

REVISTA

cartográfica

Número 97 • Julio a Diciembre • 2018



Instituto Panamericano de
Geografía e História

**AUTORIDADES DEL
INSTITUTO PANAMERICANO DE GEOGRAFÍA E HISTORIA
2017-2021**

| | | |
|-----------------------|---------------------|--------|
| PRESIDENTE | Lic. Israel Sánchez | Panamá |
| VICEPRESIDENTE | Mag. Alejandra Coll | Chile |

SECRETARIO GENERAL

Mag. César Rodríguez Tomeo
Uruguay

COMISIÓN DE CARTOGRAFÍA

(Costa Rica)

Presidente:

Mag. Max Lobo

Vicepresidente:

Mag. Álvaro Antonio Álvarez

COMISIÓN DE HISTORIA

(México)

Presidenta:

Dra. Patricia Galeana Herrera

Vicepresidente:

Mtro. Rubén Ruiz

COMISIÓN DE GEOGRAFÍA

(Estados Unidos de América)

Presidenta:

Dra. Patricia Solís

Vicepresidenta:

Geóg. Jean W. Parcher

COMISIÓN DE GEOFÍSICA

(Ecuador)

Presidente:

Dr. Mario Ruiz

Vicepresidente:

Dra. Alexandra Alvarado

MIEMBROS NACIONALES DE LA COMISIÓN DE CARTOGRAFÍA

Argentina

Belice

Bolivia

Brasil

Chile

Colombia

Costa Rica

Ecuador

El Salvador

Estados Unidos de América Guatemala

Guatemala

Haití

Honduras

México

Nicaragua

Panamá

Paraguay

Perú

República Dominicana

Uruguay

Venezuela

Agrimensor Sergio Rubén Cimbaro Sr.

Wilbert Vallejos

Cnl. DAEN José Oviedo Bustillos

João Bosco de Azevedo

Dr. Carlos Mena Frau

Fernando León Rivera

Sr. Xavier Molina

Sr. Mario Antonio Zeledón Flores

Eric Van Praag

Ing. Marcos Osmundo Sutuc Carillo Ing.

Arch. Pierre Alexilien Versaille Arq.

Marilyn Villatoro

Ing. Raúl Ángel Gómez Moreno

Lic. Israel Sánchez

Cnel. SG Antonio Gavilán Estigarribia

Mayor EP Reynaldo Flores Rivero

Geóg. Susana Hernández

Cnel. Daniel Píriz

Ing. Sergio Rodríguez

COMITÉS DE LA COMISIÓN DE CARTOGRAFÍA

Normas y estándares de Información Geográfica (Colombia)

Cartografía Temática

Infraestructura de Datos Geospaciales

(Ecuador)

INSTITUTO PANAMERICANO DE GEOGRAFÍA E HISTORIA

REVISTA **cartográfica**



NÚMERO 97

JULIO-DICIEMBRE 2018

REVISTA CARTOGRÁFICA

Publicación anual fundada en 1952
Indizada en PERIÓDICA y Latindex
Disponible en: Cengage Learning, Ebsco y ProQuest

La preparación de la REVISTA CARTOGRÁFICA
está a cargo de la editora María Ester Gonzalez
Universidad de Concepción
Juan Antornio Coloma 0201, CP 4440000
Los Ángeles, Chile
Teléfono: +56 432405244
Correo electrónico: editor_revista_cartografica@ipgh.org.

Editor invitado: *Luis M. Vilches-Blázquez*
Centro de Investigación en Computación
Instituto Politécnico Nacional (IPN), México

Canje, venta y distribución de publicaciones, escribir a:
Instituto Panamericano de Geografía e Historia
Secretaría General
Apartado Postal 18879, C.P. 11870 Ciudad de México, México
Teléfonos: (52-55)5277-5888, 5277-5791 y 5515-1910
Correo electrónico: publicaciones@ipgh.org <http://www.ipgh.org>

Las opiniones expresadas en notas, informaciones, reseñas y trabajos publicados en la REVISTA CARTOGRÁFICA, son de la exclusiva responsabilidad de sus respectivos autores. Los originales que aparecen sin firma ni indicación de procedencia son de la Dirección de la Revista.

En cumplimiento con la resolución IX de la XIV Reunión del Consejo Directivo del IPGH, celebrada en julio de 1972, en Buenos Aires, Argentina, se advierte que: "Los límites que aparecen en los mapas de esta publicación no están, en algunos casos, finalmente determinados y su reproducción no significa aprobación oficial o aceptación por el Instituto Panamericano de Geografía e Historia (IPGH)".

Diseño de portada / Cover design / Design da capa: Mônica Pimentel Cinelli Ribeiro, Instituto Brasileño de Geografía y Estadística (IBGE).

Fuente de la imagen / Source / Fonte da Imagem: Composición a partir de imagen disponible en <<https://pixabay.com/en/network-earth-block-chain-globe-3537401/>>.

© 2018 Instituto Panamericano de Geografía e Historia.

Revista Cartográfica, núm. 97, julio-diciembre 2018, es una publicación semestral editada por el Instituto Panamericano de Geografía e Historia, Ex-arzobispado núm. 29, Col. Observatorio, Delegación Miguel Hidalgo, C.P. 11860, Ciudad de México, México. Tels. (52-55)5277-5888, 5277-5791, 5515-1910, www.ipgh.org, publicaciones@ipgh.org. Editor invitado: Luis M. Vilches-Blázquez, correo electrónico: lmvilches@cic.ipn.mx. Reserva de Derechos al Uso Exclusivo núm. 04-2015-090212390600-203, ISSN 0080-2085 ambos otorgados por el Instituto Nacional del Derecho de Autor. Licitud de título y contenido: en trámite. Responsable de la última actualización de este número: Departamento de Publicaciones del IPGH, Ex arzobispado núm. 29, Col. Observatorio, Delegación Miguel Hidalgo, C.P. 11860, Ciudad de México, México. Fecha de última modificación: 13 de diciembre de 2018.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor de la publicación.

Queda estrictamente prohibida la reproducción total o parcial de los contenidos e imágenes de la publicación sin previa autorización del Instituto Panamericano de Geografía e Historia.

REVISTA **cartográfica**

NÚMERO 97

JULIO-DICIEMBRE 2018

Contenido

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Nota Editorial | 7 |
| Artículos científicos | |
| Análise exploratória da pegada digital dos turistas para a identificação de padrões espaciais e temporais em destinos urbanos <i>Luis Encalada</i> <i>Jorge Rocha</i> <i>Carlos Ferreira</i> | 17 |
| Creación de un modelo semántico para modelar los fenómenos meteorológicos de los siglos XIX y XX en Latinoamérica <i>Diana Comesaña</i> <i>Bryan Barreiro</i> | 37 |
| La calidad de las IDE desde el punto de vista de la interoperabilidad <i>Antonio F. Rodríguez</i> <i>Francisco Javier Ariza</i> | 53 |
| Cuantificación de vulnerabilidad y riesgo: las inundaciones en Motozintla de Mendoza, Chiapas, México <i>Mary Frances Rodríguez Van Gort</i> | 81 |
| Calidad en datos geográficos, geoservicios y productores de datos: análisis crítico <i>Francisco Javier Ariza</i> <i>Antonio F. Rodríguez</i> | 143 |

Reseñas

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Reproducibilidad científica: ¿Qué es y por qué debemos interesarnos en geo-ciencias? <i>Daniela Ballari</i> | 147 |
| Datos enlazados de información geográfica del Ecuador <i>Víctor Saquicela</i> <i>Fernando Baculima</i> <i>Lucia Lupercio</i> | 157 |
| Caracterización geográfica de la vivienda en México: un enfoque de Ciencia de Datos <i>Jacobo Gerardo González León</i> <i>Miguel Félix Mata Rivera</i> | 165 |
| Instructivo para autores | 177 |

Comité Editorial
Revista Cartográfica número 97

Alejandra Sánchez Maganto
Instituto Geográfico Nacional (España)

Amalia Velasco
Dirección General del Catastro (España)

Antonio Vázquez Hoehne
Universidad Politécnica de Madrid (España)

Carlos López-Vázquez
Universidad ORT Uruguay (Uruguay)

Daniel Orellana
Universidad de Cuenca (Ecuador)

Daniela Ballari
Universidad del Azuay (Ecuador)

Elena Chicaiza Mora
IPGH Sección Nacional de Ecuador
(Ecuador)

Francisco Javier López Pellicer
Universidad de Zaragoza (España)

Frida Güiza
Universidad Nacional Autónoma de México
(México)

Gregorio Pascual Santamaría
Protección Civil (España)

Israel Quintanilla García
Universidad Politécnica de Valencia
(España)

Jorge Mauricio Espinoza Mejía
Universidad de Cuenca (Ecuador)

José Giovanni Guzmán Lugo
Instituto Politécnico Nacional (México)

José Juan Arranz Justel
Universidad Politécnica de Madrid (España)

José Miguel Olivares García
Dirección General del Catastro (España)

Luis M. Vilches-Blázquez
Instituto Politécnico Nacional (México)

María Alejandra Barrera
Universidad Nacional de Catamarca
(Argentina)

María Ester Gonzalez
Universidad de Concepción (Chile)

María Victoria Álvarez Gamio
Intendencia de Montevideo
(Uruguay)

Miguel Ángel Bernabé Poveda
Universidad Politécnica de Madrid
(España)

Miguel Torres Ruiz
Instituto Politécnico Nacional (México)

Pamela Mayorga Ramos
Unidad Administrativa Especial de Catastro
Distrital (Colombia)

Paulo César Coronado Sánchez
Universidad Distrital Francisco José de
Caldas (Colombia)

Rodrigo Barriga Vargas
Universidad Bernardo O'Higgins (Chile)

Rolando Quintero
Instituto Politécnico Nacional (México)

Sergio Acosta y Lara
Ministerio de Transporte y Obras Públicas
(Uruguay)

Villie Morocho
Universidad de Cuenca (Ecuador)

Nota Editorial

El número 97 de la *Revista Cartográfica* del Instituto Panamericano de Geografía e Historia (IPGH) dedicado a los *avances, tendencias y perspectivas de la información geográfica* se enmarca en una revolución digital, que se está produciendo en nuestra sociedad desde mediados del siglo pasado, denominada Revolución Industrial 4.0 (Schwab, 2016). Esta revolución cambiará la forma en que vivimos, trabajamos y nos relacionamos entre nosotros. En este contexto se están produciendo rápidos e importantes cambios en nuestro mundo, producto de los avances en la economía compartida, *Internet of Things*, comunidades inteligentes y conectadas, y servicios personalizados basados en la ubicación (Yuan, 2017).

Este escenario actual muestra como elemento común el incremento del volumen de datos producidos en todo el mundo, producto de la cantidad de dispositivos conectados a Internet que, además, se espera crezcan exponencialmente cada año (Marr, 2016). Este hecho constituye un tema relevante para la comunidad geográfica, ya que diversos estudios han demostrado que aproximadamente el 80% de todos los datos disponibles están relacionados con una ubicación espacial (Li y Li, 2014). Esta afirmación se corrobora hoy en día por el hecho de que la Web 2.0, los dispositivos móviles, la participación ciudadana y el *crowd sourcing* se han convertido en un flujo relevante en la recopilación de datos (Salk, Sturn, See, Fritz y Perger, 2016). Estas nuevas corrientes se presentan como esfuerzos significativos y complementarios a los realizados por las organizaciones tradicionales y oficiales dedicadas a la producción de información geoespacial. Del mismo modo, los datos de las redes de sensores, los GPS en dispositivos móviles, las redes sociales (por ejemplo, Twitter) y los datos de información geográfica voluntaria (por ejemplo, OpenStreetMap) también contribuyen al aumento de la cantidad de este tipo de datos (Guo, Liu, Jiang, Wang, Liu y Liang, 2017) y a poner de manifiesto que la utilización de la información geográfica crece con rapidez (Norris, 2015).

Este nuevo marco propicia un amplio abanico de retos y oportunidades para los investigadores relacionados con el área de *GIScience* (*Geographical Information Science*), que tienen una larga tradición de reflexión, de pensar hacia el futuro y *outside the box*, así como de vislumbrar las implicaciones más allá de los desarrollos comerciales (Goodchild, 2018). Por tanto, la comunidad *GIScience* tiene un papel importante que desempeñar en la conformación de los *avances, tendencias y perspectivas de la información geográfica* y en el proceso de reflexión sobre los impactos de estos desarrollos.

Los avances, tendencias y perspectivas que están propiciando las tecnologías geoespaciales representan una herramienta invaluable, debido a la capacidad de integrar, fusionar y visualizar múltiples datos de diversas fuentes, permitiendo me-

jorar los procesos de comparación y análisis de los aspectos territoriales a diferentes escalas. En este contexto, bajo el paraguas de la Iniciativa de las Naciones Unidas sobre la Gestión Global de la Información Geoespacial (UN-GGIM, por sus siglas en inglés), un grupo de expertos reconocidos de muy diversos campos relacionados con el mundo geoespacial, junto con valiosas contribuciones de autoridades nacionales de cartografía y catastrales de diferentes países, ofrecieron una visión de cómo podría desarrollarse las futuras tendencias de la información geográfica en un escenario de cinco a diez años (Norris, 2015).

Transcurrido un tiempo desde la presentación de esta propuesta ante la comunidad internacional, este número de la *Revista Cartográfica* pone atención a las propuestas que presentan nuevos acercamientos del uso de las tecnologías geoespaciales con un enfoque especial en los Objetivos de Desarrollo Sostenible propuestos por la Agenda 2030 de Naciones Unidas, ya que el actual volumen de datos y los avances de la Revolución Industrial 4.0 pueden generar sinergias que permitan afrontar los retos que encara nuestra Región relacionados con la pobreza, cambio climático, educación, defensa del medio ambiente o el diseño de nuestras ciudades.

Este número especial se conforma de artículos relacionados con *Big Data*, Web Semántica, ciencia de datos, Infraestructuras de Datos Espaciales (IDE) y el tratamiento de la calidad a diferentes niveles. Además, se recogen una serie de reseñas, con experiencias y reflexiones en el contexto de los *avances, tendencias y perspectivas de la información geográfica*, relacionadas con la reproductibilidad científica, *Linked Data* y ciencia de datos.

Luis M. Vilches-Blázquez

Editor invitado

Editorial

The special issue 97 of the Pan American Institute of Geography and History's Cartographic Journal on *advances, trends, and perspectives of geographic information* is related to a digital revolution, which is being developed in our society since the middle of the last century, named the Industrial Revolution 4.0 (Schwab, 2016). This revolution will alter the way we live, work, and relate to each other. In this context, our world is changing rapidly with the advances in the shared economy, Internet of Things, smart and connected communities, and personalized location-based services (Yuan, 2017).

This current scenario presents a common element, namely, a worldwide data volume growth since the number of hands held devices and Internet-connected equipment, which are expected to grow exponentially every year (Marr, 2016). This fact constitutes a relevant issue for the geoscience community because diverse studies have shown that about 80% of all available data are related to a spatial location (Li y Li, 2014). This statement is nowadays corroborated by the fact that Web 2.0, mobile devices, citizen participation, and crowdsourcing have become a relevant stream in the collection of geospatial data (Salk, Sturn, See, Fritz y Perger, 2016). These new streams are significant and complementary efforts to the traditional and official organizations associated with geospatial information production. Likewise, on-site sensor networks data, GPS trace data from mobile devices, social media data (i.e. Twitter), and crowdsourcing/volunteered geographic information data (e.g. OpenStreetMap) also contribute to the increase of the amount of this data type (Guo, Liu, Jiang, Wang, Liu y Liang, 2017) and to highlight that geographical information is growing at high speeds (Norris, 2015).

This new context promotes the emergence of new challenges and opportunities for researchers related to GIScience (Geographical Information Science) area, which have a long-established tradition of reflection, of thinking ahead and “outside the box” and of thinking about the broader implications of commercially motivated developments (Goodchild, 2018). Therefore, the GIScience community has a key role to perform *advances, trends, and perspectives of geographic information* and thinks about the impacts of these developments.

The advances, trends, and perspectives encouraged by geospatial technologies are an invaluable tool, due to the ability to integrate, join and visualize different data from lots of sources, which allow improving the benchmark and analysis processes related to territorial issues at different scales. In this context, under the protection of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), a recognized group of experts from a wide range of fields related to the geospatial world, together with valuable contributions

from the National Mapping and Cadastral Authorities (NMCAs), attempts to offer some vision of how the future trends of geospatial information are likely to develop over the next five to ten years (Norris, 2015).

After some time from the presentation of this proposal to the international community, this special issue of the Cartographic Journal pays special attention to works that present new approaches related to the use of geospatial technologies with a special emphasis in the Sustainable Development Goals of the United Nations 2030 Agenda. Thus, this special issue collects efforts and synergies related to the actual data volume and the Industrial Revolution 4.0 advances, which may allow addressing the challenges of our Region associated with poverty, climate change, education, environmental protection or the design of our cities.

This special issue is composed of several papers related to Big Data, Semantic Web, data science, Spatial Data Infrastructures (SDI) and quality treatment at different levels. Moreover, different reviews are collected in this special issue, which present experiences and reflections in the context of advances, trends, and perspectives of geographical information related to scientific reproducibility, Linked Data, and data science.

Luis M. Vilches-Blázquez

Guest editor

Editorial

O número 97 da *Revista Cartográfica* do Instituto Pan-americano de Geografia e História (IPGH), dedicado aos avanços, tendências e perspectivas da informação geográfica, se moldura em uma revolução digital, que se está produzindo em nossa sociedade desde meados do século passado, denominada de Revolução Industrial 4.0 (Schwab, 2016). Esta revolução mudará a forma em que vivemos, trabalhamos e nos relacionamos. Neste contexto, se estão produzindo rápidas e importantes mudanças em nosso mundo, produto dos avanços na economia compartilhada, *Internet of Things*, comunidades inteligentes e conectadas e serviços personalizados baseados em localização (Yuan, 2017)

Este cenário atual mostra como elemento comum o incremento do volume de dados produzidos em todo o mundo, resultante da quantidade de dispositivos conectados com a Internet que, além disso, espera-se que cresça exponencialmente a cada ano (Marr, 2016). Este feito constitui um relevante tema para a comunidade geográfica, já que diversos estudos têm demonstrado que aproximadamente 80% de todos os dados disponíveis estão relacionados com a localização espacial (Li e Li, 2014). Hoje em dia, essa afirmação é corroborada pelo feito de que a Web 2.0, os dispositivos móveis, a participação dos cidadãos e o *crowd sourcing* se converteram em um fluxo relevante na recompilação dos dados (Salk, Sturn, See, Fritz y Perger, 2016). Estas novas correntes se apresentam como esforços significativos e complementares das organizações tradicionais e oficiais dedicadas à produção de informação geoespacial. Deste mesmo modo, os dados das redes de sensores, os dados GNSS dos dispositivos móveis, os dados das redes sociais (por exemplo, Twitter) e os dados de mapeamento colaborativo (por exemplo, OpenStreetMap) também contribuem para o aumento da quantidade deste tipo de dado (Guo, Liu, Jiang, Wang, Liu e Liang, 2017) e mostra que o uso da informação geográfica cresce com rapidez (Norris, 2015).

Este novo marco propicia uma ampla gama de desafios e oportunidades para os pesquisadores relacionados com a área de *GIScience* (*Geographical Information Science*), que tem uma grande tradição de reflexão, de pensar em direção ao futuro e “*outside the box*”, assim como vislumbrar as implicações para além dos desenvolvimentos comerciais (Goodchild, 2018). Portanto, a comunidade *GIScience* tem que desempenhar um papel importante na conformação dos avanços, tendências e perspectivas da informação geográfica e no processo de reflexão sobre os impactos desses desenvolvimentos.

Os avanços, tendências e perspectivas que estão proporcionando as tecnologias geoespaciais representam uma ferramenta inestimável, devido à capacidade de integrar, fundir e visualizar múltiplos dados de diferentes fontes, permitindo melho-

rar os processos de comparação e análises dos aspectos territoriais para diferentes escalas. Nesse contexto, a partir da Iniciativa das Nações Unidas sobre a Gestão Global da Informação Geoespacial (UN-GGIM, sigla em inglês), um grupo de especialistas reconhecidos de vários campos relacionados com o mundo geoespacial, junto com contribuições valiosas de autoridades nacionais de cartografia e cadastro de diferentes países, ofereceram uma visão de como poderia desenvolver as futuras tendências da informação geográfica em um cenário de cinco a dez anos (Norris, 2015).

Transcorrido um período desde a apresentação dessa proposta à comunidade internacional, esse número da *Revista Cartográfica* se concentra nas propostas que apresentam novos detalhes do uso das tecnologias geoespaciais, com um enfoque especial nos Objetivos de Desenvolvimento Sustentável propostos pela Agenda 2030 das Nações Unidas, já que o atual volume de dados e os avanços da Revolução Industrial 4.0 podem gerar sinergias que permitam enfrentar os desafios encarados em nossa Região no que tange a pobreza, mudanças climáticas, educação, defesa do meio ambiente e o planejamento de nossas cidades.

Este número especial é composto de artigos relacionados com Big Data, Web Semântica, ciência de dados, Infraestrutura de Dados Espaciais (IDE) e o tratamento da qualidade em diferentes níveis. Além disso, uma série de revisões são coletadas, com experiências e reflexões no contexto dos *avanços, tendências e perspectivas da informação geográfica*, relacionadas com a reprodutibilidade científica, *Linked Data* e ciência de dados.

Luis M. Vilches-Blázquez
Editor convidado

Bibliografia

- Goodchild, M.F. (2018). “GIScience for a driverless age”, *International Journal of Geographical Information Science*, 32:5, 849-855.
DOI: 10.1080/13658816.2018.1440397
- Guo, H.; Liu, Z.; Jiang, H.; Wang, C.; Liu, J. and Liang, D. (2017). “Big Earth Data: a new challenge and opportunity for digital Earths development”, *International Journal of Digital Earth*, 10:1, 1-12.
- Li Q. and Li D. (2014). “Big Data GIS”, *Geomatics and Information Science of Wuhan University*, 39:6, 641.
- Marr, B. (2016). “17 Predictions About The Future Of Big Data Everyone Should Read”, *Forbes*. Recuperado de

- <<https://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#44cd182c1a32>>.
- Norris, J. (2015). *Future Trends in geospatial information management: the five to ten year vision*, United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM), Second Edition, December.
- Salk, C.F.; Sturn, T.; See, L.; Fritz, S. and Perger, C. (2016). "Assessing quality of volunteer crowdsourcing contributions: lessons from the cropland capture game", *International Journal of Digital Earth*, 9:4, 410-426.
- Schwab, K. (2016). *The Fourth Industrial Revolution: what it means, how to respond*. *World Economic Forum (WEF)*. Recuperado de <<https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>>.
- Yuan, M. (2017). "30 years of IJGIS: the changing landscape of geographical information science and the road ahead", *International Journal of Geographical Information Science*, 31:3, 425-434. DOI: 10.1080/13658816.2016.1236928.

ARTÍCULOS
científicos



Análise exploratória da pegada digital dos turistas para a identificação de padrões espaciais e temporais em destinos urbanos*

Luis Encalada**
Jorge Rocha***
Carlos Ferreira****

Recebido 08 de maio de 2018; aceite 26 de setembro de 2018

Resumo

No presente trabalho, utilizaram-se as fotos georreferenciadas dos visitantes da cidade de Lisboa, partilhadas na rede ‘Panoramio’ entre 2007 e 2014, como proxy para analisar a distribuição espacial e temporal dos mesmos dentro da cidade. O conjunto total de dados (> 75.000 fotografias) foi segmentado em visitantes e locais, com base nas marcas temporais, resultando numa amostra de 17.604 fotos tiradas por > 5.000 utilizadores. A evidência empírica sugere que a distribuição espacial dos visitantes não é homogénea. Complementarmente, avaliamos a relação espacial entre o padrão observado (a aglomeração geográfica de visitantes) e um conjunto de 24 variáveis associadas à oferta turística da cidade. Através da análise de regressão linear múltipla, verificou-se que são os ‘Monumentos de interesse público’, os elementos que apresentam maior atratividade e, apurou-se, considerando os fatores explicativos selecionados, que existem áreas do centro histórico, cujo potencial turístico está subestimado.

* Este artigo é apresentado a partir dos resultados da Dissertação de Mestrado, intitulada: “‘Big Data’ e Redes Sociais na Análise Espacial de Padrões Turísticos em Contexto Urbano. O caso de Lisboa”, defendida em 26 de setembro de 2014.

** Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Rua Branca Edmée Marques, 1600-276, Lisboa, Portugal, e-mail: luisencalada@campus.ul.pt

*** Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, e-mail: jorge.rocha@campus.ul.pt

**** Departamento de Geografia e Turismo, Faculdade de Letras, Universidade de Coimbra; Colégio de S. Jerónimo, 3004-530 Coimbra, Portugal, e-mail: carlos.ferreira@uc.pt

Palavras-chave: *Geotagged fotos, Turismo urbano, Redes sociais, Big Data, Análise espaciotemporal.*

Resumen

En este artículo se utilizaron las fotos georreferenciadas, compartidas por usuarios de la red “Panoramio” entre 2007 y 2014, como un *proxy* para analizar la distribución espacial y temporal de los visitantes, en la ciudad de Lisboa. El conjunto total de datos (> 75.000 fotografías) fue segmentado en turistas y locales, con base en las marcas temporales, resultando una muestra de 17.604 fotos tomadas por > 5.000 usuarios. La evidencia empírica sugiere que la distribución espacial de los visitantes no es homogénea. Además, exploramos la relación espacial entre el patrón observado (la aglomeración geográfica de visitantes) y un conjunto de 24 variables asociadas a la oferta turística de la ciudad. A través del análisis de regresión lineal múltiple, se verificó que son los “monumentos de interés público”, los elementos que presentan mayor atractivo, considerando los factores explicativos seleccionados, y que existen áreas del centro histórico cuyo potencial turístico está subestimado.

Palabras clave: *Fotos geotiquetadas, Turismo urbano, Redes sociales, Big Data, Análisis Espacio-temporal.*

Abstract

In this paper, we use online digital footprints (i.e., geotagged photos from ‘Panoramio’) of Lisbon’s visitors, between 2007 and 2014, as a sensitive proxy for exploring tourist spatiotemporal behavior within the city. We identify the visitors’ photos from the overall dataset (> 75,000 photos) by looking at photos’ timestamps, resulting in a sample of 17,604 photos from > 5,000 users. This empirical evidence suggests that the spatial distribution of visitors is not homogenous. Then, we assess the spatial relation between the observed pattern (geographical agglomeration of visitors) and an initial set of 24 variables regarding city tourism offer. Findings from the multiple regression analysis, show that ‘Monuments of public interest’ are the elements with higher attractiveness. Furthermore, some areas located within the historic center appear underestimated, despite their significant tourism appeal considering the selected variables.

Key words: *Geotagged photos, Urban Tourism, Social networks, Big data, Spatiotemporal analysis.*

Introdução

O entendimento de como os turistas usam o espaço nos ambientes urbanos tem aplicações práticas na administração e planeamento do turismo nos locais de destino

(Lau e McKercher, 2006), podendo aproveitar-se como contributo de suporte para a organização dos serviços essenciais ao funcionamento da atividade, a criação de novos produtos, a gestão de impactos nas suas diversas formas, entre outros fins (Hayllar *et al.*, 2008).

Segundo Ashworsth e Page (2011), apesar da importância quantitativa deste tipo de turismo, é curioso a pouca atenção que tem sido dada ao entendimento de como os turistas realmente usam as cidades. Tal como referido por vários autores (Kádár, 2014; Pearce, 2001), o comportamento espacial da procura (no contexto urbano) ainda não é totalmente compreendido devido à falta de dados com este nível (local) de discriminação, justificada pelas dificuldades na sua quantificação, *i.e.*, limitações na aplicação de métodos científicos para a monitorização de turistas.

A par do desenvolvimento das novas tecnologias de informação e de comunicação, têm vindo a surgir novas fontes de dados, menos tradicionais, que podem ser aproveitadas para o estudo do turismo urbano, particularmente, para a análise do comportamento espacial e temporal dos turistas. Os serviços online de partilha de fotos (*e.g.*, Flickr, Instagram), de mensagens instantâneas (*e.g.*, Twitter), de avaliação de consumidores (*e.g.*, Tripadvisor) e outras plataformas envolvidas num ambiente colaborativo de partilha de conteúdos, formam parte integrante destas novas fontes, que se caracterizam pelos altos níveis de participação e pela quantidade de informação gerada. Estas plataformas expandem o cenário da quantidade/diversidade de fontes disponíveis, e deixam de lado a condição em que a informação era exclusivamente produzida pelas autoridades oficiais e através de métodos científicos clássicos. Estas fontes não oficiais, enquadradas no campo das *social media*, têm promovido os próprios indivíduos como geradores de informação.

Os turistas, quando utilizam as redes sociais, deixam grandes quantidades de rastros digitais, resultando num conjunto multidimensional de dados conhecido como “Big Data” (Buhalis e Amaranggana, 2014). Assim, como referido por Buhalis e Law (2008) e Munar e Jacobsen (2014), denota-se uma tendência crescente no número de turistas que partilham as suas experiências através da publicação de recomendações, opiniões, fotos ou vídeos sobre um destino, atividade ou serviço, em sítios de redes sociais.

Recentemente, a informação disponível nas redes sociais e noutras plataformas digitais tem vindo a ser aproveitada como complemento aos dados provenientes de estatísticas oficiais, nomeadamente, por serem de livre acesso e com um nível elevado de resolução espacial e temporal (Goodchild e Li, 2012). Estes dados, originados a partir da comunidade, proporcionam novas oportunidades para as ciências geográficas e para o estudo de fenómenos sociais, podendo ser utilizados como *proxy* para medir a capacidade de atracção dos locais e para analisar a distribuição espacial dos turistas num determinado local de destino.

O desenvolvimento deste artigo prende-se, por um lado, na análise do comportamento espacial e temporal dos visitantes da cidade de Lisboa através da aplicação de métodos de estatística espacial, com recurso a dados extraídos da rede social ‘Panoramio’ e, por outro lado, com a exploração das relações espaciais que podem explicar o padrão observado nos turistas.

Dados e métodos

Lisboa: A região e a cidade

De acordo com as Estatísticas do Turismo do Instituto Nacional de Estatística (INE, 2015), para o ano 2014, a Área Metropolitana de Lisboa (AML) concentrou 24,95% do total das dormidas no país. A região está entre as primeiras escolhas dos principais mercados externos, sendo o destino preferido pelos espanhóis, brasileiros, norte-americanos e italianos. Lisboa é a cidade do país que apresenta o maior número de dormidas e de hóspedes, com mais de 9 e de 3,5 milhões, respetivamente (INE, 2015). A estada média dos hóspedes estrangeiros na cidade é de 2,5 noites, abaixo da média da AML (2,6) e de Portugal Continental (3,1 noites).

Geotagged fotos de redes sociais como fonte de dados

Durante a sua visita, os turistas, em função das suas motivações, vão consumir as experiências oferecidas pela cidade. Ainda que não se verifique um consumo ativo, haverá pelo menos um consumo indiciário, através do ato de tirar fotografias (Kádár, 2014). Assim como referido por Urry e Larsen (2011), a maioria dos turistas tiram fotografias com o objetivo de criar registos da sua visita. As fotografias são agora amplamente produzidas, consumidas e partilhadas via internet, especialmente através das redes sociais (Lo *et al.*, 2011).

No contexto urbano, vários estudos têm utilizado *geotagged* fotos provenientes de redes sociais para analisar o comportamento espacial dos turistas e identificar os locais de concentração (Encalada *et al.*, 2017; García-Palomares *et al.*, 2015; Tammet *et al.*, 2013), assim como os padrões de mobilidade (Jin *et al.*, 2017). Recorrendo à informação publicada e disponibilizada na rede de partilha de fotos ‘Panoramio’, foram retiradas 75.112 fotografias localizadas no concelho de Lisboa, que correspondem as imagens carregadas pelos utilizadores (da rede) entre os anos 2007 e 2014.

O método para diferenciar as fotos tiradas por visitantes e locais assenta nos trabalhos realizados por Girardin *et al.* (2008), Kádár (2014), García-Palomares *et al.* (2015), Encalada *et al.* (2017). Os autores calcularam a diferença de dias entre a marca temporal (*timesteps*) da primeira e última foto carregadas pelos utilizadores, dentro da área de estudo. Se o número de dias entre a primeira e última foto excedesse a média do tempo de visita do destino (neste caso, < 4 dias), as fotos eram

consideradas como pertencentes a utilizadores locais. Após o processo de filtragem, apuraram-se um total de 17.604 fotos tiradas por cerca de 5.000 utilizadores considerados como visitantes (Figura 1).

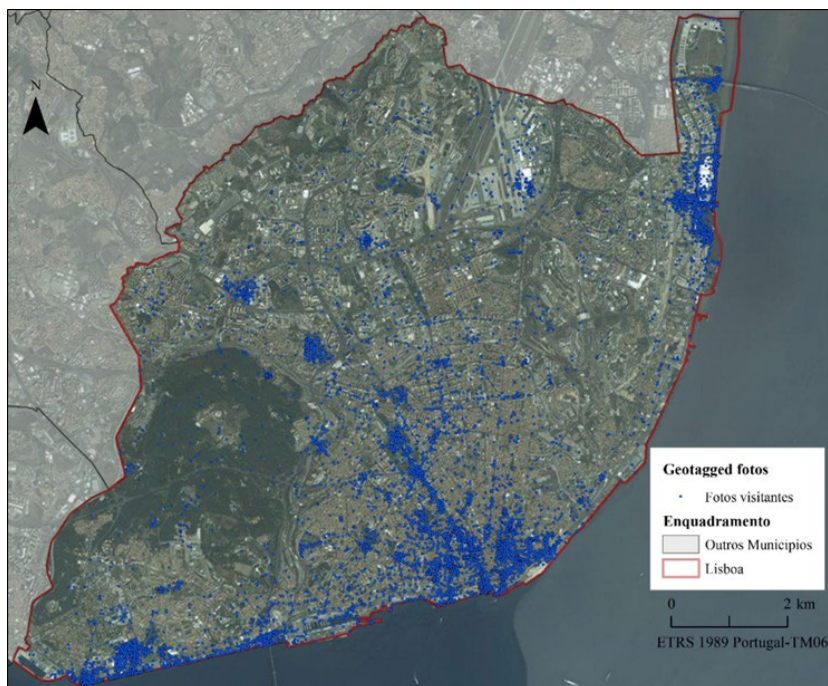


Figura 1. Fotos dos turistas de Lisboa, entre os anos 2007 e 2014.

Análise dos padrões turísticos em Lisboa

Na Figura 2, apresenta-se o esquema processual da aplicação dos métodos estatísticos, sintetizando a sequência dos procedimentos e os propósitos da utilização dos métodos.

O ponto de partida é validar a autocorrelação espacial e determinar se a distribuição dos dados é dispersa, concentrada ou aleatória. Quando avaliada a autocorrelação espacial, surge o interesse em perceber a evolução do grau de aglomeração (*clustering*), em função da variação da distância entre observações, e em comparar os padrões em diferentes períodos, com o fim de avaliar quais as suas mudanças no tempo. Apesar de, visualmente, se notar uma tendência para os visitantes se concentrarem nas áreas turísticas, procura-se, com a aplicação destes primeiros métodos, verificar se o padrão é estatisticamente significativo e, como tal, suportado por um processo espacial subjacente.



Figura 2. Esquema processual da aplicação dos métodos de análise espacial e temporal.

Sendo estas métricas derivadas das características globais da amostra, constatando-se a tendência comum das observações em se concentrarem, o que resta é identificar quais os locais de concentração (mais uma vez, considerando a sua significância estatística). Para o efeito, deve-se recorrer a métodos que consigam interpretar a heterogeneidade da autocorrelação espacial.

Assim sendo, optou-se por uma análise de *clusters* e *outliers*, baseada no *Local Moran Index* (Anselin, 1995), para determinar quais as aglomerações estatisticamente significativas, sejam compostas por valores semelhantes elevados (*High-High clusters*) ou baixos (*Low-Low clusters*), e os *outliers* espaciais. O cálculo do índice implica estabelecer a noção de proximidade entre observações, sendo alcançada, neste caso, através da criação de uma matriz de pesos espaciais.

Todavia, importa discorrer qual a influência da componente temporal na conceção dos *clusters* espaciais. É neste sentido que, através da ferramenta *Emerging hot spot analysis* (ESRI, 2016), se analisa, para além da relação espacial

entre as fotografias, a sua indexação temporal face ao período em que foram carregadas. Assim, em complemento à localização dos focos, obtém-se como resultado o seu padrão sazonal (*e.g.*, se os locais de concentração são consecutivos, esporádicos, etc.).

Previamente à identificação dos padrões espaciais, dedica-se uma parte da análise, à descrição das características da distribuição espacial dos visitantes, com recurso a duas métricas, o centro médio e a elipse do desvio padrão. Por intermédio destas, tenta-se representar a compacidade e a orientação espacial do conjunto de dados. Com o intuito de avaliar se as características da distribuição diferem de um ano para outro, as medidas foram aplicadas para cada ano, de forma segmentada.

Deve-se mencionar que a análise exploratória de dados foi operacionalizada através de várias ferramentas disponíveis no *software* ArcMap. Os parâmetros utilizados para cada uma das ferramentas são mencionados com detalhe no que resta desta secção.

Por fim, exploram-se as relações espaciais existentes entre o padrão observado dos visitantes e um conjunto de 24 variáveis selecionadas que representam parte da oferta turística da cidade (*e.g.*, Hotelaria, arquitetura religiosa, miradouros, etc.). O que se pretende com a aplicação da análise de regressão linear múltipla, é evidenciar se as áreas com potencial turístico são também as que concentram mais visitantes, *i.e.*, se são as mais cativantes para os visitantes.

Centro médio e elipse do desvio padrão

O cálculo do centro médio e da elipse do desvio padrão é simples, tanto em termos computacionais como na sua parametrização. No caso da elipse do desvio padrão, é preciso definir os desvios padrões a considerar para o seu cálculo. Neste caso, as elipses foram apuradas considerando um desvio padrão. Para esta análise, recorreu-se ao posicionamento original das fotografias sem serem agregadas.

Índice do vizinho mais próximo

Para o cálculo do índice é necessário definir a extensão da área de estudo (uma vez que é utilizada para medir a distância média observada e esperada), e o critério de proximidade. A extensão corresponde à área da cidade $\approx 86 \text{ km}^2$. No que diz respeito ao critério de proximidade, a distância entre os vizinhos mais próximos será determinada a partir do segmento de recta que os une, a distância Euclidiana. Para esta análise, utilizou-se igualmente a localização original das fotografias sem serem agregadas.

K-function (Multi-distance spatial cluster analysis)

Os dados a utilizar nesta análise correspondem às fotografias de dois dos oito anos disponíveis (2007 e 2012). Estes anos foram escolhidos por apresentarem uma

quantidade de fotos superior à da média anual. O número de intervalos (de distância), para os quais a análise foi efectuada é de 20. Quanto maior o número, mais suave será a curva dos valores observados da função $L(d)$, permitindo identificar com mais facilidade a distância à qual o grau de *clustering* é maior.

Na distância de partida estabeleceu-se a distância mínima entre as observações e o primeiro vizinho mais próximo, para assegurar que, no início da análise, todas as observações tenham pelo menos um vizinho, garantindo assim a significância dos resultados. A distância resultante foi de 50 m.

O valor do incremento para os intervalos foi estabelecido, tomando como referência a distância média resultante da análise do índice do vizinho mais próximo, *i.e.*, a distância média entre as observações e os seus vizinhos mais próximos (65 m). No que concerne aos limites de confiança -altos e baixos- estes foram determinados com base num teste de 999 permutações, obtendo assim, um nível de confiança de 99,9%. Para esta análise, fez-se novamente uso da localização original das fotografias não-agregadas.

Análise de clusters e outliers (ACO)

As observações foram previamente agregadas em unidades hexagonais de aproximadamente 2.350 m². O valor resultante da célula corresponde à soma do número de fotografias localizadas dentro da própria área. A fim de executar a análise, torna-se necessário definir: i) a área circundante (área que abrange as entidades vizinhas a comparar) a cada *target feature*; e ii) a natureza da relação espacial entre as entidades. A relação espacial foi operacionalizada através de uma matriz de pesos espaciais, sendo baseada no declínio da função da distância (euclidiana). Assim, a influência das *features* vizinhas vai decrescendo, na medida em que a distância entre elas (*target feature* e n vizinhas) vai aumentando. Assumindo que nem todas as observações dentro da área de estudo estão relacionadas espacialmente, estabeleceu-se um limite de distância para o raio de influência da vizinhança, igual a 135 m (toma-se como referência o centroide da célula).

Análise de pontos quentes emergentes (APQE)

Para executar a APQE, deve-se primeiro agregar os dados de entrada num cubo espaciotemporal (em formato NetCDF). Optou-se por uma agregação de um mês a fim de mitigar, ou mesmo omitir, o desvio/erro potencialmente presente na referência temporal das fotografias. A dimensão espacial das caixas de agregação corresponde a mesma área das unidades utilizadas para a ACO. As fotografias foram agregadas em 65.192 caixas espaciotemporais, cobrindo a área de estudo e um período de 91 meses, de Janeiro 2007 a Julho 2014.

O método APQE tem como base a estatística de *Getis-Ord Gi** (Getis e Ord, 1992). Para cada caixa de agregação do cubo, a ferramenta calcula o índice, considerando o valor da própria caixa (*target*) e das caixas na sua vizinhança (estabelecida através dos parâmetros de distância espacial e temporal). Assim, uma caixa de agregação com valor elevado será um ponto quente estatisticamente significativo, a menos que as suas vizinhas no espaço-tempo também apresentem valores elevados. Para estabelecer o limite da vizinhança, considerou-se a mesma distância da ACO, de 135 m da caixa *target* (toma-se como referencia o centroide da caixa, num contexto 2D). A vizinhança temporal foi estabelecida para considerar 12 meses para a mesma localização (caixa).

Análise de regressão lineal múltipla

Na tentativa de explorar as relações espaciais existentes entre o padrão observado dos visitantes e um conjunto de variáveis que representam parte da oferta turística da cidade (e.g., Hotelaria, arquitetura religiosa, miradouros, etc.), aplicou-se uma análise de regressão linear múltipla, podendo assim, evidenciar se as áreas com potencial turístico (ou melhor, áreas mais próximas aos elementos da oferta) são também as que concentram mais fotos.

Um dos primeiros aspetos a ser considerados na esquematização de um modelo de regressão linear é a seleção dos fatores explicativos (variáveis) a utilizar, dado que uma variável independente adicional pode melhorar a previsão da variável dependente.

De acordo com a informação disponibilizada pelo sítio internet da Câmara Municipal de Lisboa referente à atividade turística, foram selecionadas 24 variáveis (Tabela 1) para serem utilizadas como fatores explicativos da aglomeração geográfica dos visitantes. Todas as variáveis (*i.e.*, elementos da oferta representados pela sua localização espacial) foram transformadas em escalas de razão, através da aplicação do inverso da distância euclidiana ao elemento analisado mais próximo. Como a unidade de medida é sempre a mesma (distância em metros) não se revelou necessário proceder à normalização dos valores.

A melhoria do desempenho preditivo de um modelo de regressão linear não está só relacionada com a correlação de uma determinada variável com a variável dependente, mas também com a correlação dessa variável com as demais variáveis independentes do modelo. Deve-se, pois, avaliar as possíveis dependências entre os regressores. Em situações onde essas dependências forem fortes, pode-se afirmar que existe multicolinearidade.

A presença de multicolinearidade pode ser detetada de várias maneiras. Uma técnica muito usada é o valor de tolerância ou o seu inverso, designado por fator de inflação de variação (*Variance Inflation Factor* [VIF]), cujos valores elevados

indicam a existência de multicolinearidade. O VIF é uma medida do grau em que cada variável independente é explicada pelas demais variáveis independentes. Quanto maior for o fator de inflação da variância, mais severa será a multicolinearidade. Como regra geral, as variáveis explicativas com valores de VIF superiores a 7,5 devem ser removidas (uma a uma) do modelo de regressão. Das 24 variáveis pré-selecionadas, 12 foram excluídas por apresentarem valores de VIF > 7,5 (Tabela 1).

Tabela 1
Variáveis explicativas pré-selecionadas, e os respectivos valores de VIF

| Código | Variável | VIF | Código | Variável | VIF |
|--------|--------------------------------|-------|--------|---------------------------------|-------|
| V1 | Arquitetura Civil | 5,39 | V13 | Hotelaria (Alojamentos) | 6,15 |
| V2 | Arquitetura Industrial | 5,87 | V14 | Igrejas | 16,89 |
| V3 | Arquitetura Militar | 20,85 | V15 | Mercados | 10,3 |
| V4 | Arquitetura Nobre | 16,82 | V16 | Miradouros | 58,23 |
| V5 | Arquitetura Premiada | 4,5 | V17 | Monumentos de Interesse Público | 3,96 |
| V6 | Arquitetura Religiosa | 45,07 | V18 | Monumentos Nacionais | 58,77 |
| V7 | Cemitérios | 3,49 | V19 | Museus | 6,24 |
| V8 | Conjuntos de Interesse Público | 104,9 | V20 | Parques Infantis | 4,13 |
| V9 | Docas de Recreio e Marinas | 6,7 | V21 | Parques Merendas | 359 |
| V10 | Elevadores e Ascensores | 114,6 | V22 | Parques Recreativos | 442,8 |
| V11 | Feiras | 22,37 | V23 | Parques Urbanos | 2,99 |
| V12 | Geomonumentos | 5,69 | V24 | Teatros | 4,27 |

Fonte: Elaboração própria.

Posto isto, foi realizada uma análise exploratória em que foram corridos modelos que permitissem a existência de 1 a 5 variáveis independentes. Foram testadas 35.526 combinações das quais 6.866 apresentaram um R^2 ajustado superior a 0,9. Dentro deste grupo, mas restrito, os 38 modelos com melhor desempenho recorreram sempre a 5 variáveis, com desempenhos preditivos entre 0,95 e 0,96. Cruzando as variáveis que foram mais vezes selecionadas por esses 38 modelos com

as 12 variáveis selecionadas através da análise do VIF (que apresentaram valores inferiores a 7,5), obteve-se o conjunto final de 7 variáveis.

Resultados

Centro médio e elipse do desvio padrão

A localização dos centros médios evidencia a tendência geral das observações de se afastarem do interior, situando-se mais para o quadrante sul - sudoeste da cidade.

As elipses (Figura 3) apresentam uma maior cobertura do litoral do que do interior, enfatizando a importância do eixo ribeirinho e das atrações situadas nas suas proximidades (e.g., Praça de Comércio, Mosteiro dos Jerónimos, Padrão dos Descobrimentos, etc.). As elipses expandem-se mais para o sudoeste do que para o nordeste, chegando a envolver, nessa direção, a área circunvizinha ao Padrão dos Descobrimentos. De referir que nenhuma das elipses consegue abarcar, no lado oposto, os pontos de interesse do Parque das Nações, o que demonstra a disparidade existente quanto ao número de observações localizadas em ambos os extremos.

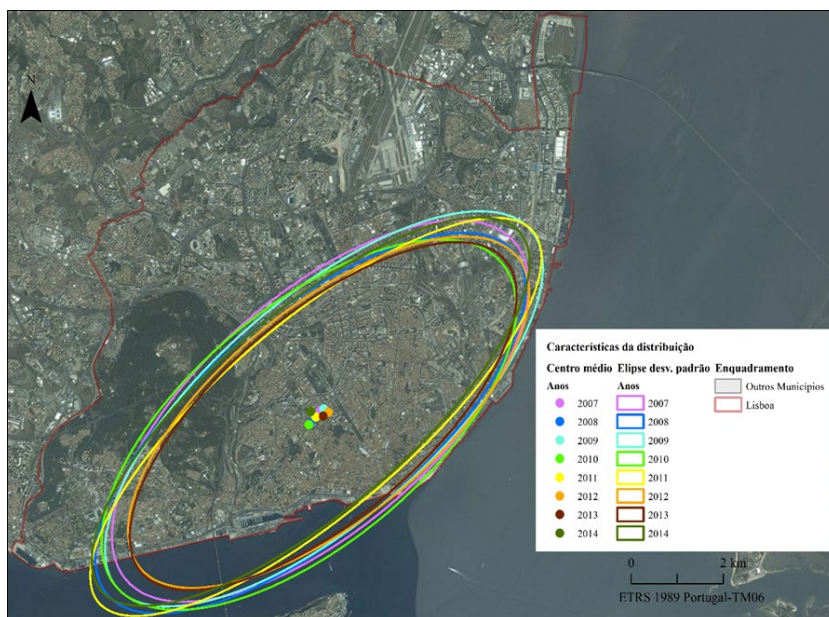


Figura 3. Características da distribuição, Centro médio e Elipse do desvio padrão.

Índice do vizinho mais próximo

O resultado do índice confirma que a distribuição espacial das observações é concentrada (sendo um padrão estatisticamente significativo, $p\text{-value} = 0,000$ e com $z\text{-score}$ inferior a $-2,56$) e que, como tal, não é produto de um processo espacial aleatório.

K-function (Multi-distance spatial cluster analysis)

Nos gráficos da função $L(d)$, para os anos 2007 e 2012 (Figura 4), é visível que a distribuição das observações é constantemente concentrada ao longo dos intervalos de distância estabelecidos (dos 50 m até os 1.300 m). Em todas as séries de distâncias, o padrão de *clustering* é estatisticamente significativo. As linhas dos valores $L(d)$, em ambos os casos, tendem a manter-se em constante crescimento até aos 1.150 m aproximadamente, indicando que após esta distância a significância estatística do padrão já é menos forte.

Os resultados confirmam que a distribuição espacial das observações é concentrada e que o padrão é estatisticamente significativo. Como se pode observar nos gráficos, a tendência de *clustering* é significativa, mesmo nas distâncias mais curtas, o que dá uma ideia sobre a estrutura da distribuição das fotografias. Em síntese, pode se concluir que as fotografias dos visitantes tendem a se concentrar em áreas de pequena dimensão e que essas áreas estão próximas umas das outras.

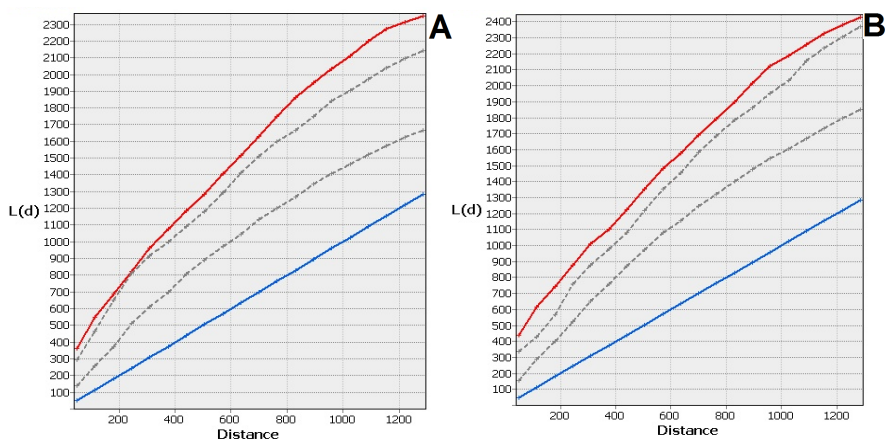


Figura 4. Função $L(d)$ para as fotos agregadas dos anos: a) 2007; b) 2012.

Análise de Clusters e Outliers

Como era de esperar, os *clusters* estão localizados nas zonas turísticas da cidade, e na circunscrição das atrações mais relevantes. Olhando para o mapa (Figura 5), o facto mais saliente tem a ver com que a quase totalidade dos *clusters* espaciais são de categoria *High-High*, ou seja, são locais onde valores elevados estão rodeados por valores semelhantes (elevados). Portanto, tem-se uma grande afluência de visitantes, tanto nesses locais como nas suas áreas contíguas.



Figura 5. Clusters e outliers espaciais do turismo em Lisboa.

Com efeito, verifica-se que as atrações turísticas funcionam como foco dos *clusters* espaciais. É sobre elas que se formam os *clusters*, e é a partir delas que estes se expandem para as áreas próximas. Em muitos casos, a extensão dos *clusters* segue a forma geométrica das atrações (*e.g.*, praças, passeios pedestres), ou dos passeios pedestres, não se expandindo além do perímetro deles.

Os *clusters* estão localizados em três grandes zonas, no Sul, na freguesia de Belém, no Centro, nas freguesias de Santa Maria Maior e Santo António e, no Norte, no Parque das Nações. Na parte sul, existem três *clusters* sobre monumentos: o da Torre de Belém, o do Padrão dos Descobrimentos e o do Mosteiro dos

Jerónimos. No centro, identificam-se 13 grupos de *clusters*, no Parque Eduardo VII, no Marquês de Pombal, nas praças dos Restauradores e Dom Pedro IV, no miradouro de São Pedro de Alcântara (incluindo a estação norte do Elevador da Gloria), no Elevador de Santa Justa e Convento do Carmo, na praça Luís de Camões (incluindo o Largo de Chiado), na estação norte do Elevador da Bica, no Convento da Graça (incluindo o miradouro da Graça), no Castelo de São Jorge, nos Miradouros das Portas do Sol e de Santa Catarina, na Igreja Sé de Lisboa e na praça do Comércio (incluindo parte da rua Augusta). E, no Norte, três *clusters*, dentro do sector do Parque das Nações.

Análise de pontos quentes emergentes

Quando a análise também é suportada pelo variável tempo, denotam-se ligeiras mudanças no esquema dos locais de concentração, que decorrem, obviamente, da junção das duas componentes (espacial e temporal). Ao analisar os resultados da APQE, em comparação com os da ACO, merece destaque o surgimento de vários locais de concentração e o desaparecimento de outros.

Olhando para o mapa (Figura 6), identificam-se cinco locais de concentração que não tinham sido descobertos no ponto anterior. Três deles estão fora das áreas com maior potencial turístico da cidade. De Sul para Norte, o primeiro local situa-se sobre o Museu da Eletricidade, e corresponde a um ponto cuja relevância aparece, nomeadamente, nos períodos mais recentes. Dentro do parque florestal de Monsanto, no antigo restaurante Panorâmico de Monsanto, desvela-se o segundo *hot spot*, cujo padrão temporal é de tipo consecutivo. Mais para o interior, sobressai o aglomerado sobre o Jardim Zoológico de Lisboa, cobrindo a quase totalidade do complexo e com uma relevância que permanece ao longo do período em análise. Na freguesia da Estrela, localiza-se um outro aglomerado, associado à Basílica da Estrela, com significância estatística apenas no último passo de tempo. Na freguesia de Santo António encontra-se o último destes novos focos, sobre o jardim do Torel, envolvendo a área do miradouro e constituindo um ponto quente nos intervalos de tempo finais (*hot spot* recente).

Por outro lado, os locais de concentração que deixaram de ser significativos, no contexto espaciotemporal, encontram-se no centro da cidade. São cinco, na freguesia de Santa Maria Maior (Miradouros das Portas do Sol e de Santa Catarina, igreja Sé de Lisboa, praças dos Restauradores e Dom Pedro IV, e Elevador de Santa Justa) e um, na freguesia da Misericórdia (Elevador da Bica). Nestes locais a significância parece não sobressair ao longo do tempo. Por exemplo, como se pode observar na visualização 3D (Figura 7), os *clusters* associados à Sé de Lisboa, em muitos intervalos de tempo, chegam apenas a um nível de confiança de 90%, sendo a sua significância estatística sem continuidade de relevo.

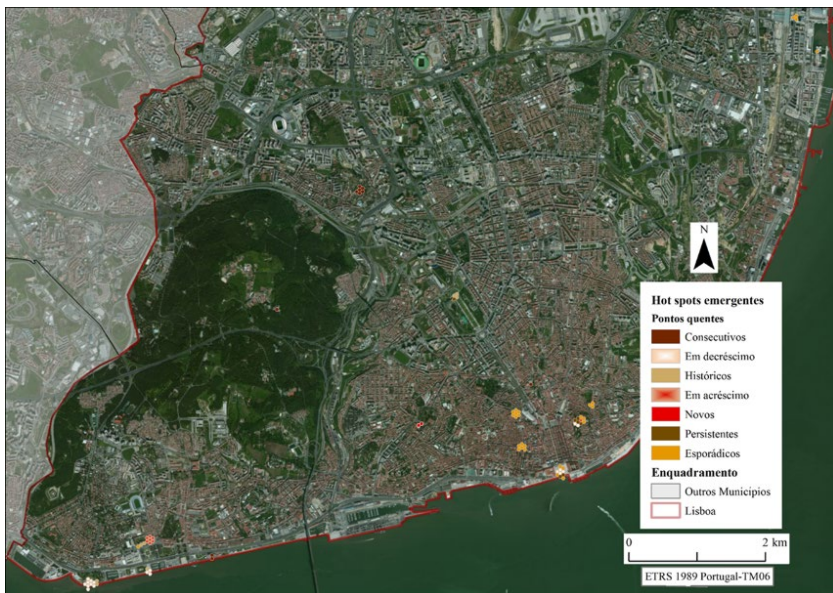


Figura 6. Tipos de pontos quentes emergentes.

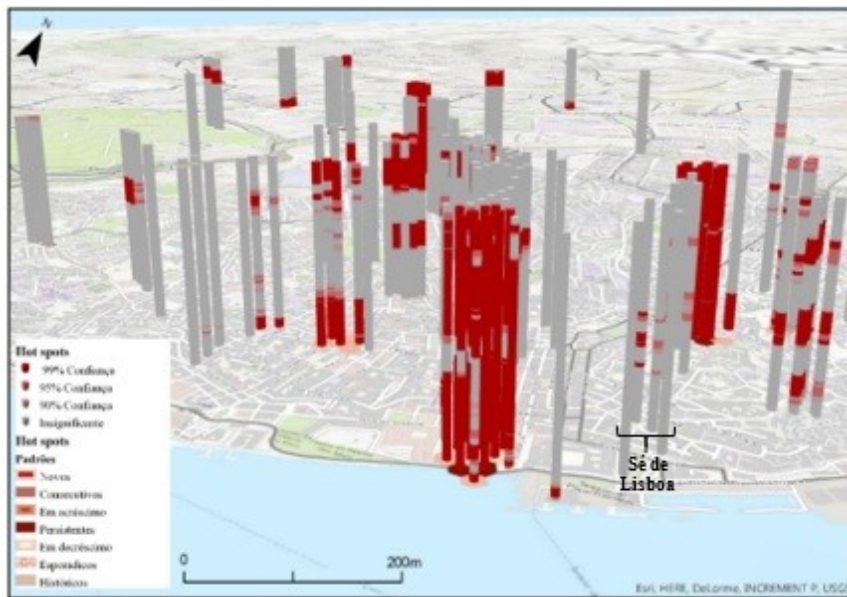


Figura 7. Visualização 3D da conformação dos pontos quentes emergentes, zona centro.

De modo geral, as áreas turísticas típicas funcionam como um foco contínuo (ou consecutivo, segundo a nomenclatura utilizada) para os visitantes. No entanto, alguns lugares são mais prominentes, mantendo frequências cada vez mais elevadas de visitação turística nos últimos anos. Além disso, também é importante o aparecimento de outras áreas turísticas, além das fronteiras do núcleo tradicional de turismo da cidade, que emergem apesar de estarem espacialmente desconectadas mas que sobressaem pela sua relevância histórica.

Análise de regressão

Da análise de regressão linear múltipla, retirou-se que os fatores mais importantes na explicação do padrão de concentração observado, são a proximidade a: Monumentos de Interesse Público (0,29); Docas de Recreio e Marinas (0,21); Arquitetura Premiada (0,15); Geomonumentos (0,12); Hotelaria (0,10); e Teatros (0,10). Como medidas de desempenho do modelo utilizaram-se os valores de R^2 e R^2 ajustado. Um valor de R^2 ajustado de 0,68 indica que o modelo explica cerca de 68% da variação do fenómeno (*i.e.*, distribuição espacial dos visitantes dentro da cidade).

No entanto, para além da análise dos fatores e da predição do modelo, o que é também interessa em termos de planeamento turístico corresponde à análise dos resíduos do modelo. Na prática, analisando a Figura 8, o que os resíduos indicam é que, face às condições observadas, *i.e.*, fatores explicativos, há zonas em que o número de fotos é superior ao expectável (resíduos positivos) e outras onde há menos fotos do que seria previsível. São precisamente estas áreas, que podem necessitar de intervenções no âmbito do planeamento para que possam explicar todo o seu potencial.

No mapa dos resíduos, para além de Belém sobressair com valores acima do que o seu potencial justificaria, destaca-se de forma inesperada a zona da Baixa, com resíduos negativos, encontrando-se num nível inferior do que potencialmente poderá atingir. A zona do Parque das Nações segue a mesma tendência, mas não com um diferencial tão grande. Contudo, são claramente duas zonas identificadas que podem ser alvo de ações, tendo em vista a captação de mais visitantes.

Conclusões

O uso de soluções tecnológicas nos processos de tomada de decisão contribuiu positivamente para avaliar as ações necessárias, o que é crucial para uma gestão sustentável do turismo e dos seus impactos. Com estas recentes soluções, potenciadas pelo envolvimento de ferramentas inteligentes, as cidades podem explorar novas abordagens e desafios em termos de planeamento do turismo, que se tornam acessíveis pelos dados disponibilizados em plataformas digitais online.

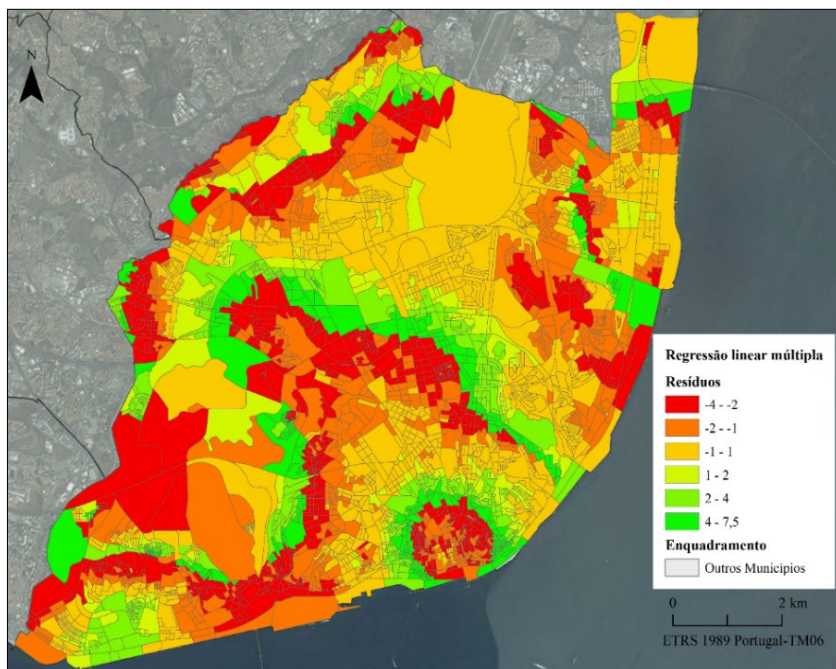


Figura 8. Resíduos do modelo de Regressão Linear Múltipla.

No caso de Lisboa, o mapeamento e consequente análise da pegada digital dos turistas levaram a constatar, por um lado, que a sua distribuição espacial não é homogénea e que, efetivamente, os turistas tendem a se concentrar nas áreas com maior potencial turístico da cidade, e, por outro lado, que são os elementos monumentais os que se mostram como focos de maior atratividade dos turistas, (*e.g.*, castelo, praças, igrejas).

O conjunto de dados (de 17.604 observações utilizadas para a análise) constitui uma amostra robusta que replica o espaço vivido pelos turistas. Os resultados da avaliação estatística da tendência geral de *clustering* evidenciou a sua aptidão para a análise espacial, deixando de lado qualquer possibilidade que sua gênese seja produto da aleatoriedade.

As áreas de interesse podem ser definidas analisando a concentração das fotografias dos turistas através de operações realizadas em SIG. Algumas delas podem nem ser consideradas atrações turísticas pelas autoridades competentes. Desta forma, podem saber quais as áreas que são mais atrativas para os visitantes e quais são as características dessas áreas (espaços verdes, monumentos, lugares com arquitetura moderna, etc.).

Esta evidência empírica sustenta o emprego destes dados, como um proxy sensível da distribuição espacial e temporal dos visitantes, para alcançar uma leitura quantitativa e geográfica sobre o comportamento da procura. As implicações dos resultados apresentados podem ser significativas para o planeamento e gestão das infraestruturas de transporte, o desenvolvimento de produtos e da imagem, e a gestão de impactos, especialmente para os destinos que são altamente dependentes do turismo.

Bibliografia

- Ashworth, G. and Page, S. (2011). “Urban tourism research: Recent progress and current paradoxes”, em *Tourism Management*, 32(1), 1-15. DOI: 10.1016/j.tourman.2010.02.002.
- Anselin, L. (1995). “Local Indicators of Spatial Association —LISA”, *Geographical Analysis*, 27(2), 93-115, <<https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>>.
- Buhalis, D. and Amaranggana, A. (2014). “Smart tourism destinations”, em Z. Xiang, & L. Tussyadiah (eds.), *Information and communication technologies in tourism 2014*, 553-564.
- Buhalis, D. and Law, R. (2008). “Progress in information technology and tourism management: 20 years on and 10 years after the internet - The state of eTourism research”, em *Tourism Management*, 29, 609-623. DOI: 10.1016/j.tourman.2008.01.005.
- Encalada, L.; Boavida-Portugal, I.; Ferreira, C. and Rocha, J. (2017). “Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city”, em *Sustainability* 9(12), 2317. <<https://doi.org/10.3390/su9122317>>.
- ESRI (2016). *Emerging Hot Spot Analysis*. Retrieved May 28, 2018, from <<http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm>>.
- Girardin, F.; Fiore, F.; Ratti, C. and Blat, J. (2008). “Leveraging explicitly disclosed location information to understand tourist dynamics: A Case Study”, em *Journal of Location Based Services*, 2(1), 41-56, <<http://dx.doi.org/10.1080/17489720802261138>>.
- García-Palomares, J.; Gutiérrez, J. and Mínguez, C. (2015). “Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS”, em *Applied Geography*, 63, 408-417, <<https://doi.org/10.1016/j.apgeog.2015.08.002>>.
- Getis, A. and Ord, J.K. (1992). “The Analysis of Spatial Association by Use of Distance Statistics”, *Geographical Analysis*, 24(3), 189-206, <<https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>>.

- Goodchild, M. and Li, L. (2012). “Assuring the quality of volunteered geographic information”, em *Spatial Statistics*, 1, 110-120. DOI: 10.1016/j.spasta.2012.03.002.
- Hayllar, B.; Griffin, T. and Edwards, D. (2008). “Urban Tourism Precincts: Engaging with the field”, em B. Hayllar, T. Griffin, and D. Edwards (eds.), *City Spaces-Tourist Places: Urban Tourism Precincts*, 3-18.
- Instituto Nacional de Estatística —INE (2015). “Estatísticas do Turismo 2014”. *Instituto Nacional de Estatística*, Lisboa, Portugal.
- Jin, C.; Cheng, J. and Xu, J. (2017). “Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility”, *Journal of Travel Research*, 57(6), 779-791, <<https://doi.org/10.1177/0047287517714906>>.
- Kádár, B. (2014). “Measuring tourist activities in cities using geotagged photography”, em *Tourism Geographies*, 16(1), 88-104. DOI: 10.1080/14616688.2013.868029.
- Lau, G. and McKercher, B. (2006). “Understanding tourist movement patterns in a destination: A GIS approach”, em *Tourism and Hospitality Research*, 7(1), 39-49.
- Lo, I.; Mckercher, B.; Lo, A.; Cheung, C. and Law, R. (2011). “Tourism and online photography”, em *Tourism management*, 32, 725-731. DOI: 10.1016/j.tourman.2010.06.001.
- Munar, A. and Jacobsen, J. (2014). “Motivations for sharing tourism experiences through social media”, em *Tourism Management*, 43, 46-54. <<https://doi.org/https://doi.org/10.1016/j.tourman.2014.01.012>>.
- Pearce, D. (2001). “An integrative framework for urban tourism research”, em *Annals of Tourism Research*, 28(4), 926-946. PII S0160-7383(00)00082-7.
- Tammet, T.; Luberg, A. and Järv, P. (2013). “Sightsmap: Crowd-sourced popularity of the world places”, em L. Cantoni and Z. Xiang (eds.), *Information and Communication Technologies in Tourism 2013*, 314-325.
- Urry, J. and Larsen, J. (2011). “The tourists gaze 3.0”, London, SAGE Publications, Ltd.

Creación de un modelo semántico para modelar los fenómenos meteorológicos de los siglos XIX y XX en Latinoamérica

Diana Comesaña*
Bryan Barreiro**

Recibido 22 de marzo de 2018, aceptado 18 de agosto de 2018

Resumen

Los fenómenos meteorológicos son eventos que se desarrollan en la atmósfera y están asociados a un lugar específico. El objetivo del proyecto “Compartiendo la historia escondida del cambio climático en América Latina a través de las TIC”, es recuperar y georreferenciar las noticias sobre eventos meteorológicos registrados en la prensa escrita de los siglos XIX y primer mitad del XX, de los países involucrados en el proyecto: Colombia, Ecuador, México y Uruguay, y desarrollar modelos semánticos, transformando la información original recopilada para incrementar la interoperabilidad conforme a los principios de Linked Data y estándares del W3C. La terminología a modelar se obtiene por un análisis del dominio basado en el paradigma de Análisis de Dominio de Birger Hjørland. Se estudia la prensa digitalizada por las Bibliotecas Nacionales de los países intervinientes y el lenguaje empleado por sus instituciones meteorológicas nacionales y por la Organización Meteorológica Mundial. El análisis muestra un lenguaje uniforme, tanto respecto a la distribución geográfica, como temporal, con poca sinonimia y escasa polisemia. Se aplica el método METHONTOLOGY para desarrollar una ontología de Eventos_Meteorológicos en el lenguaje OWL. Posibilita recuperar información no visible de las Bibliotecas Nacionales con técnicas de recuperación de información. Sus desarrollos son aplicables a otros documentos digitalizados y otras disciplinas.

* Universidad de la República – FIC, San Salvador 1944, 11200 Montevideo, correo electrónico: diana.comesana@fic.edu.uy

** Centro Regional de Profesores del Centro, Independencia y 24 de Abril, Ciudad de Florida, correo electrónico: bryanalexander03@gmail.com

Palabras clave: *Web Semántica, Semántica Geoespacial, recuperación de la información, ontología, meteorología, cambio climático.*

Resumo

Os fenômenos meteorológicos são eventos que se desenvolvem na atmosfera e estão associados a um lugar específico. O objetivo do projeto “Compartilhando a história escondida da mudança climática na América Latina através das TIC”, consiste em recuperar e georreferenciar as notícias sobre eventos meteorológicos registrados na imprensa escrita dos séculos XIX e primeira metade do século XX, dos países envolvidos no projeto: Colômbia, Equador, México e Uruguai. Trata de desenvolver modelos semânticos, transformando a informação original recompilada para incrementar a interoperabilidade conforme os princípios de Linked Data e padrões do W3C. A terminologia a ser elaborada é obtida através de uma análise do domínio baseado no paradigma de Análise de Domínio de Birger Hjørland. Se estuda a imprensa digitalizada das Bibliotecas Nacionais dos países intervenientes e a linguagem empregada por suas instituições meteorológicas nacionais e Organização Meteorológica Mundial. A análise mostra uma linguagem uniforme, tanto no que diz respeito a distribuição geográfica, como temporal, com pouca sinonímia e baixa polissemia. Se aplica o método METHONTOLOGY para desenvolver uma ontologia de Eventos_Meteorológicos na linguagem OWL. Possibilita recuperar informação não visível das Bibliotecas Nacionais com técnicas de recuperação de informação. Seus desenvolvimentos são aplicáveis a outros documentos digitalizados e outras disciplinas.

Palavras chave: *Web Semântica, Semântica Geoespacial, recuperação da informação, ontologia, meteorologia, mudança climática.*

Abstract

The meteorological phenomena are events that develop in the atmosphere and are associated to a specific place. The objective of the project “Sharing the hidden history of climate change in Latin America through ICT” is to retrieve and georeference the news about meteorological events recorded in the written press of the 19th through the first half of the 20th century, of the countries involved in the project: Colombia, Ecuador, Mexico and Uruguay and develop semantic models, transforming the original information collected to increase interoperability according to the principles of Linked Data and W3C standards. The terminology to be modeled is obtained by an analysis of the domain based on the paradigm of Domain Analysis of Birger Hjørland. The digitized press by the National Libraries of the countries involved and the language used by their national meteorological institutions and the World Meteorological Organization is examined. The analysis shows a uniform use of language, in terms

of geographical distribution, as temporary, with little synonymy and only one case of apparent polysemy. The METHONTOLOGY method is applied to develop the ontology of Eventos_Meteorologicos. It makes it possible to retrieve non-visible information from National Libraries with information retrieval techniques and natural language processing. Its developments are applicable to other digitized documents from other disciplines.

Key words: *Semantic Web, Geospatial Semantics, Information retrieval, ontology, meteorology, climate change.*

Introducción

En la década de los sesenta, Salton estableció las bases de lo que hoy conocemos como Recuperación de la Información (RI). El surgimiento de la web en los años 90, le aportó una dimensión nunca imaginada, afectando a millones de personas. La información que se encuentra disponible es inagotable, pero no siempre es fácil llegar a ella. El problema central de la RI clásica era poder determinar la relevancia de un documento ante cualquier consulta.

La web actual es lo que se denomina la web de documentos, es decir una web integrada por páginas web o documentos creados para ser interpretados por los seres humanos y es una web en transición a lo que se denomina Web Semántica.

Según el World Wide Web Consortium (W3C. 2017), “la Web Semántica es una Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante”.

Aspira a la recuperación automática de datos, a la interoperabilidad de entre comunidades y dominios y al razonamiento automático.

La web semántica está relacionada con la inteligencia artificial ya que busca que las computadoras u ordenadores “comprendan” la consulta que hace un usuario, con el fin de devolver un conjunto de resultados acordes a la misma.

Como dice Regillo (2018):

La web semántica es una extensión de la Web a través de una serie de estándares definidos por el W3C con el fin de dotar de significado a los datos. Se pretende que estos datos resulten procesables y entendibles por un sistema mediante una serie de anotaciones semánticas en formato XML, fundamentado en el concepto de tripla (sujeto, predicado, objeto) como estructura predicativa para exponer un hecho o propiedad sobre un objeto.

La web semántica aspira a combatir la *infoxicación* o sobrecarga informativa y la frustración de los usuarios frente a la avalancha de información que se encuentra en

Internet, ahorrando tiempo al usuario y devolviendo resultados de búsquedas acordes a sus necesidades.

Un gran porcentaje de las preguntas habituales que aparecen en una consulta web presentan una componente espacial: ¿Dónde se ubica...?, ¿dónde trabajas?, ¿dónde está el hospital más cercano al lugar en que me encuentro?, ¿cuál es el límite Este del Uruguay?, ¿qué clima nos espera hoy en Montevideo? Son información geográfica: datos censales, división político-administrativa, redes de transportes, fotografías aéreas, imágenes satelitales, informes meteorológicos, etc.

Este tipo de información, con los datos distribuidos, servicios compartidos, tecnologías y arquitecturas a los que se agrega el reconocimiento de información geográfica, de asociaciones espaciales y de razonamiento espacial, constituye la Web Geoespacial. Si esta información geográfica se halla enriquecida con la incorporación de metadatos semánticos y el desarrollo de ontologías que faciliten la comprensión e interoperabilidad de información, estaremos frente a una web geoespacial semántica, donde los metadatos son estándares para describir la información geográfica y las ontologías, formas explícitas y lógicas de expresar el lenguaje.

Es el escenario que estudia el proyecto IPGH N° HIST 02 2017 “Compartiendo la historia escondida del cambio climático en Latinoamérica a través de las TIC”, cuyo objetivo es obtener modelos semánticos asociados a los eventos meteorológicos recogidos en la prensa escrita de los siglos XIX y primera mitad del XX. Este proyecto fue auspiciado por el Instituto Panamericano de Geografía e Historia (IPGH), enmarcado en el contexto de los Proyectos Panamericanos de Asistencia Técnica (PAT), 2017. En él participaron investigadores de Colombia, Ecuador, México y Uruguay y se alinea con los objetivos de la Agenda 2030 de Desarrollo Sostenible del Programa de las Naciones Unidas para el Desarrollo, que en su punto N°13 “Acciones por el clima”, destaca: “Las pérdidas anuales promedio causadas solo por tsunamis, ciclones tropicales e inundaciones alcanzan los cientos de miles de millones de dólares y exigen inversiones de unos US\$ 6.000 millones anuales solo en gestión de los riesgos de desastres” (PNUD, 2018).

Todos los países sufren los efectos del cambio climático, y Latinoamérica constituye un caso particular: mientras que los países del llamado “Primer Mundo” generan la mayor parte de emisiones de gases de efecto invernadero y sufren sus consecuencias en cuanto al aumento de intensidad y frecuencia de los eventos meteorológicos extremos, la región de América Latina contribuye poco y sufre de manera desproporcionada.

Es una de las regiones más vulnerables al localizarse dentro de la franja de huracanes y tener numerosos Estados insulares y zonas costeras bajas. Depende de los deshielos andinos para el suministro de agua a sectores urbanos y agrícolas y está expuesta a inundaciones e incendios forestales.

Estudiar el comportamiento de los eventos meteorológicos del pasado, lleva a entender mejor el presente, con el fin de poder diseñar acciones tendientes a minimizar las consecuencias del cambio en el comportamiento del clima de la región.

Los fenómenos meteorológicos

Los fenómenos meteorológicos son eventos que se desarrollan en la atmósfera y están asociados a un lugar específico, puede asignarles coordenadas y constituyen por ello información geográfica.

Según el Programa Regional de Meteorología (s.f.) se pueden definir como “aque- llos procesos permanentes de movimientos y de transformaciones que sufre la natu- raleza y que pueden influir en la vida humana” Comprenden: lluvias, vientos, fríos extremos, calor, nevadas, granizo, niebla y otros que, estudiados en su conjunto y en su variabilidad en un lapso de tiempo determinado, nos muestran indicios del clima y sus variaciones.

Es decir, son cambios que suceden en la naturaleza por sí mismos. Puede tratarse de un fenómeno tan habitual e inofensivo como una llovizna o de un fenómeno poco usual y catastrófico como un huracán o tornado. La ocurrencia o no de estos fenóme- nos está relacionada a ciertas zonas geográficas y épocas del año.

Los eventos o fenómenos meteorológicos están fuertemente relacionados con la información geoespacial, ya que los mismos son la materia prima para la construcción de mapas o atlas climáticos. Dichos fenómenos son las variables representadas en estos documentos cartográficos. Se pueden elaborar, por ejemplo, mapas de riesgo de inundaciones, mapas de huracanes y mapas de sequías entre otros.

Estos mapas climáticos sirven de apoyo para la investigación y el estudio en la prevención de catástrofes naturales o para la gestión de alternativas cuando los fenó- menos son inevitables.

El surgimiento de los Sistemas de Información Geográfica (SIG) ha facilitado enormemente la construcción de este tipo de cartografía. Según la *National Center for Geographic Information and Analysis*, citada por Feito y Segura (2010) un SIG es “un sistema de hardware, procedimientos y software elaborado para facilitar la ob- tención, gestión, manipulación, análisis, modelado y representación de datos espa- cialmente referenciados y para la resolución de problemas complejos que impliquen la manipulación y gestión de dichos datos”, por ende, capaz de producir cartografía y permitir avanzar en el estudio y análisis de los fenómenos geográficos. La climato- logía es una de las disciplinas que se beneficia del uso de estas técnicas. “Un SIG puede utilizar mapas analógicos, mapas digitales, fotografías aéreas, modelos digita- les de elevaciones (MDE), imágenes satelitales, registros de GPS para, posterior- mente, elaborar cartografía de gestión de riesgos” (Ribera Masgrau, 2004).

Por lo tanto, los fenómenos naturales o meteorológicos, tienen estrecha relación con la información geográfica y con la cartografía producida, especialmente con aquella destinada a la toma de decisiones para la gestión y prevención de desastres naturales.

Para Gallegos (2011)

Todo trabajo de investigación de tipo histórico se vincula con una multiplicidad de fuentes que conforman un eje desde el cual el historiador trae hacia el presente los hechos que se analizan. La labor del investigador radica en la selección de las metodologías adecuadas y la recopilación de bibliografía que dé cuenta, por un lado, del estado de la cuestión, y por otro del contexto de estudio.

La Web se ha transformado en un banco infinito de datos, y el desafío lo constituye el poder descubrir patrones que permitan la exploración automática de los mismos. Es lo que se denomina “minería de datos”, una disciplina que reúne la estadística, la informática y la documentación, empleando métodos de inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. Si se aplican esos métodos y herramientas a los fondos de una biblioteca, entonces nos hallamos frente a una *bibliominería*.

El éxito o fracaso de la bibliominería depende de la pertinencia de las fuentes de información seleccionadas. Las características de la prensa hispanoamericana del siglo XIX, la transforman en una inmejorable fuente de información para el estudio de los eventos de este siglo.

Como dicen Mejías y Arias (1998), para América Latina, “El siglo XIX es, por excelencia, el siglo de la prensa escrita”. Muchas personalidades de la cultura y de diferentes ámbitos científicos y profesionales, estimaban de prestigio colaborar en publicaciones periódicas.

Emilio Castelar, citado por Mejías y Arias (1998), dijo: “Cuando tomo en mis manos un periódico, cuando recorro sus columnas, cuando considero la diversidad de sus materias y la riqueza de sus noticias, no puedo menos de sentir un raptó de orgullo por mi siglo...” y es que la prensa hispanoamericana estaba totalmente comprometida con su labor educadora y de difusión, de allí la variedad e importancia de los datos que recoge.

La ventaja que ofrece el análisis de las publicaciones periódicas, en particular los periódicos, es que los mismos documentan hechos cotidianos del día a día y permite rescatarlos del olvido para las generaciones futuras.

Los periódicos registran los distintos fenómenos meteorológicos ocurridos y sus consecuencias: lluvias, heladas, sequías, etc., que conforman fuentes documentales para el estudio del cambio climático en determinada región o país.

El límite cronológico se estableció en 1950, pues en 1951 se creó la Organización Meteorológica Mundial y se asume que, a partir de este momento, es factible hallar datos cuantificados y análisis de los fenómenos meteorológicos.

Análisis del dominio y de las fuentes documentales

Para llegar a un lenguaje inicial, que permitiera la recuperación automática o semi-automática de los datos buscados, se recurrió a un análisis del dominio basado en el paradigma de Análisis de Dominio de Birger Hjørland, donde, según sus palabras: “Se estudian de las relaciones entre documentos, áreas de conocimiento y discursos en relación con las posibles perspectivas de acceso de las distintas comunidades de usuarios” Birger Hjørland (2002).

Este paradigma plantea estudiar los dominios del conocimiento como comunidades discursivas. Se toma en consideración el contexto psicosocial y socio lingüístico de la sociología del conocimiento y de la ciencia. Implica que un dominio del conocimiento se estudia a través de sus representantes, inmersos en la comunidad de la que forman parte.

Así se consultó a expertos en meteorología y se estudiaron los documentos por ellos sugeridos. Se analizaron los glosarios meteorológicos empleados por las instituciones meteorológicas de cada país y de la Organización Meteorológica Mundial. Uno de los problemas que causaba más inquietud era poder cuantificar fenómenos que la prensa describía sin dimensionar, para poder llegar a una clasificación jerárquica de los fenómenos meteorológicos registrados. Por consejo de los expertos asesores, se emplearon las escalas Anemométrica de Beauford, que constituye una medida empírica para la velocidad del viento y está basada principalmente en la fuerza del mismo y el estado del mar y sus olas, y la Escala de Douglas que proporciona la altura de las mismas.

Los documentos analizados definen terminología que está en uso y es común a todos los países, permiten establecer un orden jerárquico en los términos y lo interesante es que surgen en la época que se estudia, por lo que no hay contradicciones con el lenguaje reflejado por los periódicos analizados.

Una de las observaciones de este análisis permitió determinar que, en esta disciplina, no existe gran variación terminológica, ni por los factores geográficos, ni por la coordenada temporal. Se identificaron unos pocos sinónimos y un único caso de aparente polisemia, aunque un estudio más profundo, determinó que uno de los significados era un caso específico del otro, el cual correspondía a un concepto algo más general.

El lenguaje natural en que se redactaron las noticias, aportaron sinónimos y términos relacionados con la terminología específica del área de conocimiento tratada, principalmente en cuanto a los efectos detectables de un fenómeno meteorológico.

Desarrollo de una ontología de eventos meteorológicos

Desde el punto de vista semántico, el principal problema es que la mayoría de la información que se comparte es entregada en documentos en texto plano, lo cual implica que no haya metadatos o información adicional que ayude a los buscadores web a procesar la información y sugerir a los usuarios el contenido más relevante (Caldón *et al.*, 2010).

Las nuevas tecnologías y conceptos que conforman la Web semántica permiten describir formalmente los conceptos y sus relaciones mediante lenguajes especializados evitando ambigüedades conceptuales. Los desarrollos ontológicos son los modelos por los que se ha optado en el desarrollo del proyecto, con el fin de agregar semántica, desde el punto de vista informático, al sector.

¿Qué significa esto? Es necesario hallar la forma de proporcionar la semántica más fuerte, para que los equipos puedan realizar sus inferencias.

El estudio de las ontologías disponibles mostró la falta de una que presentara suficiente especificidad para abarcar los términos correspondientes a los fenómenos meteorológicos en toda su complejidad.

Entre otras ontologías existentes se revisaron las ontologías *Falcons search*, especialmente donde la clase *meteorological_event* es sub-clase de *weather_events* y está disponible en <http://umbel.org/reference-concept/?uri=ImmediateWeatherProcess>.

Se verificó que no desarrolla los términos más allá de presentar la clase *meteorological_event*. Se estudió el conjunto de ontologías “objetos_geográficos”, desarrollados en la tesis de maestría “Modelo Conceptual de Información Geográfica para la IDE —Uruguay”, destinadas a conformar un meta-modelo mediante la extensión de lógica descriptiva ALCQM, y que prevé el desarrollo de una clase *información_meteorológica*, en la ontología “Usos_especiales”, (Comesaña, 2015).

Se decidió crear una ontología nueva, que, partiendo del término “eventos_meteorológicos”, se adaptara a los dos desarrollos anteriores e incluyera la terminología contenida en los glosarios de los institutos meteorológicos de Colombia, Ecuador, México, Uruguay, la Organización Meteorológica Mundial y las tablas Anemométrica de Beauford y de Douglas.

La ontología de eventos meteorológicos, se desarrolló en lenguaje OWL, a través del método de desarrollo de ontologías METHONTOLOGY y se implementó con el software Protege 5.2.0.

Si bien se conocen nuevos métodos como NeOn, para construir ontologías, atendiendo a la interdisciplinariedad del grupo, donde trabajaron Documentalistas, se optó por el empleo de METHONTOLOGY, por hallarse éstos más familiarizados con su uso.

En lugar de crear el meta-modelo con la extensión ALCQM, se empleó la metodología NeOn para construir una red de ontologías sumando a la desarrollada de

“eventos meteorológicos”, aquellas que cubren otras necesidades del proyecto, por ejemplo, en cuanto a nombres geográficos y fechas.

“Eventos_meteorológicos”, está disponible registrándose en la comunidad GNOSS, en el grupo “Interoperabilidad Semántica”, donde fue publicada bajo licencia Creative Commons Attribution 3.0, en el sitio: <<https://my.gnoss.com/comunidad/mygnoss/recurso/eventosmeteorologicos/cff52099-003f-4395-be54-4a528fc3553f>>.

Este desarrollo comprendió la construcción de un glosario. La Tabla 1, muestra un sector del mismo donde se indica nombre del término, sinónimos, descripción en lenguaje natural y tipo (clase, concepto, propiedad o atributo).

Tabla 1
Construcción del Glosario

| <i>Nombre</i> | <i>Sinónimo</i> | <i>Descripción</i> | <i>Tipo</i> |
|-------------------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <i>Agua Nieve</i> | Cellisca | Tipo de precipitación en la que el agua presenta dos estados, teniéndose una mezcla de agua congelada y agua líquida | Concepto |
| <i>Eventos meteorológicos</i> | Fenómenos Meteorológicos | Los <i>fenómenos meteorológicos</i> incluye todos los acontecimientos que se pueden producir ligados a la dinámica de la atmósfera. | Clase |
| <i>Meteoro</i> | | Fenómeno además de las nubes, que es observado en la atmósfera o en la superficie del globo terrestre. Los meteoros, teniendo en cuenta la naturaleza de sus partículas constitutivas o los procesos físicos que intervienen en su formación, se han clasificado en cuatro grupos principales: hidrometeoro, litometeoro, fotometeoro y electrometeoro | Subclase |
| <i>Velocidad del viento</i> | | Aire en movimiento, medido en Km/h | Atributo |

En esta Tabla 1, la diferencia entre “clase” y “concepto”, está dada porque “clase” es un término que reúne jerárquicamente un conjunto de términos-concepto simples

relacionados entre sí por cierta característica. Por ejemplo, la Figura 1, se muestra una clase “meteoros”, subdividida en Subclases “electrometeoros”, “fotometeoros”, “hidrometeoros”, etc., según el tipo de evento meteorológico de que se trate. La subclase “electrometeoros” reúne los conceptos: “truenos”, “fuegos de San Telmo”, “centellas”, “relámpagos”, “rayos” y “aurora polar”, términos simples o elementales.

Con base en este glosario se construye una taxonomía o árbol terminológico del dominio, en que se establece la jerarquía de los conceptos (Figura 1).



Figura 1 . Sector del árbol terminológico, tal como se desarrolla a través del software Protégé 5.0.0.

La jerarquía mostrada en el árbol terminológico surge del estudio de los glosarios y documentos sugeridos por los expertos en el análisis del dominio. Se separaron aquellos términos que corresponden a la climatología, de los de meteorología, y dentro de éstos, los que corresponden a los diferentes conceptos generales. Los fenómenos meteorológicos se clasificaron de acuerdo al agente que los produce.

Es necesario crear una clase nominal “efectos_detectables”, que constituye parte importante del cuerpo de las noticias que se analizan y que permite identificar y cuantificar los eventos meteorológicos documentados. También se crea la clase “registro_documental”, cuyas instancias recogen el corpus de documentos con que se trabaja.

Se establecen las relaciones binarias “documenta” y su inversa “documentado_en” para relacionar los “efectos_detectables” con el “registro_documental”; “produce” y “producido_por” relacionan los eventos meteorológicos con sus efectos, “registra” y “registrado_en” indica en que registro documental se recoge cierto evento meteorológico, y “publica” y “publicado_en” une los “avisos_meteorológicos” con su “registro_documental”. Las relaciones “localiza” y “ubicado_en” dejan la “puerta abierta” para enlazar el evento a un lugar geográfico determinado y “ocurre” y “sucede_el” con su fecha.

El árbol terminológico, lleva a obtener un Diccionario de conceptos que muestra la Tabla 2, donde se registra el nombre del concepto, las instancias correspondientes, propios de los atributos de clase, de las instancias y las relaciones que los unen a los documentos en que están registrados y los efectos que producen.

Tabla 2
Diccionario de Conceptos

| <i>Nombre del concepto</i> | <i>Instancias</i> | <i>Atributos de clase</i> | <i>Atributos de instancia</i> | <i>Relaciones</i> |
|----------------------------|-------------------|---------------------------|-------------------------------|-------------------|
| <i>Viento</i> | Pampero | Velocidad del viento | Efectos detectables | Produce |
| | Sudestada | | | Documentado_en |
| <i>Niebla</i> | — | Visibilidad | — | Produce |
| | | | | Documentado_en |

Para los Atributos de Clase e Instancia, se construyen tablas que nos muestran el nombre del atributo, concepto a que se relaciona, tipo de valor, rango y cardinalidad (Tablas 3 y 4).

Los atributos de clase surgen de la aplicación de las definiciones de los glosarios y de las tablas de Beauford y Douglas mencionadas.

Tabla 3
Atributos de clase

| <i>Atributo de clase</i> | <i>Concepto</i> | <i>Tipo de valor</i> | <i>Rango</i> | <i>Cardinalidad</i> |
|-----------------------------|-----------------|-----------------------|--------------|---------------------|
| <i>velocidad_del_viento</i> | Viento | No negativo (en km/h) | {0,n} | (1,n) |
| <i>Altura_de_la_ola</i> | Tormenta | No negativo (en mts) | {0,n} | (1,n) |

Tabla 4
Atributos de Instancia

| <i>Atributo de Instancia</i> | <i>Concepto</i> | <i>Tipo de valor</i> | <i>Rango</i> | <i>Cardinalidad</i> |
|------------------------------|-----------------|----------------------|--------------|---------------------|
| <i>efecto_detectado</i> | Tornado | cadena_de_caracteres | — | (1,n) |

Las instancias se obtienen al rescatar las noticias de la prensa digitalizada en forma automática o semiautomática, y aún manual cuando el estado de la digitalización no lo permita de otra manera. Se obtiene el periódico que la registra el evento, la fecha en que aconteció, el lugar y los efectos que detectados.

Las definiciones lógicas de los conceptos se construyeron empleando la sintaxis Mánchester, que emplea el software Protégé.

Así puede decirse que existe una clase llamada eventos meteorológicos, la precipitación es una sub-clase de los hidrometeoros, que son una sub-clase de los meteoros y a su vez son una sub-clase de los eventos meteorológicos. La precipitación tiene asociada exactamente una medida decimal y produce cierto efecto detectable:

$\text{precipitación} \subseteq \text{hidrometeoro}$
 $\text{hidrometeoro} \subseteq \text{meteoro}$
 $\text{meteoro} \subseteq \text{eventos_meteorologicos}$
 $\text{precipitación} = n.R \text{ (decimal)}$
 $\exists (\text{precipitación Produce. efecto_detectable})$

La idea de aplicar estos modelos semánticos es poder transformar la información para incrementar los niveles de interoperabilidad según los principios de Linked Data y los estándares sugeridos por el World Wide Web Consortium (W3C).

Se desarrolló la plataforma LOD-GF que es un *framework* (plataforma de trabajo) basado en Pentaho Data Integration para brindar un entorno unificado de soporte en cada una de las fases de la metodología de publicación de Linked Open Data.

En este caso el desarrollo emplea plugins o complementos para el procesamiento especializado, tanto nativos del entorno, como desarrollados y que ofrecen un marco de trabajo gráfico flexible, de forma que los usuarios puedan convertir sus datos originarios de un amplio dominio de fuentes y formatos, a datos enlazados de calidad.

Comprende cinco fases: especificación, modelado, generación y publicación de los datos.

En el modelado debe identificarse, seleccionarse o generar vocabularios que permitan describir semánticamente los datos de las fuentes disponibles. En este punto interviene la ontología anteriormente desarrollada.

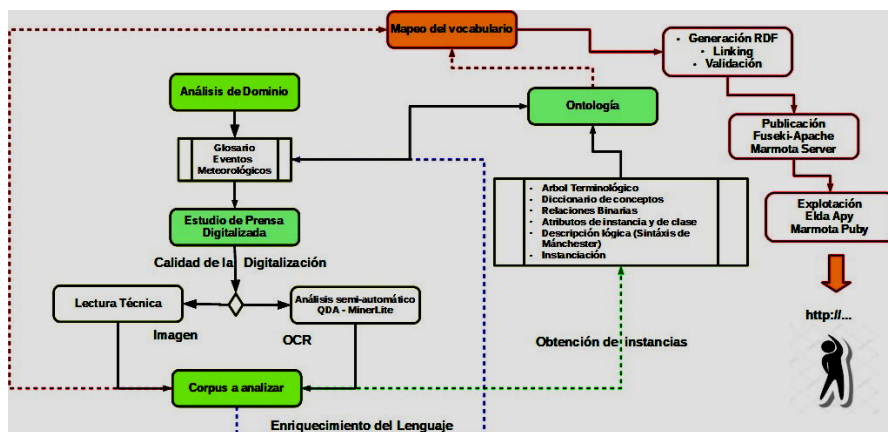


Figura 2. Procesos de desarrollo en el proyecto.

La Figura 2 muestra el proceso de trabajo del proyecto a partir del Análisis de Dominio. Es la etapa inicial del estudio de la prensa digitalizada con la obtención del vocabulario enriquecido para el desarrollo de la ontología, sus pasos y como ésta se inserta en el entorno de trabajo definido por la plataforma LOD-GF, hasta que el usuario recibe una respuesta del sistema a su consulta.

Visualizando resultados

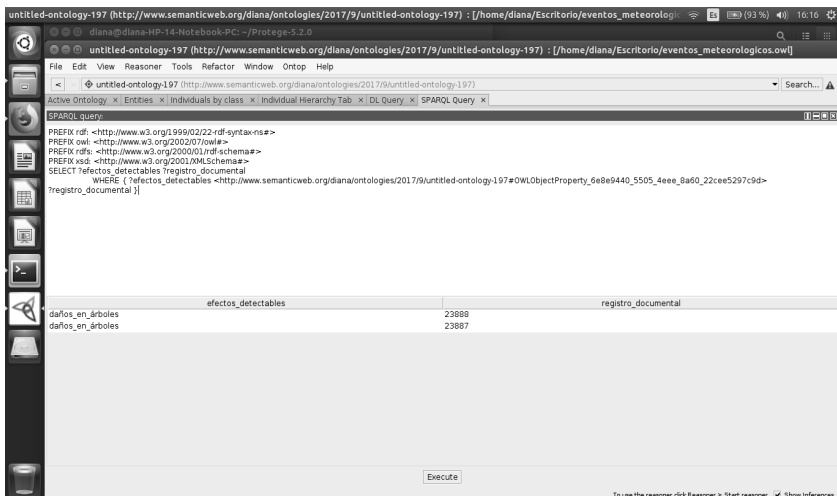
Para la publicación de los datos, se almacenan en un repositorio RDF o base de datos especialmente diseñada para el almacenamiento y recuperación de tripletas a través de consultas semánticas. Estas tripletas son entidades de datos compuestas de sujeto-predicado-objeto y disponen de un medio de acceso a los datos (SPARQL Endpoint).

En este caso es un *plugin* especializado conocido como Fuseki Loader, el cual permite configurar los parámetros básicos para el despliegue de las tripletas, es decir la información se rescata a través de consultas SPARQL en la plataforma de trabajo.

Los resultados pueden visualizarse también desde la propia ontología desarrollada en Protégé. Al desarrollar la clase “registro_documental”, cuyas instancias las conforman el corpus de prensa digitalizada con que se trabaja, establecido las relaciones binarias “documentado_en” y su relación inversa “documenta”, es posible consultar, empleando consultas SPARQL, por ejemplo, que documentos que registran eventos meteorológicos a partir de sus efectos detectables:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
SELECT ?efectos_detectables ?registro_documental
WHERE { ?efectos_detectables
<http://www.semanticweb.org/diana/ontologies/2017/9/untitled-ontology-197#
OWLObjectProperty_6e8e9440_5505_4ee8_8a60_22cee5297c9d>
?registro_documental }
```

La respuesta del sistema corresponde a las instancias ingresadas, y su resultado se observa en la copia de la descarga de pantalla de la figura 3. En ella se observa que en los documentos identificados por los números “23888” y “238887”, se registran daños en los árboles. Si recurrimos a la Escala de Bauford, esto corresponde a un temporal fuerte con vientos de 75 a 88 km/h.



The screenshot shows the Protégé SPARQL query interface. The query is as follows:

```
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
SELECT ?efectos_detectables ?registro_documental
WHERE { ?efectos_detectables
<http://www.semanticweb.org/diana/ontologies/2017/9/untitled-ontology-197#
OWLObjectProperty_6e8e9440_5505_4ee8_8a60_22cee5297c9d>
?registro_documental }
```

The results table is as follows:

| efectos_detectables | registro_documental |
|---------------------|---------------------|
| daños_en_árboles | 23888 |
| daños_en_árboles | 23887 |

Figura 3. Resultado de la consulta SPARQL a la ontología.

Conclusiones

El proyecto de investigación cuyo caso se presenta, mostró la fortaleza del trabajo multidisciplinar, al reunir especialistas de diferentes disciplinas: informática, ciencias de la información y geografía.

Rescatar información de la prensa digitalizada del siglo XIX, significó un gran desafío. Es necesario enfrentarse a documentos digitalizados a partir de originales con diferentes grados de deterioro físico y diversas condiciones de digitalización, como imagen, con software de reconocimiento de caracteres (OCR), baja resolución, lo que condujo a que no siempre se pudieran procesar los documentos en forma automática y fuera necesario recurrir al procesamiento automático y aún a la lectura técnica de los mismos.

Sería importante que las instituciones, al enfrentar un proyecto de esta envergadura, tuviesen en cuenta una estandarización en la digitalización, de forma que la calidad de su producto, permitiera un posterior procesamiento automático.

Otra dificultad se presentó con la evolución de los nombres geográficos, para solucionarlo se consultó diccionarios de nombres geográficos y *gazetteers* disponibles.

Los eventos meteorológicos que permite rescatar constituyen antecedentes históricos del cambio climático en América Latina. Nos es posible recuperar información que no estaba visible de las Bibliotecas Nacionales utilizando los beneficios de las Tecnologías de la Información y las Comunicaciones (TIC), con técnicas de recuperación de información.

La localización de la información se logra a través de consultas SPARQL en el entorno de trabajo desarrollado. El prototipo desarrollado para la consulta de artículos sobre eventos meteorológicos documentados en la prensa escrita digitalizada, es aplicable a otros documentos digitalizados y, desarrollando los modelos semánticos correspondientes, a infinidad de disciplinas, cuyos datos, por haber sido considerados menores en su momento, aún aguardan ser descubiertos en el volumen siempre creciente de datos volcados a la web.

Bibliografía

- Caldón, E.F.; Uribe, G.; López, D.M.; de Oliveira, J.P.M. y Krug Wives, L. (2010). *Mecanismos de anotación semántica de contenidos en plataformas de redes sociales*, CI. 2010 5(1), 90-99.
- Comesaña, D. (2015). “Modelo conceptual de información geográfica para la IDE – Uruguay”, tesis de maestría, Universidad de la República (Uruguay), Facultad de Información y Comunicación, disponible en: <<https://www.colibri.udelar.edu.uy/jspui/handle/123456789/5230>>.

- Fenómenos Meteorológicos (s.f.). Mendoza, Argentina: Programa Regional de Meteorología, disponible en: <<http://www.prmarg.org/fenomenos-meteorologicos>>.
- Feito Higuera, F. y Segura Sánchez, R. (2010). *Herramientas SIG 3D*. Virtual Archaeology Review, 1 (1), disponible en: <<https://polipapers.upv.es/index.php/var/issue/view/500>>.
- Gallegos, C. (2011). “Abordaje metodológico de prensa escrita: el semanario Cuba Libre”, *Latinoamérica*, núm. 53, julio-diciembre, México, disponible en: <http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-85742011000200006>.
- Hjorland, B. (2002). *Epistemology and Sociocognitive Perspective in Information Science*, JASIS, 53(4), 257-70.
- Mejías, A. y Arias, A. (1998). “La prensa del siglo XIX como medio de difusión de la literatura Hispanoamericana”, *Revista general de Información y Documentación*, 8(2), 241-257
- PNUD (2018). Objetivos del Desarrollo Sostenible. Objetivo 13: Acciones por el Clima. Disponible en: <http://www.undp.org/content/undp/es/home/sustainable-development-goals/goal-13-climate-action.html>
- Regillo, R. (2018). *Un vistazo al estado actual de la Web Semántica*. Disponible en: <https://www.beeva.com/beeva-view/innovacion/un-vistazo-al-estado-actual-de-la-web-semantica/>.
- Ribera Masgrau, L. (2004). *Los mapas de riesgos de inundaciones: representación de la vulnerabilidad y aportación de las innovaciones tecnológicas*, *Documents D'Anàlisi Geogràfica*, 43, 153-171. Disponible en <<http://www.raco.cat/index.php/DocumentsAnàlisi/article/viewFile/31812/31647>>.
- World Wide Web Consortium (2017). *Guía breve de Web Semántica*, disponible en <<https://www.w3c.es/Divulgacion/GuiasBreves/WebSemantica>>.

La calidad de las IDE desde el punto de vista de la interoperabilidad

Antonio F. Rodríguez*
Francisco Javier Ariza**

Recibido 02 de marzo de 2018, aceptado 29 de agosto de 2018

Resumen

Las Infraestructuras de Datos Espaciales (IDE) están reconocidas, bajo la etiqueta “Servicios y Sistemas de Información Geoespacial Integrada” como uno de los objetivos definidos en UN-GGIM (2018) para alcanzar los Objetivos 2030 de Desarrollo Sostenible. Sin embargo, varias fuentes indican que las IDE no están produciendo los resultados esperados. Para facilitar que se solucionen sus problemas y deficiencias, resulta especialmente adecuado profundizar en la descripción de la calidad de las IDE existentes. En este trabajo se intenta avanzar en esa línea desde un punto de vista basado en el concepto de interoperabilidad. En primer lugar, se repasan y comentan las principales metodologías formulada hasta ahora para describir y determinar la calidad de una IDE. Se propone una clasificación, meramente descriptiva, de los indicadores utilizados para ello: indicadores de acciones realizadas, de calidad de resultados, de uso, de impacto y de coste-beneficios. Y se describe un conjunto de indicadores de calidad de los resultados, es decir de los componentes de una IDE, para caracterizarlos siguiendo *el European Interoperability Framework* v2 europeo establecido en 2017. Finalmente se esbozan unas conclusiones, y se proponen líneas de mejora para optimizar la situación, muchas de las cuales suponen abrir campos de investigación muy necesarios.

Palabras clave: *calidad, interoperabilidad, calidad de recursos de una IDE.*

* Centro Nacional de Información Geográfica (CNIG), Instituto Geográfico Nacional (IGN) de España, Calle General Ibáñez de Ibero 3, 28003 Madrid, España, correo electrónico: afrodriguez@fomento.es

** Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén, Campus Las Lagunillas, s/n, 23071, Jaén, España, correo electrónico: fjariza@ujaen.es

Resumo

As Infraestruturas de Dados Espaciais (IDE) são reconhecidas como “Serviços e Sistemas de Informação Geoespacial Integrada” conforme um dos objetivos definidos no UN-GGIM (2018) para alcançar os Objetivos 2030 do Desenvolvimento Sustentável. Embora várias fontes indiquem que as IDE não estão produzindo os resultados esperados, para facilitar que se solucionem seus problemas e deficiências, resulta especialmente adequado se aprofundar a descrição da qualidade das IDE existentes. Neste trabalho se tenta avançar nessa linha desde o ponto de vista baseado no conceito de interoperabilidade. Em primeiro lugar, se repassam e comentam as principais metodologias formulada até agora para descrever e determinar a qualidade de uma IDE. Se propõe uma classificação, meramente descritiva, dos indicadores utilizados para ela: indicadores de ações realizadas, de qualidade de resultados, de uso, de impacto e de custo benefícios. E, se descreve um conjunto de indicadores de qualidade dos resultados, caracterizando os componentes de uma IDE seguindo o *European Interoperability Framework v2* estabelecido em 2017. Finalmente, se esboçam umas conclusões, e se propõem linhas de melhoria para otimizar a situação, muitas das quais se supõem abrir campos de investigação muito necessários.

Palavras chave: *qualidade, interoperabilidade, qualidade de recursos de uma IDE.*

Abstract

The Spatial Data Infrastructures (SDIs) are recognized, under the label of “Integrated Geospatial Information Systems and Services” as one of the goals defined in UN-GGIM (2018) in order to achieve the Sustainable Development Goals 2030. Nevertheless, several voices are saying that the SDIs are not yet producing the expected results. To facilitate the resolution of their problems and gaps, it would be very useful to go deep in the existing SDI quality descriptions. This paper tries to go forward in this line of work taking as a basis the concept of interoperability as a core concept in SDIs. First of all, a brief and commented analysis is provided of the methodologies proposed until now for describing SDI quality. A descriptive taxonomy of quality indicators is proposed: actions performed indicators, quality results indicators, use indicators, impact indicators and cost-benefit indicators. A set of quality results indicators (resources implemented quality indicators) is presented following the *European Interoperability Framework v2* established in 2017. Finally, a set of conclusions is outlined and some lines of work for improvement the situation are proposed, in most of cases with very necessary research areas associated.

Key words: *quality, interoperability, SDI quality.*

*Si bien sólo unos pocos son capaces de dar origen a una política,
todos nosotros somos capaces de juzgarla.*
Pericles (alrededor del 430 a. C.)

Introducción

Las Infraestructuras de Datos Espaciales (IDE) están reconocidas, bajo la etiqueta “Servicios y Sistemas de Información Geoespacial Integrada” como uno de los objetivos definidos en UN-GGIM (2018) para alcanzar los Objetivos 2030 de Desarrollo Sostenible. Sin embargo, viven un momento de desarrollo peculiar que, en cierto sentido, puede parecer paradójico. Desde sus primeras formulaciones en la década de los noventa (Orden Ejecutiva, 1994) experimentaron una primera etapa de desarrollo que condujo a la aparición temprana de herramientas tecnológicas suficientes como para iniciar su implantación a finales de los 90. La primera versión de la norma de metadatos estadounidense apareció en 1998 (FGDC, 1998) y la versión 1.0 del estándar *Web Map Service*, en el 2000 (OGC, 2000).

Sin embargo, a pesar de las dos décadas transcurridas y de que la mayoría de los países han abierto un geportal nacional de su IDE, muchos expertos expresan un cierto desaliento, que se podría llamar desaliento geográfico, porque no siempre se consigue rentabilizar la tecnología y las aplicaciones finales, basadas en servicios web, no “florece” todo lo que sería deseable.

Efectivamente, el informe titulado *Summary Report on the Status of implementation of the INSPIRE Directive in EU* (Comisión Europea, 2017b) concluye que “...se necesitan esfuerzos adicionales a nivel de la Unión Europea y de los estados miembros para salvar los huecos existentes en la implementación y recoger todos los beneficios de la Directiva”. Por otro lado, el lema de la Conferencia INSPIRE 2017, *Thinking out of the box* (“Una invitación a pensar desde una nueva perspectiva”), y el título de una de las sesiones, *What if...?* (“¿Y si...?”), parecen síntomas de una búsqueda urgente de soluciones y en un plano más técnico y concreto, el documento “*Spatial Data On the Web Best Practices*” (W3C, 2017), elaborado por un grupo de trabajo conjunto de expertos del World Wide Web Consortium (W3C) y del Open Geospatial Consortium (OGC), que trata de aconsejar sobre cómo aplicar mejor las tecnologías web en las IDE, constituye un intento explícito de encontrar esas soluciones.

Es una situación que recuerda lo que ocurría con los Sistemas de Información Geográfica (SIG) a principios de la década de los noventa: una tecnología que prometía mucho, pero que no acababa de alcanzar su madurez (Rodríguez, 1993), que llegaría una década después, en los años 2000, cuando ISO/TC 211 aportó las soluciones esperadas en forma de modelos, formatos, metadatos, calidad y especificaciones establecidos en la familia de normas ISO 19100 (Ariza y Rodríguez, 2008).

Esa posible inmadurez en el caso de las IDE, cuya complejidad resulta obvia si se considera su concepción como sistema de sistemas (Béjar, 2009), se debe en nuestra opinión no tanto a sus componentes más técnicos, ya que hay estándares, *software* libre disponible y herramientas más que suficientes para su implementación, como a sus aspectos organizativos. Si se tiene en cuenta que las IDE operan en un entorno globalizado y la amplitud y heterogeneidad de actores que deben coordinarse para su correcto funcionamiento, es lógico que factores como el apoyo político sostenido y al más alto nivel posible, la financiación, la sostenibilidad, rentabilidad, y sobre todo, su gobernanza y coordinación, resulten de gran importancia (Maganto *et al.*, 2012).

Muy probablemente haya que esperar una década más, como ocurrió con los SIG, para ver cómo las IDE se extienden y dan todos sus frutos. Sin embargo, mientras que los SIG que han ido apareciendo, han demostrado una notable flexibilidad al ir incorporando las nuevas tecnologías (Orientación a Objeto, computación en la nube, 3D...), las IDE no están incorporando con la misma facilidad las tecnologías de vanguardia (Datos enlazados, BIM, *Smart Cities*, *Big Data*, Realidad Virtual/Aumentada...) y la brecha entre la interoperabilidad en el laboratorio y en el sector productivo parece no cerrarse nunca.

En cualquier caso, el establecer un modelo de calidad IDE puede contribuir a monitorizar su progreso, identificar áreas de mejora, potenciar la sana competitividad entre proyectos y servir así de palanca para rentabilizar la tecnología.

Para considerar la calidad de una IDE, en este trabajo se va a ofrecer una panorámica de los métodos abordados hasta ahora, para tomar luego como punto de partida el modelo de interoperabilidad europeo. La interoperabilidad es el concepto clave en que se basan las IDE, tal y como se define en el *European Interoperability Framework v2* (EIFv2) (Comisión Europea, 2017). Es uno de los pilares básicos en uno de los ámbitos legislativos más complejos y desarrollados, la Unión Europea, que ha dado lugar a una IDE supranacional bajo la Directiva INSPIRE (2007/2/CE).

De esta forma, el objetivo de este trabajo es revisar de una manera crítica los conceptos actuales relativos a la calidad de una IDE, como aspecto clave para estimular y potenciar su desarrollo. Como consecuencia final, se propone profundizar en la caracterización de la calidad de todos los componentes de una IDE, para lo que se identifican líneas de investigación necesarias para poder progresar en esa línea de trabajo.

Calidad de una IDE

Existe una amplia variedad de estudios y documentación (Vancauwenberghe *et al.*, 2018, Kelm, 2017, Welle Donker y van Loenen, 2016 y Randolf Pérez *et al.*, 2015), en los que se aborda la caracterización de la calidad una IDE mediante la definición de un conjunto de indicadores. Hay que advertir que para establecer adecuadamente

un sistema de indicadores, de acuerdo a la norma UNE 66175:2003 (*Guía para la implantación de sistemas indicadores*) y a la ISO 19157: 2013 sobre calidad de datos geográficos, es necesario definir cuatro elementos para cada indicador: qué aspecto vamos a evaluar (por ejemplo, la exactitud posicional absoluta), qué medida vamos a emplear (por ejemplo, el Error Medio Cuadrático en distancia), qué metodología se va a aplicar (se toman cuatro puntos por km²...) y cómo se va a expresar el resultado final, por ejemplo, como un resultado numérico (3 m) o como un “pasa/no pasa” un umbral de calidad prefijado (si es menor o no que 3 m).

Fundamentalmente, todas las metodologías se basan en la definición de un conjunto de indicadores, como puede verse en la amplia variedad de estudios y trabajos recopilados en la obra *Multi-View Framework to Assess SDIs* (Crompvoets *et al.*, 2008) y en Morera *et al.* (2012), que se pueden agrupar a efectos meramente descriptivos en cinco clases de indicadores, que se expondrán a continuación con más detalle: indicadores de acciones realizadas, de calidad de resultados, de uso, de impacto y de coste-beneficio.

1. Indicadores de acciones realizadas

Son indicadores numéricos y objetivos que describen las acciones realizadas, como por ejemplo, el número de recursos implementados y sus características. Son útiles para fijar una línea base y poder monitorizar la evolución del esfuerzo de inversión realizado a lo largo del tiempo. Así son la mayoría de los indicadores definidos en el marco INSPIRE por medio de los llamados indicadores de seguimiento anual (Tabla 1) definidos en (Comisión Europea, 2009).

Tabla 1
Principales indicadores de acciones realizadas INSPIRE

| <i>Indicador</i> | <i>Definición</i> |
|-----------------------------------------------------------|------------------------------------------------------------------------------------------------|
| <i>Indicadores de conjuntos de datos espaciales (CDE)</i> | |
| DSv | Nº de CDE reportados |
| DSi1 | % medio de superficie nacional cubierta por los CDE |
| <i>Indicadores de servicios</i> | |
| SDs | Nº total de servicios de datos espaciales |
| NSv1, 2, 3, 4, 5 | Nº de servicios de búsqueda, visualización, descarga, transformación e invocación ¹ |

¹ Los servicios de invocación sirven para hacer llamadas a servicios no INSPIRE y transformar la respuesta al vuelo de manera que el resultado sean servicios conformes a INSPIRE.

Continuación Tabla 1

| <i>Indicador</i> | <i>Definición</i> |
|--------------------------------------------|---------------------------------------------------------------|
| NSi1.1 | % de CDE con metadatos cargados en un catálogo |
| NSi1.2 | % de servicios con metadatos cargados en un catálogo |
| NSi2 | % de CDE accesibles vía servicios de visualización y descarga |
| <i>Indicadores de existencia metadatos</i> | |
| MDi1.4 | % de CDE con metadatos |
| MDi1 | % de CDE y servicios con metadatos |

Son indicadores fáciles de determinar, simplemente inventariando los recursos implementados. Otros sistemas de indicadores de este tipo son, por ejemplo, los definidos por Eelderink y Steidler (Crompvoets *et al.*, 2008).

Tienen el gran inconveniente de que sirven para cuantificar el esfuerzo de inversión realizado, pero no si el esfuerzo va en la dirección adecuada, si los recursos implementados tienen la calidad necesaria ni si el impacto final es el deseado. Por lo tanto su utilidad depende de que se complementen con indicadores de los otros tipos.

Por otro lado, la interpretación de sus valores debe hacerse muy cuidadosamente, especialmente los que no se expresan mediante un tanto por ciento, ya que dependen de varias circunstancias. No resulta fácil comparar los valores de algunos de ellos entre unos países y otros, como por ejemplo el número de servicios, porque en algunos estados miembros los servicios pueden estar muy atomizados, por escalas e incluso por tipo de objeto geográfico, mientras que en otro puede ser que solo se reporte un servicio para la mejor resolución y para todos los datos. Algo parecido ocurre con el número de conjuntos de datos; hay países que reportan cualquier CDE de los temas INSPIRE, aunque sea de un solo municipio, mientras que otros solo reportan los CDE fundamentales para todo el país que mejor describen la realidad nacional. Para complicar más aún la situación, algunos estados han cambiado de criterio a lo largo del tiempo, con lo que resulta evidente la necesidad de unificar criterios y mantenerlos en el tiempo para que los valores y su evolución sean interpretables. Véanse Vandembroucke (2012) y *European Environmental Agency* (2014).

2. Indicadores de calidad de resultados

Se basan en la descripción de la calidad de los resultados de las acciones realizadas, en el caso de la IDE, esto se traduce en calidad de los recursos implementados: datos, metadatos y servicios. Los indicadores INSPIRE de seguimiento

anual que describen la conformidad con los Reglamentos europeos de conjuntos de datos, metadatos y servicios son indicadores de este tipo (Tabla 2).

Tabla 2
Principales indicadores de calidad de resultados INSPIRE

| <i>Indicador</i> | <i>Definición</i> |
|------------------|----------------------------------------------------|
| DSi2 | % de conjuntos de datos espaciales (CDE) conformes |
| NSi4 | % de servicios conformes |
| MDi2.4 | % de servicios con metadatos conformes |
| MDi2 | % de CDE y servicios con metadatos conformes |

En este tipo de indicadores es en el que se va a centrar este artículo, considerando la calidad de un amplio abanico de componentes de una IDE, más allá de datos, metadatos y servicios.

3. Indicadores de uso

Evalúan la utilización de los recursos de las IDE, no solo mediante indicadores cuantitativos de los volúmenes de consultas y tráfico de datos, sino también evaluando el número de usuarios, organizaciones, aplicaciones y recursos de valor añadido que se generan. En esa línea están los indicadores definidos en la Directiva INSPIRE que se recogen en la Tabla 3 y los incluidos en otros estudios similares, como Crompvoets (2008).

Tabla 3
Principales indicadores INSPIRE de uso

| <i>Indicador</i> | <i>Definición</i> |
|------------------|--------------------------------------------------------------------------|
| NSi3 | Nº medio de peticiones anuales por servicio |
| NSi3.1 | Nº medio de peticiones anuales por servicio de búsqueda |
| NSi3.2 | Nº medio de peticiones anuales por servicio de visualización |
| NSi3.3 | Nº medio de peticiones anuales por servicio de descarga |
| NSi3.4 | Nº medio de peticiones anuales por servicio de transformación (de datos) |
| NSi3.5 | Nº medio de peticiones anuales por servicio de invocación |

De nuevo, es necesario interpretar adecuadamente los valores de este tipo de indicadores y su evolución. Por ejemplo, la implementación de unos servicios de descarga excelentes puede disminuir el nº de peticiones de visualización y viceversa. Por otro lado, los *Web Map Tiles Service* (WMTS) exigen un alto número de peticiones frente a los *Web Map Service* (WMS) para un mismo caso de uso. De cualquier manera, este tipo de indicadores es aconsejable para analizar la evolución en el tiempo del uso de los servicios implementados.

4. Indicadores de impacto

Tratan de evaluar la eficacia o grado en el que una IDE alcanza sus últimos objetivos, en función del resultado o impacto esperado por los usuarios o por la sociedad (Giff, 2008). En esta categoría entrarían todos los indicadores que describen de una manera u otra la satisfacción del usuario. En este caso, la dificultad estriba en caracterizar ese impacto esperado, lo que puede hacerse a bajo nivel de abstracción eligiendo un resultado concreto, por ejemplo el aumento de la cifra de negocio global del sector de la Información Geográfica (IG) o la disminución de costes de los proyectos que explotan IG o a alto nivel de abstracción, como la aportación de las IDE a la consecución de los objetivos de la Agenda 2030 de Naciones Unidas.

Su definición depende de los objetivos y planteamientos estratégicos que se tengan y parecen, en cierta medida, más relevantes porque describen las consecuencias y efectos ocasionados por la implementación de una IDE. Al mismo tiempo, aunque se definan como algo tan simple como la satisfacción de usuario, pueden en parte reflejar la influencia de otros factores que confluyen, como las expectativas previas, de manera que resulta difícil aislar el impacto de la implementación de una IDE de otras circunstancias económicas, sociales y de todo tipo que pueden influir en sus efectos.

5. Indicadores de coste-beneficio

Son indicadores que intentan reflejar la eficiencia de los recursos implementados, es decir el resultado de un balance coste-beneficio, que a menudo es lo más convincente y significativo para tomadores de decisiones y responsables de efectuar inversiones públicas. Es ésta una línea de trabajo difícil y siempre con resultados incompletos debido a que gran parte de los beneficios de las IDE son intangibles y difusos, que muchas veces se producen a muy largo plazo y que resulta imposible comparar en la práctica cómo evoluciona económicamente un mismo país sin y con IDE. Sin embargo, hay un buen número de estudios parciales que arrojan tasas de coste-beneficio considerablemente altas, como Morena *et al.* (2012), Balciunas *et al.* (2016) y de Jong (2016).

Hay que hacer notar que, si bien los indicadores de acciones realizadas y uso tienen la ventaja de ser más fácilmente definibles como parámetros cuantitativos acompañados de métodos objetivos de determinación, son menos interesantes, en cuanto a su significado, que los indicadores de impacto y los de coste-beneficio. Efectivamente, se puede tener un amplio abanico de servicios web implementados, muy usados, por ejemplo como cartografía fundamental de fondo, pero que aportan pocos servicios de valor añadido y escasos casos de uso con un impacto real y un beneficio notable. Por el contrario, los indicadores de impacto suelen dar lugar a parámetros cualitativos de determinación difícilmente objetivable, mientras que los de coste-beneficio aunque son obviamente cuantificables, son de muy difícil evaluación, dado que la mayoría de los beneficios de las IDE son intangibles y se dan a largo plazo. Sin embargo, estos dos últimos tipos parecen más interesantes porque describen los efectos y consecuencias prácticas de la implementación de una IDE.

Los indicadores de segundo tipo, los de calidad de resultados de las acciones ejecutadas, en suma, los de calidad de los componentes de una IDE, precisan de cierta dedicación de recursos, por contemplar aspectos técnicos a veces sofisticados, sin embargo tienen la ventaja de que pueden relacionarse, como luego se verá, con la interoperabilidad, fundamento y concepto clave en las IDE.

Su determinación es más fácilmente objetivable que los indicadores de impacto y de coste-beneficio. Por otro lado, aunque a primera vista puedan parecer poco significativos, como los indicadores de acciones realizadas o los de uso, si se define un abanico suficiente amplio y comprensivo de componentes de una IDE, incluyendo tanto aspectos técnicos como organizativos, su calidad puede ser clave y tanto el impacto como el balance coste-beneficio resultarán ser consecuencias de esa calidad.

Tienen además la ventaja de que, al contrario que los indicadores de los otros cuatro tipos, no dependen tanto de la variedad de planteamientos y modelos que se pueden llevar a cabo para la implementación de la IDE, lo que evita la variedad de experiencias que dificulta la convergencia y consenso en un planteamiento común que permita comparar la calidad de varias IDE diferentes. Hay que pensar que para afinar y validar objetivamente un sistema de indicadores que no sea del segundo tipo, habría que mantenerlo en el tiempo y repetir su ejecución para poder extraer conclusiones, lo que supondría inversiones y esfuerzos considerables.

Por lo tanto, los indicadores del segundo tipo son los que permiten comparar más fácilmente IDE diferentes.

Una alternativa: la interoperabilidad como indicador de calidad

En lo que sigue se esboza un planteamiento que parte de un punto de vista alternativo basado en la interoperabilidad, una de las ideas clave de las IDE, que identifica la calidad de una IDE con ese concepto y por lo tanto, con la calidad de sus compo-

entes, expresada mediante indicadores del tipo 2, lo que permite una caracterización más objetiva y comparable, basada en aspectos técnicos que complementan a otros más organizativos.

Para ello, se tomará como base el *European Interoperability Framework v2* (Comisión Europea, 2017) (EIFv2), tal y como se plantea en Rodríguez *et al.* (2017). Ese modelo de interoperabilidad considera cuatro niveles:

- Interoperabilidad legal, que se encarga de asegurar qué organizaciones de las que trabajan en diferentes marcos, políticas y estrategias legales pueden interoperar sin encontrar barreras legislativas, independientemente de que haya o no un Marco Legal de ámbito superior que las acoja.
- Interoperabilidad organizacional, que se ocupa de la coordinación de procesos, responsabilidades y expectativas de las organizaciones para que puedan interoperar sin disfuncionalidades.
- Interoperabilidad semántica, que en realidad incluye (a) la interoperabilidad semántica propiamente dicha, que se ocupa del significado de lo que se transmite en cada momento, y (b) la interoperabilidad sintáctica, que se encarga de la interoperabilidad entre formatos físicos de datos y metadatos.
- Interoperabilidad técnica, que cubre la interconexión de la infraestructura de comunicaciones subyacente y de aplicaciones, es decir servicios web. Es la interoperabilidad más ligada a las Tecnologías de la Información y las Comunicaciones y a menudo es el único nivel de interoperabilidad que se considera en algunos estudios como en Nativi *et al.* (2012).

Cada uno de estos niveles conlleva sus propios aspectos de la calidad, conceptualmente diferentes a la hora de diseñar un modelo de calidad de una IDE. Teniendo en cuenta una de las definiciones clásicas de IDE (GSDI, 2012): “Tecnologías, políticas, estándares y recursos humanos necesarios para adquirir, procesar, almacenar, distribuir y mejorar la utilización de datos geoespaciales” (traducción propia), se van a generalizar los niveles de interoperabilidad del EIFv2 para tener en cuenta todos los componentes de una IDE en un sentido amplio e incluir así todos los posibles factores.

Interoperabilidad legal

Considerando que depende fundamentalmente de la calidad del Marco legal definido en el ámbito en el que interoperan los componentes de una IDE, aparecen situaciones problemáticas derivadas a menudo de la interacción de dos dominios tan heterogéneos en principios y métodos como son el legal y el tecnológico. Por ejemplo, los procesos de definición y reforma de disposiciones legales son largos, pesados y, a menudo, de resultado incierto comparados con los procesos de generación

de especificaciones técnicas. Otra diferencia radica en que el ámbito legal está más fragmentado por países que el tecnológico. A continuación, se enumeran las principales situaciones problemáticas que se pueden presentar en este nivel:

- Marco legal incompleto. Aparece cuando el marco legislativo de una IDE no contempla y regula claramente todos los aspectos que sólo pueden ser definidos legalmente y resultan imprescindibles para el buen funcionamiento de una IDE. Esa falta de completitud puede deberse a que los textos legales
 - sean demasiado laxos,
 - estén desenfocados y se concentren en otros temas relacionados con las IDE, pero que no resulten centrales,
 - simplemente, no existan,con lo que toda la función regulatoria recae en documentos técnicos de cumplimiento voluntario.
- Marco legal demasiado rígido. En el difícil equilibrio entre el conocimiento legal y el técnico, este riesgo aparece a menudo cuando el legislador invade campos que evolucionan con una rapidez muy grande en comparación con los tiempos necesarios para modificar una disposición legal, lo cual depende parcialmente de la mayor o menor agilidad que tenga el poder legislativo en cada país o ámbito regulatorio. Sin embargo, parece razonable que, dada la brecha temporal existente entre la evolución de la tecnología y la de las disposiciones legales, el Marco legal contemple tan sólo aspectos conceptuales, de principios muy generales, y puramente competenciales y organizativos, sin entrar nunca en aspectos relacionados con las TIC y las interoperabilidades semántica y técnica, que deben estar regulados por documentos técnicos mucho más dinámicos.

Un ejemplo de la situación descrita es la que se da en el marco de la implementación de la Directiva europea INSPIRE. Por un lado, se ha aprobado un conjunto de Reglamentos europeos, de obligado cumplimiento en todo el territorio de la UE, que descienden a especificar aspectos técnicos acerca de datos, metadatos y servicios. Por otro lado, existen una serie de Directrices técnicas que ayudan a cumplir los anteriores.

Los Reglamentos son excesivamente rígidos, no admiten una adaptación suficientemente ágil a una tecnología en continua evolución. Hay que mencionar que desde la aparición de la Web 2.0 en el año 2004, se habla de la versión beta perpetua, como filosofía de desarrollo flexible y continuamente adaptable² en oposición a la liberación de versiones del *software* cerradas y completas.

Por otro lado, las Directrices, aunque son fácilmente adaptables, no son obligatorias, con lo que tampoco está asegurada completamente la interoperabilidad.

² <https://es.wikipedia.org/wiki/Web_2.0>.

Quizás la solución ideal sería un Marco legal que estableciera la obligatoriedad de cumplir las últimas versiones en cada momento de un conjunto de especificaciones técnicas publicadas en la web y mantenidas por un organismo competente por ley para mantenerlas al día y que funcionase de manera abierta para todas las partes interesadas y tomase decisiones por consenso o por votación, a imagen y semejanza de las organizaciones de estandarización, como el OGC o el W3C.

- Contenido legal de baja calidad. Formalmente los textos legales pueden ser farragosos, ininteligibles, contradictorios, faltos de lógica, innecesariamente complejos y complicados, incluso lingüísticamente pobres y deficientes. Aspectos todos ellos que pueden dar lugar a interpretaciones erróneas o divergentes y pérdidas de tiempo invertido en encontrar la interpretación correcta de la ley. Aparece aquí la dificultad para definir parámetros de calidad y métodos de determinación que describan cuantitativamente la calidad del contenido del Marco legal.
- Mala accesibilidad del Marco legal. En ocasiones el Marco legal de una IDE puede ser fragmentario, carecer de textos consolidados o resultar poco accesible, circunstancia esta última que afecta aproximadamente el 10 % de los países³.

La fragmentación del Marco legal de una IDE podría cuantificarse mediante el número de disposiciones legales que es necesario consultar para implementar un nodo IDE completo. En el caso de la IDE de la Unión Europea (INSPIRE), el Marco legal específico está formado en el 2018 por una Directiva Europea y trece Reglamentos europeos, lo que hace un total de catorce disposiciones legales.

- Falta de garantías de aplicabilidad del Marco legal. En ocasiones se dispone de un Marco legal adecuado y completo, pero no se dispone de herramientas que permitan hacer efectivo su cumplimiento y se depende de la buena voluntad de los integrantes de la comunidad IDE para asegurarlo. Ese factor depende mucho de aspectos culturales, pero siempre es necesario disponer de algún tipo de modelo de evaluación y régimen sancionador para las organizaciones que no cumplan el Marco legal, que actúe como disuasión e incentive su cumplimiento riguroso y diligente.

Después de todo lo dicho se podría afirmar que de manera resumida las siguientes directrices podrían definir un Marco legal de calidad que facilite la interoperabilidad:

- Un Marco legal puramente teórico, organizativo y competencial, que se limite a establecer los principios y conceptos básicos (interoperabilidad, estándares abiertos, datos y servicios abiertos, servicio, metadatos, geoportal, etc.), asig-

³ <<https://index.okfn.org/dataset/law/>>.

nando una serie de competencias (coordinación de la IDE, órgano ejecutivo, procedimientos de participación y toma de decisiones) y una arquitectura territorial y temática (basada en nodos de referencia).

- Un Marco legal que no especifique aspectos técnicos, sino que establezca un conjunto de especificaciones técnicas de obligado cumplimiento, definidas por una organización creada *ad hoc*, de carácter abierto, democrático y participativo.
- Un conjunto de procedimientos técnicos claros y transparentes relacionados con su definición, mantenimiento y actualización.
- Un conjunto de mecanismos claros y razonables para asegurar su cumplimiento (modelo de evaluación y régimen sancionador).

No se conocen mecanismos para verificar si el Marco legal de una IDE es suficientemente flexible o no para su desarrollo. Una vía que parece factible, consistiría en analizar si es posible de una manera fácil, no traumática y en un tiempo aceptable su completa adaptación a una tecnología completamente nueva y diferente, como la de los datos enlazados (*linked data*).

La lista mencionada de los aspectos individuales de la calidad de un Marco legal mencionados, no se ha incluido como una relación exhaustiva y completa, sino como una primera aproximación sobre la que es necesario investigar y profundizar. Entre ellos los hay que parecen de difícil cuantificación, de modo que a lo sumo, permiten estimaciones cualitativas teñidas siempre de cierta subjetividad, por lo que parece que sería necesaria también la investigación en este campo. En cualquier caso, si no se progresa en esta dirección suficientemente, es muy probable que las IDE existentes se resientan, ya que el Marco legal es actualmente uno de sus puntos débiles, tal y como evidencian la variedad y heterogeneidad existente de este componente de un país a otro y la falta de criterios establecidos sobre su calidad. La existencia del Marco legal INSPIRE y de los logros y beneficios que ha supuesto, en una comunidad tan compleja lingüística y culturalmente como la europea, constituye un ejemplo inspirador para alcanzar el establecimiento de un Marco legal o recomendación para Latinoamérica, donde instituciones supranacionales como el IPGH ya han alcanzado resultados muy positivos en proyectos similares organizativamente, como el Proyecto SIRGAS,⁴ el Mapa Integrado de América Central (Norori Solís *et al.*, 2013) y el Mapa Integrado Andino del Norte (Martín *et al.*, 2017).

⁴ <http://www.sirgas.org/es/>

Interoperabilidad organizacional

Como ya se ha mencionado, la interoperabilidad organizacional se ocupa de la coordinación de procesos, responsabilidad y expectativas de las organizaciones para que puedan interoperar sin disfuncionalidades. El riesgo de falta de interoperabilidad organizacional en una IDE incluye todos los aspectos no estrictamente técnicos relacionado con la organización, coordinación y liderazgo de una IDE. Son aspectos comunes con otros proyectos en los que interviene una red de actores independientes que deben colaborar activamente para el éxito del proyecto. Los aspectos a considerar se pueden agrupar en tres apartados:

1. Aspectos organizativos estructurales
2. Aspectos operativos organizacionales
3. Aspectos técnico-organizativos

1. Aspectos organizativos estructurales

- Apoyo político. El primer aspecto importante a considerar es el apoyo político al más alto nivel posible. La existencia de ese apoyo claro, sostenido y efectivo a la implantación de tecnologías IDE resulta clave, puesto que de él dependen la dedicación de todo tipo de recursos, humanos, inversión en formación, equipamiento, etcétera, y la existencia de una presión externa que demanda y espera resultados relevantes y útiles en un plazo de tiempo razonable.
- Existencia de un *staff*. El siguiente aspecto a considerar es la existencia de un *staff* sostenible y estable en el tiempo, que no tiene porqué ser muy numeroso, de expertos dedicados a la coordinación y liderazgo de la IDE a tiempo completo o al menos como responsabilidad principal. Eso hace posible la planificación, el diseño de la arquitectura global y la toma de decisiones de alto nivel que luego se contrastan con los resultados y permiten aprender de la experiencia. Ese “núcleo duro” de liderazgo, debe estar rodeado de un comité en el que estén representados los responsables técnicos a alto nivel de los nodos que componen la IDE a fin de que puedan consensuar todas las decisiones a llevar a la práctica.
- Comunidad de actores. Como consecuencia del aspecto anterior, debe existir una comunidad de actores que sostiene la IDE equilibrada y participativa, formada por organizaciones públicas, empresas privadas, universidades y usuarios de la infraestructura, con un espíritu de colaboración amplio y asegurado. La colaboración debe ser ágil y exenta de burocracias entre los distintos actores.

— Dentro de esa comunidad de actores debe existir un conjunto de Grupos de Trabajo, denominados con ése o con otro nombre, que aborden y resuelvan cuantas tareas técnicas se identifiquen como necesarias para el buen funcionamiento de la IDE en su conjunto.

- Calidad de usuarios: Por último, habría que incluir aquí la calidad de los usuarios de la IDE, como actores con un papel activo de una IDE, teniendo en cuenta su implicación, la retroalimentación que devuelven, su participación activa en la toma de decisiones, incluso en procesos de *crowdsourcing*.

2. Aspectos operativos organizacionales

No solo es necesario considerar la calidad de una serie de estructuras e instancias organizativas, como se ha abordado en el punto anterior, teniendo en cuenta su composición, funcionamiento régimen de reuniones, representatividad y otros factores similares, sino también la calidad de los resultados que producen, lo que asegura la robustez del modelo de calidad al incorporar todas las perspectivas posibles, por ejemplo:

- La planificación de la progresiva implementación de la IDE, su evolución técnica y su aplicación a todos los sectores de actividad.
- El Plan de Medios o plan de sostenibilidad económica que asegure la financiación necesaria para que la IDE sea sostenible en el tiempo.
- La política de estándares, que engloba aspecto como:
 - los estándares elegidos,
 - la política de versiones y
 - las medidas para verificar y conseguir que la política de estándares establecida se cumpla.
- Las recomendaciones de armonización, que van un paso más allá de los estándares de interoperabilidad seleccionados, eligiendo algunas opciones que facilitan la interoperabilidad real y práctica, como los Sistemas de Referencia de Coordenadas, un perfil de metadatos, sistemas de teselado, escalas, palabras clave, etcétera.
- El Plan de Formación y, en general, la política de recursos humanos, considerada a nivel global de la IDE.
- Las medidas para compartir *software* y desarrollos cuando sea posible. En un mundo en el que el modelo de producción basado en el *software* libre presenta ventajas claras en cuanto a reutilización del código, las aplicaciones, compartición de las inversiones y desarrollo, escalabilidad sin coste de licencias, parece lógico planificar y coordinar el desarrollo colaborativo de soluciones.

- Los mecanismos de difusión de la información, tanto de arriba hacia abajo (información, noticias, *blogs*, redes sociales, *feeds*...) como de abajo hacia arriba (retroalimentación de usuarios y técnicos).
- Un aspecto clave en el funcionamiento de la comunidad es la eficacia con la que los proveedores de servicios capturan e incorporan la retroalimentación de sus usuarios para identificar sus necesidades y requerimientos.

3. Aspectos técnico-organizativos

Uno de los aspectos de la calidad de una IDE que entra dentro de lo organizativo, pero que es algo a la vez técnico y práctico, es la conexión y visibilidad de los nodos IDE, con una coordinación que evite la existencia de “zonas de sombra” temáticas o geográficas en las que existan recursos IDE (geoportales, datos, metadatos y servicios) que no tienen visibilidad en los Nodos IDE de Referencia de nivel superior.

Otro aspecto importante del mismo tipo es la interoperabilidad de los componentes esenciales que permiten que la IDE se comporte globalmente como un sistema único como: la navegabilidad siguiendo la red de nodos de la IDE, la coordinación de catálogos (*harvesting*), la arquitectura lógica de componentes, la política de Identificadores Persistentes (PID), o las medidas de verificación de la conformidad (validadores).

Por último, hay recursos básicos (servicios de catálogo, registros, *codelists*, sistemas de identificadores únicos, etc.), que resultan clave para que la arquitectura de la IDE pueda funcionar correctamente en la práctica.

Mención especial merece la calidad de las políticas de datos abiertos existentes. El informe (EEA-JRC, 2014) concluyó que existe una clara sinergia y correlación entre los datos abiertos y el grado de desarrollo de las IDE en un país o región. Los países europeos en los que hay datos más abiertos son los que más lejos están llegando en la implementación de INSPIRE y viceversa. Para evaluar la calidad de este aspecto pueden establecerse varios escalones de calidad creciente relacionados con la política de datos:

- Tener adoptada una política de datos de manera clara, consciente y responsable. Eso evitaría las vueltas atrás en la liberación de datos, que suelen ser mal aceptadas por los usuarios y crean inseguridad ante la sostenibilidad de las condiciones de uso.
- Mantenerla en el tiempo y adoptar medidas que garanticen su sostenibilidad, como consecuencia del punto anterior.
- Publicitarla adecuadamente, idealmente debería ser imposible que un usuario descargase un conjunto de datos sin que fuera informado de la licencia y condiciones de uso correspondientes.

- Utilizar licencias implícitas, cuyo mero uso implica su aceptación, como ocurre habitualmente en la mayoría de bienes que adquirimos. Eso facilita los trámites para el uso de los datos y elimina una barrera tan artificial como que el usuario tenga que estar de acuerdo en las condiciones de uso.
- Utilizar licencias tipo o estándar, que garanticen la interoperabilidad de licencias. El problema que genera el hecho de que los productores de datos utilicen licencias propias, descritas en un texto *ad hoc*, es que, si un usuario mezcla dos conjuntos de datos para generar una obra derivada y cada uno tiene su propia licencia no estándar, es en general extraordinariamente difícil averiguar cómo se mezclan las condiciones de uso de ambas licencias.
- Definir los datos como datos abiertos, en el sentido que define la *Open Knowledge Internacional* (2015), la definición más completa y nítida que conocemos. Eso incluye conceptos clave como no discriminar a ningún grupo de usuarios y una eliminación radical de todo tipo de barreras técnicas, legales y económicas para el uso de los datos.

Recientemente ha aparecido un estudio interesante que extiende el concepto de apertura y propone un conjunto de catorce indicadores, con tres valores de calidad (baja, media y alta) para caracterizar la apertura de una IDE en su conjunto:

1. Estrategia y visión de datos abiertos
2. Política de datos abiertos
3. Proceso de toma de decisiones sobre política de datos abierto a actores no gubernamentales
4. Apertura a incorporar datos no gubernamentales
5. Encontrabilidad de datos en buscadores estándar
6. Difusión de datos en los portales nacionales de la IDE y de datos abiertos
7. Multilingüismo de los metadatos
8. Disponibilidad de datos sin registro de usuarios
9. Gratuidad de los datos
10. Publicación de los datos mediante servicios de visualización y descarga
11. Licencias de datos abiertas e interoperables
12. Nivel de interoperabilidad de datos
13. Casos de uso no gubernamentales
14. Estudios de beneficios de los datos abiertos

El planteamiento está demasiado centrado en los datos, cuando en realidad las IDE están orientadas a servicios; mezcla en la descripción de la apertura de una IDE indicadores de coste-beneficios, que parecen demasiado heterogéneos y pertenecientes a una categoría diferente, y no considera algunos aspectos (apertura de ser-

vicios, apertura a la Información Geográfica Voluntaria, servicios de procesamiento, nomenclátor, sensores, etc.) que parecen también importantes. Por último, la evaluación de indicadores tan generales adolece de cierta subjetividad, problema que suele presentarse cuando se establecen indicadores cualitativos y no directamente medibles (Vancauberghe *et al.*, 2018).

La calidad de los aspectos relacionados con la interoperabilidad organizacional puede ser descrita con uno o más elementos de calidad como los mencionados, todos ellos difíciles de cuantificar, para los que habría que definir un sistema de indicadores especificando con precisión, como ya hemos dicho, el aspecto a evaluar, el indicador elegido y el método de medida, tal y como establecen las normas en ese campo, como la mencionada UNE 66175:2003 (AENOR, 2003).

Es necesario realizar más investigación basada en la propuesta de modelos de indicadores, basados en estos aspectos y su aplicación, para poder realizar estudios comparativos y poder progresar en la selección de los esquemas más eficaces y descriptivos.

Interoperabilidad semántica

Como ya se ha indicado, la interoperabilidad semántica incluye el aspecto sintáctico y el puramente semántico. En este subapartado se tratarán ambos.

Si bien los aspectos de interoperabilidad legal y organizacional se pueden aplicar con propiedad en el ámbito completo de una IDE, los que se van a considerar bajo el aspecto de interoperabilidad semántica pueden tenerse en cuenta para una IDE, tanto en su totalidad, como para cada uno de los nodos que la conforman, e incluso para cada recurso concreto en particular. La esencia de la interoperabilidad semántica de datos radica en que el destinatario de la información en un intercambio de datos, los pueda interpretar de manera correcta, de modo que obtenga la misma información que el productor de datos quiso representar en ellos (Comisión Europea, 2017). Como ya se ha indicado, eso depende esencialmente de dos factores: que los datos se interpreten correctamente para obtener la información deseada, con el significado pretendido, lo cual se llama interoperabilidad de significado, y que la sintaxis de formatos sea correcta, que llamaremos como es habitual, interoperabilidad sintáctica.

Interoperabilidad de sintáctica

La interoperabilidad sintáctica está asociada a la ausencia de problemas y deficiencias en los formatos en los que se codifican los datos. Estos aspectos pueden dividirse en: a) aquellos relativos a la falta de eficacia del formato, derivados de una mala elección de formato, y b) aquellos que ya están perfectamente caracterizados y

descritos en la consistencia de formato tal y como se define en la norma de calidad ISO 19157 (ISO, 2013).

La calidad del formato elegido no es fácil de cuantificar; comprende, entre otros, aspectos como: a) que no se trate de un estándar abierto; b) que conlleve pérdidas de información; c) que sea poco eficaz debido normalmente a generar ficheros demasiado voluminosos; d) que no esté bien definido, ya sea por falta de documentación o por admitir variantes que exigen un acuerdo entre el emisor y el receptor de la información.

Interoperabilidad de significado

La correcta interpretación de unos datos transferidos depende esencialmente de la adecuada comprensión del modelo de realidad asociado, ya sea implícita o explícitamente. Por lo tanto, habría que considerar los riesgos derivados de: una deficiente calidad del modelo de aplicación, deficiencias del catálogo de objetos geográficos y de algo que suele olvidarse al gestionar la información geográfica vectorial, que es la deficiente selección de fenómenos del mundo real en función de la resolución:

- Deficiencias en el modelado. El Esquema de Aplicación es el componente más importante para transformar los datos en información, es decir, para interpretar los datos y dotarlos de significado. Puesto que el Esquema de Aplicación es el resultado de un proceso de selección de lo que resulta de interés a partir de lo real, de abstracción posterior y formalización en un lenguaje formalizado, como el *Unified Modelling Language* (UML), su calidad resulta difícil de describir cuantitativamente. Sin embargo, es posible identificar deficiencias de un esquema, como: clases sobreabundantes, redundancias, relaciones mal establecidas, complejidad excesiva, errores en las clases seleccionadas, las relaciones y los atributos, errores de sintaxis en UML, falta de detalle en el esquema, mala elección de los nombres y tipos de datos erróneos.

Los errores de modelado dificultan la interoperabilidad, generan problemas en la gestión y mantenimiento de los datos, dificultan la comprensión del problema y en general, son deficiencias profundas que actúan como fuentes de todo tipo de problemas (Muller, 1997).

Un punto interesante a tener en cuenta aquí es la más que probable posibilidad de superar estas deficiencias pasando a otra etapa de modelado de los datos basada en las Ontologías. Como técnica de modelado más avanzada y en la que se incorporan aspectos conceptuales de Inteligencia Artificial, debe necesariamente mejorar la fidelidad y coherencia del modelo producido. En ese caso, una línea de investigación interesante es la que consistente en la caracterización y descripción de la calidad de una Ontología (García *et al.*, 2010), teniendo en cuenta las características particulares de las Ontologías geográficas.

- Deficiencias en el catálogo de objetos. Da lugar a aspectos de la calidad que están un nivel de abstracción por encima de la exactitud temática y tienen que ver con la definición del catálogo de objetos geográficos. Incluye errores como: catálogo incompleto, solapes y huecos entre los objetos definidos, redundancias, errores en la estructura jerárquica de los tipos de objeto, errores formales y semántica (nombres y definiciones) poco clara y confusa, entre otros.
- Deficiencias en la selección por resolución. Una fuente de error que se suele olvidar, o al menos no tenerse suficientemente en cuenta. Cuando un técnico (*surveyor*) sale al campo a hacer un levantamiento a una resolución determinada sabe perfectamente qué fenómenos del mundo real debe recoger y cuáles debe obviar, en función de su tamaño real y su importancia. La misma situación se presenta al restituir fotogramétricamente, al digitalizar datos sobre una ortofoto y en general, en todos los procesos interactivos de adquisición de datos, como por ejemplo, la generalización. Incluye errores de selección por exceso, por defecto y falta de homogeneidad al aplicar los criterios de selección.

También habría que incluir aquí los aspectos geométricos de la resolución, como superficie mínima a representar, distancia mínima entre vértices, etc., susceptibles de ser descritos en la consistencia geométrica, y la resolución métrica de las coordenadas y el tiempo, incluidas en la consistencia de formato.

Es importante no olvidar estos aspectos, en realidad clásicos y de etiología bien conocida en los entornos digitales, en los que es fácil perder la noción de lo que significa trabajar en un modelo de una resolución determinada.

- Falta de Adaptabilidad Lingüística y Cultural. En este apartado es donde hay que incluir todos los aspectos relacionados con la Adaptabilidad Lingüística y Cultural (CLA por sus siglas en inglés, *Cultural and Linguistic Adaptability*), tal y como se define en ISO/IEC (2005): capacidad de un producto para tener en cuenta las necesidades usuales de cualquier grupo de usuarios y para ser internacionalizado manteniendo sus propiedades de portabilidad e interoperabilidad, es decir para ser adaptado a las características especiales de las culturas y lenguas naturales de una región geográfica dada, incluyendo sus reglas de uso. Es decir, multilingüismo⁵ y multiculturalismo. La CLA es un aspecto que puede evaluarse tanto para una IDE completa, como para cada uno de sus recursos: geoportales, clientes, servicios, conjuntos de datos, metadatos, etc.

⁵ No es un aspecto ocioso ni marginal en absoluto, dado que en realidad en contra de lo que puede parecer a primera vista, el multilingüismo es la situación más frecuente en un país y el monolingüismo, la excepción. En el mundo hay unas 7,000 lenguas y menos de 200 países; un caso extremo es el la República de Sudáfrica, con 11 idiomas oficiales. Por un lado, es lógico que las interfaces contemplen todos los idiomas oficiales del país más el inglés, pero al hablar de conjuntos de datos, hay que tener en cuenta que los nombres geográficos están en todos los idiomas, dialectos y variantes en uso en un país.

La gestión del multilingüismo de interfaces de usuario, datos y metadatos es algo que puede solucionarse a nivel de aplicación, pero si pensamos en servicios web, tenemos que ser conscientes de que los servicios OGC no son multilingües.

A pesar de algunos intentos de considerar estos aspectos, como la introducción en las normas ISO 19100 que modelan datos y metadatos, el concepto de *locale* (un atributo que acompaña a las cadenas de texto e incluye el idioma en el que están escritas, la variante idiomática del país y el juego de caracteres empleado) parece que la CLA ha sido insuficientemente tenida en cuenta hasta ahora y es importante hacer un esfuerzo para incorporarla en los modelos de calidad.

Los aspectos indicados anteriormente afectan a los datos (modelo, catálogo, resolución, lenguaje), pero también hay aspectos semánticos en los servicios. Con estos aspectos nos referimos a entender plenamente qué es lo que ofrece un servicio. Por ello también se deben considerar:

- Interoperabilidad de significado de servicios. Aunque se puede englobar dentro de la conformidad de los servicios web con los estándares elegidos, se puede distinguir entre una conformidad meramente sintáctica, que garantiza que se cumplen todos los requisitos formales del estándar en cuanto a formatos, peticiones y parámetros, y una interoperabilidad que aseguraría además que se respeta el significado y sentido originales del estándar.

Por ejemplo, el significado y sentido originales del estándar WMS 3.1.0 (*Open Geospatial Consortium*) se pervierte, por ejemplo: a) cuando se utiliza el *abstract* (resumen) del documento de *capabilities* (capacidades) para codificar las restricciones de acceso, que deben ir en la etiqueta *AccessConstraint*; b) cuando se emplea la división en capas para organizar datos de distinta fecha, cuando para eso existe el parámetro TIME, o c) cuando se introduce una marca de aguas con un mensaje de *copyright* en el mapa que devuelve una petición *GetMap*, cuando de nuevo es la etiqueta *AccessConstraint* del *capabilities* el lugar estándar para codificar esa información.

- Falta de interoperabilidad de interfaces. Teniendo en cuenta que, más allá de la explotación automática de servicios web en remoto desde aplicaciones que los integran así en su propia lógica de aplicación, buena parte de la utilización de servicios web es llevada a cabo por usuarios humanos a través de interfaz de usuario, su calidad es un factor importante a tener en cuenta.

Interoperabilidad técnica

Desde una aproximación basada en las IDE y la Arquitectura Orientada a Servicios, la interoperabilidad técnica depende de una serie de factores relacionados con las comunicaciones, como disponer del ancho de banda suficiente, el desempeño de las transferencias de datos, la disponibilidad y estabilidad de la conexión y la calidad de los servicios.

Efectivamente toda la funcionalidad de un sistema se puede conceptualizar como un servicio, incluyendo los servicios de descarga de datos cuando no se trabaja con servicios web sino fuera de línea y en local. Por lo tanto, buena parte de los aspectos de la calidad de una IDE relacionados con la interoperabilidad técnica se pueden repercutir en la calidad de servicio, descrita en el marco de la Directiva Europea INSPIRE (Comisión Europea, 2009b) por medio de tres parámetros establecidos de la siguiente manera:

- Disponibilidad o probabilidad de que el servicio no esté caído y responda, expresado habitualmente como un tanto por ciento de disponibilidad mensual o anual. Para determinarla se realizan peticiones cada seis minutos, que deben estar perfectamente definidas y descritas. Se exige al menos un 99% de disponibilidad anual excepto tiempos de parada técnica anunciados con una semana de antelación.
- Rendimiento o tiempo de respuesta, descrito como el tiempo que se tarda en recibir el primer bit de respuesta a una petición dada. Se emplean las mismas peticiones que sirven para determinar la disponibilidad para calcular medias anuales. El marco INSPIRE (Comisión Europea, 2009b) establece un límite de 5 segundos para un servicio WMS de visualización y 3 segundos para un servicio CSW de catálogo.
- Capacidad o posibilidad de responder un pico de peticiones de prueba, como por ejemplo 20 peticiones por segundo durante un minuto, manteniendo el rendimiento deseado (Comisión Europea, 2009b).

En cuanto a las interfaces habría que tener en cuenta tanto su interoperabilidad como su calidad, y en ese sentido sí se pueden encontrar algunos estudios sobre calidad de geoportales de una IDE, como son los de Rodríguez *et al.*, 2015, 2017 y 2018. En ellos se definen una serie de aspectos de la calidad de un geoportal que atienden fundamentalmente a su apertura, es decir a la ausencia de barreras de todo tipo que limiten su uso y a la interoperabilidad, estrechamente ligada a la estandarización del recurso. Como consecuencia de los tres estudios mencionados, dedicados respectivamente a los geoportales españoles, los iberoamericanos y los europeos, lo que supone haber analizado 79 geoportales IDE nacionales o regionales, de España, Iberoamérica y Europa, y de haber realizado una encuesta para afinar la lista de

indicadores entre 33 expertos IDE de todo el mundo, se llegó a seleccionar 19 parámetros para evaluar la calidad de geoportales, que no se incluyen aquí por estar ampliamente descritos en Rodríguez *et al.* (2018).

Hasta ahora se ha prestado cierta atención a la calidad de servicios web, pero muy poca a la calidad de geoportales, ya que apenas hay estudios al respecto más allá de los citados. En ese sentido, la lista de indicadores propuestos en esa referencia puede ser un buen punto de partida para analizar geoportales y hacer propuestas de mejora.

Finalmente y a otro nivel de abstracción, podría considerarse como otro aspecto de la calidad el grado de implementación de la Arquitectura Orientada a Servicios. Así, podría definirse algún indicador que describiese cuantitativamente hasta qué punto una organización productora de datos geográficos ha implementado realmente esta arquitectura examinando sus procesos de producción y viendo por ejemplo qué porcentaje de ellos desemboca en la publicación de servicios web.

Conclusiones

El ciclo de vida de las Infraestructuras de Datos Espaciales parece que presenta anomalías respecto de la evolución en el pasado de otras tecnologías similares, como los SIG, y es posible que se esté abriendo una brecha entre sus planteamientos y las vanguardias en el campo de la información geográfica. En consecuencia, nos parecen muy recomendables las líneas de investigación que traten de integrar esas nuevas tecnologías (datos enlazados, *Big Data*, *Smart Cities*, *Internet of Things*, Realidad Aumentada/Virtual, 3D, BIM, *machine learning*, etc.) y las tecnologías IDE, haciendo evolucionar tanto a las primeras como a las segundas. Están en juego los dos grandes pilares de las IDE: la interoperabilidad y la filosofía de datos y servicios abiertos.

Por otro lado, la implantación y evolución de los modelos de calidad IDE están llamadas a ser una de las palancas que potencien su desarrollo y progreso. En este trabajo se ha presentado un análisis de las aproximaciones utilizadas hasta ahora para la caracterización de la calidad global de una IDE, aspecto sobre el que se dispone de literatura (Randolf Pérez *et al.*, 2015, Vancauwenberghe *et al.*, 2018) sobre experiencias bajo diferentes enfoques, nacidos más de modelos teóricos que de experiencias prácticas.

Resulta difícil organizativamente y costoso realizar experiencias de aplicación de algunos de tales modelos, o de una variedad interesante de indicadores, en varias IDE y durante varios años, máxime cuando las IDE son sistemas complejos, sistemas de sistemas según (Béjar *et al.*, 2009), que pueden tardar entre 10 y 15 años en alcanzar la madurez si se trata de un país o región compleja. Hay que tener en cuen-

ta que el periodo de implementación previsto en la Directiva INSPIRE en Europa es de trece años (2007-2020).

Del análisis de la literatura sobre indicadores, se desprende que se pueden clasificar en cinco categorías: indicadores de acciones realizadas, de calidad de los componentes implementados, de uso, de impacto y de coste-beneficio.

En líneas generales, son más abundantes los análisis de calidad de una IDE basados en indicadores de acciones realizadas y de uso, pero su relación causa-efecto con la calidad final de una IDE es demasiado lejana. Por otro lado, resultan más interesantes por ser efecto casi directo de la calidad de una IDE los indicadores de impacto y de coste-beneficio, pero presentan el inconveniente de ser difícilmente objetivables debido a que los beneficios de una IDE son en su mayoría intangibles y se dan a muy largo plazo. En cambio, los indicadores de calidad de los componentes implementados resultan muy interesantes, si se considera un abanico suficientemente amplio de componentes de una IDE, ya que su calidad es causa directa de la calidad global de una IDE. Además, son relativamente fáciles de cuantificar objetivamente.

Para diseñar un modelo de calidad basado en indicadores de calidad de los componentes implementados proponemos tomar como base su característica más importante, la interoperabilidad, y para caracterizarla, un esquema de interoperabilidad avanzado y lo más completo posible, como el *European Interoperability Framework v2* (Comisión Europea, 2017) y ampliarlo considerando el máximo abanico posible de aspectos de la interoperabilidad relevantes. A cada aspecto de la interoperabilidad contemplado en ese esquema ampliado, le correspondería un aspecto de la calidad.

Sin embargo, en muchos de los aspectos a tener en cuenta, se encuentran facetas poco exploradas y cuantificadas. De esas carencias nace la necesidad de abordar líneas de investigación que profundicen en los aspectos detallados en las secciones anteriores, especialmente investigación en la caracterización y determinación de la interoperabilidad legal, la interoperabilidad organizativa y la interoperabilidad de significado, toda vez que el resto de niveles y aspectos son relativamente bien conocidos.

Esa aproximación complementaría a las líneas de trabajo existentes en este campo, que siguen siendo oportunas y deseables, como los estudios de impacto y los análisis de coste-beneficio.

El Instituto Panamericano de Geografía e Historia puede ser una instancia privilegiada para fomentar este tipo de actividades y el foro idóneo tanto para compartir y discutir los resultados de las distintas líneas de investigación e innovación mencionadas, como para abordar la implementación de modelos de calidad para las IDE de la región.

Bibliografía

- AENOR (2003). “UNE 66175 Sistemas de gestión de la calidad – guía para la implantación de sistemas de indicadores”, Madrid.
- Ariza, F.J, y Rodríguez, A.F. (eds.) (2008). “Introducción a la Normalización en Información Geográfica: la familia ISO 19100. (130 pp.), Grupo de Investigación en Información Cartográfica, Universidad de Jaén, disponible en <http://coello.ujaen.es/Asignaturas/pcartografica/Recursos/IntroduccionNormalizacion_IG_FamiliaISO_19100_v2.pdf>, consultado el 13 de agosto de 2018.
- Balciunas, A. and Rozanskas, E. (2016). *National Benefits from INSPIRE Implementation. The Real Life Use Cases*, en 2016 INSPIRE Conference, disponible en <http://inspire.ec.europa.eu/pr_searchx.cfm?i=5&id_search=50026>, consultado en 13 de agosto de 2018.
- Béjar, R.; Latre, M.A.; Nogueras-Iso, J.; Muro-Medrano, P.R. and Zarazaga-Soria, F.J. (2009). “Systems of Systems as a Conceptual Framework for Spatial Data Infrastructures”, en *International Journal of Spatial Data Infrastructure Research*, vol. 4, JRC, ISPRA, Italia, 201-2017.
- Comisión Europea (2009) “Decisión por la que se ejecuta la Directiva 2007/2/CE del Parlamento Europeo y del Consejo en lo que se refiere al seguimiento y los informes”, disponible en <<https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:148:0018:0026:ES:PDF>>, consultado el 13 de agosto de 2018.
- Comisión Europea (2009b). “Reglamento (CE) 976/2009 de la Comisión, de 19 de octubre de 2009 por el que se ejecuta la Directiva 2007/2/CE en lo que se refiere a los servicios de red”, disponible en <<http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32009R0976&from=ES>>, consultado el 13 de agosto de 2018.
- Comisión Europea (2017). “*European Interoperability Framework v2*”, disponible en <http://eur-lex.europa.eu/resource.html?uri=cellar:2c2f2554-0faf-11e7-8a35-01aa75ed71a1.0017.02/DOC_3&format=PDF>, consultado el 13 de agosto de 2018.
- Comisión Europea (2017b) “*Summary Report on Status of implementation or the INSPIRE Directive in EU*”, disponible en <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC109035/jrc109035_jrc109035_jrc_inspire_eu_summaryreport_online.pdf>, consultado el 13 de agosto de 2018.
- Crompvoets, J.; Rajabifard, A.; van Loenen, B. and Delgado, T. (eds.) (2008). “*A Multi-View Framework to Assess SDIs*”, *Space for Geo-Information*, RGI, Wageningen University, Melbourne, Australia.
- De Jong, C. (2016). “*Cost-benefit analysis INSPIRE in the Netherlands*”, en INSPIRE Conference, disponible en

- <http://inspire.ec.europa.eu/pr_searchx.cfm?i=5&id_search=50025>, consultado el 13 de agosto de 2018.
- EEA-JRC (2014). “*Mid-term evaluation report on INSPIRE implementation*”, disponible en <<https://www.eea.europa.eu/publications/midterm-evaluation-report-on-inspire-implementation>>, consultado el 13 de agosto de 2018.
- García, J.; García-Peñalvo, F.J. and Therón, R. (2010). “*A Survey on Ontology Metrics*”, en Knowledge Management, Information Systems, E-learning and Sustainability Research, Springer, 22-27.
- Giff, G. and Cropvoets, J. (2008). “*Performance Indicators: a tool to Support Spatial Data Infrastructure assessment*”, *Computers, Environment and Urban Systems*, vol. 32, Elsevier, Amsterdam, Países Bajos, 365-376.
- GSDI (2012). *SDI Cookbook*, disponible en <http://gsdiassociation.org/images/publications/cookbooks/SDI_Cookbook_from_Wiki_2012_update.pdf>, consultado el 13 de agosto de 2018.
- ISO/IEC (2005). “*TR 9764:2005 Information technology - Guidelines, methodology and reference criteria for cultural and linguistic adaptability in information technology products*”, ISO Ginebra.
- Kelm K.; Probert, M. and Tonchovska (2017). “*Creating a Spatial Data Infrastructure Diagnostic Tool*”, World Bank Conference on Land and Poverty, marzo 20-24, 2017, Washington DC.
- Maganto, A.S.; Rodríguez, A.F. y Bernabé, M.A. (2012). “Componentes de una IDE”, en Bernabé-Poveda, M.A. y López-Vázquez, C.M., *Fundamentos de las Infraestructuras de Datos Espaciales*, UPM Press Serie Científica, Madrid. España, pp. 55-65.
- Mani, A. and Nagarajan, A. (2002). “*Understanding quality of service for Web Services. Improving the performance of our Web services*”, IBM, disponible en <<https://www.ibm.com/developerworks/library/ws-quality/index.html>>, consultado el 13 de agosto de 2018.
- Martín, A.; Contreras, J.J.; Salinas, E.; Valverde, P.; Balbi, R.; Medeiros, V.E.; Vianey, A.; Fajardo, A.; Guerrón, P.; Tene, E.; Sámuel, E.; Agrazal, A.; Flores, R.; Avilés, W.; Parcher, J.; Lugo, R.; Blanco, L.M. y Rodríguez, A.F. (2017). “El Mapa Integrado Andino del Norte”, en *Revista Cartográfica*, núm. 94, enero-junio, Instituto Panamericano de Geografía e Historia, 2017, 11-27.
- Morera, C.; Carrasquilla, O.; Rey, D. y Guimet, J. (2012). “Evaluación de una IDE desde su caracterización hasta su impacto en la sociedad”, en Bernabé-Poveda, M.A. y López-Vázquez, C.M., *Fundamentos de las Infraestructuras de Datos Espaciales*, UPM Press Serie Científica, Madrid, España, 443-452.
- Mueller, P.A. (1997). “Modelado con objetos UML”, editorial Eyrolles, París.

- Nativi, S.; Craglia, M. and Pealman, J. (2012). “*The Brokering Approach for Multi-disciplinary Interoperability: A Position paper*”, in *International Journal of Spatial Infrastructures Research*, vol. 4, 1-15.
- Norori, M.; Williams, C.; Torres, E.; Salazar, L.; Sáncho, E.; Flores, E.; Gómez, B. I.; Benítez, F.; Menendez, R.; Núñez, M.; Martínez, A.; Parcher, J.; Roberto, L.; Zavala, E.; Ramírez, N.; Pérez, J.; Morales, C.; Sámuels, E. y Mondragón, R. (2013). “Proceso participativo de producir un Mapa Integrado de Centroamérica y Sur de México”, en *Revista Cartográfica*, núm. 89, enero-diciembre, Instituto Panamericano de Geografía e Historia, 47-60.
- Open Geospatial Consortium (2006). “*OpenGIS Web Map Server Implementation specification v 1.3.0*”, disponible en: <<http://www.opengeospatial.org/standards/wms>>, consultado el 20 de marzo de 2018.
- Orden Ejecutiva 12906. “*Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*” (1994), Presidente de Estado Unidos, disponible en <<https://www.archives.gov/files/federal-register/executive-orders/pdf/12906.pdf>>, consultado el 13 de agosto de 2018.
- Randolf Pérez, D.; Ballari, D. y Vilches-Blázquez, L. (2015). “Participación y dinamicidad en las Infraestructuras de Datos Espaciales: una propuesta de indicadores para medir su impacto en la sociedad”, en *Revista Cartográfica*, núm. 90, enero-diciembre, Instituto Panamericano de Geografía e Historia, 175-191.
- Rodríguez, A.F.; Abad, P.; Sánchez, A.; Juanatey, M. y Cevidanes, A. (2015). “Interoperabilidad práctica de los geoportales de la IDEE” presentación realizada en las JIIDE 2015, disponible en <http://www.idee.es/resources/presentaciones/JIIDE15/20151104/18_InteroperabilidadPracticadeGeoportales.pdf>, consultado el 13 de agosto de 2018.
- Rodríguez, A.F. (1993). “Los SIG en España: un campo sin cartografiar”, *Topografía y Cartografía*, núm. 55, 25-41.
- Rodríguez, A.F.; Abad, P.; Sánchez, A.; Juanatey, M. y Cevidanes, A. (2018). “Aproximación a una metodología de evaluación de la calidad de geoportales IDE”, *Revista de Cartografía*, núm. 95, número monográfico sobre geoportales, julio-diciembre, Instituto Panamericano de Geografía e Historia, 17-42.
- Rodríguez A.F.; Sevilla C.; Abad P.; López, E. y Sánchez A. (2013). “Panorama de la estandarización en el campo de la información geográfica: la importancia de los estándares abiertos”, CIGMA, México.
- Rodríguez, A.F.; Abad, P.; Sánchez, A.; Juanatey, M. y Cevidanes, A. (2017). “*Towards a Methodology to evaluate the Quality of SDI Geoportals*” presentación realizada en el Congreso INSPIRE 2017, disponible en <<https://inspire.ec.europa.eu/sites/default/files/presentations/2017InspireConferenceGeoportalsAFRRodríguez.pdf>>, consultado el 13 de agosto de 2018.

- Rodríguez, A.F., López, E., Cabria, A., Rodríguez, J.M. y Vivas, P. (2017b). “El nuevo Esquema Europeo de Interoperabilidad y las IDE”, en *MAPPING*, vol. 26, noviembre-diciembre, 186, 6-11.
- UN-GGIM (2018). “Integrated Geospatial Information Framework. A strategic guide to develop and strengthen National Geospatial Information Management. Part 1: Overarching Strategic Framework”, disponible en <<http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Part%201-IGIF-Overarching-Strategic-Framework-24July2018.pdf>>, consultado el 13 de agosto de 2018.
- Vandenbroucke, D. (2012). “Did we approach the objectives of INSPIRE?”, INSPIRE Conference, disponible en <http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2012/presentations/183.pdf>, consultado el 13 de agosto de 2018.
- Vancouwenberghe, G.; Valeckaite, K.; van Loenen, B. and Welle Donker, Frederika (2018). “Assesing the Openness of Spatial Data Infrastructures (SDI): Towards a Map of Open SDI”, *International Journal of Spatial Data Infrastructures Research*, 13, 88-100.
- Welle Donker, F. and B. van Loenen (2016). “How to assess the success of the open data ecosystem?” *International Journal of Digital Earth*, 10 (3), 284-306.
- World Wide Web Consortium (2003). “QoS for Web Services: Requirements and Possible Approaches”. W3C Working Group Note, disponible en <<http://www.w3c.or.kr/kr-office/TR/2003/ws-qos/>>, consultado el 13 de agosto de 2018.
- World Wide Web Consortium (2017). “*Spatial Data on the Web Best Practices*”. W3C Working Group Note, disponible en <<https://www.w3.org/TR/sdw-bp/>>, consultado el 13 de agosto de 2018.

Cuantificación de vulnerabilidad y riesgo: las inundaciones en Motozintla de Mendoza, Chiapas, México*

Mary Frances Rodríguez Van Gort**

Recibido 16 de abril de 2018, aceptado 20 de septiembre de 2018

Resumen

El objetivo fundamental de esta investigación fue desarrollar una metodología para cuantificar la vulnerabilidad social global. En este estudio la metodología es aplicada para determinar el riesgo por inundación que registra Motozintla. Sin embargo, dicha metodología puede ser replicada en otras comunidades del país y ante otros peligros naturales.

La metodología consistió en (a) realizar la actualización cartográfica del lugar, (b) establecer la muestra mínima de viviendas a caracterizar y seleccionarlás con un método aleatorio simple, (c) determinar las variables fundamentales para la identificación de la vulnerabilidad por grupos (estructural, socio-económica, organizacional, por percepción de riesgo) y global (todas las anteriores por nivel de incidencia ponderado), (d) aplicar el método de interpolación de Distancia Inversa Ponderada (IDW, por sus siglas en inglés) para identificar áreas y niveles de vulnerabilidad y riesgo, y (e) presentar los resultados cartográficamente en mapas generados por un Sistema de Información Geográfica (SIG) para su análisis espacial.

Palabras clave: *vulnerabilidad, peligro por inundación, riesgo, Sistemas de Información Geográfica.*

Resumo

O objetivo fundamental desta investigação foi desenvolver uma metodologia para quantificar a vulnerabilidade social global. Neste estudo a metodologia é aplicada

* Ésta es una versión revisada y corregida de una parte del trabajo de tesis doctoral denominada "Cuantificación del riesgo asociado al peligro por inundación en Motozintla de Mendoza, Chiapas, México".

** Profesora de Tiempo Completo del Colegio de Geografía, Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México, Av. Universidad 3000, Col. Ciudad Universitaria, 04510, Alcaldía de Coyoacán, Ciudad de México, correo electrónico: francesrv@gmail.com, cls@filos.unam.mx.

para determinar o risco por inundação que registra Motozintla. Entretanto, esta metodologia pode ser replicada em outras comunidades do país e diante outros perigos naturais.

A metodologia consistiu em (a) realizar a atualização cartográfica do lugar, (b) estabelecer a amostra mínima de domicílios e selecioná-las com um método aleatório simples, (c) determinar as variáveis fundamentais para a identificação da vulnerabilidade por grupos (estrutural, sócio-econômico, organizacional, por percepção de risco) e global (todas as anteriores por nível de incidência ponderada), (d) aplicar o método de interpolação de Distância Inversa Ponderada (IDW, por sua sigla em inglês) para identificar áreas e níveis de vulnerabilidade e risco, e (e) apresentar os resultados cartograficamente em mapas gerados por um Sistema de Informação Geográfica (SIG) para sua análise espacial.

Palavras chave: *vulnerabilidade, perigo por inundação, risco, Sistemas de Informação Geográfica.*

Abstract

The central objective of this research was to develop a methodology for the quantification of global social vulnerability. The methodology is used to determine the risk of flooding in Motozintla. However, such methodology can be replicated in others communities in Mexico.

The methodology consisted in the following steps: (a) the cartographic actualization of the urban layout of Motozintla; (b) the establishment and characterization of a minimal sample of households for this study; (c) selection of variables for the identification of different types of vulnerabilities (structural, socio-economic, organizational, and perception of risk) and global vulnerability (all variables of the analyzed vulnerabilities considering a level of weighted incidence); (d) use of the interpolation method, Inverse Distance Weighting (IDW), to identify areas and levels of vulnerability and risk; and (e) use of a Geographic Information System for cartographic representation of the results and their spatial analysis.

Keywords: *vulnerability, flood hazard, risk, Geographic Information System.*

Introducción

Resulta importante señalar que las comunidades urbanas presentan un número cada vez mayor de desastres en un menor tiempo, lo cual se debe fundamentalmente a la creciente y deficiente urbanización de nuestra sociedad, así como al aumento de exposición ante diversos fenómenos. De acuerdo con los antecedentes presentados por la Organización de las Naciones Unidas (ONU, 2017) en el marco de los Objetivos para el Desarrollo Sostenible, lo anterior puede provocar graves problemas, como el

crecimiento de asentamientos marginales y la falta de servicios básicos y de infraestructura, lo que aumenta la vulnerabilidad de las comunidades a desastres. Por ello se requiere mejorar la planificación en un sentido de mayor inclusión y gestión urbana que mejore las condiciones de resiliencia y sostenibilidad de los asentamientos.

El propósito fundamental de esta investigación fue desarrollar una metodología para cuantificar la vulnerabilidad social global y a partir de ésta realizar una aproximación al riesgo por inundación en la localidad de Motozintla de Mendoza, Estado de Chiapas, México. Y que este análisis cuantitativo considerara su representación cartográfica a partir de un Sistema de Información Geográfica. Este trabajo forma parte de los resultados de una investigación de doctorado en Ciencias en el área de Geofísica.

La localidad de Motozintla de Mendoza, cabecera municipal Motozintla, se ubica en la porción sureste del estado de Chiapas, entre las coordenadas 15° 22' N y 92° 15' W, con una altitud de 1,260 msnm. De acuerdo al Censo del Instituto Nacional de Estadística, Geografía e Informática (INEGI, 2010a), contaba en 2010 con una población de 23,755 habitantes. Esta comunidad enfrentó en fechas recientes, primero en septiembre de 1998 y luego en octubre de 2005, desastres por inundación y desajamiento de laderas debido fundamentalmente a las amenazas naturales a que está expuesta, como son las precipitaciones intensas resultado de la influencia de los huracanes procedentes de diversas regiones ciclogénicas, de acuerdo con el Centro Nacional de Prevención de Desastres (CENAPRED 2001).

El régimen pluviométrico caracteriza a la zona con una moda anual de entre 1.000 y 2.000 mm con precipitaciones máximas en 24 horas de entre 200 y 400 mm (Vidal y Gómez, 2007; Vidal, 2007). Son varios los procesos de remoción en masa a los que es propenso su territorio y al combinarse con la vulnerabilidad física y social de sus habitantes se han visto favorecidas las condiciones de susceptibilidad a las inundaciones que pueden derivar en desastres.

La vulnerabilidad global se estimó considerando la vulnerabilidad estructural de las viviendas, la vulnerabilidad socio-económica de sus habitantes, la vulnerabilidad institucional (eficiencia de las autoridades correspondientes en el manejo y gestión del riesgo), y la vulnerabilidad asociada a la percepción del riesgo que la población presenta ante el peligro de inundación. También se incorporaron variables sobre la vulnerabilidad organizacional, tanto de la propia población como la relacionada con la implementación de planes de respuesta institucional.

En este trabajo se establece el nivel de riesgo de Motozintla (un territorio expuesto a inundaciones, particularmente en la zona urbana) considerando su vulnerabilidad y el peligro por inundación. Para ello se establecieron los siguientes objetivos particulares.

- Desarrollar un marco teórico básico para la cuantificación de la vulnerabilidad.
- Determinar y ponderar las variables de mayor incidencia para el establecimiento de niveles de vulnerabilidad y riesgo.
- Elaborar un instrumento de levantamiento de datos sobre vulnerabilidad en campo.
- Cuantificar los siguientes tipos de vulnerabilidades en Motozintla: estructural, socio-económica, organizacional y global, así como de la percepción del riesgo de la población.
- Determinar el riesgo por inundación en Motozintla de Mendoza a partir de la cuantificación de la vulnerabilidad y la incorporación de la caracterización de la amenaza.
- Interpretar los resultados espacialmente, a partir de su integración en un Sistema de Información Geográfica (SIG).

Cuantificación de la vulnerabilidad

Se considera al riesgo como la probabilidad de una pérdida, el cual depende de tres elementos: los peligros, la exposición y la vulnerabilidad. “Si cualquiera de estos tres elementos aumenta o disminuye, entonces el riesgo aumenta o disminuye, respectivamente” (Crichton, 1999). En términos generales el riesgo resulta de la combinación del peligro y la vulnerabilidad (Magaña, 2012). Dado lo anterior resulta prioritario la valoración de la vulnerabilidad para la reducción del riesgo (Macías, 1999; Macías y Capra, 2005; Magaña, 2012).

Las condiciones de vulnerabilidad de la población representan una parte fundamental en la ecuación del riesgo y debido a la multivariabilidad que presenta su estudio resulta también de gran complejidad. ¿Qué elementos son fundamentales?, ¿cuáles tienen un mayor peso? Así, uno de los principales retos en la valoración del riesgo resulta el logro de la estimación de una vulnerabilidad social para una población determinada (Adger, 2006; Magaña, 2012). Cabe señalar que uno de los propósitos fundamentales de identificar la vulnerabilidad se encuentra ligado directamente con el aumento de la resiliencia de las comunidades en riesgo a partir del conocimiento de su condición de riesgo y de la gestión participativa del riesgo para la reducción de los desastres.

Es así que la vulnerabilidad ha requerido el desarrollo de diferentes metodologías. Muchos autores han optado por analizar un tipo de vulnerabilidad en especial (social, económica o estructural) para un peligro en particular (Martinelli *et al.*, 2008; Novelo-Casanova y Suárez 2010; Saxena *et al.*, 2013). Por ejemplo, se han llevado a cabo diversos intentos para cuantificar la evaluación de la vulnerabilidad mediante métodos estadísticos. Por ejemplo, Martinelli *et al.* (2008) desarrollaron un procedimiento para evaluar la vulnerabilidad sísmica de los edificios residenciales en Celano, Italia, por su parte Kaynia *et al.* (2008) exploraron una nueva metodología para

estimar la vulnerabilidad a los deslizamientos en la localidad de Lichtenstein, en la cual consideraron una aproximación probabilística simple.

El reto principal para determinar la vulnerabilidad es el desarrollo de herramientas de investigación para una evaluación integrada de la vulnerabilidad (vulnerabilidad global), así como la incorporación de los estudios de gobierno con el propósito de reducir el riesgo y la promoción de estrategias de resiliencia (Adger, 2006), asimismo, este autor sugiere que es importante incorporar la percepción del riesgo, como un elemento central de la vulnerabilidad.

La percepción del riesgo es la forma en que una comunidad determinada percibe el riesgo en función de su experiencia en desastres anteriores o del nivel de exposición a estos peligros. Por ello representa un elemento fundamental en la condición de vulnerabilidad de la población, ya que influye tanto en las acciones de prevención como en las de mitigación (Macías, 1999; Bubeck, 2012). Por otra parte la percepción de riesgo tiene, por lo general un carácter temporal (De Vries, 2007). Para determinar la vulnerabilidad global es necesario tener en cuenta los elementos sociales que construyen esta vulnerabilidad. El análisis de la vulnerabilidad global puede contribuir a la implementación de acciones para la reducción del riesgo (Kappes *et al.*, 2011).

Un elemento fundamental para la metodología de este trabajo se basó en la generación de datos a partir del trabajo de campo, herramienta fundamental de acuerdo con Brid (2009), Jacob *et al.* (2013) y Banning *et al.* (2012), tanto en la actualización de la traza urbana y la generación de la cartografía básica, como en la toma de información sobre las condiciones estructurales de las viviendas, las condiciones de los servicios, la preparación ante contingencia y la percepción que la población tiene sobre sus condiciones de riesgo.

La metodología para cuantificar la vulnerabilidad se compuso de siete fases: 1) actualización georreferenciada de la traza urbana de la localidad; 2) selección de variables que condicionan la vulnerabilidad y elaboración del instrumento de encuesta; 3) cálculo de la muestra mínima estadísticamente significativa y selección aleatoria de viviendas; 4) levantamiento de encuestas en campo; 5) creación de bases de datos; 6) análisis de datos y ponderación de variables de acuerdo a su influencia en la vulnerabilidad y 7) análisis espacial de la cuantificación de la vulnerabilidad.

Actualización georreferenciada de la traza urbana de la localidad

Al visitar la zona de estudio por primera vez (octubre de 2009), se observó que la cartografía con la que contaban las autoridades municipales era obsoleta e incompleta, por lo que se procedió a la actualización de la traza urbana. Para ello, se realizó una primera actualización a partir de las imágenes de satélite de Google y se llevó a cabo una segunda visita de campo en noviembre de 2009, con los mapas obtenidos

en la red se realizaron los recorridos para el levantamiento actual de las calles, y se rediseñó el mapa base de la localidad de Motozintla de Mendoza. A partir de esta base de datos, se realizó una actualización por medio de la identificación de puntos con aparatos de geo-posicionamiento satelital por Receptores de Sistema Global de Navegación por Satélite (GNSS, por *Global Navigation Satellite System* en inglés), con lo cual se obtuvo la cartografía georreferenciada de la localidad.

Elección de variables para la encuesta

En esta fase fue importante considerar que los indicadores no fueran excesivos y guardaran significancia, lo cual es fundamental para lograr valorar cuantitativamente la vulnerabilidad (Magaña, 2012). Se observó que para la toma de datos sobre las características de la población, tanto de sus viviendas, como de los habitantes de las mismas, así como de la percepción de riesgo de la población, resulta necesario el trabajo de campo, para lo cual un instrumento importante es la entrevista a partir del uso de cuestionarios básicos. Para la elaboración del instrumento de encuesta final, se determinaron cuatro tipos de vulnerabilidad: estructural, socio-económica, institucional-organizacional y percepción de riesgo. El instrumento final se estructura por 51 reactivos, con tres tipos de preguntas (Tabla 1).

Tabla 1
Variables y formato que se incluyeron en la encuesta final (agrupados por tema)

| <i>Total de reactivos</i> | <i>Formato</i> | <i>Tema</i> |
|---------------------------|----------------|---------------------------------------------|
| 3 | Cerrado | Vulnerabilidad estructural |
| 13 | Cerrado | Vulnerabilidad social |
| 3 | Semiabierto | Vulnerabilidad económica |
| 9 | Semiabierto | Vulnerabilidad organizacional-institucional |
| 4 | Cerrado | Vulnerabilidad organizacional-institucional |
| 18 | De opinión | Percepción de riesgo |
| 1 | Cerrado | Percepción de riesgo |

Fuente: Elaboración propia a partir de dos ejercicios previos de levantamiento de datos que se realizaron en campo.

Muestra mínima y selección aleatoria de viviendas

Aplicando la fórmula (1) para el establecimiento de la muestra mínima (Mendenhall *et al.*, 2003), y el universo de viviendas en la zona de estudio que se determinó en 5,000 viviendas, se obtuvo una muestra mínima de 386. A continuación se describen los elementos del cálculo de la muestra mínima:

$$n = Z^2(p+q)N / (N-1)E^2 + Z^2(pq) \quad (1)$$

Tabla 2
Valor de la muestra mínima para Motozintla de Mendoza
a partir de la fórmula 1

| <i>Variable</i> | <i>Descripción de la variable</i> | <i>Valor de la variable</i> |
|-----------------|-----------------------------------|-----------------------------|
| Z | Nivel de confianza | 95% |
| p | Variabilidad positiva | .95 |
| q | Variabilidad negativa | .05 |
| N | Tamaño de la población | 5000 |
| E | Precisión o el error | 5% |
| n | Tamaño de la muestra | 386 |

Fuente: Mendenhall *et al.*, 2003.

En base a estos datos y con la cartografía digitalizada, se realizó una tercera visita de campo a la zona de estudio con la finalidad específica de realizar un conteo de manzanas, lotes y viviendas que sirvieran para la conformación de una base de datos que permitiría realizar, a través del corrimiento de uno de los programas de ArcGIS 3.9 la elección aleatoria simple de los lotes de la localidad y con ello lograr una mayor representatividad de las viviendas que se encuestarían en el trabajo de campo.

A partir del mapa base que se obtuvo para el levantamiento de encuestas, donde se ubicaron 375 manzanas y 5,600 lotes, se aplicó el método de estratos, donde la muestra total es X y la muestra mínima es Y. Se consideraron 40 estratos a partir de su ubicación geográfica, considerando que cada estrato contara con 140 lotes, a cada uno se le asignaron 10 lotes para encuestar al momento de correr el programa de elección aleatoria. Con esta base, la información se incorporó al programa ArcGIS 3.9 que utiliza un método de elección aleatoria simple, donde cada uno de esos 10 lotes tiene la misma probabilidad de ser elegido.

En el mapa base se consideraron los 40 estratos o áreas geográficas divididos por colores, al interior de cada manzana se aprecia el número que le fue asignado de un total de 375, y los puntos verdes (500) indican los lotes o viviendas que el programa de elección aleatoria simple que se corrió en ArcGIS 3.9 asignó para la realización de las entrevistas (Figura 1).

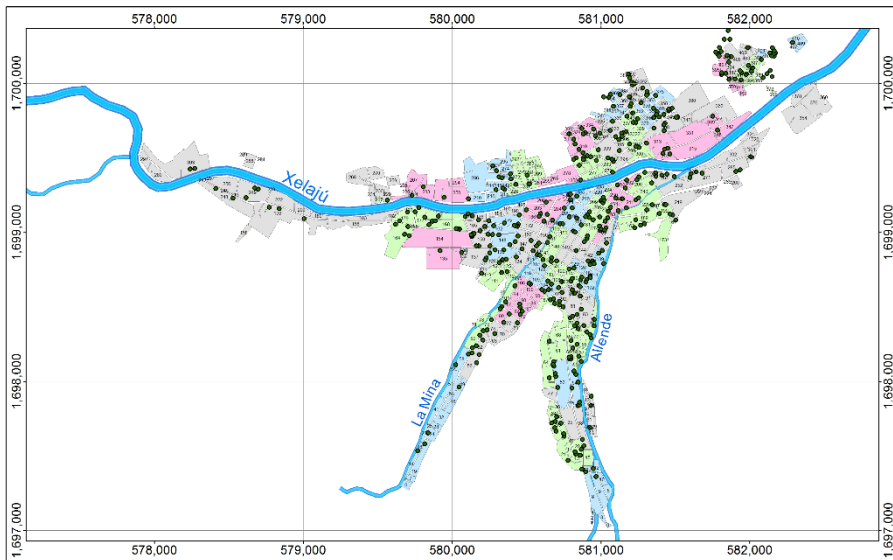


Figura 1. Mapa base de Motozintla de Mendoza (georreferenciado). Los números corresponden a cada una de las 375 manzanas en el área de estudio y los colores a cada uno de los estratos que se determinaron para que la selección aleatoria de viviendas fuera representativa de todo el territorio de la localidad.

Levantamiento de encuestas y creación de la base de datos

En campo se entrevistaron 486 viviendas, sin embargo, de éstas se descartaron 42 por diversas razones (ilegibilidad, por estar repetidas, o por tener una ubicación errónea, es decir, en vez de vivienda eran lotes baldíos u otro tipo de predio). Por lo tanto, se contó con 444 encuestas finales para la base de datos sobre la vulnerabilidad y la percepción de riesgo. Con estos resultados se creó una planilla de datos en Excel. Posteriormente, se realizaron gráficas (Figura 2) para cada una de las 30 variables seleccionadas para la valoración de la vulnerabilidad (Tabla 3).

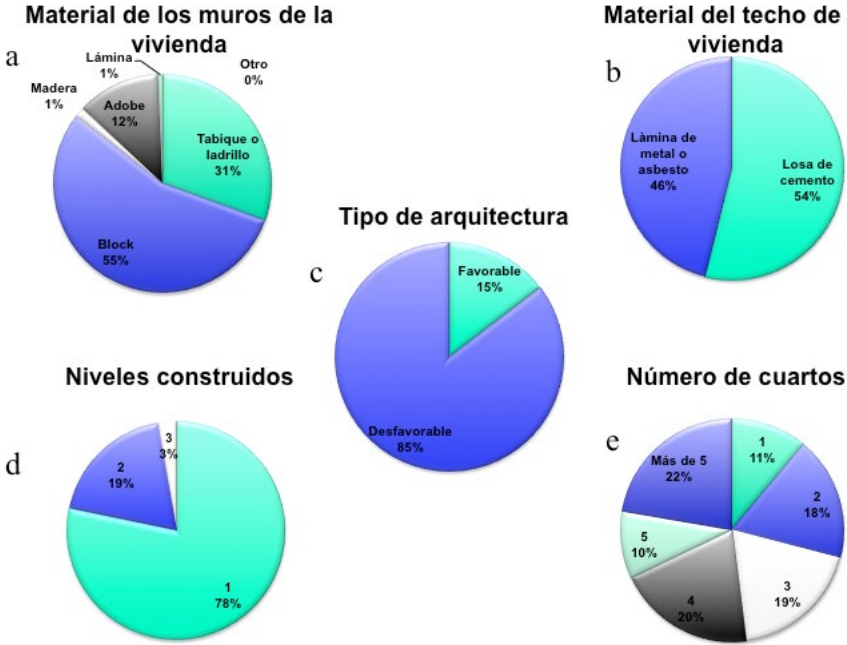


Figura 2. Gráfica de vulnerabilidad estructural, donde se muestran los porcentajes en cuanto a las características estructurales de las viviendas: a) materiales de las paredes, b) materiales del techo, c) tipo de arquitectura, d) niveles de construcción y e) número total de habitaciones en la vivienda.

Selección de variables para la cuantificación de la vulnerabilidad

A partir del modelo de árbol de decisiones (Quinlan, 1986; Blockeel y De Raedt, 1998), se seleccionaron 25 variables inicialmente, considerando que para la valoración de la vulnerabilidad, cada una de ellas constituye un elemento fundamental que aumenta el nivel de vulnerabilidad. Durante el proceso de análisis de los datos se observó que la ponderación de variables debería incorporar la mayor cantidad de datos que por su calidad de representación permitieran un mejor nivel de precisión en la cuantificación de la vulnerabilidad, por lo que se procedió a una nueva selección, en esta ocasión de 30 variables. A partir de la elección de variables se ponderó el nivel de vulnerabilidad de acuerdo al número de variables presentes para cada tipo de vulnerabilidad así como para la vulnerabilidad global (Tabla 3).

Tabla 3
Elección de variables y asignación de niveles de vulnerabilidad de acuerdo al número de variables presentes

| <i>Tipo de vulnerabilidad</i> | <i>Variables</i> | <i>Número de variables que determinan el nivel de vulnerabilidad</i> |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Estructural | Paredes de material distinto de tabique rojo (generalmente presentan paredes de block) Techo de lámina Un nivel de construcción Hasta dos habitaciones de construcción como máximo Arquitectura desfavorable | Muy bajo: 1 Bajo: 2 Moderado: 3 Alto: 4 Muy alto: 5 |
| Socio-económica | Más de seis habitantes en la vivienda Algún miembro de la familia mayor de 70 años Algún miembro de la familia menor de 5 años Algún miembro de la familia analfabeto Algún miembro de la familia con discapacidad Sin servicio médico perteneciente al IMSS o al ISSSTE Sin radio ni televisión Sin teléfono fijo o móvil Con un ingreso menor a un salario mínimo por integrante Sin posibilidad de ahorrar Sin agua potable Sin drenaje Sin energía eléctrica | Muy bajo: 1 a 2 Bajo: 3 a 4 Moderado: 5 a 7 Alto: 8 a 9 Muy alto: 10 o más |
| Organizacional-institucional | Falta de un Plan de Contingencias Falta de organización vecinal Falta de apoyo gubernamental en caso de inundación Falta de conocimiento de las rutas de evacuación Falta de conocimiento de los refugios institucionales | Muy bajo: 1 Bajo: 2 Moderado: 3 Alto: 4 Muy alto: 5 |

Continuación Tabla 3

| <i>Tipo de vulnerabilidad</i> | <i>Variables</i> | <i>Número de variables que determinan el nivel de vulnerabilidad</i> |
|-----------------------------------|--------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Desfavorable percepción de riesgo | Los fenómenos naturales suelen convertirse en desastres | Muy bajo: 1 |
| | Los desastres no se pueden evitar | Bajo: 2 |
| | Los desastres suelen ser castigo divino | Moderado: 3 a 5 |
| | No se manifiesta interesado en los planes de prevención | Alto: 6 |
| | No confía en los miembros de su comunidad | Muy alto: 7 |
| | No suele seguir las instrucciones de Protección Civil | |
| | No se manifiesta dispuesto a ser reubicado a pesar del riesgo al que se encuentra expuesto | |
| Global | | Muy bajo: 1 a 6 Bajo: 7 a 12 Moderado: 13 a 18 Alto: 19 a 24 Muy alto: 25 a 30 |

Fuente: Elaboración propia a partir de la ponderación y justificación de variables significativas con base en el trabajo de campo realizado previamente.

Análisis espacial

A partir de los cuatro grupos de vulnerabilidad analizados (estructural, socio-económica, organizacional-institucional y percepción del riesgo), se llevó a cabo la representación cartográfica de la distribución espacial. Finalmente, se identificó la vulnerabilidad global, considerando la ponderación de la vulnerabilidad en todos los grupos.

A partir de esta identificación de la vulnerabilidad se elaboraron grupos de mapas para cada tipo de vulnerabilidad con diferentes criterios espaciales que permitieran identificar la distribución de la vulnerabilidad en la localidad, tanto por grupo de vulnerabilidades, como por vulnerabilidad global. Los niveles se establecieron a partir de criterios ponderados que permitieran identificar los elementos que son de mayor o menor peso en el aumento de la vulnerabilidad y en el número de variables que se consideraron para el levantamiento de datos.

Estos mapas se generaron en tres etapas: a) puntuales, por vivienda encuestada; b) por áreas, a partir de la interpolación de datos y c) por manzanas, estableciendo niveles de vulnerabilidad. Se observó que este último análisis espacial resultó el más adecuado para el manejo y gestión del riesgo.

Representación cartográfica puntual (por vivienda encuestada)

En estos mapas podemos ubicar cada una de las viviendas encuestadas y conocer sus características de vulnerabilidad. Esto nos permite identificar las viviendas con un mayor número de variables desfavorables y “que por lo tanto”, presentan un nivel mayor de vulnerabilidad. Cada vivienda encuestada está asociada al punto georreferenciado que se incorporó al momento de la entrevista a partir de un GPS.

Para la valoración de las características de vulnerabilidad estructural, en una primera instancia se consideraron cuatro elementos fundamentales: 1) que sus paredes no fueran de tabique; 2) que el techo fuera de un material distinto de losa; 3) que el número total en la casa fuera de una o dos habitaciones y 4) que solamente tuviera un nivel de construcción. Finalmente se agregó otra variable: 5) que la vivienda presentara una arquitectura desfavorable, es decir, que la altura, el largo y el ancho de los muros la hagan menos resistente. Se consideró que estos elementos son fundamentales para que una vivienda resulte más vulnerable ante un evento de inundación. El mapa muestra que las viviendas que presentan mayor vulnerabilidad se ubican, precisamente al norte del río Xelajú (Figura3).

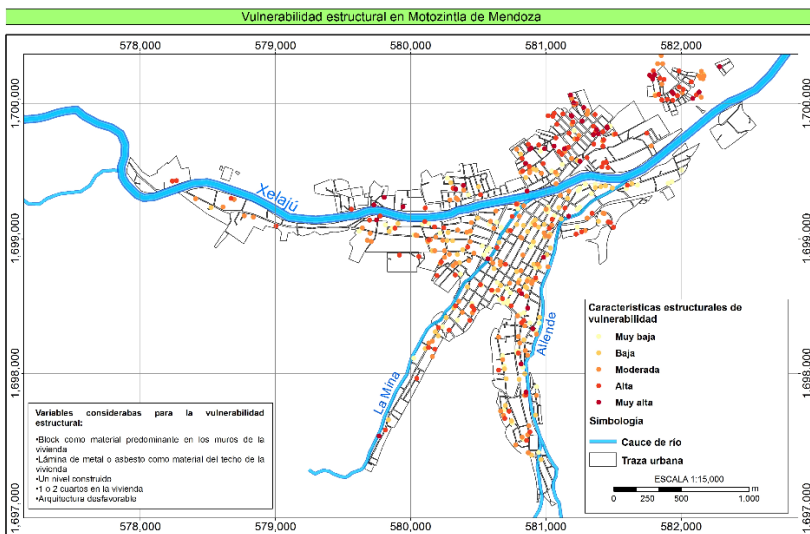


Figura 3. Mapa de viviendas en condiciones de vulnerabilidad estructural en la localidad de Motozintla de Mendoza.

Representación cartográfica por áreas a partir de la interpolación de datos

El resultado inicial arrojó puntos de mayor y menor vulnerabilidad a partir de la información obtenida sobre las viviendas y las familias levantada en campo, por lo que para identificar áreas que permitieran el análisis espacial de los resultados se utilizó el método de Interpolación de Distancia Inversa Ponderada IDW, con lo que se resaltaron áreas de vulnerabilidad para posteriormente generar una zonificación de riesgo al incorporar las áreas de peligro por inundación. Este método es de gran utilidad en el establecimiento de rangos en las condiciones que se analizan (Lahm, 1983; Watson y Philip, 1985; Philip y Watson, 1986). En un primer ejercicio se utilizó el método de Natural Neighbor, sin embargo, el resultado generó polígonos dispersos dentro del territorio, por lo que se probó con el IDW, que generó resultados de mayor congruencia con los observados en campo.

El análisis de la zonificación de los diversos elementos de la vulnerabilidad indica que las condiciones estructurales de las viviendas presentan mayor vulnerabilidad conforme se aproximan a las periferias, especialmente hacia el norte y noreste de la localidad (Figura 4).

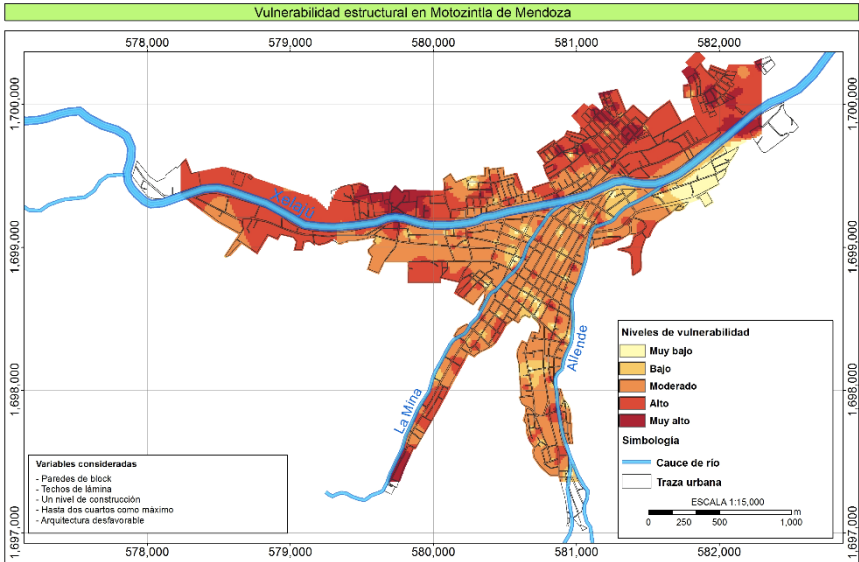


Figura 4. Mapa de vulnerabilidad estructural por áreas interpoladas en Motozintla de Mendoza.

Al considerar todas las variables seleccionadas para evaluar la vulnerabilidad global (Figura 5), se puede apreciar un patrón similar al caso de la vulnerabilidad estructural, en cuanto al aumento de vulnerabilidad hacia las periferias, sin embargo, se destaca que cuando se incorpora un mayor número de parámetros a considerar, la zonificación resultante es más homogénea en el territorio.

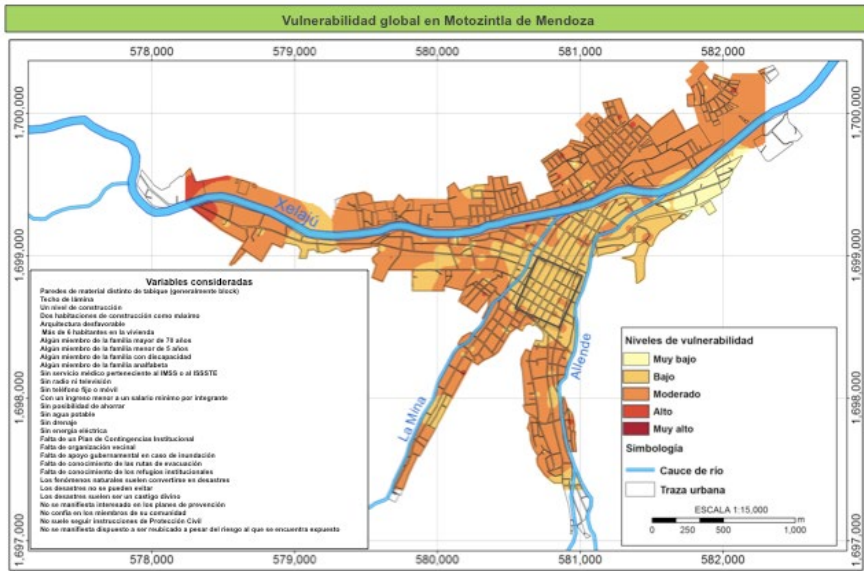


Figura 5. Mapa de vulnerabilidad global por áreas interpoladas en Motozintla de Mendoza.

Representación cartográfica por manzana a partir de la ponderación de niveles de vulnerabilidad

Con el propósito de establecer de manera más precisa la cuantificación de la vulnerabilidad, así como para favorecer el análisis espacial, se optó por incorporar la ponderación de variables de vulnerabilidad por manzana (Tabla 4). Para una mejor representación cartográfica se optó por considerar el promedio de los valores por área interpolada para ajustarlo a la traza urbana. Los resultados que se muestran en estos mapas resultan de mayor precisión, lo cual favorece la gestión de riesgo, tanto para la implementación de medidas de prevención y mitigación, como para la atención de la emergencia.

Tabla 4
Ponderación de vulnerabilidades dentro de la vulnerabilidad global

| <i>Grupo de vulnerabilidades</i> | <i>Valor ponderado (%)</i> | <i>Justificación del valor ponderado asignado</i> |
|----------------------------------|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Socio-económica | 40 | La capacidad económica y social representa el factor determinante en la vulnerabilidad, por lo que se le asignó el mayor peso. Efectivamente, cuando la población no cuenta con servicio médico institucional, se restringe la posibilidad de intervención quirúrgica, por ejemplo. Además si en la familia existe algún integrante en los grupos más vulnerables, por edad, analfabetismo o discapacidad, la vulnerabilidad aumenta. Por otra parte, si no hay suficiente ingreso para solventar la dinámica familiar, en caso de un evento de desastre, la familia queda completamente desamparada y a la espera de apoyo gubernamental, que en el caso de México suele ser deficiente |
| Estructural | 25 | Las características de la vivienda son fundamentales para la resistencia a los fenómenos naturales sin embargo, se consideró de menor influencia con respecto a la vulnerabilidad socio-económica, ya que a su vez la vulnerabilidad estructural, depende del nivel socio-económico de la familia. Así, las viviendas presentan techos de lámina y paredes poco resistentes debido a que la familia cuenta con escasos ingresos. De igual manera influye en el número de habitaciones con los que cuenta la vivienda |
| Percepción del riesgo | 20 | La percepción de riesgo impacta en todas las acciones de mitigación y prevención del riesgo de la población. Sin embargo, se consideró una vulnerabilidad menor que la socio-económica y la estructural. Fundamentalmente debido a que la mayoría de la población manifestó una percepción bastante homogénea, donde los fenómenos naturales suelen convertirse en riesgos y éstos difícilmente se pueden prevenir |

Continuación Tabla 4

| <i>Grupo de vulnerabilidades</i> | <i>Valor ponderado (%)</i> | <i>Justificación del valor ponderado asignado</i> |
|----------------------------------|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Institucional | 15 | Aunque la capacidad institucional es fundamental para la gestión de riesgo, en esta investigación se consideró como la de menor impacto de todas las que se analizaron. Debido principalmente a que los datos obtenidos en este rubro no fueron suficientemente exhaustivos para determinar con claridad su influencia en la vulnerabilidad. Por ejemplo, se consideró el conocimiento de la población acerca del plan de contingencias, a lo que los entrevistados respondieron que sí lo conocían pero no sabían en qué consistía |
| Vulnerabilidad global | 100 | En este caso se consideró la suma de las vulnerabilidades anteriores |

Fuente: Elaboración propia a partir de la ponderación de variables considerando el nivel de porcentaje de participación o incidencia en la vulnerabilidad global.

En los mapas que resultan de esta ponderación de niveles de vulnerabilidad podemos apreciar con mayor claridad las áreas de vulnerabilidad en el territorio de la localidad. Por ejemplo, el mapa de vulnerabilidad global muestra mayor diferenciación en los niveles de vulnerabilidad y por lo tanto en la identificación más clara de las áreas de mayor y menor vulnerabilidad en la localidad, lo que permite un manejo más adecuado de la información para la gestión de riesgos. La mayor parte del territorio se representa como de media vulnerabilidad en la zona centro y como de alta vulnerabilidad en el resto del territorio (Figura 6).

Peligro por inundación

Las áreas que presentan susceptibilidad de inundación (Figura 7) se determinaron a partir de la caracterización de la escorrentía en la cuenca de Motozintla realizada por Cantarero (2013), que para realizar la modelación hidráulica utilizó un SIG para integrar las capas en el modelo HEC-RAS y obtener los perfiles hidráulicos a partir de las líneas de corriente central, las orillas del canal principal y las líneas de centro de masa de flujo. Las capas se delinearon a partir de ortofotos aéreas usando el Google

Earth para su digitalización. Además incorporó una capa con el *coeficiente de rugosidad de Manning*, para describir la resistencia a la velocidad del cauce por el terreno, a partir del análisis de los usos de suelo. Además del área inundada en el evento del 4 de octubre de 2005 (debido a que considerar los eventos pasados es el primer insumo en la caracterización del peligro).

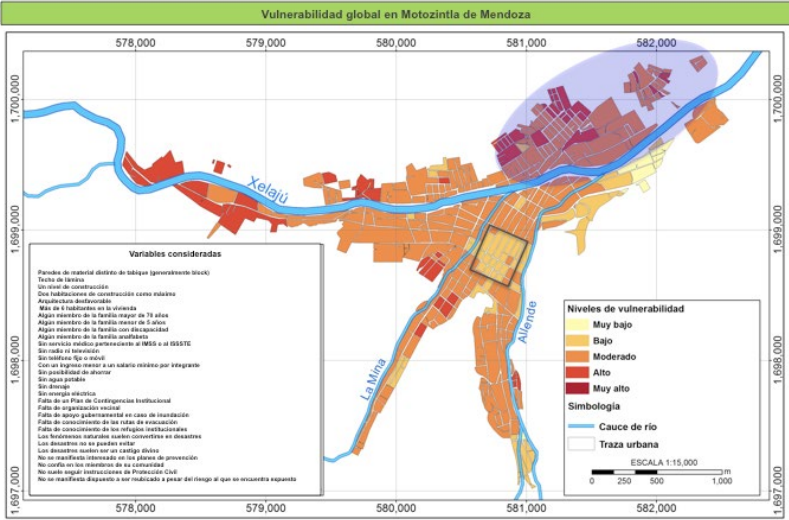


Figura 6. Mapa de vulnerabilidad global.

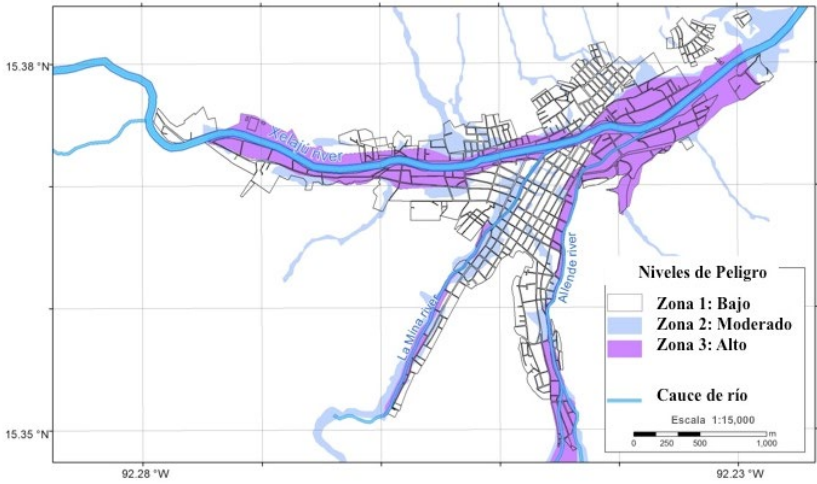


Figura 7. Mapa de susceptibilidad de peligro por inundación en Motozintla de Mendoza.

Para efectos de esta investigación se consideró únicamente la zona urbana del municipio de Motozintla, la cual corresponde a la cabecera municipal. En este mapa se puede apreciar la zona 1, con peligro bajo, como aquella donde existen asentamientos, pero donde no se tiene registro de que se hayan presentado inundaciones y que tampoco se ubican dentro del área de ninguno de los escenarios de inundación. Se designó como zona 2, peligro moderado al área que representa el escenario menos frecuente, el cual corresponde a un periodo de recurrencia de 20 años (Suárez 2006).

Finalmente, el área inundada el 5 de octubre de 2005 se designó como zona 3, peligro alto, es decir, con la mayor probabilidad de inundación. Para esta asignación se consideró que la precipitación del 4 de octubre registrada en la estación meteorológica El Novillero, localizada a unos 60 km en línea recta de Motozintla, en la parte alta de la zona, fue de 307mm en 24 horas, de acuerdo al Servicio Meteorológico Nacional (SMN) (2005) y el cual según Suárez (2006) en (Cantarero, 2013), corresponde a un periodo de recurrencia de cinco años.

Análisis espacial del riesgo

La elaboración de matrices y mapas de riesgo se realizó a partir de la identificación de las condiciones de vulnerabilidad, así como de la ponderación de las variables de la vulnerabilidad en las áreas de exposición ante inundaciones. Se consideró como área geográfica básica la manzana y se asignaron valores de vulnerabilidad considerando que los puntos encuestados por manzana representan más del 50% de las características referidas en cada mapa. Se obtuvieron cinco mapas de riesgo, uno por cada grupo de vulnerabilidades (estructural, socio-económica, organizacional-institucional y por percepción de riesgo) y el mapa final de riesgo, el cual está asociado a la vulnerabilidad global ponderada.

Para la elaboración de la matriz de riesgo (Tabla 5), se identificaron todas las posibilidades de combinación del peligro con la vulnerabilidad que generan las condiciones de riesgo en cada una de las 375 manzanas que componen el territorio de Motozintla de Mendoza. En la matriz se muestran los niveles de vulnerabilidad que se consideraron para los mapas de vulnerabilidad y para las matrices de riesgo por cada grupo de vulnerabilidad analizado.

Como se muestra en la matriz, el índice de riesgo se calculó a partir del producto de P (nivel de peligro) y V (nivel de vulnerabilidad). Este producto está expresado en la penúltima columna con un número que representa el valor ponderado del riesgo y finalmente se expresa también en forma cualitativa en la última columna.

Se construyeron mapas de riesgo por cada tipo de vulnerabilidad: estructural, socio-económica, organizacional-institucional y por inadecuada percepción de riesgo. Finalmente se elaboró el mapa de riesgo por inundación asociado a vulnerabilidad global (Figura 8).

Tabla 5
Matriz de riesgo

| <i>RIESGO</i> <i>Justificación de la escala</i> | <i>Índice de riesgo</i> | <i>Valor ponderado de riesgo</i> | <i>Nivel de riesgo</i> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------|----------------------------------|------------------------|
| Zona sin inundación y sin elementos de vulnerabilidad | 0 | 0 | Sin riesgo |
| Zona sin inundación y vulnerabilidad 1 (1x1=1) | 1 | 1 | Muy bajo |
| Zona de inundación potencial y vulnerabilidad 1 (2x1=2) Zona sin inundación y vulnerabilidad 2 (1x2=2) | 2 | 2 | Bajo |
| Zona sin inundación y vulnerabilidad 3 (1x3=3) Zona de inundación potencial y vulnerabilidad 2 (2x2=4) | 3 a 5 | 3 | Moderado |
| Zona de inundación pasada y vulnerabilidad 2 (3x2=6) Zona de inundación potencial y vulnerabilidad 5 (2x5=10) Zona de inundación pasada y vulnerabilidad 3 (3x3=9) | 6 a 10 | 4 | Alto |
| Zona de inundación pasada y vulnerabilidad 5 (3x5=15) | 11 a 15 | 5 | Muy alto |

Fuente: Elaboración propia a partir del establecimiento del índice de riesgo, considerando el peligro por inundación y el nivel de vulnerabilidad.

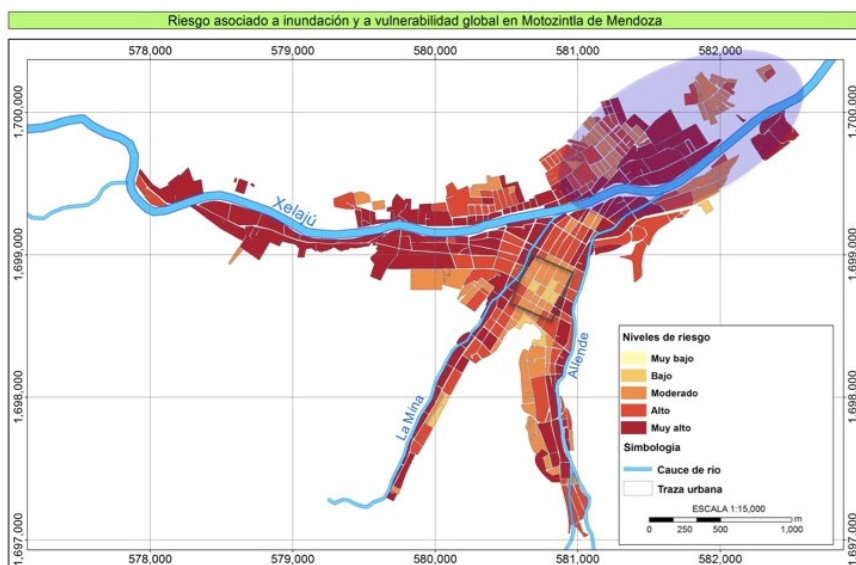


Figura 8. Mapa de riesgo por inundación en Motozintla de Mendoza.

Conclusiones

En este trabajo la valoración de la vulnerabilidad en Motozintla quedó referida a aquellos elementos característicos que generan que la población sea en mayor o menor medida susceptible a sufrir daños en su integridad física, sus bienes y su entorno ante la presencia de un evento de desastre, así como que le dificulten o favorezcan su recuperación. Por lo tanto, la vulnerabilidad global se entiende como la serie de variables que nos permiten realizar una cuantificación de la vulnerabilidad a partir de la identificación de las condiciones estructurales de la vivienda, el acceso a los servicios públicos y domiciliarios, las características socioeconómicas, la preparación ante el riesgo de desastre y la percepción de riesgo de la población. Se establece como un hallazgo el hecho de que la elección ponderada de las 30 variables consideradas fuese adecuada para lograr la diferenciación de áreas con distintos niveles de vulnerabilidad al interior de la localidad. De Vries (2011) encontró que la valoración de la vulnerabilidad global es fundamental y Puente (2009) que la ponderación de las variables que inciden en mayor o menor medida es todavía de mayor relevancia.

Los resultados indican que en cuanto a la vulnerabilidad global (Figura 6), se puede observar que aproximadamente un 60% del territorio de la localidad se ubica en el nivel alto de vulnerabilidad y un 30% en el intervalo de vulnerabilidad media. Especialmente la porción central y algunas zonas al NE presentan muy alta vulnerabilidad. En general, estas zonas corresponden a los asentamientos reubicados que el gobierno local llevó a cabo después de los eventos de desastre de 1998 y 2005.

En este sentido, se encontró que la vulnerabilidad estructural representa uno de los factores de mayor influencia en la generación de riesgo dentro de la comunidad, por ejemplo, las viviendas en general presentan la mayoría de las variables de vulnerabilidad, ya que predomina el techo de lámina, las viviendas con paredes poco resistentes y la falta de servicios. Se pudo apreciar que el nivel de riesgo que se alcanza en la mayoría del territorio con peligro de inundación es “muy alto”. Este resultado indica que en caso de una inundación, las probabilidades de que se pierda total o parcialmente la vivienda son “altas” y “muy altas”, respectivamente (Figura 4). Puente (2009) encontró que en la Ciudad de México la vulnerabilidad estructural era significativa para valorar el riesgo sísmico y dentro de los elementos que encontró como relevantes está la construcción con arquitectura desfavorable.

Al ponderar las variables de vulnerabilidad de acuerdo con su incidencia en ésta, se obtuvo una idea más clara sobre el riesgo de Motozintla ante la exposición de inundación, por lo que representa una aportación para la identificación y valoración de riesgo asociado a inundaciones, dentro de los Objetivos de la ONU para el Desarrollo Sostenible. En este sentido el Objetivo “Ciudades y Comunidades Sostenibles”, establece siete metas dentro de las cuales se encuentran el asegurar a toda la población de viviendas y servicios básicos adecuados, prestar atención especial a los grupos vulnerables, como son mujeres, niños, personas con discapacidad y adultos

mayores. Y en el marco de Sendai para la Reducción de los Desastres 2015-2030, reducir significativamente el número de muertes y daños por desastres asociados a fenómenos naturales y aumentar la gestión de riesgos de desastres en todos los niveles de acción.

En cuanto a las ventajas de la metodología presentada en esta investigación para cuantificar la vulnerabilidad y el riesgo, es factible afirmar que puede utilizarse para cualquier otro tipo de peligro o amenaza, adecuando la ponderación de variables de acuerdo con las características del nivel de peligro y de exposición. La unidad espacial seleccionada (la manzana) resultó adecuada ya que permitió manejar un nivel de acercamiento significativo en la identificación del nivel de riesgo y por otra parte resulta óptimo para la gestión del riesgo local.

La toma de datos en campo fue fundamental, ya que por una parte favoreció la corroboración de datos sobre las características estructurales de las viviendas y socio-económicas de sus habitantes y por la otra permitió la obtención de nuevos datos relacionados con la condición de la población en cuanto a estas características. La elaboración del mapa de la localidad, con la traza urbana actualizada y georreferenciada resultó la base fundamental para el análisis espacial de las condiciones de vulnerabilidad, peligro y riesgo.

El trabajo de campo también permitió obtener datos sobre las condiciones organizacionales, tanto de los habitantes, como de sus autoridades relacionadas con la prevención y la mitigación del impacto de los desastres, así como de la percepción de riesgo.

La elaboración de matrices de vulnerabilidad ponderada permitió establecer áreas de riesgo con mayor nivel de diferenciación. Finalmente, el análisis y representación de la información mediante un SIG permitió establecer niveles espaciales del riesgo dentro de la localidad. Lo cual puede favorecer la gestión integral del riesgo por parte de las autoridades locales.

Los mapas de vulnerabilidad y riesgo generados en la investigación son insumos necesarios para la gestión integral del riesgo en Motozintla de Mendoza. Esta comunidad ahora cuenta con un mapa de la traza urbana actualizado y georreferenciado, así como con mapas de vulnerabilidad diferenciada y global y mapas de riesgo asociado a estas vulnerabilidades y al peligro por inundación. Lo anterior representa una innovación y una aportación para la valoración de riesgo local en general y para Motozintla de Mendoza en particular.

Los mapas de riesgo, al considerar el peligro por inundación y la vulnerabilidad de la población, representan las bases para una adecuada gestión de riesgo, ya sea al calcular costos de las posibles pérdidas, tanto en vivienda como en infraestructura, así como en servicios y productividad, o ya sea para la atención de las emergencias, ya que aportan información valiosa para la toma de decisiones en cuanto a la evacuación y la atención prioritaria (Gobierno de España, 2013).

Considerando que la valoración del riesgo incluye cierto grado de complejidad en el reconocimiento espacial de los elementos de la vulnerabilidad especialmente para los tomadores de decisiones, se pretende que este trabajo represente una aportación metodológica para la valoración de la vulnerabilidad y el riesgo y su consecuente modificación en el patrón de gestión de riesgos, dentro de lo que cabría promover el desarrollo e implementación de medidas como: Programa de Desarrollo Urbano; Programa de Reordenamiento Ecológico Territorial; Programa de Alerta Temprana; Programa de Reducción de Riesgo; Priorización de áreas de atención en caso de contingencia o desastre; Restricción de zonas no aptas para la construcción de viviendas; Programa de Conservación Ambiental; Normas locales de cuidado y preservación ambiental; Iniciativa de Leyes y Reglamentos Municipales y el Fomento y apoyo a iniciativas de acciones comunitarias o vecinales entre otras.

Bibliografía

- Adger, N. (2006). "Vulnerability", *Global Environmental Change*, 16(3): 268-281.
- Banning, R.; Camstra, A. and Knottenerus, P. (2012). *Sampling Theory: Sampling Design and Estimation Methods*, The Hague/Heerlen: Statistics Netherlands.
- Bird, D.K. (2009). "The Use of Questionnaires for Acquiring Information on Public Perception of Natural Hazards and Risk Mitigation: A Review of Current Knowledge and Practice", *Natural Hazards and Earth System Sciences*, 9(4): 1307-1325.
- Blockeel, H. and Luc De Raedt, L. (1998). "Top-Down Induction of Logical Decision Trees", *Artificial Intelligence*, 101(1-2): 285-297.
- Bubeck P.; Botzen, W. and, Aerts, J. (2012). "A Review of Risk Perceptions and other Factors that Influence Flood Mitigation Behavior", *Risk Analysis*, 32(9): 1481-1495.
- Cantarero, F. (2013). "III.7 Zonificación de la Amenaza por Inundaciones", en *Atlas de Factores de Riesgos de la Cuenca de Motozintla, Chiapas*, coordinado por Oropeza, Oralia y Figueroa, M., Instituto de Geofísica, UNAM, 133-139.
- Centro Nacional de Prevención de Desastres (CENAPRED) (2001). *Diagnóstico de Peligros e Identificación de Riesgos de Desastres en México*, México, Secretaría de Gobernación.
- Crichton, D. (1999). "The Risk Triangle." in *Natural Disaster Management*, edited by Ingleton J. London, Tudor Rose., 102-103.
- De Vries, D. (2011). "Temporal Vulnerability in Hazardscapes: Flood Memory-Networks and Referentiality along the North Carolina Neuse River (USA)", *Global Environmental Change*, 21(1): 154-164.
- Instituto Nacional de Estadística, Geografía e Informática (INEGI) (2010a). *Censo de Población y Vivienda*, disponible en

- <<http://www.inegi.org.mx/est/contenidos/proyectos/ccpv/cpv2010/>>, consultado el 09 de mayo de 2012).
- Jacob, S.; Weeks, P.; Blount, B. and Jepson, M. (2013). "Development and Evaluation of Social Indicators of Vulnerability and Resiliency for Fishing Communities in the Gulf of Mexico", *Marine Policy*, 37: 86-95.
- Kappes, M.; Papatoma-Köhle, M. and Keiler, M. (2011). "Assessing Physical Vulnerability for Multi-Hazards Using an Indicator-based Methodology", *Applied Geography* 32(2): 577-590.
- Kaynia, A.; Papatoma-Köhle, M.; Neuhäuser, B.; Ratzinger, K.; Wenzel, H. and Medina-Cetina, Z. (2008). "Probabilistic Assessment of Vulnerability to Landslide: Application to the Village of Lichtenstein, Baden-Württemberg, Germany" *Engineering Geology*, 101(1-2): 33-48.
- Macías-Medrano, J.M. (1999). *Desastres y Protección Civil: Problemas Sociales, Políticos y Organizacionales*, Centro de Investigación y Estudios Superiores en Antropología Social (CIESAS), 101 pp.
- Macías-Vázquez J.L. y Capra, L. (2005). *Los volcanes y sus peligros: Situación actual en México y Latinoamérica*, Fondo de Cultura Económica, México, 153 pp.
- Magaña, V. (responsable) (2012). *Guía Metodológica para la Evaluación de la Vulnerabilidad ante Cambio Climático*, Instituto Nacional de Ecología y Cambio Climático (INECC), Programa de las Naciones Unidas para el Desarrollo (PNUD), México, 61 pp.
- Martinelli, A.; Cifani, G.; Cialone, G.; Corazza, L.; Petracca, A. and Petrucci, G. (2008). "Building Vulnerability Assessment and Damage Scenarios in Celano (Italy) Using a Quick Survey Data-based Methodology", *Soil Dynamics and Earthquake Engineering*, 28(10-11): 875-889.
- Mendenhall, W.; Beaver, R.J. and Beaver, M.B. (2012). "Introduction to Probability and Statistics", *Duxbury Press*, 744 pp.
- Ministerio de Agricultura, Alimentación y Medio Ambiente (2013). *Propuesta de mínimos para la metodología de realización de los mapas de riesgo de inundación. Gobierno de España*, disponible en <http://www.mapama.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/Metodologia%20mapas%20de%20riesgo%20Dir%20Inundaciones%20JULIO%202013_tcm30-98530.pdf>, Fecha de consulta: 22/06/2018).
- Novelo-Casanova, D. and Suárez, G. (2010). "Natural and Man-made Hazards in the Cayman Islands." *Natural Hazards* 55: 441-466.
- Philip, G.M. and Watson, D. F. (1986). "Automatic Interpolation Methods for Mapping Piezometric Surfaces", *Automatica*, 22(6): 753-756.
- Puente, S. (2010). "Una megalópolis en riesgo: La Ciudad de México y el desafío de la prevención de un riesgo anunciado", en *Los grandes problemas de México*, Tomo IV Medio Ambiente, coordinado por Lezama, J.L. y Graizbord, B. Centro

- de Estudios Demográficos, Urbanos y Ambientales del Colegio de México, México, 373-397.
- Quinlan, J. (1986). "Induction of Decision Trees", *Machine Learning* 1(1): 81-106.
- Saxena, S.; Geethalakshmi, V. and Lakshmanan A. (2013). "Development of habitation vulnerability assessment framework for coastal hazards: Cuddalore coast in Tamil Nadu, India —A case Study" *Weather and Climate Extremes* (2): 48-57.
- Servicio Meteorológico Nacional (SMN) (2005). "Precipitación de 04 de octubre de 2005, El Novillero, Chiapas", disponible en <<http://smn.cna.gob.mx/index.php?opcion=com>>, consultado el 10 de julio de 2012.
- Siu-Ngan Lam, N. (1983). "Spatial Interpolation Methods: A Review", *The American Cartographer*, 10(2): 129-150.
- Suárez, J. (2006). *Análisis de los problemas de erosión y deslaves en la carretera Hixtla-Motozintla, Chiapas*, Instituto de Erosión y Deslizamientos, Bucaramanga. Colombia, 34 pp.
- Vidal, R. (2007). "Carta NA IV 11 Precipitación máxima en 24 horas en milímetros", *Atlas Nacional, Instituto de Geografía*, UNAM, México.
- Vidal, R. y Gómez, R.G. (2007). "Carta NA IV 5 Precipitación Total Anual", *Atlas Nacional Instituto de Geografía*, UNAM, México.
- Watson, D. and Philip, G. (1985). "A Refinement of Inverse Distance Weighted Interpolation", *Geo-Processing*, 2(4): 315-327.

Calidad en datos geográficos, geoservicios y productores de datos: análisis crítico

Francisco Javier Ariza*
Antonio F. Rodríguez**

Recibido 02 de marzo de 2018, aceptado 29 de agosto de 2018

Resumen

En este trabajo se presenta un análisis, centrado en la calidad de datos y servicios geográficos, asumiendo que la aproximación habitual a la calidad de datos debe ser extendida a los servicios web dentro de la Arquitectura Orientada a Servicios que sirve de base a las IDE. Para tener un panorama más completo, también se considerará la calidad del productor de datos que implementa ambas clases de recursos. El análisis se realiza con un juicio crítico desde la perspectiva de las exigencias que los numerosos cambios y tendencias plantean a los tres elementos anteriores. Así, se analiza la situación actual en cuanto a estándares disponibles y prácticas habituales y se enjuician ciertas limitaciones de los estándares y métodos usuales desde la perspectiva anteriormente indicada. El trabajo ofrece líneas de mejora para superar esta situación, muchas de las cuales suponen abrir campos de investigación muy necesarios.

Palabras clave: *calidad, interoperabilidad, calidad de recursos, modelo de calidad, aspectos de la calidad.*

Resumo

Neste trabalho se apresenta uma análise, centrada na qualidade de dados e serviços geográficos, considerando-se que a aproximação habitual da qualidade de dados deve ser estendida aos serviços web dentro da Arquitetura Orientada a Serviços que serve de base às IDEs. Para se ter um panorama mais completo, também se considerará a qualidade do produtor de dados que implementa ambas as classes de recursos.

* Departamento de Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén, Campus Las Lagunillas, s/n, 23071 Jaén, España, correo electrónico: fjariza@ujaen.es

** Centro Nacional de Información Geográfica (CNIG), Instituto Geográfico Nacional (IGN) de España, Calle General Ibañez de Ibero, 3 28003 Madrid, España, correo electrónico: afrodriguez@fomento.es

A análise se realiza com um juízo crítico desde a perspectiva das exigências que as numerosas mudanças e tendências representam para os três elementos anteriores. Sendo assim, se analisa a situação atual quanto a padrões disponíveis e práticas habituais e se julgar certas limitações dos padrões e métodos usuais desde a perspectiva anteriormente indicada. O trabalho oferece linhas de melhora para superar esta situação, muitas das quais se supõe abrir campos de investigação muito necessários.

Palavras chave: *qualidade, interoperabilidade, qualidade de recursos, modelo de qualidade, aspectos da qualidade.*

Abstract

This paper presents an analysis, focused on quality of geographic data and services, assuming that the usual approach to data quality need to be extended to web services following the basic Services Oriented Architecture for SDIs. The quality of the organization implementing the two kind of resources is also considered to have a more complete picture. The analysis is carried out with a critical judgment from the perspective of the demands that the numerous changes and tendencies demand on the three previous elements. Thus, the current situation regarding available standards and usual practices is analyzed and certain limitations of the usual standards and methods are judged from the perspective indicated above. The work offers lines of improvement to overcome this situation, many of which involve opening much-needed research fields.

Key words: *quality, interoperability, resources quality, quality model, quality aspects.*

La felicidad de tu vida depende de la calidad de tus pensamientos.

Marco Aurelio

Preocúpate por la calidad, mucha gente no está preparada para un entorno donde la excelencia es lo que se espera.

Steve Jobs

Introducción

Hoy en día, es necesario concebir la actividad productiva de datos geoespaciales, o Información Geográfica (IG), en íntima conexión con el campo de las Tecnologías de la Información y las Comunicaciones (TIC). Además, se debe tener una clara orientación a la producción de recursos, ya sean bienes o servicios y, más concretamente, a la producción de modelos virtuales sobre cualquier aspecto de la realidad que nos rodea, y siempre con el objetivo de ayudar en la toma de decisiones. Estos

modelos se presentan en forma de productos de datos (p.ej. mapas o conjuntos de datos espaciales) y geoservicios (p.ej. ruterros en la web) que se utilizan en la toma de decisiones. A nivel macro son la base para el desarrollo de proyectos, planes y políticas nacionales, es decir, programación de grandes inversiones. Pero, también tienen repercusión a nivel micro, como es el quehacer de las empresas y particulares (p.ej. agricultura de precisión, manejo de vehículos no tripulados, programación de rutas y actividades de ocio, entre otros). Por tanto, se puede afirmar que generar datos espaciales es una tarea importante que tiene consecuencias reales sobre el medio y la sociedad (Oxera, 2013). Esta actividad afecta a individuos concretos en su vida, bienestar, derechos y obligaciones; por ello las administraciones encargadas de generarlos y gestionarlos deben procurar que alcancen los niveles de calidad que permitan conseguir los resultados esperados. Los datos espaciales y, en general, las tecnologías de la IG, son herramientas de la buena gobernanza y juegan un papel primordial para conseguir los objetivos de desarrollo sostenible de la Organización de las Naciones Unidas (ONU). Por todos estos motivos la calidad es un tema candente en las agendas de trabajo de las organizaciones internacionales del campo geomático (Karpik y Musikhin, 2016), apoyado por el Instituto Panamericano de Geografía e Historia (IPGH) por medio de diversos proyectos de asistencia técnica.

Al igual que otras muchas formas tradicionales de actividad (p.ej. la docencia, las relaciones sociales, las relaciones comerciales, la producción industrial y muchas otras) ha cambiado radicalmente en los últimos años y está cambiando en la actualidad. La producción de IG también se enfrenta a un nuevo marco de trabajo como consecuencia de la evolución de las TIC, la globalización y de los cambios sociales y productivos que conllevan. En esta línea, tal como muestra la Tabla 1, son numerosos los cambios radicales a los que asistimos que afectan a la producción y uso de IG, tal y como los hemos entendido hasta ahora. Estos cambios condicionan su futuro inmediato y, en particular, el de los aspectos relacionados con la calidad y, como un aspecto esencial de ella en el marco TIC, también de la interoperabilidad. En este nuevo marco aparecen multitud de nuevos usuarios (democratización del uso), pero también multitud de nuevos productores. Así se habla de democratización de la producción o Información Geográfica Voluntaria, IGV (Goodchild, 2007) y de nuevas figuras como los “prosuarios”: productores y usuarios a un mismo tiempo. Como consecuencia, están apareciendo sistemas productivos fuertemente descentralizados, en los que los esfuerzos federativos y confederativos, la hibridación de sistemas de producción (Cinnamon, 2014), etc., junto a sus consecuencias, son todavía poco claros y presentan numerosos riesgos respecto a la calidad. Además, esta situación lleva a los organismos tradicionalmente responsables de la cartografía oficial a tener que afrontar nuevos retos de cara al futuro.

En el campo industrial se habla de industria 4.0 y de revolución industrial (Oesterreich y Teuteberg, 2016), pero si bien la IG es predominantemente digital desde hace años y se encuentra estrechamente ligada a las TIC, también se requiere una cierta revolución con motivo de la nueva realidad que se está conformando. Así, según (Dans, 2010), la transformación digital de las organizaciones, ya sean públicas o privadas, es inevitable y las que no la pongan en práctica perderán oportunidades y competitividad rápidamente, hasta quedar obsoletas y verse obligadas a cerrar, en el caso de las privadas, o volverse residuos irrelevantes, en el caso de las públicas. Por ello, de una manera genérica, hablaremos de riesgo, para referirnos a todos aquellos factores clave que pueden ser decisivos, de cara al futuro, para las organizaciones no oficiales que producen IG, o que ofrecen geoservicios, y que lastran su evolución hasta impedir una transformación digital completa.

También hablaremos de determinación de la calidad, como actividad clave para prevenir esos riesgos y evitarlos siempre que sea posible o, al menos, reducirlos al mínimo.

Volviendo a la idea de la transformación digital de las organizaciones en el campo de la producción de IG, es éste un aspecto especialmente interesante en la actualidad. Por un lado, hay algunos procesos de producción cartográfica difícilmente automatizables, que hacen que la cartografía sea en parte un arte, como la generalización, la simbolización, la confluencia, la captura de datos o el case de bordes. Sin embargo, estamos asistiendo a la prometedora emergencia del aprendizaje automático (*machine learning* y *deep learning*), que ya soluciona de manera automática procesos en los que hasta ahora era imprescindible la intervención humana interactiva (véase Rodrigo, 2017), lo que permite augurar un futuro muy fructífero en esta línea.

Tabla 1
Resumen de cambios y tendencia en el campo de la IG

| <i>Cambios profundos</i> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> • Web 3.0 • La revolución del tiempo real [5] • Procesamiento paralelo, procesamiento en red, etc., OLAP [3] • Nanotecnologías, miniaturización, sensorización y nuevos materiales [4] • IPV6, Telefonía 5G, expansión de las redes [4] • <i>Blockchain</i> [7] |
| <i>Tractores</i> |
| <ul style="list-style-type: none"> • Territorios/ciudades/casas/espacios inteligentes [1] [8] • Internet de las Cosas, redes de sensores y redes de dispositivos y máquinas [1] • Máquinas y dispositivos inteligentes, autónomos y colaborativos [8] • Vehículos no tripulados [6] • Avances en internet (web semántica, IA, nube, etc.) [9] • Cambio climático y desastres [6] • Industria 4.0 [8] |

Continuación Tabla 1

| <i>Tendencias en datos</i> | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> • Web semántica [11] • BIM [8] • <i>Big Data</i> [1] [2] [5] • LiDAR [9] | <ul style="list-style-type: none"> • <i>Open Data, Linked Data</i> [11] • Ciencia ciudadana y “observation web” [6] • Coche conectado y <i>mobile mapping</i>[6] |
| <i>Tendencias en IG</i> | |
| <ul style="list-style-type: none"> • Captura masiva y barata [6] • Datos 4D nativos [9] • Desarrollo de geoservicios [9] • Administración digital [10] • Posición <i>Indoor</i> [1] | <ul style="list-style-type: none"> • Localización como servicio [5] • SIG en tiempo real y en la nube [5] • Necesidad de análisis y accesos rápidos [5] • Aplicaciones transnacionales y globales [10]. |
| <p>[1] Carpenter y Snell (2013); [2] Eldawy y Mokbel (2015); [3] Lee y Kang (2015); [4] Richardson (2017), [5] Dangermond (2017), [6] NGAC (2016), [7] Meyer, (2018), [8] Oesterreich y Teuteberg (2016), [9] Shukla (2014), [10] Kwang (2016)</p> | |

La manera en que la IG llega al usuario posee en la actualidad una esencia dual, por un lado, la IG se representa mediante datos, que se transfieren para su explotación en local en el sistema del usuario. Pero por otro, la IG se utiliza en muchos casos a través de geoservicios pensados para explotar los datos en remoto mediante peticiones que se incorporan a la lógica de la aplicación. La tendencia de cara al futuro es que esta opción última sea la dominante. Por tanto, hablar de calidad de datos geográficos desde una perspectiva de usuario, deja de tener sentido pleno si no se liga a la calidad de los metadatos y servicios que permiten accederlos. Por otro lado, aunque las organizaciones se suelen considerar como un componente más de una IDE (GSDI, 2012), no se ha considerado hasta ahora suficientemente su calidad.

Se ha de considerar que la IG tiene un valor político significativo porque permite el establecimiento integrado de políticas en diferentes sectores (agricultura, transporte, desarrollo regional, medio ambiente). El reconocimiento de esta importancia al más alto nivel político se ha desarrollado de la mano de la preocupación creciente por el desarrollo sostenible, y ya fue incluido en el Tratado de Ámsterdam en 1997 como uno de los objetivos clave de la Unión Europea. Todos los objetivos de desarrollo sostenible de la ONU tienen una vinculación con lo geoespacial, ya sea de una manera más directa y evidente (p.ej. energía asequible, acción por el clima, vida de ecosistemas terrestres, ciudades y comunidades sostenibles, etc.), como de una forma más indirecta (p.ej. la justicia social y la igualdad de género por medio de la administración de tierras). En relación con la calidad de los datos espaciales, el informe de 2010, “Iniciativa de las Naciones Unidas sobre la Gestión Global de

la Información Geoespacial” indica que, “...el incremento en las fuentes de datos, uno de los papeles clave de las agencias oficiales de cartografía para los siguientes cinco a diez años será, definir y mantener los estándares de calidad y los regímenes actuales de información con respecto de los datos que el gobierno requiriera para efectos de cumplir con sus operaciones”. Por tanto, la buena planificación y ejecución de los objetivos de desarrollo sostenible requiere datos espaciales de calidad.

Este trabajo presenta un planteamiento en tres líneas, en primer lugar, se atenderá a la calidad de los datos en sentido estricto, en segundo término, a la calidad de los servicios web basados en ellos y, por último, a las organizaciones. Todas ellas están ligadas entre sí. La calidad de un servicio depende de la calidad de los datos asociados. La calidad de un productor de datos se evidencia en parte en la calidad de los datos que produce y los servicios que implementa. A su vez, se puede considerar que la calidad de estos dos últimos se ve afectada inevitablemente por la gestión de la calidad en el productor de datos. En cada parte, obviaremos los aspectos de la calidad ya considerados anteriormente.

Las bases conceptuales de la calidad de datos son básicamente las normas ISO 19100 que establecen cómo describir de manera normalizada la calidad de la IG, sus metadatos y especificaciones. En el caso de la calidad de servicios, consideraremos esencialmente los estándares OGC y las Normas de Ejecución para servicios de la Directiva INSPIRE, un intento muy serio de caracterizar la calidad de geoservicios, que se está llevando a cabo en un universo de más de 100.000 servicios.¹ El tratamiento de la calidad en los productores de datos se ofrece desde una perspectiva centrada en la IG y no en los sistemas de gestión de la calidad (p.ej. ISO 9001 o EFQM).²

De esta forma, el objetivo de este trabajo es revisar de una manera crítica el modelo actual relativo a la calidad de datos geográficos, geoservicios y organizaciones, para considerar todos los aspectos relacionados con la calidad. Junto a esta revisión crítica, la principal aportación que realiza este trabajo es indicar un conjunto de líneas de trabajo para la mejora e, igualmente esbozar nuevos campos de investigación relacionados con la calidad.

Se ha adoptado como perspectiva dominante la centrada en los productores de datos geográficos oficiales, dado el papel preponderante y protagonista que desempeñan en el campo de la Información Geográfica.

En consecuencia, con todo lo indicado, el documento incluye las distintas perspectivas (datos, servicios, organizaciones), dedicando a cada una de ellas un apartado. Finalmente se incluyen una discusión y unas conclusiones generales.

¹ Véase <<http://inspire-geoportal.ec.europa.eu/discovery/>>.

² Fundación Europea para la Gestión de la Calidad, <<http://www.efqm.org/>>

Calidad de datos

La calidad de los datos es un aspecto que lleva recibiendo atención desde los inicios de la Cartografía como tecnología. Sin embargo, los primeros modelos o marcos conceptuales sobre todos los aspectos de la calidad que pueden concurrir en un conjunto de datos espaciales (CDE) es algo relativamente reciente. Fue en 1982 cuando en Estados Unidos se iniciaron los primeros trabajos bajo el auspicio del ACSM (*American Congress of Surveying and Mapping*). Así, se creó una propuesta de estándar (Moellering, 1987) donde se habla de adecuación al uso, de informes de la calidad y se establecieron cinco categorías de elementos de la calidad (linaje, exactitud posicional, exactitud de atributos, consistencia lógica y compleción). Es relevante indicar que todavía se sigue trabajando con estos elementos de la calidad, casi sin apenas cambios. Esta propuesta de estándar, algo modificada, fue adoptada como *Federal Information Processing Standard 173* por el *National Institute of Standards and Technology* (NIST, 1994). De esa misma época es el manual de la Asociación Cartográfica Internacional (ICA) que, editado por Guptill y Morrison (1995), fue el catalizador de una mayor preocupación internacional sobre esta temática en ámbitos universitarios e investigadores.

Casi en paralelo, el *European Committee for Standardization* (CEN) estuvo trabajando en el desarrollo de una prenorma europea (pr EN 12652:1997) centrada en la calidad de los datos (CEN, 1997). Algo después, el comité técnico 211 de ISO (ISO/TC 211) estableció las normas internacionales ISO 19113:2002, ISO 19114:2003 y la especificación ISO/TS 19138:2006. Todas ellas constituyen la base de la norma internacional sobre calidad de datos geográficos ISO 19157:2013 (ISO 2013), actualmente en vigor. Los productos de datos deben tener una calidad adecuada a su propósito que esté modelada correctamente, para lo que se requieren unas especificaciones que orienten la producción y unos metadatos que informen adecuadamente de cara a su uso. Por tanto, desde la perspectiva de la calidad se deben abarcar tanto el modelo, como las especificaciones y los metadatos. A continuación se tratará sobre cada uno de ellos.

Calidad: el modelo

El modelo de calidad de datos actualmente más extendido a nivel internacional es el propuesto en la norma internacional ISO 19157. De manera resumida, se propone un modelo basado en distintos aspectos de la calidad, los denominados elementos de la calidad (p.ej. exactitud posicional absoluta, exactitud posicional relativa, comisiones y omisiones, corrección de la clasificación, etc.). Estos elementos son cuantitativos y para su evaluación la norma internacional ISO 19157 propone un conjunto de medidas normalizadas de calidad, y unos procesos básicos de evaluación. Con elementos de calidad y ámbitos concretos en los que se aplican, se pue-

den crear las especificaciones de productos de datos (ISO 19131) e informar sobre la calidad de unos datos en sus metadatos (ISO 19115). Los elementos de calidad que se proponen son bastante limitados, pero ISO 19157 también propone un elemento denominado usabilidad, que permite soslayar, en parte, esa limitación. Además, esta norma también permite evaluar la metacalidad (la calidad de las evaluaciones de la calidad).

Este modelo presenta diferentes problemas. Desde una perspectiva centrada en las propias normas de ISO/TC211 no ha mantenido el criterio de retrocompatibilidad. Se ha cambiado un aspecto fundamental de la filosofía de ISO 19113, que era el poder definir nuevos elementos de la calidad, cosa que ya no es posible con ISO 19157. Por otra parte, este modelo está obsoleto. Desde 2003 hasta la fecha han aparecido numerosos avances en el campo de la calidad de datos, de las tecnologías de la información y de las necesidades de los usuarios que no se han tenido en cuenta, además, se han detectado numerosos fallos en ISO 19157. Empezando por lo último, es una característica común a las normas del ISO/TC211 de primera generación, las adoptadas entre el año 2000 y el 2007, el olvidar las imágenes y los datos malla. Esta circunstancia sigue presente en la norma ISO 19157, adoptada en 2013, que puede considerarse de segunda generación, pues no incluye elementos de calidad propios de imágenes. Igualmente, tampoco es totalmente adecuada para describir la calidad de metadatos, pues la calidad de campos como los de texto libre (p.ej. legibilidad, etc.) tampoco puede incorporarse adecuadamente a los elementos de calidad actualmente existentes (véase Ureña-Cámara *et al.*, 2018). ISO 19157 está centrada en productos de datos, no tanto en datos, lo que dificulta su aplicación desde la perspectiva de los procesos actuales de generación de productos y publicación de servicios por integración de datos de muy distintas fuentes. Eso supone un gran inconveniente desde la perspectiva de la trazabilidad que se requiere en conjuntos de datos complejos en los que cada dato puede provenir de una fuente distinta (p.ej. un catastro en el que cada predio procede del levantamiento de un profesional diferente). Se requiere un grano más fino en el modelo, que cada elemento (p.ej. una instancia de punto, curva o relación) lleve una “mochila” propia con sus valores de calidad y que, por medio de agregación de los valores individualizados de instancias, se pueda pasar a dar un valor de calidad general a un CDE. Esta idea es la que define Ariza-López (2013a) como micrometadatos de calidad.

Algunos aspectos adicionales relativos a la calidad de un CDE no considerados en la norma ISO 19157 son: la posesión de algún tipo de certificación de la calidad, la consistencia con otros conjuntos de datos externos y un modelo de evolución temporal de la calidad. En relación con la certificación, que es la posibilidad de mantener algún sello de “estar certificado”, este aspecto afecta tanto a las organizaciones, profesionales y productos. En ninguno de estos tres ámbitos existen esquemas de certificación. De cara al universo de aplicaciones no tripuladas que se

avecina (p.ej. coches autónomos (Wikipedia, 2018a)), tendrán especial importancia tanto la integridad de los datos como con la trazabilidad, debido a la responsabilidad que se pueda derivar de ellos (EYGM, 2017). Integridad y trazabilidad de datos serán tratadas posteriormente. En relación con la consistencia con otros conjuntos de datos externos, Giordano y col. (1994) definían la confrontabilidad como la cualidad de poder integrar conjuntos de datos procedentes de fuentes distintas, algo estrechamente ligado con la interoperabilidad de datos (Ariza López, 2013b) y que se recoge ya en algunas Normas de Implementación de INSPIRE. Por otra parte, el mundo real es dinámico y los datos son registros que pueden quedar desfasados, por ello se requiere de un modelo de evolución temporal de la calidad, ligado al ciclo de vida de los objetos geográficos y de los datos que los representan. Así, la actualidad de dos conjuntos de datos de la misma fecha, uno de Hidrografía y otro de Carreteras, por ejemplo, puede ser muy diferente en función de la tasa anual de variación de objetos geográficos. En el caso mencionado, una vez transcurridos diez años, los datos de Hidrografía pueden resultar todavía válidos, mientras que los de carreteras pueden ser completamente inservibles. En Ariza-López (2013b) se puede encontrar una primera aproximación a un modelo temporal para la calidad de datos y cuya implementación directa sobre el modelo ISO es algo compleja si no se liga a la idea de micrometadatos de calidad.

Por otra parte, dentro del propio universo ISO han surgido otras normas que se refieren a calidad de datos como, por ejemplo, la de calidad de datos de productos (ISO/TS 8000-1:2011) que desarrolla un modelo distinto. Eso es una clara limitación de cara al futuro, por ejemplo, para la integración de datos de productos con datos espaciales por medio de modelos BIM (*Building Information Model*). En campos específicos, como el de la aviación, han surgido demandas de elementos de la calidad que no se han incorporado en el modelo ISO 19157, por ejemplo, la Organización de Aviación Civil Internacional establece en su Anexo 15 requisitos de integridad (grado de garantía en que no se han perdido o alterado ninguna referencia aeronáutica o sus valores). Desde la perspectiva de los usuarios, también existe un problema en la descripción y evaluación de la calidad: el modelo actual está centrado en productos de datos de carácter general y es poco entendible para ellos. En la actualidad son muy diversos los tipos de datos que se crean (como p.ej. datos LiDAR, fotos, vídeos, multimedia, escenarios, vistas estereoscópicas, 3D, 4D, datos enlazados...), los usuarios que los usan y las aplicaciones en las que se utilizan. Junto a esta diversidad existe un problema añadido: el gran salto entre lo que se especifica y evalúa en calidad, que presenta la perspectiva de los productores tradicionales y se modela con ISO 19157, y lo que realmente necesitarían los usuarios (adecuación al uso).

La calidad de los datos es una preocupación generalizada, pero una vez que la calidad relativa a los datos se va asentando surge la necesidad de conocer y ofrecer

información más detallada y transparente sobre la calidad de los datos relativos a la calidad; este es el concepto de metacalidad: la calidad de los datos de calidad. Esta necesidad, apuntada de manera pionera por la prenorma europea de calidad ENV 12656 (CEN, 1998), se ha incorporado en la norma ISO 19157 por medio de tres elementos de metacalidad: confianza, representatividad y homogeneidad.

La evaluación de la metacalidad conlleva mayores exigencias que la de la calidad, y por ello hacen falta avances conceptuales, metodológicos y herramientas específicas, pero también la aparición de perfiles profesionales más especializados en la calidad de los datos espaciales, y también sistemas de certificación sobre la aplicación de las normas tanto a nivel de productos, procesos, organizaciones y capacidades de los profesionales. La metacalidad es algo que existe fuera del ámbito geomático desde hace mucho tiempo. Uno de los ámbitos donde la idea de metacalidad está presente es, a modo de ejemplo, el campo metrológico, donde existe una jerarquía de laboratorios que van asegurando encadenadamente la calidad de los instrumentos de medición. Incluso en la Norma Internacional ISO 9001, las exigencias sobre los equipos de medición, inspección y ensayo suponen asegurar la metacalidad instrumental y de las medidas realizadas con ellos.

Por otra parte, la falta de información relativa a la metacalidad es un riesgo en sí misma. La caracterización de unos parámetros de calidad de un CDE sin información sobre su metacalidad puede ocasionar situaciones nefastas. Por un lado, si el producto es realmente bueno, pero existe una subestimación de la calidad del CDE, al no disponer de información sobre la calidad de las estimaciones de la calidad, lo lógico será no utilizar ese recurso, con las consecuencias de despreciar una IG de valor en una toma de decisiones. Además, si realmente se pretende producir un CDE con la calidad descrita, se puede estar produciendo un derroche de recursos innecesario. Por otro lado, si el producto es realmente malo y existe una sobreestimación de la calidad del CDE, lo lógico será utilizar ese recurso, con las consecuencias de incorporarlo en una toma de decisiones. Pero aquí el productor no es consciente de que necesita aplicar mayores recursos o rigor para obtener la calidad deseada. Una producción muy descentralizada obliga a un mayor esfuerzo en la caracterización de los productores debido a su mayor número y diversidad y, como consecuencia, un esfuerzo muchísimo mayor en todo lo relativo a la metacalidad. Esto lleva, de manera directa, a la necesidad de automatizar las evaluaciones de la calidad y a tener alguna valoración sobre la confianza de los productores. Ambos aspectos son necesidades claras de cara a futuro y un campo de investigación inmediato.

Calidad de metadatos y trazabilidad

Hasta hace pocos años la cartografía era producida en forma de monopolio u oligopolio (instituciones cartográficas nacionales), y era utilizada mayoritariamente por técnicos, científicos y militares, los cuales tenían una formación que les capacitaba

para su uso, y además disponían de información relevante sobre los productos. Sin embargo, junto a la revolución TIC, el advenimiento de la globalización y la Web 2.0, se ha producido una clara democratización en el uso y producción de datos geoespaciales (Kumar, 2000). Eso ha generado que nuevos grupos sociales accedan al uso de estos productos sin la formación e información adecuadas, en un entorno de mayor abundancia y volatilidad de la oferta de productos de datos. En esta situación, los metadatos y, en particular, la descripción de la calidad son un elemento clave.

En un paradigma de producción descentralizada de IG, para que el sistema sea realmente eficaz en el uso de los recursos existentes, se requiere que toda la IG generada disponga de metadatos y que éstos estén orientados a los usuarios (Comber *et al.*, 2010). Los metadatos han de permitir conocer la existencia de un CDE y evaluar, *a priori*, la posibilidad y conveniencia de usarlo (p.ej. extensión, calidad, licencias, etc.). Podemos afirmar que lo que no está catalogado no existe, pues no se podrá encontrar en la red ni explotar correctamente. Por otra parte, consideramos que lo que no está bien catalogado constituye una fuente de riesgo, pues puede inducir usos inadecuados.

Si bien los metadatos son simples datos, se incluyen en este análisis pues tienen un carácter singular, son un elemento clave y, en general, presentan numerosos problemas específicos desde la perspectiva de su calidad. Por otra parte, en un modelo orientado a objeto, datos y metadatos deben estar encapsulados y gestionarse como un todo, en la medida de lo posible, a lo largo de todo el ciclo de vida de los datos, desde su adquisición hasta su explotación final. La norma multiparte ISO 19115 establece el modelo de metadatos actualmente más extendido en el campo de la IG. Se trata de un conjunto de normas de gran extensión y complejidad, con gran cantidad de elementos opcionales y condicionales. Con ese conjunto de partes se han normalizado, en gran medida, la estructura y formato físico de los metadatos. Sin embargo, la gran cantidad de opciones en la norma, la proliferación de perfiles y, sobre todo, el que su aplicación no haya estado acompañada de especificaciones rigurosas y criterios sobre cómo generar sus contenidos, hacen que los contenidos generados al aplicarla no estén suficientemente normalizados en la práctica.

Desde la perspectiva de informar sobre la calidad de manera más pormenorizada, la norma ISO 19157 propone un informe independiente de calidad. Sin embargo, el informe tiene formato libre y no está normalizada ni siquiera una estructura mínima para su redacción. Generar estructuras de informes para aplicaciones concretas y ayudas para su aplicación es una tarea clara a bordar en un futuro inmediato. Además de esa falta de normalización, posiblemente los cuatro problemas principales de la calidad de metadatos sean la omisión, la falta de actualidad, la falta de rigor y la falta de criterio. La omisión es la falta de metadatos, a este respecto todavía son muchos los conjuntos de datos, y en general los recursos, que existen

sin que se disponga de unos metadatos adecuados. Esa carencia también afecta a contenidos que, si bien son opcionales en el modelo ISO 19115, realmente tienen una gran importancia y utilidad, pero no se suelen incluir en aras de disponer de unos metadatos mínimos.

La falta de actualidad es consecuencia del modelo que se ha implantado para la gestión de los metadatos. Según Lopez-Pellicer (2018), poseen un modelo cuasi analógico, en el sentido de que, si bien están almacenados en digital, la concepción con la que se manejan es casi la de una ficha bibliográfica en papel, pues no se usa la potencialidad de los enlaces. Además, el modelo ISO 19115 es muy pesado, ello dificulta su aplicación a la hora de crear metadatos automatizados para conjuntos de datos creados por integración de datos de múltiples fuentes. Se requiere un modelo de metadatos a nivel de datos que sea capaz de poder agregarse de manera automatizada, y de permitir la trazabilidad.

Podemos definir la trazabilidad como la capacidad de conocer el historial, ubicación y trayectoria de un producto o lote de productos a lo largo de la cadena de suministros en un momento dado, a través de unas herramientas determinadas (Wikipedia, 2018b). La trazabilidad de los productos es algo que está asumido en muchos campos de la industria (p. ej. automovilística, espacial, alimentaria, electrónica, etc.) y que también han de llegar al campo de la IG. La trazabilidad de datos ya se ha introducido en otros campos como el *master data management* (MIKE2, 2018). Por ello cabe esperar que en el caso de la IG sea una demanda que se consolide y tome fuerza en un futuro próximo. En este caso, la generación de CDE por integración de otros CDE supone un reto tecnológico importante en la gestión de la trazabilidad. La falta de trazabilidad genera numerosos riesgos en los productos. El no conocer los suministradores, procesos, operadores, métodos, etc., deja al productor sin capacidad de exigir responsabilidades, y sin capacidad de mejorar el producto de cara al futuro. En la industria (p.ej. de componentes electrónicos) la falta de trazabilidad está estrechamente ligada al denominado mercado informal o paralelo (*grey market*), donde es usual que los productos y procesos no cumplan realmente las especificaciones del producto de marca. Por ello, las grandes multinacionales no dejan de controlar la trazabilidad de sus productos y suministros, por lo que, por analogía, una producción de IG de calidad internacional debe cuidar ese aspecto.

Por otra parte, tanto el modelo de metadatos actual, como las herramientas para su generación y gestión, transmiten una perspectiva tradicional, en la que se siguen procedimientos similares a los analógicos pero en un entorno digital. Así, no se toma ventaja de los enlaces/vínculos, de los datos enlazados, de los sistemas de sindicación, etc., que facilitan las tecnologías actuales y, por tanto, el planteamiento es poco adecuado para el mantenimiento actualizado de los metadatos a lo largo del tiempo. Esta situación genera que cualquier evolución, tanto en los datos como en

todo lo relacionado institucionalmente con ellos (p.ej. en los responsables, en las direcciones de contacto, etc.), sea muy pocas veces recogido con suficiente agilidad como para mantener los metadatos suficientemente actualizados.

Por falta de rigor nos referimos a los numerosos problemas de contenido de muy diversa índole: falta de exactitud posicional en las coordenadas de los *Bounding-Box*, inconsistencias dentro de sus propios contenidos, valores fuera de rango, equivocaciones en las clasificaciones asignadas, resúmenes y propósitos de escasa legibilidad o de valor casi nulo, faltas de ortografía, topónimos confundidos, linajes que aportan muy poco, etc. (véase Ureña-Cámara y col. 2018, que presenta una evaluación de la calidad de los metadatos de la IDE de España).

Finalmente, a menudo se detecta la falta de un criterio uniforme y homogéneo que haga que los metadatos completados por diferentes personas, o por una misma persona en distintos momentos, sean comparables y, por lo tanto, útiles para hacer búsquedas y comparar resultados. Esa situación se debe esencialmente a la falta de profesionales documentalistas especializados en la gestión de información geográfica que asuman esas tareas. Es lo que hace que se eche de menos la figura del curador de datos.³

Efectivamente, junto al modelo y las herramientas, posiblemente otro factor que ha influido en la baja calidad de los metadatos es que no hayan sido generados por profesionales en esa materia, muchas veces se han dejado para el final del proceso productivo y en manos de operadores no expertos. Además, si bien el modelo ISO 19157 se puede aplicar a la calidad de los metadatos, lo cierto es que no es algo usual por lo que las especificaciones de los metadatos muchas veces han estado huérfanas de parámetros de calidad que definieran los niveles deseados en el producto de metadatos.

Como consecuencia de las deficiencias y problemas descritos se vislumbran algunas tendencias que pueden contribuir a elevar la calidad de los metadatos en un futuro:

- Simplificación de los modelos de metadatos, reteniendo los ítems más útiles y relevantes. Por ejemplo, los metadatos INSPIRE están formados por 19 ítems para datos y 17 para servicios, frente a los más de 200 de la norma ISO 19115-1:2014.
- Aparición progresiva de la figura del curador de datos geoespaciales, encargado de la catalogación, preservación y curación de conjuntos de datos.
- Automatización del proceso de generación de metadatos, probablemente aplicando inteligencia artificial y aprendizaje automático (*machine learning*) e información de flujos de trabajo automatizados.
- Navegabilidad de metadatos de recursos acoplados, es decir que sea posible relacionar mediante los oportunos ítems de enlace, los metadatos de unos

³ <<https://www.orbemapa.com/el-bibliotecario-de-datos/>>

datos con los metadatos del servicio de visualización que los publica, éstos a su vez con los metadatos del servicio de descarga a través del cual se pueden conseguir y así sucesivamente.

- Gestión integral y encapsulada de datos y metadatos en una aproximación orientada a objeto que gestione los metadatos necesarios en cada fase del ciclo de vida de los datos, desde la adquisición hasta el almacenamiento como datos históricos.

La calidad y las especificaciones

En el campo de la industria está claro que no se puede obtener un buen producto sin unas buenas especificaciones. Es más, desde un punto de vista conceptual se puede decir que un producto no está bien definido si los usuarios no disponen de sus especificaciones. En este sentido, se dispone de la norma ISO 19131 relativa a las especificaciones de producto de datos. Consideramos que esta norma debiera ser aplicada por todas las organizaciones públicas que poseen competencia en la producción de datos geoespaciales, si bien su grado de implementación es bastante escaso. En su aplicación ocurre muchas veces como con los metadatos, que se adopta una postura de mínimos que hace que los productos no estén completamente especificados. A menudo el problema no es la propia norma, sino la falta de documentos técnicos adicionales que la complementen pues, en general, hay poca propensión a detallar y especificar bien los procesos de producción por el coste de gestión que ello supone. Puesto que se utiliza el modelo de calidad descrito anteriormente, en la aplicación de ISO 19131 también se dan las limitaciones mencionadas respecto a la adecuación al uso. Centrados en este punto, y dado que numerosos conjuntos de datos son utilizados por los usuarios en procesos concretos que pertenecen a dominios de aplicación específicos, y muy diferentes entre sí, en las especificaciones se hace cada vez más necesario el incluir determinaciones de la calidad funcional requerida. La especificación de calidad funcional supone dar el salto desde la calidad de los datos (p.ej. exactitud posicional horizontal absoluta de 5 m al 95%), a algo más aplicado y satisfactorio (funcional) en una aplicación concreta (p.ej. que el 95% de las longitudes de los cauces derivados del modelo de elevaciones son adecuadas para su uso hidrológico directo en un tipo de aplicaciones concretas).

En un marco donde las TIC aseguran la interoperabilidad de muchos tipos de datos, los datos espaciales siguen ofreciendo numerosos problemas a ese respecto. El marco desarrollado por el TC211 de ISO no deja de ser un marco orientado a la interoperabilidad, pero es muy general y se ha centrado casi en exclusiva en las interoperabilidades técnica y sintáctica. Así, se requieren más esfuerzos en busca de una mayor interoperabilidad semántica que permita el uso integrado y sinérgico de los conjuntos de datos provenientes de distintos productores. La mejor manera

de conseguir la interoperabilidad, ya sea dentro de un mismo país o en una región (p.ej. Europa o el continente americano), es partir de diseños por medio de unas buenas especificaciones. En esta línea, de cara al futuro, se requiere de mucho mayor esfuerzo en la definición de especificaciones de productos concretos para dominios de aplicación concretos, en lo que los procesos federados o confederados serán de gran relevancia.

Las especificaciones adolecen prácticamente de los mismos problemas y deficiencias que los metadatos. En esta línea, buena parte de las ideas esbozadas y conclusiones relativas a metadatos (p.ej. figura del curador, *machine learning*, etc.) son aplicables a las especificaciones.

En una primera y somera aproximación, consideramos que los aspectos más relevantes sobre los que hay que centrar la atención para que sean realmente útiles a los usuarios son: i) normalización, ii) compleción y corrección, iii) aplicabilidad y iv) inclusión de la adecuación al uso del producto (calidad funcional). En esta última línea, se requieren medidas de la calidad que no sean “datocéntricas” (como las propuestas en ISO 19157) sino “usocéntricas”, y una usabilidad definida de una manera más rica. Por tanto, se pueden considerar, al menos, dos líneas de investigación e innovación todavía no iniciadas hasta el momento: a) la profundización en medidas de calidad de datos centradas en usos concretos (casos de uso), y b) establecimiento de modelos que permitan describir y cuantificar de manera normalizada la calidad de unas especificaciones de producto, una vez que se ha descrito adecuadamente el propósito con el que se producen los datos.

Calidad de datos y su aplicación

Para concluir este apartado dedicado a la calidad de los datos, los metadatos y las especificaciones, se presenta un pequeño estudio realizado el 1 de abril de 2018 consultando la información disponible en 19 páginas web de otras tantas organizaciones nacionales responsables de la cartografía oficial (*National Mapping Agencies*) del continente americano. Así, se ha encontrado que tan sólo en 11 ocasiones ($\approx 58\%$) se publican metadatos de los datos y servicios disponibles, que sólo una organización (5,3%), publica información de calidad de sus datos más allá del linaje y que solo en 6 ocasiones (32%), se publica información descriptiva de los productos de datos geográficos disponibles, si bien en 4 ocasiones se denomina “Documentación técnica” y únicamente en un caso, se presenta como especificaciones.

La pregunta que surge de modo natural es ¿Por qué, si bien la generación y publicación de metadatos está relativamente extendida, no ocurre lo mismo con las especificaciones y sobre todo con la calidad de la IG? Consideramos que hay varias razones que pueden ser responsables de la escasa implantación de procesos de determinación de la calidad:

- Carencias en la producción de datos geográficos y en su puesta al día, lo que hace que en ocasiones los responsables prefieran invertir todos sus recursos en la producción de información y no detraer parte de ellos para hacer determinaciones de la calidad.
- Realizar determinaciones de la calidad suele ser costoso, consume recursos, requiere un modelo de calidad y expertos que no siempre existen y, además, devuelve en ocasiones resultados no deseados o al menos, digamos políticamente poco correctos, por lo que la motivación para realizar los correspondientes trabajos es baja.
- No existen, al menos en conocimiento de los autores, organizaciones que se dediquen a certificar la conformidad con la norma ISO 19157, lo que supone un desincentivo para su aplicación, al no ser un mérito reconocido por terceros.
- En la producción de otros tipos de información gubernamental (p.ej. estadística o meteorológica), no se realizan determinaciones y publicación de parámetros de calidad, por lo que los gobiernos no están habituados a aplicar modelos de calidad a la información que producen.

Sin embargo, el considerar un modelo de calidad de datos geográficos, e implementarlo, tiene ventajas estratégicas para los productores de datos oficiales:

- Puede marcar una diferencia positiva entre las fuentes de datos oficiales y la información de otros proveedores de datos y servicios no oficiales (p.ej. OpenStreetMap, GeoNames o Google Maps).
- El tener una calidad bien descrita y controlada parece uno de los requisitos exigibles a los datos fundamentales de un país, generados con el propósito de servir para georreferenciar todo tipo de datos temáticos.
- Matiza muy conveniente la responsabilidad legal de los productores de datos geográficos oficiales, e informa convenientemente a los usuarios de sus limitaciones. Sin información de calidad, los usuarios ingenuos pueden tener la impresión, y de hecho la tienen a menudo, de que los datos son perfectos y no tienen errores, con lo que pueden formular quejas e incluso emprender acciones legales si sufren daños o perjuicios debido a errores en los datos geográficos. Sin embargo, si se publican datos geográficos acompañados de una descripción de su calidad, por un lado, se difunde el mensaje de que los datos no son perfectos y tienen sus limitaciones. Por otro lado, el productor es responsable nada más (y nada menos) que de garantizar la calidad especificada.

Como se deduce del panorama mostrado, la situación no puede considerarse como avanzada o positiva. Son muchos los inconvenientes que existen y, posiblemente, pocas las recompensas. Consideramos que la introducción de esquemas de certificación específicos en el campo de la calidad de datos espaciales, para las organizaciones, productos y profesionales, podría ser un aliciente para la mejora en este campo. Siempre habría organizaciones interesadas en mostrar su excelencia y

ejercerían de ejemplo para las demás. Otra posible línea, es la implantación de métodos de comparación (*benchmarking*) que puedan ser aplicados por terceros para evidenciar las distintas situaciones de cada productor (p.ej. organización, país, etc.). El modelo de madurez país relativo a la calidad de la IG propuesto por Xabier y col. (2017), dentro del proyecto de asistencia técnica del IPGH “Diagnóstico de la situación actual sobre las metodologías y procedimientos empleados para la evaluación de la calidad de la Información Geográfica” ofrece una opción de gran interés.

Calidad de servicio

En ciencias de la computación, un servicio es un componente abierto, autodescrito y que admite la composición rápida y de bajo coste de aplicaciones distribuidas. El proveedor de servicios es la organización que implementa, brinda soporte técnico y comercial, y proporciona las descripciones de un servicio. Estas descripciones pueden incluir información sobre el servicio, su interfaz, ubicación, comportamiento e incluso calidad. Los servicios web proporcionan una manera sistemática y ampliable para la interacción entre aplicaciones, que está construida sobre protocolos web existentes y basadas en estándares abiertos.

La calidad de un servicio (QoS, de *Quality of Service*) es compleja debido a tres características exclusivas de los servicios: intangibilidad, heterogeneidad e inseparabilidad de la producción y el consumo, sin embargo, es algo muy necesario, tanto en los servicios que se nos ofrecen como personas físicas (p.ej. un servicio de restauración, o alojamiento), como en los servicios web que utilizamos junto con datos espaciales. Qué supone la calidad de servicio en el ámbito web es algo sobre lo que no existe un acuerdo generalizado. Sin embargo, como muestra la Tabla 2,⁴ son muchos los aspectos y, por tanto, términos, que se utilizan para tratar sobre este asunto.

Tabla 2
Términos relacionados con la calidad de servicio

| <i>Términos usados para la calidad</i> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Total |
|----------------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|-------|
| Availability | | x | x | x | x | x | x | x | x | | | x | x | | | 10 |
| Reliability | x | | x | x | x | x | x | x | x | x | | x | | | | 10 |
| Security | | x | x | x | x | x | x | | x | | | x | | | x | 9 |
| Capacity or throughput | | x | x | | x | x | | x | x | | | x | x | | | 8 |
| Response time | | x | x | | x | x | x | x | x | | | | | | x | 8 |

⁴ Para evitar posibles confusiones con los términos, se dejan en inglés.

Continuación Tabla 2

| <i>Términos usados para la calidad</i> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Total |
|----------------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|-------|
| Performance | | | x | x | x | | x | | | x | | x | | | x | 7 |
| Accuracy | | | | | x | | | | | | x | | x | | x | 4 |
| Scalability or Adaptability | | | x | | x | x | x | | | | | | | | | 4 |
| Transaction Integrity | | | | x | x | | | | x | | | | | | x | 4 |
| Accessibility | | | | x | x | | | | x | | | | | | | 3 |
| Assurance or consistency | x | | | | | x | | | | | | | | | x | 3 |
| Completeness | | | | | | x | | | | | x | | x | | | 3 |
| Compliance | | | | x | | | | x | | | | | x | | | 3 |
| Cost | | | | | | x | x | | x | | | | | | | 3 |
| Interoperability | | | | | x | | | | x | | | | | x | | 3 |
| Latency | | | | | | | x | x | | | | | | x | | 3 |
| Popularity | | | | | | x | | | x | | | | | | x | 3 |
| Empathy | x | | | | | x | | | | | | | | | | 2 |
| Reputation | | | | | | x | | | x | | | | | | | 2 |
| Responsiveness | x | | | | | x | | | | | | | | | | 2 |
| Successability | | | | | | | | x | x | | | | | | | 2 |
| Usability | | | | | | | x | | | | | | | | x | 2 |

[1] Zeithaml y col. (1990); [2] Menascé (2002); [3] Brahmamath y col. (2002); [4] Mani y Nagarajan (2002); [5] Lee y col. (2003); [6] DeLone y McLean (2003); [7] Kang (2007); [8] Al-Masri y Mahmoud (2008); [9] Kim y col. (2011); [10] Yang y col. (2006); [11] Subbiah y col. (2007); [12] INSPIRE (2008); [13] Gao y col. (2009); [14] Khaled y col. (2010); [15] Wu y col. (2011).

Fuente: Xabier y col. 2013.

Indudablemente, debe incluir la conformidad con el estándar seleccionado y, en una primera aproximación, lo que se suele entender como calidad de servicio operacional. En el marco INSPIRE la QoS se define mediante tres aspectos:

- Disponibilidad, que se define como la probabilidad de que el servicio de red responda a una petición correcta. En INSPIRE se mide como un porcentaje de disponibilidad anual. Para determinarlo se deben realizar peticiones cada 6 minutos, como especifica el marco INSPIRE (Comisión Europea, 2009), que deben estar perfectamente definidas y descritas. En el Reglamento INSPIRE (Comisión Europea, 2009) se exige al menos un 99% de disponibilidad anual excepto tiempos de parada técnica anunciados con una semana de antelación. Hay que hacer notar que para un mismo tanto por ciento de disponibilidad, la exigencia es mucho mayor cuanto menor es el tiempo de medida (Tabla 3).
- Rendimiento, que se define como velocidad de respuesta. Se mide en el marco INSPIRE mediante el tiempo de respuesta. Es el tiempo que se tarda en recibir el primer bit de respuesta a una consulta dada. Para calcular medias anuales de tiempos de respuesta de cada servicio se suelen emplear las mismas peticiones que sirven para determinar la disponibilidad. El marco INSPIRE (Comisión Europea, 2009) establece un límite de 5 segundos para un servicio WMS de visualización y 3 segundos para un servicio CSW de catálogo.
- Capacidad, que se define como el número máximo de peticiones simultáneas de servicio que se responden con un rendimiento garantizado. Por tanto, es la posibilidad de responder un pico de peticiones, como por ejemplo 20 peticiones por segundo durante un minuto, manteniendo el tiempo de respuesta establecido (Comisión Europea, 2009).

Tabla 3
Tiempo máximo de parada semanal, mensual y anual para una misma disponibilidad

| | <i>Semanal</i> | <i>Mensual</i> | <i>Anual</i> |
|---------|----------------|----------------|-------------------|
| 90,00 % | 17 h 5 min | 73 h | 36 días 12 h |
| 95,00 % | 8 h 31 min | 36 h 30 min | 18 días 6 h |
| 99,00 % | 1 h 45 min | 7 h 20 min | 3 días 15 h |
| 99,50 % | 51 min | 3 h 39 min | 1 día 19 h 40 min |
| 99,99 % | 1 min 3 s | 4 min 30 s | 53 min |

Hay que decir que en ocasiones faltan criterios claros y objetivos para elegir el periodo de medida tanto del rendimiento como de la disponibilidad. En ambos casos, la práctica más frecuente es determinar su valor anual, y adquirir algún tipo de compromiso de calidad con los usuarios en cuanto a mantener un umbral mínimo en cada caso. Elegir un periodo anual tiene la ventaja de que posibles fallos o episodios de baja calidad, si no son muy acusados, se compensan a lo largo del resto del año. El inconveniente es que, ante un episodio de muy baja calidad, hay que esperar todo un año para conseguir que el indicador arroje un buen resultado.

En ese sentido, para informar mejor al usuario, es siempre recomendable aplicar el principio de transparencia y publicar datos de rendimiento y disponibilidad diarios o semanales para complementar la información que dan los indicadores anuales.

Por otro lado, entendemos que, si la disponibilidad y el rendimiento se definen adecuadamente, la capacidad es redundante. Efectivamente, si un servicio en producción durante un año completo satisface el nivel de disponibilidad exigido y, al mismo tiempo, es capaz de proporcionar un tiempo de respuesta medio adecuado durante el 10% del tiempo diario que soporta mayor carga de trabajo, demuestra tener capacidad suficiente para soportar las condiciones de explotación reales.

Por último, otras fuentes (Mani y Nagarajan, 2002; Branmath *et al.*, 2002 y Lee *et al.*, 2003) consideran otros aspectos adicionales, de los que se indican los más importantes en la Tabla 4.

Tabla 4
Aspectos de la QoS adicionales a los considerados en el marco INSPIRE

| <i>Aspecto de la QoS</i> | <i>W3C (1)</i> | <i>IBM (2)</i> | <i>Brahnmath (3)</i> |
|--------------------------|----------------|----------------|----------------------|
| Escalabilidad | X | | X |
| Robustez | X | | |
| Control de excepciones | X | | |
| Exactitud | X | | X |
| Integridad | X | X | |
| Accesibilidad | X | X | |

[1] Lee y col. (2003), [2] Mani y Nagarajan (2002), [3] Brahmmath y col. (2002).

En donde (Tabla 4):

- La **escalabilidad** es la capacidad de un servicio de incrementar los recursos informáticos disponibles, de manera que su capacidad aumente cuando aumenta el número de peticiones por unidad de tiempo.
- La **robustez** representa la capacidad de un servicio de funcionar correctamente cuando le llegan peticiones erróneas o con parámetros disparatados y poco habituales.
- El **control de excepciones** alude a la calidad de un servicio de describir de manera adecuada en un mensaje de error el motivo por el que no puede dar respuesta a una petición incorrecta.
- La **exactitud** se define como la tasa de respuestas erróneas devueltas por el servicio web ante peticiones correctas.
- La **integridad** es la capacidad del servicio de evitar accesos no autorizados a los datos y *software* almacenados, lo que tiene que ver con la seguridad, y la capa-

cidad de mantener la consistencia de los datos durante las actualizaciones, lo que es importante en los servicios transaccionales.

- La **accesibilidad** representa si el servicio web es capaz o no de responder a las peticiones de los usuarios.

Puede entenderse fácilmente que no todos los aspectos de calidad de servicio mencionados son independientes entre sí y, por ello, puede ser de interés centrarse en los que presentan mayor independencia. Bajo esta idea, consideramos que pueden ser más relevantes: la robustez, el control de excepciones y la exactitud. No obstante, es necesario investigar en los aspectos de la calidad que más relevancia tengan para los servicios geográficos: estudiar qué medidas y métodos de determinación son viables y qué aspectos es razonable y útil considerar para mejorar su explotación.

Lo indicado hasta ahora en este apartado son aspectos tradicionales de calidad de servicios, pero el uso diario de los geoservicios destaca la existencia de otros muchos aspectos de su calidad que debieran ser tenidos en consideración y que, hasta la fecha, no han sido suficientemente tratados en la literatura. Así, a continuación, se presentarán algunos aspectos relevantes que ofrecen una perspectiva novedosa que debe ser abordada por la calidad.

Un aspecto no considerado hasta ahora es la calidad del contenido de la respuesta debida no a los datos, suficientemente cubierto por los aspectos considerados de calidad de datos, sino a cómo presenta los datos el servicio. Es decir, cómo está definido y configurado el servicio web. Por ejemplo, para un Servicio Web de Mapas esto incluye entre otros aspectos: a) el sistema de referencia de coordenadas elegido, b) la distribución de datos en capas, c) las escalas a las que es visible la información de cada capa, d) la calidad del documento de capacidades (*capabilities*), e) la calidad de la representación, etc. Algunos de estos aspectos pueden desplegarse muchísimo, así la calidad de representación abarcaría la leyenda, los colores elegidos, los símbolos puntuales, lineales y superficiales, los tipos de letra y su tamaño, etc. Muchos de estos aspectos son cruciales en geoservicios y, sin embargo, poco tenidos en cuenta habitualmente en los estudios.

Lo lógico sería disgregar la calidad de un servicio en calidad de cada una de las operaciones que admite y hablar así de calidad de *GetMap*, calidad de *GetCapabilities* y calidad de la operación opcional *GetFeatureInfo*, si está implementada. Especialmente importante nos parece la elección de una paleta de colores adecuada, basada no solo en el tono sino también en el brillo, adaptada para todo tipo de daltonismo, más adaptada a la visualización en pantalla que al papel y que tenga especialmente en cuenta las pantallas de los teléfonos inteligentes.

En otros servicios, como el servicio de catálogo, este aspecto incluiría también la calidad de cada una de sus operaciones, entre las que habría que tener especialmente en cuenta la cosecha de metadatos (*harvesting*), que resulta crítica porque

permite sincronizar catálogos y, en la práctica, origina muchas dificultades cuando los catálogos están implementados sobre programas diferentes.

También los servicios de descarga de ficheros exigen parámetros de calidad específicos, como la velocidad de descarga, la tasa de interrupciones, la variedad de formatos disponibles y la calidad de la descripción del servicio.

Por último, para mencionar otro ejemplo de servicio básico y muy usado en una IDE, los servicios de nomenclátor (*gazetteer*) pueden presentar un amplio abanico de características de las que depende su calidad, como: opciones de autocompletado y corrección de errores en los términos de búsqueda, búsquedas inteligentes que tengan en cuenta la posición del usuario y su historial de peticiones, posibilidad de admitir cualificadores borrosos (p.ej. cerca de, lejos de, en la dirección de...), búsquedas de términos similares y relacionados, ordenación de resultados, etc.

De nuevo estamos ante aspectos poco estudiados que constituyen candidatos a tener en cuenta a la hora de definir los programas de investigación e innovación a acometer en el campo de la IG, más concretamente en el campo de las IDE.

Calidad de productores de datos

Los apartados anteriores, centrados en los datos y servicios, tienen una perspectiva técnica centrada en la IG y en las TIC. En este apartado, incluimos uno de los factores más importantes: las organizaciones donde se produce la IG en un contexto de descentralización, como es frecuente en la actual producción de datos espaciales. Ninguno de los aspectos tratados en los demás apartados de este documento es independiente de la organización y del contexto en los que se materializan, por lo que también es relevante su análisis.

Antes de avanzar, en línea con las ideas de Cinnamon (2014), conviene aclarar que consideramos que la producción de datos espaciales, presente y futura, no se puede entender como un sistema binario con polos en IGV y agencias cartográficas tradicionales. Por el contrario, debe entenderse como un continuo que puede alentar el desarrollo de hibridaciones que aprovechen lo mejor de cada uno de los participantes. Desde esta perspectiva, los aspectos de la calidad son aún más críticos que antes y, por ello, el enfoque de adoptamos no es el que tradicionalmente se considera, centrado sobre las organizaciones responsables de cartografía oficial, donde los sistemas de gestión de la calidad (p.ej. aplicación de ISO 9001, Barrot y Pla, 2018), los modelos de calidad (p.ej. IECA 2013a, 2013b) y las herramientas (Mesterton y Kivekäs, 2018), son los aspectos sobre los que suele centrarse la atención. En este trabajo nos preocupa la IG y por ello el análisis se centrará en los que, a nuestro modo de ver, consideramos son los mayores riesgos inmediatos en este paradigma de producción descentralizada de IG en el que cualquier agente social se convierte fácilmente en productor de datos geoespaciales:

- Producción: no comprometerse con todo lo que significa.
- Acceso: no facilitar accesos de calidad.
- Trazabilidad y metacalidad: no mantener el linaje e información sobre la calidad de la calidad.
- Preservación: no asegurar el uso futuro.

A continuación se presentan en mayor detalle cada uno de estos aspectos.

Producción: no comprometerse con lo que significa

Podemos pensar que actualmente la producción de IG es sencilla pues está muy facilitada por las numerosas y cada vez más baratas tecnologías de captura. Pero esto no es del todo cierto. La producción de datos geoespaciales en la actualidad requiere más compromisos que los que se asumían anteriormente. Los modelos de datos que se están proponiendo y utilizando son cada vez más sofisticados, por ejemplo, hoy en día es usual hablar de: ciclo de vida de los fenómenos del mundo real, ciclos de vida de los datos, ciclos de vida de los productos, dato único, identificadores únicos permanentes, sellos temporales, datos enlazados, sindicación de datos espaciales, microdatos, reglas de implementación, etc.

La captura es mucho más sencilla que antes, pero ahora el valor añadido es también mucho mayor que antes. Se trata de una producción mucho más tecnificada, compleja y dependiente del conocimiento, que ha de ser escalable y también capaz de evolucionar más ágilmente que nunca para estar al nivel de exigencia de las tecnologías TIC que utilizan los usuarios. Realmente no se producen datos, se producen modelos digitales que ofrecen distintas perspectivas virtuales de la realidad.

La producción IGV presenta notables problemas en lo relativo a la calidad de los datos (Senaratne y col., 2017), pero este no es el único problema de la producción descentralizada. La producción ha de suponer, ahora más que nunca, un compromiso mantenido a lo largo del tiempo y una visión sistémica de todo el proceso de producción. Desde esa perspectiva de la producción descentralizada de los datos, los principales riesgos que vislumbramos se relacionan con el fraccionamiento y debilidad de los productores con poca experiencia, en ocasiones aquellos que se han incorporado en los últimos años a la producción de datos espaciales alentados por las actuales facilidades para la captura de datos y para ofrecer geoservicios. Entre otros, los problemas principales pueden ser:

- Relativos a la producción continuada
 - Falta de perspectiva temporal. Básicamente ocurre en agentes noveles, que consideran la creación de datos desde una perspectiva episódica y aislada. La creación de datos espaciales debe suponer un compromiso a largo plazo, pero si bien es lógico que ese compromiso no sea asumido por productores

- neófitos, ocasionales o voluntarios, también es verdad que en organizaciones oficiales no siempre se asume con el debido rigor.
- Falta de viabilidad institucional. No siempre existe un mandato legal explícito para la producción de datos, lo que no impide que se aborde. Así, hoy en día se producen datos geoespaciales por los más variados motivos, algunos coyunturales, como una bajada de costes y se considera una tendencia a seguir. Pero, en ocasiones, la organización no asume ese compromiso de cara al futuro por lo que, al menor cambio, se deja de producir.
 - Relativos a los aspectos tecnológicos
 - Exigencias tecnológicas crecientes. La generación/captura puede ser cada vez más sencilla y económica pero el mantenimiento, preservación, inclusión de ciclos de vida, creación de unos buenos metadatos, evolución tecnológica sostenible de los datos, etc., suponen cada vez más exigencias añadidas a la producción. Sólo unos pocos con financiación y personal altamente cualificado pueden acometer con éxito su incorporación al modelo de producción.
 - Acceso mediante servicios. En el paradigma de producción de la cartografía tradicional, e incluso de datos digitales sobre soportes físicos (p.ej. DVD, CD-ROM, etc.), el producto se materializaba en un recurso físico accesible, aunque el productor hubiera dejado de existir. En el paradigma actual y futuro, los datos se ofrecen y utilizan por medio de servicios web, si es el productor quien los ofrece y ese productor deja de existir la cuestión es ¿se seguirán ofreciendo esos servicios?
 - Relativos a la fragmentación de la producción.
 - Idiosincrasia. Si bien las organizaciones tradicionales productoras de datos espaciales son bien distintas y presentan un comportamiento distinto debido a sus experiencias vitales, personal, *know-how*, etc., con la llegada de nuevas formas de producción, como es la IGV, aparecen patrones muy distintos que hay que saber entender para utilizar bien sus productos. Así, si bien la IGV se ha puesto de moda como forma de crear datos espaciales (p.ej. *OpenStreetMap*, GeoNames), es de destacar que los datos creados por la mayoría de esos movimientos tienen notables limitaciones (p.ej. falta de cubrimiento homogéneo de todo el espacio geográfico, grandes niveles de omisión, no existe una clasificación única de los elementos, calidad semántica baja, etc.).
 - Interoperabilidad. La fragmentación de la producción y aparición de nuevos productores supone graves riesgos para la interoperabilidad de los datos si no se refuerzan las acciones de coordinación. En un marco de producción

descentralizada se requiere un mayor compromiso de todas las partes para que los datos sean interoperables y así conseguir la sinergia que posibilita el uso conjunto. El aspecto de la interoperabilidad abarca numerosas perspectivas y ya ha sido tratado.

Acceso: no facilitar accesos de calidad

En la actualidad los productos de datos espaciales, se ofrecen en forma de servicios vía web, por lo que también se ha de pensar en éstos como elementos críticos. En el contexto TIC es obligado considerar un paradigma descentralizado en la oferta de servicios de acceso a datos espaciales. En este sentido, dado que en otro apartado se tratará sobre la interoperabilidad, y que en el epígrafe anterior se ha esbozado el mantenimiento de los servicios a lo largo del tiempo, este epígrafe se centra en la gestión de la calidad de servicio. Muchos productores de IG, muchos de ellos con escasos recursos (p.ej. pequeñas municipalidades, organizaciones no gubernamentales, particulares, etc.), podrán llegar a generar IG de calidad e interés para la comunidad; pero la capacidad de ofrecer servicios con accesos de calidad (p.ej. disponibilidad, tiempo de respuesta, seguridad, fiabilidad, etc.), puede quedar fuera de su alcance. De esta manera, el riesgo para la calidad de los servicios de datos geoespaciales consiste en no tener una disponibilidad efectiva del recurso por limitaciones en la prestación del servicio debido a que no todos los proveedores de datos geoespaciales tienen capacidad económica ni tecnológica para ofrecer servicios de calidad.

Interoperabilidad: no posibilitar la utilización y uso legal

Bajo nuestro punto de vista, los mayores riesgos de la producción descentralizada son los relativos a la interoperabilidad. La interoperabilidad es la base de las TIC y por ello es la piedra angular de las IDE y de la Geomática 3.0 tal como se ha definido anteriormente. Según Rodríguez y col. (2013), la interoperabilidad es la idea clave hacia la que se orientan todas las actividades de normalización y estandarización en el campo de la IG. ISO la define formalmente (ISO, 2008) como: “capacidad para comunicar información, ejecutar programas o transferir datos entre unidades funcionales de manera que el usuario solo necesite tener un conocimiento escaso o nulo sobre las características particulares de tales unidades” [ISO/IEC 2382-1:1993].

En el campo de las IDE una definición de interés es la que proporciona la Directiva INSPIRE (Parlamento y Consejo Europeos, 2007): “posibilidad de combinación de los conjuntos de datos espaciales y de la interacción de servicios, sin intervención manual repetitiva, de forma que el resultado sea coherente y se aumente el valor añadido de los conjuntos y servicios de datos”.

Para conseguir la interoperabilidad se requieren acuerdos entre las partes. Estos acuerdos se materializan en forma de normas internacionales, especificaciones, estándares, reglamentos, etc. Por lo que la interoperabilidad está ligada íntimamente a las acciones de coordinación y liderazgo que resulten en documentos como los indicados anteriormente.

La interoperabilidad abarca distintos campos que van desde lo más tecnológico o duro, *hard* (p.ej. formatos, sistemas de referencia, semánticas, etc.) a lo más suave, *soft* (p.ej. aspectos institucionales, aspectos legales de licencias de uso, etc.) y no por ello menos costoso. La falta de interoperabilidad en los datos es posiblemente el riesgo más evidente y que más hemos sufrido en el trabajo con IG a lo largo de las últimas décadas. Ejemplo de esto son los problemas de intercambio de formatos, los problemas de integración de modelos de datos y semánticas distintas, los problemas de interoperabilidad debidos a discrepancias posicionales, etc. Siguiendo con lo estrictamente relativo a los datos, las técnicas de confluencia o también denominadas de integración o fusión de datos (Xavier *et al.*, 2016), son las encargadas de solventar estos problemas (p.ej. confluencia vertical, confluencia horizontal, (Ruiz-Lendínez y col. 2011)). Pero también hay otros aspectos no menos relevantes, como las limitaciones legales, que pueden surgir del uso CDE con modelos de licencia que no resulten compatibles. En el caso de la producción descentralizada llevada cabo por las administraciones, también se puede hablar de interoperabilidad institucional, la cual puede abarcar muchas facetas que impidan el uso conjunto efectivo de datos.

En principio, la descentralización se puede considerar un movimiento de incremento de la entropía y, como tal, que va en contra de alcanzar acuerdos técnicos, que son la base de la interoperabilidad. Un mayor número de productores de datos espaciales, muchos de ellos de carácter sectorial, otros amateurs, otros esporádicos, etc., tendrán tendencias naturales a crear “desorden”. Por ello, la producción descentralizada de datos geoespaciales supone notables riesgos de interoperabilidad si no existe una labor eficaz de liderazgo para alcanzar consensos que la favorezcan. Sin embargo, en contraposición a todo lo indicado previamente, ejemplos como INSPIRE nos demuestran que un liderazgo fuerte permite aprovechar la variedad y diversidad de los sistemas descentralizados para aprovechar lo mejor de cada uno creando un proyecto conjunto. Una de las claves del éxito en el liderazgo de un sistema descentralizado es facilitar especificaciones y herramientas que resulten cómodas para que nadie se salga de la corriente principal.

Preservación: asegurar el uso futuro de los datos espaciales

La debilidad del almacenamiento digital ha llevado a acuñar el término *digital dark age* como aquella posible situación futura en la que sería imposible o muy difícil

acceder a documentos del pasado por estar almacenados en soportes y formatos digitales obsoletos. A nivel mundial existe verdadera preocupación por este asunto y se están tomando iniciativas vinculadas con el desarrollo de las denominadas e-infraestructuras. La IG es un tipo de contenido digital de gran valor económico y que interesa conservar por motivos legales, administrativos, científicos, culturales, sociales y de toda índole (Ariza-López *et al.*, 2012).

La calidad está relacionada con el uso adecuado, satisfactorio y fructífero. Sin embargo, un factor limitativo del uso y beneficio futuro del uso de la IG es su preservación o curación. Para ello se requiere un conjunto de metadatos especiales (distintos de los propuestos en ISO 19115 o en Dublin Core) que son los que deben permitir la caracterización de los contenidos digitales y, a partir de ellos, permitir la gestión de los procesos de preservación. En esta línea, el TC/211 de ISO está desarrollando la norma internacional ISO 19165 sobre preservación de datos y metadatos. Junto a los metadatos para la preservación también se requiere una organización y gestión adecuada del sistema que, en parte, puede ser parecido al paradigma IDE. Este marco debe cubrir adecuadamente:

- La necesidad de incluir la perspectiva de preservación y explotación futura de los datos desde el diseño de los productos de datos espaciales mediante el diseño de un ciclo de vida de producto.
- Considerar la necesidad de establecer distintos alcances temporales en la preservación, como por ejemplo el corto (5 años), medio (20 años) y largo plazo (100 años), según la importancia de los datos.
- Establecer un conjunto de metadatos especiales pensados para la preservación e integrables con las normas ya existentes.
- Establecer las bases para la prueba de los sistemas de preservación que puedan implantarse.
- Establecer las bases para una infraestructura federada de preservación a nivel nacional.
- Evaluar la calidad (eficacia y eficiencia) de los sistemas de preservación.

La producción descentralizada ocasiona productores y producciones más dinámicas y efímeras, y heterogeneidad en la preservación. Estos dos aspectos dificultan la preservación futura de la IG y suponen un claro riesgo de ocultación o desaparición de conjuntos de datos, lo que origina, a su vez, claros retos para la investigación.

No afrontar una verdadera transformación digital

Para acabar este apartado, consideramos importante retomar la idea de la necesidad verdadera transformación digital de las organizaciones dedicadas a la producción de datos espaciales y en la manera de servirlos. El contexto que definen, por la parte de

generación de datos, la existencia de múltiples sensores, la internet de las cosas, los espacios inteligentes, los BIM, la realidad virtual y la realidad aumentada, etc., y por la parte de las aplicaciones los vehículos no tripulados para el transporte de personas y mercancías, la agricultura de precisión, la robótica y mecatrónica, etc., hace pensar que realmente se requiere otra forma de producir datos espaciales y que para ello es necesario un cambio radical. Esta transformación digital debe ir orientada a asegurar conjuntos de datos únicos (*single source of Truth*, SSOT), dotados de identificadores únicos persistentes, globales y enlazados, que den soporte continuado (espacio y tiempo) a las múltiples aplicaciones inteligentes y robotizadas (automatizadas).

Desde esta perspectiva, se podría pensar que serán exitosas aquellas organizaciones que sean capaces de implementar sistemas de servicios que ofrezcan sus datos SSOT para que sean utilizados masivamente por terceros (individuos, empresas, organizaciones, administraciones, etc.) en la sociedad más allá de las TIC.

Discusión

De manera paralela a la necesidad de un desarrollo sostenible, la sociedad actual se encuentra sometida a un continuo de revoluciones tecnológicas (nanotecnologías, miniaturización, tiempo real, sensorización, etc.), y de cambios tractores (Ciudades inteligentes, Internet de las cosas, datos enlazados, vehículos autónomos, Industria 4.0, etc.), y cambios sociales, que nos llevan a un paradigma donde los datos son el combustible del futuro. En este paradigma los datos espaciales son un elemento clave, imprescindible y de valor en alza. Estos cambios afectan a la forma de generar datos geoespaciales, a su uso y a las exigencias sobre ellos, estableciendo nuevos retos para la calidad de esos datos geoespaciales, de los geoservicios y de las organizaciones que los producen.

En los apartados anteriores se ha realizado una descripción y crítica de la situación actual, al igual que algunas propuestas sobre qué conviene investigar y trabajar de cara al futuro. Como se ha indicado, son visiones parciales, pero lo que realmente consideramos más crítico es la falta de una visión sistémica, global, y sostenible en el tiempo, de todos los aspectos intervinientes (datos, servicios y organizaciones), en la que los datos geográficos fundamentales (datos geoespaciales cuyo propósito es servir para georreferenciar otros datos, que serán temáticos) y no solo la calidad, sino también la eficacia en solucionar los problemas de calidad que los usuarios reclamen, están llamados a ser cada vez más esenciales. En una sociedad sostenible, como la que se propugna desde la ONU, hacen falta datos geoespaciales de calidad, adaptables, actualizados y coherentes entre sí, como parte de un modelo sostenible de creación, sistematización, uso y preservación de datos espaciales.

A nuestro entender, esta visión sistémica y sostenible no existe, ni con una perspectiva global ni más parcial o sectorial. Indudablemente, lo que se sugiere es complejo y será difícil de plasmar, pero no se parte de cero: como se ha evidenciado en el texto, existen numerosos elementos (p.ej. calidad de datos, calidad de servicios, calidad de los productores de datos, modelo de aseguramiento de la calidad de datos espaciales, etc.). Todos estos elementos, convenientemente extendidos y adaptados, pueden ser parte de la solución para conformar el modelo que se requiere. Así, por ejemplo, consideramos que el modelo propuesto en IECA (2013a) para el aseguramiento de la calidad en la producción de datos geoespaciales, puede ser adaptado y extendido para incluir aspectos de calidad de metadatos, geoservicios y de datos preservados. Indudablemente, realizar la propuesta queda fuera de las posibilidades de este trabajo, pero lo que sí podemos y debemos hacer es preguntarnos ¿por qué no existe aún ese modelo? La respuesta no es sencilla ni única, pero apuntando posibles problemas, también apuntaremos posibles soluciones. Bajo nuestro punto de vista son varios los factores:

- Complejidad. Se trata de un tema realmente complejo. Requiere de un trabajo interdisciplinar y de expertos. Existen numerosos perfiles de expertos, pero justamente el de expertos en calidad de datos geoespaciales, ya sea en su vertiente de evaluación de la calidad o de gestión de la calidad, no es nada abundante. Además, no se conoce ningún foro que los convoque o haya convocado para hacerlos trabajar de manera conjunta en esta temática.
- Escasos datos abiertos. Existe una escasa implantación de políticas de datos abiertos. Según establece el Global Open Data Index, tan solo el 11% de los conjuntos de datos que considera claves se publican como datos abiertos,⁵ porcentaje que baja hasta el 10% en los datos geoespaciales más esenciales.⁶ Una apertura general de datos podría ser un revulsivo para que los usuarios utilizaran y compararan masivamente los datos geoespaciales, y se promoviera su calidad.
- Idiosincrasia de la calidad. Para los productores de cartografía oficial, que son los que más necesitan de la visión que se propone aquí, la calidad es un aspecto sobre el que existe poca transparencia. Preguntarles sobre la calidad de sus datos, sobre los procesos de aseguramiento, control, etc., no siempre es bien recibido. La calidad es algo interno, oneroso y sobre lo que prefieren pasar de puntillas. No se invierte en formación específica y tampoco se dedican esfuerzos internos y, mucho menos, existe diálogo y coordinación externa para avanzar. Por su parte, para las empresas privadas que viven de los datos y servicios geoespaciales (p.ej. TomTom, Where, Google y otras), muchos aspectos de la calidad forman parte de su *know-how* y, como es lógico, no están dispuestos a

⁵ <<https://index.okfn.org/>>.

⁶ <<https://index.okfn.org/dataset/map/>>.

compartirlos. Por otra parte, en las agencias cartográficas oficiales hay una escasa implantación de una cultura de calidad de producto final. Pocas categorías de datos producidos por los gobiernos van acompañadas de información de calidad; como ejemplo significativo podemos mencionar los datos estadísticos que, a pesar de que no suelen ir acompañados de información de calidad son utilizados para todo tipo de decisiones de gran trascendencia.

- **Prioridades de normalización.** Uno de los problemas fundamentales reside en cómo se realiza la normalización y, en concreto, las normas de datos geoespaciales, y las barreras para su utilización. Los comités técnicos de normalización son muy especializados y están focalizados sobre visiones muy parciales, cuando la calidad es algo más transversal. Así, aún bajo el paraguas de un mismo cuerpo de normalización, la cooperación entre comités no siempre es tarea fácil. Centrados en el caso del TC211 de ISO, y después de 24 años de existencia, podemos afirmar que su actividad sobre la calidad de datos no ha sido todo lo amplia que podría haberse esperado, y ello no significa crear normas internacionales para todo, existen otros documentos ISO que también pueden tener gran valor. Así, la calidad de las imágenes no se ha abordado, la calidad de los geoservicios tampoco, el Anexo D de medidas de calidad apenas se ha extendido, la exactitud posicional no se ha abordado, etc. Junto a lo ya indicado, NGAC (2017) reconoce la existencia de notables barreras a la utilización de estándares, entre ellas la falta de herramientas conformes, la falta de iniciativas para hacerlos funcionar, la no inclusión en los programas académicos, etc.
- **Comunidades de productores.** Los temas de calidad son complejos y no siempre una única organización puede abordar proyectos complejos, por ello es importante trabajar con otras organizaciones que compartan los mismos problemas. Sin embargo, salvo excepciones (p.ej. Eurogeographics), no existe un alto grado de asociacionismo de instituciones productoras de cartografía oficial, en las que se aborden temas técnicos como el aquí planteado, y donde se propongan soluciones convenidas entre todas ellas. Lo mismo ocurre a nivel de profesionales especializados en calidad de datos. No tenemos conocimiento de ninguna asociación que agrupe los intereses de profesionales expertos en la materia y que realice propuestas técnicas.
- **Falta de responsabilidad legal.** Hasta la fecha de hoy, la producción de datos oficiales, salvo para los casos náutico y aeronáutico, no tiene una exigencia de responsabilidad que lleve a los productores a esforzarse en todo lo relacionado con la calidad, la trazabilidad, preservación, etc. de sus datos. Un marco legislativo más exigente sería un revulsivo para dedicar más esfuerzos a la temática con una visión más global y sistémica. Posiblemente, este marco pueda llegar de mano de la regulación de sectores como los vehículos autónomos.

- Falta de certificación. La falta de aliciente de un reconocimiento externo (p.ej. sello, acreditación, certificación, etc.) que destaque frente a la sociedad la bondad del trabajo realizado también puede actuar como factor negativo a la hora de tomar iniciativas. Así, como se ha indicado, un modelo de madurez, desarrollado a nivel de país u organización, y reconocido por tercera parte, podría ser muy positivo. Por un lado, ofrecería una guía de lo que hay que hacer (lo que se evalúa) y por otra, la oportunidad de que los esfuerzos fueran reconocidos.

Conclusiones

En este trabajo se ha presentado un análisis centrado en la calidad de los datos, la forma en que se ofrecen (los servicios) y quienes los ofrecen (las organizaciones). Cada una de estas perspectivas es complementaria y no debe abordarse sin una visión conjunta. Sin embargo, a pesar de la extensión del trabajo, son numerosos los aspectos que no han sido tratados.

La IG, su producción y puesta en “servicio”, se ven afectadas por los numerosos cambios y tendencias tecnológicos y sociales que tienen lugar, muchos de los cuales suponen verdaderas revoluciones respecto a las formas de hacer y de entender la IG de hace unos años, y suponen numerosos retos desde las perspectivas de la calidad que interesan en este trabajo.

Centrados en la calidad de datos, podemos concluir que existe un modelo de calidad consolidado conceptualmente (ISO 19157), pero escasamente aplicado, como se ha demostrado por el estudio realizado en las páginas web de diversas agencias cartográficas nacionales. Este modelo está muy limitado en sus posibilidades respecto a los retos inmediatos que se plantean tanto en tendencias de datos como de IG (Tabla 1). El grano, la inclusión del tiempo, la trazabilidad, elementos y medidas de la calidad usocéntricos, etc. son aspectos que no están bien resueltos en la actualidad.

La calidad de los datos se evalúa poco y se han apuntado posibles causas de ello. Evaluar la calidad de los datos es tanto más necesario cuanto más tecnificado y automatizado sea el entorno de los datos espaciales (p.ej. vehículos no tripulados, agricultura de precisión, etc.). Considerando los conjuntos de datos oficiales y fundamentales de cada país, se debe tender a una calidad aplicada, que realmente sea entendible por los que utilizan esos datos. Por otra parte, se puede minimizar el coste de las evaluaciones por medio de varias líneas de actuación, entre ellas, i) seleccionando el conjunto mínimo pero esencial de aspectos, medidas y métodos de la calidad que ofrezcan buen balance coste-beneficio; ii) desarrollando métodos automáticos o semiautomáticos de determinación de la calidad (p.ej. Xavier y col. 2015a y 2015b).

La calidad de los metadatos es algo poco tratado y, por tanto, desconocido, a pesar de su gran importancia para el funcionamiento de los servicios. Su creación podrá ser automatizada en breve por medio de la Inteligencia Artificial aplicada a procesos y flujos de trabajo, documentos, etc. Existen campos de metadatos cuya calidad puede ser automatizada y evaluada al 100%, pero otros campos relevantes como los resúmenes o el propósito apenas han sido evaluados (Ureña-Cámara y col. 2018). El modelo ISO 19115 es muy pesado y farragoso. Los metadatos están desligados de los datos y de su calidad, y eso es una grave limitación de cara al uso de datos simples, no de conjuntos de datos, para lo cual se requieren los que hemos denominado micrometadatos. Aspectos como la trazabilidad y navegabilidad requieren de perspectivas de implementación distintas a las actuales para ser funcionales. La creación y mantenimiento de metadatos con un verdadero carácter de enlazado es otro reto actual todavía no resuelto.

Las especificaciones de los productos de datos son un elemento clave de la calidad, pero su uso está muy limitado, muchas veces no están acompañadas de documentación técnica que les dé sentido. Además, se usan especificaciones de la calidad datocéntricas y no usocéntricas, lo que hace que los productos no queden bien especificados. Se utilizan descripciones de la calidad propias de la IG anterior a los SIG y no de descripciones de la calidad basadas en casos de uso concretos en dominios de aplicación determinados. No hay un método ni modelo para evaluar la calidad de las especificaciones de productos de datos, lo cual también impide mejorar (lo que no se mide, no se mejora).

En cuanto a la calidad de servicios web, consideramos que se han producido avances técnicos y realizaciones prácticas de monitorización continua de la calidad de servicio suficientemente aceptables. Como aspectos tradicionales de la calidad a evaluar, creemos que sería suficiente con tener en cuenta, como ya hemos dicho, la disponibilidad y el rendimiento, y quizás se podría añadir la estabilidad, que tiene que tendría en cuenta el número de caídas de un servicio, no solo el tiempo que dura cada caída. En este campo serían necesarios llevar a cabo estudios detallados que corroborasen o refutasen con datos objetivos esas conclusiones. Sin embargo, como se ha indicado, hay numerosos aspectos (p.ej. calidad de las operaciones, presentación de los resultados de las peticiones, etc.) que siendo importantes todavía no han sido difundidos ni estandarizados.

En lo relativo a las organizaciones en un marco de producción descentralizado y muy democratizado, se han identificado diversos riesgos que ponen en entredicho la calidad sostenida en cuanto a la producción de datos y su acceso a largo plazo cuando se trata de productores sin un mandato legal claro o de tipo esporádico. Centrados en las agencias cartográficas oficiales, un reto crítico es no afrontar la verdadera transformación digital de la sociedad que nos está llevando a un mundo digital que se configura como una realidad virtual, en muchos casos más allá de una

simple copia del mundo realidad. En la mayoría de los casos, todavía no se han desarrollado modelos de datos y de producción federada basados en SSOT, que permitan un espacio inteligente (territorio inteligente + ciudad inteligente + edificios inteligentes + etc.) de carácter continuo, donde los datos enlazados, los datos de sensores, los datos de la internet de las cosas, los datos estadísticos, los datos administrativos (p.ej. licencias urbanísticas, licencias de cultivos, etc.), utilicen la IG fundamental por medio de identificadores únicos globales persistentes como soporte de una web realmente semántica y basada en datos enlazados.

Como se ha indicado en el apartado de discusión, según nuestra experiencia sería necesario un modelo que aporte una visión más sistémica, global y sostenible en el tiempo para gestionar todos los aspectos aquí señalados, pero ese modelo no existe. En la sección de discusión se han indicado algunos aspectos que requieren nuestra atención para mejorar la situación.

Finalmente, queremos indicar que para avanzar en la línea de superar las limitaciones actuales y de afrontar los retos que se han indicado, se hace especialmente crítico y necesario contar con personal técnico bien formado en todos estos campos. Por ello la formación, en sus distintos niveles (p.ej. grados, postgrados, especializaciones, cursos cortos, etc.) es un elemento clave que nos debe preocupar y debe ser cuidado. En esta línea, ayudaría en gran medida el establecer un conjunto de p^énsum o planes de estudio que orientaran a los agentes formativos sobre los contenidos a desarrollar. Igualmente, el establecimiento de esquemas de certificación de personas ayudaría a dar mayor nivel a los técnicos con responsabilidades en este campo (p.ej. evaluación de la calidad). Junto a lo anterior, y dado que la base de la mejora es la comparación, consideramos necesario disponer de modelos de madurez que permitan a cada organización compararse con un estándar (que evolucionará a lo largo del tiempo), tal que en todo momento sepa a qué nivel de madurez corresponde su manejo de la calidad de datos, de los metadatos, de adopción de datos abiertos etc. Consideramos que, desde el IPGH se deberían liderar estas dos actividades tan necesarias en el marco panamericano.

Hemos de concluir indicando que trabajar y preocuparse por la calidad de los datos geográficos, de los geoservicios y productores de datos es trabajar y preocuparse para que se dispongan de mejores bases para la selección de alternativas, el diseño, ejecución, seguimiento y mejora de políticas, planes y proyectos que consoliden el alcance de los objetivos de desarrollo sostenible en cada uno de los países del mundo.

Bibliografía

AENOR (2003). UNE 66175 Sistemas de gestión de la calidad – guía para la implantación de sistemas de indicadores, Madrid, España, Asociación Española de Normalización.

- Al-Masri, E. and Mahmoud, Q. (2008). "Toward Quality-Driven Web Service Discovery", en *IT Professional*, 10, 24-28.
- Ariza-López, F.J. (2013a). *Calidad de la Información Geográfica: perspectivas de futuro*, en CIMA INEGI, México.
- Ariza-López, F.J. (ed.) (2013b). *Fundamentos de evaluación de la calidad de la información geográfica*, Jaén, España, Servicio de Publicaciones de la Universidad de Jaén.
- Ariza-López, F.J.; Ariza-López, R.M.; Ureña Cámara, M.A.; Cortés José, J. y Ureña López, L.A. (2012). "Preservación de la Información Geográfica: Perspectivas y situación en España", en *GeoFocus* 12, 171-200.
- Barrot, D. and Pla, M. (2018). "ISO 9001 for spatial data: ICGC experience", en *International Workshop on Spatial Data Quality*, Valletta (Malta), Eurogeographics, <<https://eurogeographics.org/wp-content/uploads/2018/06/3-SDQ2018-12.pdf>>.
- Brahnmath G.J.; Raje, R.R.; Olson A.; Auguston M.; Bryant B.R. and Burt C.C. (2002). "A quality of service catalog or software components", en *Proceedings of the Southeastern Software Engineering Conference*, Huntsville, AL, USA, 513-20 <<http://www.dtic.mil/dtic/tr/fulltext/u2/a492809.pdf>>, consultado el 01 de abril de 2018.
- Carpenter, J. and Snell, J. (2013). *Future Trends in geospatial information management: the five to ten-year vision*, United Nations Committee of Experts on Global Geospatial Information Management.
- CEN (1998). *ENV 12656:1998 Geographic information - Data description – Quality*. Brussels, Belgium, Comité Européen de Normalisation.
- Cinnamon, J. (2014). "Deconstructing the binaries of spatial data production: Towards hybridity", en *The Canadian Geographer* 59(1). DOI:10.1111/cag.12119.
- Comber, A.J.; Fisher, P.F. and Wadsworth, R.A. (2010). "User-focused metadata for spatial data, geographical information and data quality assessments", en 10th AGILE Conference, Aalborg, Denmark, Association of Geographic Information Laboratories in Europe.
- Comisión Europea (2009). *Reglamento (CE) 976/2009 de la Comisión, de 19 de octubre de 2009 por el que se ejecuta la Directiva 2007/2/CE en lo que se refiere a los servicios de red*, <http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32009R0976&from=ES>, consultado el 24 de junio de 2018.
- Dangermond, J. (2017). *Five GIS Trends Changing the World according to Jack Dangermond, President of Esri*. <<http://geoawesomeness.com/five-gis-trends-changing-world-according-jack-dangermond-president-esri/>>, consultado el 24 de junio de 2018.

- Dans, E. (2010). *Todo va a cambiar*. Barcelona, España, Editorial Deusto. Disponible en <<https://www.todovaacambiar.com/>>, consultado el 24 de junio de 2018.
- DeLone, W.H. and McLean E.R. (2003). "The DeLone and McLean model of information systems success: A ten-year update", en *Journal of Management Information Systems* 19, 9-30.
- Eldawy, A. and Mokbel, M.F. (2015). "The era of big spatial data", en *31st IEEE International Conference on Data Engineering Workshops*, Seoul, South Korea. DOI: 10.1109/ICDEW.2015.7129542.
- EYGM (2017). *The evolution in self-driving vehicles Trends and implications for the insurance industry*, EYGM Limited.
- Gao S.; Mioc D. and Yi X. (2009). "The measurement of Geospatial Web Service quality in SDIs", en *Proceedings of the 17th International Conference on Geoinformatics*. Fairfax, VA, USA, IEEE.
- Giordano, A. and Veregin, H. (1994). *Il controllo di qualità nei sistemi informative territoriali. come valutare e mantenere l' accuratezza del database*. Venezia, Italy, Il Cardo.
- González-Campos, M.E.; Bernabé-Poveda, M. y León Pazmiño, M.F. (2017). "Metodología para evaluar la usabilidad del visualizador de mapas del geoportail IDE de Ecuador", en *GeoFocus* 19, 109-127.
- Goodchild, M.F. (2007). "Citizens as sensors: the world of volunteered geography", en *GeoJournal* 69(4), 211-221. DOI:10.1007/s10708-007-9111-y.
- IECA (2013). *NTCA 01-002 Modelo Aseguramiento Calidad productos IG*, Instituto de Estadística y Cartografía de Andalucía, Sevilla, España.
- IECA (2013). *NTCA 01-003 Modelo de Calidad para la IG en Andalucía*. Instituto de Estadística y Cartografía de Andalucía, Sevilla, España.
- INSPIRE (2008). *Network Services Architecture, Version 3.0. Infrastructure for Spatial Information in Europe*, <http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_INSPIRE_NS_Architecture_v3-0.pdf>.
- ISO (2008). *ISO/TS 19104:2008 Geographic Information – Terminology*. Geneva, Switzerland, International Organization for Standardization.
- ISO (2013). *ISO 19157:2013. Geographic information. Data quality*. Geneva, Switzerland, International Organization for Standardization.
- ISO/IEC (2005). *TR 9764:2005 Information technology — Guidelines, methodology and reference criteria for cultural and linguistic adaptability in information technology products*. Geneva, Switzerland, International Organization for Standardization.
- Kang, Y. (2007). "Extended Model Design for Quality Factor Based Web Service Management", en *Proceedings of the Future Generation Communication and Networking*. Jeju, South Korea, IEEE, 484-487.

- Karpik, A.P. and Musikhin, I.A. (2016). "Research and practical trends in geospatial sciences", en *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B6, 2016, XXIII ISPRS Congress, 12-19 July 2016, Prague, Czech Republic.
- Khaled R.; Tayeb L. and Servigne, S. (2010). "Geospatial Web Services Semantic Discovery Approach Using Quality", en *Journal of Convergence Information Technology* 5, 28-35.
- Kim, E.; Lee, Y.; Kim, Y.; Park, H.; Kim, J.; Moon, B.; Yun, J. and Kang, G. (2011). *Web Services Quality Factors, Version 1.0*. Burlington, MA, USA. Organization for the Advancement of Structured Information Standards (OASIS). <<http://docs.oasis-open.org/wsqm/WS-Quality-Factors/v1.0/cs01/WS-Quality-Factors-v1.0-cs01.html>>, consultado el 24 de junio de 2018.
- Kumar, N (2000). "Automation and Democratization of Cartography: An Example of a Mapping System at CEM, University of Durham", en *The Cartographic Journal*, 37(1), 65-77, <<https://doi.org/10.1179/caj.2000.37.1.65>>.
- Kwang, T.W. (2016). *E-government global trends: integrated services, open data, e-participation and digital technologies*, <<http://www.enterpriseinnovation.net/article/e-government-global-trends-integrated-services-open-data-e-participation-and-digital>>, consultado el 24 de junio de 2018.
- Lee K.; Jeon J.; Lee W.; Jeong S.H. and Park S.W. (2003). *QoS for Web Services: Requirements and Possible Approaches. W3C Working Group Note*. World Wide Web Consortium (W3C), <<http://www.w3c.or.kr/kr-office/TR/2003/ws-qos>>, consultado el 24 de junio de 2018.
- Lee, J.G. and Kang M. (2015). "Geospatial Big Data: Challenges and Opportunities", en *Big Data Research* (2):74-81.
- López Pellicer, F.J. (2018). Comunicación personal.
- Mani, A. and Nagarajan, A. (2002). *Understanding quality of service for Web Services. Improving the performance of our Web services*. IBM. <<https://www.ibm.com/developerworks/library/ws-quality/index.html>>, consultado el 24 de junio de 2018.
- Menascé D.A. (2002). "QoS issues in Web services", en *IEEE Internet Computing* 6, 72-75.
- Mesterton, N. and Kivekäs, R. (2018). "Towards Automating Spatial Data Quality Evaluation in the Finnish National Topographic Database", en *International Workshop on Spatial Data Quality*, Valletta (Malta), Eurogeographics, <https://eurogeographics.org/wp-content/uploads/2018/06/14-NTDB_show_Malta_2018_NM_RK.pdf>.

- Meyer, A. (2018). 5 blockchain trends influencing the future of social media marketing, <<https://thenextweb.com/contributors/2018/04/19/5-blockchain-trends-influencing-future-social-media-marketing/>>.
- MIKE2 (2018). *Method for an Integrated Knowledge Environment*. <<http://mike2.openmethodology.org/>>, consultado el 24 de junio de 2018.
- Moellering, H. (ed.) (1987). *A Draft Proposed Standard for Digital Cartographic Data. Report 8*. Columbus, OH, National Committee for Digital Cartographic Data Standards.
- Mueller, P.A. (1997). *Modelado con objetos UML*. Paris, France, Eyrolles.
- NGAC (2016). *Merging technologies and the geospatial landscape A Report of the National Geospatial Advisory Committee*. NGAC Emerging Technologies Paper. Department of Interior.
- NGAC (2017). *Geospatial standards: a national asset*. NGAC Emerging Technologies Paper. Department of Interior.
- NIST (1994). *Spatial data transfer standard (SDTS). Federal Information Processing Standards Publication 173-1*. Gaithersburg, MD, USA, National Institute of Standards and Technology.
- Oesterreich, T.D. and Teuteberg, F. (2016). “Understanding the implications of digitization and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry”, en *Computers in Industry*, 83, 121-139.
- OGC (2006). *OpenGIS Web Map Server Implementation specification v 1.3.0*. Wayland, MA, USA, Open Geospatial Consortium, <<http://www.opengeospatial.org/standards/wms>>, consultado el 24 de junio de 2018.
- OGC (2012). *OGC Sensor Observation Service (SOS) Standard Version 2.0*. Wayland, MA, USA, Open Geospatial Consortium.
- OGC (2018). *OGC Testbed-13: Aviation Abstract Quality Model Engineering Report*. Open Geospatial Consortium. Wayland, MA, 01778, USA.
- OXERA (2013). *What is the economic impact of Geoservices? Prepared for Google*. London, United Kingdom, Oxera Consulting LLP.
- Presutti, V.; d’Amato, C.; Gandon, F.; d’Aquin, M.; Staab, S. and Tordai, A. (eds.) (2014). “The Semantic Web: Trends and Challenges”, en *11th International Conference, ESWC 2014*, Anissaras, Crete, Greece.
- Richardson, D. (2017). *Five Tech Trends Driving New Geospatial Development*. <<http://www.esri.com/esri-news/arcnews/summer17/articles/five-tech-trends-driving-new-geospatial-development>>, consultado el 24 de junio de 2018.
- Rodrigo, J.J. (2017). “Extracción de conocimiento mediante la aplicación de Inteligencia Artificial a la información espacial”, en *Jornadas de Ibéricas de Infraestructura de Datos Espaciales*,

- <http://www.idee.es/resources/presentaciones/JIIDE17/Apresentacoes_JIIDE2017_Extraccion_conocimiento_mediante_aplicacion_inteligencia_artificial_a_la_informacion_espacial_Jose_Julio_Rodrigo_Bello.pdf>, consultado el 24 de junio de 2018.
- Ruiz-Lendinez, J.J.; Ariza-López, F.J.; Ureña-Cámara, M.A. and Blázquez, L. (2011). “Digital map conflation: a review of the process and a proposal for classification”, en *International Journal of Geographical Information Science*, 25(9), 1439-1466. DOI:10.1080/13658816.2010.519707.
- Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C. and Haklay, M. (2017). “A review of volunteered geographic information quality assessment methods”, en *International Journal of Geographical Information Science*, 31(1), 139-167. DOI: 10.1080/13658816.2016.1189556.
- Shukla, Y. (2014). *AGI perspective on future trends in geospatial technologies. Association of geospatial industries*, <https://nrsc.gov.in/uim_2014_proceedings/papers_ppts/UIM2014_US2_AGI.pdf>, consultado el 24 de junio de 2018.
- Subbiah, G.; Alam, A.; Khan, L. and Thuraisingham, B. (2007). “Geospatial data qualities as web services performance metrics”, en *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. New York, NY, USA, ACM: 66:1-66:4.
- Ureña-Cámara, M.A.; Nogueras, F.J.; Lacasta, F.J. and Ariza-López, F.J. (2018). “A method for checking the quality of geographic metadata based on ISO 19157”, en *International Journal of Geographic Information Science* (pendiente de publicación).
- Wikipedia (2018a). Vehículo autónomo, <https://es.wikipedia.org/wiki/Veh%C3%ADculo_aut%C3%B3nomo>, consultado el 24 de junio de 2018.
- Wikipedia (2018b). Trazabilidad, <<https://es.wikipedia.org/wiki/Trazabilidad>>, consultado el 24 de junio de 2018.
- Wu, H.; Li, Z.; Zhang, H.; Yang, C. and Shen, S. (2011). “Monitoring and evaluating the quality of Web Map Service resources for optimizing map composition over the internet to support decision making”, en *Computers & Geosciences* 37, 85-94.
- Xavier E.M.; Ariza-López F.J.; Chicaiza-Mora, E.G. and Buenaño, X. (2017). *Modelo de Madurez-País relativo a la Calidad en IG*. Proyecto Diagnóstico de la situación actual sobre las metodologías y procedimientos empleados para la evaluación de la calidad de la Información Geográfica (Proyectos Panamericanos de Asistencia Técnica -PAT- del Instituto Panamericano de Geografía e Historia, 2016).

- Xavier E.M.; Ariza-López F.J. and Ureña-Cámara, M.A. (2016). “A Survey of Measures and Methods for Matching Geospatial Vector Datasets”, en *ACM Comput. Surv.* 49(2), 1-34.
- Xavier, E.M.A.; Ariza-López, F.J. and Ureña-Cámara, M.A. (2015a). “Web service for positional quality assessment: the WPS tier”, en *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3/W5*. La Grande Motte, France, ISPRS, 257-262.
- Xavier, E.M.A.; Ariza-López, F.J. and Ureña-Cámara, M.A. (2013). *Geospatial Web Services Quality: an Overview and a Quality Model Proposal*, Jaén, España, GIIC.
- Xavier, E.M.A.; Ariza-López, F.J. and Ureña-Cámara, M.A. (2015b). “WPS for positional quality control applying the method proposed in UNE 148002”, en *VI Jornadas Ibéricas de Infraestructuras de Datos Espaciales*, Sevilla, España, Grupo de Trabajo IDEE.
- Yang, C.P.; Cao, Y.; Evans, J.; Kafatos, M. and Bambacus, M. (2006). “Spatial Web Portal for Building Spatial Data Infrastructure”, en *Journal of Geographic Information Sciences* 12, 38-43.
- Zeithaml, V.; Parasuraman, A. and Berry, L., (1990). *Delivering Quality Service*, New York, USA, The Free Press.

RESEÑAS



Reproducibilidad científica: ¿Qué es y por qué debemos interesarnos en geo-ciencias?

Daniela Ballari*

Recibido 8 de octubre de 2018; aceito 22 de octubre de 2018

Resumen

A pesar de que en los últimos tiempos se ha resaltado la importancia de la reproducibilidad como pilar del proceso científico, en las geo-ciencias ésta no ha sido reconocida como una preocupación inminente. Esto no es por ausencia del problema, sino por falta de evidencias que así lo demuestren. Es por ello que la comunidad científica internacional, entre ellas la relacionada con las geo-ciencias, está promoviendo el posicionamiento de la problemática y generando evidencias y propuestas que permitan superar la llamada “crisis de reproducibilidad”. Así, el objetivo de este documento es aportar al avance de la discusión de la reproducibilidad científica en la comunidad científica y académica Latinoamericana de las geo-ciencias, como forma de abrir el debate crítico y propositivo sobre estrategias para mejorar la reproducibilidad de nuestras investigaciones.

Palabras clave: *reproducibilidad, replicabilidad, geo-ciencias, apertura de la investigación.*

Resumo

Embora a reprodutibilidade tenha sido considerada como um pilar do processo científico, nas geociências ela ainda não foi reconhecida como uma preocupação. Isto não ocorre devido à ausência do problema, mas sim devido à ausência de evidências que o demonstre. É por isso que a comunidade científica internacional, incluindo a das geociências, está promovendo a geração de evidências e diretrizes para superar a denominada “crise de reprodutibilidade”. Assim, o objetivo deste artigo é contribuir para o avanço da discussão da reprodutibilidade científica na comunidade científica e acadêmica latino-americana ligada às geociências, como

* Universidad del Azuay, Av. 24 de Mayo 7-77 y Hernán Malo, Cuenca, Ecuador, correo electrónico: dballari@uazuay.edu.ec. ORCID: <<https://orcid.org/0000-002-6926-4827>>.

forma de abrir um debate crítico e proativo sobre estratégias para melhorar a reprodutibilidade de nossa pesquisa.

Palavras chave: *reprodutibilidade, replicabilidade, geociências, abertura da pesquisa.*

Abstract

Although reproducibility has been regarded as a pillar of the scientific process, in geoscience it has not been recognized as a concern yet. This is not because of the absence of the problem, but instead because of the absence of evidences. The international scientific community, including the geoscience, is promoting the generation of evidences and guidelines to overcome the “reproducibility crisis”. Thus, the aim of this document is to bring to the Latin American Geoscientists the issue of scientific reproducibility, as well as to open the discussion about how we can improve the reproducibility of our research.

Key words: *reproducibility, replicability, geoscience, open research.*

Introducción

En la actualidad, al desarrollar o finalizar un proyecto de investigación, los hallazgos más relevantes son publicados en forma de artículos en revistas científicas, congresos o libros. Entre las razones para hacerlo se encuentran la necesidad de dejar un registro de la investigación para que otros puedan continuar avanzando en la misma dirección; recibir el debido reconocimiento por las ideas y los resultados; atraer el interés de otros investigadores hacia un área de conocimiento; recibir comentarios de expertos sobre las ideas y resultados documentados; y legitimar la investigación a través de la verificación independiente de métodos y resultados por parte de los revisores del artículo (Cargill y O’Connor, 2013). En ese sentido, hemos escuchado incansablemente que la descripción de los métodos en un artículo científico permite reproducir el trabajo y evaluar la validez de los resultados. Sin embargo, es evidente que en la mayoría de artículos publicados se omiten detalles metodológicos, parametrizaciones y decisiones de diseño que permitan garantizar la reproducibilidad. Por ello, los revisores y lectores encuentran dificultades para comprender los resultados, evaluar las limitaciones y, sobre todo, para reproducir el trabajo (Stark, 2018).

A pesar de ser reconocida como un pilar de la ciencia, en la práctica la reproducibilidad no es adecuadamente valorada y es difícil de implementar (Granell, Nüst, Ostermann y Sileryte, 2018). Así es que se ha generado la llamada “crisis de reproducibilidad científica”, cuya importancia puede ser evidenciada de acuerdo con una encuesta a más de 1,500 científicos (Baker, 2016) en la cual el 70% afirmó que no pudo reproducir los experimentos (publicados) de otros científicos, y el 50% no

pudo reproducir su propio experimento. En este último caso, no por falta de detalles metodológicos, sino por limitaciones o mal uso de la significancia estadística. Es decir, se trata del llamado *p-value hacking* (Head, Holman, Lanfear, Kahn y Jennions, 2015) que se caracteriza por hacer un uso inadecuado de los métodos estadísticos para encontrar y reportar selectivamente resultados y significancias favorables.

En las áreas de las ciencias de la vida y de la salud, tales como psicología, medicina, genética y biología, la preocupación por lograr la reproducibilidad científica es una preocupación reconocida e instalada desde hace algún tiempo. En las geociencias, sin embargo, no ha sido todavía reconocida como una preocupación inminente (Nüst *et al.*, 2018). Esto no es debido a la ausencia del problema, sino por la falta de evidencias que así lo demuestren. Por ello, Nüst *et al.* (2018) evaluaron la reproducibilidad de un grupo de artículos obtenidos de las conferencias de la Asociación de Laboratorios de Información Geográfica de Europa (AGILE — *Association of Geographic Information Laboratories of Europe*), una de las conferencias más destacadas en su tipo. Los resultados mostraron que, a pesar de que los autores de esos artículos reconocían la importancia de la reproducibilidad, los incentivos para hacerlo eran escasos y, por lo tanto, sus artículos no cumplían con un mínimo de reproducibilidad. A partir de este trabajo, un grupo de científicos de las geociencias se encuentra impulsando acciones en pro de mejorar la transparencia y reproducibilidad de las investigaciones en áreas de las geociencias y de las tecnologías de la geo-información (Gil *et al.*, 2016; Granell *et al.*, 2018; Konkol, Kray y Pfeiffer, 2018b, 2018a; Nüst *et al.*, 2018; Nüst, Boettiger y Marwick, 2018; Ostermann y Granell, 2017; Shannon y Walker, 2018).

En este punto es importante regresar a la pregunta fundamental ¿Por qué quisiéramos reproducir un trabajo? Voy a direccionar la respuesta desde dos puntos de vista según el tipo de investigación.

En primer lugar están los trabajos de tipo experimental, muy comunes en las ciencias de la vida y de la salud donde existe un inminente interés de la reproducibilidad. En este caso, cuando se obtienen nuevos datos, o más extensos, los trabajos se reproducen para confirmar una vez más la validez estadística de las hipótesis. Siguiendo este planteamiento, Ostermann y Granell (2017) investigaron la reproducibilidad (usando los mismos datos y métodos) y la replicabilidad (nuevos datos y métodos similares) de información geográfica voluntaria. Sus resultados mostraron que ninguno de los trabajos investigados era reproducible y solo la mitad era replicable.

En segundo lugar están los trabajos con alto contenido computacional, comunes en las geociencias. En este tipo de trabajos se manifiesta la necesidad de una reproducibilidad de tipo computacional (Konkol *et al.*, 2018a). Es decir, es muy frecuente que se utilice *software* y tipos de análisis o modelizaciones muy diferentes, que requieren de la vinculación de bases de datos espaciales, sistemas de informa-

ción geográficos, sensores remotos, modelización estadística y representación cartográfica. Así, se vuelve necesario migrar los datos de un *software* a otro, y de un formato a otro, para ejecutar las distintas partes del flujo de trabajo. Por consecuencia, el flujo de trabajo se vuelve complicado de reproducir. Además, cuando la interactividad con el *software* es manual (*clicks*), la reproducibilidad es incluso más complicada de realizar (Pebesma, Nüst y Bivand, 2012). Estas dificultades de reproducibilidad computacional son parcialmente superadas cuando se trabaja en un entorno de programación como puede ser R, Python o MATLAB, en los cuales se generan *scripts*, es decir, documentos de texto con la secuencia de funciones y órdenes que deben ejecutarse. De esta manera, todo el procedimiento se encuentra documentado en un único *script* que hace referencia a los datos y librerías externas a cargar, a la limpieza y pre-procesamiento de los datos, así como a las funciones específicas de análisis que deben ejecutarse en un orden determinado. A pesar que los entornos de programación facilitan el camino hacia la reproducibilidad computacional y se vinculan actualmente con los *softwares* más usuales en nuestra área (por ejemplo: ArcGIS, QGIS, GRASS, IDRISI, ERDAS), es necesario reconocer que su uso todavía no se encuentra lo suficientemente extendido dentro de las geo-ciencias.

Siguiendo los lineamientos de la reproducibilidad científica se intenta motivar a los investigadores para que publiquen no solo el texto y las figuras en un artículo científico, sino también sus datos, código y flujo computacional, así como cualquier otro producto intermedio junto con sus metadatos, para así facilitar investigaciones futuras (Gil *et al.*, 2016). A pesar que cada vez más revistas científicas permiten el envío de material suplementario como forma de facilitar la reproducibilidad, no existe ningún incentivo concreto para que los investigadores así lo hagan (Gil *et al.*, 2016; Granell *et al.*, 2018; Nüst *et al.*, 2018). En este contexto, el objetivo de este documento es traer a la arena de la comunidad científica y académica Latinoamericana de las geo-ciencias la discusión sobre la reproducibilidad científica como forma de abrir el debate y cuestionamiento sobre cómo mejorar la reproducibilidad de nuestros trabajos.

El resto de este documento se organiza de la siguiente manera. Primero se provee terminología, seguido de beneficios y recomendaciones para mejorar la reproducibilidad de artículos en las geo-ciencias.

Reproducibilidad *versus* Replicabilidad

A pesar que en los últimos tiempos se ha escuchado mucho sobre la reproducibilidad científica, no es fácil encontrar una definición precisa y única. En particular existe una confusión entre Reproducibilidad y Replicabilidad. Por ello, aquí se exponen estos dos conceptos a partir de la traducción de (Barba, 2018):

Reproducibilidad: los autores proporcionan todos los datos necesarios y los códigos para ejecutar nuevamente el trabajo, recreando así los mismos resultados.

Replicabilidad: es un estudio que llega a los mismos hallazgos científicos que otro estudio, pero recopilando nuevos datos (posiblemente con diferentes métodos) y aplicando análisis diferentes.

Beneficios de la reproducibilidad

Markowetz (2015) expresa que las razones fundamentales para invertir tiempo en reproducibilidad, (por ejemplo, “es la base de la ciencia” o “el mundo sería mejor si nuestros trabajos fueran transparentes y reproducibles”), suelen ser poco motivadores y no llevan a introducir cambios en los hábitos de trabajo y formas de conducir las investigaciones. Por ello, presentó cinco razones más pragmáticas que pueden despertar el interés en los investigadores:

1. Ayuda a evitar desastres (por ejemplo, permite detectar tempranamente errores en los análisis).
2. Hace más simple la escritura de los artículos (por ejemplo, se conocen y detallan con precisión todos los pasos metodológicos, parametrizaciones y decisiones de diseño).
3. Ayuda a los revisores a ver el artículo a la manera de los autores (por ejemplo, si se proveen los datos y el código como información suplementaria de los artículos, los revisores pueden, de manera más simple y rápida, reproducir o revisar el trabajo).
4. Permite la continuidad del trabajo (por ejemplo, nuevos ayudantes de investigación pueden continuar el trabajo que realizaron ayudantes previos, y lectores de un artículo pueden reproducir y extender el trabajo).
5. Ayuda a construir la reputación del investigador (por ejemplo, un investigador que transparenta sus datos y procedimientos generará más confianza que uno que no lo hace).

Recomendaciones para mejorar la reproducibilidad en las geo-ciencias

Desde hace tiempo se están generando recomendaciones para mejorar la reproducibilidad, con un fuerte impulso desde las ciencias de la computación. Así, Sandve *et al.* (2013) propusieron 10 reglas para la reproducibilidad computacional:

1. Mantener un registro de cómo se obtuvo cada resultado
2. Evitar la manipulación manual de datos
3. Documentar y archivar las versiones exactas de todos los *software* y librerías utilizadas
4. Controlar las versiones de todos los *scripts*

5. Registrar los resultados intermedios y cuando sea posible en formatos estandarizados
6. Documentar las semillas de origen de la aleatoriedad
7. Almacenar los datos crudos u originales
8. Permitir que se inspeccionen las capas de detalle en los análisis de tipo jerárquico como, por ejemplo, los *clusters* jerárquico
9. Conectar declaraciones textuales con los resultados subyacentes
10. Proporcionar acceso público a las secuencias de comandos y resultados

Con un enfoque en las geo-ciencias, Gil *et al.* (2016) plantearon estrategias y buenas prácticas para lo que llamaron el *Geoscience Paper of the Future*. En él confluyen: la reproducibilidad (posibilidad de recrear los resultados de un estudio) y el linaje computacional (documentación digital sobre los datos y métodos usados para obtener los resultados). Así identificaron tres criterios básicos:

1. Los *datos deben ser reutilizables* a través de repositorios públicos, con documentación (metadatos), y licencias claras y citables que especifiquen las condiciones de uso.
2. El *software debe ser reutilizable* a través de repositorios públicos, con documentación (metadatos), y licencias claras y citables. El software incluye tanto el modelado, como la preparación y la visualización de datos.
3. El *linaje computacional debe ser documentado* describiendo explícitamente la serie de cálculos y sus resultados en un diagrama de flujo de trabajo de alto nivel, así como almacenado en un repositorio compartido y citable con un identificador único y persistente.

Para llevar a la práctica estos tres criterios, los autores proveen un *checklist* con 20 recomendaciones para guiar a los autores sobre como incluir información en los artículos de manera que facilite la reproducibilidad. Estas 20 recomendaciones se encuentran organizadas en las categorías de accesibilidad de datos, documentación de datos, accesibilidad de software, documentación de linaje, documentación de métodos, e identificación del autor. Además, se propone una hoja de ruta para que los investigadores comiencen a implementar la reproducibilidad como un hábito de trabajo, donde se expone, primero, un enfoque simple que se puede implementar en pocas horas y está dirigido a la documentación a realizar en el artículo científico. Seguido, se presenta un enfoque avanzado que, si bien requiere de una inversión considerablemente mayor de tiempo y esfuerzo, no se enfoca sólo en el artículo sino también en los hábitos de trabajo para la preparación de datos, software y linaje computacional. Por espacio no se incluyen los detalles en este documento, pero se recomienda que los lectores se dirijan al artículo mencionado (Gil *et al.*, 2016).

Más recientemente, Nüst *et al.* (2017; 2018), propusieron la idea de “Compendio Ejecutable de Investigación – *Executable Research Compendium*”, como una alternativa de empaquetado de datos, *scripts* y documentación para la reproducibilidad científica. El mismo provee una nueva forma computacional de publicación y acceso a la investigación, combinando el texto del artículo con los datos, software y documentación. La creación puede ser *post-hoc*, es decir, posterior a la realización de la investigación con la ayuda de un servicio y bajo la supervisión del autor; u *on-the-fly* donde al mismo tiempo que se realiza la investigación se produce la generación del compendio. Tras la preparación, el compendio debe ser validado, revisado y publicado. Por lo reciente de la publicación (2017 y 2018), no se han localizado publicaciones que utilicen compendios ejecutables, sin embargo, es de esperar que en el corto plazo comience a tener una aplicación efectiva.

Desde un punto de vista de la gestión de la investigación, en el marco de las conferencias AGILE <<http://o2r.info/reproducible-agile/2018/>>, se está impulsando la creación de guías e incentivos para que los autores comiencen a implementar prácticas de reproducibilidad. Así, se prevé proponer a AGILE nuevas guías de autores en las que se explique cómo enviar y revisar artículos reproducibles e incluso premios de conferencias para el mejor artículo reproducible. A mediano plazo, también se incentivará la reproductividad en ambientes educativos (Granell *et al.*, 2018).

Conclusiones

La reproducibilidad no es todavía una práctica común en el ámbito científico en general, así como tampoco en las geo-ciencias. Razones para esto son la falta de evidencias que demuestren su necesidad, la falta de incentivos y la falta de guías y licencias ampliamente aceptadas y establecidas que faciliten la apertura y publicación de datos y códigos (Nüst *et al.*, 2017). Sin embargo, se vislumbra a futuro un cambio en la configuración del artículo científico. Según la visión del “*Geoscience paper of the future*”, ya no será un documento estático con texto y figuras, como lo es el artículo actual, sino que cada vez más será un documento dinámico e interactivo, donde los revisores y lectores podrán interactuar, re-crear y re-utilizar los datos, métodos y el código publicado junto a un artículo (Gil *et al.*, 2016).

Esto requiere fuertes cambios a nivel individual de cada investigador. No solo debemos implementar los métodos, análisis y modelizaciones novedosas para nuestras áreas científicas, sino también formas nuevas y abiertas de documentación y publicación de las investigaciones. Por supuesto, existen razones tecnológicas y culturales que impiden la apertura y publicación de investigaciones reproducibles (Gil *et al.*, 2016). De hecho, en ocasiones pueden existir restricciones de acceso a los datos, o co-autores con cierto rechazo a la apertura de los códigos. Por ello, las

agencias de investigación, de financiamiento, productoras de datos, editoriales científicas y académicas tienen un rol fundamental. Deben realizar cambios en sus estructuras para promover y motivar que la reproducibilidad sea una realidad en el corto plazo, así como reconocer los esfuerzos y avances que se realicen en esta materia (Granell *et al.*, 2018).

En el contexto Latinoamericano, la reproducibilidad científica es, sin duda, incipiente, por lo que aún no se localizan comunidades o iniciativas tendientes a divulgar y fortalecer esta práctica. Los ejemplos de reproducibilidad encontrados son aislados y parecen relacionarse con la motivación intrínseca de los propios investigadores, y no a prácticas establecidas y extendidas en la comunidad científica latinoamericana. Es por ello, que se percibe actualmente un momento adecuado para impulsar la reproducibilidad en las geo-ciencias en Latinoamérica.

Lograr que los artículos científicos sean reproducibles requiere, principalmente, de tiempo y toma de conciencia por parte de los investigadores y de las agencias vinculadas con la investigación. Por ello, es necesario introducir el tema de reproducibilidad en conferencias, en las guías de autores de las revistas científicas, en los sistemas de evaluación de la investigación e investigadores, y en cualquier otro espacio académico. La reproducibilidad debe posicionarse como un tema de preocupación y atención para la investigación latinoamericana, para así elevar los estándares científicos y caminar hacia la idea que la buena ciencia es “mostrar” y no “confiar” (Stark, 2018).

Reconocimiento

Este documento fue generado en el marco del proyecto “Representación espacial de las teleconexiones climáticas en la precipitación del Ecuador”, financiado por CEDIA—Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia. Se agradece al revisor anónimo que con sus comentarios ayudó a mejorar la legibilidad y contextualización del documento.

Bibliografía

- Baker, M. (2016). “1,500 scientists lift the lid on reproducibility”, *Nature News*, 533(7604), 452.
- Barba, L.A. (2018). “Terminologies for reproducible research”, *ArXiv Preprint ArXiv:1802.03311*.
- Cargill, M. and O’Connor, P. (2013). *Writing scientific research articles: Strategy and steps*. John Wiley & Sons.
- Gil, Y.; David, C.H.; Demir, I.; Essawy, B.T.; Fulweiler, R.W.; Goodall, J.L.; ... others (2016). “Toward the Geoscience Paper of the Future: Best practices for

- documenting and sharing research from data to software to provenance”, *Earth and Space Science*, 3(10), 388-415.
- Granell, C.; Nüst, D.; Ostermann, F.O. and Sileryte, R. (2018). *Reproducible Research is like riding a bike*.
- Head, M.L.; Holman, L.; Lanfear, R.; Kahn, A.T. and Jennions, M.D. (2015). “The extent and consequences of p-hacking in science”, *PLoS Biology*, 13(3), e1002106.
- Konkol, M.; Kray, C. and Pfeiffer, M. (2018a). “Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study”, *International Journal of Geographical Information Science*, 1-22.
- Konkol, M.; Kray, C. and Pfeiffer, M. (2018b). *The state of reproducibility in the computational geosciences*.
- Markowetz, F. (2015). “Five selfish reasons to work reproducibly”, *Genome Biology*, 16(1), 274.
- Nüst, D.; Boettiger, C. and Marwick, B. (2018). “How to Read a Research Compendium”, *ArXiv Preprint ArXiv:1806.09525*.
- Nüst, D.; Granell, C.; Hofer, B.; Konkol, M.; Ostermann, F.O.; Sileryte, R., and Cerutti, V. (2018). “Reproducible research and GIScience: an evaluation using AGILE conference papers”, *PeerJ Preprints*, 6, e26561v1.
- Nüst, D.; Konkol, M.; Pebesma, E.; Kray, C.; Schutzeichel, M.; Przibytzin, H. and Lorenz, J. (2017). “Opening the publication process with executable research compendia”, *D-Lib Magazine*, 23(1/2). Retrieved from <<http://www.dlib.org/dlib/january17/nuest/01nuest.html>>.
- Ostermann, F.O. and Granell, C. (2017). “Advancing Science with VGI: Reproducibility and Replicability of Recent Studies using VGI”, *Transactions in GIS*, 21(2), 224-237, <<https://doi.org/10.1111/tgis.12195>>.
- Pebesma, E.; Nüst, D. and Bivand, R. (2012). “The R software environment in reproducible geoscientific research”, *Eos, Transactions American Geophysical Union*, 93(16), 163.
- Sandve, G.K.; Nekrutenko, A.; Taylor, J. and Hovig, E. (2013). “Ten simple rules for reproducible computational research”, *PLoS Computational Biology*, 9(10), e1003285.
- Shannon, J. and Walker, K. (2018). “Opening GIScience: A process-based approach”, *International Journal of Geographical Information Science*, 1-16.
- Stark, P.B. (2018). “Before reproducibility must come preproducibility”, *Nature*, 557(7707), 613.

Datos enlazados de información geográfica del Ecuador

Víctor Saquicela*
Fernando Baculima*
Lucia Lupercio**

Recibido 14 de octubre de 2018, aceptado 07 de noviembre de 2018

Resumen

Actualmente, la gran cantidad de información geográfica generada por diferentes instituciones exige que se propongan nuevas soluciones de gestión y publicación de estos datos. En Ecuador, la heterogeneidad de la información geográfica generada por diversos entes y publicada en la Web no es una excepción. Con el objetivo de dar solución a esta problemática, el proyecto Geo Linked Data Ecuador integra información geográfica proveniente de diversas fuentes y la publica en formato RDF bajo estándares y recomendaciones de la comunidad de la Web Semántica, creando así el repositorio de datos enlazados geográficos del Ecuador.

Palabras clave: *Datos enlazados geográficos, Web semántica, GeoRDF.*

Resumo

Atualmente, a grande quantidade de informação geográfica gerada por diferentes instituições exige que se propague novas soluções de gestão e publicação desses dados. No Equador, a heterogeneidade da informação geográfica gerada por diversas entidades e publicada na Web não é uma exceção. Com o objetivo de solucionar essa problemática, o projeto *Geo Linked Data* Equador integra informação geográfica proveniente de diversas fontes e a publica em formato RDF por meio de padrões e recomendações da comunidade Web Semântica, criando assim o repositório de dados geográficos vinculados ao Equador.

Palavras chave: *Dados vinculados geográficos, Web semântica, GeoRDF.*

* Departamento de Ciencias de la Computación, Universidad de Cuenca, Cuenca, Ecuador
{victor.saquicela; fernando.baculima }@ucuenca.edu.ec

** Facultad de Ciencias Agropecuarias, Universidad de Cuenca, Cuenca, Ecuador
lucia.lupercio@ucuenca.edu.ec

Abstract

Nowadays, different institutions generate a large amount of geographical information, which require some new management and publication solutions to be proposed. In Ecuador, the heterogeneity of geographic information generated by different entities and published on the Web is not the exception. Aiming to solve this problem, the Geo Linked Data Ecuador project integrates geographic information from different sources and publishes it in RDF format, under the standards and recommendations of the Semantic Web community, as result, the Ecuadorian Geographic Linked Data Repository has been created.

Key words: *Geo linked data, Semantic web, GeoRDF.*

Introducción

En Ecuador existen varias instituciones del gobierno que generan y publican grandes cantidades de información geográfica (IG) referentes a una misma o diferentes temáticas (Lupercio *et al.*, 2015). A esto se le une el hecho de que las instituciones al ser autónomas utilizan sus propios modelos de datos en los procesos de generación y publicación, dando como resultado diferentes niveles de heterogeneidad en la información. Por lo tanto, se crea la necesidad de sistemas que faciliten el análisis, integración, interpretación, publicación y explotación de estos datos.

Debido a la gran cantidad de datos que se generan diariamente, varios trabajos relacionados con la gestión de los mismos se dirigen hacia la aplicación de Linked Data y la Web Semántica (Berners-Lee *et al.*, 2001). En este contexto, la utilización de RDF (*Framework de Descripción de Recursos*) permite el intercambio de datos en la Web Semántica, brindando facilidades para el establecimiento de las relaciones entre los datos provenientes de diferentes fuentes (RDF-WG, 2014). En el caso particular de la iniciativa Geo Linked Data, la adaptación de estas propuestas permite generar y publicar RDF de datos geográficos a través de tecnologías de la Web Semántica (Vilches-Blázquez *et al.*, 2010).

Durante los últimos años, en Ecuador surge la iniciativa denominada “Geo Linked Data Ecuador”,¹ vinculada a un proyecto que nació en el año 2015, cuyo objetivo es integrar información proveniente de diferentes fuentes de datos de IG del país para ser representados y publicados utilizando las tecnologías y recomendaciones de la Web Semántica. Además, esta iniciativa permitió el desarrollo de herramientas de software para llevar a cabo el proceso de generación, publicación y explotación de datos geográficos en el contexto de la mencionada Web.

¹ <<http://linkeddata.ec/>>.

Adicionalmente, “Geo Linked Data Ecuador” basa su proceso de generación, publicación y explotación de RDF de información geográfica (GeoRDF) en los principios Linked Data (Bizer, 2011) y el ciclo de vida de Linked Data propuesto en (Villazón-Terrazas *et al.*, 2011) utilizando estándares y herramientas recomendadas por la comunidad.

Proceso de generación, publicación y explotación de Geo Linked Data del Ecuador

El proceso de generación, publicación y explotación de Geo Linked Data del Ecuador se lleva a cabo a través de diversas actividades, reflejadas en la Figura 1, donde se utilizan fuentes de datos de diferentes instituciones ecuatorianas.

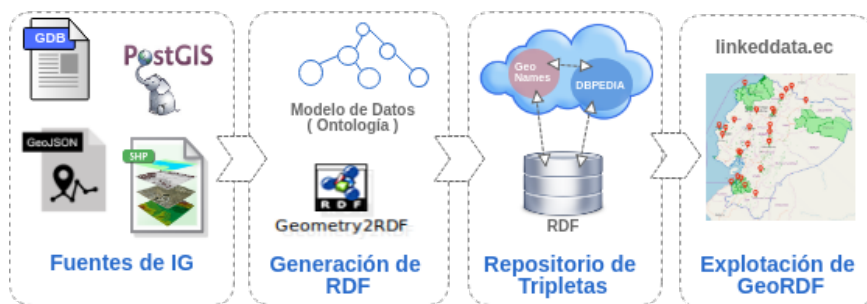


Figura 1. Proceso de generación, publicación y explotación de Geo Linked Data del Ecuador.

Fuentes de información geográfica

El análisis de las fuentes de datos permite obtener un conocimiento claro de la IG existente en diferentes instituciones en el Ecuador, así como de los medios para acceder a ellas y los formatos en los cuales se encuentran disponibles. A través de buscadores Web, se pueden encontrar servicios Web, archivos CSV y *shapefiles* alojados en diferentes repositorios de instituciones públicas del Ecuador con datos geográficos.

Los archivos de tipo *shapefile* resultan entre los formatos de datos con mayor presencia en la distribución de IG del Ecuador. Entre la información disponible se encuentran entidades espaciales con diversas geometrías (puntos, líneas y polígonos) referente a información de provincias, ciudades, instituciones educativas, ríos, lagos, entre otras. De esta manera, para la iniciativa Geo Linked Data Ecuador se abordan más de 80 entidades (*features*) con IG del país.

Algunas de las principales instituciones de las que proceden estos conjuntos de datos en el Ecuador son las siguientes:

- Instituto Geográfico Militar (IGM). La agencia que se encarga de planificar y controlar las actividades relacionadas con la cartografía ecuatoriana.
- Secretaría del Agua del Ecuador (SENAGUA). Genera y publica información geográfica de lugares de interés para la ejecución y desarrollo de proyectos relacionados con el manejo y control del agua en el país.
- PROMAS. Este centro de I+D+i de la Universidad de Cuenca, ejecuta proyectos de investigación relacionados con el manejo de los recursos hídricos del país.

Tras la obtención de los conjuntos de datos de los diversos productores mencionados, se concluye que, en Ecuador, no existe un repositorio centralizado de IG, debido a que cada institución genera su información y la almacena en sus repositorios de manera autónoma con el inconveniente de la heterogeneidad que se produce.

Generación de RDF

Para la generación de GeoRDF resulta necesario definir modelos ontológicos, los cuales permiten representar el conocimiento de un dominio. Las ontologías definen conceptualizaciones (conceptos y relaciones) de un dominio determinado, de forma compartida y consensuada, donde su representación debe ser realizada de una manera formal, legible y utilizable por los computadores (Gruber, 1993). En este trabajo específicamente se optó por la utilización de modelos ontológicos que permitan la descripción formal de la IG para alcanzar la estandarización, reutilización e interoperabilidad de la IG del Ecuador.

En este proceso se utilizan herramientas como Protégé,² que permite conformar el modelo ontológico del dominio geográfico considerado en Geo Linked Data. Para ello, se reutilizan recursos ontológicos como GeoSPARQL (Perry y Herring, 2011), hydrOntology (Vilches-Blázquez *et al.*, 2009) y WGS84.³ Estos modelos son poblados utilizando la herramienta GeoKettle,⁴ que permite realizar procesos ETL (*Extract, Transform and Load*) desde fuentes de datos geográficas hacia repositorios RDF a través de componentes creados específicamente para cubrir los procesos de generación, publicación y explotación de GeoRDF, y son consultados utilizando el lenguaje de consulta *SPARQL Protocol and RDF Query Language* (SPARQL).

Con la definición del modelo ontológico se procede a generar el RDF a partir de las diferentes fuentes de información del Ecuador. El proceso de generación de RDF se ejecuta mediante la utilización del *framework* Geo-Kettle (software basado

² <<https://protege.stanford.edu/>>.

³ <<https://www.w3.org/2003/01/geo/>>.

⁴ <<http://www.spatialytics.org/projects/geokettle/>>.

en Pentaho Integrator). Como se mencionó anteriormente, Geo-Kettle permite realizar procesos ETL, de una manera amigable para el usuario, configurar procesos de extracción, preprocesamiento, integración y transformación de datos de diferentes fuentes. En el marco de este trabajo, se utilizaron componentes para extraer los datos de las fuentes de IG y otros que se desarrollaron específicamente para el tratamiento de IG y la creación y manejo de RDF.

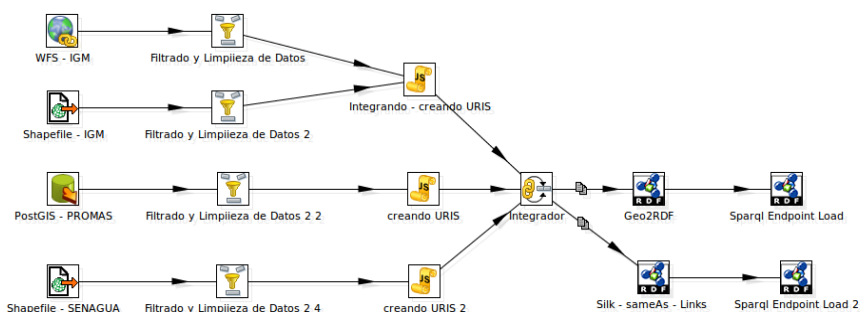


Figura 2. Proceso de generación de RDF sobre la plataforma Geo-Kettle.

Así, en la Figura 2 se muestra la configuración del proceso de transformación sobre la plataforma Geo-Kettle para la generación de RDF a partir de diferentes fuentes de IG. El proceso inicia con la utilización de los componentes de extracción de datos que posee Geo-Kettle de las fuentes de IG. En este caso, un WFS (“WFS – IGM”), dos *shapefile* (“shapefile – IGM”), una base de datos PostGIS (“PostGIS – PROMAS”).

A continuación, los datos extraídos de cada fuente son procesados con el objetivo de poseer datos de calidad, mediante la aplicación de técnicas de preprocesamiento de datos, a través de los componentes de filtrado y limpieza proporcionados por Geo-Kettle. Siguiendo las recomendaciones de la propuesta metodológica adoptada en este trabajo (Villazón-Terrazas *et al.*, 2011), se procede a crear las URI (*Uniform Resource Identifiers*) para cada fuente de datos e integrar todas ellas.

Posteriormente, basado en las ontologías utilizadas se procede a realizar el emparejamiento correcto entre los datos integrados y las ontologías utilizadas en el proceso de generación. Este emparejamiento y la consiguiente generación de RDF son posibles gracias al desarrollo del componente “Geo2RDF”. A continuación, se procede a detectar posibles recursos similares, para ello se desarrolla el componente “Silk-sameAs-Links”, el cual permite buscar recursos RDF similares en la nube de datos enlazados.

Finalmente, se procede a publicar los datos generados en formato RDF en *triple store* (base de datos semántica) para posteriormente ser explotada. Para ello, se desarrolla el componente “*SPARQL Endpoint Load*” que permite cargar el RDF a un repositorio semántico.

Publicación de GeoRDF

Los datos generados en formato GeoRDF son cargados a un repositorio semántico, que permite la consulta de los mismos a través de GeoSPARQL, donde se pueden utilizar operadores espaciales como: intersecciones, *buffers*, vecindades, entre otras.

El repositorio semántico en el que se encuentran almacenados los datos de la iniciativa “Geo Linked Data Ecuador” es Apache Marmotta.⁵ La elección de este repositorio se debe a la implicación de esta iniciativa en el desarrollo de este proyecto de Apache, ya que a través de los esfuerzos de Geo Linked Data Ecuador se ha contribuido activamente al soporte de los datos geográficos en el contexto de la Web Semántica a través de GeoSPARQL.

La iniciativa “Geo Linked Data Ecuador” refleja los resultados en el portal <<http://linkeddata.ec>> donde se encuentra información referente al proceso de generación de Geo Linked Data, acceso al SPARQL Endpoint, visualizador de mapas, contactos para soportes, links de descargas para las herramientas desarrolladas y artículos publicados, entre otros aspectos.

Explotación de datos

Los datos cargados al repositorio semántico son explotados utilizando un visualizador de mapas, el cual permite consultar toda la información generada (GeoRDF). La Figura 3 muestra un ejemplo de cómo el visualizador presenta la información publicada asociada con su representación geométrica.

El visualizador desarrollado dispone de un panel izquierdo que permite navegar sobre los recursos de los cuales se ha generado RDF (capas de información de los diversos productores mencionados con anterioridad) y el panel principal muestra los recursos graficados en un mapa a través de OpenStreetMap.

Cabe mencionar que el proceso de generación al estar basado en modelos únicos permite, de manera automática, integrar las diferentes fuentes de IG, es decir, si una fuente de datos de una institución “X” provee información de ciudades del Ecuador y una institución “Y” provee también información de ciudades del Ecuador, por el hecho de que las fuentes están modeladas de manera común, a través de una ontología, permitiría extraer todos los datos, de manera conjunta, de las dos instituciones para una ciudad específica. De esta manera, se logra un proceso de integración

⁵ <<http://marmotta.apache.org/>>.

de los conjuntos de datos, se resuelven los problemas de heterogeneidad presente en los diferentes modelos de datos y se enriquecen los datos con la información presente en las distintas fuentes.

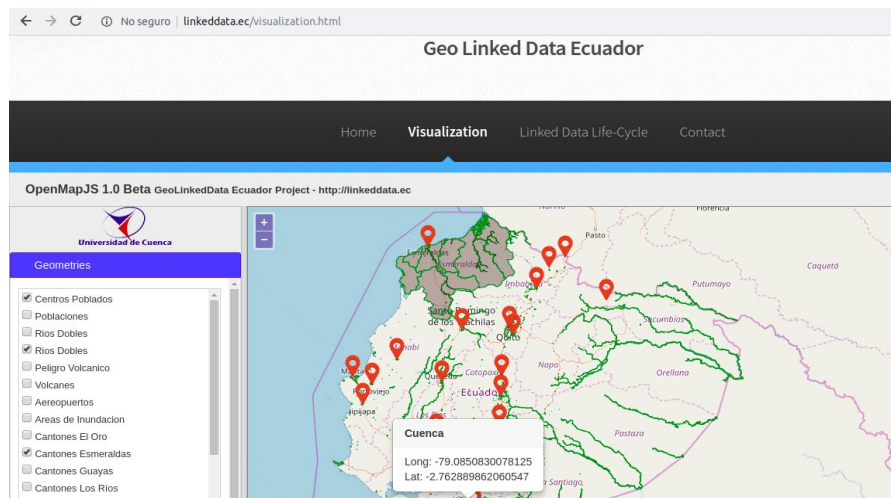


Figura 3. Visualizador de Geo Linked Data Ecuador.

Conclusiones

En este trabajo se presenta un proceso para la generación de RDF de fuentes de información geográfica, enfocando la investigación en los esfuerzos científicos actuales y desafíos abiertos en este dominio. Además, se han creado componentes que permiten tener herramientas que faciliten la publicación y explotación de datos geográficos enlazados.

El proceso descrito permitió generar un repositorio semántico de información geográfica del Ecuador (*geo.linkedata.ec*), el cual puede ser usado por instituciones interesadas en detectar fuentes de datos similares, así como por científicos interesados en proyectos de integración de datos, etc.

Es necesario continuar con la investigación, por lo que se pretende mejorar el proceso de generación a través de la automatización del proceso de emparejamiento entre la fuente de datos y la ontología utilizando recursos externos. Además, se pretende experimentar con fuentes de información geográfica en otros formatos, tales como *Geography Markup Language* (GML), *Keyhole Markup Language* (KML), etc.

Bibliografía

- Berners-Lee, T.; Hendler, J. and Lassila, O. (2001). "The semantic web" *Sci. Am.*, 284(5), 28-37.
- Bizer, C.; Heath, T. and Berners-Lee, T. (2011). "Linked Data: The Story so Far", en *Semantic Services, Interoperability and Web Applications: Emerging Concepts*. IGI Global, 205-227.
- Gruber, T.R. (1993). "A Translation Approach to Portable Ontology Specifications", *Knowledge Acquisition*, 6(2), 199-221.
- Lupercio, L.; Baculima, F.; Espinoza, M. and Saquicela, V. (2015). "Explotación de información en el dominio geo-hídrico ecuatoriano utilizando tecnología semántica", en *Maskana*, 6 (Supl.), 69-77. Recuperado a partir de <<https://publicaciones.ucuenca.edu.ec/ojs/index.php/maskana/article/view/699>>.
- Perry, M. and Herring, J. (2011). "OGC GeoSPARQL - a geographic query language for RDF data". Recuperado a partir de: <<http://www.opengeospatial.org/standards/geosparql>>.
- RDF Working Group. (2014), "Resource Description Framework (RDF)". W3C Semantic Web. Disponible Online <<http://www.w3.org/RDF/>>.
- Vilches-Blázquez, L.M.; Ramos, J.A.; López-Pellicer, F.J.; Corcho, O. and Noguera-Iso, J. (2009) "An approach to comparing different ontologies in the context of hydrographical information", in Popovich *et al.*, (eds.): IF&GIS'09. LNG&C Springer, St. Petersburg, Russia, 193-207.
- Vilches-Blázquez, L.M.; Villazón-Terrazas, B.; León, A.; Priyatna, F. and Corcho, O. (2010). *An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case*, en "Workshop on Linked SpatioTemporal Data 2010 in conjunction with the 6th International Conference on Geographic Information Science , GIScience 2010", Zurich, Suiza.
- Villazón-Terrazas, B.; Vilches-Blázquez, L.M.; Corcho, O. and Gómez-Pérez, A. (2011). "Methodological Guidelines for Publishing Government Linked Data", en *Linking Government Data*, Springer New York, 27-49.

Caracterización geográfica de la vivienda en México: un enfoque de Ciencia de Datos

Jacobo Gerardo González León*
Miguel Félix Mata Rivera*

Recibido 16 de octubre de 2018, aceptado 09 de noviembre de 2018

Resumen

Una tendencia global en los primeros años del siglo XXI ha sido la generación de plataformas y datos abiertos por parte de entidades gubernamentales para la consulta pública. Estos datos permiten a los ciudadanos generar nuevas técnicas y mecanismos para su análisis con la intención de lograr mejoras en el sistema político, social y económico de un país. En esta investigación, proponemos una metodología de Ciencia de Datos centrada en la caracterización geográfica de la vivienda mexicana, utilizando datos abiertos del período 2014-2017 y aplicando técnicas de aprendizaje automatizado no supervisado. La aplicación de esta metodología permitió descubrir, que son 7 los tipos de viviendas en México, todos ellos distribuidos a lo largo y ancho del territorio mexicano. Los resultados son desplegados en mapas que podrán ser de gran utilidad en ciencia aplicada, para el estudio de fenómenos humanos y geográficos en diferentes dominios del conocimiento, tales como: seguridad, transporte, salud, economía, industria, agricultura, entre otras.

Palabras clave: *caracterización, vivienda, clusterización, análisis de datos.*

Resumo

Uma tendência global nos primeiros anos do século XXI tem sido a geração de plataformas e dados abertos por entidades governamentais visando a consulta pública. Esses dados permitem que os cidadãos criem novas técnicas e mecanismos para sua análise com a intenção de alcançar melhorias no sistema político, social e econômico de um país. Nesta pesquisa, propomos uma metodologia de Ciência de dados

* Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Instituto Politécnico Nacional (IPN), Unidad Avenida Instituto Politécnico Nacional núm. 2580, Col. Barrio La Laguna Ticomán, 07340 Gustavo A. Madero, Ciudad de México, México, correos electrónicos: jgonzalez11007@alumno.ipn.mx, mmatar@ipn.mx

focada na caracterização geográfica de casas mexicanas, usando dados abertos do período 2014-2017 e aplicando técnicas automatizadas de aprendizado não supervisionado. A aplicação desta metodologia permitiu descobrir que existem sete tipos de casas no México; todos eles estão distribuídos em todo o território mexicano. Os resultados são exibidos em mapas que serão muito úteis em ciência aplicada, para o estudo de fenômenos humanos e geográficos em diferentes domínios do conhecimento, tais como segurança, transporte, saúde, economia, indústria, agricultura, entre outros.

Palavras-chave: *caracterização, habitação, agrupamento, análise de dados.*

Abstract

A global trend in the first years of the 21st century has been the generation of platforms and open data by government entities for public consultation. These data allow citizens to generate new techniques and mechanisms for their analysis with the intention of achieving improvements in the political, social and economic system of a country. In this research, we propose a Data Science methodology focused on the geographical characterization of Mexican houses, using open data from the period 2014-2017 and applying unsupervised automated learning techniques. The application of this methodology allowed discovering, that there are seven types of houses in Mexico; all of them are distributed throughout the Mexican territory. The results are displayed on maps that will be very useful in applied science, for the study of human and geographical phenomena in different domains of knowledge, such as security, transport, health, economy, industry, agriculture, among others.

Key words: *characterization, housing, clustering, data analysis.*

Introducción

Hoy en día, los Sistemas de Información recuperan y gestionan datos de muchas actividades humanas. Estos datos se han ido acumulando en grandes volúmenes a lo largo del tiempo. En estos almacenes de datos, no sólo se ha guardado información, también está contenida la historia, ya que los datos tienen la propiedad intrínseca de registrar consigo el paso del tiempo. Simultáneamente a esta acumulación de datos, también han surgido distintas maneras de procesarlos para obtener utilidad de ellos. Un ejemplo es el enfoque estadístico, que basándose en la aplicación de funciones de resumen ha permitido inferir comportamientos sobre los mismos, permitiendo examinar estos volúmenes de datos y encontrar información. En este enfoque resulta común buscar el ajuste de estas características (por ejemplo, considere una encuesta con la siguiente pregunta: ¿Cuántas habitaciones tiene su casa?, si alguien decide no responder esa pregunta, al final se promedia o se ajusta entre las que si se respondieron en la encuesta, para poder desplegar el resultado) a distribuciones

probabilísticas (Yang y Jargowsky, 2005). Sin embargo, hoy el enfoque ha cambiado, ya que no se busca encontrar un modelo estadístico que se ajuste a los datos, sino todo lo contrario, se busca analizar todo el conjunto de datos para poder conocer el modelo que mejor describe los mismos. Estas capacidades sin duda alguna, nos están llevando a una cuarta revolución industrial donde los datos son la materia prima (Oguro, 2016).

No obstante, los datos por sí solos carecen de significado, por lo que el problema continúa siendo ¿cómo tratar los datos? El objetivo entonces no ha cambiado y se centra en la extracción de información. En el contexto actual se habla de la Ciencia de Datos, entendida como un conjunto de técnicas que combina el procesamiento de datos, estadística aplicada, algoritmos de aprendizaje automatizado, arquitecturas distribuidas de servidores, análisis exploratorio, entre muchas otras más, y que su incorporación en cualquier área del conocimiento podría producir mejoras sustanciales (Gibert, Horsburgh, Athanasiadis y Holmes, 2018). Este enfoque, que últimamente está en *vogue*, no resulta tan nuevo como parece, puesto que tiene su primer registro en el año 1962, cuando el estadista John Tukey en su libro “*The Future of Data Analysis*” invita a la comunidad estadista a involucrarse en acelerar el desarrollo computacional, aplicar sus teoremas y generar estándares y medidas de validación, que sirvan como técnicas de interpretación, para entender y visualizar a los datos. De esta manera, según Tukey, se obtendrían resultados más precisos y certeros (Donoho, 2017).

La incorporación de la Ciencia de Datos en el estudio de la Geografía podría ser útil en el desarrollo futuro de esta disciplina, pues se podrían generar nuevas metodologías y datos para su incorporación en el análisis espacial de fenómenos y actividades humanas, tales como incidencia delictiva, comercio, movilidad, industrialización, turismo, y demás.

Considerando este contexto, en este trabajo se presenta la propuesta de una metodología para caracterizar geográficamente la vivienda mexicana, esto a través del agrupamiento de la similitud de sus características. Finalmente, se presentan los resultados del análisis de los tipos de viviendas obtenidos desplegándolos en un mapa, en donde se representan los componentes espaciales y geográficos para su posterior integración en otros estudios donde, por ejemplo, se podrían responder preguntas del tipo ¿las casas del Sur de México se parecen a las del Norte, o, por lo contrario, son más parecidas a las del centro del país?

La vivienda en México

En 1990 el Instituto Nacional de Estadística y Geografía (INEGI) realizó el XI Censo General de Población y Vivienda, donde surge una investigación enfocada a describir las características físicas y socio-espaciales de las viviendas. En esta investiga-

ción se regionaliza al país en 10 zonas ordenadas por la marginación de su población, originalmente, a través de la *Geografía de la Marginación* (1983) y, posteriormente, ajustada a través de los índices y grados de marginación publicados en *Indicadores Socioeconómicos e Índice de Marginación Municipal* por el Consejo Nacional de Población (CONAPO) en 1993 (Scheingart y Solís, 1995). A partir de la conformación de estas regiones y el tamaño de la localidad, se caracterizó a la población y las viviendas de México con una muestra del 1% del censo, y se generaron índices como el de “calidad de la vivienda”, “calidad de los materiales empleados en la construcción de la vivienda”, “distribución de las viviendas”, etc.

Recientemente, los productos desarrollados por INEGI están basados en la *Encuesta Intercensal 2015* y arrojan sólo descripción estadística de los datos crudos asociadas a variables como “promedio de ocupantes en viviendas particulares habitadas”, “porcentaje de viviendas particulares habitadas con drenaje”, “porcentaje de viviendas particulares habitadas con electricidad”.

Ante esta disponibilidad de datos asociados con múltiples variables que permiten caracterizar la vivienda en México, surge la posibilidad de realizar un análisis, utilizando técnicas de la Ciencia de Datos, que permita identificar el panorama actual de la vivienda en el país.

Metodología de Ciencia de Datos para la caracterización geográfica de la vivienda

El desarrollo de este trabajo se ha centrado en responder a la siguiente pregunta de investigación: ¿Cómo aplicar las técnicas de Ciencia de Datos para la caracterización de las viviendas mexicanas y el descubrimiento de conocimiento geográfico? A través de esta pregunta inicial se pretende descubrir las similitudes y diferencias entre las casas del norte y sur de México. Para contestar a esta pregunta se propone el procesamiento de conjuntos de datos relacionados con la temática de este trabajo a través de la metodología propuesta en la Figura 1.

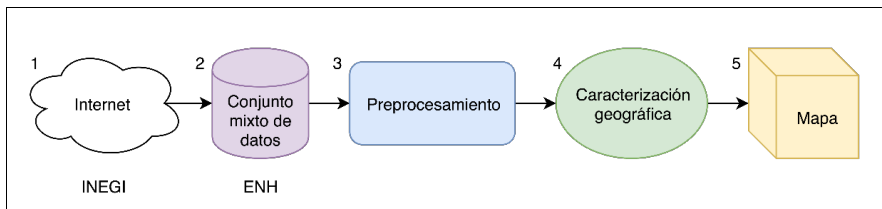


Figura 1. Metodología propuesta para la caracterización de un conjunto de datos mixto proveniente de encuestas.

La implementación de esta metodología permite que los datos estructurados sean desplegados geográficamente a través de las técnicas de aprendizaje no supervisado, lo que va a permitir, por ejemplo, identificar cuales casas de la República mexicana no tienen drenaje y conocer su distribución espacial.

La metodología consiste de 5 etapas: 1) extraer datos del portal del INEGI, 2) integrar los datos del INEGI con los datos de la Encuesta Nacional del hogar (ENH), 3) preprocesarlos para limpiarlos y adecuarlos para poder usarlos, 4) realizar una caracterización geográfica, lo que implica utilizar un algoritmo de clasificación no supervisada, ya que a priori no se conocen cuantos grupos o clases de viviendas hay en México. Finalmente, 5) visualizar los grupos identificados en un mapa.

Fuente de datos

En este trabajo se trataron datos obtenidos del repositorio de datos abiertos que publica el INEGI <www.inegi.com.mx> sobre la Encuesta Nacional del Hogar (ENH)¹ y comprenden el periodo 2014-2017. Esta encuesta está organizada en una base de datos por cada año y está compuesta por tres tablas con características de personas, hogares y viviendas, organizadas como sigue:

- **Personas:** contiene características sociodemográficas y ocupacionales de los integrantes del hogar, entre ellas, edad, sexo, parentesco, asistencia escolar, estado conyugal, tipo de trabajo, etc.
- **Hogar:** contiene características de los hogares que habitan los integrantes, entre ellas, número de integrantes del hogar, si se tiene trabajador doméstico, niño o cuidador de ancianos, etc.
- **Vivienda:** contiene características de las viviendas que habitan los integrantes de los hogares encuestados, entre ellas, tipo de vivienda, material de paredes, disponibilidad de agua, etc.

Estas tablas se concatenaron en un solo conjunto de datos, el cual se denomina mixto (contiene datos de tipos numérico y de tipo categórico) donde el tamaño de la muestra es de 64 mil viviendas por cada año, siendo un total de 256 mil viviendas. La información recopilada de las encuestas abarca la siguiente cobertura temática: educación, salud, vivienda, ocupación y tecnologías de la información.

Conjunto mixto de datos obtenido de encuestas

El conjunto de datos provenientes de encuestas del INEGI suele ser de naturaleza mixto, es decir, contiene datos de tipo numérico, de tipo categórico, así como valores cuantitativos para expresar cantidades y valores cualitativos para expresar respuestas a las preguntas que hace el encuestador. Entre las preguntas que recogen

¹ <<http://www.beta.inegi.org.mx/proyectos/enchogares/regulares/enh/2014/default.html>>.

estas encuestas se pueden encontrar, por ejemplo: si la vivienda dispone de tinaco en la azotea, si cuenta con cocina, o la forma en que se abastece de agua.

Como se mencionó con anterioridad, en este trabajo se realiza una concatenación de las tablas recopiladas, por medio de las claves de entidad, municipio y localidad. Esto permitió generar como resultado un conjunto de datos de 254,928 observaciones y 200 variables. A modo de ejemplo, en la Tabla 1 se muestran algunas de las diversas variables consideradas en este trabajo.

Tabla 1
Muestra de algunas variables del conjunto mixto de datos

| <i>Variable</i> | <i>Etiqueta</i> | <i>Tipo</i> |
|-----------------|-------------------------------|-------------|
| mat_pared | Material de paredes | Categorico |
| antigüedad | Antigüedad de la vivienda | Numérico |
| num_cuarto | Número de cuartos | Numérico |
| focos_inca | Número de focos incandescente | Numérico |
| tinaco_azo | Dispone de tinaco | Categorico |
| ... | ... | ... |

Preprocesamiento

En la etapa de preprocesamiento se realizan 4 tareas previas a la aplicación de técnicas vinculadas a la Ciencia de Datos y que van a permitir obtener resultados de calidad durante el proceso de análisis. Estas tareas se centran en limpieza, rellenado, transformación y agrupación de datos.

Con respecto a la limpieza, esta consiste en corregir los errores presentes en los datos. En el rellenado se predicen los valores faltantes en función de las variables que presentan datos completos. En la transformación se convierten los datos categóricos en numéricos para su uso en posteriores etapas. Finalmente, en la agrupación de datos se resume la información en las localidades geográficas de interés para su caracterización.

Tras la aplicación de estas tareas, se obtuvo como resultado 5,640 localidades, producto de la combinación de entidades, municipios y localidades en esta etapa de preprocesamiento. Este dato pone de manifiesto que se trabaja con una muestra de datos, ya que según el Catálogo Único de Claves de Áreas Geoestadísticas Estatales, Municipales y Localidades² se tienen 304, 221 localidades en el territorio mexicano, pero la ENH sólo se cuenta con la información de 5, 640 localidades, o sea, que, a partir de esta encuesta, sólo se podría caracterizar el 1.8% del total de localidades.

² <<http://www.inegi.org.mx/geo/contenidos/geoestadistica/catalogoclaves.aspx>>.

Caracterización geográfica

Para proceder a la caracterización geográfica de los tipos de viviendas que existen en toda una región se necesita encontrar alguna técnica para generar esta categorización. La técnica que se adopta en este trabajo es el Aprendizaje Automatizado No Supervisado a través de la Clusterización (*clustering*, en inglés). Esta técnica permite agrupar los datos en función de la similitud de sus variables (por ejemplo: qué tanto se parece la variable X a la variable Y). Para ello, se utilizó el método *k-means* (Hartigan & Wong, 1979), un algoritmo de partición que divide los datos en acumulaciones de puntos (clústeres) utilizando la distancia euclidiana. De tal modo, que las distancias entre observaciones dentro de cada clúster queden minimizadas, de tal forma que así se puede saber si un punto pertenece a un clúster o no. La dificultad estriba en que, además, debemos saber cuántos grupos de puntos se pueden formar (cuál es el número óptimo de clústeres para todos los datos que se están analizando).

Para encontrar el valor *k* clústeres (un parámetro que indica el número óptimo de clústeres, en el cual se debe dividir todo un conjunto de datos), se eligió calcularlo usando el máximo *índice de Dunn*, que se define como la razón entre la distancia más pequeña entre todas las observaciones y la distancia máxima en el clúster, también llamada el diámetro del clúster. La utilización de este índice será útil para hallar la mejor separación (agrupamiento) de los tipos de viviendas presentes en los datos, con lo cual se va a conocer si una vivienda pertenece a una cierta clase (grupo), las cuales están representadas en observaciones (datos de encuesta) (Desgraupes, 2017).

Tras ello, se ejecutó una heurística del *índice Dunn* estableciendo un rango de dos a 15 clústeres. Esto permite encontrar el valor máximo de agrupaciones, el cual se alcanza en este caso cuando existen siete clústeres, es decir, que conforme a todos los datos de viviendas considerados se detectan siete grupos con características comunes, por lo que se puede afirmar que, en términos simples, todas las viviendas analizadas se agrupan en siete tipos (véase Figura 2).

Por otra parte, para visualizar el conjunto de datos clusterizado en el espacio se requeriría de un sistema de coordenadas de 61 dimensiones, ya que ese es el número de variables que tienen las viviendas de acuerdo a la encuesta del INEGI. Sin embargo, nuestros actuales métodos de visualización no nos permitirán apreciar estas dimensiones, por lo cual, se hizo una reducción de dimensiones, aplicando Análisis de Componentes Principales (PCA, en sus siglas en inglés). Con este método (véase Figura 3) se redujo de 61 a 2 dimensiones, por medio de un cambio de coordenadas en función de los componentes no correlacionados con mayor varianza. Así, en este conjunto de datos los dos componentes que se buscan están situados en las dimensiones 1 y 2, que representan el 19.9% y el 5.6% de la varianza total.

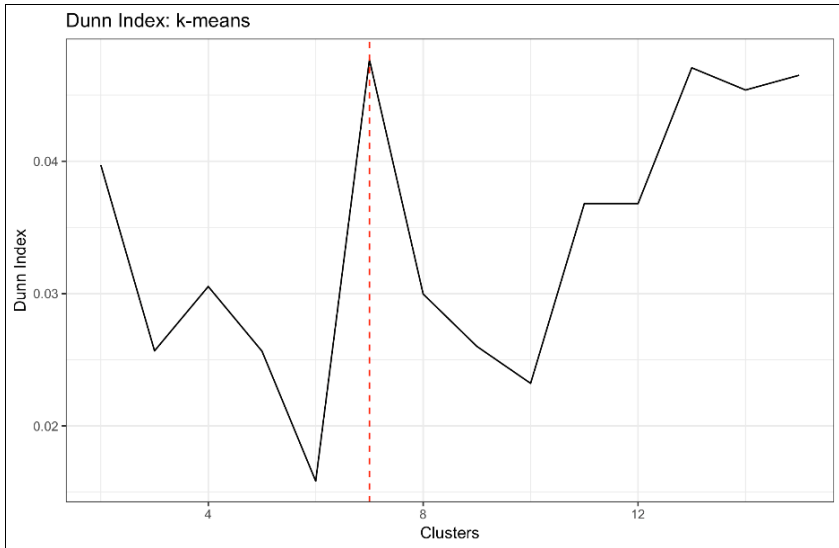


Figura 2. El índice Dunn es una medida para elegir la mejor partición en función de la cohesión de clústeres.

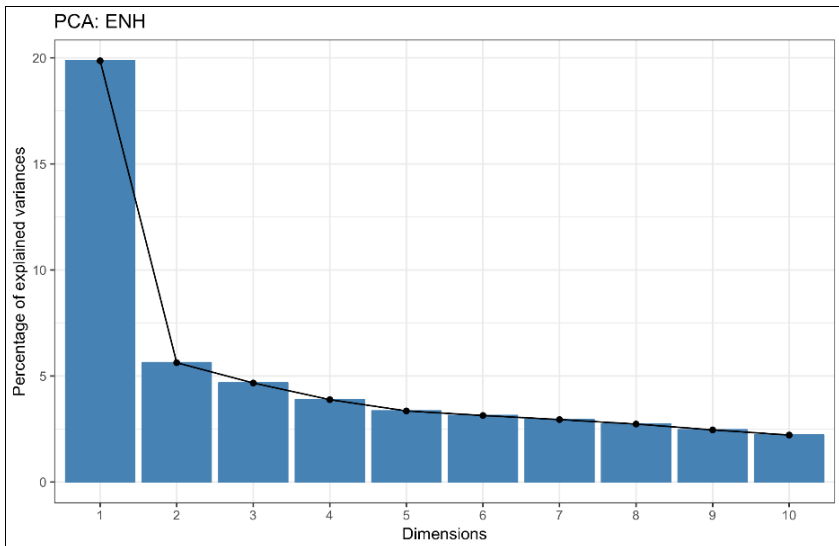


Figura 3. Análisis de Componentes Principales para el conjunto mixto de datos.

Al reducir el número de dimensiones, el resultado de la clusterización del algoritmo *k-means* con los siete clústeres encontrados con el *índice Dunn* permite generar la visualización recogida en la Figura 4, donde se pueden presentar las diferentes localidades, representadas por puntos, contenidas en estos siete clústeres de formas curvilíneas. Los clústeres son de naturaleza circular, por la *distancia euclidiana* que utiliza el método elegido. Esto se puede verificar en la forma de las agrupaciones, donde la mejor muestra está asociada con el clúster 3, que es prácticamente redondo (es decir, no es irregular). El objetivo de clusterizar los datos considerados es encontrar la estructura y organización de los datos, que a simple vista es difícil localizar, sobre todo, en altas dimensiones como es este caso. De esta manera, los clústeres conformados se interpretan como los tipos de viviendas encontrados a través del proceso de la caracterización en función de sus atributos.

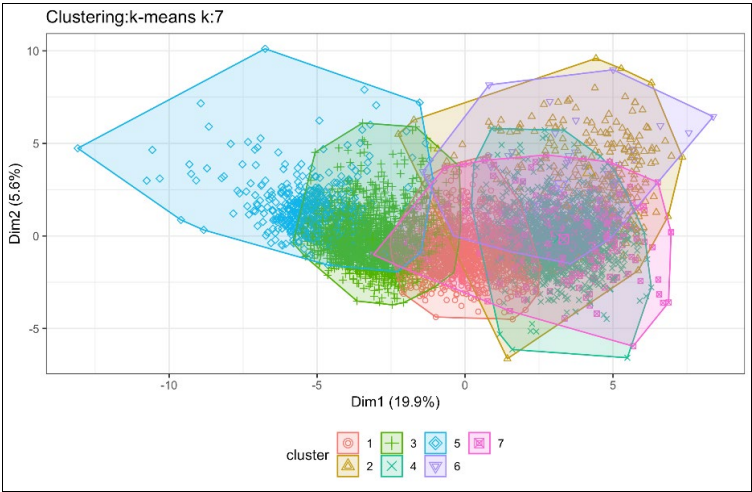


Figura 4. Los 7 clústeres encontrados por medio del algoritmo k-means.

Por lo tanto, al conjunto de datos asociados a la encuesta del INEGI, que contiene las localidades y sus características, se le puede agregar la columna del tipo de vivienda, resultado de la clusterización realizada en este trabajo. La distribución de estas localidades se encuentra en la Tabla 2, donde se pueden ver las 5, 640 localidades que poseen información y algunas variables a modo de ejemplo de la caracterización realizada. Con propósitos ilustrativos y para comprender qué características tienen en común cada clúster conformado se mencionan solo algunos de los atribu-

tos (por ejemplo, el tipo de vivienda 3 es un grupo que tiene: 4 habitaciones, drenaje de red pública, internet, entre otros aspectos).

Tabla 2
Distribución de las 5,640 localidades en los 7 tipos

| <i>Tipo</i> | <i>Localidades</i> | <i>Antigüedad</i> | <i>Núm. de Habs.</i> | <i>Drenaje</i> | <i>Internet</i> | <i>Computadora</i> | <i>...</i> |
|-------------|--------------------|-------------------|----------------------|-----------------------------------------------|-----------------|--------------------|------------|
| 3 | 1,664 | 23 | 4 | Red pública | Sí | Sí | ... |
| 4 | 1,624 | 16 | 3 | Tubería que va a dar a una barranca o grieta | Sí | Sí | ... |
| 1 | 1,328 | 19 | 3 | Red pública | Sí | Sí | ... |
| 5 | 441 | 20 | 5 | Red pública | No | No | ... |
| 7 | 283 | 16 | 3 | Tubería que va a dar a una barranca o grieta | Sí | Sí | ... |
| 2 | 246 | 18 | 3 | Una tubería que va a dar a un río, lago o mar | Sí | Sí | ... |
| 6 | 54 | 15 | 2 | Tubería que va a dar a una barranca o grieta | Sí | Sí | ... |

Visualización

Resaltando que el objetivo de esta investigación es encontrar la caracterización geográfica de las viviendas mexicanas (y agruparlas), se procede a la visualización de estos grupos o clústeres. Para ello, se realiza una visualización geográfica de la distribución de los tipos de vivienda en el país, presente en la Figura 5. En dicha representación se pueden visualizar algunos hallazgos interesantes. Un ejemplo de esto, es el hecho de que en el centro del país se observa cómo se concentran la mayoría de las casas del tipo 1 y 3, las cuales cuentan con alrededor de 20 años de antigüedad, mientras en la distribución del centro hacia el Sur del país predomina el tipo 7, donde hay el mayor número de cuartos en las viviendas, y el tipo 4, donde están presentes las viviendas que cuentan con drenaje a través de una tubería con destino a una barranca o grieta.

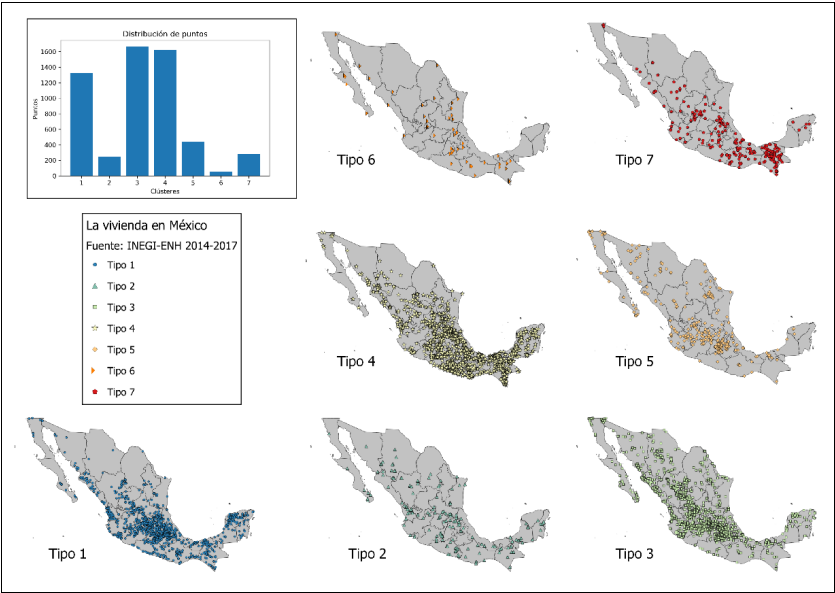


Figura 5. En este gráfico se muestran los 7 tipos de viviendas, así como su distribución de las localidades, encontrados a través de la clusterización de la ENH.

Conclusiones

La implementación de la metodología propuesta permitió encontrar la caracterización geográfica de las viviendas mexicanas, para el periodo 2014-2017, con base en la similitud de sus variables. A partir de este trabajo y la conformación de localidades agrupadas en clústeres se podría realizar un estudio más detallado del estado actual de las viviendas en México. Este estudio, por ejemplo, podría ir orientado a la detección de la brecha digital entre localidades, ya que al evaluar las viviendas del clúster 5, localizadas en la región Norte en frontera con USA, pese a que estas viviendas tienen características similares a las del centro del país, las mismas se caracterizan por no poseer computadora e Internet, por tanto, ¿serían estas las localidades más rezagadas tecnológicamente?

El aporte de este trabajo de investigación se centra en la implementación de técnicas de Ciencia de Datos mediante el preprocesamiento de los datos y Aprendizaje Automatizado No Supervisado, que permiten la agrupación de las localidades para la generación de información geográfica y su representación gráfica a través de mapas.

Con respecto al trabajo futuro, este consiste en caracterizar otras variables socioculturales y generar más capas de información para utilizar operaciones espacia-

les que permitan estudiar fenómenos humanos, tales como: el comercio, movilidad, delincuencia, entre otros.

Bibliografía

- Desgraupes, B. (2017). *Clustering Indices*. Lab Modal'X. Paris, Francia: University Paris Ouest.
- Donoho, D. (2017). "50 Years of Data Science", *Journal of Computational and Graphical Statistics*, 26, 749.
- Gibert, K.; Horsburgh, J.S.; Athanasiadis, I.N. and Holmes, G. (2018). "Environmental Data Science", *Environmental Modelling & Software*, 106, 4-12.
- Gold, J. (2009). "Behavioral Geography", en R. Kitchin and N. Thrift, *International Encyclopedia of Human Geography*, Elsevier Science, 282-293.
- Hartigan, J. A. and Wong, M.A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm", *Journal of the Royal Statistical Society*, 28(1), 100-108.
- Oguro, K. (2016). Big data—key to the 4th industrial revolution. *Japan SPOTLIGHT*, 24-27.
- Scheingart, M. and Solís, M. (1995). *Vivienda y Familia en México: Un Enfoque Socio-Espacial Tomo VIII*. Aguascalientes, México, Instituto Nacional de Estadística Geografía e Informática.
- Yang, R. and Jargowsky, P.A. (2005). "Descriptive and Inferential Statistics", en K. Kempf-Leonard, *Encyclopedia of social measurement, Volume 1*, San Diego, California: Elsevier Academic Press, 659-658.

Revista Cartográfica

Definición de la Revista

La *Revista Cartográfica* (RCA) del Instituto Panamericano de Geografía e Historia (IPGH) es una publicación semestral; incluye trabajos inéditos del estado del arte, revisión de la literatura científica, resultados de estudios e investigaciones sobre las actividades relacionadas con el campo general de la Cartografía, Geodesia y/o Información Geoespacial.

Sistema de arbitraje

La *Revista Cartográfica* cuenta con un grupo de evaluadores especialistas de notorio reconocimiento. El proceso de selección de cada artículo implica la valoración de dos o más expertos en el tema, sin identificación de la autoría, con el fin de garantizar un resultado exento de cualquier influencia. Después de la evaluación, los artículos se remiten al Comité Editorial quien lleva a cabo la selección de ellos para ser publicados.

Por otra parte, compete a la editora responder dudas y resolver cualquier situación que se suscite, así como la decisión final acerca del contenido de cada número de la *Revista Cartográfica*.

Instructivo para autores

Los lineamientos generales para presentar trabajos para su publicación, son los siguientes:

- Todo artículo sometido debe ser **original**, y no publicado, ni considerado para publicación en otra revista.
- La **extensión máxima** de los artículos debe ser de 25 páginas formadas y las llamadas de nota de 5 páginas.
- Los artículos podrán ser escritos en cualquiera de los cuatro idiomas oficiales del Instituto: **español, inglés, francés y portugués**. En el caso de artículos escritos en inglés, francés o portugués, evitar corte de palabras.
- El nombre de los autores, la institución a la que pertenecen, sus direcciones postal y electrónica se incluirán a pie de página al inicio del artículo.
- Cada artículo debe ser precedido por un **resumen** corto (máximo 110 palabras), el cual debe permitir al lector tener una idea de la importancia y campo que abarca el artículo, debe presentarse al menos en español e inglés.
- Inmediatamente después del resumen, se escribirán no más de seis **palabras clave** representativas del contenido general del artículo y características de la terminología usada dentro de un campo de estudio.
- Dentro del texto, si se trata de una cita textual que abarque como máximo dos líneas, se citará el autor, se transcribirá entre comillas y enseguida entre paréntesis se apuntará el año y número de página(s). Si la cita abarca más líneas, se transcribirá el párrafo o párrafos con una sangría, sin comillas tal como se indica en la plantilla.
- Las fotografías, figuras, gráficas, cuadros y tablas deberán ser presentadas listas para ser reproducidas y su colocación dentro del texto se indicará claramente.

- Los artículos deben ser colocados en la **plantilla** correspondiente la cual debe ser solicitada al editor responsable o al Departamento de Publicaciones en la Secretaría General.
- Se incluirá la **Bibliografía** consultada al final del artículo respetando el siguiente formato:
 Autores (apellidos, iniciales nombres.), (año entre paréntesis). “Título del artículo”, *Título de la revista*, vol. (núm.), Editorial, Ciudad, número de páginas (separadas por guión).
 Seemueller, W. y Drewes, H., (1998). “Annual Report of the RNAAC SIRGAS”, *IGS 1997 Technical Reports*, IGS CB, Pasadena, pp. 173-174.
 En el caso de tesis o libros colocar el número de páginas total al final de la referencia.
- Todos los autores deberán observar estos lineamientos.
- Los artículos deben enviarse a la Editora de la *Revista Cartográfica*, quien los someterá a dictamen anónimo de dos especialistas e informará el resultado a los autores en un plazo no mayor de un año:

María Ester Gonzalez
 Editora *Revista Cartográfica*
 Universidad de Concepción
 Juan Antonio Coloma 0201, CP 4440000,
 Los Ángeles, Chile, Teléfono:+56 432405244
 Correos electrónicos: mariaesgonzalez@udec.cl | maria.ester.gonzalez@ipgh.org

No se devolverá el material enviado.

Función editorial del Instituto Panamericano de Geografía e Historia

El IPGH publica seis revistas, impresas y distribuidas desde México. Estas son: *Revista Cartográfica*, *Revista Geográfica*, *Revista de Historia de América*, *Antropología Americana*, *Revista de Arqueología Americana* y *Revista Geofísica*.

La Secretaría General invita a todos los investigadores y profesionales de las áreas de interés del IPGH: cartografía, geografía, historia, geofísica y ciencias afines, a que presenten trabajos de investigación para que sean publicados en nuestras revistas periódicas.

Si requiere mayor información, favor de comunicarse con:

Guadalupe Romero Mayoral
 Departamento de Publicaciones
 Secretaría General del IPGH
 Ex-Arzobispado 29, Colonia Observatorio, 11860 Ciudad de México, México
 Tels.: (+52-55) 5277-5888 / (+52-55) 5277-5791 / (+52-55) 5515-1910
 Correo electrónico: publicaciones@ipgh.org

Edición del
Instituto Panamericano de Geografía e Historia
realizada en su Centro de Reproducción
Ex Arzobispado #29, Colonia Observatorio
11860, México, D.F.
Teléfono 5277-5791 5277-5888
publicaciones@ipgh.org
2018

ESTADOS MIEMBROS DEL INSTITUTO PANAMERICANO DE GEOGRAFÍA E HISTORIA

Argentina

EL IPGH, SUS FUNCIONES Y SU ORGANIZACIÓN

Belice

El Instituto Panamericano de Geografía e Historia (IPGH) fue fundado el 7 de febrero de 1928 por resolución aprobada en la Sexta Conferencia Internacional Americana que se llevó a efecto en La Habana, Cuba. En 1930, el Gobierno de los Estados Unidos Mexicanos construyó para el uso del IPGH, el edificio de la calle Ex Arzobispado 29, Tacubaya, en la Ciudad de México.

Bolivia

Brasil

Chile

En 1949, se firmó un convenio entre el Instituto y el Consejo de la Organización de los Estados Americanos y se constituyó en el primer organismo especializado de ella.

Colombia

Costa Rica

El Estatuto del IPGH cita en su artículo 1o. sus fines:

Ecuador

1) Fomentar, coordinar y difundir los estudios cartográficos, geofísicos, geográficos e históricos, y los relativos a las ciencias de interés para América.

El Salvador

2) Promover y realizar estudios, trabajos y capacitaciones en esas disciplinas.

Estados Unidos de América

3) Promover la cooperación entre los Institutos de sus disciplinas en América y con las organizaciones internacionales afines.

Guatemala

Solamente los Estados Americanos pueden ser miembros del IPGH. Existe también la categoría de Observador Permanente, actualmente se encuentran bajo esta condición: España, Francia, Israel y Jamaica.

Haití

Honduras

El IPGH se compone de los siguientes órganos panamericanos:

México

1) Asamblea General

Nicaragua

2) Consejo Directivo

3) Comisión de :

Panamá

Cartografía (Costa Rica)

Geografía (EUA)

Paraguay

Historia (México)

Geofísica (Ecuador)

Perú

4) Reunión de Autoridades

República Dominicana

5) Secretaría General (México, D.F., México)

Uruguay

Además, en cada Estado Miembros funciona una Sección Nacional cuyos componentes son nombrados por cada gobierno. Cuentan con su Presidente, Vicepresidente, Miembros Nacionales de Cartografía, Geografía, Historia y Geofísica.

Venezuela



Análise exploratória da pegada digital dos turistas para a identificação de padrões espaciais e temporais em destinos urbanos *Luis Encalada, Jorge Rocha y Carlos Ferreira* • Creación de un modelo semántico para modelar los fenómenos meteorológicos de los siglos XIX y XX en Latinoamérica *Diana Comesaña y Bryan Barreiro* • La calidad de las IDE desde el punto de vista de la interoperabilidad *Antonio F. Rodríguez Francisco y Javier Ariza* • Cuantificación de vulnerabilidad y riesgo: las inundaciones en Motozintla de Mendoza, Chiapas, México *Mary Frances Rodríguez Van Gort* • Calidad en datos geográficos, geoservicios y productores de datos: análisis crítico *Francisco Javier Ariza y Antonio F. Rodríguez* • Reproducibilidad científica: ¿Qué es y por qué debemos interesarnos en geo-ciencias? *Daniela Ballari* • Datos enlazados de información geográfica del Ecuador *Víctor Saquicela, Fernando Baculima y Lucia Lupercio* • Caracterización geográfica de la vivienda en México: un enfoque de Ciencia de Datos *Jacobo Gerardo González León y Miguel Félix Mata Rivera*

ISSN 0080-2085