Professional Information Skills and Open Data. Challenges for Citizen Empowerment and Social Change

Competencias informativas profesionales y datos abiertos. Retos para el empoderamiento ciudadano y el cambio social

Dr. María-Carmen Gertrudis-Casado is Researcher for the Ciberimaginario Group at the Rey Juan Carlos University in Madrid (Spain) (carmen.gertrudis@ciberimaginario.es) (http://orcid.org/0000-0001-6770-203X)

Dr. Manuel Gertrudix-Barrio is Professor of Digital Communication from the Communication Faculty at the Rey Juan Carlos University in Madrid (Spain) (manuel.gertrudix@urjc.es) (http://orcid.org/0000-0002-5869-3116)

Dr. Sergio Álvarez-García is Visiting Lecturer with a PhD in the Communication Faculty at the Rey Juan Carlos University in Madrid (Spain) (sergio.alvarez@urjc.es) (http://orcid.org/0000-0001-9477-2148)

ABSTRACT
The current process of social transformation is driven by the growth of the culture of transparency and accountability, the socio-technological development of the web and the opening of public data. This situation forces the media to rethink their models of social intermediation, converting the growing open data access and user participation into new instruments that facilitate citizen empowerment. Open data can only generate citizen empowerment, facilitate decision-making and democratic action if it can provide value-added information to the citizens. Therefore, the aim of the research is to analyse the competencies necessary to develop information products created with open data. The study used a qualitative methodology based on two instruments: a survey of data journalism experts (university professors of journalism, journalism professional data, and experts in transparency), and an analysis of selected cases of information products created with open data. The results allow the identification of a series of conceptual, procedural and attitudinal skills needed to perform the tasks of collection, processing, analysis and presentation of data, which are necessary for the development of this type of information product, and which should be integrated into the training of future journalists.

RESUMEN
Los actuales procesos de transformación social estimulados por el crecimiento de la cultura de transparencia y rendición de cuentas, el desarrollo socio-tecnológico de la web y la apertura de datos públicos, obliga a los medios de comunicación en el entorno digital a reorientar sus modelos de intermediación social, convirtiendo el creciente y complejo acceso a datos abiertos y los flujos de participación en nuevos instrumentos que faciliten el empoderamiento ciudadano. La investigación evalúa cuáles son las competencias profesionales necesarias para el desarrollo de productos informativos multimedia interactivos basados en datos abiertos, considerando que la apertura de datos solo generará empoderamiento ciudadano, facilitará la toma de decisiones y la acción democrática, si estos pueden proporcionar información de valor añadido para la ciudadanía. Para ello, se sigue una metodología cualitativa basada en dos instrumentos: una encuesta a expertos en periodismo de datos, relacionados con la formación superior en Periodismo, la legislación en materia de acceso a la información y los medios de comunicación, y el análisis de una muestra de productos informativos multimedia basados en datos abiertos. Los resultados permiten identificar una serie de competencias conceptuales, procedimentales y actitudinales necesarias para llevar a cabo las tareas de acopio, tratamiento, análisis y presentación de los datos, que son necesarias para el desarrollo de este tipo de productos informativos, y que deberían integrarse en la formación de los futuros comunicadores.

KEYWORDS | PALABRAS CLAVE
Professional skills, social change, open data, skills, citizen empowerment, data journalism, digital communication, multimedia, civil society.
Competencias profesionales, cambio social, datos abiertos, empoderamiento ciudadano, periodismo de datos, comunicación digital, multimedia, sociedad civil.
1. Introduction and state of the question

1.1. Changes in the media ecosystem

In today’s media ecosystem, the phenomenon of media convergence has encouraged a change in the role people play with regard to media. It has transformed them from a passive mass to an audience, and from an audience to active individuals or users in search of information (Canavilhas, 2011; Jenkins, 2008), and from mere consumer users to producers of information. This has given rise to new concepts that foster a mass collaboration among users to create products (Tapscott & Williams, 2011), in a framework which Rifkin (2014) labelled «Collaborative Commons». Such an ecosystem has led to a professional convergence (Canavilhas, 2013) as a result of workforce reductions and the emergence of new profiles to cater for new needs.

This convergence comes at a time of not only a financial crisis, but also an ethical (Álvarez, 2014) and a functional one (Dader, 2010). It has brought about a profession that is politicised and polarised in the eyes of the professionals of the future and little appreciated by civil society (CIS, 2013).

This means redefining the journalistic profession, both in terms of its situation within the media ecosystem and of its purpose, which for Kovach and Rosenstiel (2012: 18) is to «provide citizens with the information they need to be free and self-governing», the essence of which lies in the principle of verification, where transparency is not just a fundamental element but proof of a commitment to civil society.

1.2. New qualifications required for journalism professionals

Ten years after the Spanish Agency for Quality Assessment and Accreditation (ANECA) drafted its white paper (in 2005), Spain’s university qualifications in journalism are still struggling to adapt the country’s curricula to the needs of the European Higher Education Area (EHEA) in what seems like a constant search for new needs.

This renewed intermediary —not hegemonic— role must inevitably learn to coexist with the new phenomena brought about by the Internet and 2.0 tools, such as citizen journalism (Espiritusanto & Gonzalo-Rodríguez, 2011; García de Madariaga, 2006), a concept which refers to active, non-professional citizen involvement in the world of journalism (Gillmor, 2004; Meso-Ayerdi, 2005; Sampedro, 2009) and means rethinking their mediation models (Baack, 2015), but which still raises questions as to their dependence on the media, their lack of periodicity and the scarcity of reliable sources (Rich, 2008; Varela, 2005).
1.3. Open data and participation as tools for citizen empowerment

The 2.0 network and the social-technical worldview behind it (García-García & Gértrudix-Barrio, 2012) have encouraged the emergence of new citizen participation models such as crowdsourcing (Howe, 2006). In the journalistic world, this means professional journalists gathering information from a large number of citizens (Méndez-Majuelos, Pérez-Curiel, & Rojas-Torrijos, 2012), such that their contributions have become a source for the journalist’s work (Gillmor, 2004).

A digitally interconnected society demands that institutions open up democratically. The Open Data phenomenon is linked to that of Open Government and Free Culture, providing opportunities that will help public sector information and its re-use to become an asset for citizen empowerment and allowing it to be acknowledged as a basic resource in the evolution of businesses that use that information and provide it with added value.

Data journalism is rising as a new way of doing journalism, considering it not merely as a methodology that is confined to research and precision journalism (Chaparro-Domínguez, 2014; Crucianelli, 2013a; Flores-Vivar, 2012; Gray et al., 2012), but essentially as an opportunity to respond to the demand for information by creating information products using data.

These new needs require a reformulation of university qualifications so that they are capable not only of rising to current informative needs, but of adapting to those of the future.

2. Materials and methods

This study intends to assess the professional competencies needed to develop interactive multimedia information products based on open data, bearing in mind that only if data can provide added value information for citizens will such openness lead to citizen empowerment, easier decision-making, enhanced democratic action and greater social change.

The methodology, which is qualitative in nature, involves gathering information by conducting a survey of subjects who are experts in data journalism or are linked to higher education in the field, concerning regulations on access to information and the media, followed by an analysis of a sample of multimedia information products based on open data.

Considering the initial hypothesis that producing interactive multimedia information products based on open data requires a set of specific conceptual, procedural and attitudinal competencies for gathering, processing, analysing and presenting information, the aim is for the experts to reveal the needs of the profession. Regarding our product analysis, the goal is to draw a direct link between their development requirements and the skills needed for that development, as revealed by the experts.

Data have now taken on a leading role based on what they can convey and signify in financial, democratic and social terms. But more data does not necessarily mean more knowledge, democracy or development, nor does it generate empowerment or social change in itself.

In the analysis, use has been made of the procedures of the Grounded Theory in order to define a series of categories thanks to the expert statements and the product analysis, with which we will identify and characterise the conceptual, procedural and attitudinal skills linked to developing multimedia information products based on open data (Strauss & Corbin, 2002).

2.1. Expert survey

A convenience sample was chosen based on prior documentary research. To select the subjects, we analysed the most relevant benchmarks relating to data journalism in Spain (qualifications, research and dissemination) up until February 2013 with a view to detecting who was in charge of such initiatives and who else took part or collaborated. The goal was to define a population of expert subjects in the field of study: Coalición Pro Acceso (2006), Asociación Pro Bono Público (2009), Irekia. Open Government (2009), Spanish translation of Council of Europe Convention on Access to Official Documents (2009), Conference: Data journalism (MediaLab Prado, 2011), Working Group on Data journalism (MediaLab Prado, 2011), Fundación Civio (2012), Estándares de Gobierno Abierto.
In order to avoid any kind of professional bias, other subjects were included who were linked to higher education in journalism, regulations on access to information and the media.

The sample comprised 19 subjects who are expert professionals in the fields of data journalism, research journalism, information display, access to public information, and/or university professionals lecturing in journalism degrees at public and private universities in the Community of Madrid.

To gather the information, the survey used was an on-line, self-administered and standardised questionnaire of our own making that combined multiple-choice with open questions. The survey was taken by the experts in March 2013.

2.2. Analysing multimedia information products based on open data

The methodology used to analyse multimedia information products based on open data starts with the finished product and works backwards over the tasks needed to develop them (Freixa, Soler-Adillon, Sora, & Ribas, 2014). Therefore, the analysis focuses on identifying the preliminary tasks involved in gathering, processing and analysing the data. To do so, account was taken of the functional and non-functional aspects involved, but especially of the information content presented.

In view of the complexity and constant evolution of the concept, we used a convenience sample comprised of sixteen significant products owing to their relevance in developing the phenomenon, their originality and their social and media impact at a national and international level. These sixteen products were published between the year 2010 and 2015; seven of them were produced by Spanish media, while the remaining nine were taken from media based outside Spain.

To gather the information, a worksheet of our own design was created that would allow us to: a) identify the product (URL, date of publication, temporary coverage, media/author, country and language); b) classify the product (field, topic or type of media responsible), according to Crucianelli’s classification (2013a); c) specify the type of data used (source, nature of the data, whether or not they are open, depending on their availability and ease of access, re-use and dissemination, and their universal participation) (Dietrich & al., s. f.); and d) describe the tasks of gathering, processing, analysing and presenting information, as required when developing an information product of this kind.

3. Analysis and results

3.1. Needs and shortcomings in professional competencies

In line with the analysis method set out in the Grounded Theory (Strauss & Corbin, 2002), we refrained from establishing categories initially; instead, the categories arose while analysing and coding the primary documents.

Fifty-seven codes were identified, which gave rise to seven core categories, three of which were linked to the type of competency –knowledge, ability and skill, or attitudes and values– and four were associated with the tasks of gathering, processing, analysing and presenting data.

After a qualitative analysis of the content and having drawn links between categories and core categories, we went on to study the co-occurrence rates between codes.

Lastly, the diagramming technique devised by Strauss and Corbin (2002) was used. The aim of this technique is not so much to show all of the concepts that have arisen from the coding phase, but rather to represent how they are positioned with regard to the

This requires a new form of free-code journalism (Sampedro, 2014) to pick up the baton handed to it by a society that needs its services as a mediator in the face of a complex reality, and as a guarantor of control of power. In short, this service is needed to empower civil society and drive social change.

© ISSN: 1134-3478 • e-ISSN: 1988-3293 • Pages 39-47
core categories. The rest of the categories are represented according to their proximity to the core category depending on the degree of grounding, i.e. the number of references linked to that category (Gertrudis, Gértrudix & Álvarez, 2015a; 2015b; 2015c).

With regard to the open questions in the survey, the process of coding and drawing out core categories has enabled us to identify the most relevant competencies for developing interactive information products based on open data, according to their proximity to the core category.

In the professional task of developing products based on the use of open data for information purposes, it is essential to command statistical and methodological abilities, as well as the knowledge and skills to process and analyse data, and to display and verify them. A change of attitude towards working with data is also necessary, so as to go from rejection to professional demand, and towards industry and public powers, for which it is necessary to be able and willing to adapt to professional changes. The key to this transformation lies in ensuring transparency, to boost the credibility of the journalist’s work, which is hard to achieve without high levels of independence.

The results shape a map of the necessary knowledge, skills and abilities, values and attitudes, and highlight the major shortcomings that qualified journalists currently have in this regard, which, in the eyes of these experts, are extremely important.

With regard to knowledge (Gertrudis & al., 2015a), the most significant areas include statistical and methodological knowledge, knowledge of data processing and information display, which are also the areas where qualified journalists currently fall short. Other more technical aspects can also be added, such as knowledge of programming and databases, as well as general knowledge.

The most well-grounded abilities and skills are those linked to data processing and analysing, as well as displaying information and practical training. However, besides these distinctly technical assets, it is important to highlight others that involve applying knowledge linked to more theoretical training and what is considered disciplinary knowledge, as well as qualities linked to other fields of knowledge such as linguistic expression and the ability to contextualise information (Gertrudis & al., 2015b).

In terms of attitudes and values (Gertrudis & al., 2015c), journalistic work using open data requires autonomy and, above all, critical thinking, which must first be applied to the data itself and then to the sources and to the information arising therefrom.

3.2. Information products based on open data: Gathering, processing, analysing and presenting information

3.2.1. Characteristics of information products based on open data

Several of the products analysed relate to politics and social issues, and most of them (10) were presented in accordance with the Crucianelli classification (2013a) as independent interactive displays. The remaining products include one or several short articles that illustrate the phenomenon. Cartographic representations have proven to be the preferred method of graphic rendering, either because the use of «mashups» based on geolocation systems such as Google Maps is very popular or because this technique is able to combine the abstract component of data rendering with a more familiar element, namely representing them geographically on a map using proportional symbols in different sizes and colours, usually circles or bubbles. Other forms of graphic representation have also been used, such as bars, areas, lines and columns depending on the data shown.

In terms of developers, nine of the products were developed by teams that are part of the digital editorial departments of the so-called traditional media.

In all cases, the data came from secondary sources, mostly of a public nature, where they were generated as part of their usual business. Regarding private sources («Tell all phones»), the data are of a private nature, and in the case of «The top 100 papers», the database is held by Thomson Reuters. With the exception of any data obtained from leaks, most are generally available, though access to some is restricted.

Considering their availability, accessibility, re-use, dissemination and universal participation (Dietrich & al., s. f.), just 25% of the products analysed can be said to have been developed using open data.

3.2.2. Professional competencies linked to product development

The worksheets used to gather information on the products were added to the corresponding hermeneutics unit in the ATLAS.ti program for the purpose of analysing them according to the categories established by the experts. Before this, in order to link the competencies analysed to each of the stages of the development process (Bradshaw, 2011; Crucianelli, 2013a), four core categories were created: gathering, processing, analysing and presenting, and it was found that most of the competencies are not linked to just one stage or specific task, but rather they are essential requirements in the development of the entire process.
Aside from the particularities of each case, the analysis revealed a very well-defined methodology and certain needs, both in terms of competencies and technology, that were identified as the basic requirements for its development: the capacity to detect a news-worthy event; the ability and capacity to access sources; the ability and capacity to search for and retrieve information; knowledge of specialised sources; knowledge of current regulations; knowledge of how public administration works; the ability to access information; knowledge of databases; data processing; data conversion; format conversion for re-use; data analysis; the ability and capacity to filter out relevant information; interpreting information; statistical knowledge; the capacity to make data understandable; the ability and capacity to contextualise information; generating new informative content; the capacity to present information on different media and in different formats; the ability and tools to display information; the capacity to generate added value and to verify information (Gertrudis-Casado, Gétrudix, & Álvarez, 2015d).

3.3. Main results

Faced with these challenges, according to the experts, qualified journalists still show a certain reluctance to base their work on open data, firstly because they are members of a society that lacks the deep-rooted cultural custom of doing so, and secondly because it is not common practice to verify information and adapt to change, especially when it comes to using technology.

There is reason to believe that the cause of these shortcomings is a discrepancy between the curricula followed at university and the real needs of the profession on the whole and of data journalism in particular. The main reasons for this discrepancy are outdated teaching content and insufficient practical training.

Regarding the specific tasks needed to develop information products based on open data, it has been found that most are not based on open data, but rather they require specific processing in order to re-use data for informative purposes. With this, we are able to discern a series of tasks that recur with varying degrees of complexity in all of the products analysed, as well as allowing us to define a consistent methodology that reveals the need for specific professional competencies with which to carry out such tasks.

By contrasting the categories defined as a result of the product analysis, we see that the competencies pointed out by the experts as the skills needed to develop the tasks of gathering, processing, analysing and presenting information based on data are not exclusive to a given stage in the process, but rather they are extremely relevant throughout. It is essential to have a knowledge of specialised sources and, when dealing with public data, it is vital to be aware of current regulations and of how public administration works. Accessing sources of information requires skills that are not only linked to searching and selecting, but also detecting a news-worthy event.

Given that information is often not available in reusable formats, it is important to have some knowledge of computer programming and data conversion, and in order to manage data it will be necessary to know about databases and have some data management skills.

The entire process will require the ability to filter through and select relevant information in each of the data gathering, processing, analysing and presenting stages. The processing stage will often involve skills such as a command of spreadsheet software to clean up, filter and sort data. During the analysis stage, knowing how to select important information is key in order to tell a story based on data that is capable of reaching civil society.

Regarding the tasks involved in presenting information, they must be geared towards making the data understandable, which requires the ability to contextualise information and present it to civil society as a value added product. This is generally achieved by means of interactive displays and short articles that are supported by the principle of verification, which comes not only from making data available to citizens in reusable formats, but also from observing transparency by clearly and unreservedly explaining the method used.

4. Discussion and conclusions

The unfathomable reality that Lippmann referred to in 1922 (2003) still represents today a context of superabundance of public and private data. These are data which, in the best possible scenario and where the legal context allows, may be open to civil society. However, the fact of being openly accessible, despite being a desirable requirement, does not in itself generate knowledge or empower citizens, which means that transformed journalistic mediation has become even more important now than ever before, as pointed out by Baack (2015).

The main power wielded by the media is, therefore, the ability to construct these information products, to the extent that in doing so they shine the spotlight on part of reality (McCombs, 2006). The rest of that reality is occasionally cast into oblivion (Noelle-
Neumann, 1995), sometimes because the immediacy of news ostracises any other part of reality that requires a process of research which can take months or even years and which relies on the support of techniques from other disciplines. Such is the case of research journalism in general and of precision or data journalism in particular.

Experts in data journalism have long been demanding more specific training for journalistic professionals. This training must be capable of rising to the needs of a methodology that is well-defined (Bradshaw, 2011) and multi-disciplinary (Cairo, 2012; Crucianelli, 2013a; Flores-Vivar, 2012; Zanchelli & Crucianelli, 2013). However, they also demand a new vision of journalism qualifications that is open to methodologies from other sciences, such as statistics, interaction design or computing. This was already pointed out by prior studies such as those conducted by De-Maeyer & al. (2015) and Nguyen and Lugo-Ocando (2015).

Transparency in terms of data and methodology is a requirement under the principle of verification upheld by Kovach and Rosenstiel (2012). However, it is also the element that establishes the openness and re-usability of data and information in favour of a «Knowledge Society», where beyond mere verification, the consequence is providing users and citizens with the means to replicate, check, discuss or generate new knowledge, in a global society where the data culture is emerging (Álvarez-García, Gértrudix-Barrio, & Rajas-Fernández, 2014), driven by an increasingly active civil society.

We must not lose sight of the fact that technology also alters information representation models, and that the way in which information is generated proves equally important as the way it is represented (Bradshaw, 2011), which leads to new reading mechanisms. In the case of information based on open data, it is common practice to represent it in interactive multimedia displays (Crucianelli, 2013a), which are designed so as to dynamically and actively transmit to citizens the information they need, thus favouring open paths for reading and analysis that will drive empowerment in decision making. This requires a new form of free-code journalism (Sampedro, 2014) to pick up the baton handed to it by a society that needs its services as a mediator in the face of a complex reality, and as a guarantor of control of power. In short, this service is needed to empower civil society and drive social change.

Notes

1 The sample of experts is comprised of: David Cabo Calderón, Alberto Cairo, Sonia Castro, José Cervera García, Javier Davara Torrego, Roberto de Miguel Pascual, Nagore de los Ríos, Roberto Gamonal Arroyo, José María García de Madariaga, Marcos García Rey, Guzmán Garmendia Pérez, Max Römer, Gloria Rosique, Antonio Rubio Campaña, Ricardo Ruiz de la Serna, José Antonio Ruiz San Román, Juan Carlos Sánchez, Manuel Sánchez de Diego and Milena Trenta.

2 The products analysed are: ¿A dónde va el dinero contra la pobreza? (Where does money against poverty go?), Afghanistan war logs: IED attacks on civilians, coalition and Afghan troops, ¿Dónde van mis impuestos? (Where do my taxes go?), CIPPEC data, Elecciones: los de 18 a 25 años, ¿estáis ahí? (Elections: 18 to 25 year-olds, are you there?), España en llamas (Spain in flames), Gay rights in the US, state by state, La mujer en el mundo (Women in the world), Out of Sight, Out of Mind: A visualization of drone strikes in Pakistan since 2004, Patrimonio de los diputados (The estate of MPs), Tell-all telephone, Todos los papeles de Bárcenas (All of Bárcenas’ books), The top 100 papers. El tormentoso ejercicio del periodismo en...
Colombia (The troublesome practice of journalism in Colombia), and Transparency for the E.U., U.S. Gun Deaths in 2013.

Support
This study is financed by the project Digital Citizenship and Open Data Access: citizen empowerment through social media in the digital environment (CSO2012-30756, Spanish Ministry of the Economy and Competitiveness).

References
Agudelo, I. (2014). From Infocoxication to the Right to Communicate. Comunicar, 21(42), 07-08. doi: http://dx.doi.org/10.3916/C42-20-14-a1