Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms

Giacomo di Tollo [a], Stoyan Tanev [b,*], Giacomo Liotta [c], Davide De March [d,e]

[a] Dipartimento di Management, Università Ca' Foscari, San Giobbe, Cannaregio 873, 30121 Venezia, Italy
[b] Department of Technology and Innovation, Faculty of Engineering, University of Southern Denmark, Campusvej 55, Odense M, DK-5230, Denmark
[c] Center for Industrial Production, Department of Business and Management, Faculty of Engineering and Science, Aalborg University, A.C. Meyers Vaenge 15, DK-2450 Copenhagen SV, Denmark
[d] EvoSolutions S.r.l., Viale Ancona 17, 30174 Venice, Italy
[e] European Center for Living Technology, San Marco 2940, 30124 Venice, Italy

A B S T R A C T

In this paper we introduce a method that combines principal component analysis, correlation analysis, K-means clustering and self organizing maps for the quantitative semantic analysis of textual data focusing on the relationship between firms' co-creation activities, the perception of their innovation and the articulation of the attributes of their product-enabled services. Principal component analysis was used to identify the components of firms' value co-creation activities and service value attributes; correlation analysis was used to examine the relationship between the degree of involvement in specific co-creation activities, the online articulation of firms' service value attributes and the perception of their innovativeness. K-means and self organizing map (SOM) are used to cluster firms with regards to their involvement in co-creation and new service development, and, additionally, as complementary tools for studying the relationship between co-creation and new service development.

The results show that, *first*, there is a statistically significant relationship between firms' degree of involvement in co-creation activities and the degree of articulation of their service value attributes; *second*, the relationship should be considered within the context of firms' innovation activities; *third*, OS Software-driven firms are the best example in terms of co-creation and new product-enabled service development, i.e. the collaborative principles built in their customer participation platforms should be adopted by other (non-software) firms interested in enhancing their innovation capacity through involvement in co-creation and new product-enabled service development.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The recent years have witnessed an unprecedented interest in the conceptualization and development of research techniques and business intelligence tools using online textual data to generate insights about the business, marketing or innovation aspects of specific firms' or industry sectors. In some cases the focus is on data related to online communities (blogs, forums or review sites) and social networks; in other cases the focus is on company's online information which is expressed in natural language in order to be publicly disclosed to customers and other relevant stakeholders. Such information can be used to both analyze and inform specific aspects of firms' policy and strategic decision making with respect to customer relationship management, marketing communications, or open innovation management initiatives. In this article we aim to:

- demonstrate the value of public online textual data as a source of actionable information about specific product development, business and innovation practices in different industry sectors;
- validate a simple web search technique for textual data collection which is combined with principal component (PCA) to provide a systematic interpretive methodology that could be used in conceptualizing the relationship between different business and innovation practices;

* Corresponding author.
  E-mail addresses: giacomodt@gmail.com (G. di Tollo), tan@iti.sdu.dk (S. Tanev), gl@business.aau.dk (G. Liotta), demarch@evosolutions.it (D. De March).

- apply a combination of cluster analysis (CA) and an artificial neural network (ANN) approach using self-organizing maps (SOMs) as complementary methods enabling a more generic classification of firms in terms of the degree of their involvement in the specific business and innovation practices.

The analytical approach that was adopted in this article could be categorized as a modified and more intuitive version of latent semantic analysis (LSA) – a method for extracting the meaning from passages of text, based on a series of linear algebraic operations and statistical computations over a collection of documents or other textual sources such as e-mail messages or web pages [1–4]. The adoption of the LSA technique has visibly increased over the last 20 years. It is however a relatively complex approach with high barriers to adoption by the majority of scholars and practitioners. The analytical method suggested here aims at validating a more intuitive version of quantitative content analysis that could become more easily accessible through the use of open source software web search and computational tools as well as standard business statistics software packages such as SPSS, SAS, Microsoft Excel, etc. Unfortunately, user-friendly intelligent systems for business intelligence purposes are not widely available yet, thus entailing labor intensive and time consuming tasks for business analysts [5]. Our article addresses this issue by using a keyword-based web search on the online text of companies' websites to generate the number of occurrences of the keywords and a combination of well-known techniques such as PCA, correlation analysis, K-means CA and ANN to process the collected data and interpret it in a meaningful way. The set of keywords is preliminary constructed on the basis of subject domain expertise in a way that it could cover in a relatively complete way the various issues, aspects or activities associated with a particular business or innovation aspect. In this sense, there is a key difference between LSA and our approach to the selection of the keywords that are going to be used in the search process. In LSA, the initial set of keywords is usually coming from the documents themselves through a pre-processing step focusing on the identification of the terms that would be most relevant in the research process. In other words, LSA is looking to discover or uncover what is already in the documents without any preliminary focus on a specific theme or topic. In our approach we start with a specific theme of interest such as, for example, firms' involvement in co-creation or innovation practices, and then construct a set of keywords that is used as a probing tool or an exploratory lens to look into how the firms included in the research sample articulate their association with the particular theme or topic. The main sources of the initial set of keywords are research and practitioner publications focusing on the topic of interest, complemented by a parallel alignment of the terminology with selected websites of firms included in the research sample. One could say therefore that in our approach the set of keywords is external to the online text and not necessarily coming out of it. Such approach allows for some more flexibility in the construction of the logical combination of keywords including different words with similar or overlapping meanings. This is another important difference since it shows that the exploratory power of our approach is based on the business domain subject expertise and not so much on the expertise required for the development and the operation of the business intelligence tool.

A approach similar to the one used in this article was previously applied to the identification of the technology commercialization strategies of high-tech small firms [6] and the value co-creation practices of firms using open source (OS) technologies as part of their business model, showing that it allows for the classification of the firms in terms of the relationship between their co-creation and innovation activities [7,8]. Recently it was also applied to the identification of the product-enabled value attributes of top

Research and Development (R&D) spenders in Canada and Europe [9,10]. The present article represents a first attempt to apply the methodology to the simultaneous analysis of three different aspects – the degree of firms' involvement in value co-creation practices, the articulation of their product-enabled service value attributes and their own perception of the innovativeness of their products, processes and services. While the various innovation aspects of new and established firms have been traditionally the subject of significant interest, value co-creation practices and product services are in the process of attracting particular attention.

Value co-creation has emerged as a new business, marketing and innovation approach in which customer and end users are seen as active part in the design and shaping of personalized products, services and experiences [7,11–16]. It is based on the design and development of customer participation platforms, which provide firms with the technological and human resources, tools and mechanisms to benefit from the engagement experiences of individuals and communities [17–19]. The ability of such platforms to enable the personalization of new products and services challenges the operational presuppositions of traditional marketing segmentation techniques by promoting a new service-dominant logic [20] which allows firms to address broader heterogeneous markets aiming at a better fit between what customers need and what the firm makes and offers.

Product-enabled services or product-service systems (PSS) are considered as a way for firms' to be more innovative by offering more value to customers as well as to attain business differentiation and sustainability [21–24]. More specifically, PSS are usually associated with a specific type of value proposition consisting of a mix of tangible products and intangible services designed and combined in a way that they jointly are capable of fulfilling specific customer needs [23]. In this article the term product-enabled service will be used as an inclusive way of addressing the different ways of adding value by combining products and services within a specific business context. The application of the suggested methodology within the context of three different aspects demonstrates its ability to handle a variety of aspects related to business and innovation management.

The rest of the article is structured as follows. We outline our method is Section 2 and review the relevant literature on value co-creation practices and product-enabled services in Section 3. Results will be shown in Section 4 and discussed in Section 5. Finally, in Section 6 we summarize our concluding remarks and make suggestions for future research.

2. Methodology

The objective of this article is to examine the relationship between the degrees of firms' involvement in value co-creation practices, the articulation of the value attributes of their product-related services and the perception of their innovativeness. The study adopts a quantitative methodology which is based on: (i) online textual data that could be found on companies' websites; (ii) web search techniques for textual data collection combined with PCA to provide the key components of their value co-creation practices and the articulation of their service value attributes; (iii) cluster and ANN analysis to provide a generic classification of the firms in terms of the degree of their involvement in value co-creation and the degree of articulation of their service value attributes.

2.1. Research sample

The results are based on a sample of 230 firms that were selected for being representative of the breadth of their value co-creation activities. This is a subset of the research sample that

Table 1

Breakdown of the firms in the research sample: ECL – non open source firms members of the Eclipse Foundation, OSS – open source software firms, GEN – firms that are not open source and non-members of the Eclipse Foundation.

Type of firm	Frequency	Percent
ECL	94	40.9
GEN	58	25.2
OSS	78	33.9
Total	230	100.0

was used previously by Tanev et al. [7]. It is however subjected to a completely new data collection which enables the validation of the methodology by looking into the temporal evolution of specific research insights. 25.2% of the firms in the sample can be found as examples in the reviewed value co-creation literature. The rest 74.8% of the firms are engaged (i.e. having revenue from involvement) in OS Software (OSS) projects. The majority of them were selected between the members of the Eclipse OS Foundation[1] (Table 1). The rest of the OSS firms were selected by using two sources of information about OS firms – the Open Source Experts[2] and the Canadian Companies Capabilities Directory of OS Companies database maintained by Industry Canada.[3]

The focus on firms involved in OSS projects was driven by the realization that value co-creation platforms can be considered as a type of participation architectures similar to firm-driven open source platforms [25]. One of the key underlying assumptions is that, as part of their customer interactions and marketing communications, firms do openly articulate the specifics of their co-creation activities, the add-on value of their product-related service offerings and the innovativeness of their products, processes and services. The focus on the marketing communication aspect is a key characteristic of our research approach since it provides a perspective that would be highly relevant for the online articulation of the differentiation and the positioning of the firms with respect to competitors. This is what makes the suggested methodology relevant to the study of the efficiency of web-based marketing platforms in industries such as tourism and airline transportation [26].

2.2. Keyword selection

The web search was based on preliminary designed combinations of keywords derived from a systematic examination of the literature and validated through the visual examination of selected company websites. The keyword selection process started with the selection of keywords from research publications focusing on the specific topic of interest and was complemented by a keyword verification and enrichment procedure based on a visual inspection of particular websites including their customer blogs, developer sites and/or community forums. We have specifically focused on publications suggesting frameworks or models describing the building blocks of co-creation or the value attributes of product-enabled services. The identification of the keywords related to co-creation was done on the basis of the following publications: Nambisan and Baron [17], Nambisan and Nambisan [18], Etgar [13], Payne et al. [14], Jaworski and Kohli [27], Prahalad and Ramaswamy [16], and Ballantyne [28]. The selection of the keywords related to product-enabled services was done on the basis of publications by Velamuri et al. [29], Vladimirova et al. [22], Tukker and Tischner [23], Tukker [24], and Oliva and Kallenberg

Table 2

List of the initial 29 keywords focusing on value co-creation activities.

No.	List of keyword combinations focusing on co-creation
1	Customer OR user AND dialog OR dialog OR conversation OR feedback OR call OR interact OR "information exchange" OR "information sharing" OR "information access" OR engage
2	Customer OR user OR forum OR connect OR network OR networking
3	Lease OR rent OR license OR "self serve" OR "self service"
4	Customer OR user AND cooperate OR cooperation OR collaboration OR partnership
5	Customer OR user AND suggest OR suggestion OR input OR request OR demand
6	Internal AND expertise OR resource
7	Customer OR user AND risk AND manage OR management OR control OR assess OR reduce OR reduction OR potential OR exposure
8	Customer OR user AND IP OR "intellectual property"
9	Customer OR user AND learn OR learning
10	Product OR process OR service AND evolution OR evolve
11	Customer OR user AND experience
12	Customer OR user AND test OR trial OR beta
13	Integrated AND online AND services
14	Simulation OR simulate OR model OR modeling OR "virtual world" OR "reference design" OR "reference flow" OR "demo application" OR toolkit OR tutorial OR sdk OR "software development kit"
15	Product OR process AND modularity OR modular OR module
16	Customer OR user AND produce OR assemble OR manufacture
17	Customer OR user AND options OR choice OR choose
18	Design OR process AND flexibility OR flexible OR adaptable
19	customer AND partnerships OR interaction OR relationship OR participate OR participation OR activity OR action
20	Cost AND reduce OR reduction OR saving
21	Customer OR user AND survey OR review OR voting OR vote OR rate OR rating
22	Trust OR honesty OR integrity OR transparency
23	Customer OR user AND disclose OR inform OR disseminate OR reveal
24	Customer OR user AND dashboard OR statistics
25	Customization OR customize OR customized OR personalize OR individualize OR "add feature" OR "added feature"
26	Customer OR user AND negotiate OR negotiation
27	Ecosystem OR "value network" OR "value constellation" OR "multiple partners" OR "external contributor" OR "external source"
28	Customer OR user AND language AND translation
29	Customer OR user AND address AND concern

Source: Tanev et al. [7].

[30]. The detailed analysis of the selected frameworks resulted in two larger sets of single keywords corresponding to the various types of co-creation activities and service attributes. The single keywords were then combined on the basis of the proximity of their sematic meaning to provide the final keyword combinations (Tables 2 and 3) that were used in the web search of companies' websites. The selection of the keywords associated with the innovativeness of the firms was based on the *Oslo Manual* which is the foremost international source of guidelines for the collection and use of data on innovation activities in industry.[4] The Oslo manual suggests the use of composite metrics containing multiple dimensions of innovation such as the number (for a given period of time) of new products, new processes, new services, new patents, as well as several financial metrics focusing on investment in new product development activities and market success. Given the fact that we are dealing with online data we focused on keywords related to new products, new processes and new services. A preliminary web search test of the different innovation keywords on the company websites indicated that we should combine all the innovative aspects into one composite keyword. In this way the

Table 3
List of the initial 51 keywords focusing on product-enabled service attributes.

No.	List of keyword combinations focusing on product-enabled services
1	Hybrid AND value AND creation OR create
2	Hybrid OR integrated AND value AND offering OR offer
3	Product AND service AND system OR bundle
4	Complex AND product AND system
5	Product AND enabled AND OR related AND service
6	Additional OR add OR added AND value AND service
7	Service AND manufacturing AND business
8	Service AND oriented OR focused OR enabled AND business AND model
9	Increase OR raise OR enhance AND finance OR economic OR environmental AND benefits OR return OR revenues
10	Service AND raise OR elevate OR increase AND profit OR income OR return OR revenue OR revenues OR benefit
11	Service AND minimize OR optimize AND assets
12	Service AND improve OR upgrade OR enhance OR support OR supporting OR supported OR strengthen AND product
13	Service AND success AND product
14	Hard OR difficult OR impossible AND imitate AND service AND competitors OR competitor OR rival OR rivals
15	Differentiate OR extend AND product AND life AND cycle
16	Service AND enhance OR increase OR grow OR gain AND market AND share
17	Service AND increase OR enhance AND product AND usage OR use OR utilization
18	Service AND increase AND firm OR company AND flexibility OR resilience
19	Decrease OR diminish OR lessen OR abate OR reduce OR reduction AND product AND development AND time
20	Service AND improve OR enhance OR enable OR help OR product AND adoption
21	Service AND generate OR provide OR offer OR add AND value
22	Service AND improve AND customer AND trust OR loyalty OR fidelity OR allegiance
23	Service AND create OR creating AND additional AND value
24	Service AND information AND sharing OR interaction
25	Increase OR enhance OR improve AND strategic OR competitive OR business AND advantage
26	Service AND facilitate AND sales OR goods OR production
27	Service AND operational OR operation OR technical OR technology AND integration
28	Service AND marketing OR market OR commercialization OR commercial AND integration
29	Service AND creating OR create AND dependency
30	Create AND new OR complementary AND offering OR solution
31	Product AND individualization OR personalization OR customization AND service
32	Product AND modernization OR improvement OR innovation OR innovative OR innovating
33	Service AND lengthen OR sustain OR sustainable OR strengthen OR improve OR enhance AND customer AND relationships OR interaction
34	Service AND quality OR satisfaction
35	Service AND increase OR enable OR improve AND communication
36	Improve AND customer AND satisfaction OR contentment
37	Hybrid AND offering OR offer AND save OR optimize AND customers AND time OR effort
38	Service AND create OR establish OR make OR enable OR create AND supplier AND customer AND collaboration OR interaction OR "co-creation" OR cocreation OR "co-creating" OR co-creating
39	Product OR system AND service AND availability
40	Product OR system AND service AND effectiveness
41	Product AND service AND performance
42	Product AND service AND modernization OR modernization
43	Product AND service AND innovation OR innovative
44	Product AND service AND competitive OR competitiveness
45	Product AND service AND lifecycle AND cost AND reduction OR reduce OR decrease OR lower
46	Product AND service AND support AND customer OR technical
47	Product AND service AND partner OR partnership OR cooperation OR collaboration OR collaborative OR cooperative OR alliance
48	Service AND training AND customer OR user
49	Product AND service AND maintenance
50	Product AND service AND support AND logistics OR logistic OR logistical
51	Product AND service AND level AND agreement OR turnkey

Source: Modified from Ansevics et al. [31].

new keyword became a measure of firms' degree of online articulation of the innovative aspects of their products, processes and services. This is just one of the possible ways to introduce an online innovation metric and it was found satisfactory at the present stage of our research. There is a need however for more studies that should focus on developing more sensitive online innovation metrics that are able to discern different types of innovation such as product or service, radical or incremental, disruptive or sustaining.

The process described above resulted in a set of 29 keyword combinations focusing on co-creation (Table 2), 51 keywords focusing on product-enabled services (Table 3) and one composite keyword focusing on the innovative aspects of companies' products, processes and services (Table 4). The co-creation and the innovation keywords were previously used by Tanev et al. ([7], pp. 142–143) A subset of the 51 keyword combinations shown in

Table 2 was first tested in a preliminary study aimed at validating the research methodology in the case of product-enabled services [31]. They were however refined and enriched for the present study by adding another 13 keywords with additional service value aspects that were found also to be relevant for our study.

Table 4
Keyword combination focusing on the innovative aspects of companies' new products, processes and services.

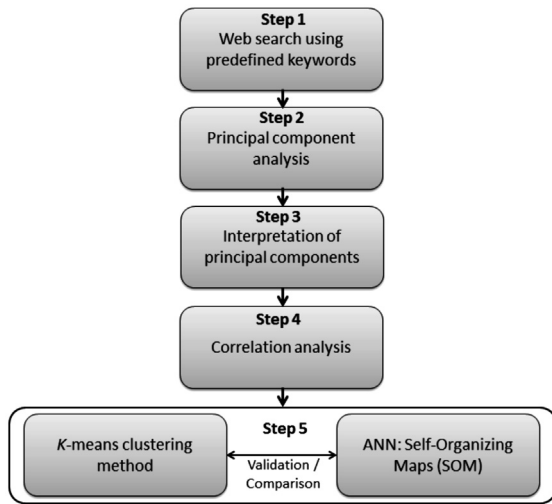| Keyword combinations focusing on the innovative aspects of companies new products, processes and services |
| --- |
| New AND product OR service OR process OR application OR solution OR feature OR release OR version OR launch OR introduction OR introduce OR "new product" OR "new service" OR "new process" OR "new solution" OR "product launch". |

Source: Tanev et al. [7].

Fig. 1. A schematic representation of the overall research process.

Fig. 2. A visual representation of the difference between PCA (a) and cluster analysis (b).

2.3. Web search technique

The set of keyword combinations was used as part of an Advanced Google web search process[5] resulting in the number of hits of each keyword combination on each company website normalized by the number of web pages that could be found under the main company website address. The search process was automated by the development of a software tool using two input data files containing the list of keyword combinations and the list of company websites. The output of the tool is a Microsoft Excel sheet containing a 230 × 29 (or 230 × 51 in the case of product-enabled services and 230 × 1 in the case of the online innovation metric) matrix containing the number of web hits of all the keyword combinations on all 230 company websites. The number of hits was normalized by the total number of webpages that could be found under the main company website.

2.4. Data processing and analysis

The rest of the research process is visualized in Fig. 1. The next step (step 2) consisted in the application of PCA to the company website-vs-keyword matrix to identify the number and the composition of the components including specific sets of keyword combinations that tend to appear together on companies' websites (see the upper half of Fig. 2).

Each PCA component contains a number of keywords which are ranked in terms of their loading values, i.e. the relevance of the specific keyword within the specific component. Having the keyword structure of the PCA components allows for their interpretation in terms of specific co-creation activities and service value attributes (step 3 in Fig. 1). The interpretation is done by assigning a consistent semantic meaning of the entire component based on the meaning conveyed by the keywords having the highest loading values ([32], p. 422). The keyword structure of the components allows also for their quantification as numerical variables by adding the normalized number of web hits corresponding to each of the keyword combinations included in a given PCA component.

The newly created component variables are then used for the ranking or classification of the firms with respect to the degree of their involvement in co-creation and service value articulation.
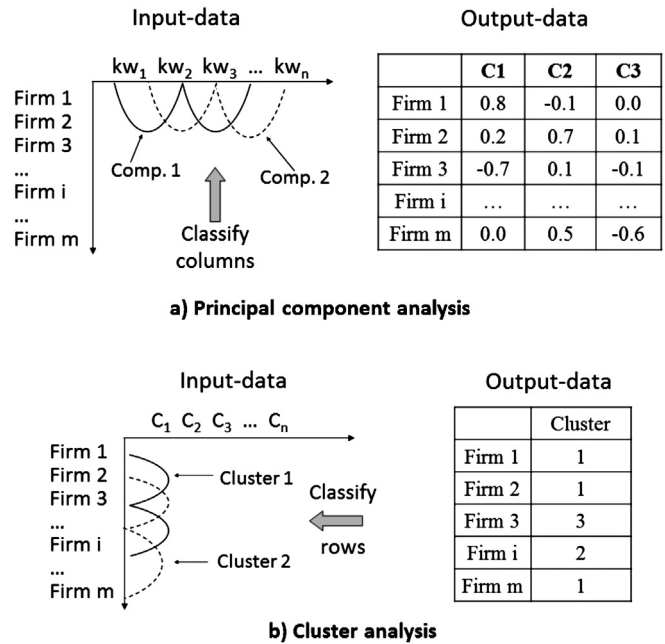
2.5. Classification of the firms

The ranking/classification of the firms was done in two ways (step 5 in Fig. 1) by applying K-means cluster analysis, which is a data reduction technique identifying relatively homogeneous groups of cases based on some preselected characteristics [33], and the SOM approach [8,34,35].

3. Theory

This section will describe the theoretical background of value co-creation and product enabled services. It will also describe the logic of the SOM classification algorithm which is applied in parallel to the more traditional K-means CA.

3.1. Value co-creation practices

Prahalad and Ramaswamy [16] suggested a generative framework describing the fundamental building blocks of value co-creation practices, including Dialog, Access, Risk management and Transparency (DART framework). The open *Dialog* between the multiple actors within the value network encourages knowledge sharing and mutual understanding [27,28]. The initiation of dialog requires a forum with clearly defined rules of engagement leading to an orderly, productive interaction within emerging thematic communities. The focus on *Access* challenges the notions of openness and ownership. Providing customer access to resources, information, tools, assets and processes at multiple points across the value network provides companies with innovative ideas about new products and services, new business opportunities and new potential markets. As customers become co-creators of value, they become more vulnerable to *Risk* and demand more information about the potential risks associated with the design, manufacturing, delivery and consumption of particular products and services. Proactive risk communication and management offers companies with new opportunities for competitive differentiation. *Transparency* enables a creative dialog which helps the emergence of trust between firms' representatives and individual customers [28,36]. By sharing vital business process information with consumers

---

companies hand over part of the control of the value creation process. In addition to the DART framework, Prahalad and Ramaswamy [16] identified four dimensions of choice that could enable personalized co-creation experiences: (i) co-creation across multiple channels, (ii) co-creation through multiple options, (iii) co-creation through multiple transactions at multiple points of access across the value network, and (iv) co-creation through the ability to influence the relationship between price and experience where customers could associate their specific choice with the type of experiences they are willing to pay for. While the existing literature within this stream provides multiple examples of firms that have adopted co-creation principles and useful insights about the specific business and marketing issues that need to be addressed, there is relatively little research on the specific groups of activities that should be undertaken in order to enable the value co-creation processes [14]. There is a clear need for research studies that would contribute to the development of value co-creation platform design rules, transition pathways and maturity implementation models [37]. In addition, there is an opportunity to explore the relationship between the degree of firms' engagement in co-creation and their focus on product-enabled services as a way of business differentiation on the market place.

The study of the innovation-related outcomes of value co-creation activities has gradually emerged as one of the most relevant topics in value co-creation research [7,12,17,19,38,39] since the new paradigm entails a new vision about the relationship between marketing and innovation. However, the growing interest in the co-creation paradigm has been so far predominantly focused on qualitative case studies. Existing literature emphasizes that the customer participation in value co-creation activities has an impact on their innovation outcomes, such as innovation cost, time-to-market, new product or service quality and development capacity [11,12,17,38,39]. It suggests the existence of a tendency for scholars to measure the performance of co-creation practices from an innovation perspective alone, neglecting such remarkable 'side effects' as brand perception, customer satisfaction, customer-firm relationship quality or the value of product-enabled services [10,40]. It has been pointed however that such additional benefits may even exceed in value the actual innovation performance [17]. Last but not least, existing literature fails in pointing out the inherent complex systemic and emerging nature of value co-creation systems as well as how complexity and emergence would affect business model design, pricing models and management practices [7,41]. At the same time, however, there are not many quantitative studies focusing on the co-creative sources of innovation. A first investigation of this specific aspect was done by Tanev et al. [7] who focused on online Internet data and linear regression analysis to examine the relationship between the degree of involvement of firms in value co-creation activities and the frequency of their online comments about new products, processes and services.

3.2. Product-enabled services

The literature on hybrid value offerings has discussed several advantages that could be attained by companies which successfully combine products with services [29]. The advantages could be divided into *economic benefits* and *strategic benefits* [30]. For example, in periods of economic downturns, service revenue streams are less likely to be affected as compared to revenue streams coming from sales of capital goods [42]. In addition, hybrid offerings usually move the firm downstream in the value chain, thereby creating a possibility for differentiation and de-commoditization. Last but not least, a stronger interaction between customers and suppliers through the delivery of services can lead to improved solutions and more innovative hybrid offerings [43].

Other studies focus on the challenges associated with the development of product-enabled services [44–46]. The motivations of product firms to engage into explicit servitization strategies have been also discussed [43]. However, the different value attributes of product related services have not been studied extensively. Existing research is mostly qualitative and needs to be further developed by contextualizing the research studies within a specific industry, firm type or business context. For example, Velamuri et al. [47] discuss four main value creating attributes of hybrid offerings: individualization of the offer; marketing integration, operational integration and firm–customer interaction.

The *individualization* attribute is related to the possibility for customers to individualize or customize a certain offering. Hybrid offerings foster customization as services by nature are delivered through the interaction with specific individual customers. At its best, individualization can be considered as a co-creation activity which essentially enables the customer to co-create and contextualize value in cooperation with the focal firm [11]. Services are always co-created and require real-time involvement of the end users, as compared with products where it is much more difficult to involve users in the pre-launch testing phase [45]. What is more important however is that customers are usually willing to pay a premium for an individualized offering [47]. *Marketing integration* denotes the ability of an offering to deliver value beyond the sum of all of its included components. It is related to product-service bundling where a customer can benefit from consolidated purchases of maintenance, repair or installation from a single vendor, thus capturing more value in terms of saved time, effort and money [47]. Marketing integration can be related to the hybrid characteristic of low complementarity offers since the customer will generally benefit from various aspects based on convenience. *Operational integration* is about seamlessly linking and/or engineering the various components of an offering in order to avoid effort, time and risk on the customer side [47]. Hybrid offerings enable such kind of value by presenting total operational solutions for complex problems based on products and services available from a single vendor [29]. Operational integration will predominantly be associated with high complementarity since the solution can only work and bring value if the products and services are integrated tightly together [47]. The quality of *firm–customer interaction* is considered as a critical source of additional value for the customer while at the same time constituting a critical strategic resource for the firm [11,16,47]. The focus of this specific type of service value is not on individual opportunities for value exchange, but instead on a continuous value extraction process which is distributed over the entire product lifecycle.

Existing literature suggests several ways of combining products with services depending on which value attributes the firm wants to achieve. Shankar et al. [48] propose two dimensions to determine how customers will value and use a hybrid offering: *complementarity* – to what extent the customer value increases when the customer uses the product in combination with the service; *independence* – to what extent the customer can extract value from one part of the offering without the availability of the other part. The two dimensions suggest four types of product-service combinations: multi-benefit bundle, flexible bundle, one-stop bundle and peace-of-mind bundle. All four combinations offer the customer a different set of value attributes. In the multi-benefit bundle (low independence, high complementarity) customers are offered a hybrid solution that will dramatically increase in value when the product and service are used together. The product and service cannot, however, be purchased separately as it is a highly integrated offering. A flexible bundle (high independence, high complementarity) offers similar value in terms of degree of complementarity, but the products and services can be bought separately. The true value in these kinds of offerings is found in the

ability of the supplier to handle complex problems for the customers by providing full operational solutions. Low complementarity hybrid offerings such as the one-stop and peace-of-mind bundle (with low and high independence, respectively) create value based on convenience and quality assurance as the products and services do not generate any significant extra value when combined together. Oliva and Kallenberg [30] associate these combinations with less advanced hybrid offerings relating them to product-centric service activities associated with product installation and maintenance.

4. Results

4.1. PCA components of value co-creation practices and product-enabled service attributes

4.1.1. Value co-creation components

The data collection method resulted in a matrix having two dimensions: *one*, containing the list of the firms in the research sample and, *two*, containing the number of hits of each keyword combination on each company website normalized by the number of web pages under the main company website address. The first step in the data analysis consisted in validating the results of the factor analysis by examining the correlation table determinant, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's significance test [49]. The second step was to examine the Scree plot of the PCA test in order to determine the optimum number of PCA components [50]. The PCA extraction method was preferred among all other methods available in the SPSS software since it provided the highest loadings for the individual keyword combinations in the PCA components. Table 5 shows the four components corresponding to the value co-creation practices. The four components explain 59.44% of the variance of the data. The minimum possible loading value was set to 0.45. The results show that there were only 21 keywords (out of

the initial set of 29 keyword combinations) that remained in the final analysis.

Component 1 could be interpreted as including activities "Providing mutual learning mechanisms by offering multiple options for customers and end users through their involvement in tests, beta trials, user networking forums and surveys leading to the articulation of specific suggestions, demands and requests." Component 2 could be interpreted as including activities enabling "Customer and user risk management through the involvement in partnerships and cooperation aiming at cost reduction and better experiences based on design and process flexibility." Component 3 could be interpreted as activities leading to "Customization or personalization of companies' products through modular design and process flexibility that could enable the efficient use of customer intellectual property." Component 4 could be interpreted as providing "Multiple language customer and end user self-service access to internal resources, tools and simulation toolkits through leasing, renting and licensing."

4.1.2. Product-enabled service attribute components

Table 6 shows the three components corresponding to the value attributes of product enabled services. The three components explain 68.13% of the data variance. The results show that there were only 15 keywords (out of the initial set of 51 keyword combinations) that were included in the analysis.

The first service value component (attribute) could be formulated as "Enhancing company flexibility, resilience and asset efficiency to achieve stronger competitive advantage and better communication with customers aiming at more sales, increased revenue and product-service innovation." The second service value attribute could be formulated as the "Integration of technology and operations management, technology commercialization and marketing leading to higher value of training for customers and end users." The third service value attribute could be formulated as "Additional value in terms of higher profitability and revenue."

Table 5
PCA components corresponding to value co-creation practices. The loading value indicates the relevance of every specific keyword combination in the interpretation of each component.

Combination of keywords included in co-creation Component 1		Loading
Customer OR user AND test OR trial OR beta		0.75
Customer OR user AND suggest OR suggestion OR input OR request OR demand		0.74
Customer OR user AND options OR choice OR choose		0.69
Customer OR user AND learn OR learning		0.68
Customer OR user AND survey OR review OR voting OR vote OR rate OR rating		0.64
Customer OR user OR forum OR connect OR network OR networking		0.53
Customer OR user AND dashboard OR statistics		0.48
Combinations of keywords included in co-creation Component 2		Loading
Customer OR user AND risk AND manage OR management OR control OR assess OR reduce OR reduction OR potential OR exposure		0.75
Customer OR user AND experience		0.74
Customer AND partnerships OR interaction OR relationship OR participate OR participation OR activity OR action		0.64
Cost AND reduce OR reduction OR saving		0.64
Customer OR user AND cooperate OR cooperation OR collaboration OR partnership		0.62
Design OR process AND flexibility OR flexible OR adaptable		0.61
Combinations of keywords included in co-creation Component 3	Loading	C2 loading
Customization OR customize OR customized OR personalize OR individualize OR "add feature" OR "added feature"	0.71	
Product OR process AND modularity OR modular OR module	0.70	
Customer OR user AND IP OR "intellectual property"	0.60	
Design OR process AND flexibility OR flexible OR adaptable	0.51	0.61
Combinations of keywords included in co-creation Component 4		Loading
Customer OR user AND language AND translation		0.69
Lease OR rent OR license OR "self serve" OR "self service"		0.65
Internal AND expertise OR resource		0.60
Simulation OR simulate OR model OR modeling OR "virtual world" OR "reference design" OR "reference flow" OR "demo application" OR toolkit OR tutorial OR sdk OR "software development kit"		0.52

Table 6

PCA components corresponding to the value attribute of product-enabled services. The loading value indicates the relevance of every specific keyword combination in the interpretation of each component.

Combinations of keywords included in service value Component 1		Loading
Service AND increase AND firm OR company AND flexibility OR resilience		0.83
Service AND minimize OR optimize AND assets		0.77
Increase OR enhance OR improve AND strategic OR competitive OR business AND advantage		0.72
Service AND increase OR enable OR improve AND communication		0.71
Service AND facilitate AND sales OR goods OR production		0.68
Product AND service AND innovation OR innovative		0.66
Product AND service AND maintenance		0.58
Increase OR raise OR enhance AMD financial OR economic OR environmental AND benefits OR return OR revenues		0.56
Decrease OR diminish OR lessen OR abate OR reduce OR reduction AND product AND development AND time		0.46
Combinations of keywords included in service value Component 2	Loading	C3 loading
Service AND marketing OR market OR commercialization OR commercial AND integration	0.87	
Service AND operational OR operation OR technical OR technology AND integration	0.86	
Service AND training AND customer OR user	0.65	
Additional OR add OR added OR high OR higher AND value AND service	0.57	0.67
Combinations of keywords included in service value Component 3	Loading	C2 loading
Service AND create OR creating AND additional AND value	0.84	
Service AND raise OR elevate OR increase AND profit OR income OR return OR revenue OR revenues OR benefit	0.73	
Additional OR add OR added OR high OR higher AND value AND service	0.67	0.57

Table 7

Cross-correlation table including all important variables: C1_ccr, C2_ccr, C3_ccr and C4_ccr; C_ccr = C1_ccr + C2_ccr + C3_ccr + C4_ccr; C1_pes, C2_pes and C3_pes; C_pes = C1_pes + C2_pes + C3_pes; Innovation.

Correlations – Spearman's rho coefficients[*]									
	C1_ccr	C2_ccr	C3_ccr	C4_ccr	C_ccr	C1_pes	C2_pes	C3_pes	C_pes	Innovation
C1_ccr		0.68	0.62	0.56	0.93	0.60	0.65	0.63	0.67	0.71
C2_ccr			0.66	0.50	0.81	0.86	0.84	0.78	0.87	0.62
C3_ccr				0.58	0.77	0.65	0.65	0.61	0.67	0.52
C4_ccr					0.71	0.55	0.53	0.55	0.57	0.51
C_ccr						0.73	0.76	0.73	0.79	0.73
C1_pes							0.83	0.83	0.92	0.55
C2_pes								0.81	0.96	0.61
C3_pes									0.91	0.61
C_pes										0.64

[*] All correlation tests had *p*-values (two tail significance level) less than 0.0001.

4.2. Relationship between value co-creation, product-enabled service value and innovation

The results presented in the previous section were used to construct quantitative variables corresponding to each of the components shown in Tables 5 and 6. The variables were constructed by adding the normalized web hits of all the keywords included in a given component weighted by their specific loadings. The processes resulted in four variables (C1_ccr, C2_ccr, C3_ccr and C4_ccr) corresponding to each of the four co-creation components, and three variables (C1_pes, C2_pes and C3_pes) corresponding to each of the three service value components. We also introduced a total co-creation variable C_ccr = C1_ccr + C2_ccr + C3_ccr + C4_ccr, and a total service value variable C_pes = C1_pes + C2_ pes + C3_pes. In addition, we also have the variable (Innovation) corresponding to firms' perception of the innovativeness of their products, processes and services. Table 7 shows the cross-correlation table including all these variables. The specific results will be discussed in Section 5.1.

4.3. Classification of the firms

4.3.1. Classification of the firms on the basis of K-means CA

4.3.1.1. K-means clustering based on the co-creation and innovation variables. The K-means clustering procedure uses preliminary selected variables to group the firms in several clusters that could be characterized by different mean values (centers) of the

corresponding variables [33]. Fig. 2 illustrates the difference between PCA and cluster analysis. While the PCA analysis uses cross-correlation analysis of the keywords to group them into components (Fig. 2a), the cluster analysis uses the newly generated PCA component variables to group the firms into several clusters (Fig. 2b). The specific distribution of the firms in the different clusters depends on the number and the type of the preliminary selected PCA component variables. Table 8 shows the distribution of the firms in four clusters on the basis of the co-creation and the

Table 8

K-means cluster centers and distribution of the firms in the clusters based on the co-creation and innovation variables.

K-means cluster centers (co-creation + innovation)				
	Clusters			
	1	2	3	4
C1_ccr	2.16	1.24	0.75	0.26
C2_ccr	0.94	1.15	0.26	0.08
C3_ccr	0.90	0.24	0.16	0.06
C4_ccr	0.90	0.60	0.22	0.09
Innovation	0.62	0.49	0.25	0.10
Average	1.10	0.74	0.33	0.12

Clusters based on co-creation and innovation variables	Number of cases in each K-means cluster
1	4
2	4
3	99
4	123

Table 9
Cross tabulation of the type of firms vs the K-means clusters based on the co-creation and innovation variables.

Cross tabulation: K-means cluster number × type of firm (co-creation + innovation)					Total
		Type			
		ECL	GEN	OSS	
Cluster number	1	2	0	2	4
	2	3	0	1	4
	3	45	18	36	99
	4	44	40	39	123
Total		94	58	78	230

innovation variables together with the cluster centers. One can see that the clusters are ranked in terms of the average cluster centers. Table 9 shows the distribution of the different types of firms within the four K-means clusters based on the co-creation and the innovation variables. Fig. 3 is a graphical representation of the results shown in Table 9. The specific will be discussed in Section 5.2.

4.3.1.2. K-means clustering based on the service value and innovation variables. Table 10 shows the distribution of the firms in four clusters on the basis of the service value and the innovation variables together with the cluster centers. The clusters are ranked in terms of the average cluster centers. Table 11 shows the distribution of the different types of firms within the four K-means clusters based on the service value and the innovation variables. Fig. 4 is a graphical representation of the results shown in Table 11.

Fig. 3. Graphical representation of the results shown in Table 9.

Table 10
K-means cluster centers and distribution of the firms in the clusters based on the service value and innovation variables.

K-means cluster centers (service value + innovation)				
	Clusters			
	1	2	3	4
C1_pes	0.38	0.29	0.14	0.04
C2_pes	2.47	0.59	0.29	0.05
C3_pes	0.92	0.67	0.15	0.04
Innovation	0.55	0.59	0.29	0.12
Average	1.08	0.54	0.22	0.06

Clusters based on service value and innovation variables	Number of cases in each K-means cluster
1	1
2	7
3	60
4	162

Table 11
Cross tabulation of the type of firms vs the K-means clusters based on the service value and innovation variables.

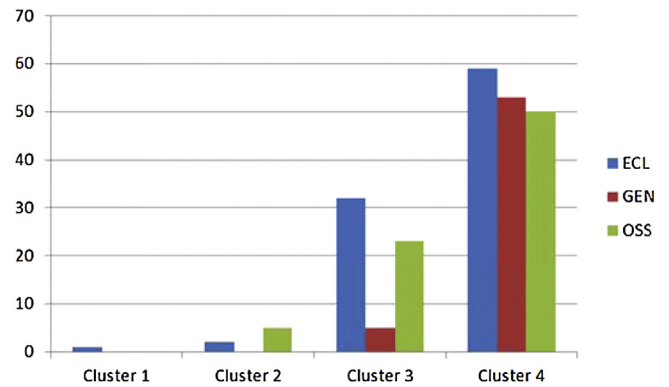| Cross tabulation: K-means cluster number × type of firm (service value + innovation)					Total
		Type of Firm			
		ECL	GEN	OSS	
Cluster Number	1	1	0	0	1
	2	2	0	5	7
	3	32	5	23	60
	4	59	53	50	162
Total		94	58	78	230

Fig. 4. Graphical representation of the results shown in Table 11.

4.4. Classification of the firms on the basis of SOMs

Artificial neural networks (ANNs) are algorithms modeled as elementary units or neurons connected in a way that forms a network capable of solving complex nonlinear problems. SOM is a type of ANN approach using unsupervised learning training to produce a two-dimensional, discretized representation of the input data samples, called a map, and they are different from other ANNs since they use a neighborhood function to preserve the topological properties of the input space [35]. We have decided to use SOMs since they allow for the simultaneous recognition of groups of similar input vectors (clustering) and to project the data into lower-dimensional spaces (2D graphs) [51]. Tables 12 and 13 show the results of the four cluster centers obtained by the SOM clustering based on the co-creation and innovation variables (Table 12) and the service value and innovation variables (Table 13). The clusters are ranked in terms of their average center values.

Table 14 shows the distribution of the firms in the clusters based on the SOM analysis of the co-creation, service value and innovation variables. The table includes the corresponding results in the case of K-means clustering.

Table 12
SOM cluster centers based on the co-creation and innovation variables.

SOM cluster centers (co-creation + innovation)				
	Clusters			
	1	2	3	4
C1_ccr	1.87	1.09	0.65	0.24
C2_ccr	1.16	0.30	0.25	0.07
C3_ccr	0.70	0.20	0.15	0.06
C4_ccr	0.78	0.35	0.19	0.09
Innovation	0.59	0.39	0.22	0.09
Average	1.02	0.47	0.29	0.11

Table 13
SOM cluster centers based on the service value and innovation variables.

SOM cluster centers (service value + innovation)				
	Clusters			
	1	2	3	4
C1_pes	0.38	0.27	0.11	0.03
C2_pes	2.47	0.58	0.22	0.04
C3_pes	0.92	0.47	0.12	0.04
Innovation	0.56	0.43	0.29	0.10
Average	1.08	0.44	0.19	0.05

5. Discussion

5.1. Relationship between degree of involvement in co-creation, articulation of product enabled service value and perception of innovation

The analysis suggested in this section is based on the results shown in Table 7 containing the cross-correlation table including all the variables related to value co-creation, product-enabled service value and the perception of innovation. The first observation based on the examination of Table 7 is the fact that there is a relatively high degree of statistically significant correlation between all the variables (all p-values of the statistical tests for correlation were less than 0.0001). Almost all correlation coefficients are larger than 0.5. Statistics literature provides some guidelines for the interpretation of the degree of correlation between two variables, according to which coefficients larger than 0.5 could be interpreted as corresponding to strong or large correlation ([32], pp. 93–95). However, our correlation results should be interpreted with a sober optimism since the numerical variables are based on measuring the frequency of particular keywords in online textual data and not on the answers to specific survey questions. In addition, the search terms include some overlap of words, and hence, some correlation is expected. In this sense, the specific correlation coefficients should not be interpreted in absolute terms and the difference in the correlation coefficients should be rather considered as an indication of a tendency that should be further explored by more detailed qualitative methods [9,10]. This however does not weaken the value of the suggested methodology. It just shows the opportunity to use it as a valuable probing device that could indicate potential hidden relationships that might not be able to be detected in other ways.

One can start by pointing out that the correlation between the C_ccr (the total co-creation variable) and C_pes (the total service value variable) is 0.79. We believe that this could be the first quantitative empirical indication that there is a relationship between firms' degree of involvement in co-creation activities and the degree of articulation of the value attributes of their product-enabled services. In this sense, this finding could be considered as the key contribution of our article. Although it should not come as a surprise, it provides an empirical insight about the complementary relationship between co-creation activities and product-enabled services as well as about some

of the ways companies could enhance the relevance of both their co-creation and new service development activities. In addition, it offers an opportunity to emphasize the distinction between the co-creative nature of services and user involvement in co-creation activities. It is true that services are always co-created since they require the real-time involvement of the end users [45]. This is however different from user involvement in co-creative activities which requires a completely different setting enabling the active and deliberate participation of customers and end users with respect to specific market offers. Such participation may be enhanced by specific complementary services but cannot be based just on these services.

The results shown in Table 7 allow for a more detailed discussion of the degree of complementarity between product-enabled services and market offers based on co-creation with customers and end users. For example, the highest correlation coefficient (0.86) is between C2_ccr (customer risk management through the involvement in partnerships and cooperation aiming at cost reduction and better experiences based on design and process flexibility) and C1_pes (company flexibility, resilience and asset efficiency to achieve stronger competitive advantage and better communication with customers aiming at more sales, increased revenue and product-service innovation). C2_ccr focuses on partnerships and cooperation with customers and identifies design and process flexibility as a source of better customer experiences. C1_pes suggests that the introduction of product-enabled services makes a company more flexible and more resilient which leads to better communication with customers and more economic benefits. The high degree of correlation between the two variables suggests that, *first*, the combination of a stronger customer integration with appropriate product-enabled services could enhance the benefits for both the customers and the company. *Second*, the introduction of product-enabled services enhances the flexibility and the resilience of the companies which, on the other hand, enhances the value of customer experiences. It should also be pointed out that there is a high degree of correlation between C2_ccr and the other two service value components – C2_pes (integration of technology and operations management, technology commercialization and marketing leading to higher value of training for customers and end users) and C3_pes (additional value in terms of higher profitability and revenue). This is just another illustration of the complementary nature of co-creation activities and product-enabled services as drivers of both customer and company value. At the same time it should be pointed out that the coefficient of correlation between C4_ccr (multiple language customer and end user self-service access to internal resources, tools and simulation toolkits through leasing, renting and licensing) and C2_pes is much lower (0.53). Such lower degree of correlation could be interpreted in several ways. For example, it might suggest that services enabled through the use of more technical solutions could be more difficult to commercialize and may need the introduction of additional customer training services. Or, one may suggest that a more active involvement of customers and end users in self-services could help the

Table 14
Distribution of the firms in the clusters based on the SOM analysis of the co-creation, service value and innovation variables. The table also shows the corresponding results in the case of K-means clustering.

	SOM clustering based on the co-creation and innovation variables	SOM clustering based on the service value and innovation variables	K-means clustering based on the co-creation and innovation variables	K-means clustering based on the service value and innovation variables
Cluster 1	6	1	4	1
Cluster 2	21	14	4	7
Cluster 3	89	68	99	60
Cluster 4	114	147	123	162

commercialization of co-created offers trough the development of proper customer training solutions. In all cases the low value of the correlation coefficient demonstrates the sensitivity of the methodology to differentiate specific business and innovation aspects in terms of their relevance and frequency.

The second most valuable trend in the results of the correlation analysis is the relationship between co-creation, service value and the perception of innovation. The correlation between C_ccr (total co-creation variable) and Innovation (the frequency of online articulation of the innovative aspects of firms' products, processes and services) is 0.73. On the other hand, the correlation between C_pes (total service value variable) and Innovation is 0.64. These results indicate that firms' involvement in co-creation and the introduction of new product-enabled services are positively related to firms' ability to differentiate through a better articulation of the innovative aspects of their products, processes and services. It however appears that the relationship between value co-creation and Innovation is stronger. This is another important finding that deserves a closer attention since it uses empirical data to affirm the understanding of co-creation as part of the innovation management practices of the firms [39]. Interestingly, the highest correlation is between C1_ccr (providing mutual learning mechanisms by offering multiple options for customers and end users) and Innovation (0.71). In other words, user involvement in tests, beta trials and networking forums leading to the articulation of specific customer suggestions, demands and requests appears to have a direct relation to firms' ability to articulate its innovativeness. On the one hand, such relation could be interpreted as an expression of a mere marketing emphasis on what customers care about. On the other hand, a better articulation of the innovative aspects of new products, processes and services could be associated with a stronger dedication to a more articulated innovation management strategy. In this sense, the article identifies the potential link between firms' degree of online articulation of the innovative aspects of their products, processes, and services and their commitment to a more focused innovation management strategy.

In what it concerns product-enabled services, C2_pes (higher value offered through customer training) and C3_pes (additional value in terms of higher profitability and revenue) are equally correlated to Innovation (0.61). These results suggest that, first, customer training is associated with a better articulation of the innovative aspects of companies' products and services and, second, more innovative products and services are associated with higher profitability. At the same time, customer training should be considered as a valuable source of revenue on its own. The interpretation of the lower degree of correlation between C1_pes (focus on flexibility, resilience and asset efficiency to achieve stronger competitive advantage and better communication with customers aiming at more sales, increased revenue and product-service innovation) and Innovation could be a bit more challenging and deserves more attention in future studies. It could be said however that companies' efforts on achieving more flexibility, resilience and asset efficiency might have a lower impact on their ability to demonstrate their innovativeness.

5.2. Ranking of the firms in terms of the degree of their co-creation activities, articulation of service value attributes and perception of innovation

The ranking of the firms was done by two different approaches – K-means CA and SOM approach. The two types of clustering are not identical but show the same tendency in the distribution of the firms in four different clusters (Table 14). The similarity in the K-means and SOM clustering is of particular interest since it is based on two completely different ways of grouping the firms in terms of the articulation of their co-creation, new service development and

innovation activities. A detailed analysis however is out of the scope of this article and will be addressed in future studies. What seems to be more interesting is to deepen the examination of the relationship between firms' involvement in co-creation activities and the introduction of new product-enabled services. To do that we compared the K-means clustering results provided by the two different sets of variables – (the four co-creation variables together with the Innovation variable) vs (the three service value variables and together with the Innovation variable). This was a way to see, first, to what extent the ranking of the firms would depend on the specific choice of the input variables and, second, to offer the opportunity to interpret the similarity in the two rankings in terms of a potential relationship between co-creation activities and product-enabled services. Table 15 shows a combination of the results shown in Tables 9 and 11 which allow to compare the two different K-means clustering results including the distribution of the different types of firms in the different clusters.

The two rankings show a certain degree of similarity. The degree of similarity could be examined by means of a statistical test [52] to assess the proximity between the two partitions obtained on the basis of the two sets of variables (in our case: "co-creation and innovation" and "service value and innovation"). The statistical test measures the proximity between the two partitions by defining a parameterized family of metric functions that includes Shannon entropy as a special case. The reported proximity is 0.35, which supports our finding about the potential relationship between firms' involvement in co-creation with customers and their ability to introduce new product-related services. The value indicates a more moderate relationship which could be associated with a more realistic view about the extent to which product-enabled services contribute to value co-creation and innovation. The link however appears to be unquestionable.

One can also see that in both cases the firms included in the first two clusters (having the highest center means) are members of the Eclipse Foundation (ECL) and OSS firms. All general type (GEN) firms (non-Eclipse and non OSS firms) can be found in the clusters 3 and 4 which include firms with a lower degree of involvement in co-creation and a lower degree of articulation of their service value attributes.

This finding shows that OSS and Eclipse firms offer good examples of best practices in terms of co-creation and new product-enabled service development. It could be explained by the fact that such firms offer well designed technology-driven customer participation platforms aiming at offering a stronger degree of involvement in co-creation and supporting services [7,8,25]. Furthermore, it raises interesting research questions such as: What could other (non-Eclipse and non-OSS) firms do to enable customer participation? Should they emulate OSS and Eclipse firms, and if so, how difficult would that be for them? The present study suggests a positive answer to the last question, but a deeper study focusing on translating the OSS collaborative principles within the context of firms interested in adopting a customer-driven co-creation strategy is still missing.

Table 15
Comparison of the distribution of the firms in the four clusters based on the K-means analysis.

Cluster	Clustering based on co-creation and Innovation			Clustering based on service value and innovation		
	Type of firm			Type of firm		
	ECL	GEN	OSS	ECL	GEN	OSS
1	2	0	2	1	0	0
2	3	0	1	2	0	5
3	45	18	36	32	5	23
4	44	40	39	59	53	50

Table 16
Ranking of the first 25 firms in the sample on the basis of the C_ccr (total co-creation variable) values.

#	Company name	Company type	Total co-creation variable	Total service variable	Co-creation K-means clustering	Service K-means clustering
1	http://www.tensilica.com/	ECL	6.08	0.38	1	2
2	http://www.vyatta.com/	OSS	5.10	2.52	1	1
3	http://www.birt-exchange.com/	ECL	4.28	1.77	1	1
4	http://www.enterprisedb.com/	OSS	4.14	1.38	1	1
5	http://www.iwaysoftware.com/	ECL	4.12	3.78	1	1
6	http://www.borland.com/	ECL	3.38	1.18	1	1
7	http://www.progress.com/	OSS	2.87	1.32	1	1
8	http://www.blackducksoftware.com/	ECL	2.55	0.60	1	2
9	http://www.ddci.com/	ECL	2.54	0.64	2	2
10	http://code.google.com/android/	OSS	2.51	1.66	2	1
11	http://www.mozilla.com/en-US/	OSS	2.38	0.24	2	2
12	http://www.compiere.com/	OSS	2.24	1.14	2	2
13	http://www.android.com/	OSS	2.13	0.07	2	2
14	http://www.project-open.com/	OSS	2.05	0.49	2	2
15	http://www.serena.com/	ECL	1.95	0.55	2	2
16	http://www.prosyst.com/	OSS	1.90	1.36	2	2
17	http://aws.typepad.com/	GEN	1.89	0.52	2	2
18	http://www.curl.com/	ECL	1.89	0.28	2	2
19	http://www.latticesemi.com/	GEN	1.87	0.12	2	3
20	http://www.orangehrm.com/	OSS	1.86	0.42	2	2
21	http://www.cleversafe.com/	OSS	1.85	0.58	2	2
22	http://developer.amd.com/	ECL	1.82	0.19	2	3
23	http://www.arm.com/	ECL	1.80	0.45	2	2
24	http://www.timesys.com/	OSS	1.80	0.63	2	2
25	http://www.pentaho.com/	OSS	1.79	1.03	2	2

Table 16 provides the ranking of the first 25 firms in terms of their C_ccr (total co-creation variable) value. The list of firms contains 13 OSS, 10 ECL and 2 GEN firms. The last two columns on the right-hand side indicate the association of the firms with a specific cluster number. One can see that to a great extent the two clustering methods agree on the types of firms included in the top 2 clusters. The association of specific types of firms with particular clusters is a valuable finding with a great potential for future studies. The value of this finding consists in the possibility to relate the clustering of firms to their type and not just to their ranking with respect to one of the key performance variables. The results shown in Table 16 could be also used as an illustration of the correlation between the degree of firms' involvement in co-creation activities and the degree of their focus on product-enabled service value. This is another illustration of the relationship between value co-creation and product-enabled services that could be considered as one of the major contributions of this article.

6. Conclusion

In this article we introduced a method for semantic analysis of online textual data that is similar to latent semantic analysis but could be used as part of user-friendly business intelligence tools. The method employs a preliminary constructed set of keywords that were designed on the basis of scholarly and professional expertise in the field of value co-creation and product-enabled services. It combines web search techniques, principal component analysis, correlation analysis, and K-mean CA which was complemented by a neural network approach based on SOMs. We have used this method to identify a relationship between firms' degree of involvement in co-creation activities, the degree of articulation of their service value attributes and their innovativeness.

The article has three key contributions. The first one is methodological and consists in the validation of a business intelligence method based on the combination of web search techniques and easily accessible statistical packages. The ease-of-use of the method will allow for more scholars and practitioners to make sense of publicly and freely available online textual data.

The second key contribution consists in the identification of a relationship between firms' degree of involvement in co-creation activities, the degree of articulation of their service value attributes and their innovativeness. We have found that a stronger customer integration through appropriate product-enabled services could enhance the benefits for both customers and firms: it enhances the flexibility and the resilience of the firms, and the customer experiences, leading to additional value for firms in terms of higher profitability and revenue.

Finally, the article identifies OSS-driven firms as an example of best practices in terms of co-creation and new product-enabled service development. The question how other (non-OSS) firms could translate the collaborative principles built in their customer participation platforms within the context of their own operational and business environment is of particular interest and will be considered in future studies. Future studies should also focus on seeking a more seamless integration and automation of the different techniques included in the present article. Last but not least, the suggested analytical method should be further developed through its application in other business and innovation domains.

References

[1] N. Evangelopoulos, Latent semantic analysis, WIREs Cognit. Sci. 4 (November–December) (2013) 683–692, http://dx.doi.org/10.1002/wcs.1254.
[2] N. Evangelopoulos, X. Zhang, V. Prybutok, Latent semantic analysis: five methodological recommendations, Eur. J. Inf. Syst. 21 (2012) 70–86.
[3] D. Martin, M. Berry, Mathematical foundations behind latent semantic analysis, in: T. Landauer, D. Mcnamara, S. Dennis, W. Kintsch (Eds.), Handbook of Latent Semantic Analysis, Lawrence Erlbaum Associates, Mahwah, NJ, 2007, pp. 33–57.
[4] T. Landauer, P. Foltz, D. Laham, An introduction to latent semantic analysis, Discourse Process. 25 (2–3) (1998) 259–284, http://dx.doi.org/10.1080/01638539809545028.
[5] W. Chung, BizPro: extracting and categorizing business intelligence factors from textual news articles, Int. J. Inf. Manage. 34 (2014) 272–284.
[6] D. Libaers, D. Hicks, A. Porter, A taxonomy of small firm technology commercialization, Ind. Corp. Change 6 (2010) 1–35.
[7] S. Tanev, T. Bailetti, S. Allen, H. Milyakov, P. Durchev, P. Ruskov, How do value co-creation activities relate to the perception of firms' innovativeness? Special issue "Rethinking the boundaries of innovation", J. Innov. Econ. 1 (7) (2011) 131–159.
[8] G. Di Tollo, S. Tanev, D. De March, Z. Ma, Neural Networks to model the innovativeness perception of co-creative firms, Expert Syst. Appl. 39 (16) (2012) 12719–12726.
[9] S. Tanev, G. Liotta, A. Kleismantas, A business intelligence approach using web search tools and online data reduction techniques to examine the value of product-enabled services, Expert Syst. Appl. 42 (21) (2015) 7582–7600.

[10] A. Kleismantas, S. Tanev, G. Liotta, The value of product-enabled services in top R&D spenders in Canada and Europe, in: 21st International Product Development Management Conference, Limerick, Ireland, June 15–17, 2014.

[11] V. Ramaswamy, F. Gouillart, Building the co-creative enterprise, Harv. Bus. Rev. (October) (2010) 100–109.

[12] C.K. Prahalad, M.S. Krishnan, The New Age of Innovation, McGraw Hill, New York, 2008.

[13] M. Etgar, A descriptive model of the consumer co-production process, J. Acad. Mark. Sci. 36 (1) (2008) 97–108.

[14] A. Payne, K. Storbacka, P. Frow, Managing the co-creation of value, J. Acad. Mark. Sci. 36 (2008) 83–96.

[15] R. Lusch, S. Vargo (Eds.), The Service Dominant Logic of Marketing. Part III. Co-production, Collaboration, and Other Value-Creating Processes, M.E. Sharpe, New York, 2006, pp. 105–179.

[16] C.K. Prahalad, V. Ramaswamy, The Future of Competition – Co-creating Unique Value with Customers, Harvard Business School Press, Boston, 2004.

[17] S. Nambisan, A. Baron, Virtual customer environments: testing a model of voluntary participation in value co-creation activities, J. Prod. Innov. Manage. 26 (2009) 388–406.

[18] S. Nambisan, P. Nambisan, How to profit from a better 'virtual customer environment', MIT Sloan Manage. Rev. 49 (3) (2008) 53–61.

[19] M. Sawhney, V. Gianmario, E. Prandelli, Collaborating to create: the internet as platform for customer engagement in product innovation, J. Interact. Mark. 19 (4) (2005) 4–17.

[20] S. Vargo, R. Lusch, Evolving to a new dominant logic for marketing, J. Mark. 68 (January) (2004) 1–17.

[21] S. Cavalieri, G. Pezzotta, Product–service systems engineering: state of the art and research challenges, Comput. Ind. 63 (2012) 278–288.

[22] D. Vladimirova, S. Evans, V. Martinez, J. Kingston, Elements of change in the transformation towards product service systems, in: J. Hesselbach, C. Herrmann (Eds.), Functional Thinking for Value Creation, Springer, Berlin, 2011, pp. 21–26.

[23] A. Tukker, U. Tischner, Product-services as a research field: past, present and future. Reflections from a decade of research, J. Clean. Prod. 14 (2006) 1552–1556.

[24] A. Tukker, Eight types of product–service system: eight ways to sustainability? Experiences from Suspronet, Bus. Strateg. Environ. 13 (2004) 246–260.

[25] J. West, S. O'Mahony, The role of participation architecture in growing sponsored open source communities, Ind. Innov. 15 (2) (2008) 145–168.

[26] W.-H. Tsai, W.-C. Chou, J.-D. Leu, An effectiveness evaluation model for the web-based marketing of the airline industry, Expert Syst. Appl. 38 (2011) 15499–15516.

[27] B. Jaworski, A. Kohli, Co-creating the voice of the customer, in: R. Lusch, S. Vargo (Eds.), The Service Dominant Logic of Marketing, M.E. Sharpe Inc., Armonk, NY, 2006 (Chapter 8).

[28] D. Ballantyne, Dialog and its role in the development of relationship specific knowledge, J. Bus Ind. Mark. 19 (2) (2004) 114–123.

[29] V. Velamuri, A. Neyer, K. Möslein, Hybrid value creation: a systematic review of an evolving research area, J. Betr. 61 (1) (2011) 3–35.

[30] R. Oliva, R. Kallenberg, Managing the transition from products to services, Int. J. Serv. Ind. Manage. 14 (2) (2003) 160–172.

[31] E. Ansevics, T. Iversen, S. Tanev, G. Liotta, Product enabled services' potential for sustainable growth: an empirical study, in: Proceedings of the XXIV ISPIM Conference "Innovating in Global Markets: Challenges for Sustainable Growth", Helsinki, Finland, 16–19 June, (2013), p. 11.

[32] J. Reinard, Communications Research Statistics, Sage Publications, Thousand Oaks, 2006.

[33] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recogn. Lett. 31 (2010) 651–666.

[34] T. Kohonen, M.R. Schroeder, T.S. Huang (Eds.), Self-organizing Maps, Springer-Verlag New York, Inc., Syracuse, NJ, USA, 2001.

[35] T. Kohonen, Self-organized formation of topologically correct feature maps, Biol. Cybern. 43 (1) (1982) 59–69.

[36] D. Romero, A. Molina, Value co-creation and co-innovation: linking networked organisations and customer communities, in: L.M. Camarinha-Matos, I. Paraskakis, H. Afsarmanesh (Eds.), Leveraging Knowledge for Innovation in Collaborative Networks, IFIP Advances in Information and Communication Technology, vol. 307, Springer-Verlag, Berlin Heidelberg, 2009, pp. 401–412.

[37] P. Warnke, M. Weber, K.-H. Leitner, Transition pathways towards user-centric innovation, Int. J. Innov. Manage. 12 (3) (2008) 489–510.

[38] B. Bowonder, A. Dambal, S. Kumar, A. Shirodkar, Innovation strategies for creating competitive advantage, Res. Technol. Manage. 53 (3) (2010) 19–32.

[39] P. Kristensson, J. Matthing, N. Johansson, Key strategies for the successful involvement of customers in the co-creation of new technology-based services, Int. J. Serv. Ind. Manage. 19 (4) (2008) 474–491.

[40] M. Etgar, Co-production of services, in: R. Lusch, S. Vargo (Eds.), The Service Dominant Logic of Marketing, M.E. Sharpe Inc., Armonk, NY, 2006.

[41] I. Ng, The future of pricing and revenue models, J. Revenue Pricing Manage. 9 (3) (2010) 276–281.

[42] T. Brady, A. Davies, D. Gann, Can integrated solutions business models work in construction? Build. Res. Inf. 33 (6) (2005) 571–579.

[43] S. Vandermerwe, J. Rada, Servitization of business: adding value by adding services, Eur. Manage. J. 64 (1988) 314–324.

[44] S. Brax, A manufacturer becoming service provider – challenges and a paradox, Manag. Serv. Qual. 15 (2) (2005) 142–155.

[45] J. Ettlie, S. Rosenthal, Service innovation in manufacturing, J. Serv. Manage. 23 (3) (2012) 414–440.

[46] H. Gebauer, Identifying service strategies in product manufacturing companies by exploring environment-strategy configurations, Ind. Mark. Manage. 37 (3) (2008) 278–291.

[47] V. Velamuri, A. Neyer, K. Möslein, Hybrid value creation: understanding the value creating attributes, in: Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI) 2010 – Integration von Produkt und Dienstleistung – Hybride Wertschöpfung, Göttingen, Germany, 23–25 February, (2010), pp. 2003–2016.

[48] V. Shankar, L. Berry, T. Dotzel, A practical guide to combining products + services, Harv. Bus. Rev. 87 (11) (2009) 94–99.

[49] A. Field, Discovering Statistics Using SPSS., Sage Publications, Thousand Oaks, CA, 2005.

[50] A. Costello, J. Osborne, Best practices in exploratory factor analysis, Pract. Assess. Res. Eval. 10 (7) (2005) 1–9.

[51] J. Vesanto, E. Alhoniemi, Clustering of the self-organizing map, IEEE Trans. Neural Netw. 11 (3) (2000) 586–600.

[52] D. Weisman, D.A. Simov, Several remarks on the metric space of genetic codes, Int. J. Data Mining Bioinform. 6 (1) (2012) 17–26.

Giacomo di Tollo holds a PhD Europaeus from the University "G. D'Annunzio" in Pescara (I). He has served as PostDoc researcher in different Universities in Italy and France, and as visiting researcher in Poland, Norway, Denmark, Tunisia, and Estonia. He is currently PostDoc researcher at the Dipartimento di Economia, Università Ca' Foscari, Venezia, Italy.

Stoyan Tanev is an associate professor in the Department of Technology and Innovation at the University of Southern Denmark, Odense, Denmark, as well as adjunct professor in the Department of Systems and Computer Engineering at Carleton University in Ottawa, Canada, where he was previously a faculty member in the Technology Innovation Management Program. He has a MSc and a PhD in Physics jointly from the University Pierre and Marie Curie, Paris, France and the University of Sofia, Bulgaria, a PhD in Theology from the University of Sofia, Bulgaria, an ME in Technology Innovation Management from Carleton University, Canada, and an MA from the University of Sherbrooke, Canada. He has multidisciplinary research interests with a focus on the fields of technology entrepreneurship and innovation management, born global technology startups, business model design, business intelligence and value co-creation. Dr. Tanev is senior IEEE member, and he is a member of the editorial boards of the Technology Innovation Management Review and the International Journal of Actor-Network Theory and Technological Innovation.

Giacomo Liotta is associate professor in Supply Chain Management and IT at Aalborg University, Center for Industrial Production, in Copenhagen, Denmark. He previously served as assistant professor in the Department of Technology and Innovation at the University of Southern Denmark, and as contract researcher in the National Research Council of Italy, Institute of Industrial Technologies and Automation. He received the Ph.D in Economics and Management Engineering, Master of Management/MBA, M.Sc. and B.Sc. in Management and Industrial Engineering from the University of Rome "Tor Vergata", Italy. He is author of more than 30 papers published in international journals, books, and refereed conference proceedings. His current research interests include: design and management of sustainable supply chains; modeling and simulation of logistics, production, and service systems; product lifecycle management and value-added services.

Davide De March is a data scientist in Insurance, and previously collaborated as research fellow at Ca' Foscari University of Venice and as professor at University of Bozen. He obtained a Ph.D in applied statistics at the University of Florence where he cultivated his interests in computational statistics, machine learning and optimization. During his research career, he developed new statistical computational algorithms for the design of experiments in very high and sparse dimensional settings, He extensively approached real case studies with machine learning techniques, and with evolutionary optimisation methods and contributed to the development and application of new approaches, mainly based on neural network models, to several industrial studies. He is also a co-funder of Evosolutions, a statistical consulting company to transfer the research knowledge to the private sector.

Co-author Statement

Co-author statement in connection to the publication made by:	Names: Giacomo di Tollo

Concerning the publication:	Authors, title, journal: Giacomo di Tollo, Stoyan Tanev, Giacomo Liotta, Davide De March: Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. Computers in Industry, Vol 74, pag. 16-28 (2015).

The undersigned who is:

☐ Corresponding author
☒ Co-author

on the abovementioned publication, hereby confirms that Giacomo di Tollo has contributed to the work as stated below:

---

1. **Intellectual input:**

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

---

2. **Experimental results: (indicate contribution to individual figures, tables and supplementary data)**

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

3. Writing process:

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

Name of co-author (capital letters): GIACOMO LIOTTA

Date and signature of co-author: 09.06.2016

Co-author Statement

Co-author statement in connection to the publication made by:	Names: Giacomo di Tollo

Concerning the publication:	Authors, title, journal: Giacomo di Tollo, Stoyan Tanev, Giacomo Liotta, Davide De March: Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. Computers in Industry, Vol 74, pag. 16-28 (2015).

The undersigned who is:

☐ Corresponding author
☒ Co-author

on the abovementioned publication, hereby confirms that Giacomo di Tollo has contributed to the work as stated below:

---

1. Intellectual input:

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

---

2. Experimental results: (indicate contribution to individual figures, tables and supplementary data)

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

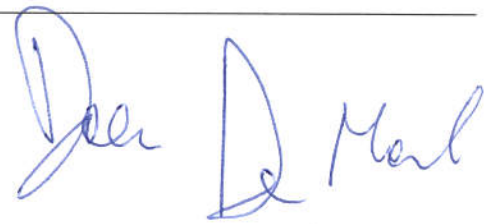3. **Writing process:**

☐ Less than 25%   ☒ 25%-50%   ☐ 51%-75%   ☐ 76%-100%

Comments:

Name of co-author (capital letters): DAVIDE DE MARCH

Date and signature of co-author: 09.06.2016

Co-author Statement

Co-author statement in connection to the publication made by:	Names: Giacomo di Tollo

Concerning the publication:	Authors, title, journal: Giacomo di Tollo, Stoyan Tanev, Giacomo Liotta, Davide De March: Using online textual data, principal component analysis and artificial neural networks to study business and innovation practices in technology-driven firms. Computers in Industry, Vol 74, pag. 16-28 (2015).

The undersigned who is:

☒ Corresponding author
☐ Co-author

on the abovementioned publication, hereby confirms that Giacomo di Tollo
has contributed to the work as stated below:

---

1. Intellectual input:

☐Less than 25%   ☒25%-50%   ☐51%-75%   ☐76%-100%

Comments:

---

2. Experimental results: (indicate contribution to individual figures, tables and supplementary data)

☐Less than 25%   ☒25%-50%   ☐51%-75%   ☐76%-100%

Comments:

3. Writing process:

☐Less than 25%   ☒25%-50%   ☐51%-75%   ☐76%-100%

Comments:

Name of co-author (capital letters): Stoyan Tanev

Date and signature of co-author: 09.06.2016