ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

Keskusteluaiheita
Discussion Papers
24 May 2012

No **1272**

The Role of Data and Knowledge in Firms' Service and Product Innovation

Heli Koski*

* ETLA – The Research Institute of the Finnish Economy, heli.koski@etla.fi

The author is grateful for TEKES (The Finnish Funding Agency for Technology and Innovation) and The Federation of Finnish Technology and Industries for their financial support, and Mika Pajarinen for his helpful comments.

ISSN 0781–6847

Contents

Abstract

The importance of data and different sources of knowledge in the development of new services and products, and further in the creation of new markets, has dramatically increased during the past few decades. This empirical study uses data from 531 Finnish firms to explore the determinants of generation of new data-based products and services. The empirical findings emphasize the role of a firm's absorptive capacity and its ICT competence in data-based innovation. It seems that generally a firm's external information sources play a more prominent role than internal information sources. Particularly customer involvement in innovation process positively relates to the production of new data-based products and services. The reported empirical findings further indicate that data-based product and service innovation tends to be rather strongly demand-driven.

Key words: Firm performance, innovation, data-based products and services, ICT

JEL: D22, L20, O31

Tiivistelmä

Raportoidussa tutkimuksessa analysoidaan 531 suomalaisesta yrityksestä koostuvan aineiston perusteella uusien tietoperusteisten tuotteiden ja palveluiden syntymiseen vaikuttavia tekijöitä. Yrityksen kyky hyödyntää sen ulkopuolella tuotettua tietoa ja osaamista (eli ns. absorptiivinen kapasiteetti) sekä yrityksen kompetenssi koskien tieto- ja viestintäteknologian hyödyntämistä liiketoiminnassa ovat kaksi keskeistä tekijää, jotka vaikuttavat tietoperusteisten innovaatioiden syntyyn. Yrityksen ulkoisilla tietolähteillä näyttäisi olevan enemmän merkitystä kuin sisäisillä tietolähteillä. Erityisesti yrityksen asiakkaiden osallistuminen innovaatioprosessiin edistää uusien tietoperusteisten tuotteiden ja palveluiden tuotantoa. Raportoitu empiirinen tutkimus yleisemminkin viittaa siihen, että tietoperusteiset tuote- ja palveluinnovaatiot syntyvät verrattain vahvasti kysyntälähtöisesti.

Asiasanat: Yritysten innovaatiotoiminta, tietoperusteiset tuotteet ja palvelut, tieto- ja viestintäteknologia

1 Introduction

The importance of data and different sources of knowledge in the development of new services and products, and further in the creation of new markets, has dramatically increased during the past few decades[1]. Public sector agencies form the largest single provider of data in the European Union. Also, policy makers have reacted to the growth possibilities hiding in the under-utilized data pools managed by the public sector agencies. For instance, European Commission aims at encouraging its member countries to promote utilization of data via research programs and infrastructure projects (e.g., INSPIRE focusing on an infrastructure for spatial information in Europe[2]) and policy actions of which most recent example is the December 2011 launch of an Open Data Strategy for Europe to facilitate the re-use of different sources of information.[3]

The utilization of vast data sources in a firm's regular business activities, and particularly in innovation activities, may become a key factor affecting firm performance and competitiveness in various industrial sectors. It seems credible that those companies that are able to detect and exploit data-based business opportunities first benefit from the first-mover advantage in the markets for new data-based products and services. Currently, very little is known, based on the systematic data analysis, about the origins or determinants of data-based innovation (e.g., which industrial sectors contribute most prominently to the creation of new data-based products and services).

Researchers in the fields of applied economics have investigated the importance of various internal and external sources of knowledge of a firm (e.g. own R&D, customers and business partners) for the firm's innovative performance but, so far, the reported empirical studies have not explicitly explored the role of data use in entrepreneurial innovation process. Closest to this study is the empirical work of Brynjolfsson et al. (2011) investigating how data-driven decision-making affect firm performance. Their study using survey data on the business practices and information technology investments of 179 large publicly traded firms indicates that more extensive use of data in a firm's business practices and decision-making relates positively to the firm's output and productivity.

This study addresses the question of firms' use of different types of data (i.e. spatial, meteorological, demographic, business and traffic data) in product and service innovation. We use a survey data concerning 531 Finnish firms' data use and innovation patterns collected in the autumn 2011, and combined with financial firm-level information obtained from Asiakastieto[4], to empirically explore the determinants of data-based product and service innovation. We cover not only innovation based on public sector information (PSI) but also new products and services for which the firm has used data obtained from other sources (i.e. self-production, other firms, free services).

---

1 See, e.g., McKinsey Global Institute's report "Big data: The next frontier for innovation, competition and productivity". June 2011. (http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation).

2 For further information concerning INSPIRE, see http://inspire.jrc.ec.europa.eu/.

3 More information concerning European Commission's digital agenda and Open Data Strategy for Europe is available on the following web page: http://ec.europa.eu/information_society/policy/psi/index_en.htm.

4 Asiakastieto is the leading provider of business and credit information in Finland.

This paper is organized as follows. Section 2 introduces the data and gives a brief overview of its key characteristics. Section 3 discusses potential determinants of data-based product and service innovation and introduces the explanatory variables used in the empirical estimations. Section 4 reports and discusses the estimation results of the econometric models. Section 5 concludes.

2 Data

In the autumn 2011, a web survey was conducted to collect data concerning the Finnish firms' use of data and particularly public sector information, and further its importance in the firms' creation of new products and/or services. The idea was to shed light on the firms' use of data (particularly PSI) and its economic impacts through the economy, without limiting the survey to certain sectors or firms that are well-known, based on case examples, to use certain data types (such as firms offering navigation solutions). Fonecta's[5] firm catalogue was used to identify the e-mail addresses of over 30,000 Finnish firms widely from different sectors of economy. An invitation and a web link to participate to the questionnaire were then sent to these companies. We obtained a response from 531 firms of which majority came from the service sector: 80 percent of the respondents were active in the service sector (see Table 1 for the distribution of sample firms across different sectors). The majority of the respondents firms – 93 percent of them – were either micro firms employing less than 10 people or small firms employing 10–49 people corresponding relatively well the size distribution of the Finnish firms in general.

We were particularly interested in the firms' use of five major data types[6] in their new products and services: spatial data, meteorological data, demographic data, business data and traffic data. Spatial data refers to "*any data with a direct or indirect reference to a specific location or geographical area*" (European Parliament and Council, 2007[7]).[8] Spatial data and socio-economic data form the most valuable databases held by public sector agencies. Potential for the creation of high-growth business via new products and services utilizing spatial data are particularly strong for geographical information systems (GIS), navigation and location-based services (e.g. information on the local attractions and events) and geomarketing (e.g. real estate consulting). The survey responses reveal that meteorological data is not only used in production control and planning in the sectors such as agriculture, forestry and construction but firms also develop new weather-data based products and services (e.g., heating systems utilizing local weather forecast information). The sample firms use demographic data typically in research (e.g., for generating new educational brochures and other educational material), and also in advertising and marketing services. Business data has various uses, for instance, in data mining products, firm-level analysis needed for providing financial services and industry-analysis for consulting firms in their strategic corporate decisions such as mergers and acquisitions. The respondents of our survey report that traffic data is utilized, for instance, in emission calculations and logistics services (e.g. escort vehicle services).

---

5 Fonecta is part of the European Directories Group, which has business operations in eight European Union countries. (http://www.fonecta.com)

6 The respondents were also given three open lines to report other types of data they have used in new products and/or services.

7 Directive 2007/2/EC of the European Parliament and Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE).

8 According to the prior studies, the most important categories of spatial data for re-users are topographic data, cadastral information (including address coordinates) and aerial photography (see. e.g., Fornefeld et al., 2009).

Table 1 The distribution of industrial sectors* of sample firms

Industrial sector	*Frequency*	*Percentage*
Agriculture, forestry and fishing	5	0.94
Manufacturing	67	12.62
Water supply; sewerage, waste management and remediation activities	5	0.94
Construction	29	5.46
Wholesale and retail trade; repair of motor vehicles and motorcycles	54	10.17
Transportation and storage	10	1.88
Accommodation and food service activities	7	1.32
Information and communication	80	15.07
Financial and insurance activities	12	2.26
Real estate activities	12	2.26
Professional, scientific and technical activities	214	38.41
Administrative and support service activities & Other service activities	31	5.84
Sector unknown	15	2.82

* Standard Industrial Classification 2008.

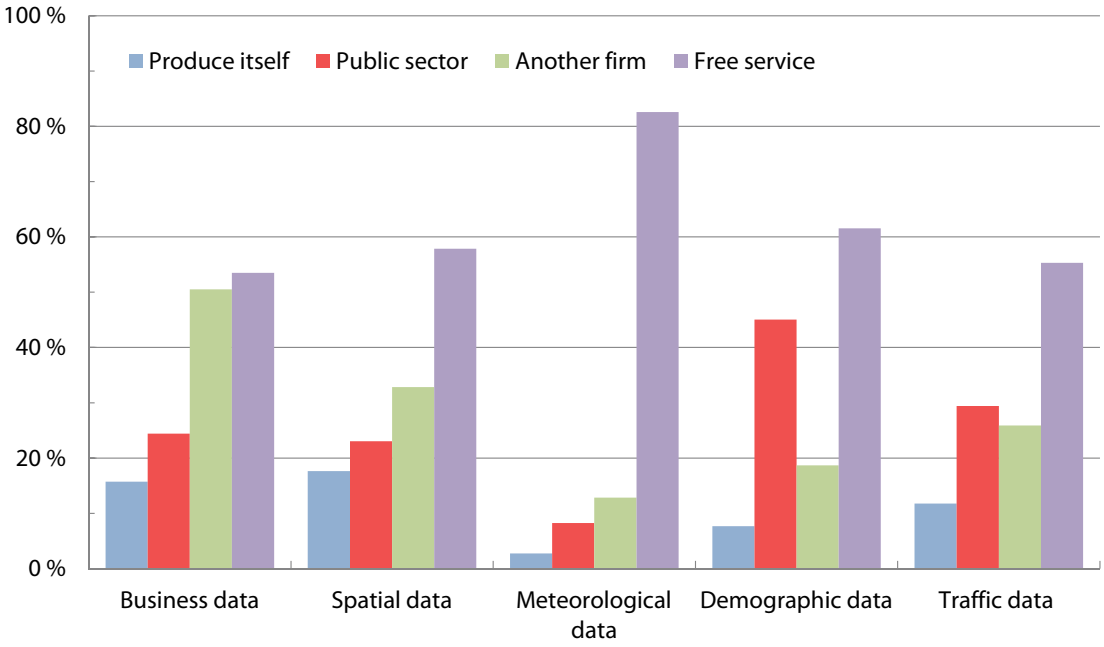
Two-thirds of the respondents reported that their company had during the past three years used one or multiple forms of PSI in their new products and/or services. Business and spatial data were most commonly used forms of PSI: 57 percent of firms had used business data and 39 percent spatial data in their new products or services. Meteorological data was applied by 20 percent, demographic data by 17 percent and traffic data by 16 percent of the respondent firms.

We also asked from where a firm obtains data it uses. The four non-exclusionary options for the data sources were: i) self-production, ii) public sector agency, iii) other firm, and iv) free service. Figure 1 shows that the most commonly used data sources were generally those that provided the data free of charge. This concerns particularly meteorological data: over 80 percent of the users of meteorological data reported that they obtain weather data from free services, while very few firms (i.e. only 3 percent of sample firms) produced such data by themselves. Business data and spatial data were relatively often also bought from other companies: half of the users of business data and almost one third of the users of spatial data obtained it from other firms. Also, both business and spatial data were most commonly, among all data types, produced by the firms themselves. Instead, the public sector was for a relatively large share of firms (i.e 45 percent of user firms) the source of demographic data.

We also inquired the use and relative importance of different internal (i.e. R&D, production, marketing, management) and external (i.e. customers, competitors, suppliers and subcontractors, research institutes and universities and public sector) sources of information in a firm's innovation process generally. The descriptive statistics of the importance of a firm's internal sources of information tells that about half of the respondents regarded R&D, marketing and management and 38 percent production unit either important or very important for the firm's innovation process.

Customers seem to be the most important external source of information for sample firms' innovation process: 62 percent of the respondents reported that customers form either im-

Figure 1 Sources of data firms have used in their new products and/or services within the past years



portant or very important source of information for their innovation process. This descriptive finding is in line with the strategic management and industrial organization literature addressing the importance of knowledge flows from a firm's customers to the firm's innovative performance. Suppliers and subcontractors were regarded as an important or very important source of information for the firm's innovation by 38 percent of respondents and competitors by one third of the respondents. Also, universities and research organizations (i.e. private or public research institutes) are relatively important for data-based innovation: about 28 percent of the respondents reported that they are either important or very important. Though the public sector agencies hold massive data pools with a great potential for re-use, they provide an important source of information for innovation only for a relatively small portion of firms. Only 13 percent of the respondents evaluated that public sector agencies or institutes form an important or very important source of information for their innovation activities.

3 Determinants of data-based service and product innovation

This paper approaches the question concerning the determinants of firms' data-based product and service innovation empirically by testing the importance of various factors[9] that the economics, industrial organization and strategic management literature has addressed in the context of entrepreneurial innovation. Generally, technological development takes place and innovation are created through the knowledge production processes combining a firm's in-house R&D activities – directly generating new products and services – and knowledge ab-

[9] See Annex for the description of variables used in the estimations.

sorbed from other internal (e.g. marketing, management) and external (e.g. customers, business partners, competitors, universities and research institutes) sources (see, e.g., Belderbos et al., 2004). A firm's own R&D investments and accumulated intellectual capital facilitate the firm's capacity to utilize external information and knowledge production or its "ability to recognize the value of new information, assimilate it, and apply it to commercial ends", i.e. its absorptive capacity (Cohen and Levinthal, 1990)[10]. We measure the role of a firm's own R&D activities in its innovation process by the dummy variable OWN_RD that gets value 1 if the respondent reports that the firm's own R&D section is either important or very important source of information for the firm's innovation process, and 0 otherwise.

The influence of a firm's both internal and external sources of knowledge on entrepreneurial innovation behavior has gained substantial attention in the management and economic literature (see, e.g., Cohen and Levinthal, 1990; Antonelli et al., 2000; Berry, 2006; Vega-Jurado et al., 2008). The *open innovation* concept launched by Chesbrough (2003, 2006) further emphasizes the importance of firms' use of both intra-firm and external flows of knowledge for innovation: "*Open innovation is the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively. [This paradigm] assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as they look to advance their technology*." (Chesbourgh, 2006).

The resource based view of the management literature argues that the organizational resources and capabilities of a firm's internal functional units and further the firm's ability to utilize them affect product innovation (Verona, 1999). The importance of internal information sources – other than R&D activities – in a firm's innovation process are measured the by the four dummy variables, PRODUCTION, MARKETING, MANAGEMENT, MULTI-PLANT and getting value 1, if the respondent reports that the firm's production unit, marketing unit or management or other firm(s) in the same corporate group – respectively – is used as a source of information for the firm's innovation process *generally*, and 0 otherwise. Whether these different internal information sources matter for the creation of new data-based products and services is an empirical question we aim at investigating.

Due to complexity of innovation process and knowledge used in it, particularly small firms' internal competencies alone may not be sufficient to produce innovative products and service solutions but external sources of knowledge are required for successful innovation. The strategic management and industrial organization literature stresses the important role of a firm's *customers* for its innovation performance, particularly when new services or products are complex and require adaptations by the customers for their use (see, e.g., Tether, 2002; Thomke and von Hippel, 2002). A firm's collaboration with its customers in development and testing enables creation of new products and services that better fit to the needs of the firm's customers and, for the firm, reduce risks related to the commercialization of an innovation. Also, other firms (i.e. competitors, suppliers and subcontractors), universities and research institutes as well as public sector agencies may convey technological know-how and other useful information for a firm's innovation process, e.g., via strategic R&D partnerships.

[10] Various empirical studies suggest that a firm's investments in R&D promote its productivity through increase in the firm's absorptive capacity (see, e.g., Eaton, Gutierrez, and Kortum, 1998, and Griffith, Redding, and Van Reenen, 2004).

The importance of external information sources for a firm's innovation process are measured by the five dummy variables, CUSTOMER, COMPETITOR, SUPPLIER, RESEARCH, and PUBLIC, getting value 1, if the respondent reports that the firm's customers, competitors, suppliers/subcontractors, universities and/or other research institutes (i.e. public or private research organizations or consulting companies) or public sector agencies – respectively – are used as a source of information for the firm's innovation process, and 0 otherwise.

Yet, another critical factor that may prominently influence on how successfully a firm utilizes data for its innovation – as well as both its internal and external information sources – is the firm's technological competence, particularly the firm's ability to accrue benefits from the use of ICT. The strategic management literature emphasizes the role of ICT as an enabler of complementary organizational changes – or that full benefits from the use of ICT can only be obtained via organizational re-structuring – further facilitating a firm's access to complementary external competencies and use of both external and internal information flows to generate new knowledge-based services and products (see, e.g. Bresnahan, Brynjolfsson, and Hitt, 2002; Brynjolfsson and Hitt, 2000). It seems credible that the costs of generation of new data-based products and services are lower, and further the benefits from the new data-based products and services for a firm higher, for the information technology intensive firms with better capabilities and experience to generally utilize ICT.

We measure the order of magnitude of a firm's utilization of ICT by the variable ICT_use that is a sum of six likert-scale[11] variables measuring the importance of different forms of ICT for a firm's business: i) Internet, ii) E-commerce, iii) Customer Relationship Management (CRM), iv) Business Intelligence (BI), v) Supply Chain Management (SCM), and vi) Enterprise Resource Planning (ERP). These information and communication technologies enable firms to acquire and process large data volumes, and to further integrate data-based activities to their regular business operations. Thus, greater ICT intensiveness credibly relates to a better capability and competence to collect, manage and process data, and also potentially contributes to the generation of new data-based products and services.

In addition to a firm's utilization of ICT, the absorptive capacity arising from the firm's intangible resources more generally is likely to affect its innovation performance. The variable INTANGIBLE – that is the order of magnitude of a firm's intangible assets (i.e. intangible rights, other capitalized expenses, and advances paid) in relation to the order of magnitude of its all assets, both tangible and intangible – controls for the firm's absorptive capacity arising from its intellectual capital.

It is an interesting question as such whether there are industrial sectors – and if yes, which are those sectors – that are more often than others the source of different data-based innovation. The industrial sectors are measured by the dummy variables using Standard Industrial Classification (SIC) codes at the 2 digit level (see Table 1). Furthermore, we control for a set of factors that prior empirical studies have suggested to matter for a firm's innovation performance: firm's size, age, profitability, and location (see, e.g., Cohen and Klepper, 1996; Martinez-Ros and Labeage, 2002; Koski et al., 2012). A firm's size is controlled by a set of dummy vari-

[11] The variables get value 1 for "not important at all", 2 for "slightly important", 3 for "quite important", 4 for "very important", 5 for "essentially important" for a firm's business. If all six forms of ICT use are not important at all for a firm's business, the variable gets value 6 and if all six forms of ICT use are of essential importance, the variable gets value 30, otherwise some value between the lower and upper bound.

ables SMALL (i.e. firms with 10–49 employees), MEDIUM (i.e. firms with 50–249 employees) and LARGE (i.e. firms with at least 250 employees), while the group of firms with less than 10 employees is used as a reference group in the estimations. The variable AGE measures the number of years elapsed since the establishment of the firm. The variable PROFITABILITY is, quite typically used measure of a firm's profitability, return on its total assets. The estimations also include 17 province dummies to capture inter-regional variation in the generation of data-based products.

4 Empirical analysis

The reported empirical part of the analysis aims at detecting what are the determinants of new data-based products and services via the estimations of a set of standard probit models[12] with the following dependent variables: i) INNO_DATA, the dummy variable that gets value 1 if a firm has during the past three years used spatial, meteorological, demographic, business, traffic or other type of data in their new products and/or services, and 0 otherwise ii) five separate dummy variables (INNO_SPATIAL, INNO_METEO, INNO_DEMOG, INNO_BUSIN and INNO_TRAFFIC) that get value 1 if a firm has during the past three years used – respectively – spatial, meteorological, demographic, business or traffic data – in their new products and/or services, and 0 otherwise. The idea is first, by using the dependent variable INNO_DATA, to explore how different internal and external factors of a firm affect data-based innovation production generally. Then, the individual dummies for innovation based on five different data types are used as dependent variables for detecting the origins of different data based products and services.

As the used dependent variables describing a firm's introduction of data-based product and service innovation may get positive values only for firms having innovation activities, there is clearly a potential sample selection problem that the empirical analysis needs to take into account. We therefore estimated the equations with the above defined dependent variables jointly, using the Heckman's probit model, with the selection equation of which dependent variable was a dummy variable getting value 1 if firm reported that it has innovative activities[13] generally, and 0 otherwise. Correlation coefficients between the equations were not statistically significant, and the Wald-test failed to reject the null hypothesis of the independence of equations. As the sample selection seems not to be a problem in this respect, we used regular probit models in the reported final estimations.

Table 2 reports the estimation results of the six probit models for the data-based innovation. The explanatory power of the estimated models is relatively high. The used independent variables capture generally between 75–85 percent of variation in the dependent variables though the estimated model for the creation of business data based innovation succeeds in correctly classifying about 72 percent of the dependent variables.

[12] We also estimated the instrumental variables probit model, with the Wald –test of endogeneity, to explore potential endogeneity of the variable ICT_USE. The Wald test statistic was not statistically significant indicating that we cannot reject the null hypothesis that the variable ICT_USE is not endogenous. Therefore, as our data indicates that the ICT_USE variable is exogenous determinant of innovation, we report the estimation results of a regular probit regression.

[13] In the questionnaire, innovation activities were defined to be all actions that were producing or aimed at producing new or improved products, services or processes.

Table 2 The estimation results of the probit model for data-based innovation

	Dependent variable INNO_					
Variable	*DATA*	*SPATIAL*	*METEO*	*DEMOG*	*BUSIN*	*TRAFFIC*
OWN_RD	-0.14	-0.12	0.02	-0.05	-0.59*	-0.29
	(0.33)	(0.33)	(0.39)	(0.35)	(0.34)	(0.35)
PRODUCTION	-0.12	0.61**	0.23	0.17	-0.06	-0.47
	(0.30)	(0.30)	(0.36)	(0.35)	(0.28)	(0.33)
MARKETING	-0.08	-0.18	-0.32	-0.50	0.40	1.12***
	(0.32)	(0.36)	(0.43)	(0.39)	(0.34)	(0.47)
MANAGEMENT	0.24	0.13	0.79**	0.54	0.13	-0.58**
	(0.32)	(0.35)	(0.37)	(0.44)	(0.30)	(0.30)
MULTI-PLANT	0.25	-0.23	-0.32	0.14	0.35	0.17
	(0.20)	(0.19)	(0.22)	(0.21)	(0.20)	(0.22)
CUSTOMER	2.35***	1.85***	0.10	0.32	1.49***	1.21
	(0.66)	(0.66)	(0.82)	(0.71)	(0.59)	(0.77)
COMPETITOR	-0.11	-0.05	0.48	-0.22	-0.12	-0.23
	(0.46)	(0.40)	(0.53)	(0.46)	(0.41)	(0.42)
SUPPLIER	-0.47	-1.04***	0.34	-0.38	0.20	0.56
	(0.43)	(0.38)	(0.61)	(0.41)	(0.39)	(0.48)
RESEARCH	0.87**	0.51	0.18	0.82*	0.56	0.03
	(0.39)	(0.42)	(0.45)	(0.49)	(0.39)	(0.44)
PUBLIC	-0.14	0.23	0.16	0.27	0.02	0.50*
	(0.24)	(0.22)	(0.25)	(0.24)	(0.23)	(0.27)
ICT_USE	0.03**	0.03*	0.01	0.04***	0.03**	0.00
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)
INTANGIBLE	0.72**	0.33	0.19	-0.44	0.25	-0.53
	(0.32)	(0.28)	(0.33)	(0.37)	(0.30)	(0.38)
SMALL	-0.35	-0.20	0.02	-0.11	-0.30	-0.34
	(0.22)	(0.22)	(0.23)	(0.22)	(0.22)	(0.25)
MEDIUM	-0.51	-0.85**	-0.58	-0.80*	-0.14	-0.80**
	(0.43)	(0.42)	(0.52)	(0.45)	(0.40)	(0.42)
LARGE		1.08	0.73	1.08	0.52	1.15
	(omitted)	(0.75)	(0.80)	(0.74)	(0.80)	(0.70)
AGE	0.01	0.00	0.00	0.02*	0.01	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
PROFITABILITY	-0.01	0.39**	0.52**	0.31	-0.04	0.25
	(0.17)	(0.19)	(0.23)	(0.19)	(0.19)	(0.19)
Agriculture, forestry and fishing	(omitted)	(omitted)	0.80			
			(0.85)	(omitted)	(omitted)	(omitted)
Water supply	1.49***	0.55		1.23	1.71***	0.26
	(0.50)	(0.77)	(omitted)	(0.76)	(0.55)	(0.62)
Construction	0.31	0.98***	1.37***	0.37	0.18	0.71
	(0.39)	(0.38)	(0.45)	(0.48)	(0.37)	(0.45)
Wholesale and retail trade	0.23	0.39	0.56	-0.08	0.41	0.05
	(0.34)	(0.36)	(0.38)	(0.42)	(0.34)	(0.39)
Transportation and storage	0.39	0.08	1.10	0.92	-0.42	1.07
	(0.63)	(0.55)	(0.68)	(0.73)	(0.63)	(0.60)
Accommodation and food service	(omitted)	0.11	0.59	1.46		(omitted)
		(0.79)	(1.12)	(1.29)	(omitted)	
Information and communication	0.46	0.29	0.23	0.52	0.44	0.26
	(0.30)	(0.31)	(0.34)	(0.35)	(0.30)	(0.38)
Financial and insurance activities	2.40***	0.48	0.82		1.10	(omitted)
	(0.77)	(0.72)	(0.72)	(omitted)	(0.75)	
Real estate activities	1.64***	2.30***	0.86	0.85	1.74***	0.65
	(0.47)	(0.61)	(0.84)	(0.78)	(0.51)	(0.73)
Professional, scientific and technical activities	0.34	0.47*	0.52*	0.68**	0.59**	0.24
	(0.26)	(0.27)	(0.30)	(0.31)	(0.26)	(0.32)
Administrative and support service	0.63	1.22***	0.97**	0.27	0.46	0.27
	(0.45)	(0.38)	(0.43)	(0.48)	(0.44)	(0.46)
Constant	-3.21	-3.14	-3.09	-3.68	-3.38	-3.22
	(0.71)	(0.69)	(0.77)	(0.79)	(0.73)	(0.75)
+ Locational dummies						
Observations	**357**	**367**	**343**	**361**	**357**	**344**
Correctly classified (%)	**75.07**	**74.11**	**81.05**	**80.42**	**71.75**	**86.05**
Log pseudo likelihood	**174.92**	**184.12**	**134.54**	**134.51**	**194.10**	**126.37**

The robust standard errors are reported in the parentheses. Significance levels are reported on superscripts, where *** denotes significance level of 1%, ** significance level of 5% and * significance level of 10%.

The estimated coefficients for industry dummies indicate that there are certain sectors that appear more often than others as a source of different data-based innovation. The estimated coefficients of the dummy variables for professional, scientific and technical activities, real estate activities and construction are positive and statistically significant in three of the five different data based innovation equations suggesting that in these sectors different data are used more broadly for innovation than in others (see Table 3).

The estimation results indicate that there are also clear sector-specific differences in the creation of new data-based products and services. Spatial and weather data based innovation are more likely originating from firms functioning in construction, professional, scientific and technical activities, and administrative and support service activities. Spatial data based innovation also tend to take place in the firms active in real estate activities. The demographic data based new services and products tend to appear more often in the firms involved in the professional, scientific and technical activities. Three sectors appear as statistically significant origins of business data innovation: water supply, real estate activities and professional, scientific and technical activities. Traffic data based new products and services are, quite logically, more likely generated in firms functioning in transportation and storage.

The estimation results explaining variation in the variable INNO_DATA indicate that, generally, for the generation of new data-based products/services both certain internal and external factors of a firm matter. The firms with relatively higher order of magnitude of intangible assets tend to create new data-based products and/or services more often than other firms. The estimated coefficient of the intangible assets variable is not, however, statistically significant in any of the estimated equation explaining variation in different data based innovation. This hints that firms that are involved in data-based product or service innovation have generally higher intangible asset base than other companies but the differences among data-based innovation producers concerning their intangible asset base are not statistically significant.

The estimations also suggest, quite surprisingly, that the firm's own R&D does not play a significant role in the creation of data-based products and services. However, when the models are estimated without the dummies for the use of various internal and external sources of information in innovation, the variable OWN_RD gets positive and statistically significant co-

Table 3 Sector-specific origins of data-based innovation

Sector (statistically significant origin)	*Product and service innovation using*				
	Spatial data	*Weather data*	*Demographic data*	*Business data*	*Traffic data*
Professional, scientific and technical activities	x	x	x		
Real estate activities	x	x		x	
Construction	x	x		x	
Administrative and support service activities	x	x			
Transportation and storage				x	
Water supply					x

efficient in all of the estimated models.[14] Given these empirical results, we can't conclude that a firm's own R&D is not important for data-based innovation but rather it seems credible it positively relates to the use of certain important internal and external sources of information, and due to this correlation structure, does not appear statistically significant in the final estimations. Another potential explanation is that, when the firm's use of various internal and external information sources are controlled for, the firm's own R&D as such does not matter but rather whether it uses certain important information sources for its innovation activities.

It seems that the creation of different data based products benefits from different internal information sources of a firm. The involvement of a firm's production unit to its innovation process seems to enhance spatial data based innovation. Our estimation results further indicate that firms using internal information provided by the management of a firm are more often than otherwise the source of weather-based product/service innovation. The variable MARKETING positively and significantly relates to the creation of traffic data based new products indicating that those firms that collaborate in innovation with their marketing section are more often than otherwise the source of traffic data based product/service innovation. A firm's market section holds information on the market and customer needs, and it seems credible that it can well contribute to the development of, e.g., new data-based transportation services.

The ICT competence of a firm seem to matter for the production of data-based innovation generally, and particularly for those product/service innovation comprising spatial, demographic and business data, while the estimated coefficient of the variable ICT_USE is also positive but not statistically significant for weather and traffic data based innovations.

Among the external information sources, it seems that a firm's customers provide a very important source of information for data-based product/service innovation. The estimated coefficient of the variable CUSTOMER is statistically significant in the estimated equation for data-based innovation generally but it appears also as a highly significant explanatory variable for the creation spatial, demographic and business data based innovation. The empirical findings emphasizing the role of customers and customer-related information (such as in the case of traffic data based innovation) in firm's innovation production are rather interesting. It seems that many new data-based products and services origin in close collaboration with a firm's customers.

The statistically significant estimated coefficient of the variable RESEARCH in the equation explaining variation in the variable INNO_DATA suggests that another important external source of information for the generation of data-based products are universities, other research organizations and/or consultant firms. However, if we look at the determinants of five different data-based innovation, the variable RESEARCH does not get a statistically significant coefficient in the individual equations, except that it appears as weakly statistically significant (at p=0.10) determinant for the creation of demographic data based innovation. It thus seems that the firms producing data-based innovation generally are more often in collaboration with different research organizations than other firms.

The micro-sized firms (i.e. firms that employed less than 10 people) were used as a reference group for the size dummy variables in the estimations. The dummy variable for medium

[14] These estimations results are not reported in the paper but are available from the author.

firm size negatively and statistically significantly relates to the generation of new spatial, demographic and traffic data based products and services, though the relationship with demographic data base innovation is rather weak (p=0.07). This empirical finding is consistent with some earlier empirical studies suggesting that the relationship between firm size and innovation is U-shaped such that innovation takes place less likely in medium-sized companies than in small and large firms (see, e.g., Bound et al., 1984). The estimated positive and statistically significant coefficient of the age variable in the probit regression explaining the creation of demographic and traffic data based innovation indicates that older firms are more likely to provide certain data based products and services than other firms. Further, we find that the firms that tend to create more often spatial and weather data based innovations seem to be more profitable than others.

We further expanded the empirical exploration by using the count of different data types a firm has used (i.e. the variable DATA_COUNT) as the dependent variable, and i) explaining its variation with the same set of explanatory variables as in the probit equations, and ii) by replacing the internal and external sources of information by the counts or number of internal and external sources of information firm uses in its innovation process (i.e. the variables INTER_COUNT and EXTER_COUNT). The idea here is to explore which factors affect the scope of different data use in innovative products and services. Table 4 reports the estimation results of the Poisson model for data-based product and service innovation.

The empirical results of the Poisson model confirm that use of a greater variety of external sources of information, and particularly firm's customers as a source of information, contributes also to the scope of different data used in product and/or service innovation. Also, the firms with greater ICT competence tend to generate innovation based on a larger variety of data types. Interestingly, the order of magnitude of a firm's intangible assets does not seem to matter for the scope of data use in innovation.

5 Conclusions

This empirical study represents one of the first systematic analyses of the economic impacts of a firms' utilization of data in their business activities. The reported empirical exploration uses a survey data collected from 531 Finnish firms, and combined with the firm-level financial data, to explore the determinants of generation of new data-based products and services. The data suggest that there are certain sectors (i.e. professional, scientific and technical activities, real estate activities and construction) that appear more often than others as a source of different data based innovation, but it indicates that there are also clear sector-specific differences in the creation of new data-based products and services.

The empirical findings emphasize the role of a firm's absorptive capacity and its technological competence in the generation of new data-based products and services. Firms that are involved in data-based product or service innovation have generally higher intangible asset base, and thus absorptive capacity, than other companies but it seems that the producers of different types of data-based innovations do not differ significantly in terms of their intangible assets from the sample companies. Neither a firm's scope of data use in new products and services is affected by the order of magnitude of its intangible assets. Furthermore, ICT intensive firms tend to be the producers of both data-based innovation generally and certain types of

Table 4 The estimation results of the Poisson model for data-based innovation

	Dependent variable			
Variable	*DATA_COUNT*		*DATA_COUNT*	
OWN_RD	-0.16	(0.23)	-0.06	(0.22)
PRODUCTION	0.13	(0.21)		
MARKETING	0.05	(0.23)		
MANAGEMENT	0.17	(0.24)		
MULTI-PLANT	0.03	(0.13)		
INTER_COUNT			0.12*	(0.08)
CUSTOMER	1.60**	(0.66)		
COMPETITOR	-0.09	(0.28)		
SUPPLIER	-0.13	(0.23)		
RESEARCH	0.36	(0.32)		
PUBLIC	0.18	(0.14)		
EXTER_COUNT			0.27***	(0.07)
ICT_USE	0.02**	(0.01)	0.02**	(0.01)
INTANGIBLE	0.00	(0.18)	-0.05	(0.17)
SMALL	-0.13	(0.16)	-0.12	(0.16)
MEDIUM	-0.44	(0.28)	-0.47	(0.27)
LARGE	0.75**	(0.38)	0.69	(0.36)
AGE	0.01	(0.01)	0.01	(0.01)
PROFITABILITY	0.24***	(0.09)	0.23***	(0.09)
Agriculture, forestry and fishing	0.53***	(0.22)	0.48*	(0.22)
Water supply	0.50*	(0.29)	0.58**	(0.30)
Construction	0.62**	(0.26)	0.63**	(0.26)
Wholesale and retail trade	0.30	(0.28)	0.35	(0.28)
Transportation and storage	0.46	(0.39)	0.39	(0.42)
Accommodation and food service	0.55	(0.50)	0.53	(0.52)
Information and communication	0.33	(0.23)	0.36	(0.22)
Financial and insurance activities	0.51	(0.38)	0.36	(0.32)
Real estate activities	1.04**	(0.31)	1.11***	(0.31)
Professional, scientific and technical activities	0.50*	(0.21)	0.55***	(0.21)
Administrative and support service	0.54*	(0.25)	0.55**	(0.24)
Constant	-2.75	(0.65)	-2.27	(0.63)
+ Locational dummies				
Observations	**370**		**370**	
Log pseudo likelihood	**-536.70**		**-542.51**	

The robust standard errors are reported in the parentheses. Significance levels are reported on superscripts, where *** denotes significance level of 1%, ** significance level of 5% and * significance level of 10%.

data-based innovation (i.e. spatial, demographic and business data based new products and/or services). The scope of data use in product and service innovation is also higher for the ICT intensive firms.

It seems that generally a firm's external information sources play a more prominent role than a firm's internal information sources in the generation of data-based innovation. Customer involvement in innovation process positively relates to the production of new data-based products and services, and facilitates particularly spatial and business data based innovation. This finding is in line with the previous economic and management literature stressing the importance of a firm's engagement with customers, in addition to its technological competence, to the firm's innovation performance (see, e.g., von Hippel, 1998; Danneels, 2002; Fabrizio and Thomas, 2012). Furthermore, the data-based products and services are generated, by and large, in the Internet-based environments that enable firms to more efficiently engage with a large number of customers, possibly also real-time getting feedback fast and frequently (see, e.g., Sawhney, et al., 2005).

The reported empirical findings further indicate that data-based product and service innovation tend to be rather strongly *demand-driven*. Our survey data further indicates that over 50 percent of the respondents have developed new data-based products for their own use, i.e. there are also substantial *user-driven* innovation based on data taking place. One interesting question for our future research – that the reported empirical work cannot provide answer to – is to better understand the relative roles of user-driven and demand-driven innovation activities in the generation of new data-based products and services.

References

Antonelli, C., Marchionatti, R. and Usai, S. (2000): Productivity and external knowledge: The Italian case. Mimeo July 2000.

Belderbos, R., Carree, M. and Lokshin, B. (2004). Cooperative R&D and firm performance. *Research Policy* 33, 1477–1492.

Berry, H. (2006). Leaders, laggards, and the pursuit of foreign knowledge. *Strategic Management Journal* 27, 151–168.

Bound, J., Cummins C., Griliches, Z. Hall, B. H. and Jaffe, A. (1984). Who does R&D and who patents?, in Z. Griliches (ed.), *R&D, Patents, and Productivity*, University of Chicago Press. Chicago, USA.

Bresnahan, T. F., Brynjolfsson, E. and Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117, 339–376.

Brynjolfsson, E. and Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *The Journal of Economic Perspectives*, 14, 23–48.

Brynjolfsson, E., Hitt, L. M. and Kim, H. H. (2011). Strength in Numbers: How does data-driven decision-making affect firm performance? Working Paper Available at SSRN: http://ssrn.com/abstract=1819486.

Chesbrough, H. (2003). *Open innovation: The New Imperative for Creating and Profiting from Technology.* Harvard Business School Press. Boston.

Chesbrough, H. (2006). *Open Business Models*. Harvard Business School Press, Boston.

Cohen, W. M. and Klepper, S. (1996). A reprise of size and R&D. *Economic Journal* 106, 925–951.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128–152.

Danneels, E. (2002). The dynamics of product innovation and firm competencies. *Strategic Management Journal* 23, 1095–1121.

Eaton, J., Gutierrez, E. and Kortum, S. (1998). European technology policy. *Economic Policy*, 27, 403–438.

Fabrizio, K. R. and Thomas, L. G. (2012). The impact of local demand on innovation in a global industry. *Strategic Management Journal* 33, 42–64.

Fornefeld, M., Boele-Keimer, G., Recher, S. and Fanning, M. (2009). Assessment of the re-use of public sector information (PSI) in the geographical information, meteorological information and legal information services. Micus Management Consulting.

Griffith, R., Redding, S. and van Reenen, J. (2004). Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *Review of Economics and Statistics*, 86, 883–895.

von Hippel, E. (1998). *The Sources of Innovation.* Oxford University Press: Oxford, UK.

Koski, H., Marengo, L. and Mäkinen, I. (2012). Managerial practices, performance and innovativeness: Some evidence from Finnish manufacturing. Forthcoming in *International Journal of Technology Management*.

Martinez-Ros, E. and Labeage, J. M. (2002). The relationship between firm size and innovative activity: A Double decision approach and an application to Spanish manufacturing firms. *Economics of Technological Innovation* 11, 35–50.

Sawhney, M., Verona, G. and Prandelli, E. (2005). Collaborating to create: the internet as a platform for customer engagement in product innovation. *Journal of Interactive Marketing* 19, 4–17.

Tether, B. (2002). Who co-operates for innovation, and why, an empirical analysis. *Research Policy* 31, 947–967.

Thomke, S. and von Hippel, E. (2002). Customers as innovators, A new way to create value. *Harvard Business Review* 80(4), 73–81.

Vega-Jurado, J., Gutierrez-Gracia, A., Fernandez-de-Lucio, I. and Manjarres-Henriquez, L. (2008). The effect of internal and external factors on firms' product innovation. *Research Policy* 37, 616–32.

Verona, G. (1999). A resource-based view of product development. *Academy of Management Review* 24, 132–142.

Annex Description of variables

Description of variable	*Variable name*	*Mean (S.D.)*
Dependent variables:		
Dummy variable that gets value 1 if a firm has during the past three years used spatial, meteorological, demographic, business, traffic or other data in new products and/or services, and 0 otherwise.	INNO_DATA	0.67 (0.47)
Dummy variables that get value 1 if a firm has during the past three years used – respectively – spatial, meteorological, demographic, business or traffic data in new products and/or services, and 0 otherwise.	INNO_SPATIAL INNO_METEO INNO_DEMOG INNO_BUSIN INNO_TRAFFIC	0.39 (0.49) 0.20 (0.40) 0.17 (0.37) 0.57 (0.50) 0.16 (0.37)
Number of data types a firm has during the past three years used in new products and/or services, and 0 otherwise.	DATA_COUNT	1.49 (1.46)
Explanatory variables:		
Internal information sources: Dummy variables that get value 1 if firm's own R&D, production, marketing, management and other firm(s) in the same corporate group – respectively – is used as a source of information for the firm's innovation process, and 0 otherwise.	OWN_RD PRODUCTION MARKETING MANAGEMENT MULTI-PLANT	0.71 (0.46) 0.68 (0.47) 0.72 (0.45) 0.72 (0.45) 0.24 (0.43)
Number of internal information sources a firm has used as a source of information in its innovation process.	INTER_COUNT	2.62 (1.29)
External information sources: Dummy variables that get value 1 if firm's customers, competitors, suppliers/subcontractors, research institutes (i.e. consulting companies, public or private research institutes or universities) or public sector agencies – respectively – is used as a source of information for the firm's innovation process, and 0 otherwise.	CUSTOMER COMPETITOR SUPPLIER RESEARCH PUBLIC	0.77 (0.42) 0.74 (0.42) 0.73 (0.44) 0.73 (0.45) 0.62 (0.49)
Number of internal information sources a firm has used as a source of information in its innovation process.	EXTER_COUNT	3.59 (2.04)
Variable that is a sum of six likert-scale variables measuring the importance of different forms of ICT for a firm's business: i) Internet, ii) E-commerce, iii) Customer Relationship Management (CRM), iv) Business Intelligence (BI), v) Supply Chain Management (SCM), and vi) Enterprise Resource Planning (ERP).	ICT_USE	15.10 (5.78)
The order of magnitude of a firm's intangible assets in relation to the order of magnitude of its total assets.	INTANGIBLE	0.13 (0.25)
Dummy variable that gets value 1 if a firm has 10–49 employees, and 0 otherwise.	SMALL	0.21 (0.40)
Dummy variable that gets value 1 if a firm has 50–249 employees, and 0 otherwise.	MEDIUM	0.04 (0.19)
Dummy variable that gets value 1 if a firm has at least 250 employees, and 0 otherwise.	LARGE	0.01 (0.09)
Firm's age.	AGE	16.46 (12.77)
Return on total assets.	PROFITABILITY	0.12 (0.37)
+ industry dummies		
+ locational dummies		

Aikaisemmin ilmestynyt ETLAn Keskusteluaiheita-sarjassa

Previously published in the ETLA Discussion Papers Series

No 1257 *Terttu Luukkonen – Matthias Deschryvere – Fabio Bertoni – Tuomo Nikulainen*, Importance of the Non-financial Value Added of Government and Independent Venture Capitalists. 2.9.2011. 28 p.

No 1258 *Ari Hyytinen – Mika Pajarinen – Pekka Ylä-Anttila*, Finpron vaikuttavuus – Finpron palveluiden käytön vaikutukset yritysten kansainvälistymiseen ja menestymiseen. 15.9.2011. 32 s.

No 1259 *Kari E.O. Alho*, How to Restore Sustainability of the Euro? 19.9.2011. 27 p.

No 1260 *Heli Koski*, Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth? 28.9.2011. 15 p.

No 1261 *Valeriy Naumov – Olli Martikainen*, Method for Throughput Maximization of Multiclass Networks with Flexible Servers. 13.12.2011. 19 p.

No 1262 *Valeriy Naumov – Olli Martikainen*, Optimal Resource Allocation in Multiclass Networks. 14.12.2011. 17 p.

No 1263 *Jari Hyvärinen*, Innovaatiotoiminta: Suomi globaalitaloudessa. 30.12.2011. 49 s.

No 1264 *Jari Hyvärinen*, Productivity: An International Comparison. 30.12.2011. 20 p.

No 1265 *Jukka Lassila – Tarmo Valkonen – Juha M. Alho*, Fiscal Sustainability and Policy Rules under Changing Demographic Forecasts. 21.12.2011. 32 p.

No 1266 *Reijo Mankinen – Olavi Rantala*, Ulkomaanliikenteen palveluiden arvonlisäverotuksen käyttöönoton vaikutukset laiva- ja lentoliikenteeseen. 11.1.2012. 29 s.

No 1267 *Ville Kaitila – Pekka Ylä-Anttila*, Investoinnit Suomessa. Kehitys ja kansainvälinen vertailu. 30.1.2012. 34 s.

No 1268 *Valeriy Naumov – Olli Martikainen*, Queueing Systems with Fractional Number of Servers. 21.3.2012. 11 p.

No 1269 *Kari E.O. Alho – Nuutti Nikula*, Asevelvollisuus ja työmarkkinat. Varusmiespalveluksen vaikutus koulutukseen, työllisyyteen ja palkkaan. 29.3.2012. 25 s.

No 1270 *Hannu Hernesniemi*, Merenkulun toimintaedellytykset, kilpailukyky ja julkisen talouden sopeuttamistoimet. Taustaselvitys valtiovarainministeriölle. 18.4.2012. 53 s.

No 1271 *Petri Böckerman – Mika Maliranta*, Outsourcing, Occupational Restructuring, and Employee Well-being. Is There a Silver Lining? 11.4.2012. 28 p.

Elinkeinoelämän Tutkimuslaitoksen julkaisemat "Keskusteluaiheita" ovat raportteja alustavista tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista. Tässä sarjassa julkaistuja monisteita on mahdollista ostaa Taloustieto Oy:stä kopiointi- ja toimituskuluja vastaavaan hintaan.

Papers in this series are reports on preliminary research results and on studies in progress. They are sold by Taloustieto Oy for a nominal fee covering copying and postage costs.

Julkaisut ovat ladattavissa pdf-muodossa osoitteessa: www.etla.fi/julkaisuhaku.php
Publications in pdf can be downloaded at www.etla.fi/eng/julkaisuhaku.php

ETLA
Elinkeinoelämän Tutkimuslaitos
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B
00120 Helsinki

Puh. 09-609 900
Fax 09-601 753
www.etla.fi
etunimi.sukunimi@etla.fi

ISSN 0781–6847