A user-centric framework to improve the reusability of open geodata in cities

Dissertation submitted by Manuel Fernando Benitez Paez to apply for the joint doctorate degree from the Universitat Jaume I, Universida de Nova de Lisboa and Universität Münster.

UNIVERSITAT
JAUME·I

European Joint Doctorate Marie Sklodowska-Curie in Geoinformatics

Doctoral School from Universitat Jaume I

Manuel Fernando Benítez Paez

Supervisors:

Dr. Joaquin Huerta   Dr. Roberto Henriques   Dr. Christian Kray
(Universitat Jaume I)   (Universidade Nova de Lisboa)   (Universität Münster)

Castelló de la Plan (Spain), November 2018

Every word in this dissertation is dedicated to my mom, Flor Paez, my aunt Cenaida Paez, my patient and lovely, Carolina Jimenez, and finally to my little brother Linton Paez who from heaven is guiding all my steps. All of this is for each of you.

Acknowledgments

There are many people who contributed in certain ways to achieve the results presented in this research. First, to all, I would like to thank my main supervisor Joaquin Huerta, to the guide and all support along this research also to introducing me to the European geospatial community and its challenges, projects, communities and overall to help me to start an academic carrier in the European research community. Special thanks to Michael Gould who always provided me with new insights and different ways to tackle different open data issues. Special thanks to professor Christian Kray for his support, new ideas, comments and always helping to get better results in my publications and the GeoC outcomes. I am very grateful to Sergi Trilles for his support in all my research, and pushing me to move forward and reach always better results. Special thanks to professor Auriol Degbelo of the Institute of geoinformatics of the University of Münster for his patience, guidance, support and especially to helping me to understand how the scientific contributions should be and the right method to follow. Special thanks also to my co-supervisor Roberto Henriques for helping me in my external semester in Lisbon finding the way to reach the initial publication presented in this research. I am very grateful to professor Alexis Comber from the school of the geography of the University of Leeds for the guide in my second publication presented in this research.

Besides my advisors, I would like to thank the two international experts Dr. Rolf de By and Dr. Çetin Cömert for their comments, suggestions, and questions that make of this thesis a better contribution. To my colleagues and friends in the GeoTec research group and the GeoC project, their interesting comments, and help with the activities that this research carried out. I want to express my acknowledgement to the four data authorities who allow me to perform the participatory workshops and always be interested in the topic presented in this investigation. To Pamela Mayorga in IDECA - Bogotá, Julio Muñoz in IDECS -

Resumen

Open Data cuenta hoy en día con un profundo impacto en un ambiente donde la información es creada y compartida en todos los niveles. En los gobiernos locales, las iniciativas de datos abiertos han resultado en altos niveles de transparencia. Un importante nivel de empoderamiento entre los tomadores de decisiones y los ciudadanos ha cambiado la forma en que los datos, su reuso y como son usados por gobiernos locales. Sin embargo, el actual reto de las iniciativas de datos abiertos de las ciudades va mucho más allá de la accesibilidad, y el desarrollo de marcos de trabajo que permitan evaluar el impacto de los actuales proyectos de datos abiertos se ha convertido en un elemento destacado en cualquier estrategia, desde amplios catálogos de datos hasta incrementables niveles de reuso. No obstante, a pesar de contar cada día con más catálogos de datos liberados, solo algunos estudios se han centrado en el problema de la reusabilidad, cuáles son las barreras que evitan que los datos geográficos abiertos se usen efectivamente, tomando en cuenta principalmente el punto de vista de los usuarios de datos. Aunque algunos "frameworks" ilustran el nivel de madurez de las iniciativas de datos abiertos a nivel nacional, parece que se requieren marcos de trabajo a nivel local, los cuales ciertamente no se toman en cuenta de forma adecuada en la literatura. Se necesita un marco de trabajo que guie a las autoridades de datos locales a incluir o enganchar a los actuales usuarios además de permitirles mover hacia una estrategia ascendente.

Esta investigación contribuye con tres elementos en este sentido. El primero es un estado actual del nivel del reuso de los datos geográficos abiertos en las ciudades. Segundo por la delimitación de una taxonomía de barreras que los usuarios de datos en Colombia y en España deben enfrentar. Y finalmente la tercera contribución es un conjunto de elementos que forman una estrategia centrada en los usuarios enmarcados en un marco de trabajo o "Framework" qué

ayuda a las autoridades de datos a mejorar el nivel de reuso de los datos en geográficos publicados en las actuales iniciativas locales de datos abiertos. La taxonomía propuesta y framework están basados en la revisión literaria, encuesta online y un conjunto de talleres participativos conducidos en cuatro ciudades (Bogotá, Medellín, Cali en Colombia y Valencia en España), con autoridades de datos y comunidades de usuarios de diferentes perfiles y experiencia en datos abiertos.

La taxonomía presentada en esta investigación destaca variedad de obstáculos que evitan que los usuarios puedan hacer un uso efectivo de los datos geográficos disponibles, mencionado algunas de estas barreras, encontramos que los datos desactualizados, baja integración de los productores de datos, y un acceso complejo a los datos publicados, son las barreras más mencionadas y que más preocupan a las comunidades de usuarios.

Una vez las barreras se han identificado y validado con las comunidades de usuarios de datos a través de las ciudades seleccionadas esta investigación definió los elementos que deberán ser incluidos en un marco conceptual de trabajo que permita a las autoridades de datos usar como guía, para mejorar el nivel del reuso de sus actuales iniciativas de datos abiertos. El framework presentado cuenta con partes esenciales en la correcta implementación de los elementos presentados. Inicialmente el framework requiere de una base definida por tres elementos denominados "Impact Enablers" o habilitadores de impacto. Estos habilitadores se han considerado relevantes por ser identificados en la literatura validada como elementos que permiten un reuso sostenibles en otras iniciativas de datos abiertos. Los habilitadores de impacto son: A) Comunidades de usuarios, pero en especial sus requerimientos en datos abiertos. B) Iniciativas de datos abiertos a nivel local o ciudad como una forma mas efectiva de promover con usuarios el proyecto de open data. Finalmente, C) Un enfoque geográfico, como un potencializado de reuso de datos en los usuarios, gracias a ser el tipo de datos que mas se consulta y se solicita por parte de esta comunidad.

La segunda parte del marco centrado en el usuario es compuesta por cuatro recursos/elementos que se proponen deben estar conectados. 1) Una completa y constante identificación de las comunidades de usuarios de datos y su demanda de datos. 2) La comunidad del reuso. Un conjunto de herramientas para promover la reusabilidad de los datos geográficos publicados. 3) Metadatos centrados en el usuario y 4) términos de reuso centrados igualmente en el usuario. Estos

elementos mencionados fueron agrupados e incluidos por su relevancia para las comunidades de usuarios en los cuatro casos de éxito de esta investigación. El "framework" propuesto en esta investigación provee una clara hoja de ruta para autoridades de datos locales que buscan reformar su actual iniciativa de datos abiertos hacu un enfoque ascendente que incluya los usuarios de datos como un elemento primordial. Esta investigación finaliza con una discusión y conclusiones además de las posibles limitaciones que los resultados de nuestras contribuciones puedan llegar tener lugar.

Introducción

El número de iniciativas, colaboraciones y proyectos relacionados con datos abiertos aumenta cada día, y la mayoría de los gobiernos nacionales y locales actualmente se esfuerzan por avanzar con implementaciones de datos abiertos. Para estas administraciones, los problemas críticos en el éxito de los portales de datos abiertos han sido la accesibilidad y los datos que contienen. Para crear portales web más útiles y fáciles de usar, se han desarrollado soluciones tecnológicas para respaldar los requisitos de accesibilidad de los organismos públicos nacionales y locales. Estas soluciones han sido impulsadas por regulaciones locales y nacionales para garantizar procesos de publicación de datos abiertos dentro de los organismos públicos (Zuiderwijk and Janssen, 2014). Sin embargo, aún quedan algunos problemas críticos y nuevos desafíos que el movimiento de datos abiertos tiene que abordar. Esta tesis abarca el desafío de la reutilización con el que muchas autoridades de datos están luchando después de la implementación de sus portales de datos abiertos. De manera similar, después de varios años de implementación, la comunicación y la colaboración entre los usuarios de datos y las autoridades no ha progresado tan bien como se esperaba. Por un lado, los usuarios aún solicitan datos más útiles o han perdido confianza en los nuevos portales, y su percepción es que los datos proporcionados no son adecuados para sus necesidades. Por otro lado, las autoridades de datos están continuamente tratando de capacitar a sus usuarios y aumentar los niveles de reutilización de datos.

Motivación

Nuestro trabajo está motivado por un entorno de datos abiertos donde el enfoque descendente (desde los productores de datos hasta los usuarios) se ha estudiado y aplicado en muchos contextos durante varios años. Sin embargo, después de 15 años de implementaciones de datos abiertos, parece que este enfoque no es suficiente para alcanzar los beneficios que se recomendaron. Se necesita un enfoque ascendente (desde los usuarios hasta los productores de datos) como una forma eficaz de desarrollar una iniciativa de datos abiertos sostenible y eficiente que funcione en tres direcciones. El primero es el papel aislado de las comunidades de usuarios de datos en proyectos de datos abiertos en curso. El segundo involucra la creciente atención de las ciudades y su potencial para desarrollar estrategias efectivas entre las necesidades de los ciudadanos y las autoridades locales. Finalmente, el tercero es la posibilidad de incluir datos geográficos como combustible para impulsar iniciativas efectivas. En la siguiente sección, se define el problema abordado en esta investigación y se abordan las preguntas de la investigación.

Problema

En un entorno de datos abiertos descendente, la demanda de datos de los usuarios se considera solo brevemente (The European Data Portal, 2016) y el productor de datos es el que establece qué recursos, herramientas o métodos de retroalimentación deben incluirse. Este escenario unidireccional considera los portales de datos abiertos como el puente que permite a los usuarios de diferentes orígenes acceder a los datos publicados (consulte la parte A de la Figura 1.2). Sin embargo, las autoridades de datos buscan activamente mejorar la reutilización de los datos abiertos disponibles, a fin de permitir que las comunidades de usuarios de datos formen parte de un escenario mejorado, es decir, el enfoque de ascendente. Esta investigación abarca el desafío de la reutilización con el que muchas autoridades de datos están luchando en la implementación de sus portales de datos abiertos. Del mismo modo, la comunicación y la colaboración entre los usuarios de datos y las autoridades no han progresado bien. Por un lado, los usuarios aún requieren datos más útiles o han perdido confianza en los portales nuevos o ya implementados, y perciben que los datos proporcionados no son adecuados para sus necesidades. Por otro lado, las autoridades de datos intentan continuamente capacitar a sus usuarios y aumentar los niveles de reutilización de datos.

Nuestro objetivo es definir un enfoque ascendente que considere que los portales de datos abiertos están más conectados con las necesidades de los usuarios y que no tiene una comunicación unidireccional del productor de datos únicamente. El enfoque sugerido es circular, donde el rol y las contribuciones de las comunidades de usuarios de datos definen los pasos posteriores que se deben tomar en las iniciativas de datos abiertos en curso (consulte la parte B de la Figura ref fig: Problema). Estos pasos son: (a) Metadatos centrados en el usuario que están concebidos para proporcionar información que ayuda a los usuarios a evaluar la utilidad de un conjunto de datos con respecto a su problema; (b) Términos legales centrados en el usuario que ayudan a los usuarios a comprender el nivel de reutilización, en lugar de tener que leer términos de licencia complicados; (c) Desarrollar y fomentar una comunidad de reutilización en la que los usuarios con diferentes perfiles y necesidades puedan ver fácilmente otros ejemplos, casos de uso, comentarios o sugerencias de otros sobre la reutilización de datos abiertos; finalmente, (d) una identificación completa de los usuarios de datos y su demanda de datos para tener una mejor idea de los servicios más solicitados, problemas o requisitos de calidad de los datos, y crear eventos, metodologías y herramientas centradas en el usuario que sean relevantes para el Iniciativa local de datos abiertos.

En resumen, la siguiente lista establece las preguntas de investigación definidas en esta tesis y mencionadas anteriormente.

- **RQ1:** ¿Cuál es el estado actual de la reutilización de datos abiertos a nivel local (ciudades)?

- **RQ2:** ¿Qué barreras impiden la reutilización de datos geográficos abiertos por parte de los usuarios de datos locales?

- **RQ3:** ¿Qué estrategia se necesita para mejorar el nivel de reutilización de los datos geográficos abiertos a nivel local (ciudad)?

Contribuciones

- **Estado de reutilización actual de datos abiertos en iniciativas locales**
  Conocer el estado actual del nivel de reutilización de datos abiertos es el
  principal requisito para poder contribuir luego a mejorarlo. La literatura con-
  tiene mucho trabajo sobre el impacto, las implicaciones y el valor agregado,
  pero aún falta una definición precisa de lo que significa la reutilización.

- **Taxonomía de las barreras de los usuarios de datos:** Una vez que se
  definió claramente el estado actual, nuestro segundo objetivo fue determinar
  las barreras o fricciones desde la perspectiva del usuario a nivel local que
  impiden el uso adecuado y la reutilización de los datos publicados. Esta ruta
  nos llevó a investigar el estado actual de los portales de datos abiertos en
  las ciudades mediante la recopilación de barreras antiguas, los formatos de
  datos requeridos, los tipos de datos más utilizados y las recomendaciones
  de usuarios de diferentes orígenes para establecer una taxonomía de las
  barreras de los usuarios de datos (Benitez-Paez et al., 2017).

- **Un marco conceptual para mejorar la reutilización de los datos geo-
  gráficos abiertos en las ciudades:** La tercera contribución es un marco
  centrado en el usuario para ayudar a las autoridades locales de datos a
  reformular su estrategia de datos abiertos en curso hacia una estrategia
  centrada en el usuario. Validado con usuarios de datos en la ciudad de Valen-
  cia (España), este marco se basa en un conjunto de recomendaciones de
  usuarios de datos en cuatro ciudades (Bogotá, Cali y Medellín en Colombia
  y Valencia en España) (Benitez-Paez et al., 2018).

Marco Teórico

El número y la variedad de iniciativas de datos abiertos han aumentado a nivel
mundial. Varios estudios se han centrado en las barreras, las características
clave, los desafíos de implementación, el uso y los casos de impacto. Peled
(2013) analizó los problemas actuales de la primera versión de datos abiertos
en los EE. UU. después del lanzamiento del programa de Gobierno Abierto de
la administración de Obama. Propusieron una nueva etiqueta *Open Data 2.0*
para incluir estrategias de datos liberadas. Señalan el potencial temporal de la
política de transparencia, es decir, situaciones en las que las políticas solo se

implementan durante cortos períodos de tiempo. Sieber and Johnson (2015) analizó el rol cambiante del gobierno en la "cultura" de datos abiertos y presentó cuatro modelos en los que los ciudadanos, el sector privado y las autoridades realizan cambios para determinar la efectividad de las futuras estrategias de datos abiertos.

El potencial de los datos abiertos se relaciona no solo con el gobierno o la autoridad local, sino también con partes interesadas con diferentes roles, intereses y contribuciones. Una consideración crítica es cómo el valor agregado se crea a través de la cadena de datos y análisis Carrara et al. (2017a). Kassen (2013) exploró el potencial empoderador de los datos abiertos a nivel local como una plataforma para la participación ciudadana, basada en un caso de uso en Chicago. Kassen señaló que los esquemas de descendentes a nivel local pueden proporcionar una plataforma efectiva para recopilar y comprender los requisitos de los ciudadanos. El papel del gobierno local en el movimiento de datos abiertos también es cada vez más reconocido. Por ejemplo, la comision euroepa señaló que *datos abiertos tiene el potencial de convertirse en el activo mas valioso para los ciudadanos, las empresas y las autoridades públicas* (The European Data Portal, 2016).

En general, hay una gran cantidad de trabajos publicados sobre la reutilización de datos abiertos debido a que es uno de los desafíos para las iniciativas actuales. Barry and Bannister (2014) seleccionó las ocurrencias de algunos temas relacionados con datos abiertos, mencionando que el intercambio y la reutilización de datos son dos temas con un gran número de ocurrencias, lo que demuestra un enfoque en aprovechar al máximo el recurso de información del sector público. La literatura ha analizado los angulos economicos Johnson et al. (2017); Jetzek et al. (2012); Ahmadi Zeleti et al. (2016), técnicos, institucionales Yang et al. (2015); Cranefield et al. (2014), factores políticos y regulador Nug (2015); Jetzek et al. (2013) que influyen en la cadena de valor de los datos abiertos, lo que sugiere que los beneficios teóricos no se ha visto como se esperaba en las ciudades Carrara et al. (2017b); Janssen et al. (2012); Conradie and Choenni (2014); Barry and Bannister (2014); Attard et al. (2015); Cranefield et al. (2014).

Con respecto a la intersección de iniciativas entre esfuerzos nacionales y locales, ambos trabajando para mejorar los esfuerzos de su iniciativa de Gobierno Abierto, en muchos casos, los conjuntos de datos se ofrecen en varios sitios web Yang et al. (2015) de forma fragmentada, que es difícil de encontrar. También se

necesitan metadatos adecuados para mejorar la reutilización de datos Janssen et al. (2012). La cooperación entre los departamentos municipales para promover y gestionar el proceso de intercambio de datos es un desafío diario relevante para muchos organismos públicos. Crear un marco para la colaboración podría ser una de las formas en que una IDE local puede contribuir a proyectos locales de datos abiertos, debido a su experiencia en el desarrollo de acuerdos legales, técnicos e institucionales. No obstante, las IDE locales también tienen muchas barreras que superar con respecto al intercambio de datos. Carrara et al. (2017a) identificó las características de los datos geográficos como barreras técnicas para la reutilización de las iniciativas de datos abiertos. Las diferencias entre estándares, formatos, tamaño de archivos, estándares de calidad de datos, derivadas de Directivas como INSPIRE [1].

Para resumir, el trabajo realizado por otros autores (mediante entrevistas, encuestas y/o talleres) ha identificado un conjunto de barreras principalmente desde el punto de vista del productor de datos, donde las autoridades nacionales desempeñan el papel principal de las iniciativas de datos abiertos. En cuanto a la reutilización de datos abiertos, no hay mucho trabajo realizado; encontramos solo cuatro referencias relacionadas, ninguna de ellas ha considerado el posible potencial de los datos geográficos o el papel que pueden desempeñar las ciudades. El número de artículos que examinan los obstáculos de reutilización desde la perspectiva de un usuario de datos también es limitado.

Metodologia

Durante este capitulo 5, se describen las etapas consideradas en esta investigación. La figura 3.8 muestra inicialmente que la revisión de la literatura incluía la validación del estado actual del estado de los datos abiertos a nivel local. Esta revisión de la literatura se realizó para validar cómo otros autores han considerado el tema de la reutilización, analizando otros casos de uso, las implicaciones y las barreras encontradas que impiden la reutilización efectiva de los datos abiertos.

Una vez que la revisión de la literatura revela algunas barreras y consideraciones tales como la importancia del nivel local, la brecha entre las actividades de los productores de datos y los requisitos de los usuarios de datos y, finalmente, el tipo de datos que podría alentar a las comunidades de usuarios de datos a

---

[1] Website INSPIRE, disponible en el enlace https://inspire.ec.europa.eu/ consultado el 5 de febrero de 2018

reutilizar más activamente los datos de publicación. La encuesta en línea fue diseñada para preguntar directamente a las personas en las ciudades acerca de su percepción sobre la reutilización de datos abiertos. La encuesta en línea compartida públicamente permite a esta investigación definir qué ciudades eran más participativas, teniendo a Bogotá, Medellín, Cali en Colombia y València en España como ciudades con más participantivas, por lo tanto, fueron las ciudades seleccionadas para la siguiente etapa.

Con más de 113 usuarios de datos en las cuatro ciudades, lideramos una discusión y un conjunto de actividades adecuadamente descritas en el capitulo 3 donde la reutilización, el impacto de los datos abiertos y los requisitos de los usuarios de datos fueron los temas clave que recopilamos para validar las barreras identificadas en La encuesta mencionada. Dos talleres participativos adicionales fueron conducidos a obtener la retroalimentación requerida sobre el resultado preliminar y los supuestos presentados en esta investigación. La conferencia AGILE 2017 fue la manera de discutir los resultados de la encuesta en línea y algunas de las recomendaciones de los usuarios de datos recopilados en los talleres anteriores.

Los resultados de estas actividades iniciales se describieron correctamente en el artículo de investigación Benitez-Paez et al. (2017), en el que definimos una taxonomía de las barreras de los usuarios de datos para ayudar a las autoridades de datos a identificar fácilmente dónde están teniendo más problemas sus comunidades de usuarios de datos.

Sin embargo, la definición de una taxonomía de barreras parece no ser suficiente, se requiere un tipo de guía para ayudar a las autoridades de datos a reformular sus estrategias actuales considerando un enfoque ascendente e incluir las recomendaciones hechas por usuarios de datos en los talleres participativos.

Incluyendo el conjunto de recomendaciones de los talleres participativos anteriores, la revisión de la literatura y las preocupaciones más mencionadas en la taxonomía, se diseña y valida un marco conceptual en un taller participativo adicional con usuarios de datos de la ciudad de València. La descripción completa de los elementos de este marco de trabajo se describe adecuadamente en el capitulo 5

Barreras que evitan la reutilización de datos geográficos en las ciudades, desde la perspectiva de los usuarios de datos

A través de tres actividades principales, nuestro trabajo se dedicó a analizar en toda la literatura el estado actual de la reutilización de datos geográficos abiertos. En segundo lugar, mediante una encuesta en línea exhaustiva, preguntamos a los usuarios de datos, especialmente en las ciudades, sobre sus impresiones, desafíos, barreras, recomendaciones y su opinion general sobre el nivel de reutilización de los geodatos abiertos en sus ciudades. Esta encuesta nos permite tener una mejor imagen de los problemas potenciales que los usuarios locales enfrentan actualmente en relación con la reutilización. A pesar de la cantidad de respuestas válidas (195 respuestas válidas, consulte el Capítulo anterior 3), la validación y la discusión cara a cara fueron necesarias para corroborar lo que las personas dicen y lo que las personas hacen.

Teniendo en cuenta solo las ciudades más participativas, hemos seleccionado cuatro ciudades para organizar un conjunto de talleres participativos. Durante esos talleres participativos, nos reunimos con autoridades de datos locales, comunidades de usuarios de datos con diferentes antecedentes y experiencia en datos abiertos, en torno al concepto de reutilización. Recolectando barreras y recomendaciones, esta discusión fue la manera de validar las barreras identificadas. El resultado de esas actividades es la taxonomía de las barreras de los usuarios de datos y el estado actual del nivel de reutilización de geodatos abiertos que describe el capitulo 4.

Marco para mejorar el nivel de reutilización de datos geográficos abiertos en las ciudades

A través de la taxonomía de las barreras, describimos los problemas más comúnmente mencionados que enfrentan los usuarios de datos en las iniciativas en curso de datos abiertos. Las categorías y las barreras descritas en el Capítulo 4 se basaron en el documento de investigación de Benitez-Paez et al. (2017). Actualmente, autoridades como IDECA en Bogotá o Valencia lideran la campaña de datos abiertos en sus respectivas ciudades. Sin embargo, buscan activamente incluir más usuarios en sus iniciativas de datos abiertos, dado que el enfoque descendente no ha tenido éxito en los últimos años.

Aunque la descripción de las barreras que enfrentan los usuarios de datos es una herramienta útil para los autoridades locales de datos abiertos, esta

investigación propone una forma centrada en el usuario para superar los problemas identificados y ayudar con acciones específicas también sugeridas por sus usuarios de datos. El capítulo 5 está dedicado a explicar el marco centrado en el usuario y sus elementos. Este marco tiene dos partes: la primera parte se define como 'Habilitadores de impacto', que son tres recursos que las autoridades locales de datos abiertos pueden usar para mejorar el impacto esperado de la implementación de datos abiertos.

Cuatro acciones concretas y conectadas sugeridas por los usuarios de datos en las cuatro ciudades para mejorar la reutilización de los datos geográficos abiertos se presentan como la segunda parte de este marco. Se describen en la Sección 5.2, de la siguiente manera: 1) Metadatos enfocados en el usuario, 2) Comunidad de reutilización, 3) Identificación de los usuarios de datos y su demanda de datos abiertos y, por último, 4) Términos legales enfocados en la reutilización. Las cuatro acciones incluidas en este marco centrado en el usuario se validaron con usuarios de datos y expertos en datos abiertos en dos eventos, con diferentes enfoques. Este capítulo se basa en el reciente trabajo de investigación publicado en la revista Transactions in GIS Benitez-Paez et al. (2018).

Discusión

Estudios anteriores han abordado las iniciativas de datos abiertos a nivel nacional como la escala con la que se pueden configurar los posibles beneficios y los obstáculos que dificultan la implementación de datos abiertos cuando se reutilizan. La mayoría de ellos ha considerado el proceso de publicación de datos y la opinión del productor de datos como el núcleo de cualquier sistema de datos abiertos.

Esta investigación presenta otra perspectiva al tomar las necesidades del usuario de los datos como el foco para mejorar la reutilización de datos geográficos abiertos a nivel local. El estudio incluye tres elementos para una mejor perspectiva de datos abiertos. Estos tres elementos son (1) datos geográficos, como el tipo de datos más comúnmente solicitados por los usuarios de datos; (2) utilizar el nivel local o el nivel de la ciudad como una forma de mejorar la comunicación entre los usuarios de datos y los editores; y (3) la parte más relevante de esta investigación ha sido el papel que las comunidades de usuarios de datos pueden desempeñar en las estrategias actuales de datos abiertos, lo que implica identificar sus necesidades para definir acciones, funcionalidades y conjuntos de datos adicionales. La sección 6.1 presenta algunas observaciones sobre el estado

actual del nivel de reutilización en los cuatro casos de uso, y analiza nuestros hallazgos al respecto. La sección 6.2 resume las barreras mencionadas por los usuarios de datos de cuatro ciudades con iniciativas abiertas basadas en diferentes enfoques. La sección 6.3 ofrece una serie de comentarios sobre el papel de las comunidades locales de usuarios de datos y cómo las autoridades de datos enfrentan problemas similares relacionados con las licencias, el proceso de identificación de usuarios de datos, sus necesidades y las estrategias actuales para mejorar el compromiso de los usuarios.

Palabras clave: *Datos Geograficos abiertos*, *Reusabilidad*, *Requerimientos de usuarios de datos abiertos*.

Abstract

Open data has a profound effect in working environments in which information is created and shared at all levels. At the local government level, open-data initiatives have resulted in higher levels of transparency as regards policies. Greater engagement between decision-makers and citizens has changed the way data analysis and evidence are used to support local governance. Initiatives on open data are currently playing an essential role in local governments. However, the current challenge of local open data that authorities are facing has gradually changed from accessibility issues to measures of the impact of the ongoing open-data projects, from more data catalogs to sustainable and increasing levels of reuse of released data, and better reusability of open data. Despite an increasing amount of data being made open, few studies have looked into its level of reusability, and the barriers that hamper the reuse of open geodata from a data consumer's perspective are an issue that most communities of data users are currently faced with. Some frameworks are showing how the level of maturity in national open-data initiatives is either increasing or decreasing, but there is still a need for a specific framework to guide local data authorities to engage their current users and also help them to move toward a bottom-up approach.

This research contributes with three elements in this regard. The first is the current status of the level of reuse of open geodata in cities. This is followed by a taxonomy of the barriers faced by data users in Colombia and Spain, and the third is a set of elements that shape a user-centric framework to help data authorities improve the level of reuse of published open geodata in their ongoing local initiatives. The proposed taxonomy and framework are based on a literature review, an online survey, and a set of participatory workshops conducted in four selected cities (Bogotá, Medellín, Cali in Colombia and Valencia in Spain), with local data authorities and user communities from different backgrounds and with experience

in the field of open data. The taxonomy presented in this research highlights a number of issues such as outdated data, low integration of data producers, and difficulty to access data, the most relevant from the data consumer's point of view being misinterpretation and misuse of released data and their terms of use. Once the barriers had been identified and validated with data users across the selected cities, this research defined the elements included in a conceptual framework that local authorities could use as a guideline to improve the level of reuse in their ongoing open data initiatives. The core elements of this framework are what are defined as 'Impact Enablers', which consist of three aspects considered by the literature reviewed as relevant to improve the positive impact of current initiatives. These three factors are: A) the requirements of data-user communities; B) open data at city level as a way to promote and engage users; and finally, C) a geographic approach to improving the level of reusability of released data due to its potential to engage more users. The second part of the proposed framework is made up of four connected elements: 1) The complete identification of data-user communities and their needs; 2) The community of reuse as a set of technological tools to promote the reusability of released data; 3) User-focused metadata; and 4) Reuse-focused legal terms. The elements mentioned earlier were compiled and included due to their relevance for data-user communities in the four use cases included in this research. This framework provides a clear path for local data authorities to reshape their current open data strategies so as to include data-user requirements and move toward a bottom-up approach. The research ends with a discussion and some concluding points, in addition to several limitations in the application of our findings. At the end of this dissertation, a roadmap for future research and implementations are presented, taking into account some reflections on the framework.

Keywords: *Open Geodata*, *Reusability*, *Users' requirements*, *Impact*.

Index

Index of tables

Index of figures

Acronyms

GIS   Geographic information systems

SDI   Spatial Data Infrastructure

OGD   Open Government Data

IDECA   Infraestructura de Datos Espaciales para el Distrito Capital

IDESC   Infraestructura de Datos Espaciales de Santiago de Cali

EDP   European Data Portal

API   Application Programming Interface

RSS   Really Simple Syndication

INSPIRE   Infrastructure for Spatial Information in Europe

ODB   Open Data Barometer

OECD   The Organisation for Economic Co-operation and Development

1 | Introduction

Data should be able to flow freely
between locations, across borders,
and within a single data space.

Andrus Ansip

Open data holds the promise of "*dramatically reduc[ing] the time and money citizens need to invest to understand what government is doing and to hold it to account*" (The World Wide Web Foundation, 2015). The word "*open*" can be interpreted in many ways (for a recent review, see Pomerantz and Peek (2016)), but in this work it is used in line with the Open Definition: "*Open means anyone can freely access, use, modify, and share for any purpose*"[1]. Providing datasets freely for access and reuse has received an increasing amount of attention from public bodies and society, which see it as a means to improve governance and stimulate knowledge-driven economic growth (Ubaldi, 2013). The concept of open data is now entering the mainstream, with 51 countries (i.e., about 25% of all the countries in the world) having an Open Government Data (OGD) initiative according to The World Wide Web Foundation (2015). OpenDataSoft [2] has listed more than 2,600 open-data portals around the world, which include national, regional and local implementations. The European Data Portal[3], which acts as a gateway to access open data from public administrations in Europe, has listed +860,000 metadata sets to date. Furthermore, the number of countries that

---

[1] The Open Definition, available online: http://opendefinition.org/, accessed on 22 June 2018.

[2] OpenDataSoft, available online: https://www.opendatasoft.com/, accessed on 22 June 2018.

[3] European Data Portal, available online: https://www.europeandataportal.eu/, accessed on 22 June 2018.

endorsed the Open Government Partnership (OGP)[4] is increasing with more than 70 members pursuing the same goal, that is, to make their governments more open and accountable.

The number of initiatives, collaborations, and projects related to Open data is increasing daily, and most national and local governments are currently striving to advance with open-data implementations. For these administrations, the critical issues in the success of open-data portals have been accessibility and the data they contain. To create more useful and user-friendly web portals, technological solutions have been developed to support the national and local public bodies' accessibility requirements. These solutions have been driven by local and national regulations to guarantee open-data release processes within public bodies (Zuiderwijk and Janssen, 2014). However, there remain some critical issues and new challenges the open-data movement has to tackle. This thesis embraces the reusability challenge that many data authorities are struggling with after the implementation of their open-data portals. Similarly, after several years of implementation, communication and collaboration between data users and authorities have not progressed as well as expected. On the one hand, users still call for more useful data or have lost confidence in the new portals, their perception being that the data provided is not suitable for their needs. On the other hand, data authorities are continually trying to empower their users and increase the levels of data reuse.

The sections that follow cover the motivation underlying the work that has been carried out in this thesis, a description of the problem to be dealt with, the research questions addressed, the contributions made throughout this research, an outline of the context of our work in order to highlight where our contributions are located in the current state-of-the-art, and close with an outline of the structure of this dissertation.

---

4   Open Government Partnership, available online: https://www.opengovpartnership.org/, accessed on 22 June 2018.

1.1 | **Motivation**

Without a doubt, the open-data movement is currently holding the attention of the most influential stakeholders at the national and local levels. Since the Public-Sector Information (PSI) Directive in 2003 in Europe,[5] Obama's executive order about open data in 2009,[6] the creation of the OGP initiative,[7] in 2011, the Open Data Charter defined by the G8 in 2013,[8] and the guidelines on recommended standard licenses by the European Commission in 2014,[9] the world has been witness to a remarkable amount of progress regarding open-data principles. Many countries have currently implemented their national policies to follow the trendsetter countries and to gain some of the expected benefits of allowing public access to data that was previously kept private or less accessible by public bodies.

Despite open data being a broad concept that might be applied in several contexts, the spotlight has been focused on the intersection between the concept of open data and public information (Ubaldi, 2013). However, the concept of open data is often confused with the OGD definition. These are different concepts and how we address the definition could be the way to improve the expected outcomes. On the one hand, in the article by Ubaldi (2013), government data is defined as "any data and information produced or commissioned by public bodies". On the other hand, the Open Knowledge Foundation (OKF) established that "Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share-alike". In a study on the impact of reuse of public data prepared by Capgemini Consulting as part of the European Data Portal, the boundaries of open data and Public Sector Information are defined by an intersection of "big data", Public Sector Information (PSI), and open data (see Figure 1.1). This picture offers a general perspective of the amount

---

[5] Public-Sector Information (PSI) Directive, available online: https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information, accessed on 22 July 2018.

[6] Obama's executive order, available online: https://obamawhitehouse.archives.gov/open, accessed on 22 July 2018.

[7] Open Government Partnership, available online: https://www.opengovpartnership.org/, accessed on 22 July 2018.

[8] The international Open Data Charter, available online: https://opendatacharter.net/, accessed on 22 July 2018.

[9] European Commission's guidelines on recommended standard licenses, available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014XC0724(01), accessed on 22 July 2018.

of data currently released. Open data could refer to data created by public bodies associated with the PSI directive or that published by private companies or directly by empowered citizens.
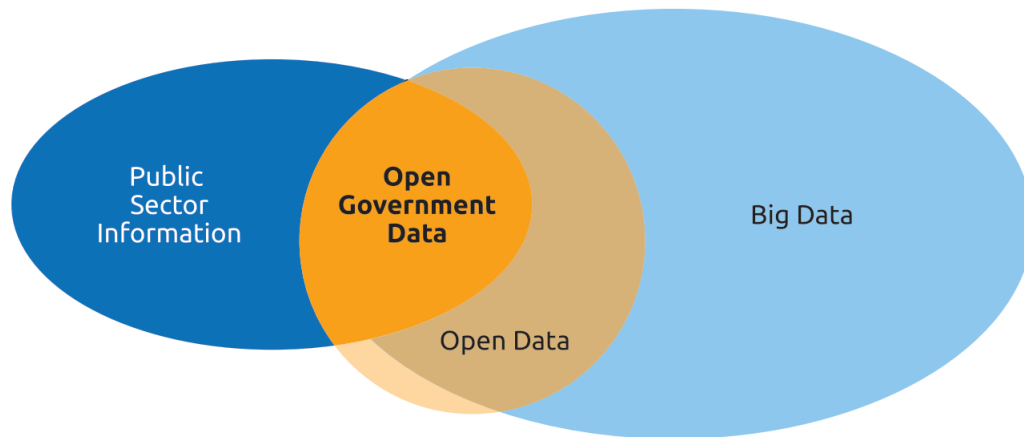
Figure 1.1: Boundaries of Open Data and Public Sector Information. Source: Carrara et al. (2015)

Overall, the concept of Open Data attracts the most attention when comes to OGD initiatives, as mentioned above (Ubaldi, 2013; Attard et al., 2015). The expected results are mostly related to more transparent and efficient governments (Huijboom and den Broek, 2011), getting to grips with growing corruption (Attard et al., 2015; Karolis.Granickas, 2013), improving citizens' participation (Jetzek et al., 2013, 2012), attempting to solve social challenges (Kassen, 2013; Gurstein, 2011; Young and Verhulst, 2016), increasing the economic value and empowering entrepreneurship (Carrara et al., 2017b; Young and Verhulst, 2016; Janssen and Zuiderwijk, 2014), and fostering innovation, efficiency, and effectiveness in government services (Ubaldi, 2013; Carrasco and Sobrepere, 2015; Williamson and Eisen, 2016). With more social and also economic evidence from national and local governments around the globe, it is difficult for any national or local authority to remain isolated from the OGD initiatives. However, Worthy (2015) argued that the impact of open data and especially the release and use of open government data is more complex, more unpredictable, and more political than its advocates claim.

Nowadays, after almost 15 years of open data implementations, most OGD initiatives, data authorities, public bodies, and users' communities are starting to consider what the next steps in open data projects should be. Janssen et al. (2012)

listed five myths about open data and open government, sketching a clear picture of the possible expectations that OGD initiatives may not accomplish. Just the implementation of open data is not enough to reach most of the expected benefits; the role of reuse needs to be clear, and only when the released data is being used can stakeholders obtain the expected value.

In more recent work, Ubaldi (2013) established a value chain of OGD and delimited the stages of data generation and use required to have a positive impact on the changing role of governments, expand the role of private actors, and empower citizens and contribute to civil societies. Attard et al. (2015) presented the OGD life-cycle, in which the pre-processing, exploitation and maintenance steps are defined. Moreover, the impact is also considered in terms of motivations to use open data and a set of challenges in OGD initiatives. Welle Donker and van Loenen (2017) defined a conceptual framework to assess the success of open-data ecosystems. Peled (2013) presented lessons learned from the initial open-data implementations, and proposed a new approach for further open-data initiatives defining what the second version of open data should be like.

Most of what has been written or studied regarding open data has focused on benefits, adoption or implementation barriers, isolated use cases, or positive impact. Little is known about its use, reuse or impact whether negative or positive. According to Andrus Ansip, Vice-President of the Digital Single Market of the European Commission, "*Data should be able to flow freely between locations, across borders, and within a single data space. In Europe, data flow and data access are often upheld by localization rules or other technical and legal barriers. If we want our data economy to produce growth and jobs, data needs to be used. However, to be used, it also needs to be available and analyzed*" (European Commission, 2017). A poor understanding of the impact and the circumstances in which the open-data implementations have been most effective could be a relevant risk to mismatch the current data supply and demand requirements. It is essential to gain better a understanding of how open data, its demand, level of reuse, and the possible impacts through several levels. Young and Verhulst (2016) proposed a methodological attempt to define a framework that establishes a taxonomy of impact for open-data initiatives through 19 use cases from around the world. In another recent example in Spain, reusability is considered a relevant milestone in ongoing open-data initiatives. Thus, Abella et al. (2014) suggested a metric to assess the reuse of open data, considering the relationship between a "smart city"

and open data as the fuel enabling services to be provided directly to citizens.

The latest edition of the Open Data Barometer[10], developed by the World Wide Web Foundation[11] includes four findings after studying the current status of open data for more than 113 countries. This report claimed that Governments are not publishing the data needed to restore citizens' trust, an aspect considered increasingly more frequently in the literature, where data-user requirements were not part of the initial studies or reviews. Gurstein (2011) criticized the possible effect on marginalized citizens, who have more restricted access to technological resources, highlighting the reduced social impact of open data and open data for everyone, both of which are definitions that many initiatives want to establish. Moving in the same direction, Kassen (2013) noted that the inclusion of data users is one of the critical aspects that needs to be considered in open-data projects, and stressed the role of local communities and their knowledge in becoming efficient and productive in open-data projects.

But data users and their demands are not the only relevant aspects that have been considered in recent studies. The level of implementation could have a specific impact on subsequent results or on the sustainability of the open-data project. Although most of the literature addresses the national level, there are a few studies that include the local level as the core of the initiative and, therefore, examine what role could be played by cities in an open-data environment. The fourth analytic report of the European Data Portal (EDP) (The European Data Portal, 2016) established the vital role cities can play in a national open-data strategy. After investigating eight medium-sized European cities, the report found a strong relationship between the local open-data initiatives and smart city initiatives, which allow cities and the reuse of their open data to tackle the current urban challenges and consequently citizens' most important concerns. Kassen (2013) mentioned that the intersection between citizens, but especially citizens' needs, and local level authorities could be more useful than the traditional means of communication, where the outcomes of the strategy, such as an open-data portal, have already been built, thus reinforcing the need to collect data users' needs as primary input of any open-data strategy.

---

[10] Open Data Barometer, available online: https://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf, accessed on 21 June 2018

[11] World Web Wide Foundation, available online: https://webfoundation.org, accessed on 21 June 2018

To sum up, our work is motivated by an open-data environment where the top-down approach has been studied and applied in many contexts for several years. However, after 15 years of open-data implementations, it seems that this approach is not enough to reach the benefits that were advocated. A bottom-up approach is needed as an effective way to develop a sustainable and efficient open-data initiative that works in three directions. The first is the isolated role of data-user communities in ongoing open-data projects. The second involves the growing attention of cities and their potential to develop effective strategies between citizens' needs and local authorities. Finally, the third is the possibility of including geographic data as fuel to drive effective initiatives. In the next section, the problem addressed in this research is defined, and the research questions are addressed.

1.2 | **Problem**

Two widely recognized and very frequently mentioned concepts in recent years are, on the one hand, open data, where the overlap with government data has been the most studied approach (Attard et al., 2015). On the other, there is the concept of a smart city, which is still emerging and has been interpreted by different stakeholders in a variety of ways. Sometimes the concept of smart city is judged regarding technological and sensor-connected solutions throughout cities, while others associate it with an effective integration among citizens' needs, local administrations, industry, and academia (Degbelo et al., 2016b).

In Degbelo et al. (2016b), the authors outlined the trends in technologies, architectures, and infrastructures for smart cities, suggesting that cities are the current focus for many disciplines from social and technological angles. Moreover, Degbelo et al. (2016b) claimed that regardless of the focus of the smart-city implementation, there is a common request, along with some recent experiences, concerning the sharing and use of available data. Although the open-data movement is a key factor in the smart-city concept, the integration with citizens' requirements is also considered another success factor in current use cases[12].

The initial idea to make public data available to promote the transparency and accountability of citizens was essentially a promising idea initially promoted by two well-known administrations (the United States and the United Kingdom). Nowadays, this idea is being criticized due to the lack of impact as well as the amount of effort required to carry out those projects in legal, technical, and political terms. Peled (2013) discussed the idea of having a second and improved version of open data, where politicians must understand that implementing an effective open-data program requires time, patience, and mid- or long-terms goals. The question that arises from this is: Do current or coming administrations have enough patience and commitment to keep supporting open-data initiatives? Is there any strategy that can help today's open-data authorities to demonstrate the effectiveness of those open-data projects?

---

[12] Smart cities: technologies challenges, available online: http://urbanopus.net/smart-cities-technology-challenges/, accessed on 24 June 2018

The sustainability of open-data initiatives has depended mostly on national administrations. Because they have a strong commitment and political support, national authorities have led a national coordinated effort to develop many open-data portals and later populate them with the data that needs to be published. However, the national efforts invested in the open-data strategies do not seem to be enough to reach the desired impact of open data and its portals (The European Data Portal, 2016; Carrara et al., 2017b). One way that some governments have found to achieve a more profound and more tangible result has been through local administrations (i.e., at city level), which see open-data initiatives as an excellent opportunity to move toward the open government approach and gain some of the expected benefits for their local administrations. Moreover, most of the ongoing local open-data initiatives have been implemented and designed by a politician leadership without coordination with the national open-data authorities. The number of national open-data strategies that include local administrations and how cities need to be integrated into the whole ecosystem is significantly low. (The European Data Portal, 2016) listed a set of European cities (local level) that have implemented open-data strategies and included the lessons learned from national implementations. Some of the cities listed have had, or are having, remarkable results in terms of open-data portals, events, and accessibility (i.e., New York, Madrid, London, Puebla). However, other cities have had to tread a more problematic path and invest a significant amount of effort in order to accomplish the results, due to the required coordination between the national and local administrations.

From National Open Data to Local Open Data: Several national governments (i.e., Canada, Australia, Argentina, Uruguay) have robust open-data portals and mature strategies with some lessons learned from previous or initial implementations of open-data initiatives. Cities, especially those with open-data initiatives at a local level (i.e., city level), are attracting more attention due to the potential to reach data-user communities and the increasing amount of work they are doing on well-known topics like smart cities, mobility, air pollution, green living and, above all, on urban challenges (Ubaldi, 2013; The European Data Portal, 2016). However, the transition from national open-data initiatives to local open data is not always the only way that authorities have found to establish a city-level open-data strategy. The European Data Portal (2016) illustrated the link between open-data

strategies and broader digital or smart-city strategies. In the emerging "wave" of smart-city initiatives and their close relationship with open-data projects, several politicians and mayors have seen open data as a new way to promote their results among their voters (Peled, 2013). Regardless of the way that local level open-data strategies have been defined, the aim is changing and nowadays the reusability and impact that the published data could have seems to be one of the priorities. A better understanding of the level of reuse of open data is required. To address this emerging challenge, this dissertation includes the following question *What is the current status of open-data reuse on a local level (cities)?*

Better and early integration of data users' demands and requirements in the local open-data strategy: As mentioned above, the current top-down approach of many open-data initiatives around the world does not seem to be enough to reach the expected benefits, especially regarding effective reuse, added value, and impact in data-user communities.

Although several national and local governments have adopted open-data projects, there is an increasingly strong call from data users, who see current initiatives as a gap between the published data and the real data demand. Although, in the literature, some frameworks include data users as part of the open-data ecosystem, most of them only consider users' communities to evaluate and test the final results and obtain the feedback required to enhance the process. However, many elements cannot be built upon or modified once the open-data initiatives are released, even when data users provide extensive feedback.

Early integration of data users' requirements should be considered in the conceptualization process of any open-data strategy. Even though accessibility is still an important aspect that needs to be part of the priorities, it should not be the principal purpose in current open-data strategies. This dissertation claims that reusability needs to be the primary goal of the open-data strategies and early integration of data users' needs might have a positive effect on future open-data initiatives. Therefore, identifying the current frictions that prevent data users from taking full advantage of published data is the initial step to later suggest what ways could be used to improve the reuse level in local open-data initiatives.

Validation of current barriers considering only the data-user perspective and which of them are related to geographic data leads us to the second research question included in this dissertation: *What barriers prevent the reuse of open geographic data by local data users?*

The potential of geographic data: Welle Donker and van Loenen (2017) proposed an assessment framework to evaluate the success of open-data ecosystems and determined that open geographic data could provide useful tips for other OGD initiatives. (Welle Donker and van Loenen, 2017) compared different kinds of datasets in the Netherlands and found that national geodata was rated significantly higher in all the indicators evaluated, such as being usable, findable, and recognizable, together with reliability, in which data users rated geodata significantly better than non-geodata, or even than open data. In the same direction, in a study on companies transforming open data into economic benefits, (Carrara et al., 2017b) reported that geospatial and statistical data were consulted and reused the most by their respondents. The European Data portal published another analytic report (The European Data Portal, 2016), in which it establishes that geodata is one of the most consulted by data users, especially those from the business sector. This kind of data was grouped in the description of technical barriers mentioned that are due to specificities of the geographic data. Some of these particularities include size, standards, and web services, which are possible sources of friction hampering the reuse of this kind of data in open-data communities that are not familiar with the geographic formats. Being aware of the importance of and the growing attention to geographic data at the local and national levels, the European Commission (2017) outlined an improved set of guidelines on recommended standard licenses, datasets, and charging for reuse across European countries. This guideline defines five categories to prioritize the release and inclusion of geospatial data as the initial datasets that need to be published, due to the high demand for it from reusers across the EU.

Despite the fact that geographic data attracts growing interest from data users, and also the scholars reviewed show that it is valuable to users and therefore should be more open, data authorities nevertheless continue to ignore the potential of geographic initiatives. In a recent study, Reynard (2018) revealed that despite the relevant and significant contributions of a Public Participatory Geographic Information System (PPGIS) developed to work in mobility issues throughout Edmonton (Canada), the city hall ignored its results and took a top-down approach

without considering the data generated by citizens and the empirical results that showed where the city infrastructure could be improved.

Countries have been working on developing National Spatial Data Infrastructures (NSDIs) since 1994, with the goal of establishing metadata standards, developing clearing houses, and enhancing availability among national departments that collect, manage, and analyze geospatial data (Esri, 2010). Although the concept of SDI has been established in many countries, full implementation and its success are still limited (Harvey and Tulloch, 2006). These NSDIs have been transferred to the city level through local Spatial Data Infrastructures (local SDI), in which accessibility, standardization, and high data quality are the primary goals. By developing complicated legal and technical frameworks, local SDI authorities want to be the end-point of all geographic data across the city, thus making it easier for citizens and experts looking for spatial data about the city. However, the integration and standardization of all geographic data from several city departments is often a mid-term task, due not only to the legal barriers but also the high degree of data quality control that most of those authorities need to accomplish as an official data provider. Many of those local SDI have trodden a long, tough path to accomplish, in most cases, a geo-portal and a metadata catalog with geographic data related to many topics. Particularly the process and experience of many of those local SDI might be an adequate aid for open-data initiatives in cities. At the same time, this process faced by those authorities is entirely related to the process that most of the open-data authorities need to lead, especially in cities. Therefore, lessons learned from local SDIs could be transferred to ongoing open-data initiatives at the city level (Welle Donker and van Loenen, 2017).

It is not a case of having to reinvent the wheel, but of listening to other local data authorities, especially those that have previously spent years on formulating the legal and technical framework to guarantee a certain level of standardization and integration. However, as mentioned above, this does not mean that local SDIs do not have any particular reusability issues. In fact, those authorities are facing similar challenges, and the reuse of geographic data is limited for the time being. Notwithstanding the combination of experiences, lessons learned, and data users' requirements, the potential to reuse geographic data should be added for joint development that allows the creation of user-focused open-data initiatives.

Regarding the limited reuse of geographic data, it is important to bear in mind the important role of cities in the implementation of open-data initiatives, and the

increasing need to include data-user communities as part of the whole system, adding their needs to the conceptualization of any open-data strategy. The research question to be put forward in this respect is the following: *What strategy is needed to improve the level of reusability of open geographic data at a local (city) level?*

In a top-down open-data environment, the data demand from users is considered only briefly (The European Data Portal, 2016) and the data producer is the one who establishes what resources, tools or feedback methods should be included. This one-way scenario considers open-data portals as the bridge that allows users from different backgrounds to access the published data (see part A of Figure 1.2). However, data authorities are actively looking to improve the reusability of available open data, so as to allow data-user communities to be part of an improved scenario, that is, the bottom-up approach. This research embraces the reusability challenge that many data authorities are struggling with in the implementation of their open-data portals. Similarly, communication and collaboration between data users and authorities have not progressed well. On the one hand, users still require more useful data or have lost confidence in the new or already-implemented portals, perceiving that the data provided is not suitable for their needs. On the other hand, data authorities are continually trying to empower their users and to increase levels of data reuse.

Our goal is to define a bottom-up approach that considers open-data portals as being more connected to users' needs, and that does not have a one-way communication from the data producer only. The suggested approach is more circular, where the role and contributions from the data-user communities define the subsequent steps to be taken in the ongoing open-data initiatives (see part B of Figure 1.2). These steps are: (a) User-focused metadata that is conceived to provide information that helps users to assess the usefulness of a dataset with respect to their problem; (b) User-focused legal terms that help reusers to understand the level of reuse, instead of their having to read complicated license terms; (c) Developing and fostering a reuse community in which users with different backgrounds and needs could easily see other examples, use cases, comments or suggestions from others about open-data reuse; finally, (d) a full identification of data users and their data demands helps to gain a better idea of the most requested services, data quality issues or requirements, and to create user-focused events, methodologies, and tools that are relevant to the local open-data initiative.

To sum up, the following list establishes the research questions defined in this dissertation and mentioned above.

- **RQ1:** What is the current status of open-data reuse on a local level (cities)?

- **RQ2:** What barriers prevent the reuse of open geographic data by local data users?

- **RQ3:** What strategy is needed to improve the level of reusability of open geographic data at a local (city) level?
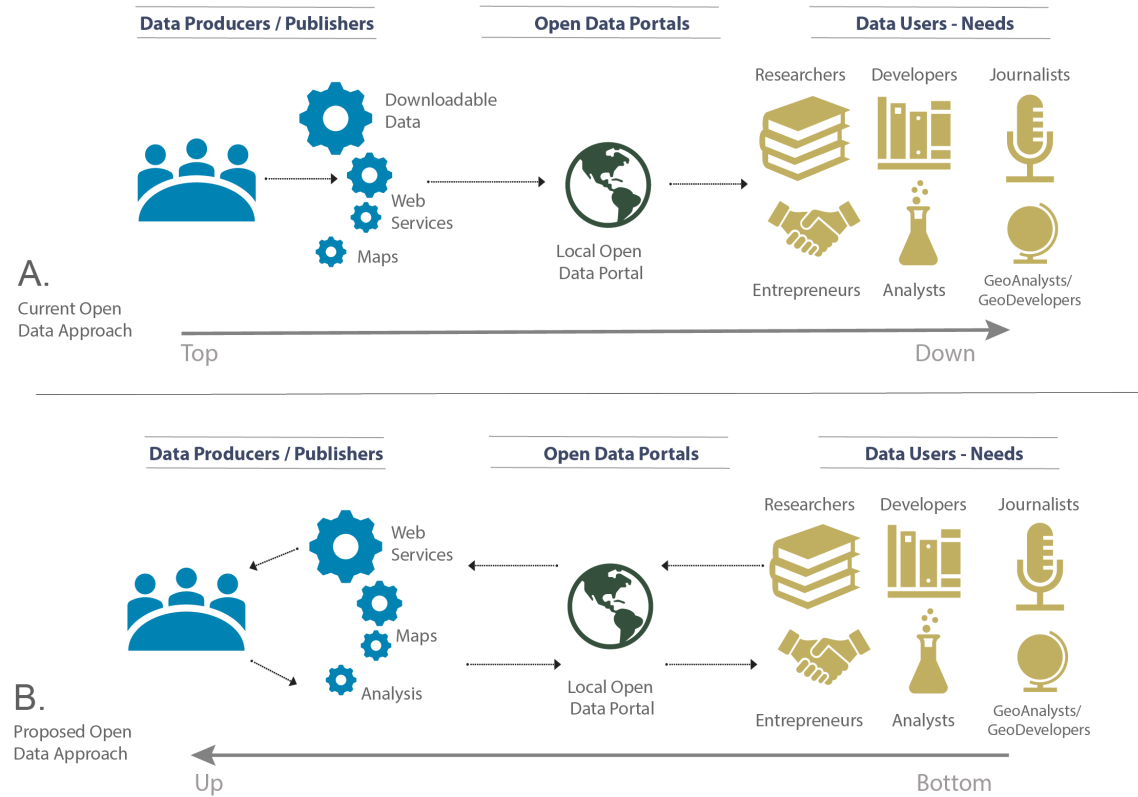
Figure 1.2: A. Current Open-Data Approach (Top-Down), B. User-Centered Open-Data Approach (Bottom–Up)

1.3 | Contributions

After holding discussions, congresses, and forums about open data, and a series of participatory workshops with data authorities, user communities, and private companies that work in the data-releasing process, we have identified a recurring comment that refers to improving the level of reuse of current open-data initiatives.

Nevertheless, if the community of data users comes from Europe or South America, there is a similar concern: the existing open-data portals and strategies do not follow our needs, and the reuse of published data is limited to just short periods or events where the initiatives are promoted. The literature has studied and validated obstacles and suggestions, mostly from the heads of open-data projects or open government initiatives, who shared their experience and lessons learned about how open data should be addressed. However, a set of barriers together with some recommendations considering only the data-user perspective and providing a better understanding of how data users expect to reuse the published data in cities is a rather limited amount of information. There are some isolated use cases where datasets published are included in projects or external applications with particular relevance, for instance, how the public budget was spent, bus and bike station distributions, or air quality measurements. However, the amount of published data is significantly higher than the number of use cases that are expected to be created.

Although some data authorities are aware of this challenge that their open-data initiatives are facing, there is still a lack of knowledge about how the reusability challenge should be tackled. A user-focused strategy is needed to help current open-data leaders to reshape their activities and guide them over the most relevant concerns from data users, and mainly the elements or modifications required to integrate the data-user requirements within the whole open-data system. Therefore, this dissertation presents a set of contributions in this regard.

- **Current reusability status** Knowing the current status of the level of reuse of open data is the principal requirement to be able to later contribute to improving it. The literature contains a great deal of work on the impact, implications, and added value but a precise definition of what reusability means is still lacking.

- **Taxonomy of data users' barriers:** Once the current status had been clearly defined, our second goal was to determine the barriers or frictions from the user's perspective at a local level that prevent the proper use and reuse of published data. This path led us to investigate the current status of open-data portals in cities by gathering old barriers, required data formats, the kinds of data that are most used, and recommendations from users from different backgrounds to draw a taxonomy of data users' barriers (Benitez-Paez et al., 2017). The delimitation of the different categories of obstacles and their recommendations to overcome the current issues of many local open-data projects in cities will help public bodies to define a better path to deal with the data users' level of engagement, as well as to determine where they might focus their ongoing efforts (Benitez-Paez et al., 2017).

- **A conceptual framework to improve the reusability of open geographic data in cities:** The third contribution is a conceptual framework proposed to help local data authorities to reshape their ongoing open-data strategy toward a user-focused strategy. Validated with data users in the city of Valencia (Spain), this framework is based on a set of recommendations from data users in four cities (Bogotá, Cali and Medellín in Colombia, and Valencia in Spain). The proposed framework suggests three impact enablers of sustainable open-data strategies that include data users, the city level, and a geographic approach as the basis for any open-data initiatives that will be launched in a city. Four related and interconnected elements define the conceptual framework: *(1)* Full identification of data users and their data demands; *(2)* Creation of the community of reuse as a set of tools to promote effective and fluent feedback from data users by making a suitable effort in use-case galleries and technical resources like forums and user-ranking stars; *(3)* Implementation of linked data and web semantics, as part of the technological tools required by most of the users; and finally: *(4)* User-focused legal terms, proposing a traffic light system to illustrate to data users what the level of reuse for released datasets is, instead of describing complicated license terms that could lead to some confusion (Benitez-Paez et al., 2018).

1.4 | **Related Projects**

This section describes some projects that are closely related to our work. The lack of concrete results places open-data projects in the spotlight, in which sustainability, citizen engagement, and reusability are the new issues faced by open-data authorities, rather than just accessibility.

European Open Data Portal (EDP): The EDP is a project developed by the European Commission with the support of a consortium led by Capgemini Consulting, including INTRASOFT International, Fraunhofer Fokus, con.terra, Sogeti, the Open Data Institute, Time.Lex and the University of Southampton. EDP *"harvests the metadata of Public Sector Information available on public data portals across European countries. Information regarding the provision of data and the benefits of reusing data is also included"*[13]. The EDP project has been relevant in our work due to its indicators for assessing the maturity of open-data portal initiatives throughout European countries and the extensive studies and analytic reports developed by its consortium.

Open-data maturity is defined as *"A series of indicators [that] have been selected to measure Open Data maturity across Europe. These indicators cover the level of development of national policies promoting Open Data, an assessment of the features made available on national data portals as well as the expected impact of Open Data"*.[14]. This report presents a set of indicators to evaluate the maturity and level of readiness of national open-data initiatives across the EU. Although this report offers an overview of the status of open-data initiatives in European countries, it does not include and has not studied the local level, which means there is no indicator to evaluate the maturity or impact that the local open-data initiatives have across Europe. Furthermore, geographic data is not analyzed or included in this report. Nonetheless, the geographic approach is considered in one of the analytic reports where the barriers to reuse are analyzed. These indicators are based on a survey and interviews with leaders of national open-data authorities who present their results and perspective regarding their internal strategy (but, as mentioned above, this perspective is one of the most analyzed, and it does not

---

[13] European Open Data Portal Website, available online: https://www.europeandataportal.eu/, accessed on 30 June 2018

[14] European Maturity Dashboard, available online: https://www.europeandataportal.eu/en/dashboard, accessed on 20 June 2018

reflect the data users' perception).

Besides the maturity dashboard, EDP has also published a set of reports included in the literature review in this research.

- **Reusing open data portal**: An economic perspective on the potential reuse of open data in Europe.

- **Creating value through Open Data**: The delimitation of the value chain of open data.

- **Analytic Reports 4 and 6 about Open Data in cities**: A general overview of the efforts being made in European cities and their coordination with national open-data strategies, which illustrates the most successful activities and highlighted cities.

Although the EDP has drafted the above-mentioned reports associated with cities, their primary focus is on the national strategies and the collaboration among European institutions. The local impact of open data and the issues faced by local open-data users and their data authorities are still overlooked.

Open Data Barometer (ODB): A study conducted by the World Wide Web Foundation[15] that shows what 115 countries are doing about the principles of the open-data charter[16]. The Open Data Charter is a framework to promote and encourage openness in governments that want to bring about a political change driven by users' demands.

The most relevant principle in our research is the third, which establishes open data as accessible and usable. In the latest ODB report, 73% of datasets were relatively easy to find. Nonetheless, 10% of the datasets surveyed were not available free of charge, and only a quarter of the analyzed datasets were available under an open license. This means that, despite the efforts of many governments to make license terms easier to read, the legal terms are still a substantial barrier hindering the use of open data.

Two additional insights from this report contribute to our goal, bringing our contributions more into context. Initially, the fourth finding of the ODB report is that

---

[15] World Wide Web Foundation website https://webfoundation.org/, accessed on 30 June 2018.

[16] Open Data Charter Principles, available online: https://opendatacharter.net/principles/, accessed on 30 June 2018.

"Governments are not publishing the data needed to restore the citizens' trust". In section 3.3 we will discuss how users are requesting more data associated with their local needs. The second relevant insight is illustrated in Figure 1.3, considering the data users' requirements (Data People Need) and the level of use (Data people can use) as the main parts of an open-data strategy. Finally, the third topic of the ODB is defined as "Data for everyone", which refers to the fact that accessibility still needs to be improved. This sentence is discussed in depth by Gurstein (2011), who presents a critical look at the concept of "data for everyone" and claims that the current approach to open data only empowers those with digital access, while others who have no easy access to technology infrastructure or technology knowledge are not effectively included.

Figure 1.3: The three principles of the Open Data Barometer. Source: 4th Open Data Barometer Edition Report

The Organization for Economic Co-operation and Development (OECD): Barbara Ubaldi is the author of working paper No. 22 on Public Governance, towards empirical analysis of open government data initiatives (Ubaldi, 2013).

This report highlighted the main principles, concepts, and criteria framing open government data initiatives. Mainly focused on accessibility, this report underlines the importance of the conditions for using, reusing, and adding value to published data. The Tim Berners-Lee Five Star Schema (see Figure 1.4) is presented as one way to assess reusability. However, this schema considers reusability as a matter from the data producer perspective. It is a good way to assess the level of reuse but lacks a mechanism to help data authorities to understand where their open-data strategy needs adjusting.
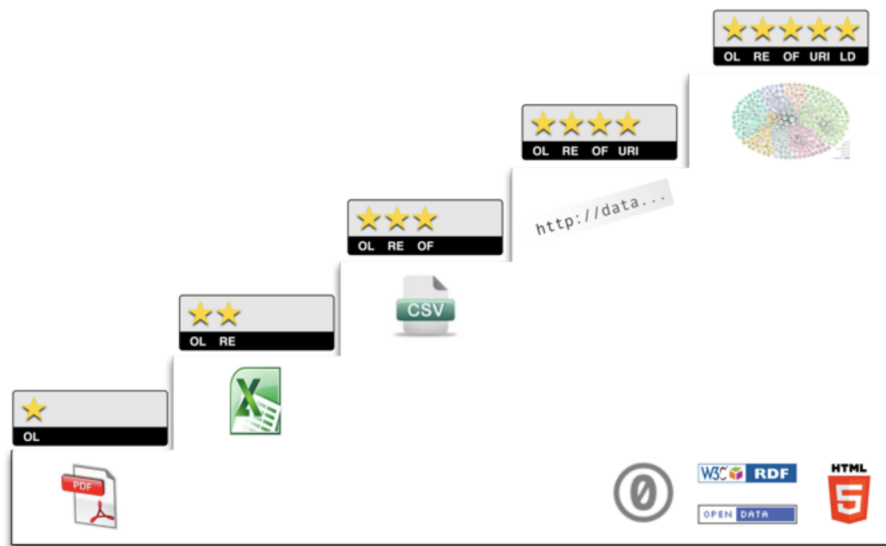
Figure 1.4: 5-star deployment scheme for Open Data. Source: https://5stardata.info/en/

Fundación Cotec para la innovación: This is a Spanish non-profit organization that promotes innovation in economic and social improvement. COTEC has recently published a study called "The reuse of open data: An opportunity for Spain" (The original title in Spanish is: *La reutilización de datos abiertos: Una oportunidad para España*). This report focuses on reusability as the core of assessment. (COTEC, 2017) conducts a diagnosis of the general knowledge of reusability based on the opinions and statistics of several leaders of open-data portals in Spain. Respondents were asked about activities to promote reuse in each portal and a SWOT analysis was performed to explore possible opportunities to improve the ongoing portals in Spain so as to gain more economic and social benefits. Although this report deals with aspects of reusability, it does not include the data users' perspective and, like the EDP reports, it is based on a

survey among leaders of open-data authorities who listed their recommendations and barriers according to their experience. Nevertheless, this is a perspective that might have significant differences with what data-user communities consider relevant to improve reusability.

Open Data Index: An initiative run by the Open Knowledge Foundation that presents an annual benchmark for the publication of open government data. This project aims to deliver insights about the current status of openness of government data released by several countries by comparing and setting a ranking by datasets and countries. Although this index offers a good picture of many datasets that can be deemed as relevant for open government projects, it does not offer information about the current status in terms of what datasets are the most used at city level. Furthermore, it does not include the data users' perspective.

From Academia: There are two relevant studies that recognize the importance of reuse, data-user requirements, or the local level in the improvement of open data.

- Janssen et al. (2012) published the first academic study highlighting the potential of open data and the need to reuse it. This paper describes the myths surrounding open government data, barriers, and implications.

- Abella et al. (2014) proposes "Meloda", a metric to assess open-data reuse. Meloda is an initial indicator to measure the level of reuse in four aspects: technical standards, accessibility, legal framework, and data model. This index, which has 100 as its highest value, establishes four levels of reusability where from 75 to 100 is advanced reuse, from 50 to 75 improved reuse, from 25 to 50 represents basic reusability, and from 0 to 25 is not suitable for reuse. Despite the attempts of the index to cover reusability in four dimensions, it tends to focus more on assessing the technical perspective of open-data portals, without considering the data users' feedback or the opinions of current users, which are in fact the final goal of any open-data strategy.

1.5 | **Thesis Structure**

This thesis is structured in six chapters. Chapter 1 has introduced the problem on which this work is centered. It describes the rationale underlying the research, the problem addressed based on the literature review, the contributions it makes, and finishes with some related projects. It also considers the current challenges of the open-data movement, in which the level of reuse is currently the center of attention for open-data initiatives. Chapter 2 presents the concepts used to portray the background underlying this research. These concepts include Open Data, Open Government Data, the current status of the level of reuse of open government data, some frameworks to measure the impact of national open-data initiatives, the role of cities, and data users but especially geographic information. Chapter 3 covers the methodology applied, and the main activities carried out in this research. Thus, it describes the online survey that was administered to evaluate the current status of local open data in cities, the selection of local data authorities, the use cases, and data-user communities. This same chapter also includes a description of how the participatory workshops on the reusability of local open data were set up and how information about data users' barriers was collected, while also offering some recommendations on how to overcome the current challenges in terms of local open data. Chapter 4 covers the structure of the proposed taxonomy of barriers by explaining the most relevant concerns and recommendations given by data users in the four use cases on how to overcome the issues that have been found. However, the description of current open-data users' barriers might have a positive impact on helping local authorities to identify where they need to pay more attention in order to engage more users. This research sought to develop a clearer path to help local data authorities to get data users properly involved in the ongoing open-data initiatives. Based on the recommendations and the most relevant barriers, Chapter 5 describes the proposed user-centered framework for improving the level of reusability of open geodata in cities. Initially, a set of 'Impact Enablers' is established to describe what factors are required in any local open data in order to make initiatives more sustainable and then it goes on to include four elements that work together to reinforce the level of reuse of published geodata in cities. Data users and their data demands, the community of reuse, user-focused metadata, and user-centered terms of use are all described in some depth in this chapter. At the end of chapter 5 there is a description of how

data users in the city of Valencia have validated the elements of the framework. Chapter 6 discusses the conclusions and limitations of this research, and suggests some avenues for future work.

2 | **Background**

Open Data is data that can be freely
used, reused and redistributed

OKF

This chapter points to concepts taken into consideration for this research. It
touches upon Open Data, Open Government data, the value of open data reuse,
as well as the relevant role of open geodata reuse and the implications to include
data users' perspective at a local level.

2.1 | **Open Data**

Open Data is a concept that has been in the spotlight for more than 10 years
and it has influence lot of public and private initiatives around the globe, most of
the following concepts in this chapter are based on what open data means. The
Open Knowledge International foundation (OKF) [1] defined *"Open data is data that
can be freely used, re-used and redistributed by anyone - subject only, at most,
to the requirement to attribute and sharealike."* The OKF has built this definition
considering the full requirements of the open definition and seeing open data as
the building blocks to develop open knowledge: *Open knowledge is what open
data becomes when it's useful, usable and used.* There are three key features
established as the core of openness according to OKF:

- Availability and accessibility: Regarding to data must be available as a whole
  preferably by downloading over the internet.

---

[1] Open Knowledge International Foundation, available online: Open Knowledge International,
accessed on 25 July 2018.

- Reuse and Distribution: Establishing data must be available in terms that permit reuse and redistribution. Including its distribution in a machine-readable format.

- Universal Participation: Considering that everyone must be able to use, reuse and redistribute, including the commercial use [2]

There are many kinds of data that could have the potential to become an open data, however data produced or guarded by government or public bodies is one of the most requested to be open. Laura James the CEO of the Open Knowledge Foundation, and Co-Founder and Director of Makespace claims *"Open can apply to information from any source and about any topic. Anyone can release their data under an open licence for free use by and benefit to the public. Although we may think mostly about government and public sector bodies releasing public information such as budgets or maps, or researchers sharing their results data and publications, any organisation can open information (corporations, universities, NGOs, startups, charities, community groups and individuals)."*

Currently there are many kinds of open information in culture, transport, environment, weather, statistic, financial, business, urban planning, science and geodata that often is considered as another category even when most of the aforementioned categories might have a spatial component. Open data could be any kind of data created by public bodies, private companies, researchers or individuals in several size from small data to big data. [3]

Another brief definition is described by Open Data Institute - ODI [4] which defined open data as *"data that anyone can access, use or share"*. The argue of this important institute in Open Data established its definition in terms of the license. ODI has developed the Data Spectrum (see Figure 2.1) illustrating that does not matter the size of data, whether the data is from government data or private data, the relevant aspect that really matters is how it is licensed. Despite the ODI considered the legal openness is the foundation stone of any open data, technical openness is the second relevant aspect and it is demanding available data in a

---

[2]   What is open?, available online: http://opendatahandbook.org/guide/en/what-is-open-data/, accessed on 25 July 2018.

[3]   What is 'open data' and why should we care?, available online: https://theodi.org/article/what-is-open-data-and-why-should-we-care/, accessed on 25 July 2018.

[4]   Open Data Institute, available online: https://theodi.org/, accessed on 25 July 2018.

machine-readable and bulk format. Thus having data in formats like PDF does data extremely difficult to work with.
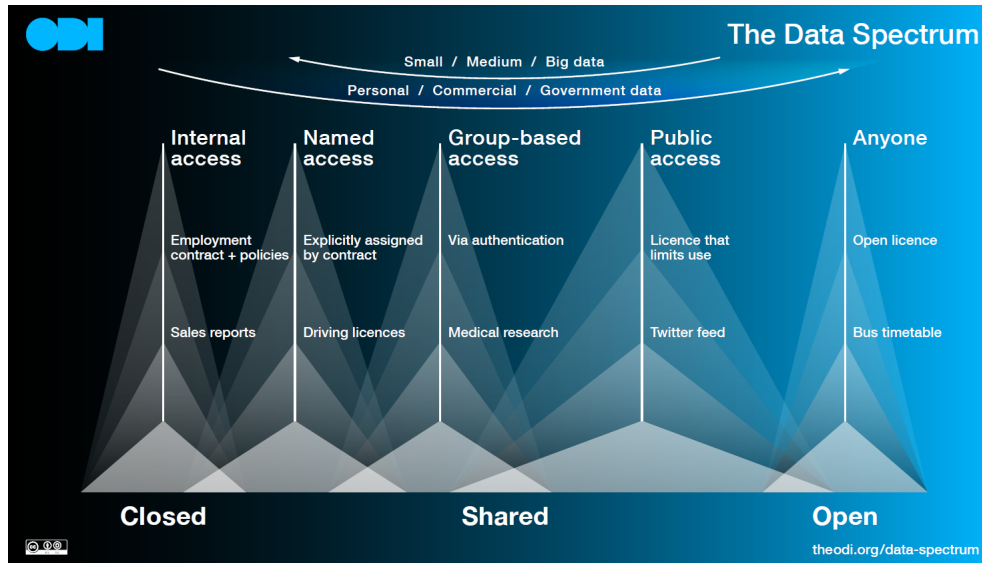
Figure 2.1: The Data Spectrum by ODI. Source: https://theodi.org/about-the-odi/the-data-spectrum/

This Data spectrum could be useful to define the boundaries required to avoid the confusion between concepts of closed, shared and open data making special emphasis in the license. However, accomplish the full requirements of the definition of open data is often challenging for many organizations who appear with the requirements of open data. (Bargh et al., 2016) proposed a new paradigm called "Semi-Open Data" to acknowledge and encourage organizations that are making important efforts towards open data but do not comply with the requirements completely and still are progressing to improve the accessibility to get some of the expected benefits.

The data has been refereed as the new oil [5], having a intrinsic value, the real potential of the open data is when is transformed being reused, being the centre piece of final applications. But why open data is valuable? Currently people has access to many technological ways to capture and process enormous among of data, finding trends, connections and lot of variables to affine the decisions making process, in ways that we have never imagined. The philosophic underpinning behind open data is, knowledge is power and having shared knowledge allow us to

---

[5] Is data the new oil?, available online: https://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/, accessed on 25 July 2018.

promote dispersal of power in a way that benefits the many, not the few. Therefore the potential benefits of make data available could contribute to improve the accountability, empowerment, impact measurement, scientific and technological advancement. The entrepreneurship, and developing a new services economic policy where transparency, public efficiency, innovation and economic value are broadly some of the most relevant expected benefits that open data might have when is combined with government data.

The McKinsey Global Institute [6] has estimated in October 2013 that seven sectors alone could generate more than \$3 trillion a year in additional value as a result of open data (see Figure 2.2). In the next subsection, we will cover the most relevant examples of estimated economic benefits of open data when is applying in open government initiatives.

Potential value in open data, \$ billion

Education	890–1,180
Transportation	720–920
Consumer products	520–1,470
Electricity	340–580
Oil and gas	240–510
Health care[1]	300–450
Consumer finance	210–280
Total	**3,220–5,390**

The values here are drawn from examples of open-data potential and are not a comprehensive sizing of potential value across the two sectors.
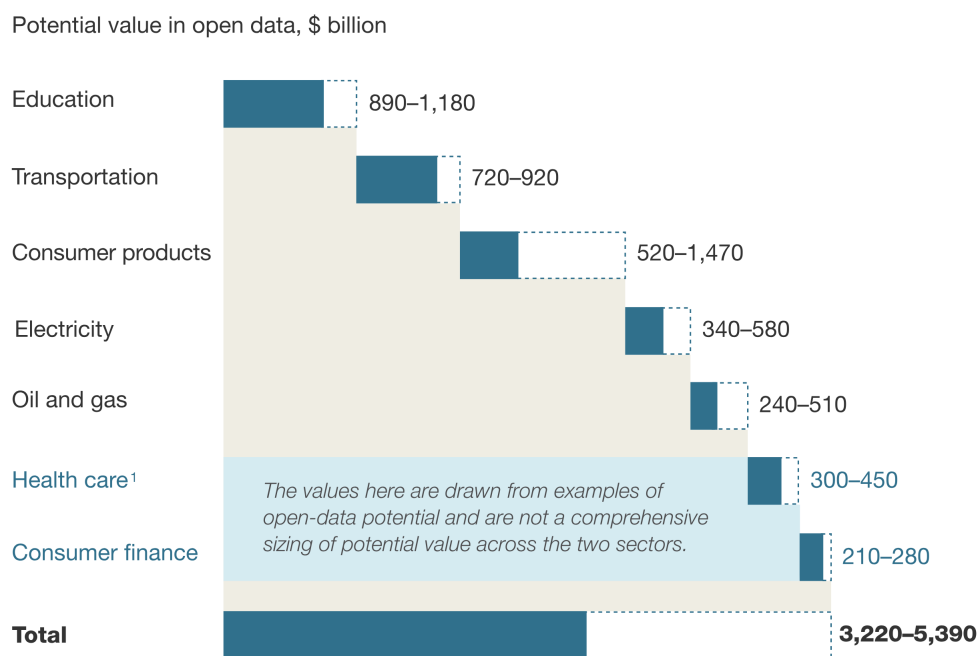
Figure 2.2: Open Data Economic Value annually across seven sectors. Source: Open data: Unlocking innovation and performance with liquid information

Open data as a contested concept, and it could be defined from several angles, using a licence to use perspective, using the economic added value. So far we have not agree a generic and single definition despite the efforts of OKF to make a clear pathway to cover the open data definition and therefore establish a map route to tackle the current challenge, especially in the awareness of open data. Recently, the GovLab [7] has lead an comprehensive analysis of competing meanings used for Young and Verhulst (2016) working in a wider definition that involved the current challenges and specificities of open data. This dissertation will considered (Young and Verhulst, 2016) definition to enclose what we mean by open data, due to clear reference to the importance of the reusability and impact of the released data. *"Open Data is publicly available that can be universally and readily accessed, used and redistributed free of charge. It is structured for usability and computability."*

---

[7] GovLab, available online: http://www.thegovlab.org, accessed on 25 July 2018.

2.2 | Open Government Data

Open Government Data (OGD) as one of the most influential forms of open data, establish its importance through the potential benefits that public authorities, citizens, and private companies might have from the promotion of available public sector information without technical and legal restrictions. Making government data available to the public in easy and massive ways to read and process fostering the transparency, accountability, and citizen participation. Several national governments have deployed open data portals to make accessible data that is collected or processed. The general idea of OGD is to develop methodologies and tools that allow a cultural shift around the participation, and transparency of open governments.

To start, it is important to make a clear definition of the Open Government (OG) concept. Due to the enormous attention that OG concept has attracted, authors from different areas have defined under their perspective or experience the definition of OGD leading a very heterogeneous comprehension of this concept. In this dissertation we will use the definition proposed by Wir (2015) deducing an integrative definition based on a systemic review of several papers about OG. Wir (2015) defines OG open government as *"a multilateral, political and social process, which includes in particular transparent, collaborative and participatory action by government and administration, To meet these conditions, citizens and social groups should be integrated into political processes with the support of modern information and communication technologies, which together should improve the effectiveness and efficiency of governmental and administrative action."* Having also external drivers like accountability, regulations, technology and the trust in government. Figure 2.3 illustrate the components of the OG framework where the Government to Citizens (B2C) and Government to Business (B2B) relationship are also included.

Once we have a clear definition of OG, the next step is to establish the definition of Open Government Data (OGD). Kučera et al. (2013) claimed OGD as government-related data that is created and published in the way that meets with the Open Definition[8]. OGD is seen as a current trend and a key factor in cities with intersection with Open Data initiatives. Ubaldi (2013) defined OGD as the

---

[8] The Open Definition, available online: http://opendefinition.org/, accessed on 29 July 2018.

combination of government data (defined above) and Open Data (see sub-section 2.1).

Despite, transparency, collaboration, and participation are defined as the most desired benefits from OGD implementations, is the relationship between the business and OGD another perspective that arise the attention of advocated in open data and indeed determine the further steps of the ongoing open government initiatives. According to Open Government Partnership (OGP) economic gains will happen in three scenarios:

- Business Innovation: Just in the European Union, the implementation of OGD will return €87 billion annual investment in Research and Development, due to the rapid access to data and scientific papers making easier for business and research.

- Business Creation: New service-based economy is under way once the OGD is available in national, regional and local levels, promoting the creation of innovating applications which the cost to get the data is minimal.

- Business Efficiency: Getting a better understanding about users, citizens, possible issue or challenge could contribute to perform better processes fitting to the real needs of final users. Improving the way that authorities deliver their outcomes and making easy for citizens the interaction with the government.

The European Data Portal (EDP) [9] has also estimated the economic value of OGD calculating direct and indirect benefits of the implementation of open data initiative. In terms of direct benefits EDP estimated 25.000 jobs created, €1.7 billions of government cost savings, in a possible direct market in 2020 of 325 billion for EU28+ countries. Regarding the indirect benefits, EDP established 16% less energy used, more than 6000 lives saved due to quick response and 5.5% less road fatalities, having more open data to make better decisions (Carrara et al., 2015). But not only the European Commission has estimated it economic benefits. Australia has conducted an study of the aggregate economic impacts of spatial data suggesting that open spatial data can increase productivity by billions of Australian

---

[9] European Data Portal, available online: https://www.europeandataportal.eu, accessed on 29 July 2018.

dollars. For instance spatial information industry revenue in 2006-07 could have added cumulative gain of AUD 6.43-12.57 billion. Denmark, has estimated the potential national market for energy improvements drawing on various government data sources is €0.54-2.7 billion. Spain, through a government-commissioned study found that open data released by national government was €550-650 million and more than 5,500 employees were assigned to re-using information.

Meanwhile, in order to understand the different meanings of OGD from bureaucratic, political, technological, and economic perspectives, Gonzalez-Zapata and Heeks (2015) used the definition of Yu and Robinson (2012) through three main foundations—open, government, and data, illustrating three intersection points to determine what OGD means. Going in the same direction Carrara et al. (2015) claims open government data as the overlaying among public sector information, big data and open data initiative, delimiting the boundaries of open government data (see Figure 1.1). The result of this combination is government data, open data, and Open Government (defined above).
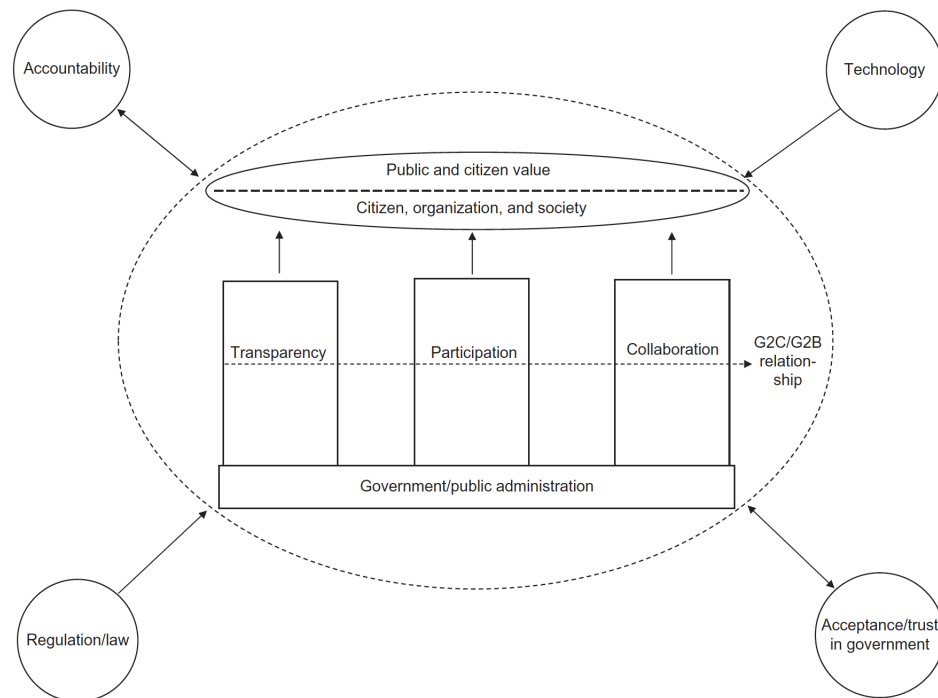
Figure 2.3: Open Government Framework proposed by Wir (2015). Source: Wir (2015)

In general, local and national administrations, civil society organizations, the private sector, and overall several stakeholders are taking advantage of the intersection among open and government data. The impact of this combination can be positive in many ways. For instance, the continuous online access to government data is positively associated with knowledge absorption according to Lee et al. (2016), who indicated that government data openness could positively affect the formation of knowledge bases in a country; therefore, the level of knowledge base even positively affects the global competitiveness.

Furthermore, not only the level of knowledge has been identified. When Open Government initiatives are on the table of public agencies, the expectations to improve the governance processes are certainly high. Moreover, increasing transparency, expanding the public engagement, and improving responsiveness and accountability are the desired goals of most governments. Determining whether the Open Government initiative is effective or successful could be a challenge for many public agencies. The ambitious aims and expectations of these sorts of initiatives could lead to some failed activities that yield some immediate success but then run the risk of losing steam over time. Identifying the participant, their roles, and including them in the current Open Government initiatives and the way that data is released is illustrated by Williamson and Eisen (2016) as the key to successful Open Government initiatives. Through a rubric of six questions, Williamson states that even high accessibility levels and well-publicized data are not enough to transform the government processes if people or participants do not have channels to influence it.

Due to the major role of participants, a better understanding of citizens' motivators for engaging in Open Governments actions could guide the current initiatives to get the expected outcomes. Wijnhoven et al. (2015) demonstrated that when citizens feel that their contribution is significant, they are more open for contributing in Open Government projects. Besides, Wijnhoven et al. (2015) found that there is no evidence to suggest that socio-economic factors could affect the participation in those Open Government projects, whereby projects that appear to be well-implemented have a better reaction from citizens than others that only focus their attention on some stakeholders. In the same direction, a well-detailed description and the way in which data is shared could have a positive impact on the participation of citizens. Afful-Dadzie and Afful-Dadzie (2017) collected the preferences of media practitioners in five countries in Africa, and observed that

online journalists see metadata as the most important factor in a functioning OGD, followed by data format and data quality.

In particular, there are some factors that can influence the level of success of the implementation of Open Government initiatives, especially for authorities that require a solution beyond the current administration. Wang and Lo (2016) looked into some of those factors in Taiwan; their study disclosed that perceived benefits, organizational readiness, and external pressures have a positive effect on OGD adoption. On the contrary, perceived barriers seemed not to have any significant effect on OGD initiatives adoption.

Overall, the participation and interaction with the general public, the identification of their needs, and the sustainability of OGD actions are a particular focus of the current studies. Beyond accessibility, the scope is now moving to determine factors and barriers to using or reusing the available public sector data (e.g., crime rates, gas emission, mobility, air quality, or security). Since many available datasets and likewise many cities are in the middle of the implementation stage of open data government initiatives, the aim of authorities is to motivate users to reuse the published data and create a new set of services, generating value for data-opening projects. Local experiences of cities, local needs, and kinds of data could bring a differential factor in the way to reuse the available OGD. In the following sections, we will discuss why the reuse of open geographic data is becoming necessary in current open data initiatives and why barriers from a data users' perspective are preventing full advantage being taken of the data release process in which many cities are involved.

2.3 | **Reuse of OGD**

Much published work is related to open data and desired benefits that this trend might bring to governments and its stakeholders. There are several authors Barry and Bannister (2014); Conradie and Choenni (2014); Martin and Foulonneau (2013); Janssen et al. (2012); Beno et al. (2017); Cranefield et al. (2014); Europe and Foundation (2011) who have worked on open data barriers from different perspectives; most of the work done has been focused on national governments, OGD initiatives applied for data producers, integrators, or suppliers. Beyond promoting a sustainable reuse of Open Government data in cities, a constant and circular reuse should be considered in OGD initiatives.

In Janssen et al. (2012), a set of benefits, adoption barriers, and five myths of open data initiatives are defined; most of them are still present in current initiatives. For instance, myth number five is about open data and the incorrect interpretation that will result as Open Government. In Section 2.4 it was explained that releasing open data is only the first stage in getting the expected benefits of Open Government, especially collaboration and participation. The process can only start when the published data is used. Janssen also suggested a set of adoption barriers from a national government perspective; however, at the local level, barriers, data user communities, and even the expected benefits may vary. Likely, national and local levels are both pursuing the improvement of accessibility, legal issues, and technical integration concerns, but the contact with data users could be easier at the city level. Factors such as reuse, feedback, channels to influence, and integration requirements create a solid way to work towards the benefits mentioned.

Barry and Bannister (2014) also worked on the definition of open data barriers when the data is published from a data integrator perspective. They took Ireland as a use case, creating a detailed comparison among the current literature about open data barriers, and proposed a new barrier schema as a taxonomy of release barriers from senior managers in this country.

Yang et al. (2015) suggested factors that could reduce the possible impact of published data, using several authorities in Taiwan as a use case. Thus, their work presents those factors as barriers from data producer perspective and at the national level. Another related paper about barriers—but at a local level—was published by Conradie and Choenni (2014). They found that the ways in which

data is stored, obtained, and used by local departments are crucial indicators of open data release. Conradie and Choenni (2014) suggests take incremental steps to explore and learn about the data release, avoiding releasing data for political or internal purposes.

In the literature, there are also some reports created by the European Commission and its project EDP; the initial and related report taken into account is the reuse of open data Carrara et al. (2017b) from a business perspective. This report presents a study of several companies—most of them from the private sector—around their business model built using open data. It lists a set of factors that European countries or corporations should consider to promote the reuse of open data. Internal and external barriers that do not allow the standardization and automatization of open data are defined, and at the same time, some recommendations for the public and private sectors are illustrated.

Another report related to the last one by the European Commission is the fifth analytic report of the EDP Carrara et al. (2017a), where barriers are seen as a core of the problem to reusing open data, basically from two perspectives: data producers or suppliers and data consumers. However, this report is based on the same findings as the reuse and maturity level report Carrara et al. (2016) that the European Commission studied as well. The description of the barriers are listed according to the national level in the EU28+ countries of Europe and their open data initiatives.

Mentioned above, current literature has studied the traditional way to describe the open data initiatives, considering data producers as the key stone of all the strategy and taking users as the final result but including them as part of the process. This way of thinking, therefore has been lead several studies where data users perspectives is limited or mostly ignored. Our work is an attempt to demonstrate the important role of data users' needs, and their point of view to contribute in the design of improved and more inclusive open data strategies in which the reusability should the final goal.

2.4 | Added Value, Impact, and Reusability of Open Data

Beyond access to OGD, the creation of value is perhaps the most interesting part of open data systems, in which economic, social, and political benefits are being established in local governments. Official entities are trying to increase the transparency of their processes and empower their citizens by publishing a vast list of relevant data. Ubaldi (2013) provide a work about OGD, in which a list of commonly recognized main beneficiaries of OGD can be found, where the wider economy, the private sector, and the public service marketplace provide the opportunity to increase the innovation expected by official authorities. Access to data by itself does not offer new services or make a difference with other private data provider companies, per se. New value-added services must come in addition to data to bring more opportunities to developer companies to pursue the commercial exploration of OGD. This commercial approach and a new set of value-added services are possible when data is reused Carrara et al. (2017b).

Assuming that greater openness automatically creates value Ubaldi (2013), is a common mistake in many governments. The OGD systems should include the value chain as part of the initiatives, where conditions to develop value-added services and indicators to measure the impact of released data are included as a relevant part of the systems. The data value chain is the mechanism to move towards from data production to impact generation., European Data Portal in its report (Carrara et al., 2017c) about the reuse of open data described the workflow from data creation to aggregated services. However, this value chain established only one way where the data is produced and is offering its benefits. There is no a circular system in which data users have a important role to contribute. Through data creation, aggregation, analysis and ending up with the generation of aggregated services, this value chain represent a typical top-down approach. This scope consider data users only as final users in the end of the chain, showing the preference for the data producer point of view (see Figure 2.4) but leaving the data users' perspective limited or with not possibilities to influence the system.
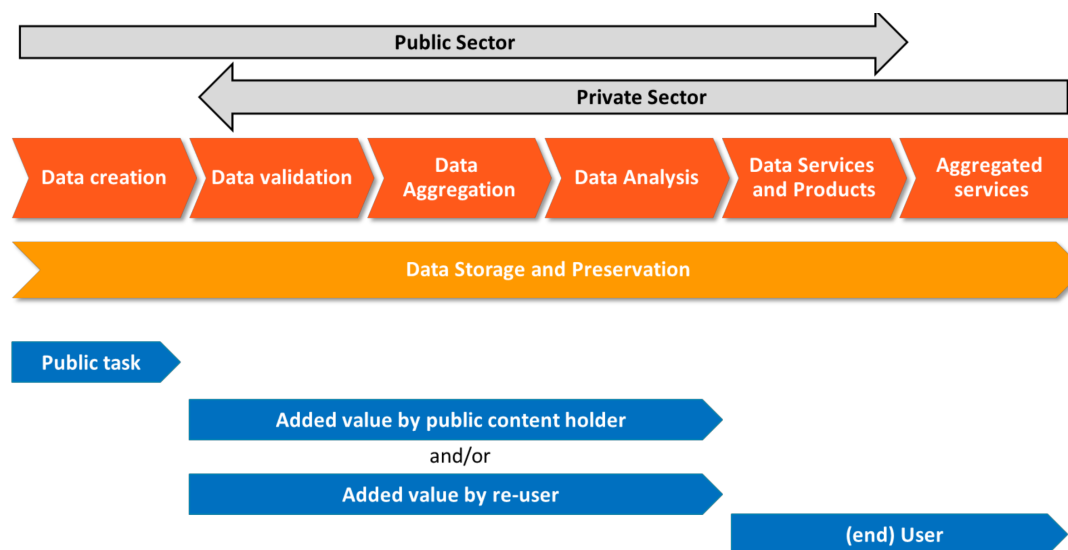
Figure 2.4: Open Data Value Chain. Source: Carrara et al. (2017c)

Open Data Watch [10] a non-profit, non-governmental organization dedicated to monitors progress and provides assistance to guide implementations of open data systems around the globe, proposed an more accuracy value chain of open data (see Figure 2.5) presenting a workflow to add value across the open data stakeholders.

The inclusion of the impact, and uptake stages allow the use and reuse having the important role along the whole process. Besides, the constant feedback is also another component that Open Data Watch included and is indeed another example about the requirement of the constant feedback and the reuse and the final goal of any open data strategy.

---

[10] Open Data Watch, available online: https://opendatawatch.com/, accessed on 29 July 2018.
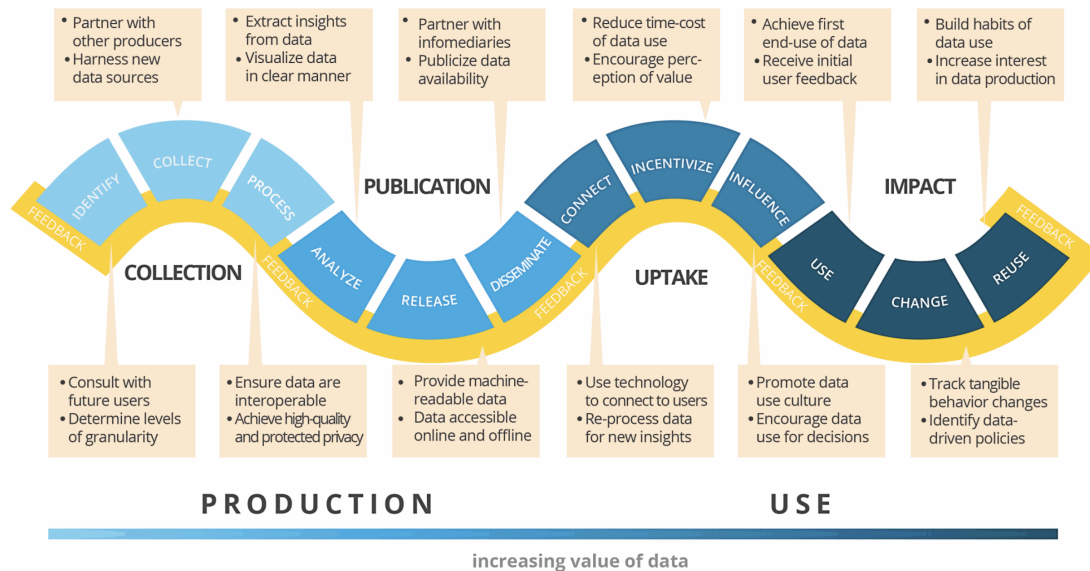
Figure 2.5: Open Data Value Chain. Source: https://opendatawatch.com/reference/the-data-value-chain-moving-from-production-to-impact/

The way of the aforementioned value chain is applied also has a consequence in the effectiveness of the open data strategy. The implementation at national level is different than the implementation at city level, the stakeholders, roadblocks, results and most important the data users communities would have differences. Therefore, local governments overall have an essential role to play in the value chain. Connecting published data with data user communities, local authorities are not only playing the role of providers; they also become a partner, facilitator, convener, and enabler of easy reuse. At the same time, empowering data user communities tackling local issues and deal with reuse barriers on a daily basis, local level could be the key to transforming the current actions into concrete results. This integrated scenario is only possible when data authorities behind open data initiatives incorporate the reusability as part of their priorities. The transformation in concrete and sustainable results is, therefore, the generation of the positive impacts of open data initiatives across cities. Thus, deployment of data portals, events, or workshops around them will not generate the expected impact, most likely it is just the first half of the whole process where the access is a guarantee, but in an establishment of impact, the accessibility needs to end up in reusability.

Although OGD is a common topic in local governments, most of them have not understood the benefits and value of open data, but mainly the expected

benefits for cities' stakeholders. Yang et al. (2015) illustrated that data authorities should not only consider data users as the general public, also they should include in their open data scope, data users inside internal departments or other related agencies, where discoverability and accessibility of published data is also a relevant challenge. (Worthy, 2015) evidenced some political disruption when the local government in UK released the spending data, instead to drive some accountability. The evidences from (Worthy, 2015) suggested that assessments of the reuse it is more complex, unpredictable and more political than advocates of open data indicate. Not accomplish in mid-term the expected benefits of open data could bring some danger in the ongoing initiatives, leading some disappointment from supporters.

A better understanding of the impact of open data brings about a set of challenges. It is difficult to create a generic schema that allows open data stakeholders to assess progress or to determine where they need to put in more effort. The impact has been analyzed under some hypothetical scenarios, based on specific experiences, sporadic use cases or have been implemented during one administration. Comprehension over how the open data is impacting on people's lives is studied by (Young and Verhulst, 2016) using 19 use cases around the world and developing the open data taxonomy impact (see Figure 2.6). The taxonomy of impact could help to understand which dimensions the open data has been effective and to determine what conditions are required to enable the impact.
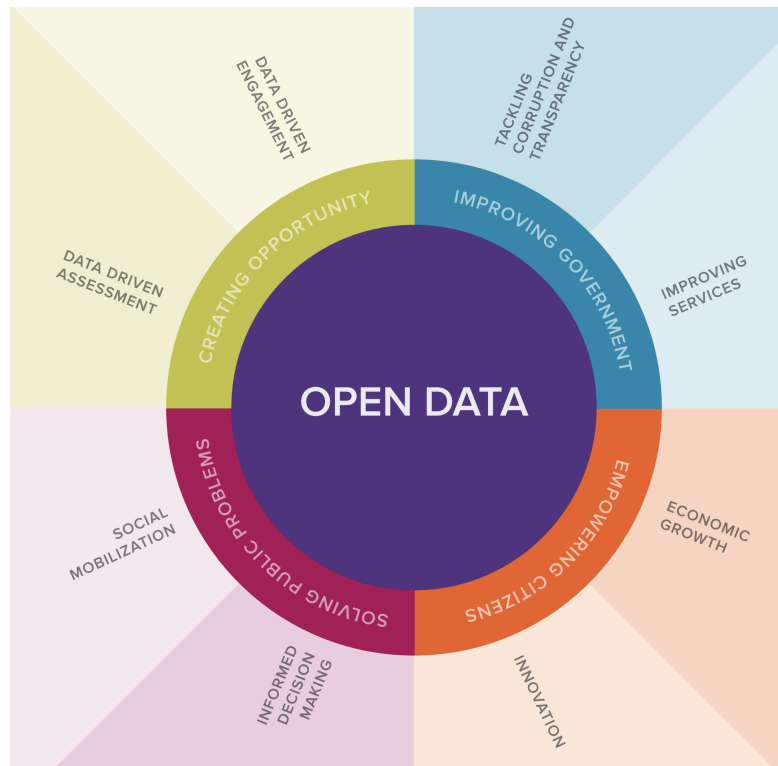
Figure 2.6: Taxonomy of Open Data Impact . Source: Young and Verhulst (2016)

According to Young and Verhulst (2016) there are four ways in which open data is providing influence and developing impact in people's lives. (1) Improving government. (2) Empowering Citizens. (3) New economic opportunities for citizens and organizations. (4) Solve several big public problems. Despite, there are documented several use cases in the four directions mentioned above, in government the impact is presented in different nuances, for instance OGD initiatives offers an opportunity for local agencies to carefully survey and identify what datasets they have, which are the most used ones, and what they can share with other departments to improve internal collaboration. Thus, local governments need to educate and empower not only the general public; the first step should be to promote the open data initiatives inside their departments. Yang et al. (2015) presented the concept labeled interagency as a foundation of OGD which is the positive impact on cross-boundary information sharing among cities' agencies, where the continuous information sharing is a spiral process to reinforce the communication and at the same time reinforce the OGD initiatives.

The Open Data Barometer Org (2015) considers impact at three levels: polit-

ical, economic, and social. However, deeper insights are needed to understand the multiple impacts of open data Huijboom and den Broek (2011) including its effectiveness and reuse in order to establish successful policies. Meng (2014) investigated the social impacts of open data and analyzed how marginalized groups access data. Gurstein (2011) criticized the popular claim that "Open data is for everybody" and noted that guaranteeing open data access does not equate to efficient use as accessibility frequently focuses on specific educated groups with access to technologies. Zuiderwijk and Janssen (2014) proposed a framework to compare open data policies. The authors claimed that collaboration among authorities and organizations is the key for real and continued progress. The democratic impacts for the the UK coalition government's transparency agenda was studied by Worthy (2015). The current European Data Portal [11] linked all of the EU+28 open data portals and analyzed the economic impacts and how businesses were re-using open data to get economic benefits Carrara et al. (2017a).

2.5 | Geographic Approach

Considering an open data value chain where data user needs are fundamental and cities are a relevant piece of the OGD initiatives (see part B of Figure 1.2), the kind of data also has a role to play. The foundations of OGD developed by Yu and Robinson (2012) (see Section 2.2), data is also a large concept that could be considered from a specific point of view to identify data user requirements. The nature of data can influence future barriers, needs, and strategies of OGD initiatives. Based on the concept of Yu and Robinson (2012)—the intersection of data seen as geographic data (data with a spatial or geographic component), government initiatives, and the definition of open—could be more efficient and interesting in light of city data user requirements. However, is geographic data a relevant type of data that might bring more effective benefits in local government, and why the reuse of this kind of data support authorities in their engagement strategy?, Does the geographic approach offer any advantage in the design of open data strategies, differently that current approaches has not considered yet?

For geographic data and its relationship with local open data projects, the literature is still limited. Although some authors have mentioned the importance

---

[11] European Data Portal, available online https://www.europeandataportal.eu/, accessed 28 January, 2018

of the spatial data in open data projects or use cases, formal consideration of geographic data provision and access is still only thought of as an extension. In the work mentioned above Young and Verhulst (2016) three of 19 use cases that are geographically focused, where the impacts include economic growth, the creation of opportunities and improved services. Johnson et al. (2017) claimed that "spatial is special" and that the consideration of contributions from GIScience to open data research could provide insights into the assumptions and issues of open data projects. Some of the listed challenges are the most demanding stages faced by local SDIs as standardization, integration, local policies, technical integration obstacles and institutional contributions. Harvey and Tulloch (2006) has mentioned that availability of local governments data is the foundation of the current local SDIs.

According to the *Reusing Open Data* report of the European Data Portal (EDP) (Carrara et al., 2017b), geographic data (25.8%) is the second category only surpassed by the statistical (27%) category that is most reused and also consulted by companies (among 128 domains mentioned) that try to generate revenue from open data reuse in EU member states. This report Carrara et al. (2017b) also illustrates a strong correlation among open data categories, where "*region & cities*", "*transport*", "*environment*", and "*population & society*" suggest a trend of organizations using those categories together.

Geographic data is considered to be one of the most economically relevant data domains for its high demand from re-users across the EU, according to the analytic report of EDP Carrara et al. (2017a). In Young and Verhulst (2016) work three use cases analized involved the geospatial sector, with public authorities in Denmark, Great Britain, United States. Additionally, more use cases geographically related in Singapore, and Uruguay where the impact is assessed in terms of improving services, economic growth, and data-driven engagement. Geographic data has specific characteristics that also demand specific needs from data users; therefore, the identification of those requirements might contribute to improving the OGD initiatives in cities.

The previous cited report Carrara et al. (2017a) showed an insight that is relevant to this research—the role of the geospatial domain in the open data movement. Presenting some barriers from a data producer perspective, the authors discuss why geospatial data plays a major role in an open data strategy for any country.

2.6 | Barriers and the Importance of Data Users' Perspective

Many of the identified barriers to improving the reuse level of OGD has been already tackled from the geographical community several years ago before the open data movement has started being recognized by public administrations and research field. In cities but especially in countries issues like standardization, accessibility or integration of several data sources has been a constant headache for many geographic institutions. Since 1993, the term SDI was coined by the U.S. National Research Council to define a framework of technologies, policies, and institutional arrangements working together to facilitate the creation, sharing, and use of geospatial data and related information resources across an information-sharing community Esri (2010); Douglas D. Nebert (2014). Such a framework can be implemented at local, national, regional or even international levels to allow different stakeholders have the effective and easy access to official, high data quality, and standard geographic information. Taking into account the important role of cities, Harvey and Tulloch (2006) presented a typology of local-government data sharing arrangements in the US in times where the local SDI was moving to a second generation. Authors Harvey and Tulloch (2006) suggested that political, institutional and economic factors need be considered in local governments to guarantee the effectiveness of the sharing-data process and likewise a continuous reuse of geodata in cites.

Janssen et al. (2012) suggested the creation of open data infrastructure as a possible way to guarantee a constant support around all political, institutional and even technical issues that are involved in the sharing data process. At the same time, current local SDI projects have a significant challenge regarding the way that geographic data user communities are using and re-using the available data, leading both projects with a common problem, which could be tacked working together. Both in local open data initiatives and local SDI, the role of data user is fundamental, a better understanding of their needs or requirements could be the key factor to refine the current initiatives and find the way to be more effective.

Additionally, there are few authors considered a data user's perspective at the local level. Janssen and Zuiderwijk (2014) described an infomediary business model that connects data users with providers. They noted the importance of

identifying the roles of data users to adequately understand their needs, bridging between provider services and user needs. The identification of end users is also mentioned by Williamson and Eisen (2016) as the first step in structures for successful open government initiatives. In Zuiderwijk et al. (2014) work a particular emphasis on the components of the open data ecosystem where users' pathways reveal the direction of how open data can be used, then the initiative can use this direction to move towards data user requirements. Based on the work's conclusions, three aspects are especially salient: (1) More and clear information related to license or terms of use. Data re-users get confused more often than data producers think (despite the fact that the terms of use are included in most of the open data portals). Additionally, most of them are difficult to read; (2) More statistical and geographic context.

This means that raw data is important and is considered to be a requirement to consider published data as open Sunlight Foundation (2010); however, it is necessary to include statistical and spatial relationships to guide users to understand what this data is about. Including comparison with other regions, or neighbourhoods with different geographic features, a comparison during the time or even with the inclusion of basic statistics, published data can reach more users' attention offering an enriched perspective, than only a list of downloads. Finally, (3) Feedback for both data providers and data users; providing ways to discuss, both sides can learn and enhance the value of available data. This component is likely one of the most forgotten resource in current open data initiatives, where the feedback resource is limited to email contact or a questionnaire to end-users to express some issue. Only few open data portals have a proper systematic way to discuss issues, use cases, best practices or suggestions from end-users, and also show updated data or features to their community.

2.7 | **Summary**

The number and variety of Open Data initiatives have increased globally. Barriers, key features, implementation challenges, use and impact cases have been the focus of several studies. Peled (2013) analyzed the current issues of the first version of open data in USA after the Obama administration's Open Government program was released. They proposed a new label *Open Data 2.0* to include released data strategies. They note the potential for transparency policy bubbles, that is situations where policies are only implemented for short periods of time. Sieber and Johnson (2015) analyzed the shifting role of the government in the open data "culture" and presented four models in which citizens, private sector, and authorities make changes to determine the effectiveness of future open data strategies.

The potential of open data relates not only to the government or local authority, but also to stakeholders with different roles, interests and contributions. A critical consideration is how added value is created through the data supply and analysis chain Carrara et al. (2017a). Kassen (2013) explored the empowering potential of open data at local levels as a platform for citizen engagement, based on a use case in Chicago. Kassen noted that bottom-top schemas at a local level can provide an effective platform to collect and understand the citizen requirements. The role of the local government in the open data movement is also increasingly recognized. For example, the European Committee of the Regions noted that *open data has the potential to become valuable assets for citizens, businesses, and public authorities* (The European Data Portal, 2016).

In general, there is a great deal of published work regarding the reuse of open data and why it is one of the challenges for current initiatives. Barry and Bannister (2014) have selected the occurrences of some themes surrounding open data, mentioning that data sharing and reuse are two themes with a high number of occurrences, demonstrating a focus on making the most of the resource of public sector information. Literature has analyzed economics Johnson et al. (2017); Jetzek et al. (2012); Ahmadi Zeleti et al. (2016), technical, institutional Yang et al. (2015); Cranefield et al. (2014), political and policy factors Nug (2015); Jetzek et al. (2013) that influence the value chain of open data, suggesting that theoretical benefits have not been seen as cities' stakeholders expected Carrara et al. (2017b); Janssen et al. (2012); Conradie and Choenni (2014); Barry and Bannister (2014);

Attard et al. (2015); Cranefield et al. (2014).

Regarding the intersection of initiatives between national and local efforts, both working to improve their Open Government's efforts, in many cases datasets are offered on several websites Yang et al. (2015) in a fragmented way, which is in some cases difficult to find. Adequate metadata are also necessary to improve data reuse Janssen et al. (2012). The cooperation among city, municipality and local government departments to promote and manage data sharing process is a relevant daily challenge for many local public bodies. Creating a framework for collaboration could be one of the ways where a local SDI can contribute to local open data projects, due to its experience developing a legal, technical and institutional agreements. Notwithstanding, local SDIs also have many barriers to overcome regarding sharing data. Carrara et al. (2017a) identified the geographic data features as technical barriers to the re-usability of open data initiatives. The differences between, standards, formats, size of files, data quality standards, stemming from Directives like INSPIRE [12], create an additional effort for open data initiatives to implements the geographic data domain.

To summarize, previous work—using interviews, surveys, workshops, or sets of references—has identified a set of barriers mostly from a data producer point of view, where national authorities are having the main role of open data initiatives. Regarding the reuse of open data, there is not too much work done; we found only four related references, none of them have considered the possible potential of geographic data, or the role that cities can play. The number of articles that examine reuse obstacles from a data user's perspective is also limited.

---

[12] INSPIRE , available online, https://inspire.ec.europa.eu/ accessed 5 February, 2018

3 | Methodology

> Open data on its own has little intrinsic value; the value is created by its use.

> *Marijn Janssen*

This chapter is devoted to cover the methodology used during this research. The applied method used to get the findings presented in this dissertation is a set of five stages, that includes (1) Literature review. (2) The current status of the reuse of open data through a online survey. (3) Set of participatory workshops to validate the preliminary findings. (4) Definitions of the conceptual framework and its elements. (5) Finally the validation of the conceptual framework's elements and discussion with Valencia data users was the last stage presented this dissertation.

3.1 | Overview

Many authors Zuiderwijk et al. (2014); Ubaldi (2013); Barry and Bannister (2014); Europe and Foundation (2011); Open Data Institute (2015); Young and Verhulst (2016); Benitez-Paez et al. (2017) have mentioned that the potential value of open data is in its use. Improving the reusability level of open data at local level (city level) might be a critical factor to truly make an positive impact through city stakeholders. This method is an attempt to empirically demonstrate the potential of the reusability and the need to define a bottom-up in the current open data strategies moving forward to sustainable actions that allow a more useful and trusted reuse of open data.

This research took place from August 2016 until February 2018 based on multiple use cases and a combination of structured online survey, hands-on

activities (participatory workshops) and a comprehensive literature review. The established method was defined to addressed the proposed research questions in section 1.2. During this research in an open data environment, two components were considered as the key stones, data consumers' requirements and city level, both were covered from three angles: **(1)** A literature review, **(2)** Structured online survey (what people say regarding the reuse of open geodata), **(3)** Outcomes from a set of participatory workshops (what people do, discussion with several city stakeholders regarding, barriers and actions to overcome those barriers), to later define a conceptual framework defining the elements that users considered relevant to make more reusable the released data and consequently validate the proposed framework with data users to get the required feedback. Considering that our assumptions were:

- Despite, the several studies that have documented the potential of open data in terms of economic and social terms, the little knowledge about how open data actually works, who is really using the ongoing initiates. We assumed that reusability level of many of local open data initiatives (i.e. city level) is considerably low.

- Open data at the local level might bring more effective impact that at the national level. Based on our preliminary findings, in the literature review and what we found around some open data events, the number of local open data initiatives are getting more attention and relevance that national open data initiatives.

- The keystone that is missing in current open data strategies is the consideration of data user requirements. Re-shaping the ongoing open data strategies towards a bottom-up approach will lead to a new set of activities and initiatives allowing data authorities improve the level of use and reuse of current open data projects.

- The geographic approach, seen in this dissertation as the spatial data in-
cluded in the published data included in local data portals. Due to the
potential to increase the impact of current open data initiatives, spatial data
is one of the most requested kind of data by users communities, business
sector and academia. Including the integration and standardization experi-
ence of local spatial data infrastructure might contribute to emerging open
data initiatives.

Figure 3.1 displays an overview of the stages that this research has done. **Stage
(1): A systematic literature review** was conducted, initially to understand the
state-of-the-art reusability of open data on several levels. The topics considered
in this review were, the impact of open data, the current status of reuse of open
geodata, implementation levels of open data projects. Learned lessons, implica-
tions, improvement frameworks and use cases around the word related to open
data especially OGD. Furthermore, we also wanted to determine what frictions
or barriers do not allow data users to reuse the published data and the level of
reuse of open geodata in several cities. **Stage (2): Online survey**, once the initial
barriers were identified, and considering that most of them were validated not con-
sidering the data users perspective at local level or the spatial scope. The second
stage was about a comprehensive online survey to get a better understanding of
what people say in term of reuse, barriers, and recommendations to overcome the
identified barriers. **Stage (3): Set of participatory workshops**, once we have
validated the barriers and collected all the initial opinions of our respondents (195
valid responses), four cities and their local data authorities where contacted. With
the support of open data leaders in each city, we led a set of discussions with data
users communities and enthusiasms (over 115 data users across the four selected
cities)., in which the reusability, barriers to reuse and reasons about why local
users are not currently using the published data. Moreover, the challenges in each
open data initiative and recommendations from data users' experience to improve
the level of reuse were also gathered. **Stage (4): Framework design** Taking into
account the previous stages and the gathered data, we defined four elements in
an user-center framework for data authorities could use as a guideline to re-shape
their ongoing open data initiatives. The conceptual framework elements are above
three 'impact enablers' which were selected based on the results of the initial liter-
ature review looking for the resources that could contribute for a more sustainable
strategy that make a differential factor of the current ones. **Stage (5):** Finally in

the last stage the validation of those four elements and three impact enablers as the keystones of the framework were validated and discussed with data users in Valencia.
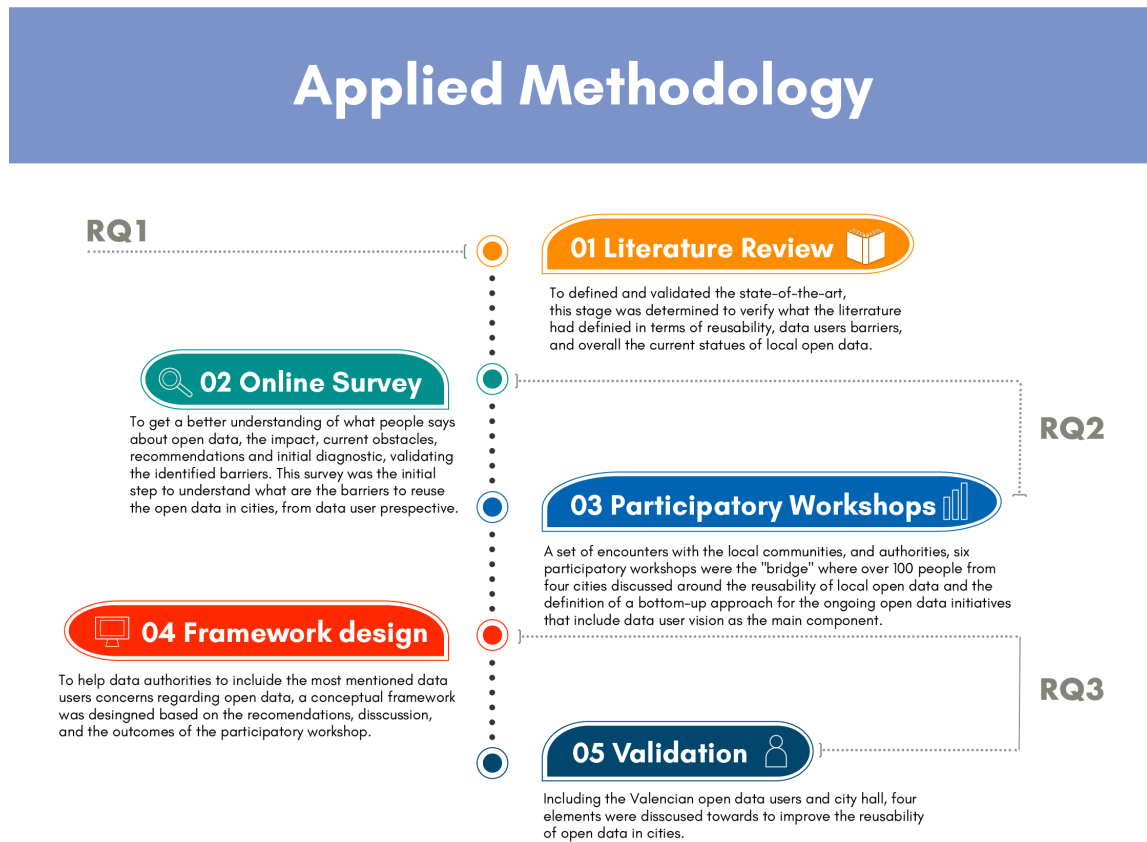
Figure 3.1: Workflow used. Literature review, identifying initial barriers and perceptions, then a publicly shared online survey (n = 195 valid responses). Using the preliminary results, four cities were selected and the identified barriers were contrasted with the participatory workshop in each city.

As a result of stages, 1, 2, and three a data users' barriers taxonomy (see Chapter 4) was defined and published in Benitez-Paez et al. (2017). This taxonomy presents six obstacles' categories to the reuse of open geographic data in cities (e.g., local level). It can be used to inform data authorities about weaknesses of the current city open data systems, thereby enabling them to design better and more effective strategies to improve the reuse of their data.

Along with this research, four use cases were considered, with local authorities in the three principal cities in Colombia (Bogotá, Medellín, Cali) and the third

central city of Spain (València). Initially, this research studied the current status of their open data initiatives, considering that cities have different Open Government data approaches Beno et al. (2017); Williamson and Eisen (2016). To enriching the discussion and reduce a possible bias of the findings considering only one city, the selected cities have distinct progress and perspectives from legal, technical, institutional, political, and awareness points of view (see Table 3.4). Beno et al. (2017) worked in the delimitation of barriers to use open data in Austria at a national level, and claimed that "*caution must be applied as the findings might not be transferable to other countries*", because there may be differences in terms of maturity of their open data "*culture*" and the datasets that official authorities offer. The available datasets in each city have an essential role to understand possible frictions to use or reuse the data in each city.

All selected cities have their own data portal. Valencia and Medellín have a central portal called "Transparency and open data portal" [1] and OpenData Alcaldía de Medellín" [2] respectively, with considerable number of web services, mostly are geo-services, related to several city domains, such as mobility, education, environmental, urban planning, demographic and culture. In general, topics that each local authority considered relevant to users and the city.

Another aspect that also contributes to the diversity of the selected cities in the current role of the local authorities contacted regarding open data "*culture*" in each city. Initially, both Bogotá and Cali were contacted by the local SDIs, whose principal objective is to facilitate the production and access to geographic information across the city, thus placing the importance of open geographic data considerably high. On the other hand, in Medellín and València the authorities contacted were the City Halls, where the open data initiative is assessed and created in terms of Open Government, or Smart City initiative (see Section 2.2; therefore, geographic information is taken as another type of data, and its relevance is moderately less than in Bogotá and Cali.

The combination of official local authorities, data user communities, and open data experts allow this research to take a bottom-up view of the open data schema to understand what the data user requirements are and their contributions to

---

[1] Open Data València, available online: http://gobiernoabierto.valencia.es/en/, accessed on 29 May 2018

[2] Open Data Medellin, available online: https://geomedellin-m-medellin.opendata.arcgis.com/, accessed on 29 May 2018

improving the reuse of open geographic data in cities. As it was mentioned by Nug (2015), a better relationship between local data authorities and data user communities stimulates the provision of data and increases the involvement of data users.

The local authorities contacted were, Bogotá SDI (Infraestructura de Datos Espaciales para el Distrito Capital - IDECA [3]), Cali SDI (Infraestructura de Datos Espaciales de Santiago de Cali - IDESC [4]), City Hall of Medellín [5] and City Hall of València [6]. The next two sections provide some background information about the online survey and the conducted participatory workshops.

3.2 | Literature Review

A literature review was defined with two scopes, 1). Initially to get a better understanding of the current status of the reuse of open geodata, complementing with what other authors have done regarding reuse and its importance. Secondly 2). Validate and identified what barriers or fictions the scholars have already identified and under what perspectives other work is covering this potential issue in an open data environment.

Collecting journals, conference papers, and governmental or non-governmental reports in several databases: Science Direct (eight papers related), Scopus (four papers related), and Emerald Insights (eight papers related). The words used to find related articles were, *'barriers in open data'*, *'barriers in Open Government data'* and *'barriers in Open Government'*.

Only papers that addressed barriers, challenges, issues to reuse, adoption, and releasing data were taken into account. Additionally, use was made of the cited references in papers where barriers were identified in order to enrich the discussion and literature review. The number of articles was filtered by year, choosing only articles from the last five years (2012 to present) in order to have a current approach, and only journals related to governments, open data, geography and economics were taken into account. The literature review was classified in two

---

[3] IDECA Website, available online: https://www.ideca.gov.co/, accessed on 30 May 2018

[4] IDESC Website, available online: http://www.cali.gov.co/planeacion/publicaciones/3560/idesc/, accessed on 30 May 2018

[5] Medellin City Hall Website, available online: https://www.medellin.gov.co/, accessed on 30 May 2018

[6] València City Hall Website, available online: http://www.valencia.es/, accessed on 30 May 2018

groups: barriers from data producers' and users' perspectives. In total, 12 relevant papers were selected and related to barriers to reuse. The relevance of those papers was determined by scanning and manually reviewing their title and abstract. These related papers can be found in Tables 3.1 which sum up the barriers considered by authors but that only include the data producer perspective and table3.2 which list the papers where the the data users perspective is considered and the categories barriers identified.

Table 3.1: This table represents the type of barriers to release data considering national or local use cases of open data initiatives, mentioned by each author. Due to there is no standard classification, barriers columns illustrate the barriers mentioned in each work, and at the same time the geographic context used for the use case. Note that mostly the national level is considered.

Author(s)	Barriers							Geographic Context
Yang et al. (2015)	Technological	Organizational	Legal and policy					New York State
Janssen et al. (2012)	Institutional	Task complexity	Use and Participation	Legislation	Information quality	Technical		The Netherlands
Martin and Foulonneau (2013)	Governance	Economic issues	Licenses and legal frameworks	Data characteristics	Metadata	Access	Skills	Rennes, France, Berlin, Germany, and UK
Barry and Bannister (2014)	Economic	Technical	Cultural	Legal	Administrative	Risk related		Ireland
Conradie and Choenni (2014)	Fear of false conclusions	Financial effects	Opaque ownership and unknown data locations	Priority (i.e., local government has more important things to do first)				Rotterdam
Wang and Lo (2016)	Data findbility and collecttion	Data layout and format selection	Personal privacy	Data licensing	Data Description			Taiwan
Attard et al. (2015)	Technical	Policy/Legal	Economic/Financial Budget	Cultural				N/A
Schmidt et al. (2016)	Desire to publish results before releasing data	Legal constraints	Loss of credit or recognition	Misinterpretation or misuse	Loss of control over intellectual property	Organizational constraints		N/A
Carrara et al. (2017b)	Poor quality Open Data	A lack of standardization or heterogeneity	Difficulties in obtaining the data with the right information (metadata) for the purpose of its usability					European National level
Carrara et al. (2017a)	Political	Legal	Technical	Financial	Others			European National level

Table 3.2: This table represent the mentioned barriers by some authors to release and reuse open data, considering the perspective of data users.

Author(s)	Barriers			
Carrara et al. (2017b)	Low quality of Open Data	Lack of standardization	Availablity of open data, poor discoverability	Incorrect metadata
Carrara et al. (2017a)	Little awareness	Low availability	Legal	Technical Financial
Zuiderwijk et al. (2012)	Fragmentation of data	Lack of access to data	Lack of interoperability	Difficulties in processing the data
Janev et al. (2014)	Lack of standard procedures for querying government portals	The low quality of metadata	Low reliability and incompleteness of public datasets	The heterogeneity of formats used to publish open data
Schmidt et al. (2016)	Paying for data	Varying degrees of data quality in different datasets	Varying standard in how data has been gathered	Varying data formats

But not only the barriers or frictions to reuse the open data was the goal of this literature review, as we mentioned above, the review was also undertaken to shape the concepts and critical factors to improve the re-usability level, considering papers where authors suggested recommendations, best practices or results. Table 3.3 disclosing the elements that other authors have illustrated as relevant or should be intensely developed if the goal includes the data user perspective. These elements provided by Table 3.3 are the fundamental part to later define the conceptual framework elements, which are also combining with the data users perceptions gathered the next two stages in this research.

Table 3.3: The references used where authors have proposed improvements to enhance the re-usability level of open data.

Improvement Proposed	Literature
Promote the released data reuse	Benitez-Paez et al. (2017), Zuiderwijk et al. (2014), Janssen et al. (2012), Org (2015), Welle Donker and van Loenen (2017), Johnson et al. (2017)
Identification of users and re-users needs	Kassen (2013), Janssen and Zuiderwijk (2014), Benitez-Paez et al. (2017), Williamson and Eisen (2016)
User-focused metadata	Carrara et al. (2017a), Comber et al. (2008), Comber et al. (2006)
Easy to read terms of use	Zuiderwijk et al. (2014), Beno et al. (2017), Attard et al. (2015), Zuiderwijk and Janssen (2014)

The literature reviewed has guided us to validate the first and the fourth assumption (see Section 3.2). We have assumed that the level of reuse is currently an essential challenge in open data strategies and likewise, the considerations of the data users' perspective is a point of view rarely considering. Therefore, this stage helps us to validate that the data producer perspective is included in much more opportunities for other authors, than the number of studies that included data users perspective as the core of the investigation. At the same time, throughout this literature review, we gather the recommendations from authors that have covered the importance of data users in the open data strategy and the suggestions or elements mentioned as required factors in improving open data initiatives. This suggestions also put some light on the way that this research should consider to improve the reuse level including the point of view and experience of data user communities.

3.3 | Online Survey: Current status of open geodata reuse

Taking into consideration the potential data users' barriers obtained in the literature review (see above section 3.2), and considering that citizens access to data through the official open data portals, an online survey was designed with the public Google Forms web application.

The survey aim was to know the barriers, errors, or problems that users have encountered while using cities' open data portals and its shared datasets, especially geographic data web services. The questionnaire was released in three different languages (Spanish, English, and Portuguese) to gather more responses from several cities. The survey was a modular form with seven sections, including general information about the respondents (working country, city and age), their work (employment role and industry), perception of open data, possible barriers faced, most-used features in well-known cities' open data portals, and finally method(s) used to find open data in a city—especially geographic data. The survey took about five minutes to complete, and was anonymous (i.e., no information about the name of the participant or email was collected). Participation in the survey was voluntary, and it was not necessary to answer all questions. The Appendix A.2 presents the questions formulated in the survey.

The survey was launched in August 2016 and remained active until December 2016. The survey was shared in several ways:

- Through social networks (e.g., Facebook, Twitter, Linkedin)

- E-mail list

- Several open data and smart cities events during spring–winter 2016 (e.g., International Open Data Conference 2016 [7], Open Cities Summit 2016 [8], Inspire 2016 [9], Geo Mundus Conference 2016 [10], Esri User Conference

---

[7] International Open Data Conference, available online: https://www.opendatacon.org, accessed on 30 May 2018

[8] Open Cities Summit, available online: http://opencitiessummit.org/, accessed on 30 May 2018

[9] Inspire, available online: https://inspire.ec.europa.eu/conference2018, accessed on 30 May 2018

[10] Geo Mundus Conference, available online: http://www.geomundus.org/2018/, accessed on 30 May 2018

2016 [11], Esri Spain User Conference [12] and Esri Colombian User Conference 2016 [13]).

The survey received replies from data users from cities in South America and Europe, but especially cities in Colombia and Spain (see Figure 3.2).
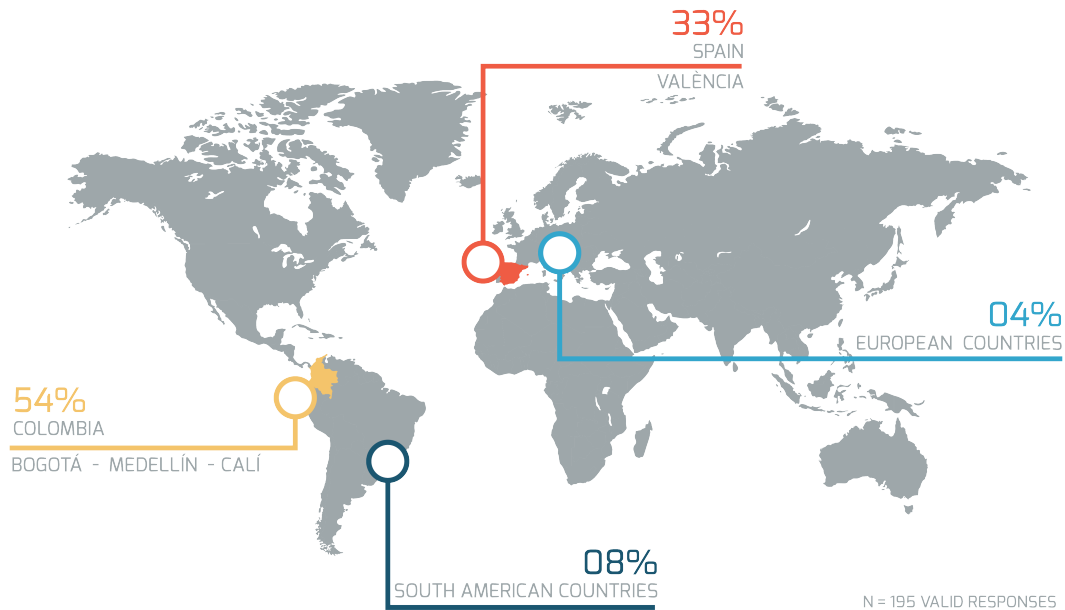
Figure 3.2: Countries and cities registered by participants that participated in the online survey. n=195 valid responses

Overall, a total of 195 participants completed the survey. However, some of them did not completely answer the questions; therefore, some questions have a smaller sample. Only responses that were fully completed were considered. Concerning the employment role (n = 195), 25% (48) of participants saw themselves as geographic data analysts, and 19% (n = 37) as part of academia (e.g., professor, researcher, or student). It could be argued that the high prevalence of participants with a geographical background and from academia was due to the way that survey was promoted with university colleagues that helped to distribute the survey

---

[11] Esri User Conference, available online: https://www.esri.com/en-us/about/events/uc/overview, accessed on 30 May 2018

[12] Esri Spain User Conference, available online: http://conferencia.esri.es/, accessed on 30 May 2018

[13] Esri Colombian User Conference 2016, available online: http://esri.co/, accessed on 30 May 2018

and organizations that work with geographic data. Regarding managers and project leaders, about 18% of participants (a third of respondents) were part of this group of open data users. About 17% ($n = 33$) saw themselves in multiple roles, as developer and analyst at the same time. Finally over 21% of participants were developers of any type of application, exclusive geographic developers, or had a different role (see Figure 3.3).
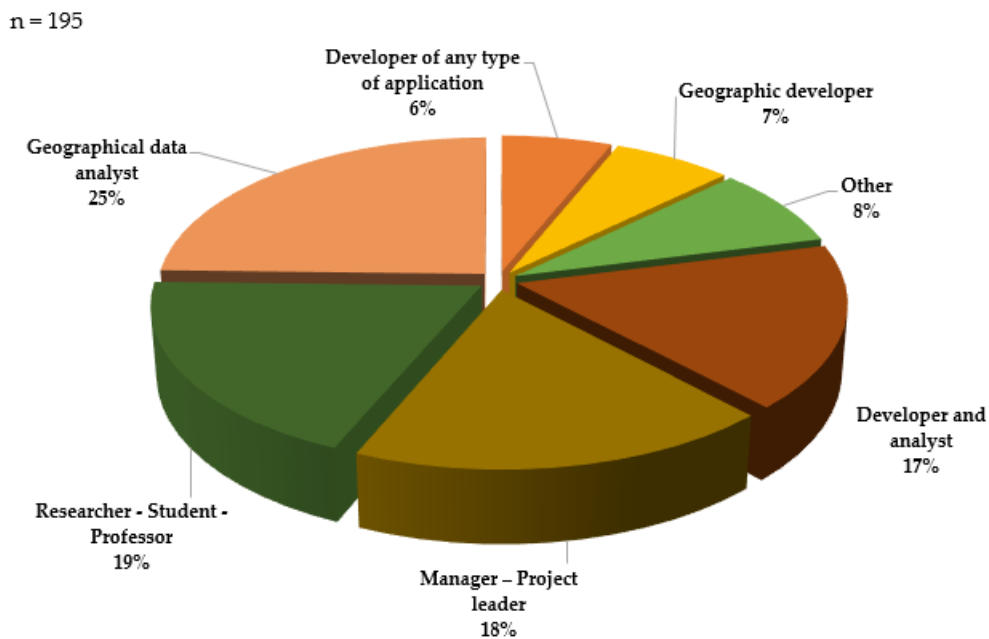
Figure 3.3: Employment role of respondents.

In relation to the question "In which industry do you work?", the 27.2% of our respondents have mentioned Local Government as the industry where they are currently working on. The following two categories, Business/Private sector and Education with 22.1% and 19% respectively (see Figure 3.4) were also considered by the participants. Due to most of our respondents are from Local Government also we have included the years of experience in the industry, to validate that our participants are familiar either with the data at city level and also with the the challenges that ongoing initiatives in open data regard. Actually we have asked to our participants if they have explored or use the cities' open data portals in which the 76% have mentioned Yes and only 24% replied Not (see Figure 3.5). Figure 3.6 shows the year of experience of our participants in the industry.

Most of the respondents have 2 to 6 years of experience with the 43,5% and 19,4% with 7 to 10 years of experience in the industry.
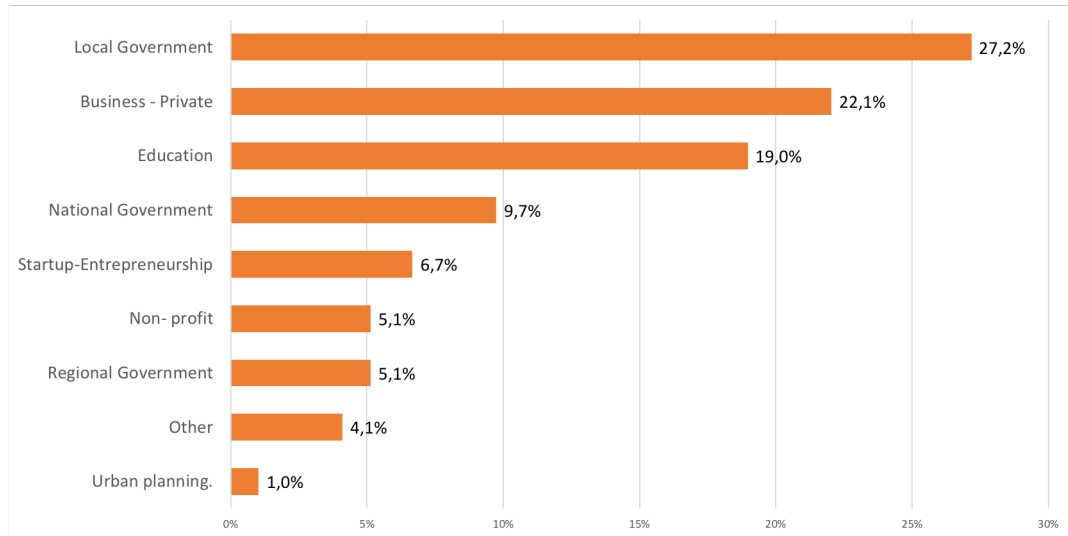
Figure 3.4: Industry registered by participants that participated in the online survey. n=195 valid responses
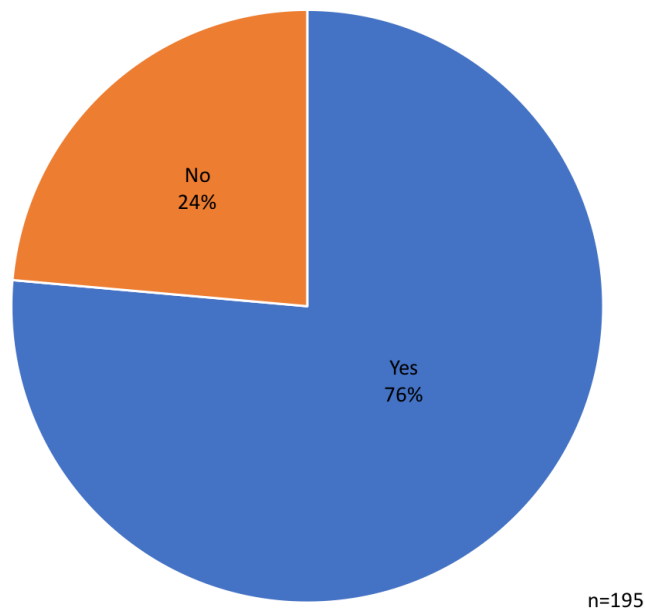
Figure 3.5: Knowing and exploration of local open data portals. n=195 valid responses
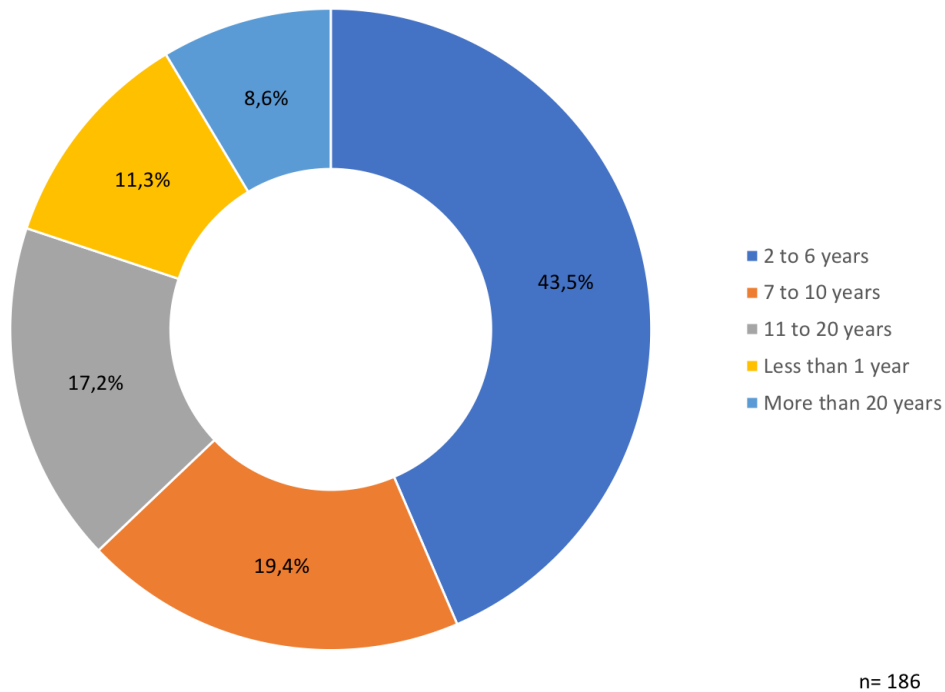
Figure 3.6: Years of experience registered by participants, having 43,5% of our respondents with 2 to 6 years and 19,4% with 7 to 10 years of experience. n=186 valid responses

Overall, the aforementioned figures are offering an overview of the participants' experience, their knowledge regarding local open data portal, the industry where they are currently working on and finally their background. Having a better perspective of data users participating in this stage was a relevant part of this research, due to the need to have e heterogeneous sample, that contribute with valid responses that could represent the reality that most of the cities are facing with the reuse of open data. It could be argued that despite the reduced sample (195 valid responses) and also the possible bias offered by the geographic community the gathered responses offer a valid perspective in terms of reuse, barriers to the ruse and valid recommendations to overcome the current challenges that most of data user communities are dealing with.

However, the survey's responses led us to get an initial data users' perception in open data reuse regard, besides to validate some of the barriers already mentioned in the literature (see section 3.3). To reach one extra level of validation and also to test what people say against what people do, next part will is devoted to talking about the participatory workshops. Considering the initial results of the survey mentioned above and selecting only four cities with the most higher participation in the survey, we have organized in collaboration with local data authorities the workshop called Open Data for Open Cities.

3.4 | Selecting the use cases (Cities and Users communities)

As we have mentioned in the last section, having responses from an online survey is not sufficient to understand the whole picture of open data users' issues regarding use or reuse. To shed some additional light on users' barriers, participatory workshops were conducted with participants from different backgrounds and experience in open data regard. The participatory workshops were called Open Data for Open Cities. Figure 3.7 shows that the participants in these participatory activities were developers, entrepreneurs, analysts, journalists, professors, researchers, open data experts, or data authorities who also consider themselves as data users Zuiderwijk et al. (2014).We organized four participatory workshops in the selected cities and three extra in different context with the academic sector in the Agile Conference [14] and open data experts where we could validate the initial results from the online survey or from the previous workshops made in the selected cities, the final workshop aim to validate the elements presented in the conceptual framework and the final result of this research.

In a total of 7 Open Data for Open Cities (OD4OC) workshops took place during this research, and with over 213 participants the reusability and the recommendations to overcome the current barriers to reuse the released data were the most mentioned topics. The last participatory workshop was made in Valencia with data users from different industries to validate the the last contribution of this research. During this stage the four initial workshops aim was to observe, confirm the mentioned barriers in the survey, allowing to data users to express their concerns to

---

[14] Agile Conference, available online: https://agile-online.org/, accessed on 20 August 2018

effectively use or reuse of the available datasets in each city through the open data portals that they consider relevant for their external application or analysis. During the two extra OD4OC in the AGILE conference the aim was to present the results of the section 3.4 and also get feedback from open data experts or others authors to enrich our results with more experience opinions. Finally in the last participatory workshop the aim was the validation of the elements of the conceptual framework that enclose the recommendation from the previous discussion with data users from several cities.

Bringing together the data user profiles that have been working in the same city give this research a broad view of the current data user barriers at a local level. To consider the cities with the most collaboration (see Figure 3.2), this research has chosen the aforementioned cities (see Section 3.1) for the workshops. Likewise, two more workshops were conducted in Wageningen, The Netherlands with students of a Master's Geographic Information Science (GIScience) (33 participants) and open data experts (11 participants) and Lund-Sweden in the last AGILE conference 2018 (25 Participants) to have better insights into barriers faced by geographic data users, but also the elements included in the conceptual framework to improve the reusability level at local level. Collecting the opinions and impressions from people from EDP, master and PhD students, and open data experts like Andrew Turner from Esri Research and Development centre in Washington with lot of experience in the implementation of open data portals in United states, or Professor Alexis Comber who has lot experience and set of publications regarding the importance of a improved metadata that cares of the users' requirements. The last OD4OC workshop took place in Valencia city with over 23 participants from several local institutions and with different roles and open data experience, the goal was to discuss and validate the elements presented in the conceptual framework suggested in the Chapter 5.
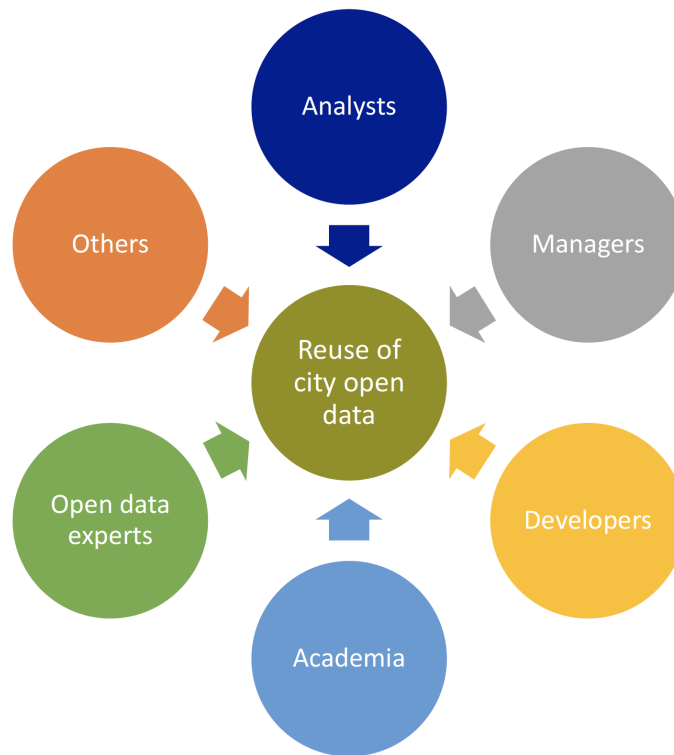
Figure 3.7: Workshop participant roles. There were 56 from academia, 49 analysts, 20 managers or project leaders, 11 developers, 3 Open data experts and 16 counted as others that include politicians, journalists, entrepreneurs and others roles. In total there were 155 workshop participants, only into the OD4OC organized in the four selected cities.

All the participatory workshops lasted approximately four hours, split into two-hour sessions. The methodology applied in the four initial OD4OC workshop had the same two parts. The initial session was about finding suitable city open data in the official or well-know open data portal, using the main data domains defined by the contacted data authorities (e.g., mobility, education, urban planning, air pollution, crime, and others) depending on the priority of each city (see Table 3.4). Participants were required to create groups with three members at most, then choose one category which they found more interesting or related to their work. Once the groups were created and the category was selected, participants were required to think about a general idea, analysis, or application that included published data of each city. During the mentioned first session participants were looking and evaluating datasets and its properties to use or reuse for external projects. During this activity, participants were able to bring their own laptops or use computers provided by the organizers with access to the Internet. Details

of this research were not included at the beginning of the activity in order to reduce a possible bias presenting the initially identified barriers in literature or the online survey. Participants were able to use any search method that they considered appropriate to find the open data in each city (e.g., search engines like Google, official open data portals, or any web portal). The second session was a discussion, where participants could express all the found barriers to reuse, their requirements, or common issues when they need to include open data in their work. Local authorities were also part of this discussion, but only obstacles from a data user point of view were collected.

Table 3.4 illustrates the general aspects of each selected city, such as population, context, current data policy regulation, main data thematics, terms of use, and identified users. In addition to the workshops mentioned above, there were two participatory workshops in different cities with other kinds of participants.

During the workshops, a follow-up questionnaire was administered regarding methods used to find data, barriers found, and users' suggestions to overcome the obstacles found. Participants' personal information such as name, email, and organization were collected in analogue form, but not in a way that would identify them personally. This information was used to share the workshop report with participants to explore the results and insights collected. Participation in those workshops was also voluntary.

Table 3.4: General aspects of selected cities data authorities.

	Bogotá	Medellín	Cali	València
Country	Colombia	Colombia	Colombia	Spain
Population	8.080.734 inhabitants (2017)	2.508.452 inhabitants (2017)	2.420.013 inhabitants (2017)	790.201 inhabitants (2016)
Context	The most populated and capital of Colombia	The second most populated city of Colombia	The third most populated city of Colombia	The third most populated city of Spain
Autority(ies) Contacted	SDI Bogotá (IDECA)	Medellín City Hall and Ruta N	Cali City Hall and SDI Cali (IDESC)	València City Hall and Las Naves
Main Open Data Theme of Interest	Urban Planning, Economic Development, and Infrastructure	Security, Environment and Urban Planning around a sustainable smart city strategy	Mobility, Security, and Health	Environment, Transport, Society, and Wellbeing are the themes more used andconsulted of the Open Data catalogue
License or Terms of Use of Open Data	IDECA license	License Attribution-Share Alike 4.0 International	No open data license, only IDESC web site terms of use	All the datasets offered by the City of València, unless otherwise indicated, are published under the terms of the Creative Commons license-Recognition (CC-By 4.0)
Open Data Portal or Official Portal	IDECA website	GeoMedellin Website	IDESC website	València Open data website

Continued on next page

Table 3.4 – *Continued from previous page*

	Bogotá	Medellín	Cali	València
Current Engagement Activities	Strategies implementation to facilitate the discovery, use, and reuse of available open data	Engagement activities with the community and identified users. Creation of the platform of open data, dynamic visualizations, and analysis with the data of the different dependencies of the Mayor's Office of Medellín	Create channels of communication with citizen initiatives related to open data in the city. Promote the publication of open data of utility by the agencies of the Mayor of Cali. Promotional events for the open data available in Cali	The position of the City Council in relation to Open Government is that the technologies serve for the citizens to have more knowledge of municipal action and to make possible participation and collaboration with the management of the city; actively listen to citizens in social networks or any other media. They also work on the creation and application of standards as well as the use of transmedia to bring important issues to citizens
Developer Companies Identified as Open Data Users	A few companies identified. Note that this identification is not done periodically	There was one company identified	There were three companies identified	The policy of the City Council in Open Government, does not see as relevant to collect data of entities or individuals who have used the datasets

Continued on next page

Table 3.4 – *Continued from previous page*

	Bogotá	Medellín	Cali	València
Universities or Colleges Identified as Open Data Users	There were several universities identified	There are several universities identified	There were several universities identified	Public Valencian universities collaborate with the city council in organizing activities and events on open data
Internal and Official Authorities Identified as Open Data users	There are 73 local entities integrated and identified	City Hall, Metropolitan and regional authority	Utilities, Transportation, Urban planing and Environmental, and Economical authorities	Representatives of the regional government have collaborated in some of the events of the Open Government Chair with the Polytechnic University of València, and both policies—local and regional
Urban Observatories or Analysis Groups Identified as Open Data Users	Several urban observatories were identified	Only one Urban observatory was identified	Several urban observatories were identified	N/A
Others Identified Open Data Users	Several cities stakeholders considered relevant	Several cities stakeholders considered relevant	Several cities stakeholders considered relevant	N/A

Along the two participatory workshops made in the context of the AGILE conference, the method was quite different, due to the background of the participants, the aim of those workshops was to get the required feedback. Collecting ignored elements in the preliminary results or considerations in the online survey or the four initial participatory workshops, the discussion led during these two conference bring us a extra level of understanding about the importance of reuse of open data in cities and to understand the difficulty of measures the impact of open data. Finally the last participatory workshop made along this research had the goal to validate with data users from different experience level with open data the four elements presented in the conceptual framework to get feedback and include possible ignored elements that others data users have not suggested. The last OD4OC was important due to the amount of discussion around the recommendations to overcome the identified barriers, the importance of an improved metadata option in local open data portals, the issued faced by data users in term of knowledge about the right to reuse the released data published in local open data portals.

The participatory workshops Open Data for Open Cities - OD4OC were a fundamental part of this research. Using these meetings with data users in four selected cities and academic congress the barriers identified in the online survey were validated. Users and local authorities had the opportunity to express their particular concerns either with the released data process and also with the consumption of released data by data user communities side. Although the goal of this stage was the set space where data users could express why they are facing the reuse issues or what are the reasons to do not rely on in the ongoing open data initiatives. Local Authorities especially in Medellín, Bogotá a Valencia could see the importance to include data user requirements in their strategy in open data regard.

3.5 | **Summary**

During this section, the stages considered in this research were described. Figure 3.8 shows initially the literature review included validating the current status of the open data status at the local level. This literature review was made to validate how other authors have considered the issue of the reusability, analyzing other use cases, implications and found barriers that prevent the effective reuse of open data. Once the literature review reveals some barriers and considerations such as the importance of local level, the gap between the data producers activities and the data user requirements and finally the kind of data that could encourage data user communities to reuse more actively the release data, an online survey was designed to directly ask people in cities about their perception in open data reuse regard. The online survey publicly shared allow to this research to define which cities were more participative, having Bogotá, Medellín, Cali in Colombia and València in Spain as cities with more participants, therefore, were the selected cities for the next stage. With over 113 data users in the four cities, we led a discussion and set of activities properly described in the next section where the reusability, impact of open data and the data user requirements were the key topics that we collected to validated the identified barriers in the above-mentioned survey. Two extra participatory workshops were led to getting the required feedback around the preliminary result and the assumptions presented in this research. The AGILE conference was the way to discuss the online survey' results and some of the gathered data user recommendations in the previous workshops.

The results of this initial activities were properly described in the research article Benitez-Paez et al. (2017), in which we define a taxonomy of data users' barriers to help data authorities to easy identified where their data user communities are having more issues to effectively reuse the released data in their cities, besides a set of recommended were also included. The whole description and details of the definition of the taxonomy of data users barriers are properly described in the next section.

However a definition of a taxonomy of data users' barriers Benitez-Paez et al. (2017) seems to be not enough, a kind of guideline is required to help data authorities to re-shape their current strategies considering a bottom-up approach and include the recommendations made by data users in the participatory workshops.

Including the set of recommendations from previous participatory workshops,

literature review and the most mentioned concerns in the taxonomy of barriers, a conceptual framework is designed and validated in an additional participatory workshop with data users from València city. The whole description of the elements of this frameworks is properly described in chapter 5

Figure 3.8: Method Overview. Literature review, identifying initial barriers and perceptions, then a publicly shared online survey (*n* = 195 valid responses). Using the preliminary results, four cities were selected and the identified barriers were contrasted with the participatory workshop in each city.

4 | Roadblocks hindering the re-use of open geodata in cities

Open data must be data people can easily use.

Word Wide Foundation

In the previous chapter 3 we have presented the methodology used during this research. Through three main activities our work was devoted to analized across the literature the current status of the reuse of open geodata, secondly using a comprehensive online survey we asked data users especially in cities about their impressions, challenges, barriers, recommendations and overall their option about the level of reuse of the open geodata in their cities. This survey allow us to have a better picture regarding the potential issues that local users are currently facing in reuse regard. Despite the number of valid responses (195 valid responses, see previous Chapter 3), the validation and face-to-face discussion was needed to corroborate what people say and what people do. Considering only the most participatory cities, we have selected four cities to organize a set participatory workshops. During those participatory workshops, we meet local data authorities, data user communities with different backgrounds and open data experience, around the reusability concept. Collecting barriers and recommendations this discussion was the way to validate the identified barriers in the scholar and the survey. The results of those activities is the data users barriers taxonomy and the current status of the reuse level of open geodata that this chapter is describing. In the chapter 5 an additional workshop is developed in Valencia-Spain to validate and test the conceptual framework elements that were designed taking the most mentioned recommendations from data users in the initial participatory workshops

as an idea to compile a guideline to help data authorities to improve the reusability level of their ongoing open data initiatives. Including the perceptions, suggestions of Valencia data users the conceptual framework was validated and disputed and is properly described in Benitez-Paez et al. (2018).

4.1 | Overview

Open data holds the promise of "*dramatically reduc[ing] the time and money citizens need to invest to understand what government is doing and to hold it to account*" The World Wide Web Foundation (2015). The word "*open*" can be interpreted in many ways (for a recent review, see Pomerantz and Peek (2016)), but throughout this research it is used the definition based on the Open Definition and mentioned in section 2.1. During this dissertation we are attempting to demonstrate the important role that reusability currently has, along the open data strategies. Due to open data is now entering the mainstream, with 51 countries (i.e., about 25% of all countries in the world) having an open government data (OGD) initiative according to Org (2015). Empowering citizens to take full advantage of available open data, is a promising way to foster innovation and citizens-centric solutions for cities (see Degbelo et al. (2016b)).

The geographic community has carried out considerable effort from international to the local level, developing and implementing an integrated way to promote the sharing process of geographic data. Local Spatial Data Infrastructures (SDI) were since 1992 Douglas D. Nebert (2014) a way to tackle issues such as standardization, integration, and accessibility shaping a framework that combines institutional arrangements, several technologies, and new policies around geodata. SDIs have been only attractive for limited and specialized geographic communities, and as discussed in Díaz et al. (2012), could benefit from existing trends (one of these being open, distributed and linked data).

Opening up data is valuable, but using available open data to provide useful services to citizens is equally important. According to Andrus Ansip, Vice-President of the Digital Single Market of the European Commission, "*Data should be able to flow freely between locations, across borders, and within a single data space. In Europe, data flow and data access are often held up by localization rules or other technical and legal barriers. If we want our data economy to produce growth and jobs, data needs to be used. However, to be used, it also needs to be available*

and analyzed"European Commission (2017).

Along the same lines, Janssen et al. Janssen et al. (2012) stated: "*Open data on its own has little intrinsic value; the value is created by its use. Supporting use should not be viewed as secondary to publicizing data*". Previous work has investigated various aspects of Open Government data initiatives. These aspects include a business model for Open Government data (OGD) Ahmadi Zeleti et al. (2016), a measurement framework to quantitatively assess the quality of OGD Vetrò et al. (2016), an index to measure the maturity of e-government openness Veljković et al. (2014), the use of semantic application programming interfaces as a way of improving access to OGD Degbelo et al. (2016b), and the motivations of citizens to participate in OGD projects Wijnhoven et al. (2015), to name but a few. Complementary to these, this work takes a user-centric view, and investigates barriers faced by people when interacting with existing open data portals.

Next sections will explain the identified barriers in the three ways used along with this research and described in the previous chapter. Initially the identified barriers through the literature (see section 4.2), looking for what authors have covered the data producer perspectives and what related work is covering the data user point of view and under what conditions the barriers were identified or validated. Secondly and based on the preliminary results of the literature, the validated barriers on the online survey with a participation of 195 valid responses are described in the section 4.3. As we have mentioned in the methodology chapter 3., the initial barrier validation through the survey was not enough. The next step to fully identify what frictions data users are facing in the reuse of open geodata in cities included a set up of participatory workshops in selected cities or use cases, the identified barriers in this stage are described in section 4.4. Once we have enough data regarding the associated barriers considering only the data users perspective at a local level the final step was to set up a taxonomy of the most mentioned and collected barriers during the three ways but making particular emphasis in the opinion of participants in the workshops. The definition of the taxonomy of barriers to reuse open geodata identified in cities in Colombia and Spain, particularly in Valencia is described in section 4.5 and is available in the research article Benitez-Paez et al. (2017).

4.2 | **Identified Barriers in Literature**

Much of the documentation that this research has reviewed mostly referred to the three following aspects.

1. **Benefits of Open Government implementations**, through several countries, explaining the ways to reach the economic, social, or political benefits of releasing government data;

2. Regarding barriers, challenges, or issues in the literature, there are mainly two categories of open data barriers: **from data producer or data user perspectives**;

3. Most of the **published papers discuss national governments**, but the local governments are briefly mentioned in addition to the possible barriers from the data user point of view.

Martin and Foulonneau (2013) demonstrated through local and national cases that the sustainability of the open data initiatives needs to be considered regarding risks, challenges, and limitations, having in mind the evolution of the stakeholders involved (re-users, data creators, and national aggregators). Related to the role of users in open data systems, Janssen et al. (2012) suggest that feedback and insights from this point of view must be considered in order to continuously improve. Janssen also established a list of adoption barriers of open data, presenting "Use and Participation" as part of those obstacles in the open data implementation process. Barry Barry and Bannister (2014) and Conradie Conradie and Choenni (2014) consider the process of releasing open data as the center of attention in Open Government initiatives; they examined the barriers to open data release at national and local levels from the perspective of senior managers and six local public sector organizations.

Wang and Lo (2016) examined three factors that influence the adoption of OGD, where perceived benefits of OGD are more significant than other determinants of OGD; however, looking into the perceived barriers in this work, the participants mentioned data findability, personal privacy, data layout, and licenses as potential barriers of OGD adoption in their organizations. However, not only the official governments have been consulted, in Schmidt et al. (2016), global environmental

data research and data infrastructure communities were considered in a survey to highlight users' perceptions in terms of open data, and also barriers to share data. The survey revealed that "paying for data", "varying degrees of data quality in different datasets", and "varying standards in how data is gathered" are seen as the most significant burdens. Attard et al. (2015) presented a systematic review of OGD initiatives describing 15 challenges where citizen participation is an essential factor to promote innovation among developers and other stakeholders. However, a number of barriers prevent public participation—most of them are included as cultural challenges in this work.

In terms of the private sector or organizations that have the skill to transform open data in a new bunch of innovative services is likewise a relevant group of users considered in the literature. Although Beno et al. (2017) mentioned that the barriers faced by the private sector have not been sufficiently studied, the EDP project performed a study with 76 organizations across Europe Carrara et al. (2017b) to understand how they use open data and what business models have been developed based on the reuse of the available data, finding that there is a mismatch between the available data sets that public organizations are releasing and the data sets that are most reused. Meanwhile, another report also from the European Data Portal project Carrara et al. (2017a) presents a set of barriers faced by open data suppliers and users considering the study above in EU28+ countries. For open data publishers, the most frequently encountered obstacles are financial and legal; however, for re-users of open data, lack of awareness and low availability are the barriers most mentioned in this report. An important remark of this report is that geographical data is counted as a technical barrier; according to Carrara et al. (2017a), a significant part of all information used and published by public administrations and exchanged with citizens has a spatial component. Thus, aspects such as different standards, level of geographic knowledge, lack of metadata, and even file size are significant barriers that prevent users and publishers from efficiently working with geospatial data.

Notwithstanding that the benefits of open and government data have been mentioned in most of the literature, there is also some work that has been done analyzing determinants of the success or failure of open data projects, especially involving government authorities. Yang et al. (2015) illustrated through four perspectives of the impact of open data initiatives in Taiwan that legislation and policy have the most significant impact. Additionally, Keefe et al. (2013) used

a case study of an e-Government project to explore the key factors of an open data project's success. Revealing that the development of a management and measurement framework of all the objectives and aims can bring some success, at the same time the lack of clarity about aims and specific objectives from the side of partners could affect the project development.

In Bargh et al. (2016) work, the definition of Semi-Open Data paradigm is presented to define and frame initiatives and efforts that publish data but do not entirely accomplish the open data requirements. The authors presented a method to assess the level of implementation of the semi-open data in organizations, acknowledge their effort and guide them to reach the open data requirements. In fact, public agencies like Great Britain's Ordnance Survey from geospatial services sector, got realistic economic benefits partially releasing data, developing a mixed- cost model, with some free data and also some paid data Young and Verhulst (2016).

To review the barriers found in the literature and categorize what barriers belong to the data producer's perspective and what barriers belong to the data user's point of view, Table 3.1 illustrates authors, types of barriers, and the geographic context that proves that most of the work done has not considered the local level. Additionally, Table 3.2 presents the references where data users' barriers were included. Finally, because most of the obstacles cited were not mentioned in the same way and there was no generic categorization found, Table 4.1 summarizes the number of occurrences to determine what barriers have been most analyzed. There are five relevant findings listed regarding the literature, as follows:

1. Seven relevant categories of barriers considering the data producer's point of view were most mentioned in the literature:

   - Technical
   - Organizational
   - Legal and Policy
   - Data quality
   - Financial issues
   - Cultural
   - Use and Participation

2. It seems that Use and Participation barriers are still not significant barriers; only two authors mentioned the user perception and active participation as an important issue to release or use open data.

3. Regarding the previously mentioned barriers experienced by data users, the categories that were not included are as follows:

   - **Standardization**: Included as another category where fragmentation of data, lack of interoperability, and many standards in how data is gathered are seen as issues from data re-users.

   - **Accessibility**: It is seen as heterogeneity of formats and lack of access to re-users.

   - **Discoverability**: Defined as how easy it is to find the data that is required. Related to other barriers such as standardization of data quality (metadata) but categorized as a remaining challenge by users.

4. Categories such as legal, financial, and technical were also mentioned from a data user point of view, but were less cited.

5. Data quality is still a significant burden from data producer and user perspectives.

Table 4.1: The highlighted rows correspond to data users' barriers mentioned in the literature. The remaining rows were barriers mentioned as data producers' barriers.

Category	Occurrences
Data quality	5
Standardization	5
Accessibility	3
Awareness (cultural)	2
Technical	2
Financial	2
Discoverability	1
Legal and policy	1

4.3 | Identified Barriers in On-line Survey

Although the online survey had 21 questions for this stage, we have considered questions related to barriers regarding the reuse of open geodata in cities. The first question described was: *From your experience with cities' open data portals, what do you consider to be barriers when using those portals?* Using a Likert scale Nemoto and Beglar (2014) with three options (Major barrier, Moderate barrier, Not a barrier) respondents provided their option regarding barriers listed (see Figure 4.1). Overall, the top five obstacles considered by respondents as the most significant obstacles for the whole sample are **lack of update on published data** with 68.04% (Update data) and **low integration of data sources** with 53.09% (Standardization). Barriers related to Accessibility such as **low relevance to access** for re-users and Published data is **hard to access** with 47.94% and 47.42%, respectively. Finally, there was Discoverability barriers related to **time spent searching for data** with 43% (see Figure 4.1).

We now turn to the top five barriers mentioned by data users in the selected cities. Table 4.2 shows that **Lack of updated data** and **low integration** among data producers are the major barriers mentioned by data users in each city except Bogotá, where time spent finding data was the second major burden. A possible explanation is that data users in Bogotá (46 respondents, 23%) did not mention integration as a problem, possibly due to the existence and continuous progress of their local SDI (IDECA), which integrates more than 73 local entities (see Table 4.2). **Misunderstanding the reuse of available data** and the **terms of use** were also relevant burdens chosen by respondents in all cities. Although in the whole sample those barriers are not considered within the top five concerns, the users of cities show a significant concern with understanding how the data can be used, and under what terms of use they are available. Finally, access to data through URL to establish a direct connection to available data in external applications or analysis processes (probably to get updated data) was chosen as another relevant barrier for data users in Bogotá, Medellín, and València.

Table 4.2: Top five of barriers mentioned by data users along the online survey first question, for the entire sample and also group by each selected city

Category	Barriers Most Mentioned in Online Survey	Percentage	n
Entire survey	Lack of updates of published data	68%	
	Varying and low integration of data sources or data producers	53%	
	Nonexistence or low relevance of URL to access to data	48%	195
	Published data is hard to access	47%	
	Time spent searching for data	43%	
Bogotá	Lack of updates of published data	74%	
	Time spent searching for data	54%	
	Understanding terms of use	52%	46
	Nonexistence or low relevance of URL to access to data	48%	
	Published data is hard to access	46%	
Medellín	Varying and low integration of data sources or data producers	68%	
	Lack of updates of published data	64%	
	Nonexistence or low relevance of URL to access to data	60%	25
	Time spent searching for data	44%	
	Misinterpretation and misuse of data	44%	
Cali	Lack of updates of published data	71%	
	Misinterpretation and misuse of data	71%	
	Varying and low integration of data sources or data producers	54%	41
	Published data is hard to access	54%	
	Understanding terms of use	46%	
València	Understanding terms of use	68%	
	Lack of updates of published data	63%	
	Varying and low integration of data sources or data producers	53%	19
	Misinterpretation and misuse of data	47%	
	Nonexistence or low relevance of URL to access to data	37%	

Regarding the low relevance of URL to access data, we also gathered users' opinions about the format or service they consider most useful for their work. This was achieved through the question *what format/service do you consider most useful for your work?* (see Appendix A.2) in the survey. We found that for 186 respondents, the shapefile (80.11%) is the most useful format, secondly the downloadable formats like .zip (64.52%) and CSV (58.60%) in third place. This can explain that despite the effort in open data initiatives to promote formats like RDF or JSON and promoting access through web RESTful services, users still consider most useful having the data in their own computers and manipulate as they want.

This result may be due to the fact that in our sample 25% were geographical data analysts (probably a cultural aspect could have had an influence, since shapefile is a well-known format by this community, see Figure 3.3). Others geospatial services, such as OGC (WMS, WFS, WMTS), KML and others services based on GeoJSON format were mentioned by the participants, but had a lower percentage of occurrence (54.30%, 50.0%, and 46.24% respectively). The surprising finding was that typical machine-readable formats, such as RDF or JSON have been mentioned as less useful for our respondents (see Figure 4.2).
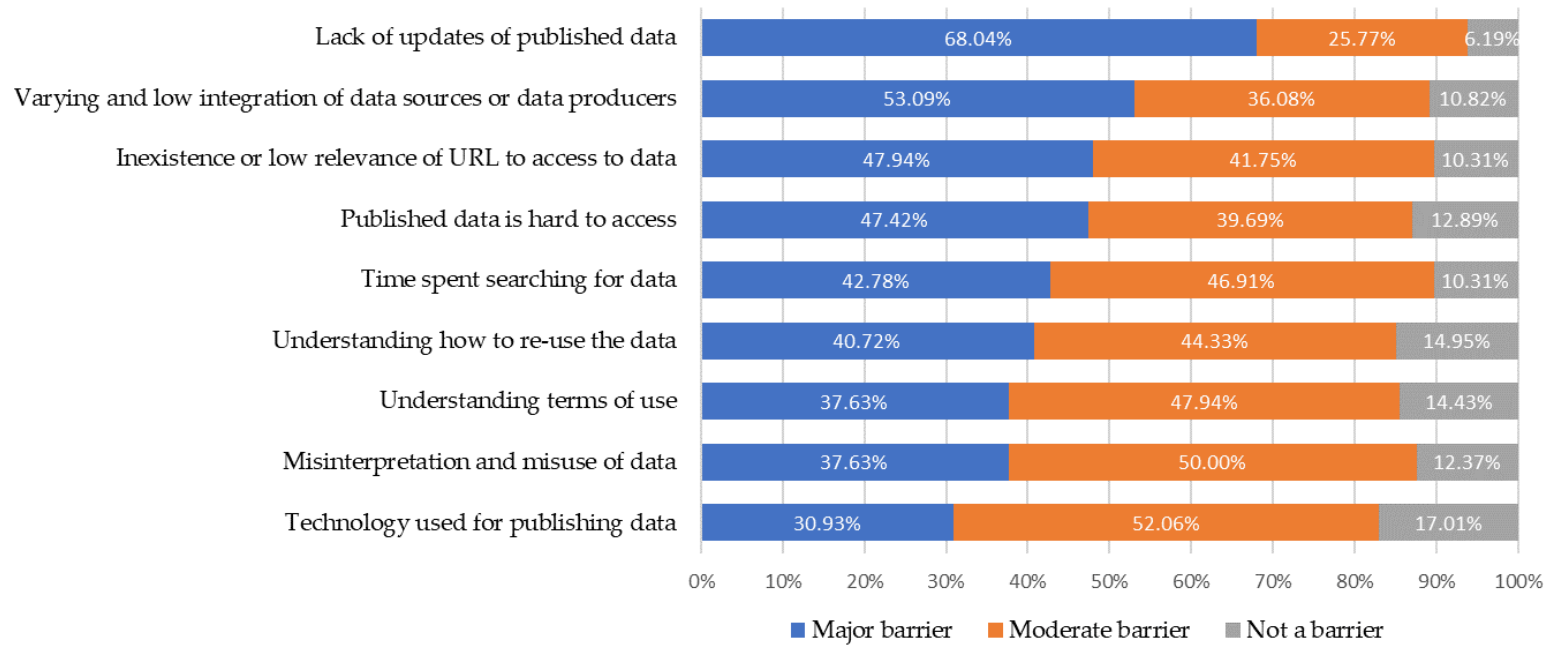
Figure 4.1: Barriers mentioned by respondents. Lack of updates of published data, varying and low integration of data sources, and low accessibility were considered to be major barriers (n = 195).
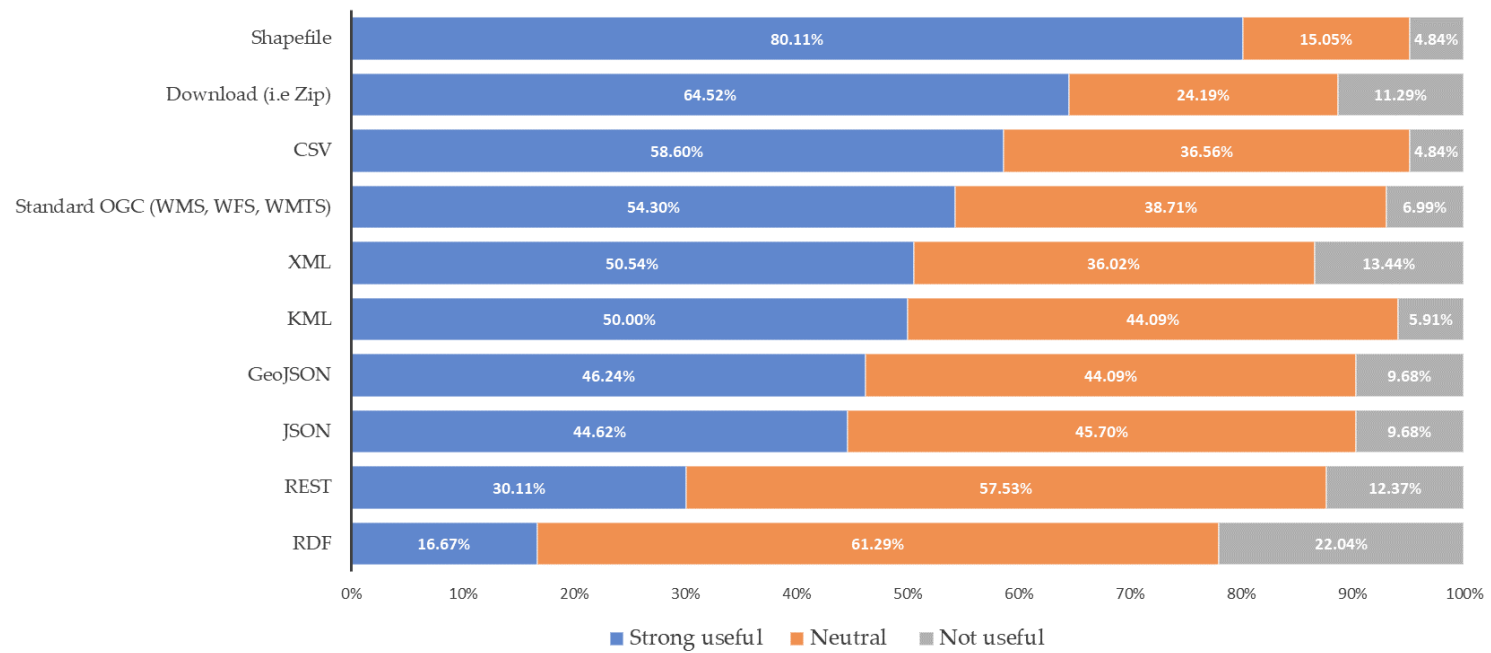
Figure 4.2: Formats or services mentioned by respondents as most useful for their work. Shapefile, .zip and CSV are considered strong useful (*n* = 186).

The third considered question related to barriers was: *From your experience, which was the most common error/barrier you have faced (not have faced) when searching or using data from city open data portals?* It was an open question, and respondents were able to enter barriers from their own standpoint. The aim of this question was to identify any barriers that were not categorized or included in the question mentioned above but which are still an issue from the data users' point of view. This question was answered by only 164 people. Some participants' answers were not related to barriers or were challenging to interpret, and were excluded from the analysis, leaving a total of 151 valid responses for this question. Since most of the replies were in Spanish, it was necessary to translate to English, then group by categories and summarize the occurrences along the replies.

Table 4.3 illustrates the number of occurrences and the frequency of all barriers mentioned by data users, clustered by the categories as stated earlier. Currency is disclosed as the most mentioned category, related to available data, but not updated, 24% (36 occurrences) were reported for data users. This means that users not only expect a vast list of data from data providers, but the possibility of having access to current data is also a constant user requirement. Barriers related to categories such as Usability (15%, 22 occurrences), Data Quality (14%, 21 occurrences), and Standardization (13%, 20 occurrences) are also described by users as the most common errors when the available data is being used or searched. It was surprising that in this question Legal and Policy (3%, 5 occurrences) and Awareness (3%, 5 occurrences) were categories with fewer occurrences. It could be argued that current cities' open data portals have unclear and complicated licensing schema (where sometimes it is better not to use the available data to avoid any legal trouble, as also mentioned by Beno et al. (2017)).

Taking into account the responses to both questions, Table 4.4 summarizes the most mentioned categories. Barriers related to Currency and Usability are two significant obstacles that are not considered in the literature (see Section 4.2); however, in this section they are validated as one of the main requirements from a data user point of view.

Table 4.3: Number of occurrences of the mentioned barriers by data users in the open question regarding the most common barrier when searching or using data from cities' open data portals

Barrier Category	Occurrences	Percentage
Currency	36	24%
Usability	22	15%
Data Quality	21	14%
Standardization	20	13%
Accessibility	16	11%
Technical	16	11%
Discoverability	10	7%
Legal and Policy	5	3%
Awareness	5	3%

Table 4.4: Summary of most mentioned category barriers by data users along the used questions in the online survey.

Category	Example of Barrier
Currency	Lack of updates of published data
Accessibility	Varying and low integration of data producers. Nonexistence or low relevance of URL to access to data.
Discoverability	Published data is hard to access. Time spent searching for data
Usability	Misinterpretation and misuse of data
Data Quality	Data catalogs with poor descriptions
Standardization	Many formats, difficulty in searching the data

4.4 | Identified Barriers in Participatory Workshops

During this activity, over 113 data users in selected cities (see Figure 3.8) discussed the data reuse and filled out over 46 follow-up questionnaires, where we asked participants about found reuse barriers and suggestions to overcome them. Since most of the replies in the questionnaires were in Spanish, it was also necessary to translate to English. Data users mentioned over 60 barriers grouped and filtered by six categories mentioned above during this activity. Table 4.5 groups these issues described in the selected cites; **Accessibility**, **Usability**, **Data Quality**, and **Currency** were the most frequently pointed out categories.

The lack of a relationship (direct or indirect) among the available datasets, defined as **non-existent geographic or statistical context**, was expressed as one the aspects to improve the usability and discoverability by data users, most of them economic analysts in the city of Medellín, geographical analysts and professors in urban planning in the city of València, and entrepreneurs who were looking for open geographic data to establish a new way to understand the education rates and their relationship with cultural indicators in city of Bogotá. In terms of accessibility barriers, two points of view have been described: user accessibility (in terms of an analyst, for whom a download option is necessary to have full control of the datasets) and re-user accessibility (in terms of developers or data enrichers, where automatic and machine access is the most relevant way to connect for their applications) Ubaldi (2013). Barriers related to this category were mentioned in all cities, but having most of the mentions in Cali, where GIScience master students cited the need to download the data in a suitable format to develop analysis processes concerning mobility and safety issues inside the city. Analysts have claimed, for instance, the following: "*there is no download option*", "*lack of mobility data*", "*data only for visualization but not able to download*", and "*many data related to events in the city but not suitable for analysis*".

Table 4.5: Most mentioned barriers by data users in selected cities in the participatory workshops.

Category	Cali	València	Medellín	Bogotá
Usability	Data difficult to understand	No suitable for reuse data format	Misunderstanding of available data	No relationship among published datasets
		No applications to validate the reuse of data	No categories for available data	No apparent usability of available datasets
		No relationship among the datasets available	No relationship among available datasets	There are no examples of reuse
		Reduced usability		
Accessibility	No download option	Only one dataset for education	No downloaded option	
	Official data web sites have no data	No transportation data is available	No georeferenced data available	Available data in PDF format
	Lack of data for transportation	Lack of important attributes	Lack of accessibility for some datasets	
	Lack of accessibility	Reduced discoverability, to find data it was necessary to spend a great deal of time	Data in PDF format	
	More marketing of current initiatives			
	Information related to events, but no data related			
	Data only for visualization, not downloadable option			
Data Quality	No metadata	Not enough metadata	No suitable format for open data	Duplication of data
	Gaps in available data	Generalization of data, only for regional or national approach, Not local level		Attribute inconsistency
	No georeferenced data			Gaps in published data
	No raw data, the available data is processed			No updated metadata
	No metadata is related to the data source			Generalization of data, Nor for local reuse
	Processed data			Published data not georeferenced
Technical		No API documentation or examples	No advanced search option	Some web sites based on Flash technology
		Language issues among datasets		User authentication for some portals
		No advanced searching options to find datasets		
		JSON file with issues		
Legal and Policy				Misunderstanding regarding terms of use
				License not clear
				Available data, but no open
				The terms of reuse are not clear
				Lot of available data, but not truly open
Currency	Not up to date data	Some datasets are not up to date	Not up to date data	Data not up to date
		No up to date apps in official websites		

Other accessibility barriers were mentioned by data users in Valencia, Medellín and Bogotá; the data download option was sometimes complicated and included web log-in. Often the available data was not in a suitable format to reuse (e.g., PDF). Having data in pdf format not only restricts the automatic extraction that results in low reuse level, it is also considered to be poor open data The World Wide Web (2017); Carrara et al. (2017a). Regarding data quality, "*gaps of data*", "*duplication of data*", "*no-clear metadata*", and "*no spatial resolution for local analysis*" were mentioned by the journalists and analysts in each city—especially in València, where the generalization level of available data (data at regional or national scales not suitable for local analysis—e.g., air pollution). The level of updated metadata was also considered by participants as an obstacle to understanding how the published data was gathered. Technical issues which were less mentioned but also cited by developers complaining that there is not enough information to understand how to use or apply the development resource. The multi-language option in some portals is not entirely supported according to València data users. Lastly, regarding terms of use, Bogotá's users mentioned a misunderstanding over the policy of available data.

A frequent issue mentioned by entrepreneurs and managers throughout the workshops was related to commercial use allowed in published data. This activity found a lack of clear terms of reuse in selected cities; some of them have created a specific license to use their data (e.g., IDECA in Bogotá), and other cities only have open data portal terms of use or they do not have a clear reuse policy. In general, after the set of participatory workshops and the interaction with data users from several backgrounds, Accessibility , Usability, and Currency categories have been reinforced as constant concerns from a data user point of view. Terms of use were less mentioned by data users, though this does not mean that licenses of available data are well-defined for re-users (i.e., developers). Many data users did not consider the available data "fully reusable" once the data was found in the cities' open data portals.

4.5 | Defining the data users' taxonomy barriers

A total of six categories have been identified and validated during this stage considering the barriers from the literature review, using an online survey, and confirming through a set of participatory workshops; each category corresponds to barriers mentioned and identified from a data user's perspective. Figure 4.3 is a fishbone diagram that represents the groups and its obstacles that prevent the reuse of open geographic data in cities.

- **Currency**: Category refereed mostly to "recency", "freshness" or "current-ness" of released data, the lack of updated data in the local open data initiatives was considered by data users as the major barrier to reusing the available data. Outdated data and services or broken links were mentioned in both online survey and workshops as most disappointing when analysts, entrepreneurs, geospatial developers, journalists, and other data users need to include data in their processes or external applications. Having updated data is a common requirement for all kind of open data, regarding geodata, data users mentioned currency also due to the difference among the available data by paid versus the available data accessible through geo-portals. Considering that there is much work to do to get full accessibility to updated data. A possible precedent associated to this issue if the way that some geographical authorities found having a mixed-open data model, releasing only a certain among of data but keeping the most updated as a premium service Young and Verhulst (2016).

- **Accessibility**: Although all selected cities have their own data portal initiat-ives, with several available data sets, accessibility barriers were mentioned over and over again by the data user communities. The most mentioned obstacles were the nonexistent or difficult way to download data for users that need full access to make a local analysis Section 4.2. As well as the low relevance of the developers' resources for re-users that need to link the published data in external applications. However, the URL access was not the only concern; in cases where the API resources were included, the lack of documentation and guidelines to use was also cited by re-users. Ultimately, there was a lack of datasets with specific geographic component (e.g., *air quality*, *local mobility*, *education*, and *urbanization*) that was not accessible

through current cities open data portals.

- **Data Quality**: This category is a large topic and was mentioned by the literature review (see Section 4.2) and is included in the empirical analysis that this research carries out. However, the criteria of data quality from a data user point of view could be more specific. Based on the findings of this study, the lack of metadata (especially for geographic data) was one of the major barriers mentioned by re-users. Attribute-inconsistent or gaps in published data is also a relevant feature to improve. According to data users, the possibility of predicting published data that are not complete or data which has specific characteristics (e.g., local reference system) might help them to save time. Generalization of data was cited for many users when they found relevant data which was not appropriate for local analysis or development. As an example of this issue, users mentioned an environmental use case that could be considers as an accessibility issue—the air quality data found in most of the selected cities have a regional or national scale. As another example data users in Valencia mentioned that education rates were published only in a regional or national scale which not contribute at all to analyze the local issues. Once cities become involved in open data initiatives, they need to consider extracting, processing, and integrating the correct information for the city's needs, not only integrating any open data from several national or local departments with any local propose.

- **Usability**: Further barriers—especially in the participatory workshops—were related to the lack of reuse examples. Many city portals limit their actions to publishing data, but there are no examples or use cases that users can use as a guideline to understand how the data is applied or how it could be integrated with other applications. Based on the data user's opinions, *many open data portals are a vast list of data, but there is no context to understand how data could be relevant to the city.* Likewise, besides the data category, there is no relationship among the available services. This lack of context creates a misunderstanding of data and misuses about how data can be applied or reused.

- **Discoverability**: During this stage we have identified that although all selected cities have an ongoing open data project, when users need to find the required data they search in several websites but not in the local open data

initiative. Using search engines (e.g., Google, Yahoo, Bing) or in the best case the open data national initiative websites, when users were asked to find specific data such as bike routes in their city, they encountered several issues in obtaining the required data. In some occasions, users went to the data authorities' website to find the current open data initiative, but most of them did not have the expected emphasis on the initiative. It seems that the lack of open data centralization could be a relevant usability barrier from data users' point of view. Another mentioned obstacle was the low integration between city departments regarding the data release process—especially in Cali and València. Data users claimed that the existence of several city department websites—sometimes all of them offering a different kind of data about the same topic—could confuse and reduce the reliability of the releasing process. This minor integration could result in a significant amount of time required to find relevant or useful data. The integration of current local SDIs with local open data initiatives mentioned in section 2.5 provides an opportunity to move forward in terms of integration with several city departments, using geographic data as a starting point.

- **Terms of Use**: The least-pronounced but still a common category barrier among three data sources used in this research was legal and policy concerns. Many data user communities manifested a significant misunderstanding of the terms of use or reuse of available data. Most of the open data policies around cities depend on national legal implementation; many countries have been involved in their own open data policy, and the transition to the local level could affect the way that the published data is being reused. Currently, to have a successful national open data initiative, cities have a determinant role to play in this value chain Carrara et al. (2017a). Having a consistent, clear, and integrated open data policy could attend to re-users to understand what kind of use is allowed and how they should include the published data in their external process or applications. Regarding terms of use in cities' open data, portals are not clear and easy to read, and the reliability to reuse could be affected. As was mentioned by Beno et al. (2017), potential users may feel misled when they find that available data have legal restrictions. Some entrepreneurs in the participatory workshop in Bogotá referred to the need to include whether commercial use is included or not to avoid future legal issues. This research notes that many of the terms of use

available in cities' open data portals are related to websites or portals rather than data per se. Having specific terms of reuse and use for published data might avoid any misunderstanding.
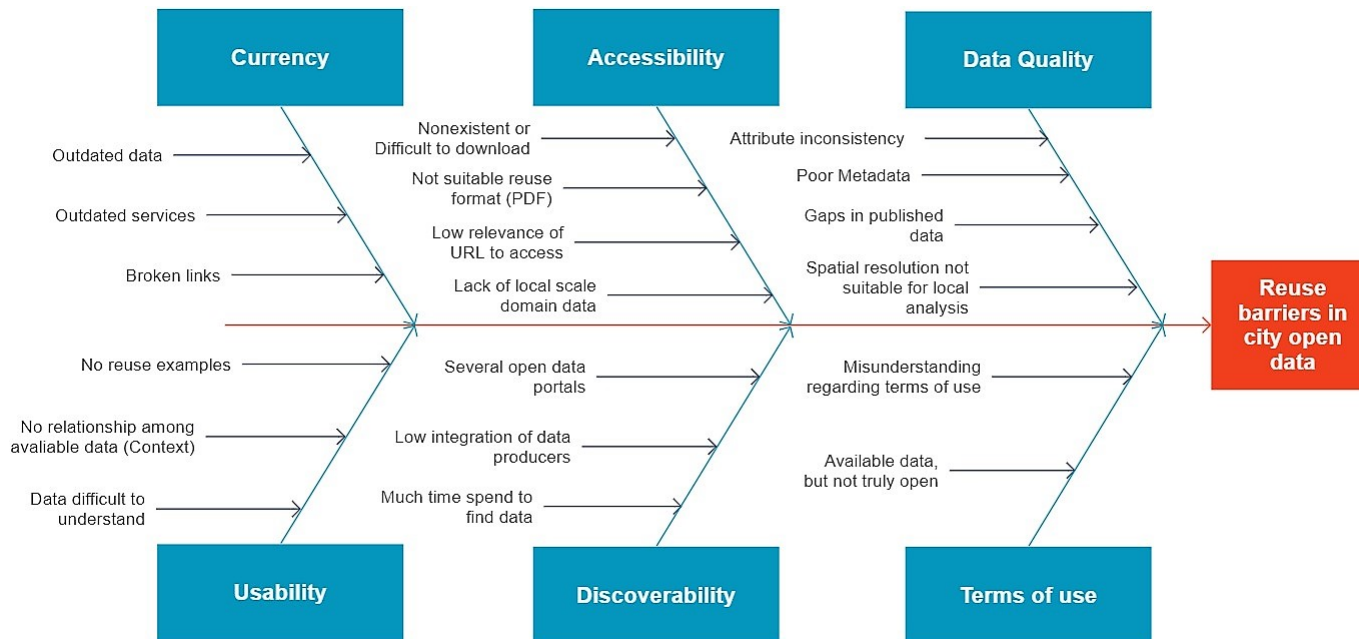
Figure 4.3: Fishbone diagram of barriers identified from a data user point of view.

4.6 | Recommendations

Based on the findings of this stage and the data users' opinions collected during the participatory workshops, there are suggestions that local data authorities can apply to integrate their data user groups into current open data initiatives.

Identifying data user groups: We have noted that most of local authorities need to clear identification of their data user groups. Some universities have been contacted and they were included in activities (e.g., hackathons or workshops) along the open data strategy. Other cities (local data authorities) have identified software development companies or organizations that continuously work with open data; however, the identification of those users, their needs, or their requirements are not part of the strategy. All data authorities that we included have an interest in engaging more users and adjust their strategy to data user requirements (see Table 3.4). However the inadequate identification of their needs and kind of data they are demanding, resulting in an inadequate or nor effective strategy. Data users need to be integrated during the whole open data initiative, not just included in the last stage of the strategy. The current research has listed barriers which inform data authorities about aspects to focus on while working towards the higher integration of users' wishes in their strategies.

Continuous services tracking: Other suggestions that were cited in the literature Conradie and Choenni (2014); Zuiderwijk et al. (2014) is related to the analysis and continued tracking of the available services. The accessibility and data quality concerns mentioned by data users in participatory workshops (see Table 4.5) might be addressed by a comprehensive understanding of the most requested services. What services users require to download, and what services need more accurate and complete metadata were questions mentioned by users that unfortunately local data authorities were not able to reply. Overall, this data users' suggestion is all about a continuous tracking of services, helping data authorities to yield an improvement of the published services.

Notification of further released data: In cities such as Valencia and Bogotá, data users have mentioned the need to know through an automatic way what services or new data have been released in the portal of their interest. Data producers can put more efforts into including notifications or alerts regarding the state of available services—especially services that have been identified as the most frequently used (another reason to include a services tracking). At the same

time, syndications like Really Simple Syndication (RSS) can also be used for future services or data that will be included as part of open city data.

Clear and straightforward terms of use or license: During the survey and set of workshops the terms of use were mentioned for data users as one of the obstacles to reusing the current data in cities. This barrier reduced the reliability of the open data in the selected cities. We consider that creating a simple and specific set of terms with natural language will help to reduce any misunderstanding regarding the utilization allowed of the available data.

More examples or basic reuse kit: Regarding usability barriers, data users cited that the lack of examples and basic guidelines to use/reuse available data have a negative impact on the reusability. Most of data users in our participatory workshops took an important among of time to understand initially where the data was published, secondly, once the data was found (if that was the case), most of the users were not able to understand how to reuse. Even some users with high experience, were facing issues due to the poorness of the metadata or a tool to help them to visualize the data before download. Regardless of the data portal and the tools to explore the available data, in the four use cases included in this research, users complaint over the lack of examples/guides to get an initial insight about why the data was released and how users can properly use them. Creating guidelines (as suggested by Degbelo et al. (2016a)) to reuse the basic technical elements of the released data as part of the local open data initiative might have a positive effect and reduce the misunderstandings of published data. This research suggests that data producers and the ongoing open data strategies should not limit their work to the provision of an extended list of available datasets through data catalogs. Creating a "basic kit to promote the reuse" that includes, for instance, a guideline to downloading, connecting, enriching, and displaying the released data. Explaining how to read the included metadata, why the data was relevant to publish, and possible uses can help newcomers and other reusers to get better insights into the released data.

4.7 | **Summary**

During this chapter, we have identified and validated several barriers that prevent data users take full advantage of released data in a city. In three different ways, we determined, what other authors have reviewed or disclosed through a comprehensive literature review (see Section 4.2). We found that mostly the work done can be grouped by barriers considering the data producer perspective but just a few authors are investigating the obstacles that data users have to deal with. We found that Standardization, Accessibility, and Discoverability are barriers that scholars have identified considering the data user perception (see Table 4.1). Throughout an online survey, we wanted to validate the identified barriers and also find other obstacles that literature has not considered but local data users have described as an obstacle (see Section 4.3). With 195 valid responses, we identified, Currency (seen as "recency", "freshness" or "currentness" of released data), Usability, Data Quality as most mentioned barriers along the responses of our sample (See table 4.4).

Finally, in order to verify what people say against what people do, we organized a set of participatory workshops (see Section 4.4). Selecting four cities were our online survey got more participation, we contact the local data authorities or the had of the ongoing open data project in Bogotá, Medellín, Cali in Colombia, and Valencia in Spain. Open Data for Open Cities - OD4OC was the name of our workshops where the discussion around reusability and current barriers to reuse the published data at local level was the keystone. Collecting data from two technical sessions, data users were asked to looking for data and afterwards using the identified data users were asked to include the data in some practical exercises. Having a comprehensive discussion around the "easiness" and resusability of the requested data, we have complemented the identified barriers and collecting some recommendations to overcome the obstacles from the participants. A taxonomy of barriers is designed and presented in a fish-bone diagram (see Section 4.5) to understand what were the most mentioned and validated obstacles from data users point of view considering the released data at local level.

The aforementioned taxonomy could be a useful tool for those local data authorities that currently are looking for a bottom-up approach to engage data users but the lack of knowledge about what are the specific users' requirements do not allow them to follow the right path. Identifying what are the most painful

obstacles and the categories where they need to shed some additional light, help data authorities to improve the efforts of their ongoing initiatives towards a more user-center strategy. Considering challenges like usability, discoverability, terms of reuse or data quality but tackled from the point of view of users will provide a new level of communication between the data producer and the data consumer. This taxonomy is an attempt to not only describe the current issues considering only the leaders of open data initiatives, our contribution is an attempt to provide a description of the most mentioned issues and grouped by categories to present some recommendations (see Section 4.6) also from real data users that daily required reuse the released data in cities.

5 | Framework to improve the re-usability of open geodata

> Metadata should help users to assess the usefulness of a dataset relative to their problem.
>
> *Alexis Comber*

In the previous chapter, we described the obstacles hindering the reuse of open geodata in four cities. Through the taxonomy of barriers, we described the most commonly mentioned issues facing data users in the ongoing open-data initiatives. The categories and the barriers described in Chapter 4 were based on the research paper by Benitez-Paez et al. (2017). A better understanding of what obstacles users encounter when the reuse of open data is required can be beneficial for local data authorities. Currently, authorities like IDECA in Bogotá or Valencia city hall are leading the open-data campaign in their respective cities. However, they are actively looking to include more users in their open-data initiatives, in view of the fact that the top-bottom approach has not been entirely successful in the last few years.

Although the description of the barriers faced by data users is a useful tool for local open-data publishers, this research seeks to propose a user-centric way to overcome the identified issues and to help with specific actions also suggested by their data users during the participatory workshops. This chapter is devoted to explaining the user-centric framework and its elements. This framework has two parts: the first part is defined as 'Impact Enablers', which are three resources that local open-data authorities can use to improve the expected impact of the open-data implementation. These resources take into account what other authors

have suggested and have been validated by data users in the four use cases. Local level (city level), Data-user communities, and Geographic approach were included as the initial considerations that data publishers need to incorporate before taking any further action toward achieving a bottom-up approach (see Section 5.2).

Four concrete and connected actions suggested by data users in the four cities to improve the reusability of open geodata in order to develop a bottom-up strategy are presented as the second part of this framework. They are described in Section 5.2, as follows: 1) User-focused metadata, 2) Community of reuse, 3) Identification of data users and their open-data demands, and lastly, 4) Reuse-focused legal terms. The four actions included in this user-centric framework were validated with data users and open-data experts at two events, with different approaches. The first one was the Open Data for Open Cities workshop held in Valencia, attended by over 20 data users. The second event was the full-day pre-conference workshop at the last AGILE Conference in Lund, Sweden. With participants including open-data experts from EDP, Esri R&D, geospatial master's degree students, PhD candidates and professors, the discussion revolved around how the framework should be implemented, how data publishers could integrate the framework elements within their ongoing strategy, and what role citizens play in a bottom-up strategy. This chapter is based on the recent research paper published in the journal Transactions in GIS (Benitez-Paez et al., 2018).

5.1 │ Overview

Many local governments around the world have adopted the Open Data principles. Their major concern has been to establish and populate open-data portals with data related to different topics and themes Attard et al. (2015). As we mentioned above in section 1.1, for many years, the critical issue in the success of open-data portals has been their accessibility and the data they contain. A number of technological solutions have been developed by national and local public bodies to create portals and to populate them rapidly. This has been driven by local and national regulations to guarantee open-data release processes Zuiderwijk and Janssen (2014).

This research embraces the reusability issue that many data authorities are still struggling with years after implementing open-data portals. Similarly, after several years of implementation, the communication and collaboration between data users

and authorities has not progressed as well as expected. On the one hand, users are still requesting useful data or have lost confidence in the portal, perceiving that the data provided is not suitable for their needs. On the other hand, data authorities are continually trying to empower their users and to increase levels of data reuse.

Nevertheless, some attempts have been made to measure the impact based on compilations of use cases (Young and Verhulst, 2016) and qualitative analyses, but the generalizability of such impact indicators is limited due either to the environment studied or to the specificities of each case, which results in impact measures that only apply when the conditions of the local case are replicated.

Notwithstanding, reusability (defined as the level of reuse of released data in an open-data initiative) has another role to play and has started to be added as a part of many open-data agendas and OGD initiatives; some research agendas are also seeking to gain new insights that enable them to better understand how open data can be made a more effective support for data users. In order to avoid a possible bias regarding the specificities of one use case, we have selected four cities and four different local open-data authorities. In Bogotá and Cali, the local SDIs were selected, but in Medellín and Valencia we contacted the open-data department directly. Having two different perspectives regarding levels of usefulness and the way that data is shared helps this research by providing different points of view about how the open-data strategy needs to be shaped.

Considering what we found in the literature review (see section 3.2) and also what data users have mentioned in the online survey and participatory workshops, there is clearly a need to actively empower more data users. Likewise, the data publishers need to prove themselves to an increasing number of users of the open-data portal and to offer them the expected results. In the four use cases, we have seen the following features regarding the ongoing open-data projects: 1) All four use cases are involved in a national open-data initiative, which means that national administrations both in Spain and in Colombia have an open-data project at national level. 2) The four use cases have also been working recently on local open-data projects (for not more than 5 years), with the local open-data portal as the primary mechanism to engage data users and promote an open-data catalog. In the four cities used in this research, the integration between national level and local level was mentioned by data users as a poor integration, resulting in a great deal of confusion related to what initiative they need to consider or what kind of

data they can find. 3) Each of the four cities considered has a local open-data portal included in a local open-data strategy. However, we found the same pattern of implementation, in which the aim of the open-data initiative was limited to having a web portal with an open-data catalog. 4) In the four cities explored, data authorities are looking to engage more users, and are also interested in improving the level of reuse. They are aware of this challenge, although we could not see any specific strategy to improve the level of reusability. Limited to just events or workshops with users, data authorities are promoting the open-data portal using traditional means of promotion, and users have the feeling that the portal created does not represent what they need. The four findings above raise the need to have a consistent and user-centric framework to improve the level of reuse of data by combining what data users are having problems with and also including the kinds of data that are most commonly required by data users, namely, geographic data.

In the sections that follow, a user-centric framework with four specific actions to help data authorities to improve the level of reusability of open geodata is described. The four steps and the way to implement the framework is also included. This proposed framework is based on what data users in the four uses cases suggested and the current status of reusability described in section 3.3. Three elements are introduced as the initial steps that local data authorities need to include if they wish to move toward a bottom-up approach in their open-data initiative. These three elements are defined as "Impact Enablers", due to their potential to improve the way that users can reuse the released open data. The impact enablers suggest what data authorities need to change in the ongoing strategy before actually undertaking any action to engage data users and promote additional actions to be carried out by data producers. Local level, Data users' requirements, and Geographic data are presented as the initial considerations for further actions toward improving the level of reuse of open data at a city level. Each impact enabler will be explained in section 5.2. Now that we have a proper background that includes actions to improve the level of reusability in the open-data strategy, four elements are then presented as the framework components. These elements need to work together and in a combined manner in order to achieve better results. According to what users presented as the most relevant barriers of current open-data portals and what they have also suggested, User-focused metadata, Community of Reuse, Identification of data users and their data demands, and finally reuse-focused legal terms are introduced as the specific actions required to move toward a better level

of reuse in released open geodata. The four elements outlined in section 5.3 are connected elements that data authorities can apply to accomplish better results across all the elements in the framework. For instance, by using the results and data collected in the Community of Reuse element, the User-focused metadata will have the required information to be able to match the released metadata to what users are demanding. Having identified the data-user communities and the way that they are using the released data can be used in the reuse-focused legal terms to fine-tune the color-based schema and provide more user-centric legal terms, thereby reducing the confusion about how to reuse the data properly.

This conceptual framework is the result of a better identification of the main barriers described in the last chapter and also a set of recommendations from data users from our use cases. By so doing, the conceptual framework provides a "bridge" between data authorities and users to support the release of relevant data, to build citizens' trust, and to support the easy reuse of geodata.

5.2 | Impact Enablers

Open-data impact assessment is considered one of the most challenging tasks in the implementation of open-data projects (Org, 2015). Although some studies have sought to develop critical points of view for rigorous analysis, many evaluations of open-data reuse and potential impacts have been limited, sometimes by the number of use cases or on other occasions due to the methodology applied. In this section, we present the three 'Impact Enablers' to improve the level of reusability of open geographic data which underpin the user-centric framework presented in this research.

To help local data authorities to establish an improved open-data strategy that embraces the reusability issue and promotes the initial integration of more data users, we found three initial considerations that most of the local data authorities have access to and can integrate within the ongoing strategies. In section 2.4 some examples of measures of the impact of open data at the national level are described, but in general we could not find any concrete user-based frameworks for improving the level of reuse of released geodata. Although the impact is considered in the literature, few authors mention the important role of improving reusability and including data-users' requirements as the core of further open-data strategies.

We present the Impact Enablers as the initial actions that any local open-data authorities need to consider if they are looking to implement a bottom-up approach in their ongoing open-data strategy. We found three elements that have the potential to improve the level of reuse before carrying out any dissemination activities or including more data in the data catalog. According to the literature (see Table 5.1) and also the opinion of data users in our use cases, a local open-data initiative that is well-integrated with the national efforts has more opportunities to engage and empower data users, especially those that do not have a lot of experience with open-data viewers or tools. Likewise, better integration with other local data authorities that have been working on the standardization and legal framework of geographic data (SDI initiatives) for some years can contribute to a more precise scope that includes previous issues and challenges in the local administration. The inclusion of data-user communities, but especially their data requirements, arises as the main component that most of the local open-data initiatives have not considered adequately, and it was found to be the weakest part across the four use cases in this research. Lastly, the important role of geographic data and its potential to provide new insights into released data needs to be recognized, and users are therefore demanding this sort of data. The impact enablers presented in this section are the initial actions in the user-centric framework, which guide local data authorities in the three basic elements that need to be considered, before modifying or including any new features or capabilities in current open-data portals. Local open-data leaders can apply these Impact enablers once reusability becomes a priority in the open-data strategy.

5.2.1 Local Level

Based on the literature reviewed (see Table 5.1), local open-data initiatives present a better scenario to empower more data users, to promote institutional integration, and overall to see the expected benefits of open data in the medium term. First, cities can reach data-user communities more easily and efficiently. Second, cities frequently work on issues that affect the daily lives of citizens (e.g., mobility, security, pollution, urban planning, economic development, among others). Third, cities provide the economic and innovation driving force for many national governments, and act as stakeholders of policy transformations through their innovations (The European Data Portal, 2016). Currently, there are many

'innovation-labs' in different cities, and data users see these local policies as a way to get institutional data and to avoid bureaucratic procedures. They provide a modern environment, with fewer tax charges, more technical and administrative teaching activities, and the promise to rapidly develop the proposed business model or expected social contribution. Such start-up incubators are becoming the way to deal with the current challenges, transforming the traditional economic model of many cities. During this research, two of these 'innovation-labs' took part in our workshops, due to their leading role in the local open-data initiatives. In Medellín, *Ruta-N*[1] is empowering academia, entrepreneurs, and the industry sector to work collaboratively and take Medellin into a new economic model based on the city's priorities and the services that new entrepreneurs can provide. In Spain, there are many crowdsourcing and co-working spaces located throughout its cities. In Valencia, through *Las Naves*[2] entrepreneurs, academia, and civil organizations are able to use modern facilities supported by the local government (Valencia city hall) to access lectures in current technological trends, engage in co-working, attend participatory workshops, and work toward developing innovative projects which, in this case, have a social component that must provide benefits to the inhabitants of Valencia.

In Bogotá and Cali, however, the local SDI are leading the open-data initiatives. Taking a geographic approach and having experience in the integration/standardization of spatial data from several local departments (e.g, mobility, education, culture, security, etc.) can contribute to the ongoing local open-data strategy. According to Ubaldi (2013); The European Data Portal (2016), smart-city projects and open-data initiatives at local levels increase citizen engagement and collaboration. To validate these claims, this research used the question illustrated in section 3.4 to determine what sector most data users are currently working in. We found that Local government (27.2 %), Private sector (22.1 %), and Education (19 %) are the top three sectors in a sample of 195 valid responses (see Figure 3.4). These findings reinforce the need to focus on the local open-data initiatives, while also establishing that most of the data users included in this research currently work at a local level. By using the participatory workshops, we sought to determine whether data users are familiar with the open-data challenges and with the local open-data portals provided by the local administration.

---

[1]   Ruta N, available online: https://www.rutanmedellin.org/en/, accessed on 27 August, 2018
[2]   Las Naves, available online: https://www.lasnaves.com, accessed on 27 August, 2018

Table 5.1: Authors that have mentioned local level, data users, and geographic approach to obtain a better impact in open-data initiatives.

Impact Enabler	Literature
Local Level	Carrasco and Sobrepere (2015), Benitez-Paez et al. (2017), Kassen (2013), The European Data Portal (2016)
Data-User Communities	Kassen (2013), Janssen and Zuiderwijk (2014), Johnson et al. (2017) Benitez-Paez et al. (2017), Williamson and Eisen (2016), Zuiderwijk et al. (2014), van Schalkwyk et al. (2016),
Geographic Approach	European Commission (2004), Welle Donker and van Loenen (2017), Reynard (2018), Van Loenen et al. (2017), Budhathoki et al. (2008), Young and Verhulst (2016)

5.2.2 Data-User Communities

This impact enabler covers more especially the requirements of data-user communities. Our approach uses a bottom-up schema (see Figure 1.2), where the needs of data users are the focus of any open-data implementation. Having an effective feedback system helps data producers to gather information about possible reuse issues and provides an effective way to collect opinions or ratings from data users. A "bridge" between data-user demands and the data providers' perspective could make a relevant difference to the services published in the initiative. With a diverse group of data users (e.g., journalists, developers, entrepreneurs, analysts, academic sector, and social organizations) and their needs in terms of open data, publishers have the opportunity to adjust ongoing open-data initiatives. The integration of data-user communities or their data demands should be included within the whole strategy, taking into consideration the design, deployment, and implementation stages so as to have a positive impact on the downstream results (Martin and Foulonneau, 2013; Janssen et al., 2012; Young and Verhulst, 2016; The European Data Portal, 2016) and provide a mechanism for dissemination.

Another question from the online survey described in Chapter 3 that is used at this stage and explained in section 3.3 asked data users *What is your primary concern when using the geographic data available in the open-data portals of the city?* (see Appendix A.3). It was an intentionally open question and 159 valid answers were obtained. Outdated data was the most significant concern mentioned by respondents, followed by Confidence in data creation, Licensing issues, such as *limitations to reuse*, *not for commercial use* or *legal issues once*

the available data is included in third applications. Finally, 15 categories of further concerns were also cited by survey participants, including reference system, data quality, and adequate data descriptions about data creation. Table 5.2 illustrates the occurrences of the aforementioned concerns as indicated by the participants in the survey.

Overall, most of the concerns with more than six occurrences were about metadata and how to obtain information about open-data integration, and use or reuse in their projects. This reflects the relevance of this impact enabler and the reason for including it, namely, to have a better understanding of the main issues faced by users. A better way to promote metadata, especially metadata that can provide a context to guide users in how to use the published data, is included in the conceptual framework presented below, as a means to support data authorities in shaping their efforts toward meeting data users' priorities. As noted by Comber et al. (2008), *metadata and its specifications in standards should be expanded* to accommodate current *de facto* standards and data users' needs. Metadata should support use, and not only data discovery and cataloging. Additionally, the Community of Reuse element within the framework, described in the next section, proposes some tools to allow users to subscribe to any dataset, to score data quality, and to add comments regarding currency issues, which was reported by data users as the main issue in open-data catalogs.

Table 5.2: The number of times that concerns were mentioned by respondents on being asked the question about their primary concerns when using geographic data from city open-data portals. 1.26% did not mention any concerns.

| Concerns mentioned	Occurrences	Percentage
Outdated data	45	28.30%
Confidence (about how data was created or gathered)	23	14.47%
License issues	18	11.32%
Quality of data issues	17	10.69%
Reference system or scale issues	11	6.92%
Completeness	10	6.29%
Format issues	8	5.03%
Accessibility issues	6	3.77%
Usability issues	5	3.14%
Privacy issues	4	2.52%
Metadata issues	3	1.89%
Accuracy (data/metadata record correctly described)	3	1.89%
Technical issues	2	1.26%
Discoverability	1	0.63%
Interoperability	1	0.63%

5.2.3 Geographic Approach

This research has found that a geographic approach in local open-data initiatives has a positive impact on their efforts to improve the level of reusability. The geographic scope in this impact enabler refers to the spatial data released in the open-data portals. The geospatial domain has an important role to play in a reuse-focused open-data strategy due to the fact that it is the most frequently consulted by data users and the business sector, with a positive influence on the success of open-data initiatives according to Carrara et al. (2015). In Young and Verhulst (2016), three case studies out of 19 in the geospatial services sector were found to improve services and promote economic growth in Denmark, Great Britain, and the USA. Welle Donker and van Loenen (2017) claimed that open geographic data policies can provide useful advice for other OGD initiatives. We have also noted this in our case studies, where users were more reactive to geo-portals than traditional open-data portals. The European Commission (2004) established geospatial data as one of the five categories with the highest demand for reuse across the EU, and recommends making data reuse a priority.

In this study, data users were consulted about the level of importance of several aspects with regard to the reuse of open data in cities. The question used was: *Please state the level of importance of each option when using city open data*. The responses indicated that *Geographic information accessibility* (95.38%), *High-quality geographic information* (92.82%), and *Academic and research improvement* (69.23%) were the most important categories (see Figure 5.1). Regarding the third category, we can say that data users consider local open data to be a very important resource that supports better results and methods in both academia and research. Further reflection on previous results illustrates that despite the number of geo-portals and the increasing access to data through open-data portals, users are still asking for better accessibility of geographic data. In a similar way, the quality of geodata is considered a priority for open data, which suggests a modification to the sentence in most open-data projects *data for everyone* by replacing it with *Data for easy and effective reuse*. It is also worth noting that users were not too concerned about the economic benefits of open data, despite being a common topic in the literature.

Although geographic data might have a positive impact by increasing the level of reuse of data in cities, this does not mean that publishers of spatial data do

not have any reusability issues. Some cities have been making substantial efforts regarding the integration and standardization of geographic data for a number of years, through local SDIs. The creation of tools to display integrated data (e.g., geo-viewers) are common goals in SDIs, as are legal frameworks and data standards that promote the sharing process among local public bodies Carrara et al. (2017a). Local open-data projects can take advantage of ongoing SDI initiatives to overcome integration or standardization barriers, which were frequently identified by data users Schmidt et al. (2016). In the same direction, Johnson et al. (2017) concluded that the experience of SDI projects could help open-data researchers to gain a better understanding of how the value of both initiatives is generated. During the participatory workshops held as part of this research, we also noted that most of the users are not really familiar with geographic initiatives, despite having important data catalogs that, in most cases, are bigger than the data catalogs of recent open-data projects.

Figure 5.1: Levels of importance of different aspects of open data in cities. Accessibility of geographic data, data quality, and improvements for academia and research were mentioned as being very important. n=195.
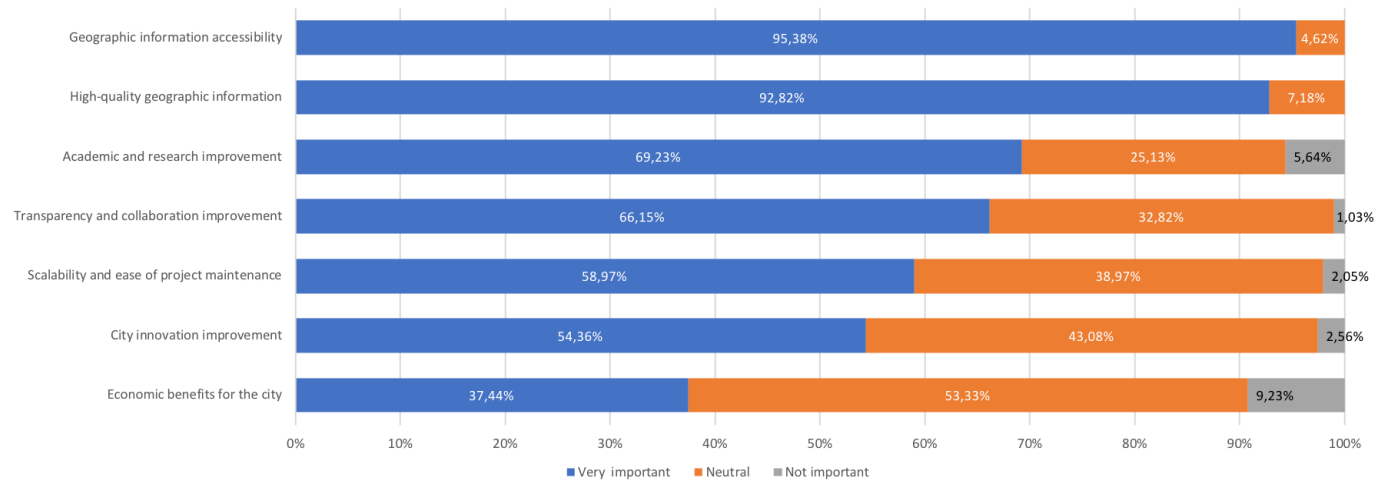
5.3 | Framework Elements

Once the impact enablers had been established and the local data authorities had the concepts so that subsequent actions could run a specific element, the next step was to select the elements that need to be included in order to make specific suggestions that any local data authority can apply to increase the level of reuse of open geodata. Through a set of participatory workshops described in section 3.4, this research has gathered the suggestions that users considered relevant to be able to overcome the barriers identified. Appendix A.3 illustrates the most frequently cited suggestions in each city where the workshops took place.

The most commonly mentioned suggestions throughout the workshops were the inclusion of concepts like linked data, semantic web, feedback tools, an improved open-data strategy that includes documentation and guidelines to reuse the available data, user-focused metadata, standardization or harmonization of data catalogs, and integrations among the city hall departments.

Based on the suggestions mentioned above (see Appendix A.3), plus recommendations from Benitez-Paez et al. (2017) to overcome the barriers currently hindering data users' reuse of open geographic data in cities, described in section 4.6, we have selected a list of recommendations (see Table 5.3). This list was validated and ranked during the participatory workshop held in Valencia in collaboration with the Transparency and Good Governance Office and 'Las Naves', the innovation lab mentioned above. With the participation of 19 data users, the aim was to select the top four elements that any open-data strategy should include in order to increase the level of reusability (see Table 5.4). The most voted suggestions were the following.

1. **Notifications system (F)**, the open-data portal should include the possibility that users might subscribe to a particular dataset in order to receive any updates to a dataset of interest to them.

2. With the same number of votes, **identification of data-user communities (A)** and the **inclusion of linked data (J)** were the second most voted options. Users claimed that having the possibility of seeing what other datasets are related to in each case is fundamental to be able to understand how the data could be reused.

3. The third place, in terms of number of votes, was for **improved documentation that includes guidelines on reusing the available data (H)**, and the **integration of geographic tools (I)** like a geo-viewer to display related data and overlaps with another dataset of interest.

4. **Redefinition of the terms of use (D)**, which is required to make it easier to read legal terms that include possible limitations for commercial reuse. Adding a traffic-light system to help data users rapidly understand whether the released data is completely reusable or has restrictions was another idea that was referred to quite frequently.

Table 5.3: List of recommendations from the data users' perspective, assessed during the last workshop in Valencia

	Recommendations by Valencian Data Users
A	Identification of the communities, groups, organizations, and departments of a city that are using open data, together with their needs.
B	Reuse community. Gallery of use cases, forum, recent datasets, tools allowing interaction among users.
C	Metadata oriented toward the user. Identification of the metadata most consulted by users and their deployment.
D	Terms of reuse. Color convention in order to know whether the data is reusable or has restrictions.
E	Continuous monitoring of the services that are requested in the open-data initiative.
F	Notification system, which alerts the user when a new dataset is included, improved or updated (e.g., RSS).
G	Gallery of examples where open data is used or deployed by the community.
H	Reuse guidelines or manuals, explaining how the published data can be reused or handled.
I	Geographic viewer, in order to display or overlay the published data.
J	Linked Data or Related data: The published data must have a geographic and statistical context. Possible relations with other published data.

Table 5.4: Votes by data users during the last participatory workshop. A notification system, identification of data users, better documentation and guidelines on reusing data, linked data, and easy-to-read legal terms using a traffic light system were the most voted recommendations. See the full list of recommendations used in Appendix A.3. n=19.

Recommendation	Priority				Votes	Percentage
	1	2	3	4		
F	1	6	4	3	14	73.7%
A	4	2	1	3	10	52.6%
J	4	1	1	4	10	52.6%
H	0	5	1	3	9	47.4%
I	2	0	6	1	9	47.4%
D	4	1	2	0	7	36.8%
C	2	3	0	1	6	31.6%
B	0	1	1	3	5	26.3%
G	1	0	2	1	4	21.1%
E	1	0	1	0	2	10.5%

Based on the impact enablers mentioned in section 5.1, a literature review (see Section 3.2, and Table 5.1), and the suggestions (see Table 5.3), concerns (see Table 5.2), and requirements (see Figure 5.1) from the open data-user communities in four case studies (see Section 3.4), this research has designed the conceptual framework presented in Figure 5.2. The framework elements are:

1. User-Focused Metadata

2. Community of Reuse

3. Identification of Data Users and Demands

4. Reuse-Focused Legal Terms

Figure 5.2: Components of User-centric Framework to Improve the Reusability of Open Geodata in Cities.

Framework elements

User-Focused Metadata

- Linked Data
- Attributes based on the data users needs
- Keywords

← Semantic →

Community of Reuse

- Efficient feedback resources
- Comments by data users
- Notifications system
- Data quality rating by users

Data Users Identification and Demand

- What is the users' data demand
- Who are the data users communities
- Use-case gallery

Reuse-Focused Legal Terms

- Color- based schema
- Easy to read at the glance
- Commercial issues included

Impact Enablers

| Local Level | Data User Communities | Geographic Approach |

The user-centric framework illustrated above aims to guide local data authorities in shaping their ongoing strategies to enhance the level of reusability of released data, taking into consideration the data users' needs collected during this research, and addressing the obstacles most frequently mentioned by authors in the literature but also validated during this study (see section 3.3). This framework presents a set of four elements to increase the usefulness of open data in cities, especially data with a spatial component.

There are two components presented in Figure 5.2 that should be emphasized. We found that most local data authorities have to isolate actions to engage more users in their open-data portals but, according to Table 3.4, the four local data authorities have no reusability strategy. However, they are aware of the current situation, and they want to improve the number of success stories using the released data. Most open-data authorities undertake dissemination activities, such as "datathons," "hackathons", and workshops with some data users, but we could not find any joint actions that include tackling the reusability issue from different angles. Data users in four different cities suggested the four proposed elements but, more importantly, this framework suggests a collaboration between them. The "roundabout" in the center of the figure represents the need to have connected activities that help data authorities to use the result in each framework element as an input for the rest of the elements. For instance, the outcomes from the Identification of Data Users and Demands component will help data authorities to know what kind of legal terms are required. By knowing what kind of data services data users are looking for or are trying to reuse, data authorities can include a more suitable "pool" of keywords for a more "natural" search activity. In addition, having correctly worded commercial legal terms can also contribute to having user-focused metadata that includes not only how the data was produced but also how the data can be used.

The second component to be included in the proposed framework was suggested by the data users in the participatory workshops, and also mentioned by Comber et al. (2008), who claimed that *there is more valuable geographic information embedded in the data semantics than in the released data itself*. The integration of the semantic web within the current open-data portals has the potential to offer the context and useful meaning that data users are looking for. Especially in Valencia and Bogotá, data users mentioned that most of the released data have no context. The semantic web is presented in the user-centric frame-

work as the option to link the User-Focused Metadata and Community of reuse components. Through the semantic web implemented inside the open-data portal, authorities will able to see what data users are requesting and, by using their comments, for instance, they can obtain the most recurrent issues concerning the datasets. The user-rated schema allows them to understand why some datasets are more comfortable for reuse and why others are not. They can observe what categories are more frequently requested by data users. By including the semantic web, they can also define a user-focused object catalog which lists the relationship between the datasets, thereby developing a linked data ontology. Here, the role of geographic data inside the reuse-focused open-data strategy has more impact by demonstrating the advantages that encoding topological relations among geographic datasets has over traditional spatial queries (Goodwin et al., 2008; Fonseca et al., 2002), thereby helping data users to find the spatial relationship among the released datasets and encoding them using a natural language in current open-data portals. Instead of having a technical support team reading the comments, the semantic web offers an innovative and up-to-date way of obtaining this information so as to make the open-data portal more user-centered.

In the following sections each component is described in depth to explain its objectives and how data users have suggested they could contribute to improving the reusability of open geographic data.

5.3.1   User-Focused Metadata

Comber et al. (2008) identified several ways in which metadata could be made more relevant to data users. Experimental metadata, expert opinions on relations with other datasets, and descriptions of the social-political context of data creation were some options put forward. Taking into consideration the ways mentioned above plus the suggestions listed by data users in the four use cases, in the last workshop in Valencia, this research asked participants what the most relevant metadata would be. After a group discussion, participants suggested ways to improve the current "metadata tab" in open-data portals, based on their experience. The implementation of related keywords in each dataset, adding comments by data users, possible uses by the data producer or publisher, and data quality ratings by users were the most popular improvements discussed.

Linked data and having information about how to exploit the data were persist-

ently mentioned by users during this study. Linked data is the way that released data could have a geographic or statistical context. Across the four case studies, only the Valencia open-data portal allows users to filter the available datasets by the openness score (5-star Open Data)[3]. However, there are no available datasets with five stars, and only 4.2% of the released dataset in this portal have a four-star rating. Though Medellín Open Data Portal allowed data users to explore some "related" datasets, which were in fact part of the same category inside the portal, no linked data were deployed, and users did not obtain any context until they downloaded the data and explored it for themselves.

IDECA, in Bogotá, understood that the way to promote the relation among several datasets is through Linked Data definitions. IDECA is currently working on the definition of a set of ontologies for the official basemap of Bogotá. At the same time, they are building a database of the semantic web for the geographic data of the reference basemap. The proposed framework has validated that data authorities need to start to work on the definition of linked data projects that include the development of ontologies, data vocabularies and the semantic web, as is currently being carried out by IDECA (Bogotá) or The Ontology Engineering Group (OEG) at the Universidad Politécnica de Madrid[4] (Spain). Including this component will have a remarkable impact on the way that users are using, understanding, and reusing the released data in cities.

Taking into consideration the definition of metadata cited in Comber et al. (2006), *information that helps the users assess the usefulness of a dataset relative to their problem*, this research proposes an integrated way to obtain that kind of information. Once open-data portals have implemented the feedback resources suggested in the previous element, users will provide a particular level of usefulness of the available dataset. Determining the number or kinds of uses for a specific dataset is quite difficult or almost impossible. Therefore, combining the possible uses from the data producer perspective plus the use case or comments provided by other users about their experience using the data, whether positive or negative, could be the way to understand whether the dataset is suitable for other particular problems. Related to linked data, there was also another resource suggested by data users: including the semantic web in open-data portals. The feedback provided by other

---

[3] 5-Star open data, accessible online: http://5stardata.info/en/, accessed 30 March, 2018

[4] The Ontology Engineering Group (OEG), accessible online: http://5stardata.info/en/, accessed on 2 April, 2018

users can guide newcomers in understanding how the data can fit their problem. That is also the reason why Figure 5.2 includes an arrow between the community of reuse and user-focused metadata elements. Semantic integration will cooperate to obtain the information required by the user-focused metadata, and users will be encouraged to participate by writing comments and rating the released datasets.

5.3.2   Community of Reuse

This element was a constant requirement and suggestion from all participants in the four case studies, especially in the last workshop. Valencian data users chose a notification system as the primary priority for any open-data strategy that seeks to increase the levels of reusability. This element is essentially a set of tools that allow data users to be part of the open-data strategy, and provide the local data authorities with information about how the data is being used. Such approaches are more efficient than a contact form to get feedback, comments, ratings, and suggestions, and allow users to ask other users about a specific dataset. A Subscribe tool like RSS was also suggested by users who want to know when some dataset is not available, has been changed or updated, especially in the case of developers who need to integrate the available dataset in external applications. Likewise, creating technological tools such as forums where users can explore, comment, highlight, and suggest success cases or relevant issues was often promoted.

5.3.3   Data Users Identification and Demand

Full identification of data users' requirements and their data demands was one the most voted recommendations by data users in the last participatory workshop in Valencia (see Table 5.4). Collecting what the data users' demands are, the most used web services, what communities are likely to use open data, and also how data users search for data in portals should be the initial task that any open-data strategy undertakes. We have noted that most of the local open-data authorities contacted do not consider the specific requirements of any particular user group. As summarized in Benitez-Paez et al. (2017) (see Table 3.4), cities like Bogotá, Cali, and Medellín in Colombia contacted developers to establish events and projects with local universities but, beyond that, we did not find any collaboration between data users and local authorities. In fact, in Valencia, the local open-data

technical office does not consider it relevant to collect information about who is using the released datasets.

A gallery of use cases, explaining how the available data is used by other users, was often requested during the participatory workshops in the four selected cities. Only in Valencia and Medellín were the local authorities able to describe two recent collaborations. In Medellín, local authorities started a group called "City Council Open Data" with the aim of describing the steps of the open-data strategy to data users and getting feedback from their side. In Valencia, the recent collaboration between a local university (Universitat Politècnica de València)[5] and the local open government office at the city hall established the project (Catedra de Gobierno Abierto)[6] to promote citizen participation, transparency, and open government.

5.3.4 Reuse-focused Legal Terms

One of the main barriers to reusing open data (Benitez-Paez et al., 2017; Carrara et al., 2015; Beno et al., 2017) is the lack of information about what kind of license each dataset has. There is a generalized lack of knowledge about the meaning of each license, or sometimes the available datasets do not have any associated license. In the latter case, the agency that makes the datasets publicly available must define the license under which that data is released, so that the rights of reuse are apparent to the potential users.

The proposed conceptual framework offers a solution to these barriers by establishing a color scale (green, yellow and red) to define each of the existing licenses. In this way, users can quickly determine whether they can use a specific dataset freely or if it has some restrictions. According to the findings of the final workshop, Valencian data users voted on their priority list (see Table 5.4) for the development of a traffic light system as an easy-to-read resource that allows users to understand whether the data is fully reusable, including commercial use, or if it has some limitations.

Starting with the definition of Open "*A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to*

---

[5] Universitat Politecnica de Valencia, accessible online: https://www.upv.es, accessed on 30 March, 2018

[6] Catedra de Gobierno Abierto, accessible online: http://catgo.webs.upv.es/, accessed on 30 March, 2018

attribute and/or share-alike.", there are licenses that fulfill the concept of Open data, namely: *Creative Commons Attribution*, *Creative Commons Attribution Share-alike*, *Creative Commons Zero*, *Open Data Commons Public Domain Dedication and License*, *Open Data Commons Attribution License* and *Open Data Commons Open Database License*. Table 5.5 shows the equivalence of the most important current licenses with the defined color scale. Green represents a reusable dataset, yellow represents reusable datasets but not for commercial use, and red represents a dataset that has restrictions on reuse. This schema does not replace the current license standards, but instead what we are proposing is an easy way to cluster the current licenses, especially in geographic data, where the number of terms of reuse could be an important challenge for open-data authorities.

Table 5.5: Proposed schema, based on a color scale for the licenses under consideration

Reuse Level	Color	License	Domain	BY	NC	ND	SA
Reusable	Green	CC BY	Content	Yes	No	No	No
	Green	CC Zero	Content and Data	N/A	No	No	No
	Green	ODC – BY	Data	Yes	No	No	No
	Green	ODC – PDDL	Data	N/A	No	No	No
Limitations	Yellow	CC BY-SA	Content	Yes	No	No	Yes
	Yellow	ODC – ODbL	Data	Yes	No	No	Yes
Restrictions	Red	CC BY-ND-*	Content	Yes	No	Yes	N/A
	Red	CC BY-NC-*	Content	Yes	Yes	No	N/A

5.4 | Validation

The validation of the elements mentioned above took place during two participatory workshops: first, with data users in one of our use cases, and then with open-data experts at an international conference. The last Open Data for Open Cities workshop that included data users was organized with the collaboration of Valencia city hall through its Transparency and Good Governance Office, attended by over 20 participants with different backgrounds and open-data experience (e.g., journalists, students, entrepreneurs, and local SDI officials).

Having the framework's elements and the impact enablers available to shape the ongoing open-data strategy, during the workshop data users validated the list of recommendations from the previous participatory workshops, (Benitez-Paez et al.,

2017) by rating what the top four tasks for any local administration or open data initiative should be. During a discussion about the reusability issue in Valencia and the potential of geographic data, participants voted according to their experience. Table 5.4 illustrates how the participants in Valencia voted, with a ranking of the recommendations to determine their top four priorities.

Although the title of this validation workshop was the same as the previous editions, the approach was different. The aim was not to collect how data users are looking for or reusing open geodata. Instead, the objective of this workshop was to collect all the opinions from participants, and to this end it staged a discussion about the proposed framework elements and the impact enablers, which can later be included within our main contribution.

The second validation activity that we carried out was another participatory workshop, but which was open only to open-data experts inside an academic event that involved master's and PhD students also working on open data. In the full-day pre-conference workshop at the International AGILE conference[7] in June 2018, the framework elements, including the impact enablers, were also part of the debate.

This validation workshop was held with people who have years of experience in the implementation or analysis of open-data initiatives around the world. Two experts were particularly important in this validation, and shared their insights and validated the framework components. One was Radu Cosmina, who is currently working at the European level assisting the European Commission in the implementation and monitoring of its digital transformation and open-data policies in the Member States. As part of the European Data Portal project, Cosmina is in charge of the annual European Commission Flagship Benchmark measuring the Open Data Maturity in Europe. The second expert was Andrew Turner, who is currently Director and Chief Technical Officer (CTO) of ESRI's Research and Development Center in Washington, DC. His work focuses on cross-domain collaboration and democratizing the process of map-making, and creating open tools for cartography and analysis. His team develops new technologies for government and citizen collaboration to build communities. This includes the global ArcGIS Open Data network for open access to authoritative data, as well as open-source and interactive tools and applications to build and share insights and

---

[7] AGILE conference, available online: https://agile-online.org, accessed on 28 August, 2018

solutions.

In addition to the open-data leaders mentioned above, over 23 participants from several universities, including master's degree and PhD students, and professors, also took part in this workshop. The discussion was centered on the reusability of open geodata, spatial analysis, and the current challenges of open-data initiatives in cities. Additional points that had not been considered in previous workshops were discussed, including the role of data created by citizens, traditional tools to deploy open-data portals in cities with more emphasis on the data catalog than the way the data is presented, and which do not include at least a geo-viewer. The role of the 'Community of Reuse' was one of elements that participants highlighted the most, together with the need for more User-centered metadata and terms of reuse.

Full implementation of the proposed user-centric framework elements may take local open-data authorities a certain amount of time, although we are beginning to see some of our use cases applying some of those elements. In Medellín, the open-data project "MedData"[8] included a set of participatory workshops with data-user communities, in which the technologies to deploy the data portal, the data catalog, and the topics of interest were discussed with stakeholders from several cities in order to have an open-data portal that is more related to Medellín data users. The "MedData" project is currently being implemented: a Beta version is available, but it includes tools to provide a geographic and statistical context. "Medellín en Cifras" (Medellín in Figures) is an option that allows users to explore general topics such as mobility, demographics, and urban planning by means of a set of dashboards. Moreover, users have the possibility of running their own analytic views, and loading the available layers and exploring them with statistical tools. In Bogotá, the local SDI (IDECA) has recently been assigned to lead the open-data project in the city, with the aim of encouraging data users from different backgrounds and experience to use and reuse the current geoportal. IDECA has integrated and published geodata from more than 70 local departments in Bogotá. Initially it started with a strong land register emphasis, but it now has a geo-portal with data in more than 20 categories related to the city. After the participatory workshop in Bogotá, components like feedback for each dataset, and easy-to-read legal terms were adjusted to reduce the confusion and promote the

---

8   MedData, available online: http://medata.gov.co/, accessed on 2 October, 2018

participation of more data users in Bogotá. In Valencia, with the collaboration of the Transparency and Good Governance Office and the Universitat Politècnica de València (UPV) promoting the participatory workshops and involving users and "Civic hackers" from other approaches to the reusability issue, we performed two open-data promoting activities with cyclists and journalists to support the available data and obtain feedback about how the released data can be improved. Currently, the open-data project is under revision. However, the main issue that we found in this use case with regard to increasing the level of reuse of open geodata is the number of initiatives that data users have access to. These initiatives are led by several offices; the open-data project is led by Valencia city hall, as is the geo-portal but through a different office, and the local SDI is driven by another agency. Having a common approach to promote a single level of reusability, and increasing the data users' level of confidence was particularly complicated in this city. Data users recommended better integration and communication between local departments, geared toward establishing a single open-data strategy to combine the efforts of those local departments.

5.5 | Summary

In this chapter, we have introduced a set of elements for a user-centric framework to help local data authorities to know what kind of actions they need to consider in order to improve the level of reuse of open data. By using the elements of the framework, local data authorities can define a set of activities in ongoing open-data initiatives or enhance the capabilities of current open-data portals to include what data users are suggesting. Several users in this research mentioned the inclusion of "User-focused metadata" as one of the top four priorities. This means the number of attributes should be displayed based on what data users have specified as their current barriers. For example, currency was selected as one the most required attributes in released data. Therefore, open-data portals can display the data catalog filter by date of issue and creation, instead of using the category of data or the name. Including the date of creation and/or updating in each published dataset will also provide the information that current data users are looking for. A common example from local SDIs is to include a link to obtain extended metadata if data users require more specific information but this is poorly implemented in local open-data portals. Making a standard and consistent "pool" of keywords was

seen by data users as another method to create a "natural" way to find the data inside the portal. Here is where the web semantic and the community of reuse have an essential role to play in the implementation of the proposed framework. The community of reuse is the component that local authorities can use to gather what users are looking for, what kind of data they are requesting the most, but more important how they rated the dataset used or what topics are more relevant for data users. Through an active 'community of reuse' implementation, in which data users can rate each dataset and include comments or suggestions, data publishers can get accurate information to extend in a web semantic link for each dataset. An important remark from data users, especially those with experience in geographic data, was about the need for a linked data approach. Unfortunately, in our use cases, only Bogotá was working on the definition of an ontology that allows them to define linked services. However, the project is still in progress and users do not have the tools to explore any related data. In Valencia, although the open-data tool can allow users to filter data by level of openness, if users filter the datasets by five stars, there are no datasets with this level of openness.

In terms of the identification of data users, having a clear picture about what data users are requesting as regards the published dataset was found in our use cases as an aspect in need of improvement and can also be seen as the initial step toward the implementation of this user-centric framework. Although at the beginning of this research Valencia mentioned that it had no interest in collecting what users are requesting from the released datasets, during the last participatory workshops Valencia started to create a network of open-data users, including students, journalists, entrepreneurs, and developers, among others, and this list is helping them to create more user-focused tools, like the budget-viewer tool[9], which was recently made available in the open-data portal, and helps users to understand how the Valencia public budget is being spent.

In section 5.2 we introduced a set of 'Impact Enablers', labelled in this way due to their potential to develop a more sustainable and effective impact according to the previous work that was reviewed throughout this research. Data-user communities, Geographic approach, and Local level were found as important directions for successful open-data initiatives, and are the keystones of the proposed framework. In section 5.3, we described the four elements of our main contribution in this

---

[9] Valencia Budget Viewer, available online: http://gobiernoabierto.valencia.es/es/, accessed on 2 October, 2018

research, namely, User-focused Metadata, Community of Reuse, Identification of Data Users and Demands, and finally Reuse-Focused Legal Terms were suggested by data users in this research as the key elements to improve the level of reuse in the ongoing initiatives. Finally, in section 5.4 we explained how our contributions, which consider the taxonomy of barriers from the data users' side, our recommendations, and also the elements proposed in the framework were validated and discussed by other data users and open-data experts in recent participatory workshops in Valencia and at the AGILE Conference 2018.

6 | Discussions, Conclusions and Future work

Previous studies have addressed open-data initiatives at the national level as the scale with which to shape the possible benefits and obstacles hindering the implementation of open data when it is reused. Most of them have considered the process of publishing data and the opinion of the data producer as the core of any open-data system. This research presents another perspective by taking the data user's needs as the focus to improve the reuse of open geodata at a local level. The study includes three elements for an improved open-data perspective. These three elements are (1) geographic data, as the type of data most commonly requested by data users; (2) using the local level or city level as the way to improve communication between data users and publishers; and (3) the most relevant part of this research has been the role that data-user communities can play in current open-data strategies, which involves their needs to define further actions, functionalities, and datasets. Section 6.1 presents some remarks on the current status of the level of reuse in the four use cases, and discusses our findings in this regard. Section 6.2 summarizes the barriers mentioned by data users from four cities with open initiatives based on different approaches. Section 6.3 offers a number of comments on the role of local data-user communities and how data authorities are facing similar issues regarding licenses, the process of data-user identification, their needs, and current strategies to enhance users' engagement. Section 6.4 considers the impact enablers and framework elements presented in 5 and how this framework could be a useful tool for data authorities that are currently working on transforming their ongoing open-data strategy into a bottom-up approach that includes more user-centered tools and a growing level of reuse.

6.1 | Level of reuse in the use cases

Regarding the initial research question about the current status of open-data reuse on a local level (cities), after conducting this research we can claim that the level of reuse of many local open-data initiatives is still being implemented and in most cases it is only a concern of some leaders in local data authorities. Despite the relevant efforts of many public administrations to improve the engagement process with open-data portals that include wider data catalogs, data users still do not have a level of confidence that is high enough to reuse the published data. According to our findings in the four use cases (Bogotá, Cali, Medellín and Valencia), the implementation of the open-data portal is still underway although, in some cases, the open-data portal was launched several years ago. It seems that local open-data authorities are still seeking to gain a better understanding of how users could utilize the published data, and accurate data or knowledge about how data users are exploring or reusing the released data is also limited. Overall, the four cities are facing issues as regards the level of reusability, and the opinions from their users were clear after the participatory workshops: "They do not trust the published data in terms of its quality, currency and usability".

A positive remark after conducting this research and exploring four use cases in two countries is that we found that most local open-data authorities are aware of the challenge of reusability. This means that our workshops were arranged with the goal of reaching more users but, more importantly, of finding a new way to ensure that the open-data portal and the data catalog provided will be a useful tool for data users which they can trust and use for their projects. Once we had explained how important the bottom-up approach is and the relevance of data-user requirements, they were willing to listen and follow any suggestions, and even to reshape their ongoing open-data strategy, as was the case of Valencia city hall. Another point to mention regarding the current status of reusability was the need to include data-user communities not only when the open-data portal or the initiative is running. After we had explained the important role of data-user requirements, a set of meetings with several users' communities were held in Medellín and Valencia to explain what new versions of the open-data portal should be like, and a new set of requirements were outlined that could be taken into account to enrich future open-data portals. At the same time, the Bogotá and Cali data authorities found several points in our approach to work on once they had listened their users

and they are currently working to include the data users' suggestions.

6.2 | Summary of Barriers

As mentioned earlier, most of the barriers to open-data reuse from the literature were determined from the data producer's perspective. In Section 4.2, it was shown that most of the authors directed their efforts toward analyzing the possible benefits, adoption barriers, implementation limitations, and determinants affecting the chances of their data initiatives being successful or not. None of the references mentioned in that section considered the role geographic data could play in the strategies of local open-data initiatives to tackle OGD challenges. Likewise, the context used in the work illustrated that national efforts and the process of releasing data play an important role.

In Table 3.2, we have listed some work carried out taking the data user's viewpoint into consideration and presenting possible obstacles that could prevent full advantage being taken of open data: **Discoverability**, **Accessibility**, and **Standardization** being the categories that were less identified. Most of the barriers related to these categories were confirmed in our online survey (see Section 4.3). Furthermore, we extended the barriers mentioned above and found that **Currency**, **Usability**, and **Data quality** are additional, relevant concerns of data-user communities when open geodata is being searched or reused at a local level. These barriers were highlighted in the participatory workshops, where geographic data was the kind of data most frequently requested by users, although it was also the most criticized throughout the activity. Out of date web services, lack of or gaps in metadata, data available without any quality control, and a lack of standardization of the reference systems (some services even had their own custom reference systems) all contribute to make the task of reusing the data more difficult.

Most of the discussion in the literature is centered on accessibility issues, and indeed most of the official organizations at national or local levels take the data release process as the primary task. Data users are currently demanding not only accessibility Ubaldi (2013); Carrara et al. (2017b), but in fact want to go beyond access. According to our findings (presented in Section 4.3), a constant concern in data-user communities is the currency of published data. The "*lack of updated published data*" was selected as a significant burden for 68.04% of

the 195 participants in our survey. Furthermore, *"misinterpretation and misuse of data"* were also considered by data users as an obstacle to the efficient reuse of published data. Data catalogs with long lists of data without any statistical or geographical relationship or context may confuse data users and lead them to spend too much time searching for the relevant data.

Two of the eight OGD principles are related to the format that data is released in, and the way that data should be open to the public in a machine-readable format which is also non-proprietary. This research found that, for the sample considered (see Section 4.3), the shapefile as the most useful format (see Figure 4.2). The respondents considered typical geographic services like OGC services, KML, GeoJSON as more valuable than the promoted open-data formats like RDF. A possible explanation of this result could be the data ambiguity that exists in local open government initiatives, where formats like RDF have an inadequate description. Moreover, in the geographic community, they are not represented or used to a significant extent in the analysis process.

In Beno et al. (2017), the lack of harmonization between portals was considered a severe burden that makes data users confused about similar data available in different portals. This research has confirmed this finding and grouped it under the usability category "barriers" mentioned by data users (in all the selected cities), such as "*data difficult to understand*", "*no relationship among published data*", or "*no applications to validate the usability of available data*" (see Table 4.5). In Valencia, participants in the workshops placed special emphasis on improving data integration between local departments in the city. The mobility data found in the open-data portal was different from that found on the mobility department website or that found on the SDI website. In this case, users could not recognize what the official data was, due to the fact that the three portals came from public bodies. The quality of data is also a constant burden for data-user communities—in particular for the data users included in this research. Although this category has already been considered in the literature from a data-producer perspective, it is still an aspect for improving an open-data chain Carrara et al. (2016). According to our survey and workshop participants, having data with issues like "*no metadata*", "*published data not geo-referenced*", or "*not enough or clear metadata*" considerably reduces the reliability of the data source, and thus the effectiveness of the open-data initiative. Data which is not machine-readable (e.g., PDF) was another barrier mentioned in our workshops in the selected cities—especially in Cali and Medellín. At the

same time, issues like "*data only for visualization*" or "*no download option*" were mentioned by users that require full control over data for local analysis. As an example of this situation, in Valencia and Bogotá data about the different layers of the local roads were required to perform a mobility analysis. However, data users stated that there was no option to download such information and it could only be viewed through geo-portals.

To conclude, we revisit the research question presented in Section 1.2 (What barriers prevent the reuse of open geographic data by local data users?) and summarize the discussion in chapter 4 with the following observations. We identified and explored 19 barriers, sorted into six categories (see Section 4.5). We then identified the most commonly mentioned concerns and requirements from data users in the four cities—particularly those that work with open geographic data on a daily basis. Currency was the concern that was most frequently mentioned by data users from different backgrounds. Accessibility and Data quality were commonly mentioned during this research. Usability, Discoverability, and Terms of use were also included in this taxonomy of reuse barriers, while the low integration of city departments, misunderstanding of terms of use, and no geographical or statistical relationship were constant issues faced by data users in the selected cities.

6.3 | The Role of Cities and Their Data-User Communities

The open-data chain European Commision (2013) is presented by the European Commission in its strategy as an interaction between official departments and open-data stakeholders. Carrara et al. (2017b) illustrated how raw data is transformed into economic value considering the creation of data up to the aggregated services. At the same time, this report categorized the roles of open-data stakeholders into four types of actors: Suppliers and Aggregators in charge of the creation and aggregation process, and Developers and Enrichers performing analyses and a new set of data services or products. Ubaldi (2013) presented a similar scheme, but included one additional step, called "final data use", as the last stage in promoting the sustainability of the process of public data creation. Correspondingly, Ubaldi also suggested the identification of an "ecosystem of users" that responds

to specific user demands to promote the creation of value. In terms of open-data ecosystems, Zuiderwijk et al. (2014) proposed the essential elements of a multidimensional system where the feedback from data users is one of the key elements. Likewise, Janssen et al. (2012) suggest that open-data systems must consider the data users' feedback, mentioning that "*there is no insight into the users' perspective and users' needs*".

Data-user communities and their feedback are becoming more important in the current open-data value chain, but the geographic context in which those users are located is also important. Indeed, cities have a relevant role to play here. During this research, several cities were considered to compare the actions currently being implemented by local authorities in charge of leading the open-data initiative and their data-user communities. In Table 3.4, several aspects regarding open data in those cities are compared. The initial element concerned what open-data topic is of interest to them; all the cities have mentioned topics like mobility, urban planning, economic development, or security. Medellín mentioned that data on urban planning about a sustainable and smart-city strategy are of interest to them. This is an interesting claim. According to (Carrara et al., 2017a), open data could allow a smart-city initiative to be reinforced or implemented, as a more "connected" city, and the development of new services related to sensors around the city could result in an important amount of data that users can use to enhance the quality of life in the city.

In relation to terms of reuse, Medellín and Valencia have adopted a creative commons license (Attribution 4.0 International CC-BY 4.0) for their published data (see Table 3.4); however, it seems that this does not guarantee the prevention of misunderstandings from a data-user point of view (approximately 68% chose this as one of the major or moderate barriers in this city; see Figure 4.1). Cali does not have any defined open-data terms of use, but the local authority follows a coordination mechanism in force at the national level. Likewise, 46% of their respondents mentioned "the understanding of terms of use" as a barrier. Only Bogotá, which has a local authority in charge of the open-data initiative and is at the same time the Local SDI (IDECA), has its own license (IDECA License), acting as a kind of barrier for its users, who get confused when they need to understand what use or reuse is allowed. Fifty-two percent of the participants in our survey chose "*understanding terms of use*" as the third barrier hampering use of the open-data portal in Bogotá.

According to the open-data value chain European Commision (2013); Carrara et al. (2017a), developers play an important role in any open-data initiatives. At the same time, they have the skills to enrich the available services and transform them into new kinds of innovative services or applications that show the real potential of open data Attard et al. (2015). Thirty percent of our respondents were developers (see Figure 3.3). However, Valencia does not consider it relevant to collect information about the entities or organizations that have used the datasets that are available. In Medellín, Bogotá and Cali, the identification of those stakeholders is quite poor (see Table 3.4). Nonetheless, Valencia , Medellín, and Bogotá users chose the low relevance of URL access to data as a major obstacle, and in the set of workshops *"API documentation"* and *"JSON files with issues"*, among others, were some of the technical barriers that were most frequently mentioned.

Finally, the internal departments in each city were also compared in this research. It was found that the barrier "varying and low integration of data sources or data producers" was chosen as a major burden not only in Bogotá. In the other selected cities, this obstacle had an important percentage (68% for Medellín, 54% for Cali, and 53% for Valencia). A possible explanation of this result could be the work done by IDECA, which is a well-known authority among their data users (especially those who work with geographic data), and the integration of the spatial information of more than 70 local entities. Although Cali also has a local SDI, which was the authority contacted, this SDI is in an initial phase and Cali data users are still getting used to knowing what IDESC is doing and what kind of data it is publishing.

6.4 | Toward improving the level of reusability

In chapter 5 this research defines the elements and the basic concepts to help data authorities enhance the level of reuse of their current open-data initiatives and portals in a sustainable fashion. It considered open-data reusability and then proposed a user-centered framework. The focus of this research was on empowering data-user communities to participate and to be part of the ongoing open-data projects as their needs should be the driving force of any strategy. During this research we noted that despite many local efforts to promote the open-data movement through concepts like transparency, citizen participation, and new web portals, users still consider that a lot of work needs to be done and, importantly, that they do not feel that those projects are fulfilling their key role as

data intermediaries.

Using the findings from an online survey (see section 4.2) and a set of participatory workshops (see Section 4.3), as well as interviews with specific local open-data authorities and groups of users, four integrated elements were proposed in the framework. This framework is a guide to help data authorities to identify where they should focus their current efforts to shape ongoing initiatives, including what users require and the barriers preventing them from taking full advantage of available data. A sustainable open-data initiative is possible when the stakeholders are identified, when data accessibility is promoted, and finally when the levels of usefulness and reusability are included in the whole open-data picture.

We have presented several impact assessment methods or use cases. However, we did not find any existing frameworks that sought to improve the reusability and usefulness of open geographic data at the local level. Three impact enablers were defined and included as the basis for reshaping ongoing local open-data efforts. These promote a bottom-up approach that includes a geographic perspective and its potential to improve the reuse of the released data allowing collaboration between local departments.

6.5 | Conclusions

This thesis has focused mainly on issues concerning reusability of open geodata at a city level, and discusses some difficulties open-data authorities face when it comes to empowering and engaging data users to the expected degree in their current open-data initiative. This dissertation has considered data use requirements as the keystone in any open-data strategy and has sought to demonstrate why a better understanding of what data users are demanding regarding open data could be one of the most effective ways of reaching the expected benefits of open-data portals and their strategies.

This thesis began by looking at the current status of reusability of open data and how other authors have suggested measuring the impact of ongoing or past open-data implementations. Initially, one of the most important lessons learned was that most of the work carried out to date is focused mainly on the data producer's point of view. Overall, the data users' perspective is considered to a lesser extent, and possible recommendations to overcome the issues identified include a top-down approach. Accessibility issues are mentioned as one of the most important

challenges, although once the data is released open-data authorities are faced with reusability issues and a bottom-up approach seems to be required. Chapter 2 described the key concepts in which open government data is seen as one of the most relevant intersections between open data and public data. Nevertheless, a clear understanding of why some open-data initiatives have been successful and why others are still looking for the expected results has been the focus of attention in many studies. Measuring the impact of open-data initiatives has been challenging, due to the number of variables included and the differences in implementing each open-data initiative.

Several attempts are being made to measure the impact in Europe, led by the EDP, or international foundations like the OKF, in which they assess the level of maturity of a country or a whole region based on a comprehensive survey conducted with the heads of national open-data initiates. However, the current frameworks to measure the impact either do not consider the city level or the geographic approach is only briefly included. Chapter 3 included a description of the methodology used in this research. It started by considering how other authors have covered the reusability issue, and then went on to look at the barriers that prevent full advantage being taken of the open data that is released. It was found that Discoverability, Accessibility, and Data Quality are already issues that have been identified from the data-producer perspective and that some authors have considered data users' opinions in their studies. An extensive online survey was used to collect the initial perceptions of a large number of data users. However, what data users say is not necessarily what they do. Therefore, to confirm what they are actually doing regarding the reuse of open data in their cities, we organized a set of participatory workshops, in which they were presented with the problems and proposed some recommendations to overcome the validated barriers according to their experience.

Most of the frameworks, initiatives or projects designed to assess the impact and levels of maturity and progress have been developed for open-data projects at the national level (see Section 1.4), and we could not find any properly developed frameworks for assessing the impact of open data in cities. How open data is progressing on a city level was only briefly commented on in some projects. The EDP has isolated analytic reports in which some important remarks have been addressed to cities due to their potential to implement national open-data efforts. However, after the participatory workshops it became clear how data users are

looking for more integrated, consistent, and local efforts than national open-data initiatives. Once the city level has gained the data users' full trust and reusability has reached sustainable levels, national efforts regarding open data are likely to be able to reach new levels of confidence and reusability with the expected positive impact.

Throughout the use cases considered in this research, the level of reuse of released data was a important concern for local open-data authorities, although the actions to improve it were limited to some announcements on their website, events like hackathons, workshops to explore the data catalog or the recently launched open-data portal. However, the existence of comprehensive analyses of what data users are looking for and how data users are demanding data was scarce. In fact, after conducting the initial survey to determine the current status of the open-data initiative in each selected city (see Table 3.4), the city of Valencia mentioned that 'they do not see the collection of data, users or entities that are using their published data as being relevant'. In the other three cities, only a few universities and users were identified, but since this research began local authorities have not considered an initial identification of their users, the data demand, or the level of reuse as one of their priorities, and thus reusability was not part of their open-data initiatives either.

Conducting the Open Data for Open Cities workshops and the results from two research papers (Benitez-Paez et al. (2017, 2018)) help us to make the data authorities in the selected cities more aware of the importance of reusability and the important role of data users. From that point on, the entities that were contacted were able to realize that their data users are stuck on issues like accessibility, terms of use, standardization, usability, and above all on the barriers categorized in the proposed taxonomy of data users' barriers (see Section 4.5 in Chapter 4). The importance of a better identification of what data-user communities are and what data demands they have starts to become an important aspect to be included. In fact, the Valencia authority was one of the most interested in moving forward to this bottom-up approach, and that is why the validation workshop was organized with other Valencian data users. As a result, there were more specific topics to discuss and more users were involved in our discussion on reusability.

Currently the Transparency and Good Governance Office in the city hall of Valencia has a list of all the participants in our workshop and is holding other kinds of events to engage different communities with related open-data interests.

Likewise, the current strategy of open data is on the way to being restructured so as to include the level of reuse, the geographic approach, and data-user requirements as part of the core strategy following our results. Medellín city hall, which is working toward the launching of an improved open-data portal, has held several meetings with open-data communities to present the technology that the new portal will perhaps use, the kind of data to be release, or the data catalog to be published. Details of a more user-centered open-data initiative that includes universities, civil organizations, entrepreneurs, and other stakeholders were shared with data users last year in Medellín. The initiative is called 'MeData' and it has been socialized several times, including specific meetings with data-user communities that this research has included in the participatory workshop discussed in section 4.4

Regarding the importance of geographic data inside the local open-data initiatives, only those authorities with a geographical approach (Bogotá and Calí) were aware of the potential of geographic data to increase the impact of open-data initiatives. Currently the authority selected in Bogotá (IDECA) is leading the open-data project, which means that the whole experience in standardization and the legal framework to publish geospatial data from several city departments is now leading the open-data strategy in that city. This new approach, however, does not mean that IDECA and Bogotá open-data users are communicating with each other in the most effective way and that the level of reuse does not need to be improved. Through our OD4OC workshop, IDECA has identified that users have issues regarding legal terms, even after having developed their own license[1]. At the same time, our participatory workshops were useful to discuss the usability and currency issues in the Bogotá open-data project. User-focused metadata was another important remark and recommendation in Bogotá, included within the suggestions from data journalists that took part in our activity. During the coming years we will see how Bogotá might lead the new approach in local open data by considering geographic data as the core component of the open-data strategy, thus validating whether our approach and results lead to more confidence and positive impact at the local level. Cali was another of the selected cities that is leading the open-data movement in the city through the local SDI. After identifying data demand through our survey and workshop, the selected authority is leading a set of user-centered events focused on the potential of geographic data available

---

[1] IDECA license, available online: https://www.ideca.gov.co/es/licenciaabierta, accessed on 28 August 2018

in the open-data portal, and at the same time it is collecting information about all data users that are making daily use of the data catalog already available in Cali.

In Chapter 4 we described the most frequently mentioned barriers that data users throughout four use cases are facing to reuse local open data. Currency, which refers to 'recency' or 'freshness' of published data, was the most relevant of the concerns that do not allow data users to have full confidence in data catalogs. Data Quality is an issue already identified by some other authors, although mostly included from the data-producer perspective, which was also mentioned by data users in all the selected cities as one of the priorities to be solved. Released data that was published without any data quality control, especially in terms of metadata, results in data users not trusting the data catalog included in the open-data portals.

Although Chapter 4 is focused on the barriers and their classification, during the participatory workshop we also collected the most important and commonly mentioned recommendations from the data users' perspective. Our aim was to develop a new strategy that helps data authorities to improve the level of reuse of the open data that is released by considering what data users are demanding. Therefore the suggestions from the experiment and the way that users are tackling current issues define how we design the main contribution to this dissertation. In this respect, in Chapter 5 a conceptual user-centered framework was introduced to take a step forward in this direction. In accordance with the findings from literature, we suggested three 'Impact Enablers' as the key elements that any data authority can implement in order to find a sustainable open-data strategy. In addition to the impact enablers, four other elements that need to work together in the proposed conceptual framework make up the whole strategy that this research is presenting as its main contribution. Full identification of data users and their data demands was one of the most frequently lacking areas of knowledge in the four selected cities, in addition to the data demand and how the data catalog is being used. The four use cases have not recognized their users or what kind of uses the published data might be put to. Regarding communities of reuse, most of the ongoing open-data portals only considered accessibility as the main barrier to be overcome, but the way that data users are using it is often not included as part of the open-data initiatives. Our use cases have little contact with their reusers, and tools to obtain useful feedback are hardly ever included. Thus, a better feedback component and new tools that provide data users with a better way to communicate their needs to data authorities are required. There was no gallery of use cases, user-rated

systems, blogs or tools to write specific comments for a specific dataset. Only Valencia data users were able to like or unlike a specific dataset in the official open-data portal.

User-focused metadata was another element selected in this framework, due to the frequent request by data users to have adequate descriptions about what the released data is about and how it can be used or reused. Most of the users across the four cities have requested a semantic relationship in the published data catalogs. At the same time only Bogotá and Cali have a proper metadata catalog as well as a geo-viewer to assist data users in understanding what kind of spatial relationships the published data could have. Medellín and Valencia have taken a different approach, and users in those cities are requesting more integration with the geographic tools, and a comprehensive metadata catalog that guides them in finding data that can meet their needs. Finally, user-focused terms of reuse was also frequently mentioned by data users in the four cities and included in the survey with a large number of participants. Although many cities are following standard licenses like Creative Commons, misunderstandings about the proper rights to be applied in the reuse of the released data reduces the level of reuse and increases the lack of confidence, thus preventing the published data from being included in external applications.

This research has considered a number of ways to obtain an initial perception of the issues that data users are facing on the local level once they need to reuse the open data, although we found several barriers that are currently reducing the level of reuse of the ongoing open-data initiatives. We held four participatory workshops to validate what people do when they need to reuse the data but, more importantly, we created a small "bridge" where the data-users' requirements were the center of the discussion and data authorities could observe what kind of problems their users are facing. As part of our contributions, we have developed a user-centered framework that can be employed by any local data authority as a guideline to move toward a bottom-up approach that takes into consideration the requirements of other communities of users.

Limitations

In this research, we have identified a taxonomy of barriers (see Chapter 4) from a data users' point of view. We aimed to identify what obstacles data users in cities face when they are looking for data, but especially when they want to reuse and incorporate the data made available by a city into their projects, analyses, or external applications. In Section 2.5, we highlighted the consideration of geographic data in open and government initiatives due to its relevance to the reuse of available data according to the Reuse of Open Data report of the European Data Portal Carrara et al. (2017b). We gathered opinions, requirements, and barriers to the reuse of open data in cities through participatory workshops, and contacted over 100 people from different backgrounds.

Most of our respondents and participants had a geographic background or had worked with spatial data. Therefore, there are possible limitations that need to be acknowledged. A possible bias of the barriers that were identified could be that they are not applicable to other open-data users from other backgrounds such as journalists, analysts, or developers who work with any kind of data, but might be particularly interested in open data. The barriers identified here are based on the opinions of the respondents and the workshop participants in the four use cases, with an emphasis on including opinions in Spanish. We therefore recommend that these conclusions be taken with caution, as these barriers might not apply to other cities.

During this research, we found that there are some significant differences between the open-data initiatives led by data authorities in charge of spatial data integration in the city and entities that consider the open-data initiative as another project of Open Government. For example, Bogotá has IDECA, which is the data authority that is currently leading the open-data initiative, while at the same time it is the local SDI. In Medellín or Valencia, open-data leadership is the responsibility of city halls, such as the "Alcaldía de Medellín" or the "Ayuntamiento de Valencia". An explanation of the differences is not really possible at this point, since this requires information about the open-data agendas and working processes of the different institutions represented, which is not currently available. An extension of this study could thus investigate the actual interplay between the strategy of the local SDI/open-data initiatives and the way that data is being released, searched, and used.

In chapter 5 we described the framework elements that were validated in the final participatory workshop with over 19 participants, and hence there are some possible limitations that also need to be acknowledged. The recommendations from our sample might not necessarily be the same recommendations as those put forward by other data users in other cities. We recommend using the elements in this dissertation with caution, bearing in mind our sample and the fact that there could also be some variation in data-users' perception depending on the local environment and the kind of open-data portal that is operating in a particular city.

6.6 | Future Work

From a research point of view, we have analyzed several local open-data initiatives in two countries and found that in some cities the local authority that leads the open-data movement is the local SDI. This illustrates data-user engagement based on the SDI approach, where geographic data and standardization issues are the top-priority task. However, in cities where the open-data initiative is led by open government offices inside the city hall, the strategy and the way that data is released could have different impacts in data-user communities. On comparing current open-data strategies and SDI projects across cities, some similarities are observed between the two approaches. For instance, SDIs had to deal with standardization barriers in the past, and through geographic viewers they also wanted to tackle accessibility issues, while at the same time having to face barriers related to providing high quality data services. In this sense, we suggest that further research should explore the role of local SDIs in times of open data so that lessons learned from years of work on SDIs could flow into current open-data projects.

A roadmap for future research can be sketched out, as follows. Each component proposed by the conceptual framework will be implemented using a different CKAN extension and some of them will be reused from available extensions. By so doing, we define five different extensions:

- Data-User Communities: In this first component, an extension is proposed to categorize the users or reusers of the released datasets.

- Community of Reuse: Two extensions are expected to solve this component. The first of them is a Feedback extension. For this feature we will reuse an

extension called Disqus [2], which includes the possibility of enabling users to create comments and rate the data quality for each dataset inside the CKAN catalog. The main goal of this extension is to enrich each dataset with feedback from each user. The second extension is the Showcases feature. By using this Showcase extension, a CKAN that is already available, users can see what datasets are used in different sites.[3]

- User-Focused Metadata: By including a semantic web the aim of this extension is to allow data authorities to link data-users' comments with the metadata view available for the released data. Users will be able to filter the user-focused metadata to explore how the consulted dataset has been used by other users.

- Reuse-Focused Legal Terms: The last extension establishes a color legend to guide users in knowing whether the dataset is fully reusable or has a restriction. Data authorities will be able to easily choose the most suitable license from a dropdown list, and the extension will determine which color is chosen according to the license features.

The results of this research provide significantly more depth than the suggestions from previous research on this topic. Valencia City Hall and IDECA in Bogotá have already shown an interest in participating in the evaluation of the conceptual framework, as well as the functional tests that the proposed components require. A generic definition of open-data ontologies that integrate local geographic data for the most important datasets in cities is required as an extension of this research. Spain has been involved in similar efforts and has drawn up a list of 100 datasets that any cities require[4]. Another way to extend this research is to integrate those vocabularies into the data release process in a local open-data initiative following the conceptual framework presented here.

---

[2] Extensions CKAN, avaiable online http://extensions.ckan.org/extension/disqus, accessed on 17 October 2108

[3] Showcase of CKAN, avaliable online http://extensions.ckan.org/extension/showcase accessed on 17 October 2018

[4] Onpen City Data, avaiable online http://opencitydata.es/web/guest, accessed on 2 April 2018

Bibliography

(2015). «A comparison of national open data policies: lessons learned». *Transforming Government: People, Process and Policy*, **9(3)**. ISSN 1750-6166. doi: 10.1108/TG-03-2014-0008.

(2015). «Open Government: Origin, Development, and Conceptual Perspectives». *International Journal of Public Administration*, **38(5)**, pp. 381–396. ISSN 15324265. doi: 10.1080/01900692.2014.942735.

ABELLA, A.; ORTIZ-DE URBINA-CRIADO, M. and DE-PABLOS-HEREDERO, C. (2014). «Meloda, a metric to assess open data reuse | Meloda, métrica para evaluar la reutilización de datos abiertos». *Profesional de la Informacion*, **23(6)**, pp. 582–588. doi: 10.3145/epi.2014.nov.04.

AFFUL-DADZIE, ERIC and AFFUL-DADZIE, ANTHONY (2017). «Open Government Data in Africa: A preference elicitation analysis of media practitioners». *Government Information Quarterly*. ISSN 0740624X. doi: 10.1016/j.giq.2017.02.005.

AHMADI ZELETI, FATEMEH; OJO, ADEGBOYEGA and CURRY, EDWARD (2016). «Exploring the economic value of open government data». *Government Information Quarterly*, **33(3)**, pp. 535–551. ISSN 0740624X. doi: 10.1016/j.giq.2016.01.008.

ATTARD, JUDIE; ORLANDI, FABRIZIO; SCERRI, SIMON and AUER, SÖREN (2015). «A systematic review of open government data initiatives». *Government Information Quarterly*, **32(4)**, pp. 399–418. ISSN 0740624X. doi: 10.1016/j.giq.2015.07.006.
http://www.sciencedirect.com/science/article/pii/S0740624X1500091X

BARGH, MORTAZA S; CHOENNI, SUNIL and MEIJER, RONALD (2016). «Meeting Open Data Halfway». pp. 199–206. ISBN 9781450336406. doi: 10.1145/2910019.2910037.

BARRY, EMILY and BANNISTER, FRANK (2014). «Barriers to open data release: A view from the top». *Information Polity*, **19(1-2)**, pp. 129–152. ISSN 18758754. doi: 10.3233/IP-140327.

BENITEZ-PAEZ, FERNANDO; COMBER, ALEXIS; TRILLES, SERGIO and HUERTA, JOAQUIN (2018). «Creating a conceptual framework to improve the re-usability of open geographic data in cities». *Transactions in GIS*.

BENITEZ-PAEZ, FERNANDO; DEGBELO, AURIOL; TRILLES, SERGIO and HUERTA, JOAQUIN (2017). «Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain: A Data User's Perspective». *ISPRS International Journal of Geo-Information*, **7(2)**, p. 6. ISSN 2220-9964. doi: 10.3390/ijgi7010006. http://www.mdpi.com/2220-9964/7/1/6

BENO, MARTIN; FIGL, KATHRIN; UMBRICH, JÜRGEN and POLLERES, AXEL (2017). «Open Data Hopes and Fears Determining the barriers of Open Data», pp. 69–81. doi: 10.1109/CeDEM.2017.22.

BUDHATHOKI, NAMA RAJ; BRUCE, BERTRAM and NEDOVIC-BUDIC, ZORICA (2008).. doi: 10.1007/s10708-008-9189-x.

CARRARA, WENDY; CHAN, WAE SAN; FISCHER, SANDER and VAN STEENBERGEN, EVA (2015). «Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources». doi: 10.2759/328101.

CARRARA, WENDY; NIEUWENHUIS, MARGRIET and VOLLERS, HELEEN (2016). «Open Data Maturity in Europe 2016». *Technical Report*, European Data Portal.

CARRARA, WENDY; VOLLERS, HELEEN and BERENDS, JORN (2017a). «Analytical Report 5: Barriers in working with Open Data». *Technical Report*, European Data Portal.

—— (2017b). «Re-Using Open Data». *Technical Report*, European Data Portal.

—— (2017c). «Re-Using Open Data». *Technical Report*.

    ABSTRACT: A study on companies transforming Open Data into economic & societal value

CARRASCO, CARLOS and SOBREPERE, XAVIER (2015). «Open Government Data: An Assessment of the Spanish Municipal Situation». *Social Science Computer Review*, **33(5)**, pp. 631–644. ISSN 15528286. doi: 10.1177/0894439314560678.

COMBER, ALEXI J; FISHER, PETER F and WADSWORTH, RICHARD A (2008). «Semantics, metadata, geographical information and users». *Transactions in GIS*, **12**. ISSN 13611682. doi: 10.1111/j.1467-9671.2008.01102.x.

COMBER, LEX; FISHER, PETE; HARVEY, FRANCIS; GAHEGAN, MARK and WADSWORTH, RICHARD (2006). «Using Metadata to Link Uncertainty and Data Quality Assessments».

CONRADIE, PETER and CHOENNI, SUNIL (2014). «On the barriers for local government releasing open data». *Government Information Quarterly*, **31(SUPPL.1)**. ISSN 0740624X. doi: 10.1016/j.giq.2014.01.003.

COTEC (2017). «LA REUTILIZACIÓN DE DATOS ABIERTOS: UNA OPORTUNIDAD PARA ESPAÑA», p. 66.
http://informecotec.es/media/INFORME{_}REUTILIZACION-DE-DATOS.pdf

CRANEFIELD, JOCELYN; ROBERTSON, OLIVER and OLIVER, GILLIAN (2014). «Value in the mash: Exploring the benefits, barriers and enablers of open data apps». In: *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*, pp. 1–15. ISBN 9780991556700.

DEGBELO, AURIOL; BHATTACHARYA, DEVANJAN; GRANELL, CARLOS and TRILLES, SERGIO (2016a). *Toolkits for Smarter Cities: A Brief Assessment*. pp. 431–436. Springer International Publishing, Cham. ISBN 978-3-319-48799-1. doi: 10.1007/978-3-319-48799-1_47.
https://doi.org/10.1007/978-3-319-48799-1_47

DEGBELO, AURIOL; GRANELL, CARLOS; TRILLES, SERGIO; BHATTACHARYA, DEVANJAN; CASTELEYN, SVEN and KRAY, CHRISTIAN (2016b). «Opening up Smart Cities: Citizen-Centric Challenges and Opportunities from GIScience». *ISPRS International Journal of Geo-Information*, **5(2)**, p. 16. ISSN 2220-9964. doi: 10.3390/ijgi5020016.
http://www.mdpi.com/2220-9964/5/2/16

DÍAZ, LAURA; REMKE, ALBERT; KAUPPINEN, TOMI; DEGBELO, AURIOL; FO-
ERSTER, THEODOR; STASCH, CHRISTOPH; RIEKE, MATTHES; SCHAEFFER,
BASTIAN; BARANSKI, BASTIAN; BROERING, ARNE and WYTZISK, ANDREAS
(2012). «Future SDI – Impulses from Geoinformatics Research and IT Trends».
doi: 10.2902/.
http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/297

DOUGLAS D. NEBERT (2014). «Developing Spatial Data Infrastructures: The SDI
Cookbook», pp. 1–128.

ESRI (2010). «Spatial Data Infrastructure A Collaborative Network». *Esri*, p. 8.
http://www.esri.com/library/brochures/pdfs/
spatial-data-infrastructure.pdfhttp://www.esri.com/library/
brochures/pdfs/spatial-data-infrastructure.pdf[Accessed:
1December2015]

EUROPE, ACCESS INFO and FOUNDATION, OPEN KNOWLEDGE (2011). «Beyond
Access : Open Government Data & the Right to (Re) use Public Information»,
(January), p. 89.

EUROPEAN COMMISION (2013). «Elements of a data value chain strategy | Digital
Single Market». *Technical Report*.
https://ec.europa.eu/digital-single-market/en/news/
elements-data-value-chain-strategy

EUROPEAN COMMISSION (2004). «Guidelines on recommended standard licences,
datasets and charging for the reuse of documents». *European Journal of Health
Law*, **11(3)**, pp. 309–317. ISSN 0929-0273. doi: 10.1163/1571809042388581.

—— (2017). «Commission outlines next steps towards a European data eco-
nomy».

FONSECA, FREDERICO T.; EGENHOFER, MAX J.; AGOURIS, PEGGY and CMARA,
GILBERTO (2002). «Using ontologies for integrated geographic information
systems». *Transactions in GIS*, **6(3)**, pp. 231–257. ISSN 13611682. doi:
10.1111/1467-9671.00109.

GONZALEZ-ZAPATA, FELIPE and HEEKS, RICHARD (2015). «The multiple mean-
ings of open government data: Understanding different stakeholders and their

perspectives». *Government Information Quarterly*, **32(4)**, pp. 441–452. ISSN 0740624X. doi: 10.1016/j.giq.2015.09.001.

GOODWIN, JOHN; DOLBEAR, CATHERINE and HART, GLEN (2008). «Geographical linked data: The administrative geography of Great Britain on the semantic web». In: *Transactions in GIS*, volume 12, pp. 19–30. Wiley/Blackwell (10.1111). ISBN 1467-9671. ISSN 13611682. doi: 10.1111/j.1467-9671.2008.01133.x.

GURSTEIN, M.B. (2011). «Open data: Empowering the empowered or effective data use for everyone?» *First Monday*, **16(2)**. doi: 10.5210/fm.v16i2.3316.

HARVEY, F and TULLOCH, D. (2006). «Local-government data sharing: Evaluating the foundations of spatial data infrastructures». *International Journal of Geographical Information Science*, **20(7)**, pp. 743–768. ISSN 13658816. doi: 10.1080/13658810600661607.

HUIJBOOM, NOOR and DEN BROEK, TIJS VAN (2011). «Open data : an international comparison of strategies». *European Journal of ePractice*, **12(March/ April 2011)**, pp. 1–13. ISSN 1988-625X. doi: 1988-625X.

JANEV, VALENTINA; MIJOVIĆ, VUK; PAUNOVIĆ, DEJAN and MILOŠEVIĆ, UROŠ (2014). «Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data Tools». pp. 208–221. Springer, Cham. doi: 10.1007/978-3-319-10178-1_17.
http://link.springer.com/10.1007/978-3-319-10178-1{_}17

JANSSEN, MARIJN; CHARALABIDIS, YANNIS and ZUIDERWIJK, ANNEKE (2012). «Benefits, Adoption Barriers and Myths of Open Data and Open Government». *Information Systems Management*, **29(4)**, pp. 258–268. ISSN 1058-0530. doi: 10.1080/10580530.2012.716740.

JANSSEN, MARIJN and ZUIDERWIJK, ANNEKE (2014). «Infomediary Business Models for Connecting Open Data Providers and Users». *Social Science Computer Review*, **32(5)**, pp. 694–711. ISSN 15528286. doi: 10.1177/0894439314525902.

JETZEK, THORHILDUR; AVITAL, MICHEL and BJORN-ANDERSEN, NIELS (2013). «Generating Value from Open Government Data». In: *ICIS 2013 Proceedings - International Conference on Information Systems*, July 2016, pp. 1737–1756. ISBN 9781629934266. doi: 10.1007/978-3-662-43459-8_5.

JETZEK, TORNHILDUR; AVITAL, MICHEL and BJORN-ANDERSEN, NILES (2012). «The Value of Open Government Data». *Openarchive.Cbs.Dk*, **(2002)**, p. 12. doi: 10.13140/RG.2.1.2413.3608.
http://openarchive.cbs.dk/handle/10398/8621

JOHNSON, PETER A.; SIEBER, RENEE; SCASSA, TERESA; STEPHENS, MONICA and ROBINSON, PAMELA (2017). «The Cost(s) of Geospatial Open Data». *Transactions in GIS*, **21(3)**, pp. 434–445. ISSN 13611682. doi: 10.1111/tgis.12283.
http://doi.wiley.com/10.1111/tgis.12283

KAROLIS.GRANICKAS (2013). «Open Data as a Tool to Fight Corruption». *Open Government Data and Service*, pp. 1–104.

KASSEN, MAXAT (2013). «A promising phenomenon of open data: A case study of the Chicago open data project». *Government Information Quarterly*, **30(4)**, pp. 508–513. ISSN 0740624X. doi: 10.1016/j.giq.2013.05.012.

KEEFE, TERRY; BIKFALVI, ANDREA; BEER, MARTIN and LA ROSA, JOSEP (2013). «A Case Study Analysis of Factors Determining Success or Failure for Participants in Collaborative Innovation Projects in e-Government». *Proceedings of the 13Th European Conference on Egovernment*, pp. 276–282.

KUČERA, JAN; CHLAPEK, DUŠAN and NEČASKÝ, MARTIN (2013). «Open Government Data Catalogs: Current Approaches and Quality Perspective». *Technology-Enabled Innovation for Democracy, Government and Governance*, **8061**, pp. 152–166. doi: 10.1007/978-3-642-40160-2.

LEE, JAE-NAM; HAM, JUYEON and CHOI, BYOUNGGU (2016). «Effect of government data openness on a knowledge-based economy». *Procedia Computer Science*, **91**, pp. 158–167. ISSN 18770509. doi: 10.1016/j.procs.2016.07.053.

MARTIN, SÉBASTIEN and FOULONNEAU, MURIEL (2013). «Open Data: Barriers, Risks and Opportunities». pp. 301–309. ISBN 978-1-909507-22-7.

MENG, AMANDA (2014). «Investigating the R oots of Open Data 's Social Impact». *JeDEM - eJournal of eDemocracy and Open Government*, **6(1)**, pp. 1–13. ISSN 2075-9517.
http://www.jedem.org/article/view/288

NEMOTO, TOMOKO and BEGLAR, DAVID (2014). «Likert-Scale Questionnaires».

OPEN DATA INSTITUTE (2015). «The Value of Open Data».
http://theodi.org/the-value-of-open-data

ORG, WWW OPENDATABAROMETER (2015). «ODB Global Report Third Edition».
http://opendatabarometer.org/doc/3rdEdition/
ODB-3rdEdition-GlobalReport.pdf

PELED, ALON (2013). «Re-Designing Open Data 2.0.» *eJournal of eDemocracy & Open Government*, **5(2)**, p. 187. ISSN 20759517.

POMERANTZ, JEFFREY and PEEK, ROBIN (2016). «Fifty shades of open». *First Monday*, **21(5)**. ISSN 1098-6596. doi: 10.5210/fm.v21i5.6360.
http://journals.uic.edu/ojs/index.php/fm/article/view/6360/5460

REYNARD, DARCY (2018). «Five classes of geospatial data and the barriers to using them». *Geography Compass*, p. e12364. ISSN 17498198. doi: 10.1111/gec3.12364.
http://doi.wiley.com/10.1111/gec3.12364

SCHMIDT, BIRGIT; GEMEINHOLZER, BIRGIT; TRELOAR, ANDREW; SEEGER, B; GEMEINHOLZER, B and BENDIX, J (2016). «Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey». *PLOS ONE*, **11(1)**, p. e0146695. ISSN 1932-6203. doi: 10.1371/journal.pone.0146695.
http://dx.plos.org/10.1371/journal.pone.0146695

SIEBER, RENEE E. and JOHNSON, PETER A. (2015). «Civic open data at a crossroads: Dominant models and current challenges». *Government Information Quarterly*, **32(3)**, pp. 308–315. ISSN 0740624X. doi: 10.1016/j.giq.2015.05.003.

SUNLIGHT FOUNDATION (2010). «Ten Principles For Opening Up Government Information : Sunlight Foundation».
https://sunlightfoundation.com/policy/documents/
ten-open-data-principles/

THE EUROPEAN DATA PORTAL (2016). «Analytical Report 6: Open Data in Cities 2». *Technical Report 860*, European Data Portal.

THE WORLD WIDE WEB (2017). «Open Data Barometer Global Report Fourth Edition», p. 36.
http://opendatabarometer.org/doc/4thEdition/ ODB-4thEdition-GlobalReport.pdf

THE WORLD WIDE WEB FOUNDATION (2015). *Open Data Barometer - Global Report*. The World Wide Web Foundation, 3rdedition.
http://opendatabarometer.org/doc/3rdEdition/ ODB-3rdEdition-GlobalReport.pdf

UBALDI, BARBARA (2013). «Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives». *OECD Working Papers on Public Governance*, **22(22)**, p. 61. ISSN 1993-4351. doi: 10.1787/5k46bj4f03s7-en.

VAN LOENEN, BASTIAAN; CROMPVOETS, JOEP and POPLIN, ALENKA (2017). «Open data portals as part of the open data ecosystem? Lessons learned from geoportal research». ISSN 16130073.

VAN SCHALKWYK, FRANÇOIS; WILLMERS, MICHELLE and MCNAUGHTON, MAURICE (2016). «Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem». *Information Technology for Development*, **22**, pp. 68–83. ISSN 15540170. doi: 10.1080/02681102.2015.1081868.

VELJKOVIĆ, NATAŠA; BOGDANOVIĆ-DINIĆ, SANJA and STOIMENOV, LEONID (2014). «Benchmarking open government: An open data perspective». *Government Information Quarterly*, **31(2)**, pp. 278–290. ISSN 0740624X. doi: 10.1016/j.giq. 2013.10.011.

VETRÒ, ANTONIO; CANOVA, LORENZO; TORCHIANO, MARCO; MINOTAS, CAMILO OROZCO; IEMMA, RAIMONDO and MORANDO, FEDERICO (2016). «Open data quality measurement framework: Definition and application to Open Government Data». *Government Information Quarterly*, **33(2)**, pp. 325–337. ISSN 0740624X. doi: 10.1016/j.giq.2016.02.001.

WANG, HUI-JU and LO, JIN (2016). «Adoption of open government data among government agencies». *Government Information Quarterly*, **33(1)**, pp. 80–88. ISSN 0740624X. doi: 10.1016/j.giq.2015.11.004.
http://linkinghub.elsevier.com/retrieve/pii/S0740624X15300198

WELLE DONKER, FREDERIKA and VAN LOENEN, BASTIAAN (2017). «How to assess the success of the open data ecosystem?» *International Journal of Digital Earth*, **10(3)**, pp. 284–306. ISSN 17538955. doi: 10.1080/17538947.2016.1224938.

WIJNHOVEN, FONS; EHRENHARD, MICHEL and KUHN, JOHANNES (2015). «Open government objectives and participation motivations». *Government Information Quarterly*, **32(1)**, pp. 30–42. ISSN 0740624X. doi: 10.1016/j.giq.2014.10.002.

WILLIAMSON, VANESSA and EISEN, NORMAN (2016). «The Impact of Open Government: Assessing the Evidence». *Brookings Institution*, **(December)**, pp. 1–30.

WORTHY, BEN (2015). «The impact of open data in the Uk: Complex, unpredictable, and political». *Public Administration*, **93(3)**, pp. 788–805. ISSN 14679299. doi: 10.1111/padm.12166.

YANG, TUNG-MOU; LO, JIN and SHIANG, JING (2015). «To open or not to open? Determinants of open government data». *Journal of Information Science*, **41(5)**, pp. 596–612. ISSN 0165-5515. doi: 10.1177/0165551515586715. http://journals.sagepub.com/doi/10.1177/0165551515586715

YOUNG, ANDREW and VERHULST, STEFAAN (2016). *The Global Impact of Open Data: Key Findings from Detailed Case Studies Around the World*. ISBN 9781491964675.

YU, HARLAN and ROBINSON, DAVID G. (2012). «The New Ambiguity of 'Open Government'». *SSRN Electronic Journal*. ISSN 1556-5068. doi: 10.2139/ssrn. 2012489. http://www.ssrn.com/abstract=2012489

ZUIDERWIJK, ANNEKE and JANSSEN, MARIJN (2014). «Open data policies, their implementation and impact: A framework for comparison». *Government Information Quarterly*, **31(1)**, pp. 17–29. ISSN 0740-624X. doi: 10.1016/J.GIQ.2013. 04.003.

ZUIDERWIJK, ANNEKE; JANSSEN, MARIJN; CHOENNI, SUNIL; MEIJER, RONALD and ALIBAKS, ROEXSANA SHEIKH (2012). «Socio-technical Impediments of Open Data.» *Electronic Journal of e-Government*, **10(2)**.

ZUIDERWIJK, ANNEKE; JANSSEN, MARIJN and DAVIS, CHRIS (2014). «Innovation with open data: Essential elements of open data ecosystems». *Information Polity*, **19(1-2)**, pp. 17–33. ISSN 18758754. doi: 10.3233/IP-140329.

A | Appendix

A.1 | Publications, Congresses and External Semester

Journals with impact factor:

1. **Fernando Benitez-Paez**, Auriol Degbelo, Sergio Trilles, Joaquín Huerta. *Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain: A Data User's Perspective.* MDPI: ISPRS International Journal of Geo-Information. (Published)

2. **Fernando Benitez-Paez**, Alexis Comber, Sergio Trilles, Joaquín Huerta. *Creating a conceptual framework to assess and improve the re-usability of open geographic data in cities.* Transactions in GIS. (Published)

International Conferences:

1. **Fernando Benitez-Paez**, Sergi Trilles, Joaquin Huerta. *City data 3.0 - A generic initiative to promote and assess the reuse of geographic information in cities - early steps.* in International Conference on Connected Smart Cities 2017 (Lisbon). July 2017. ISSN 1863-2246.

2. **Fernando Benitez-Paez**. Sergi Trilles, Joaquin Huerta, 2017. *City data 3.0 - A generic initiative to promote and assess the reuse of geographic information in cities - early steps.* In: Bregt, A., Sarjakoski, T., Lammeren, R. van, Rip, F. (Eds.). Societal Geo-Innovation : short papers, posters and poster abstracts of the 20th AGILE Conference on Geographic Information Science. Wageningen University and Research 9-12 May 2017, Wageningen, the Netherlands. Selected as the Best Poster. ISBN 978-90-816960-7-4.

Guidelines:

1. **Fernando Benitez-Paez** Open Data Reusability Handbook. May 08 2018. Open City Tool Kit - Guidelines. Available online: https://elcano.init.uji.es/documents/qaBFbFxqPB2tQrf97

Workshops:

1. **Fernando Benitez-Paez**, Adeoluwa Akande, Michael Gould, Joaquin Huerta. Open Data for Open Cities, The reuse of open data through spatial analysis – Second edition of AGILE 2018 pre-conference Workshops. June 12 2018. Available online: http://opendata4opencities.uji.es/

2. **Fernando Benitez-Paez**, Mohamad Medhi Moradi, Joaquin Huerta. Open Data for Open Cities, reuse and discovery level applied to spatial point process on linear network. First edition in AGILE 2017 Workshop. May 9 2017. Available online: http://opendata4opencities.uji.es/

External stay in University:

- Universidade de Lisboa Nova ISM , under supervision of Professor Roberto Henriques and Marco Painho from February 7, 2017 to July 31, 2017.

Special Issue in Open Data:

- Guest Editor in a special issue of Information (ISSN 2078-2489). Open Data for Open Cites (OD4OC): Reuse of Open Data through Spatial Analysis. This special issue belongs to the section "Information and Communications Technology".

Intership:

- Esri Portugal, Implementation and development of Amadora Open Data Portal using ArcGIS Open Data, under supervision of Professor Marco Painho and Rui Santos from February 7, 2017 to March 31, 2017.

A.2 | **Online Survey Questions**

This appendix illustrate the questions and sections included in the online survey that was publicly shared. The following format was used to guide respondents through the survey's sections.

1. **Personal information:** Tell us a little about yourself. We will not share or publish this information.

   (a) Which country are you currently working? Open Question.

   (b) Which city/cities are you working or using geographical data? Open Question.

   (c) How old are you? Open Question.

2. **Your work**: In this section we are interested in aspects of your work and your experience level in the sector or industry to which you belong or have belonged to in the past. You can mention the elements that are the most relevant.

   (a) What is your employment role? Multiple choice: Geographical apps developer, Geographical data analyst, Developer and analyst, Data Science analyst, Manager—Project leader, Researcher-Student-Teacher, Other.

   (b) In which industry do you work? Multiple choice: Local Government, National Government, Education, Non- profit, Media, Startup—Entrepreneurship, Business, Other.

   (c) How much experience do you have in the industry? Multiple choice: Less than 1 year, 2 to 6 years, 7 to 10 years, 11 to 20 years, More than 20 years.

3. **City Open Data**: It is important for us to know your opinion about open data available in the cities. In particular geographic data. In this section we will ask you about your reasons for use this data and your knowledge of those current initiatives.

   (a) Please indicate the level of importance for each option when using city open data? Multiple choice grid, with Very important, Neutral and Not important as choices: Geographic information accessibility, High-quality geographic information, Scalability and ease of project maintenance, City innovation improvement, Transparency and collaboration improvement, Economic benefits for the city, Academic and research improvement

   (b) Do you know or use the cities' open data portals? Multiple choice with yes or not as choices.

4. **Cities' open data portals**: Please provide specifics on data portals, adding a URL where possible. If your previous answer was Yes, please specify which city open data portals you know or have used

5. **Barriers and features**: We would like to know the barriers, errors, and problems that you have encountered while using cities' open data portals. Also, we would like to know the features and aspects that you consider positive and that should be kept within these initiatives.

   (a) Which functionalities do you think are not useful in city open data portals? Open Question.

   (b) From your experience with city open data portals, what do you consider to be barriers when using those portals? Multiple choice grid with Not a barrier, Moderate barrier and Major barrier as choices: Published data is hard to access, Misinterpretation and misuse of data, Time spent searching for data, Understanding how to re-use the data, Understanding terms of use, Nonexistence or low relevance of URL to access to data, Technology used for publishing data, Varying and low integration of data sources or data producers, Lack of updates of published data.

(c) From your experience, which was the most common error/barrier you have faced (not have faced) when searching or using data from city open data portals? Open Question.

(d) Which of following do you think are the most needed features of city open data portals? Multiple choice grid with Highly necessary, Neither necessary nor unnecessary and Unnecessary as choices: Filters for advanced search, URL to Access data, URL to Access data, Data Categories, Table view and graphs, Terms of use and re-use, Details on how the data has been produced, Viewers and interface to explore the data, Feedback from other users.

(e) Which of following functionalities, is your frequency of use in cities' open data portals? Multiple choice grid with Every time, Occasionally/Sometimes, and Never as choices: Filters for advanced search, Access data URL, Data Categories, Table view and graphs, Terms of use and re-use, How the data has been produced? Viewers and interface to explore the data, Viewers and interface to explore the data, Feedback from others users.

6. **City open data portals usability**: We'd like to know about the level of use of city open data portals and the available geographic data. In this Section, we will ask your frequency of use and we want to determine the usability level of those portals.

   (a) When you need to use city geographical information which portals do you normally use? Multiple choice grid with Often, Sometimes and Not used as choices: Government data portals. (National), Government data portals. (City-Local), Private repositories, Pay or collect data, International repositories, Other.

   (b) Indicate your agreement level regarding these statements on current city open data portals: Multiple choice grid with Agree, Neither agree or disagree and Disagree as choices. I would like to use these portals frequently, I found the portals unnecessarily complex, These portals were easy to use, I would need the support of a technical person to be able to use the portals, I found the various functions in the portals were well integrated, There was too much inconsistency in the portals,

I would imagine that most people would learn to use the portals very quickly, I found the portals very cumbersome to use, I felt very confident using the portals, I needed to learn a lot of things before I could get going with the portals.

7. **Searching for geographical data**: We'd like to know which criteria and formats you use when searching and choosing geographical data.

   (a) Tell us about your data quality criteria when choosing available data in city open data portals? Multiple choice grid with Desirable, Neutral and Undesirable as choices. Accuracy: data/metadata record correctly described, Completeness: the number of completed fields in a data/-metadata record, Consistency: discrepancy between data published and entire data catalogs, Currency: data or metadata is up date, Technical accessibility, Openness.

   (b) Which of the following are main features that you consider when choosing available data in city open data portals. Multiple choice grid with Definitely consider, Might or might not consider and Would not consider as choices. Data quality, how data was produced, Geometry (Point, Lines, Polygons, raster, other), Lack of information (Incomplete fields), Terms of use and re-use, Technology used for the publication process, Creation/Publication date, Author (Public agency, Private), Cost, Openness.

   (c) What of the following output formats do you consider most useful for your work? Multiple choice grid with Strong useful, Neutral and Not useful as choices. KML, OGC Standard (WMS, WFS, WMTS), REST, CSV, Shapefile, GeoJSON, JSON, RDF, XML, Download files (i.e., Zip).

   (d) If you had the chance to improve city open data portals, which are the improvements/features or tools will you would add and why? Open Question.

   (e) In your industry, how do you think we might increase the usage of geographical data on current city open data portals? Open Question.

A.3 | List of data user suggestions in use cases

Use case	Data user suggestions	Category
Bogotá - IDECA	Do not start from scratch. Many current initiatives ignore their predecessors. It should be at least have a link to them. Linked data is required. Making visible the connections between the data and its data sources.	Web Semantic - Linked Data
	Having a link between the current community project and the open data initiative	
	Data license must be available in each dataset	Legal terms concerns
	Current metadata is not appropriate to understand the potential use of available data	User-focused metadata
	Standardization is required. Better communication channels with users	Data users communities
	The identification of the data users' demand is required	
	Data opening-level need to be improved	Interoperability
	Integration between the local SDI and the open data initiative is required	
	Promote the open data initiative internally in the city hall and its departments	Open Data Strategy
	Strengthening the open data area	
	Linked data to exploring data category, illustrating the related data	Web Semantic - Linked Data
	Web Semantic is required, including keywords for an efficient discoverability	

Continued on next page

Table A.1 – *Continued from previous page*

Use case	Data user suggestions	Category
Valencia City Hall	Lack of communication among the city hall departments	Open Data Strategy
	Released data is on a regional scale and do not allow the local analysis	Accessibility
	Better integration among the city hall departments	Standardization
	Duplicate data, in several portals with a different publication date	
	Use case documentation	Open Data Strategy (Documentation)
	API and examples about how to use the released data	
	Make clear what is the format of released data	Technical
	Make clear which data is only for visualization and which data if for download	
Medellin	Allow data users communities are part of the open data initiative, including citizen science projects	Data users communities
	Linked data, including some examples	Web Semantic - Linked Data
	Better communication channels with data users' communities, including blogs, forums, and promoting the released data through social networks exploring the potential use	Feedback improvements – User-focused metadata
	Include useful feedback to make adjustments and improvements in the current initiatives of open data	
	Collection data process should be available, to allow users explore how the data was created	User-focused metadata
	Data opening-level need to be improved	Interoperability

Continued on next page

Table A.1 – *Continued from previous page*

Use case	Data user suggestions	Category
	Better integration among the city hall departments	Standardization
Cali - IDESC	Develop better communication channels between the open data initiative and data users	Feedback improvements
	Integration with current data initiatives like Local SDI	Standardization
	Dashboard to display the statistical and geographic context of the released data	Web Semantic - Linked Data
	More emphasis on downloading the data	Technical
	Make the inventory of current data users and re-users	Open Data Strategy
	Promote the city hall departments integration, releasing an integrated dataset, instead of having several portals with different levels of data	