The role of the Data Scientist within Smart Cities

De Obeso-Orendain, A., Lopez-Neri, E., Donneaud-Bechelani, C.

Abstract— The role of the data scientist has gained considerable importance in recent years. This paper explores images, conceptions and phenomena regarding data scientists from different perspectives, both theoretical and practical, guided by the premise that Smart Cities require a particular type of data expert. A survey was distributed among data scientists, and its results are presented to provide insight regarding the perception of data science in the community surrounding the Ciudad Creativa Digital (CCD). Data scientists may become a key factor in the successful operation of a Smart City, provided that they are able to reach and empower its most critical element: the citizen.

Index Terms—Data Scientist, Smart Cities

1 Introduction

A Smart City is a city that performs well in the following characteristics: economy, people, governance, mobility, environment and living [1]. The high degree of datification and connectivity embedded in a Smart City demand tools and mechanisms for data manipulation and representation that facilitate the extraction of meaningful insight. However, this process of extracting knowledge from data is mediated by people. The purpose of this paper is to identify characteristics of a role that has gained importance in recent years: the data scientist [2]. Furthermore, we focus this analysis on those particular abilities required by a data scientist living in a Smart City. We consider this topic particularly relevant considering the shortages of data scientists that the labor market is facing in other domains [3].

This approach is aligned with the strategy of the data analysis and visualization (DAV) team defined previously [4], in which the Smart People component occupies a prominent role. In this context, we deem data scientists as critical for the ability of a Smart City to promote information sharing and foster collaboration. As described in our previous work, citizen engagement is a fundamental factor in the successful implementation of Smart Cities in conjunction with the technological infrastructure [5] [6] [7]. For this reason, we put special emphasis on describing the relationship between data scientists and citizens' engagement.

The Ciudad Creativa Digital (CCD) master plan envisions the CCD as a unique place to live and work [8]. With this exercise, the DAV team continues its works of paving the way for the success of CCD by identifying the characteristics of a key player of this ecosystem.

2 The Data Scientist

The emergence of data scientists is a natural response to the radical revolution in the way data are generated, represented, stored, analyzed, visualized and used. New technologies and techniques are required to address the challenges posed by Big Data, commonly encapsulated by its several V's: volume, variety, velocity, etc. In this section, we explore the definition of a data scientist from two different points of view. From a theoretical standpoint, we explore definitions from different domains, whilst, for a practical approach, we conduct an empirical exercise.

2.1 A theoretical approach

A definition of a data scientist extracted from p. 9 of a textbook states that "… scientist in these areas face unprecedented volumes of data, such as in the human genome project with three billion base pairs per human." [9]. A definition extracted from the web articulates that "A data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician" [10]. Shlomo Aragmon [11] identifies the components of a data scientist: statistician, programmer coach, storyteller and artist. Drew Conway proposed the data science Venn diagram [12], which comprises three sets of skills: hacking skills, math and statistics knowledge and substantive expertise. Hacking skills refer to the ability to manipulate digitalized data, e.g., text files, and to have abstraction abilities able to decompose problems into algorithms. To extract meaningful patterns and insight from data, it is also necessary to apply mathematical, statistical and machine learning methods. However, a third element is compulsory to be an effective data scientist: considerable expertise. This last element strongly suggests that we need to identify the areas of expertise related to Smart Cities.

Beyond these definitions, we conducted two literature reviews. The first regarded existing research of recognized people in the data scientist area and the main issues addressed in conferences related to this topic. The second was focused on the vision that the industry has about the data scientist concept; for this reason, job posting web announcements linked to data scientist positions were checked. The survey designed to encapsulate the role of a data scientist is based on this analyzed information.

2.1.1The Data Scientist from the scientific community standpoint

Figure 1 shows the different main ideas and concepts regarding the data scientist profession. Each concept (blue globe) is associated with an author or authors, and the previous connection with another concept means that the author accepts all the previous definitions. There exist general [13] [9], modern [14] [15] [12] and detailed [16] [11] definitions, but the main concepts are (a) the use of data to understand the current state of a system and predict outcomes; (b) the ability to discover a previously hidden insight; (c) hacking, coaching and storytelling skills; (d) substantive expertise; and (e) a data visualization component.

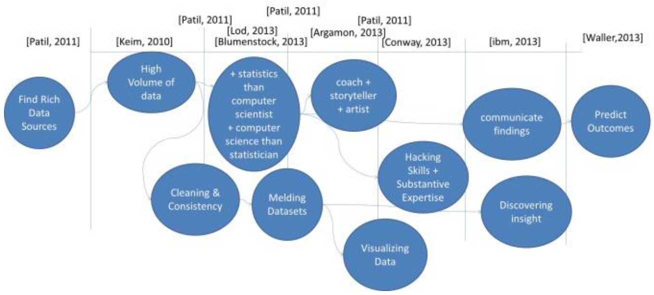


Figure 1: concepts related to the data scientist

Table 1 shows the main conferences related to themes and issues of data science. Technical conferences (for instance, on particular tools[1]) were not taken into account because they are very specialized.

Conference / Event	Objective	Issues
[2]The 2014 International Conference on Data Science and Advanced Analytics (DSAA'2014),	Communicate Data Science	Analytics science Big data Advanced analytics
[3]KDD-LESI 2014	Structure and dynamics of social and information behavior	Large-scale emergencies • Natural disasters • Terrorist attacks
[4]Workshop on Data Science for Social Good	How data can be used to make decisions	Data science to be used effectively for social good • Education • Healthcare • Energy • Transportation • Climate change • Public safety • Disaster relief • Poverty • Community development.
[5]Enterprise Data World conference	Data Management. The Transformation to Data-Driven Business.	• Data Governance & Quality Program Implementation. • EIM & Data Architecture • Big Data Trends • Analytics & Business Intelligence for New Demands
[6]Strata. Making Data Work	Data Science and Big data (training)	• Cloud Computing • IT • Big Data • Data Science
[7]CITE Conference & Expo	New Technologies to drive your business.	Innovation with mobile, social, cloud and Big Data.
[8]Data 360 Conference	Big Data Revolution	• Financial services driven by big data • Future of healthcare through the lens

[1] http://hadoopsummit.org/
[2] http://datamining.it.uts.edu.au/conferences/dsaa14/
[3] https://sites.google.com/site/kddlesi2014/
[4] http://dssg.uchicago.edu/kddworkshop/
[5] http://edw2014.dataversity.net/
[6] http://strataconf.com/
[7] http://www.citeconference.com/ehome/index.php?eventid=78152&
[8] http://www.data-360.com

		of big data & population analytics • Data driven insights for formation services • Harness the power of big data in retail.
[9]HotCloud	Hot topics in Cloud Computing	Cloud Computing

Table 1: Main conferences related to data science

2.1.2 The Data Scientist from the Industry Standpoint

Based on job postings for data scientists that were posted on the web, we generated Table 2. This table identifies the main soft and hard skills required by the industry from a data scientist[10,11,12,13,14,15]. In our sample, we found some variations, such as a Statistical Programmer/Developer/Analyst[16,17,18] and a particularly interesting one that focuses on an aspect of the data science domain: Data Curator[19].

Soft Skills
Knowledge of industry business
Financial Planning
Leadership and Project management
Hard Skills
Statistical Software: • R • SAS • SPSS • MATLAB
SQL Server and Oracle, MySQL
Big Data Analytics Tools: • Hadoop • Mahout • Map/Reduce • HBase • NLP methods (OpenNLP, LingPipe) • Lucene Core • Solr • Spark • Impala • Hive • Pig
Machine learning
Graph/network analytics
Programming in the Linux environment: • Java • C++/C • Perl • Python.
Text mining
Social media data (Twitter, Facebook, LinkedIn, Google+)
Probabilistic Models • Parametric • Non-parametric • Bayesian Statistics
Regression models
Pattern recognition
Recommender systems
Constraint-based optimization
Scalable programming on big data

Table 2: List of hard and soft skills required by a data scientist

The required backgrounds identified in these job descriptions were: Master of Science (MS), MS in mathematics, MS in statistical modeling, and PhD in Computer Science (Engineering). The industries requiring services from data scientists included: advertising, marketing, broadcasting, radio and TV, research and development, technology services, social communications, tech support, electronic health record systems, healthcare and robotic automation.

Table 3 describes the main tasks executed on data or information and the purpose of these activities. These relations were recovered from the expected functions of a data scientist as specified in each of the offered jobs.

<table>
<tr><th>Task</th><th>Over…</th><th>Purpose</th></tr>
<tr><td>Fusion</td><td rowspan="2">Disparate Large Scale Media Data Sets</td><td>Gain beneficial insights</td></tr>
<tr><td rowspan="2">Modeling</td><td></td></tr>
<tr><td rowspan="3">Internal Operational Data</td><td>Reduce uncertainty</td></tr>
<tr><td>Identify [Hybrid Systems]</td><td>Improve operations</td></tr>
<tr><td>Develop new ontologies and new methods of data</td><td></td></tr>
</table>

[9] https://www.usenix.org/conference/hotcloud13

[10]https://careersen-bbm.icims.com/jobs/1487/data-scientist/job?mode=job&iis=Indeed&iisn=Indeed.com&mobile=false&width=1024&height=475&bga=true&needsRedirect=false

[11]https://www.kaggle.com/forums/t/10039/sysomos-senior-data-scientist-toronto-can

[12]http://mckessonjobs.com/ca/richmond/project-management/jobid5691907-sr.-data-scientist-clm-_-bi-(predictive-and-prescriptive-modeling)

[13] http://www.cwjobs.co.uk/JobSearch/JobDetails.aspx?JobId=60336135

[14] http://www.respondhr.com/35491972

[15] https://jobs.lever.co/counsyl/87683bc8-37c8-4b20-82cd-eda2fde15048

[16]http://www.cwjobs.co.uk/JobSearch/JobDetails.aspx?JobId=60336135

[17] https://www.recruit.ox.ac.uk/pls/hrisliverecruit/erq_jobspec_version_4.display_form?p_company=10&p_internal_external=E&p_display_in_irish=N&p_applicant_no=&p_recruitment_id=114670&p_process_type=&p_form_profile_detail=&p_display_apply_ind=Y&p_refresh_search=Y

[18]http://brightrecruits.com/job/6348/data-scientist-statistical-developer?utm_source=Indeed&utm_medium=organic&utm_campaign=Indeed

[19]http://explorys.theresumator.com/apply/job_20140813173858_UQLZI26XMVSXYGPO/Data-Scientist-Data-Curation.html?source=INDE

Formulate validation strategies and methods	Internal Operational Data	Predetermine outcome
		Ensure accurate and reliable data
Solve (Problem Solving)	Seemingly intractable problems	Address previously unidentified market needs
Monitoring and Analytics	Social Content	
Create reports	Data Business	Communicate To business head
Define rules for data analysis		Identify data trends
Integration and further enhancement	Business and Sales Databases	Better understanding
Standardize	Data from business	
Coordinate	Business units	Optimize the cost effectiveness of the business
Cross-functional analysis (business sales)	Large corpus of historical sales and business data	Optimize revenue
Statistical modeling and process debugging	Data lab process	Develop new assays and new technology
Build structured reports	Critical laboratory parameters	Monitor for process control and regulatory compliance

Table 3: data-centric tasks

2.2 A practical approach: the survey

Beyond the literature review described previously, we gathered first-hand appraisals related to data scientists with the aim of shedding light on this role.

2.2.1 Method

An online survey was administered to a group of people who actually work in areas related to data science, such as information management with great volumes of data, data modeling, statistical modeling, etc. We were looking for important information regarding the required skills in these types of positions.

2.2.2 Participants

The survey was administered to 11 people, 10 men and one woman, all from different enterprises (ranging from small businesses to large corporations) and from different lines of business, industry and education. Their ages ranged from 25 to 45 years old, and their professional experience in data science was between 2 and 5 years. All the participants work in the city of Guadalajara and were contacted by email. The survey sample design distribution was composed as shown in Table 4:

Computer Science PhD.	**18%**
Researchers	**9%**
Supply chain or data analyst employees	**55%**
Own Business	**18%**

Table 4 : Sample Design Distribution

2.2.3 Survey

The survey was applied online, and its rationale was disclosed to the applicants. It is important to note that this survey is comprised of closed (Yes or No) and open questions to capture the free expression of each of our participants. These are the questions of the survey:

1.- Do you know what a Data Scientist is?
2.- Have you ever communicated to senior management the results of a data analysis process?
3.- Have you ever designed/implemented a data-centric system that impacted the business strategic objectives or generated extra profits?
4. - Have you ever participated in a machine-learning workshop?
5.- Have you ever faced a situation where you had to access different information sources (for instance, from other departments) to gain knowledge about a topic?
6.- From the data that you have collected, have you considered the legitimacy of the information source, the value of the data and its validity?
7.- How often do you use math equations (for instance, statistical modeling) to explore or understand the phenomena observed in the data?
8.- Do you ponder the value of information according to its source and the context of the application? Can you mention an example?
9.- From your own experience in information management and analysis, have you found any correlations between data? Can you mention any example?
10.- When you analyze information, have you ever had the feeling that, from the observed data, you can "predict" the system behavior or understand the distribution of data?
11.- Do you consider yourself to have good communication skills? What tools do you usually use to communicate your own ideas?
12.- Do you use a special tool to analyze and visualize the data? (We provide a list here)
13.- Have you used one of the following data mining techniques? (Random forest, support vector machines)
14.- What are your usual information sources? (Surveys, experiments, content analysis)
15.- What have been your data analysis objectives? (For instance, recommendation systems or business optimization)
16.- Which of the following tasks is more frequent in your job: coding or data analysis?
17.- How do you respond when working under pressure?
18.- How do you set priorities in your work on a daily basis?
19.- What do you consider when making a decision?
20.- Do you participate in the executive meetings in your organization?
21.- Are you involved in the short-term planning of your

company?
22.- How do you react to unexpected problems?
23.- Would you like to propose any change in your schema of work?

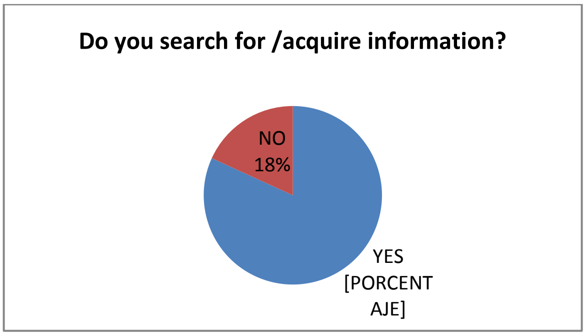
2.2.4 Analysis and results

The data from the surveys were analyzed via a grouping method that classified questions on "hard Skills" (related to data analysis) and "soft skills" (business vision, working under pressure, future planning, communication and decision making). Some questions about the personal profile of the respondent were considered independently.

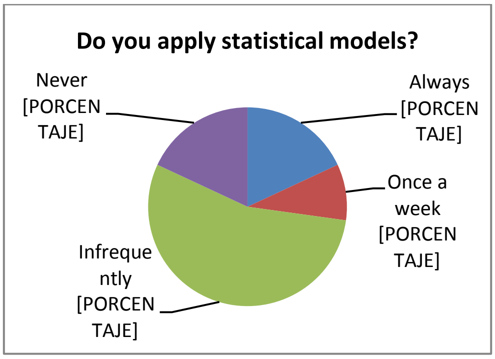
We found the participants to vary both in their abilities and in their duties. For this reason, we grouped hard skills into two categories: how information is acquired and how information is analyzed. For the first category, we considered questions 5, 7 and 15; the results are presented in Figure 2.

According to the charts below, respondents have a tendency to search for information and rarely use statistical or mathematical models. Additionally, their data sources are diverse.

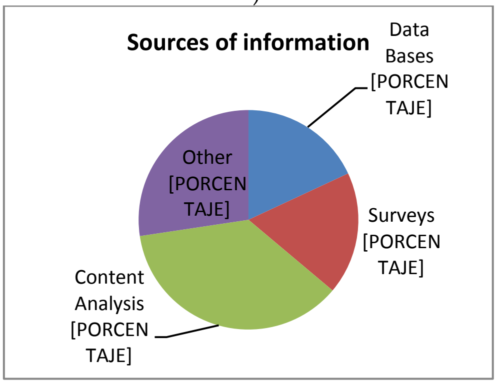
Regarding the data analysis category, we grouped answers to questions 6, 8, 9, 13, 14 and 16. From our sample, 100% answered affirmatively to the question about pondering the value of information according to its source and the context of the application; 73% from our statistical sample had the feeling that, through the data observed, you can "predict" the system behavior. To analyze data, our participants primarily use SQL, R and Python (Figure 3).



a)



b)



c)

Figure 2: a) Tendency to acquire information b) Use of statistical models c) sources of information

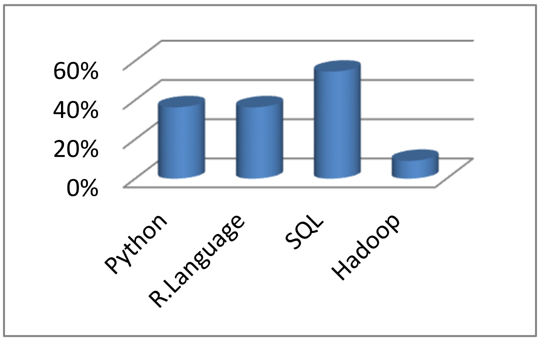


Figure 2: data analysis tools

From our sample, 73% do not know the mining techniques that were specified (random forests, support vector machines), and 45% stated that, in their work, they are used to analyzing data instead of only coding (18%).

The objectives from the data analysis reported by our participants are as follows: recommendations (18%), business optimization (36%), both (36%).

Overall, we found more data analysts than programmers in our sample; all of them understand the importance of information and consider the importance of making predictions.

Soft skills are summarized as follows: 60% answered that they work better under pressure than normal conditions, 89% propose changes in the work schemas and only 45% participate in strategic business meetings for future planning (49%). Regarding communication skills, 73% consider themselves to have good communication skills. Only 55% of our statistical sample knows what "Data Scientist" means. Our conclusion regarding the soft skills is that it is first necessary to put special emphasis on capabilities related to business acumen.

3 A DATA SCIENTIST FOR SMART CITIES

A data scientist can be conceptualized as a problem solver/solution provider who uses data as his/her primary tool. In this section, we want to explore the nature of his/her role when facing the challenges posed by Smart Cities. Different tasks have been associated with Smart Cities. For instance, the UK government identifies (a) unemployment (mainly in young people) as a consequence of dramatic changes in the economy; (b) housing and transport; (c) climate change, energy efficiency, reduction of carbon emissions, and pricing; and (d) social care [17]. A proficient data scientist should be able to provide integrated solutions to these challenges. Circumventing the idea that Smart Cities are limited to a transactional relationship between citizens and a technological infrastructure, the DAV group is particularly interested in developing citizens' engagement by promoting environmental and technological facilitators to foster collaboration and dialog. The data scientist, as an information-driven professional, can greatly contribute to these goals. For instance, reducing unemployment is particularly relevant for CCD, as one of its aims is to foster the growth of the creative industry in Mexico by attracting and developing local talent, which is mainly comprised of young people.

Overall, the key responsibilities of a data scientist should be to (a) ensure and promote open access to data; (b) design and develop tools for data re-use; (c) promote the creation of data models with a citizen-centric representation of data and (d) ensure that the right data are used to inform decision-making processes. In the following sections, we explore the DS in the particular context of CCD. We want to focus our analysis on the first stages of the CCD, when a few companies are based on the premises and the aim is to attract more investment, promote citizens' engagement and create a knowledge society.

3.1 Data challenges in Smart Cities

Smart Cities demand a precise set of data skills. In addition to the ability to identify, extract and format data, a data scientist should have enough expertise and domain knowledge to achieve meaningful integration of data, as several different services (online and offline), which speak different protocols, need to be consolidated. Additionally, the data scientist must design data models and delivery mechanisms to allow those applications to enable citizens to make informed decisions [17].

One of the directives of a comprehensive study conducted in the UK recommends the development of a strategic approach to open standards, data management and data sharing [17]. Data Governance includes standards and best practices to capture and communicate the meaning and format of information. Another critical aspect is to define and enforce standards and mechanisms to ensure data quality.

3.2 Talent management

A pervasive problem that a data scientist working in CCD must understand and solve is how to match the right person with the right job. The city of Guadalajara has been growing an increasingly larger mass of highly trained professionals that will ultimately power the creative industries located in CCD. The problem of finding, classifying and assigning this rich set of professional profiles to a diverse set of job specifications is both complex and interesting. Job specifications and candidate profiles present a mixture of structured and unstructured data, which may demand the use of text mining techniques. Once jobs and candidates are properly encoded, optimization techniques to find the best match need to be applied. Keeping track of how specialized professionals grow in skill and knowledge as they gain experience working in CCD can be used to identify opportunities for collaboration.

3.3 Interacting with the uOS

Within CCD, data are created, managed and consumed by means of a large set of components distributed over several layers, including a physical infrastructure, network capabilities and business processes. This is the Urban Operating System (uOS). This uOS supports a comprehensive set of digital services. The data scientist is both a contributor and consumer of the uOS. Therefore, sufficient knowledge of data transport protocols, the physical infrastructure (hardware) and the network topology and configuration is useful to grasp how the different uOS modules are interconnected. It is particularly important to understand how to make good use of the underlying structure of the data layer to exploit it and run data intensive analyses properly. For instance, if a NoSQL database, such as Cassandra, is implemented, the data scientist must understand how its data are distributed in nodes and actively participate in the development of data models and selection of primary keys. This means that the DS must be an expert of the data taxonomy of CCD. The main data streams involved in CCD include energy, water, mobility, waste, fiber and telecoms, security and street lighting, industry statistics, and building occupancy and utilization [18].

3.4 Analyzing Citizens' Engagement

A citizen of CCD is self-decisive and aware. Engagement can be measured in three dimensions: (a) behavioral, based on the notion of participation and involvement in activities beyond a person's role within his/her group; (b) emotional, oriented to either positive or negative reactions towards other people and institutions; and (c) cognitive, which denotes the required effort to comprehend complex ideas or to gain specialized skills [4]. A data scientist must be familiar with the different metrics related to citizen engagement.

3.4 Complex Event Processing

A proposal for CCD is to implement an event-driven module [19]. In this context, smartness is defined in terms of the ability of actors to handle events [20]. This framework provides a set of tools for citizens to make the most of CCD by providing intelligence and fostering collaboration.

An event is not just data but rather a product of an analytic process that provides added value to a set of raw data. Events can be created by following a standard structure that can be understood by all actors (people, services, sensors and devices) because they all share the same knowledge.

Listeners can also be created to execute actions in response to events. In this way, the members of CCD can define pre-programmed actions in response to events they create or subscribe to events created by other users.

4 Conclusions

By merging the information we have presented thus far, we can provide a depiction of the data scientist: he/she is able to find rich data sources, work with large volumes of data, clean, meld, and visualize data, and generate tools for working with data.

A curriculum to form effective data scientists for smart cities must include, at its foundation, computer science, analytics and math, with a strong emphasis on understanding societal challenges; in this way, data scientists will be able to pick the right problems and solve them. Such a program must also include subjects such as the use of tools, data ingestion and cleaning, data exploration, statistical modeling and machine learning. Another relevant component is knowledge about product development and communication. These data scientists should understand that their ultimate professional goal is to influence decision-making and to drive change.

We are living in a world driven by open policies, where information is available and accessible to everyone, from financial statements of listed companies to both the income and expenditure of government offices. The possibility of not only consulting but also understanding and interpreting these data are achieved through appropriate data management, which is the main function of a data scientist.

A current trend is related to transparency, meaning that anyone has the right to access information previously considered confidential. To the extent that ordinary citizens realize these rights, for instance, it is possible to extract detailed data about the government and transform it into a concrete and manageable format to promote understanding, the creation of public policies substantiated by hard data is conceivable. As a consequence, greater reliability in leaders can be generated, which will in turn promote greater engagement of citizens to continue contributing to society.

As stewards of the data, equipped with a highly specialized set of skills and a clear understanding of the challenges and opportunities of the city they are inhabiting, data scientists may become a key factor in the successful operation of a Smart City, provided that they are able to reach and empower its most critical element: the citizen.

Acknowledgment

The first author wishes to thank the OpenSource Technology Center at Intel. The second and third authors wish to thank the UVM Laureate International Universities for all of the support provided.

References

[1] R. Giffinger, C. Fertner, H. Kramar, R. Kalasek, N. Pichler and E. Meijers, "Smart cities - Ranking of European medium-sized cities," Centre of Regional Science (SRF), Vienna University of Technology, 2007.

[2] T. H. Davenport and D. J. Patil, "https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/," *Harvard Business Review,* no. October 2012, 2012.

[3] P. Davis, "McKinsey Report Highlights the Impending Data Scientist Shortage," July 2013. [Online]. Available: http://blog.pivotal.io/pivotal/news-2/mckinsey-report-highlights-the-impending-data-scientist-shortage.

[4] A. De-Obeso-Orendain, E. Lopez-Neri, J. Dominguez and J. Gutierrez-Garcia, "Smart people: enhancing CCD citizens' engagement through data utilization," 2014. [Online]. Available: http://smartcities.ieee.org/home/whitepapers.html .

[5] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. A. Pardo and H. J. Scholl, "Understanding Smart Cities: An Integrative Framework," in *45th Hawaii International Conference on System Sciences*, 2012.

[6] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions.," in *Proceedings of the 12th Annual International Digital government Research Conference: Digital Government Innovation in Challenging Times*, 2011.

[7] H. Shaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson and A. Oliveira, "Smart Cities and the Future Internet: Towards cooperation frameworks for Open innovation," in *Future Internet Assembly*, 2011, pp. 431-446.

[8] C. Ratti, D. Frenchman, A. C. srl, E. s. Arup, F. Metrópoli, M. M. i. Chain and S. F. M. srl, "Ciudad Creativa Digital Master Plan," 2012.

[9] D. Keim, J. Kohlhammer, G. Ellis and F. (. Mansmann, Mastering the Information Age Solving Problems with Visual Analytics, 2010.

[10] J. Blumenstock, "INFX 598 C/D: Introduction to Data Science," 2013. [Online]. Available: http://www.jblumenstock.com/teaching/course=infx598.

[11] S. Argamon, "The Well-Rounded Data Scientist," 2014. [Online]. Available: http://www.predictiveanalyticsworld.com/patimes/well-rounded-data-scientist/.

[12] D. Conway, "The Data Science Venn Diagram," 2013. [Online]. Available: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

[13] D. J. Patil, "Building data science teams," 2011. [Online]. Available: http://radar.oreilly.com/2011/09/building-data-science-teams.html.

[14] M. Waller and F. S. E., Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management, Wiley, 2013.

[15] IBM, "What is a data scientist," 2013. [Online]. Available: http://www-01.ibm.com/software/data/infosphere/data-scientist/.

[16] C. Miller, "Data Science: The Numbers of Our Lives," 2013. [Online]. Available: http://www.biostat.jhsph.edu/courses/bio624/misc/lod2013/universities.pdf.

[17] Department For Business, Innovation & Skills, "Smart Cities: Background paper," 2013.

[18] C. Ratti, D. Frenchman, C. srl, Accenture, Arup, E. srl, F. Metrópoli, M. M. i. Chain and S. F. M. srl, "Plan Maestro Guadalajara Ciudad Creativa Digital," 2012.

[19] L. Lei, Z. Hui, L. Jianhui, W. Runqiang, Y. Luqing, Y. Jianjun and P. Li, "Building a community of data scientists: an explorative analysis," *Data Science Journal,* vol. 8, no. 24, 2009.

[20] G. Liviu, "Smart Cities Design using Event-Driven Paradigm and Semantic Web," *Informatica Economica,* vol. 16, no. 4, 2012.

[21] X. Nie, "Constructing Smart Campus Based on the Cloud Computing Platform and the Internet of Things," in *Proceedings of the 2nd International Conference on Computer Science and electronics Engineering (ICCSEE 2013)*, 2013.

[22] J. A. Fredricks, P. C. Blumenfeld and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research,* pp. 59-109, 2004.

[23] Mexican Federal Government, "Estrategia Diginal Nacional," www.cidge.gob.mx, 2013.

De Obeso-Orendain, A. (alberto.de.obeso.orendain@intel.com**)** is a data scientist at Intel focused on providing Business Intelligence tools to the marketing area. He obtained his PhD degree in Informatics from the University of Sussex for his research on the cognitive modelling of complex dynamic tasks, and he teaches a Master's course at the Tecnológico de Estudios Superiores de Occidente (ITESO).

Lopez-Neri, E. (emmanuel.lopezne@uvmnet.edu) is a full professor at the Research, Innovation and Technological Development Center at UVM Guadalajara Sur Campus (CIDETEC-UVM). His research interests include crowdsourcing, multi-agent system modelling, and event-oriented modeling and simulation. Website: http://www.cidetec-uvm.com.

Donneaud-Bechelani, C. (Claudia.donneaudbe@uvmnet.edu) is a full professor at the Research, Innovation and Technological Development Center at UVM Guadalajara Sur Campus (CIDETEC-UVM). His research interests include strategies for profiling jobs. Website: http://www.cidetec-uvm.com.