Visual Exploration of Formal Requirements for Data Science Demand Analysis

Conference or Workshop Item

Visual Exploration of Formal Requirements for Data Science Demand Analysis

Aba-Sah Dadzie and John Domingue

KMi, The Open University
Milton Keynes, UK
{aba-sah.dadzie, john.domingue}@open.ac.uk

Abstract. The era of Big Data brings with it the need to develop new skills for managing this heterogenous, complex, large scale knowledge source, to extract its content for effective task completion and informed decision-making. Defining these skills and mapping them to demand is a first step in meeting this challenge. We discuss the outcomes of visual exploratory analysis of demand for Data Scientists in the EU, examining skill distribution across key industrial sectors and geolocation for two snapshots in time. Our aim is to translate the picture of skill capacity into a formal specification of user, task and data requirements for demand analysis. The knowledge thus obtained will be fed into the development of context-sensitive learning resources to fill the skill gaps recognised.

Keywords: big data, visual exploration, visual analytics, demand analysis, RtD, data-driven decision-making, ontology-guided design

1 Introduction

We are in the middle of a technological, economic, and social revolution. How we communicate, socialise, occupy our leisure time, learn and run a business has slowly moved online. In turn, the Web has entered our phones, our newspapers and notebooks, our homes and cities, and the industries that power the (digital) economy. The resulting explosion of data is transforming enterprise, government and society.

These developments have been associated with a number of trends of which the most prominent is *Big Data*. As noted recently by Google [1], what is important about data is not volume, but its contribution to innovation and, thereby, value creation. We agree with this assessment of the situation today: we are in the midst of a Data Driven Innovation (DDI) revolution. The benefits for DDI will be significant. Studies suggest that companies that adopt data driven decision-making have an output and productivity 5-6% higher than would be expected given their IT investments alone [5]. This assessment is backed up by Cisco, who report [4] that the Internet of Everything (the confluence of people, processes, data and things) will create $14.4 trillion in value globally through the combination of increased revenue and cost savings. McKinsey [15] makes similar predictions. Big Data has an estimated value of $610 billion across four sectors in the US (retail, manufacturing, healthcare and government services), with open data alone raising more than $3 trillion per year in seven key business sectors worldwide – education, transport, retail, electricity, oil & gas, healthcare, consumer finance [17].

According to a recent OECD report EU governments could reduce administrative costs by 15-20% by exploiting public data, equivalent to savings of €150-300 billion [5].

A major blocker for these promising prospects is the lack of *Data Science* skills within the workforce, be that technical specialists, managers or public servants. A well-known McKinsey study [16] estimated in 2011 that the US would soon require 60% more graduates able to handle large amounts of data effectively as part of daily work. With an economy of comparable size (by GDP) and similar growth prospects, Europe will most likely be confronted with a talent shortage of hundreds of thousands of qualified data scientists, and an even greater need for executives and support staff with basic data literacy. The number of job descriptions and increasing demand in higher-education programmes and professional training confirm this trend,[1] with some EU countries forecasting an increase of over 100% in demand for data science positions in less than a decade.[2] For example, a recent report by e-Skills UK/SAS [7] notes a tenfold rise over the past five years in demand for Big Data staff in the UK, and a 41% increase in the number of such jobs posted during the 12-month period from Oct 2013–Oct 2014, with over 21,000 vacancies in 2013. The study also estimated that 77% of Big Data roles were "hard-to-fill" and forecast a 160% increase in demand for Big Data specialists from 2013–2020, to 346,000 new jobs. Similar trends may be extrapolated to other EU countries. The European Commission's (EC) communication on '*Towards a thriving data-driven economy*' highlights an "*adequate skills base is a necessary condition of a successful data driven economy*"[3], while the Strategic Research and Innovation Agenda (SRIA) for the Big Data Value contractual Public-Private Partnership (cPPP) lists Big Data skills development as their top non-technical priority.[4]

The *European Data Science Academy* (EDSA)[5] is a new EU-funded project which will deliver the learning tools that are crucially needed to close this problematic skills gap. EDSA will implement cross-platform, multilingual data science curricula which will play a major role in the development of the next generation of European data practitioners. To meet this ambitious goal, the project will constantly monitor trends and developments in the European industrial landscape and worldwide, and deliver learning resources and professional training that meets the present and future demands of data value chain actors across countries and vertical sectors. Thus, a core part of our work is focused on demand analysis. We need to ensure that the data science curricula and associated learning resources that we create meet the needs of industry across Europe, recognising that this will vary by sector, job role and geographical region. In this paper we describe some of the visual tools we are developing to support our *Demand Analysis*. Through our visual analysis, we aim ultimately to make visible to a wider audience data we are collecting through a combination of interviews with key stakeholders, online surveys and data mining of job websites.

We continue in section 2 with a discussion of related work. We then describe in section 3 the methodology we are following, through data exploration (detailed in sec-

[1] Government calls for more data scientists in the UK: `http://bit.ly/1RLztP8`
[2] Demand for big data IT workers to double by 2017 ... `http://bit.ly/1Ntwcm8`
[3] Communication on data-driven economy: `http://bit.ly/1pNmxQq`
[4] SRIA on Big Data Value for Europe: `http://bit.ly/1LDSR1C`
[5] European Data Science Academy: `http://edsa-project.eu`

tion 4), to uncover design and task requirements. We discuss our findings in section 4.1 and feed these into the definition, in section 5, of target users, typical user tasks and the data necessary for users to meet their goals. We conclude in section 6 with pointers to the next stage in our study. We envisage, through this process in which we use dynamic, living data to guide our investigation, to identify intuitive, expressive approaches that will aid our analysis and serve as pointers to our targets, mainly Data Scientists, to a picture of capability and demand for their skills in today's data driven economy.

2 Related Work

Ontologies provide a useful framework for capturing, sharing and guiding (re)use of knowledge about an object, a domain or a situation [9,13,18,22]. Devedzic [6] in an early paper forecast the utility of ontologies for the Semantic Web and to improve competition in industry. The survey ([6]) illustrates how knowledge modeling and acquisition using ontologies aids collaboration and interoperation within and across disciplines, by providing standardised references to, and, therefore, interpretation of knowledge extracted from independent sources. The ESCO (European Skills, Competences, Qualifications and Occupations) vocabulary [8], for instance, was built to reduce recognised mismatch between demand in employment sectors across the EU and expertise in the current and future workforce. ESCO aims to help reduce unemployment by matching also to up-to-date training for each market. A final version is to be released in 2017 as Linked Open Data (LOD) to increase reusability in, e.g., statistical and demand analysis.

Ontologies may also be used to directly influence design, development and use of technology [13,22]. Grimm *et al.,* [9] describe their use to guide design choices during software development, to generate metadata about design and other intermediate artefacts created during the software development process and, therefore, improve communication between developers. Paulheim *et al.,* [18] survey work carried out to enhance capability in employing user interfaces (UIs) for specified tasks using ontologies, e.g., for filtering, clustering, visualising and navigating through information, as well as customising the UI itself for a task, user type or the user's environment. In our case, we aim to employ ontologies to guide: (i) knowledge capture – about demand for Data Science skills and capability to meet this demand, (ii) (re)use of this knowledge for context-driven, analytical and decision-making support, (iii) through an interface that supports context- and user-centred exploration and extraction of information about skill gaps and training resources for plugging them.

Visual analytics provides an intuitive, interactive approach for extracting task-based knowledge from complex data such as in our use case [12]. Both visualisation technique and how it is applied influence where visual and cognitive attention are directed and how data content is perceived and interpreted [10,11]. Especially for abstract, dynamic, large scale, multi-dimensional data, therefore, it is useful to provide alternative perspectives on the same dataset. These, used in isolation or in concert, allow different patterns and relationships to be revealed, triggering insight and resulting in more comprehensive exploration. Further, integrating (highly advanced) human perception into the analytics loop, to guide data processing, analysis and visualisation widens the scope for exploration and increases confidence in decision-making based on the results obtained [10,12].

Reusable, extensible libraries and APIs for already proven visualisation and analytics techniques are particularly useful in such cases, as they ease development of and interaction with visual analytics tools [3,12], allowing a focus on research into novel solutions and their application and evaluation. However, identifying the optimal technique(s) for a task is influenced by a variety of factors, including user skill, data and domain, the task itself and whether and what subsequent use will be made of the results [10,11,12]. Taxonomies and ontologies play a useful role here by providing a formalism for specifying requirements and translating them into design. They may be used to document best practice for specific use cases, thereby providing context-based design guidelines [13,14,18]. We aim to harness ontologies to drive and document our design activities, to guide the development of an intuitive, reusable, extensible solution that serves the user's particular and evolving needs and context.

3 Methodology

Keim *et al.,* [12] extend the "visual information seeking mantra" [20] by placing first analysis of the data and/or situation, before an overview that highlights salient information, followed by exploration and further, detailed analysis of regions of interest (ROIs). In line with user-centred design (UCD) principles and recommended practice in visual analytics [10,11,12], we must ensure an intuitive UI and interaction methods that allow a focus on user tasks rather than the tool or its interface. We follow the principles of research through design (RtD) [2,19], using the process of exploring the knowledge and design space, during iterative data exploration, to probe initial and reveal additional requirements – see Fig. 1.

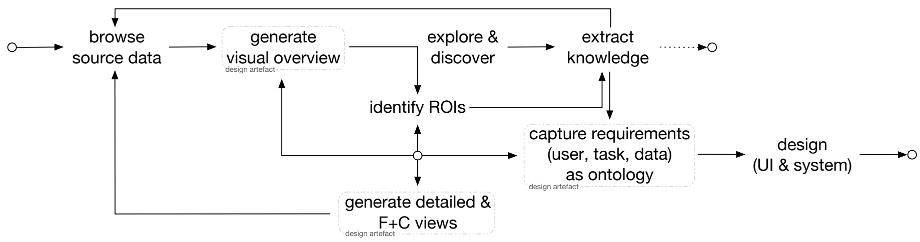


Fig. 1. Methodology followed in exploratory study, highlighting (with a faint border) where design artefacts are generated.

This study bears some similarity to [11]; however, we do not seek to formalise design patterns for composite visualisations. We aim, from the knowledge exploration exercise detailed in sections 4 and 5, to identify a range of intuitive visual analytics options, and as a result design that will guide customisation for use, individually and in concert, to meet the needs of different user types and tasks within the scope of demand analysis as described in sections 1 and 2.

4 Initial, Visual Exploration of Demand Data

We present a case study in demand analysis: investigating capacity against capability in key industrial sectors within the EU for expertise in Data Science. The dataset used comprises summary data crawled from LinkedIn, the first of a number of target domain expert networks and web services advertising job postings. Search terms are core and specialised skills grouped into seven skillsets (e.g., as listed in Fig. 2), identified through interviews with policy and decision-makers in industry across the EU and a focus group in the UK. Each skillset is translated into the official or dominant working language (37 in all) in each of 47 European countries. Due to restrictions to data extraction on LinkedIn only term frequency in job adverts is currently extracted, collected daily across three top-level dimensions: (i) industrial sector (captured as skill), (ii) geolocation and corresponding language and (iii) time. (Term selection and the data collection process is detailed in ([21].) While the dataset size is currently relatively small, complexity is introduced by large differences in scale and cross-linking within it. A key requirement is therefore scalability, to handle growth in size with time, and also complexity as additional context is mined from LinkedIn and other relevant sources.

We focus in this study on the temporal element in the data, treating the spatial attribute (geolocation and, by derivation, language) as an additional lens (filter) through which we examine the non-spatial data (skillsets). The aim here is two-fold: to identify effective methods for revealing, first, temporal patterns in the data, and then relating these to our target audience, using techniques that speak to them [10,12]. Another key requirement is therefore learnability, and to a point, customisability. The second goal is, in the process of exploring the data, to clarify our target end user characteristics and identify key tasks users would expect to be able to perform. We expect as a result, also, to identify additional data requirements (structure and content) for completing these tasks.

This necessitates data exploration from different perspectives, to identify where and how insight is triggered and which views reveal ROIs and answer key questions. While our overall goal includes the exploration of novel (visual) analytics approaches, this exploratory exercise focused on obtaining an initial, broad picture of demand and the identification of ROIs – anomalies, peaks and islands – within the data. We therefore made use of web-based solutions able to support quick prototyping of simple, yet informative overviews. A number of research prototypes and working visualisation tools, graphics libraries and APIs exist, implementing one or more of a range of visual analysis techniques (examples can be found in [10,11,12,14,20]). These include (not considering 3D for practical reasons): for high-dimensional data – parallel coordinates and small multiples (e.g., scatterplot matrices); techniques useful for temporal or dynamic data such as timelines and theme rivers; cartographic or geographical plots; statistical charts (e.g., line and scatter plots, bar and pie charts); and finally, techniques typically applied to non-spatial data such as word maps, tree and node-link graphs, and space-filling techniques such as tree maps and sunbursts. Freeware and open source tools such as *p5.js*, *Cytoscape.js*, *Raphaël*, *D3.js* and *Leaflet DVF* vary with regard to scalability, stability, author support, user community and compliance with web standards. Tools backed by commercial organisations, such as *Visual.ly*, *Tableau Public*, *IBM Watson Analytics* and *Google Charts* typically make available a limited set of features as free to use and/or open source, often with restricted licenses. Such services may also require data upload to company servers.

For this exercise we use *D3.js*, a relatively well-established JavaScript library developed for "data-driven", interactive visualisation [3]. D3.js was built to overcome challenges encountered by its authors using existing web-based libraries, due to, among others, reliance on custom features with inconsistent compliance with web standards and browsers, or with high complexity. (Server-side) data input and initial, basic parsing was carried out with PHP, reading from the demand data written to CSV. The visualisations have been tested in Firefox, Chrome and Safari. It should be noted that not all events are triggered in all browsers, e.g., onChange in drop-down lists. We report our findings, in section 4.1, from the first three visual analytics techniques we employed:[6]

(i) ***line*** and ***dot timeline plots***, to obtain an overview of trends and variation in demand (patterns) over time;
(ii) ***small multiples***, employing a ***matrix plot***, to compare variation in patterns in attributes of interest for each data point and the whole dataset;
(iii) ***aster plots***, to examine skill demand by location.

We then discuss, in section 5, formal requirements specification and the implications for design for intuitive, interactive demand analysis and decision-making.

4.1 Findings

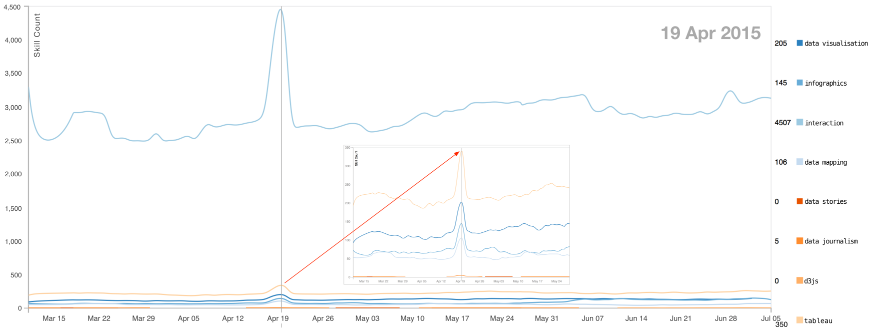
Summary statistical analysis showed some skew[7] with counts several times higher across all skillsets for *language* == *'gb'* (English) and also for the UK only. However, further investigation showed more uniform relative distribution across countries. We therefore normalise the data or exclude the UK or all 'gb' countries where necessary to reveal trends suppressed in the remaining data.

Fig. 2 shows the trend in three *timeline* plots from the start of the collection period, 11 Mar 2015, to 05 Jul 2015 for demand for the skillset *visualisation* for 'gb' countries. There is a small peak early in the plot, and a sharp rise from 17th Apr to the 20th, peaking on the 19th. Beyond this there is in general a gradual rise for the rest of the period, with a few small dips. Trends are similar across all skills but *D3.js*, which records no counts till 11 Jun, rising to five on the 30th.

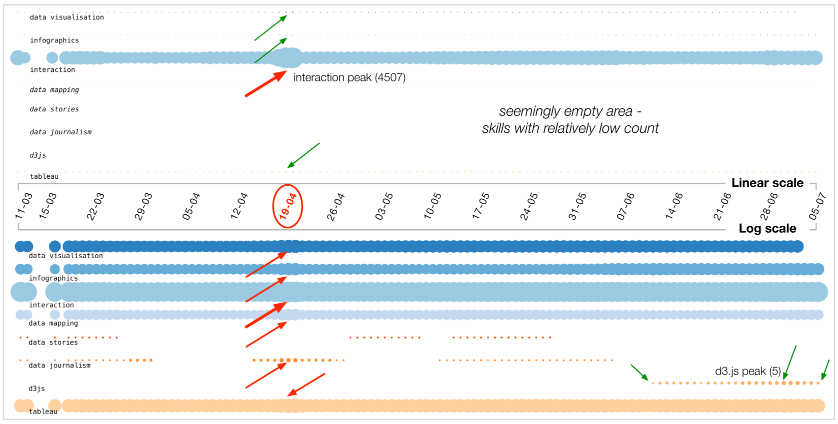
One skill, *interaction*, records counts more than 20 times greater than all others, suppressing, as a result, trends in the latter (see Fig. 2a). We looked at two options for revealing this detail. In the line chart hiding the outlier and modifying, correspondingly, the range of the y-axis, provides more space for the remaining attributes (see inset, Fig. 2a). We show also in Fig. 2b the second option: the bottom plot contrasts a $\log_n$ scale with the linear scales used in the other two plots. By stacking the journal plot with the linear scale on top of the normalised plot we obtain two gains. We are able to examine relative trend for each skill, still within the context of overall demand patterns, but with little increase in cognitive load.

[6] Additional, high-resolution snapshots (including video) at: `http://bit.ly/1FLevZE`

[7] The skew may be due in part to differences in terminology usage and interpretation across regions and/or the translator used. Further, the data collected to this point comes from a single (web) source. While we take this into consideration for further analyses, a full investigation of the cause of the skew is out of scope for this paper.



(a) A multi-line chart showing the demand trend for a selected skillset



(b) Normalisation using a $\log_n$ scale to reveal relative trend for each skill

Fig. 2. Daily demand for *'gb' visualisation*; large variation across skillsets for the four month period requires filtering (2a) and/or normalisation (lower plot, 2b) to reveal suppressed patterns.

We looked next at a data snapshot, taken in Apr 2014,[8] aggregating for five countries (excluding the UK), demand categorised under eleven key topics in Data Science (see [21] and bottom, right, Fig. 3). We use *small multiples* to examine multiple attributes simultaneously. Fig. 3 stacks, from left-right and top-bottom, mini plots showing in descending order counts for demand for each skill for each country, followed by skill and country percentage. Dynamic sliders are used to examine each skill as a percentage of the total demand for all countries (in the dataset, including the UK – olive inner border), and skill distribution within each country (faint blue outer border). The snapshot highlights values for skill and country percentage greater than 30 and 25% respectively. Blue borders are found predominantly at the top, showing top-heavy demand for selected skills. This is mirrored in the colour-coded line plot for the full dataset overlaid

[8] While the picture of demand continues to change the variables to be examined remain the same. Comparing earlier findings with current demand allows us to revisit the initial project requirements defined with respect to data structure and content.

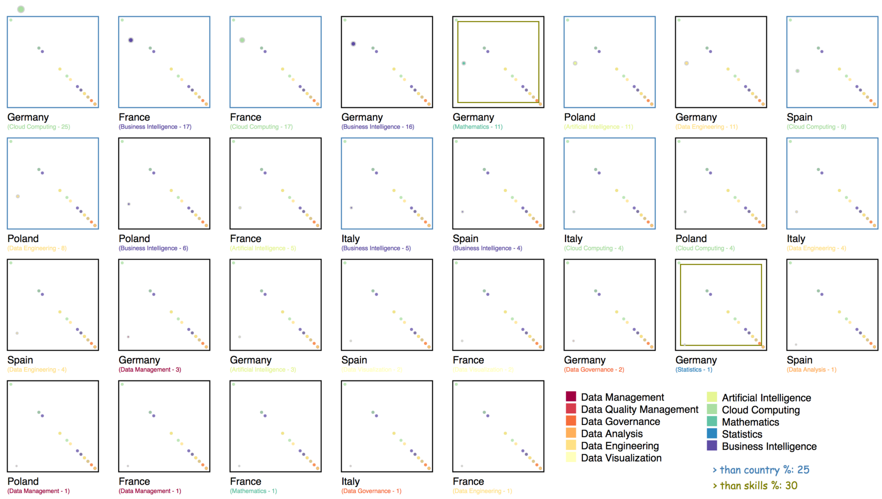


Fig. 3. Small multiples used with dynamic filters to investigate trends across three attributes – skill count and distribution (%) by country and skill percentage per country. Data for the UK, which is ∼70% of the total (see Fig. 4) is filtered out.

on each mini plot, which shows a long tail with very small counts per skill and country. Olive borders are more randomly distributed – e.g., Germany sees 50% of all *Statistics*, with a count of one, near the tail end of the chart (the other 50% (one) is in the UK).

We used, next, a space-filling technique, *aster plots*, a variant of a nested pie chart, to examine further skill distribution within each country, including the UK, using again the small multiples technique. Fig. 4 shows on the far left an overview of total demand for each country, then distribution by skill. Area maps to count for each slice. For the first aster, height maps to skill count per country (up to 11), and for all others skill percentage (over all countries). While the UK dominates all others, the individual country plots reveal a degree of similarity in skill distribution. *Business Intelligence*, *Data Engineering* and *Cloud Computing* are in demand across all, followed by *Artificial Intelligence*, which is highest in Poland. One skill, *Data Quality Management*, records one count in only one country, the UK – so slim that only by thickening the borders of each inner slice, to provide an additional visual cue, is it recognised. Here, we see the

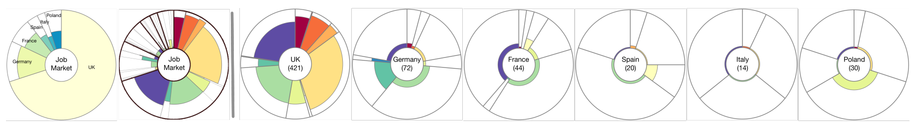


Fig. 4. The overview (first two from left) shows total demand per country and distribution of skills, respectively, for the six countries that follow, clockwise, from 12:00. As in Fig. 2, juxtaposing the two sets of charts enables *focus* on the detail for each country within the *context* of the overview (F+C). Colour coding is as in Fig. 3.

power in multiple perspectives – this is highlighted in the matrix plot (e.g., in Fig. 3) for the full dataset with the skills slider set to the maximum (100%).

5 Core Requirements for Demand Analysis

A key point reiterated throughout this exercise is the importance of letting the data drive our exploration of the knowledge and design space. By examining the questions raised as we explored this initial dataset, we have identified four areas that we must address if we are to meet our goals for demand analysis: (i) definition of target users and (ii) the tasks relevant to each; (iii) data, and therefore, knowledge requirements for effective task performance and completion, all of which lead to: (iv) effective support for decision-making. Fig. 1 shows three points at which we expect to generate design artefacts; we focus in this section on requirements specification, a living artefact that will evolve with the project. To allow use also as a communication tool with end users and within the project team, and to feed into the design, development and evaluation cycle, we aim to specify our requirements formally using an ontology. At this early stage we use simple structure diagrams, as in Fig. 5, to map this knowledge space. We document in the process, also, existing standards that we may reuse.

5.1 Target User Types

In line with our aim to match skill gaps in Data Science with context-sensitive training, at the start of the study we had two key targets identified: *Data Scientists*, practising

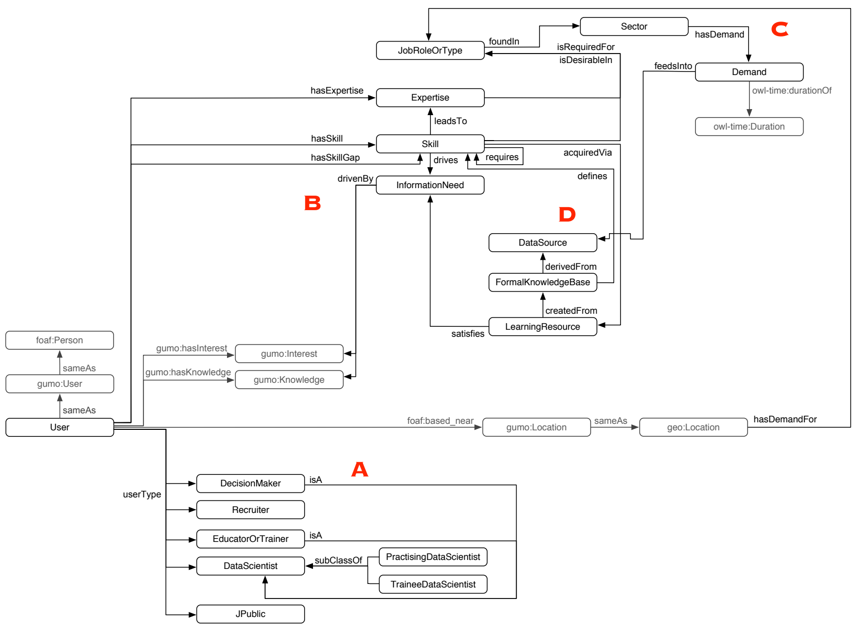


Fig. 5. High level definition of knowledge structure for design space

or new, and the institutions and personnel (*Educators* or *Trainers*) who will develop learning resources. In exploring the data we recognised that the picture of demand as extracted from recruitment sites such as LinkedIn may not adequately define the resources required to fill the gaps identified. A third user is critical in ensuring that training resources meet the needs of the market: the *Decision Maker* ultimately responsible for the definition of new roles and essential skills for them, who will also influence training of new and existing employees. Decision makers may be data scientists or other technology or domain experts. Finally, we recognise two additional user types. Recruiters may influence the language in and the interpretation of job adverts. In line with requirements for communicating the results of research to the (interested) public, we include the non-domain end user, who may or may not be a technology expert.

5.2 Task Identification and Requirements Analysis

Sections B and D of Fig. 5 focus on task requirements for data scientists, looking at what factors drive the acquisition of new skills and what is required to identify which skills are in demand and where. For clarity, Fig. 5 omits the detailed breakdown of skills as shown in, e.g., Fig. 2; the demand data will however be mapped along these categories (classes) to related information captured from, among others, LOD and OERs (Open Educational Resources). Our exploration raised three questions which impact the development of training resources: (1) which skills are typically required together? (2) how are skills ranked in and (3) how transferable are skills across sectors and regions? We must provide means for users to answer these and other questions, and also filter the data on different criteria, with a baseline for filtering on *sector*, *skill*, *location* and *time*, before matching to resources recognised to meet their needs. For technology experts (such as data scientists) functionality for complex query formulation should be useful for more complex analysis and knowledge retrieval.

5.3 Input Data Requirements

The summary data gives a broad overview of demand. However, our analysis is limited by the lack of context; while distinct patterns can be seen, what led to them cannot be ascertained. To complete the picture of demand we need to fill in the gaps resulting from content access restrictions on web and other services. We require detail that answers questions such as raised in section 5.2, and that may be used to enrich information extracted from other target sources. Data requirements include: (i) ideally, full text for role descriptions, containing, among others: (ii) term frequency per advert, (iii) weighting of terms, i.e. required vs. desired; and (iv) other detailed views on the sectors and skills of interest, from, e.g., related ontologies and vocabularies, LD and OERs.

With this additional detail we can begin to build a knowledge base (KB) that our target may draw from to make informed decisions, based on current market data and context-specific skills training. In section D of Fig. 5 we use a catch-all – *Data Source* – to represent both the data mined specifically for the project and other third party KBs. In the next stage of our study we will define more fully target KBs and how we will link to and reuse their content.

6 Discussion & Conclusions

Big Data presents a challenge for industry, due not just to its scale and the rate at which it continues to grow, but because its inherent heterogeneity and complexity present additional challenges for mining and reusing its valued content, value which contributes to gaining competitive advantage. Making effective use of Big Data starts with the specification of the skills required of the *Data Scientist*, for roles specific to and that span industrial sector and local context.

Our exploratory study has provided an initial picture of the demand for Data Science skills in key sectors across the EU. We have, in the process, uncovered questions with implications for the design of effective, intuitive knowledge exploration and analytical support for our target users. Demand is in turn sparse and large, within each and across skillsets. Relative distribution, conversely, is fairly uniform across location, but with instances where a specific skill is isolated to a small pocket. We must bear in mind, however, a key limitation in our study – the loss of context in the summary data. A second is that our current picture of demand comes from a single source, albeit extracted using search terms specified by a range of technology and domain experts. These impact the depth of analysis and the determination of optimal techniques for doing so, both for our requirements and those of our target users. We must therefore design for scalability, to manage continued increase in size and complexity and the potential for even greater variability. This demands alternative perspectives from which to examine data content and structure. Following the methodology used in this study, we will investigate additional approaches to ensure the generation of intuitive, informative overviews, with lenses for detailed analysis tailored to the user's particular context.

The knowledge structure summarised in Fig. 5 is a living document that will evolve as we obtain a more balanced and complete picture of demand and corresponding skill gaps. We have started to map the concepts and relationships defined so far to other related knowledge sources such as the ESCO vocabulary and new information obtained from further interviews with industry experts. This is to enable more detailed examination of the relationships between skills within and across skillsets and industry sectors. The aim is to refine our current skill definitions and map these to role descriptions. We will then be able to return to our target users, to review the formal, updated requirements and discuss further design for the analytical and decision-making tools they require.

The ultimate aim is to map the picture of demand to the user's specific requirements and feed the knowledge thus obtained into developing effective learning resources. This is to aid data scientists and decision-makers in industry and academia to identify optimal paths to acquiring and updating skills that meet the requirements for managing Big Data in the modern digital economy.

Acknowledgments. The work reported in this paper was funded by the EU project EDSA (EC no. 643937).

References

1. Andrade, P.L., Hemerly, J., Recalde, G., Ryan, P.: From Big Data to Big Social and Economic Opportunities: Which policies will lead to leveraging data-driven innovation's potential. In:

GITR 2014: The Global Information Technology Report 2014: Rewards and Risks of Big Data, pp. 81–86 (2014)
2. Bardzell, J., Bardzell, S., Koefoed Hansen, L.: Immodest proposals: Research through design and knowledge. In: CHI '15: 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 2093–2102 (2015)
3. Bostock, M., Ogievetsky, V., Heer, J.: D^3: Data-driven documents. IEEE Transactions on Visualization and Computer Graphics 17(12), 2301–2309 (2011)
4. Bradley, J., Barbier, J., Handler, D.: Embracing the Internet of everything to capture your share of $14.4 trillion. Tech. rep., Cisco (2013)
5. Brynjolfsson, E., Hitt, L.M., Kim, H.H.: Strength in numbers: How does data-driven decision-making affect firm performance? Social Science Research Network (2011)
6. Devedzic, V.: Knowledge modeling – state of the art. Integrated Computer-Aided Engineering 8(3), 257–281 (2001)
7. Big data analytics: Assessment of demand for labour and skills 2013–2020. Tech. rep., e-Skills UK/SAS (2014)
8. ESCO: European classification of skills/competences, qualifications and occupations: The first public release – a Europe 2020 initiative. Tech. rep., Luxembourg: Publications Office of the European Union (2013)
9. Grimm, S., Abecker, A., Vlker, J., Studer, R.: Ontologies and the Semantic Web. In: Handbook of Semantic Web Technologies, pp. 507–579. Springer (2011)
10. Heer, J., Bostock, M., Ogievetsky, V.: A tour through the visualization zoo. Communications of the ACM 53(6), 59–67 (2010)
11. Javed, W., Elmqvist, N.: Exploring the design space of composite visualization. In: PacificVis: 2012 IEEE Pacific Visualization Symposium. pp. 1–8 (2012)
12. Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: Definition, process, and challenges. In: Information Visualization: Human-Centered Issues and Perspectives, pp. 154–175. Springer (2008)
13. Kitamura, Y.: Roles of ontologies of engineering artifacts for design knowledge modeling. In: Design Methods for Practice. The Design Society (2006)
14. Lohse, G.L., Biolsi, K., Walker, N., Rueter, H.H.: A classification of visual representations. Communications of the ACM 37(12), 36–49 (1994)
15. Lund, S., Manyika, J., Nyquist, S., Mendonca, L., Ramaswamy, S.: Game changers: Five opportunities for US growth and renewal. Tech. rep., McKinsey Global Institute (2013)
16. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity. Tech. rep., McKinsey Global Institute (2011)
17. Manyika, J., Chui, M., Farrell, D., Kuiken, S.V., Groves, P., Doshi, E.A.: Open data: Unlocking innovation and performance with liquid information. Tech. rep., McKinsey Global Institute (2013)
18. Paulheim, H., Probst, F.: Ontology-enhanced user interfaces: A survey. International Journal on Semantic Web and Information Systems 6(2), 36–59 (2010)
19. Pierce, J.: On the presentation and production of design research artifacts in HCI. In: DIS '14: the Designing Interactive Systems Conference. pp. 735–744 (2014)
20. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: 1996 IEEE Symposium on Visual Languages. pp. 336–343 (1996)
21. Tarrant, D., Bullmore, S., Costello, M.: Deliverable D1.1: Study design document. Tech. rep., European Data Science Academy (EDSA), EC. project 643937 (2015)
22. Zlot, F., de Oliveira, K.M., Rocha, A.R.: Modeling task knowledge to support software development. In: SEKE '02: Proc., 14th International Conference on Software Engineering and Knowledge Engineering. pp. 35–42 (2002)