Abstract: Digital footprints (DF) offer relevant information about educational activities and processes related to strategies of academic assessment, identification of skills and psychological traits of students, prediction of the actions of the different educational actors, and trends of permanence and dropout, among others. The present study aims to analyze the scientific evidence on the use of DF in the educational field and shows, both the process and the results of a systematic review of literature carried out based on the analysis of documents published in the last fifteen years (2005-2019) in two databases: Scopus and Web of Science (WoS).

Among the most relevant findings of this review, it is found that educational research on DF is focused on learning analytics, the study of digital presence and psychometric modeling. Likewise, the article reports on the scarce investigation of DF in MOOC environments and highlights the multiple meanings of DF as an action and as a service, beyond the generalized conception of data. These findings suggest the importance of preparing educational institutions in general, and educational actors in particular, so that they can implement processes of use and management of their DF in order to favor processes that range from DF curation, to cognitive evaluation, prediction of school success and/or failure, and the identification and attraction of talent, among others.
revisión sistemática de literatura llevada a cabo con base en el análisis de documentos publicados en los últimos quince años (2005-2019) en dos bases de datos: Scopus y Web of Science (WoS). Entre los hallazgos más relevantes de esta revisión, se encuentra que la investigación educativa sobre DF está enfocada en las analíticas de aprendizaje, el estudio de la presencia digital y el modelado psicométrico; al tiempo, se subraya la escasa investigación de DF en entornos MOOC. Igualmente se destacan los múltiples significados de las DF pues en los estudios se las define indistintamente como acción y como servicio, más allá de la concepción generalizada de dato. Estos hallazgos sugieren la importancia de preparar a las instituciones educativas en general y a los actores educativos en particular, para que puedan implementar procesos de uso y gestión de sus huellas digitales con el objeto de favorecer procesos que van desde la curación de huellas, hasta la evaluación cognitiva, la predicción del éxito o fracaso escolar, y la identificación y atracción del talento, entre otros.

**Keywords:** digital footprints, digital footprints management, learning analytics, social networks, educational data mining, educational big data.

**Palabras clave:** huellas digitales, gestión de huellas digitales, learning analytics, redes sociales, minería de datos educativos, big data educativo.

1. **Introducción**

   Today's society is characterized by its high level of hyperconnectivity, which has generated circumstances that affect the communicative, economic, educational and emotional ways in which we relate as people. In a way, we –humans– are the result of hyperconnection mediated by the emergence and development of technologies such as mobile devices and the Internet of Things (IoT). These technologies have made possible the development of new geolocation, accommodation, mobility and education products and services. New dynamics have emerged, such as the creation of transport companies that do not have a single vehicle, hotel companies that do not have own physical properties and educational companies with hundreds or thousands of students who do not attend any physical campus. These cases are striking not only because they involve the use of cutting-edge digital technologies, but also because they are established under the breakdown of paradigms that seem to point towards the consolidation of a society shaped and configured from its digital interactions (Brown & Duguid, 2001; Castells, 2009; Cobo & Moravec, 2011; Lévy, 2007). Such digital interactions are the departure point for people and organizations to produce traces of digital information, which, reflect and configure their DF.

   Thus, the study of this hyperconnection phenomenon as a result of the massive interaction of users, who generate, consume and produce DF on a daily basis, is only incipient when compared with the great effects that its use can produce. This is proven in cases such as that of the past elections of the American government, spotted by the scandal of the use of thousands of Facebook accounts that were bombarded and leaked with messages that ended up influencing the 2016 presidential election. Other reported cases include the 2010 Tunisian Revolution and the 2011 Arab Spring, where hundreds of thousands of civilians revolved physically and digitally against the government, by using the power of massive and
Immediate communication that social networks have to generate DF. These traces became an avalanche of demonstrations of such civil uprising, which ended deposing the leadership of the Egyptian government.

These cases add evidence that constitute one of the greatest risks identified by the World Economic Forum in terms of probability (2019), and suggest that by 2020 there will be more than 50 billion interconnected digital devices (Figure 1), and that cyber dependence as well as massive incidents of fraud or data theft (Figure 2). Hence, it is necessary for people to learn to manage their lives and their digital identity. This implies the need to educate people and organizations in training activities around the management processes of their DF, which goes beyond their generation, consumption and production.

Figure 1

*Global number of interconnected mobile devices 2012-2020. Source: Statista Research Department (2016)*

![Global number of interconnected mobile devices 2012-2020. Source: Statista Research Department (2016)](image)

Figure 2

Digital Footprints

DF are defined as the set of digital trails that we leave as a result of our interaction in the different digital environments. These include social networks (Buchanan, Southgate, Smith, Murray, & Noble, 2017; Hoel & Xiao, 2018b; Vivakaran & Maraimalai, 2019), learning management platforms (Azcona, Corrigan, Scanlon, & Smeaton, 2017; Wan & Tsai, 2014; Zielke, Bernst, Crombach, Becker, & Lippel, 2018), and massive online courses (Leon Urrutia, Vazquez-Cano, & Lopez Meneses, 2017). DF also include spaces and digital information systems, which include forums, blogs and wikis (Kosinski, Wang, Lakkaraju, & Leskovec, 2016). For instance, Madden, Fox, Smith y Vitak (2007) state that these DFs are considered traces of data that a person leaves voluntarily or involuntarily online and can be of two types: passive or active. Passive DFs relate to those personal data accessible online without the deliberate intervention and knowledge of an individual. Active DFs are built voluntarily, often in specific contexts with specific audiences in mind (Madden et al., 2007), such as in the case of DFs generated by the deliberate publication of messages or the exchange of information.

2. Materials and Methods

The research study emerges from a systematic literature review that included the analysis of studies done from 2005 to 2019 related to the use of DF in the educational field, and specifically, in activities and learning or teaching processes. The methodology takes into account the phases recommended by Ke (2009) and Kitchenham et al. (2010) from which a review protocol is developed, including the statement of the purpose of the systematic review,
the formulation of the research questions (RQ), the definition and application of the inclusion and exclusion criteria, an in-depth reading and the presentation of results.

2.1 Purpose of the review

As indicated by Staples and Niazi (2007), a first process in a systematic review has to do with establishing an approach that allows for an effective research study, mainly through the formulation of guiding questions. In the particular case, the RQ that guided the review were:

RQ1: What are the conceptions of DF?
RQ2: What are the educational activities and processes in which DF are used?
RQ3: What are the resources used to collect DF?
RQ4: What kind of technologies are used for the analysis of DF?

Then, some related initial concepts were determined from these questions. These are displayed in Table 1.
**Research Questions and Initial related concepts**

<table>
<thead>
<tr>
<th>Research Questions (RQ)</th>
<th>Initial related concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the conceptions of DF?</td>
<td>Data</td>
</tr>
<tr>
<td>What are the educational activities and processes in which DF are used?</td>
<td>School success and/or failure, analysis of learning, school dropout and retention, psychometric modeling</td>
</tr>
<tr>
<td>What are the resources used to collect DF?</td>
<td>Social networks, virtual learning environments, massive open online courses (MOOC), digital environments and IT systems (forums, blogs)</td>
</tr>
<tr>
<td>What kind of technologies are used for the analysis of DF?</td>
<td>Machine learning, data analytics, specialized software, cloud computing</td>
</tr>
</tbody>
</table>

**2.2 Definition and application of the inclusion and exclusion criteria**

In this phase, the inclusion/exclusion criteria were determined. This allowed excluding studies that were not relevant to answer the RQ and instead, choose the bibliographic sources that would constitute the documentary basis (Kitchenham et al., 2010). For this purpose, it was determined to include all kinds of documents published in Scopus and Web of Science (WoS) databases, including articles, conference papers, books, book chapters, reviews and notes. In general, four inclusion/exclusion criteria were determined, as displayed in Table 2.

**Inclusion/exclusion criteria**

<table>
<thead>
<tr>
<th>Inclusión criteria</th>
<th>Exclusión criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents that fit the search equation</td>
<td>Documents that do not fit the search equation</td>
</tr>
<tr>
<td>Documents published between 2005-2019 in the Scopus and WoS databases</td>
<td>Documents published in Scopus and in WoS that are outside the range 2005-2019</td>
</tr>
<tr>
<td>Studies that are oriented to the analysis of DF</td>
<td>Studies that are not oriented to the analysis of fingerprints</td>
</tr>
<tr>
<td>Studies on DF analysis in the educational field or that have a possible application in education</td>
<td>Studies that do not address DF analysis in the educational field or that do not have a possible application in that field</td>
</tr>
</tbody>
</table>

As a first inclusion/exclusion criteria, the fit to a search equation used in the Scopus and WoS databases was determined, as shown in Table 3. This equation combined the key words taking into account aspects of synonymy, thus as the etymological root of the words: education, teaching and learning. This was done with the aim of including the largest number of documents that had a direct or indirect relationship with the use of DF in educational activities or processes.
The second inclusion/exclusion criteria that was set was time. The selected period was 2005-2019. Initially, this period corresponded to \([-\infty, 2019]\), but in the review in the two databases, there was only one document published in 1978 that did not meet the rest of the inclusion/exclusion criteria. This implied that the range of time was exclusively set for 2005-2019, which is practically a period that studies the use of DF in education in depth. The third inclusion/exclusion criteria included all studies in any area of knowledge that would have used DF. Finally, the fourth inclusion/exclusion criteria was formulated based on the need to include studies that, even though did not have an educational purpose, provided relevant information on a possible application of DF in any educational activity or process.

Subsequently, a pre-reading activity was carried out, which consisted of reviewing and studying all the titles, abstracts and keywords of the documents that responded to the search equation in the Scopus databases \(n=96\) and WoS \(n=38\), displaying a total of 134 documents. From this number, duplicate documents were eliminated \(n=29\), as well as those which were not available in the databases \(n=3\). Then, the four inclusion/exclusion criteria were applied to the resultant number. Finally, the documents that made up the review base were read in depth \(n=46\). This procedure is explained in Figure 3.
2.3 In-depth reading

46 documents were read and analyzed in depth. Then, key ideas were processed, and analysis categories were identified. The researchers grouped and unified the categories through the application of a Cohen’s Kappa coefficient (K = 0.66). From this process, it was possible to identify the documents that were related to the four research questions (RQs) as well as to the concepts that were defined as related initial concepts, displayed in Table 2 and Table 4.
Table 4  
Analysis categories a priori vs. Documents read in depth

<table>
<thead>
<tr>
<th>Research Questions (RQ)</th>
<th>Initial related concepts</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the conceptions of DF?</td>
<td>Data (Abdelnour-Nocera, Oussena, &amp; Burns, 2015; Azcona, Hsiao, &amp; Smeaton, 2019; Buchanan et al., 2017; Chretien, Tuck, Simon, Singh, &amp; Kind, 2015; Simon Cleveland, Jackson, &amp; Dawson, 2016; Eberlin, 2018; Hoel &amp; Xiao, 2018b; Hwong, Oliver, Van Kranendonk, Sammut, &amp; Seroussi, 2017; Laleh &amp; Shahram, 2018; Robinson &amp; Gran, 2018; Scanlon &amp; Smeaton, 2017; Shafie, Yacocob, &amp; Paramjit Kaur, 2015; Surmelioglu &amp; Seferoglu, 2019; Vivakaran &amp; Maraimalai, 2019; Youyou, Kosinski, &amp; Stillwell, 2015)</td>
<td></td>
</tr>
<tr>
<td>What are the educational activities and processes in which DF are used?</td>
<td>School success or failure, analysis of learning, school dropout and retention, psychometric modeling (Burr &amp; Cristianini, 2019; Chamorro-Premuzic, Akhtar, Winsborough, &amp; Sherman, 2017; Gelbard, Ramon-Gonen, Carmeli, Bittmann, &amp; Talyansky, 2018; Grover &amp; Mark, 2017; Hinds &amp; Joinson, 2019; Hoel &amp; Xiao, 2018b; Laleh &amp; Shahram, 2018; Leon Urrutia et al., 2017; Schoedel et al., 2018; Shafie et al., 2015; Zielke et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>What are the resources used to collect DF?</td>
<td>Social networks, virtual learning environments, massive open online courses (MOOC), digital environments and IT systems (forums, blogs) (Azcona et al., 2017; Benson &amp; Filippaioes, 2010; Chretien et al., 2015; Simon Cleveland et al., 2016; Eberlin, 2018; Galimova, Konyshcheva, Kalugina, &amp; Sizova, 2019; Gruzd, 2009; Hwong et al., 2017; Kosinski et al., 2016; Laleh &amp; Shahram, 2018; Nechaev, Corcoglioniti, &amp; Giuliano, 2017; Parks, Lowry, Wigand, Agarwal, &amp; Williams, 2018; Polignano, Basile, Rossello, De Gemmis, &amp; Semeraro, 2017; Scanlon &amp; Smeaton, 2017; Shafie et al., 2015; Surmelioglu &amp; Seferoglu, 2019; Kulkarni Varsha &amp; Monica, 2018; Vivakaran &amp; Maraimalai, 2019; Wan &amp; Tsai, 2014; Ye, Du, &amp; Zhao, 2017; Youyou et al., 2015; Zielke et al., 2018)</td>
<td></td>
</tr>
<tr>
<td>What kind of technologies are used for the analysis of DF?</td>
<td>Machine learning, data analytics, specialized software, cloud computing (Chamorro-Premuzic et al., 2017; Gruzd, 2009; Hinds &amp; Joinson, 2019; Hoel &amp; Xiao, 2018b; Hwong et al., 2017; Laleh &amp; Shahram, 2018; Pardo, Ellis, &amp; Calvo, 2015; Shafie et al., 2015; Kulkarni Varsha &amp; Monica, 2018; Vivakaran &amp; Maraimalai, 2019)</td>
<td></td>
</tr>
</tbody>
</table>

2.4 Presentation of results

The presentation of the results was made based on the analysis of the responses to the RQs.

3. Results

The analysis of results was carried out from the extraction and organization of data in an analysis matrix in which the documents read in depth were recorded. In this matrix, the
key ideas were recorded from the four RQs proposed. From these ideas, the text fragments that partially or completely answered the RQs were extracted. Then, categorization and qualitative and quantitative data analysis processes was carried out, leading to a systematic interpretation of data collected.

3.1 What are the conceptions of DF?

The review corroborates what was said by Madden et al. (2007) who define the DF as a data set (n=30.43%) derived from the digital interaction of individuals. In this sense, the data is understood as an entity that carries information about a concept, a fact, a phenomenon or a person. These studies refer to DF as data available in digital interaction spaces in a broad way, and more specifically, in social networks, as indicated in Table 5. In addition to the abovementioned arguments, it was found that there are two other meanings that emerge as possible meanings of the DF: Action (n=8.70%), and Service (n=4.34%). Finally, as indicated in Table 6, it is found that a significant number of the studies and documents reviewed (n=56.52%) do not define a specific conception for the DF.

Table 6

DF as data

<table>
<thead>
<tr>
<th>Meanings of DF</th>
<th>Documents</th>
<th>Emerging related concepts</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data in unspecified digital spaces</td>
<td>(Abdelnour-Nocera et al., 2015; Azcona et al., 2019; Robinson &amp; Gran, 2018; Sjöberg et al., 2017; Surmelioglu &amp; Seferoglu, 2019)</td>
<td>Personality traits, emotionality, identity, learning analytics</td>
<td>10.87%</td>
</tr>
<tr>
<td>Data on social networks (Facebook, Twitter, Snapchat, Youtube, etc.)</td>
<td>(Chretien et al., 2015; Eberlin, 2018; Hoel &amp; Xiao, 2018a; Horesh, Varshney, &amp; Yi, 2016; Hwong et al., 2017; Laleh &amp; Shahram, 2018; Shafie et al., 2015; Vivakaran &amp; Maraimalai, 2019; Youyou et al., 2015)</td>
<td>Likes, followers, content analysis</td>
<td>19.56%</td>
</tr>
</tbody>
</table>
Within the set of documents that consider DF as action, these are classified as a set of interactions related to the possibilities of the data left in digital environments (Sjöberg et al., 2017; K Varsha & Monica, 2018). On the other hand, new perspectives, especially those that have to do with DF produced by organizations, define DF as a service (Connelly & Osborne, 2017; Osborne & Connelly, 2016), which suggests that they are projected as an asset, given that its management can result in the benefit and increase of resources for organizations. This is the case of some primary and secondary education institutions (Cladis, 2018; Martin et al., 2018) and higher education (Abdelnour-Nocera et al., 2015; S Cleveland et al., 2016; Osborne & Connelly, 2015, 2016; Scanlon & Smeaton, 2017). In this situation, DFs are used to examine issues such as academic success (Leon Urrutia et al., 2017; Zielke et al., 2018), analysis of learning (Hoel & Xiao, 2018b), psychometric modeling of students (Grover & Mark, 2017; Laleh & Shahram, 2018), identification of students at academic risk (Azcona et al., 2019), detection of plagiarism in the learning tests (Parks et al., 2018), and planning and development of courses where students and parents are taught about curation activities of their own DF (Buchanan et al., 2017; Osborne & Connelly, 2016). This is either made to make their profiles more attractive to future employers, or to protect personal data, so there is a low risk of data misuse.

3.2 What are the educational activities and processes in which DF are used?

The activities and processes in which DF are used are diverse. These are organized into four large groups: (1) analytical learning, (2) digital presence and life, (3) psychometric modeling, and (4) other. In the latter category, activities include the use of DF in the socialization and preservation of the cultural practices of ethnic and indigenous groups. In this regard, Table 7 displays these activities and/or processes, the authors, the documents and the frequency of appearance of those activities in the review.
### Table 7

Educational activities and processes in which DF are used

<table>
<thead>
<tr>
<th>Activities/processes</th>
<th>Documents</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning analytics (commitment, cognitive activity, learning acquisition and evaluation, learning feedback, prediction of school success or failure, student recruitment, identification and attraction of talent, collaboration in learning processes, detection of psycholinguistic profiles, study of relationship behavior vs. performance, design of environments and learning experiences, acquisition of a foreign language, career management, improvement of the curriculum)</td>
<td>(Abdelnour-Nocera et al., 2015; Azcona et al., 2019; Benson &amp; Filippaio, 2010; Chamorro-Premuzic et al., 2017; Galimova et al., 2019; Hoel &amp; Xiao, 2018b; Hwong et al., 2017; Kosinski et al., 2016; Leon Urrutia et al., 2017; Pardo et al., 2015; Paredes et al., 2018; Porouhan &amp; Premchaiswadi, 2017; Raybourn, 2017; Rossetti et al., 2016, 2015; Scanlon &amp; Smeaton, 2017; Shafie et al., 2015; Sjöberg et al., 2017; Kulkarni Varsha &amp; Monica, 2018; Vivakaran &amp; Maraimalai, 2019; Zielke et al., 2018)</td>
<td>47.83%</td>
</tr>
<tr>
<td>Digital Presence and Digital Life (DF management, DF healing, digital awareness development, digital privacy, cyberbullying, digital presence, digital inequality, digital skills, effects of using digital technologies, legislation on life and digital presence)</td>
<td>(Buchanan et al., 2017; Cladis, 2018; Simon Cleveland et al., 2016; Eberlin, 2018; Gruzd, 2009; Kay, 2015; Martin et al., 2018; Osborne &amp; Connelly, 2015, 2016; Robinson &amp; Gran, 2018; Surmelioğlu &amp; Seferoğlu, 2019)</td>
<td>23.91%</td>
</tr>
<tr>
<td>Psychometric modeling (creation of student profiles, studies of personality traits, emotionality and/or empathy)</td>
<td>(Burr &amp; Cristianini, 2019; Gelbard et al., 2018; Grover &amp; Mark, 2017; Hinds &amp; Joinson, 2019; Laleh &amp; Shahram, 2018; Nechaev et al., 2017; Polignano et al., 2017; Schoedel et al., 2018; Ye et al., 2017; Youyou et al., 2015)</td>
<td>21.74%</td>
</tr>
<tr>
<td>Others (detection of traps, use of social networks, preservation of indigenous wisdom and culture, detection of social ties, development of digital skills)</td>
<td>(Chretien et al., 2015; Parks et al., 2018; Wan &amp; Tsai, 2014)</td>
<td>6.52%</td>
</tr>
</tbody>
</table>

### 3.3 What are the resources used to collect fingerprints?

Among the media and resources mentioned in the studies (Table 8), more than half of them correspond to studies that have used social networks as collection spaces for DF (52.17%). From these networks, the most used are Facebook (Laleh & Shahram, 2018; Shafie et al., 2015; Vivakaran & Maraimalai, 2019; Youyou et al., 2015) and Twitter (Chretien et al., 2015; Simon Cleveland et al., 2016; Nechaev et al., 2017; Kulkarni Varsha & Monica, 2018), noting that content analysis, likes and re-tweets are the most common elements in the analysis of these social media sources.

The group of online spaces is also a salient category (23.91%). This includes virtual learning environments (Azcona et al., 2017), MOOCs (Leon Urrutia et al., 2017) and platforms educational programs of the LMS type (Wan & Tsai, 2014), especially linked to learning analytical processes. The third group is that of digital devices (8.70%) in which resources such as smartphones, tablets and interactive tables, are the main tools used (Wan & Tsai, 2014).

Following this categorization, email corpus (2.17%) is found. Only one study is cited (Gelbard et al., 2018) from a company that sought to analyze the relationship between email...
messages and the levels of performance, commitment, leadership, workplace dynamics, support for organizational development and learning and knowledge creation of workers. Finally, there are the studies in which the DF were collected manually through interviews, surveys or forms (6.52%) (Azcona et al., 2019; Pardo et al., 2015; Robinson & Gran, 2018). Only 6.52% of the studies did not mention the resource they used to collect the DF.

Table 8

Spaces and resources for the collection of DF

<table>
<thead>
<tr>
<th>Spaces/Resources</th>
<th>Documents</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social networks (e.g. Facebook, Twitter, Youtube, Instagram, Snapchat)</td>
<td>(Benson &amp; Filippaio, 2010; Chamorro-Premuzic et al., 2017; Chretien et al., 2015; Cladis, 2018; Simon Cleveland et al., 2016; Eberlin, 2018; Grover &amp; Mark, 2017; Gruzd, 2009; Hwong et al., 2017; Kosinski et al., 2016; Laleh &amp; Shahram, 2018; Martin et al., 2018; Nechaev et al., 2017; Osborne &amp; Connelly, 2015, 2016; Parks et al., 2018; Polignano et al., 2017; Rossetti et al., 2016, 2015; Shafie et al., 2015; Kulkarni Varsha &amp; Monica, 2018; Vivakaran &amp; Maraimalai, 2019; Ye et al., 2017; Youyou et al., 2015)</td>
<td>52.17</td>
</tr>
<tr>
<td>Online spaces and learning environments (eg Moodle, MOOCs like FutureLearn, Educational Platforms like Eduroam and DigitalMe) and non-formal</td>
<td>(Abdelnour-Nocera et al., 2015; Azcona et al., 2017; Burr &amp; Cristianini, 2019; Galimova et al., 2019; Kay, 2015; Leon Urrutia et al., 2017; Paredes et al., 2018; Raybourn, 2017; Scanlon &amp; Smeaton, 2017; Sermelioglu &amp; Seferoglu, 2019; Wan &amp; Tsai, 2014)</td>
<td>23.91</td>
</tr>
<tr>
<td>Digital devices (cell phones, tablets, interactive tables)</td>
<td>(Buchanan et al., 2017; Hinds &amp; Joinson, 2019; Porouhan &amp; Premchaiswadi, 2017; Schoedel et al., 2018)</td>
<td>8.70</td>
</tr>
<tr>
<td>Email</td>
<td>(Gelbard et al., 2018)</td>
<td>2.17</td>
</tr>
<tr>
<td>Others (surveys, questionnaires, forms, interviews)</td>
<td>(Azcona et al., 2019; Pardo et al., 2015; Robinson &amp; Gran, 2018)</td>
<td>6.52</td>
</tr>
<tr>
<td>Not mentioned</td>
<td>(Hoel &amp; Xiao, 2018b; Sjöberg et al., 2017; Zielke et al., 2018)</td>
<td>6.52</td>
</tr>
</tbody>
</table>

3.4 What kind of technologies are used for the analysis of fingerprints?

The studies report technologies ranging from those based on the use of specialized software in the analysis of qualitative data (Shafie et al., 2015), to those supported by big data (Kosinski et al., 2016). It is also worth clarifying that, in some studies, these technologies are not mentioned or do not refer to actual technologies, but rather allude to resources related to concepts that have guided the development and implementation of these, such as data mining, which is why they are usually associated to different resources or technologies. However, in some studies, these technologies or resources are used assiduously in the educational field. Such is the case of educational big data (EBD) (Hoel & Xiao, 2018b), educational data mining (EDM) (Vivakaran & Maraimalai, 2019), and learning analytics (Abdelnour-Nocera et al., 2015; Azcona et al., 2017; Pardo et al., 2015).

Some of these resources and technologies are also confirmed in the word count of the keywords that were made from the analyzed studies, as shown in Figure 4. Finally, it is worth
clarifying that the answer to this RQ is shown in terms of technologies and resources on which the DF analysis is supported, which implies that several of them appear associated to several studies simultaneously displayed in Table 9.

Table 9

*Technologies and resources for DF analysis*

<table>
<thead>
<tr>
<th>Technologies and resources</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Data (BD)</td>
<td>(Chamorro-Premuzic et al., 2017; Hinds &amp; Joinson, 2019; Hwong et al., 2017; Laleh &amp; Shahram, 2018; Pardo et al., 2015)</td>
</tr>
<tr>
<td>Educational Big Data (EBD)</td>
<td>(Hoel &amp; Xiao, 2018b)</td>
</tr>
<tr>
<td>Data mining (DM)</td>
<td>(Gelbard et al., 2018; Grover &amp; Mark, 2017; Laleh &amp; Shahram, 2018; Porouhan &amp; Premchaiswadi, 2017; Kulkarni Varsha &amp; Monica, 2018; Wan &amp; Tsai, 2014)</td>
</tr>
<tr>
<td>Educational Data Mining (EDM)</td>
<td>(Gruzd, 2009; Vivakaran &amp; Maraimalai, 2019)</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>(Burr &amp; Cristianini, 2019)</td>
</tr>
<tr>
<td>Learning Analytics</td>
<td>(Abdelnour-Nocera et al., 2015; Azcona et al., 2017; Grover &amp; Mark, 2017; Leon Urrutia et al., 2017; Pardo et al., 2015; Vivakaran &amp; Maraimalai, 2019; Zielke et al., 2018)</td>
</tr>
<tr>
<td>Specialized software</td>
<td>(Laleh &amp; Shahram, 2018; Shafie et al., 2015)</td>
</tr>
<tr>
<td>Other (Model of collective cybernetic action, survey and interview analysis, focus groups, psychological traits tests BIG5, smart devices)</td>
<td>(Buchanan et al., 2017; Grover &amp; Mark, 2017; Kosinski et al., 2016; Laleh &amp; Shahram, 2018; Martin et al., 2018; Parks et al., 2018; Robinson &amp; Gran, 2018; Schoedel et al., 2018)</td>
</tr>
<tr>
<td>Not mentioned</td>
<td>(Azcona et al., 2019; Benson &amp; Filippaios, 2010; Burr &amp; Cristianini, 2019; Chretien et al., 2015; Chadi, 2018; Simon Cleveland et al., 2016; Eberlin, 2018; Galimova et al., 2019; Kay, 2015; Nechaev et al., 2017; Osborne &amp; Connelly, 2016, 2015; Paredes et al., 2018; Polignano et al., 2017; Raybourn, 2017; Rossetti et al., 2016, 2015; Scanlon &amp; Smeaton, 2017; Sjoberg et al., 2017; Surmelinoglu &amp; Seferoglu, 2019; Ye et al., 2017; Youyou et al., 2015)</td>
</tr>
</tbody>
</table>
4. Discussion of results

The studies reviewed reflect an increasing interest of educational institutions to take advantage of the DF that people produce in a hyperconnected digital society. Studies such as the Global Risk Report of the World Economic Forum (2019) indicate that society needs to be prepared for the outburst of data that people generate on a daily basis. This not only unveils the risks posed in terms of computer and digital security, but also highlights the myriad possibilities that DF offer to favor better learning experiences. Hyperconnectivity is a trait of individuals pertaining the current digital society (as coined by Lévy, 2007), fact that encompasses not only the set of risks indicated in the report, but also offers a set of possibilities in social, cultural and educational domains (Cobo, 2016; Moravec, 2011). Thus, DFs, as an expression of this hyperconnection, constitute a resource that must be studied beyond mere simplistic perspectives that assume them as cold data, and instead, should be envisioned as indicators of the way in which people and organizations are configuring such a digital society.

On the other hand, the threefold nature of DF (data, action and service) helps supporting its value as a resource that can be used in educational research, since it allows both the analysis of the nature of the data linked to these traces of information, and the analysis of the ways in which fingerprints are generated, consumed, and used. The analysis of DF will have to attend to the multiple conceptions embedded in the data-action-service triad, which denotes the meanings that complement the concept of DF. If it is accepted that
a digital footprint is constructed from a cognitive, motivational and reflective perspective as Galimova, Konysheva, Kalugina and Sizova (2019) suggest, then, its conception as data-action-service can be useful in the educational field for the analysis of processes such as the acquisition of concepts, the interaction of people with knowledge, and its use. In this way, if the conceptions are examined in parallel with these dimensions, one can more likely find a relationship among these and some theoretical, axiological and experiential factors linked to the processes of generation and use of DF, which we propose as elements of analysis (Table 10).

Table 10

Relationship between conceptions, dimensions and factors associated with the processes of generation and use of DF

<table>
<thead>
<tr>
<th>Conceptions of the DF</th>
<th>DF configuration dimensions (Galimova et al., 2019)</th>
<th>Factors associated with the processes of generation and use of DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data (Madden et al., 2007)</td>
<td>Cognitive</td>
<td>Theoretical</td>
</tr>
<tr>
<td>Action (Sjöberg et al., 2017; K Varsha &amp; Monica, 2018)</td>
<td>Motivational-practical</td>
<td>Praxic</td>
</tr>
<tr>
<td>Service (Connelly &amp; Osborne, 2017; Osborne &amp; Connelly, 2016)</td>
<td>Reflective</td>
<td>Axiological</td>
</tr>
</tbody>
</table>

Additionally, some studies affirm that DFs allow the analysis of at least three components that can help determine the characteristics of the person or organization that produces them. These are: profile, content and network. Research such as that of Jainy, Kumaraguruy and Joshi (2013) and Garg, Raghuwanshi and Singh (2019) mention these three components, which seem to be linked to what Cobo (2016) refers to in his theory of 3C (container, content, and context), as it describes the elements from which any learning process can be studied and analyzed. We argue that DF, as manifestations of data, actions and services that people and organizations produce or consume in their lives and in their digital interactions, constitute themselves as potential learning and knowledge instruments, since DF reflect what people and organizations know (cognitive-theoretical dimension), do (motivational-experiential dimension) and value (axiological-reflexive dimension). As a consequence, the analysis that is made of DF in the educational field, forces us to always think about three elements. Firstly, the “containers” as the space in which DF are generated, consumed and produced (that is to say, spaces for digital interaction such as social networks). Secondly, the “context”, which makes reference to the hyperconnected digital society as a set of physical and symbolic conditions that surround its production and use. Thirdly, the “contents”, constituted by the DFs themselves, which are presented in the containers in the form of flat alphabetical text, of an image or multimedia content, or even in numerical form, displaying the activity of a person in any social network, translated for example in number of likes, number of followers, number of visits, etc.
Another phenomenon that is interesting to examine is the increase in DF studies in the organizational context where educational institutions are located, especially those of higher education such as universities. In any case, this phenomenon does not occur in all DF production spaces, as the amount of DF research in MOOC is scarce and even paradoxical (only Coursera involves more than 11 million people). All things considered, it can be argued that there is a trend towards DF studies focused on large groups of students, probably as a result of the emergence of technologies and resources such as educational data mining, learning analytics and educational big data, which facilitate and allow the analysis of enormous amounts of DF, once envisioned as an impossible mission. Hence, despite the development of these technologies, their use is largely oriented towards issues such as psychometric modeling, which does not necessarily require them, to the detriment of DF research in activities such as analysis of learning and formation of the students. The latter are pointed out in studies such as those by Abdelnour et al. (2015), Leon Urutia et al. (2017), Osborne and Conelly (2016) and Shafie et al. (2015), studies that confirm the challenges for universities in demographic, curricular, administrative, academic monitoring, and training strategies for students and even, for the making of policies concerning the normative, ethical and legal management of DF.

To conclude, it is worth highlighting that educational institutions (i.e. universities) are interested in the use of DF, not only because they understand them as a resource to reinforce their institutional image in the digital environment, but because they have begun to understand their potential to favor research in education, in the redesign of educational activities and processes. Such is the case of University of Southampton, University of Edinburgh, Malasya University, Universidad Autónoma de Madrid, Universidad de Madrid, and University of Morocco, among others. For these institutions, DF are now a resource that should be not only used, but also taught to manage students and the educational community in general. Tasks such as the aforementioned ones are examples of how to promote institutional development from the use of DF.

Lastly, it is worth noting that the studies coincide that DFs provide limited information on the interactions that take place in the physical world and in the physical spaces of the universities, for example. Hence, their analysis should always take place in a broader social and cultural framework than does not limit the understanding of DF as indicators of some characteristics and human behaviors, both individual and collective.

…if the changes are driven solely by a fingerprint, there is a risk of focusing only on factors that are directly related to numerical methods. However, if the changes are complemented by an understanding of how students approach their learning, the quality of the evidence used in the redesign increases significantly (Pardo et al., 2015, p. 305).

In this way, our DFs should serve to reveal who we are, and how we learn. DF should also serve to identify the ways of learning we should favor, the content we prefer, the learning paths we take in our digital interactions, the ways to represent content, the selection of content, the conditions under which content is produced and consumed, and of course, the side effects derived from these overarching actions.
References


M. E. Buitrago-Ropero Et Al.


https://www.scopus.com/inward/record.uri?eid=2-s2.085060009934&partnerID=40&md5=e2edf7b94d9c9343d884e114a08d96de


