A framework for Linked Data business models

Michalis Vafopoulos
Mathematics Department, Aristotle University of Thessaloniki, Greece
e-mail: vaf@aegean.gr

***Abstract*— Linked Data (LD) are emerging as the potential "killer application" of the third Web era. If they actually become mainstream in the next years depends on how useful will be for the average User. In economic terms, the creation of successful business models to place LD in the market and the resulting revenue streams will support further developments. In this article, the major sources of value and a business model taxonomy for LD are identified. Network effects exist in LD by enabling Web users to build bidirectional and massively processable interconnections among online data and second, these data are critical enablers for existing infrastructure in the government and business spheres. The proposed eleven distinct business model categories are: Public service, Community service, Intelligence, Subscriptions, Sponsorships, Advertising, Marketplace, Affiliate program, Multi-sided platforms, Traffic generation and Data branding. Section I briefly introduces the advent of LD in the Web ecosystem. The next section is about the main question, which is addressed in the present paper. Related literature is reviewed in Section III. The encompassing value source of the Web is specified for LD in Section IV. Section V describes a proposed framework for LD business models and Section VI concludes.**

Keywords- network externalities; value creation; infrastructure; two magics of Web science

I. LINKED DATA: FROM READ-WRITE TO PROCESS WEB

The Web began as an information project at CERN, where Tim Berners-Lee developed the vision of a common information space [1]. The Web is a system of interlinked, hypertext documents accessed via the Internet. With a Web browser a User views and edits Web pages that may contain text, images, videos, and other multimedia and navigates between them using hyperlinks. Web technologies are successful because are based in an architecture (e.g. HTTP, URI, HTML), which is simple, networked, based on open standards, extensible, tolerant, universal, free or cheap, fun and powerful.

Social Web or Web 2.0 is a term depicting the trend in the use of Web technology (especially, after 2001) that facilitates naïve user creativity, information sharing and, most importantly, massive collaboration among Users. Web 2.0 includes social-networking sites (i.e. Facebook), wikis (i.e. Wikipedia), blogs and micro-blogs (i.e. twitter). Despite the fact that the term Web 2.0, misleadingly suggests a new version of the Web, it does not refer to an update to any technical specifications, but to changes in the ways software developers and Users exploit the Web.

The Semantic Web is an evolving extension of the Web in which content can be expressed not only in natural language, but also in a format that can used by software agents, thus permitting them to find, share and integrate information more easily. In few words, the vision of the Semantic Web is as an extension of Web principles *from documents to data*. Web hides computers and analogously, Semantic Web hides documents based on a set of architectural principles and enabling technologies. Some of these include Resource Description Framework (RDF) [2], which unifies a variety of data interchange formats (e.g. N3, Turtle) using XML as the interchange syntax and notations such Ontology Web Language, all of which are purposed to provide a formal description of concepts, terms and relationships within a specific knowledge domain. In order to accelerate the realization of the Semantic Web vision, W3C announced the Linked Data community project [3]. Linked Data (LD) is for spreadsheets and databases what the Web of hypertext documents is to word processor files and could be considered as crucial for the Semantic Web as hypertext has been for the Web [4]. It is a solution to the complex and not flexible "top-down" model of designing an ontology or schema before developing the data at the scale of the Web. LD is a recent movement to build the Semantic Web grass-roots-style by extending the Web with data commons, by publishing various open datasets as RDF on the Web and by setting RDF links between data items from different data sources. Openness, modularity and scalability characterizing LD have recently enabled many private and government institutions to publish billions of RDF triples [5] including domains such as geographic information, people, companies, online communities, films, music, books and scientific publications. From the ability to read (Web 1.0) and write (Web 2.0) online, we are heading towards the discovery and processing of information in massive scale (Web 3.0).

II. RESEARCH QUESTION

Despite the transformative power of the current Web, the existence of unstructured information limits Users' exploring capabilities (Figure 1). The innovative idea to overcome this issue is based on LD. The acceptance of LD in the Web ecosystem depends both on technical (i.e. URI, HTTP, RDF) and social factors (i.e. link incentives). *Will Users embrace LD in micro scale to collaboratively enrich the Web system with LD?* In this article, we address the economic aspect of this question by describing the sources of value of LD and by providing a modeling framework of LD existing and potential markets. Briefly, LD is an attempt to simplify and accelerate the network externalities of Web 3.0 based on two sources of value. First, it enables Web users to build bidirectional and massively processable interconnections among online data and second, these data are critical enablers for existing infrastructure.

Electronic copy available at: http://ssrn.com/abstract=1850365

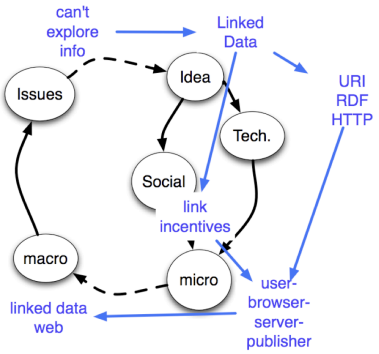

Figure 1. Two magics of Web Science: the case of Linked Data [6]

III. Related Literature

The commercialization of the Web began in mid-90s with the introduction of browsers and search engines that enabled user-friendly navigation. The excessive global optimism for quick fortunes and endless business growth in the late-90s impelled early entrants to monopolies [7] and drove economy in 2001 to the burst of the dot-com bubble [8]. After this hard lesson, the new business models in mid-00s were updated to include advertising revenue from Web navigation and provision of value added services. Today, the Web economy is bigger and more robust with new services ranging from search to social networking, virtual entertainment and giant multi-stores. In the demand side, most of the population in the western world is involved in the Web economy. In 2009, the Web contributed 115 billion euros to the British economy (7.2% of GDP) overcoming traditional sectors like construction, transport or utilities industries. It is also predicted that by 2015, UK's Web economy is likely to grow up to 10% of GDP, eclipsing the financial sector. Much of the growth is driven by consumption, the majority of it online spending, while the rest comes from government spending, private investment and exports [9].

While Silicon Valley is currently focused on social networking [10], Google embodies LD technologies (e.g. RDFa in search [11]) and new businesses are providing LD-based services (e.g. Garlik [12], Infochimps). The fast growing awareness for LD has motivated scholars and practitioners of the Web business to discuss about the emerging business models.

There is a broad literature about business models in general and in connection to the Web (for an extensive review see [13]). First, Timmers [14] identified eleven categories of e-business based on their degree of innovation and functional integration. The models are divided to e-shops, e-procurement, e-malls, e-auctions, virtual communities, collaboration platforms, third-party marketplaces, value chain integrators, value-chain service providers, information brokerage, trust and other third-party services. Tapscott *et al.* [15] introduced five types of value networks (Agora, Aggregation, Value Chain and Distributive Networks) based on their degree of control (hierarchical or self-organizing) and value integration. Weill and Vitale [16] proposed eight business models in order to compactly describe all possible practices of conducting business in the Web. Their taxonomy includes Content Providers, Direct to Customer Services, Full-Service Providers, Intermediary, Shared Infrastructure, Value net Integrators, Virtual Communities and Whole-of-Enterprise/Government. In his influential work Rappa [17] created a taxonomy based on the position of e-business in the value chain. The distinct value proposition of generating revenues defines eight business models, namely: Brokerage, Advertising, Infomediary, Merchant, Manufacturer, Affiliate, Community, Subscription and Utility.

Servant [18] discussed the potential benefits of LD for enterprises and Latif *et al.* [4] provided the LD Value Chain as a lightweight model for business engineers. [4] defined the entities, which act in different roles (Raw Data Provider, LD Provider, LD Application Provider and End User) and both consume and provide different types of data (Raw, Linked and Human-Readable). They exemplarily applied the proposed Value Chain within a concrete case study from BBC to demonstrate the underlying business risks. Brinker initiated the dialogue about LD business models [19] and Erickson [20] provided a comprehensive understanding of potential revenue streams and customers.

IV. Network Externalities and Effects: The Source of Value for the Web

Our analysis is initiated by the fundamental concept of network externalities and how is related to the Web. *Pure network goods* are defined to be goods that derive their entire value from network externalities. Pure network goods have no value in a network of zero size (e.g. telephony, Internet, the Web) [21]. *Externalities* in economic theory are defined to be the indirect effects of consumption or production activity, that is, effects on agents other than the originator of such activity, which do not work through the price system. In a private competitive economy, equilibria will not be in general Pareto optimal since they will reflect only private (direct) effects and not social (direct plus indirect) effects of economic activity [22]. If this indirect effect (or transaction spillover) is beneficial to the other agents is called a *positive* externality and in the opposite case of a cost is called a *negative* externality. For instance, uploading and interconnecting online content in the Web may result positive externalities if it is for instance, educational or joyful, or may cause negative externalities if it is privacy threating or libelous. Furthermore, positive network *effects* characterize a good when more usage of the good by any user increases its value for other users. These effects are also called *positive consumption or demand side externalities*. In a significant part of literature, the concepts of network externalities and effects are misleadingly used as synonyms. Network externalities in economic analysis are defined to be the externalities involved in network effects. Usually, in the presence of network effects, a user only takes into account his own utility in his decision to join or not the network. The additional utility his joining provides on all other users is overlooked in his decision [23]. Briefly, network externalities are determined by four factors: *Expectations* and *Coordination* of consumers, *Switching*

Electronic copy available at: http://ssrn.com/abstract=1850365

Costs and *Compatibility* of the network good under consideration with the rest of goods [24].

The major value source of the Web is the ability to link resources. In Web 1.0 the dominant resource is documents, in Web 2.0 is Users and their contribution [25] and in Web 3.0 is structured data [26]. According to Hendler [27] Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data (Figure 2). A potential joint exploitation of all linking spaces in the Web will create enormous social and commercial value [28].

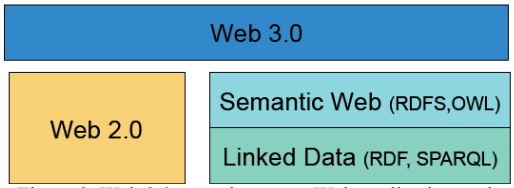

Figure 2. Web 3.0 extends current Web applications using Semantic Web technologies and graph-based, open data [27]

V. The Network Effects of LD

LD is an attempt to simplify and spread horizontally throughout the Web the network externalities that exist in Web 3.0. Specifically, two sources of value have been identified for LD technology. First, it enables Web users to build bidirectional and massively processable interconnections among online data and second, these data are critical enablers for existing infrastructure in the government and business spheres. In particular:

A. Building bidirectional and massively processable interconnections among online data

In contrast to Web APIs, LD mashups are statements that link items in related datasets. As Heath [29] explains: "*Crucially these items are identified by URIs starting "http://", each of which may have been minted in the domain of the data publisher, meaning that whenever anyone looks up one of these URIs they may be channeled back to the original data source. It is this feature that creates the business value in Linked Data compared to conventional Web APIs. Rather than releasing data into the cloud untethered and untraceable, Linked Data allows organisations and individuals to expose their data assets in a way that is easily consumed by others, whilst retaining indicators of provenance and a means to capitalise on or otherwise benefit from their commitment to openness.*"

B. Linked Data as an enabler for existing infrastructures

Public and private entities produce and collect tremendous amounts of data as part of their daily operations. At the same time, increasing investments on IT infrastructure and skills are needed in order to maintain and operate these data on complex hardware and software systems. LD enable the creation of better and massive services for use and reuse for many of these data, driving existing infrastructure in its full potential. For government bodies, LD adoption is focused on open, transparent, collaborative and more efficient governance. For enterprises, the core issue is about effective knowledge management and the implementation of new business models that enable more energetic involvement and collaboration between producers and consumers. There is also significant economic potential in open government LD, which can be used by business as an input to improve existing and to create added value services. This potential can be realized to useful business projects if a certain threshold of data quantity and quality and relevant knowledge is reached. Today, it seems that we are about to approach the triggering point of a virtuous cycle for better services and more involved consumers in the Web economy. In the next section, we attempt to identify under a common framework, the functional aspects of value creation mechanisms in LD.

VI. A Framework for LD Business Models

The next step, after defining the core value sources of LD, is to identify critical functional components that can place LD services into the market. In this context, we follow Stähler's [30] four-components analysis. These components are the following:

1) *Value Proposition:* what value the business creates
2) *Product/Services:* which service and to whom
3) *Architecture:* how the value is created
4) *Revenue Model:* sources and types of income

The fourth component is further analyzed and depicted in Figure 3, on how direct and what is the source of revenues of LD business [19]. In particular:

A. How direct is the revenue?

At the bottom of the pyramid are situated models that are based on direct revenues streams like public subsidies, donations and payments for business projects. Toward the top levels are more indirect revenue sources, such as using LD to attract traffic to a Web site or to strengthen business brand and reputation [19].

B. Who pay for the LD?

For any business model, the end User of the service and the source of revenue are not necessarily the same. For instance, in the subscription model Users and revenue source are identical, but in the case of advertisement and public service models, Users are enjoying LD without making direct payments [19].

C. Who are the consumers of LD?

Web Users can be divided into three segments: Professional Editors, Amateur Editors and Navigators. Let us describe the basic characteristics of the above-mentioned categories. First, Users are partitioned to Navigators and Editors of the Web. Navigators are consuming information by navigating (browsing, surfing or accessing) the Web. Editors are creating, updating or deleting online content and links in the Web network. Editors can be further elaborated to Amateurs and Professionals, based on their production incentives. In contrast to Amateur Editors (e.g. Wikipedia editors), Professional Editors are profit maximizers and take into account direct financial compensations in producing Web content (e.g. the Facebook Corp.).

Public service is the base of the LD business stack because government bodies are the biggest producers and

consumers of public data and the exclusive issuers of basic trust certificates (e.g. passport, digital signatures) that facilitate the consumption of public data. Web 2.0 has underlined the importance of collective intelligence and inter-creativity in the progress of the Web ecosystem. Thus, the two first layers (Public and Community) are open to all Web Users. Subscriptions, customized services, affiliate programs and marketplaces are mainly (but not exclusively) focused to attract Professional Editors. Multi-sided platforms, traffic generation and data branding are basically targeting Navigators. The five classes of revenue sources are identified by their different color: orange for government or non-profit, light orange for donation and community contribution, magenta for (professional and amateur) Editors, green for Advertisers and blue for Navigators.

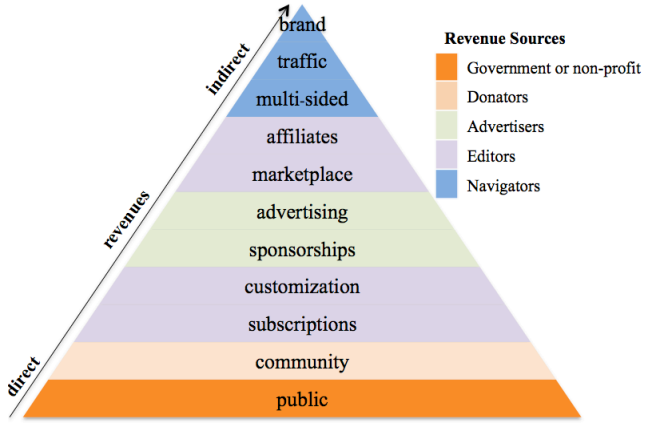

Figure 3: A framework of LD business models

The above framework supports any possible categorization that is based on how direct and what is the source of revenues and who are mainly the end users of LD. In Figure 3 eleven business models have been described as a starting point for exploring LD monetization and wide provision. These are the following:

1) Public service: The basic LD infrastructure needs to be initiated and maintained by government bodies (e.g. geodata.gov.gr, data.gov.uk). This service includes fundamental data that are mainly produced by the public sector (e.g. legislation, street addresses, demographics etc.).

2) Community Service: As in the collaborative Web 2.0, a group of Users may create, edit and publish royalty-free data in a voluntary basis. Revenue stems mainly from donations (e.g. Wikipedia, Open Knowledge Foundation) and ancillary products. The viability of the community model is based on User loyalty and their high investments in time and emotion [17].

3) Subscriptions: As in offline and online markets, Users are charged a periodic fee to subscribe to a different level of access to data:

3.1) Full access: Payments for access to detailed and denser data (e.g. richer links to other datasets) [20], including licensing fees to let developers use data in other computing environments [19].

3.2) Timely Access: Fees for access to the most recent version of the data (e.g. delayed stock prices) [20].

3.3) *On-demand access:* Payments for on-demand access for individual queries or particular data sets (i.e. micro-transactions) [19].

3.4) *Block access:* Payments for access to a dataset based on time (e.g. daily pass) or based on the number, frequency and concurrency of accesses [20].

3.5) Archival access: Since publishing and maintaining large archival datasets on-line can be expensive in terms of data conversion, digitization, interlinking and service provision, paid access will only be available to subscribers [20].

3.6) Convenient Access: Payments for access to the data through a specific mechanism (e.g. the Talis platform) [20].

3.7) Freemium: Users may enjoy free but limited access to data to sample, but a fee is charged for extended premium access [19]. Commonly, users provide some personal data (e.g. email) and their traffic patterns.

4) Customized Service: Some companies charge fees to provide professional services, intelligent tools and customized solutions to create, edit, publish and re-use LD for particular business needs (e.g. statistical analysis of LD traffic, integration to ERPs, ontology customization).

5) Sponsorships: Advertisers can pay fees to be included in the LD set or attributed valuable meta-data or a small number of advertisers can be charged for brand visibility of sponsoring the data [20].

6) Advertising: Contextual advertising in Web sites and sponsored search can be a strong revenue force for LD, as in the current Web economy. As Dodds [20] argues "data-layer advertisements" are not seem to be viable or useful in practice because they will be easily identified and ignored by applications like RSS advertising.

7) Marketplace: LD producers and administrators can supply data to a partner in exchange for an opportunistic royalty [19].

8) Affiliate Program: As it happens in the case of Web APIs, LD producers will supply data streams to affiliates who deliver them in other applications in exchange for commissions on related sales [19]. In addition, existing successful affiliate's programs (e.g. Amazon, Flickr) may adopt LD technologies in order to upgrade their services. Partners in affiliate programs can associate affiliate product links with data to earn commissions on related sales [19].

9) Multi-sided platforms: Multi-sided platforms (or two-sided network effects) are evidenced in cases where increases in usage by one set of consumers increases the value of a complementary good to another distinct set of consumers, and vice versa. For instance, Google is offering

for free (and some are also advertisement-free) a rich set of services in order to promote the sales of its advertising platform. Currently, LD-based services are included in the bundle of free Google services (e.g. Google Refine [31] and Google public data explorer [32]).

10) Traffic Generation/Search Engine Optimization (SEO): LD publication might help Web sites to receive higher positions in search engines and other directories to attract more Users [19].

These eleven generic categories of business models are often enjoying a symbiotic relation in practice. For instance, subscription services may come with discounted fee if they are combined with contextual advertisements.

VII. Conclusions and Further Research

In the first twenty years of its existence, the World Wide Web has proven, to have a fundamental and transformative impact on all facets of our society. While the Internet has been introduced 20 years earlier, the Web has been its "killer" application with more than 2 billion users worldwide accessing more than 1 trillion Web pages. Searching, social networking, broadcasting, photo sharing and blogging have become part of everyday life whilst the majority of software and business applications have migrated to the Web.

Although, LD is still an immature set of Semantic Web technologies, has proven to be fast evolving and efficient in revealing the potential of existing infrastructure and information. As it happened with the current Web, it is needed to incentivize Users in order to become part of the next Web era. In this article, we are identifying the major sources of value and a business model taxonomy for LD.

The major challenges for LD is to understand the real practical experience in the field and to incorporate it in business modeling and to investigate further, more efficient ways of co-evolution between the current Web and LD.

Acknowledgment

The Municipality of Veria and the Mathematics Department of Aristotle University of Thessaloniki supported our research.

References

[1] T. Berners-Lee, "Information management: A proposal," 1989.

[2] "W3C rdf - http://www.w3.org/RDF/ - date accessed: May 5th, 2011."

[3] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *International Journal on Semantic Web and Information Systems*, vol. 5, 2009, p. 1.

[4] A. Latif, A.U. Saeed, P. Hoefler, A. Stocker, and C. Wagner, "The linked data value chain: A lightweight model for business engineers," *Proceedings of I-SEMANTICS'09 International Conference on Semantic Systems*, p. 568–575.

[5] K. Moeller, M. Hausenblas, R. Cyganiak, G. Grimnes, and S. Handschuh, "Learning from Linked Open Data Usage: Patterns & Metrics," 2010.

[6] T. Berners-Lee, "The process of designing things in a very large space - http://www.w3.org/2007/Talks/0509-www-keynote-tbl/#%281%29 - date accessed: May 5th, 2011," 2007.

[7] C. Shapiro and H.R. Varian, *Information rules: a strategic guide to the network economy*, Harvard Business Press, 1999.

[8] R. Clarida, "G7 Current Account Imbalances: Sustainability and Adjustment," 2006.

[9] "Internet accounts for 7.2% of economy: study - http://www.google.com/hostednews/afp/article/ALeqM5j9KJ1aijg0ECa0R6Dl7Pg2dOQhZg?docId=CNG.cc58070f3f3b3257db047a1b6231a166.421 - date accessed: May 5th, 2011."

[10] "Facebook & Silicon Valley - http://online.wsj.com/article/SB10001424052748704050204576218863163692064.html - date accessed: May 5th, 2011," *Wall Street Journal Online*.

[11] T. O' Brien, "Google Announces Support for Microformats and RDFa - http://radar.oreilly.com/2009/05/google-announces-support-for-m.html - date accessed: May 5th, 2011."

[12] S. Harris, T. Ilube, and M. Tuffield, "Enterprise Linked Data as Core Business Infrastructure," *Linking Enterprise Data*, 2010, p. 209–219.

[13] A. Osterwalder and Y. Pigneur, "An e-business model ontology for modeling e-business," *15th Bled Electronic Commerce Conference*, Citeseer, 2002, p. 17–19.

[14] P. Timmers, "Business Models for Electronic Markets," *Electronic Markets*, vol. 8, 1998, pp. 3-8.

[15] D. Tapscott and A. Lowy, "Harnessing the Power of Business Webs," *Harvard Business School Press, Boston*, 2000.

[16] P. Weill, *Place to space: Migrating to eBusiness Models*, Harvard Business School Press, 2001.

[17] M. Rappa, "Business models on the web," *Available at Managing the Digital Enterprise website*, 2003.

[18] F.P. Servant, "Linking enterprise data," *Linked Data on the Web (LDOW 2008)*, 2008.

[19] S. Brinker, "Business models for Linked Data and Web 3.0 - http://www.chiefmartec.com/2010/03/business-models-for-linked-data-and-web-30.html%20 - date accessed: May 5th, 2011."

[20] L. Dodds, "Thoughts on Linked Data Business Models - http://www.ldodds.com/blog/2010/01/thoughts-on-linked-data-business-models/ - date accessed: May 5th, 2011."

[21] N. Economides and F. Flyer, *Compatibility and market structure for network goods*, New York University, Leonard N. Stern School of Business, 1998.

[22] S.N. Durlauf and L. Blume, *The new Palgrave dictionary of economics*, Palgrave Macmillan, 2008.

[23] A. Rapoport, "What is the difference between a network externality and a network effect? - http://www.quora.com/Network-Effects/What-is-the-difference-between-a-network-externality-and-a-network-effect%20 - date accessed: May 5th, 2011."

[24] M.L. Katz and C. Shapiro, *Systems competition and network effects*, JSTOR, 2000.

[25] "Network effects from user contributions are the key to market dominance in the Web 2.0 era - http://oreilly.com/web2/archive/what-is-web-20.html?page=2 - date accessed: May 5th, 2011."

[26] T. Berners-Lee, "Welcome to the Semantic Web," *The Economist - The World in 2007*, 2006.

[27] J. Hendler, "Web 3.0 Emerging," *Computer*, vol. 42, Apr. 2009, p. 111–113.

[28] J. Hendler and J. Golbeck, "Metcalfe's law, Web 2.0, and the Semantic Web," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, Feb. 2008, pp. 14-20.

[29] T. Heath, "Where is the business value in linked data? - http://tomheath.com/blog/2008/09/where-is-the-business-value-in-linked-data/ - date accessed: May 5th, 2011."

[30] P. Staehler, "Business Models as an Unit of Analysis for Strategizing," *International Workshop on Business Models, Lausanne, Switzerland*, 2002, p. 4–5.

[31] "Google refine - http://code.google.com/p/google-refine/ - date accessed: May 5th, 2011."

[32] "Google Public Data Explorer - http://www.google.com/publicdata/home - date accessed: May 5th, 2011."