Management Letters / Cuadernos de Gestión 22/1 (2022) 193-213

FESIDE
Fundación Emilio Soldevilla para la Investigación y Desarrollo de la Economía de la Empresa

enpresa institutua
Instituto de Economía Aplicada a la Empresa

Management Letters / Cuadernos de Gestión

journal homepage: http://www.ehu.eus/cuadernosdegestion/revista/es/

ISSN: 1131-6837 / e-ISSN: 1988-2157

Big Data, Accounting and International Development: Trends and challenges

Big Data, Contabilidad y Desarrollo Internacional: Desafíos y Tendencias

Sonia Arroyo Esteban*, Elena Urquía-Grande[a], Alberto Martínez de Silva[b], Raquel Pérez-Estébanez[c]

[a] *Associate Professor, PhD at Faculty of Economics and Business. Complutense University of Madrid, Campus de Somosaguas. Carretera de Húmera s/n, Pozuelo de Alarcón, Madrid 28223 Spain. – eurquiag@ucm.es – https://orcid.org/0000-0003-4724-7185*

[b] *Lecturer, PhD at Faculty of Economics and Business. Complutense University of Madrid, Campus de Somosaguas. Carretera de Húmera s/n, Pozuelo de Alarcón, Madrid 28223 Spain. – albermar@ucm.es – https://orcid.org/0000-0003-1885-2781*

[c] *Associate Professor, PhD at Faculty of Economics and Business. Complutense University of Madrid, Campus de Somosaguas. Carretera de Húmera s/n, Pozuelo de Alarcón, Madrid 28223 Spain. – raperez@ucm.es – https://orcid.org/0000-0002-8203-5594*

* ***Corresponding author:*** *PhD Student in Business Administration and Management. Faculty of Economics and Business, Complutense University of Madrid, Campus de Somosaguas. Carretera de Húmera s/n, Pozuelo de Alarcón, Madrid 28223 Spain. – soniarro@ucm.es – https://orcid.org/0000-0002-8734-8303*

ARTICLE INFO

Received 02 June 2021,
Accepted 22 November 2021

Available online 10 February 2022

DOI: 10.5295/cdg.211513sa

JEL: O1, M41

ABSTRACT

This article aims to show how the Big Data techniques application in accounting to monitor international cooperation projects are a green-field in the academic world. To obtain an exhaustive vision of the state of the art in academic research in this field, a bibliometric analysis has been carried out, based on multiple Web of Science searches, with focus on international development, Big Data and accounting, adding the holistic vision of the 17 SDGs or "Sustainable Development Goals" of the UN Agenda 2030. Research on Big Data, international development and accounting is a new field that has started in 2015 although academic literature is still scarce. Publications related to SDGs also begin on that date, but with much more prolific academic literature, without explicit references to the use of Big Data in accounting. The article finds deficiencies in existing academic research compared to other enterprise fields in which Big Data techniques are much more developed, and international organization reports lead this line of research, as opposed to the scholarly world. The main practical implication derived from the paper is the need to deepen in real cases of use outside the academic sphere as a starting point to develop this line of research. The development of this research area will help NPOs and governments to have a better accounting to evaluate the impact of their initiatives and cooperation projects. In addition to the bibliometric techniques used for the analysis of main publications, authors and relevant topics focused on this area of study, the authors consider a challenge and an opportunity to take the plunge into this field from academic world, which will undoubtedly improve decision-making in international development, emphasizing the need to gain momentum given the current state of greenfield.

Keywords: Big Data, International Development, Accounting, Monitoring, Sustainable Development Goals, Bibliometrics.

R E S U M E N

El objetivo de este artículo es mostrar cómo las técnicas de Big Data aplicadas a la contabilidad para el seguimiento y monitorización de los resultados y avances de los proyectos de cooperación internacional son un campo novedoso e incipiente en el mundo académico. Con el fin de obtener una visión exhaustiva del estado del arte de la investigación académica en este campo, se ha realizado un análisis bibliométrico detallado, basado en múltiples búsquedas en la Web of Science, con foco en los proyectos de cooperación internacional junto con Big Data y contabilidad, añadiendo la visión holística de los 17 Objetivos de Desarrollo Sostenible de la Agenda 2030 de la ONU.

La investigación sobre Big Data, cooperación internacional y contabilidad es un campo en desarrollo con referencias desde 2015 pero siendo la literatura académica todavía escasa. Las publicaciones relacionadas con los ODS también comienzan en esa fecha, pero con una literatura académica mucho más prolífica, sin embargo, sin referencias explícitas al uso de Big Data y contabilidad de forma conjunta. El artículo encuentra deficiencias en la investigación académica existente en comparación con otros campos donde las técnicas de Big Data están mucho más desarrolladas, siendo informes de organizaciones internacionales como la ONU los que lideran esta línea de investigación, frente al mundo académico. La principal implicación práctica que se deriva del artículo es la necesidad de profundizar en casos reales de uso fuera del ámbito académico como punto de partida para desarrollar esta línea de investigación. El desarrollo de esta línea de investigación ayudará a las ONG, gobiernos y administraciones públicas a tener una mejor contabilidad y evaluación del impacto de sus iniciativas y proyectos de cooperación. Además de las técnicas bibliométricas utilizadas para el análisis de las principales publicaciones, autores y temas relevantes enfocados a esta área de estudio, los autores consideran un reto y una oportunidad profundizar en este campo desde el mundo académico, lo que sin duda mejorará el soporte a la toma de decisiones en materia de desarrollo internacional.

Palabras clave: Big Data, Cooperación Internacional, Contabilidad, Control, Objetivo de Desarrollo Sostenible, Bibliometría.

1. INTRODUCTION

Big Data represents a new era in data exploration, achieving a revolution in many fields with a lot of applications in the business world. However, beyond the term "Big Data" there is a collection of data sources, technologies and methodologies that have emerged from, and to, exploit the exponential growth in data creation and management over the past decade (UN DESA, 2015) where the main purpose is to transform data into information and this information into insight to take better decisions in real time.

Although governments and Non-profit organizations (NPOs) usually delay in adopting new management techniques, they are already aware of the value of Big Data, and its possibilities to help support decisions in social initiatives (Kim *et al.*, 2014; Landefeld, 2014) and to enhance citizens´ life quality including economics, health care, job creation, natural disasters and terrorism.

To advance in this line, for example, the United Nations (UN) have boosted "Global Pulse", an initiative to promote the opportunities that big data presents for sustainable development and humanitarian action. National Statistical Institutes with UN Global Pulse are developing policies to measure the 17 goals and 169 targets of the 2030 Agenda for Sustainable Development. Another example is GSMA organization (Global System for Mobile communications association), which launched the Big Data for Social Good (BD4SG) initiative to leverage mobile operators' big data capabilities and help tackle health, environmental and other humanitarian challenges (*The State of Mobile Data for Social Good Report*, 2017) pointing the interests of mobile operators worldwide in this area. Hence, GSMA is focused on the development of initiatives that help to achieve the Sustainable Development Goals 2030 (SDGs) supported by the potential of mobile data. In this sense, the use of BD4SG can help in issues as varied as capturing feelings and concerns of the population via social networks (Helft, 2008; Roy *et al.*, 2020), explaining price changes in products as necessary as food (Baldwin *et al.*, 1984), predicting human migrations (Dekker *et al.*, 2018; Gillespie *et al.*, 2016; Hilbert, 2016; Oliver *et al.*, 2015; UN Global Pulse & UNHCR, 2017), ensuring financial inclusion (Óskarsdóttir *et al.*, 2018; Protopop, 2016) or improving the access to healthcare with an early diagnosis of diseases, helping to embrace decisions in case of humanitarian emergencies and reducing costs (Amankwah-Amoah, 2016; Hay *et al.*, 2013; Thara *et al.*, 2016).

Thus, mobile communications are a key source of data, due to its accelerated penetration in developing countries with respect to other basic goods. New information technologies (ITs) based in Big Data should help to provide a better life in these countries making available new data sources that serve both for analysis and modeling of their main needs and to fight poverty. The Data revolution will help monitor the progress of the programs of development, but also to advance evidence-based policies and programs that will enhance the most vulnerable.

In addition, the development of Open Data is also an important factor in international cooperation. Open data can be freely used, re-used and redistributed by anyone where the only requirement is to attribute and share (Bertot *et al.*, 2014; Davies *et al.*, 2019; Linders, 2013). Countries such as Singapore or Indonesia are worldwide references (UN DESA, 2015) and many other governments around the world are taking this type of initiatives. Open data complement massive sources such as mobile and sensor information as a part of Big Data ecosystem, helping to gather more information from the most disadvantaged areas with the aim of improving their living conditions.

The aim of this research work is to show the importance and gap of deepening the scientific knowledge in the areas of Big Data, Accounting and Sustainable Development Goals (SDGs). This main objective arises two sub-objectives. The first sub-objective is to analyze the state of art in academic research in the field of Big Data for International development linked with accounting, controlling, monitoring and planning (from now on whenever we refer to the word accounting including the other concepts). In complementary way, the second sub-objective it is to analyze from an academic perspective the evolution of this field in the 2030 SDGs framework, reference and guide of projects and actions in international cooperation in the coming years, where Big Data will be a keystone. Therefore, a bibliometric analysis has been performed with different techniques identifying the most relevant academic publications, the main interrelations between authors, articles, journals, countries and universities, as well as determining the most important themes within the area: Big Data, Accounting and international development. Additionally, this work defines trends and challenges in the area researched.

Studies about big data in accounting have increased (Arnaboldi *et al.*, 2017; Bhimani, 2020; Dagilienė & Klovienė, 2019; Janvrin & Weidenmier Watson, 2017; McKinney *et al.*, 2017). When linked also to international cooperation, academic research begins to be scarcer, even within the scope of the UN 2030 SDGs, which by their very nature seem to be a representative area of application of Big Data for measurement and accounting.

Bearing all this in mind the research questions could be defined as:

RQ1: Is academic literature linked with international development and big Data or SDGs?
RQ2: Is Big Data academic research related with SDGs?
RQ3: Is Big Data academic research linked with international development and accounting?

This article is structured as follows. As the methodology used is the bibliometric analysis a section for bibliographic framework is not needed. Thus, the next section will explain the data, instrument and methodology used. The methodology includes two phases: 1) conducting article searches on WoS using relevant search terms including Big data, accounting and international development and 2) importing the search results into the programming tool R to perform bibliometric analysis such as bibliographic coupling, co-word analysis and word cloud analysis among others. Next section will show the results together with its discussion. Finally, conclusions are defined highlighting some interesting insights. Firstly, Big Data development in international cooperation is less used than in other fields such as marketing, operations, social networks, so there is an opportunity to develop it. Secondly, the academic literature associated with the three fields studied (Big Data, international cooperation and accounting) is practically nonexistent, with reports from international organizations being the current spearhead. Thirdly, health and climate change are the most relevant themes. Finally, Anglo-Saxon countries play a fundamental role, although

China´s research is increasing remarkably about Big Data. The purpose of this research and its main contribution in the chosen field of study, in addition to the methodology and techniques applied, is to show the current decoupling of the academic world in the use of Big Data in the accounting and international cooperation frameworks. The authors argue that, in contrast to the fast incorporation of these techniques in both the public and private sectors (Gandomi and Haider, 2015), there is a gap in the academic field in which the principal current references for study are the reports of international organizations.

This bibliometric analysis aims to underline to the scientific community the need for more in-depth research related to Big Data, accounting and international cooperation, which although a very specific field can have a great impact when it comes to tackling problems in developing countries.

The future lines of research will be to analyze whether NPOs use the Big Data to improve their performance, launch projects or comply with the 2030 SDGs, even if the academic literature is scarce, identifying real cases that can serve as a basis for further scientific research.

2. DATA, INSTRUMENTS AND METHOD

Our central analysis is based on bibliometric methodology. Bibliometric methods are treated by an extensive scientific literature from 60´s (Kessler, 1963; De Solla Price, 1965; Pritchard, 1969; Nalimov *et al.*, 1971; Small, 1973; Garfield, 1979), they could be defined by the application of statistical analysis to the bibliographic production reviewing the majority of the literature collecting publications on a specific subject (Bufrem & Prates, 2005).

Bibliometric methods are used to find relationships among authors, universities, countries, sources or journals of the publications along a time horizon, to calculate impacts, to analyze trends in the subjects or to identify emerging areas considering the title words, abstracts or keywords assigned by authors or sources.

The methodology adopted to review the current state of academic research in the selected field is based on the following steps. First, the appropriate search terms are defined for the purpose of our research, and then the results are retrieved in different academic databases, performing a benchmark and assessing the best option among them. After, dataset chosen is imported in the programming tool R, using mainly its bibliometrix library and a preliminary exploratory analysis is completed. The aim is to accomplish a bibliometric analysis, applying different methods as bibliographic coupling, co-citation, collaboration, historical map in order to understand the academic interconnections, co-word and conceptual analysis, thematic evolution, strategic maps, "word-cloud" to deepen in the scientific knowledge in the research field of Big Data, Accounting, International Development and Sustainable Development Goals (SDGs).

2.1. Data

We have analyzed the three main academic databases, taking into account decision variables such as the number of results, search facilities and the quality of the cited references in each article with analogous searches. After this comparison, both Scopus and Google Scholar were discarded, selecting Web of Science (WoS) as the source for the bibliographic analysis, although the three databases have advantages and disadvantages.

Google Scholar has been discarded due to its limitations of search delimitation, especially for complex searches, exporting data and difficulties for the co-citation analysis. It is not a database but a search engine, and there is no data structure for abstract or information of the citations that are referenced in each article. However the main reason for discarding Google Scholar is its limited output, added to his significant weaknesses in the accuracy of the search (see Table 1) (López-Cózar *et al.*, 2014). In any case, Google Scholar is an important tool, suitable for quick searches, first contact with a subject or limited analyses (Adriaanse & Rensleigh, 2011, 2013; Boeker *et al.*, 2013), standing out for its major alternative sources that enrich and complement academic analysis, especially in Social Sciences (Kousha & Thelwall, 2008).

Table 1
Examples of records number by search and database

Google "Sustainable development goals"	Scopus "Sustainable development goals"	WoS "Sustainable development goals"
Search records number	Search records number	Search records number
69700	4829	4271
Google "Big Data"	Scopus "Big Data"	WoS "Big Data"
Search records number	Search records number	Search records number
417000	66861	41195
Google "Sustainable development goals", "Big Data"	Scopus "Sustainable development goals", "Big Data"	WoS "Sustainable development goals", "Big Data"
Search records number	Search records number	Search records number
7880	39	36
Google "Sustainable development goals", "Big Data", "Accounting"	Scopus "Sustainable development goals", "Big Data", "Accounting"	WoS "Sustainable development goals", "Big Data", "Accounting"
Search records number	Search records number	Search records number
3740	1	2

Source: Compiled by the authors from Wos, Scopus and Google Scholar.

Regarding Scopus dismissal, it is mainly explained by the cited references in each article. Scopus and WoS have both advantages and disadvantages, the results are similar, with the largest number of publications in Scopus.

The main factor in selecting WoS versus Scopus, is the consistency of the information of cited references in each of the publications recalled. Scopus was rejected due to the lower standardization of some references in articles (Leydesdorff *et al.*, 2010), not only with international reports or UN resolutions so important in the domain of SDGs, but also in other type of articles. While WoS carries out a normalization of the references, Scopus, does not encode them, which implies dates in different positions, with slightly different names of authors or titles, different page numbers, references without date, authors or name of the article (Franceschini *et al.*, 2016). For example, UN Resolution "Transforming our world: The 2030 Agenda for sustainable development (2015)" is referenced in more than two hundred different ways, hindering the co-citation and coupling

analysis. WoS also have these problems, but to a lesser extent due to a better standardization of references.

Although we selected WoS for analysis, the searches were conducted analogously in the three databases, including articles up to August 2019. The collecting of retrieved articles contains the following words and logic relationships between them in the tittle, abstract or keywords: ("*big data**" and "*international development**") or ("big data*" and "social good*") or ("*big data**" and "*international cooperation**") or ("*millennium development goals**" *and "big data**") or ("*sustainable development goals**") or ("*sdg**" and "*indicators**" and "*target**"). The number of articles found in the base search of WoS was 4,323 (see Table 2).

Even though one final purposes of our research is the use of Big data in international cooperation, in baseline search it has been decided (Search RQ1) not to filter the subquerys associated with the SDGs ("*sustainable development goals**") or ("*sdg** and "*indicators** and "*target**") by this term, to have a general and complete knowledge of the academic literature dealing with these indicators and goals. It is in the world of targets, key performance indicators (KPIs) and SDGs where Big Data should help the most. However, we want to understand the full context of current academic research within the scope of "Sustainable Development Goals". Then, a complementary analysis will be carried out by restricting the subquerys related to the SDGs/ Sustainable development goals also with the concept of Big Data (Search RQ2, 77 records). Finally, articles containing some of the terms related to accounting in addition to Big Data will be analyzed, obtaining a very small number of publications (see Table 2, Search RQ3, 21 records).

Table 2
Records in WoS for the baseline search and restricted searches

WoS: TS= ("*big data* *" and "international development*") OR TS=("sustainable development goals*") OR TS= ("sdg*" and "indicators*" and "target*") OR TS= ("*big data* *" and "social good*") OR TS=("*big data* *" and "international cooperation*") OR TS= ("Millennium Development goals*" and "*big data* *")

	until 2015	2016-2017	2018-2019	Total
Total publications	437	1.588	2.298	4.323

RQ1
(Baseline, referential, or initial search)

WoS: TS= ("*big data* *" and "international development*") OR TS=("sustainable development goals*" and "*big data* *") OR TS= ("sdg*" and "indicators*" and "target*" and "*big data** ") OR TS= ("*big data* *" and "social good*") OR TS=("*big data* *" and "international cooperation*") OR TS= ("Millennium Development goals*" and "*big data* *")

	until 2015	2016-2017	2018-2019	Total
Total publications	11	22	44	77

RQ2
RQ1+ Big Data Mandatory in Sustainable Development Goals

WoS: (TS= ("*big data* *" and "international development*") OR TS=("sustainable development goals*" and "*big data* *") OR TS= ("sdg*" and "indicators*" and "target*" and "*big data** ") OR TS= ("*big data* *" and "social good*") OR TS=("*big data* *" and "international cooperation*") OR TS= ("Millennium Development goals*" and "*big data* *")) and TS=("accounting*" or "controlling*" or "planning*" or "monitoring*")

	until 2015	2016-2017	2018-2019	Total
Total publications	3	7	11	21

RQ3
RQ3 + some term related to accounting (accounting or controlling or planning or monitoring) mandatory

Source: Compiled by the authors from WoS.

The analysis flow will include three searches: a first search we have denominated "Search RQ1" because it is the one referred to research question 1 which examines literature about international development and Big Data, and SDGs in general. For the SDGs, Big Data has not been a mandatory concept, understanding that the natural trend to measure these concepts will be to apply Big Data techniques. The second search denominated "Search RQ2" include "Big Data" in SDGs too. Finally, the issue "Accounting" is added to the search and will be denominated "Search RQ3". The baseline query (Search RQ1) is composed by two premises (see Figure 1). It is worth noting that most of the obtained literature in RQ1 is related to Sustainable Development Goals (4,283 publications) and a minimum part is associated directly to Big Data applied to international development (42 publications, with 2 publications in common). In addition, it highlights the high number of articles associated with the term Big Data in WoS (41,195 publications) and its minimum relationship with concepts associated with both international cooperation and SDGs. More than 70% of articles associated with Big Data deal with technical issues, far from social sciences, economics, health and environment. Secondly, the number of publications associated

with international cooperation, including the former UN MDMs program (with the terms: *"international cooperation", "international development", "social good", "Millennium Development Goals"),* is also important (16,850 publications). The same as those articles related to SDGs (4,283 publications). The lack of relationship between the publications associated with the SDGs or international cooperation and the term Big Data is a key conclusion regarding RQ2.

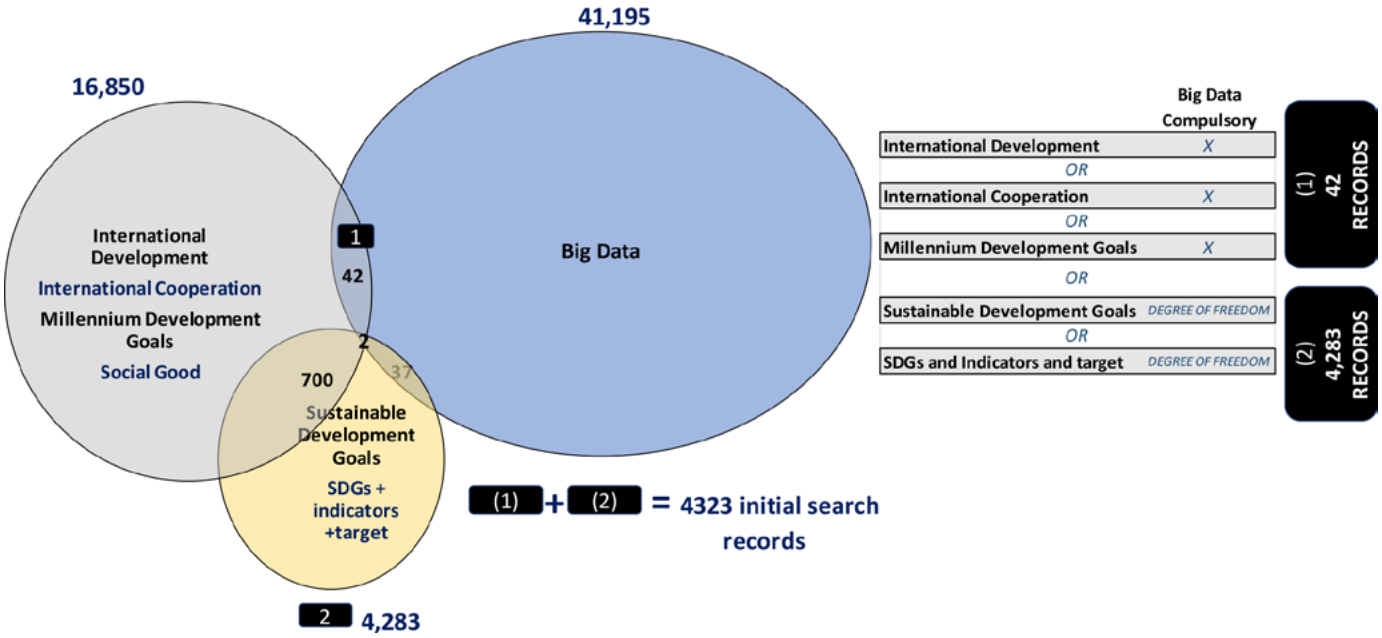

Figure 1
Number of records in the search restrictions
Source: Own study based on data retrieved from WoS.

2.2. Instruments

This bibliometric study has been programmed mostly in the language R, an open sourced programming language specialized for statistical computing and graphic visualization. R has the Bibliometrix package (Aria & Cuccurullo, 2017) an open source tool for quantitative research in scientometrics and bibliometrics that we have used.

This package provides routines for importing bibliographic data from SCOPUS, WoS, and other databases, performing bibliometric analysis and building data matrices for co-citation, coupling, scientific collaboration analysis and co-word analysis. One of the main advantages of bibliometrix is its flexibility and that it can be rapidly upgraded and integrated with other statistical R-packages.

2.3. Methods

Different analyses will be carried out, from a basic exploratory analysis, based on analysis of the sample, to more complex bibliometric techniques. To analyze the structure and relationships of our scientific research field, bibliometric techniques such as co-citation, bibliographic coupling will be employed, and to discover the main concepts and topics covered in our research field, techniques such as co-word, conceptual or strategic maps will be applied. Furthermore, this research is going to use science mapping techniques (Boyack *et al.*, 2013; Calero Medina & Van Leeuwen, 2012; Small, 1999) to represent the cognitive structure of our research field based on relationships and graphs determined by the aforementioned analyses such as co-citation, coupling or co-word, similar to the analysis in social networks.

2.3.1. Network techniques

To understand relationships between different authors or level aggregations (sources, universities, countries...) co-citation (Marshakova, 1981; Small, 1973) and bibliographic coupling (Kessler, 1963) are key analysis in Bibliometrics. Co-citation analysis establishes relationships between pairs of papers that are jointly cited in the source articles. When the same pairs of papers are co-cited by many authors, it is an indicative of a possible research cluster. Bibliographic coupling is based on the frequency of shared cited references: if two works often cite the same literature, it is likely that they are related to the same topic. Co-citation allows studying the evolution, maturity and growth of research areas from several perspectives, and even though bibliographic coupling is also a good indicator to the relationship between sources, countries and authors, its impact is more difficult to quantify, because the relationships are static, while the co-citation changes over time. Cited references used by bibliographic coupling are fixed, whereas similarity between

documents determined by co-citation can change as new citing papers are published. In addition, the most cited articles are an indicator of scientific relevance and their clusters enable the identification and evaluation of relationships of topics, authors and journals over a period of time. Recently, there has been a revival in the use of bibliographic coupling that is challenging the historical preference for co-citation analysis proposing it in a complementary way (Bichteler & Eaton, 1980; Jarneving, 2007; Meyer *et al.*, 2014). Thereby, bibliographic coupling is able to cluster more recent papers instead of older ones, while co-citation clustering does the opposite, it clusters the older documents but cannot cluster the more recent works that have not yet been cited (Boyack & Klavans, 2010; Kuusi & Meyer, 2007).

Moreover, the collaboration networks are based on the co-authorship of articles and are important in the development of subject areas and in the dissemination of research results (Melin & Persson, 1996). It is interesting to explore the dynamics of scientific collaboration between higher levels, such as universities or countries cited (Alcaide & Ferri, 2014; Bergmann *et al.*, 2017; Boyack, 2009).

2.3.2. Content and Conceptual techniques

A co-word analysis along with other dependent methods is conducted to understand the thematic evolution in the research field chosen, deepening in the conceptual framework, evaluating knowledge status and identifying trends and emerging topics (Callon *et al.*, 1983; Cobo *et al.*, 2011; Williams & Plouffe, 2007). Co-word analysis is based in the association between words or terms shared by documents, exploring with co-ocurrence analysis the relationships among keywords, subject headings or abstract´s words. With this analysis we can build conceptual maps of thematic areas or show strategic diagrams (Callon *et al.*, 1991; López-Herrera *et al.*, 2009; Muñoz-Leiva *et al.*, 2012). These analyses can be done at a specific time periods or along a time horizon complemented with clustering methods.

Thematic evolution and strategic maps are performed based on co-word network analysis and clustering following the definitions of Cobo *et al.*, 2011, where the words are grouped by themes, whose density and centrality, classifying them in a two-dimensional diagram.

3. EXPLORATORY ANALYSIS RESULTS

To deepen in the selected field of research, an exploratory analysis of the 4,323 registers of our dataset will be carried out before the bibliometric analysis. For this exploratory analysis we loaded in R three datasets: First, baseline dataset shaped by the initial search (Search RQ1), described in section 2.1, where the term Big data is mandatory in all sub-sentences associated to international cooperation and development, but not in those associated to the SDGs, since we want to have a complete overview of these. The second dataset (Search RQ2) is based on the previous one but includes the term Big Data obligatory in all the sentences, including SDGs sub-sentences. The third (Search RQ3), is based on Search RQ2, and in addition includes as compulsory terms those related to accounting.

Results showed that the number of publications about our baseline search (Search RQ1) have grown considerably since 2013, with 2015 to 2018 being years with the highest growth rates and volume of articles. Big Data explosion, in all areas of study, occurred earlier, in the years 2012 to 2016, reaching 8,977 articles in WoS in 2017, and with flat growth in 2018. This leads us to conclude that there is no clear correlation between the "Big Data Search" and our Search RQ1, being better correlated with Search RQ2 and related to international cooperation and the SDGs. In fact, the higher increase of registers in our search coincides with the launch of the UN program: "Sustainable development goals 2030" in 2015. If we review search RQ2, where Big data is always a mandatory term, including SDGs sentences, we can appreciate the low number of papers obtained. This indicates that although the application of Big Data in the world of the SDGs should be important, it is a scholar green-field and an opportunity to deepen the study. With the search further restricted to terms of related to accounting we obtain 21 articles meaning there is an open field in full development.

Regarding the countries, results show that U.S and China have emerged as two Big Data superpowers in the business world. However, in our baseline query the countries of reference in the academic world in number of articles (first authors) are USA and UK, followed by Australia, Germany, Switzerland, Canada, China and Netherlands (see Figure 2, Search RQ1). Switzerland has important organizations like WHO, and Netherlands stands out by EU institutions and the Wageningen University Research, which focuses on the field of life sciences and natural resources, highlighting both in the world of SDGs. Nevertheless, China, UK and USA (11 articles each one) stand out as reference countries when the more restrictive search is made, and the word "Big Data" is mandatory in all sentences (see Figure 2, Search RQ2). When we filter restricting the search also to accounting concepts (see Figure 2, Search RQ3) the number of articles is very small and unrepresentative, it is more difficult to draw conclusions being China and Japan the most productive countries. Regarding academic collaboration only Switzerland, UK, Canada, Sweden and France exceed 50% of articles with multi-country co-authors in Search RQ1 and USA, China present 41% of multi-country publications. In search RQ3, the low rate of collaboration stands out, probably explained by the low relevance of the Big Data issue associated with accounting in the world of international cooperation, and the few existing publications. The most relevant issue of the productivity by countries (first authors) in search RQ3 is that USA does not appear as one of the most relevant countries. However, if we analyze the articles related to the entire Big Data scope in WoS, 55% of the articles are produced in the USA (28%) and China (27%), while UK produces only 7% of the articles. Our Search RQ1 is not aligned with the most prolific countries in Big Data publications, though it is with the most prolific countries in academic research in international cooperation, explained by the weight of the number of articles related to the SDGs in Search RQ1.

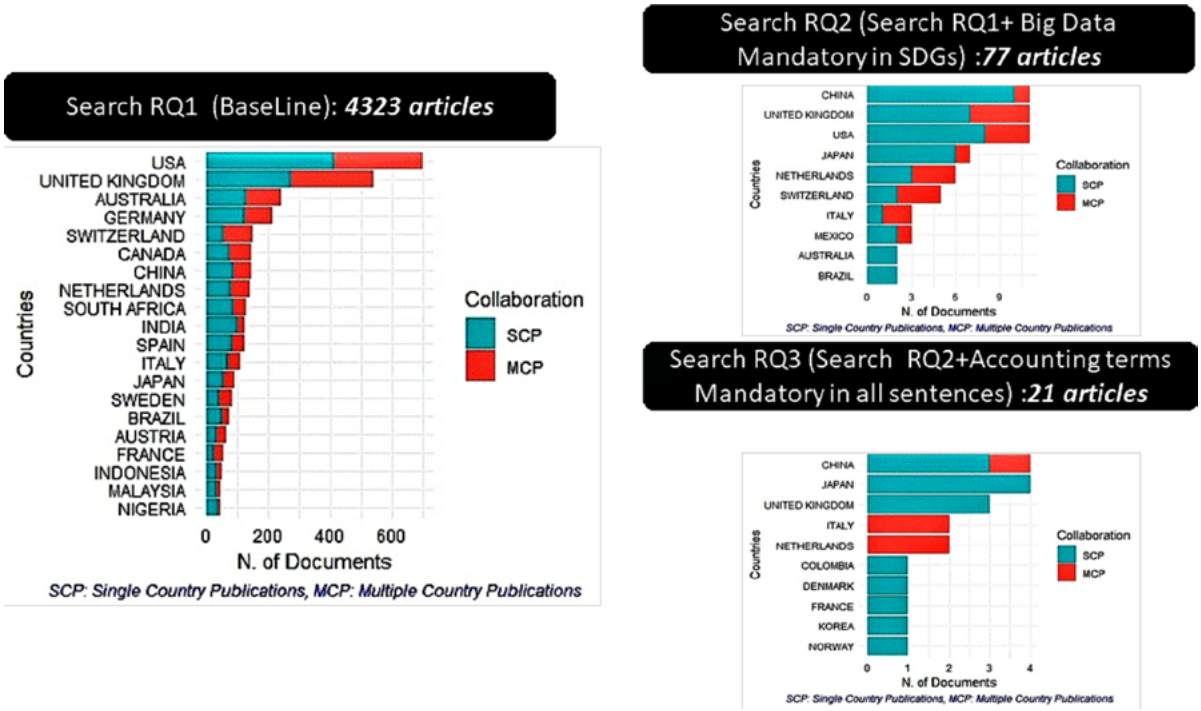

Figure 2

Most Productive Countries based in Corresponding Authors

Source: Own study based on data retrieved from WoS using Bibliometrix R package

Regarding the most productive universities Anglo-Saxon Universities are leaders in number of papers (see Table S1 — supplementary online material— for further details, Search RQ1). Most of the top 15 universities are in UK and USA. Australia, Canada and South Africa enter in the Top 15. The previously mentioned Wageningen University Research, located in Netherlands is considered world class in agricultural and environmental science, being in the Top 15 with 66 articles published as the only European representative. The importance of organizations such as WHO or World Bank in the academic development of this research field is another outstanding issue. In Search RQ2 (see Table S1), the universities that appear are very different, highlighting some such as MIT and California with a more technical profile. It is important to note the fact that no Chinese organization appears among the most relevant in number of articles by university/organization, despite being among the best ranked countries according to the first author. In Search RQ1, this is due to its relative position with respect to UK and USA, while in Search RQ2 or RQ3, it is due to the dispersion of articles published among the different Chinese universities and a lower rate of inter-university collaboration.

The most cited references by the articles collected in Search RQ1 (see Table 3). The "local citation" concept has been marked with an "x" when the cited reference is in the list of articles of the search, meaning it appears both in the search and referenced by other papers with more articles locally cited in the last periods in proportional terms. Other important consideration is that the reports/resolutions of institutions like UN or WHO may be anonymous and not normalized. In fact, the article "Transform our world 2030" appears twice among the most cited until 2015, since it is referenced in two different ways in the search. Most of the more cited articles that appear in the table, that are not local citations, are resolutions of international organizations not indexed in WoS (neither in Scopus and nor even in Google Scholar).

Table 3

Most cited references by year in RQ1

Year	Most cited references (before 2016)	Frecuency	Local citation
2015	UNITED NATIONS (UN), 2015, TRANSF OUR WORLD 203.	352	
2013	GRIGGS D, 2013, NATURE, V495, P305, DOI 10.1038/495305A.	143	x
2012	SACHS JD, 2012, LANCET, V379, P2206, DOI 10.1016/S0140-6736(12)60685-0.	133	x
2015	STEFFEN W, 2015, SCIENCE, V347, DOI 10.1126/SCIENCE.1259855.	121	
2009	ROCKSTROM J, 2009, NATURE, V461, P472, DOI 10.1038/461472A.	113	
2015	[ANONYMOUS], 2015, MILL DEV GOALS REP 2.	102	
2015	UNITED NATIONS, 2015, SUST DEV GOALS.	77	
2015	LE BLANC D, 2015, SUSTAIN DEV, V23, P176, DOI 10.1002/SD.1582.	74	x
1999	SEN A., 1999, DEV FREEDOM.	65	
2015	UNITED NATIONS GENERAL ASSEMBLY, 2015, TRANSF OUR WORLD 203.	61	
Year	**Most cited references (2016-2017)**	**Frecuency**	**Local citation**
2016	NILSSON M, 2016, NATURE, V534, P320, DOI 10.1038/534320A.	111	x
2016	HAK T, 2016, ECOL INDIC, V60, P565, DOI 10.1016/J.ECOLIND.2015.08.003.	50	x
2016	LIM SS, 2016, LANCET, V388, P1813, DOI 10.1016/S0140-6736(16)31467-2.	38	x
2016	WANG HD, 2016, LANCET, V388, P1459, DOI 10.1016/S0140-6736(16)31012-1.	31	
2016	KEESSTRA SD, 2016, SOIL-GERMANY, V2, P111, DOI 10.5194/SOIL-2-111-2016.	28	x
2016	WHO, 2016, WORLD HLTH STAT 2016.	26	
2017	PRADHAN P, 2017, EARTHS FUTURE, V5, P1169, DOI 10.1002/2017EF000632.	25	x
2016	UNITED NATIONS, 2016, SUST DEV GOALS.	25	
2017	BIERMANN F, 2017, CURR OPIN ENV SUST, V26-27, P26, DOI 10.1016/J.COSUST.20	23	x
2017	STAFFORD-SMITH M, 2017, SUSTAIN SCI, V12, P911, DOI 10.1007/S11625-016-03	23	x
Year	**Most cited references (2018-2019)**	**Frecuency**	**Local citation**
2018	NERINI FF, 2018, NAT ENERGY, V3, P10, DOI 10.1038/S41560-017-0036-5.	18	x
2018	UNITED NATIONS, 2018, SUST DEV GOALS.	13	
2018	COWIE AL, 2018, ENVIRON SCI POLICY, V79, P25, DOI 10.1016/J.ENVSCI.2017.10.0	11	x
2018	LEAL W, 2018, INT J SUST DEV WORLD, V25, P131, DOI 10.1080/13504509.2017.1	11	x
2018	UNITED NATIONS (UN), 2018, SUST DEV GOALS REP 2.	11	
2018	MCCOLLUM DL, 2018, ENVIRON RES LETT, V13, DOI 10.1088/1748-9326/AAAFE3.	10	x
2018	O'NEILL DW, 2018, NAT SUSTAIN, V1, P88, DOI 10.1038/S41893-018-0021-4.	10	
2018	WEITZ N, 2018, SUSTAIN SCI, V13, P531, DOI 10.1007/S11625-017-0470-0.	10	x
2018	HOGAN DR, 2018, LANCET GLOB HEALTH, V6, PE152, DOI 10.1016/S2214-109X(17)	9	x
2018	KRUK ME, 2018, LANCET GLOB HEALTH, V6, PE1196, DOI 10.1016/S2214-109X(18)	9	x

Source: Own elaboration based on data retrieved from WoS.

The low number of local citations in the most limited searches (Search RQ2 and Search RQ3) stands out. In search RQ2, many of the most cited references deal with smart cities topics (see Here Table 4), revealing the low development of the studied research field, once we have related the terms SDGs and Big Data to each other. Search RQ3 given the dispersion of references does not present illustrative results (see Here Table 4, Search RQ3).

Table 4
Most cited references in the RQ2 and RQ3

RQ2 (RQ1+ Big Data Mandatory in SDGs) :*77 articles*				
Year	*Most cited references*	*Frecuency*	*Local citation*	*Titte*
2014	KITCHIN R, 2014, GEOJOURNAL, V79, P1, DOI 10.1007/S10708-013-9516-8.	7		The real-time city? Big data and smart urbanism
2013	MAYER-SCHONBERGER V, 2013, BIG DATA REVOLUTION.	6		Big data: A revolution that will transform how we live, work, and think
2012	BATTY M, 2012, EUR PHYS J SPEC TOP, V214, P481, DOI 10.1140/EPJST/E2012-01703-3.	5		Smart cities of the future
2011	BENGTSSON L, 2011, PLOS MED, V8, DOI 10.1371/JOURNAL.PMED.1001083.	5		Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti
1999	SEN AMARTYA, 1999, DEV FREEDOM.	5		Development as freedom (1999)
2015	AL NUAIMI E, 2015, J INTERNET SERV APPL, V6, DOI 10.1186/S13174-015-0041-5.	4		Applications of big data to smart cities
2015	ALBINO V, 2015, J URBAN TECHNOL, V22, P3, DOI 10.1080/10630732.2014.942092.	4		Smart cities: Definitions, dimensions, performance, and initiatives
2015	BLUMENSTOCK J, 2015, SCIENCE, V350, P1073, DOI 10.1126/SCIENCE.AAC4420.	4		Predicting poverty and wealth from mobile phone metadata
2012	BOYD D, 2012, INFORM COMMUN SOC, V15, P662, DOI 10.1080/1369118X.2012.678878.	4		Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon
2016	HILBERT M, 2016, DEV POLICY REV, V34, P135, DOI 10.1111/DPR.12142.	4	x	Big data for development: A review of promises and challenges
RQ3 (RQ2+Accounting terms Mandatory in all sentences) :*21 articles*				
Year	*Most cited references*	*Frecuency*	*Local citation*	*Titte*
2012	BATTY M, 2012, EUR PHYS J-SPEC TOP, V214, P481, DOI 10.1140/EPJST/E2012-01703-3.	3		Smart cities of the future
1999	SEN AMARTYA, 1999, DEV FREEDOM.	3		Development as freedom (1999)

Source: Own elaboration based on data retrieved from WoS.

Regarding the most productive authors of our search such as Bhutta, Liu or Bartram (Figure 3) these are not necessarily the most relevant. Relevant authors are those that have many citations and somehow are a reference in the research issues. Only researchers Nilsson, Gupta, Li and Bhutta are locally referenced more than 100 times in Search RQ1 followed by Salomon, Allen y Murray. On the other hand, many of the publications of the most productive authors are quantitative studies, more or less directly related with the SDGs in different areas (Health is treated by Buttha, Water and Health is treated by Bartram, and Environment and Sustainability by Liu). In addition, many of these authors share a high number of publications. In the bibliographic analysis we will analyze if these authors are a reference in our line of research or if they are just prolific authors. In the most constrained searches RQ2 and RQ3, given the resulting small sample, it becomes more difficult to draw conclusions. In Search RQ2, the most productive authors are Taylor, whose articles deal with models and applicability of big data in international development, and Heeks, who presents two recent and remarkable theoretical articles focusing on justice, governance, or ethics in data. Sekimoto presents articles closely related to urban mobility based on mobile data and appears as prominent in both search RQ2 and RQ3.

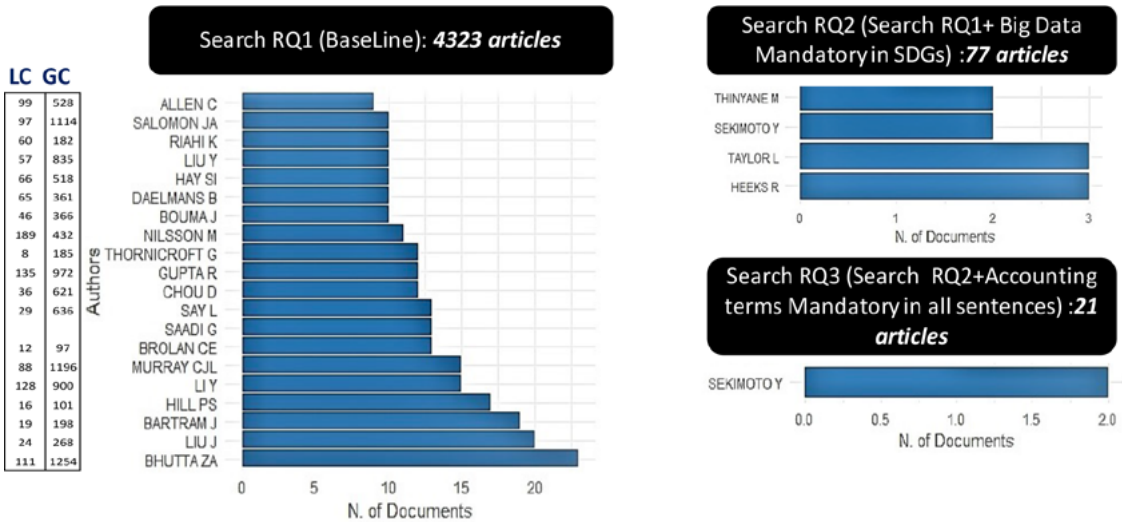

Figure 3
Most productive authors in each search
(*) LC=Local Citations, GC=Global Citations in WoS Database
Source: Own study based on data retrieved from WoS using Bibliometrix R package.

Most of the authors' production start after 2015, when our line of research begins to develop according to the launch of UN program (see Figure S1-supplementary online material). The greater productivity and the greater number of citations per year in WoS comes from studies linked to the area of health, promoted by organizations such as the UN, WHO, Global Burden of Disease, with a large number of participants in the authorship of the same, many citations in WoS, and framed in the SDG 3 (Health). Thus, the most relevant authors by number of global citations per year are Bhutta, Murray, Salomon, Gupta, Li, authors not involved in Big Data and accounting research, and specifically related to the SDGs field.

The most relevant articles in number of citations in WoS in search RQ2 and RQ3 are listed in Table 5. In search RQ2 the journal Development Policy Review appears two times. Only the article of Bebbington, 2018 is published in an explicit accounting journal and a second article, Di Bella, 2018, is published in a Social Sciences area with focus in Quality of Life Measurement. The main problem of this scarcity of academic literature is the joint filter with the concept Big Data because the world of international cooperation and accounting presents more publications. Removing Big Data, we found 2,410 papers in WoS related to accounting, controlling, monitoring or planning and international development. Undoubtedly the confluence of the three worlds is still an incipient field.

Table 5
Top Articles by WoS Citations in RQ2 and RQ3

RQ2 (RQ1+ Big Data Mandatory in SDGs): ***77 articles***

Paper	DETAIL TITLE	Nº Citations in WoS
BIBRI SE, 2017, SUST CITIES SOC	SMART SUSTAINABLE CITIES OF THE FUTURE: AN EXTENSIVE INTERDISCIPLINARY LITERATURE REVIEW	90
HILBERT M, 2016, DEV POLICY REV	BIG DATA FOR DEVELOPMENT: A REVIEW OF PROMISES AND CHALLENGES	86
KISSLING WD, 2018, BIOL REV	BUILDING ESSENTIAL BIODIVERSITY VARIABLES (EBVS) OF SPECIES DISTRIBUTION AND ABUNDANCE AT A GLOBAL SCALE	44
TAYLOR L, 2015, GEOFORUM	IN THE NAME OF DEVELOPMENT: POWER, PROFIT AND THE DATAFICATION OF THE GLOBAL SOUTH	22
TAYLOR L, 2015, GEOJOURNAL	IS BIGGER BETTER? THE EMERGENCE OF BIG DATA AS A TOOL FOR INTERNATIONAL DEVELOPMENT POLICY	22
BEBBINGTON J, 2018, ACCOUNT AUDIT ACCOUNT	ACHIEVING THE UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS: AN ENABLING ROLE FOR ACCOUNTING RESEARCH	12
VARSHNEY KR, 2015, BIG DATA	TARGETING VILLAGES FOR RURAL DEVELOPMENT USING SATELLITE IMAGE ANALYSIS	11
BECK EJ, 2016, GLOB HEALTH ACTION	PROTECTING THE CONFIDENTIALITY AND SECURITY OF PERSONAL HEALTH INFORMATION IN LOW- AND MIDDLE-INCOME COUNTRIES IN THE ERA OF SDGS AND BIG DATA	9
HALEWOOD M, 2018, NEW PHYTOL	PLANT GENETIC RESOURCES FOR FOOD AND AGRICULTURE: OPPORTUNITIES AND CHALLENGES EMERGING FROM THE SCIENCE AND INFORMATION TECHNOLOGY REVOLUTION	8
HEEKS R, 2018, INF DEV	DATA JUSTICE FOR DEVELOPMENT: WHAT WOULD IT MEAN?	8
SEELE P, 2016, SUSTAIN SCI	ENVISIONING THE DIGITAL SUSTAINABILITY PANOPTICON: A THOUGHT EXPERIMENT OF HOW BIG DATA MAY HELP ADVANCING SUSTAINABILITY IN THE DIGITAL AGE	7
WILLIAMS C, 2017, INT J HUM RIGHTS	NEGLECTING HUMAN RIGHTS: ACCOUNTABILITY, DATA AND SUSTAINABLE DEVELOPMENT GOAL 3	6
BONILLA SH, 2018, SUSTAINABILITY	INDUSTRY 4.0 AND SUSTAINABILITY IMPLICATIONS: A SCENARIO-BASED ANALYSIS OF THE IMPACTS AND CHALLENGES	5
TAYLOR L, 2016, PHILOS TRANS R SOC A-MATH PHYS ENG SCI	THE ETHICS OF BIG DATA AS A PUBLIC GOOD: WHICH PUBLIC? WHOSE GOOD?	5
LATIF S, 2017, FUTURE INTERNET	HOW 5G WIRELESS (AND CONCOMITANT TECHNOLOGIES) WILL REVOLUTIONIZE HEALTHCARE?	4
XU Y, 2017, J R SOC INTERFACE	COLLECTIVE BENEFITS IN TRAFFIC DURING MEGA EVENTS VIA THE USE OF INFORMATION TECHNOLOGIES	4
FLYVERBOM M, 2017, INF SOC	BIG DATA AS GOVERNMENTALITY IN INTERNATIONAL DEVELOPMENT: DIGITAL TRACES, ALGORITHMS, AND ALTERED VISIBILITIES	4
PFEFFER K, 2016, EUR J DEV RES	PERSPECTIVES ON THE ROLE OF GEO-TECHNOLOGIES FOR ADDRESSING CONTEMPORARY URBAN ISSUES: IMPLICATIONS FOR IDS	4
BOGOMOLOV A, 2015, BIG DATA	MOVES ON THE STREET: CLASSIFYING CRIME HOTSPOTS USING AGGREGATED ANONYMIZED DATA ON PEOPLE DYNAMICS	4

RQ3 (RQ2+Accounting terms Mandatory in all sentences) :***21 articles***

Paper	DETAIL TITLE	Nº Citations in WoS
BIBRI SE, 2017, SUST CITIES SOC	SMART SUSTAINABLE CITIES OF THE FUTURE: AN EXTENSIVE INTERDISCIPLINARY LITERATURE REVIEW	90
KISSLING WD, 2018, BIOL REV	BUILDING ESSENTIAL BIODIVERSITY VARIABLES (EBVS) OF SPECIES DISTRIBUTION AND ABUNDANCE AT A GLOBAL SCALE	44
TAYLOR L, 2015, GEOJOURNAL	IS BIGGER BETTER? THE EMERGENCE OF BIG DATA AS A TOOL FOR INTERNATIONAL DEVELOPMENT POLICY	22
BEBBINGTON J, 2018, ACCOUNT AUDIT ACCOUNT	ACHIEVING THE UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS: AN ENABLING ROLE FOR ACCOUNTING RESEARCH	12
VARSHNEY KR, 2015, BIG DATA	TARGETING VILLAGES FOR RURAL DEVELOPMENT USING SATELLITE IMAGE ANALYSIS	11
WILLIAMS C, 2017, INT J HUM RIGHTS	NEGLECTING HUMAN RIGHTS: ACCOUNTABILITY, DATA AND SUSTAINABLE DEVELOPMENT GOAL 3	6
FLYVERBOM M, 2017, INF SOC	BIG DATA AS GOVERNMENTALITY IN INTERNATIONAL DEVELOPMENT: DIGITAL TRACES, ALGORITHMS, AND ALTERED VISIBILITIES	4
BOGOMOLOV A, 2015, BIG DATA	MOVES ON THE STREET: CLASSIFYING CRIME HOTSPOTS USING AGGREGATED ANONYMIZED DATA ON PEOPLE DYNAMICS	4
RAVANELLI R, 2018, REMOTE SENS	MONITORING THE IMPACT OF LAND COVER CHANGE ON SURFACE URBAN HEAT ISLAND THROUGH GOOGLE EARTH ENGINE: PROPOSAL OF A GLOBAL METHODOLOGY, FIRST APPLICATIONS AND PROBLEMS	3
BATRAN M, 2018, ISPRS INT GEO-INF	INFERENCING HUMAN SPATIOTEMPORAL MOBILITY IN GREATER MAPUTO VIA MOBILE PHONE BIG DATA MINING	3
XIE P, 2019, INT J ENERGY RES	A REVIEW ON CHINA'S ENERGY STORAGE INDUSTRY UNDER THE ``INTERNET PLUS'' INITIATIVE	1
VIRTO LR, 2018, MAR POL	A PRELIMINARY ASSESSMENT OF THE INDICATORS FOR SUSTAINABLE DEVELOPMENT GOAL (SDG) 14 ''CONSERVE AND SUSTAINABLY USE THE OCEANS, SEAS AND MARINE RESOURCES FOR SUSTAINABLE DEVELOPMENT''	1
LWIN KK, 2018, J DISASTER RES	DEVELOPMENT OF GIS INTEGRATED BIG DATA RESEARCH TOOLBOX (BIGGIS-RTX) FOR MOBILE CDR DATA PROCESSING IN DISASTERS MANAGEMENT	1

Source: Own elaboration based on data retrieved from WoS.

4. BIBLIOMETRIC ANALYSIS RESULTS

Bibliometric encompasses several empirical methods, such as bibliographic coupling and co-citation analysis together with other analysis more conceptual such as co-word analysis and longitudinal analysis of thematic evolution.

The number of publications is helpful to make basic comparisons, but co-citation analysis allows to know the impact the articles have had on others, showing also what authors, journals, organizations, and countries have high impact in different fields of research.

In addition to the analysis of the relationships between authors and associated levels, conceptual analysis focused on content and thematic will be elaborated using other bibliometric techniques such as co-word, conceptual and strategic maps and word-cloud. The bibliometric analysis undertaken is based on studying the relationships between the publications obtained in the baseline search query (Search RQ1). In addition, two more complementary analyzes will be performed based on the queries with the word Big Data as mandatory in the sentences including those associated with SDGs (Search RQ2), or also added the restriction of the terms associated with accounting (Search RQ3).

4.1. Network techniques based on relationship between authors, countries, universities, papers

4.1.1. Coupling Network

We use the routine biblioNetwork of bibliometrix package (Aria & Cuccurullo, 2017) to calculate the coupling network, where each element of the matrix indicates how many bibliographic couplings exist between the different manuscripts of each search. It has been selected the Salton cosine similarity measure to normalize this matrix, transforming absolute frequencies in relative frequencies and Fruchterman's layout algorithm to visualize the nodes.

Figure S2 in the supplementary material shows the 100 authors most connected to each other, or with the highest degree in the network represented, i.e., those with the most shared references. Authors who already stood out in productivity such as Bhutta, Murray, Gupta and Salomon, also stand in the coupling graph, being co-authors of articles promoted by Global Burden of Disease (GBD).These articles are characterized by a high levels of collaboration, with co-authorships of more than 600 authors in most cases, such as Kassebaum *et al.* (2016) with 624 citations in WoS or Lim *et al.* (2016) with 153 citations in WoS. This leads us to think that in articles with many authors, the coupling network ends up resembling the collaboration network (based on co-authoring relationships) as shows Figure S2.

In order to improve the coupling analysis, eliminating the effect of the high degree of co-authorship, we selected the articles whose number of authors is less than 25 with at least 5 citations, enhancing the representativeness of the relationships into researchers based on coupling or sharing of citations (see Figure S3). With these premises we obtain three clusters with thematic areas well defined, which are: (1) Health (Childhood development, mortality studies related to SDGs etc..), (2) Environment (Energy, Climate Change, Land-use in the SDGs scope) and (3) SDGs in general, their conceptual framework, governance, purposes, measurement possibilities, mainly focusing on any of the topics of the previous clusters. However, there is a higher bias towards environment and sustainability issues.

Given the size of the RQ2 and RQ3 searches, with articles limited to the world of Big Data and/or accounting, the clusters identified are not representative.

4.1.2. Co-citation Network

Two articles are co-citated when both are cited in a third article. Each element from co-citation matrix indicates the frequency with which two documents are cited together in the search (Aria & Cuccurullo, 2017).

Most of the articles cited in Search RQ1 (see Figure 4) correspond to reports/resolutions from international organizations such as the UN or WHO. Many of them are not within the initial search, nor are they in WoS database, since they do not correspond to scientific articles, for example: *"Transform our world 2030, 2015"*, *"Millennium development goals, 2015"*, *"Sustainable development goals , 2015"* (see Table 3 . The blue cluster has mainly UN-Reports related to SDGs, and other relevant articles such as *"Map the interactions between Sustainable Development Goals, Nilsson, 2016"*, which characterizes SDG interactions and *"From millennium development goals to sustainable development goals, Sachs, 2012"*. The green cluster focuses on the entire conceptual framework of the SDGs with a bias in the environment and especially in sustainability, and its main node is *"Sustainable Development Goals for People and the Planet, Griggs, 2013"*, an academic reference about the UN program. Finally, the red cluster is mainly related to health issues. There are no explicit references to Big data and Accounting research in the articles of the co-citation network associated to the Search RQ1.

The co-citation network in Search RQ2 (see Figure 4) determines three clusters, although the themes are highly interspersed, for example the blue cluster has use cases of smart cities but also theory of Big Data and sustainability. The red cluster treats of general themes, not necessarily focused on Big Data. Finally, the green cluster is more theoretical, explaining the possibilities of Big Data in Social Sciences with a bias in international development and use cases of mobile phones. In any case, the sample is too small to establish clear relationships and is not applicable for Search RQ3.

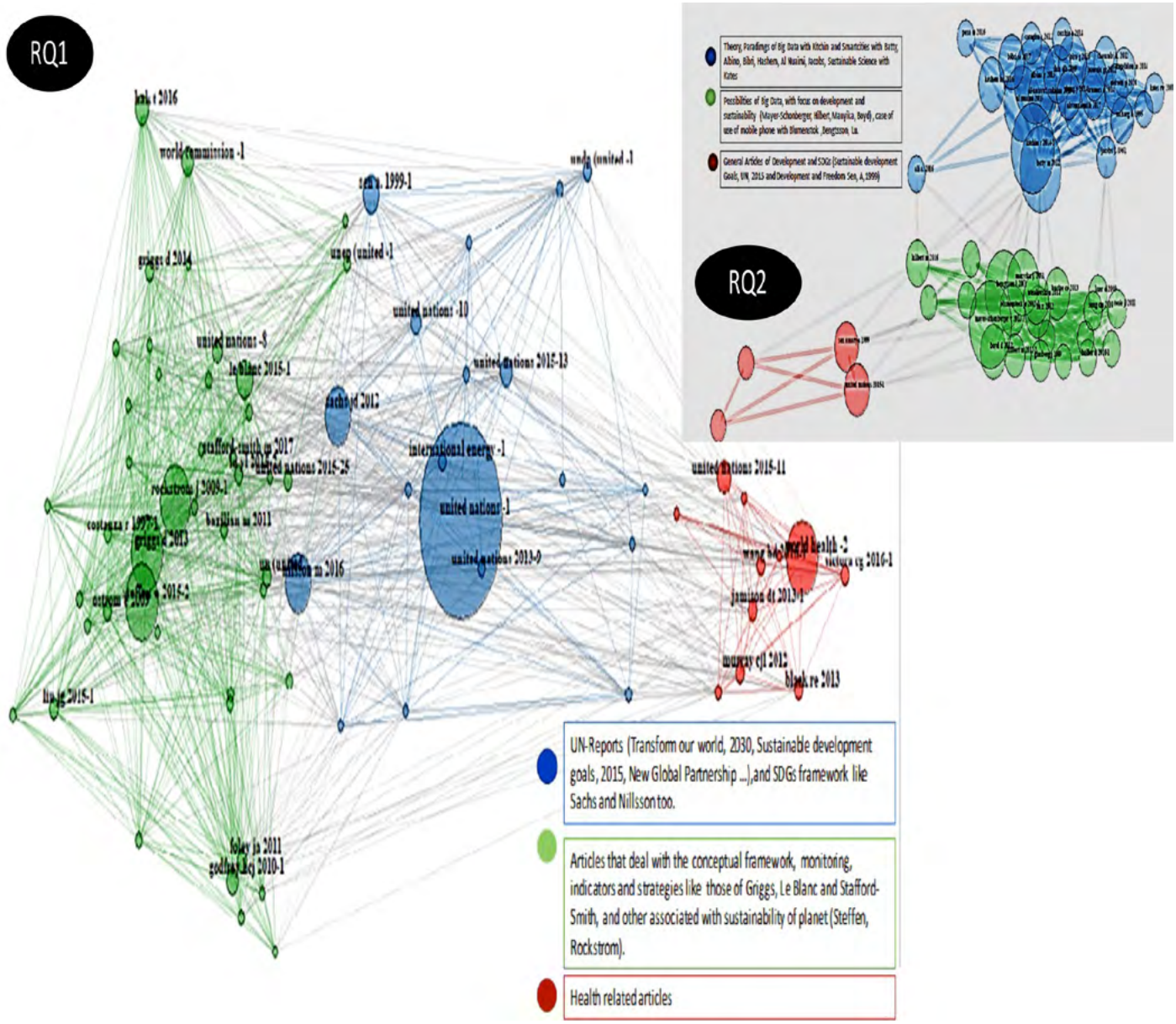

Figure 4
Co-citation Network Articles

Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

Two tables are included in the supplementary material that complement the co-citation network with additional information. Table S2 lists the articles with the greatest number of citations according to WoS, the most referenced articles within the initial search (Search RQ1), and Top 10 of papers with the highest number of local citations. Many of the articles that appear in any of these rankings stand out in the co-citation network, especially those that belong to the ranking of references most cited, regardless of them being local citations.

In Search RQ2 there are only 77 articles, genuinely associated with Big Data (1.8% of the total of Search RQ1). This is because most of the articles in Search RQ1 are related to the SDGs, but do not have the word Big Data in any of the search fields (title, keywords, abstract).Anyway, in Search RQ2, the dispersion of the references is high, and the maximum number of times that some references are cited is very low, no more than 7 (see Table S3). Noteworthy is Bebbington's article, since it is the only one published in an accounting journal that refers to Big Data and international development.

According to the editorial line of the journals, the co-citation graph shows the two most relevant worlds in academic research associated with International Development/Big Data and SDG: Health and Environment (see Figure 5, RQ1). The journals such as Lancet, Plos Med, BMC Public Health, Social Science & Medicine are related to Health. Science, Nature, Global Environment Change and other specialized publications such as Climate Change of Nature are related to Environment. Both themes define two well-differentiated clusters Health and Environment, considering the

number of papers where these sources are referenced together. Publications like Plus One or Nature are multidisciplinary science journal. In any case, the greater number of scientific articles related to the environment is clearly evidenced. In the B and C searches (see Figure 5, RQ2 and RQ3), more specialized journals appear (Cities, Urban studies, Energy, Big data and Society, Information, Communication & Society). No accounting journal appears in search RQ3 and only the journal Ecological Indicators whose aim is to monitor and evaluate ecological and environmental indicators appears in this co-citation graph.

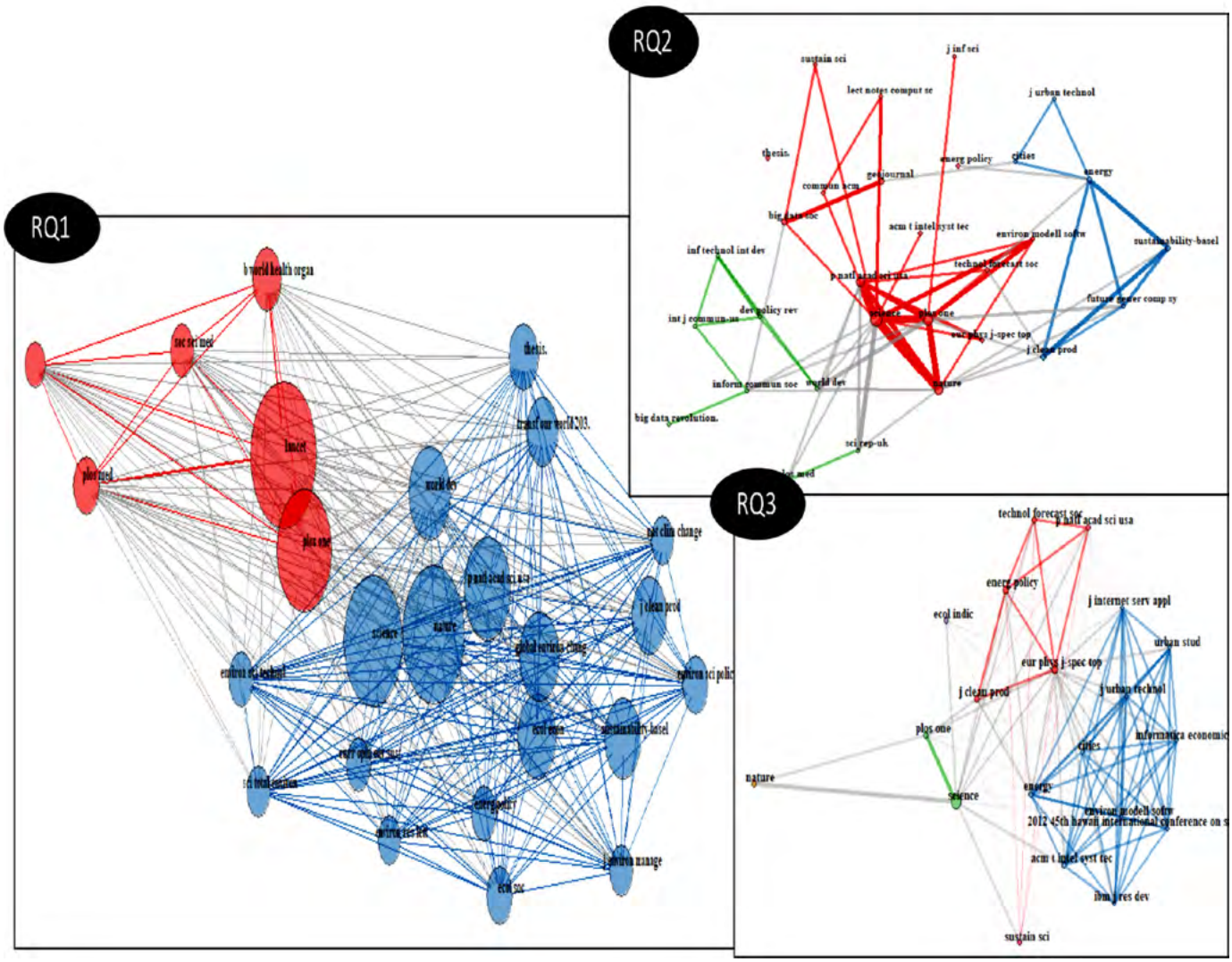

Figure 5
Co-citation Network Journals by search
Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

4.1.3. Collaboration

Scientific collaboration is a network where the nodes are either authors, countries or universities establishing the links between them through co-authorship (Glänzel & Schubert, 2004; Price & Beaver, 1966).

In our research and Search RQ1, we have core countries very strongly related (USA, UK, South Africa, Netherlands, Switzerland, Australia, Canada, Germany, France, Netherlands, Kenya, Italy and Sweden) (see Figure S4 in supplementary material) and as a second tier the rest of the world, very interrelated too with tier one (Japan, Peru, Colombia, Ethiopia, Spain, Finland, China…). In most cases, the core countries are the most productive countries. In any case, it stands out the high number of links in each node. China presents collaboration with other countries on papers related to SDGs, which is logical, since it is a program of the United Nations.

4.1.4. Longitudinal Analysis of citations

Finally, the historiographic map of bibliometrix is a graph (Garfield, 2004) that represents a chronological network map

of most relevant direct citations resulting from a bibliographic collection.

Unfortunately, none of these authors (see Figure 6) have any research associated with the world of international development and big data jointly, focusing on initiatives that are directly or indirectly linked to Sustainable Development Goals.

Researchers such as Griggs, Sachs and Nilsson are key authors in the literature related to "Sustainable development goals". In 2015 Le Blanc also defines another referenced thread in this scope.

The evolution of the direct citation when we select Big Data like mandatory in the sentences of SDGs too (Search RQ2) do not have any thread and it is completely discontinuous, without temporal thread connections, without citation line that demonstrates the evolution, progress and academic development of our selected field. Another fact that shows that we are facing a field which is still to discover from the academic level. There are no researchers whose work contributes in a relevant way to the evolution of the subject. Only Hilbert, 2016 is referenced by Heeks (2018) and Lwin (2018), and Taylor is self-reference with two articles in 2015. In Search RQ3 there are no direct citations. One more, a demonstration that we are dealing with an academic greenfield.

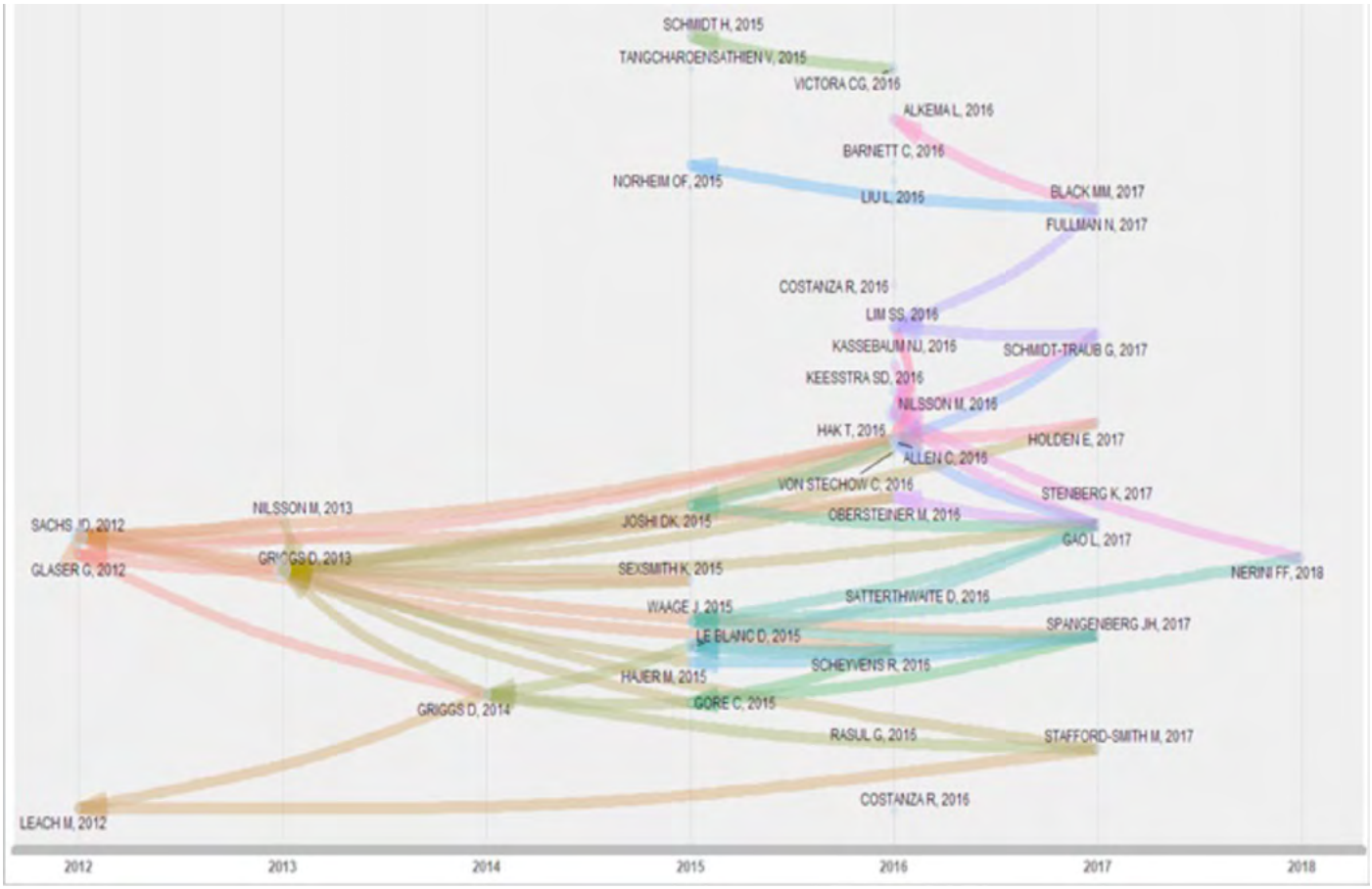

Figure 6
Historical Direct Citation Network in baseline search
Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

4.2. Co-word and Conceptual analysis

4.2.1. Co-word Analysis

We use co-word graphs and bibliometric maps to show in a visual way the associations between the main concepts treated. It also helps to understand the current state of the art of the research about Big Data, Accounting and International Development and predict where future research could be led.

The co-word analysis, based on the sharing of words between articles, defines two big clusters in Search RQ1, with topics similar to the co-citation clusters (health and environment). These clusters are considered like the "big themes" in our Search RQ1

When we run Search RQ2 the referenced words change from the previous search. Health and environment no longer weigh so much, and new relationships are shown associated with governance, policies, management, technologies such as mobile, cloud or internet, as well as mathematical concepts that are applied in machine learning such as decision trees, vector machines etc. The word "city" is a relevant word. In search RQ3 similar terms stand out with respect to search RQ2: framework, cities, governance, big data, mobile phones but no cluster associated with accounting appears except the word indicators, accountability and

control program, but in different clusters. Detailed figures are available in the supplementary material (see Figure S5).

4.2.2. Conceptual Analysis

With the function conceptualStructure from Bibliometrix R package we perform a Correspondence Analysis and the results are plotted on a two-dimensional map.

Taking account of key-words of WoS we have three clusters (see Figure 7) (same clusters that the co-occurrence network): Health and Climate Change, and other specifics related with Sustainable Development Goals. Note relevant issues such as education (Goal 4) , gender equality (Goal 5), industry, innovation and infrastructure (Goal 9), Peace and Justice (Goal 16), and partnerships (Goal 17) do not appear in the graph when keywords are taken into the consideration (see Table S4)

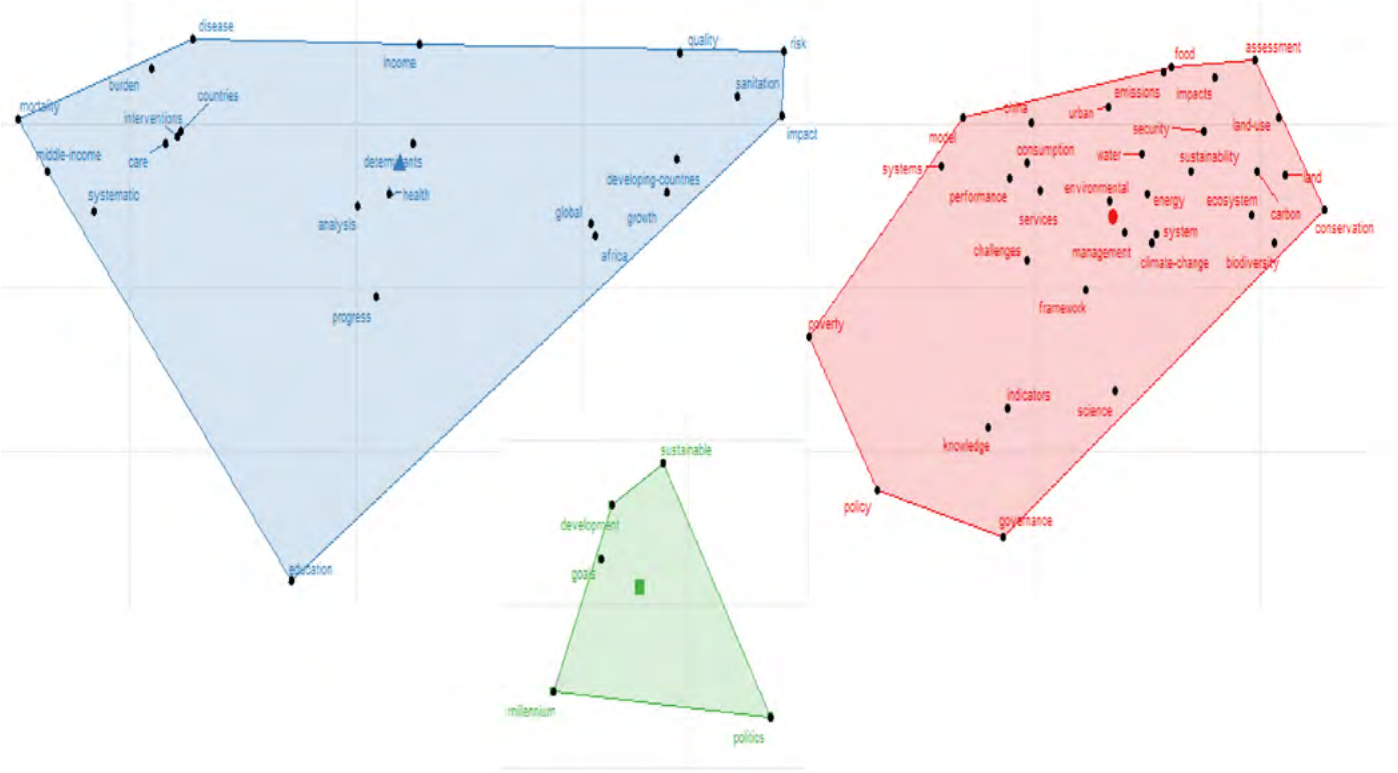

Figure 7
Correspondence Analysis baseline search
Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

4.2.3. Strategic Analysis

A strategic diagram is a two-dimensional space by plotting themes according to their centrality and density (Cobo *et al.*, 2011). We can find four kinds of themes according to the quadrant in which they are placed. Strategic diagrams along with thematic networks (López-Herrera *et al.*, 2016; Muñoz-Leiva *et al.*, 2012), show the thematic focus of the chosen research field.

- Density is defined as the proportion of ties existing in the network to all probable ties. The Callon´s density measure the internal strength of the network.
- Centrality measures the importance of node within a network. The Callon´s centrality is referred to the degree of interaction of a network with other networks. Centrality measures the importance of a theme in the development of the research field.

Our strategic diagram analysis is based in the key-words of WoS and the Search RQ1, with the volume of the spheres proportional to the number of published documents associated with each theme (based on clusters of words). The analysis shows interesting conclusions: 1) Health, Sustainable Development Goals and Water are motor themes; 2) Policy, Adaptation or Framework are emerging or declining themes; 3) Growth, Impact and Death are highly development or isolated themes; 4) Management, Climate-change and Challenges are basic and transversal themes; 5) (see Figure 8). Neither accounting nor big data appear as relevant when we take the total articles sample (search RQ1). Not only Climate Change is related to Environment. There are issues related to environment in the four quadrants: Water, Management, Growth, Framework and Climate-Change.

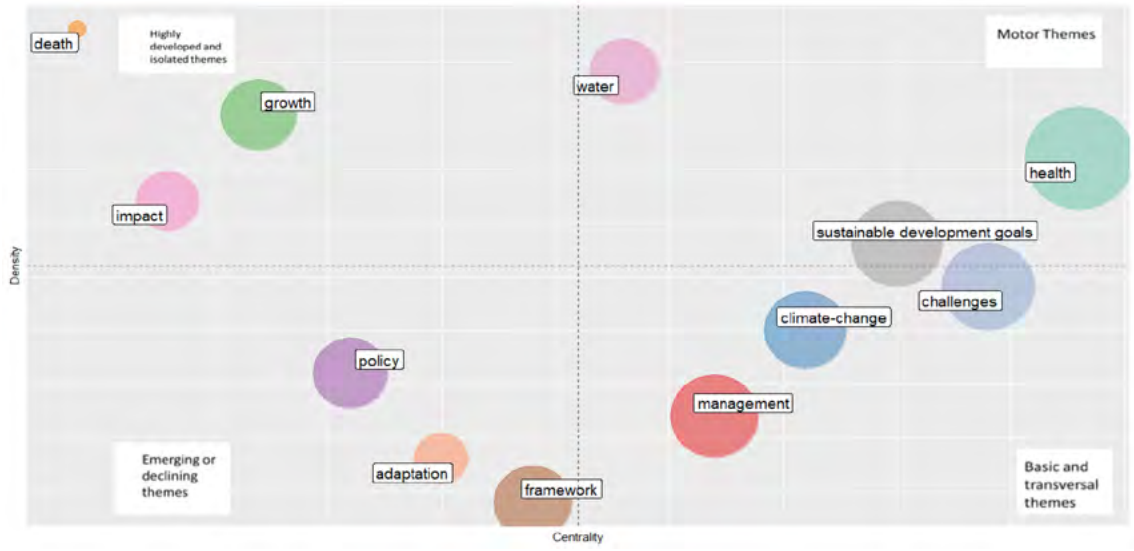

Figure 8
Strategic diagram Baseline search
Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

4.2.4. Thematic Evolution

Thematic evolution identifies the main themes over time. It is based on co-word network analysis and clustering (Cobo *et al.*, 2011, 2014; Moral-Munoz *et al.*, 2018). Nodes represent themes being cluster composed of several keywords. The edges are the relationships between themes.

In each subperiod the keywords are not the same (see Figure 9). New topics with their associated keywords appear and other disappears. For example, in the search RQ1: Climate change appears in all periods. Health, management, governance, sustainable development and conservation are also a mainstream along the analyzed time. General Assembly adopted the 2030 Agenda for Sustainable Development with its 17 goals (SDGs) in September 2015, in 2017 appears "Sustainable development goals" like main theme. Health, care, disease, global burden and mortality are more relevant until 2015, then other concepts such as policy, impacts, challenges, performance, strategy, sustainability and consumption become more relevant. In the case of search RQ2 and RQ3 it is not possible to achieve a topic evolution given the incipient nature of the subject.

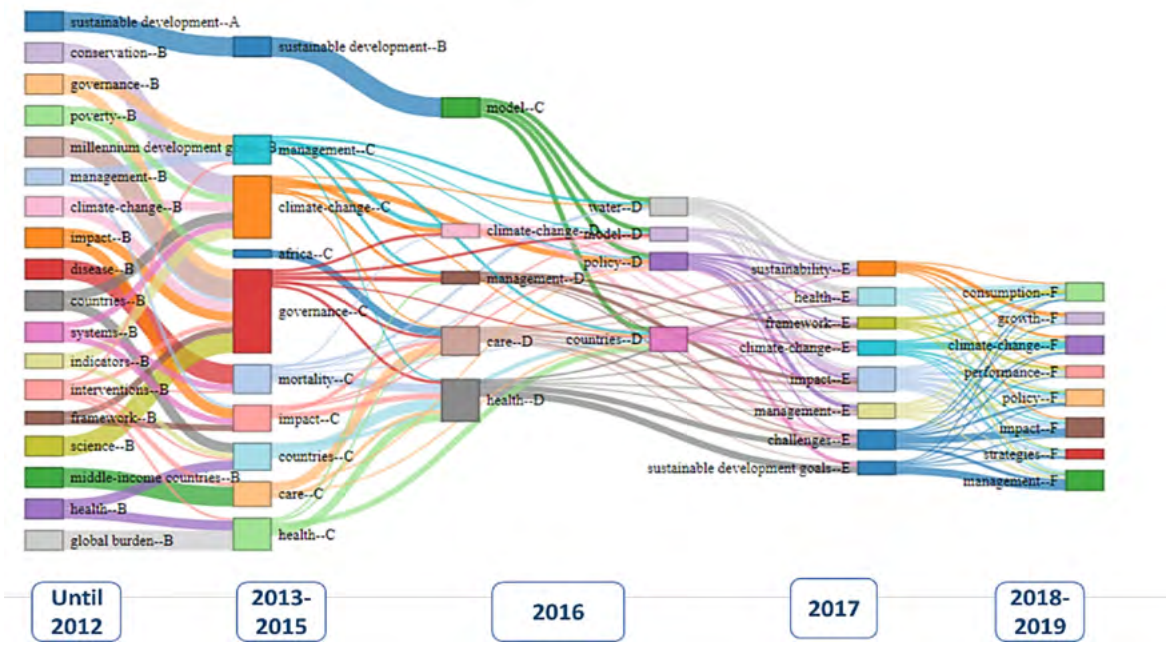

Figure 9
Thematic Evolution baseline search
Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

4.2.5. Word Cloud Analysis

WordCloud is a text mining visual representation method that allows us to highlight the most frequently used keywords in a paragraph (see Figure 10).

We use the inverse frequency method, in order to avoid irrelevant words such as articles, adverbs, etc., applying it to the words of the abstract, unlike the rest of the analyzes are based on WoS keywords. The most relevant words in the Abstract are Health, SDGs, countries, water, energy, food, global, environment, mortality, indicators, sustainability, social, policy, urban, research, study… The word Data appears as relevant when we analyze the words of the abstract, but nevertheless it did not appear in any of the keyword-based analyzes.

Filtering the articles with the mandatory Big Data in the SDGs subquerys, Search RQ2, we found three remarkable points: 1) Smarcities and urban development like a trend, 2) data, research, information, urban, sustainable, health and indicators like important words 3) the word accounting appears, which may make sense, since we know that 21 of the 77 articles are related to our core subject of research. The planning word also appears. In search RQ3, most of words are related with sources of data (mobile, geospatial, CDR), models, indicators, cities and urban development, rural, accounting, planning etc… Biodiversity, Sustainability and Research are words of greater relevance in search RQ3. The reality is that in searches restricted to Big Data and Accounting, the bias on sustainability, environment and development is greater than in search RQ1, where health-related concepts play an important role.

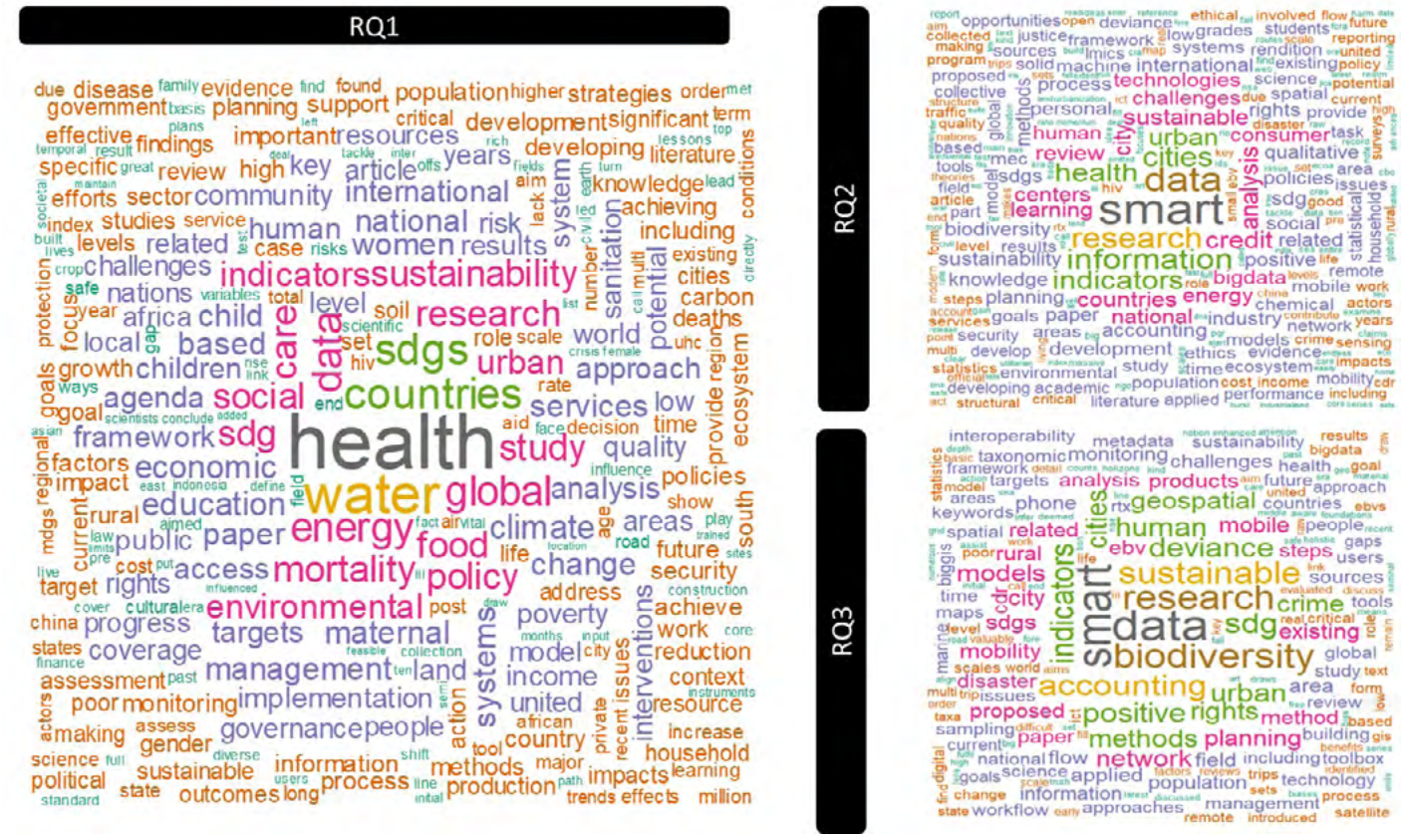

Figure 10
Word Cloud diagrams

Source: Own elaboration based on data retrieved from WoS using Bibliometrix R package.

5. CONCLUSIONS

Big Data is revolutionizing many fields, allowing companies to gain greater understanding and actionable intelligence for their businesses, but also the public sector and the NPOs have realized its value, understanding that data is essential for decision making and its accountability. The penetration of mobile phones use, sensors, satellite data along with new technologies and the democratization of data analytics with open source and open data environment help to adopt these techniques in development initiatives, international organizations and NPOs. In this research, our objective has been to know the degree of development of Big Data academic research applied to accounting in international cooperation, in combination with the academic literature associated with the SDGs.

Our study concludes that this is a "green-field" or emerging field, with little associated academic literature, with only 21 articles in WoS when screened simultaneously for the terms Big Data and accounting in addition to international development and SDGs concepts. In the case of the SDGs, there is a significant

volume of articles that tackle in an overall way, without restrictions of any kind, but there are very few papers that include academic research where Big Data is directly applied, despite being a suitable field for its application due to its own nature of objectives and targets. Furthermore, this field is currently led by international organizations such as UN, WHO and World Bank that are attempting to incorporate Big Data as key tool, promoting initiatives such as the UN Global Pulse, a platform to use of new data streams and stimulate collaborative research and supported by associations of mobile telecom operators (GSMA).

To understand the starting point of the selected research field, three complementary publication searches have been defined, with different levels of restrictions in their queries, in order to have a 360-degree vision that respond to the research questions. The research development after 2015 is greatly influenced by the launch of the UN 2030 SDGs program and the predominance of Anglo-Saxon countries.

Regarding the literature associated with SDGs along with international development and Big Data, it can be concluded that the most referenced articles are reports from international organizations that do not appear in academic databases. Restricting more the initial search, with the terms Big Data mandatory also in the sentences related to the SDGs (77 articles), it stands out the diversity in the cited references and the low index of local references, which denotes that we are facing an emerging research field, without impact references, without contributors of great relevance, many fronts and open topics, and without a defined research horizon. If we further consider the concepts related to accounting as compulsory (21 articles), we hardly obtain any interrelations between the references cited, which is a consequence of the lack of development of the field of research in the academic sphere. Examining the most restricted searches, only 4 articles have more than 10 citations in WoS and deal with issues associated with governance, politics, strategy and indicators performance in the world of Big Data, and international cooperation or SDGs (articles related to specific use cases are excluded) which are Hilbert (2016), Taylor (2015a) and Taylor (2015b). Bebbington (2018) is the only article published in an accounting journal.

The bibliometric analysis has been completed both from a structural viewpoint, with methods as co-citation, coupling, collaboration or evolution direct citation to determine the most prominent relationships and clusters between authors, articles, countries or journals, and from a conceptual perspective, identifying which topics are the most significant in our searches, distilling the co-word and strategic map as relevant methods.

Co-citation and coupling analysis reveal relationships and prominent groups. The clusters of articles or authors obtained in both analyses, although they present similarities, are not the identical. There is coincidence in the number of clusters, but they differ in some issues, especially because in the analysis of co-citation many relevant cited references are international reports, which, not being of an academic nature, do not form part of the search itself, and therefore cannot appear in the bibliographic coupling. In both analyses, the health topic together with the governance and conceptual framework of the SDGs, where Griggs is the most relevant author, draw up two very similar clusters. The third cluster differs in co-citation of articles and bibliometric coupling of authors, since in the case of the co-citation it is basically formed by the SDGs international reports; however, in the coupling analysis it is oriented to environment, energy, water studies.

In any case, we highlight that there is a clearer silo in health issues than in the general vision of the SDGs (conceptual framework, governance and politics), more intermingled with the topics of sustainability and environment both in authorship and in thematic.

Considering the journal referenced in the articles of the baseline search, the co-citation method distinguishes two clusters, one formed by journals focused on Environment and Sustainability, highlighting Nature, Science and World Development and another determined by journals whose theme is Health and Medicine with Lancet and Plos One as representative actors. These clusters are mainly determined by the volume of articles associated with the SDGs that are in that search, which are the majority. For the most restrictive searches, with Big Data or Big Data and Accounting in all sub-sentences, besides the journals oriented to the environment and sustainability, others whose editorial line is associated with new technologies (Big data and Society , Information, Communication and Society, Software and Environmental Modelling, Information technologies and International development) or specialized in specific topics (Cities, Urban Studies, Energy, Cleaner Production Journal) also stand out. No accounting journal turns up in the relevant clusters, and "Ecology indicators" journal is the only one with the greatest focus on KPIs monitoring.

The direct citation history shows us the great contributors to the academic development of the research associated with the SDG (governance, conceptual framework, target monitoring, interlinkages, implementation information tracking models) that are Griggs, Nilsson, Sachs, Le Blanc and Allen. Nevertheless, we do not have a starting point in the knowledge and development of our research line when we have the restricted searches with Big Data and Accounting as mandatory words without relevant direct citations. Only Hilbert, 2016 receives two citations in 2018.

In the conceptual analysis we apply different methods such as co-word, CA/MCA to define conceptual structure, strategic map and word-cloud obtaining similar insights. The main topics are Environment/Climate Change and Health. The first theme is more closely related to sustainability, performance, policy, indicators, and governance concepts of the SDGs, core elements of our research, as well as those issues related to the Climate Change (energy, emissions, etc…). However, in the analysis of the most restrictive searches appear other concepts and clusters, some more technical or mathematical own Big data technology (support vector, neural network, cloud data-center, algorithms, systems), source data (internet, twitter, mobile phones) and others that have to do with specific development problems but above all with its governance (framework, indicators, politics, sustainability). The urban topic and smart-cities seem to be a relevant topic when we focus on Big Data and Accounting. In the most restrictive search, with accounting mandatory, the most frequent terms are those related to the final objective of our line of research: Big Data, governance, cities, framework, information or poverty, although clusters are determined by a small number of articles, which makes it difficult to achieve significant conclusions.

Finally, the wordcloud analysis, based on abstract´s words, presents the following insights: Health and Water the two most prominent words. Regarding to the most restrictive searches, technological and methodological terms appear, words related to

sources of information, and with a prominent place for smart cities, urban development, data, accounting, biodiversity and research.

The main contribution of our research is not only the complete bibliometric analysis in R, applying different techniques that allow a 360 degree vision to assess the state of the art of academic development in the field of big data, accounting and international development with relevant insights; but also, the authors of this paper want to emphasize the time for developing this field from an academic perspective and the challenge and need to enhance this area of research, which will help to improve the lives of the most disadvantaged countries after the gap in the chosen field of study has been shown. The momentum that UN 2030 SDGs are having in both the developed and developing world should be a catalyst to stimulate the deployment of the use of Big Data techniques associated with accounting and economic planning in international cooperation, and a fundamental conclusion of this paper, however, is that the trend is not yet as expected despite the extensive generalist literature on this global program. Just as international organizations are promoting in an important way the development of this field, where the usefulness of Big Data can change the way of measuring and controlling development projects, this must be reflected in a clearer manner in the field of scholarly research.

Our future research lines will analyze NPOs and its Big Data to improve its performance with the aim of contributing to the academic development of this field of research, comparing in a more empirical way the decoupling between the real use of these new tools and the development of academic literature on the same. Real use cases can be also studied to understand which is the main problem that these organizations have for the faster development of Big Data (data ownership in the hands of a few, little development of open data in these countries, digitization problems, etc...). In addition, these real applications can be a starting place for more in-depth academic studies, and can help to develop this field of research, which is currently less attractive in the academic community than others related to Big Data. We expect accounting, Big data, NPOs, companies, governments, and researchers to be part of a great community that allows the academic community to strengthen and enhance this field of research with international organizations. In parallel, a ground-breaking may be taken and will make a qualitative thematic review or a mixed-methods review of the existing literature on Big Data and international development to identify the key features and complexities of this nexus, its interconnections with accounting and related lines of inquiry.

6. SUPPLEMENTARY FILES

A supplementary file with several figures and tables quoted throughout this paper can be accessed at the following URL: http://www.ehu.eus/cuadernosdegestion/documentos/Supplementary-File-21A1513.pdf

7. BIBLIOGRAPHY

Adriaanse, L. S., & Rensleigh, C. (2011). Comparing Web of Science, Scopus and Google Scholar from an Environmental Sciences perspective. *South African Journal of Libraries and Information Science*, *77*(2), 169-178. https://doi.org/10.7553/77-2-58

Adriaanse, L. S., & Rensleigh, C. (2013). Web of science, scopus and google scholar a content comprehensiveness comparison. *Electronic Library*, *31*(6), 727-744. https://doi.org/10.1108/EL-12-2011-0174

Alcaide, G. G., & Ferri, J. G. (2014). La colaboración científica: Principales líneas de investigación y retos de futuro. *Revista Espanola de Documentacion Cientifica*, *37*(4), e062. https://doi.org/10.3989/redc.2014.4.1186

Amankwah-Amoah. (2016). Emerging economies, emerging challenges: Mobilising and capturing value from big data. *Technological Forecasting and Social Change*, *110*, 167-174. https://doi.org/10.1016/j.techfore.2015.10.022

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Arnaboldi, M., Busco, C., & Cuganesan, S. (2017). Accounting, accountability, social media and big data: revolution or hype? *Accounting, Auditing and Accountability Journal*, *30*(4), 762-776. https://doi.org/10.1108/AAAJ-03-2017-2880

Baldwin, C. Y., Tribendis, J. J., & Clark, J. P. (1984). The Evolution of Market Risk in the U.S. Steel Industry and Implications for Required Rates of Return. *The Journal of Industrial Economics*, *33*(1), 73. https://doi.org/10.2307/2098425

Bergmann, T., Dale, R., Sattari, N., Heit, E., & Bhat, H. S. (2017). The Interdisciplinarity of Collaborations in Cognitive Science. *Cognitive Science*, 41(5), 1412-1418. https://doi.org/10.1111/cogs.12352

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*, *19*(1–2), 5-16. https://doi.org/10.3233/IP-140328

Bhimani, A. (2020). Digital data and management accounting: why we need to rethink research methods. *Journal of Management Control*, *31*(1-2), 9-23. https://doi.org/10.1007/s00187-020-00295-z

Bichteler, J., & Eaton, E. A. (1980). The combined use of bibliographic coupling and cocitation for document retrieval. *Journal of the American Society for Information Science*, *31*(4), 278-282. https://doi.org/10.1002/asi.4630310408

Boeker, M., Vach, W., & Motschall, E. (2013). Google Scholar as replacement for systematic literature searches: Good relative recall and precision are not enough. *BMC Medical Research Methodology*, *13*(1). https://doi.org/10.1186/1471-2288-13-131

Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, *79*(1), 27-44. https://doi.org/10.1007/s11192-009-0402-6

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389-2404. https://doi.org/10.1002/asi.21419

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology*, *64*(9), 1759-1767. https://doi.org/10.1002/asi.22896

Bufrem, L., & Prates, Y. (2005). O saber científico registrado e as práticas de mensuração da informação. *Ciência Da Informação*, *34*(2), 9-25. https://doi.org/10.1590/s0100-19652005000200002

Calero Medina, C. M., & Van Leeuwen, T. N. (2012). Seed journal citation network maps: A method based on network theory. *Journal of the American Society for Information Science and Technology, 63*(6), 1226-1234. https://doi.org/10.1002/asi.22631

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155-205. https://doi.org/10.1007/BF02019280

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191-235. https://doi.org/10.1177/053901883022002003

Cobo, M. J., Chiclana, F., Collop, A., De Ona, J., & Herrera-Viedma, E. (2014). A bibliometric analysis of the intelligent transportation systems research based on science mapping. *IEEE Transactions on Intelligent Transportation Systems*, *15*(2), 901-908. https://doi.org/10.1109/TITS.2013.2284756

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, *5*(1), 146-166. https://doi.org/10.1016/j.joi.2010.10.002

Dagilienė, L., & Klovienė, L. (2019). Motivation to use big data and big data analytics in external auditing. *Managerial Auditing Journal*, *34*(7), 750-782. https://doi.org/10.1108/MAJ-01-2018-1773

Davies, T., Walker, S. B., Rubinstein, M., & Perini, F. (2019). *Estado de los datos abiertos: historias y horizontes | Universo Abierto*. https://universoabierto.org/2019/11/14/estado-de-los-datos-abiertos-historias-y-horizontes/

De Solla Price, D. J. (1965). Networks of scientific papers. *Science*, *149*(3683), 510. https://doi.org/10.1126/science.149.3683.510

Dekker, R., Engbersen, G., Klaver, J., & Vonk, H. (2018). Smart Refugees: How Syrian Asylum Migrants Use Social Media Information in Migration Decision-Making. *Social Media and Society*, *4*(1). https://doi.org/10.1177/2056305118764439

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). The museum of errors/horrors in Scopus. *Journal of Informetrics*, *10*(1), 174-182. https://doi.org/10.1016/j.joi.2015.11.006

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359-375. https://doi.org/10.1007/BF02019306

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. In *Journal of Information Science*, *30*(2), 119-145. https://doi.org/10.1177/0165551504042802

Gillespie, M., Ampofo, L., Cheesman, M., Faith, B., Iliadou, E., Issa, A., Osseiran, S., & Skleparis, D. (2016). Something About Refugee Media. In *The Open University/France Medias Monde* (Issue May). https://www.open.ac.uk/ccig/sites/www.open.ac.uk.ccig/files/Mapping Refugee Media Journeys 16 May FIN MG_0.pdf

Glänzel, W., & Schubert, A. (2004). Analyzing Scientific Collaboration through Co-Authorship. In *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies on S&T systems* (pp. 257-276).

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big Data Opportunities for Global Infectious Disease Surveillance. *PLoS Medicine*, *10*(4). https://doi.org/10.1371/journal.pmed.1001413

Helft, M. (2008). Google Uses Searches to Track Flu's Spread. *The New York Times*, 1-5. http://www.nytimes.com/2008/11/12/technology/internet/12flu.html?_r=1&th&emc=th&oref=slogin

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, *34*(1), 135-174. https://doi.org/10.1111/dpr.12142

Janvrin, D. J., & Weidenmier Watson, M. (2017). "Big Data": A new twist to accounting. *Journal of Accounting Education*, *38*, 3-8. https://doi.org/10.1016/j.jaccedu.2016.12.009

Jarneving, B. (2007). Complete graphs and bibliographic coupling: A test of the applicability of bibliographic coupling for the identification of cognitive cores on the field level. *Journal of Informetrics*, *1*(4), 338-356. https://doi.org/10.1016/j.joi.2007.08.001

Kassebaum, N. J., Arora, M., Barber, R. M., Brown, J., Carter, A., Casey, D. C., Charlson, F. J., Coates, M. M., Coggeshall, M., Cornaby, L., Dandona, L., Dicker, D. J., Erskine, H. E., Ferrari, A. J., Fitzmaurice, C., Foreman, K., Forouzanfar, M. H., Fullman, N., Goldberg, E. M.,… Zuhlke, L. J. (2016). Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1603-1658. https://doi.org/10.1016/S0140-6736(16)31460-X

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10-25. https://doi.org/10.1002/asi.5090140103

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85. https://doi.org/10.1145/2500873

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*, *74*(2), 273-294. https://doi.org/10.1007/s11192-008-0217-x

Kuusi, O., & Meyer, M. (2007). Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm. *Scientometrics*, *70*(3), 759-777. https://doi.org/10.1007/s11192-007-0311-5

Landefeld, S. (2014). *Uses of Big Data for Official Statistics: Privacy, Incentives, Statistical Challenges, and Other Issues Discussion Paper United Nations Global Working Group on Big Data for Official Statistics*. 1-20. https://unstats.un.org/unsd/trade/events/2014/beijing/Steve Landefeld - Uses of Big Data for official statistics.pdf

Leydesdorff, L., De Moya-Anegón, F., & Guerrero-Bote, V. P. (2010). Journal maps on the basis of scopus data: A comparison with the journal citation reports of the ISI. *Journal of the American Society for Information Science and Technology*, *61*(2), 352-369. https://doi.org/10.1002/asi.21250

Lim, S. S., Allen, K., Dandona, L., Forouzanfar, M. H., Fullman, N., Goldberg, E. M., Hay, S. I., Holmberg, M., Kutz, M. J., Larson, H. J., Lopez, A. D., McNellan, C. R., Mokdad, A. H., Mooney, M. D., Naghavi, M., Olsen, H. E., Pigott, D. M., Vos, T., Wang, H.,… Zonies, D. (2016). Measuring the health-related Sustainable Development Goals in 188 countries: a baseline analysis from the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1813-1850. https://doi.org/10.1016/S0140-6736(16)31467-2

Linders, D. (2013). Towards open development: Leveraging open data to improve the planning and coordination of international aid. *Government Information Quarterly*, *30*(4), 426-434. https://doi.org/10.1016/j.giq.2013.04.001

López-Cózar, E. D., Robinson-García, N., & Torres-Salinas, D. (2014). The google scholar experiment: How to index false papers and manipulate bibliometric indicators. *Journal of the Association for Information Science and Technology*, *65*(3), 446-454. https://doi.org/10.1002/asi.23056

López-Herrera, A. G., Cobo, M. J., Herrera-Viedma, E., & Herrera, F. (2016). A bibliometric study about the research based on hybridating the fuzzy logic field and the other computational intelligent techniques: A visual approach. *International Journal of Hybrid Intelligent Systems*, *7*(1), 17-32. https://doi.org/10.3233/his-2010-0102

López-Herrera, A. G., Cobo, M. J., Herrera-Viedma, E., Herrera, F., Bailón-Moreno, R., & Jiménez-Contreras, E. (2009). Visualization and evolution of the scientific structure of fuzzy sets research in Spain. *Information Research*, *14*(4).

Marshakova, I. V. (1981). Citation networks in information science. *Scientometrics*, *3*(1), 13-25. https://doi.org/10.1007/BF02021861

McKinney, E., Yoos, C. J., & Snead, K. (2017). The need for 'skeptical' accountants in the era of Big Data. *Journal of Accounting Education*, *38*, 63-80. https://doi.org/10.1016/j.jaccedu.2016.12.007

Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*(3), 363-377. https://doi.org/10.1007/BF02129600

Meyer, M., Grant, K., Morlacchi, P., & Weckowska, D. (2014). Triple Helix indicators as an emergent area of enquiry: A bibliometric perspective. *Scientometrics*, *99*(1), 151-174. https://doi.org/10.1007/s11192-013-1103-8

Moral-Munoz, J. A., Arroyo-Morales, M., Herrera-Viedma, E., & Cobo, M. J. (2018). An Overview of Thematic Evolution of Physical Therapy Research Area From 1951 to 2013. *Frontiers in Research Metrics and Analytics*, *3*. https://doi.org/10.3389/frma.2018.00013

Muñoz-Leiva, F., Sánchez-Fernández, J., Liébana-Cabanillas, F. J., & López-Herrera, A. G. (2012). Applying an automatic approach for showing up the hidden themes in financial marketing research (1961-2010). *Expert Systems with Applications*, *39*(12), 11055-11065. https://doi.org/10.1016/j.eswa.2012.03.017

Muñoz-Leiva, F., Viedma-del-Jesús, M. I., Sánchez-Fernández, J., & López-Herrera, A. G. (2012). An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality and Quantity*, *46*(4), 1077-1095. https://doi.org/10.1007/s11135-011-9565-3

Nalimov, V. V., & Mul´chenko, Z. M. (1971). Measurement of science: Study of the development of science as an information process. Washington , DC: Foreign Technology Division. In *Naukometriya, Izucheniye Razvitiya Nauki kak Informatsionnogo Protsessa* (p. 196).

Oliver, N., Matic, A., & Frias-Martinez, E. (2015). Mobile Network Data for Public Health: Opportunities and Challenges. *Frontiers in Public Health*, *3*. https://doi.org/10.3389/fpubh.2015.00189

Óskarsdóttir, M., Sarraute, C., Bravo, C., Baesens, B., & Vanthienen, J. (2018). Credit scoring for good: Enhancing financial inclusion with smartphone-based microlending. *International Conference on Information Systems 2018, ICIS 2018*.

Price, D. J., & Beaver, D. D. (1966). Collaboration in an invisible college. *The American Psychologist*, *21*(11), 1011-1018. https://doi.org/10.1037/h0024051

PRITCHARD, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*, 348.

Protopop, I. (2016). Big Data and Smallholder Farmers: Big Data Applications in the Agri-Food Supply Chain in Developing Countries. *International Food and Agribusiness Management Review*, *19*, 173-190. https://doi.org/10.22004/ag.econ.240705

Roy, M., Moreau, N., Rousseau, C., Mercier, A., Wilson, A., & Atlani-Duault, L. (2020). Ebola and Localized Blame on Social Media: Analysis of Twitter and Facebook Conversations During the 2014-2015 Ebola Epidemic. *Culture, Medicine and Psychiatry*, *44*(1), 56-79. https://doi.org/10.1007/s11013-019-09635-8

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, *50*(9), 799-813. https://doi.org/10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Thara, D. K., Premasudha, B. G., Ram, V. R., & Suma, R. (2016). Impact of big data in healthcare: A survey. *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016*, 729-735. https://doi.org/10.1109/IC3I.2016.7918057

The State of Mobile Data for Social Good Report (Issue June) (2017). https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/06/Mobile-Data-for-Social-Good-Report_29June.pdf

UN DESA. (2015). Inequality and the 2030 Agenda for Sustainable Development. In *Development Issues* (Vol. 4). https://www.un.org/development/desa/dpad/publication/no-4-inequality-and-the-2030-agenda-for-sustainable-development/

UN Global Pulse, & UNHCR. (2017). *Rescue Patterns in the Mediterranean Partners :* https://doi.org/Project Series, no. 29, 2017

Williams, B. C., & Plouffe, C. R. (2007). Assessing the evolution of sales knowledge: A 20-year content analysis. *Industrial Marketing Management*, *36*(4), 408-419. https://doi.org/10.1016/j.indmarman.2005.11.003