Online Information Review

Figshare: a universal repository for academic resource sharing?
Mike Thelwall Kayvan Kousha

Article information:

Users who downloaded this article also downloaded:

(2016),"Review on event detection techniques in social multimedia", Online Information Review, Vol. 40 Iss 3 pp. 347-361 http://dx.doi.org/10.1108/OIR-08-2015-0281

(2016),"Can two-sided messages increase the helpfulness of online reviews?", Online Information Review, Vol. 40 Iss 3 pp. 316-332 http://dx.doi.org/10.1108/OIR-07-2015-0225

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Figshare: a universal repository for academic resource sharing?

Mike Thelwall and Kayvan Kousha
School of Mathematics and Computer Science,
University of Wolverhampton, Wolverhampton, UK

Abstract

Purpose – A number of subject-orientated and general websites have emerged to host academic resources. The purpose of this paper is to evaluate the uptake of such services in order to decide which depositing strategies are effective and should be encouraged.

Design/methodology/approach – This paper evaluates the views and shares of resources in the generic repository Figshare by subject category and resource type.

Findings – Figshare use and common resource types vary substantially by subject category but resources can be highly viewed even in subjects with few members. More active subject areas do not tend to have more viewed or shared resources.

Research limitations/implications – The view counts and share counts analysed may reflect author accesses or may be spammed.

Practical implications – Limited uptake of Figshare within a subject area should not be a barrier to its use. Several highly successful innovative uses for Figshare show that it can reach beyond a purely academic audience.

Originality/value – This is the first analysis of the uptake and use of a generic academic resource sharing repository.

Keywords Data sharing, Digitization, Repository, Figshare

Paper type Research paper

Introduction

Scientific resource sharing is important in research for reasons of efficiency, power and rigour. At the most basic level, making experimental data available allows others to check calculations or to replicate a study, which is central to rigorous science (*Nature*, 2015; Sieber, 1991). Moreover, some studies need data on a scale that requires sharing. For example, identifying diseases from brain scans requires access to large numbers of healthy and diseased examples obtained from organised data sharing (Poline *et al.*, 2012) using common standards (e.g. Demir *et al.*, 2010). More generally, sharing any kind of academic resource can aid efficiency by ensuring that scholars do not have to needlessly repeat prior work. An important example of this is the software R, which contains many statistical procedures written by academics (e.g. Rosseel, 2012) and freely shared for others to use.

Scholars can use the web to disseminate electronic resources, including software, datasets, internal reports, and digitised art and cultural artefacts (Schubert *et al.*, 2013; Schopfel *et al.*, 2014). In genetics and environmental science, for example, datasets are significant research outputs and are often shared (Anagnostou *et al.*, 2013). Data sharing can have practical challenges (Borgman, 2012) and researchers may be cautious (Huang *et al.*, 2012; Vogeli *et al.* 2006; Walport and Brest, 2011), but there is a strong tradition of sharing resources in some fields (e.g. software in computer science), for some types of general data (e.g. surveys: UKDA, 2007) and for specialist information, such as species records in biodiversity research (Faith *et al.*, 2013; Moritz *et al.*, 2011; see also: Barve, 2014) and human biological samples (Chen, 2013).

Moreover, there is an argument that all publically funded research should publish its data (Arzberger, *et al.*, 2004; Walport and Brest, 2011), and many funding agencies, including the European Commission (2015), the National Science Foundation (Hswe and Holt, 2011) and the UK's Wellcome Trust (Wellcome, 2015), require scientific data sharing when possible for their funded projects. Some, like the National Institutes of Health in the USA, support scientific data repositories, including its Cancer Imaging Archive (NLM, 2015). About half of all academic journals seem to have a data sharing policy for submitted articles and a quarter have data sharing mandates (Sturges *et al.*, 2014) – including PLOS journals (PLOS, 2015; Figshare, 2013) and *Nature* (2015). Finally, there are also peer reviewed data journals, including the *Geoscience Data Journal*, although they may describe rather than host resources (Chavan and Penev, 2011; Costello *et al.*, 2013; Gorgolewski *et al.*, 2013). Given the wide variety of ways in which data can be shared it is fortunate that researchers seem to be willing to use others' data (Tenopir *et al.*, 2011), although it is not clear how widespread data reuse is.

Although academics can use their web CVs to publicise their outputs (Kousha and Thelwall, 2014), resource sharing is supported by general purpose scientific repositories, such as Figshare, which began in 2011 and supports researchers from all disciplines to deposit any type of electronic information online for use by others. Collections of files deposited in Figshare are given a digital object identifier to encourage users to credit the originators of the resources used by citing it in formal publications. It is not clear whether general repositories can be as successful as sites that target specific user groups or that have additional features to support specific uses, however. Dryad, for example, targets evolutionary biologists, archiving data associated with their publications (Greenberg, 2009). In contrast, the National Geophysical Data Center supports the collaborative creation of large scale comprehensive databases, including an international map of magnetic anomalies (Maus *et al.*, 2007). Similarly, SourceForge and Google Code provide tools to support the collaborative construction of computer software (Thelwall and Kousha, 2016; Nyman and Mikkonen, 2011). Within the arts and humanities, individual research may rarely generate sharable data, but digitisation initiatives (Gorman, 2007) create shared resources of various types, such as photographs of artworks, music recordings, historical records or cultural artefacts (e.g. Abd Manaf and Ismail, 2010; Alonso Gaona García *et al.*, 2014).

Although there have been studies of individual scholarly resource sharing sites, as cited above, they have focused on individual disciplines and types of resource, such as datasets. Investigations of general multidisciplinary repositories are also needed in order to obtain advice for researchers and research managers about who should use them and how useful they are for the different types of resource. There seem to have been no such articles published yet, with the exception of opinion pieces (e.g. Singh, 2011). Figshare is the focus of the current paper because it does not target a specific discipline, allows multiple types of resource to be uploaded, and seems to be the main current example of this type of universal scientific repository. Although counts of citations to Figshare data could be used as indication of their intellectual impact (see, Chavan and Ingwersen, 2009, 2011), this is not a good choice because datasets are rarely cited (e.g. Piwowar *et al.*, 2007) and the same is probably true for other types of resource. View counts and shares are used here because they are published by Figshare and alternative online metrics, such views, saves and recommendations are more informative for digital resources (Konkiel, 2013). Usage or impact indicators derived from the web are widely used for academic articles (Thelwall and Kousha, 2015a, b) and

other academic outputs (Kousha and Thelwall, 2015) and so it seems reasonable to also use them for shared electronic resources. The following research questions drive this Figshare study:

RQ1. Are more uploaded resource types more viewed and more shared?

RQ2. Are the different resource types allowed by Figshare (papers, figures, data, files, media (e.g. sound, video), posters, presentations, theses, code) uploaded more in different subjects?

RQ3. Are resources from the most used subjects more viewed and more shared?

RQ4. Does the average level of viewing and sharing of specific resource types vary by subject category?

Methods
The overall research design was to gather a sample of Figshare members and then to compare the number of views of their deposits by field and by resource type. Figshare does not provide comprehensive lists of members and so an indirect method was chosen to identify them. Each Figshare member has a profile page that lists up to five outputs together with their view counts. These pages are indexed by search engines and Bing was used to search for them. Although Google may have a larger index of the web, Bing was chosen because it allows automatic searching and no gaps were found in its results for Figshare. A separate query was used for all view counts between 0 and 20,000 (larger than the maximum number found using Figshare's browse option and selecting "Most viewed" in the "Sort by" field) in order to get a complete set of active Figshare profiles, at least in theory. For example, the query for 48 views would match all profile pages containing at least one resource that had been viewed 48 times: "48 views" site: figshare.com/authors. One of the results from this query was the page http://figshare.com/authors/Estrella_Lopez_martin/524824, which had 48 views at the time but has subsequently attracted additional views.

The queries were submitted by the free software Webometric Analyst (http://lexiurl. wlv.ac.uk) on 27 May 2015. This process produced a list of Figshare profile pages that had a least one resource with less than 20,000 views, excluding profiles not indexed by or not reported by Bing. Additional ad-hoc tests based upon browsing Figshare suggested that the list of profiles from Bing was close to comprehensive because all pages found by browsing Figshare were also in the Bing search results. Although search engines do not comprehensively index the web (Lawrence and Giles, 1999) the goal of Figshare is to publicise the shared resources and so it seems reasonable to assume that they would take steps to ensure timely and comprehensive indexing by the major search engines.

All of the 2,753 profile pages found were downloaded by the free web crawler SocSciBot (http://socscibot.wlv.ac.uk; Thelwall, 2001) on 27 May 2015. A new menu item was then added to Webometric Analyst (citations menu: extract information from downloaded Figshare profiles) to extract the subject categories and resources from each downloaded profile page, including view counts and share counts. This programme exploits the standardised structure of the downloaded profile pages to automatically extract their information. Each share corresponds to a visitor, or the owner, clicking on a button in Figshare to post a dataset link in Twitter, Facebook or Google+. If a user had more than five of one particular type of resource then only the

most viewed five were displayed in their profile page. The maximum view count and share count was calculated for each user and resource type from the profile pages downloaded. A few (78) of the profiles listed no resources, leaving 2,675 for analysis.

Figshare users do not declare a specialist field but are asked to enter one or more categories for each uploaded resource. These fields predominantly describe the subject area of the resource, although the category computer software is also used despite it not being an academic field description. The five most frequent categories are displayed on profile pages. These were taken to be the specialist subject areas of the user even though they varied from broad (e.g. education) to narrow (e.g. interstellar and intergalactic matter). Figshare has 14 categories on its main resource browsing page (figshare.com/articles/browse), but allows uploading authors to select additional categories from a longer list and 157 were reported in the profiles downloaded. The top 20 categories were selected for further analysis.

The view count and share count data was highly skewed with many small numbers and so the arithmetic mean is not a good measure of central tendency and the median is too crude to differentiate between areas. Hence the geometric mean was calculated instead (the mean of the log of the data), with an offset of 1 due to the presence of 0. This data was approximately normally distributed. Geometric means are reported as the antilog of the arithmetic mean of 1+views, with 1 subtracted from the antilog (for details see: Thelwall and Fairclough, 2015).

Results

The analysis of the 2,675 Figshare member profiles with resources is organised by research question.

RQ1: uploading vs viewing and sharing

There are statistically significant differences between the geometric means of the number of views of each resource type, as evidenced by the non-overlapping 95 per cent confidence intervals (Table I). For example, media resources tend to be viewed more often than all the others and datasets send to be viewed less than all other types. Answering *RQ1*, more uploaded resource types do not tend to be more viewed. There is a weak trend in the opposite direction (Spearman correlation −0.283, $n = 9$, not statistically significant).

Resource	Users uploading 1 or more	Min. views	Max. views	Lower 95% mean views	Geometric mean views	Upper 95% mean views
Media	111 (4%)	14	2,571	110.1	138.8	174.9
Posters	457 (17%)	4	15,356	82.9	92.7	103.6
Files	727 (26%)	0	28,383	78.0	87.2	97.5
Presentations	343 (12%)	8	1,985	67.0	75.7	84.6
Papers	840 (31%)	2	34,182	66.4	73.4	82.1
Theses	90 (3%)	8	29,492	50.4	67.7	90.8
Figures	497 (18%)	1	16,024	52.5	61.2	71.2
Code	113 (4%)	4	1,129	37.5	44.6	53.1
Datasets	1,283 (47%)	0	7,336	29.0	31.5	34.2
All	2,753	0	75,700	6.7	6.9	7.2

Notes: Resources are listed in decreasing order of views. There were 4,894 resources in total

Table I.
Number of uploading members and geometric mean views (with 95 per cent confidence intervals) of each resource type

The general pattern of sharing resources (Table II) is similar to that for viewing resources. There is no evidence that more uploaded resource types get shared more (Spearman correlation −0.517, $n = 9$, not statistically significant). The extremely high number of shares for a thesis (and one of the highest overall for Figshare) is for "Irregularidades Temporo-espaciales E Hipótesis De Los Cuerpos Desacelerados" by Pedro Javier Villanueva Hernández. This 2008 thesis was published in Figshare on 20 April 2015 and had received 838 shares and 29,500 views within three months by 6 June 2015 (Villanueva Hernández, 2015) but had received only 149 more views and two more shares after an additional five months on 16 November 2015. It is about designing strategies and solutions to research problems, which may help new researchers even though it was uncited in Google Scholar by 16 November 2015. The unusual pattern of early apparent interest suggests that it either generated a huge amount of publicity in the mass media when it was published or its figures were spammed. Web and social media searches for the thesis suggested that the latter was more likely. Both views and shares in Figshare could be derived from the owner or other visitors.

RQ2: uploading frequency by resource type and subject category
The percentage of members uploading each resource type varies substantially by category (Table III). For example, 68 per cent of members declaring the education category uploaded "data" resources in contrast to 33 per cent from molecular biology. Some of the differences are to be expected from the nature of the subject, such as code being more uploaded by users tagging with computer software than with other categories, although in this case the tag may be a resource description rather than an academic subject category. The relevant academic fields could be computer science, computing or software engineering, for example. Most strikingly, however, 68 per cent of economics users uploaded papers in contrast to 18 per cent of computational biology users despite papers being highly relevant to both groups.

RQ3: viewing and sharing by category
The Spearman correlation between the geometric mean total number of views for a category's members (i.e. members recording that category for at least one of their resources) and the number of users in the category ($n = 157$ categories) is 0.019, which is not significant at the 0.05 level. Similarly, the correlation between the geometric mean

Resource	Members	Min. shares	Max. shares	Lower 95% mean shares	Geometric mean shares	Upper 95% mean shares
Media	111 (4%)	0	63	1.29	1.79	2.40
Posters	457 (17%)	0	214	0.97	1.16	1.36
Files	727 (26%)	0	162	0.93	1.08	1.23
Theses	90 (3%)	0	838	0.87	1.40	2.09
Presentations	343 (12%)	0	56	0.77	0.94	1.12
Papers	840 (31%)	0	793	0.72	0.83	0.95
Figures	497 (18%)	0	173	0.66	0.81	0.97
Datasets	1,283 (47%)	0	76	0.35	0.40	0.46
Code	113 (4%)	0	14	0.21	0.35	0.50
All	2,753	0	1,905	0.4	0.4	0.5

Notes: Resources are listed in decreasing order of shares. The share count is the number of times that somebody has used Figshare to promote a resource, such as through Facebook, Twitter or Google+

Table II.
Number of uploading members and geometric mean shares (with 95 per cent confidence intervals) of each resource type

Category	Papers (%)	Figures (%)	Data (%)	Files (%)	Media (%)	Posters (%)	Pres. (%)	Theses (%)	Code (%)
Education	39	**9**	*68*	*13*	*2*	9	9	*2*	1
Ecology	25	23	59	30	4	29	21	4	4
Bioinformatics	22	20	44	41	4	*36*	22	5	*10*
App. comp. sci.	45	17	36	40	6	28	28	4	8
Biological sci.	**19**	*74*	*65*	*50*	6	13	**9**	3	6
Evolut. biology	30	31	44	*50*	6	26	25	4	5
Economics	*68*	**9**	52	**12**	**2**	**5**	**8**	3	1
Medicine	28	25	49	33	4	26	12	4	**1**
Environ. science	41	21	45	24	3	30	24	6	5
Genetics	23	31	37	46	*7*	32	18	*7*	7
Neuroscience	22	32	39	45	*11*	23	12	4	4
Science policy	44	*35*	52	42	*7*	20	23	4	4
Comput. biology	**18**	18	38	35	6	*41*	29	*7*	9
Lib. and info. stud.	42	15	46	32	2	22	*29*	6	4
Sociology	*51*	22	43	29	3	15	13	4	3
Microbiology	21	25	41	43	5	22	**9**	4	**2**
Molec. biology	25	30	**33**	36	5	35	14	4	6
Comp. software	37	14	**34**	31	**2**	21	27	*7*	*21*
Info. systems	50	20	39	27	5	21	25	4	6
Statistics	39	25	35	36	**1**	19	*29*	**2**	9

Note: The two highest percentages in each column are italic and the two lowest are bold

Table III.
Percentage of users uploading at least one of the resource types, by category, for the top 20 categories

number of shares and the number of users in a category is 0.135, which is also not significant at the 0.05 level. Thus there is no evidence that the larger categories tend to attract views at a different rate to smaller categories.

RQ4: viewing and sharing differences by subject for different resource types
There are differences between subjects in the extent to which a specific type of resource is viewed or shared (see Tables IV and V). The differences between some pairs of categories are statistically significant, at least for the most common types of resource, in the sense that their 95 per cent confidence limits do not overlap (Figures 1-4). For example, papers from education and economics members tend to get viewed and shared substantially less than papers from science policy members. These results cannot be explained by individual prolific figures uploading many papers because the results are averaged by uploader rather than by uploaded resource. It may be, however, that a field has systematically encouraged paper uploading, such as for a conference, and this has naturally produced high view counts. The confidence intervals tend to overlap for less used resources and so statistical methods would not reveal differences between categories for these due to a lack of statistical power.

Discussion and conclusions
Although in answer to the first research question the more uploaded resources are not more viewed or shared, this does not mean that they are not more valuable. Datasets are the most common type of resource, perhaps because Figshare is a natural site for data and because of encouragement from journals and funders to share data, but datasets are the least viewed and second least shared. Nevertheless, an average of over 30 views per dataset is substantial, given that they are presumably only of interest to people wishing for detailed knowledge of any associated paper or wanting to

Category	Geo. mean paper views	n	Geo. mean figure views	n	Geo. mean data views	n	Geo. mean file views	n	Geo. mean media views	n	Geo. mean poster views	n	Geo. mean pres. views	n	Geo. mean thesis views	n	Geo. mean code views	n
Education	**57**	141	88	34	**12**	248	152	46	294	8	114	33	82	34	112	8	40	3
Ecology	169	76	60	70	44	182	92	91	113	12	99	88	87	64	113	11	70	12
Bioinformatics	188	63	152	56	79	126	160	116	123	12	116	100	91	61	81	13	48	27
App. comp. sci.	169	96	156	35	*90*	75	161	86	328	12	*203*	59	*117*	59	80	9	43	16
Biological sciences	103	39	20	155	**22**	150	46	107	161	12	158	27	79	18	50	6	*109*	12
Evolut. biology	218	55	164	57	76	80	120	92	161	11	145	47	100	46	*132*	7	50	12
Economics	**52**	124	60	17	32	95	98	22	67	3	**63**	10	**41**	14	54	6	91	1
Medicine	103	51	**46**	47	35	92	90	61	147	8	95	47	60	21	37	7	**32**	1
Environ. science	118	72	56	37	48	82	123	42	102	6	78	52	68	43	*142*	11	58	9
Genetics	*232*	39	*193*	52	78	62	*182*	77	236	12	123	53	103	30	42	11	51	11
Neuroscience	89	37	85	53	43	65	**85**	75	216	18	98	38	107	19	47	6	47	7
Science policy	*249*	72	55	57	60	87	138	69	*511*	11	*204*	33	*108*	38	46	6	*91*	7
Comput. biology	213	29	120	29	85	60	109	55	109	9	122	65	99	46	99	11	44	14
Lib. and info. stud.	187	61	61	21	*101*	67	121	47	825	3	90	31	59	42	72	9	84	6
Sociology	93	70	81	30	42	64	133	41	95	4	115	20	**43**	18	135	5	38	4
Microbiology	65	27	138	32	57	53	109	55	67	6	112	28	77	12	89	5	46	3
Molecular biology	109	31	165	38	65	41	124	46	448	6	113	44	71	18	42	5	49	8
Comp. software	110	37	**52**	14	57	34	**86**	31	**40**	2	**63**	21	72	27	**37**	7	**40**	21
Info. systems	136	48	64	19	57	37	86	26	278	5	66	20	59	24	47	4	44	6
Statistics	224	37	*239*	24	76	33	*200*	34	**31**	1	91	18	100	28	**26**	2	42	9

Notes: The two highest numbers in each column are italic and the two lowest are bold. For each user with a particular resource type, the maximum views for any of their resources of that type is used

Table IV.
Geometric mean
views for resource
types, by category

Category	Geo. mean paper shares	n	Geo. mean figure shares	n	Geo. mean data shares	n	Geo. mean file shares	n	Geo. mean media shares	n	Geo. mean poster shares	n	Geo. mean pres. shares	n	Geo. mean thesis shares	n	Geo. mean code shares	n
Education	**0.7**	141	1.5	34	**0.1**	34	*2.6*	247	2.8	46	2.1	8	1.5	33	*3.9*	8	1.0	3
Ecology	1.5	76	0.9	70	0.5	70	1.4	180	2.2	90	1.1	12	1.2	88	1.2	11	0.2	12
Bioinformatics	1.7	63	1.1	55	0.8	55	1.6	124	1.8	116	1.1	12	1.4	100	2.1	13	0.2	27
Applied comp. sci.	1.4	96	*2.9*	35	0.8	35	2.1	75	2.9	12	*3.1*	12	1.1	59	0.8	9	0.4	16
Biological sciences	1.2	39	**0.2**	154	1.0	154	**0.5**	136	2.6	105	1.6	12	**0.3**	27	**0.0**	6	0.5	12
Evolut. biology	2.2	55	1.3	57	1.0	57	1.2	80	3.6	92	1.6	11	1.1	47	0.6	7	0.3	9
Economics	**0.3**	124	0.9	17	**0.2**	17	1.1	95	0.6	22	1.2	3	0.6	10	0.8	6	**0.0**	1
Medicine	1.0	51	0.4	46	0.6	46	1.2	89	1.8	60	1.1	8	0.5	47	0.8	7	*3.0*	1
Environ. science	1.3	72	0.7	37	0.4	37	1.3	80	1.3	42	0.8	6	1.3	52	2.0	11	0.1	9
Genetics	2.2	39	1.4	52	0.9	52	2.0	61	3.0	76	1.5	12	1.1	53	0.7	11	0.4	11
Neuroscience	0.9	37	1.4	53	0.6	53	0.9	64	3.8	75	1.4	18	*1.7*	38	0.6	6	0.2	7
Science policy	*3.0*	72	1.3	57	*1.2*	57	1.8	86	*4.6*	69	2.8	11	1.0	33	**0.1**	6	0.9	7
Comput. biology	2.2	29	1.1	29	0.6	29	1.0	60	2.6	55	1.2	9	1.4	65	1.7	11	0.3	14
Lib. and info. stud.	2.8	61	1.1	21	*1.4*	21	2.1	66	3.3	46	1.6	3	0.7	31	2.1	9	0.9	6
Sociology	1.1	70	1.2	30	0.5	30	1.9	59	0.4	40	2.3	4	0.4	20	*3.1*	5	*1.2*	4
Microbiology	1.3	27	1.3	32	0.6	32	1.8	52	0.8	55	1.4	6	*3.0*	28	0.6	5	**0.0**	3
Molecular biology	1.3	31	1.2	38	0.4	38	2.2	41	*4.8*	45	1.1	6	**0.2**	44	0.7	5	0.8	8
Comp. software	1.1	37	**0.1**	14	0.4	14	0.9	34	**0.0**	31	**0.5**	2	1.1	21	0.9	7	0.5	21
Info. systems	1.8	48	0.9	19	0.6	19	**0.8**	37	1.4	26	0.7	5	0.6	20	2.2	4	0.6	6
Statistics	2.4	37	*2.6*	24	0.5	24	2.3	33	**0.0**	34	**0.6**	1	0.9	18	0.4	2	0.3	9

Notes: The two highest numbers in each column are italic and the two lowest are bold. For each user with a particular resource type, the maximum shares for any of their resources of that type is used

Table V.
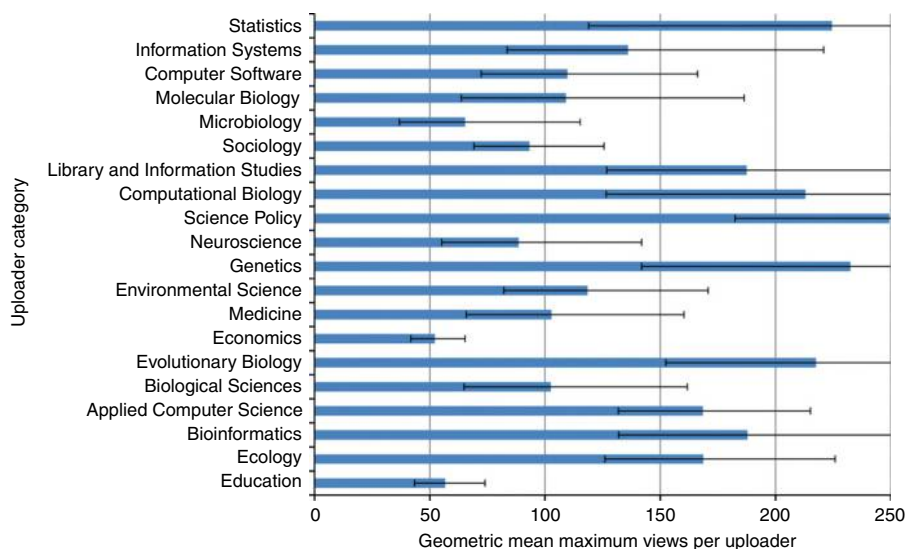Geometric mean shares for resource types, by category

Figure 1.
Geometric mean
maximum views
per member of
papers for the 20
most common
categories, together
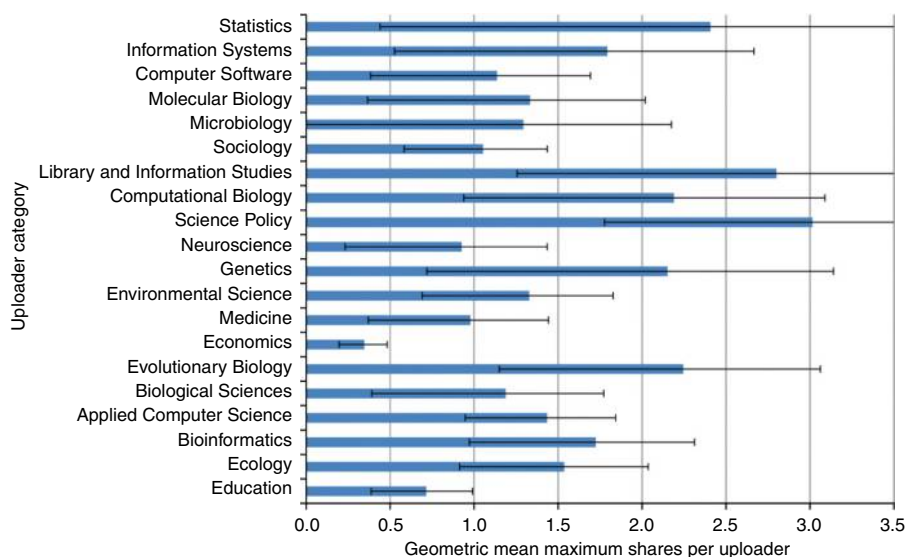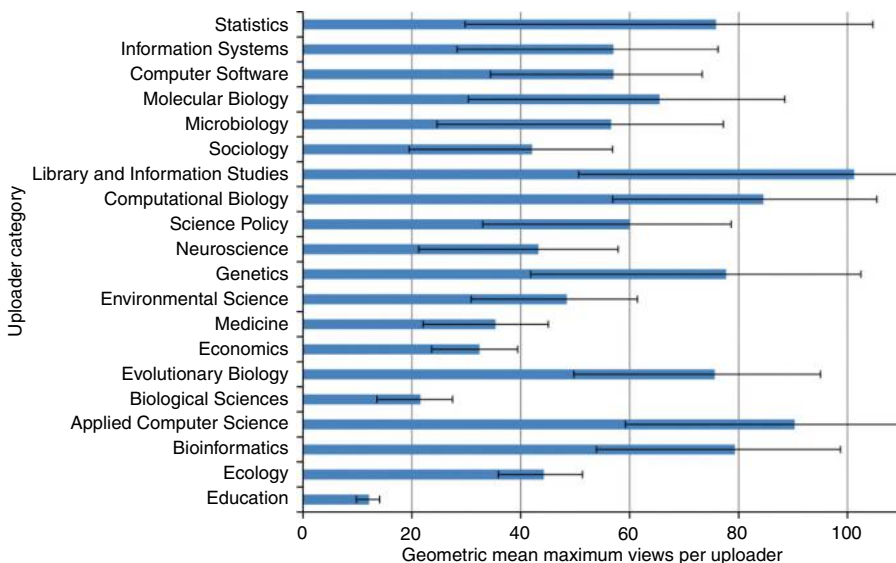with 95 per cent
confidence limits

Figure 2.
Geometric mean
maximum shares
per member of
papers for the 20
most common
categories, together
with 95 per cent
confidence limits

investigate the data for reuse. In this context, the raw view count and share count figures are misleading because datasets target a much narrower audience than, for example, a video, which could aim at the general public and attract millions of views (Sugimoto and Thelwall, 2013). As previously argued, context is important when interpreting impact data for non-standard academic outputs (Thelwall and Delgado, 2015). Moreover, an individual dataset or piece of software seems more likely to provide substantial help to future researchers than the other resource types and so each individual use may be more valuable. In this context, the average of 30 views per

Figure 3.
Geometric mean
maximum views
per member of
datasets for the
20 most common
categories, together
with 95 per cent
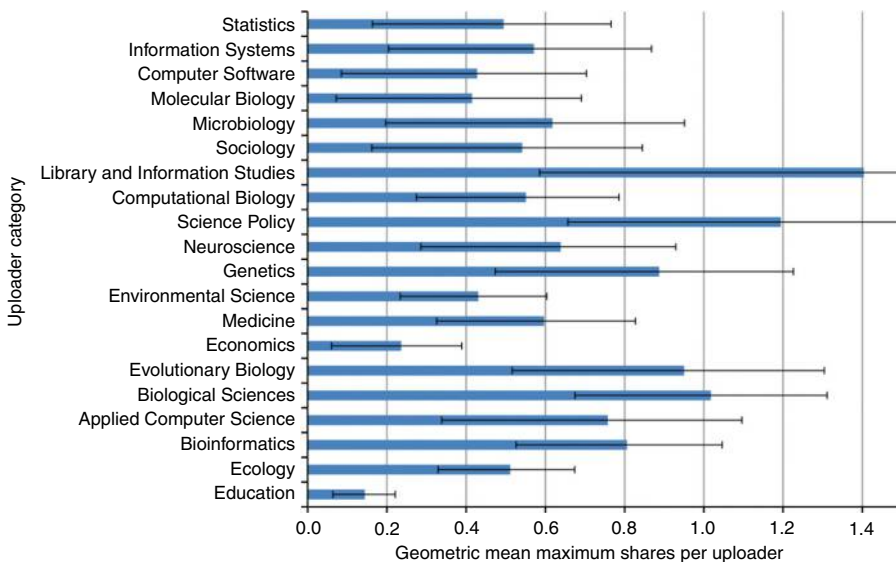confidence limits

Figure 4.
Geometric mean
maximum shares
per member of
datasets for the
20 most common
categories, together
with 95 per cent
confidence limits

dataset is encouraging although the lack of a way to assess how many of these views are genuine and useful makes this a tentative conclusion. Overall, however, the results give no suggestion that any particular resource type is ignored in Figshare and so it seems reasonable for funders and journals to continue to encourage dataset sharing and for academics to consider sharing wider types of outputs.

In answer to the second research question, the extent to which particular types of resources are uploaded varies substantially by subject, even though there is at least one

resource of each type in all of the top 20 subjects. The hosting of a particularly high percentage of economics papers, despite the existence of a specialist repository for economics research (Research Papers in Economics) suggests that disciplinary cultures influence what researchers share. Researchers might therefore consider field norms when deciding on the types of resource that they might be expected to share. From the fourth research question, the extent to which a resource type is viewed or shared varies significantly by category, and so researchers should also consider uploading less used resource types if they are popular within their subject category, when uploaded.

For the third research question, resources in subject categories with the most uploads do not tend to be the most viewed or shared because even resources in small categories can be highly shared. The differences between subjects found could be due to the utility of the specific types of resources uploaded or disciplinary cultures of data reuse, although the subject categories used here probably only very broadly reflect the home disciplines of the scholars concerned. For example, Graham Steel, a neuroscientist and open science advocate is categorised as library and information studies for a comedy paper (apparently published only in Figshare) about blank pages in academic publications (Wright *et al.*, 2014), with 6,992 views. The descriptions of the resources are also not always correct, which may also affect the results to some extent. For example, at least one dataset is categorised as code.

Some of the uploaded resources have generated particularly much interest. Within library and information studies, a dataset of information about UK university journal subscription costs attracted 3,947 views (Lawson *et al.*, 2014) and a similar dataset of open access spending by the Wellcome Trust funding agency (Kiley, 2014) attracted 3,352 views. In addition, one (beautiful) single page poster with the title, "101 Innovations in Scholarly Communication – the Changing Research Workflow" had attracted 5,629 views (Kramer and Bosman, 2015). Finally, a set of three dimensional images of a dinosaur skeleton attracted 28,494 views (Lacovara, 2014). These numbers are large and plausible enough to suggest wider professional or educational contributions to their subjects. Thus, although Figshare is ostensibly an academic site, the seamlessness of the internet apparently allows non-academic uptake of the resources that it hosts.

Overall, successful use of Figshare is not limited to any particular discipline, resource type or audience. Although there are differences between subject areas in the average popularity of their uploaded resources, there is not a simple pattern that some disciplines have many users and their resources tend to attract many views. Instead, it seems that people from many different subject areas have found effective uses for the repository and so innovation in its use should be encouraged across disciplines and even for non-academic audiences. The suggestions given above should be taken as general guidelines about what types of resource to share but individual researchers should be guided by their own understanding about which of their outputs might prove useful to other researchers. Finally, nothing in the analysis suggests that any type or subject category of resource tends to be ignored, when shared, which gives some evidence to support current attempts to encourage sharing between researchers, and not just for data.

References

Abd Manaf, Z. and Ismail, A. (2010), "Malaysian cultural heritage at risk? A case study of digitisation projects", *Library Review*, Vol. 59 No. 2, pp. 107-116.

Alonso Gaona García, P., Martín-Moncunill, D., Sánchez-Alonso, S. and Fermoso García, A. (2014), "A usability study of taxonomy visualisation user interfaces in digital repositories", *Online Information Review*, Vol. 38 No. 2, pp. 284-304.

Anagnostou, P., Capocasa, M., Milia, N. and Bisol, G.D. (2013), "Research data sharing: lessons from forensic genetics", *Forensic Science International: Genetics*, Vol. 7 No. 6, pp. e117-e119.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P. and Wouters, P. (2004), "Promoting access to public research data for scientific, economic, and social development", *Data Science Journal*, Vol. 3, pp. 135-152.

Barve, V. (2014), "Discovering and developing primary biodiversity data from social networking sites: a novel approach", *Ecological Informatics*, Vol. 24 No. 1, pp. 194-199.

Borgman, C.L. (2012), "The conundrum of sharing research data", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 6, pp. 1059-1078.

Chavan, V. and Penev, L. (2011), "The data paper: a mechanism to incentivize data publishing in biodiversity science", *BMC Bioinformatics*, Vol. 12 No. S15, p. S2.

Chavan, V.S. and Ingwersen, P. (2009), "Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community", *BMC Bioinformatics*, Vol. 10 No. S14, p. S2. doi: 10.1186/1471-2105-10-S14-S2.

Chen, H. (2013), "Governing international biobank collaboration: a case study of China Kadoorie Biobank", *Science Technology & Society*, Vol. 18 No. 3, pp. 321-338.

Costello, M.J., Michener, W.K., Gahegan, M., Zhang, Z.Q. and Bourne, P.E. (2013), "Biodiversity data should be published, cited, and peer reviewed", *Trends in Ecology & Evolution*, Vol. 28 No. 8, pp. 454-461.

Demir, E., Cary, M.P., Paley, S., Fukuda, K., Lemer, C., Vastrik, I., Wu, G., D'Eustachio, P., Schaefer, C., Luciano, J., Schacherer, F., Martinez-Flores, I., Hu, Z., Jimenez-Jacinto, V., Joshi-Tope, G., Kandasamy, K., Lopez-Fuentes, A.C., Mi, H., Pichler, H., Rodchenkov, I., Splendiani, A., Tkachev, S., Zucker, J., Gopinath, G., Rajasimha, H., Ramakrishnan, R., Shah, I., Syed, M., Anwar, N., Babur, O., Blinov, M., Brauner, E., Corwin, D., Donaldson, S., Gibbons, F., Goldberg, R. *et al.* (2010), "The BioPAX community standard for pathway data sharing", *Nature Biotechnology*, Vol. 28 No. 9, pp. 935-942.

European Commission (2015), "Guidelines on open access to scientific publications and research data in Horizon 2020", available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (accessed 18 November 2015).

Faith, D., Collen, B., Ariño, A., Koleff, P., Guinotte, J., Kerr, J. and Chavan, V. (2013), "Bridging the biodiversity data gaps: recommendations to meet users' data needs", *Biodiversity Informatics*, Vol. 8 No. 2, pp. 41-58.

Figshare (2013), "Figshare partners with open access mega journal publisher PLOS", available at: http://figshare.com/blog/figshare_partners_with_Open_Access_mega_journal_publisher_PLOS/68 (accessed 9 June 2015).

Gorgolewski, K.J., Margulies, D.S. and Milham, M.P. (2013), "Making data sharing count: a publication-based solution", *Frontiers in Neuroscience*, Vol. 7 No. 9. doi: 10.3389/fnins.2013.00009.

Gorman, G.E. (2007), "The DeLorean or the Mini? Digital imperatives for publishers, digital dilemmas for repositories", *Online Information Review*, Vol. 31 No. 6, pp. 741-743.

Greenberg, J. (2009), "Theoretical considerations of lifecycle modeling: an analysis of the dryad repository demonstrating automatic metadata propagation, inheritance, and value system adoption", *Cataloging & Classification Quarterly*, Vol. 47 Nos 3-4, pp. 380-402.

Hswe, P. and Holt, A. (2011), "NSF data sharing policy", available at: www.arl.org/focus-areas/e-research/data-access-management-and-sharing/nsf-data-sharing-policy (accessed 9 June 2015).

Huang, X., Hawkins, B.A., Lei, F., Miller, G.L., Favret, C., Zhang, R. and Qiao, G. (2012), "Willing or unwilling to share primary biodiversity data: results and implications of an international survey", *Conservation Letters*, Vol. 5 No. 5, pp. 399-406.

Ingwersen, P. and Chavan, V. (2011), "Indicators for the data usage index (DUI): an incentive for publishing primary biodiversity data through global information infrastructure", *BMC Bioinformatics*, Vol. 12 No. S15, p. S3.

Kiley, R. (2014), "Wellcome Trust APC spend 2012-13: data file", Figshare. doi:10.6084/m9. figshare.963054.

Konkiel, S. (2013), "Tracking citations and altmetrics for research data: challenges and opportunities", *Bulletin of the American Society for Information Science and Technology*, Vol. 39 No. 6, pp. 27-32.

Kousha, K. and Thelwall, M. (2014), "Disseminating research with web CV hyperlinks", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 8, pp. 1615-1626.

Kousha, K. and Thelwall, M. (2015), "Web indicators for research evaluation, Part 3: books and non-standard outputs", *El Profesional de la Información*, Vol. 24 No. 6.

Kramer, B. and Bosman, J. (2015), "101 innovations in scholarly communication – the changing research workflow", Figshare. doi:10.6084/m9.figshare.1286826.

Lacovara, K. (2014), "Dreadnoughtus Schrani 3D PDF images", Figshare, available at: http://dx.doi.org/10.6084/m9.figshare.1130885

Lawrence, S. and Giles, C.L. (1999), "Accessibility of information on the web", *Nature*, Vol. 400 No. 6740, pp. 107-109.

Lawson, S., Meghreblian, B. and Brook, M. (2014), "Journal subscription costs – FOIs to UK universities", Figshare. doi:10.6084/m9.figshare.1186832 (accessed 10 May 2016).

Maus, S., Sazonova, T., Hemant, K., Fairhead, J.D. and Ravat, D. (2007), "National geophysical data center candidate for the world digital magnetic anomaly map", *Geochemistry, Geophysics, Geosystems*, Vol. 8 No. 6. doi: 10.1029/2007GC001643.

Moritz, T., Krishnan, S., Roberts, D., Ingwersen, P., Agosti, D., Penev, L., Cockerill, M. and Chavan, V. (2011), "Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF data publishing framework task group", *BMC Bioinforma*, Vol. 12 No. S15, p. S1.

Nature (2015), "Availability of data, material and methods", available at: www.nature.com/ authors/policies/availability.html (accessed 9 June 2015).

NLM (2015), "NIH data sharing repositories", available at: www.nlm.nih.gov/NIHbmic/nih_data_ sharing_repositories.html (accessed 9 June 2015).

Nyman, L. and Mikkonen, T. (2011), "To fork or not to fork: fork motivations in sourceforge projects", *International Journal of Open Source Software and Processes*, Vol. 3 No. 3, pp. 1-9.

Piwowar, H.A., Day, R.S. and Fridsma, D.B. (2007), "Sharing detailed research data is associated with increased citation rate", *PLoS ONE*, Vol. 2 No. 3, p. e308. doi: 10.1371/journal.pone.0000308.

PLOS (2015), "Data availability", available at: http://journals.plos.org/plosone/s/data-availability (accessed 9 June 2015).

Poline, J.B., Breeze, J.L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Haselgrove, C., Helmer, K.G., Keator, D.B., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J. and Kennedy, D.N. (2012), "Data sharing in neuroimaging research", *Frontiers in Neuroinformatics*, Vol. 6 No. 9. doi: 10.3389/fninf.2012.00009.

Rosseel, Y. (2012), "Lavaan: an R package for structural equation modelling", *Journal of Statistical Software*, Vol. 48 No. 2, pp. 1-36.

Schopfel, J., Chaudiron, S., Jacquemin, B., Prost, H., Severo, M. and Thiault, F. (2014), "Open access to research data in electronic theses and dissertations: an overview", *Library Hi Tech*, Vol. 32 No. 4, pp. 612-627.

Schubert, C., Shorish, Y., Frankel, P. and Giles, K. (2013), "The evolution of research data: strategies for curation and data management", *Library Hi Tech News*, Vol. 30 No. 6, pp. 1-6.

Sieber, J.E. (1991), "Openness in the social sciences: sharing data", *Ethics & Behavior*, Vol. 1 No. 2, pp. 69-86.

Singh, J. (2011), "FigShare", *Journal of Pharmacology & Pharmacotherapeutics*, Vol. 2 No. 2, p. 138.

Sturges, P., Bamkin, M., Anders, J. and Hussain, A. (2014), "Access to research data: addressing the problem through journal data sharing policies", *Proceedings of the IATUL Conferences, Helsinki, 2-5 June*, available at: http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2012&context=iatul (accessed 9 June 2015).

Sugimoto, C.R. and Thelwall, M. (2013), "Scholars on soap boxes: Science communication and dissemination via TED videos", *Journal of the American Society for Information Science and Technology*, Vol. 64 No. 4, pp. 663-674.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M. and Frame, M. (2011), "Data sharing by scientists: practices and perceptions", *PLoS ONE*, Vol. 6 No. 6, p. e21101. doi: 10.1371/journal.pone.0021101.

Thelwall, M. (2001), "A web crawler design for data mining", *Journal of Information Science*, Vol. 27 No. 5, pp. 319-325.

Thelwall, M. and Delgado, M. (2015), "Arts and humanities research evaluation: no metrics please, just data", *Journal of Documentation*, Vol. 71 No. 4, pp. 817-833. doi: 10.1108/JD-02-2015-0028.

Thelwall, M. and Fairclough, R. (2015), "Geometric journal impact factors correcting for individual highly cited articles", *Journal of Informetrics*, Vol. 9 No. 2, pp. 263-272.

Thelwall, M. and Kousha, K. (2015a), "Web indicators for research evaluation, Part 1: citations and links to academic articles from the web", *El Profesional de la Información*, Vol. 24 No. 5, pp. 587-606. doi: 10.3145/epi.2015.sep.08.

Thelwall, M. and Kousha, K. (2015b), "Web indicators for research evaluation, part 2: social media metrics", *El Profesional de la Información*, Vol. 24 No. 5, pp. 607-620. doi: 10.3145/epi.2015.sep.09.

Thelwall, M. and Kousha, K. (2016), "Academic software downloads from Google code: useful usage indicators?", *Information Research*, Vol. 11 No. 1, available at: www.informationr.net/ir/21-1/paper709.html

UKDA (2007), "Across the decades: 40 years of data archiving", available at: http://data-archive.ac.uk/media/54761/ukda-40thanniversary.pdf (accessed 9 June 2015).

Villanueva Hernández, P.J. (2015), "Irregularidades Temporo-espaciales E Hipótesis De Los Cuerpos Desacelerados", Figshare. doi:10.6084/m9.figshare.1385169.

Vogeli, C., Yucel, R., Bendavid, E., Jones, L.M., Anderson, A., Louis, K.S. and Campbell, E.G. (2006), "Data withholding and the next generation of scientists: results of a national survey", *Academic Medicine*, Vol. 81 No. 2, pp. 128-136.

Walport, M. and Brest, P. (2011), "Sharing research data to improve public health", *The Lancet*, Vol. 377 No. 9765, pp. 537-539.

Wellcome (2015), "Data sharing", available at: www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/ (accessed 10 May 2016).

Wright, G., Coudert, F.X., Bentley, M., Deville, S. and Steel, G. (2014), "This study is intentionally left blank – a systematic literature review of blank pages in academic publishing", Figshare. doi:10.6084/m9.figshare.1230110.

Corresponding author

Mike Thelwall can be contacted at: m.thelwall@wlv.ac.uk

This article has been cited by:

1. Michael Thelwall. 2016. Web Indicators for Research Evaluation: A Practical Guide. *Synthesis Lectures on Information Concepts, Retrieval, and Services* **8**:4, i-155. [CrossRef]