

Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs

ERIK LAKOMAA¹ AND JAN KALLBERG²

¹Department of Marketing and Strategy, Stockholm School of Economics, Stockholm 113 83, Sweden

²Cyber Security Research and Education Institute, Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, Richardson, TX 75083-0688, USA

Corresponding author: E. Lakomaa (erik.lakomaa@hhs.se)

ABSTRACT Public open data access has a direct impact on future IT entrepreneurs' perception of ability to execute their business plans. Using high quality (50%–98% response rate) survey data from 138 Swedish IT-entrepreneurs, we find that access to public open data is considered very important for many IT-startups; 43% find open data essential for the realization of their business plan and 82% claim that access would support and strengthen the business plan. The survey also indicates a significant interest in, and willingness to pay for, public sector information data from companies that do not intend to commercialize data themselves but intend to use it to support or test other business models. From the survey, it is possible to infer that the previous discourse on open data, viewing it as a means for government accountability or e-government, or as the foundation for the commercialization of public sector information data is too limited. Open data should instead be seen as an enabler of innovation outside these traditional sectors. This also indicates that the previously calculated societal values of open data might be underestimated.

INDEX TERMS Open data, entrepreneurs, innovation, public sector information, PSI, European Union, entrepreneurial incentive.

I. INTRODUCTION

Existing research on access to Public Sector Information (PSI) is predominantly related to e-government inquiries addressing aspects of democratic theory, voter participation, democratic deliberation, and open government in a broader context [1]. A secondary stream in the literature follows an older discourse pursuing administrative excellence and public cost reduction.

Governments accumulate massive amounts of digital information that serves the government to manage, allocate, administrate public programs, and collect tax revenue. In the early 2000s a second perspective became vocal – the democratic perspective [2]. Access to public sector information is there seen as vital for the vitality of the civic society.

These two earlier directions in the literature are either focused on the interests of the government or the citizens as a political constituency. We introduce a third direction – the importance of public sector Open Data, the free dissemination of PSI, as a foundation and catalyzer of innovation.

The conducted inquiry is providing evidence for a direct correlation between feasible internet startup company activity and access to Open Data.

For data to be considered “Open” it has to be free, downloadable, machine-readable, and structured without prior processing. In this paper Open Data, to avoid excessive repetition, also includes downloadable public sector raw data accessible at marginal cost (which in most cases is close to zero). The important factor is whether an unfunded potential entrepreneur with limited means is able to utilize the resource, or not.

II. OPEN DATA FOR TRANSPARENCY AND ACCOUNTABILITY

Further aims of government's online activity are to better serve citizens and bring government closer to the people. The Internet empowers people through transparency, e-voting, collecting opinions on public matters, and increasing political self-efficacy among citizens. Since knowledge of the future

is unknown, researchers create scenarios for the future of e-government [3]; the key question is whether the Open Data can increase accountability and transparency. The amount of information the government can publish is immense; however, publication itself does not automatically translate to trust and confidence from citizens.

Open Data can also be seen as a proxy for democracy and bring the government closer to the citizenry. According to its proponents, e-government increases efficiency in service offerings and saves money for the public sector [4].

III. THE PUBLIC SECTOR'S RESISTANCE TO OPEN DATA

The internal agenda of a bureaucracy is efficiency, the purpose of which is to process, expedite, and administrate as efficiently as possible. The state bureaucracy has limited, if any, incentive to usher forward or be a proponent of dissemination of Open Data public sector information.

There are significant differences between the Anglo-Saxon common law countries, such as USA, UK, Australia, and New Zealand, compared to civil law and centralized government countries, such as Denmark, France, and Sweden, when it comes to the degree of access to government public sector information [4]. The Anglo-Saxon countries are more open to the voluntarily dissemination of Open Data compared to civil law non-Anglo-Saxon countries.

In the last decade, challenges for public administration increased with growing complexity, uncertainty, and scarcity of resources [4]. Governments have also in general been slow to embrace the internet. The driving force has been the ability to cut public costs, especially in the latest years when austerity measures were implemented in several nations [4]. As the market has grown for information services, government entities started to see an opportunity to capitalize on their public sector information [5].

The Internet serves several purposes in e-government such as accepting online payments, taking service requests, tracking the development of a case, providing information, supplying forms, filing information, and providing online services that help the user determine eligibility for government services.

If the government can receive payments for service online, it can also retail the data as an information service. The Swedish cadastral and land survey authority (*Sw: Lantmäteriverket*), an national agency that both handles geo-data and real estate information in Sweden, is one example of a government agency that sees the opportunity to retail public service data as if it was any private enterprise while blocking competitors from buying raw data at wholesale or marginal cost. At present the authority receives a significant portion of its funds through commercialization of data. Then Open Data becomes a competitor to the bureaucratic interest to sell the data at a retail price. The competition between bureaucratic commercialization and free Open Data can explain partly regional delays in the dissemination of Open Data.

IV. THE POLITICAL DILEMMAS OF OPEN DATA

The four ways of disseminating public information described by Piotrowski [6] – public meeting, leaks, voluntarily dissemination and freedom of information requests – are driven by other actors than the bureaucracy itself. The dissemination described by Piotrowski sees the information sharing as a result of political processes. The voluntarily dissemination, which freely accessible Open Data would be, has been historically rarely seen at a global level. The voluntarily dissemination is a political decision.

The first countries and states in a federal framework to actively pursue open up for citizens' access to Open Data are mainly North America, the UK, and Oceania. One reason why the Anglo-Saxon countries are more active could be the conflict between bureaucratic interest and the interest of the civic societies where Anglo-Saxon countries have a weaker bureaucratic culture in relation to politicians than centralized governments. The bureaucracy can create a case for increased costs for information services to provide access to Open Data and, if the government agencies are already retailing data, a loss of tangible revenue. The decision to release data is a political decision, and politicians are exposed to the bureaucracy's agenda in several different ways. Politicians are assumed to seek to optimize the financial benefit for the state, or the entity they are elected to represent, and the release of Open Data without compensation can be seen as losing revenue – especially when the bureaucracy already generate revenue but prevents access to raw data and retails the data after predetermined processing.

The European Union identified already in year 2000 the potential for dissemination of public sector information [7] and resulted in EU Directive 2003/98/EC [8] that should increase the commercialization and dissemination of public sector information. The EU-PSI directive is primarily promoting commercialization and re-utilization of public sector information. The political and bureaucratic resistance in Sweden and Poland has forced the EU to take action to increase the dissemination [9]. A major point of conflict between the EU and member states has been the price for the data [10]. The European development, with the exception of the UK, is in stark contrast to for example the Australian approach where commercialization has been replaced with an Open Data approach [11]. These presented debates over the utilization of public sector information have two points of view, the view from the government or the view from the citizens. The potential internet startup entrepreneur's point of view is not considered, and as these innovators have limited political interest as a group; their interest is noted but not heard.

We decided to conduct an inquiry to seek evidence that the bureaucratic resistance or retailing of processed data, while not disseminating Open Data, might have had a significant alternative cost for society.

Our inquiry seeks to answer to the question: what is the impact on Internet startups and entrepreneurial innovation for the government's refusal to disseminate public sector information as Open Data?

Once that question is answered it could be compared to the revenues generated by the government agencies trying to retail the data, meanwhile denying Open Data access. The working hypothesis was that the revenues generated by the government agencies were marginal compared to the impact on innovation. The inquiry was designed as a survey.

V. SURVEY, INQUIRY AND METHODS

The survey data collection was performed during three start-up conferences held in Sweden, with competitive business plan contests, during 2011 and 2012. The conferences were 24 Hour Business Camp 2011 (24HBC2011), 24 Hour Business Camp 2012 (24HBC2012) and Geek Girl Meet Up 2012 (GGM2012). None of these conferences was aimed exclusively towards open data, the dissemination of public sector information, or creating access to these information sources. The targeted audience was IT-entrepreneurs seeking to create a startup business plan during the event. Access to the event was based on the ability to participate in the creation of a business plan and potential startup with 24 hours at a designated venue. These events had no additional participants that were respondents. The 24HBC2011 and 24HBC2012 were sponsored by the Swedish Internet Infrastructure Foundation (IIS) which is the administrator and registrar of the Swedish country domain .SE. The IIS foundation is (by the bylaws) bound to support the development and utilization of the Internet in Sweden. The IIS sponsorship did not only include actual resources for the startup business camp, but also marketing support utilizing the IIS to promote the event to potential entrepreneurs and existing IT-entrepreneurs.

The GGM2012 resembled the 24HBC2011 and 24HBC2012 in the purpose and aims, but was focused on female IT-entrepreneurs. GGM2012 was added to the data set to increase validity and mitigate gender bias. The 24HBC2011 and 24HBC2012 had 94 respondents and GGM2012 had 43 respondents.

The 24HBC2011 response rate was 98% which in any survey is extraordinarily high, and the survey was distributed at a general gathering during the conference and the survey collection was supported by staff from the IIS foundation. The response rate at 24HBC2012 and GGM2012 was around 50%, also a very high survey response rate.

The limited number of surveys is taken in consideration, but the business plan camps represent a highly focused sampling pool as these beginner or serial entrepreneurs are at the entrance of their new IT-entrepreneurial endeavor.

Of the 137 respondents of the survey, 86 were owners or founders of an IT business. The mean age was for early stage entrepreneurs 30 years old and for entrepreneurs with an existing IT business 31 years old.

The questionnaire was identical at all three events and contained five questions about the linkage between their business plan and Open Data, a question about which country they considered most entrepreneurially friendly, twelve questions about their interest in twelve different data sources that are public sector information in Sweden but not accessible as

Open Data, and a final question regarding the willingness to pay for access to public sector information.

The twelve data sources surveyed were geo-GIS data from the cadastral and land survey agency, incorporation and company data from the bureau of incorporation, vehicle registration data from department of motor vehicles, residence data from the tax authority, taxation data from tax authority, crime data from law enforcement, commute transit data from different transit authorities, real estate data from the cadastral authority, depersonalized health data, state and municipal purchasing data, weather from national weather services, and court data. The respondents rated their need for the different data sources. In this paper we have focused primarily on the demand for geo-data.

VI. EVIDENCE OF OPEN DATA'S IMPACT ON INNOVATION

A large portion (43%) of all the respondents, and 46% of the owner or founder respondents, claimed that access to public sector information open data was *required* to be able to execute their business plans. A majority (82%) of all respondents said that access to open data from the public sector would *support and strengthen* their business plans.

The survey was conducted with startup IT-entrepreneurs, of which not everyone had set out to create applications using data sources, but also B2B e-commerce, networking, and hardware solutions, thus the responses indicate an even higher interest for Open Data among software start-ups.

The Swedish cadastral and land survey authority (*Sw: Lantmäteriverket*) sell data at close to retail prices and prefer to retail the data through their own government website [12]. This retailing is a revenue stream for the agency, but from a wider and broader societal perspective this Open Data preventive stance might come at a far higher cost. Of the 24HBC2012 and 24HBC2011 respondents, 47% considered that their business plan relied on access to geo-data, and 23% would see their potential business venture to be jeopardized if not given access to geo-data. The Swedish cadastral and land survey authority (*Sw: Lantmäteriverket*) also the central authority for registration and issuing deeds for properties. In the conducted inquiry the 24HBC2011 and 24HBC2012 surveys, which had a 98% and a 50% response rate, one of the questions asked was the importance of access to cadastral and land registry data. Of the 94 respondents, 12% considered the access to cadastral and land survey data to be essential for the launch of their startup.

The question is the effect of multiple denial of access to affordable or free public sector information. In the 24HBC2012 and 24HBC2011 surveys, 16% stated that the access to geo-data and cadastral information was either essential or of high importance to the launch of the business plan. The Swedish cadastral and land survey authority (*Sw: Lantmäteriverket*) approach to denial of access to affordable or free public data, as a single government entity, could directly influence, as evidenced, the failure of one out of six business plans and early Internet ventures. The same pattern is repeated in the Swedish Incorporation Authority (*Bolagsverket*). As a

reference and comparison, the State of Montana, USA, provides cadastral open data at no cost [13].

The respondents stated as their primary use of public sector information was to resell the data after the raw data had been processed and merged with additional data to create new information products, test drive their business models, and seek to provide marketing and sales opportunities by analyzing the public sector's procurement.

VII. OPEN DATA AS CATALYST FOR INNOVATION

The conducted inquiry has shown that public sector Open Data is in many cases essential in the innovation for web-based applications and information services. Open Data influences the innovative process in many ways, and the absence of Open Data slows the innovative process or prevents the initiation of entrepreneurial innovation. The societal benefits from this economic activity and contribution to information markets could translate to economic growth and increased efficiency.

The importance of Open Data as an enabler and foundation for entrepreneurial activity and innovation can be presented as five categories.

A. SIMULATE POTENTIAL VIABILITY TO ENSURE FUNDING

Internet startups are based on a set of ideas, as any revolution or creative advancement, and are initiated in the minds of the founder or founders. The main obstacle to convey any idea of change is to present it to others in a way it could be understood and actually lead to change. In the traditional political revolution, it was a tangible event, such as the storming of the Bastille in the French Revolution, and in technology by a pilot project or technical test that visualize the functionality and abilities of the potential startup.

Open Data will provide data that can serve as a test bed to simulate an application and build evidence for the factual viability of the proposed project. The Open Data will feed the technical design with relevant data to ensure the best possible outcome of pilot projects.

The utilized Open Data would then create a case for further funding from business angles, venture capitalists and early startup funding, thus enabling the startup to move faster up the value chain and generate more accurate and better services.

There is a high risk for potential online entrepreneurial endeavors to fail already as prototypes. Open Data can feed data to test the feasibility and functionality of innovation projects that can detect flaws, tune, and optimize the value generated. Open Data will then serve the purpose as both a test bed and quality control before the innovation is presented to either investors or the market.

The inquiry has shown a high interest from the entrepreneurs to acquire relevant data sets for early testing purposes. The respondents are prepared to pay mean average \$1,500 for testing data sets, but only \$25 when the median is calculated. The spread in the answers can be seen as the majority of the respondents have considered the need for a large number of datasets in an early development phase,

meanwhile a few are ready to pay for more for a limited number of data sets for an application that is further in its development.

The willingness to pay for data is far lower than the expected price for data made available by the government. The Swedish government has been reluctant to make data available and, in the few cases data has been offered, priced at levels that are not financially reachable by early stage entrepreneurs [14], [15].

B. PROVIDE INFORMATION ABOUT POTENTIAL MARKET

Startup companies seek niche markets in which they can operate. Open Data creates an abundance of niche information through all sectors. Modern, industrialized advanced democracies allocate significant amounts to gather and systemize public data every year. The amount of available data in the public sector is by its sheer size difficult to quantify, but it is enormous compared to any private data repository.

Each sector of the society has a government agency or department assigned to manage public program, monitor developments, and ensure compliance with laws and regulations. The information gathered by these agencies and departments will give potential entrepreneurs ability to gauge the market and the size of the opportunity they seek to seize.

The ability to gauge the market increases the likelihood for business angles and venture capitalists to comprehend the potential market.

C. REDUCE DEVELOPMENT LEAD TIME TO APPLICATION MARKET

Public sector information is often of high quality and well-structured and thus requires less processing before it can be used in applications. Public sector data sources utilize to a high degree identical formatting, data interface, and share commonalities that remove hurdles to merge data. There are variations between countries, but a centralized non-federal government structure increases the standardization. Examples of non-federal centralized governments are the Government of France and the Government of Sweden.

D. DRIVE INNOVATION BEYOND APPLICATIONS

The 24HBC2012 and 24HBC2011 events were entrepreneurial business plan creation events that attracted entrepreneurs in information technology and the peripheral markets. The aforementioned 43% of the respondents that stated Open Data to be required for their execution of their business plan should be compared with the 82% of the respondents who answered that open data access would strengthen their business plans and the 42% of respondents who also indicated willingness to pay for PSI data, despite having no stated intention to commercialize data themselves. These results indicate that a large proportion of entrepreneurs intend to use data indirectly. Thus, the focus on the public sector information creating a repackaging and reselling information market through

development of data processing and applications could be a serious underestimation of the total impact the release of Open Data.

E. ENHANCE EXISTING ONLINE SERVICES AND OFFERINGS

Based on logics there can be an assumption that the identical opportunity seen by new entrepreneurs can be captured by existing online services and offerings. The existing services have already established a structure, understanding of the market, processes, financial foundation, and are operating in the online market place. In the study the respondents that were most eager to utilize Open Data were already entrepreneurs and owners of an IT-business – by either being a serial entrepreneur that started a new business or enhanced an existing business by creating additional user and customer value. The existing businesses are likely faster to utilize Open Data as they are already operating.

E-commerce companies, and new startup challengers, need data to calculate freight alternatives and ensure that delivery addresses are correct. In Sweden geo data, address data, and zip code data are government property and resold as a permanent data link to major businesses with high entrance and user fees. This established level of charges is prohibiting innovators and early entrepreneurs to have access to these data as it is beyond their financial reach.

If instead geo, address, and zip code data were Open Data, the data would not only usher innovation forward, but also increase the quality in existing services and enhance the service offerings to the benefit of the businesses and society in general.

VIII. CONCLUSION

The release of the public sector information data as Open Data has a tangible direct payoff for the disseminating government in increased entrepreneurial activity, creating triggers for internet startup realizations, and enhanced business services. The conducted inquiry found evidence that the bureaucratic resistance or retailing of processed data, while not disseminating Open Data, had a significant societal cost in lost innovation, entrepreneurial incentive, and disabled a vast number of new business plans to be executed. The Open Data impact on Internet startups and entrepreneurial innovation is so high that political executives could be left with two choices – feasible climate for Internet startup's innovation or bureaucratic data retailing – and they cannot have both.

ACKNOWLEDGEMENT

The authors would like to thank Pernilla Rydmark, Joakim Jardenberg, Jan Elvelid, and Karina Töndevold for their help with the data collection at the different events.

REFERENCES

- [1] Y. Amichai-Hamburger, K. McKenna, and S. A. Tal, "E-empowerment: Empowerment by the internet," *Comput. Human Behavior*. vol. 24, no. 5, pp. 1776–1789, Sep. 2008.
- [2] J. C. Bertot, P. T. Jaeger, and J. M. Grimes, "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies," *Government Inf. Quart.*, vol. 27, no. 3, pp. 264–271, 2010.
- [3] K. Janssen, "The influence of the PSI directive on open government data: An overview of recent developments," *Government Inf. Quart.*, vol. 28, no. 4, pp. 446–456, 2007.
- [4] J. Kallberg, "The Internet as a proxy for democratic accountability and transparency—A comparative test of Waldo's five problem areas in five advanced democratic societies," Ph.D. dissertation, Dept. Public Affairs, Univ. Texas at Dallas, Richardson, TX, USA, 2011.
- [5] European Commission. (2011). *Review of Recent Studies on PSI Re-Use and Related Market Developments*, Brussels, Belgium [Online]. Available: <http://epsiplatform.eu/content/review-recent-psi-re-use-studies-published>
- [6] S. J. Piotrowski, *Governmental Transparency in the Path of Administrative Reform*. New York, NY, USA: SUNY Press, 2007.
- [7] European Commission. (2000). *Commercial Exploitation of Europe's Public Sector Information*, Brussels, Belgium [Online]. Available FTP: ftp.cordis.europa.eu/pub/econtent/docs/2000_1558_en.pdf
- [8] European Commission. (2003, Nov. 17). *Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the Re-Use of Public Sector Information*, Brussels, Belgium [Online]. Available: <http://eu.vlex.com/vid/directive-parliament-use-public-information-37791265>
- [9] European Commission. (2008). *Press Release: Re-Use of Public Sector Information: Commission Launches Infringement Proceedings Against Poland and Sweden*, Brussels, Belgium [Online]. Available: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/1524&type=HTML&aged=0&language=EN&guiLanguage=en>
- [10] European Commission. (2011). *Pricing of Public Sector Information Study*, Brussels, Belgium [Online]. Available: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/report/11_2012/models.pdf
- [11] Creative Commons. (2012). *Attribution 2.5 Australia (CC BY 2.5)*, Mountain View, CA, USA [Online]. Available: <http://creativecommons.org/licenses/by/2.5/au/>
- [12] Lantmateriet. (2012). *Kartförsäljningen*, Gävle, Sweden [Online]. Available: <http://www.lantmateriet.se/Kartor-och-geografisk-information/Kartor/>
- [13] (2012). *Montana Cadastral* [Online]. Available: <http://svc.mt.gov/msl/mtcadastral/>
- [14] S. Jonson. (2012). *Internet Infrastructure Foundation (IIS). The Öppna Data—En Utdragen Historia* [Online]. Available: <https://www.iis.se/blogg/oppna-data-en-utdragen-historia/>
- [15] J. Kallberg and E. Lakomaa. (2012, Mar. 12). *Det Behövs Verklig Öppenhet i Sverige*. Svenska Dagbladet. Stockholm, Sweden [Online]. Available: http://www.svd.se/opinion/brannpunkt/det-behovs-verklig-oppenhet-i-sverige_6925111.svd



ERIK LAKOMAA is a Research Fellow with the Department of Marketing and Strategy and the Director of the Institute for Research in Economic and Business History, Stockholm School of Economics, Stockholm, Sweden. His work has focused on business history and decision making processes in companies and public sector entities.



JAN KALLBERG is a Researcher and the Head for Cyber Operations Laboratory, Cyber Security Research and Education Institute, University of Texas at Dallas, Dallas, TX, USA, an Assistant Professor with Arkansas Tech University, Russellville, AR, USA, and part-time Faculty with George Washington University, Washington, DC, USA. He received the Ph.D. degree in public affairs and the M.A. degree in political science from the University of Texas at Dallas, in addition to a JD/LLM from Stockholm University, Stockholm, Sweden. He has published in Strategic Studies Quarterly, Joint Forces Quarterly, IEEE Professional, and IEEE Security and Privacy.

...