Crowdmapping Digital Social Innovation with Linked Data

Harry Halpin[1](✉) and Francesca Bria[2]

[1] World Wide Web Consortium/MIT, 32 Vassar Street, Cambridge, MA 02139, USA
hhalpin@w3.org

[2] Nesta, 1 Plough Place, London EC4A 1DE, UK
francesca.bria@nesta.org.uk

Abstract. The European Commission recently became interested in mapping digital social innovation in Europe. In order to understand this rapidly developing if little known area, a visual and interactive survey was made in order to crowd-source a map of digital social innovation, available at http://digitalsocial.eu. Over 900 organizations participated, and Linked Data was used as the backend with a number of valuable advantages. The data was processed using SPARQL and network analysis, and a number of concrete policy recommendations resulted from the analysis.

Keywords: Digital social innovation · Linked data · Crowdsourcing · Policy

1 Introduction

In order to understand the emerging field of digital social innovation (DSI), the DigitalSocial.eu website[1] was created to crowdsource a map of organizations and projects involved in digital social innovation.[2] This crowd-sourced map is a part of a larger study by the European Commission to create new data-driven policy in order to support social innovation throughout Europe. The back-end of the crowd-mapping website was a triple-store that exposed the data as both Linked Data and a SPARQL endpoint, with sophisticated visualizations being dynamically generated "on the fly" from the RDF data to allow the inspection of network of digital social innovation. Using SPARQL and network analysis, a number of detailed questions from the European Commission could then be answered about the current structure of digital social innovation. So far, 967 organizations working on DSI in Europe were mapped as of January 18th 2015, including 609 projects between the organizations, with more organizations being added dynamically to the data-set every day. This is one of the first studies to

[1] http://digitalsocial.EU.

[2] This work was funded by the European Commission via the DSI contract and the DCENT EC project. More information on the DCENT project is here: http://dcentproject.eu/.

F. Gandon et al. (Eds.): ESWC 2015, LNCS 9088, pp. 606–620, 2015.
DOI: 10.1007/978-3-319-18818-8_37

show off the power in Linked Data to create *data-driven policy* via the use of semantic technologies to bootstrap a data-set, increase the scope of a survey, and so help structure a complex data analysis. By allowing diverse sets of flexible data to be inter-connected and combined, and then having new data being added with a focused crowd-sourcing survey campaign, a large set of empirical data could be boot-strapped easily on an entirely new topic.

To summarize, our contributions are:

1. Defining the emerging field of digital social innovation.
2. A website based on RDF that crowd-mapped a large data-set about digital social innovation in Europe where semantic technologies were a crucial advantage.
3. A detailed network analysis of digital social innovation in Europe.
4. Data-driven policy recommendations to the European Commission that were driven by the analysis of the collected data.

2 What Is Digital Social Innovation?

The new field of "Digital Social Innovation" (DSI) has emerged over the last few years, pointing to radically new ways of organizing many of the essentials of life from money and health to democracy and education. We tentatively defined Digital Social Innovation (DSI) as "a type of social and collaborative innovation in which innovators, users and communities collaborate using digital technologies to co-create knowledge and solutions for a wide range of social needs and at a scale and speed that was unimaginable before the rise of the Internet."[3] To provide some intuition, communities as diverse as fablabs, hackerspaces, and community wifi networks are part of this new trend.

This definition has a number of different elements to be clarified. Innovation in general is any change to the process of production, consumption, or distribution of commodities necessary for the social reproduction of society. Innovation is first and foremost a collective process undertaken by a large number of actors. Innovation is no longer seen as a linear step-by-step process in which research activities or technology pushes automatically lead to innovation and the commercialization of new products, but rather as a complex, dynamic, and interdependent process involving different stakeholders, including "bottom-up" communities that connect over the Web. Some innovations involve big discontinuities - 'radical' or 'disruptive' innovations - and others involve continuous small 'incremental' innovations [5]. In particular, social and institutional change, called "social innovation," often takes place as gradual and incremental change outside of revolutionary periods. Yet it is precisely social innovation that seems to be desperately needed in an era where many governments and services are viewed as ineffective or failed. As innovation happens as part of larger collective process,

[3] DSI Interim Report: http://content.digitalsocial.eu/wp-content/uploads/2014/09/FINAL-2ND-INTERIM-STUDY-REPORT.pdf.

mapping the social networks of innovators and their users who then co-produce the innovation is crucial.

Second, while we can demarcate the study insofar as the definition of digital social innovation to focus on "digital" technologies that scale using the Web, we still have a very wide range of possible areas DSI can be applied to, from health to democracy. After all, today almost all facets of life are touched by digital technologies. While many areas of focus in digital social innovation are still emergent, some are growing very fast, and others are quite marginal. In order to narrow our study somewhat, we will focus on innovation where the digital aspect of scaling depends on bringing together a community of innovation over the Web, which would simply not be possible in the era before the Web. Simply put, this eliminates from our study organizations for social innovation that do not use the Web or organizations that do not depend crucially on the Web for their scaling.

Note in terms of using the internet to "scale" we want to distinguish between two levels: (1) the level of the technical networking infrastructure itself provided by the Internet and (2) the level of online services built on top of these networks on the Web. Metcalfe's Law, (i.e. that the value of the network is in proportion to the number of members squared, so that the value of the network goes up for all users the more users are added) clearly applies to the value of technical networks given by examples such as widespread smartphone usage. One open question for DSI is that we can also scale social innovation via services using the Web, and thus we need to pay close attention to the structural dynamics of the social network of DSI that should be understood.

What forms of digital social innovation are emerging? What are their characteristics and needs? How they can scale? What role can the European Commission play in this context? These are some of the over-arching questions that this research is trying to answer. The general thesis the European Commission-funded study on digital social innovation is that it provides a new path out of the larger economic crisis that is neither reliant on traditional business models or austerity measures in the public sector, but that the power of innovation over the Web can allow people to self-organize in order to solve their myriad problems. Yet until now, digital social innovation has been almost entirely invisible to policy makers and has had none of the extensive support that has gone into digital technologies for the military, government, or business. Yet DSI has the potential to contribute to three of the most important challenges facing Europe: reinventing public services, often in less costly ways; reinventing community, and how people collaborate together; and reinventing business in ways that are better aligned with human needs. The evidence gathered here enables us to recommend how best to combine research, strategy, and policy for DSI in relation to the Digital Agenda for Europe and under the Horizon 2020 Work Programme, and in particular, but not limited to, the Collective Awareness Platforms (CAPS) Programme.[4]

[4] http://ec.europa.eu/digital-agenda/en/collective-awareness-platforms-sustainability-and-social-innovation.

We identified six primary domains of digital social innovation. (1) *The Collaborative Economy*: the rise of digital marketplaces for people to transact and share skills, assets and money is fast becoming a key economic trend that enables people to share skills, knowledge, food, clothes, housing and so on. The Collaborative economy has been documented by organizations like Couchsurfing and OuiShare.[5] (2) *New ways of sensing* is a vibrant ecosystem of makers is developing across Europe and globally. Low-cost home 3D manufacturing tools (3D printers, CNC (computer numeric control) machines), free CAD/CAM software like Blender, 123D or Sketchup and open source designs are now giving innovators better access to the enabling infrastructures, products, skills and capabilities they need to enhance collaborative making. (3) *Open democracy* strives to create opportunities for all members of a population to make meaningful contributions to political decision-making. Organizations and projects pioneering open democracy, large-scale feedback, and citizen participation through crowdsourcing legislation such as Open Ministry or Liquid Feedback are transforming the traditional models of representative democracy. (4) *Awareness networks*, including cities including Vienna and Santander, are transforming governments, businesses and society by pioneering new practices in open data and open sensor networks that are changing the provision and delivery of public services. (5) *Open Access* exploits the power of open data, open APIs, and citizen science to provide citizens better public services, with projects like CitySDK defining interoperable interfaces for city-scale applications. (6) In terms of *funding and incubation*, as has been the case with the support for innovative businesses, social innovations often need support in the early idea stages to refine their business models and grow their venture. Incubators like Nesta typically support innovators in exchange for equity, at pre-seed or seed stage. There are nearly 100 incubators/accelerators in Europe.

In terms of methods, we identified four primary methods. *Open networks* includes wireless sensor networks, community (bottom-up) networking (such as Freifunk and Guifi), and privacy-aware open networks as well as hardware such as Arduino. *Open data* includes how governments and other large organizations and companies that hold or generate data about society can release their data to enable citizens to hold government to account for what it spends, the contracts it gives, and the assets it holds. The Linked Data community is a prime example. *Open knowledge* covers the variety of ways in which citizens can use online services and platforms for mass scale social collaboration. Ordinary people today use blogs, wikis, social networks and hundreds of other collaborative platforms to manage their daily lives, solve social challenges, and to participate in e-campaigns, crowdfunding, and the like. *Open hardware* consists of hardware whose blueprints are made publicly available so that anyone can study, modify, distribute, make, extend, and sell the design or hardware based on that design.

[5] Note that DSI entities with proper names such as Ouishare mentioned are defined in the DSI case study document: http://content.digitalsocial.eu/wp-content/uploads/2014/05/DSI-report-casestudies.pdf.

The work by organization like Raspberry Pi and Arduino illustrates the potential in open hardware.

3 Crowd-Mapping with Linked Data

As DSI is a relatively new field of study, there is little existing knowledge on who the digital social innovators are; what types of activities they are involved in and how they are using digital tools to achieve a social impact. Therefore, the first task of European Commission was to look in more detail at the different types of organizations involved with DSI, and the activities these organizations are involved in. This led to the creation of a formal vocabulary to capture the characteristics of DSI organizations, which was formalized as an RDF(S) vocabulary, as can be browsed[6] along with various necessary pieces of data such as name, address, size (number of employees), and so on. The DSI ontology is available also for re-use.[7] It crucially depends on inter-linking with other vocabularies such as the Datacube vocabulary[8] and FOAF. In particular, we needed to capture the following characteristics for each DSI organization (where the categories for each role are given in parenthesis):

1. A typology of organizations (government and public sector organizations, businesses, academia and research organizations, social enterprises, charities and foundations; and grassroots communities);
2. The way these organizations are supporting DSI (undertaking research, delivering a service, organizing networking events and festivals, etc.);
3. The main technological trends the organizations and their activities fit under (open data, open networks, open knowledge, open hardware);
4. The area of society the organizations and their activities operate and seek an impact in: The DSI field does not have fixed boundaries; it cuts across all sectors (the public sector, private sector, third sector and social movements) and cuts across domains as diverse as (1) health, wellbeing and inclusion; (2) innovative socio-economic models (3) energy and environment; (4) participation and open governance, (5) science, culture and education; and (6) public services.

In order to understand the DSI landscape, large amounts of technical work was needed in order to build the foundations for a solid map. First, a number of diverse data-sets already collected about digital social innovation needed to be transformed into a common format, in particular data from traditional Excel spreadsheets and SQL databases. Yet this data often was incomplete and partial in terms of the kinds of information we needed for our study. Also, we considered that as we began our study, we would want to capture increasingly large amounts of data as well as perhaps pruning unnecessary or under-utilized categories.

[6] http://data.digitalsocial.eu/data/organizations-and-activities.
[7] http://xmlns.com/foaf/spec/.
[8] http://www.w3.org/TR/vocab-data-cube/.

Lastly, we wanted to do complex and open-ended data analysis over our data on-line before commencing the final offline data analysis. While SQL databases were considered, their inability to deal with dynamic schemas easily, and the difficulty of querying traditional "NoSQL" databases is well-known. In this regard, Semantic Web technologies were considered the best fit, as they easily coped with incomplete data, dynamic schemas, and complex querying via SPARQL. We then used off-the-shelf standards such as RDB2RDF and CSV2RDF[9] to convert existing data over to RDF, and where possible supplemented the data for well-known organizations with RDF-compatible data from DBPedia and OpenStreetMaps.[10]

Pre-existing data on DSI rarely captured the relationships, in particular the social network of which DSI actors were working with what other actors on particular DSI projects. For example, the city of Helsinki was working with both international networks like Open Knowledge Foundation as well as national groups like Open Ministry, and European-level groups like "Code for Europe," to open their data. Thus, we thought that the social networks of DSI actors would naturally map onto RDF's network-like data structure, and we could use a survey to both add new DSI organizations as well as gather information about their collaborative projects. While a graph database would have allowed a similar analysis, it would not have allowed data to be added 'on the fly' without manual intervention but only via the survey. In order to gather this data, we created a RDF-backed website, built using Ruby on Rails, that combined a Sesame-based native RDF backend that was initially structured according to our DSI RDF(S) vocabulary and seeded with Nesta's information as transformed to RDF. This allowed us to create a "map" by projecting the RDF data with geolocation information onto our own instance of OpenStreetMaps, which then could present the social network of DSI actors visually. An example of this is shown in Fig. 1. When new visitors came to the website, they were asked to "Get on the Map" by filling out a simple survey that fit their organization and projects into the various categories presented earlier. When they listed partners in projects, we asked for contact information in order to get in touch with partners.

4 Semantics for Surveys

One crucial question is what advantages does Semantic Web technologies provide for this kind of survey-based data collection? When the work was first started, the initial group of contacts came from Nesta's pre-existing DSI work, which led to a map of 32 organizations that were studied using traditional case-study based methodology. From the case-study methodology, an initial RDF(S) vocabulary and online survey was created that was sent to contacts found from Nesta and partners (Waag Society, IRI, ESADE Business School) contacts database, converting them to use RDF using the aforementioned semantic pipeline. Starting in October 2014, this traditional online survey-making work continued till May

[9] http://data-gov.tw.rpi.edu/ws/csv2rdf.html.
[10] http://wiki.openstreetmap.org/wiki/OSM_Semantic_Network.

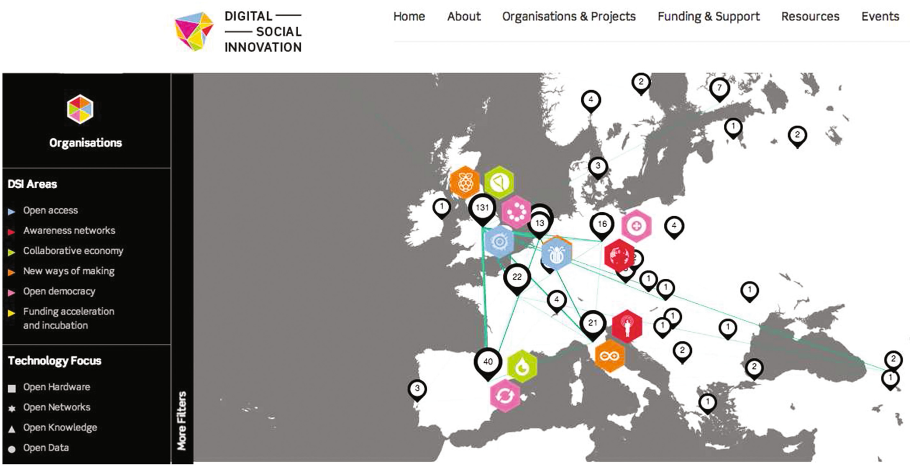

Fig. 1. Screenshot of RDF-backed DSI map

2014. This led to a total of 32 case studies from the contacts of the 8 original organizations, and increase of 300.25 %. However, only "friends of friends" of our partners were being added to the data-set.

We embedded our RDF data-set and interlinked with existing data via named entity recognition [6] and then sent the results to DBPedia Lookup.[11] This allowed the discovery of organizations, names, and e-mails in the data-set via SPARQL queries. This was then interlinked via missing geographical areas using LinkedGeoData[12] (based on OpenStreetMaps) via pre-made DBPedia mappings in order to analyze the geographic distribution of organizations in the data to determine organizations that were missing. Although we could not check the precision of Edinburgh LT Named Entity Recognition system on this particular data-set, the accuracy over a 100 entity sample was 97 %. Furthermore, an initial categorization of DSI into nine areas of society impacted was made, which took advantage of the schema-less nature of RDF to see what kinds of DSI were missing. Then the survey was closely linked into the new schema, and outreach was done to these geographic areas, domains, and organizations via names and contacts discovered in the larger interlinked data-set. Outreach with the semantic-enabled survey far surpassed the non-enriched survey, as the number of organizations in the data-set grew to 581 by August 2014. This was an increase of 480 %, in comparison with the increase of 300.25 % without semantic technologies.

After the initial burst of activity, we added four new categories of DSI to the schema to help guide the outreach. However, at this point we had reached the limit of using semantics to help pull new organizations in and so had reached the

[11] https://github.com/dbpedia/lookup.
[12] http://linkedgeodata.org.

limit of our data-set, so we mapped 967 DSI organizations and 609 collaborative projects as of January 2015 (an increase of only 166.44 %). From our domains (given that a single organization could be in more than one category), there were 412 open knowledge, 269 open networks, 258 open data, and 105 open hardware projects. In terms of areas of society impacted, there were 254 organizations focusing on education, 251 on democracy, 164 on arts, 163 on health, 162 on employment, 138 on neighborhood regeneration, 130 on the environment, and 110 on science, and 104 on finances.

5 Network Analysis

We have hypothesized that the success of DSI is located in the social structure of the larger network of social innovation. Social networks are formally defined as set of nodes (or network members) that are tied by one or more types of relations. In the case of the DSI social network collected in this study, the nodes in a graph are organizations, and the edges represent joint projects. The results of this analysis have informed the recommendations on a policy. The data-set and resulting policy changes are needed for the EC to knit the map of DSI actors into a coherent single integrated EC DSI network, and thus achieve the "critical mass" necessary to harness the collective intelligence of DSI organizations to solve large-scale European social problems. At the time of our data analysis,[13] there are a total of 930 organizations with a total of 588 shared projects in the DSI data-sets. This data-set likely fairly represents the empirical phenomena at hand with two caveats (1) It has a bias towards English speakers as the survey was not translated into other European languages, and (2) as outreach was directed by the partners it reflects their social networks likely more in-depth than disconnected social networks. However, it is a large sample and thus worth exploring in detail. A number of questions were determined by the European Commission, which we then answered with a combination of SPARQL, network analysis tools such as Gephi,[14] and custom scripts written in Python and Perl.

In the DSI data-set, nodes (also called vertices) are organizations and shared projects are edges (also called links) between nodes. Although the edges are directed in the native RDF data-structure, we assume that if an organization included another organizations in its project and the aforementioned project did not, that was an oversight of the latter project, not an error on the behalf of the former. Thus, the edges are undirected. In this paper, we do not provide detailed mathematical definitions of the common network analysis constructs used but quick qualitative definitions, as quantitative definitions can be found in final report.[15]

[13] Nov. 13th 2014.

[14] http://gephi.github.io/.

[15] http://content.digitalsocial.eu/wp-content/uploads/2014/05/DSI-report_final_19.05.2014.pdf.

5.1 What Is the Distribution of Social Innovation Across Europe?

Is social innovation done by a few large actors or evenly distributed between various actors? The answer is social innovation in Europe is currently done often by a few large actors in concert with a large mass of smaller organizations, but the majority of social innovation actors in Europe are disconnected from these networks.

In order to determine this, for all the organizations we mapped their degree, which is for a given node (organization) the number of connections (links) it has with other nodes (organizations). There are 243 organizations with connections to other organizations (26 %). The average number of connections per organization is almost 3 organizations. There are a few organizations that are hyper-connected "hubs," including the Waag Society, Nesta, Fondazione Mondo Digitale, and Institute for Network Cultures.

5.2 What Communities of Social Innovation Exist in Europe?

A community exists when a network is partitioned in such a manner that nodes within a community are more densely interconnected than those outside of the network. The clustering coefficient is way to understand this quantitatively. A triplet consists of three nodes that are connected by either via two nodes (a straight line) or three links undirected, where a triangle is defined by three closed triplets, where each node a triangle is part of a triplet. The global clustering coefficient is the number of triangles over the total number of triplets. This measure gives an indication of the clustering in the whole network. The global clustering coefficient of the DSI network is 0.875 with undirected links.

Modularity is the fraction of number of the edges that fall within a given community minus the expected fraction of the number if edges were distributed at random. The modularity of the DSI network is .65, where modularity is the percentage of the connections that fall within the given community minus the expected such links if they were random [7]. There are approximately 115 distinct disconnected communities of social innovation (measured in terms of connected components that are connected to each other but not connected to anyone else). Although there is one large pan-European network, there are also many smaller communities do not have connections to the larger cross-European digital social innovation "super-community." The vast amount of disconnected communities is visualized in Fig. 2.

Communities can also be automatically identified are mostly small by optimizing around modularity [1]. The results are given in Fig. 3, where a few large communities stand out from each other. These inter-connected communities only count for 28 % of the total amount of connected DSI activities. The largest community (green 10.29 %) is focused around open hardware and open networks and includes organizations such as iMinds, Fairphone, the City of Amsterdam, and FabLab Barcelona, despite it being the smallest category of DSI methods. Its most interconnected member is the Waag Society, and there is a large focus on awareness networks and new ways of making. The collaborative economy

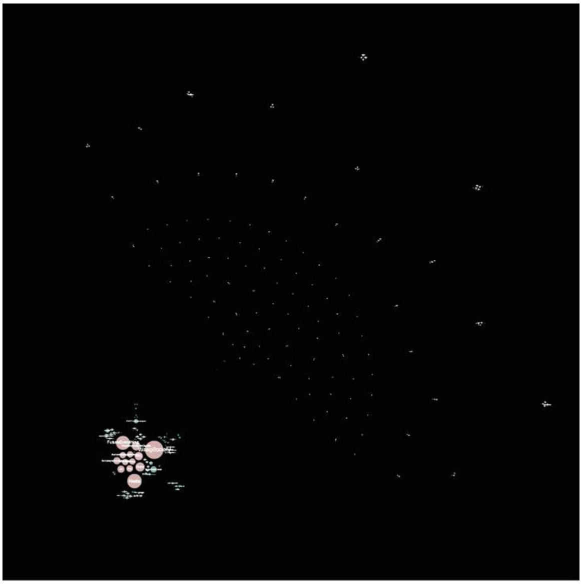

Fig. 2. Connected and dis-connected communities in the DSI network.

and open knowledge is the specialty of the second largest but also more scattered community (red - 7.41 %), consisting of ESADE, IRI, European Institute for Participatory Media, and the Institute for Network Cultures. A third large communities is grouped around Nesta (blue - 5.35 %) and is focused on funding, acceleration, and open democracy, although it has a very diverse technology focus, containing groups such as Open Ministry, Nominet, and Mozilla. Open data for open access is the last large community (purple 4.95 %), with a centre on FutureEverything, but also containing Open Knowledge and its local chapters as well and city councils working on open data such as that of Salford. Interestingly, although the open hardware network is the smallest overall, it is the most highly-interconnected and intermixed with open networks. Open knowledge is the most popular technological focus of DSI but it also the most spread out and disconnected. Other communities, such as those around open data, are developing connected communities. Nonetheless, the vast majority of communities are not interconnected.

5.3 Which Organizations Currently Bridge the Various Communities?

How can we determine which organizations act as crucial "bridges" between different organizations in DSI? This can be measured by using betweenness centrality, where the centrality of an organization is measured by counting the number of times a node occurs as a shortest number of links between any other nodes [3]. Betweenness centrality is equal to the number of shortest paths from all nodes to all others that pass through that node. The betweenness centrality was done using Brandes' algorithm [3]. It also calculated the total network diameter of 7

Fig. 3. DSI organizations in automatically discovered (colored) communities

and average path of 2.75. By this metric, these central organizations are: Waag Society, Nesta, Future Everything, Fondiazione Mondo Digitale, Kreater Social Innovation Agency, Forum Virium Helsinki, Swirrl, Open Knowledge Finland, IRI, BettterPlaceLab, Alfamicro, Amsterdam Smart City, Alfamicro, European Institute for Participatory Media, and ESADE. Each bridging of these organizations brings over 70 organizations together using the shortest possible number of links.

Who connects the diverse DSI communities, such as those of open data, open knowledge, open hardware, and open networks? What is more interesting than just being well-connected is whether an organization bridges diverse parts of the network with the greatest connectivity. Even if a organization is not central and so has only a few links, it may be these few important links that connect communities of well-connected organizations. Eigenvector centrality is a measure of reach (or influence) of a node in the network, and is thus important to quantify this idea of bridging diverse communities. It differs from betweenness centrality insofar as links to heavily ranked nodes contribute more, and so is closely related to the well-known Google Pageranking algorithm. If we look at the this kind of centrality, we see that a number of new organizations are crucial in bridging diverse communities (over .6 eigenvector centrality) outside of the original list of central organizations bubble up to the top: Institute of Network Cultures, iDROPSzw, Elva Community Engagement, Arduino, and Fing. To encourage cross-hybridization of different kinds of social innovation, special effort should be made by the European Commission to strengthen these digital crucial connectors between diverse DSI communities. Interdisciplinary European projects that force diverse communities to work together would strengthen the overall resilience of DSI in Europe by combining open hardware, open data, open knowledge, and open networks.

5.4 What Are the Conditions for Scaling DSI?

Successful actors in DSI have managed to leverage large networks using the Web in order to accomplish innovation at scale by the network effect. We can define scale in terms of scale-free, namely that the distribution of DSI should undergo the phase shift typical of complex systems from a disconnected network to a highly interconnected and self-similar small world network, where communities are clustered. This scale free network is often seen in organically developing eco-systems and is thought to be a sign of efficiency and resilience [2]. Note that power-laws have an important property called "scale invariance," which means that when scaling the network by a constant (i.e. multiplies the original power-law relation by the constant) only causes proportionate scaling of the function itself, which shows the underlying distribution is stable.

We used a set of techniques that test to see if a power-law existed in our data using the Kolmogorov-Smirnov distance (a non-parametric information-theoretic measure of the similarity between distance) on a set of simulated data of a power-law and the actual data. We used techniques outlined by Clauset et al. [4] to find a power-law possibly after $x = 3$ with and an α of 3.13. and a σ (standard deviation) of .19. We found a Kolmogorov-Smirnov statistic D of .1004. There was a p-value of less than .01. When tested rigorously, a power-law was indeed a strongly better fit ($p < 0.01$) than an alternative distribution without such a long-tail, such as the exponential distribution [4]. An exponential distribution does not have the majority of its strength in the long tail. However, the power law appears truncated, so that the power-law behavior may only actually apply to organizations with above 3 connections. Figure 4 shows the comparison between the power law distribution and the exponential function of the DSI network.

The reason digital social innovation has not yet scaled is because the long tail of smaller European DSI Networks is still heavily disconnected, with 687 organizations out of 930 (74 %) have no links to other organizations! Many of these organizations are also in countries without much support, such as Eastern Europe. If we want a single scaling European DSI network, an additional magnitude more of links (approximately 350 links) is needed to gather all the disconnected organizations to a single European network. This is probably too many connections that can be made via traditional European projects, but via a recommendation system a future version of the Digital Social Innovation website could introduce innovators to both other local innovators and innovators sharing similar interests across Europe to boot-strap these connections. By connecting the currently isolated innovators, we should be able to achieve the necessary phase shift so that the scaling power of the heavily interconnected innovators is replicated across Europe by currently isolated innovators and communities. Globally, we are already interested outside of Europe (such as Harvard Ash Center) to re-use the ontology and eventually link data-sets.

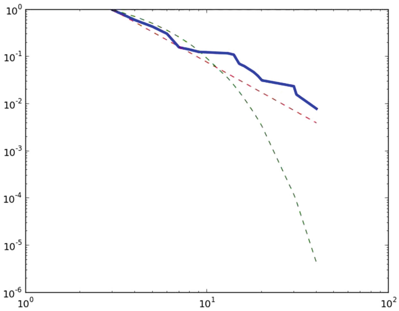

Fig. 4. Comparing the power law distribution (red) to exponential distribution (green) against the real actual network data (blue) (Color figure online).

6 Data-Driven Policy Recommendations

Based on the data, a number of recommendations were made to the European Commission. The full report, with much more detail, is available online.[16] In order to implement future DSI policy goals and strategies, several tools and instruments have to be deployed. Although most policy influencing DSI will be at national, regional and local levels, it is clear the European Commission has an important role in networking these new actors in the field due to vast disparities in connections. Rather than through traditional networking events, the European Commission can then use the network data to find the small disconnected communities and introduce them based on national and sector-specific networking i.e. around health, money, and education.

Crowdfunding is a promising collaborative approach for such bottom-up networks in comparison to traditional European large-scale research projects, which are often too "heavy-weight" and require larger networks than most of the DSI organizations in Europe have. However, while traditionally crowdfunding depends on individuals donating funding, there is no reason why crowd-funded projects cannot receive in-kind funding from the public sector or have their general structure be replicated using traditional research funding. The European Commission should start promoting more of these crowdfunding tools for DSI, involving users in choosing the best projects to be funded in a bottom-up fashion, as part of their funding allocations already in place for ICT research and development programmes in. Note that the CHEST project has already begun in this fashion using a traditional EC-funded project structure to start a crowd-sourcing campaign.[17] Another example is the 'European Social Innovation Challenge that in 2013, the European Commission launched in memory of Diogo Vasconcelos, to encourage new social innovations from all over Europe. The Competition invited Europeans to come up with new solutions to reduce unemployment with the three winning projects will be awarded financial support. By leaning on the

[16] http://content.digitalsocial.eu.
[17] http://chest-project.eu.

large DSI networks already existing around funding and acceleration, progress could be made quickly, but care needs to be taken to help disconnected DSI organizations. Given the large network of DSI on open hardware, more focus on how open hardware is necessary for trusted local hardware would be of use. More work is needed in opening public sector information, as the smaller but still substantial DSI communities focused on open data.

Open knowledge is comparatively doing better than open data in terms of DSI, and copyright reform is necessary to ensure its growth. The European Commission recently published its 'Report on the responses to the Public Consultation on the Review of the EU Copyright Rules.' This report summarizes the responses (over 11,000) that the Commission received in response to the copyright consultation held between December 2013 and March 2014. The results show conflicting positions between citizens and institutional users on one side and corporate rights holders on the other. Such patterns in public policy consultations show that stakeholders involvement is crucial, and that the Commission should take advantage of the increased user involvement in open knowledge, in particular to channel energy towards DSI.

Perhaps one of the largest recommended changes is, given the large number of DSI organizations in Europe, is to open public procurement to these organizations. In January 2014, the European Parliament adopted new procurement directives on PPI (Public procurement of innovative solutions) featuring increased flexibility and simplification on the procedures to follow, negotiations and time limits; clearer conditions on how to establish collaborative or joint procurement; and the creation of innovation partnerships. A review of procedures in public procurement is needed in order to include actors from grassroots DSI communities in procurement.

Lastly, the preservation of net neutrality to allow DSI organizations to use the Web to scale their services. It is a crucial to define and make public how network operators manage traffic volumes and restrict applications usage. Not only does net neutrality protect the freedom of expression and freedom of information online, it reasserts the principle of fair competition and guarantees that users may freely choose between services online, and thus is crucial to help DSI actors reach scale. Otherwise, large existing organizations that may not be as innovative could lock out smaller DSI actors rather than working with them collaboratively. One final note is that we may even eventually see DSI applied to policy-making directly. For example, Tim Berners-Lee (the inventor of the Web) is advocating for a sort of Magna Carta for the Web to establish basic rights and freedoms where the Magna Carta for all Web users could be directly crowd-sourced from the Web itself, engaging effectively in multi-stakeholder processes.

7 Conclusion and Next Steps

The DSI works shows how Linked Data can be used to combine traditional sources of data with survey-collected data in order to create a crowd-mapped data-set around new and innovative areas like digital social innovation. Through the website digitalsocial.eu, we have shown how visualizations and mapping can make Linked Data comprehensible by policy-makers. Through our network analysis how the various characteristics of DSI can be both understood structurally and how this understanding can produce concrete policy-level recommendations.

As a site digitalsocial.eu is still active, and but it is not in general interactive enough to get existing organizations to up-date their activities. For this to be the case, there would have to be concrete benefits. One idea is to tie new funding initiatives either directly or indirectly (via announcements) to the site. If recommender algorithms could be combined with the community clustering algorithms already used, we could effectively create a "LinkedIn" of DSI that directed new and existing organizations to like-minded organizations for partnership in projects and applications for funding. Combined with appropriate internationalization and reach outside of Europe, Linked Data could provide the infrastructure for growth of digital social innovation not only in Europe, but in the entire world.

References

1. Blondel, V., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech.: Theory Exp. **10**, 1000 (2008)
2. Boisot, M., McKelvey, B.: Connectivity, extremes, and power laws: towards a power-law science of organizational effectiveness. J. Manag. Inq. **20**(2), 119–133 (2011)
3. Brandes, U.: A faster algorithm for betweenness centrality. J. Math. Sociol. **25**(2), 163–177 (2001)
4. Clauset, A., Shalizi, C., Newman, M.: Power-law distributions in empirical data. SIAM Rev. **54**(4), 661–703 (2009)
5. Freeman, C., Soete, L.: The Economics of Industrial Innovation, 3rd edn. MIT Press, Cambridge (1997)
6. Mikheev, A., Moens, M., Grover, C.: Named entity recognition without gazetteers. In: Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics, EACL 1999, pp. 1–8. Association for Computational Linguistics, Stroudsburg (1999)
7. Newman, M.: Modularity and community structure in networks. Proc. Nat. Acad. Sci. U.S.A. **103**(23), 8577–8696 (2006)