Food Data Driven Business Innovation

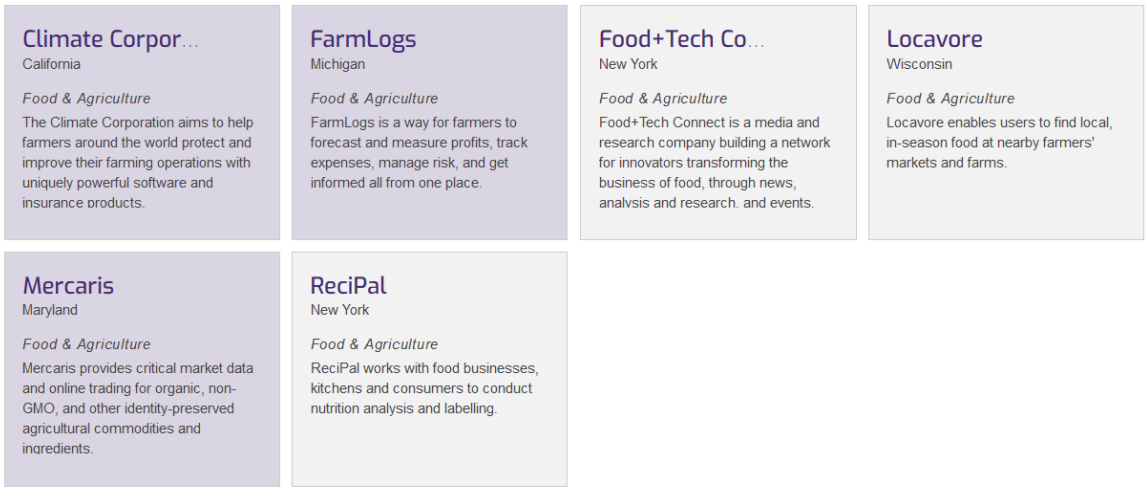
Heidi Gautschi, Gianluigi Viscusi, Christopher Tucci

1. Introduction

Digital developments are disrupting industries in both mundane and surprising ways. The potential for unlocking meaningful information in "Big Data" is one way in which industries are being transformed. In this specific report, we focus on the role of open data in the food industry as a potential catalyzer for change. The food industry is very broad and includes everything from transportation and logistics, mass production of food, local farms and farmers markets, retail outlets and hospitality and tourism. The level of disruption these various subsectors is facing is different from one another and therefore, the readiness for open data is also different. This report offers a broad overview of how open data is being used at the moment and perhaps will be used in the future.

For the past few years, open data as a source of business innovation has gained traction. Many of the reports and studies published take a macro-level view of the impact of open data on GDP and innovation. One initiative that is providing a more granular look at who is using open data and for what purpose is OpenData500. "The OpenData 500 is very useful," according to Daniel Castro, Director at the Data Innovation Center in Washington D.C. "What we have been struggling to do is to make the connections very clear between economic and social benefits. What the OpenData 500 does is, it creates a map of where the innovation is occurring. For example, it lets Members of Congress see that companies are creating jobs in their districts using open data and lets them ask these businesses about how their business model works" (Boyd 2014). Still, there are only five food & agriculture-related datasets available on the OpenData500 website, showing that there is still room for growth in the food industry.

Figure 1 Food & Agriculture datasets on opendata500.com



This report is structured as follows: We begin by discussing what data is and what type of value it can bring to business. We then provide snapshots of how open data is being used by different companies to bring added value to their business propositions. We then discuss the Open Food Data Hackdays and the winners of this event. We provide some high level recommendations on how the Hackdays and the support given to the winning teams could be more turned towards business innovation and business development. Finally, based on focus group results, we provide insight into how open data can be a catalyst for business innovation and what applications are being used by start-ups.

2. Data

The interest in open data has increased in the last ten years with a consequent growing number of initiatives. These initiatives are mainly tied to the idea of open government, symbolically enforced at the global level by the memoranda and directives signed by Barack Obama in the early years of his first mandate (Chignard 2013; Obama 2009). In particular, open data has represented a key issue for the digitalization of the public sector and the accessibility of governments' information assets. As pointed out by Janssen et al. (2012), open data implies changes at a systemic and institutional level, due to the moving from closed to open systems and a consequent reinforcement and transformation of institutional structures. Economic benefits have been recognized as related to open data such as

stimulation of innovation and competitiveness, improvement of processes, products, and/or services as well as the development of new ones related to the creation of new sectors adding value to the economy with the availability of information for investors and companies. Myths and barriers still prevent the full exploitation of the value of open data, as well as the development of appropriate and effective initiatives. Some that have been identified by Janssen et al. (2012) for open government data are also relevant for open data in the private sector and business domains, such as the publicizing of data. All public information should be unrestrictedly publicized and it is a matter of simply publishing the data so that every constituent (user) can make use of open data.

Data spectrum (open to closed)

The definition of open data by Open Knowledge International states "*Open data is data that can be freely used, re-used and redistributed by anyone—subject only, at most, to the requirement to attribute and share alike"* (Open Knowledge International 2017b). Notwithstanding the clarity and appeal of this definition as well as the number of resources made available for detailing the issues at stake and improving data science skills, the main barriers for laypersons are understanding the types of data and the capacity of current technological infrastructures and human resources to maintain, produce, and use open data. According to the Open Data Institute's data spectrum different types of data (ranging from *closed* to *open*) are available with differences in terms of *volume* (small, medium, big data), *ownership* (personal, commercial, and government data), *access* (internal, named, group based, public access, and open license). An important issue is related to the *license* associated to each dataset, ranging from the contracts typical of *closed data* to the open license of *open data* through authentication required by *shared data* (The Open Data Institute 2017).

Licensing is related to the requirement for open data to be *legally open*; a further requirement for open data is to be *technically open* (Open Knowledge International 2017a).

Thus, besides licensing, another relevant issue is related to the different types of data (text, statistics, images, maps, videos, etc.) and data types available in the spectrum ranging from

unstructured (not having a pre-defined data model such as textual data), structured (organized in relational databases) to semi-structured data (markup languages such as Extensible Markup Language—XML or open standards formats such as JavaScript Object Notation—JSON). The appreciation of these different types of data is important to open up data that are in formats not only understandable and readable by humans but also by machines.

The issues related to the Data Spectrum are strictly connected to three key challenges for open data exploitation by laypersons and management by public as well as private organizations: quality of open data, digital asset management, and value drivers. In what follows we briefly discuss each of these.

Quality dimensions of Open Data

We provide a brief overview of dimensions worth considering when dealing with open data and open linked data, the latter generating semantic connections among datasets (Tim Berners-Lee 2006). In particular, we consider the following cluster of dimensions as identified by Batini & Scannapieco (2016, 99–110) and Batini et al. (2015), focusing on the ones relevant to open data: *accuracy*, *completeness*, *readability*, *accessibility*, and *trustworthiness*.

Accuracy refers to the extent to which entities and facts correctly represent a real-life phenomenon (Carlo Batini and Scannapieco 2016), and can be classified as *syntactic accuracy* (the degree to which a document conforms to specifications in terms of format, etc.) and *semantic accuracy* (the degree to which data values correctly represent real-world facts). Other dimensions, related to the temporal facets of accuracy are *currency* (how promptly data are updated) and *timeliness* (how current the data are for the task at hand).

Completeness refers to the degree to which all required information is present in a particular dataset*,* encompassing *schema completeness* (the degree to which the classes and properties of an ontology are represented); *property* completeness (the amount of missing values for a specific property); *population completeness* (the percentage of all real-world

entities of a particular type represented in the datasets)*; linkability completeness,* (the degree to which instances in the dataset are interlinked) (Carlo Batini and Scannapieco 2016).

Relevance refers to the provision of information which is in accordance with the task at hand and important to the users' query. As for *readability,* we consider *understandability* as a key dimension referring to the "ease with which data can be comprehended without ambiguity and be used by a human information consumer" (Carlo Batini & Scannapieco, 2016, p. 106).

While the discussion about the data spectrum has already provided insights into the dimensions related to *accessibility* such as *licensing*, *availability, and interoperability*, here we also consider dimensions for *trustworthiness*. The following dimensions are actually relevant when information is provided in the "wild" of Big Data ecosystems (C Batini et al. 2015): *believability* refers to certainty that data is true and credible, for example acquiring trusted content from users, based on associations that transfer trust to resources (Carlo Batini & Scannapieco, 2016, p. 424); *verifiability* is the degree a data consumer can assess the correctness of the data set; *reputation* is a judgment made by a user to determine the integrity of a source, associated with a data publisher, a person, organization, group of people or community of practice, or it can be a characteristic of a dataset (C Batini et al., 2015, p. 75).

To summarize, data and information quality are relevant to published open data, but they also represent a strategic issue for the management of the information asset by public and private organizations producing the data worth opening.

Open data and Information Capacity of organizations

As pointed out by Janssen et al. (2012), among the myths of open data we can include the belief that all information should be unrestrictedly publicized and that it is a matter of simply publishing public data. While the former belief concerns the issues discussed above about the data spectrum and open data quality, the second myth will be further developed

in this section from the perspective of management of information assets within an organization and in the following Section from a policy perspective, with a specific focus on value drivers for open data.

Let us discuss the management of information assets within an organization as a basis for the production and maintenance of open data worth publishing. In particular, we consider the role of information capacity, which can be defined as *the current stock of understandings informed by a given installed base* (Viscusi and Batini 2014). In fact, data worth opening are the source for a set of understandings by the final users. As argued by Viscusi & Batini (2014, p. 81) the information capacity of an organization in managing represents "*the potential of a digital information asset that can be defined and evaluated independently from the usage, on the one hand, determining the economic utility of a digital information asset; on the other hand, enabling capabilities providing a social or else public value perspective.*"

Accordingly, as noted by Viscusi & Batini (2014) information capacity is strictly connected to the capabilities within an organization that are necessary to produce and maintain the digital asset which is the source for open data worth publishing. On the other hand, it is influenced by other key components of an organization's information system that have already been mentioned: data information quality dimensions, the information structure (the degree of integration of the available data structure), and information infrastructure (e.g., actual data base management systems, enterprise systems, data integration technologies, communication technologies, etc.).

Value drivers and digital governance

Open data can provide value to organizations, either private or public, not only in terms of *economic value* (through the development of new products and services, availability of information for investors and companies, optimization of administrative processes and consequent savings, etc.), but also in terms of *public value*: "*What does the public most value?*" and *social value*: "*What adds value to the public sphere?*" (Benington 2011; G. Viscusi, Castelli, and Batini 2014).

The three kinds of value depend on a corresponding set of drivers (*performance*, *openness*, and *inclusion*, respectively associated with economic, public, and social value), which should be considered when assessing the requirements and the current impact of an open data initiative. Furthermore, as shown in Table 1, each value driver can be measured by a set of quality dimensions which have an impact at different levels of analysis. They should be applied at a legal framework, service, organization, and technology level. However, it is worth noting, on one hand, that the above-discussed accuracy and completeness dimensions are relevant to leverage efficiency and effectiveness enforcing performance as value driver; on the other hand, accessibility of data is key at the information level for openness as a value driver as well as for equity, together with readability and trustworthiness.

Table 1. Value drivers, quality dimensions and levels of analysis

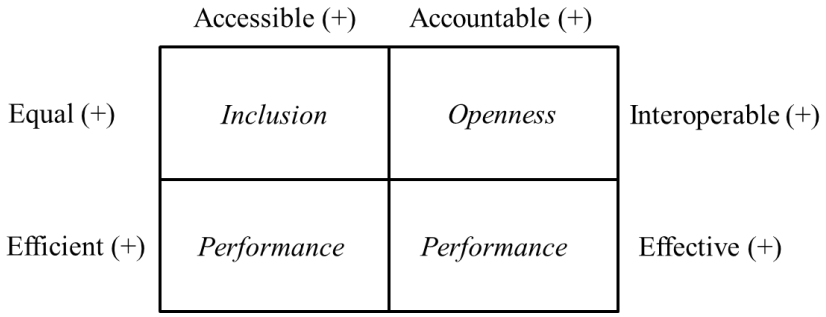
Value Drivers	**Quality Dimensions**	**Levels of analysis**
Performance	Efficiency	Legal framework
		Service
		Technology
	Effectiveness	Service
		Organization
Openness	Transparency	Service
		Organization
	Accountability	Organization
		Legal framework
	Accessibility	Organization
		Information
		Technology
Inclusion	Accessibility	Service
		Technology
	Equity	Organization

		Information

Source: Adapted from Misuraca & Viscusi, 2014a

The diverse value drivers once identified define a specific "attitude" of digital governance for the organization opening or willing to open their own data to the public (Misuraca and Viscusi 2014b), as shown in Figure 2.

Figure 2. A typology of digital governance attitudes



Source: adapted from Misuraca and Viscusi, 2014b

The performance driven types represent the basis for the other types due to their focus on having accurate and/or complete data sets, as in the case of the "efficient and accessible" type. The "effective and accountable" type has a further focus on the trustworthiness of data and their provenance can increase the degree of transparency and accountability of an organization. The two performance driven types have a degree of accessibility which can yet be named or group based, according to the data spectrum, thus neither necessarily public nor shared. The openness-driven type adds a higher degree of interoperability to the trustworthiness of data allowing sharing and eventual opening to diverse user. That said, the openness driven type still has an orientation moving from the organization point of view rather than the one of the potential final users: data are opened to the public, possibly with higher standards of accessibility and/or interoperability, but without a real understanding of the capacity of final users to exploit them and without a clear sense of their needs either from a social or economic value perspective. To this end the inclusion-driven types allow for a change of orientation in the opening of data considering readability and understandability; coupling accessibility with an equal access to the largest number of available users, thus a

better understanding of their capabilities, needs as well as the potential value of open data in the last end of the data spectrum.

3. The open data ecosystem: illustrative use cases

In this section, we provide a snapshot of how various stakeholders in the food industry are using and/or creating open data. Government agencies, international initiatives and multinationals are all opening some of their data. And this makes sense. Given the vast amount of data the US Department of Agriculture (USDA) collects, for example, the agency itself would be hard pressed to analyze and interpret all of it. By making many of their datasets available to the public, the USDA is incentivizing businesses, individuals, and non-profits to use the data instead of having it collect virtual dust on servers somewhere. The same argument can be used for large multinationals. These types of companies are also collecting vast amounts of data that are not being used by the organization. Opening certain datasets may make sound business sense.

The examples below are not exhaustive. There are many more examples, but the ones chosen below illustrate the different types of open food data now available.

Organization: GS1

Brief description: GS1 is known as the "barcode company", but it is much more than that. GS1 is a global non-profit organization that develops supply chain standards for retail, healthcare, transportation, and logistics. By adopting GS1 standards, companies can identify, capture, share and use data about products and sales, for example. In the foodservice industry, GS1 standards gives one the capability to trace a fish from the place it was fished to its arrival on one's plate.

Examples:

1. Customers at global retailing company METRO Cash & Carry can now scan barcodes for seafood products to understand where the fish was caught, how it was caught,

the best-by-date, and other important data – all information provided directly by fish suppliers ("Wildcatch — GS1 Discover Foodservice").

2. Swiss grocery retailer Coop was able to reduce inventory in its stores by 8% and at the same time managed to increase on-shelf availability by automating the ordering of its fresh produce. ("Inventory Management – GS1 Discover Foodservice").
3. With the growth of smartphone apps and online food ordering, coupons can now be used in many more channels to reach consumers. To keep things simple, the GS1 Digital Coupons Management standard provides a common secure and reliable process for managing digital coupons—by foodservice operators, restaurants or takeaways. ("Digital Coupons—GS1 Discover Foodservice").

Open data initiatives: Based on the information on the GS1 website, it would appear that opening, or closing the data is the company's choice. It would seem that some data is shared with certain companies along the supply chain. More than anything, GS1 illustrates how standards for identifying, capturing, sharing, and using information about products along the supply chain can improve efficiency, reduce costs and improve traceability.

Organization: Syngenta

Brief description: Syngenta is a Swiss agribusiness multinational. The company focuses on six core values: make crops more efficient, recue more farmland, help biodiversity flourish, empower smallholders, help people stay safe, and look after every worker. They use cutting edge science and technology to improve crop yield and food safety. Syngenta is present in 90 countries. In 2016, the company had $12.8 billion in sales.

Open data initiative: Syngenta has a history of using openly available public data in its research and development. Data the company have used include information about soil and weather conditions, as well as biochemical data about crops. In 2014, Syngenta created its Good Growth Plan, which is built around its six core values. In collaboration with the Open Data Institute, Syngenta worked on opening six of its datasets that will be updated every year. By making these data open, Syngenta sought to improve trust in the company and increase innovation. Consequently, the company sees opening their data as part of their push towards collaborative innovation.

Syngenta works with the Open Data Institute (ODI) to improve their data skills and make sure that the data published is of the highest quality and follows established best practices. Syngenta's data quality is verified by Price Waterhouse Cooper (PwC) and evaluated by the ODI.

Organization: The Kraft Heinz Company

Brief description: The Kraft Heinz Company is the fifth largest food and beverage company in the world and had $26.5 billion in net sales in 2016. It produces a wide variety of processed foods and drinks for consumption at home and in restaurants. Some the company's brands include Kraft, Heinz, Caprisun, Philadelphia and Weight Watchers.

Open data initiative: The Kraft Heinz Company has made available an API for recipes using Kraft products called Kraft Recipe. Third parties can use the data to incorporate recipes using Kraft ingredients into applications they may develop. The recipes are categorized by meal type and can be retrieved based on a classification which includes categories such as kid-friendly, budget, health-conscious, and ethnic cuisine. Recipe data is broken down by ingredient, number of servings, preparation time, and other specifics. The API also supports user polls, shopping lists and favorite recipes, for example. ("Kraft Recipe API | ProgrammableWeb")

Organization: AB InBev

Brief description: AB InBev is a global brewing company with its headquarters in Leuven, Belgium. It began over 600 years ago as a brewer in Leuven and has grown to include over 500 beer brands that are sold in 150 countries. It is the leading beer company in the world with 28% of the market and an estimated $55 billion in sales in 2017. AB InBev brands include Stella Artois, Budweiser, Corona, and Löwenbräu.

Open data initiative: AB InBev launched an open API called BeerGarage to pair beer with food, much like the more traditional wine and food pairings. AB InBev views the API as a testbed for ways to convince customers to engage with its brands in-stores from their smartphones. "An API allows our data to be present in many consumer facing apps," said a

spokesperson (Joseph 2015). In February 2015, AB InBev hosted a BeerGarage hackathon in London with the purpose of encouraging participants to use the new API in creative ways.

Organization: Campbell's

Brief description: Campbell's is a soup and related products company based in Camden, New Jersey. Campbell's products are sold in 120 countries. The company entered pop culture when Andy Warhol produced what is now an iconic series of silkscreens depicting Campbell's soup cans. The company was founded in 1869.

Open data initiative: Campbell's Kitchen is a dedicated website that provides recipes and nutritional information to consumers using Campbell's brands. The open API was developed to share information from this website. The company hopes that developers will integrate Campbell's Kitchen data into the services they build to make it easier to cook healthy meals.

The API provides access to

- Access to thousands of proven family favorite recipes
- Extensive recipe filtering by key ingredients, product UPC, keywords and more
- Professional food photography
- Reader-generated recipe reviews & comments
- Recipe search results through superior tagging
- Well-known food brands people know and trust

The company also offers suggestions as to how to use the data:

- Enhance websites with related recipes & delicious looking photographs
- Create food-related apps (for websites and the latest devices and toys) and helpful shopping and cooking tools
- Augment social media sites like Facebook, Twitter & Google+
- Raise visibility for third parties' brands
- Drive more traffic to third parties' sites and gain new readers from a wider audience
- "The sky's the limit"

Organization: USDA Open Data Catalog

Brief description: The United States Department of Agriculture is a government body that, according to its website, "We provide leadership on food, agriculture, natural resources, rural development, nutrition, and related issues based on public policy, the best available science, and effective management" ("About the U.S. Department of Agriculture | USDA").

Open data initiatives: The USDA also promotes innovation. One of the ways it does so is by making available an extensive list of datasets for anyone to use. The USDA is one of the founding members of the Global Open Data for Agriculture and Nutrition (GODAN) initiative. Here are the categories of datasets available (https://www.usda.gov/topics/data) to give an idea of the scale of the open data initiative:

- Economic Data on Food, Agriculture, and the Rural Economy
- Agricultural Markets and Trade
- Diet and Health
- Food Safety
- Natural Resources, Environment, and Conservation
- Rural America
- Commodity Outlook
- Food Consumption (Per Capita) Data System
- Farm Income
- Agricultural Resource Management Survey (ARMS)
- Trade Data
- Production, Supply, and Distribution Online Database (PSD Online)
- Commodity Data and Statistics
- Current World Production, Market and Trade Reports
- US Export Sales Reports
- Attaché Reports
- Crop Explorer
- Today's Reports
- Census of Agriculture
- Data and Statistics
- Annual Agricultural Statistics

- Statistics by State
- Charts and Maps by Commodity
- Statistics by Subject
- RBS Service Reports
- Cooperative Directory
- World Agricultural Supply and Demand Estimates
- USDA Agricultural Projections to 2020
- Weather and Climate

Among these various data sets are the following:[1]

- The USDA National Farmers Market Directory connecting farmers and consumers, communities, and businesses around the US. For smartphone developers and users there is a directory API, in which for example consumers can search for farmers markets (see related "Meet Me at the Market" - The Evolution of a Farmers Market blog).
- The VegScape tool, which provides data and mapping capabilities from satellite-based assessments and monitoring of U.S. crop conditions.
- GRIN Global (Germplasm Resource Information Network), a web-based information management system for the world's plant genebanks, providing access to plant genetic resources in the face of challenges such as crop diseases and pests.
- New statistical products supporting the U.N.'s Global Strategy for Agriculture and Rural Statistics.
- An API / widget for access to daily visualization series Charts of Note (see related USDA, ERS Moving Down the Track to Open Data blog),
- APIs to select data sets (including GIS applications).
- A dynamic API for the best source of agricultural census and survey data published by the National Agricultural Statistics Service (NASS), the Quick Stats data.

[1] https://www.usda.gov/media/blog/2013/11/29/data-decisions-using-data-improve-public-access-and-knowledge

Organization: GODAN

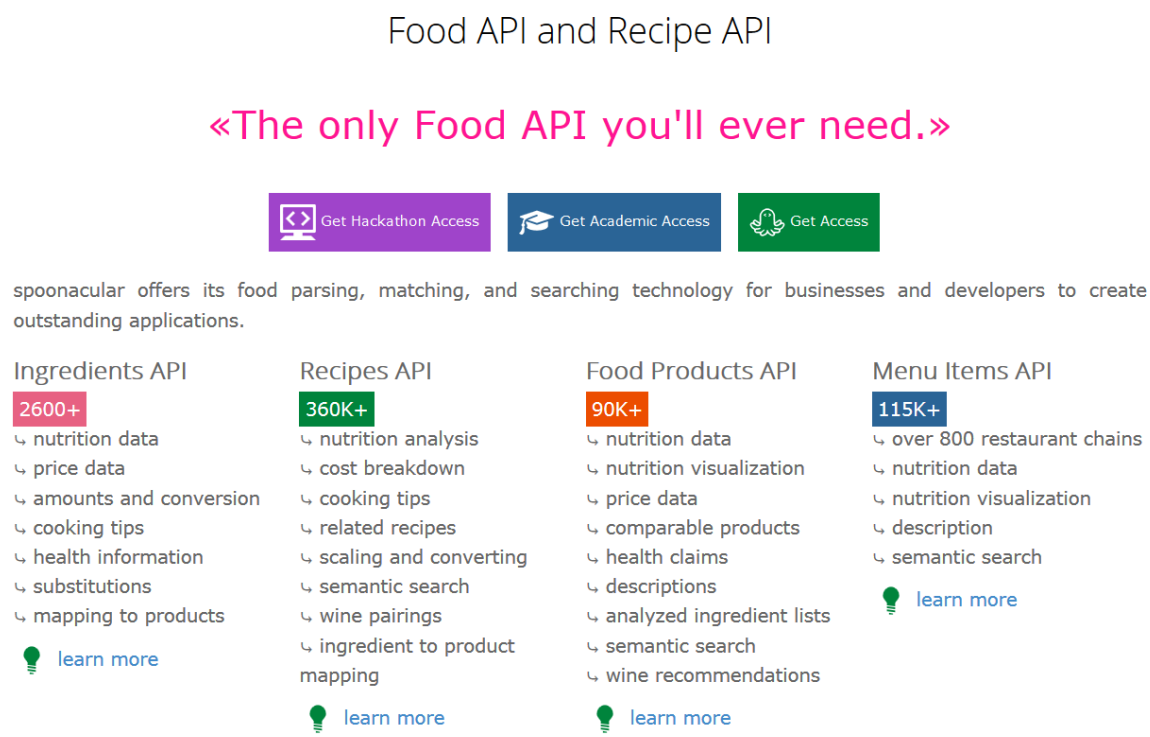
Brief description: GODAN was launched at the G-8 meeting in 2012. Its purpose is to promote the opening of agriculture and nutrition datasets and the unrestricted sharing of this data worldwide. It seeks to solve urgent nutrition and food related problems through the accessibility and usability of relevant data. GODAN seeks to influence policy around opening data worldwide.

Open data initiatives: GODAN currently has 629 members. The organization regularly publishes "success stories" on its website in pdf format. The most recent issue was published in May 2017 and contains 11 examples from all over the world, including CTA data for farmers in Africa, the Caribbean, and the Pacific; crowdsourced video production (Digital Green) on best practice farming techniques in Asia and Africa; crop yield data worldwide; and an Open Data Institute, which supports open data in food, nutrition, and agriculture worldwide. For more information on the success stories, please visit: http://www.godan.info/sites/default/files/documents/Success_Stories_2.pdf

Organizatons: Yummly and Spoonacular

Brief description & open data initiatives: These recipe APIs serve as aggregators for developers, taking the concept of open food data one step further. Both are primarily recipe websites and apps, but both also make data available to developers. Yummly API lists the sources of their data on their website. Spoonacular makes 2 APIs available for free (see Figure 3), recipes and food, but the company uses a freemium model so that if you want more data, you have to pay for it.

Figure 3: spoonacular's APIs



As these examples illustrate, some large companies in the food industry are beginning to open some of their data. The motivations for providing APIs to developers is clear. If more and more developers integrate product information into their applications, this will drive more consumers to buy these products. It will also raise awareness of the associated brands. Government agencies and international organizations have different motivations. By circulating data widely, the hope is that solutions to difficult problems will be more readily found. Food will be safer and more traceable; Farmers can better prepare for the coming season; Nutritional information can promote health

On the website Programmeableweb, there is an amazing wealth of open food data from all different sources. The site currently lists 168 APIs under the "food" category. These include nutritional information, food composition from the USDA, recipes, restaurant and bar localization, manufacturing information, food pairings data, pet food composition, grocery checklists, restaurant delivery APIs, agriculture grain prices, to list but a few. The question to ask then is whether or not this data is being used. And, if it is, who is using it? These questions will remain unanswered for the time being. The next two sections, however, will provide insight into how researchers and entrepreneurs foresee using and producing open food data and why this data can be a source of business innovation.

4. Open Food data hackdays 2017

The first Open Food Data Hackdays were held on February 10 and 11, 2017 at two venues in Switzerland: the Ecole Polytechnique Fédérale in Lausanne and the School of Art in Zurich. 192 people signed up to attend the event in Lausanne and 141 in Zurich.[2] The underlying purpose of this event was to promote the use of open food data for businesses innovation. The Open Food Data Hackdays is part of a three year project that "aims to build a publicly available base of nutrition data, to create new innovative and value adding solutions, and to further develop the use of open data for entrepreneurial purposes" ("Open Food Data Program," 2017). Given that the aim of the overall project that the hackdays event was part of, to use open food data as a driver for business innovation, the organizers invited individuals and groups to propose projects for hackday participants to work on that could become viable businesses. A team of three researchers from the College of Management at EPFL attended the hackdays. Two of the researchers observed the Zurich event and one researcher observed the event in Lausanne.

The Open Food Data Hackdays event was similarly organized in both venues. On the morning of the 10th of February, the organizers in both venues presented the event to the participants. Teams would have 24 hours to work on a project and present a prototype at the end of the event. A jury made up of members of the organizing and funding teams would chose a small number of projects to be incubated and coached over the next 2-3 months. The research team from EPFL also gave their input in the form of evaluation criteria[3] and their evaluations of the projects.

After the initial presentation, the projects selected by the organizers prior to the event were pitched to the participants. The floor was then opened to additional projects. Crucially, each pitch included a problem to solve, or a goal to reach within the time frame of the hackdays. Hackathons are generally organized with both a high level theme and specific challenges and the Open Food Data Hackdays was no exception. All the projects revolved

[2] Not everyone who signed up attended. We do not have data on actual attendance since people came and went during the Hackdays but the numbers appeared to be close to the number of registrations,

[3] Please see the annex for the evaluation criteria.

around using open food data sets and/or creating open food data sets. The projects were diverse. Some were serious, some were humorous, some were already businesses and others were barely in the ideation phase. Therefore, the challenges that were presented ranged from the specific to the more general.

Based on observation and discussions with participants, the teams were created based on a number of different factors, such as motivations, academic and professional backgrounds, and interpersonal affinities. A small number of individuals joined two teams.

The winning projects announced a few days after the event reflect the diversity of opportunities presented by open food data. **Open receipts** (https://hack.opendata.ch/project/74), **Nutrimenu** (https://hack.opendata.ch/project/68) and **Jarvis the Nutritionist** (https://hack.opendata.ch/project/60) were selected from Lausanne. **Open receipts** seeks to transform your supermarket receipts into "actionable data" that would give you information about the calories in the food you purchased and allergens, for example. **Nutrimenu** was already collaborating with the city of Lausanne prior to the hackdays. Nutrimenu is an application that helps you create healthier and tastier meals. Currently, the city of Lausanne uses it to improve the meals for schools and other restaurants in public administration offices. It won the 2017 award for Swiss health enterprises. **Jarvis the Nutritionist** is intended to be a chatbot that consults with the users on grocery choices. **Meat Story** (https://hack.opendata.ch/project/73) and **Foodimmune** (https://hack.opendata.ch/project/79) came out of Zurich. **Meat Story** seeks to make the meat you buy traceable from farm to fork through the use of a mobile app and possibly QR codes. **Foodimmune** leverages the medicinal properties of food to help you stay healthy.

Who were the participants?

After the Open food data hackdays event, a survey using google forms was sent out to the participants and we received 44 responses (roughly 13%). The majority of the participants were between 19 and 45 years old (19-27: 29.5%; 28-35: 27.3%; 36-45: 31.8%) and 70.5% of the respondents identified as male. 86.4% had received had received either a bachelor, master, doctoral, or law degree. Most respondents studied computer science and engineering (75% combined). It is interesting that despite the open food data theme so few

respondents had social science or health science degrees. While hackathons are traditionally oriented towards computer scientists, attracting more non-computer scientists may increase innovative outputs.

We were interested in trying to better understand the motivations of the participants and asked a series of questions using a 7 point Likert scale (1=strongly disagree and 7=strongly agree) to ascertain what motivates people to attend hackathons. Three motivators stood out: attending the hackathon was a way to enhance skills (68.2% chose 5 or above) and participating in the hackathon allowed participants to explore their strengths and limitations (72.8% chose 5 or above); most participants (77.3% chose 5 and above) attended the hackathon to learn about open food data challenges. Given the sponsor and the overall goal of this event, it should come as little surprise that **most survey responders care about the open data movement** (**81.3% chose 5 or above** with 37.2% choosing 7—strongly agree) and saw the hackday event as a way of participating in the open data cause (65% chose 5 or above). 46% of respondents strongly agreed that it is important to participate in initiatives like the hackdays event.

The winning teams

After the hackdays, as mentioned above, the winning teams received a stipend, workspace and coaching for 2-3 months. At the end of October 2017, the teams were asked to evaluate their experience. Below is a snapshot of their feedback.

Table 3 Winning teams' feedback

Team	vision	goals	accomplishments	support	Next steps
Jarvis	Chatbot is a new way to interact with products	1. Build a recommender system 2. Build a chatbot interface	1. Scraped recipes from allrecipes 2. Built chatbot prototype 3. Information gathering about chatbot and recommender system	1. Financing 2. Meeting with Hannes Gassert 3. Meeting with Thomas Rippel	1. Complete recommender system prototype 2. Complete chatbot 3. Test bot with real users 4. Launch strategy (March 2018)
Open Receipts	Unlock the data stored in receipts and turning into insights for consumers	Develop a minimal viable product (MVP)	1. Developed a minimal viable product 2. Identified consumer needs 3. Developed a business use case	1. Coaching (could have been more timely) 2. financing	Test current product and further develop its functionality
Food Immune	Mobile application inspire healthy eating based on herbs and local ingredients	1. Limits of the concept 2. Develop, design the app 3. Test prototype 4. Fully functioning downloadable MVP	1. Narrowed down concept 2. created website 3. tested early prototype 4. reiterated design 5. market research 6. further refined product	1. Able to attend global summit 2. coaching 3. financing	Rethinking project
Nutrimenu	Extend client base for creating healthier menus	1. Assess demand 2. Adapt the product to new customers 3. Present the product at conferences & expos 4. Look for funding	1. Sent a survey to 6000 restaurants 2. Met with 12+ catering companies 3. Market research for product in Swiss German cities 4. Presented product at 6 venues 5. Applied for funding, support & collaboration	1. Financing 2. Helped with outreach to 3. German speaking cities 4. Coaching 5. Offered visibility and credibility	1. Translate product in German 2. Continue seeking new clients 3. Continue developing the product
MeatStory	Trace origin of meat	1. Build a MVP 2. Test product on real customers 3. Gather feedback and reiterate	1. Press releases in 2 newspapers 2. Access to data 3. Meetings with potential partners 4. Prototype design sessions 5. Market research 6. Presentation at Opendata event 7. Developed 2 versions of product	1. Hackdays 2. Inspiring network 3. Regular calls and meetings 4. financing	1. Marketing 2. Testing

What the table above shows is that not all teams were focused on developing viable businesses. The two teams that stand out in this regard are Nutrimenu and MeatStory. The lack of a business perspective is something that struck the EPFL team observing the Hackdays. It was clear based on observation and conversations with participants that the many participants attended the Hackdays to have fun and test their technical skills, which is typically what a hackathon is about. In situations where hackathons are being held to foster business innovation, some changes can be made to better achieve this goal:

- Make sure the challenges have a clear business orientation by vetting them with this in mind prior to the hackathon.
- While hackathons typically focus on developing a prototype, ask teams to also produce a simple business plan using a template distributed to all attendees prior to the event.
- Provide a business bootcamp either before or after the hackathon. If provided before, then this will possibly limit the number of attendees. If provided afterwards, the bootcamp would be required for winning teams and an opportunity for other attendees to participate in for a small fee, for example.
- Based on the responses from the winning teams in the table above, more coaching is needed to really propel these teams from the ideation phase to at least developing a viable business plan.
- For those teams who decide to move forward with their projects, additional incubation and support is necessary so that innovative businesses come out of the open food data hackdays.

5. Focus group insights

In November 2017, we held a focus group on open food data at EPFL. A focus group is a small group of people who are led through an open discussion by a moderator. Focus groups are structured around two or three clearly defined questions, but the aim is to get participants talking. People should feel comfortable enough to speak freely, but should not know one another. Focus group participants should not know what questions you are going

to ask prior to the event. In our case, it was important that participants be knowledgeable about the subject to be discussed. Therefore, participants were recruited from the open food data hackdays and through recommendations.

Our focus group participants were drawn from the research, business and non-profit sectors. Our primary interest was to see how these individuals currently applied open data in their endeavors and how they foresaw the future of open food data.

Everyone in attendance was a proponent of open data and more specifically open food data. Of the five participants, two were actively producing open food data. One participant was helping companies use open food data. One participant was active in a non-profit related to the food industry that collects a lot of data and is pushing to make it more open. And, one participant was a researcher in nutrition and had a clearly defined need for open food data. Despite actively using, or producing open food data, participants also recognized that making open food data more acceptable on the business side and increasing the general population's knowledge of it was a challenge. Data, as one participant mentioned, is not the challenge. Most people are aware of how data can improve profitability and increase efficiency. The crux of the challenge is openness. Unless you are able to demonstrate to companies how opening data can help them make money, or save money, they will not see why data should be opened.

A rejoinder to this comment is that consumers are becoming increasingly interested in food labels and nutrition and will begin to push for more accurate information. Companies that provide this information will be favored. Additionally, people are beginning to want to work for companies that are making a difference. So, if companies in the food industry want to recruit and retain the most innovative and talented employees, they may need to become more transparent and opening data may be one way of demonstrating a willingness to be transparent. Furthermore, in a world where everyone is inundated with unsecure data, having a source of secure, verified data may be a major benefit. This is especially the case with food. People want to make sure that the food they consume is safe and more and more individuals have added requirements, such as ethical treatment of animals, organic

food, or locally grown food. There is a push for farm to fork traceability. Open data is a way to achieve this.

However, one main question remains: who is going to drive this transformation? The focus group participants recognized the importance of this question, but had difficulty answering it. In any case, it is a complex question that resists a simple answer.

6. Conclusion

As we hope to have shown, open food data can drive business innovation. From startups to large multinationals, the role of open data in creating innovative products and services in the food industry is increasingly recognized. There is, however, work to be done in raising awareness about the benefits of using and creating open food data. It is not enough to show what open data can bring, there needs to be a clear financial benefit for larger companies for using, or making available their data. There is also a lack of awareness among the general population as to how open data can help them in their daily lives.

In terms of hackathons for open data, we see hackathons as a new tool in the innovation manager's toolkit, a kind of live crowdsourcing exercise that goes beyond traditional ideation. Of course, the size of the crowd may be much more limited in a hackathon relative to a large, public crowdsourcing challenge; on the other hand, the level of interactivity and the parallel structure of the teamwork can lead to interesting solutions, better communication of those solutions via prototyping, and inspiration for companies and other organizations to take some of the ideas further. In addition, sponsor organizations may gain insight into aspects of the challenge that resonate with external innovation teams, input that is often hard to come by via other means such as focus groups or external consultants.

Who is going to drive this transformation so that open food data is widely accepted by industry and civil society? The transformation should be led from the top and the bottom. Without buy-in by governments and large industries, change will happen, but more slowly. A wide reaching open food data strategy would doubtless spur on business innovation; increase food security and help citizens make healthier food choices.

References

"About the U.S. Department of Agriculture | USDA." n.d. Accessed January 16, 2018. https://www.usda.gov/our-agency/about-usda.

Batini, Carlo, Matteo Palmonari, and Gianluigi Viscusi. 2014. "Opening the Closed World: A Survey of Information Quality Research in the Wild." In *The Philosophy of Information Quality SE - 4*, edited by Luciano Floridi and Phyllis Illari, 358:43–73. Synthese Library. Springer International Publishing. https://doi.org/10.1007/978-3-319-07121-3_4.

Batini, Carlo, and Monica Scannapieco. 2016. *Data and Information Quality: Dimensions, Principles and Techniques*. Springer.

Batini, C, A Rula, M Scannapieco, and G Viscusi. 2015. "From Data Quality to Big Data Quality." *Journal of Database Management* 26 (1):60–82. https://doi.org/10.4018/JDM.2015010103.

Benington, John. 2011. "From Private Choice to Public Value?" In *Public Value - Theory and Practice*, edited by John Benington and Mark H. Moore, 31–51. Palgrave Macmillan.

Berners-Lee, T. 2006. "Linked Data - Design Issues." 2006.

Berners-Lee, Tim. 2006. "Design Issues: Linked Data."

Boyd, Mark. 2014. "OpenData 500 Launch Demonstrates Open Data API Value | ProgrammableWeb." ProgrammableWeb. 2014. https://www.programmableweb.com/news/opendata-500-launch-demonstrates-open-data-api-value/2014/04/09.

Chignard, Simon. 2013. "A Brief History of Open Data." Paris Innovation Review. 2013.

Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. "Benefits, Adoption Barriers and Myths of Open Data and Open Government." *Information Systems Management* 29 (4):258–68.

Joseph, Seb. 2015. "AB InBev's Mobile Marketing Plan to Bring Beer and Food Closer Together." *The Drum*, March 2015. http://www.thedrum.com/news/2015/03/02/ab-inbev-s-mobile-marketing-plan-bring-beer-and-food-closer-together.

"Kraft Recipe API | ProgrammableWeb." n.d. Accessed December 20, 2017. https://www.programmableweb.com/api/kraft-recipe.

Misuraca, G., and G. Viscusi. 2014a. "Digital Governance in the Public Sector: Challenging the Policy-Maker's Innovation Dilemma." In *8th International Conference on Theory and Practice of Electronic Governance (ICEGOV2014)*.

———. 2014b. "Is Open Data Enough? E-Governance Challenges for Open Government." *International Journal of Electronic Government Research* 10 (1):19–36.

Obama, B. 2009. "Transparency and Open Government. Memorandum for the Heads of Executive Departments and Agencies." 2009.

"Open Food Data Program." n.d. Accessed November 3, 2017.

https://food.opendata.ch/#supported-projects.

Open Knowledge International. 2017a. “How to Open up Data.” Open Data Handbook. 2017.

———. 2017b. “What Is Open Data?” Open Data Handbook. 2017.

The Open Data Institute. 2017. “The Data Spectrum Helps You Understand the Language of Data.” The Data Spectrum. 2017.

Viscusi, G., M. Castelli, and C. Batini. 2014. “Assessing Social Value in Open Data Initiatives: A Framework.” *Future Internet* 6 (3):498–517. https://doi.org/10.3390/fi6030498.

Viscusi, Gianluigi, and Carlo Batini. 2014. “Digital Information Asset Evaluation: Characteristics and Dimensions.” In *Smart Organizations and Smart Artifacts SE - 9*, edited by Leonardo Caporarello, Beniamino Di Martino, and Marcello Martinez, 7:77–86. Lecture Notes in Information Systems and Organisation. Springer International Publishing. https://doi.org/10.1007/978-3-319-07040-7_9.

“Wildcatch — GS1 Discover Foodservice.” n.d. Accessed January 16, 2018. http://discover.gs1.org/foodservice/wildcatch/.

HACKATHON FRAMEWORK FOR EVALUATION

A panel of judges will assess each proposal based on the criteria listed below. Each judge will assign a proposal a score between 1 and 10 inclusively for each of the four criteria, with 1 being unexceptional and 10 being exceptional.

1. Innovation / Ambition / Creativity
2. Technical competence and capabilities
3. Overall Quality
4. Quality of Pitch / Impact

Innovation / Ambition / Creativity

- Was the idea unique, or a different take on an existing/similar service/product?
- Does the proposal incorporate creative design and innovative capabilities?
- Was the idea behind the service/product ambitious? Creative?

Technical competencies and capabilities

- Does the proposal address the primary goals of the hackathon?
- Does the UI of the service/product look professional/fun, or is it sloppy?
- Is the product/service easy to use?

Overall Quality

- How many and how severe were any bugs encountered?
- Did the team scope their proposal's features well given the time frame of the hackathon?

Quality of Pitch / Impact

- Does the service/product seem like something people would want to buy and use outside of the Hackathon?
- Was the team able to explain their idea and motivation clearly?