQcCQiAyWOgOGG	Freenor	Michael	Freenor, Michael	N	https://images.ctfassets.net/3cttzl4i3k1h/5wUEGfGbbG6oomgsCYmsaW/f517b896f0a80ce68fc4f081a0762b9/headshot_michael_freenor.jpg
43	336esrta1WmIOqa06m0qog	French	Whitney	French, Whitney	N	https://images.ctfassets.net/3cttzl4i3k1h/5Jkhg0xi12WWkQy6Sq0MaE/f2ba4b01b05a7a60c30a5e26c5684d2b/blog-featured-In_Memoriam_Whitney_French.jpg
44	5udlAbCug8Y8yYMSaM8w0u	Friedel	Rich	Friedel, Rich	N	https://images.ctfassets.net/3cttzl4i3k1h/2VVBxdMyas06mW6EWW4wqy/960f7416dfe03de9031528488b08f2f/headshot_rich_friedel.jpg
45	6wFgHqSdkA4ywiuqwQogKq	Fry	Anne	Fry, Anne	N	https://images.ctfassets.net/3cttzl4i3k1h/UpvYJ4aJYOO4WGgOl4lEm/0db7c6c5dd06290e30ceadf3f84ae759/IMG_0287.jpg
46	29Xgxeqdm AO86mGelmwQQ	Frye	Ben	Frye, Ben	N	https://images.ctfassets.net/3cttzl4i3k1h/3SQLlq0Y36oYyaiwCSwOY8/a65ae6620c8041b2773c1915156261d7/headshot_ben_frye.jpg
47	3fgaigpnJYGYGWGWA6IMAy	Frysinger	Robert	Frysinger, Robert	N	https://images.ctfassets.net/3cttzl4i3k1h/1RF2rNFlPqo4Eu0aWsmmag/3e95edc1fcf65e6d83a725f013a23/robert.jpg
48	4NCJTL13UkK0qEIAAcg4lQ	Garrett	Joel	Garrett, Joel	N	https://images.ctfassets.net/3cttzl4i3k1h/4Mv2CONANym46UwuuClgK/cbeb43c93a843a43c07b1de9954795e2/headshot_joel_garrett.jpg
49	4Dm4Je4S0gg6iyUkK00e2	Gatewood	Patrick	Gatewood, Patrick	N	https://images.ctfassets.net/3cttzl4i3k1h/1rBIGKBikUsMYGw0QiyUv/936555b504d1ae795188282a3996dc5c/patrick_360.png
50	1QJiy9TGBWueKEi6yqMUku	Giacherio	Paul	Giacherio, Paul	N	https://images.ctfassets.net/3cttzl4i3k1h/3Viqobgn0AyCoGm444O8Wo/216db30b4378e7d7b286fdbee986983d/Paul.jpg
51	6JmEiXMWGcggAo4W0UKeG	Glidewell	Todd	Glidewell, Todd	N	https://images.ctfassets.net/3cttzl4i3k1h/1MHPEnSoKcKglwiSygms/4f1b33ae54f4c14ea0f984c098002a62/headshot_todd_glidewell.jpg
52	4WJSbbtNn2W02ygcQQ2o80	Gordon	Steve	Gordon, Steve	N	https://images.ctfassets.net/3cttzl4i3k1h/3NwKAPP6MwCO0Wyg4qsmuc/674013ab1db26d941b332e89884a7faa/headshot_steve_gordon.jpg
53	6hKgliUEeclO2zccykOguG	Gray	Matt	Gray, Matt	Y	https://images.ctfassets.net/3cttzl4i3k1h/50WnQfUf7qSAMiUqUEOkYE/d08884f88d6e997615d599ae230e184/Matt_Gray.jpg
54	ZkH72Rj3aegcA6QEeilQY	Greve	Edward	Greve, Edward	N	https://images.ctfassets.net/3cttzl4i3k1h/6lRgN8vjtC6ySQ04eo8M2u/5f86a698b8c2746385d5c7ea0369c3/20170113.png
55	6uDdi3KAggcOiACGckiyMw	Grigsby	Ryan	Grigsby, Ryan	N	https://images.ctfassets.net/3cttzl4i3k1h/4bT2XiacbumqWuAYC4cY4m/53acabf5dfbf0e0b8703243f85778f35/headshot_ryan_grigsby2.jpg
56	3tGbSlO6vK8qKEeayIYS0c	Guevara	Adrian	Guevara, Adrian	N	https://images.ctfassets.net/3cttzl4i3k1h/AJiKnzF8HiColsKayK2/336955258da0f333223a613cd77327da/adrian.png
57	4gMPr6QczCugU8uis0wq4E	Hannan	Walker	Hannan, Walker	N	https://images.ctfassets.net/3cttzl4i3k1h/4GYNOsaS4Em002QmiQu4o6/c6479f8d19e917e5eeb1f7f5dcf92ba1/headshot_walker_hannan.jpg
59	1N1fKFmUuMG6YwiKEowcsw	Harris	Andrew	Harris, Andrew	N	https://images.ctfassets.net/3cttzl4i3k1h/33GORUYsZ2GW26wQGeMoEK/9209c49185789402ea67bf495906b586/headshot_andrew_harris.jpg