

## European Patent Office (EPO) en Twitter: A content analysis

Víctor Herrero-Solana<sup>a,\*</sup>, Cristina Faba-Pérez<sup>b</sup>

<sup>a</sup> Dept. Information & Communication, University of Granada, Granada, 18071, Spain

<sup>b</sup> Dept. Information & Communication, University of Extremadura, Badajoz, 06071, Spain

### ARTICLE INFO

#### Keywords:

European Patent Office  
EPO  
Twitter  
Content analysis

### ABSTRACT

Twitter accounts can be a relevant (in addition to websites) source of information, particularly for studying the public profile of an institution. In this study, a content analysis of the @epoorg account was performed. The description of the EPO on social networks has never been researched thus far. Using the Tractor software, 19,565 tweets were downloaded and uploaded to the Graphext platform, which uses machine learning to perform topic analysis. Twenty main clusters were identified, which, once organized, engendered the following four main content groups: awards, training, events, and thematic. The thematic contents that the EPO addresses include industry 4.0, cancer, COVID-19, battery innovation, clean energy, climate change. The content is balanced: no country stands out.

### 1. Introduction

Patent offices are a crucial element within the global scientific-technical innovation system. Searching for any of them in large bibliographic databases such as Web of Science (WoS) and Scopus yields thousands of records. However, most papers address patents registered in these offices, with hardly any studies on the offices themselves as sources of information.

One of the most mentioned offices is the European Patent Office (EPO). The EPO is usually mentioned in comparative studies of databases [1–4] or for the development of its own information system [5] termed EPOQUE [6,7].

Although the EPO has a sizeable and well-developed website, unfortunately, no research on its vast content is available in the literature other than a fleeting (and outdated) description of its website [8], e-learning resources [9] and Epoline service [10].

The description of the EPO on social networks has never been researched thus far. However, there is precedent for some studies linking patents with information from Twitter [11].

In the present study, we aim at studying the main EPO Twitter account (@epoorg) and to analyze its content.

### 2. Material and methods

First, a dataset was created with all tweets of the @epoorg account, from its inception in 2008 to July 2022. Data were collected using the application Tractor<sup>1</sup> to download a large amount of data without having to use the Twitter application programming interface (API). The retweets were filtered to select only posts generated by the EPO, with the final dataset totaling 19,565 rows (tweets) and 52 columns (variables).

The dataset was uploaded to Graphext,<sup>2</sup> an online platform for big data analysis using machine learning techniques. Graphext is a platform used to create, analyze, and contextualize information from different digital feeds, including Twitter feeds. Through its analyses, this tool considerably helps find relevant data from such sources [12]. Graphext analyzes texts of tweets to identify thematic communities, known as topics. For this purpose, Graphext parses the text through ngram analysis seeking coincidences toward a thematic clustering [13]. A sentiment analysis algorithm was also applied, namely the Cardiff NLP Model. The dataset is open and available online (see Data Availability section).

### 3. Results and discussion

The temporal evolution of the 19,565 tweets is shown in Fig. 1, which excludes 2022 because the data on this year is not complete in the dataset. The number of tweets steadily increased from 2009 to 2015,

\* Corresponding author.

E-mail addresses: [victorhs@ugr.es](mailto:victorhs@ugr.es) (V. Herrero-Solana), [cfabper@unex.es](mailto:cfabper@unex.es) (C. Faba-Pérez).

<sup>1</sup> <https://www.graphext.com/tractor>.

<sup>2</sup> <https://www.graphext.com/>.

subsequently stagnating and decreasing slightly each year but increasing again in 2021 with more than 3000 tweets in a year. The decline is in line with a worldwide drop in tweets.<sup>3</sup> We will have to wait until the end of 2022 to determine whether this 2021 was simply an outlier or the account is now being used more intensively by the EPO. Annual variation rates range from -39.1% (2011–2012) to 183.9% (2011–2012), and the average is 34.3%.

Regarding the content of the tweets, the room for sentiment analysis is limited because 73% of the tweets have a neutral sentiment, slightly more than 24% of tweets are positive, and only 1.55% of the tweets are negative. The negative tweets are almost impossible to track, but the positive tweets tend to accumulate around awards, a thematic community that we will observe below.

Fig. 2 shows the graph with the 19,565 tweets connected by their content. The graph is available online from the following address (select the “graph” option at the top): <https://public.graphext.com/c3d88475>

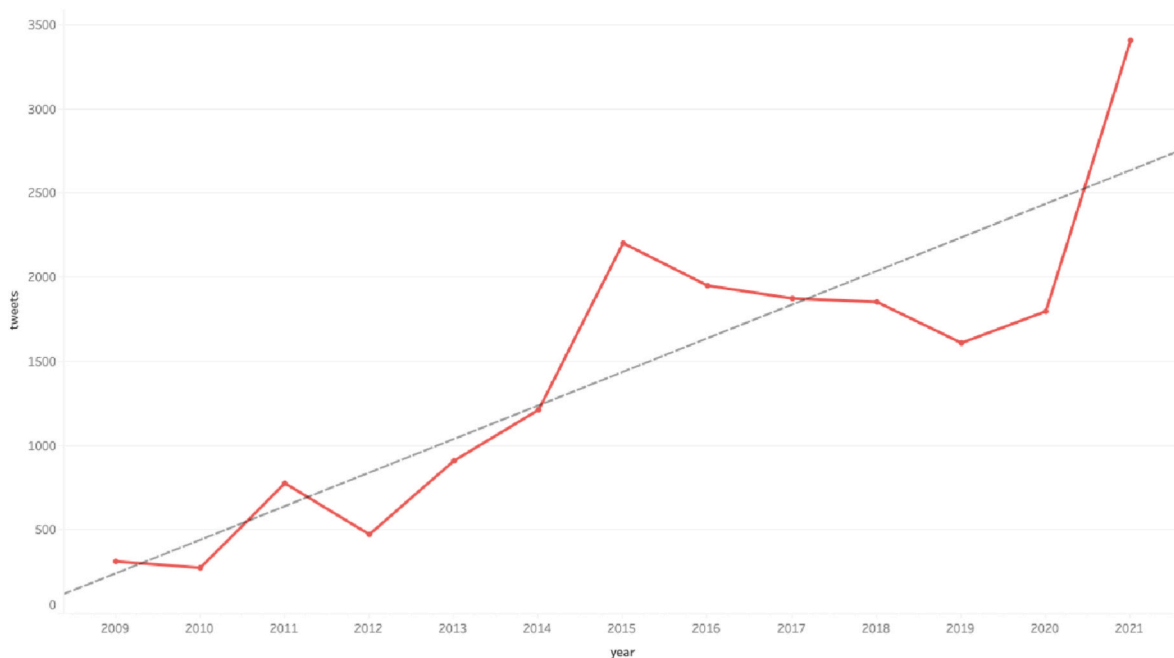


Fig. 1. Temporal evolution of @epoorg. Source: Twitter.

[70da2d85/](https://public.graphext.com/c3d88475). In this graph, the first 20 clusters were identified by topic analysis in Graphext. To better understand them, we have listed these clusters in Table 1 by importance (column one). Column two outlines the number of tweets associated with that cluster, and column three outlines the percentage of variance of each cluster. The fourth column presents the most representative words of the cluster. The fifth column shows the color of the cluster, and the sixth column displays its classification based on our analysis.

The cluster with the highest variance (7%) is related to awards, especially nominations for the European inventor award,<sup>4</sup> followed by clusters 3 (award ceremony<sup>5</sup>) and 9 (popular prize<sup>6</sup>). All these awards are an alluring subject to communicate about on Twitter, and the EPO undoubtedly strives to share such information since the three clusters account for 15% of the variance.

An equally important EPO Twitter activity can be grouped under the word **training** and consists of cluster 2 (European Patent Academy<sup>7</sup> and European Patent Register<sup>8</sup>), cluster 6 (IP Portfolio<sup>9</sup>), cluster 14 (Patlib Centres<sup>10</sup>), and especially cluster 16 (e-learning centre<sup>11</sup>).

Strongly related to the group above is a group of clusters mentioning **events**. Sometimes, these events refer to some form of training, but they typically occur at a specific time. For example, they are found in cluster 5 (EPOPIC<sup>12</sup>), cluster 7 (EPO User Day<sup>13</sup>), cluster 13 (PKWeek<sup>14</sup>) and 18 (Search Matters Conference<sup>15</sup>).

As an online magazine with EPO news, cluster 4 cannot be classified as training or an event and was therefore labeled as **news**.

An intriguing group of clusters termed **thematic** provides information on specific themes, albeit particularly relevant to the EPO, such as Industry 4.0,<sup>16</sup> which is found in two clusters, 8 and 11. The EPO has published at least two reports [14,15] on the subject of patents and Industry 4.0.

Another key **thematic** cluster is related to health, and its main topics

<sup>7</sup> <https://www.epo.org/about-us/services-and-activities/academy.html>.

<sup>8</sup> <https://www.epo.org/searching-for-patents/legal/register.html>.

<sup>9</sup> <https://www.epo.org/applying/online-services/myepo.html>.

<sup>10</sup> <https://www.epo.org/searching-for-patents/helpful-resources/patlib.html>.

<sup>11</sup> <https://e-courses.epo.org>.

<sup>12</sup> [https://www.oepm.es/cs/OEPMSite/contenidos/Revista\\_InfoPYM/2016/junio/ingles/noticia-en-02.html](https://www.oepm.es/cs/OEPMSite/contenidos/Revista_InfoPYM/2016/junio/ingles/noticia-en-02.html).

<sup>13</sup> <https://www.epo.org/news-events/news/2021/20210421.html>.

<sup>14</sup> [https://intellectual-property-helpdesk.ec.europa.eu/news-events/events/patent-knowledge-week-2022-epo-2022-10-04\\_en](https://intellectual-property-helpdesk.ec.europa.eu/news-events/events/patent-knowledge-week-2022-epo-2022-10-04_en).

<sup>15</sup> <https://www.epo.org/news-events/events/conferences/search-matters.html>.

<sup>16</sup> <https://www.epo.org/news-events/in-focus/ict/fourth-industrial-revolution.html>.

<sup>3</sup> <https://blog.gdeltproject.org/visualizing-eight-years-of-tweets-evolution-2012-2019/>.

<sup>4</sup> <https://www.epo.org/news-events/events/european-inventor.html>.

<sup>5</sup> <https://inventoraward.epo.org>.

<sup>6</sup> <https://popular-prize.epo.org>.

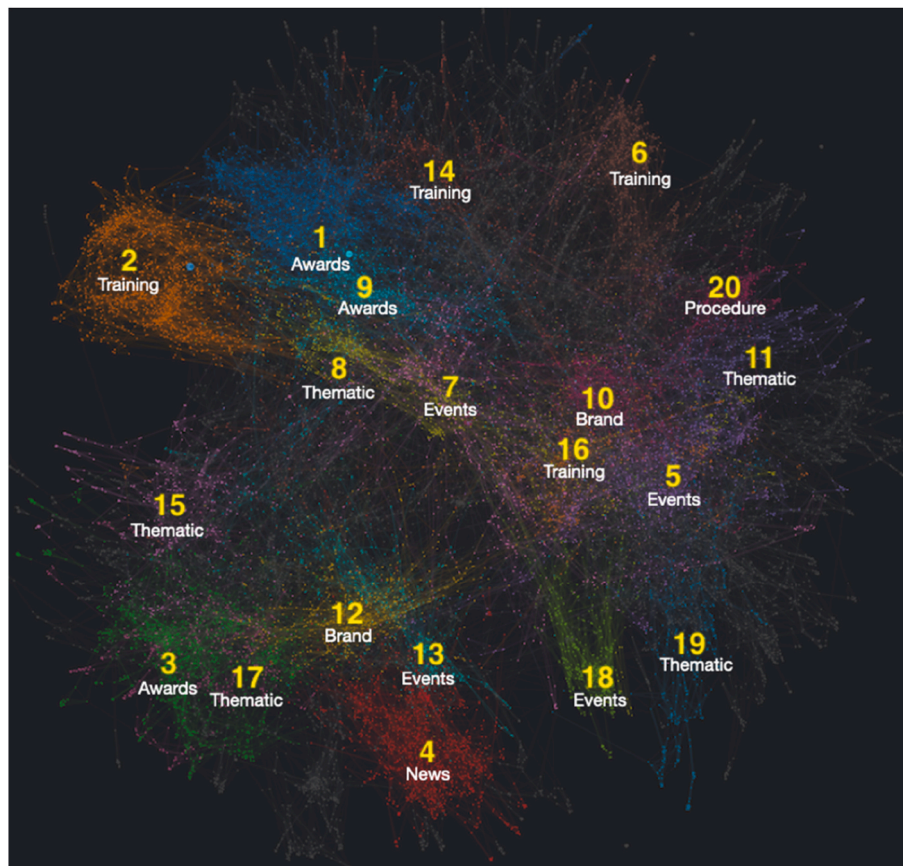


Fig. 2. Graph of tweets. Source: Twitter.

are cancer,<sup>17, 18</sup> coronavirus,<sup>19</sup> COVID-19<sup>20, 21</sup> and vaccines<sup>22</sup> (cluster 15). With roughly the same size and importance, a cluster (number 17) is identified with climate change,<sup>23</sup> albeit focusing on greentech<sup>24</sup> (battery innovation [16], solar energy, clean energy [17], etc.). Lastly, the **thematic** cluster (19) does not indicate any specific topic but refers to the section of the website in which the EPO publishes its patent insight reports<sup>25</sup> (quantum tech, CART cell, hydrogen, cosmonautics, etc.).

Of the remaining two groups, one is formed by clusters 10 and 12 and termed **brand** for working on the EPO brand, with cluster 10 including tweets on the impressive new EPO building in Rijswijk, near The Hague, which was inaugurated in 2018 by the king of The Netherlands.<sup>26</sup> The second cluster includes the tweets that directly mention their president: Antonio Campinos. The last cluster (20) is much more technical and related to the functioning of the board of appeal.<sup>27</sup>

<sup>17</sup> <https://new.epo.org/en/news-events/european-inventor-award/meet-the-finalists/axel-ullrich>.

<sup>18</sup> <https://new.epo.org/en/news-events/european-inventor-award/meet-the-finalists/jerome-galon>.

<sup>19</sup> <https://www.epo.org/news-events/in-focus/fighting-coronavirus.html>.

<sup>20</sup> <https://www.epo.org/about-us/annual-reports-statistics/annual-report/2020/covid19-pandemic.html>.

<sup>21</sup> <https://new.epo.org/en/news-events/press-centre/press-release/2020/451888>.

<sup>22</sup> <https://www.epo.org/news-events/in-focus/fighting-coronavirus/vaccine-s-and-therapeutics.html>.

<sup>23</sup> <https://www.epo.org/mobile/news-events/news/2015/201501208.html>.

<sup>24</sup> <https://www.epo.org/news-events/news/2020/20200922.html>.

<sup>25</sup> <https://www.epo.org/searching-for-patents/business/patent-insight-reports.html>.

<sup>26</sup> <https://www.epo.org/news-events/news/2019/20190411.html>.

<sup>27</sup> <https://www.epo.org/law-practice/case-law-appeals/eba.html>.

**Table 2** summarizes the groups of clusters identified in this study. The main issues communicated, almost to the same extent, are awards, training, events, and specialized thematic content. Most strikingly, no European country is mentioned as a topic in any of the clusters. This finding suggests that the EPO avoids emphasizing the countries that finance the organization, most likely to maintain some political balance.

#### 4. Conclusions

We believe that @epoorg is a well-managed Twitter account with balanced contents and the ability to hold the attention of its followers, in addition to a great advantage, that is, an excellent website. The Twitter account often functions as a guided access to the powerful set of EPO websites. For this reason, we have illustrated clusters with a long list of links.

The EPO has primarily tweeted on diverse topics such as awards, events, training, and various thematic content related to Industry 4.0, cancer, COVID-19, battery innovation, clean energy, and climate change.

Even though the objective of the investigation was not to analyze the decision-making processes of the EPO, we can affirm that the behavior of the @epoorg account shows signs of solvency and efficiency. It does not seem like a random Twitter account with irrelevant content. It appears to be part of the EPO's decision-making process and the engagement indicators seem to be satisfactory. This matter could be part of future research lines.

As future lines of research, we propose at least two avenues. The first would be to assess the actual overview of EPO tweets on the social network (followers, retweets, and favorites, among other parameters). The second would be to perform a similar analysis with other offices, such as the United States Patent and Trademark Office (USPTO), World Intellectual Property Organization (WIPO) and others, to ascertain

**Table 1**  
List of topic clusters. Source: Graphext.

#	twts	% V	topics	colour	group
1	1361	7%	european inventor award, nomination for the european, award 2015, 2016, finalist, ceremony	blue	awards
2	1017	5%	european patent register, european patent convention, european patent academy, attorney	orange	training
3	976	5%	inventor award, nominate, vote, amazing, work, scientist, world, reward, woman, brilliant, finalist	green	awards
4	845	4%	patent information, patent knowledge, news, unitary patent, newsflash, webinar, pknew, magazine	red	news
5	807	4%	epo patent information conference, epopic, datum, open patent services, attorney, grants, epo experts	purple	events
6	708	4%	ip strategy, ipforsme, case studies, business, ip management, startup, ip assets, ip portfolio, small	brown	training
7	654	3%	online services user, epo online, user day, newsletter, online filing, fortnightly, filing software, user	pink	events
8	644	3%	innovation case studies, digital technologies, 4IR, wireless, fourth industrial revolution, field, market	light yellow	thematic
9	575	3%	popular prize, vote, finalist, eia12, eia13, eia14, eia15, eia16, winner, ceremony, win the popular prize	light green	awards
10	556	3%	hague, munich, building, guidelines2day, epoinauguration, new building, building in the hague	dark red	brand
11	544	3%	patent applications, 4ir, industry40, epopatentindex, eporesult, anualreport, growth, patent filings	dark purple	thematic
12	351	3%	epo president, antonio campinos, blog, representative, environmental, plan, sustainability, talk to epo	yellow	brand
13	495	3%	thank, booth, epopic, pkweek, exhibitor, welcome, day, looking forward, live, happy, question, join	light green	events
14	487	2%	patlib centres, ip advice, unitary patent, recording, watch, know, provide, find, thank, new, patent	light brown	training
15	458	2%	cancer, coronavirus, disease, dna, vaccine, fight, treatment, medicine, patient, breast, test, covid19	light pink	thematic
16	446	2%	elearning platform, course, webinar, online, browse, free, service, knowledge you need, easy	dark orange	training
17	410	2%	energy, battery innovation, electricity, renewable, solar, clean energy, greentech, climate change	red	thematic
18	394	2%	search, searchmatters conference, workshop, search strategies, prior art, tool, patent searching	green	events
19	358	2%	patentinsight, patent protection, secure, exclusivity, ipforsme, product, strong patents aligned, R&D	light blue	thematic
20	333	2%	enlarged boards of appeal, decision, fee, case law, rule, petition, petition for review, payment, law	dark red	procedure

whether they perform similarly.

## Funding

This article was supported by the Ministry of Science, Innovation and Universities (Spain) Grant N°: RTI2018-098966-B-I00 and the University of Granada: PP2022.PP.32.

**Table 2**  
Grouped clusters. Source: The author.

group	tweets	%V
awards	2912	15%
training	2658	13%
thematic	2414	12%
events	2350	12%
brand	907	6%
news	845	4%
procedure	333	2%

## Author statement

Víctor Herrero-Solana: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing, Visualization, Project Administration, Funding acquisition.

Cristina Faba-Pérez: Methodology, Formal análisis, Investigation, Resources, Data Curation, Writing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Herrero-Solana, Víctor (2022). Tweets from the European Patent Office account (@epoorg): April 2009-July 2022. [Data set]. **Zenodo**. <https://doi.org/10.5281/zenodo.6991473>.

## References

- [1] E.M. Davis Smith, A comparison of some patent databases, *World Patent Inf.* 10 (1988), [https://doi.org/10.1016/0172-2190\(88\)90211-6](https://doi.org/10.1016/0172-2190(88)90211-6).
- [2] B. Jürgens, V. Herrero-Solana, Espacenet, patentscope and depatisnet: a comparison approach, *World Patent Inf.* 42 (2015) 4–12, <https://doi.org/10.1016/j.wpi.2015.05.004>.
- [3] G. McKiernan, esp@cenet: Europe's network of patent databases, *Sci. Technol. Libr.* 22 (2001) 57–88, <https://doi.org/10.1300/J122v22n01>.
- [4] M.J. White, esp@cenet Europe's network of patent databases, *Issues Sci. Technol. Librarian.* 47 (2006) 5062, <https://doi.org/10.5062/F4F47M2H>.
- [5] C. Jonckheere, EPOQUE (EPO QUERy service) the inhouse host computer of the European patent office, *World Patent Inf.* 12 (1990) 155–157, [https://doi.org/10.1016/0172-2190\(90\)90226-B](https://doi.org/10.1016/0172-2190(90)90226-B).
- [6] Y.T. Demey, D. Golzio, Search strategies at the European patent office, *World Patent Inf.* 63 (2020), 101989, <https://doi.org/10.1016/j.wpi.2020.101989>.
- [7] J. Van Thielen, Search tools at the European Patent Office: a new approach, *World Patent Inf.* 16 (1994) 137–139, [https://doi.org/10.1016/0172-2190\(94\)90126-0](https://doi.org/10.1016/0172-2190(94)90126-0).
- [8] T.C. Melvin, European patent office, *J. Bus. Finance Librarian.* 8 (2002) 55–62, [https://doi.org/10.1016/0306-4573\(91\)90105-u](https://doi.org/10.1016/0306-4573(91)90105-u).
- [9] R. Feinäugle, Distance learning from the European patent office, *World Patent Inf.* 28 (2006) 63–74, <https://doi.org/10.1016/j.wpi.2005.07.012>.
- [10] D. Delporte-Vermeiren, P. Vervest, E. Van Heck, In search of margin for business networks: the European Patent Office, *Eur. Manag. J.* 22 (2004) 167–182, <https://doi.org/10.1016/j.emj.2004.01.013>.
- [11] X. Li, Q. Xie, J. Jiang, Y. Zhou, L. Huang, Identifying and monitoring the development trends of emerging technologies using patent analysis and Twitter data mining: the case of perovskite solar cell technology, *Technol. Forecast. Soc. Change* (2019) 687–705, <https://doi.org/10.1016/j.techfore.2018.06.004>.
- [12] J. Pérez-Rey, V. Izquierdo, Starting up, Graphext: nunca más "lost in translation" en la gran conversación, *Ctrl Control Strateg* (2016) 30–31.
- [13] P. Suddon, How to study brand conversations with advanced text analysis? Graphext blog 22 July. <https://www.graphext.com/post/how-to-study-brand-conversations-with-advanced-text-analysis>, 2021. (Accessed 20 November 2022).
- [14] Y. Ménière, I. Rudyk, J. Valdes, Patents and the fourth industrial revolution: the inventions behind digital transformation. <https://epo.org/4IR>, 2017. (Accessed 20 November 2022). Munich.
- [15] Y. Ménière, J. Philpott, J. Pose Rodríguez, I. Rudyk, S. Wewege, N. Wienold, Patents and the Fourth Industrial Revolution: the global technology trends

- enabling the data-driven economy. <https://epo.org/trends-4IR>, 2020. (Accessed 20 November 2022). Munich.
- [16] G. Gregori, S. Meini, Y. Ménière, J. Pose Rodríguez, I. Rudyk, S. Bennett, N. Johnstone, L. Munuera, Innovation in batteries and electricity storage: a global analysis based on patent data, in: European Patent Office, International Energy Agency, Munich, 2020. <https://epo.org/trends-batteries>. (Accessed 20 November 2022).
- [17] Y. Ménière, C. Rossatto, I. Rudyk, J. Pose Rodríguez, M. Bonora Ortega, N. Johnstone, S. Bennett, Patents and the energy transition: global trends in clean energy technology innovation, in: European Patent Office, International Energy Agency, Munich, 2021. <https://epo.org/trends-energy>. (Accessed 20 November 2022).

Víctor Herrero-Solana, Professor: Department of Información and Comunicación, University of Granada, Granada, Spain. Subject: the thematic specialization has been directed towards the data visualization, evaluation of scientific activity, impact on media, and technological surveillance through patents. Teaching: 29 years of teaching in four universities of three different countries: Universidad Nacional de Mar del Plata, Argentina

(1991–1997), Universidad Nacional Autónoma de México (1996–1998), Universidad de Extremadura (1998–1999), and Universidad de Granada (from 1999 and remains). Ten research and teaching stays in international institutions in Norway, Germany, Italy, Portugal, Cuba, Chile, Brazil, Mexico and Argentina). Publications: 20 books/chapters, 90 scientific articles y 30 national and international meeting communications.

GS profile: <https://scholar.google.com/citations?user=OKIleUEAAAAJ>  
 Scopus ID: <https://www.scopus.com/authid/detail.uri?authorId=55667326400>  
 Publons: <https://publons.com/wos-op/researcher/347737/victor-herrero-solana/>  
 ORCID ID: <https://orcid.org/0000-0003-1142-5074>

Cristina Faba-Pérez, Associated professor: Department of Información and Comunicación, University of Extremadura, Badajoz, Spain. She has specialized in Information Visualization, Data Journalism, Scientometrics and Patent Information. PhD in Information Science from the University of Granada.

Scopus ID: <https://www.scopus.com/authid/detail.uri?authorId=6507765282>  
 Publons: <https://publons.com/wos-op/researcher/4844977/cristina-faba-perez/>  
 ORCID ID: <https://orcid.org/0000-0002-0537-3231>