

# Milestone M2.19

## Report on the options for a biodiversity data citation metric for data published through ViBRANT

This report will detail the options available to ViBRANT to enable the assessment of an individual's contributions to a project and the impact of those contributions. An initial test case will be developed using the Scratchpads, with the principals being valid for all mechanisms of publishing data from ViBRANT, and also to non-ViBRANT projects.

### Authors

E.Baker, S.Rycroft, D.Roberts

Combination of Collaborative Project and Coordination and Support Action  
Virtual Biodiversity Research and Access Network for Taxonomy  
Grant Agreement No: 261532  
Project Co-ordinator: Dr Vince Smith  
Project Homepage: <http://vbrant.eu/>  
Duration of Project: 36 months  
Start Date: Dec 2010  
End Date: Nov 2013

## Introduction

Metrics are used in academic disciplines as a proxy to measure the scholarly output of individual researchers, the relevance of publications and the relative quality of journals. As such they are often used by authors to judge where is best to publish and by managers as a factor in who to promote or recruit.

Traditionally a number of metrics have been used as tools to measure scholarly output. As the vast majority of published academic output to date has been in the form of articles (paper or PDF) published in journals the accepted metrics have focussed on the number of citations articles produced by researchers (h-index) and journals (impact factor) receive. The pressure has been on scientists to publish highly cited articles in high impact journals. Numerous authors have both criticised these measures and offered improvements (Ruane & Tol, 2008) or alternatives (Bourne and Fink, 2008).

This citation-is-king methodology, even where it can be argued as a good proxy for overall scholarly output, misses a large section of activity. Modern scholarly practices often include a wide range of other activities including public engagement, conference presentations, posters and the collection of datasets. Whereas public engagement is difficult to quantify, many of these scholarly activities can be published through online tools, such as SlideShare, and some quantitative judgment made about their impact.

The need for a broader, more inclusive, metric is perhaps best illustrated by a researcher working on a study that requires a long period of intensive data collection and curation which results in one well regarded and highly cited paper. In terms of h-index (and perhaps career) improvement they may have been better, at least in the short term, producing a small number of lower quality papers that each get cited a handful of times.

After a review of the presently available alternative metric options in the 'Alternative Metrics' section our solutions to these issues, and the problems we have encountered, in the sections Biodiversity Data Journal and Scratchpads Metrics.

In addition to these researcher/publication measurements there is a need for those involved in large scale projects such as ViBRANT to measure the performance of projects. There are a number of ways of assessing such projects, including the traditional methods of assessing the project itself via the publications it has resulted in, and using presentations, posters and training sessions as a proxy for the hard to define concept of "impact". ViBRANT has indeed collated these data, which are available at <http://vbrant.eu/biblio> and <http://vbrant.eu/talks> respectively. The nature of ViBRANT in general, and in particular the Scratchpads, as data-centric online projects also allows for a finer-grained level of data collection relating to usage of these projects. These are discussed in the section Scratchpads Statistics.

For this work we separate these two concepts by using the term "metrics" to refer to people

centric measurements and “statistics” to refer to project centric measurements.

## Alternative Metrics

In recent years a number of projects, nested under the ‘alternative metrics’ or ‘altmetrics’ banner have attempted to broaden the way impact is measured - using social media data to widen the definition of impact.

The altmetrics movement (not be confused with the alternative metrics provider called altmetric”) recognises that public interest and engagement in research is crucial, and that increasingly scientists communicate using online tools for sharing data and content (e.g. Slideshare for presentations; figshare for images and other content). See Weller (2011).

Alongside the altmetrics movement, services such as figshare now allow academics to assign digital object identifiers (DOIs) to work other than traditional publications, including diagrams, posters and slide presentations. As DOIs have become the de facto method of tracing citations these services allow for a wealth of previously uncitable content to gain the recognition it deserves.

Combining altmetrics with an expanded scope of citable content should allow for a more comprehensive measure of scholarly production.

## Addressing metrics and statistics in ViBRANT

As well as allowing scholars to be valued on the entire output of their work there is a need in ViBRANT to assess the usage of the products we have developed or created. Scratchpads are at the core of the ViBRANT services ecosystem, with X many users across X many sites.

## Biodiversity Data Journal

The Biodiversity Data Journal (BDJ - Smith et al, 2013) was created under the ViBRANT project and launched in September 2013 and provides a publication venue for datasets and the long-tail of biodiversity data: single species descriptions, local floristic treatments, etc. All papers published in the BDJ are assigned a Digital Object Identifier (DOI). DOIs are the standard identifiers used for scholarly papers, and increasingly for books (Weller, 2011) and presentations (Baker, 2013). The use of DOIs allows for easy citation analysis, so by assigning DOIs to BDJ papers authors may gain credit for publishing their data. The BDJ also allows for publishing metadata descriptions of online datasets, allowing researchers to get a DOI citable publication from their online databases.

By using predefined templates and an online-only workflow the cost to the publisher of producing articles is significantly reduced. The BDJ is therefore an affordable, quick and easy avenue for people to publish and get rewarded for the masses of ‘small data’ that they produce.

## Scratchpads Metrics

Scratchpads (Smith et al, 2011) are an online tool for collaboratively publishing biodiversity information on the web used by researchers worldwide. In developing the Metrics module we have attempted to answer two questions: how much has a person contributed to this particular site, and how reputable are they as a researcher? The first question is easy to address using data already available to the Scratchpads team. The second question is more problematic. We have attempted an implementation of the Scholar Factor suggested by Bourne and Fink (2008) which combines the traditional h-index with other factors including grant.manuscript reviews, commits to software repositories, datasets and web outputs such as scholarly blog posts. Each of these signals is weighted to balance the influence of, for example, publications against manuscript reviews. Since this was proposed in 2008 there has been scant progress in making much of this data available, and we have been limited to a handful of data sources (social media likes, GitHub commits, Google Scholar h-index and contributions to the Scratchpad site).

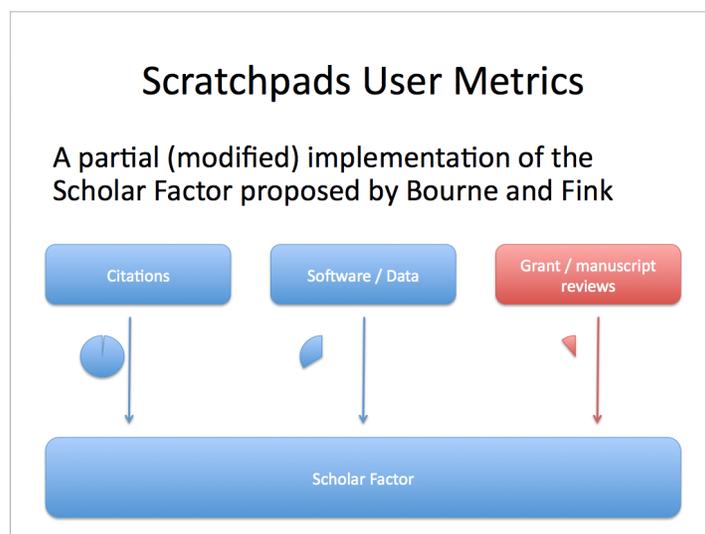
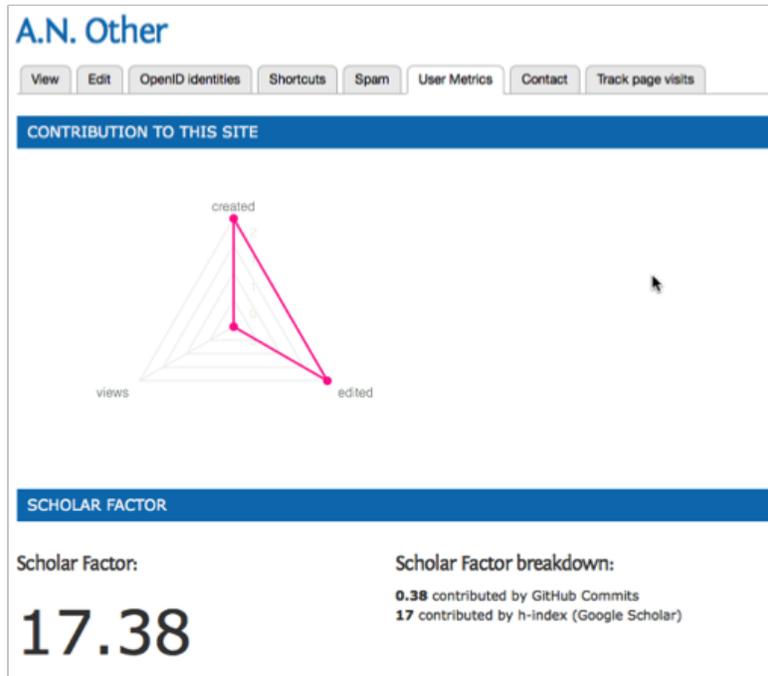


Figure showing selected signals and their weightings used in the Scratchpads User Metrics tool (from Baker, 2013).

We recognise that there is not a one size fits all metric that can be applied to all of our communities, therefore the Metrics module has been developed as a modular framework that allows community maintainers to tailor what sources contribute to the metrics on their site. As an example the Phasmid Study Group site mainly comprises of interested amateurs who in general do not have a scholarly publication record so the h-index component is likely to be of little use compared to an individual's contribution of content to that site. The modular framework also allows for future inclusion of other data sources as they become available, and for graceful retirement of services that cease to exist.



*The Scratchpads user metric display - showing that (top) A.N.Other has acted equally as a content creator and a content editor and (bottom) our partial implementation of Bourne and Fink's Scholar Factor, including the sources used to derive this score.*

It is hoped that a more widespread adoption of the Open Researcher & Contributor ID (ORCID - Haak et al, 2012) will encourage a more coherent method of accessing the contributions made by individual workers. In the interim period we are working with Pensoft (publishers of the BDJ, ZooKeys, etc) on a method of accessing the number of reviews completed by an individual for inclusion in our Scholar Factor implementation.

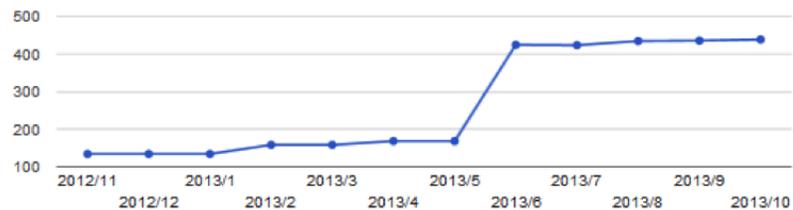
## Scratchpads Statistics

With the release of Scratchpads 2, a greater quantity of detailed statistics on data creation and data use is now being collected on each Scratchpad. Where possible, an item of data is associated with the user that created it, and the users that have edited it. Also where possible, an item of data is associated with a taxonomy term, which helps, categorise statistics, and the discovery of which particular areas individuals and groups are working on. The number of edits and views an item receives per week is also recorded. Weekly values for all statistics are recorded, which enables us to show how the data within Scratchpads is growing, week by week, and also allows us to analyze particular time periods (e.g. we can compare January 2012 with January 2013). The collection of this data helps us to create statistics on data usage and data creation. It is possible for us to show, for example, who is the most prolific author of data on a Scratchpad, or whose data has the greatest number of views on a Scratchpad.

## Scratchpads 2.0 statistics

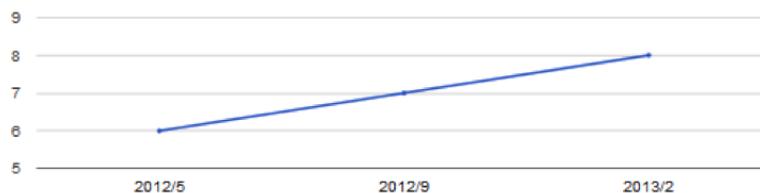
### CONTENT

The number of different objects in Scratchpads (incl. taxon descriptions, bibliographic entries, localities, specimens and images)



### ACTIVE USERS

Registered users logged-in at least once to their Scratchpad sites



### VIEWS

The number of times Scratchpads objects have been viewed

## Conclusions

The concept of comprehensive metrics for both publications (i.e. those metrics beyond traditional citation analysis), and there is not yet enough community agreement on what should be measured, and how to access and combine data from numerous potential sources. The work we have done with the BDJ makes data a first class citizen of the citation world, alongside traditional publications. We consider this to be the best available compromise in the current environment, and it is robust mechanism for continued use in the alternative metrics movement as it develops over the coming years.

Metrics for researchers beyond the h-index are defined, and the model of Bourne and Fink's Scholar Factor is one we particularly like as it can be expanded to incorporate any number of measurable scholarly activities, each of which can be appropriately weighted. The issue at present is the lack of available data beyond the h-index and newer collaborative tools such as GitHub. The wider adoption of ORCID's to identify authors would help to facilitate the sharing of data about grant and manuscript reviews.

It is difficult to give a significant value to metrics based on website usage statistics. Website usage statistics are often heavily skewed by factors unrelated to the content on a single page,

and should only be compared with statistics for similar sites, or preferably, the same site. If trying to provide metrics across many different websites, then the metrics data for each individual site would need to be normalised before being compared. Therefore, we can say that the Scratchpads statistics we are generating are useful for measuring of usage and effort across the Scratchpads, but are not of worth when comparing against other non-Scratchpad websites. Additional work, over and above what has been carried out under ViBRANT, would be required to be able to meet this requirement, and even then, with current technology, this would be a difficult product to deliver.

## Future Work

The development of metrics and statistics will be continued post-ViBRANT to continually assess the Scratchpads project, and to allow people and communities to measure their impact and influence. The following are a list of projects that we will be considering in the near future:

- Working with our ViBRANT partner Pensoft to investigate how to get access to the number of manuscript reviews completed by a scholar
- Investigating the COUNTER project (<http://www.projectcounter.org/>) to see whether we can make data about the use of our electronic resources available for others to access
- Implementing ORCID integration into the User Metrics functionality to access an increasing amount of an individual's academic output as the service obtains wider adoption

## References

Baker E (2013): Measuring Impact. figshare. <http://dx.doi.org/10.6084/m9.figshare.865022>

Bourne PE, Fink JL (2008) I Am Not a Scientist, I Am a Number. PLoS Comput Biol 4(12): e1000247. doi:10.1371/journal.pcbi.1000247

[Frances Ruane](#) & [Richard S. J. Tol](#) (2008). "Rational (successive) h -indices: An application to economics in the Republic of Ireland". [Scientometrics](#) 75 (2): 395–405. [doi:10.1007/s11192-007-1869-7](https://doi.org/10.1007/s11192-007-1869-7).

Haak L, Fenner M, Paglione L, Pentz E, Ratner H (2012) ORCID: a system to uniquely identify researchers. *Learned Publishing* 25:259-264

Smith V, Rycroft S, Brake I, Scott B, Baker E, Livermore L, Blagoderov V, Roberts D (2011) Scratchpads 2.0: a Virtual Research Environment supporting scholarly collaboration, communication and data publication in biodiversity science. *ZooKeys* 150: 53-70. doi: 10.3897/zookeys.150.2193

Smith V, Georgiev T, Stoev P, Biserkov J, Miller J, Livermore L, Baker E, Mietchen D, Couvreur T, Mueller G, Dikow T, Helgen K, Frank J, Agosti D, Roberts D, Penev L (2013) Beyond dead

trees: integrating the scientific process in the Biodiversity Data Journal. Biodiversity Data Journal  
1: e995. DOI: [10.3897/BDJ.1.e995](https://doi.org/10.3897/BDJ.1.e995)

Weller M (2011) The Digital Scholar. Bloomsbury Academic. 256pp.  
<http://dx.doi.org/10.5040/9781849666275>