M2.11 – Report on Scratchpad usage statistics options

Usage statistics are a very important tool to help project management. Usage statistics help to prioritise further development of a project like ViBRANT. For this reason, we have put a lot of emphasis on collecting and analysing to a fine detail as many user statistics as possible. The following report relies heavily the paper Who Learn from Whom[1], written for the ViBRANT special issue of ZooKeys.

Collection

Scratchpad usage statistics will be collected using the Scratchpad registry. The registry is a ViBRANT deliverable, due to be delivered at the end of month 18. The registry is intended to run as a service at the NHM[2], and will allow Scratchpads hosted by both the NHM and other institutions or individuals, to report their presence, and to also report statistics about themselves. This process will be fully automated, and will be enabled by default on all Scratchpads. The exact configuration of the Scratchpad registry is yet to be finalised, although it is likely to be very similar to the GBIF[3] IPT[4] registry system[5].

We will also continue to collect usage statistics using the Google Analytics tool[6]. This currently aggregates statistics across all Scratchpads, and is therefore of as much benefit to us as it could be. The registry, could be integrated with an open source analysis tool such as Piwik[7] to enable us to be much more flexible with the usage statistics we collect.

What?

Wherever possible, as much detail of Scratchpad usage will be collected as possible. There are privacy concerns regarding this, although users will be advised in the terms and conditions of the Scratchpad that site usage statistics will be collected, and every effort will be taken to ensure that statistics are anonymized before publication or distribution.

It is our intention to produce a module registry system that is capable of storing any data sent to it from a Scratchpad. This will enable us to quickly and easily enable the collection of additional statistics as and when they are required. A default system will should be expected to collect, at the very least, the following statistics:

- Number of users, and statistics regarding the login of those users.

- Quantity of data entered into the site, including statistics regarding the types of data entered, and by whom.

- Statistics regarding the use of the sites be anonymous users (those not logged in to the site and providing content to it) will also be collected. These will be in addition to the data collected from Google Analytics or a similar tool.

- Metadata about the site, helping to group analysis of multiple Scratchpads together (e.g. Are bird Scratchapds more popular than banana Scratchpads?).

---

1 I. Brake, D. Duin, I. Van de Velde, V. Smith, S. Rycroft – In print – Who learns from whom? Supporting users and developers of a major biodiversity e-infrastructure.
2 Natural History Museum, London
3 Global Biodiversity Information Facility - http://www.gbif.org/
4 Integrated Publishing Toolkit - http://ipt.gbif.org/
5 http://code.google.com/p/gbif-registry/
6 http://www.google.com/analytics/
7 http://piwik.org/