

The problem:

The web development team wanted to use data on how people use ebi.ac.uk to guide how we made changes to the website <links to case studies once done>. We were already gathering website analytics. However, it's [unlikely that web data is squeaky-clean](#). Thus as we used the data we had to keep a few things in mind.

- 1) Multiple links to the same URL on one page
- 2) Internal IP ranges
- 3) The edges of our tracking abilities

Likely these are issues that other EBI teams are likely to run into regardless of web analytic platform. (Don't believe me? Check out our comparison of platform choices<add link>). So here's the rundown of those three issues with details on how to figure out if you have that issue and what our solution was for ebi.ac.uk.

Multiple links to the same URL on one page

Do you have this issue?

You do if one page has multiple links to another page, and you want your website tracking to differentiate between on which link the user clicks.

Why ebi.ac.uk has this issue

Frequently links are repeated in the header, footer, and body of a web page. Take for example the home page, before scrolling down the page I have four different links to ebi.ac.uk/research I could click on.

The screenshot shows the EMBL-EBI website home page. The top navigation bar includes links for Services, Research, Training, and About us. The Research link is highlighted with a red box. Below the main header, there is a section titled 'The European Bioinformatics Institute' with a sub-header 'Part of the European Molecular Biology Laboratory'. A paragraph describes the institute's mission, with the words 'basic research' highlighted in a red box. A search bar is present with the text 'Find a gene, protein or chemical:'. Below the search bar, there are several tiles for different sections: Databases and tools, Services, Research (highlighted with a red box), Bioinformatics training, Training, Industry Programme, ELIXIR, European Coordination, EMBL Alumni Relations, and EMBL. To the right, there is a 'Popular' section with links to Services, Research (highlighted with a red box), Training, News, Jobs, Visit us, EMBL, and Contacts. Below this is a 'Visit EMBL.org' section with the EMBL 40th anniversary logo. Further down, there is an 'Upcoming events' section featuring the SWAT4LS International Conference. At the bottom, there is a 'News from EMBL-EBI' section with three articles: 'Unpacking embryonic pluripotency', 'Allergy: the price of immunity?', and 'Enzyme research made easier'.

If we want to know which of the links visitors are using to navigate, we need our web analytics to differentiate between the links on our pages.

The solution

See the post by Ken Hawkins about two possible solutions for Piwik <link once posted on confluence>. If you use Google Analytics, Google recommends [link enhancement](#). We ended up using JS tracking for ebi.ac.uk's Piwik analytics.

Internal IPs

Do you have this issue?

If a large number of EBI or Sanger employees using your website it's likely that you have internal IPs appearing in your log data. The effect of these internal IPs is likely to be noticeable if a large percentage of your visitors are internal (e.g. CTTV <[link to Niki's DRAFT doc](#)>) or they use the website differently from external visitors.

Why ebi.ac.uk has this issue

The way that employees use our website is likely different from how non-employees use our site. Employees have reported being largely interested in finding different information (e.g. contact information) and are likely to be power users of only a few services. External visitors tend to be academics or from industry and could be looking for directions to our campus or be using a service new to them that discovered through a search on EBI.

The solution

Segment the data based on the [internal IP ranges](#). This is the filter we use on ebi.ac.uk's Piwik to do data analysis on our internal IPs:

The screenshot shows the Piwik segmentation interface. On the left, a sidebar lists various categories: Events, Actions, Visit (metrics), Referrers, Visit Location, Visit, and Custom Variables. The main area is titled 'Name: EBI private network' with an 'edit' link. It displays a complex filter rule for 'Visitor IP' with three conditions connected by 'AND' operators. The first condition is 'Visitor IP' 'Less than' '172.31.255.255'. The second condition is 'Visitor IP' 'Greater than' '192.168.0.0'. The third condition is 'Visitor IP' 'Less than' '192.168.255.255'. Each condition has a '+ Add OR condition' button below it. At the bottom, there is a '+ Add AND condition' button. The interface also includes a 'Search' box, a 'Delete' button, a 'Close' button, and a 'Save & Apply' button.

We also have a filter for all external IPs. Set up these two filters and analyse the sets of data separately. Be aware that the geo-location data of internal IP addresses could be incorrect. It's a known issue for Piwik data and suspected to be correct for Google Analytics data based on an analysis of intranet web analytics <link once written up>.

The edges of tracking data

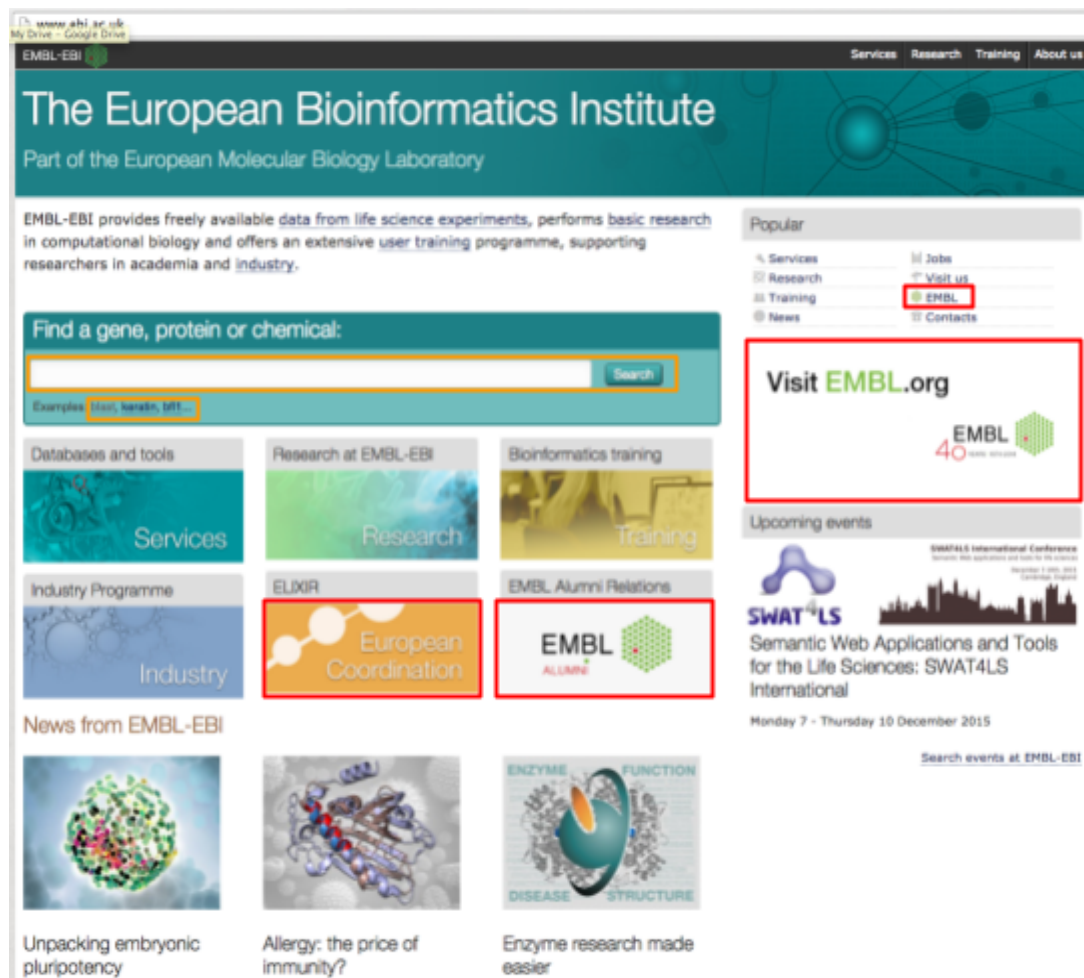
Do you have this issue?

If you're part of a collaboration with links to other services it's likely that you should look into whether this impacts your data. It could increase the number of recorded outlinks and return visitors. It could also increase the number of recorded bounces. The more users navigate from your website to pages without your tracking code and back again, the bigger the impact there could be on the data.

Why ebi.ac.uk has this issue

EBI offers many services and websites. Although the URL may be ebi.ac.uk/<something>, the ebi.ac.uk domain has websites on it owned by the different teams and running different technologies. Most visitors going to several pages on ebi.ac.uk will eventually use our search bar or one of the EBI services -- pages that exclude our tracking code.

On the homepage alone there are 8 places a user can click without scrolling that would take them to a webpage without our tracking code.



Orange boxes are links to EBI services that uses different data tracking while red boxes are to non-EBI websites (e.g. EMBL or Elixir).

The solution

Talk with your collaborators about installing the same tracking code across the main pages that visitors use. If you can't extend your tracking code to encompass your visitor's movements (likely the case), you have several options:

- 1) Use the data with care (more details on how in next paragraph)
- 2) Use the EBI server logs (solution for orange boxed links)

For the first option, you use your website data but should avoid measuring something that's impacted by "edges" of data tracking (e.g. the exact number of visitors or bounces). Rather, you can measure changes in those numbers. This assumes that you have not introduced a change to the number of "edges crossed" in the period you're analysing. For instance our drupal responsive fix project <link once written up> proved there was a significant reduction in the number of bounces before and after changing the website to be more mobile friendly. We could do this because the website update didn't change any links or extend our tracking code to new webpages.

The [server logs](#) record all transactions on EBI services. Contact the web production team to see if they can provide you with the visitor data you need.