Title	Piwik: background information
Version	1
Author(s)	Dieter Van Uytvanck
Date	2015-06-04
Status	Draft
Distribution	NCF
ID	CE-2015-0528

1   Status of stats.clarin.eu

The currently running Piwik installation (stats.clarin.eu) is running as a test service in Nijmegen. It is planned to be migrated to a computing centre as soon as possible (e.g. Jülich, in combination with the monitoring). After that migration we will officially invite all CLARIN centres to use this service.

2   Using PiWik is not obligatory

It is not obligatory to use Piwik as a CLARIN centre – one can still use another solution. We do however recommend it, as it is a handy tool and provides more or less standardized metrics.

3   What main metric to use

Proposal: count *visits* per month.

A *visit*[1] entails a subsequent series of interactions with a website, by a single visitor who did not access the website in the last 30 minutes. This corresponds roughly to a *session*[2] in Google Analytics.

Overall advantages of counting visits:

- Seems to be a widely used metric (and makes statistics comparable to Google Analytics and other Piwik installations)
- It is a pretty intuitive metric
- It does not expose technical implementation differences (e.g. pageviews on classic website vs a single page in combination with ajax-driven page updates)

4   Who has access to the statistics

For every centre one or more contact person(s) will get access to:
- the monitoring (Icinga) dashboard
- the Piwik dashboard

To keep things simple this will be the same persons. This will be managed via the Centre Registry. Both web applications will be accessible via federated login (in first instance via the CLARIN IdP).

---

[1] See http://piwik.org/faq/general/faq_36/

[2] See https://support.google.com/analytics/answer/2731565?hl=en

5   What to measure

Some broader discussion is needed to define what exactly constitutes a CLARIN service, and thus could be tracked via Piwik. A first (non-exhaustive) proposal is:

- CLARIN ERIC website
- National CLARIN consortium websites
- Centrally developed CLARIN web applications, targeting end-users: VLO, (Federated) Content Search engine, Virtual Collection Registry
- Other user-directed web applications, provided via national consortia or centres: WebLicht, Language Resource Inventory, WebMaus web application, …

6   How to measure

The most convenient and complete way to collect user statistics is to use the Piwik javascript-based tracker[3]. This works well for all browser-based access: web sites and web applications.

The second-best option – suitable for web services – is to use one of the programming-language dependent APIs[4]. In this case some additional work needs to be done by the application developer to pass on enough information to the Piwik server. If this does not happen, the resulting metrics will be incomparable with the ones produced by the javascript tracker (e.g. lacking page view counts).

7   Privacy concerns

As Do-Not-Track requests are respected (see next section) and IP-addresses are anonymized, the use of Piwik is compliant with the current EU-regulations on data protection (see also document CLARIND-AP3-015[5]).

8   Limits

The numbers that Piwik will report will certainly be underestimations. This is because Piwik honours the Do-Not-Track feature[6], respecting the wish of certain users not to be tracked/counted. Visits of such users (including all those using Internet Explorer 10 and Windows 8 – where Do-Not-Track is the default setting[7]) are not counted.

9   Technical background information

For developers[8] that want to learn more about how to use Piwik the following wiki page has been setup: https://trac.clarin.eu/wiki/Piwik

---

[3] See http://piwik.org/docs/tracking-api/#javascript-client-for-tracking-api
[4] See http://developer.piwik.org/guides/tracking-api-clients
[5] See http://www.clarin-d.de/en/documents/technical-documents.html
[6] See http://en.wikipedia.org/wiki/Do_Not_Track
[7] http://en.wikipedia.org/wiki/Do_Not_Track#Internet_Explorer_10_default_setting_controversy
[8] See http://clarin.eu/svn for instructions about getting access to the Trac