

Some issues relating to data metrics



Mike Taylor

Senior Product Manager, Informetrics

<http://orcid.org/0000-0002-8534-5985>

@herrison

Empowering
Knowledge

Data (and other) metrics are on our roadmap

- New metrics group in Elsevier
- Citation metrics, alternative metrics, social metrics, data metrics
- Proposed data metrics to Snowball Metrics consortium
- Broad agreement on need, business case, types of metrics
- Stumbling block is definition of data

Tempting thought!

- Borrow a bit from bibliometrics
- Borrow a bit from citations
- Borrow a bit from altmetrics
- But do it for numbers and code – et voila!

Tempting but a false friend

- The scale is citation-sized, not altmetric-sized
- E.g. 80% of articles get citations, 20% of articles get altmetric activity
- Complexity is more like altmetrics order
- 2.3M papers published in 2013 (according to Scopus), what percentage used data?
- Differences – citation is part of the way we work – altmetrics happens as a consequence of activity
- New forms of citation require activity *everywhere* and by *everyone*
- Authors, funders, infrastructure, publishers, consumers

What is data? Or “what is not data”?

- Almost everything is data
- Data-data
- Real-time data
- 2.3M articles published last year
- Audio transcripts
- Video
- Art



irsmend	readersciteu	facebookpost	twitterpostsco	redditpostsco	googlepluspo	videopostsco	f1000postsco	blo
223767	21801	44567	423311	5553	16840	1486	5804	
8168	2795	3978	13683	1504	2624	283	866	
28904	2734	5363	44766	931	1808	174	536	
2870	200	1327	10104	85	175	15	20	
53246	7081	8853	81808	1403	3877	388	2202	
3078	381	471	5903	17	110	12	0	
7452	572	1295	12786	102	355	137	124	
18749	1412	2464	33306	247	662	69	245	
5177	1458	632	8887	57	534	103	34	
861	118	92	1632	4	64	8	1	
7405	1361	1071	12670	103	315	91	13	
2755	168	432	5410	47	87	7	4	
2625	169	375	3961	52	96	14	1	
8116	651	1226	14401	104	417	111	62	
11735	1112	2111	18687	228	569	58	80	
11107	935	2012	18296	147	736	149	560	
8430	489	1035	15338	86	375	52	28	
3112	873	382	6755	34	353	49	30	
106828	7454	21699	203992	2005	8159	552	2912	
13775	2000	2563	22263	604	1380	154	458	
5691	257	1279	12321	147	353	17	41	
8332	426	1694	16497	298	514	43	187	
11877	1182	1652	22402	263	808	118	116	
10507	672	1874	18404	315	841	46	28	
16025	1208	3539	35817	222	606	57	19	
1951	49	330	3781	19	122	4	4	
1440	13	202	2391	4	29	0	7	
3882	137	1103	7459	45	505	14	17	

More prosaically

- The type of data posted on Figshare – no curation, no authority
- Is different from data posted on Dryad (requires authority)
- Is different from data in data journals (requires curation)
- Is different from data in a community repository (requires on-going curation)
- Is different from primary data sources
- We need definitions
- We need infrastructure / standards