

The Europa Analytics (EA) is a single reporting environment with the common metrics solution for web analytics run in the European Commission.

EA is dealing with about 70GB/day of raw data.

History

The Europa Analytics is available since 2003. Originally called Europa Web Analytics (or web analytics tool for [Europa](#)).

In 2010 this service has been upgraded from the SAS Web Hound to SAS Web Analytics. Although from Web Analytics 5.2 on SAS supports the page tagging capability, the system is actually uses the weblog analysis. Feasibility studies to implement tagging technology were performed for mobiles' analysis (June 2012).

Condition of Use

Prerequisites

The weblogs have to be provided in the [BlueCoat](#) format. Other formats might be possible but not guaranteed, and have to be analysed case by case.

The responsible of the website (web-master, web-manager, communication officer ...) has to [request the delivery of the weblogs](#).

Prior notification of the DG COMM

DG COMM must be informed prior to the launching of the website that you wish to use the EA. DG COMM has to configure the system in order to process your data - there is no possibility to process data backwards.

Europa Analytics Reports

The EA reporting tool offers a long list of the [generic reports](#). Together with a list of provided standard template reports, reports can also be customised by users and saved for future use. Access to the tool and other information with a practical user's guide can be accessed via the ['Reporting Tool'](#).

A simplified interface has also be introduced end 2011. The simplified interface (aka '[Kiosk](#)') provides standard reports easily accessible and open (no account is needed) to all EC users.

Indicators, Reports and Statistics offered by the EA

The EA does not produce statistics in stricto sensu, but just [indicators](#) (as we do not assess an error margins of the methods used to collect the data, nor the error margin of the used metrics as such, nor the methods used to produce the indicators).

Thus these indicators produced should be interpreted only as trends within a wider time span and/or

combined with other indicators. Even in combination, much caution should be taken, bearing in mind external factors.

To make these trends more obvious the indicators are presented in collections as [reports](#).

Software Details

The Europa Analytics is currently powered by the SAS Business Intelligence WebAnalytics 5.3.3 on SAS 9.2 M3. The previous version used SAS WebHound 4.1. (c.f. also [EA on IPG](#)).

ETL of the EA System

[Extract, Transfer and Load Procedure](#)

Backup of the EA System

[Backup of EA](#)

Configuration

Webmart ID

It is composed in total by 10 numbers composed by 6 digits root for the [Webmart](#) itself plus 4 digits suffix for section (in other words a section ID include also the ID of the webmart to which it belongs).

- continuous numbering - note that once used ID could not be used again
 - 1st free ID in the sequence available
 - starting with 000000 + 0000: or all Europa nest
 - next 000001 + 0000 including the a-z (ANSI) characters (that is, after 000009 the next is 00000a). Attention: ID is case sensitive - caps should be avoided!
 - from 7777777777 to 9999999999: for other specific purposes (redirections etc)
- Legacy (based on the IPG classification [\[1\]](#)) (should be abandoned as we have not the information stored in the specific meta field)
 - 01xxxx - Institution-Independent Websites
 - 02xxxx - European Commission Websites
 - 03xxxx - Offices and Agencies, and Delegation Websites
 - 05xxxx - Court of Justice Websites

Section ID

They follow the same rules as the webmarts. There is the "0000" by default and then after using 1st 10 digits we continue with "a-z" (ANSI - see above).

Ignored File Types

Following file types are ignored by EA:

- images with extension:
 - jpeg
 - jpg
 - gif

- bmp
- ico
- png
- jpe
- non pages with extension:
 - inc
 - css
 - rss
 - js
 - eot (since February 2014)
 - txt (since February 2014)
 - ttf (since mid-June 2014)
 - swf (since mid-June 2014)
 - json (since mid-June 2014)

In addition, webmaster can add to their called page or file with special file type the keyword : ea-ignore=true. This allow them to avoid pages counted involved in Ajax-calls or iframes.

Webmaster should pay attention to the above mentioned situation in order not to get uncoherent results.

Content Languages

(used in the EA report "[Languages](#)")

Following languages are defined in the EA:

- bg
- cs
- da
- de
- el
- en
- es
- et
- fi
- fr
- ga
- hr
- hu
- it
- lt
- lv
- mt
- nl
- pl
- pt
- ro
- sk
- sl
- sv

languages keys in static URL

following url characteristics will be treated as language ; text containing (*xx=language*) not case sensitive

- *_xx*. (default in [IPG](#))
- */xx/* (originated from some WCMs, like DRUPAL, WordPress ...)
- */xx?*
- */xx_*
- *:xx:* (EUR-Lex URI)
- *_xx/*
- *_xx?*
- *_xx* (for DRUPAL)

languages keys in dynamic urls (uppercase and lowercase letters)

- 'lang='
- 'lng='
- 'lg=' (preferred variable name)
- 'ihmlang='
- 'language='
- 'CL=' (from CELEX)
- 'L_id='
- 'langue='
- 'locale='
- 'target=' (from bookshop) using the 4th value (separator is :)

Content Dynamic URL keywords

(used in the EA reports "[Pages](#)" and everywhere a URL can be specified)

Dynamic urls contains keywords and values that have a meaningful value for the application server. Many keywords are session related or even random values and have no value for web analytical tools, in many cases they are all deleted. Of course there are a big number that are meaningful, but unknown for the EA administrators. Currently in EA we keep those that are requested by our users. This helps for limiting the number of unique URLs.

keywords that are rejected

- all related to date of request
- all related to userid and password
- all session related
- all random values

keywords kept in dynamic urls

- catid=

- cv_ed=
- doc=
- firstletter=
- form=
- gamename=
- method=
- pcp=
- rechtype=
- ref=
- reference=
- screen=
- template=
- userinput=
- val=
- videoref=
- cid=
- cv_id=
- dosid=
- filename=
- format=
- fuseaction=
- name=
- prodid=
- semaine=
- vardate=
- parentid=
- s_ref=
- uri=
- cl=
- id=
- ihmlang=
- l_id=
- lang=
- langid=
- language=
- langue=
- lg=
- lng=
- lng_id=
- locale=
- plang=
- obj_id=
- orgid=
- porgid=
- sid=
- uid=
- L_id=
- dt_code=
- intpageid=
- subject=
- comp_id=

- formid=
- phf=
- itemid=
- keywords=
- root=
- tag=
- thematic=
- type=
- query=
- submit=

- ea-ignore=

More information

- [Europa Analytics Documentation](#)
- [Europa Analytics - Frequently Asked Questions](#)

Errors/Bugs...

[Bugs on Europa Analytics](#) contains all known bugs and/or needed or wanted improvements