

# ¿Active? Web listening What our users are saying

José Jabier Zurikarai, Mikel Bilbao

Eustat, Vitoria-Gasteiz, Basque Country – Spain, [jjzurikarai@eustat.eus](mailto:jjzurikarai@eustat.eus)  
Eustat, Vitoria-Gasteiz, Basque Country – Spain, [mikel@eustat.eus](mailto:mikel@eustat.eus)

---

## Abstract

In the world of official statistics we continuously need to know how our users are, what they needs are and whether we fulfill their expectations with our products and services. It's a long time ago that user satisfaction surveys were launched, but we know better than others about their limitations and boundaries. Our data publication web services have big amounts of users or web-traffic with unknown origin. But that anonymous traffic creates itself more information in the web that can be identified. We can identify not only traditional Media companies making news with our data. Thanks to web 2.0 and social there are more and more web resources where our data can be disseminated. So we have all those intermediaries helping us disseminating our data results. But they also are active agents that may give value to our information or even undervalue it. So, in this scene web listening becomes the tool to be used, the tool that makes possible to identify and assess what is being said about our organization, our statistics, or about our product and services.

With this active listening we have the ¿proper? features to build the image of those users that are using and sharing our information with others, so that we can improve our work in the dissemination areas.

We will show the results obtained for EUSTAT during the last months. First, a brief idea of how we should organize the job, the construction of keywords. This is essential regarding official statistics, because it shows the difference that can be found between our technical use of language and how those concepts are used in the internet. Then some comments about things that were unexpected and that we found out thanks to this technique or about the expansive wave effect of social network.

**Keywords:** Web listening, Users knowledge, Tool

## PAPER

A young girl is walking on the street and she can see an older man in a Café reading a newspaper (Oh, my god, he's reading what happened yesterday!)

A boy is in trouble because he's all day long with his smartphone and his parents are not paying for his phone line any more.

One girl from Bilbao is sending a whatsapp to her friend in Berlin and they are planning to meet next month in London.

Boy from the Gijon is recently 18. He will vote for first time in next elections. He follows in twitter Pablo Iglesias. Will he vote Podemos? Or will he vote PSOE as his older sister did?

Are these young people out of this world? Do they know what is happening in politics? Do they know what economic crisis is? Are they worried about their future?

Who knows, maybe they are, or maybe not. But do they read newspapers and try to get information from them?

Definitely no, they don't.

Last data of time use survey show that young people hardly read any printed newspaper. There is a constant reduction of time dedication to read printed media. In twenty years, the average percentage of young readers has dramatically fallen.

In this context we are facing big changes in the way mass-media are disseminating our statistics to the society. Antonio Caño, Editor-in-chief of EL PAIS (one of the Spanish leader newspapers) recently wrote a letter to the newspaper's staff "to explain the imminent transformation of the newspaper into a media outlet that is, above all, digital" he also says *"The revolution that is affecting the media has still not concluded, and the outlook is very confusing. The crisis, most likely, has not yet hit bottom. The migration of readers from print to digital is constant. We can already assume that the habit of buying a newspaper at a*

*newsstand is now the thing of a minority. The majority of people, in particular youngsters, look for their information on a range of different devices and they consume it in a different way.”*

This is something that will also affect to Official Statistics. We can no longer focus our attention just on printed media, not even on our webpage only but we must check also our impact in the digital world on the whole: other webs, blogs, forums, social media,. That will show us whether we are succeeding, it will give us an idea of our performance and it can also show us the perception of people about our job, so that we can collect not only quantitative user data, but qualitative too. Actually, this listening should (and will) provide valuable inputs for us to get better in our duty.

And then the question arises: how do we proceed?

There are several solutions in the market. We can buy software solutions, web based applications that can give us the information we need.

They usually offer scalable solutions, and they share at least two things:

- 1.-They check digital media a social network: Media, Blogs, FB and twitter
- 2.- A number of key words must be chosen to “listen” to them.

### **Main targets**

We seek three main targets with this project:

- .- Track activity related to our products and services.
- .- Compare performance with other institutions in order to change strategies, if needed.
- .- Find fields of opportunity, be it new niches or influent persons/institutions to work along.

### **Setting things up**

Once the software solution which best suits us is chosen, it’s time to decide which words should be monitored.

First things first, the obvious: branding. “Eustat”, “Instituto Vasco de Estadística” and “Euskal Estatistika Erakundea” are the first must-follow. We want to know what’s been said about us as institution and/or when/where are we being cited as data sources.

Besides, we want to get a grip of the use of our data, even when we are not cited as source (yes, it happens). To achieve that we must monitor keywords related to our activity that should be general enough not to miss anything, and specific enough not to be flooded with noise and unrelated data. It’s a tough compromise and we are learning by the minute. Fine tuning the keyword pool is probably a never-ending task.

First of all, we have the need of accuracy. It is absolutely necessary to choose the right words. For instance, GDP related news tend to be from official statistics, but Tourism is a subject much more used in the web, so we can hardly know the impact of our tourism related news releases. We must combine it with some other word.

Besides, if we choose “too much” words, then we have to clear up the information we collect and it can be tricky. So in the first steps we’d better choose significant words that we presume they will have a clear impact.

Our strategy, so far, consists in using the same keywords we’ve been using in our SEO efforts, meaning, key-phrases like “paro AND Euskadi”, “P.I.B. AND Euskadi” and the likes.

If we are to compare performance with other institution, we could add keywords accordingly (INE, IECA,...)

We are aware that we are missing a lot, for sure, but we hope to get better over time and, all in all, we think it always will be better than nothing.

### **Limitations**

The proper tuning of the system is essential because researching all possible mentions in every topic is a very time (and money) consuming task. Actually, that could be the main limitation: how much time and effort are we willing to put in this task.

Besides, there are other limitations more difficult or even impossible to override, for example, those related to closed/private environments, be it a password protected forum, pay-walled media or private posts in social media.

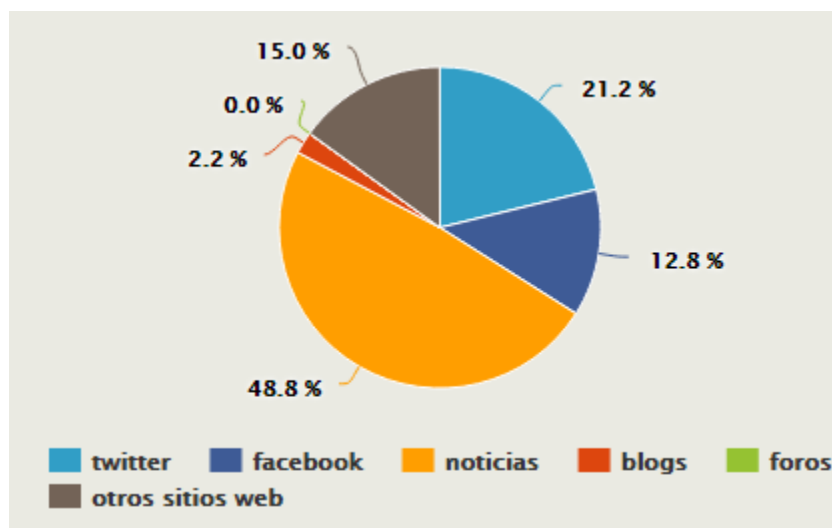
Other limitation found relates to language. The tool we've tested doesn't seem to track basque language properly.

### **Our findings, so far**

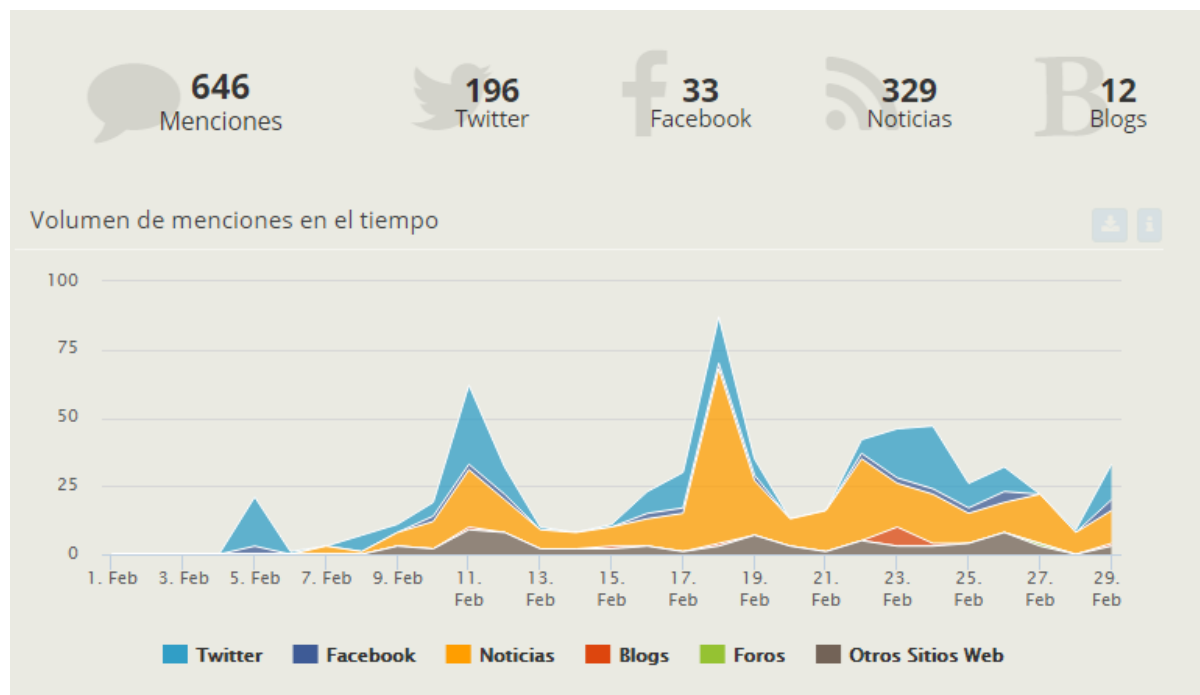
In our recent experience we started with web-listening and despite it's for a few months, there are some conclusions.

### *Quantitative*

First, what we knew and/or expected: we are more mentioned in the digital versions of traditional newspapers than anywhere else on the www. Those are followed by twitter and facebook, in that order. Even though we must consider that many of the Facebook posts are private and we can't keep track of them.



Spikes in mentions follow the publication our press releases, but (and we didn't know this) some topics seem to have longer reach, longer eco, than others. It's the case with GDP related news.



IT related topics have more impact in digital world than other topics. People who interact with our press releases are more likely to be interested in IT related statistics. It's their "natural" environment, so they have a personal interest about what's going on with R+D, the use of Internet, and similar topics.

### **Qualitative**

As previously said, with this listening we try to track not only quantitative data, but qualitative too. For that purpose, the tools provide automatically assigned "sentiment" (it must be checked for context). This sentiment is tagged positive, neutral and negative.

As for now, the vast majority of our sentiment records have been "neutral". Sure, we'd like them to be positive, but it's not bad considering our *business* which obliges us to be as neutral as we possibly can.

Nevertheless, it's worth mentioning that if this tool is to serve us, the "negative" mentions are probably the most interesting. Those which will help us better improving.