# follow.up: University of Porto news aggregator

**Tiago Devezas**[1] • **Sérgio Nunes**[2] • **Bruno Giesteira**[3]
[1] Communication Sciences Program, University of Porto, Portugal
[2] Department of Informatics, Faculty of Engineering, University of Porto, Portugal
[3] Department of Design, Faculty of Fine Arts, University of Porto, Portugal

## Motivation and goals

The University of Porto (UP) is a large organization, with dozens of organic units, each one publishing independently several articles per day. This hinders information consumption and potentiates content duplication.

We aim to minimize both these problems by:

- Collecting, storing and consolidating all the information on a database.
- Developing a centralized access point to all the published articles.

## Methodology

The data collection process started on June 2012. A database stores all the articles published through the RSS channels available on the SIGARRA platform of 15 UP organic units. Associated data, like the topic, channel and source the article was published under is also stored.

So far we've collected more than 6,000 articles published under 56 distinct topics through 108 channels from 15 sources.

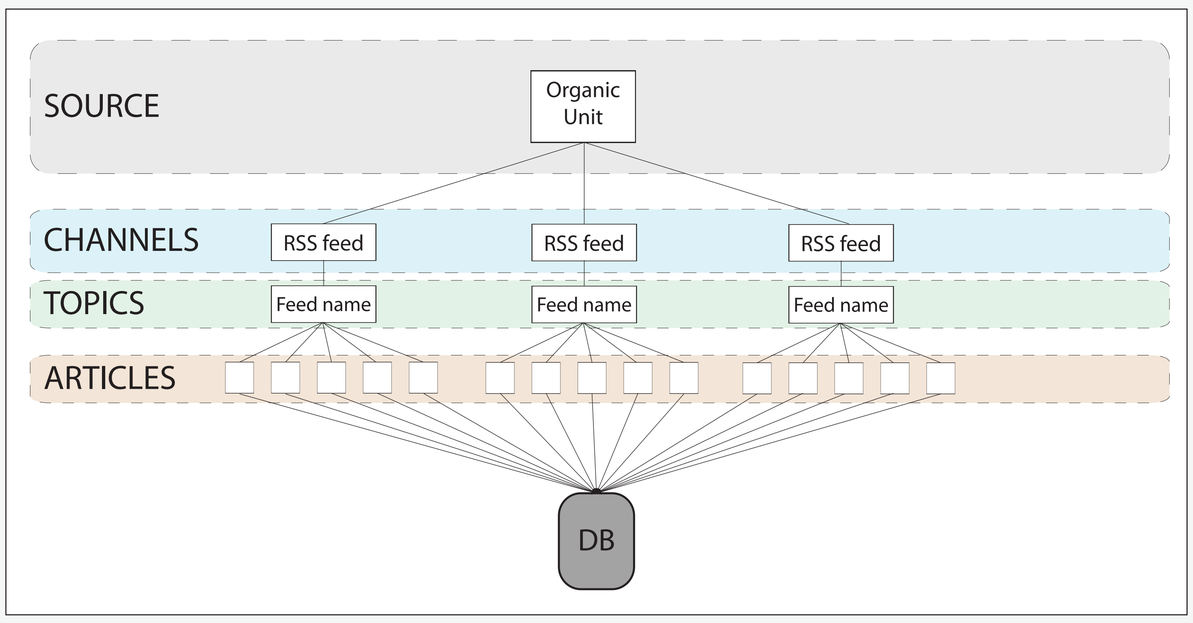


Figure 1: Data collection process (exemplified here with only one source)

## Analysis

Between June 1st 2012 and January 31st 2013, we analyzed the information collected to understand the volume and dynamics of the data published.

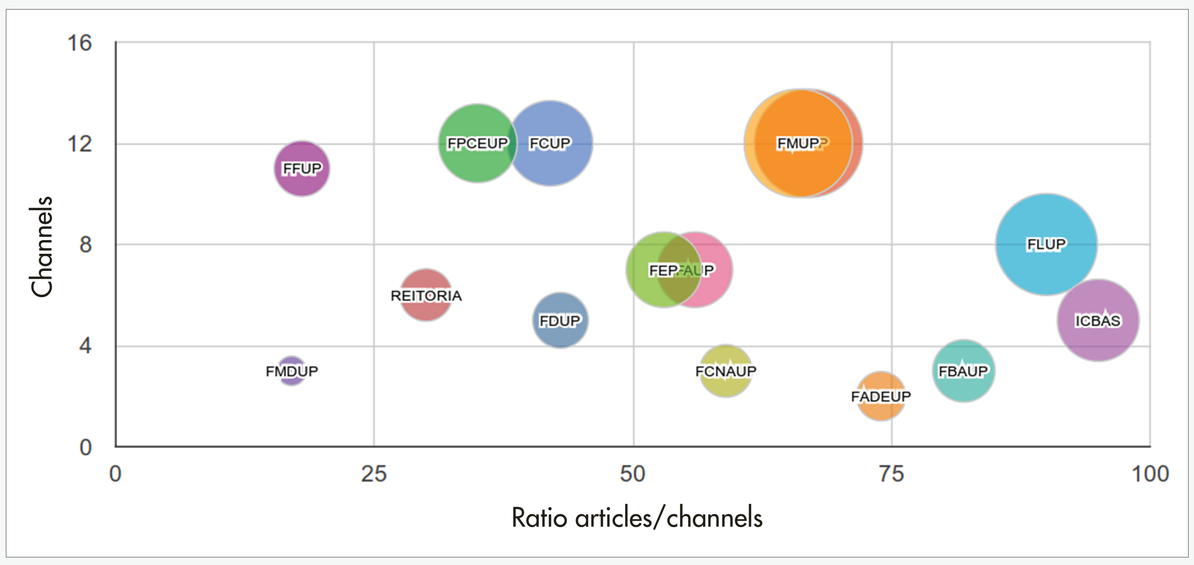


Figure 2: Channels, articles and ratio articles/channels by source (circle size represents articles published)

### Sources

Almost half (49%) of all the 5755 articles published during the period in question belonged to five sources: FEUP, FMUP, FLUP, FCUP and ICBAS. The five least productive sources - FMDUP, FADEUP, FCNAUP, REITORIA and FFUP - accounted for about 13% of all published articles.

| Source | Channels | Articles | Ratio articles/channels | Daily average | Monthly average |
|---|---|---|---|---|---|
| FEUP | 12 | 812 | 67.67 | 5.93 | 101.5 |
| FMUP | 12 | 802 | 66.83 | 6.27 | 100.25 |
| FLUP | 8 | 722 | 90.25 | 5.6 | 90.25 |
| FCUP | 12 | 513 | 42.75 | 5.29 | 64.13 |
| ICBAS | 5 | 478 | 95.6 | 4.43 | 59.75 |
| FPCEUP | 12 | 429 | 35.75 | 4.21 | 53.63 |
| FAUP | 7 | 398 | 56.86 | 3.69 | 49.75 |
| FEP | 7 | 375 | 53.57 | 3.87 | 46.88 |
| FBAUP | 3 | 248 | 82.67 | 2.56 | 31 |
| FDUP | 5 | 216 | 43.2 | 2.45 | 27 |
| FFUP | 11 | 203 | 18.45 | 2.57 | 25.38 |
| REITORIA | 6 | 180 | 30 | 3.83 | 22.5 |
| FCNAUP | 3 | 178 | 59.33 | 2.83 | 22.25 |
| FADEUP | 2 | 149 | 74.5 | 2.04 | 18.63 |
| FMDUP | 3 | 52 | 17.33 | 1.68 | 13 |
| All | 108 | 5755 | 53.29 | 3.82 | 48.39 |

Table 1: Channels, articles, ratio articles/channels, daily and monthly average by source

### Topics

The number of topics identified (56) doesn't match the number of channels (108) because there are several repeated ones. About 71% of all articles were published under 10 topics while the bottom 10 topics didn't even achieve 1%.

| Topic | Sources with topic | Articles |
|---|---|---|
| Eventos | 11 | 882 |
| Geral | 9 | 887 |
| Ensino | 5 | 118 |
| Eventos Científicos | 4 | 212 |
| Informações | 4 | 158 |
| Eventos Culturais | 4 | 108 |
| Provas Académicas | 3 | 592 |
| Alunos | 3 | 321 |
| Eventos Académicos | 3 | 189 |
| Recrutamentos | 3 | 66 |

Table 2: The 10 topics belonging to more sources and number of articles published by them

| Topic | Articles | Sources with topic |
|---|---|---|
| Geral | 887 | 9 |
| Eventos | 882 | 11 |
| Provas Académicas | 592 | 3 |
| Notícias FEUP | 385 | 1 |
| Alunos | 321 | 3 |
| Notícias Gerais | 280 | 1 |
| Eventos Científicos | 212 | 4 |
| Eventos Académicos | 189 | 3 |
| Destaques | 185 | 1 |
| Informações | 158 | 4 |
| I&D | 6 | 1 |
| GEEA | 5 | 1 |
| GEEA - Eventos | 4 | 1 |
| FAUP | 3 | 1 |
| Boletim Informativo | 3 | 1 |
| Serviços Administrativos | 2 | 1 |
| Provas Académicas - Agregação | 2 | 1 |
| Edições | 2 | 1 |
| I&D - Programas | 1 | 1 |
| Divulgação | 1 | 1 |

Table 3: The 10 topics with the most and least articles published and the number of sources with that topic

### Articles published by month and by hour

November 2012 was the month with the most articles published (982), while August had only 431. The majority of the sources showed similar behavior.

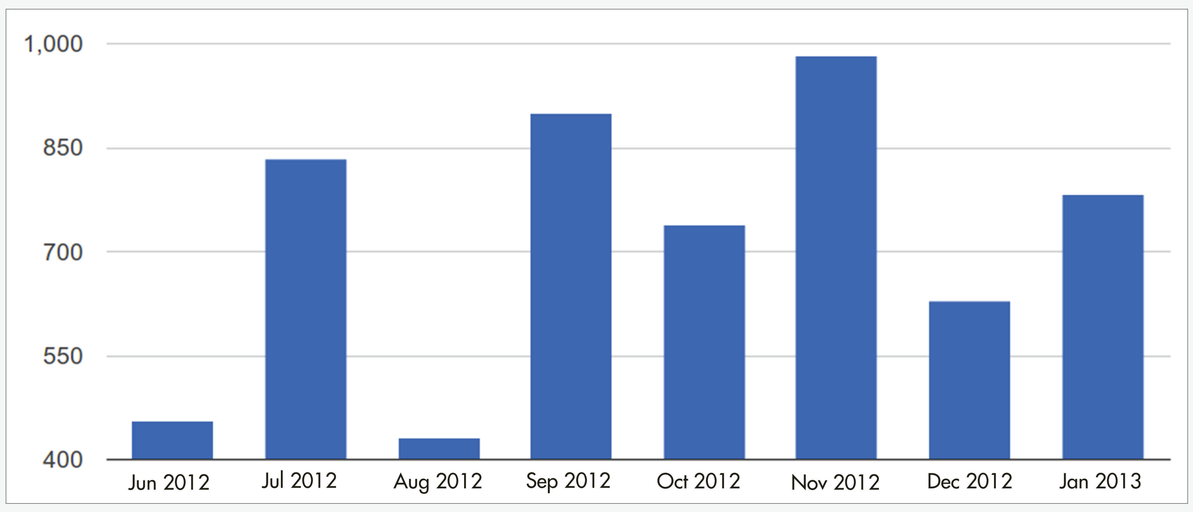


Figure 3: Articles published by month (all sources)

As for the hourly data, we identified a main publishing period that goes from 06:00 to 21:00, with some residual publishing activity outside this period.

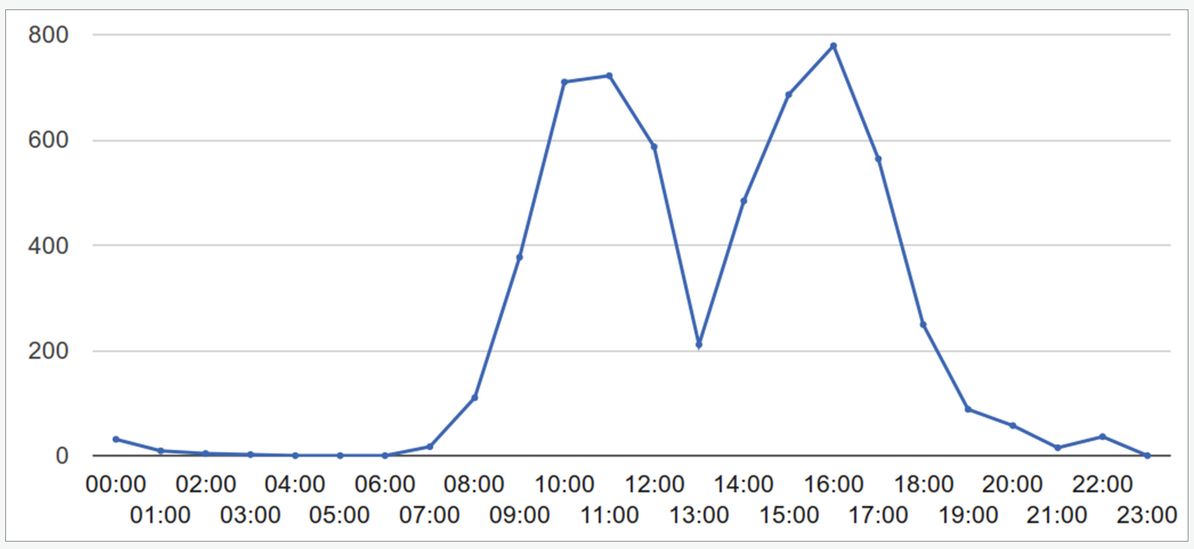


Figure 4: Articles published by hour (all sources)

## Future work

Based on the data collected and subsequent analysis, we are currently working on the design of the interface for the web-based system. We already have an initial prototype online.