



Universitat d'Alacant
Universidad de Alicante

Marco de referencia para la publicación
de datos abiertos comprensibles
basado en estándares de calidad

Susana Graciela Cadena Vela

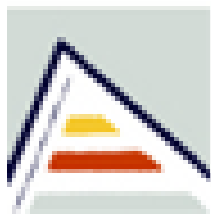


Tesis **Doctorales**

UNIVERSIDAD de ALICANTE

Unitat de Digitalització UA

Unidad de Digitalización UA



Universitat d'Alacant Universidad de Alicante

Instituto Universitario de Investigación en Informática

Escuela Politécnica Superior

Marco de referencia para la publicación de datos abiertos comprensibles basado en estándares de calidad

SUSANA GRACIELA CADENA VELA

Tesis presentada para aspirar al grado de
DOCTORA POR LA UNIVERSIDAD DE ALICANTE

DOCTORADO EN INFORMÁTICA

Dirigida por:

Dr. Andrés Fuster Guilló

Dr. José Norberto Mazón López

Tesis doctoral realizada en el seno del proyecto de investigación "Publi@City: plataforma para la publicación y consumo de datos abiertos para una ciudad inteligente" (TIN2016-78103-C2-2-R) financiado por el Ministerio de Economía, Industria y Competitividad de España.

Alicante, noviembre 2019

AGRADECIMIENTO

«Lo importante en la ciencia no es tanto obtener nuevos datos, sino descubrir nuevas formas de pensar sobre ellos» William Lawrence Bragg

A Dios por la fuerza para emprender,
A mi esposo y mis hijas por dame el aliento y las ganas de seguir,
A papá Ulpiano (+) y mi familia por su ejemplo de lucha y constancia.

A Andrés y Norberto, tutores de este trabajo por su guía, amistad y apoyo total en este proceso,

A la Universidad Central del Ecuador por su decisión de aportar al desarrollo de la investigación en el Ecuador,

A Jorge, Miguel y José por su gestión para realizar el programa doctoral.

A los amigos por acompañarme en este trayecto de la vida para lograr este sueño.

RESUMEN

Los portales donde se publican datos abiertos son una oportunidad para democratizar el acceso a la información generada por las instituciones del sector público, así como potenciar su reutilización. Esta reutilización de datos abiertos permite el desarrollo de nuevos productos y servicios digitales, creando oportunidades de desarrollo social y económico. Sin embargo, la reutilización de estos datos se enfrenta con diversas barreras en su expansión, debido a diversos problemas relacionados con la calidad de datos que van desde la incompletitud de los datos hasta la falta de actualización de los mismos.

En concreto, un criterio relevante de calidad de datos abiertos es la comprensibilidad, ya que un proceso de interpretación errónea de los datos abiertos provocará ambigüedades o malentendidos que desmotivarán su reutilización. Por esta razón, un aspecto importante en el acceso a los datos abiertos son sus metadatos, es decir aquellos datos que describen el contenido del conjunto de datos abiertos, cuya finalidad es facilitar su comprensión.

Con el fin de mejorar la comprensibilidad de los datos abiertos, las propuestas actuales se centran en la evaluación de los datos ya publicados, no existiendo propuestas que evalúen cómo se puede mejorar la comprensibilidad de los datos abiertos desde los sistemas de información de origen del dato.

Esta tesis doctoral revisa y analiza los trabajos de investigación en datos abiertos mediante el desarrollo de un mapeo sistemático, incluyendo un análisis de barreras en la reutilización de datos abiertos, así como buenas prácticas y estándares aplicables al proceso de publicación de datos en formatos abiertos. Este análisis sistemático de trabajos previos de investigación ha servido como punto de partida para la definición de un marco de referencia que permita la publicación de datos abiertos comprensibles. En este marco de referencia se considera la comprensibilidad de los datos abiertos mediante la gestión de los metadatos desde las mismas fuentes de datos en los sistemas de información de origen, incluyendo el concepto de fichero de datos maestros de apertura. Finalmente, cabe destacar que, a través de la aplicación del método de investigación-acción, se ha validado el marco de referencia, utilizando entrevistas y grupos de discusión con varios administradores de portales de datos abiertos.

RESUM

Els portals on es publiquen dades obertes són una oportunitat per a democratitzar l'accés a la informació generada per les institucions del sector públic, així com potenciar la seua reutilització. Aquesta reutilització de dades obertes permet el desenvolupament de nous productes i serveis digitals, creant oportunitats de desenvolupament social i econòmic. No obstant això, la reutilització d'aquestes dades s'enfronta amb diverses barreres en la seua expansió, a causa de diversos problemes relacionats amb la qualitat de dades que van des de la *incompletitud de les dades fins a la falta d'actualització d'aquests.

En concret, un criteri rellevant de qualitat de dades obertes és la comprensibilitat, ja que un procés d'interpretació errònia de les dades obertes provocarà ambigüitats o malentesos que desmotivarà la seua reutilització. Per aquesta raó, un aspecte important en l'accés a les dades obertes són les seues metadades, és a dir aquelles dades que descriuen el contingut del conjunt de dades obertes, la finalitat de les quals és facilitar la seua comprensió.

Amb la finalitat de millorar la comprensibilitat de les dades obertes, les propostes actuals se centren en l'avaluació de les dades ja publicades, no existint propostes que avaluen com es pot millorar la comprensibilitat de les dades obertes des dels sistemes d'informació d'origen de la dada.

Aquesta tesi doctoral revisa i analitza els treballs de recerca en dades obertes mitjançant el desenvolupament d'un mapatge sistemàtic, incloent una anàlisi de barreres en la reutilització de dades obertes, així com bones pràctiques i estàndards aplicables al procés de publicació de dades en formats oberts. Aquesta anàlisi sistemàtica de treballs previs d'investigació ha servit com a punt de partida per a la definició d'un marc de referència que permeta la publicació de dades obertes comprensibles. En aquest marc de referència es considera la comprensibilitat de les dades obertes mitjançant la gestió de les metadades des de les mateixes fonts de dades en els sistemes d'informació d'origen, incloent el concepte de fitxer de dades mestres d'obertura. Finalment, cal destacar que, a través de l'aplicació del mètode d'investigació-acció, s'ha validat el marc de referència, utilitzant entrevistes i grups de discussió amb diversos administradors de portals de dades obertes.

ABSTRACT

The portals where open data are published are an opportunity to democratize access to information generated by public sector institutions, as well as enhance its reuse. This reuse of open data allows the development of new digital products and services, creating opportunities for social and economic development. However, the reuse of these data faces several barriers in its expansion, due to various problems related to the quality of data ranging from incomplete data to the lack of updating them.

In particular, a relevant criterion for the quality of open data is comprehensibility, as a process of misinterpretation of open data will lead to ambiguities or misunderstandings that will discourage their reuse. For this reason, an important aspect of access to open data is its metadata, i.e. those data that describe the content of the open data set, the purpose of which is to make it easier to understand.

In order to improve the comprehensibility of the open data, the current proposals focus on the evaluation of the data already published, and there are no proposals that evaluate how the comprehensibility of the open data can be improved from the data origin information systems.

This doctoral thesis reviews and analyses open data research through the development of systematic mapping, including an analysis of barriers to the reuse of open data, as well as good practices and standards applicable to the process of publishing data in open formats. This systematic analysis of previous research work has served as a starting point for the definition of a reference framework that allows the publication of understandable open data. This reference framework considers the comprehensibility of open data through the management of metadata from the same data sources in the source information systems, including the concept of open master data file. Finally, it should be noted that, through the application of the action-research method, the reference framework has been validated, using interviews and focus groups with several administrators of open data portals.

ÍNDICE

Agradecimiento	2
Resumen	3
Resum.....	4
Abstract.....	5
Índice.....	6
Índice de Tablas	10
Índice de Figuras	11
1. INTRODUCCIÓN	13
1.1 Motivación y Contexto.....	14
1.2 Objetivos	17
1.2.1 General.....	17
1.2.2 Específicos	17
2. METODOLOGÍA DE INVESTIGACIÓN	19
2.1 Mapeo sistemático.....	20
2.1.1 Amenazas a la validez del mapeo sistemático	22
2.2 Investigación en acción.....	22
2.2.1 Proceso de la aproximación de “investigación en acción”	22
3. ESTADO DEL ARTE EN DATOS ABIERTOS	25
3.1 Estructura del Estado del Arte	26
3.1.1 Alcance.....	26
3.1.2 Proceso de búsqueda.....	27
3.1.3 Esquema de clasificación	29
3.2 Amenazas a la validez	35
3.3 Mapeo Sistemático	35

3.3.1	Faceta 1: Lugares de publicación.....	36
3.3.2	Faceta 2: Impacto.....	39
3.3.3	Faceta 3: Dominio	39
3.3.4	Faceta 4: Tópico.....	41
3.3.5	Faceta 5: ciclo de vida de los datos	43
3.3.6	Faceta 6: tipo de investigación.....	45
3.3.7	Combinación de facetas. Mapa(s) sistemático(s)	47
3.4	Discusión de resultados del mapeo sistemático.....	55
3.4.1	Lugares en los que se ha publicado investigaciones de datos abiertos	55
3.4.2	Impacto que han tenido estos estudios.....	55
3.4.3	Dominio al que pertenecen las investigaciones realizadas	56
3.4.4	Ciclo de vida del dato en los cuales han trabajado las investigaciones	57
3.4.5	Tipo de investigaciones que se realizan.....	59
3.4.6	Tópicos que se trataron en la investigación.....	60
3.5	Más allá del mapeo sistemático: proyectos de innovación de datos abiertos.....	62
3.5.1	Portales de datos abiertos y búsqueda de conjuntos de datos	62
3.5.2	La innovación de datos abiertos como facilitador del negocio.....	64
3.5.3	Calidad de datos abiertos.....	65
3.6	Aspectos relevantes del estudio sistemático	69
4.	CALIDAD DE DATOS Y SU IMPACTO EN LA PUBLICACIÓN DE DATOS ABIERTOS.....	72
4.1	Calidad de datos	73
4.1.1	Planificación de la calidad de datos.....	74
4.1.2	Estándares para conseguir calidad de datos	75
4.1.3	Estándares para evaluar la calidad de datos.....	76
4.1.4	Comprensibilidad de los datos.....	79
4.1.5	Metadatos y diccionario de datos	80

4.1.6	Vocabulario de Catalogo de datos (DCAT)	81
4.2	Gobierno de datos	82
5.	MARCO DE REFERENCIA PARA APERTURA DE DATOS COMPENSIBLES	86
5.1	Contexto	87
5.2	Nivel estratégico	89
5.3	Nivel táctico	90
5.3.1	Alcance - Portafolio de Conjunto de Datos	91
5.3.2	Informe de inicio del proyecto y actualizaciones.....	92
5.3.3	Acciones de gestión para el uso de datos abiertos (usuarios reutilizadores)....	92
5.3.4	Definir herramientas de publicación de datos abiertos	93
5.3.5	Medir, monitorear y comunicar uso de los datos publicados.....	94
5.4	Nivel operativo	95
5.4.1	Definición del archivo maestro de datos abiertos.....	95
5.4.2	Definición de metadatos.....	97
6.	RESULTADOS Y DISCUSIÓN.....	102
6.1	Diseño de la metodología de evaluación.....	103
6.2	Población participante.....	103
6.3	Procesamiento de la información de las entrevistas.....	105
6.3.1	Incluir a los potenciales reutilizadores en el proceso.....	106
6.3.2	Considerar aspectos de calidad de datos y archivos de origen.....	106
6.3.3	Concepto de comprensibilidad.....	108
6.4	Discusión de resultados de las entrevistas	108
6.5	Procesamiento de información de grupos de discusión.....	109
6.5.1	¿Es aplicable el marco de referencia a su institución?	110
6.5.2	¿El marco de referencia favorece la comprensibilidad por parte del reutilizador?	112

6.6	Aspectos importantes de la evaluación.....	113
7.	CONCLUSIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS.....	115
7.1	Líneas de investigación futuras.....	117
8.	CONTRIBUCIONES.....	118
8.1	Publicaciones científicas.....	119
8.2	Transferencia de conocimiento.....	120
9.	BiBLIOGRAFÍA.....	121
10.	ANEXOS.....	134
10.1	Anexo 1: Ejemplo de un catálogo de conjuntos de datos a publicarse.....	135
10.2	Anexo 2: Transcripción de entrevistas realizadas.....	136
10.3	Anexo 3: Transcripción de las observaciones de los grupos focales.....	152
10.4	Anexo 4: Organización de las mesas de trabajo para ANIE (Agenda Nacional de Investigación Educativa).....	159
10.5	Anexo 5: Certificado de aportación en la co-creación de la Política Nacional de Datos Abiertos.....	161

ÍNDICE DE TABLAS

Tabla 1: Motores de búsqueda.....	27
Tabla 2: Listado de acrónimos por conferencia.....	38
Tabla 3: Descripción de características de calidad.....	78
Tabla 4: Descripción de indicadores de comprensibilidad	94
Tabla 5: Estructura de metadatos.....	97
Tabla 6: Ejemplo de aplicación de la estructura de metadatos.....	100



Universitat d'Alacant
Universidad de Alicante

ÍNDICE DE FIGURAS

Figura 1: Estructura del marco de referencia propuesto	17
Figura 2: Proceso llevado a cabo para la realización del mapeo sistemático en la presente tesis doctoral, basado en [29], [30]	21
Figura 3: Etapas de la investigación en acción.....	24
Figura 4: Revistas con mayor número de publicaciones sobre datos abiertos	37
Figura 5: Conferencias con mayor número de trabajos publicados sobre datos abiertos	37
Figura 6: Porcentaje de publicaciones según el repositorio científico donde están accesibles	39
Figura 7: Publicaciones por impacto.....	39
Figura 8: Porcentajes de publicaciones según su dominio	40
Figura 9: Distribución de las publicaciones por año y dominio.....	41
Figura 10: Porcentaje de publicaciones según su tópico	42
Figura 11: Distribución de publicaciones según su tópico entre el 2006 y 2017	42
Figura 12: Distribución de las publicaciones según el ciclo de vida de los datos	43
Figura 13: Distribución de las publicaciones según el ciclo de vida de los datos de 2006 a 2017	44
Figura 14: Porcentajes de publicaciones por tipo de investigación	45
Figura 15: Distribución de las publicaciones por tipo	46
Figura 16: Clasificación de las publicaciones de acuerdo al dominio y ciclo de vida de los datos.	48
Figura 17: Clasificación de las publicaciones de acuerdo al dominio y tópico	49
Figura 18: Clasificación de las publicaciones de acuerdo al dominio y tipo	50
Figura 19: Clasificación de las publicaciones de acuerdo al ciclo de vida del dato y tópico	51
Figura 20: Clasificación de las publicaciones de acuerdo al tópico y tipo	53
Figura 21: Clasificación de las publicaciones de acuerdo al ciclo de vida del dato y tipo de investigación.....	54

Figura 22: Modelo DCAT tomado de [124].....	82
Figura 23: Visión general del marco de trabajo para la publicación de datos abiertos comprensibles.....	88
Figura 24: Procesos a ejecutarse con los responsables adaptado de [131].....	88
Figura 25: Estructura del metadato basado en DCAT [124].....	99
Figura 26: Percepción de los gestores de los portales acerca de incluir a los reutilizadores.....	106
Figura 27: Percepción de los gestores de los portales acerca de la a calidad de los datos..	107
Figura 28: Percepción de los gestores de los portales acerca si la institución cuenta con estándar de calidad de datos.....	107
Figura 29: Concepto de comprensibilidad de los gestores de los portales.....	108
Figura 30: Resultados grupo focal 1 y 2.....	111
Figura 31: Resultados grupo focal 3.....	112
Figura 32: Validación de criterios de comprensibilidad.....	113

1. INTRODUCCIÓN

Este capítulo presenta el contexto en el cual se desarrolla el trabajo de investigación, la motivación del mismo, así como las contribuciones que aporta. Además, se presentan los objetivos generales y específicos del trabajo de investigación desarrollado.

Universitat d'Alacant
Universidad de Alicante

1.1 Motivación y Contexto

Los datos abiertos se caracterizan por ser “usados, modificados y redistribuidos libremente”[1]. En la última década, la cantidad y variedad de datos abiertos publicados por los gobiernos ha sido creciente. Esta tendencia responde a la constante evolución de la tecnología (marcada fuertemente por la expansión de Internet) y el crecimiento en las capacidades de los sistemas de información, que ha posibilitado un incremento en la generación de datos y la consecuente demanda de acceso a la información en formatos reutilizables [2].

De hecho el uso de datos abiertos es considerado una oportunidad sin precedentes para los gobiernos y las organizaciones, ya que su reutilización ha permitido el desarrollo de soluciones tecnológicas en áreas como biología, salud o turismo, entre otras [3], contribuyendo al desarrollo económico impulsado por el sector infomediario mayoritariamente.

Por tanto, datos abiertos significa mucho más que tener la posibilidad de acceder a determinada información [4], sino que son una oportunidad para que la sociedad pueda innovar mediante la reutilización de la información del sector público (RISP) [5]. Esta realidad ha impulsado a organizaciones de todo el mundo a continúen proponiendo portales donde publicar en abiertos sus datos [6], [7].

Emprender un proyecto para crear un portal de datos abiertos, puede ser complejo para las organizaciones, ya que se enfrentan a problemas de diversa naturaleza [1]. Por ejemplo, a nivel técnico la dispersión de las fuentes de datos de origen dificulta la publicación de manera integrada de datos abiertos teniendo en cuenta sus respectivos metadatos. A nivel organizativo, el mayor problema reside en la falta de gobernanza en el proceso de publicación de datos abiertos. Además, cabe destacar que las organizaciones que desarrollan un proyecto de datos abiertos, generalmente, deciden publicar conjuntos de datos sin considerar estrategias que permitan apoyar la toma de decisiones sobre cómo hacerlo [11]. De hecho, la estrategia más ampliamente utilizada para la apertura de datos obedece al cumplimiento de imperativos legales, como la legislación existente sobre transparencia. Es decir, se publican en abierto aquellos conjuntos de datos que establece como obligatorios una determinada legislación relacionada generalmente con el concepto de transparencia, ignorándose así otras cuestiones relevantes de carácter técnico a la hora de priorizar la publicación de datos en abierto como la gestión y publicación adecuada de metadatos [9]. Los metadatos son un elemento clave para la interpretación de los datos abiertos [2], ya que permiten describirlos y

facilitar, de esta manera, su comprensión. Por tanto, una gestión adecuada de los metadatos de los datos abiertos hace que sean comprensibles por parte de los usuarios [8], [9] y, en consecuencia, más fácilmente reutilizables.

La usencia o la mala calidad de estos metadatos, dificulta la reutilización por cuanto no se interpreta adecuadamente el contenido de los conjuntos de datos publicados en portales de datos abiertos [10].

Si los datos publicados en los portales de datos abiertos no son comprensibles, pueden ser fácilmente malinterpretados y las ambigüedades y los malentendidos desalentarán a los reutilizadores. Además, los servicios y productos creados mediante el uso de datos abiertos incomprensibles pueden contener incorrecciones, no generando suficiente confianza en los usuarios potenciales, no alcanzando el éxito esperado. En definitiva, este escenario hace que aumente el riesgo de reutilización de un número limitado de conjuntos de datos abiertos, lo que, a su vez, hace que los portales de datos abiertos se abandonen poco tiempo después de su lanzamiento por falta de uso de los datos que contienen [3].

De hecho, varios trabajos de investigaciones previos han definido una serie de barreras en la adopción de proyectos de datos abiertos relacionadas con la reutilización de los mismos [5]. Dentro de estas barreras, se pueden destacar aquellas que tienen que ver con la comprensibilidad de los datos abiertos:

- Ausencia o mala calidad de los metadatos, que dificulta la reutilización de los datos [11], [12].
- Poca o ninguna estandarización de columnas o atributos, unidades de medida, datos de catálogos, entre otros [13], [14], [15].
- Conjuntos de datos incompletos, datos faltantes, falta de atributos relevantes, datos incorrectos, [16], [17], [18].
- Falta de especificación de la licencia de los datos publicados, indicando si el uso puede ser comercial o si se puede compartir [19], [20].
- Difícil acceso y disponibilidad de datos [21].
- Falta de especificación en el propósito, fuentes y responsable de los datos [22], [23].

Al especificar las fuentes de datos, otra problemática encontrada es la información que se encuentra almacenada en silos (fuentes diversas de datos) que no necesariamente están relacionados y al momento de publicarlos en formato abiertos, ocasiona que sean difíciles de

entender por no contar con información contextual adicional, lo que limita el valor de los datos por no poder reutilizarlos [24]. En este contexto la gestión de datos maestro (MDM), tiene como objetivo gestionar los datos de una organización para proporcionar un punto de vista único (conocido como archivo maestro), permitiendo la eliminación de registros duplicados, normalizando datos, incorporando normas y reglas que impidan el ingreso al sistema de datos incorrectos, favoreciendo a que los datos publicados en abierto se comprendan.

La norma como ISO/IEC 25024 define la comprensibilidad como la capacidad de los datos para ser leídos e interpretados por los usuarios teniendo en cuenta los lenguajes, símbolos y unidades [4] apropiados, según los estudios realizados para este trabajo de investigación, hasta ahora no se ha considerado ningún mecanismo para abordar la comprensibilidad de los datos abiertos.

Con el fin de superar estos vacíos, este documento introduce un marco que amplía y adopta varias buenas prácticas y normas como la ISO 8000, para considerar los datos de diferentes fuentes y la ISO 25000 para mejorar la comprensión de los datos abiertos correspondientes (que serán publicados). La Figura 1 , muestra una visión general de la estructura de la propuesta. Las fuentes de datos son el punto central del proceso de publicación de los datos, ya que las organizaciones deben considerar la integración de una variedad de fuentes de datos dentro de un proyecto de apertura. Por lo tanto, uno de los elementos a desarrollar a la hora de publicar datos es un archivo maestro de apertura de datos, como estrategia para integrar metadatos heterogéneos de diferentes fuentes de datos. Este nuevo artefacto (junto con un diccionario de datos abierto) contiene toda la información necesaria para que los datos sean comprensibles. Se propone identificar a los grupos de interés y definir sus funciones y tareas dentro del marco de referencia.

Para evaluar el marco de referencia se han llevado a cabo la aplicación de entrevistas y grupos de discusión considerando indicadores extendidos de la norma ISO 25000.

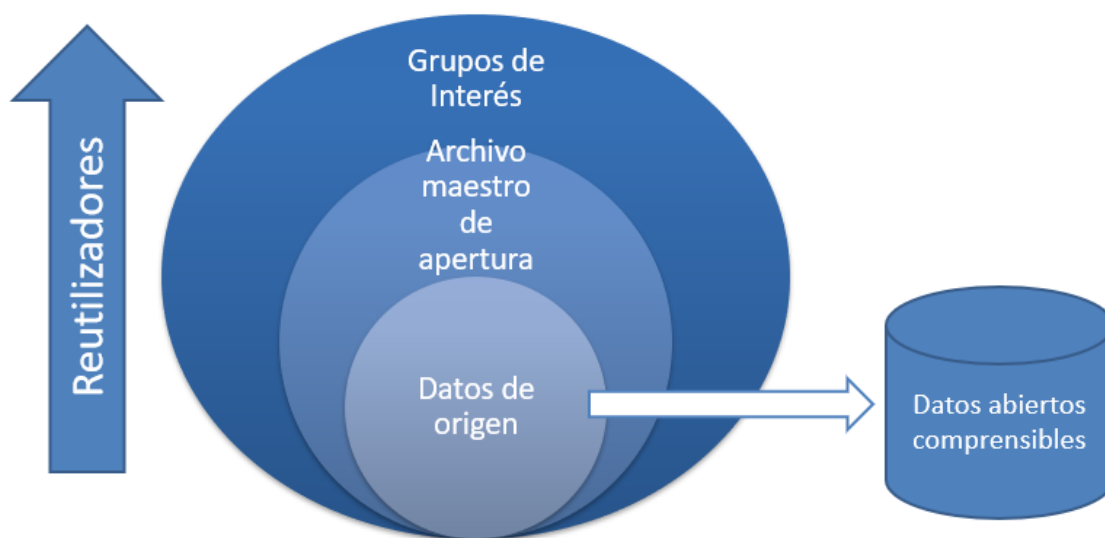


Figura 1: Estructura del marco de referencia propuesto

1.2 Objetivos

1.2.1 General

Definir un mecanismo para la publicación de datos abiertos comprensibles que proponga una solución para la diversidad de fuentes de los datos dentro de la organización considerando la adopción de estándares y buenas prácticas.

1.2.2 Específicos

- Realizar un estudio de las investigaciones relacionadas con datos abiertos clasificándolas desde un punto de vista tecnológico, identificando las barreras, buenas prácticas y estándares aplicables al proceso de apertura de datos.
- Diseñar un marco de referencia para apertura de datos que incluyen una propuesta de solución a la problemática identificada en el estado del arte, considerando buenas prácticas y estándares, la configuración del archivo de metadatos, basado en el concepto de datos maestros y con criterios de comprensibilidad.
- Realizar una validación basada en la metodología de investigación en acción mediante iteraciones con los responsables directos en la publicación de los portales de datos

abiertos, analizando prácticas aplicadas en los procesos de apertura de datos y contrastándolos con el marco de referencia propuesto.

La tesis doctoral se estructura de la siguiente manera: en el capítulo II, se presenta el resultado del estudio del estado del arte, realizado a través de una metodología de mapeo sistemático, en el siguiente capítulo se describe los conceptos relacionados a la gestión de la calidad de datos y su impacto en la publicación de datos en formatos abiertos. Con estos los resultados presentados en estos dos capítulos, se diseña un marco de referencia para apertura de datos comprensibles que es presentado en el capítulo V, en el siguiente capítulo se presentan los resultados de la validación del marco de referencia diseñado seguido por las conclusiones y contribuciones.



Universitat d'Alacant
Universidad de Alicante

2. METODOLOGÍA DE INVESTIGACIÓN

Este capítulo presenta una descripción de los métodos utilizados en cada una de las etapas de investigación. El mapeo sistemático es el método utilizado para la clasificación y análisis de los trabajos de investigación relacionados con datos abiertos. Para el desarrollo y evaluación del marco de referencia se ha utilizado la aproximación metodológica de “investigación en acción”.

Universitat d'Alacant
Universidad de Alicante

2.1 Mapeo sistemático

Un mapeo sistemático tiene como objetivo encontrar y clasificar los trabajos de investigación de un área temática específica siguiendo un método bien definido y repetible [25]. Consiste en varios pasos que son: obtención de un esquema de clasificación, recopilación de estudios pertinentes, realización de la clasificación y análisis de los resultados [26]. El análisis se centra en responder a preguntas específicas de investigación, generalmente relacionadas con la identificación y cobertura del área temática, en la evolución a lo largo del tiempo, así como en la discusión adicional sobre los desafíos y las tendencias en relación con dicha área [27].

Para el mapeo sistemático realizado en esta tesis doctoral, se aplica el proceso introducido en [27]. Este proceso está basado en la adaptación de los mapeos sistemático de estudios realizados en el campo médico y su aplicación a la ingeniería de software propuestos en [26]. Específicamente, este proceso consiste en la realización de varios pasos:

1. Definir el alcance de la investigación a llevarse a cabo con el mapeo sistemático.
2. Definir el proceso de búsqueda de estudios (es decir, de publicaciones científicas) a incluir en el mapeo sistemático.
3. Definición de un esquema de clasificación.
4. Realización de un mapeo de los estudios según el esquema de clasificación definido que incluya el análisis y discusión de los resultados obtenidos.

Una descripción detallada de este proceso se muestra en la Figura 2 y se detalla a continuación:

- Paso 1. Alcance de la investigación. Este paso consiste en la definición de las preguntas que guiarán la investigación. Estas preguntas deben ser específicas y permitir profundizar en el área temática del mapeo sistemático. Además, cabe resaltar que estas preguntas de investigación deben responder a un proceso reflexivo previo a iniciar la revisión de los estudios que se incluirán en el mapeo sistemático. Estas preguntas se utilizarán en el último paso de análisis y discusión de resultados. Un ejemplo de pregunta de investigación para un mapeo sistemático sobre el control de calidad en sistemas de crowdsourcing sería: “¿qué mecanismos para controlar la calidad de los resultados se han propuesto en el crowdsourcing? [28].
- Paso 2. Proceso de búsqueda. Después de seleccionar las cadenas de búsqueda, se realiza una búsqueda de estudios (que incluye la selección de los resultados y la validación de la

búsqueda). Este proceso tiene por objetivo identificar los estudios primarios, que son relevantes dentro del ámbito específico de la investigación, para obtener resultados de los repositorios de investigación de una manera bien definida y repetible. Los criterios de inclusión y exclusión pueden ser variados y dependerán de la investigación que se esté realizando, un ejemplo de criterios de inclusión considerados en una gran parte de mapeos sistemáticos es: “publicaciones en inglés, revisadas por pares para las cuales el texto completo esté disponible y publicadas en revistas, conferencias, simposios o talleres”.

- Paso 3. Definición de un esquema de clasificación. El objetivo de este paso es definir las facetas (o aspectos), que permiten establecer categorías de clasificación relacionadas con las preguntas de investigación. Estas categorías deberán definirse con criterios objetivos debidamente justificados, permitiendo la clasificación de los estudios.
- Paso 4. Mapeo sistemático de estudios. Este paso consiste en obtener una clasificación de los estudios según el esquema de clasificación definido en el paso anterior, así como realizar un análisis detallado de los datos.

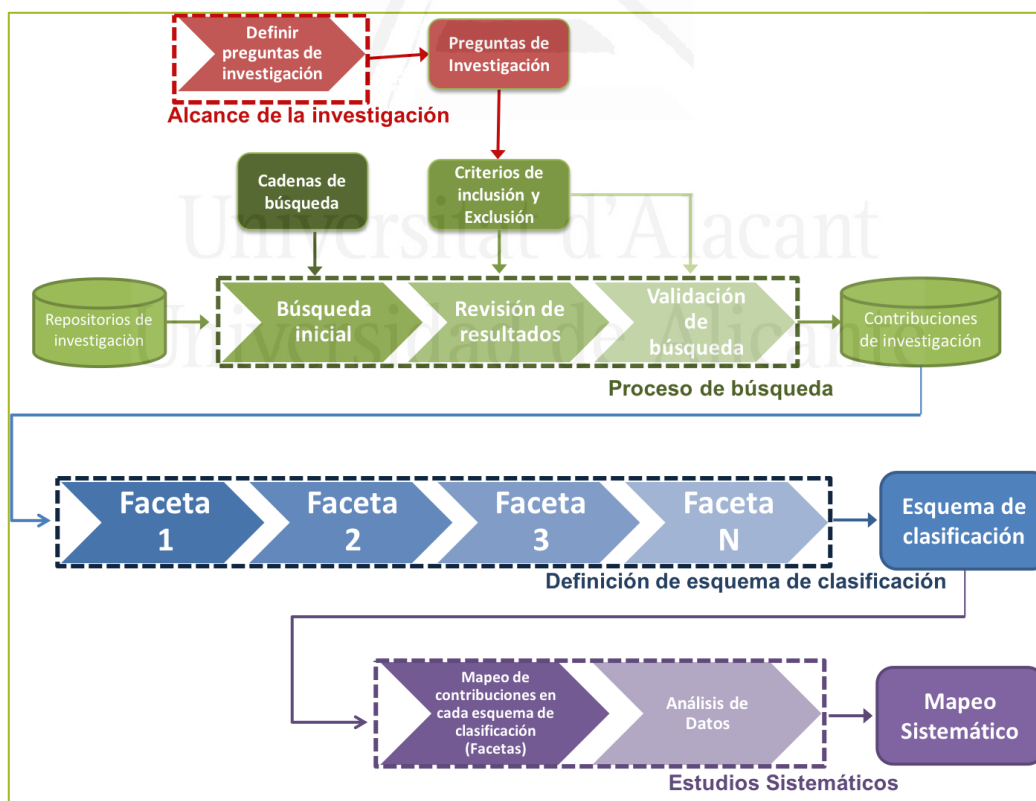


Figura 2: Proceso llevado a cabo para la realización del mapeo sistemático en la presente tesis doctoral, basado en [29], [30]

2.1.1 Amenazas a la validez del mapeo sistemático

En la literatura, las principales dificultades de los mapeos sistemáticos se han identificado de la siguiente manera: sesgo en la selección de publicaciones [31], errores en la categorización de las publicaciones en categorías detalladas [32] y errores duplicados o contribución adicional débil de una publicación a otra (los llamados “artículos delta” o “delta papers”, es decir, documentos que proporcionan sólo adiciones menores en comparación con trabajos publicados anteriormente por los autores). Durante el proceso detallado anteriormente se debe tener en cuenta estas amenazas a la validez del mapeo sistemático con el fin de obtener unas conclusiones acordes a la realidad en el área de investigación objeto de estudio.

2.2 Investigación en acción

La “investigación en acción” (o Action Research en inglés), es una aproximación metodológica aplicada al desarrollo de investigación en campos eminentemente prácticos como la Ingeniería de Software. De hecho, mediante el uso de “investigación en acción”, se facilita la participación de todas las partes involucradas en la investigación, examinando la situación existente (que sienten como problemática), con los objetivos de cambiarla y mejorarla, desde un punto de vista eminentemente de aplicación práctica [33].

Cabe destacar que la aproximación de “investigación en acción” está fuertemente orientada hacia la colaboración, por tanto, exige de un cambio que involucra tanto a los investigadores como a los sujetos, en un proceso en dos fases:

- La fase de diagnóstico implica un análisis colaborativo de la situación social por parte del investigador y de los sujetos de la investigación.
- La fase terapéutica implica un cambio colaborativo. En esta fase se introducen cambios y se estudian los efectos [34].

2.2.1 Proceso de la aproximación de “investigación en acción”

Las etapas que plantea la aproximación de “investigación en acción” según lo descrito en [34] son las siguientes:

- Establecer el propósito de cualquier acción, involucra que la teoría debe ser explícita antes de que se tome la acción, de lo contrario existe el riesgo de que la acción no tenga sentido.

- La existencia de una acción práctica en la resolución de problemas. Esto es necesario para revelar el valor relativo de la verdad de los conceptos teóricos que subyacen a la acción.
- La acción práctica debe informar la teoría. La teoría debe ser ajustada de acuerdo con el resultado práctico de la acción.
- La acción debe estar socialmente situada. Esta situación social implica que los investigadores deben ser observadores participantes.

La “investigación en acción” combina teoría y práctica (incluyendo tanto a investigadores como a profesionales) a través del cambio y la reflexión en una situación problemática inmediata dentro de un marco ético mutuamente aceptable. Esta aproximación metodológica es un proceso iterativo que involucra a investigadores y profesionales que actúan juntos en un ciclo particular de actividades, incluyendo el diagnóstico de problemas, la intervención para la acción y el aprendizaje reflexivo.

Un estudio cuidadoso de la literatura de “investigación en acción” permite establecer que esta aproximación metodológica tiene cuatro características comunes [35]:

- Orientación a la acción y al cambio.
- Enfoque de problemas.
- Proceso "orgánico" que implica etapas sistemáticas y a veces iterativas.
- Colaboración entre los participantes.

Las etapas de esta aproximación metodológica “investigación en acción” [35], se presentan en la Figura 3, y se describen a continuación:

- Articulación del problema. Se establece la problemática la cual es objeto de investigación y que se pretende resolver.
- Hipótesis dinámica. Se plantea la hipótesis que guiará el proceso de investigación.
- Diagnóstico. Se investiga el estado de la problemática identificada.
- Diseño de la acción. Se realiza el diseño de una acción que será evaluada mediante iteraciones
- Pruebas y resultados. Se analizan los resultados de cada etapa para realizar las mejoras respectivas.



Figura 3: Etapas de la investigación en acción

Por otro lado, según [36], los participantes en la investigación deben ser:

- Investigadores.
- Objeto investigado.
- Grupo crítico de referencia.
- Partes implicadas o beneficiarios.

Para el caso del presente trabajo de investigación se define los siguientes actores:

- Investigadores: la autora de esta memoria de investigación, así como directores de tesis doctoral.
- Objeto investigado: marco de referencia basado en estándares de calidad para la publicación de datos abiertos comprensibles.
- Grupo crítico de referencia: responsables de la publicación de datos relacionados con el gobierno, academia y sociedad civil en Ecuador.
- Partes implicadas o beneficiarios: ministerios, universidades y sociedad civil.

Los instrumentos que se utilizan dentro de este proceso son grupos de discusión y entrevistas que permitirán recolectar la información necesaria para la validación del marco de referencia diseñado.

3. ESTADO DEL ARTE EN DATOS ABIERTOS

El capítulo presenta un estudio del estado del arte en cuanto a la investigación desarrollada en torno al concepto de datos abiertos. Este estudio se realiza utilizando la metodología de mapeo sistemático. El objetivo es determinar el estado de las investigaciones en datos abiertos clasificándolas por impacto, dominio, tópico, ciclo de vida y tipo de investigación.

Universitat d'Alacant
Universidad de Alicante

3.1 Estructura del Estado del Arte

El estudio de estado del arte se desarrolla utilizando una metodología de mapeo sistemático (detallado en 2.1), que permite encontrar y clasificar las publicaciones científicas para conocer el estado de los datos abiertos con un enfoque tecnológico. Para este proceso se realizaron las siguientes definiciones orientadas a los datos abiertos:

- Alcance de la investigación.
- Proceso de búsqueda.
- Esquema de clasificación

Luego de obtener estas definiciones se realiza el proceso de búsqueda y se presentan los resultados obtenidos.

3.1.1 Alcance

El objetivo general del mapeo sistemático, realizado como parte de esta tesis doctoral, es proporcionar una visión general consolidada de la investigación en el campo de los datos abiertos desde una perspectiva tecnológica, a través de sus lugares de publicación de trabajos de investigación, dominios de impacto, fases de publicación de los datos, nivel de impacto y su evolución a lo largo del tiempo, para identificar las áreas con mayores oportunidades de investigación. El proceso responde a las siguientes preguntas de investigación relacionadas con datos abiertos y su evolución en el tiempo:

- ¿Dónde se han publicado los trabajos de investigación?
- ¿Qué impacto han tenido los trabajos de investigación?
- ¿Qué ámbitos, dentro del campo de los datos abiertos, recibieron mayor cobertura?
- ¿Cuáles son los temas que se abordan en los trabajos de investigación?
- ¿Qué fases del ciclo de vida de los datos se han considerado en los trabajos de investigación?
- ¿A qué tipos de investigación corresponden los trabajos realizados?

Estas preguntas permiten definir el alcance de la investigación de acuerdo a la metodología detallada en el capítulo 2 de este documento.

3.1.2 Proceso de búsqueda

Se realizó un proceso de búsqueda en tres etapas para identificar las publicaciones de trabajos de investigación (o estudios primarios) en el campo de los datos abiertos. A continuación, se detallan cada uno de los pasos de este proceso.

PASO 1: Búsqueda inicial. Para obtener el conjunto inicial de estudios primarios potencialmente relevantes (es decir, “publicaciones”), se realiza una búsqueda exhaustiva, utilizando motores de búsqueda científica de los repositorios más importantes de editoriales dentro del campo de la tecnología de la información (de acuerdo con la perspectiva tecnológica de este estudio); por lo tanto, se selecciona los motores de búsqueda incluidos en la Tabla 1 .

Todas las búsquedas se realizaron entre el 1 de septiembre de 2016 y el 20 de enero de 2018.

Tabla 1: Motores de búsqueda.

<i>Repositorios científicos.</i>	<i>Url</i>
Springer RD	http://rd.springer.com/
IEEE Xplore	http://ieeexplore.ieee.org/ .
ACM Digital Library	http://dl.acm.org
IOS Press	http://content.iospress.com
Science Direct	http://www.sciencedirect.com/

Dado que el objetivo es estudiar las publicaciones sobre datos abiertos y cómo este término se utiliza específicamente en el área de investigación, se definió que el título debe contener la cadena de búsqueda “open data”.

Se incluyen publicaciones entre 2006 y 2017. Sólo se tiene en cuenta las publicaciones en inglés. Los motores de búsqueda general de publicaciones científicas como Google Scholar no son considerados por dos razones:

- Indexan las publicaciones de los repositorios mencionados anteriormente
- Enumeran una gran cantidad de literatura gris (es decir, trabajos de investigación producidos fuera de los canales tradicionales de publicación y distribución académica,

tales como informes, documentos de trabajo, documentos gubernamentales, libros blancos, etc.).

Una vez realizada la búsqueda con la cadena propuesta (“open data”), los motores de búsqueda producen 671 resultados de búsqueda. Estos son posteriormente clasificados, cuantificados y presentados en el análisis de resultados, basados en las diferentes facetas derivadas de las preguntas de la investigación.

PASO 2: Revisión de los resultados. El conjunto inicial de publicaciones contiene resultados irrelevantes que hay que descartar. Se examina la relevancia de cada publicación con respecto a los objetivos de investigación con el fin de proceder a su selección (cabe destacar que en un primer momento se debe eliminar los resultados duplicados). Para ello, se definen criterios de inclusión y exclusión, tanto en la forma como en el contenido que se describen a continuación:

- ***Criterios de inclusión.*** Todas las publicaciones revisadas por pares en inglés para las cuales el texto completo estaba disponible y que en su título tenga la cadena “open data”. Esto incluye todos los trabajos de investigación (tanto cortos como completos) publicados en revistas, conferencias, simposios o talleres revisados por pares. Los documentos deben haber sido publicados entre 2006 y 2017 (doce años de análisis).
- ***Criterios de exclusión.*** Publicaciones procedentes de fuentes que no pasan por un proceso de revisión por pares o que no constituyen contribuciones puras de investigación, por ejemplo: libros, tesis doctorales y de maestría, descripciones de patentes, normas y recomendaciones, resúmenes de libros o tesis, informes técnicos, libros blancos, charlas invitadas, documentos de demostración, documentos de tutoría, publicaciones de pósteres, editoriales, prefacios, artículos o columnas en revistas que no han sido revisadas por pares, boletines informativos, entradas de enciclopedias, resúmenes o entradas de blog. Además, los estudios para los que no se publicó el texto completo, como los resúmenes, los resúmenes ampliados y las presentaciones (presentaciones de diapositivas). También se excluyen documentos que no se centran en la publicación o reutilización de datos abiertos, por ejemplo, los relacionados con la publicación de datos institucionales dirigidos únicamente a usuarios internos o aquellos en los que se utilizaban datos abiertos en algunos experimentos, es decir, el contenido no identificaba un problema de investigación relacionado con la reutilización o publicación de datos abiertos, sino que sólo mencionaba de forma incidental los datos abiertos.

PASO 3: Validación de la búsqueda. Para validar el proceso de selección, los criterios de inclusión y exclusión se verifican para cada uno de los resultados de la búsqueda de la siguiente manera: para asegurar decisiones correctas, los autores se dividen en dos grupos; cada publicación del conjunto inicial se asignó a un grupo mientras que el otro grupo revisa todo el proceso. En caso de desacuerdo, ambos grupos discuten los resultados hasta que llegan a un consenso. De esta manera, la cantidad de publicaciones relevantes se reduce finalmente en no más de un 5%.

3.1.3 Esquema de clasificación

Una vez obtenido el conjunto final de publicaciones, se elabora un esquema de clasificación correspondiente a las cuestiones planteadas en el alcance de la investigación. Para esto se consideran seis facetas: lugar, impacto, dominio, temas, fase del ciclo de vida de los datos y tipo de investigación. A continuación, se detalla cada una de estas facetas.

FACETA 1: Lugar. Se identifica donde se ha realizado la publicación. Para ello, tras aplicar el proceso de búsqueda, cada motor de búsqueda identifica la fuente de la que procede la publicación, clasificándola por congreso o revista científica. También se cuantifica el número de publicaciones por congreso o revista científica identificadas.

FACETA2: Impacto. Describe el alcance de cada publicación identificada en el proceso de búsqueda. La clasificación se realiza de la siguiente manera:

- Impacto local: describe un área pequeña y focalizada a la que se refieren los datos abiertos dentro de la publicación, por ejemplo: una ciudad, un museo, una universidad o un hospital.
- Impacto regional: la publicación describe un área, mayor que la local, relacionada con los datos abiertos, como una provincia o un estado.
- Impacto nacional: la publicación describe un país.
- Impacto internacional: los datos abiertos dentro de la publicación tienen un carácter internacional (es decir, no restringido a un área determinada).

FACETA 3: Dominio. Esta faceta está relacionada con la cobertura en la investigación de datos abiertos, es decir la temática de las publicaciones científicas a considerar según el dominio temático de los datos abiertos correspondientes. Los dominios considerados provienen de las 14 categorías de datos sugeridas en la Carta de Datos Abiertos del G8 [37] y [38], se describen a continuación:

- Agricultura: publicaciones que tratan de seguridad alimentaria, agricultores y consumidores finales, desarrollo agrícola sostenible, nutrición
- Biología: artículos que tratan del estudio de la vida y de los organismos vivos.
- Química: publicaciones que tratan sobre la materia, la composición material y la reacción.
- Cultura: publicaciones sobre el comportamiento social y las normas que se encuentran en las sociedades humanas.
- Periodismo de datos: publicaciones sobre enfoques para la narración y el periodismo, actividades periodísticas.
- Economía: movimientos financieros, ingresos, gastos y presupuesto.
- Educación: aplicaciones educativas, rendimiento escolar, habilidades digitales.
- Medio ambiente: medio ambiente, cambio climático, recursos naturales, información medioambiental, paisajes forestales, meteorología/clima, agricultura, silvicultura, pesca y caza, niveles de contaminación, consumo de energía.
- Geoespacial: topografía, códigos postales, mapas nacionales, mapas locales, seguridad alimentaria, lagos, zonas geográficas.
- Salud: publicaciones relacionadas con el estudio del estado de completo bienestar físico, mental y social y no simplemente la ausencia de enfermedad.
- Humanitario: asistencia humanitaria, gestión de desastres, actividades de socorro y reconstrucción, desastres, epidemias.
- Infomediarios: empresarios e inversores que crean empresas para identificar y aprovechar el valor de mercado de la información al consumidor.
- Innovación: publicaciones en torno al objetivo de obtener un impacto económico o social a partir de datos abiertos.
- Ciencia: investigación y descubrimientos científicos, métodos innovadores para abrir datos científicos y crear nuevas herramientas para manipularlos, financiación de proyectos científicos, colaboración entre grupos de interesados.
- Transporte: horarios de transporte público, puntos de acceso y penetración de la banda ancha.

- Energía: aplicaciones orientadas al desarrollo energético.
- Turismo: aplicaciones de datos abiertos aplicadas al desarrollo turístico.

FACETA 4: Tópico. Se relaciona con los aspectos que se abordan en la investigación. Estos temas se recopilan a partir de las “Solicitudes de Artículos” (o "Call for Papers", CFP) de dos de las conferencias científicas más importantes sobre datos y web semántica (es decir, dos de los tópicos más importantes dentro del ámbito de los datos abiertos), a saber, la International Conference on Very Large Data Bases (VLDB) y la International Semantic Web Conference (ISWC). En relación con estas dos conferencias, se realiza una cuantificación y se establecen los tópicos más recurrentes en las convocatorias de ambas conferencias en los últimos tres años (2015, 2016 y 2017), dando lugar al resultado siguiente:

- **Emprendedor:** uso empresarial de datos abiertos. Uso e impacto de los datos abiertos en países o sectores específicos con el fin de potenciar nuevos modelos de negocio o conocer su impacto económico.
- **Gobierno:** elaboración, aplicación e institucionalización de políticas de datos abiertos; creación de capacidad para una mayor disponibilidad y utilización de datos abiertos; conceptualización de los ecosistemas e intermediarios de datos abiertos; vínculos entre la transparencia, la libertad de información y las comunidades de datos abiertos; medición de las políticas y prácticas de datos abiertos, incluidos los métodos para evaluar el impacto de los datos abiertos; y ubicación de los datos abiertos en el contexto de la gobernanza mundial y el desarrollo.
- **Recuperación de información:** base de datos, recuperación de información, extracción de información, procesamiento del lenguaje natural y técnicas de inteligencia artificial para la web semántica, búsqueda y consulta de la web semántica, bases de datos difusas, probabilísticas y aproximadas, recuperación de información, texto en bases de datos.
- **Infraestructura:** Web Semántica y datos enlazados para entornos “en la nube”, métodos de acceso, control de concurrencia, recuperación, transacciones, indexación y búsqueda, gestión de datos en memoria, aceleradores de hardware, procesamiento y optimización de consultas, gestión de almacenamiento, ajuste, benchmarking, medición del rendimiento, administración y gestión de bases de datos, base de datos como servicio.

- Sistemas inteligentes: gestión de datos gráficos, redes sociales, sistemas de recomendación, inteligencia de negocio.
- Internet de las cosas (IoT): flujos de datos e IoT, crowdsourcing, bases de datos integradas y móviles, bases de datos en tiempo real, sensores, bases de datos de flujos.
- Calidad: limpieza, aseguramiento de la calidad y procedencia de los datos, servicios y procesos de la web semántica, limpieza de datos, filtrado y difusión de información, integración de información, gestión de metadatos, descubrimiento de datos, gestión de datos web, web semántica, sistemas de bases de datos heterogéneos y federados, minería y análisis de datos, almacenamiento.
- Seguridad: confianza, privacidad y seguridad de la web semántica, privacidad y seguridad en la gestión de datos, retos críticos para los datos abiertos: privacidad, exclusión y abuso.
- Web semántica: creación de gráficos de conocimiento, razonamiento, uso, representación del conocimiento y razonamiento en la web, gestión escalable de semántica y datos en la web, incluyendo datos enlazados, análisis de datos de web semántica, lenguajes, herramientas y metodologías para representar y gestionar semántica y datos en la web, arquitecturas y algoritmos para volúmenes extremos, heterogeneidad, dinamicidad y descentralización de datos de web semántica, acceso a datos basados en ontologías e integración/intercambio en la web, ingeniería ontológica y patrones ontológicos para la web, modularidad ontológica, mapeo, fusión y alineación para la web, apoyo al multilingüismo en la web semántica, interfaces de usuario e interacción con la semántica y los datos en la web, visualización de información y métodos de análisis exploratorio para los datos de la web semántica, acceso personalizado a los datos y aplicaciones de la web semántica, métodos y aplicaciones de la semántica social.
- Ingeniería de software: Tecnologías semánticas para plataformas móviles, sistemas de bases de datos distribuidos, gestión de datos en nube, NoSQL, análisis escalables, transacciones distribuidas, consistencia, p2p y gestión de datos en red, desarrollo de software y redes de entrega de contenidos.
- Visualización: modelos de datos y lenguajes de consulta, gestión y diseño de esquemas, usabilidad de bases de datos, interfaces de usuario y visualización.

FACETA 5: Ciclo de vida de los datos abiertos. Esta faceta se encarga de describir el proceso de convertir datos en datos abiertos, es decir, preparar los datos que se van a publicar, utilizar los datos publicados y conservarlos. Por lo tanto, se ocupa principalmente de tres cuestiones: el pretratamiento, la explotación y el mantenimiento.

Para este propósito se tuvieron en cuenta varios trabajos de investigación [30], [39] y [37] que definen el ciclo de vida de los datos abiertos. A continuación, se describen los diferentes conceptos que permiten clasificar las publicaciones dentro de esta faceta:

- Creación de datos: se refiere a la generación de datos, así como a la recopilación de datos con el propósito específico de publicarlos.
- Selección de datos: es el proceso de selección de los datos a publicar. Esto requiere la eliminación de cualquier dato privado o personal, así como la identificación de las condiciones bajo las cuales estos datos serán publicados, potencialmente a través de la especificación de políticas de datos abiertas.
- Armonización de datos: este proceso implica la preparación de los datos que se publicarán para que se ajusten a los estándares de publicación.
- Publicación de datos: es el acto concreto de abrir los datos mediante su publicación en portales de datos abiertos.
- Interconexión de datos: Se refiere al proceso de vinculación de datos. Este proceso permite que los datos publicados tengan un valor adicional, relacionado con la característica de estar conectados y consultados considerando sus diferentes fuentes.
- Descubrimiento de datos: la publicación de los datos no es suficiente para permitir su reutilización. Los consumidores de datos deben poder ser conscientes de la existencia de datos abiertos para poder consumirlos (es decir, deben poder encontrarlos fácilmente).
- Exploración de datos: este paso es la forma más trivial de consumir datos. Aquí, un usuario examina pasivamente los datos abiertos visualizándolos o escudriñándolos.
- Explotación de datos: este paso es una forma más avanzada de consumir datos. La explotación de datos permite a un usuario utilizar, reutilizar o distribuir los datos abiertos de forma proactiva mediante la realización de análisis, el emprendimiento o la innovación basada en los datos abiertos.

- Curado de datos: aunque no necesariamente ocurre en una etapa fija, la curaduría de datos es vital para asegurar que los datos publicados sean reutilizables. Esto implica una serie de procesos, incluida la actualización de datos obsoletos, el enriquecimiento de datos y metadatos, la limpieza de datos, etc.

FACETA 6: Tipo de investigación. Considerando que no es específico para el tema particular de los datos abiertos, sino que es de aplicación general. Como en el caso de otros mapeos sistemáticos realizados en ingeniería de software, se utiliza el esquema de clasificación propuesto en [40]. El esquema de clasificación incluye:

- Propuesta de solución: describe una solución normalmente ilustrada con un ejemplo, un caso de estudio, un ejemplo de ejecución, etc. El trabajo está poco o nada validado (véase el siguiente punto); la propuesta sólo se explica y se describe su aplicación.
- Investigación de validación: validación de la investigación que no se lleva a cabo en un entorno práctico. Se incluyen experimentos, realización de tipos de pruebas, estudios de laboratorio, etc. Normalmente sigue a una propuesta de solución. Responde a la pregunta: ¿La solución propuesta es "buena"?
- Investigación de evaluación: una evaluación de la investigación, generalmente observando cómo funciona la solución en la práctica o comparándola con otras soluciones, señalando puntos positivos y negativos. Es más extensa que la validación y a menudo se lleva a cabo en un entorno industrial. Responde a la pregunta: ¿Es la solución propuesta la solución "correcta"?
- Propuestas filosóficas o conceptuales: esbozan una nueva forma de ver las cosas existentes, proporcionando una visión holística sobre un tema.
- Documento de opinión: describe la opinión de los autores, que suelen adoptar una postura positiva o negativa. También puede presentar una visión general de un campo o una comparación de técnicas desde el punto de vista del autor. Generalmente no se basa en un trabajo relacionado o en una metodología de investigación.
- Documento de experiencia: describe la experiencia de los autores, generalmente en la práctica, utilizando un determinado método, tecnología, etc. Los autores suelen ser personas que trabajan en la industria [27].

Basándose en el esquema de clasificación mencionado, se clasifica cada una de las 671 publicaciones relevantes de acuerdo con cada faceta. Para asegurar la corrección y consistencia del proceso de clasificación, los investigadores (según la metodología de investigación descrita en el capítulo 2 de esta memoria de tesis doctoral) son divididos en dos grupos y cada una de las 671 publicaciones se asigna a un grupo para ser revisada y clasificada dentro de cada faceta. En el caso de resultados conflictivos, se pide al otro grupo que realice una clasificación y los resultados se discuten hasta que se llega a un consenso final. Cabe reseñar que un total del 10% de las 671 publicaciones requirieron discusión.

3.2 Amenazas a la validez

Las amenazas a la validez consisten en varios factores que pueden poner en riesgo la representatividad de los hallazgos del mapeo sistemático desarrollado. Para mitigar este riesgo se realizan una serie de acciones en el mapeo sistemático. En primer lugar, se tiene en cuenta una selección adecuada de motores de búsqueda para cubrir el área de investigación específica (datos abiertos desde una perspectiva tecnológica). En segundo lugar, se tiene que asegurar que se encuentran todas las publicaciones relevantes con la búsqueda realizada. Para ello, se lleva a cabo un paso de validación de la búsqueda (ver 3.1.2) diseñado para identificar las publicaciones que faltan después de revisarlas. Además, el proceso de revisión formal que se lleva a cabo según se comenta en la sección anterior, minimiza los riesgos en la validez de los resultados del mapeo sistemático.

Finalmente, para mitigar el riesgo de incluir estudios que no tengan suficientes contribuciones respecto a otros o que incluso, tengan contribuciones duplicadas (los denominados “artículos delta” tal y como se comentan en el capítulo 2), se trabaja en los resúmenes de cada publicación, discutiendo las contribuciones de las mismas con el fin de determinar si se debe incluir en el mapeo sistemático. Se encuentran 6 artículos delta de un total de 671 publicaciones (= 1%), sobre todas las categorías en cada faceta (de ellas, sólo una fue encontrada como duplicada). Consecuentemente, se considera un porcentaje insignificante en el global de publicaciones consideradas en el mapeo sistemático, determinando que esta amenaza no atenta contra la validez del trabajo realizado.

3.3 Mapeo Sistemático

La metodología de mapeo sistemático utilizada permite identificar todas las publicaciones relacionadas con datos abiertos, definir un esquema de clasificación, categorizar, analizar y

discutir la investigación seleccionada desde una perspectiva tecnológica. Se realiza con un análisis de los datos de las 671 publicaciones, se utilizaron diferentes tipos de gráficos para diferentes propósitos, respondiendo así a diferentes preguntas de investigación.

Se utilizan gráficos de barras apiladas para representar los resultados por año. Estos gráficos permiten: visualizar las publicaciones más frecuentes y la periodicidad de las diferentes categorías dentro de una misma faceta, visualizar el peso relativo de cada faceta y la evolución de cada categoría en el tiempo.

Los gráficos de barras tienen por objeto visualizar la distribución de las publicaciones por lugar de celebración y por año.

Los gráficos de burbujas se utilizan para mostrar las relaciones entre las diferentes facetas y, como tales, representan los mapas sistemáticos de la investigación de datos abiertos. Estos gráficos se utilizan tradicionalmente para este propósito porque representan cuatro dimensiones de datos donde cada burbuja representa la frecuencia de publicación con respecto a dos facetas.

3.3.1 Faceta 1: Lugares de publicación

En la Figura 4, se muestran los lugares donde se publican la mayor parte de los, de los 671 artículos detectados en el mapeo sistemático, incluyendo revistas y conferencias. Se pueden identificar 11 revistas con 3 o más publicaciones sobre datos abiertos. Las revistas “Semantic Web” y “Government Information Quarterly” fueron las más importantes, según el número de publicaciones, para la difusión de la investigación de datos abiertos desde una perspectiva tecnológica.

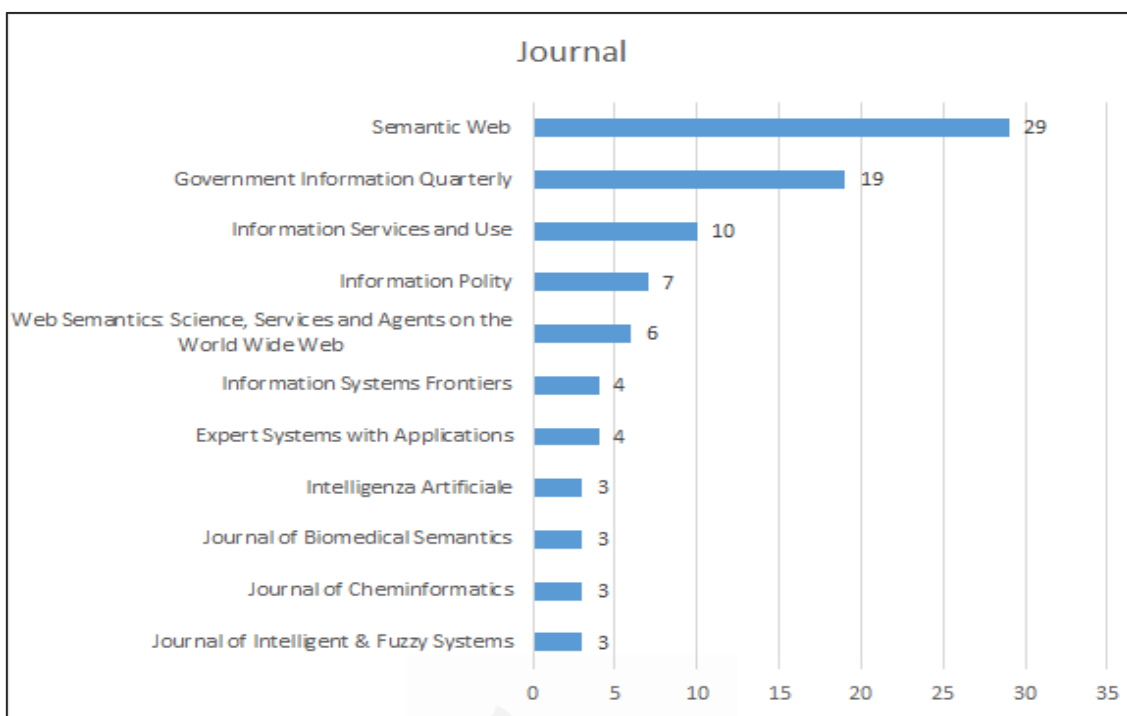


Figura 4: Revistas con mayor número de publicaciones sobre datos abiertos

En la Figura 5, se pueden visualizar 13 conferencias donde se publica el mayor número de los 671 artículos revisados. Se destacan las dos primeras conferencias: DG. O (Conference on Digital Government Research) y ICEGOV (International Conference on Theory and Practice of Electronic Governance).

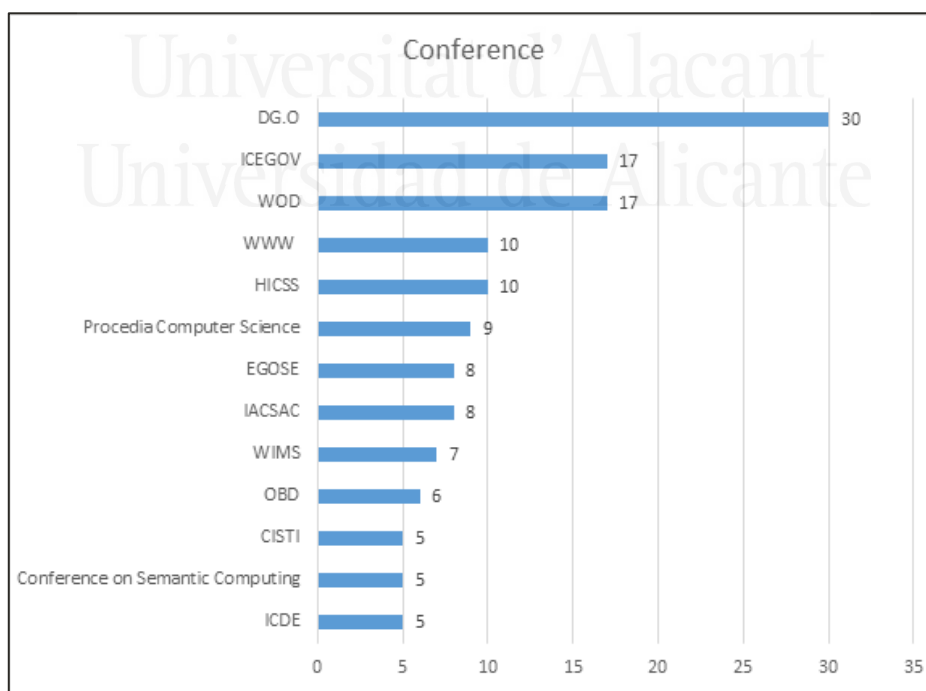


Figura 5: Conferencias con mayor número de trabajos publicados sobre datos abiertos

Tabla 2: Listado de acrónimos por conferencia

<i>Conferencia</i>	<i>Acrónimo</i>
International Digital Government Research Conference	DG. O
International Conference on Theory and Practice of Electronic Governance.	ICEGOV
Proceedings of the International Workshop on Open Data	WOD
International Conference on the World Wide Web Companion	WWW
Hawaii International Conference on System Sciences	HICSS
Electronic Governance and Open Society	EGOSE
IEEE Annual Computer Software and Applications Conference	IACSAC
International Conference on Web Intelligence, Mining and Semantics	WIMS
International Conference on Open and Big Data	OBD
Iberian Conference on Information Systems and Technologies	CISTI
IEEE International Conference on Data Engineering	ICDE

La Figura 6 muestra un gráfico circular que representa la distribución de las publicaciones consideradas en este mapeo sistemático según el repositorio desde donde son accesibles. El repositorio científico que más contribuye es IEEE con un 37%, seguido por ACM con un 33%. El resto representa en conjunto el 30% de las publicaciones. Notablemente, los dos primeros repositorios son las más relacionadas con las tecnologías de información y comunicaciones (TIC): esto confirma la relevancia de una perspectiva tecnológica cuando se examinan los estudios de datos abiertos.

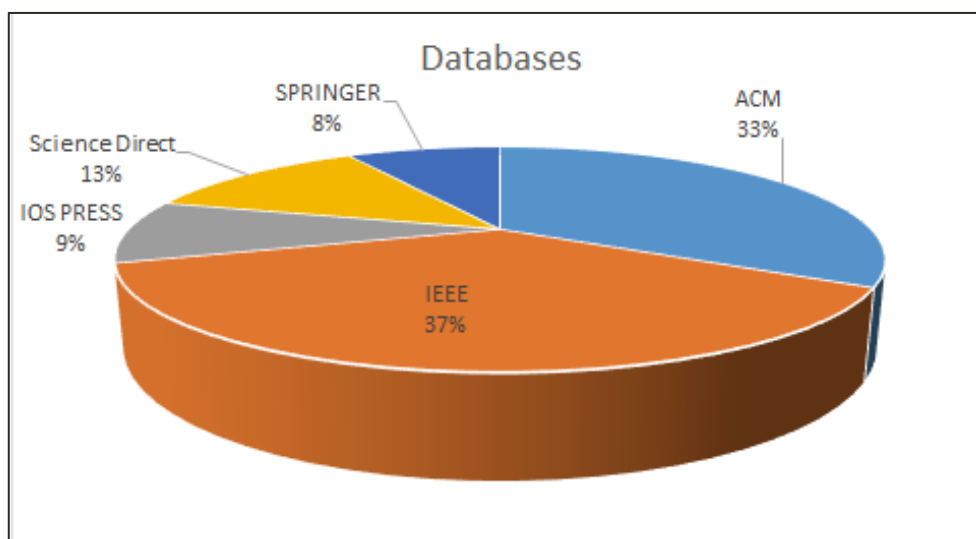


Figura 6: Porcentaje de publicaciones según el repositorio científico donde están accesibles

3.3.2 Faceta 2: Impacto

En la Figura 7 se muestran los porcentajes relativos al impacto de la publicación según lo detallado en la faceta 2; las publicaciones enmarcadas en un ámbito internacional fueron las más frecuentes con un 69% del total.

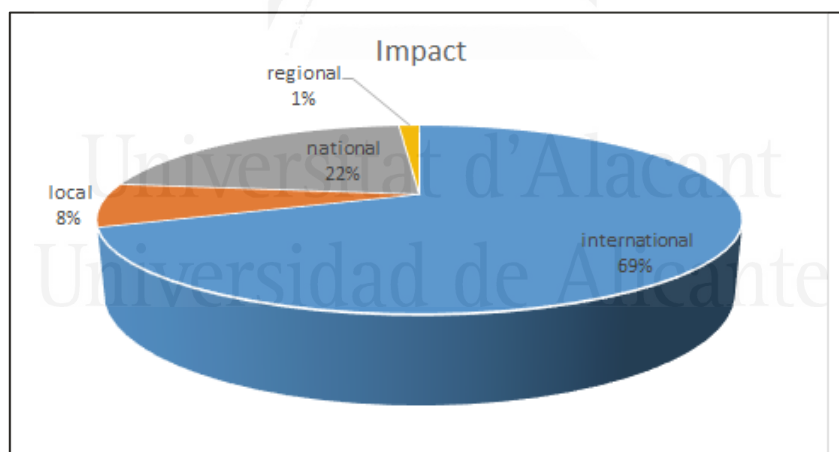


Figura 7: Publicaciones por impacto

3.3.3 Faceta 3: Dominio

En la Figura 8 y Figura 9 se presenta la distribución de las publicaciones consideradas en este mapeo sistemático según el dominio (faceta 3), mostrando así el progreso de las publicaciones entre 2006 y 2017. El dominio infomediario es el más relevante, con un 49,9% de publicaciones pertenecientes al dominio. Los otros dominios contribuyeron con menos del 8% cada uno, lo que sugiere que las publicaciones se dirigen principalmente a solventar problemas desde la perspectiva de los consumidores de datos abiertos.

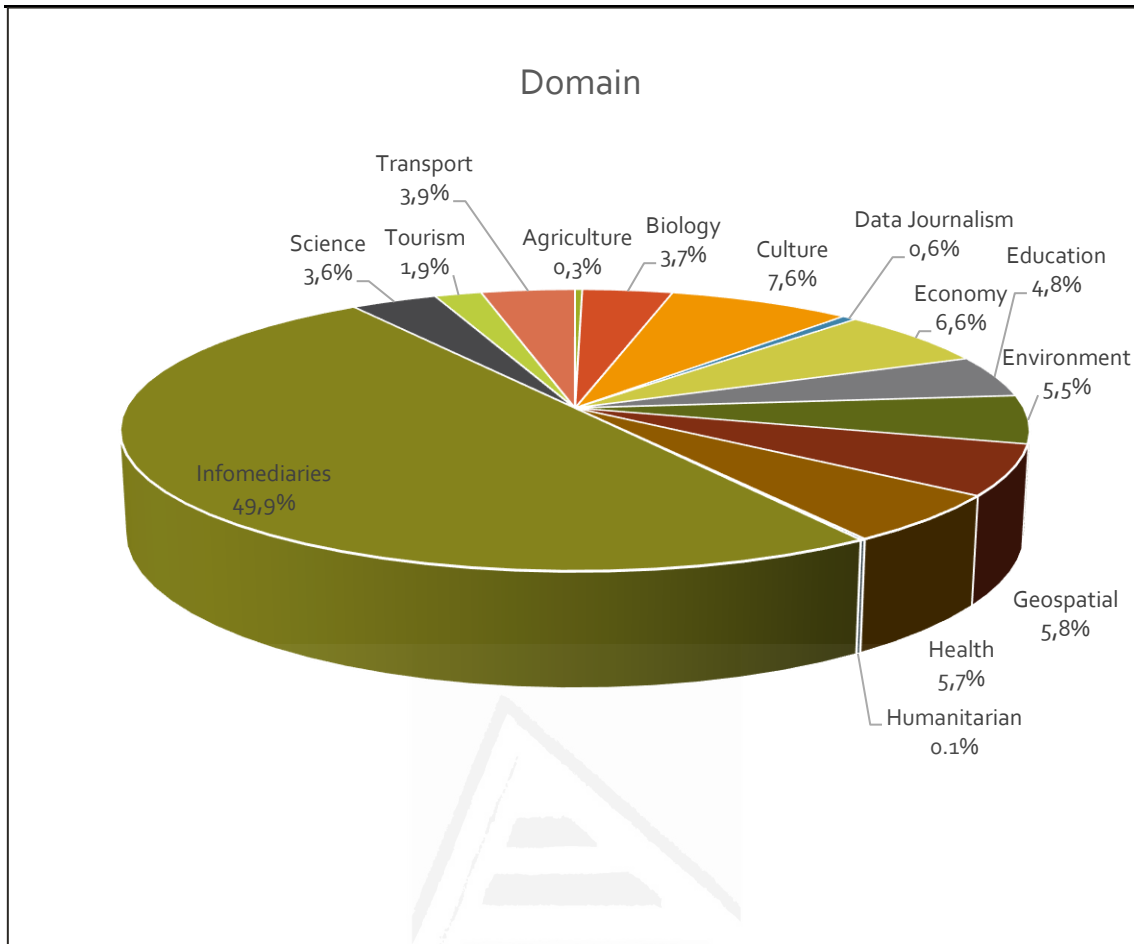


Figura 8: Porcentajes de publicaciones según su dominio

Las cifras muestran que las publicaciones relacionadas con los campos de transporte, economía, educación, medio ambiente, cultura, salud y geoespacial han aumentado significativamente en la misma proporción en los últimos siete años. No hay publicaciones que reflejen que se ha llevado a cabo investigaciones en estos campos entre 2006 y 2009. Los datos del campo del periodismo y el ámbito humanitario no han generado interés en la comunidad investigadora.

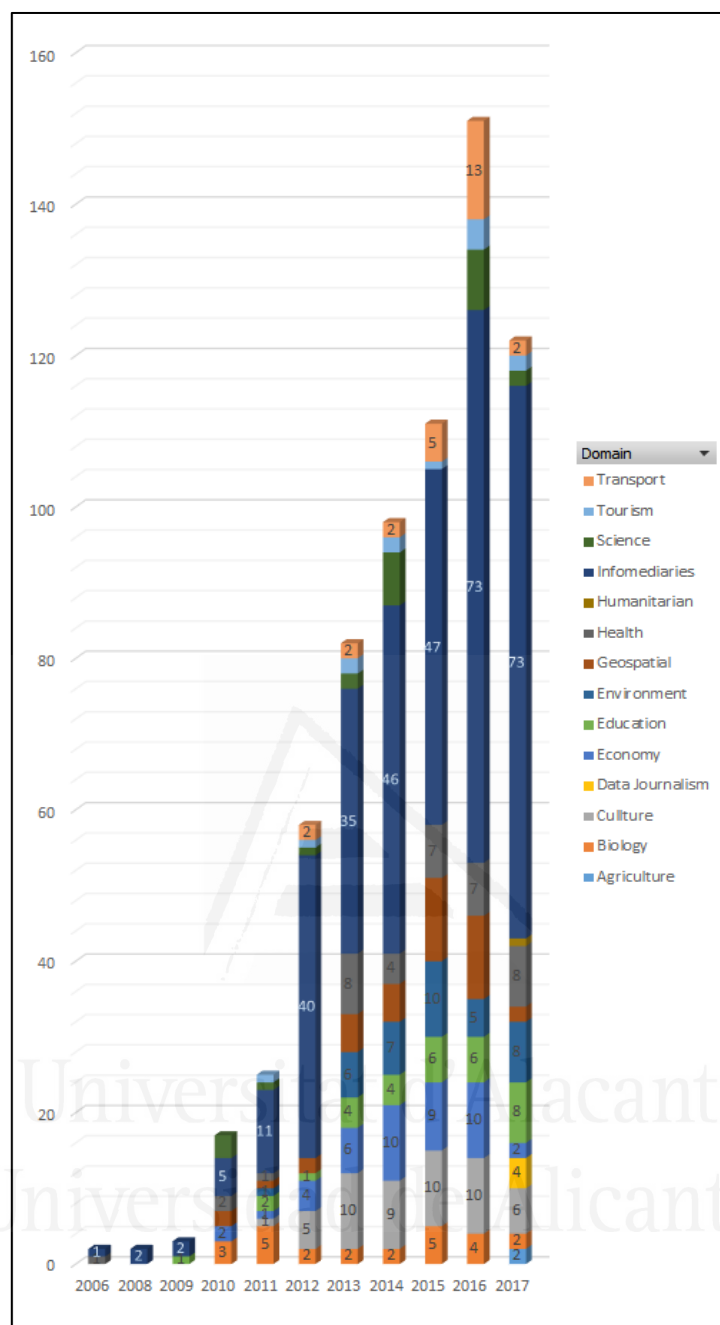


Figura 9: Distribución de las publicaciones por año y dominio

3.3.4 Faceta 4: Tópico

La Figura 10 y Figura 11 ilustran la distribución de las publicaciones según el tópico y muestran cómo evolucionan las publicaciones entre 2006 y 2017. El tema de la web semántica recibió mayor atención (24% de las publicaciones), seguido por la ingeniería de software (22%) y el gobierno (16%).

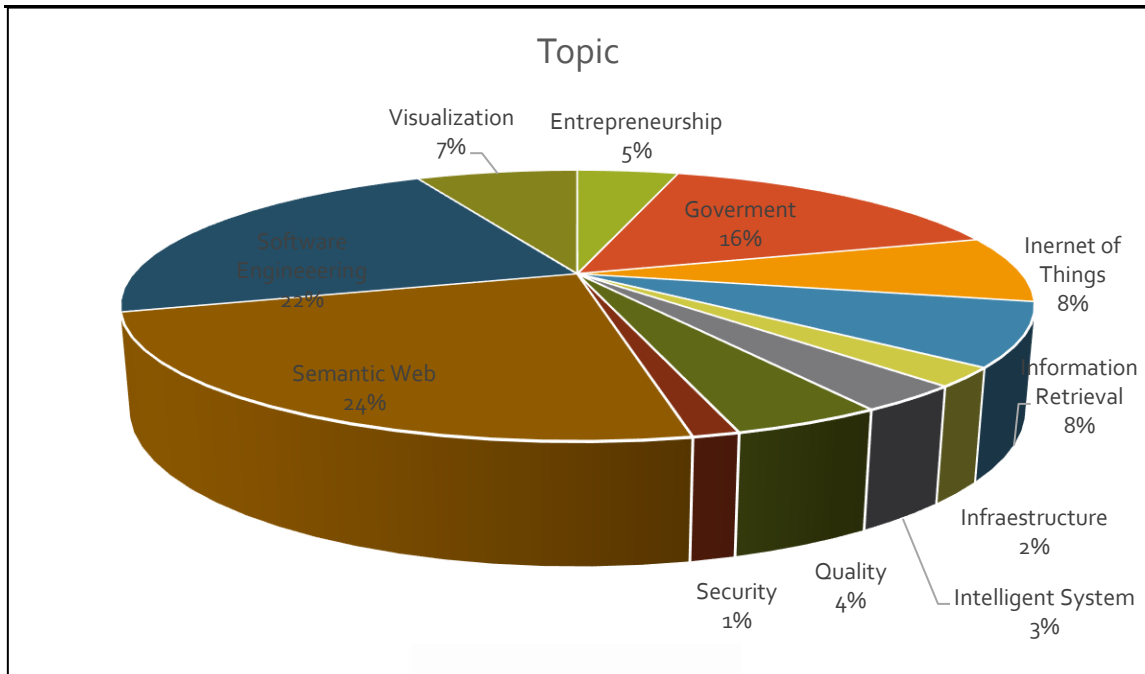


Figura 10: Porcentaje de publicaciones según su tópico

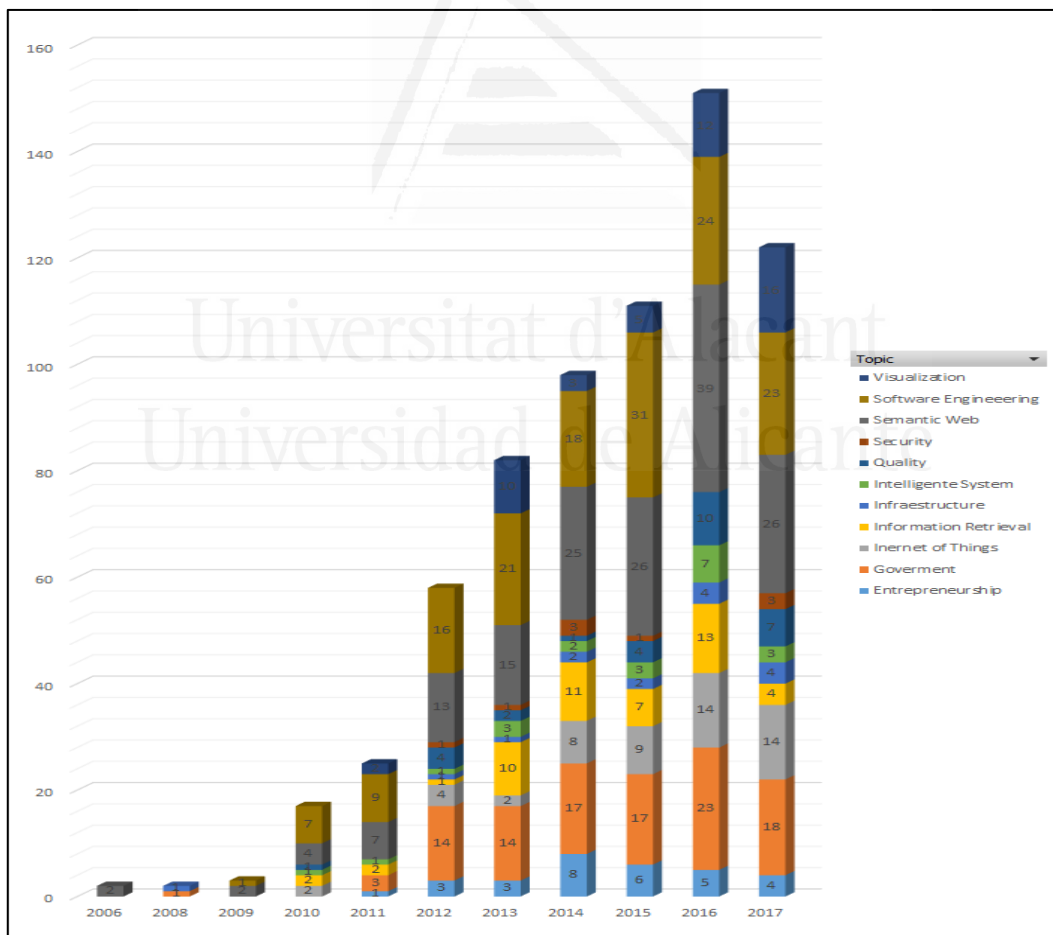


Figura 11: Distribución de publicaciones según su tópico entre el 2006 y 2017

Sin embargo, los temas de calidad y seguridad son los menos abordados con un 4% y 1% de publicaciones, respectivamente.

3.3.5 Faceta 5: ciclo de vida de los datos

La Figura 12, representa la distribución de las publicaciones según el ciclo de vida. La mayoría de los estudios se centran en la fase de explotación de datos (43% de las publicaciones). La siguiente fase es la exploración de datos (19% de las publicaciones). Esto sugiere que las publicaciones que buscan extraer valor de los datos y los resultados aplicables son las más numerosas.

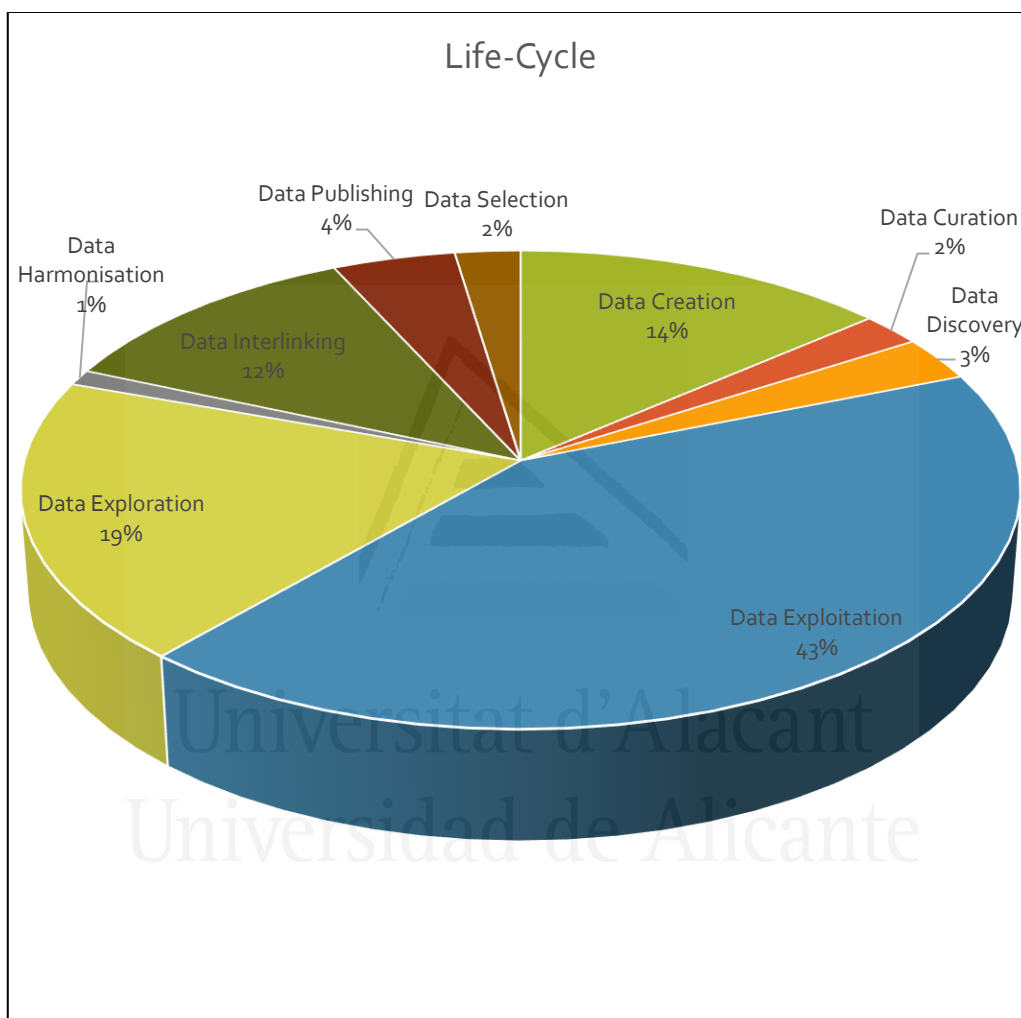


Figura 12: Distribución de las publicaciones según el ciclo de vida de los datos

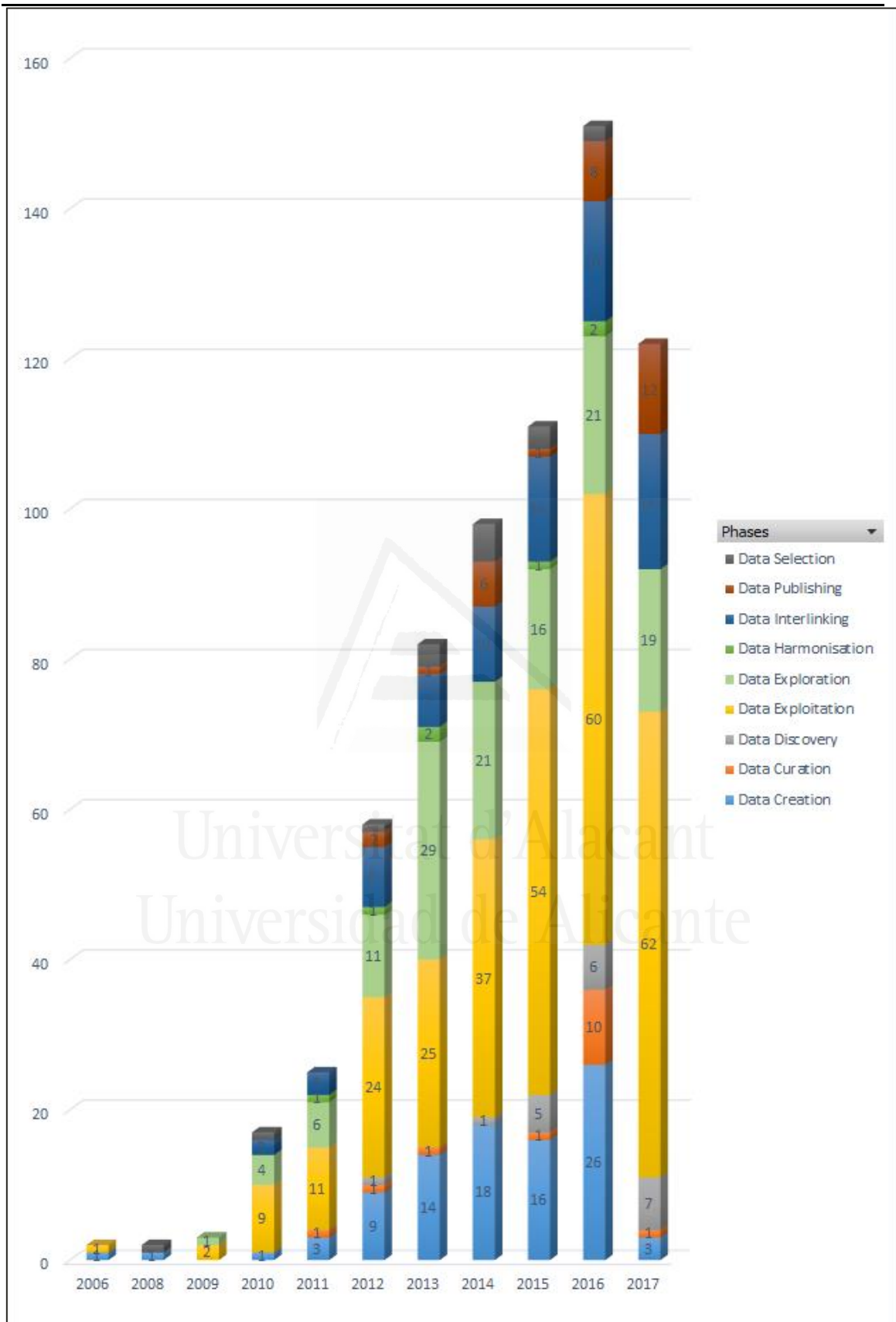


Figura 13: Distribución de las publicaciones según el ciclo de vida de los datos de 2006 a 2017

La Figura 13 muestra la distribución anual de las publicaciones por año. Se observa que las publicaciones sobre la fase de explotación crecen notablemente desde 2009 donde existen dos artículos científicos hasta un total de 62 publicaciones en 2017. Las fases de armonización de datos, selección de datos y curación de datos tienen una tasa de progresión de menos del 3%.

3.3.6 Faceta 6: tipo de investigación

La Figura 14 y la Figura 15 presentan la distribución de las publicaciones según el tipo de investigación, mostrando el progreso de las publicaciones desde 2006 hasta 2017. Se puede observar que el tipo más frecuente es el de publicaciones que proponen soluciones con un 53% de las publicaciones, que representa una mayoría, seguido por publicaciones sobre “validación de la investigación con un 27%, un porcentaje muy inferior pero importante.

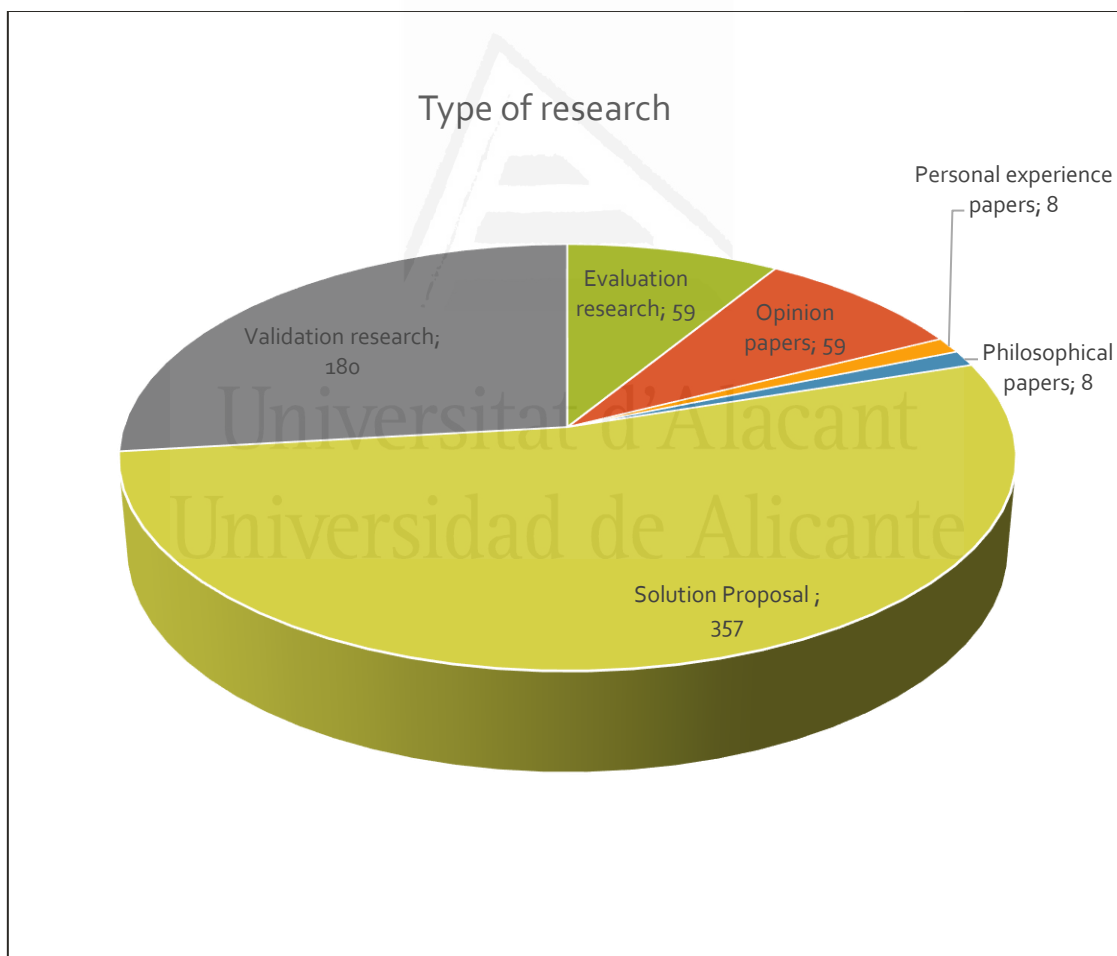


Figura 14: Porcentajes de publicaciones por tipo de investigación

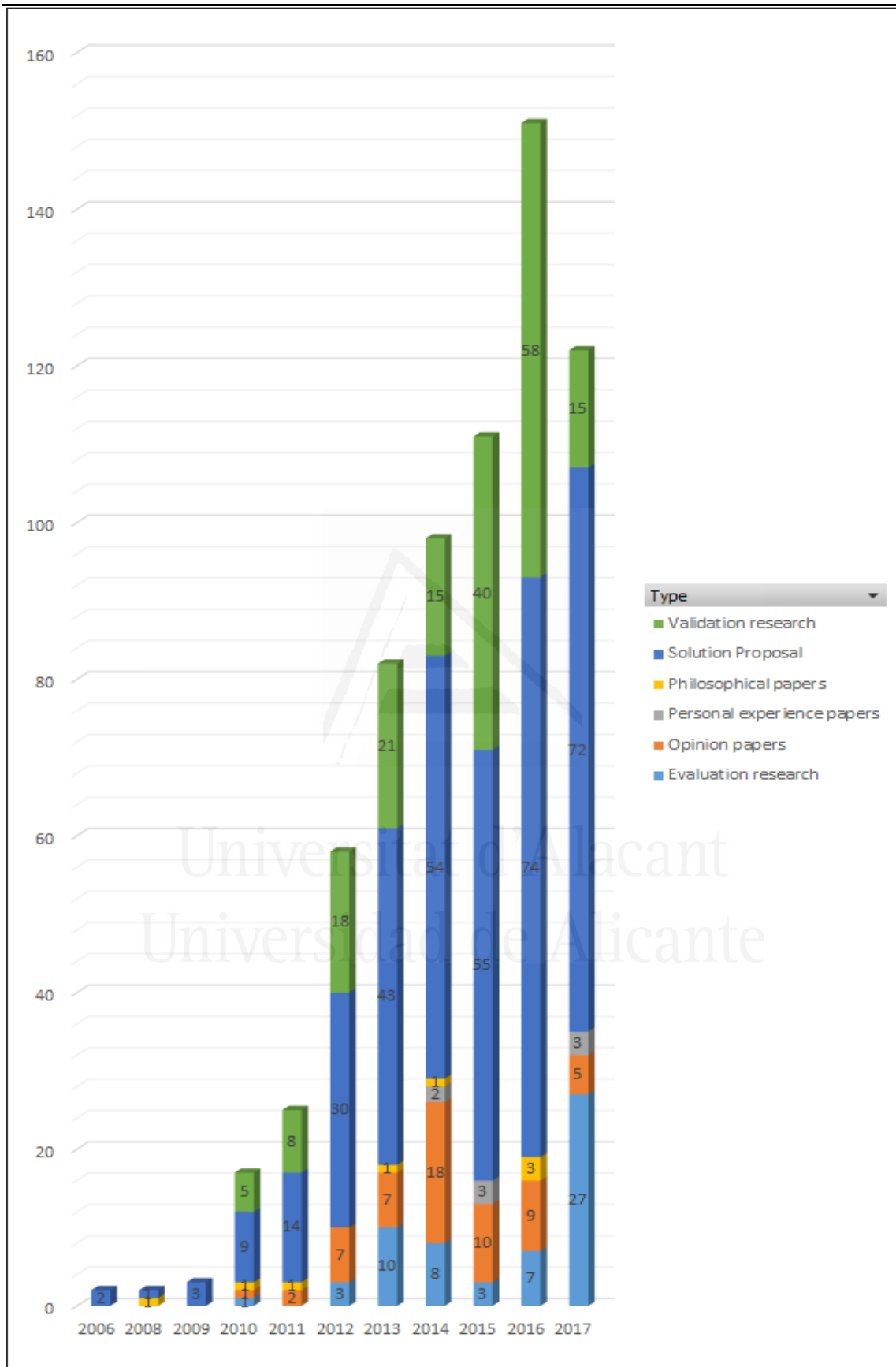


Figura 15: Distribución de las publicaciones por tipo

Además, las publicaciones sobre evaluación de la investigación han crecido en los últimos años, a medida que se ha establecido el campo de los datos abiertos.

3.3.7 Combinación de facetas. Mapa(s) sistemático(s)

Para un análisis completo de los resultados del mapeo sistemático, se realizó una visualización de los datos mediante la combinación de diferentes facetas. Los resultados se presentan en 6 gráficas, cada una de las cuales representa diferentes combinaciones de facetas posibles.

La **¡Error! No se encuentra el origen de la referencia.** es el resultado de combinar la faceta dominio con la faceta ciclo de vida de los datos. La mayoría de las publicaciones de explotación de datos (123) y de exploración de datos (74) estaban relacionadas con el dominio de los infomediarios, también se observa que las publicaciones que incluyen una fase de explotación de datos se distribuyen en casi todos los dominios.

La **¡Error! No se encuentra el origen de la referencia.** combina la faceta de dominio y tópico. El mayor número de publicaciones de la web semántica (77) están relacionadas con el dominio de los infomediarios. También se encuentra que el principal dominio en las publicaciones gubernamentales es infomediarios (79). Clasificados en el dominio infomediarios existen varias publicaciones (26) en calidad de datos y unas pocas (8) en seguridad.

En la Figura 17, la faceta dominio se combina con el tipo de investigación. La mayor cantidad de publicaciones son clasificadas como propuestas de solución (182), y están relacionadas con el dominio de los infomediarios. Las publicaciones clasificadas como validación de la investigación (71) están relacionadas con el ámbito de los infomediarios. Las propuestas de soluciones y las publicaciones sobre validación de la investigación contribuyen en casi todos los ámbitos. Sin embargo, no hay publicaciones de ningún tipo sobre agricultura, química, periodismo de datos, humanitarismo, innovación y energía.

La Figura 19, combina las facetas de ciclo de vida y de tópico. el mayor número de publicaciones clasificadas como ingeniería de software (88) están relacionadas con la fase de explotación de datos, mientras que la mayoría de las publicaciones clasificadas como web semántica (51) están relacionadas también con la fase de explotación de datos. el tópico de la web semántica contó con un total de 159 publicaciones; ingeniería de software con 150, y gobierno con 107, siendo estos tres tópicos los más relevantes. En el ciclo de vida de los datos, los tópicos más relevantes son: explotación de datos (285 publicaciones) y exploración de datos (128).

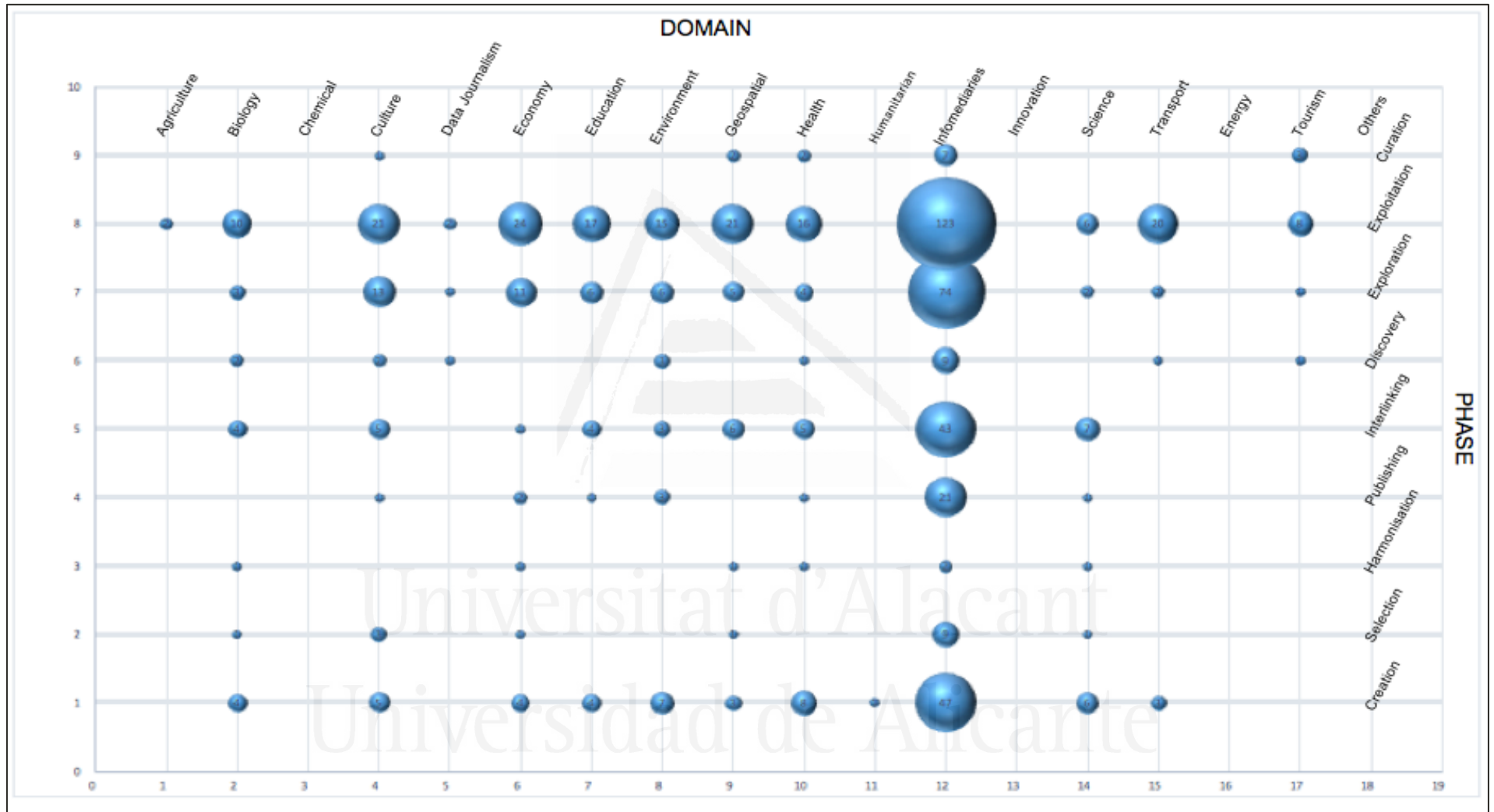


Figura 16: Clasificación de las publicaciones de acuerdo al dominio y ciclo de vida de los datos.

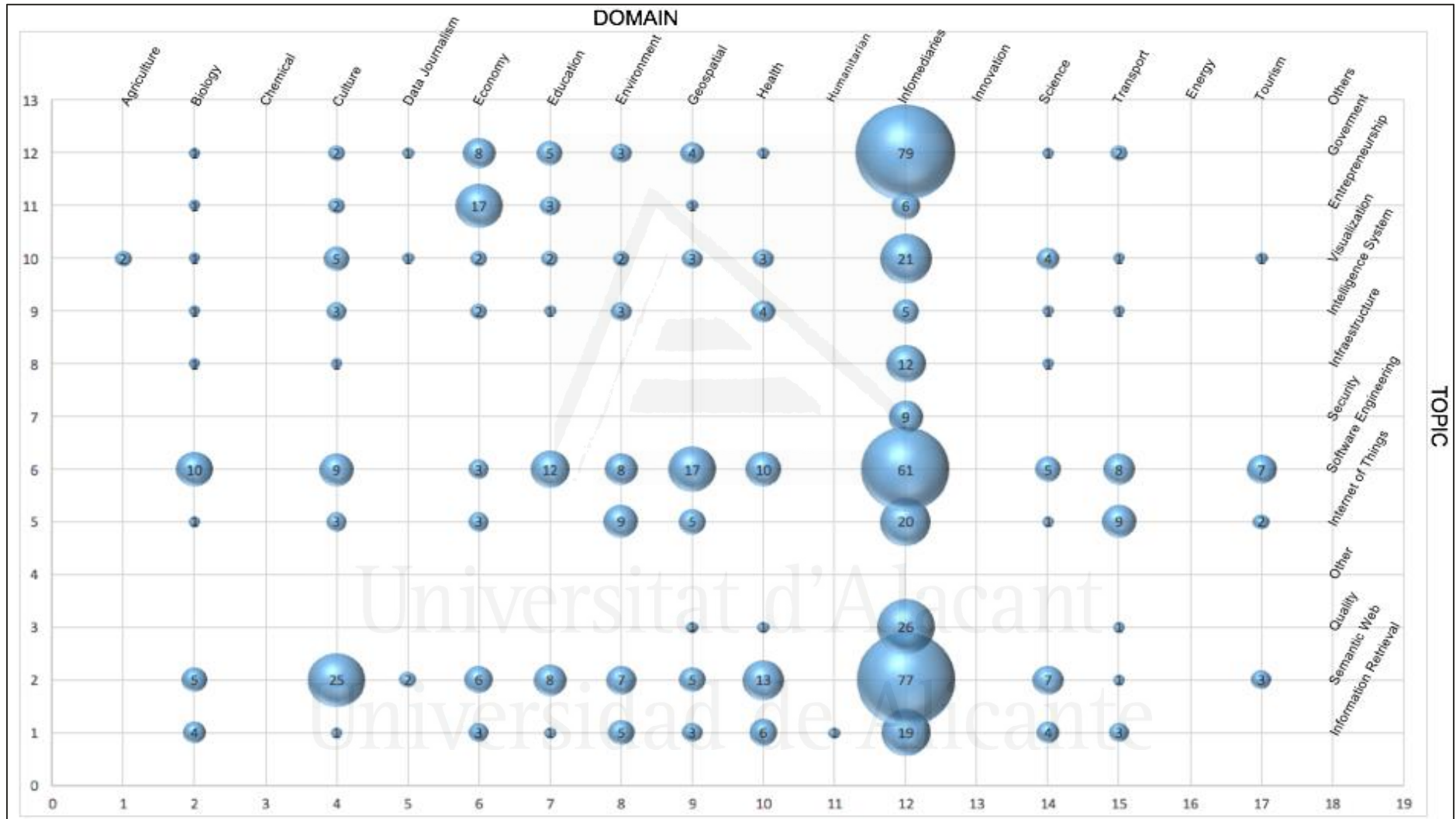


Figura 17: Clasificación de las publicaciones de acuerdo al dominio y tópico

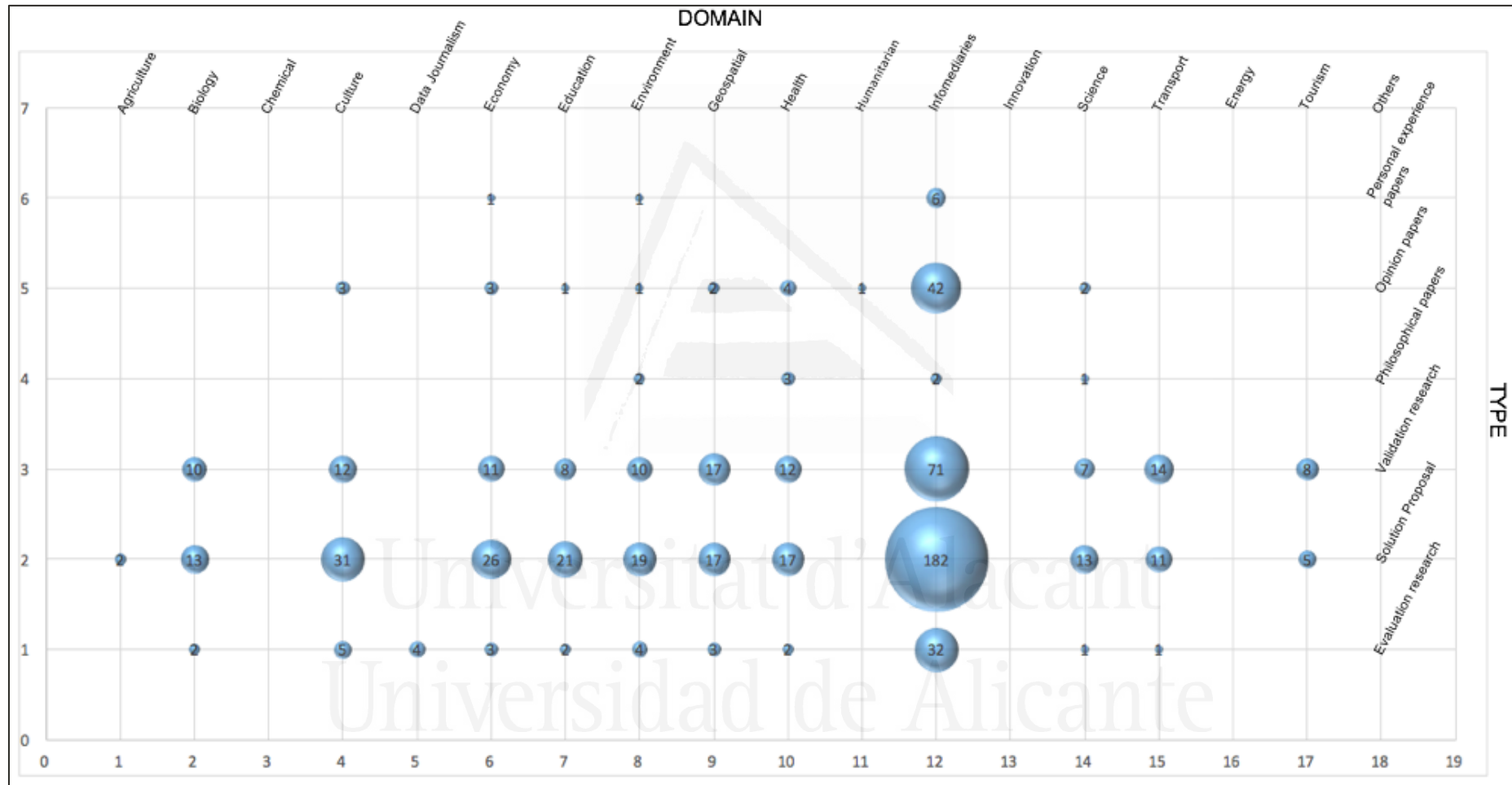


Figura 18: Clasificación de las publicaciones de acuerdo al dominio y tipo

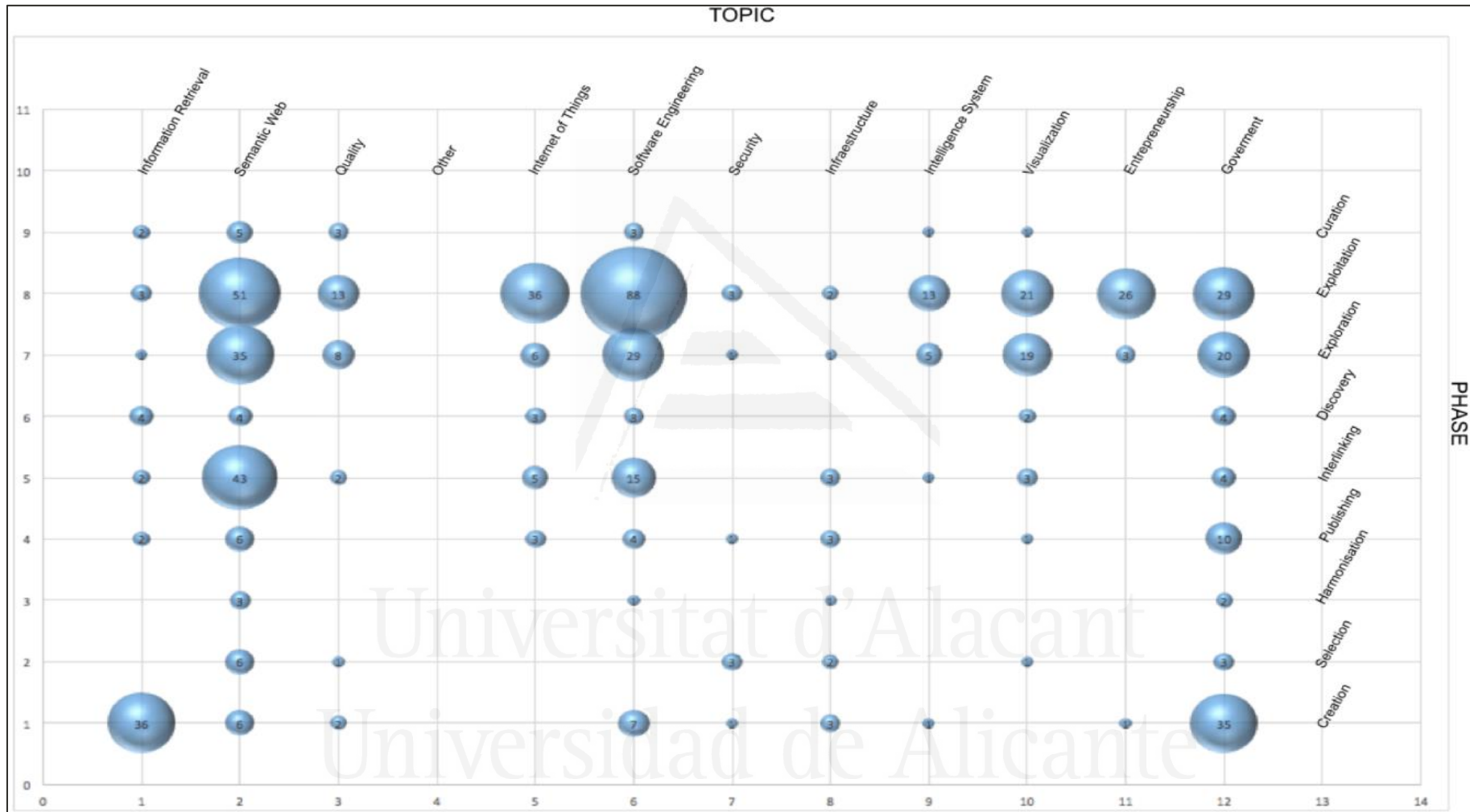


Figura 19: Clasificación de las publicaciones de acuerdo al ciclo de vida del dato y tópico

La Figura 20 combina la faceta t3pico y del tipo de investigaci3n, el tipo de investigaci3n m3s frecuente se refiere a las propuestas de soluciones (con 357 publicaciones contabilizadas), donde 94 art3culos est3n relacionados con el t3pico web sem3ntica, mientras que el t3pico ingenier3a del software se asigna a 92 publicaciones.

La Figura 21 combina la faceta del ciclo de vida de los datos con la del tipo de investigaci3n. un total de 285 publicaciones est3n relacionadas con la fase de explotaci3n, seguidas de 128 publicaciones sobre la fase de exploraci3n. en cuanto al tipo de investigaci3n, las propuestas de soluci3n son predominantes, con 357 publicaciones. en cuanto a esta combinaci3n de facetas, el mayor n3mero de publicaciones pertenec3a a la fase de explotaci3n, clasific3ndose tambi3n como propuesta de soluci3n (149 publicaciones). adem3s, un alto n3mero de publicaciones en la fase de explotaci3n de datos se clasifican, a su vez, como validaci3n de la investigaci3n (94).



Universitat d'Alacant
Universidad de Alicante

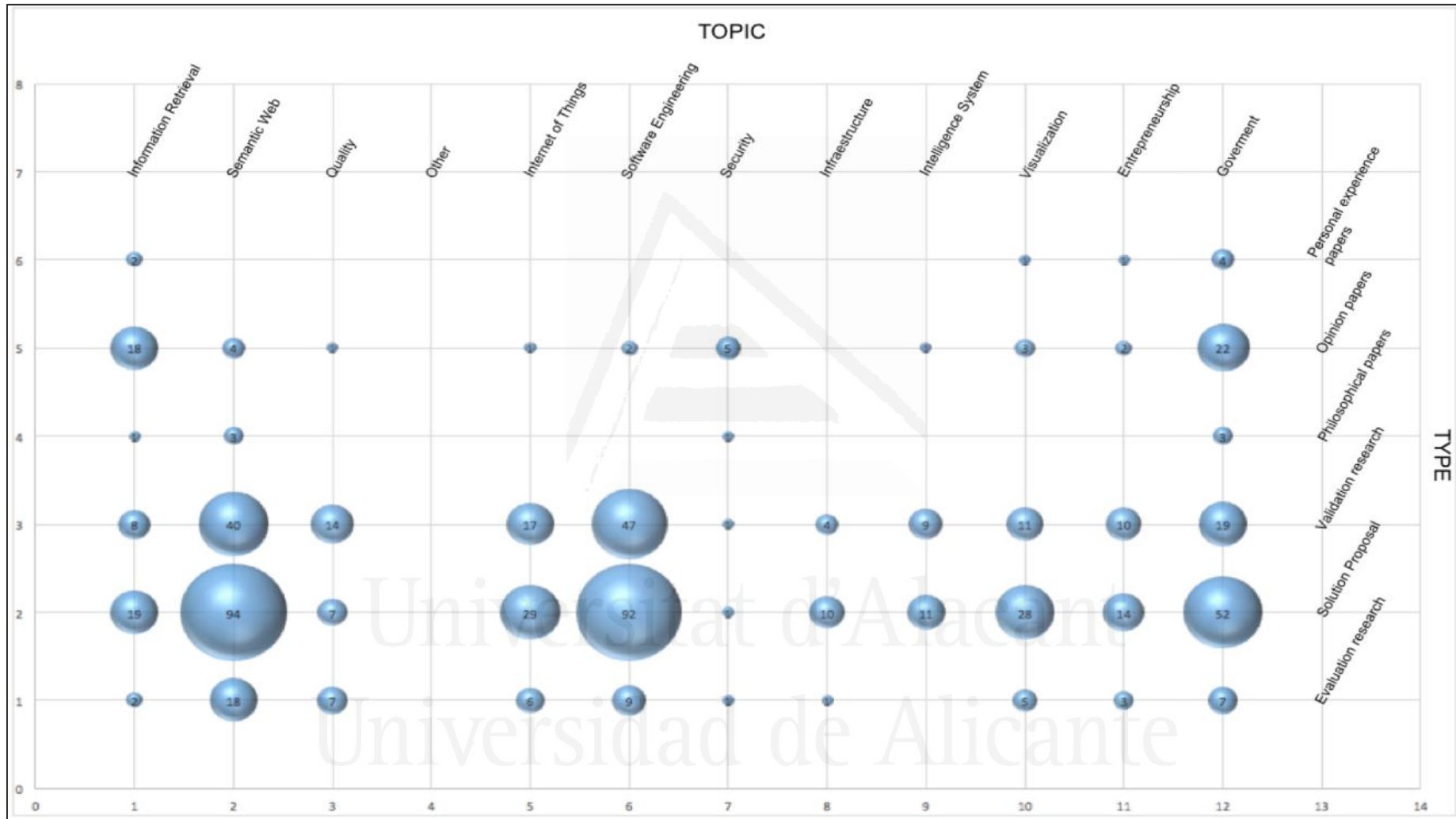


Figura 20: Clasificación de las publicaciones de acuerdo al tópico y tipo

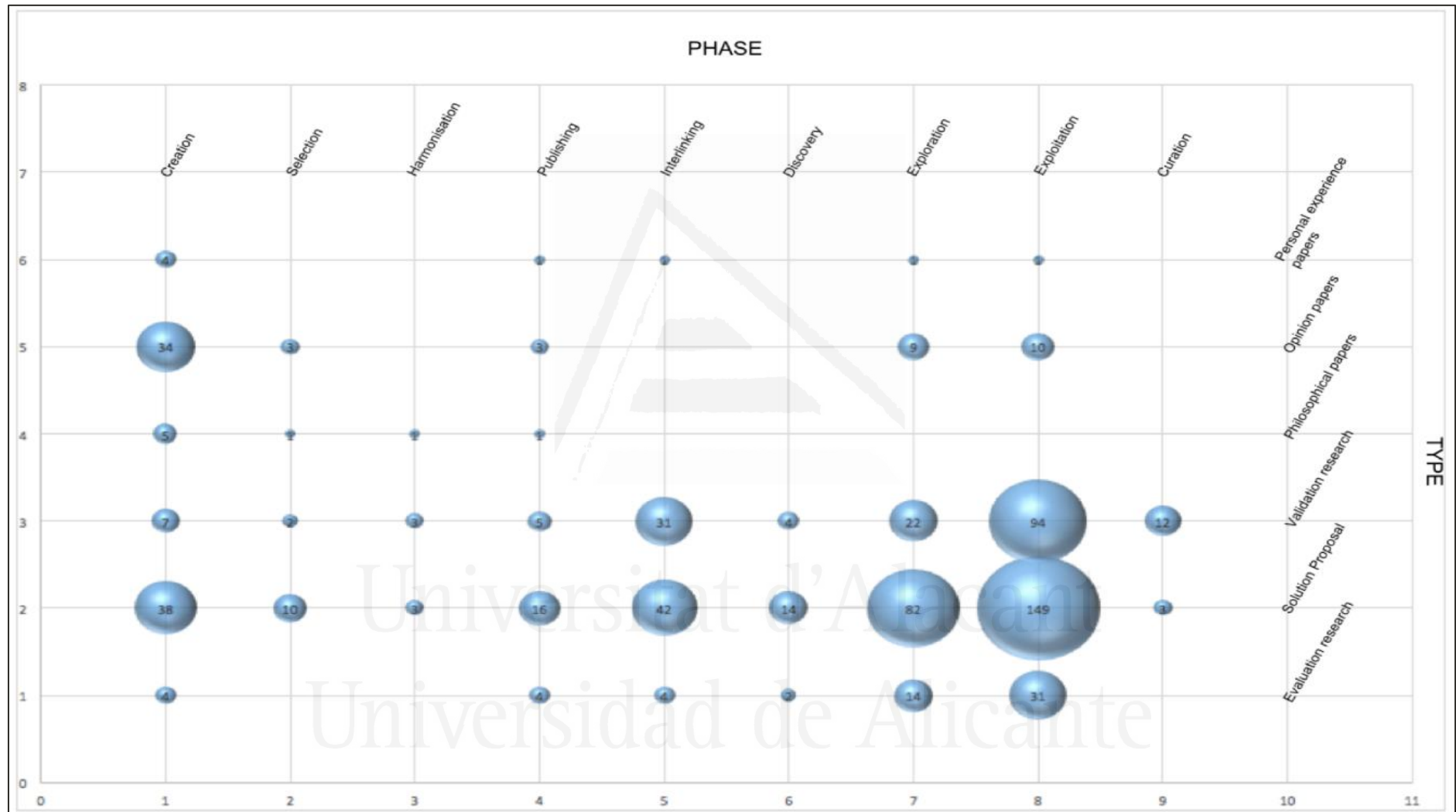


Figura 21: Clasificación de las publicaciones de acuerdo al ciclo de vida del dato y tipo de investigación.

3.4 Discusión de resultados del mapeo sistemático

En la sección anterior, el análisis de los datos describe objetivamente el número de publicaciones por faceta, se basa en la clasificación aplicada y tiene en cuenta la evolución a lo largo del tiempo. En la presente sección se presenta una discusión de los resultados en relación a las preguntas de investigación y se identifican tópicos de investigación abiertos y tendencias.

3.4.1 Lugares en los que se ha publicado investigaciones de datos abiertos

Después de analizar los resultados, se puede observar que dos comunidades importantes están trabajando en la investigación sobre datos abiertos:

- Una dedicada a tópicos relacionados con la web con especial énfasis en la web semántica (así como tópicos relacionados) como la revista “Semantic Web” o la conferencia “Hawaii International Conference on System Sciences”.
- Otra dedicada a la administración electrónica y su relación con los datos abiertos, es decir, la intersección entre la tecnología de la información y las administraciones públicas “Government Information Quarterly” o conferencias como la “DG. O” [16]. Por otra parte, hay otros escenarios interesantes con comunidades emergentes, como la ingeniería de datos, los sistemas de información o la inteligencia artificial. Estas comunidades enriquecerán el campo de los datos abiertos más allá de las tecnologías de la web semántica, como la integración de datos abiertos para la Inteligencia de Negocios [17].

3.4.2 Impacto que han tenido estos estudios

La Figura 7 muestra que el 70% de las publicaciones clasificadas tienen un impacto internacional, seguido por el 21% de las publicaciones con impacto local y el 9% con impacto nacional.

Por lo tanto, aunque los datos abiertos están estrechamente relacionados con instituciones públicas (que tienen regulaciones locales, regionales o nacionales), la mayoría de las publicaciones tienen una aplicabilidad internacional. Por ejemplo, las investigaciones en torno a los trabajos de biblioteconomía realizados en el ZBW-Leibniz Information Centre for Economics y en reuniones internacionales de bibliotecas [41], hacen especial énfasis en temas como Linked Open Data (LOD), la integración de datos y la integración de esquemas,

la gestión de datos distribuidos, entre otros; los que pueden utilizarse no sólo a nivel local de una biblioteca sino también internacional. En [42], se trabaja con datos recopilados de dos servicios relevantes proporcionados por MeteoGalicia -la generación de informes climáticos con descripciones mensuales del comportamiento climático y la generación de predicciones meteorológicas- que se convirtieron en el núcleo de un marco para generar descripciones lingüísticas del tiempo no sólo en Galicia, sino que el marco puede ser utilizado por cualquier servicio meteorológico del mundo. En [43], las tecnologías de la web semántica se utilizaron para comprender la descripción de las entidades asociadas con la creación y el mantenimiento del patrimonio histórico en Bangladesh, pero estos resultados de investigación pueden utilizarse en otras regiones del mundo con otros datos de similar naturaleza.

Para concluir, en base al análisis realizado, las publicaciones sobre datos abiertos a menudo se refieren a una ubicación geográfica específica donde se encuentran las administraciones públicas que son el caso de estudio (o incluso las impulsoras) de la investigación. Sin embargo, estas publicaciones suelen tener un alcance internacional, ya que se convierten en una referencia para otros contextos geográficos más amplios.

3.4.3 Dominio al que pertenecen las investigaciones realizadas

Tras un fuerte aumento en 2012, el número de trabajos de investigación relacionados con el dominio de los infomediarios se mantuvo constante hasta 2015, aumentando de nuevo en 2016 y estabilizándose en 2017. De hecho, la mayoría de las publicaciones se centran en el objetivo de buscar el valor de los datos abiertos, es decir, buscan los usos prácticos que pueden tener los datos abiertos para beneficiar a la sociedad (por ejemplo, propuestas de investigación que resuelven los problemas experimentados por los infomediarios, facilitando la reutilización de los datos y estudios sobre cómo realizar esta reutilización). Por ejemplo, el movimiento de los Recursos Educativos Abiertos (Open Educational Resources, OER) plantea retos inherentes al descubrimiento y reutilización de materiales educativos digitales procedentes de repositorios digitales altamente heterogéneos y distribuidos. En [44], los autores presentaron las especificaciones de una plataforma de datos orientada al consumidor para datos abiertos, el Data-TAP. Esta plataforma proporciona una interfaz fácil de usar y entender para hacer que los datos abiertos sean más amigables para los consumidores [45].

Sobre la base de los resultados de los datos clasificados, aproximadamente el 50% correspondió a infomediarios, el 8% a cultura, el 7% a economía, el 6% a geoespacial y el 6% a salud.

Los trabajos de investigación relacionados con las esferas del transporte, la economía, la educación, el medio ambiente, la cultura, la salud y la geoespacial han aumentado considerablemente en la misma proporción en los últimos siete años, pero entre 2006 y 2009 no hay investigación en dichas áreas. Este crecimiento podría deberse al hecho de que estos ámbitos utilizan datos abiertos para proporcionar soluciones y beneficios a los ciudadanos, es decir, retos sociales que requieren investigación. Los datos de los campos del periodismo y el ámbito humanitario no generan interés en la comunidad investigadora hasta 2017, aunque son campos relevantes:

- Existen muchos datos heterogéneos (no sólo textuales sino también multimedia) en el ámbito humanitario (como los datos de gestión de desastres).
- El periodismo de datos está relacionado con las tecnologías de visualización de datos y la gestión de datos fáciles de usar, lo que merece ser abordado más a fondo por la comunidad investigadora.

La **¡Error! No se encuentra el origen de la referencia.** muestra la relación entre el dominio y la fase de datos del ciclo de vida, con 123 publicaciones que relacionan la explotación de datos abiertos con los infomediarios, revelando así la necesidad de resolver problemas reales de la sociedad.

En el ámbito de la cultura, los avances se derivan de la política de apertura de los datos procedentes de fuentes bibliográficas que se reutilizan en las aplicaciones museísticas. El ámbito de la economía está creciendo, y este aumento se atribuye a la utilización de datos por parte de las empresas. El dominio geoespacial se ha desarrollado debido a su creciente uso en aplicaciones marítimas, de seguridad y de salud. Estos temas corresponden a mejoras en la calidad de vida, especialmente en el ámbito de la salud. El ámbito de la salud se ocupa de las necesidades básicas de la sociedad, por lo tanto, existe una investigación considerable basada en datos abiertos.

En la **¡Error! No se encuentra el origen de la referencia.**, predominan las propuestas de solución (357 de 671) y la validación de la investigación (180 de 671), distribuyéndose en todos los dominios, pero principalmente al dominio de los Infomediarios.

3.4.4 Ciclo de vida del dato en los cuales han trabajado las investigaciones

En la Figura 12 se muestran las contribuciones de los trabajos clasificados por ciclo de vida de los datos. el 43% correspondió a explotación de datos, el 19% a exploración de datos, y en conjunto, estas dos fases del ciclo de vida representaron el 62% de los trabajos publicados.

la Figura 13 muestra que, desde 2010, ambas fases del ciclo de vida del dato crecieron, respondiendo a la necesidad de obtener información a partir de datos publicados en aplicaciones de la vida real. este proceso ha sido liderado por las necesidades de la comunidad de infomediarios con respecto a la reutilización de datos (lo que concuerda con la discusión anterior sobre el dominio de investigación más importante sobre datos abiertos, a saber, los infomediarios). las primeras fases del ciclo de vida de los datos son la publicación, la creación, la interconexión y la armonización; al examinar los resultados, se observa la necesidad de investigación adicional sobre estas primeras etapas del ciclo de vida de los datos.

La Figura 21 muestra las fases del ciclo de vida de los datos según el tipo. La mayoría de los trabajos (149) se refieren a propuestas de solución dentro de la fase de explotación de datos, confirmando que la comunidad científica se está centrando en presentar soluciones para la reutilización de datos.

La Figura 12 muestra que el 14% de los artículos están dedicados a la creación de datos y el 4% a la publicación de datos. esto contrasta con el 43% de los documentos relacionados con la explotación de datos, el 19% con la exploración de datos y el 12% con la interconexión de datos. esto lleva a la conclusión de que la investigación sobre datos abiertos se centró principalmente en la explotación y la exploración (es decir, en temas directamente relacionados con el consumo de datos), requiriendo un mayor esfuerzo de la comunidad investigadora en las fases del ciclo de vida de los datos relacionadas con su publicación. De hecho, hasta la fecha, ante la falta de trabajos de investigación en estrategias de publicación de datos, los datos abiertos no son de suficiente calidad, ya que se publican sin seguir mecanismos formales de aseguramiento de la calidad. Por lo tanto, se requiere investigación adicional sobre la calidad de los datos abiertos [30] y la fase de publicación de datos del ciclo de vida debe ser enfatizada en la investigación de datos abiertos. Cubrir las fases del ciclo de vida de los datos propuestas en [30], es hoy muy relevante, ya que un proceso de preparación previo a la publicación garantiza una mejor reutilización posteriormente.

Además, hoy en día, las fases relacionadas con la publicación de los datos son cruciales debido al rápido proceso de digitalización de los datos, que plantea problemas de normalización de formatos y de conexión a fuentes de datos heterogéneas [46] [47] [48] [49]. Estas contribuciones son importantes para entender que es necesario trabajar todas las etapas propuestas en el ciclo de vida de los datos, con especial énfasis en las fases que preparan los datos para su publicación. De hecho, según varios autores, una de las razones por las que no los datos de una institución pública no se abren es debido a la compleja naturaleza técnica

inherente a su publicación. Un ejemplo de esta problemática se puede encontrar en la publicación de datos de investigación en abierto [50].

Por otra parte, cabe destacar que el mapeo sistemático llevado a cabo muestra que Interconexión de datos representó el 12% de todas las publicaciones revisadas. Esta fase de datos crece constantemente con el paso del tiempo debido al auge de la web semántica. Estas aportaciones de la investigación sobre la interconexión de datos podrían mejorar los procesos de interoperabilidad necesarios para la publicación de datos abiertos. Por último, las fases de armonización, selección y conservación de datos tuvieron una tasa de desarrollo inferior al 3%. La disminución del número total de trabajos en 2017 se debe básicamente a la disminución del número de trabajos relacionados con la conservación y creación de datos. También, el número de publicaciones relacionadas con el Descubrimiento de datos y Publicación de datos crecieron a un ritmo más lento.

Por tanto, se establece la necesidad de mejorar los procesos de publicación considerando los diferentes orígenes de datos, ya que las aportaciones en estas temáticas son pocas.

3.4.5 Tipo de investigaciones que se realizan

La Figura 14 muestra que el 53% de los trabajos de investigación consisten en propuestas de soluciones, seguido por trabajos de validación de la investigación con un 27%, Documentos de Opinión con un 9% e evaluación de investigaciones con un 9%. Los tipos de investigación menos frecuentes son Experiencia personal y Artículos filosóficos, ambos con un 1%, ya que los esfuerzos se dirigen hacia soluciones prácticas destinadas a ser validadas y evaluadas.

La Figura 20 relaciona el tópico con el tipo de investigación. Una vez analizados los datos, un gran número de publicaciones se clasifican como: Propuesta de Solución o validación de la investigación. Muchas propuestas de soluciones se orientaron hacia la web semántica (94), respondiendo a la necesidad de relacionar datos abiertos y crear aplicaciones más eficientes. Las soluciones propuestas tienen su origen en las necesidades de los infomediarios (182). Los infomediarios proporcionan soluciones basadas en datos abiertos, pero, según la el mapeo sistemático realizado, cada vez se necesitan más datos de calidad, por lo que se tienen que aportar propuestas de solución para conseguir datos abiertos de mejor calidad. De hecho, según [51], la calidad de los datos abiertos es una de las principales amenazas para alcanzar los objetivos del movimiento de datos abiertos.

De cada uno de estos tipos de investigación se pueden encontrar ejemplos en el mapeo sistemático desarrollado. Un ejemplo de evaluación de datos abiertos es la propuesta de

análisis de informes de ensayos clínicos, donde la evaluación de la investigación juega un papel importante antes de que la solución sea implementada en la sociedad [52]. Otro ejemplo de una propuesta de solución usando tecnologías de la web semántica trata de un marco unificado para convertir datos heredados abiertos en imágenes gráficas y en un formato legible por máquinas y reutilizable mediante crowdsourcing [53].

En [53] se propone una solución de ingeniería de software que responde a las necesidades de datos abiertos de las personas mayores aplicando tamaños de letra más grandes y comprensibles. Por otra parte, otros autores han abordado las necesidades de información de las personas con discapacidad motriz en relación con las distancias de marcha y la existencia de rampas, peldaños, asientos entre otros [54].

Estos ejemplos permiten entender que las soluciones propuestas son importantes y están estrechamente relacionadas con el desarrollo del concepto de datos abiertos. Sin embargo, el rápido aumento asociado de los volúmenes de datos digitales ha hecho que se propongan nuevas soluciones en el ámbito de los datos abiertos pero ello no ha ido correlacionado automáticamente con la validación o evaluación de los nuevos conocimientos y avances [55]. No obstante, los datos recabados en el mapeo sistemático son optimistas: cuando se iniciaron las investigaciones sobre datos abiertos es obvio que se necesitaba la creación de propuestas para solucionar diversos requisitos en torno a los datos abiertos. Una vez desarrolladas estas soluciones, surge la necesidad de realizar evaluaciones y validación de las propuestas, por lo que la cantidad de documentos de evaluación y validación, aun estando lejos de las propuestas de solución, aumentó a partir de 2010. Esto demuestra que la comunidad investigadora en datos abiertos ha alcanzado un cierto grado de madurez: la comunidad está cada vez más centrada no sólo en el desarrollo de soluciones, sino también en la evaluación y validación de las mismas.

3.4.6 Tópicos que se trataron en la investigación

La Figura 10 muestra que la web semántica representa el 24% de la investigación en datos abiertos según el mapeo sistemático realizado. El segundo tópico más popular es la ingeniería de software con un 22%, seguido por el gobierno con un 16%. Significativamente, el tópico de la recuperación de información representa el 8% y el internet de las cosas el 8% de las publicaciones.

La **¡Error! No se encuentra el origen de la referencia.** relaciona el dominio y el tópico. La web semántica (77), el gobierno (79) y la ingeniería de software (61) representan

conjuntamente el 32 % del total de publicaciones. El notable desarrollo de las publicaciones de web semántica y desarrollo de software se basa en la tendencia a explotar y explorar datos, en los diferentes dominios (discutidos en la pregunta de investigación relacionada con el dominio). El gran número de publicaciones relacionadas con el gobierno revela la necesidad de un marco jurídico con políticas y reglamentos para la gestión de datos abiertos, pero desde una perspectiva tecnológica. Por lo tanto, la investigación multidisciplinar sobre cuestiones jurídicas y tecnológicas de datos abiertos es muy necesaria para establecer políticas que faciliten el desarrollo de datos abiertos y generen servicios tecnológicos que utilicen datos abiertos.

La razón principal por la que el tópico de la Web Semántica puede haber atraído la mayor parte de la investigación es la necesidad de esquemas y vocabularios potentes para el intercambio de datos. Este proceso permite generar automáticamente información para desarrollar aplicaciones y facilitar plataformas, por ejemplo, en las áreas de salud [56], o la visualización de tópicos ambientales, mapeo de gestión de servicios públicos, evaluación de cabildeo político, beneficios sociales, cierre de la brecha digital, biología y otros [57].

El desarrollo de la investigación en ingeniería del software aplicada a diversas áreas es notable en aplicaciones de captura de datos para la enseñanza o desarrollo de APIs para aplicaciones turísticas [34], análisis de datos generados por smartphones o recolección de datos abiertos para presentar información útil a los ciudadanos para eventos públicos [58] [59] [55] [60][61][62].

La ingeniería de software representa 148 artículos y es el segundo tópico más frecuente. El mayor número de contribuciones se produce en 2015, con un 21% del total. Sin embargo, este resultado contrasta con el desarrollo de proyectos empresariales. Las aplicaciones informáticas se orientan hacia los servicios de reutilización, no hacia el espíritu empresarial. Se debe recordar que el espíritu empresarial no depende necesariamente de la investigación. Esto explica por qué anteriormente hay pocos estudios científicos de este tipo.

Por último, la investigación y el desarrollo en materia de seguridad han sido limitados, a pesar de que el tópico es relevante desde el punto de vista de los datos abiertos. El tópico de la internet de las cosas (IoT) crece constantemente hasta 2017, en respuesta al desarrollo de la IoT en el mundo de las tecnologías de información (TI). A pesar de que siguen incrementándose las aplicaciones en ingeniería de software existen pocas investigaciones que trabajan sobre el tópico de calidad de datos (4%).

3.5 Más allá del mapeo sistemático: proyectos de innovación de datos abiertos

Además de un análisis cuantitativo de la investigación sobre datos abiertos de los últimos años, presentada en las secciones anteriores, ahora se va más allá del estudio de mapeo sistemático y se presenta una revisión de los proyectos de innovación de datos abiertos. Este estudio se centra en dos fases del ciclo de vida de los datos: la publicación (por ejemplo, portales de datos abiertos) y el consumo (por ejemplo, datos abiertos como facilitador del negocio) de datos abiertos.

3.5.1 Portales de datos abiertos y búsqueda de conjuntos de datos

El primer paso crítico del ciclo de vida de los datos cuando se consideran los agentes externos que pueden consumir los datos es su publicación. Normalmente, esto se hace mediante la creación de un portal de datos abiertos, donde la división administrativa en particular, o la organización pública que produce o recopila los datos los hace accesibles al público. En [20] se propone una clasificación de los portales de datos abiertos de acuerdo con el número de funcionalidades que proporcionan, ordenadas por la cantidad de esfuerzo que se requiere para su creación. Esta clasificación se distribuye en niveles, según se enumera a continuación:

1. Un "registro de conjuntos de datos", es decir una simple lista de enlaces hacia conjuntos de datos, no necesariamente alojados en el portal. Un registro responde a la pregunta "¿Quién tiene qué conjunto de datos (abierto) y dónde puedo encontrarlo?"
2. Un "proveedor de metadatos", es decir un registro de conjuntos de datos que también contiene metadatos sobre conjuntos de datos, por ejemplo, licencias, contexto espacio-temporal, frecuencia de actualización.
3. Una "plataforma de co-creación", es decir un proveedor de metadatos que incluye herramientas de participación de los ciudadanos y los consumidores de datos para actuar sobre los conjuntos de datos: generar ideas, plantear cuestiones, contribuir con ejemplos de reutilización, y/o participar en debates.
4. Una "plataforma de publicación de datos", es decir una plataforma de co-creación que permite a múltiples proveedores de datos publicar sus propios conjuntos de datos. También soporta la fase de interconexión entre los conjuntos de datos alojados.

5. Un "centro de datos común", es decir una plataforma de publicación de datos que también soporta la implementación de otras fases del ciclo de vida de los datos, permitiendo a los editores de datos implementar sus propios ciclos de vida del dato.

Los sistemas de software de código abierto y comerciales como CKAN, Socrata y OpenDataSoft permiten a los proveedores de datos de código abierto configurar fácilmente un portal entre los niveles 3 y 4. Esto permite proporcionar una visión unificada que facilite la búsqueda a nivel nacional y supranacional desde los portales regionales y locales, ya que los denominados metaportales rastrean los conjuntos de datos y los indexan hasta una ubicación central en la que también se pueden realizar enlaces adicionales. Un ejemplo de formalización de tal modelo es el marco MODA (Middleware for Open Data), definido en [63]. Muchos portales de datos abiertos nacionales proceden de esta manera, agregando datos de portales regionales y de agencias gubernamentales. Existen varios ejemplos de proyectos en este sentido. Por ejemplo, el proyecto Europeana [64], proporciona un único punto de acceso a millones de libros, pinturas, películas, objetos de museo y archivos que han sido digitalizados en toda Europa. Otro ejemplo es el Portal Europeo de Datos [65], que recoge los metadatos de la información del sector público disponibles en los portales de datos públicos de todos los países europeos. También se incluye información sobre el suministro de datos y los beneficios de su reutilización.

Una vez publicados los conjuntos de datos, el siguiente paso es proporcionar una funcionalidad de búsqueda adecuada. La búsqueda de conjuntos de datos es un campo de investigación relativamente nuevo que se encuentra en la intersección de la recuperación de información, las bases de datos, la web semántica y la gestión de datos empresariales. El primer esfuerzo fue la búsqueda internacional de conjuntos de datos de gobierno abierto propuesto en [66], una interfaz de navegación facetada para buscar en más de un millón de conjuntos de datos de gobierno abierto de todo el mundo. Los desafíos para la búsqueda de conjuntos de datos se esbozaron en la propuesta "Challenges for Dataset Search" [67] en el contexto de los datos científicos, proponiendo un proceso Crawl-Read-Extract similar a los motores de búsqueda de documentos. Recientemente, se han realizado esfuerzos para comprender las sutilezas de la búsqueda de conjuntos de datos desde la perspectiva del usuario, ya sea entrevistando a profesionales [68] o analizando el análisis de los registros de consultas y las solicitudes de datos [69]. La investigación actual se centra en el uso de estos conocimientos para desarrollar modelos de aprendizaje automático para clasificar conjuntos de datos en un portal según una consulta [70], [71].

3.5.2 La innovación de datos abiertos como facilitador del negocio

Varios trabajos han identificado el potencial de los datos abiertos como catalizadores de la innovación, permitiendo así la creación de valor y de servicios que, en última instancia, benefician a los ciudadanos. En el contexto de Open Government Data (OGD), en [72] se identifican 4 mecanismos para la generación de valor socioeconómico mixto: transparencia (que mejora la visibilidad), participación (compromiso con todos los grupos de interés), eficiencia (reducción de costes y tiempo) e innovación (generación de nuevas ideas), destacando la apertura de datos como facilitador tanto de la generación como de la apropiación de valor. En [71] se encuesta a 138 empresarios suecos de TI y se encuentra que el acceso a datos públicos abiertos se considera muy importante para muchos de ellos; el 43% encuentra que los datos abiertos son esenciales para la realización de su plan de negocios y el 82% afirma que el acceso apoyaría y fortalecería el plan de negocios.

El empresariado también muestra interés y voluntad de pagar por los datos de información del sector público para apoyar o probar otros modelos de negocio. En [73] se analiza datos de 500 empresas con sede en EE.UU. que utilizan datos de gobierno abierto en su modelo de negocio y propone una taxonomía de arquetipos de modelos de negocio: facilitadores de la recopilación, gestión y divulgación de datos públicos; facilitadores que apoyan o aceleran el acceso y el intercambio de datos entre el lado de la oferta y el de la demanda; e integradores que hacen uso de los datos de gobierno abierto combinándolos con datos internos u otros tipos de datos de propiedad exclusiva a fin de aumentar sus capacidades empresariales. Por otro lado, en [74] se identifica como un bloqueador percibido la falta de ventaja competitiva, ya que los datos abiertos son accesibles a todos (incluyendo a los competidores), su estudio sugiere que la generación de una ventaja competitiva con datos abiertos requiere que una compañía tenga capacidades y recursos internos para el uso de datos abiertos.

Por lo que se refiere a las actividades de innovación, dos recientes acciones de innovación financiadas por la UE se centran en cómo liberar el potencial de los datos abiertos como herramienta de apoyo a las empresas, proporcionando a las PYME las capacidades necesarias para procesar datos abiertos, además del apoyo general a las empresas, de acuerdo con las recomendaciones de [19]. A continuación, se describe brevemente su funcionamiento y sus resultados. En primer lugar, FINODEX estuvo en marcha desde mediados de 2014 hasta 2016. Su objetivo era la promoción y el apoyo de servicios innovadores de TIC que reutilizan datos abiertos, utilizando como anclaje técnico la plataforma FI-WARE. FINODEX se organizó en dos convocatorias abiertas para que las PYME presenten sus ideas sobre

productos y servicios. Se proporciona financiación, apoyo, formación a medida, oportunidades de creación de redes y conexión con los inversores a determinadas PYME. El tiempo y la financiación de las PYMES dentro de la incubadora son dictados utilizando un enfoque de "embudo". Inicialmente, todas las PYMES reciben una cierta cantidad de financiación para realizar un hito inicial, que se evalúa para decidir qué PYMES pasan a la siguiente fase, en la que reciben una cantidad adicional de financiación en función de un nuevo hito, y así sucesivamente, se ha alcanzado hasta 4 hitos. Los importes más elevados de financiación directa: 170.000, 135.000 y 115.000 euros, se asigna al primer, segundo y tercer proyecto en ambas rondas de aceleración.

Por otro lado, la acción ODINE funcionó desde mediados de 2015 hasta mediados de 2017, con el objetivo de incubar ideas empresariales centradas en datos abiertos. En comparación con FINODEX, no utiliza el método de incubación en embudo, sino un período de incubación de 6 meses para todas las empresas seleccionadas, que es seleccionado en una convocatoria abierta de 8 iteraciones y no obliga a las PYME a utilizar la pila FIWARE, y con una financiación máxima de 100.000 euros por PYME. ODINE incubó 57 empresas que creó 278 nuevos puestos de trabajo y 23,7 millones de euros en ventas e inversiones. El impacto de ODINE en las perspectivas de crecimiento de las empresas financiadas es relevante, resultando en unos 110 millones de euros de ingresos acumulados en el periodo 2016-2020, más 784 puestos de trabajo creados. Un estudio independiente de evaluación de impacto sobre ODINE preparado por IDC [75] descubre que la mayoría de las empresas financiadas son jóvenes y desempeñaban el papel de "experimentadores", es decir, combinaban varias fuentes de datos abiertas para mejorar sus productos y servicios. En la evaluación también se observa una correlación positiva entre el nivel o la madurez a nivel de país del mercado de datos abiertos y el número de solicitantes exitosos de ODINE por país, lo que sugiere que un rico entorno de datos abiertos ofrece condiciones favorables para los innovadores en este campo.

3.5.3 Calidad de datos abiertos

De acuerdo con los resultados del mapeo sistemático se encontraron 29 artículos de investigación que fueron clasificados con el tópico de calidad, de los cuales se realizó un análisis de los resultados de investigación, para determinar los aspectos de calidad que afectan el proceso de publicación de datos.

A continuación, se presenta el resultado del análisis de los 13 documentos que incluyen estudios de varios portales describiendo resultados de investigación en torno a la calidad de datos en portales abiertos.

En [76], se definen métricas que demostraron que los beneficios de la divulgación de datos centralizada en contraposición con la publicación de datos descentralizados. Además, los problemas más comunes identificados por las métricas aplicadas fueron: falta de metadatos, datos incompletos, falta de información sobre actualizaciones de datos que afectan directamente a la comprensibilidad de los datos.

Por otro lado en [77], se concluye que el hecho de que los metadatos tengan una baja calidad (incompletos o poco comprensibles) tiene un impacto negativo drástico en los resultados del análisis o reutilización de los datos abiertos que también se convierten en una problemática para soluciones de Big Data.

En la propuesta presentada en [78], se realizó un estudio de calidad de datos de Dbpedia y se determina diferentes tipos de problemas de calidad. En este artículo se propone el crowdsourcing como alternativa de evaluación de calidad integrado con marcos de gobernanza de datos existentes.

En la investigación llevada a cabo en [79], se implementó la plataforma "Open Data PortalWatch" que monitorea los metadatos de más de 261 portales de datos abiertos, con un conjunto de métricas propuestas por este estudio. Se concluye que varias propiedades del estándar de catalogación DCAT todavía no pueden asignarse automáticamente a los metadatos existentes. Por ejemplo, las propiedades relacionadas con información espacial o temporal (representada en DCAT por el dct:temporal y dcat:spatial_properties), que sería particularmente importante para la capacidad de búsqueda de conjuntos de datos relevantes para un contexto local o temporal específico, no pueden ser mapeados a los metadatos existentes predeterminados. El problema, de hecho, reside en que existe mucha heterogeneidad, entre estos metadatos adicionales como los relacionados con el contexto espacial y temporal. Además, es necesario contar con los URLs de recursos sin errores ni restricciones, especificaciones de licencia e información sobre la procedencia. De acuerdo a este trabajo, no contar con especificaciones de licencia y procedencia, implican un riesgo de transparencia y perjudica el uso de los conjuntos de datos haciendo prohibitivo la retroalimentación valiosa de los usuarios.

En el estudio [80], realizado a los metadatos de 4.4 millones de recursos, que son vinculados a través de los datos de varias fuentes sigue siendo un desafío y actualmente está

obstaculizando severamente la visión de creación de tecnologías para publicar datos legibles por aplicaciones informáticas en funcionamiento. En 261 portales se concluye que para los portales Socrata y OpenDataSoft, es necesario implementar una propuesta para analizar automáticamente la actualización de los conjuntos de datos publicados [81]. Este resultado es coincidente con la conclusión obtenida en [82] acerca de la integración de bases de datos o coincidencia de esquemas.

En la investigación [80], se presenta un enfoque para la conciliación de metadatos entre los portales de datos abiertos. Se analizó el uso de etiquetas y se encontraron varios problemas, como una baja tasa de reutilización de etiquetas. En el análisis también se encontró que varios portales comparten las mismas etiquetas, lo que demuestra que las etiquetas tienen un buen potencial para vincular elementos entre conjuntos de datos. La conversión de etiquetas en identificadores semánticos también se mostró como una opción viable, aunque se deben investigar métodos más sofisticados [12].

Mejorar la calidad de grandes conjuntos de datos, al involucrar a los usuarios, es un desafío pero esencial para mantener fuentes de datos geoespaciales actualizadas según las conclusiones aportadas en [83].

El estudio [84], concluye que se debe desarrollar un marco de calidad de datos que aborde la clasificación de datos en tipos de datos, la definición de calidad de datos por tipo de datos, y el desarrollo de conceptos y métodos para comprobar y mejorar estática y / o dinámicamente la calidad de los datos. Estas características solo se podrán cumplir si se integran enfoques sobre gestión de datos, gobierno de datos, mantenimiento y limpieza de los mismos [85].

En el estudio [86], indica la importancia de mejorar los metadatos, para ayudar a los usuarios a navegar rápidamente a través de un sitio web desconocido e inspeccionar de manera efectiva su contenido para evaluar su utilidad [86]. Este resultado coincide con la investigación [86], que aplica un marco de referencia para comparar más de 250 portales de datos abiertos, impulsados por organizaciones en 43 países diferentes. Los resultados del estudio revelan que las organizaciones de hoy no prestan suficiente atención a la gestión de conjuntos de datos, recursos y metadatos asociados que están publicando actualmente en su portal [87].

En la investigación [86], se analizó un corpus de recursos de datos abiertos tabulares de 200k de 232 portales desde el punto de vista de los consumidores. El análisis destaca algunas características muy específicas para los datos abiertos tabulares y los desafíos para los consumidores de datos. Los principales hallazgos del estudio se pueden resumir de la

siguiente manera: 200k (10%) de los recursos en 232 portales de datos abiertos están etiquetados como CSV, de los cuales solo se pueden analizar 100k archivos, solo el 50% de los archivos CSV reales especifican el formato correcto en el encabezado de respuesta HTTP, solo el 10% de los valores de encabezado en las tablas de encabezado único podrían asignarse a entradas en el diccionario de WordNet en inglés, la mayoría de los archivos CSV usan el delimitador de coma (,) es correcto, un archivo CSV de datos abiertos promedio contiene 365 filas y 14 columnas. Sin embargo, el 10% de las tablas tienen solo una fila, lo que posiblemente indica que estas son tablas ficticias / de prueba, el 50% de las columnas en nuestras tablas contienen valores numéricos o ID (valores de la misma longitud, baja selectividad) [88].

El estudio [17], indica que la calidad es importante desde el origen del dato en las organización, porque los datos publicados a menudo no son familiares para el usuario y pueden carecer de metadatos [89].

Por otro lado, el estudio [90], reveló que hay muchas fuentes de datos con información valiosa que desafortunadamente no se pueden usar debido a los estrictos términos de uso o políticas que prohíben el uso de cualquier técnica para la recuperación automática de datos publicados en una fuente determinada (casi el 28% de todas las fuentes evaluadas), o que requieren autorización previa por escrito para usar sus datos [90].

En [91], se concluye que una calidad adecuada en los conjuntos de datos publicados favorece la utilización de los datos por parte de los usuarios, pero se debe incluir orientaciones de la forma de utilizar estos datos. Además dentro de los riesgos descritos en [92], se indica que en los últimos años, cada vez hay más datos abiertos disponibles y utilizados como parte del movimiento de datos abiertos, pero hay problemas con la calidad de los metadatos en los portales de datos y los datos en sí. Este es un riesgo grave que podría frenar los proyectos de datos abiertos, así como las iniciativas de gobierno electrónico, ya que la calidad de los datos debe gestionarse para garantizar la fiabilidad del gobierno electrónico al público.

Relacionado con este mismo riesgo el estudio [93], concluye que la calidad que un proceso de minería de datos depende en gran medida de la calidad de los datos que procesa [93].

Los datos abiertos tienen como objetivo desbloquear el potencial de innovación de las empresas, los gobiernos y los empresarios, por tanto existe incertidumbre sobre la calidad de los datos de dichos conjuntos de datos, y esta realidad se convierte en una amenaza para el valor que se puede generar a partir de los datos publicados [94].

La calidad deficiente y desconocida ha sido ampliamente reconocida como uno de los principales obstáculos para la utilización de datos abiertos, en esto concluyen varias investigaciones como [95], [96], [97].

La importancia que representa la calidad de los datos publicados, se evidencia en el estudio realizado en [98], en el cual a través de datos abiertos genera elementos que permiten comparar proveedores de servicios hospitalarios. En este caso la calidad de los conjuntos de datos de hospitales publicado es valioso siempre y cuando tenga las condiciones necesarias para el reutilizador.

3.6 Aspectos relevantes del estudio sistemático

El mapeo sistemático tiene como misión reunir, clasificar y analizar toda la investigación sobre datos abiertos desde una perspectiva tecnológica realizada entre 2006 y 2017 (ambas inclusive) por la comunidad científica con el objetivo de: (i) proporcionar una visión general consolidada del campo de investigación, y (ii) identificar, entre otros, temas bien establecidos, tendencias y cuestiones de investigación abiertas. Este mapeo sistemático reveló varios hechos interesantes:

- La mayor parte de la investigación sobre datos abiertos desde una perspectiva técnica proviene de los repositorios científicos IEEE y ACM.
- Los trabajos sobre datos abiertos publicados antes de 2009 no son significativos. Por cierto, fue en 2009 cuando Estados Unidos establece una estrategia de Gobierno Abierto bajo la administración de Barack Obama [99].
- Las publicaciones de 2006 a 2009 son incipientes (de hecho, no hay publicaciones sobre datos abiertos en 2007).
- La web semántica, ingeniería de software y gobierno son algunos de los tópicos más importantes abordados en la investigación. Esto es lógico, porque (i) las tecnologías de web semántica están intrínsecamente relacionadas con la reutilización de datos abiertos, (ii) desarrollo de software requiere nuevos enfoques para resolver problemas técnicos relacionados con la apertura de datos, y (iii) se necesita legislación y estandarización para introducir datos abiertos en la sociedad.
- Los tópicos de Internet de las cosas aún no se han desarrollado, pero dadas las evoluciones tecnológicas actuales, estos tópicos podrán adquirir una gran importancia en un futuro próximo con la implantación de ciudades inteligentes y la

importancia de una calidad de datos suficiente para apoyar la toma de decisiones óptima.

- Infomediarios es el dominio más desarrollado en las publicaciones. Los otros dominios representaban menos del 8% cada uno; esto sugiere que las publicaciones estaban dirigidas especialmente a los canales de información; sin embargo, los dominios geoespaciales, de salud y cultural resultaron ser un objeto de investigación en curso y estos estudios continúan creciendo.
- Al analizar las fases, explotación y exploración son las más frecuentes, lo que justifica la necesidad de la comunidad de aplicaciones prácticas de los datos abiertos. Sorprendentemente, las fases relacionadas con la publicación de datos están poco desarrolladas. Estas fases no han recibido suficiente atención y deberían investigarse más en el futuro.
- De la misma manera, los tipos de investigación de propuesta de solución y validación son los más frecuentes, mostrando que el campo está alcanzando la madurez.
- En cuanto al impacto, las publicaciones internacionales son las más frecuentes porque el desarrollo de software y las publicaciones de web semántica son aplicables y aceptadas internacionalmente.
- Además del análisis cuantitativo proporcionado por el estudio de mapeo sistemático, se lleva a cabo una encuesta sobre las perspectivas de la investigación en relación con los proyectos de innovación de datos abiertos, con especial énfasis en la publicación (por ejemplo, portales de datos abiertos) y el consumo de datos abiertos (por ejemplo, datos abiertos como facilitador del negocio).
- En 2017, el interés de la investigación por los datos abiertos desde una perspectiva tecnológica en general disminuye. Este hecho puede indicar que la investigación se está estabilizando, que el punto más alto de la investigación de datos abiertos ha llegado a su fin y que este campo de investigación está alcanzando la madurez. Aunque también debe tenerse en cuenta otros factores para evaluar el nivel de madurez, por ejemplo, el tipo de publicación, puede considerarse que la cantidad de investigación realizada se está consolidando, ya que el número de soluciones propuestas disminuyó en 2017, mientras que al mismo tiempo se están llevando a cabo más estudios de validación y evaluación.

- Existe solamente 28 artículos publicados que están relacionados con tópicos de calidad de datos abiertos considerando que los datos se publican y explotan sin pasar por ciclos de vida que puedan añadir calidad. Por lo tanto, se requiere investigación adicional sobre la calidad de los datos abiertos. Considerando además que según [30] un proceso de preparación previo a la publicación garantiza una mejor reutilización. Sin embargo, también cabe señalar que la investigación sobre datos abiertos se centra principalmente en la explotación y la exploración, es decir, en tópicos directamente relacionados con el consumo de datos equivalente al 43%. Estos datos nos llevan a la conclusión de que se requiere un mayor esfuerzo de la comunidad investigadora en las fases del ciclo de vida de los datos relacionadas con su publicación.
- El resultado del estudio de mapeo sistemático permite identificar que solamente el 4% de las investigaciones corresponden al tópico calidad.
- Los documentos analizados, se centran en la explotación de datos, pero se ha descuidado las primeras fases del ciclo de vida de los datos, que corresponden a la publicación, la creación, la interconexión y la armonización.
- Se identifica la necesidad de analizar mecanismo de divulgación de datos centralizada e investigar en problemática comunes identificados como: falta de metadatos, datos incompletos, falta de información sobre actualizaciones de datos, que afectan directamente a la comprensión de los datos por parte del usuario.
- Uno de los elementos coincidentes es que la calidad es importante desde el origen del dato en las organizaciones, investigar en esta temática junto con una adecuada estructura de metadatos, permitirá que los datos publicados sean comprensibles para los usuarios.

4. CALIDAD DE DATOS Y SU IMPACTO EN LA PUBLICACIÓN DE DATOS ABIERTOS

Este capítulo presenta un análisis de estándares y buenas prácticas que permiten mejorar la calidad de datos de manera general. Se incluye la descripción de los conceptos de metadatos y de comprensibilidad de los datos, ya que están relacionados con la calidad de los datos publicados en abierto. Bajo este escenario, también se describe el concepto de datos maestros como un esquema de solución a la problemática de la calidad de datos cuando existen de diferentes fuentes datos heterogéneas.

Universitat d'Alacant
Universidad de Alicante

4.1 Calidad de datos

Los portales de datos abiertos son la interfaz más utilizada por los reutilizadores (tanto compañías como ciudadanos) para acceder a la información que una organización publica en la Web. Sin embargo, la falta de calidad de los conjuntos de datos publicados es una barrera para que el proceso de reutilización sea masivo. En esta sección se describirán aquellos aspectos de la calidad de datos de interés para la publicación de datos abiertos, incluyendo la propia definición de calidad de datos.

Para un usuario de los datos, la calidad de los datos implica que estos cumplan un propósito definido por este mismo usuario [100], es decir, la "adecuación al uso". También se puede entender que los datos son de calidad cuando cumplen los requisitos para los que fueron diseñados [76]. Por otro lado estos datos se crean, almacenan, procesan, intercambian, comparten, agregan y reutilizan, siendo de calidad si son útiles, es decir si se obtiene el valor esperado que se deriva de ellos [101].

Para entender la calidad de los datos de una organización, se aplicará la analogía de considerarlos como un producto de la propia organización. En este contexto, entonces, se requiere lo siguiente: (i) tener en cuenta las necesidades de los usuarios; (ii) definir un proceso de generación de datos claro; (iii) identificar el ciclo de vida de los datos; (iv) nombrar a una persona que gestione los datos y, finalmente, (v) evaluar la calidad del producto de datos. Solo así la organización podrá entender que los datos son importantes y tienen que cumplir procesos objetivos para asegurar su calidad [102], [103].

Considerando estos elementos aplicables al proceso de apertura de datos, en donde el usuario es la comunidad reutilizadora, la calidad de los datos abiertos se logra únicamente si los datos publicados en un portal de datos abiertos cumplen con los requisitos de esta comunidad. En caso contrario, los usuarios perderán interés en el portal de datos abiertos y el portal quedará abandonado [76]. De hecho, el esfuerzo que hacen las organizaciones y administraciones públicas para publicar datos abiertos se ve afectado cuando no existe una gestión adecuada de la calidad de los datos abiertos publicados en un portal, ocasionando que se desechen proyectos de datos abiertos al poco tiempo de ponerse en marcha [104]. La calidad de datos en un portal debe ser un proceso multidimensional que se inicia desde el momento que se digitaliza la información y considera los requisitos de la comunidad reutilizadora según varias dimensiones [105], [106] que se describen a continuación:

1. Exactitud, es decir, datos que representan correctamente la entidad o evento del mundo real.
2. Puntualidad, datos que representan la situación real y la agilidad en la publicación.
3. Consistencia, no contienen contradicciones.
4. Exhaustividad, incluyen todos los elementos de datos que representan a la entidad o evento.
5. Disponibilidad, accesibles ahora y con el tiempo.
6. Conformidad, siguen las normas aceptadas.
7. Credibilidad, son de fuentes confiables.
8. Accesibilidad, son legibles por máquina.
9. Relevancia, cantidad adecuada de datos.

Las cuatro primeras dimensiones (1-4) de calidad citadas, dependen totalmente de los sistemas de información de origen. Las últimas dimensiones (5-9) dependen de la gestión implementada en los portales de datos. Desafortunadamente, las investigaciones actuales se han centrado exclusivamente en analizar la calidad de los datos ya publicados [107], es decir los elementos de calidad del (5-9), no existiendo trabajos de investigación que profundicen en las dimensiones (1-4) con el fin de vincular la calidad de los sistemas de origen y el proceso de apertura de datos para conseguir datos abiertos publicados, con la calidad adecuada, en un portal.

4.1.1 Planificación de la calidad de datos

La calidad de datos requiere una planificación que tenga en cuenta el posible proceso posterior de apertura de datos y de reutilización de estos datos una vez publicados. La gestión de la calidad de datos involucra personas, procesos y tecnología, siendo necesario definir un programa de gestión de calidad de datos [101]. Los elementos que deben ser considerados son los siguientes:

- Definición de las dimensiones de la calidad de los datos relacionadas con los valores de los datos, formatos y su semántica.
- Definición de estándares para evaluar la calidad de datos.
- Definición de estándares para metadatos

- Considerar aspectos de gobernanza de datos

En este contexto es valioso mencionar que, en [108] se manifiesta que los problemas que enfrentan los portales de datos abiertos están relacionados con aspectos relacionados con la gestión de la calidad de datos, a saber: la falta de estándares de datos, la dificultad en el acceso a los datos y la poca capacidad de comprensión. En el caso de los datos abiertos, es la comunidad reutilizadora la que utiliza de manera directa los datos, convirtiendo a dichos datos en un activo para la organización que los publica, en cuanto a que se está generando un impacto en la sociedad.

4.1.2 Estándares para conseguir calidad de datos

El concepto que permite implementar la gestión de la calidad de datos es el archivo maestro de datos o MDM (de sus siglas en inglés Master Data Management), el cual se define en la norma ISO 8000, que considera la calidad como un proceso multidimensional. Este proceso es afectado por el origen disperso de los datos, para lo cual se propone el concepto de MDM como estrategia de creación y apropiación de los datos críticos del negocio, considerando sus diferentes orígenes [109].

La estructura del dato maestro deberá contemplar las características planteadas en la ISO 8000 [7] que son: identificación de estado, origen de los datos almacenados, así como relaciones con otros datos. Esta norma propone que existe una relación directa entre la calidad y la creación y apropiación del concepto de datos maestros dentro de una organización, por cuanto se trata de la recopilación de datos de diferentes orígenes [109].

La creación del archivo maestro responde a la identificación de la necesidad de información por parte de la organización, realizando luego un proceso de combinación, identificación y eliminación de duplicados si fuese necesario. El propósito es entregar la información requerida por el usuario [110]. Esta norma se organiza en varios apartados, desde el 100 al 140, que se describen a continuación:

- Los apartados 100 establecen requisitos sobre cómo intercambiar datos maestros a través mensajes de datos y las recomendaciones para adjuntar información relacionada con la calidad.
- El apartado 110 describe el formato para implementar un diccionario de datos [105], este debe tener los siguientes elementos: la fuente de datos, se refiere a "historia o genealogía / linaje" del dato.

- El apartado 120 proporciona un modelo de datos para adjuntar información sobre la procedencia de los datos considerando el ciclo de vida de los datos, indicando quién, cuándo y qué operación se hizo con los datos [111]. El ciclo de vida del dato permite determinar entradas, entidades, sistemas y procesos que influyen en los datos [30], con el propósito de conocer el registro histórico de los datos y sus orígenes [112]. Al definir el llamado por la ISO “data provenance”, también se logrará detectar datos duplicados para determinar la credibilidad, actualidad o valor de los datos. Esta información se agregará a los datos maestros generados por la institución [111].
- El apartado 130 describe cómo agregar información acerca de la precisión.
- El apartado 140 describe cómo agregar información sobre la completitud.

Además, este estándar sugiere diseñar un formato de intercambio de datos [113], esquema para el registro, identificador de organizaciones e identificador de personas con la sintaxis de identificadores, afirmación de nivel de precisión y exactitud de los datos y mecanismo de reclamo para el incumplimiento de los niveles de calidad [114][115].

4.1.3 Estándares para evaluar la calidad de datos

La calidad de datos, en un portal de datos abiertos, se puede evaluar según diversos criterios, como por ejemplo, la disponibilidad (por ejemplo, como el número de conjuntos de datos y metadatos disponibles en un portal específico) [76]. En este contexto se analizarán los estándares de calidad existentes con el fin de determinar qué criterios pueden ser aplicados al concepto de datos publicados en un portal de datos abiertos.

El estándar ISO / IEC 25000: 2014, es un conjunto de normas internacionales para sistemas y requisitos de calidad y evaluación de software (SQuaRE). También contiene una explicación del proceso de transición entre el antiguo ISO / IEC 9126 y la serie ISO / IEC 14598 y SQuaRE¹. Esta norma se organiza en los siguientes apartados:

- ISO/IEC 25000 - Guide to SQuaRE: contiene el modelo de la arquitectura, la terminología de la familia, un resumen de las partes, los usuarios previstos y las partes asociadas, así como los modelos de referencia.

¹ <https://www.iso.org/standard/35683.html>

- ISO/IEC 25001 - Planning and Management: establece los requisitos y orientaciones para gestionar la evaluación y especificación de los requisitos del producto software.
- ISO/IEC 25010 - System and software quality models: describe el modelo de calidad para el producto software y para la calidad en uso. Esta norma presenta las características y subcaracterísticas de calidad para evaluar el producto software.
- ISO/IEC 25012 - Data Quality model: define un modelo general para la calidad de los datos, aplicable a aquellos datos que se encuentran almacenados de manera estructurada y forman parte de un Sistema de Información.
- ISO/IEC 25020 - Measurement reference model and guide: presenta una explicación introductoria y un modelo de referencia común a los elementos de medición de la calidad. También proporciona una guía para que los usuarios seleccionen o desarrollen y apliquen medidas propuestas por normas ISO.
- ISO/IEC 25021 - Quality measure elements: define y especifica un conjunto recomendado de métricas base y derivadas que puedan ser usadas a lo largo de todo el ciclo de vida del desarrollo software.
- ISO/IEC 25022 - Measurement of quality: define las métricas para realizar la medición de la calidad en uso del producto.
- ISO/IEC 25023 - Measurement of system and software product quality: define las métricas para realizar la medición de la calidad de productos y sistemas software.
- ISO/IEC 25024 - Measurement of data quality: define las métricas para realizar la medición de la calidad de datos

La presente tesis doctoral se orienta hacia los datos, por tanto, se toma como referencia la norma ISO 25012, que propone la evaluación de la calidad del dato en dos dimensiones:

- Calidad de Datos Inherente. Se refiere al grado con que los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas. Desde el punto de vista inherente, se refiere a: valores de los datos con sus posibles restricciones, relaciones entre valores de datos y metadatos.
- Calidad de Datos Dependiente del Sistema. Se refiere al grado con el que la calidad de Datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas. Desde el punto de vista dependiente del sistema,

depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático.

Dentro de las características inherente y dependientes se dispone de las siguientes:

Tabla 3: Descripción de características de calidad

<i>Características</i>	<i>Descripción</i>
Accesibilidad	Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.
Conformidad	Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.
Confidencialidad	Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información (junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.
Eficiencia	Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico
Precisión	Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.

<i>Características</i>	<i>Descripción</i>
Trazabilidad	Grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico.
Comprensibilidad	Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.
Disponibilidad	Grado en el que los datos tienen atributos que permiten ser obtenidos por usuarios y/o aplicaciones autorizados en un contexto de uso específico
Portabilidad	Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico
Recuperabilidad	Grado en el que los datos tienen atributos que permiten mantener y preservar un nivel específico de operaciones y calidad, incluso en caso de fallos, en un contexto de uso específico.

4.1.4 Comprensibilidad de los datos

De acuerdo a los estándares ISO, y tal y como se ha descrito en la sección anterior, la comprensibilidad es uno de los elementos de la calidad. La comprensibilidad involucra varias características como: el grado de utilización de los símbolos comprensibles, los metadatos y el diccionario de datos que deben utilizar un vocabulario común, el modelo de datos y los

datos maestros vinculados [116]. Se revisan a continuación acciones a considerar para el aseguramiento de la comprensibilidad en los conjuntos de datos abiertos publicados en un portal [18]:

- Proporcionar un esquema/ontología/modelo para sus metadatos que se adapte a estándares.
- Derivar los valores de los metadatos directamente de los datos de forma automatizada (por ejemplo, tamaño de archivo, formato, disponibilidad).
- La restricción de ciertos valores de metadatos a una lista predefinida de opciones (por ejemplo, para descripciones de licencias, formatos de campo) y comprobar/validar la conformidad de ciertos valores de metadatos (por ejemplo, URLs, correos electrónicos).

El estudio realizado por [117] revela que las organizaciones actuales no prestan suficiente atención a la gestión de los conjuntos de datos, recursos y metadatos asociados que están publicando actualmente en un portal de datos abiertos [51]. De hecho, los problemas que enfrentan los portales de datos abiertos están relacionados con: la falta de estándares de datos, la dificultad en el acceso a los datos y la poca capacidad de comprensión, lo que está relacionado con la capacidad de las organizaciones para gestionar los problemas de datos, favoreciendo que los datos cumplan las necesidades del negocio y se conviertan en activos de la organización [108].

Por lo tanto, la calidad de los datos y, en particular, su comprensibilidad, es una de las barreras que impide el crecimiento de proyectos de portales de datos abiertos.

4.1.5 Metadatos y diccionario de datos

Los metadatos son "información estructurada que describe, explica, localiza o de alguna otra manera facilita la recuperación, uso o administración de un archivo" o en una manera sencilla la datos sobre los datos [118]

La importancia que tienen los metadatos en los procesos de apertura de datos está estrechamente ligada a la reutilización, ya que ésta podría mejorarse proporcionando metadatos estructurados considerables que permita la comprensibilidad de los datos a reutilizar [119]. El uso adecuado de los conjuntos de datos abiertos muchas veces se ve obstaculizado porque los metadatos están mal documentados, es decir son incompletos y/o imprecisos, y por lo tanto la semántica de los datos puede ser ambigua [120]. Una función clave de un portal de datos es la gestión de los metadatos incluyendo la armonización.

En este escenario el reto está relacionado con la forma de abordar una arquitectura de la información y facilitar la interoperabilidad entre los catálogos de datos publicados en abierto por distintos portales en la Web. El Consorcio World Wide Web (W3C) publicó el Vocabulario de Catálogos de Datos (DCAT), un vocabulario RDF (Marco de Descripción de Recursos) para describir catálogos de datos, basado en 3 conceptos clave: catálogo, conjunto de datos y distribución [121].

4.1.6 Vocabulario de Catalogo de datos (DCAT)

Dentro de varias investigaciones como [122]: se concluye en la importancia que tienen los vocabularios estándar en los conjuntos de datos publicados en abierto, por cuanto independientemente de su cobertura, son inherentemente opacos y requieren metadatos externos para facilitar la catalogación, comprensibilidad y por tanto su descubrimiento. El vocabulario del esquema RDF para catálogos de datos, llamado DCAT, fue desarrollado a partir de una encuesta de los esquemas de metadatos de siete catálogos de datos diferentes, como Dublin Core, SKOS y FOAF [122]. Actualmente para continuar con el desarrollo de DCAT, se ha trasladado al grupo de trabajo en Administración Electrónica del W3C, este catálogo ha sido aplicado en países como de EE. UU, Australia y Europa.

Para la implementación de DCAT se tiene los siguientes requisitos:

- 1) Los metadatos deben almacenarse y manejarse en base a una sintaxis y una semántica bien definidas, es decir, un esquema documentado
- 2) El software para la gestión del catálogo debe ofrecer tanto una interfaz de usuario como una aplicación ampliamente aceptada para el acceso de aplicaciones y portales de datos [123].

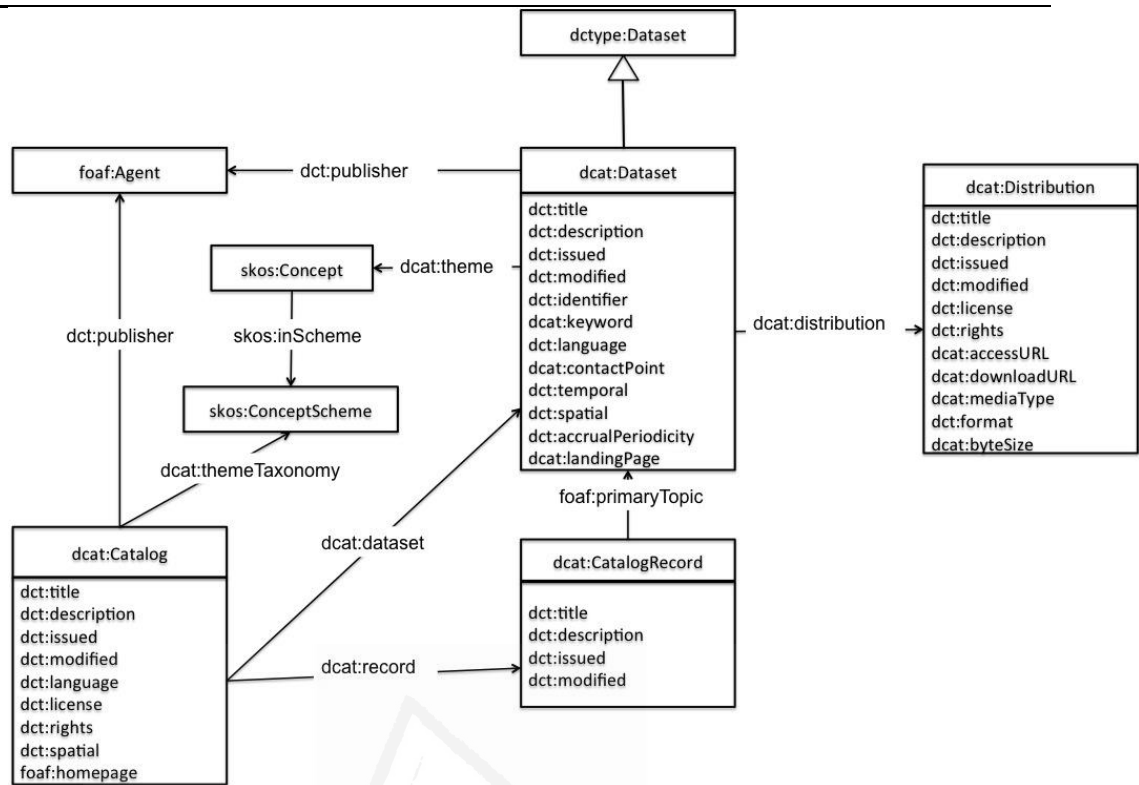


Figura 22: Modelo DCAT tomado de [124]

Para los fines de esta tesis doctoral no se abordará la configuración del modelo, y se centrará en lo que representa un catálogo, el conjunto de datos y la distribución, no se profundizará en foaf y skos por ser conceptos relacionados con la Web Semántica, ya que el objetivo es tratar temas de comprensibilidad de los conjuntos de datos publicados.

El catálogo en sí es una descripción de los conjuntos de datos, el cual describe es un título del catálogo y una descripción, además de metadatos como el lenguaje, la licencia (que en la última versión se puede poner a nivel de distribución), la fecha de publicación, la de modificación.

4.2 Gobierno de datos

Del resultado del estudio del mapeo sistemático presentado anteriormente en este documento, se evidencian problemas recurrentes en los procesos de apertura como: portales abandonados, metadatos erróneos, conjuntos de datos incompletos, falta de información acerca del responsable del dato, entre otros. Para mitigar estos obstáculos, la utilización de las mejores prácticas procedentes del gobierno de datos podría presentar una alternativa de solución [125].

Por otro lado, dentro del manejo de datos emerge una problemática relacionado con que, “las soluciones de gestión de datos por sí mismas se están volviendo muy caras y no son capaces de hacer frente a la realidad de la complejidad duradera de los datos” [126], el problema de los datos, es recurrente en las instituciones y debe ser analizado desde el momento que se emprenden procesos de automatización, es decir, desde el origen mismo del dato. De hecho existen estándares al respecto que podrían usarse como ISO 38550 [108], [126]–[133].

El gobierno de datos es un enfoque hacia la organización, orientado a la gestión de datos e información que utilizando un conjunto de políticas y procedimientos buscar abarcar todo el ciclo de vida de los datos, desde su adquisición hasta su uso y eliminación [134].

El gobierno de datos es abordado desde marcos de referencia, buenas prácticas y/o estándares que dan orientan en la forma de abordarlo, proponiendo principios, procesos, estructuras organizativas, que permitan a la organización implementarlo. Se ha realizado una revisión de la literatura científica y se han identificado los siguientes: DAMA BOOK, DGI Data Governance Framework, IBM Data Governance, ISO 33000, ISO 8000: 61-62, ISO 8000:150. A continuación, se describen los dominios, dimensiones y estructuras organizativas que proponen los diferentes enfoques. De acuerdo a la propuesta definida en [135], se propone los siguientes dominios:

- Principios de datos. Relacionado con establecer el papel de los datos como activo.
- Calidad de los datos. Establecer requisitos para el uso de los datos.
- Metadatos. Establecer la semántica o el "contenido" de los datos.
- Acceso a los datos. Especificar los requisitos para el acceso a los datos
- Ciclo de vida de los datos. Determinar la definición, producción, retención y eliminación de datos.

La gobernanza de datos incluye: la definición de calidad de datos, estructura organizativa sugerida, roles y responsabilidades, y flujos de trabajo para las actividades de los participantes con el fin de proporcionar una supervisión integral de la calidad de la información empresarial [136]. A continuación, se describe cada uno de los elementos que se toman como referencia para el desarrollo del presente trabajo de investigación:

- En el primer elemento se incluye la definición de calidad explicada en la sección 4.1, donde la calidad de datos en procesos de apertura tiene lugar cuando los datos

publicados son aptos para el uso, en este caso para generar procesos automáticos de reutilización por parte de usuario del portal de los datos abiertos.

- En cuanto a los roles y responsabilidades, se tomará en cuenta los resultados procedentes de [137], en donde se analizó una serie de estudios, casos de estudio e informes de analistas y consultores concluyendo las siguientes funciones dentro de un marco de gobernanza de datos:
 - Patrocinador ejecutivo.
 - Consejo de Gobierno de Datos (o Junta de Calidad de Datos).
 - Jefe oficina de Datos /Chief Data Officer.
 - Administrador de datos comerciales /Data Steward.

El Patrocinador ejecutivo y el consejo de Gobierno de Datos (o Junta de Calidad de Datos), son los que cumplen el rol de gobierno dentro de la organización, su importancia radica en la definición de la estrategia de manejo de datos. Dentro de este enfoque se define que los procesos de apertura deben tomar un rol estratégico dentro de la organización como se explicará en el siguiente capítulo.

El rol del jefe de la oficina /Chief Data Officer (CDO), será quien asume varias responsabilidades que se describen a continuación:

- Explotar el valor crítico que los datos abiertos pueden proporcionar.
- Enfocarse en la exploración de nuevas oportunidades estratégicas para la organización.
- Establecer un marco que optimiza la colaboración entre las unidades de negocio internas (enfoque hacia el interior).
- Desarrollar una política de datos interempresarial para la estrategia empresarial y la colaboración externa (hacia el exterior y enfoques estratégicos)

De acuerdo a [138], se establecen los procesos dentro de la gestión de datos que son necesarios implementar, y que se utilizaron para evaluar el nivel de implementación de gobierno de datos en entidades públicas de Brasil:

- Políticas y estándares de datos.
- Gestión de la arquitectura de datos.

- Desarrollo de los datos.
- Gestión de las operaciones de la base de datos.
- Gestión de la seguridad de los datos.
- Gestión de datos maestros y de referencias.
- Gestión de almacenamiento de datos (DW) y Business Intelligence (BI).
- Gestión de documentación y contenidos.
- Gestión de metadatos.
- Gestión de la calidad de los datos.
- Cumplimiento y auditoría de datos.

Por ello, es necesario contar con un modelo de gestión de la calidad de datos abiertos que tenga en cuenta las mejores prácticas del gobierno de datos tal y como se ha visto en esta sección.

Universitat d'Alacant
Universidad de Alicante

5. MARCO DE REFERENCIA PARA APERTURA DE DATOS COMPENSIBLES

Este capítulo presenta el marco de trabajo desarrollado, el cual está compuesto de tres niveles: estratégico, táctico y operativo. Cada uno de estos niveles incluye una serie de procesos que deben ejecutarse para completar un procedimiento de publicación de datos abiertos teniendo en cuenta criterios de comprensibilidad

Universitat d'Alacant
Universidad de Alicante

5.1 Contexto

Un marco de referencia es una estructura que establece los aspectos a considerar en la consecución de un objetivo. En el caso que nos ocupa en el desarrollo de esta tesis doctoral, el objetivo es la publicación de datos abiertos teniendo en cuenta criterios de comprensibilidad. Por ello, el marco de referencia propuesto describe los aspectos a desarrollar en la publicación de datos abiertos con el fin de que sean comprensibles para la comunidad reutilizadora, priorizando que los datos publicados en abierto representen la realidad de la información almacenada en la organización. Este marco de referencia sirve de modelo para apoyar a las organizaciones a publicar sus datos en formatos abiertos, considerando sus sistemas de información de origen y teniendo en cuenta criterios que permitan la fácil comprensión de los datos por parte de la comunidad reutilizadora a través de los metadatos publicados. Para ello se usarán estándares de calidad de datos existentes, adaptándolos las particularidades de los datos abiertos. De manera general, el marco de referencia diseñado posee las siguientes características:

1. Se basa mejores prácticas en el ámbito de las tecnologías de información y comunicaciones (TIC), como las procedentes del ámbito del gobierno de datos.
2. Presenta una adaptación del concepto Master Data Management (procedente del estándar ISO 8000) para considerar las características de la publicación de datos abiertos.
3. Propone una estructura de metadatos que favorece la comprensibilidad de los datos publicados en abierto y propone métricas para su evaluación (adaptando el estándar ISO 25000).
4. No se centra en mejorar la comprensibilidad de los datos abiertos ya publicados, sino que el aseguramiento de la comprensibilidad se realiza desde los sistemas de información de origen (donde se encuentran los datos a publicar en abierto).
5. La propuesta se estructura tres niveles (estratégico, táctico y operativo) cada uno de los cuales se compone de una serie de procesos con el fin de facilitar su aplicación en cualquier tipo de organización.

En la Figura 23, se presenta los niveles con sus correspondientes procesos, describiendo los objetivos de cada nivel dentro del marco de referencia diseñado.

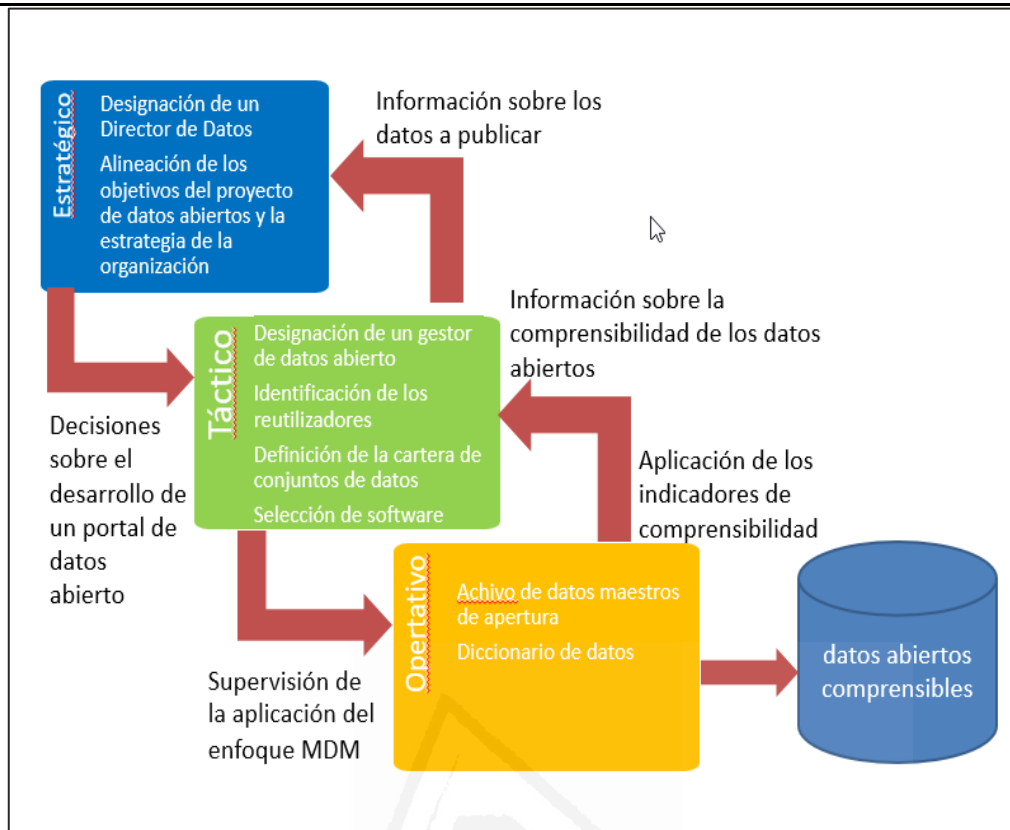


Figura 23: Visión general del marco de trabajo para la publicación de datos abiertos comprensibles.

En la Figura 24, se detalla cada uno de los procesos dentro del marco de referencia.



Figura 24: Procesos a ejecutarse con los responsables adaptado de [131]

Cabe destacar que la definición de este marco de referencia ha sido posible gracias a la aplicación de la metodología de “investigación en acción” comentada anteriormente en este

documento. Esta metodología ha permitido que todas las partes implicadas en la investigación trabajen colaborativamente para evaluar cada una de las propuestas de investigación acerca del marco de referencia y así poder evolucionar el marco de referencia hasta su versión final aquí presentada.

En las siguientes secciones se describirá cada uno de los niveles del marco de referencia y sus procesos asociados. Para una mejor descripción del marco de referencia se utilizará un ejemplo de aplicación, tomando como referencia a una organización gubernamental que está a cargo de la publicación de datos en formatos abiertos con un impacto nacional.

5.2 Nivel estratégico

Considerando el concepto de Gobierno de Datos, se establece que dentro de la organización el desarrollo de un proyecto de datos abiertos que incluya la implementación de un portal de datos abiertos deber ser parte de la estrategia institucional. Es necesario, por tanto, su aprobación al máximo nivel de responsabilidad de la institución.

Una vez presentado un proyecto de apertura de datos, el rol de la parte estratégica es considerar los siguientes aspectos:

- La designación de un responsable del proyecto de datos abiertos, quien posee una visión institucional para gestionar los datos que deben publicarse en formatos abiertos. El rol asignado a esta persona se corresponderá con el cargo directivo de Chief Data Officer (CDO) [139], [140], pero adaptado a los objetivos de un proceso de apertura de datos. Concretamente, estará a cargo de desarrollar estrategias para la implementación, gestión y supervisión del proyecto de datos abiertos
- Orientar a la organización en el cumplimiento de la legislación vigente relacionada con el concepto de datos abiertos, así como otro tipo de normativa o regulaciones. Se debe tener en cuenta la legislación relacionada con la seguridad de datos, con proyección de datos y otra legislación en el ámbito de la transparencia y de la reutilización de la información del sector público. Además, se deben considerar regulaciones o políticas dentro del contexto en donde se ejecutará el proyecto de datos abiertos.

Los procesos a realizarse en este nivel relacionado con un proyecto de datos abiertos son los siguientes:

1. **Aprobación**, el objetivo es establecer como estratégico al proyecto de datos abiertos, este proceso deberá iniciar con la entrega de una propuesta del proyecto y las salidas serán la definición del responsable y los lineamientos jurídicos y de seguimiento.
2. **Seguimiento**, el objetivo es conocer el estado de proyectos de datos abiertos la entrada para este proceso es el informe estado del proyecto de datos abiertos que se explica en la sección 5.3.5, y la salida se encuentran las políticas estratégicas ajustadas con el fin de fortalecer los procesos de reutilización.
3. **Marco Jurídico**, el objetivo es establecer las condiciones jurídicas internacionales, nacionales y locales que el proyecto de datos abiertos deberá considerar, por tanto, el resultado de este proceso son los lineamientos para la implementación del proyecto.

Como ejemplo de aplicación, una organización nacional ecuatoriana que ejecuta un proyecto de apertura de datos debe definir los siguientes procesos:

- **Aprobación a nivel estratégico**. Se define la necesidad de implementar un proyecto de datos abiertos, por lo que se necesita la aprobación de la política estatal para apoyar la gestión de un proyecto de datos abiertos. Los lineamientos del proyecto se emiten directamente de la Presidencia de la República.
- **Definición del responsable del proyecto de apertura de datos**. En una segunda instancia se define al ministerio "Organización", como responsable del proyecto de apertura de datos y se crea una Gerencia de Apertura de Datos.
- **Orientación legal del proyecto**. Se define como marco jurídico lo definido en "Open Government Partnership" (OGP) -en español, "Alianza para el Gobierno Abierto"- y se asume lo indicado por la carta de datos abiertos [141]. Por otro lado, también se define la necesidad de adoptar la ley de protección de datos vigente en el país.

Con estos insumos que da institucionalidad a un proyecto de apertura de datos, que permitirá trabajar en los siguientes niveles: táctico y operativo.

5.3 Nivel táctico

En este nivel se coordinan las acciones que se emprenden en el nivel operativo, siendo el nivel responsable de las decisiones que se tomen para promover la implementación de un proyecto de datos abiertos. Las actividades a ejecutar se orientan a obtener mayor valor de

los datos publicados en formatos abiertos. A continuación, se describen cada uno de los procesos de los que se compone este nivel táctico.

5.3.1 Alcance - Portafolio de Conjunto de Datos

Definir el alcance del proyecto significa establecer cada una de las etapas a desarrollarse siguiendo una planificación anual. Esta planificación es necesaria para una completa comprensibilidad de los conjuntos de datos a tener en cuenta en el proyecto de datos abiertos que serán publicados en el portal de datos abiertos, por cuanto el usuario que accede a esta información conoce las acciones emprendidas en el portal, y que no responde a un proceso ocasional sino continuo.

Dentro de este proceso se deberá definir el portafolio de conjuntos de datos, es decir, determinar los datos que deben publicarse y la fecha tentativa de publicación. Para esto se realiza un análisis de compensación entre los requisitos de los reutilizadores y los conjuntos de datos disponibles de tal manera que se prioricen los conjuntos de datos que se van a publicar en abierto. Además, este portafolio debe incluir la especificación de la licencia de reutilización de los datos abiertos. El propósito del portafolio de conjuntos de datos es definir toda la información necesaria para facilitar la publicación de conjuntos de datos considerados como estratégicos. La estructura propuesta para este portafolio es la siguiente:

- Código del conjunto de datos.
- Nombre del conjunto de datos a publicar.
- Descripción.
- Departamentos dentro de la organización que generan la información de origen.
- Fecha de la publicación de los datos en abierto.
- Estado actual del proceso de publicación de datos en abierto (iniciado, en construcción, previo a publicación).
- Descripción de la licencia para publicación de los datos en abierto.
- Departamento solicitante de la publicación.

En el Anexo 1: Ejemplo de un catálogo de conjuntos de datos a publicarse.

5.3.2 Informe de inicio del proyecto y actualizaciones

Se debe crear un informe de un proyecto de datos abiertos que permita describir las acciones para iniciar, planificar, ejecutar, monitorear y controlar la publicación de los datos en abierto. En este informe de inicio se incluye una especificación de forma de creación del proyecto de datos abiertos y la estructura del equipo para generar los conjuntos de datos a publicarse. También se define los usuarios internos y externos identificados, así como el alcance, los tiempos y el costo del proyecto.

Es necesario generar los informes periódicos de avance del proyecto con el fin de detallar el estado actual de la publicación de los conjuntos de datos, los porcentajes de uso y la evaluación de comprensibilidad aplicadas de acuerdo a lo definido en la Sección 5.3.5

5.3.3 Acciones de gestión para el uso de datos abiertos (usuarios reutilizadores)

Un proyecto de apertura de datos requiere un cambio cultural en la institución, ya que se debe entender cómo se pueden utilizar los datos publicados y las condiciones de uso que tienen. Esto involucra a usuarios internos y externos (considerando a los potenciales reutilizadores).

La gestión del cambio cultural es una actividad importante, ya que favorece a la reutilización de la información publicada. Para esta etapa se establecen cuatro procesos:

1. Identificación de usuarios: se debe trabajar en la identificación de los usuarios externos e internos del portal de datos abiertos. Los usuarios de un portal de datos abierto son los reutilizadores de dichos datos y pueden ser tanto internos de la organización como externos. Mientras que los reutilizadores internos son fáciles de detectar, los reutilizadores externos son infomediarios o ciudadanos que generan valor a partir de los datos publicados y su apoyo es importante para identificar cómo los datos que son requeridos para la publicación. El responsable de datos abiertos deberá impulsar iniciativas con los reutilizadores externos, dedicando tiempo a crear estrategias y coaliciones para lograr el cambio. Los potenciales reutilizadores identificados deben ser capacitados para utilizar los datos abiertos publicados².
2. Diseño de metas: establecer las metas a corto y largo plazo del proyecto de apertura de datos, incluyendo acciones que permita el cambio y cultura en el uso de los datos

² <https://www.europeandataportal.eu/elearning/en/module3/#/id/co-01>

publicados. Establecer claramente las metas del proyecto de datos abiertos, permitirá definir acciones de planificación, ejecución y seguimiento para que los conjuntos de datos publicados sean utilizados por los potenciales reutilizadores, objetivo primordial de un proceso de apertura. Dentro de este tipo de proyectos la gestión del cambio orientado al uso de datos abiertos, será un elemento esencial para favorecer la reutilización de los conjuntos de datos publicados, para lo cual será útil y necesario implementar técnicas estructuradas para respaldar la captura y el análisis de datos relacionados con la gestión del cambio [142], en este caso orientado a evaluar el nivel de reutilización de datos.

3. Reglas y normas: proveen una guía sobre el proceso de apertura tanto para los usuarios internos como externos.
4. Diseño de catalizadores: Un catalizador (se define en este trabajo como una condición precursora favorable en el proceso de diseño e instrumentación [142]). Un catalizador es un proceso que permitirá acelerar la reutilización de los datos publicados. Por ejemplo, las acciones de relación con los potenciales reutilizadores para diseñar los conjuntos de datos requeridos. Los usuarios pueden ser internos o externos.

5.3.4 Definir herramientas de publicación de datos abiertos

La ejecución de un proyecto de apertura de datos incluye la implementación del portal de datos abiertos, que es un espacio en la web, que presenta datos de manera que facilite su reutilización [143]. Por tanto, el responsable del proyecto de portal de datos abiertos deberá ejecutar las siguientes acciones:

- Identificar las herramientas de software existentes para la implementación de un portal de datos abiertos. Existen varios trabajos de investigación, relacionados con esta temática, en donde se identifican tres plataformas más utilizadas para publicar datos en abierto [79]: (i) Socrata que es software propietario y tiene un API muy potente, (ii) una plataforma de código abierto denominada CKAN que fue desarrollada por OKFN (Open Knowledge Foundation), y (iii) la plataforma de publicación de datos OpenDataSoft.
- Realizar un análisis de las características técnicas y el presupuesto de la organización logrando un informe técnico de la selección adoptada.
- Definir el plan de implementación una vez realizada la selección de la plataforma adecuada para la publicación de datos en abierto.

5.3.5 Medir, monitorear y comunicar uso de los datos publicados

El proceso de monitoreo y evaluación, permiten establecer el desempeño real del proyecto de datos abiertos y contrastarlo con lo planificado. Uno de los objetivos, cuando se desarrolla un portal de datos abiertos, es conocer cómo se han utilizado los conjuntos de datos publicados. Esto está relacionado directamente con el plan de publicación de datos y con la identificación de usuarios internos y externos. En este contexto, para el presente marco de referencia, se plantea la evaluación de los conjuntos de datos publicados siguiendo tres indicadores:

- Porcentaje de publicación en función del catálogo planificado, es decir la comparación de lo ejecutado en función de los planificado.
- Porcentaje de descargas o accesos con relación al periodo anterior, es decir, se trata de un contador automático que se deberá implementar en el portal.
- Niveles de comprensibilidad en un periodo establecido. Se propone utilizar los indicadores de comprensibilidad explicados en la Tabla 4, los cuales fueron diseñados para identificar la comprensión del usuario (potencial reutilizador). Estos indicadores se aplicarán periódicamente, quedando definida la periodicidad en este proceso. Los indicadores propuestos son una adaptación de los incluidos en el estándar ISO 25000.

Tabla 4: Descripción de indicadores de comprensibilidad

<i>Indicador</i>	<i>Descripción</i>
Porcentaje entre el número de valores de datos presentados por símbolos desconocidos y el número de datos que se solicita la revisión de los símbolos.	Comprensión de los símbolos. Porcentaje de uso de símbolos comprensibles.
Porcentaje del conjunto de datos que se entiende la descripción frente a los que no se entienden.	Comprensión semántica. Relación entre el vocabulario común reconocido y los términos o definiciones utilizados en el diccionario de datos.
Porcentaje de comprensión del diccionario de datos. Es una relación entre el número de valores de datos	Comprensión del diccionario de datos. Relación de comprensión de los términos utilizados en un vocabulario común.

<i>Indicador</i>	<i>Descripción</i>
definidos en el diccionario de datos utilizando identificados con el uso de un vocabulario común y los valores de datos definidos en el diccionario de datos.	
Porcentaje entre el número de metadatos entendidos frente a los metadatos de interés para los reutilizadores potenciales en un contexto específico de uso.	Comprensión de los metadatos. Relación entre la comprensión del significado de los elementos de datos.
El porcentaje de conjuntos de datos que son entendidos por la licencia, su aplicación en relación a los que no son entendidos.	Comprensión de la licencia de uso. Relación entre el conocimiento del uso que se le puede dar a una licencia y el espectro de aplicación.
Porcentaje de criterios de anonimato entendidos frente a los no entendidos.	Comprensión de las normas de anonimato y su aplicación.

5.4 Nivel operativo

Este nivel tiene como responsabilidad la ejecución de la publicación en abierto de los conjuntos de datos según su grado de comprensibilidad. Este nivel contiene tres procesos: definición del archivo maestro de apertura de datos, creación de los metadatos y generación del diccionario de datos.

5.4.1 Definición del archivo maestro de datos abiertos

Dado que los datos que se van a publicar en abierto proceden de diferentes fuentes de datos de una organización, se necesita un repositorio centralizado de metadatos para que los datos abiertos sean comprensibles. Para ello, proponemos utilizar un enfoque de Gestión de Datos Maestros (MDM) como ISO 8000, el cual permite a una organización relacionar todos sus

datos críticos en un único archivo llamado archivo maestro. En este sentido, el archivo maestro de datos abiertos se define en el sistema de información de origen con el fin de contener toda la información que será posteriormente necesaria para publicar los datos en abierto de manera comprensible. Las actividades a realizar son las siguientes [144]:

- Definición del conjunto de datos requerido para su publicación en abierto.
- Identificación del origen de cada uno de los datos.
- Diagnóstico de calidad de los datos requeridos específicamente se evaluará los siguientes indicadores: completitud (relacionado con la evaluación de los valores nulos en los campos requerido) y exactitud (asociado con el porcentaje de datos nulos, blancos, incompletos, duplicados, datos sin formato, datos sin estándar).
- Proporcionar una clara comprensión de quién es responsable de los datos.
- Definir las reglas de negocio y los valores de dominio válidos para sus datos comunes (con fines de publicación en abierto).
- Descubrir dónde no se siguen las reglas de negocio y los valores de dominio válidos y retroalimentar al responsable del dato.
- Habilitar la gestión de datos maestros de apertura, considerando que tendrá fines de publicación en abierto.
- Definir estándares de anonimización.
- Definir licencia de uso.

No obstante, en primer lugar y de manera previa, ese debe establecer la situación de los datos en la organización, es entender la forma de gestión de los datos implementadas dentro de la organización como: gobierno de datos, gestión de metadatos, arquitectura de datos, gestión de calidad de datos, archivos maestros entre otros, o la inexistencia de una política de gestión de datos. A partir de este diagnóstico la organización podrá implementar la creación del archivo maestro de datos abiertos, el cual deberá cumplir con las siguientes actividades:

- Estructuración de conjunto de datos requerido. este proceso se realiza en base a lo definido en la sección 5.3.1.
- Consulta de datos: tiene por objeto identificar los datos en el sistema de información de la organización, de acuerdo con los requisitos levantados. Se crea un documento

único que contiene la siguiente información: identificación de la fuente de datos, responsable del dato y fecha de acceso.

- Integración de datos: evaluar la calidad de los datos de las fuentes de datos, así como el proceso de integración necesario para vincular las fuentes de datos antes de publicarlas como datos abiertos, incluyendo los metadatos de origen y el diccionario de datos.
- Definición de datos: aplicación de estándares de anonimización, establecimiento de formatos de publicación de datos y licencia de uso.
- Generación del archivo maestro de datos abiertos: es la generación automatizada del archivo maestro de datos abiertos en formatos establecidos.

Por lo tanto, una vez que los datos a publicar se estructuran, identificando su origen y su nivel de calidad, se crea el archivo maestro de datos abiertos que es una estructura que admite la vinculación de fuentes de datos heterogéneas, para definir un dato unificado dentro del conjunto de datos, luego de este proceso se aplica la metodología de anonimización y se establece la licencia de uso.

5.4.2 Definición de metadatos

El archivo maestro de datos abiertos contiene todos los metadatos de las fuentes de datos que se necesitarían para publicarlos como datos abiertos. Los metadatos son el puente entre el conjunto de datos y el usuario (potencial reutilizador), el mismo permitirá entender la estructura del conjunto de datos, así como otros datos relacionados al manejo de la información publicada. El grado de utilización de los símbolos comprensibles permitirá que la persona que accede al conjunto de datos, pueda establecer el alcance del conjunto de datos. Dentro de las buenas prácticas, para definir los metadatos existe DCAT (Data Catalog Vocabulary) [145] y OAI-PMH (Open Archive Initiative Protocol for Metadata Handling) [146], por tanto luego de creado archivo maestro de datos abiertos, se estructura los metadatos de acuerdo a la propuesta descrita en la Tabla 5.

Tabla 5: Estructura de metadatos

<i>Campo</i>	<i>Descripción</i>	<i>Propiedad</i>
Identificador	Identificador único de un conjunto de datos de las	dct:identifier

<i>Campo</i>	<i>Descripción</i>	<i>Propiedad</i>
	fuentes de datos de la organización.	
Nombre	Título del conjunto de datos.	dct:title
Descripción	Descripción textual completa del conjunto de datos.	dct:description
Fechas_de_creación	Fecha de creación del conjunto de datos.	dct:issued
Frecuencia_de_actualización	Frecuencia con la que se actualizan los datos del conjunto de datos.	dct:accrualPeriodicity
Archivo maestro	La actualización se realiza a través de un archivo maestro de datos abiertos.	fw:openmasterdata
Responsable_conjunto_datos	Contacto de la persona responsable del conjunto de datos.	fw:responsibledata
Anonimización	Estado de anonimización del conjunto de datos.	fw_anonymization
Gestor_datos	Contacto del gestor de datos abierto de la organización.	fw_managerportal
Licencia	Licencia aplicable al conjunto de datos (es decir, condiciones de reutilización).	dct:license
Lenguaje	Idioma de publicación.	dct_language
Organización	Sitio web de la organización.	dct_organization

<i>Campo</i>	<i>Descripción</i>	<i>Propiedad</i>
Palabras_clave	Palabras clave de búsqueda para el conjunto de datos.	dct_keywords

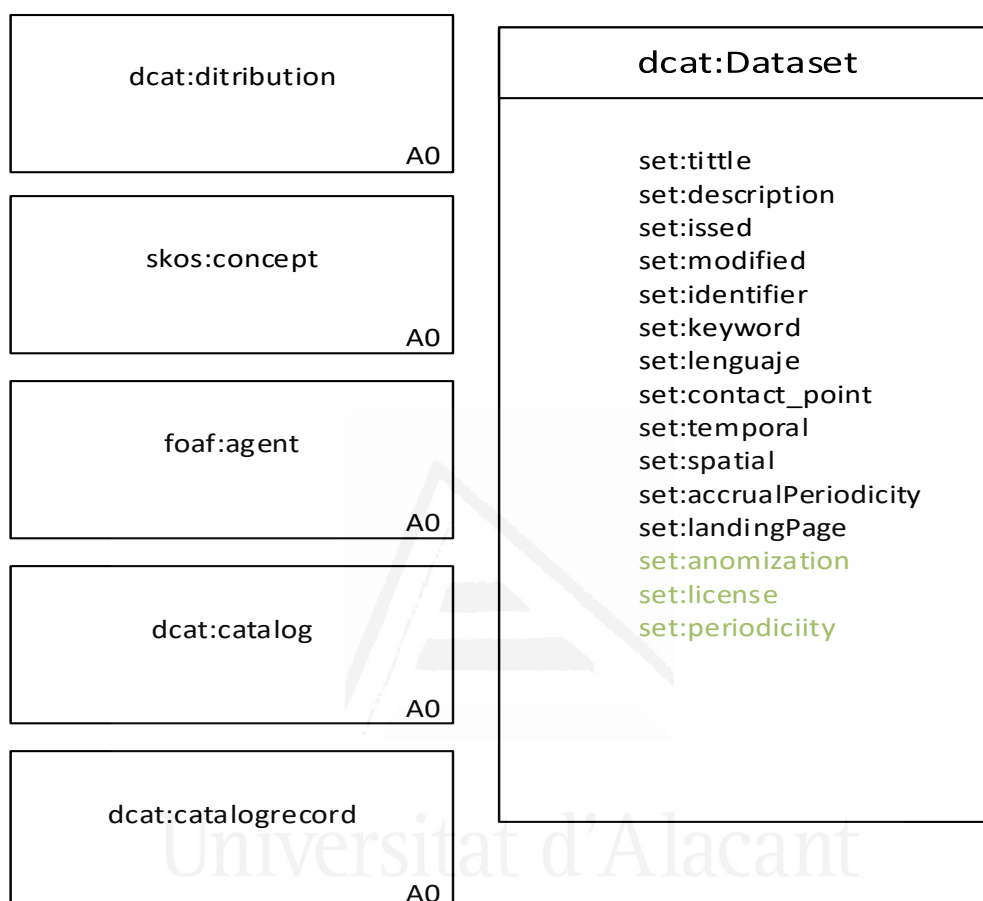


Figura 25: Estructura del metadato basado en DCAT [124]

De acuerdo a la Figura 25, se presenta una estructura del metadato. En concreto, dcat:dataset describe la estructura del metadato, se agrega un nivel de anonimización, licencia y periodicidad del dato, por cuanto estos datos son requeridos para entender la semántica del conjunto de datos publicado. Se considera a la estandarización como una característica de la comprensibilidad, por tanto, se sugiere siguientes buenas prácticas:

- Registro del historial para conocer los niveles de actualización, considerando la ISO 8000 (procedencia) o PROV [147], [145] que propone una ontología para poder generar el metadato de la procedencia del dato, como una información relacionada al identificar el responsable, fecha y la forma de exponerlo.

- Nivel de organización, es la descripción de las jerarquías de conceptos, se puede utilizar skos para clasificar como un sistema de árbol relacionado con el esquema de organización de conocimiento, describiendo los temas se tratan en el conjunto de datos.
- Estandarización, establecer decisiones de estandarización nacional que permita comprender mejor los datos, se podría utilizar al estructura foaf:agent donde cada entidad tiene un identificador único, facilitando una posterior generación de datos abiertos enlazados.
- La definición de licencia específica, describe como el usuario podrá utilizar los conjuntos de datos para generar valor (uso por parte de los infomediarios) , estos deberán tener el máximo de acceso posible, es necesario considerarla licencia Creative Commons [148].

A continuación, se presenta un ejemplo de aplicación de la estructura de metadatos.

Tabla 6: Ejemplo de aplicación de la estructura de metadatos

<i>Campo</i>	<i>Descripción</i>
Identifier	30339983
Name	Información de viaje de las autoridades públicas y funcionarios de una organización.
Description	Contiene información sobre las dietas de viaje o asignaciones nacionales de las autoridades y funcionarios de la organización "xxx" en enero de 2018
Creation date	Junio 8, 2018
Update_frequency	Mensualmente
Opendatamaster	Si
Owner_dataset	managerportal@organizacion.gob.ec

<i>Campo</i>	<i>Descripción</i>
Anonymization	Si
Organization_manager_portal	datamanager@organizacion.gob.ec
License	Attribution 4.0 International (CC BY 4.0)
Language	Español
Organization	Organization.gob.ec
Keywords	subsídios, salarios.

El diccionario de datos contiene información sobre los conjuntos de datos y sus descripciones (llamados atributos). En aras de una mejor comprensión, debe describirse explícitamente el significado de los datos que se van a abrir, así como su estructura. Por lo tanto, se considera la siguiente información de cada conjunto de datos: nombre, descripción, dominio y tipo de datos.

En el siguiente capítulo se presenta la evaluación del marco de referencia y los resultados obtenidos.

Universitat d'Alacant
Universidad de Alicante

6. RESULTADOS Y DISCUSIÓN

Este capítulo presenta la evaluación del marco de referencia. La evaluación se realizó en dos etapas, la primera con una serie de entrevistas que permitió realizar un diagnóstico de las prácticas de publicación de datos, mientras que la segunda etapa consistió en la aplicación de grupos de discusión, en tres iteraciones con el fin de mejorar el marco de referencia.

Universitat d'Alacant
Universidad de Alicante

6.1 Diseño de la metodología de evaluación

En esta sección se presentan los resultados de la evaluación del marco de referencia diseñado, mediante la comparación de las prácticas de publicación de datos actuales y la aplicación del propio marco de referencia para mejorar dichas prácticas. Se usan dos técnicas diferenciadas para obtener datos cualitativos sobre las realidades de la publicación de datos abiertos a nivel nacional en Ecuador (mediante la realización de entrevistas), así como de la aplicabilidad del marco de referencia propuesto con el fin de realizar su evaluación (mediante grupos de discusión). Para cumplir los objetivos propuestos mediante el uso de técnicas (entrevistas y grupos de discusión), se propone un conjunto de preguntas. En primer lugar, en lo que respecta a las entrevistas para recolectar y analizar información relacionada con las prácticas de apertura de datos, la cuestión a resolver es la siguiente:

- ¿Cómo se relacionan las prácticas de publicación de datos utilizadas por las organizaciones con la estructura del marco de referencia diseñado?

En segundo lugar, con respecto al grupo de discusión con el fin de indagar acerca de la aplicabilidad del marco de referencia y su aporte a publicar datos con criterios de comprensibilidad, las cuestiones a analizar son las siguientes:

- ¿Es aplicable el marco de referencia en las organizaciones del sector público ecuatoriano?
- ¿El marco de referencia diseñado favorece la publicación de conjuntos de datos comprensibles?

6.2 Población participante

La población interviniente en el proceso de evaluación (tanto para las entrevistas como para el grupo de discusión) se conformó por responsables de portales de datos del gobierno central de Ecuador, responsables de portales de datos abiertos del sistema de educación superior en Ecuador y representantes de portales de la sociedad civil ecuatoriana.

- Los responsables de portales del gobierno central fueron identificados a través de la institución encargada de coordinar el proceso de publicación en abierto del país (Gerencia de Gobierno Electrónico del Ministerio de Telecomunicaciones y Sociedad de la Información).

- Para determinar la participación académica, se aplicó una encuesta a todas las universidades reconocidas en el país, obteniendo una participación del 70%, de las cuales 14 contestan que tienen un proceso de implementación de un portal de datos abiertos, pero solamente 5 tienen el proceso aprobado. Estas últimas se incluyeron en el presente estudio.
- De la sociedad civil se incluyó a los responsables de portales de datos abiertos con fines de participación ciudadana en los cuales no tiene injerencia el estado.

Las técnicas de investigación cualitativa (entrevistas y grupo de discusión) incluidos en la investigación se desarrollaron de la siguiente manera:

1. Se realiza una entrevista a cada uno de los expertos para indagar los procesos implementados en la publicación de los datos abiertos en su organización. Esta entrevista se estructuró de la siguiente manera:
 - Siete preguntas cerradas orientadas a contextualizar las acciones tomadas en torno a la implantación del portal de datos abiertos por parte de la persona entrevistada.
 - Cuatro preguntas abiertas que tienen la finalidad de identificar las acciones en el portal para lograr la comprensibilidad de los datos publicados por parte de la persona entrevistada.

En el Anexo 2, se adjuntan todas las entrevistas realizadas luego de digitalizarlas. La guía para la recopilación de datos de cada una de las entrevistas realizadas fue la siguiente:

- Se dio prioridad a que los entrevistados proporcionen respuestas espontáneas.
- Las entrevistas duraron veinte minutos y fueron grabadas.
- El periodo que se realizó la entrevista fue entre enero y junio del 2019.
- Las grabaciones se transcribieron y codificaron.
- En el primer paso de "codificación abierta", se evaluó cada cita textual como significativa y fue anotada, todas las entrevistas fueron codificadas considerando la cita textual y los sinónimos.
- Inicialmente se codificaron tres entrevistas de forma independiente por dos investigadores, para luego comparar los resultados.
- Se realizaron ajustes al proceso de tal manera que el análisis de las transcripciones restantes se estandarizaría y sin sesgo.

- Una vez que todas las transcripciones fueron analizadas y la codificación final doblemente verificada por los dos investigadores, se procedió a completar las observaciones para cada entrevista con "notas analíticas para explicar los detalles"
 - Luego las tablas de codificación individuales se fusionaron en tablas maestras, cada una incluyendo una compilación de todos los objetivos determinados para cada uno de los aspectos analizados [149].
2. Con el fin de llevar a cabo el grupo de discusión en relación a las dos cuestiones planteadas en la sección anterior, se siguen las siguientes directrices:
- El objetivo es evaluar el marco de referencia propuesto por parte de expertos.
 - Se utilizó el método de grupo de discusión con expertos identificados del gobierno ecuatoriano, así como procedentes de la academia y sociedad civil de Ecuador.
 - La duración fue de una hora y se realizó en los meses de abril y julio del 2019 en tres grupos de 4 a 7 participantes por grupo.
 - Se realizó una presentación de los resultados de las entrevistas previas y del marco de referencia propuesto.
 - Se presentaron las cuestiones planteadas y se receptaron todas las observaciones a partir de la discusión de los expertos dirigida por la autora de esta tesis doctoral.

Se codificaron los aportes de cada uno de los grupos para proceder a fusionarlos y generar conclusiones. En el Anexo se puede observar los aportes digitalizados de los grupos de discusión realizados.

6.3 Procesamiento de la información de las entrevistas

Se analizan los resultados de las entrevistas realizadas a los gestores de los portales de datos abiertos con el fin de identificar las prácticas de publicación de datos que se han implementado en los portales de datos abiertos. De las 16 entrevistas realizadas, 8 entrevistados son responsables de los portales de datos abiertos ya implementados con datos publicados (específicamente, un total de 512 conjuntos de datos abiertos publicados), 6 entrevistados que ya implementaron los portales (sin datos publicados) y 2 entrevistados sólo han aprobado el proyecto para la creación del portal.

En las siguientes secciones, se presenta los resultados del análisis de la información recabada por las entrevistas realizadas.

6.3.1 Incluir a los potenciales reutilizadores en el proceso

En cuanto a los proyectos de datos abiertos cuyos responsables son entrevistados, el 63% no considera a los reutilizadores en su proceso inicial de apertura de datos. Las razones para ellos son variadas, pero se podrían destacar dos de ellas: el desconocimiento y no existencia de la necesidad, así como la falta de una determinación interna, porque el proyecto se consideró una prueba piloto, o por ausencia de legislación o solicitud directa. También, hubo responsables que argumentaron que no se consideró necesario y otros que argumentaron un cambio de personal y autoridades.

Por otro lado, los responsables que sí consideraron a los reutilizadores lo hicieron porque estaban convencidos de que aportaba valor y utilidad, ya que se trata de los usuarios del portal de datos abiertos, además de por motivos de transparencia y calidad de los datos.

Además, se realizó una pregunta cerrada relacionada con conocer la opinión de la importancia de incluir a los potenciales reutilizadores y el 75% están de acuerdo en que es muy importante. Los resultados se pueden observar en la Figura 26: Percepción de los gestores de los portales acerca de incluir a los reutilizadores.

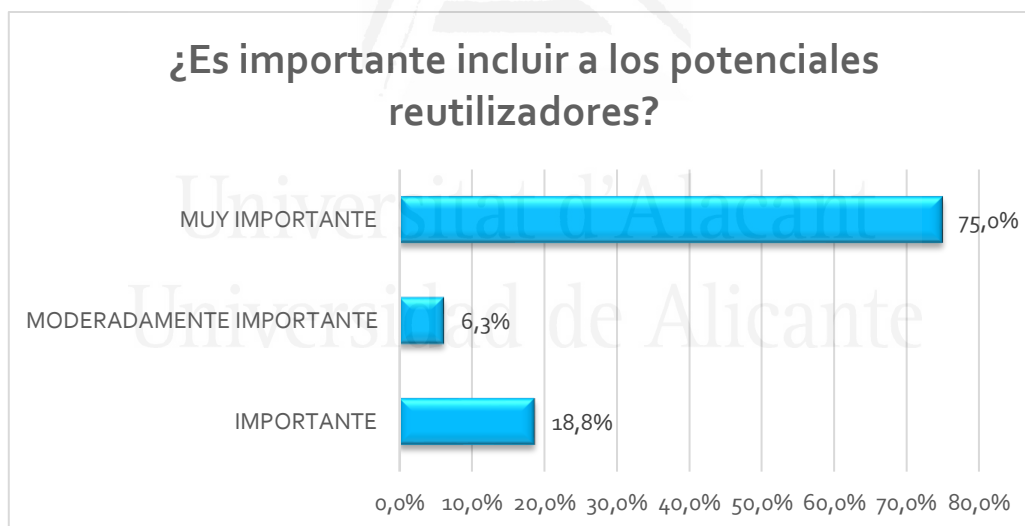


Figura 26: Percepción de los gestores de los portales acerca de incluir a los reutilizadores.

6.3.2 Considerar aspectos de calidad de datos y archivos de origen

El 32% de los entrevistados no consideraron la calidad de las fuentes de datos antes de la publicación de los datos abiertos. De acuerdo al análisis cualitativo se identificaron las varias razones. En primer lugar, se argumenta desconocimiento de la consideración de criterios de calidad en datos abiertos. Además, otro argumento es que se considera la publicación de datos abiertos como un resultado por lo que se asume que la calidad de datos se debe

considerar una vez obtenido los propios datos abiertos. Finalmente, otro argumento es que la publicación de los datos abiertos se hizo motivada por la legislación vigente y no se consideraron criterios de calidad de datos en absoluto.

Se realizó una pregunta cerrada relacionada con la percepción del entrevistado sobre la dependencia de la calidad de los datos publicados en abierto y los sistemas informáticos de origen. El 56,25% considera importante considerar la calidad desde los sistemas de origen. Los resultados se presentan en la Figura 27.

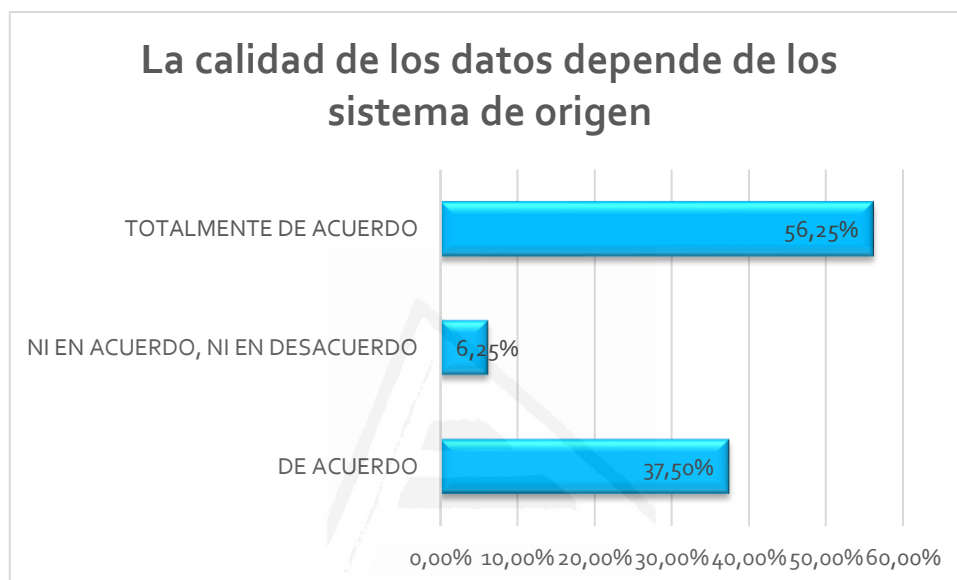


Figura 27: Percepción de los gestores de los portales acerca de la a calidad de los datos

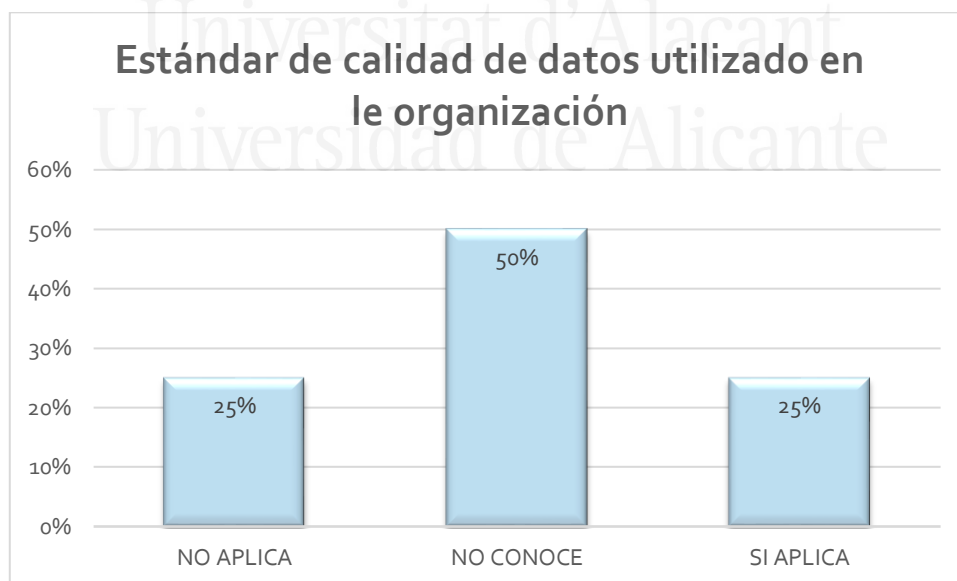


Figura 28: Percepción de los gestores de los portales acerca si la institución cuenta con estándar de calidad de datos

Como complemento a esta información y, de acuerdo con la Figura 28, los entrevistados indican que dentro de la institución no se conoce o no se aplican estándares de calidad de datos, con el 25% en los dos casos.

6.3.3 Concepto de comprensibilidad

Luego del análisis cualitativo, en relación con concepto comprensibilidad se pueden identificar los elementos graficados en la Figura 29, donde se indica que el 50% del total entiende que la comprensión debe estar enfocada a la información que se presenta y el 18,8% indica que la comprensibilidad debe ser entendible en cualquier segmento que se presente la información.

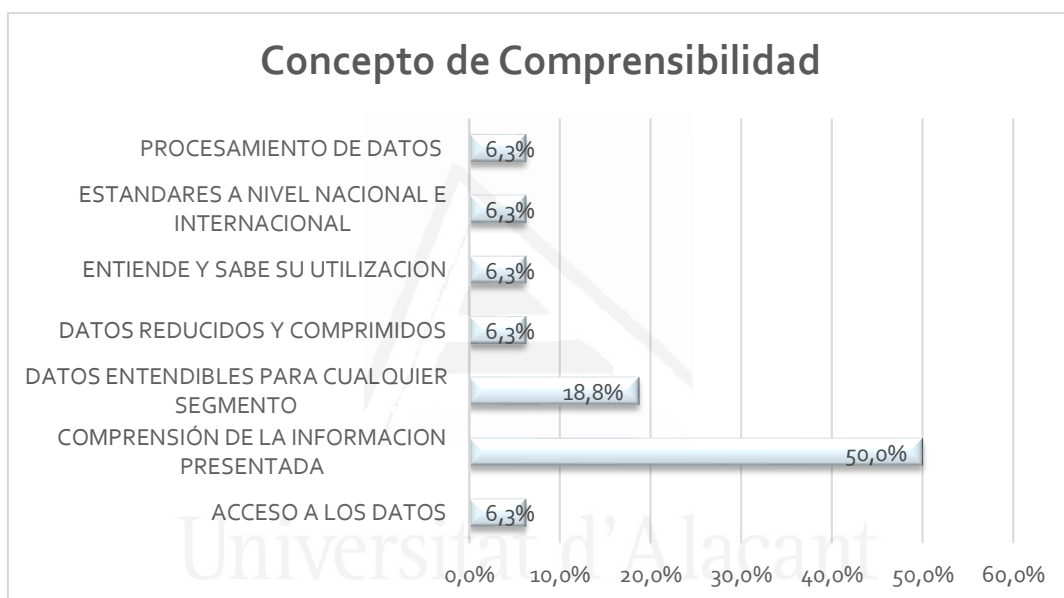


Figura 29: Concepto de comprensibilidad de los gestores de los portales.

De la misma manera se realizó una pregunta relacionada con la forma que se implementaban los procesos de comprensibilidad en la organización (en los sistemas de información de origen) identificando que se usaban los siguientes mecanismos: diccionarios y metadatos, talleres/información/socialización, así como la publicación de datos importantes con herramientas adecuadas.

6.4 Discusión de resultados de las entrevistas

En [89], se realiza un estudio sobre la necesidad de publicar datos abiertos con unos mínimos de calidad, aunque se publiquen menos conjuntos de datos abiertos en un primer momento. También, se establece que es importante considerar criterios de calidad de datos desde el

origen del dato en las organización. Esto se alinea con los resultados de las entrevistas, donde el 56,25% coincide que es importante considerar la calidad de los datos a publicar en abierto desde los sistemas de origen. Sin embargo, en los aportes cualitativos se indica que se desconoce los metadatos de los sistemas de origen y, por tanto, a la hora de publicar los datos en abierto no se consideran los metadatos de origen con el fin de mejorar el criterio de calidad de la comprensibilidad. Un aspecto importante es que el 75% no aplica un estándar de calidad de datos, el 50% no conoce el concepto. Además, no se ha considerado en el 63% de los portales implementados a los potenciales reutilizadores.

En los resultados del mapeo sistemático (presentado en el capítulo 2 de esta tesis doctoral) se identificó que la comprensibilidad se relaciona con la estructura de metadatos, esto es coincidente con las respuestas de los entrevistados que indican que la comprensibilidad de relaciona con los metadatos y el diccionario de datos. Sin embargo, existen otros conceptos que los entrevistados identificaron como característica de comprensibilidad que deben ser tenidos en cuenta, como son, la estandarización y el acceso a los datos.

Con estos resultados de las entrevistas, también se puede resaltar que las prácticas actuales de publicación de datos en abierto no consideran estándares de calidad de datos y tampoco consideran ni los sistemas de origen, ni a los potenciales reutilizadores. De hecho, e objetivo fundamental de los portales de datos abiertos fue cumplir la normativa y legislación referida a transparencia y a la publicación de datos en abierto.

No obstante, a esto como aspecto positivo, los entrevistados reconocen la importancia de incluir a los potenciales reutilizadores en el proyecto de datos abiertos, tener en cuenta estándares de calidad de datos que deben ser aplicados a los sistemas de origen.

6.5 Procesamiento de información de grupos de discusión

En esta sección se presentan los resultados de los tres grupos de discusión realizados, organizados de la siguiente manera:

- Grupo de discusión 1: instituciones que tienen a su cargo la prueba nacional ecuatoriana de admisión a la universidad (8 personas).
- Grupo de discusión 2: instituciones ecuatorianas de gobierno encargadas de definir la política de apertura de datos y la consolidación de los conjuntos de datos a publicarse en la República del Ecuador.

- Grupo de discusión 3: instituciones académicas encargadas de publicación de datos e investigadores que realizan trabajos de investigación en la temática de datos abiertos.

En cuanto a los resultados, en primer lugar, se presentan los derivados de grupos más alineados (grupos 1 y 2), mientras que los resultados del grupo focal 3 se presentan de manera independiente. Los resultados sirvieron para evolucionar el el marco de referencia diseñado con el fin de que la propuesta final de marco de referencia estuviera alineada con los expertos y fuera el resultado de un trabajo de investigación cooperativo siguiendo la metodología de “investigación en acción” anteriormente definida en el capítulo 3 de este documento. A continuación, se describen las preguntas directrices que guiaron los grupos de discusión.

6.5.1 ¿Es aplicable el marco de referencia a su institución?

En primer lugar, todos los expertos contestaron que es aplicable el marco de referencia en la institución que trabajan. Adicionalmente, se solicitó analizar las ventajas y la estructura del marco de referencia diseñado, las aportaciones fueron codificadas y cuantificadas logrando los resultados presentados en la Figura 30. En esta figura, se presentan 5 ejes principales (codificaciones encontradas) relacionados con el marco de referencia, como son los datos ordenados y la consolidación de datos que proporciona en una sola base de datos (dato maestro de apertura). Estos ejes son importantes porque algunos de los sistemas de información no tienen documentación adecuada. Otro elemento que favorece su aplicación es la inclusión de la máxima autoridad y el cambio cultural en la organización que permita conocer la aplicabilidad de los datos abiertos en las propias organizaciones.

Además, otro eje a considerar es la estructura ordenada que se tiene del marco de referencia, lo que resulta de importancia para la adaptación de normas de calidad y la orientación a usuarios finales. En este eje se considera de manera prioritaria la comprensibilidad de los conjuntos de datos publicados.

Por otro lado, una de las barreras que tiene el marco de referencia, según los expertos, es la falta de especificación de estándares de los metadatos en el país lo que podría complicar su uso en las instituciones sino se establece primero un plan de especificación y uso de estándares a nivel nacional.

Por lo tanto, se puede concluir que los grupos 1 y 2 consideran que el marco de referencia es totalmente aplicable a las organizaciones participantes, pero, como desventaja, se considera la falta de estándares de metadatos a nivel nacional. En este sentido, la

recomendación ha sido agregar al marco de referencia un estándar como el DCAT, incluyendo las características necesarias para responder a los indicadores de comprensibilidad planteados.

El grupo 3 recibió la explicación del marco de referencia diseñado, agregando los procesos de cambio cultural en el uso de datos, obteniendo los resultados que se detallan a continuación (y que se muestran en la figura 31).

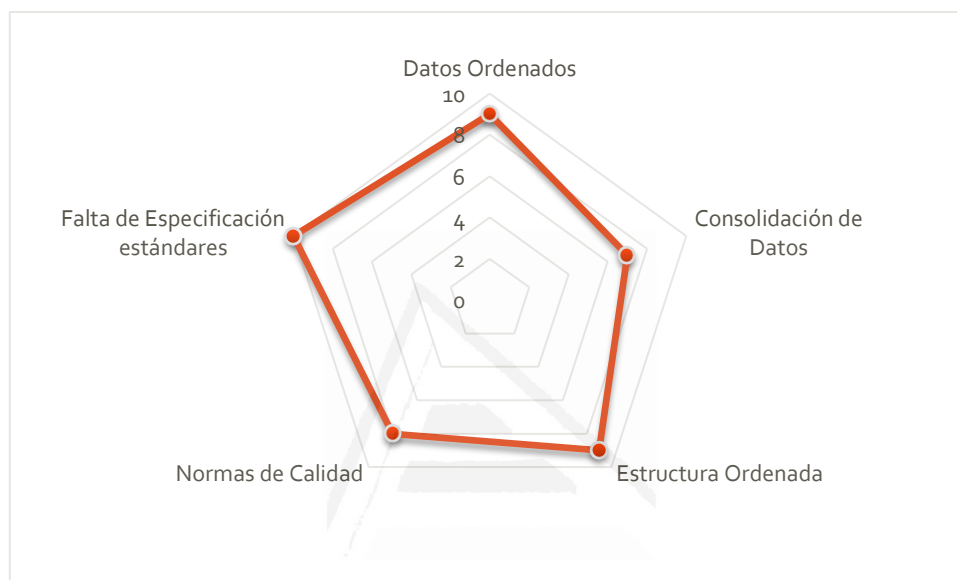


Figura 30: Resultados grupo focal 1 y 2

El grupo de discusión 3, considera que los ejes que son aplicables para el marco de referencia propuesto son los siguientes:

- Profundización en temas de estandarización. El grupo de discusión, considera que es una responsabilidad del gobierno por lo que no es una tarea fácil tener en cuenta este aspecto. No obstante, se deben incluir buenas prácticas o tener en cuenta estándares internacionales, con el fin de incluir datos de origen en los sistemas que no tienen documentación.
- Sistematización de actividades, que se considera una de las principales ventajas del marco de referencia. Esta sistematización, permite fomentar el conocimiento de aspectos relativos a la publicación de datos, favoreciendo la comprensión del mismo y permitiendo la consolidación de la información y la preparación para la publicación de los datos. La desventaja que considera este grupo de discusión acerca del marco

de referencia es acerca de la inclusión de estándares de metadatos. Se establece la posibilidad del uso de DCAT y Dublín Core.

- La inclusión para crear estrategias con el fin de trabajar con los reutilizadores mediante estándares reconocidos y que puedan aplicarse en las instituciones descentralizadas es considerada, por el grupo de discusión, un reto importante relacionado con el cambio cultural en las organizaciones.

En la Figura 32, se indica que el marco de referencia es aplicable totalmente, sin embargo, la participación ciudadana es un proceso que se debe trabajar mucho, empezando por el nivel técnico, y extendiéndose a la máxima autoridad en los órganos de dirección, con el fin de difundir el concepto de datos abiertos.

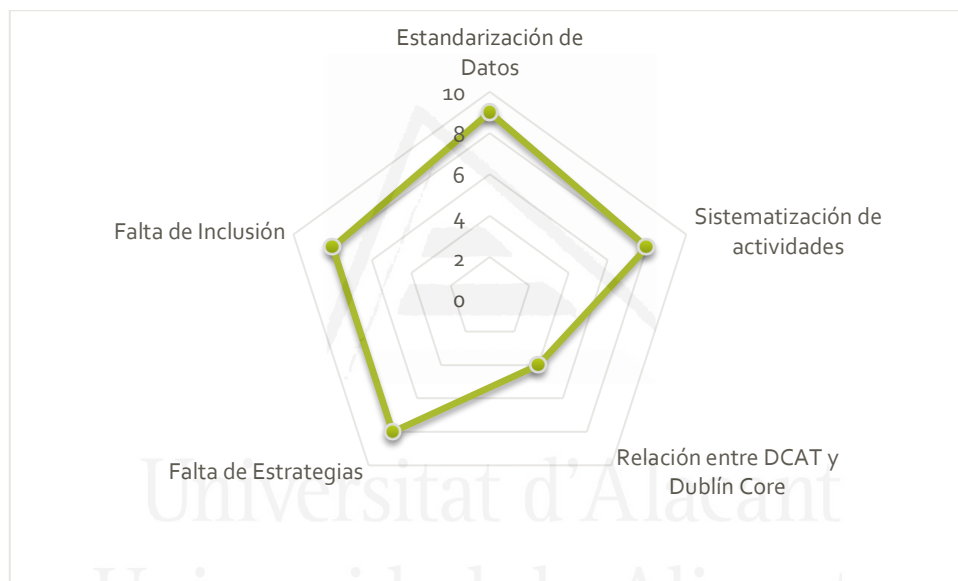


Figura 31: Resultados grupo focal 3

6.5.2 ¿El marco de referencia favorece la comprensibilidad por parte del reutilizador?

Esta pregunta es considerada de la misma manera por los grupos de discusión 1, 2 y 3, coincidiendo que la comprensibilidad de los datos es un concepto novedoso e importante en los datos abiertos y que el marco de referencia ayuda a favorecer la comprensibilidad de los datos abiertos.

Todos coinciden que la estructura del marco de referencia favorece la publicación con criterios de comprensibilidad, debido a su estructuración ordenada y sistematizada. Esto hace que se ayude a consolidar los procesos desde una perspectiva estratégica para la publicación

de los datos en abierto, facilitando la comprensión de los datos abiertos por parte del usuario final, estos se presentan al lado derecho de la Figura 33. En el lado izquierdo de la Figura 33 se presente los ejes que pueden resultar en barreras para la aplicabilidad del marco de referencia son el cambio cultural, la mala calidad de los datos en origen y la falta de aplicación de la norma ISO 25000. Por ello, se solicita establecer un plan de trabajo para la implementación a nivel país de una estandarización de metadatos y aplicación de la gestión de la calidad, no solo en los procesos de apertura, sino también en el origen del dato.

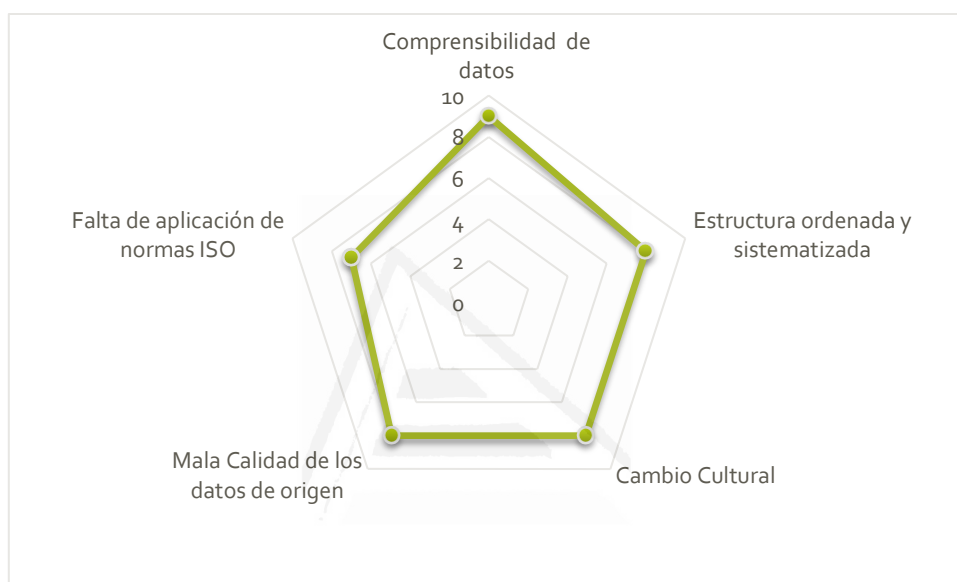


Figura 32: Validación de criterios de comprensibilidad

6.6 Aspectos importantes de la evaluación

Dentro del proceso de revisión de mapeo sistemático se evidenció la necesidad de considerar el origen de los datos para mejorar la calidad de los datos publicados en abierto, realizándose una ampliación y adaptación de los estándares ISO 8000 e ISO 25000 para el marco de referencia. Como resultado de la evaluación se identificó como ventajas la aplicación de los estándares y la estructura de gobierno de datos, considerando que los datos abiertos deben ser parte de la estrategia institucional.

En las tres iteraciones realizadas se dio importancia al cambio cultural en la organización de tal manera que se considere el uso e importancia de los datos abiertos. Este proceso fue incluido en el marco de referencia como parte de las actividades del responsable del proyecto de apertura de datos. Otro elemento importante fue la necesidad de que un proyecto de apertura de datos involucre a la máxima autoridad y se definan catálogos institucionales.

Todos los participantes en los grupos de discusión coinciden que es necesario el uso de un estándar de metadatos basado en DCAT y crear catálogos a nivel nacional.

Realizando un cruce de los datos obtenidos en la entrevista y los grupos de trabajo, se visualiza que el desconocimiento del concepto de datos abiertos y sus correspondientes procesos implicó la apertura de datos sin considerar criterios de calidad, estandarización y orientación al potencial reutilizador.



Universitat d'Alacant
Universidad de Alicante

7. CONCLUSIONES Y LÍNEAS DE INVESTIGACIÓN FUTURAS

En este capítulo se presentan las conclusiones del trabajo de investigación realizado y se describen las líneas de investigación futuras.



Universitat d'Alacant
Universidad de Alicante

Luego de la realización de actividades de recolección, clasificación y análisis de las investigaciones sobre datos abiertos desde una perspectiva tecnológica, enmarcadas en una metodología de mapeo sistemático, en donde se consideraron publicaciones científicas entre el 2006 y 2017 (ambas inclusive), se evidencia que las mismas provienen de los repositorios científicos del IEEE y la ACM. Dentro del análisis de desarrollo de los datos abiertos se visualiza que las publicaciones de 2006 a 2009 son incipientes (de hecho, no hay publicaciones sobre datos abiertos en 2007). Analizando el dominio en los cuales se trabaja se concluye que el más desarrollado es el relacionado con los infomediarios. Al analizar las fases, la explotación y la exploración de datos abiertos son las que producen resultados de investigación más frecuentes. De la misma manera, el tipo de investigación que más se ha realizado son las propuestas de solución y la validación de propuestas de investigación. Dentro de las temáticas abordadas en las investigaciones, se observa que la web semántica, ingeniería de software y gobierno son las que más publicaciones tienen. De hecho, existen solamente 28 artículos publicados en calidad de datos abiertos, representando el 4% de las investigaciones. Los problemas relacionados a la calidad son: falta de metadatos, datos incompletos, falta de información sobre actualizaciones de datos, que afectan directamente a la comprensión de los datos por parte del usuario. Las fases relacionadas con la publicación de datos están poco desarrolladas y, además, los temas de calidad de datos en origen no han sido abordados por lo cual el presente trabajo de investigación se centró en estas temáticas.

Se ha realizado un análisis de estándares (ISO 8000, ISO 25000) y buenas prácticas (gobierno de datos), ampliándolas y adaptándolas para construir un marco de referencia que permita la apertura de datos con criterios de calidad (en concreto, teniendo en cuenta la comprensibilidad de los datos), mediante la definición de dos artefactos novedosos que se aplican a las fuentes de datos: un archivo maestro de datos abiertos y una propuesta de metadatos. Estos artefactos se encuentran en el nivel operativo, considerando que el marco de referencia consta de tres niveles estratégico, táctico y operativo.

Se ha realizado una evaluación del artefacto diseñado, con gestores de portales abiertos a nivel de gobierno, academia y sociedad civil, demostrando la viabilidad de su uso, encontrando que el archivo maestro de datos abiertos, los metadatos y el diccionario de datos son elementos útiles para la apertura de datos con criterios de comprensibilidad. Además, se identificó que el cambio cultural en el uso y adopción de los datos abiertos es un elemento importante adicionando requisitos de adopción de estándares de metadatos y catálogos a nivel institucional y nacional.

7.1 Líneas de investigación futuras

Una vez que se ha logrado definir y evaluar el marco de referencia, se continuará en las siguientes líneas de investigación:

- Desarrollo de un modelo de sistema experto que permita a las instituciones evaluar las afectaciones en la reutilización en dependencia de la capacidad de aplicación de diferentes niveles de calidad de datos.
- Validación empírica de los indicadores de comprensibilidad con una comunidad de reutilizadores amplia, de tal manera que se puedan considerar como parte de un estándar internacional en los procesos de publicación de datos abiertos.



Universitat d'Alacant
Universidad de Alicante

8. CONTRIBUCIONES

En este capítulo se describen las contribuciones científicas y las actividades realizadas de transferencia de conocimiento relacionadas con la investigación llevada a cabo.



Universitat d'Alacant
Universidad de Alicante

8.1 Publicaciones científicas

A continuación, se enumeran las principales publicaciones realizadas durante el desarrollo de la investigación del presente trabajo:

- S. Cadena-Vela, A. Fuster-Guilló, J. N. Mazón, “Publicando datos abiertos con criterios de calidad”, 2019, Revista Ibérica de Sistemas y Tecnologías de la Información, RISTI indexada en SCOPUS (SJR Q3).
- S. Cadena-Vela, A. Fuster-Guilló, J. N. Mazón, Using a Master Data Management approach for increasing open data quality, 2019, Workshop on Methods, Evaluation, Tools and Applications towards a Data-driven e-Society (Meta4eS’19), publicado en Springer Lecture Notes in Computer Science (LNCS) in the OTM 2019 post-conference Workshop Proceedings, indexado por ISI, ACM Digital Library, DBLP and Google Scholar
- S. Cadena-Vela, R. Enríquez-Reyes, A. Fuster-Guilló, L.Ibáñez, J.N. Mazón, E. Simperl, “Systematic mapping of open data studies: classification and trends from a technological perspective”, (enviado para revision en febrero de 2019), Information Systems Frontiers , JCR Journal Impact Factor: 2,539 (JCR Q2).
- S. Cadena-Vela, A. Fuster-Guilló, J. N. Mazón, Framework for publishing data with understandability criteria, 2019 (en proceso de envío), Computer Standards & Interfaces, JCR Journal Impact Factor: 2,441
- S. Cadena-Vela, R Enríquez-Reyes. Portal de datos abiertos de la Universidad Central del Ecuador, 2017, FIGEMPA: Investigación Y Desarrollo, 1(2), 85-90. <https://doi.org/10.29166/revfig.v1i2.73>.
- S. Cadena-Vela, R. Enríquez-Reyes. Diseño e implementación de una universidad abierta: caso Universidad Central del Ecuador, 2017, FIGEMPA: Investigación Y Desarrollo, 1(2), 27 - 32. <https://doi.org/10.29166/revfig.v1i2.865>

8.2 Transferencia de conocimiento

Luego del proceso de evaluación realizado al marco de referencia, se ha solicitado por parte de varias instituciones gubernamentales ecuatorianas a la posibilidad de trabajar en la aplicación y ampliación del mismo en los siguientes aspectos:

- Gerencia de Gobierno Electrónico, del Ministerio de Telecomunicaciones y Sociedad de la Información (MINTEL). Uso del marco de referencia para el levantamiento e identificación de las prácticas de calidad de datos en origen en cada uno de los ministerios que publican datos en formatos abiertos.
- Gerencia de Gobierno Electrónico, del Ministerio de Telecomunicaciones y Sociedad de la Información (MINTEL). Definición de una guía técnica para publicación de datos en formatos abiertos como complemento al uso del marco de referencia
- Secretaría Nacional de Ciencias y Tecnología SENESCYT. Se solicita la coordinación de la comisión de formalización de estrategias de calidad para la publicación de datos en abierto dentro de la Agenda Nacional de Investigación Educativa ANIE. Ver Anexo 3, en sección 10.4.
- Ministerio de Telecomunicaciones y Sociedad de la información (MINTEL). Se solicita participación en la creación de la política de datos abiertos y en la legislación de protección de datos como parte de la academia y con aportes desde la investigación. Ver Anexo 5, en sección 10.5.

Universidad de Alicante

9. BIBLIOGRAFÍA

En este capítulo se encuentra la lista de referencias bibliográficas utilizadas en la tesis
doctoral

Universitat d'Alacant
Universidad de Alicante

- [1] Info@opendatacharter.net, “Open Data Charter,” 2016. [Online]. Available: <https://opendatacharter.net>.
- [2] L. F. Ramos Simón *et al.*, “De la reutilización de información del sector público a los portales de datos abiertos en Europa,” *BiD Textos Univ. Bibliotecon. i Doc.*, vol. 29, 2012.
- [3] D. Pirozzi and V. Scarano, “Support Citizens in Visualising Open Data,” *2016 20th Int. Conf. Inf. Vis.*, pp. 271–276, 2016.
- [4] I. Hrynaszkiewicz, S. Busch, and M. J. Cockerill, “Licensing the future: report on BioMed Central’s public consultation on open data in peer-reviewed journals,” *BMC Res. Notes*, vol. 6, no. 1, p. 318, 2013.
- [5] J. Attard, F. Orlandi, S. Scerri, and S. Auer, “A systematic review of open government data initiatives,” *Gov. Inf. Q.*, vol. 32, no. 4, pp. 399–418, 2015.
- [6] M. Garriga-Portolà, “¿Datos Abiertos? Sí, Pero de Forma Sostenible,” *El Prof. la Inf.*, vol. 20, no. 3, pp. 298–303, 2011.
- [7] P. Benson, “ISO 8000 the International Standard for Data Quality,” *MIT Inf. Qual. Symp.*, vol. July 16-17, no. 2008, pp. 1–4, 2008.
- [8] I. Rafique, P. Lew, M. Q. Abbasi, and Z. Li, “Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model,” *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 6, no. 5, pp. 568–573, 2012.
- [9] J. Kucera and D. Chlapek, “Benefits and Risks of Open Government Data,” *J. Syst. Integr.*, pp. 30–41, 2014.
- [10] J. Debattista, C. Lange, and S. Auer, “Representing Dataset Quality Metadata using Multi-Dimensional Views,” no. August, 2014.
- [11] P. Conradie and S. Choenni, “On the barriers for local government releasing open data,” *Gov. Inf. Q.*, vol. 31, no. SUPPL.1, pp. S10–S17, 2014.
- [12] A. Tygel, S. Auer, J. Debattista, F. Orlandi, and M. L. M. H. Campos, “Towards Cleaning-Up Open Data Portals: A Metadata Reconciliation Approach,” *Proc. - 2016 IEEE 10th Int. Conf. Semant. Comput. ICSC 2016*, pp. 71–78, 2016.
- [13] S. Moore, “How to Create a Business Case for Data Quality Improvement,” *Gartner*, 2017. [Online]. Available: <https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/>.

- [14] A. G. Carretero, F. Gualo, I. Caballero, and M. Piattini, “MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000,” *Comput. Stand. Interfaces*, vol. 54, no. September 2016, pp. 139–151, 2017.
- [15] V. Merchán, “Estrategia de evaluación de calidad de datos orientados a Web en Sistemas de Soporte a Negocios de Interconexión,” *Rev. Tecnológica Espol*, vol. 28, no. Agosto, pp. 110–123, 2015.
- [16] E. Oviedo, J. N. Mazon, and J. J. Zubcoff, “Quality and maturity model for open data portals,” *2015 Lat. Am. Comput. Conf.*, pp. 1–7, 2015.
- [17] S. Sadiq and M. Indulska, “Open data: Quality over quantity,” *Int. J. Inf. Manage.*, vol. 37, no. 3, pp. 150–154, 2017.
- [18] J. Umbrich, S. Neumaier, and A. Polleres, “Quality assessment and evolution of Open Data portals,” *Proc. - 2015 Int. Conf. Futur. Internet Things Cloud, FiCloud 2015 2015 Int. Conf. Open Big Data, OBD 2015*, pp. 404–411, 2015.
- [19] A. Zuiderwijk, M. Janssen, K. Poulis, and G. van de Kaa, “Open Data for Competitive Advantage: Insights from Open Data Use by Companies,” in *Proceedings of the 16th Annual International Conference on Digital Government Research*, 2015, pp. 79–88.
- [20] P. Colpaert, J. Sarah, P. Mechant, E. Mannens, and R. Van de Walle, “The 5 stars of open data portals,” *7th Int. Conf. Methodol. Technol. tools enabling e-Government*, 2013.
- [21] S. Mirri, C. Prandi, P. Salomoni, F. Callegati, and A. Campi, “On combining crowdsourcing, sensing and open data for an accessible smart city,” *Proc. - 2014 8th Int. Conf. Next Gener. Mob. Appl. Serv. Technol. NGMAST 2014*, pp. 294–299, 2014.
- [22] S. Acharya and H. W. Park, “Open data in Nepal: a webometric network analysis,” *Qual. Quant.*, pp. 1–17, 2016.
- [23] L. F. Luna-Reyes, J. C. Bertot, and S. Mellouli, “Open Government, Open Data and Digital Government,” *Gov. Inf. Q.*, vol. 31, no. 1, pp. 4–5, 2014.
- [24] G. Flouris, Y. Roussakis, M. Poveda-Villalón, P. N. Mendes, and I. Fundulaki, “Using provenance for quality assessment and repair in linked open data,” *CEUR Workshop Proc.*, vol. 890, pp. 1–12, 2012.
- [25] B. Kitchenham *et al.*, “Systematic literature reviews in software engineering—A tertiary study,” *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.
- [26] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic Mapping Studies in

-
- Software Engineering,” *12Th Int. Conf. Eval. Assess. Softw. Eng.*, vol. 17, p. 10, 2008.
- [27] S. Casteleyn, I. Garrigós, and J.-N. Mazón, “Ten years of Rich Internet applications: A systematic mapping study, and beyond,” *ACM Trans. Web*, vol. 8, no. 3, 2014.
- [28] P. A. Vélez and A. Rey Piedrahita, “Control de calidad en sistemas crowdsourcing: un mapeo sistemático,” *Sci. Tech.*, vol. 22, no. 1, p. 73, 2017.
- [29] J. Zubcoff, I. Garrigós, S. Casteleyn, J. N. Mazón, J. A. Aguilar, and F. Gomariz-Castillo, “Evaluating different i*-based approaches for selecting functional requirements while balancing and optimizing non-functional requirements: A controlled experiment,” *Inf. Softw. Technol.*, vol. 000, no. January 2017, pp. 1–17, 2018.
- [30] J. Attard, F. Orlandi, S. Scerri, and Soren Auer, “A systematic review of open government data initiatives,” *Gov. Inf. Q.*, vol. 32, no. 4, pp. 399–418, 2015.
- [31] S. Barney, K. Petersen, M. Svahnberg, A. Aurum, and H. Barney, “Software quality trade-offs: A systematic map,” *Inf. Softw. Technol.*, vol. 54, no. 7, pp. 651–662, 2012.
- [32] E. R. P. M. Engström, “A systematic review on regression test selection techniques,” *ACM Int. Conf. Proceeding Ser.*, vol. 53, no. 1, pp. 14–40, 2010.
- [33] R. L. Baskerville and Wood-Harper, “Diversity in information systems action research methods,” *Oper. Res. Soc. Ltd.*, pp. 90–2017, 1998.
- [34] R. Baskerville and M. D. Myers, “Special Issue on Action Research in Information Systems: Making is Research Relevant to Practice Foreword,” *Inf. Syst.*, vol. 32, no. 2, pp. 205–225, 2008.
- [35] R. L. Baskerville, “Investigating Information Systems with Action Research,” *Commun. Assoc. Inf. Syst.*, vol. 2, no. October, 2018.
- [36] Y. Wadsworth, “What is Participatory Action Research? The research framework,” *Action Res. Int.*, p. Paper 2, 1998.
- [37] <https://opendatacharter.net>, “G8 Open Data Charter,” *G8 Lough Erne 2013*, no. June, pp. 1–10, 2013.
- [38] J. Hagel III, “‘The Coming Battle for Customer Information’,” *Harvard Business Review*, 1997.
- [39] Open Data Charter, “International Open Data Charter,” no. September, p. 8, 2015.
- [40] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, “Requirements engineering paper

- classification and evaluation criteria: A proposal and a discussion,” *Requir. Eng.*, vol. 11, no. 1, pp. 102–107, 2006.
- [41] A. Latif, A. Scherp, and K. Tochtermann, “LOD for Library Science: Benefits of Applying Linked Open Data in the Digital Library Setting,” *KI - Künstliche Intelligenz*, vol. 30, no. 2, pp. 149–157, 2015.
- [42] A. Ramos-Soto, A. Bugarín, S. Barro, and F. Díaz-Hermida, “Automatic linguistic descriptions of meteorological data,” *Proc. Cist.*, pp. 1–6, 2013.
- [43] S. Chakraborty, M. H. H. Rahman, and M. H. Seddiqui, “Linked open data representation of historical heritage of Bangladesh,” *16th Int’l Conf. Comput. Inf. Technol. ICCIT 2013*, no. March, pp. 242–248, 2014.
- [44] E. Piedra, N.; Chicaiza, J.; Lopez, J.; Tovar Caro, “Towards a Learning Analytics Approach for Supporting discovery and reuse of OER,” no. March, pp. 978–988, 2015.
- [45] C. Millette and P. Hosein, “A consumer focused open data platform,” *2016 3rd MEC Int. Conf. Big Data Smart City, ICBDS 2016*, pp. 101–106, 2016.
- [46] a O. Erkimbaev, V. Y. Zitserman, G. a Kobzev, V. a Serebrjakov, and K. B. Teymurazov, “Publishing scientific data as linked open data,” *Sci. Tech. Inf. Process.*, vol. 40, no. 4, pp. 253–263, 2013.
- [47] A. Callahan, J. Cruz-Toledo, and M. Dumontier, “Ontology-Based Querying with Bio2RDF’s Linked Open Data,” *J. Biomed. Semantics*, vol. 4 Suppl 1, no. Suppl 1, p. S1, 2013.
- [48] S. O’Riain, E. Curry, and A. Harth, “XBRL and open data for global financial ecosystems: A linked data approach,” *Int. J. Account. Inf. Syst.*, vol. 13, no. 2, pp. 141–162, 2012.
- [49] D. S. Sayogo and T. A. Pardo, “Exploring the motive for data publication in open data initiative: Linking intention to action,” *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, no. 2011, pp. 2623–2632, 2011.
- [50] J. N. Rouder, “The what, why, and how of born-open data,” *Behav. Res. Methods*, vol. 48, no. 3, pp. 1062–1069, 2016.
- [51] P. Ciancarini, F. Poggi, and D. Russo, “Big Data Quality: A Roadmap for Open Data,” *2016 IEEE Second Int. Conf. Big Data Comput. Serv. Appl.*, pp. 210–215, 2016.

-
- [52] P. Doshi *et al.*, “Open data 5 years on: a case series of 12 freedom of information requests for regulatory data to the European Medicines Agency,” *Trials*, vol. 17, no. 1, p. 78, 2016.
- [53] S. Oyama, Y. Baba, I. Ohmukai, H. Dokoshi, and H. Kashima, “Crowdsourcing chart digitizer: task design and quality control for making legacy open data machine-readable,” *Int. J. Data Sci. Anal.*, 2016.
- [54] N. B. Hounsell, B. P. Shrestha, M. McDonald, and A. Wong, “Open Data and the Needs of Older People for Public Transport Information,” *Transp. Res. Procedia*, vol. 14, pp. 4334–4343, 2016.
- [55] A. Lausch, A. Schmidt, and L. Tischendorf, “Data mining and linked open data - New perspectives for data analysis in environmental research,” *Ecol. Modell.*, vol. 295, pp. 5–17, 2015.
- [56] H. Demski, S. Garde, and C. Hildebrand, “Open data models for smart health interconnected applications: the example of openEHR,” *BMC Med. Inform. Decis. Mak.*, vol. 16, no. 1, p. 137, 2016.
- [57] M. Kassen, “A promising phenomenon of open data: A case study of the Chicago open data project,” *Gov. Inf. Q.*, vol. 30, no. 4, pp. 508–513, 2013.
- [58] W. Brunette *et al.*, “Open data kit sensors: a sensor integration framework for android at the application-level,” *Proc. 10th Int. Conf. Mob. Syst. Appl. Serv. - MobiSys '12*, p. 351, 2012.
- [59] T. Silva, V. Wuwongse, and H. N. Sharma, “Disaster mitigation and preparedness using linked open data,” *J. Ambient Intell. Humaniz. Comput.*, vol. 4, no. 5, pp. 591–602, 2013.
- [60] F. G. De Andrade and R. José, “Semantic Annotation of Geodata Based on Linked-Open Data,” vol. 2, pp. 9–16, 2015.
- [61] Y.-A. Lai, Y.-Z. Ou, J. Su, S.-H. Tsai, C.-W. Yu, and D. Cheng, “Virtual disaster management information repository and applications based on linked open data,” *2012 Fifth IEEE Int. Conf. Serv. Comput. Appl.*, pp. 1–5, 2012.
- [62] N. Kobayashi and T. Toyoda, “BioSPARQL: Ontology-based smart building of SPARQL queries for biological Linked Open Data,” *ACM Int. Conf. Proceeding Ser.*, no. 1, pp. 47–49, 2012.

- [63] X. Masip-Bruin, G.-J. Ren, R. Serral-Gracia, and M. Yannuzzi, “Unlocking the Value of Open Data with a Process-Based Information Platform,” *IEEE 15th Conf. Bus. Informatics*, 2013.
- [64] B. Isaac, Antoine | Haslhofer, “Europeana Linked Open Data – data.europeana.eu,” *Semant. Web*, vol. 4, no. 3, pp. 291–297, 2013.
- [65] Europeandataportal.eu, “Portal Europeo de Datos,” 2019. [Online]. Available: <https://www.europeandataportal.eu/es/homepage>.
- [66] E. Rozell, J. Erickson, and J. Hendler, “From international open government dataset search to discovery: a semantic web service approach,” *ICEGOV '12 Proc. 6th Int. Conf. Theory Pract. Electron. Gov.*, pp. 480–481, 2012.
- [67] D. Maier, V. M. Megler, and K. Tufte, “Challenges for Dataset Search,” in *Database Systems for Advanced Applications*, 2014, pp. 1–15.
- [68] L. M. Koesten, E. Kacprzak, J. F. A. Tennison, and E. Simperl, “The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1277–1289.
- [69] E. Kacprzak, L. Koesten, L. D. Ibáñez, T. Blount, J. Tennison, and E. Simperl, “Characterising dataset search—An analysis of search logs and data requests,” *J. Web Semant.*, no. xxxx, 2018.
- [70] M. de Rijke, “Learning to Search for Datasets,” in *Companion Proceedings of the The Web Conference 2018*, 2018, p. 1483.
- [71] L. Mlynárová, J. P. Nap, and T. Bisseling, “The SWI/SNF chromatin-remodeling gene AtCHR12 mediates temporary growth arrest in Arabidopsis thaliana upon perceiving environmental stress,” *Plant J.*, vol. 51, no. 5, pp. 874–885, 2007.
- [72] T. Jetzek, M. Avital, and N. Bjorn-Andersen, “Data-driven innovation through open government data,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 9, no. 2, pp. 100–120, 2014.
- [73] E. Lakomaa and J. Kallberg, “Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs,” *IEEE Access*, vol. 1, pp. 558–563, 2013.
- [74] G. Magalhaes, C. Roseira, and L. Manley, “Business Models for Open Government

-
- Data,” in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 2014, pp. 365–370.
- [75] IDC, “Impact Assessment of Oline Programme,” 2017.
- [76] A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, “Open data quality measurement framework: Definition and application to Open Government Data,” *Gov. Inf. Q.*, vol. 33, no. 2, pp. 325–337, 2016.
- [77] D. Rousidis, E. Garoufallou, P. Balatsoukas, and M. A. Sicilia, “Metadata for Big Data: A preliminary investigation of metadata quality issues in research data repositories,” *Inf. Serv. Use*, vol. 34, no. 3–4, pp. 279–286, 2014.
- [78] M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, F. Flöck, and J. Lehmann, “Detecting Linked Data quality issues via crowdsourcing: A DBpedia study,” *Semant. Web*, vol. 0, pp. 1–33, 2016.
- [79] S. Neumaier, J. Umbrich, and A. Polleres, “Automated quality assessment of metadata across open data portals,” *J. Data Inf. Qual.*, vol. 8, no. 1, 2016.
- [80] S. Auer *et al.*, “Diachronic linked data,” *Proc. First Int. Work. Open Data - WOD '12*, p. 31, 2012.
- [81] S. Auer *et al.*, “Diachronic linked data: Towards long-term preservation of structured interrelated information,” *ACM Int. Conf. Proceeding Ser.*, no. October 2014, pp. 31–39, 2012.
- [82] J.-C. Kalo, S. Homoceanu, J. Rose, and W.-T. Balke, “Avoiding Chinese Whispers: Controlling End-to-End Join Quality in Linked Open Data Stores,” *WebSci '15 Proceedings 2015 ACM Conf. Web Sci.*, 2015.
- [83] R. Karam and M. Melchiori, “A crowdsourcing-based framework for improving geo-spatial open data,” *Proc. - 2013 IEEE Int. Conf. Syst. Man, Cybern. SMC 2013*, pp. 468–473, 2013.
- [84] I. Schieferdecker, “(Open) Data Quality,” *2012 IEEE 36th Annu. Comput. Softw. Appl. Conf.*, pp. 83–84, 2012.
- [85] I. Schieferdecker, “(Open) data quality,” *Proc. - Int. Comput. Softw. Appl. Conf.*, pp. 83–84, 2012.
- [86] G. Cheng, C. Jin, W. Ding, D. Xu, and Y. Qu, “Generating illustrative snippets for open data on the web,” *WSDM 2017 - Proc. 10th ACM Int. Conf. Web Search Data Min.*,

- pp. 151–159, 2017.
- [87] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon, “Comparison of metadata quality in open data portals using the Analytic Hierarchy Process,” *Gov. Inf. Q.*, vol. 35, no. 1, pp. 13–29, 2018.
- [88] J. Mitlohner, S. Neumaier, J. Umbrich, and A. Polleres, “Characteristics of open data CSV files,” *Proc. - 2016 2nd Int. Conf. Open Big Data, OBD 2016*, vol. 838, pp. 72–79, 2016.
- [89] “Open data: Quality over quantity,” *International Journal of Information Management*, vol. 37, no. 3, pp. 150–154, 2017.
- [90] M. Stróżyńska, G. Eiden, W. Abramowicz, D. Filipiak, J. Małyszko, and K. Węcel, “A framework for the quality-based selection and retrieval of open data - a use case from the maritime domain,” *Electron. Mark.*, vol. 28, no. 2, pp. 219–233, 2018.
- [91] P. Hitzler *et al.*, “Quality Assessment for Linked Data: A Survey A Systematic Literature Review and Conceptual Framework,” *Semant. Web*, vol. 1, pp. 1–5, 2012.
- [92] S. Kubler, J. Robert, Y. Le Traon, J. Umbrich, and S. Neumaier, “Open Data Portal Quality Comparison using AHP,” *Proc. 17th Int. Digit. Gov. Res. Conf. Digit. Gov. Res.*, pp. 397–407, 2016.
- [93] H. Paulheim and J. Fürnkranz, “Unsupervised generation of data mining features from linked open data,” *ACM Int. Conf. Proceeding Ser.*, 2012.
- [94] S. Sadiq and M. Indulska, “Open data: Quality over quantity,” *Int. J. Inf. Manage.*, vol. 37, no. 3, pp. 150–154, Jun. 2017.
- [95] P. Ciancarini, F. Poggi, and D. Russo, “Big Data Quality: A Roadmap for Open Data,” *Proc. - 2016 IEEE 2nd Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2016*, pp. 210–215, 2016.
- [96] D. Pirozzi and V. Scarano, “Support citizens in visualising open data,” *Proc. Int. Conf. Inf. Vis.*, vol. 2016-Augus, pp. 271–276, 2016.
- [97] A. Immonen, E. Ovaska, and T. Paaso, “Towards certified open data in digital service ecosystems,” *Softw. Qual. J.*, vol. 26, no. 4, pp. 1257–1297, 2018.
- [98] C. A. Locoro, and C. Batini, “A User Study to Assess the Situated Social Value of Open Data in Healthcare,” *Procedia Comput. Sci.*, 2015.
- [99] O. Barack, “Transparency and Open Government,” *21-01-2009*, 2009. [Online].

-
- Available: <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>.
- [100] B. Carlo, B. Daniele, C. Federico, and G. Simone, “A Data Quality Methodology for Heterogeneous Data,” *Int. J. Database Manag. Syst.*, vol. 3, no. 1, pp. 60–79, 2011.
- [101] D. Loshin and D. Loshin, “Business Impacts of Poor Data Quality,” *Pract. Guid. to Data Qual. Improv.*, pp. 1–16, Jan. 2011.
- [102] D. Loshin, “Praise for Master Data Management,” p. 302, 2009.
- [103] M. Hubert Ofner, K. Straub, B. Otto, and H. Oesterle, “Management of the master data lifecycle: a framework for analysis,” *J. Enterp. Inf. Manag.*, vol. 26, no. 4, pp. 472–491, 2013.
- [104] F. Benitez, A. Comber, and J. Huerta, “IMPROVE THE REUSABILITY OF OPEN How much data is generated every minute?”
- [105] I. Caballero *et al.*, “SLA4DQ-I8K: Acuerdos a Nivel de Servicio para Calidad de Datos en Intercambios de Datos Maestros regulados por ISO 8000-1x0,” *X Jornadas Ciencias e Ing. Serv. (JCIS 2014)*, pp. 157–166, 2014.
- [106] D. Quality, “Best Practice: Enable quality assessment of open data,” no. July, 2016.
- [107] D. W. Coleman, A. A. Hughes, and W. D. Perry, “The role of data governance to relieve information sharing impairments in the federal government,” *2009 WRI World Congr. Comput. Sci. Inf. Eng. CSIE 2009*, vol. 4, pp. 267–271, 2009.
- [108] and F. B. Juan Ribeiro Reis, José Viterbo, “A rationale for data governance as an approach to tackle recurrent drawbacks in open data portals,” *Proc. 19th Annu. Int. Conf. Digit. Gov. Res. Gov.*, [1] F. B. Juan Ribeiro Reis, José Viterbo, “A Ration. data Gov. as an approach to Tack. Recurr. Draw. open data portals,” *Proc. 19th Ann*, 2018.
- [109] “Chapter-7-Identifying-Master-Metadata-and-Master-Data_2009_Master-Data-Management.” .
- [110] A. Reichert, B. Otto, and H. Österle, “A Reference Process Model for Master Data Management,” *Int. Conf. Wirtschaftsinformatik*, no. March, pp. 817–830, 2013.
- [111] H. Ye, D. Wu, and S. Chen, “An open data cleaning framework based on semantic rules for Continuous Auditing,” *ICCET 2010 - 2010 Int. Conf. Comput. Eng. Technol. Proc.*, vol. 2, pp. 158–162, 2010.

- [112] S. I. Organization, “ISO/TS 8000-120:2016,” *ISO/TS 8000-120*, 2016. [Online]. Available: <https://www.iso.org/standard/66234.html>. [Accessed: 10-Jan-2018].
- [113] S. I. Organization, “ISO/TS 8000-130-2016,” *Standar*, 2016. [Online]. Available: <https://www.iso.org/standard/62393.html>.
- [114] B. T. Hazen, F. K. Weigel, J. D. Ezell, and B. C. Boehmke, “Toward understanding outcomes associated with data quality improvement,” *Int. J. Prod. Econ.*, 2017.
- [115] B. Rivas, J. Merino, I. Caballero, M. Serrano, and M. Piattini, “Towards a service architecture for master data exchange based on ISO 8000 with support to process large datasets,” *Comput. Stand. Interfaces*, vol. 54, no. September 2016, pp. 94–104, 2016.
- [116] ISO, “International Standard Iso ISO/IEC 25024,” *2015*, vol. 2015, 2013.
- [117] S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon, “Comparison of metadata quality in open data portals using the Analytic Hierarchy Process,” *Gov. Inf. Q.*, vol. 35, no. 1, pp. 13–29, 2018.
- [118] R. J. NISO, *Understanding Metadata What Is Metadata*, ., 2014.
- [119] A. Zuiderwijk, K. Jeffery, and M. Janssen, “The Potential of Metadata for Linked Open Data and its Value for Users and Publishers,” *JeDEM - eJournal eDemocracy Open Gov.*, vol. 4, no. 2, pp. 222–244, 2012.
- [120] P. Conradie and S. Choenni, “Exploring process barriers to release public sector information in local government,” *ACM Int. Conf. Proceeding Ser.*, no. October, pp. 5–13, 2012.
- [121] J. Debattista, C. Lange, and S. Auer, “Representing dataset quality metadata using multi-dimensional views,” *ACM Int. Conf. Proceeding Ser.*, vol. 2014-Sept, no. September, pp. 92–99, 2014.
- [122] R. Cyganiak, F. Maali, and V. Peristeras, “Self-service linked government data with dcat and gridworks,” *ACM Int. Conf. Proceeding Ser.*, 2010.
- [123] E. Lapi, N. Tcholtchev, L. Bassbouss, F. Marienfeld, and I. Schieferdecker, “Identification and utilization of components for a linked open data platform,” *Proc. - Int. Comput. Softw. Appl. Conf.*, pp. 112–115, 2012.
- [124] A. W. P. Albertoni, Riccardo; Browning, David ; Cox, Simon;González, Alejandra; Perego, “Vocabulario del catálogo de datos (DCAT).” [Online]. Available: <https://www.w3.org/TR/vocab-dcat-2/#references>.

-
- [125] J. R. Reis, J. Viterbo, and F. Bernardini, “A Rationale for Data Governance As an Approach to Tackle Recurrent Drawbacks in Open Data Portals,” in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 2018, pp. 73:1--73:9.
- [126] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, “A systematic literature review of data governance and cloud data governance,” *Pers. Ubiquitous Comput.*, vol. c, pp. 1–21, 2018.
- [127] P. P. Tallon, “Corporate governance of big data: Perspectives on value, risk, and cost,” *Computer (Long. Beach. Calif.)*, vol. 46, no. 6, pp. 32–38, 2013.
- [128] R. L. Trope, E. M. Power, V. I. Polley, and B. C. Morley, “A coherent strategy for data security through data governance,” *IEEE Secur. Priv.*, vol. 5, no. 3, pp. 32–39, 2007.
- [129] D. Proença and J. Borbinha, “Maturity Models for Information Systems - A State of the Art,” *Procedia Comput. Sci.*, vol. 100, no. 2, pp. 1042–1049, 2016.
- [130] S. Soares, “Part I: Big Data Governance,” 2013.
- [131] Informatica, “Data Governance Framework Walkthrough,” no. February, 2014.
- [132] IBM, “Data Governance for Master Data Management,” *Data Manag.*, no. November, pp. 67–86, 2007.
- [133] E.-K. Lee, J.-A. Park, A. Cole, and J. Mestre-Ferrandiz, “Data Governance Arrangements for RWE in South Korea Data Governance Arrangements for Real-World Evidence: South Korea Consulting Report,” no. September, 2017.
- [134] J. J. Korhonen, I. Melleri, K. Hiekkanen, and M. Helenius, “Designing Data Governance Structure : An Organizational Perspective,” vol. 2, no. 4, pp. 11–17, 2013.
- [135] V. Khatri and C. V Brown, “Designing Data Governance,” *Commun. ACM*, vol. 53, no. 1, pp. 148–152, 2010.
- [136] D. Loshin and D. Loshin, “Data Governance,” *Pract. Guid. to Data Qual. Improv.*, pp. 115–128, Jan. 2011.
- [137] Y. Lee, “ESD Working Paper Series A Cubic Framework for the Chief Data Officer : Succeeding in a World of Big Data ESD-WP-2014-34,” no. November, 2014.
- [138] A. M. Barata and E. P. V. Prado, “Data Governance in Brazilian Organizations,” *XI*

- Brazilian Symp. Inf. Syst.*, no. Sbsi, pp. 267–274, 2015.
- [139] A. E. Prieto, J.-N. Mazon, and A. Lozano-Tello, “Framework for Prioritization of Open Data Publication: An Application to Smart Cities,” *IEEE Trans. Emerg. Top. Comput.*, vol. 6750, no. c, pp. 1–1, 2019.
- [140] M. Kassen, “A promising phenomenon of open data: A case study of the Chicago open data project,” *Gov. Inf. Q.*, vol. 30, no. 4, pp. 508–513, 2013.
- [141] T. Berners-Lee, “Linked Data,” *W3C*, 2009. [Online]. Available: <https://www.w3.org/DesignIssues/LinkedData.html>. [Accessed: 10-Feb-2017].
- [142] J. C. Nwokeji, T. Clark, B. Barn, V. Kulkarni, and S. O. Anum, “A Data-centric Approach to Change Management,” *Proc. - IEEE Int. Enterp. Distrib. Object Comput. Work. EDOCW*, vol. 2015-Novem, pp. 185–190, 2015.
- [143] C. Rica, E. Oviedo, and J. J. Zubcoff, “Modelo de calidad y madurez para portales de datos abiertos.”
- [144] D. Marco, “Understanding data governance and stewardship, Part 1,” *DM Rev.*, vol. 16, no. 9, p. 28, 2006.
- [145] P. Missier, K. Belhajjame, and J. Cheney, “The W3C PROV Family of Specifications for Modelling Provenance Metadata,” in *Proceedings of the 16th International Conference on Extending Database Technology*, 2013, pp. 773–776.
- [146] R. Devarakonda, G. Palanisamy, J. M. Green, and B. E. Wilson, “Data sharing and retrieval using OAI-PMH,” *Earth Sci. Informatics*, vol. 4, no. 1, pp. 1–5, 2011.
- [147] M. Szomszor and L. Moreau, “Recording and Reasoning over Data Provenance in Web and Grid Services,” pp. 603–604, 2003.
- [148] network-support@creativecommons.org, “CC Global Network Community Site,” 2018. .
- [149] N. Fabbe-Costes, C. Roussat, and J. Colin, “Future sustainable supply chains: What should companies scan?,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 41, no. 3, pp. 228–252, 2011.

10. ANEXOS

En este capítulo se incorporan los anexos correspondientes a: ejemplos de aplicación del marco de referencia y la digitalización de las entrevistas y grupos de discusión.

Universitat d'Alacant
Universidad de Alicante

10.1 Anexo 1: Ejemplo de un catálogo de conjuntos de datos a publicarse.

Código	1	2
Descripción	Archivo que contiene las obras adjudicadas a nivel nacional con adjudicatario	Archivo que publica las paradas de autobuses con geoposicionamiento
Organizaciones que generan la información de origen	Dirección Nacional de Compras Públicas	Municipio de Quito
Fecha de publicación	10/01/2020	10/01/2020
Estado actual (iniciado, en construcción, previo a publicación)	Iniciado	Construcción
Descripción de la licencia para publicación	Creative Commons Reconocimiento 3.0	Creative Commons Reconocimiento 3.0
Organización solicitante	Organización Anticorrupción	Asociación de Transportistas

10.2 Anexo 2: Transcripción de entrevistas realizadas

Entrevista N: 1

Nombre del entrevistado: Giovanni García – Escuela Superior Politécnica Agropecuaria de Manabí, **Correo Electrónico:** tecnologia@espam.edu.ec, **Fechas de entrevista:** 2019-06-18 13:50:23, **En qué etapa esta su portal:** Implementado, **Número de conjunto de datos publicados:** 2 (está en proceso de preparación).

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Sí.

¿Consideró la calidad de los datos en sus procesos de publicación? Sí

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos?, No, no se ha visto la necesidad

¿Su organización sigue algún modelo de Gobierno de Datos?, No, no se conocía

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?, No conoce

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?, Estándares usados a nivel nacional e internacional, pienso que son importantes pero no lo hemos aplicado

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Ninguno, no se ha visto la necesidad de hacerlo

¿Qué características tienen estos mecanismos?, Ni idea

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Ninguna

Entrevista N: 2

Nombre del entrevistado: Roberto Moya- Instituto Nacional de Evaluación, **Correo Electrónico:** roberto.moya@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-06, 13:22:34, **¿En qué etapa esta su portal?** Publicado, **Número de conjunto de datos publicados:** 10.

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?, No, por qué?, Por desconocimiento de normativa y estándares.

¿Consideró la calidad de los datos en sus procesos de publicación?, No, por qué?, Por desconocimiento de normativa y estándares.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos?, No

¿Su organización sigue algún modelo de Gobierno de Datos?, No

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?, No

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?

Si los usuarios pueden comprender la información presentada

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?

Diccionarios y Metadatos, se ha trabajado en la definición de estos mecanismos en todos los conjuntos de datos que se han publicado.

¿Qué características tienen estos mecanismos?

Se ha tratado de trabajar en todos los conjuntos de datos los formatos abiertos.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?

Aplicar estándares que se conocen.

Entrevista N: 3

Nombre del entrevistado: Cristian Alvarez- Instituto Ecuatoriano de Evaluación, **Correo Electrónico:** cristian.alvarez@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-06 14:29:54, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 3

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? No, por qué?, No se evidenció la necesidad en su momento.

¿Consideró la calidad de los datos en sus procesos de publicación?, Si, Se realizó una validación y depuración de la información previo a la publicación de los datos.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? De acuerdo

¿Su organización sigue algún estándar de calidad de datos? No, no se conocía

¿Su organización sigue algún modelo de Gobierno de Datos? No, la prioridad era publicar los datos.

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? No conoce

¿Qué entiende por comprensibilidad de los datos en formatos abiertos? Los datos deben ser entendibles por cualquier segmento de la población.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Diccionario de datos.

¿Qué características tienen estos mecanismos? Características en abierto, metadatos.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Talleres con investigadores.

Entrevista N: 4

Nombre del entrevistado: Patricio Toledo- Instituto Nacional de Evaluación, **Correo Electrónico:** patricio.toledo@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-06 14:48:14, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 10

¿Incluyó a los potenciales reutilizadores en la implementación de su portal ?, No, por qué? ya que se solicitó que sea para público en general como una acción inmediata.

¿Consideró la calidad de los datos en sus procesos de publicación ?, Si claro, por que si no eso lleva a estar en continua modificación.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos?, No porque Desconocíamos

¿Su organización sigue algún modelo de Gobierno de Datos?, No

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?, No

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?, Si se entiende la información que está publicado y para que la información se puede utilizar.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Diccionarios y metadatos

¿Qué características tienen estos mecanismos?, Se encuentran junto a los archivos para que puedan identificar los campos presentados.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Talleres

Entrevista N: 5

Nombre del entrevistado: Alejandra Espinosa-Instituto Nacional de Evaluación, Correo Electrónico: alejandra.espinosa@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-06 14:43:09, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 50

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?, No, por qué? se publicó con ese fin. Sino porque era una ley.

¿Consideró la calidad de los datos en sus procesos de publicación?, Si

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, De acuerdo

¿Su organización sigue algún estándar de calidad de datos?, No

¿Su organización sigue algún modelo de Gobierno de Datos?, No

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?,No

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?, Los usuarios puedan entender los datos cargados

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Talleres de validación de informes

¿Qué características tienen estos mecanismos?, Oportunidad de opinión sin barreras

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Socialización de los portales

Entrevista N: 6

Nombre del entrevistado: Michelle Lopez-Instituto Nacional de Evaluación, **Correo Electrónico:** michelle.lopez@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-06 15:50:28, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 2

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?, No, por qué? El cambio de personal y autoridades, no estaba diseñado para temas de usabilidad.

¿Consideró la calidad de los datos en sus procesos de publicación?, No, por qué? estaba diseñado solo para mostrar los resultados, no para usabilidad.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, De acuerdo

¿Su organización sigue algún estándar de calidad de datos?, Si, Potenciales usuarios de información, Leyes para publicación de datos, Geoportal, Portal de Resultados de la Evaluación Educativa

¿Su organización sigue algún modelo de Gobierno de Datos?, no

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?, Si

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?, Que todas las personas pueden acceder a los datos

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Socializar la información publicada

¿Qué características tienen estos mecanismos?, Talleres de participación

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Talleres con grupos focales destinados a tipos de usuarios

Entrevista N: 7

Nombre del entrevistado: Darwin Padilla, Intituto Nacional de Evaluaciòn, **Correo Electrónico:** darwin.padilla@evaluacion.gob.ec, **Fechas de entrevista:** 2019-06-07 08:32:04, **En qué etapa esta su portal:** Implementado, **Número de conjunto de datos publicados:** 1

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?, No, Principalmente por desconocimiento.

¿Consideró la calidad de los datos en sus procesos de publicación?, No, por Desconocimiento

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, De acuerdo

¿Su organización sigue algún estándar de calidad de datos?, No

¿Su organización sigue algún modelo de Gobierno de Datos?, No

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?, No conoce

¿Qué entiende por compresibilidad de los datos en formatos abiertos?, Que tengan una estructura fácil de entender y no haya suposiciones de los atributos que cada una de las variables presente.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Ninguno

¿Qué características tienen estos mecanismos?, N/A

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Reuniones y revisión bibliográfica

Entrevista N: 8

Nombre del entrevistado: Elena Chicaiza, **Correo Electrónico:** elena.chicaiza@evaluacion.gob.ec. **Fechas de entrevista:** 2019-06-07 09:43:08. **En qué etapa está su portal:** Publicado, **Número de conjunto de datos publicados:** 23

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?

No, por cuanto desconozco, no trabaja en ese momento en la institución y no existen documentos que prueben que han sido incluidos.

¿Consideró la calidad de los datos en sus procesos de publicación?, No, No existe documentación que los pruebe.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, Ni en acuerdo, Ni en desacuerdo

¿Su organización sigue algún estándar de calidad de datos?, No, No existen registros de si existió un modelo

¿Su organización sigue algún modelo de Gobierno de Datos?, No

¿Aplica el concepto de datos maestros dentro de su institución (MDM)?, No conoce

¿Qué entiende por comprensibilidad de los datos en formatos abiertos?, Facilidad para comprender los datos dispuestos en los portales

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Estamos trabajando en la usabilidad del portal y los metadatos.

¿Qué características tienen estos mecanismos?, Comprender a través de métricas cómo los perfiles de usuarios acceden y usan los datos que están dispuestos en el portal.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Talleres de usabilidad.

Entrevista N: 9

Nombre del entrevistado: Juan Castillo-Ministerio de Telecomunicaciones y Sociedad de la Información, **Correo Electrónico:** juan.castillo@mintel.gob.ec, **Fechas de entrevista:** 2019-06-07 10:43:45, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 128

¿Incluyó a los potenciales reutilizadores en la implementación de su portal?

No, por que se inicio como un piloto con datos que pueden ser liberados de acuerdo a la Ley de transparencia y acceso a la información, la nueva versión incluye un proceso participativo para publicar datos en función de la demanda

¿Consideró la calidad de los datos en sus procesos de publicación?, Si, Son datos oficiales apalancados en una Ley

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio?, Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen?, De acuerdo

¿Su organización sigue algún estándar de calidad de datos?, Si, mediante la guía política de datos abiertos, se establecen los criterios y formatos en los cuales los datos deben ser publicados

¿Su organización sigue algún modelo de Gobierno de Datos?, Si

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ?,No

¿Qué entiende por compresibilidad de los datos en formatos abiertos?, que los datos pueden ser reducidos o comprimidos

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados?, Ninguno

¿Qué características tienen estos mecanismos?, No aplica

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Emisión de Políticas de datos abiertos, Guía de datos abiertos, donde se plasma un proceso participativo con la ciudadanía, donde basado en una necesidad se priorizaría la publicación de datos abierto.

Entrevista N: 10

Nombre del entrevistado: Tania Gualli, **Correo Electrónico:** tania.gualli@epn.edu.ec

Fechas de entrevista: 2019-06-16 21:38:22, **En qué etapa esta su portal:** Aprobado,

Número de conjunto de datos publicados: 0

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Si,

porque se busca que los datos que se proporcionen aporten valor y sean de utilidad.

¿Consideró la calidad de los datos en sus procesos de publicación? Si, porque si los datos no cumplen con un mínimo de calidad no serán de utilidad para el potencial usuario.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Muy importante.

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente en desacuerdo.

¿Su organización sigue algún estándar de calidad de datos? No, aún no existe un estándar totalmente establecido, pero al menos se ha realizado una evaluación inicial para identificar en que dimensiones se debe trabajar para mejorar la calidad de los datos.

¿Su organización sigue algún modelo de Gobierno de Datos? No

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? No

¿Qué entiende por comprensibilidad de los datos en formatos abiertos? Que los datos que aperturen, se puedan analizar e interpretar, por cualquier usuario, no solamente por expertos en análisis de datos

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? La iniciativa de datos abiertos aún está en proceso, por diversos cambios institucionales, sin embargo se ha procurado adoptar herramientas como gráficos, dashboards de información, etc., que permitan una fácil visualización e interpretación.

¿Qué características tienen estos mecanismos? Estos mecanismos deben fomentar la reutilización de los datos

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Identificar las necesidades de información más recurrentes así como aquella información que más aporta a la toma de decisiones en los diferentes niveles organizacionales.

Entrevista N: 11

Nombre del entrevistado: María José Andrade – Universidad Central del Ecuador, **Correo Electrónico:** mjandrade@uce.edu.ec, **Fechas de entrevista:** 2019-06-18 07:39:10, **En qué etapa esta su portal:** Implementado, **Número de conjunto de datos publicados:** 0

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Si

¿Consideró la calidad de los datos en sus procesos de publicación? Si

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos? No

¿Su organización sigue algún modelo de Gobierno de Datos? No

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? Si

¿Qué entiende por comprensibilidad de los datos en formatos abiertos? Que los datos deben ser entendibles a la vista de cualquier usuario.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Mostrar los datos más importantes que se generan en la Institución.

¿Qué características tienen estos mecanismos? Creación de vistas dentro de la Base de Datos, de tal manera que la información que se refleja se actualice en tiempo real.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Brindar un mayor número de opciones de datos para toma de decisiones.

Entrevista N: 12

Nombre del entrevistado: uan Carlos Garcia – Universidad Ténica del Norte, **Correo Electrónico:** jcgarcia@utn.edu.ec, **Fechas de entrevista:** 2019-06-18 14:59:35, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 18

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Si

¿Consideró la calidad de los datos en sus procesos de publicación? Si

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos? Si, un modelo a la medida

¿Su organización sigue algún modelo de Gobierno de Datos? Si

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? Si

¿Qué entiende por compresibilidad de los datos en formatos abiertos? Aplicar a conjuntos grandes de datos un protocolo de compresión para mejorar la rapidez en el acceso

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Tabularlos de manera organizada, por facultades, carreras, genero, etnia, etc

¿Qué características tienen estos mecanismos? Mejoran la comprensibilidad porque nos ayudan a clasificar la información en base a indicadores de gestión.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?, Mejorar la accesibilidad a los datos y la eficiencia en la consulta de los mismos

Entrevista N: 13

Nombre del entrevistado: Victor Hugo Saquicela Galarza Universidad de Cuenca, **Correo**

Electrónico: victor.saquicela@ucuenca.edu.ec, **Fechas de entrevista:** 2019-06-18 17:30:23.

En qué etapa esta su portal: Implementado, **Número de conjunto de datos publicados:**

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Si, porque la participación de las diferentes unidades es crucial puesto que permite definir el uso de los datos.

¿Consideró la calidad de los datos en sus procesos de publicación? No, por la urgencia de mostrar los datos solicitados por entes de control.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos? No, actualmente estamos en proceso de análisis. Esstamos analizando la nueva versión de COBIT

¿Su organización sigue algún modelo de Gobierno de Datos? Si

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? Si

¿Qué entiende por compresibilidad de los datos en formatos abiertos? Que permita ser procesados por las maquinas y por humanos.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Estamos en proceso de generacion de datos en RDF

¿Qué características tienen estos mecanismos? Procesable por humanos y máquinas

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Utilización de tecnologías de la Web Semántica

Entrevista N: 14

Nombre del entrevistado: Egdar Rivadegneira, Universidad Nacional de Bolívar, **Correo Electrónico:** eribaden@yahoo.com, **Fechas de entrevista:** 2019-06-19 09:38:40, **En qué etapa esta su portal:** Implementado, **Número de conjunto de datos publicados:**

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? No, hasta determinación final de roles

¿Consideró la calidad de los datos en sus procesos de publicación? Si

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Moderadamente importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? De acuerdo

¿Su organización sigue algún estándar de calidad de datos? No

¿Su organización sigue algún modelo de Gobierno de Datos? No

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? No conoce

¿Qué entiende por comprensibilidad de los datos en formatos abiertos? Datos claros y transparentes

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Aún no, en análisis

¿Qué características tienen estos mecanismos? Aún no, en análisis

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? Aún no, en análisis

Entrevista N: 15

Nombre del entrevistado: Stefani Naranjo- Secretaria de Gobierno Abierto, **Correo Electrónico:** steph_naranjo@yahoo.com, **Fechas de entrevista:** 2019-06-22 08:28:25, **En qué etapa esta su portal:** Publicado, **Número de conjunto de datos publicados:** 128

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? No, se tuvo apoyo de entidades públicas pero no se buscó el apoyo de academia y sociedad civil en su construcción. Se contaba con tiempos cortos.

¿Consideró la calidad de los datos en sus procesos de publicación? Sí, se elaboraron algunos lineamientos y directrices como herramientas para la publicación de datos para las entidades públicas... sin embargo existía aspectos de fondo que debían ser resueltos.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos? No, existe ningún estándar de calidad que se siga actualmente como instituciones, sin embargo existen áreas que manejan datos que creería que manejan su estándar de calidad.

¿Su organización sigue algún modelo de Gobierno de Datos? No

¿Aplica el concepto de datos maestros dentro de su institución (MDM)? No conoce

¿Qué entiende por compresibilidad de los datos en formatos abiertos? Pues entiendo que es reducir el manejo de los datos a un grupo de datos más relevantes (datos maestros) y que cuenten con la calidad necesaria para ser reutilizados.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Ninguno

¿Qué características tienen estos mecanismos? Ninguno

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos? El portal se encuentra paralizado de momento. Se planea fortalecer el portal.

Entrevista N: 16

Nombre del entrevistado: Jorge Pantoja-Organizaciòn San Francisco, **Correo Electrónico:** jorgepantoja@sanfranciscoglobal.org, **Fechas de entrevista:** 2019-06-28 11:27:00, **En qué etapa esta su portal:** Aprobado, **Número de conjunto de datos publicados:** 0

¿Incluyó a los potenciales reutilizadores en la implementación de su portal? Si, como un esfuerzo desde la sociedad civil en la Fundación San Francisco Global, nos encontramos implementando el Observatorio Productivo, que tiene como fin fomentar una correcta transparencia y calidad en datos públicos, para lo cuál estamos articulados desde la sociedad civil para poder organizar este portal.

¿Consideró la calidad de los datos en sus procesos de publicación? Si, uno de los objetivos principales de este proyecto es garantizar la calidad e integridad en los datos públicos, enfocados al sector productivo.

¿En la implementación de un proyecto de datos abiertos debe considerarse a los infomediarios o potenciales reutilizadores desde el inicio? Muy importante

¿En la implementación del portal de datos abiertos la calidad de datos es un proceso que depende de los sistemas informáticos de origen? Totalmente de acuerdo

¿Su organización sigue algún estándar de calidad de datos? Si, principalmente se utiliza el estándar DCAT, <https://www.w3.org/TR/vocab-dcat/>

¿Su organización sigue algún modelo de Gobierno de Datos? No

¿Aplica el concepto de datos maestros dentro de su institución (MDM) ? No

¿Qué entiende por comprensibilidad de los datos en formatos abiertos? Que los datos se encuentren en una codificación adecuada (ej. UTF8) y estén correctamente descritos para su uso, con un correcto preprocesamiento para garantizar la calidad.

¿Qué mecanismos ha implementado en su portal para favorecer la comprensibilidad los datos publicados? Publicar diccionarios de datos en todos los conjuntos de datos.

¿Qué características tienen estos mecanismos? El diccionario de datos ayuda a describir correctamente los campos de los datasets publicado.

¿Qué acciones está emprendiendo para mejorar la reutilización del portal de datos abiertos?

Articulación directa con la academia desde el inicio del proyecto.

10.3 Anexo 3: Transcripción de las observaciones de los grupos focales

Grupos de Trabajo N. 1

Número de participantes: 8

Instituciones participantes: Instituto Nacional de Evaluación y Senescyt

Agenda:

1. Justificación
2. Normas de Calidad
3. Marcos de Referencia
4. Estructura
 - Nivel Estratégico
 - Nivel Gerencial
 - Nivel Operativo

Desarrollo y aportes:

Roberto Moya roberto.moya@evaluacion.gob.ec

Cristian Álvarez cristian.alvarez@evaluacion.gob.ec

Patricio Toledo patricio.toledo@evaluacion.gob.ec

Alejandra Espinosa alejandra.espinosa@evaluacion.gob.ec

Michelle López michelle.lopez@evaluacion.gob.ec

Darwin Padilla darwin.padilla@evaluacion.gob.ec

Elena Chicaiza elena.chicaiza@evaluacion.gob.ec

Una vez presentado el marco de referencia, se realizaron las siguientes participaciones:

Preguntas directrices:

- ¿Es aplicable el marco de referencia a su institución?

Participante 5 indica que, si es aplicable el marco de referencia para su institución en especial porque es ordenado para no dejar ningún aspecto descuidado, sin embargo, indica que las

máximas autoridades comprendan el propósito de la publicación porque en el portal se publicó los datos por decreto presidencial y se tuvo que hacerlo como se podía.

Cuando se realizó el proceso de publicación se invirtió mucho tiempo en la instalación de la herramienta y aprender su utilización nunca se realizó un trabajo con los datos.

Por otra parte, el participante 4 indica que, si el estándar ISO 8000 de master data, no se conocía si este aprobado, a lo que se contesta las condiciones del mismo. Manifiesta que es importante para consolidar los datos que por la historia del instituto se ha desarrollado poco a poco y no tenía una sola base de datos, lo que lo hace pertinente para este caso.

El participante 1 y 2 manifiestan que si sería aplicable pero que es importante considerar que de mucho sistema no existe ni documentación.

El participante 7 informa que es ordenado e incluye características no consideradas.

- Ventajas

Los participantes manifiestan que es la estructura ordenada, que se alinee a la máxima autoridad porque muchas veces no se entiende a los datos generados.

Además, la adaptación de las normas de calidad y orientado al usuario final les parece que es una característica importante.

En la publicación de datos anterior no se consideró a los usuarios finales, simplemente se publicó el dato. Actualmente se piensa trabajar en temas de usabilidad del dato

- Desventajas

Indican que una de las debilidades podría ser la especificación de estandarización que es importante a nivel de país, no se maneja un estándar para los metadatos lo que complica la comprensión por parte del usuario. Por conocimiento interno se está utilizando Dublín Core, pero no es una estándar nacional, por tanto, no se puede compartir la información con otros ministerios.

Además, no existen catálogos únicos a nivel de país. El marco de referencia debería ser tácito en este requisito, por cuanto la catalogación institucional es básica al momento de publicar los datos.

Otra de las posibles desventajas está asociada al cambio cultural de las personas, por cuanto la alineación a la estrategia podría ser una limitante.

- ¿El marco de referencia favorece la comprensibilidad por parte del reutilizador?

- Ventajas

Los participantes opinan, que el marco de referencias al estructurarse por capas, consolida los procesos de desde el nivel estratégico para la publicación, y lo más importante es el metadato para que el usuario pueda comprender la información publicada.

- Desventajas

Otra vez relacionada al cambio cultural, a pesar de la publicación en abierto muchos de los usuarios insisten en solicitar por escrito la información para que sea entregada con un CD. Este problema tendremos por varios años posteriores.

Además, se agregaron una consulta acerca la lógica del funcionamiento del archivo maestro de datos abiertos a detalle y la norma de calidad ISO 25000 relacionada con datos.

Se realizó una aplicación de este concepto, indicando que era un concepto nuevo, pero totalmente aplicable a la apertura por cuanto muchas veces preparar el dato separadamente del origen y cuando desean volver actualizar se complica.

Además, la calidad de los datos en origen es un tema que muchas veces no es analizado ni tomado en cuenta previa a la publicación, inclusive al interior de la organización mismo.

Grupos de Trabajo N. 2

Número de participantes: 4

Instituciones participantes: Gobiernos Central encargado del portal de datos abiertos del Ecuador datos.ec y un investigador de datos abiertos

Agenda:

5. Justificación
6. Normas de Calidad
7. Marcos de Referencia
8. Estructura
 - Nivel Estratégico
 - Nivel Gerencial
 - Nivel Operativo

Desarrollo y aportes:

Juan Carlos Castillo Juan.castillo@mintel.gob.ec

Jhoana Pazmiño jhoana.pazmino@mintel.gob.ec

Carlos Javier Pacha jpacha@senplades.gob.ec

Sandra Gaibor sgaibor@senplades.gob.ec

Grupos de Trabajo N. 3

Número de participantes: 4

Instituciones participantes: Universidades del Ecuador y sociedad civil

Agenda:

9. Justificación

10. Normas de Calidad

11. Marcos de Referencia

12. Estructura

- Nivel Estratégico
- Nivel Gerencial
- Nivel Operativo

Desarrollo y aportes:

Edison Loza Edison.loza@epn.edu.ec

Tania Gualli Tania.gualli@epn.edu.ec

Jorge Pantoja patricio.toledo@evaluacion.gob.ec

Nelson Piedra npiedra@utpl.edu.ec

Una vez presentado el marco de referencia, se realizaron las siguientes participaciones:

Preguntas directrices:

- ¿Es aplicable el marco de referencia a su institución?

El participante 1 tiene una inquietud acerca de cómo identificar a los infomediarios, y que se incluyan en el marco de referencia la metodología para poder hacer el contacto y levantar sus requerimientos y hacer un seguimiento. Se propone hacer la conferencias y desconferencias.

Yo creo que al menos el marco de referencia es muy aplicable, y la mayor ventaja que se ve, es que nace del nivel estratégico, que sea un proyecto que las máximas autoridades deben estar en conocimiento y entendimiento pleno de todo el proyecto para que tenga

sostenibilidad, ya que si nace del nivel técnico no se ve la utilidad, y sobre todo que existe el apoyo necesario para que exista el proyecto.

Para los siguientes pasos del marco de referencia es importante incluir el cambio cultural del uso de los datos abiertos, el desconocimiento de un tema de esta naturaleza, a pesar que el marco habla de anonimización, pero los usuarios cuando ven un proyecto de datos abiertos la ciudadanía piensa en seguridad, por tanto, es importante trabajar acerca del uso de los datos abiertos por cuanto es importante saber

Es un marco de referencia muy interesante, y el aporte es desde la experiencia ya que he implementado de varios portales de datos abiertos, uno de los fines no es solo la reutilización sino la comparación con otros datos que tengan un contexto similar, a nivel nacional o regional, desde este entorno es importante incluirse la forma de como compararse datos nacionales o regionales, por ejemplo, los de compras públicas o las áreas administrativas, por tanto, uno de los fines debe ser la comparación por tanto DCAT es un estándar ya conocido. Uno de los fines debe ser la estandarización y comparación.

El proceso que se ha tratado de implementar es similar al propuesto, en cuanto al responsable del portal pero que ocasionó que el miedo de uso de los datos es una barrera sino viene amparado por la máxima autoridad, se coincide que se incluye el cambio y cultura, es necesario pensar en la sostenibilidad por cuanto es necesario contar con personal calificado y existe una alta rotación esto puede afectar a la sostenibilidad del portal. Además, es necesario contar con las guías necesarias para la participación ciudadana, para conocer desde la ciudadanía entendié o no los datos publicados. Los metadatos es un elemento fundamental del marco de referencia, ya que muchos de los campos que constan en el marco no se utilizan actualmente, la política pública también incluye una guía, pero no se incluye todos los campos propuesto. Los metadatos es lo importante y no es clara peor el diccionario de datos que es importante para la apertura. Se creo una guía a nivel de gobierno existe los campos adicionales.

Es importante incluir la fuente de los datos eso da garantía que la integridad de los datos al publicar por cuanto muchas veces pasa por varios usuarios y no llega como están el origen. Esto es necesario incluir caso contrario se piensa que los datos fueron manipulados o no, y si fueron afectados en su integridad.

Se pide realizar una ampliación del archivo maestro de datos abiertos que podría mejorar la relación con los archivos de origen

Se agrega SKOS que para ciencias de conocimiento y se agrega estos elementos. Es necesario establecer los catálogos a nivel nacional que permita cosechar los datos.

El tercer participante indica que es super interesante el marco de referencia, sin embargo, una de las problemáticas es la falta de estandarización, la descripción a nivel táctico, estratégico y operacional por tanto hay que hacer una guía de implementación. A futuro debería hacerse un análisis de impacto de tomar una decisión dentro del portal a nivel táctico, operativo y estratégico y como esto afecta a la metadata si afecta, que implica cada una de estas decisiones. La idea sería generar un sistema experto para que te asesore que decisiones tomar.

La medición de la comprensibilidad me parece que es lo correcto, por tanto, es aplicable para todos los datos abiertos, dentro del framework el tema de anonimización es interesante y hay que dar fuerza dentro del marco de referencia por cuanto es importante manejarlo.

De la definición de comprensibilidad está orientada a un usuario común por tanto es importante considerarlo dentro de una comparativa.

Piensan que la aplicación del marco de referencia

Los aspectos que se incluyen en la comprensibilidad están muy bien identificados, pero se sugiere el glosario de términos de negocio que aclaren la información que se está publicando. Porque puede dar lugar a malinterpretaciones si el dato no es completamente comprendido.

Jorge indica que estuvo a cargo de la guía de publicación de datos abiertos del Ecuador, por tanto, se considera que el valor agregado es la comprensibilidad es nuevo y responde a una necesidad.

Están bien direccionados, los tres indicadores son los más importantes dentro de un proceso de apertura y su relación al diccionario. Es importante que los metadatos son claros y lo importante es la licencia de uso.

Se entiende que el medio de publicación es el portal de datos abiertos, por tanto, es importante sugerir que deberían tener los portales de datos abiertos, desde el punto de vista del software, además descarga de formatos y se toma como referencia el nivel de madurez de los portales por tanto es importante incluir un desglose de esto como características técnicas. Los que quieren publicar los datos deben conocer.

Hay muchas limitaciones para publicar datos, la mayoría son técnicas desde que no están sistematizados, por tanto, se debería generar hojas de ruta para saber cuándo se abrirá el dato


y la fechas, por las limitantes para tomar conciencia del proceso que involucra generar el dato en abierto.

Se sugiere incluir características de DCAT como skos y data provenance para robustecer el concepto datos en origen. La importancia de los marcos de referencia es en tomar partido en varios temas uno de ellos podría ser este, data provenance y skos como organizador de conocimiento.




Universitat d'Alacant
Universidad de Alicante

10.4 Anexo 4: Organización de las mesas de trabajo para ANIE (Agenda Nacional de Investigación Educativa)


 GOBIERNO DE LA REPUBLICA DEL ECUADOR

Agenda Nacional de Investigación Educativa - ANIE
 Comisiones Técnicas


 ANIE

* En la parte de la normativa incluir los contactos de la parte jurídica
 ** En comunicación por favor los contactos de comunicación

Comisiones de Trabajo	Productos	Líder	Corresponsables
1.- Documentos	1.1. Plan estratégico (metas a mediano y largo plazo)	Mineduc	UASB, Senescyt, IAEN
	1.2. Agenda Nacional de Investigación Educativa (documento resumen de difusión nacional)	Presidencia y Secretaría de la comisión	UNAE, Mineduc, Fidal, UASB, Enseña Ecuador
	1.3. Documento consolidado (trabajo de toda la construcción de la ANIE)	UTI	Ineval, Senescyt, UASB, CACES, Enseña Ecuador
	1.4. Plan Temático (Planificación de actividades durante el primer año)	Presidencia y Secretaría de la comisión	UNAE, UTI; UTMach
2.- Normativa*	2.1. Convenio Marco	Senescyt	Todos
	2.2. Convenio Modificatorio (ingreso de nuevos integrantes)	Senescyt	Todos
	2.3. Marco normativo (Reglamento interno) de la Comisión Interinstitucional	Presidencia y Secretaría de la comisión	Todos
	2.4. Marco normativo de bases de datos	Presidencia y Secretaría de la comisión	Todos
3.- Roles de nuevos actores	3.1. Definición de roles de nuevos actores	FIDAL	Unesco, Ineval, Senescyt, UASB
4.- Redes, contactos y financiamiento	4.1. Propuesta de gestión para el financiamiento	PUCE	Grupo Faro, Unesco, CES, UNAE, Ineval, Senescyt, VVOB,SEE
	4.2. Propuesta de factibilidad de creación de red	PUCE	Grupo Faro, Unesco, CES, UNAE, Ineval, Senescyt, VVOB,SEE
	4.3 Propuesta de estrategia	Mineduc	MIES, Senescyt, PUCE, UNAE, CES, Enseña Ecuador, FIDAL
5.- Acceso y calidad de BDD	5.1. Elaboración de modelo de gestión de información	UCE	MIES, Mineduc, Ineval CACES, Senescyt, apoyo técnico DINARDAP
	5.1. Elaboración de propuesta de metodología de interoperabilidad	UCE	MIES, Mineduc, Ineval CACES, Senescyt (Líderes en las bases de datos cada institución en su ámbito)
	5.1. Elaboración de propuesta de protocolos, normas, mecanismos de acceso y calidad de los datos.	UCE	MIES, Mineduc, CACES, Flacso, Unesco, Senescyt, Grupo Faro; UDLA
	5.3. Talleres de difusión de la información	Ineval	MIES, Mineduc, Ineval, CACES, Senescyt.



UNIVERSIDAD CENTRAL DEL ECUADOR
RECTORADO

Oficio No. R-760-2019
Quito D.M., 24 de julio de 2019

Magister
María Dolores Benavides Ordoñez
COORDINADORA DE GESTIÓN DE LA INFORMACIÓN
SECRETARÍA DE EDUCACIÓN SUPERIOR, CIENCIA,
TECNOLOGÍA E INNOVACIÓN (SENESCYT)
En su Despacho

De mi consideración:

La Universidad Central del Ecuador tiene como misión institucional el ofrecer acceso al conocimiento y cultura universal, y generar investigación de excelencia integrada al desarrollo humano del Ecuador. Esta misión la cumple a través de la formación de grado y posgrado, de la investigación social y experimental y de la vinculación con la sociedad.

En este contexto la UCE informa del interés de ser incluida en la Agenda Nacional en Investigación Educativa (ANIE), impulsada por la SENESCYT, que busca desarrollar y fortalecer la investigación científica de pertinencia en el ámbito educativo ecuatoriano, mediante la integración de diversos actores del sector público y privado, con el fin de incidir en el desarrollo de políticas públicas acordes a la realidad nacional, que permitan el mejoramiento de la calidad educativa en los diferentes niveles de formación.

Desde las fortalezas de nuestra institución, en la que se está apoyado en la formación doctoral de docentes en el campo de la Investigación Educativa, consideramos que podemos ser actores activos en la construcción y ejecución de esta agenda, comprometiéndonos específicamente en la coordinación técnica de la comisión de acceso y calidad de bases de datos.

La persona a cargo de esta coordinación, y quien conformará la Comisión Interinstitucional con voz, será la Ing. Susana Cadena, Docente de la Facultad de Ciencias Administrativas. La Ingeniera Cadena informará mensualmente de los avances y actividades realizados en el marco de la ANIE al Vicerrectorado de Investigación, Doctorados e Innovación.

Seguros de contar con una respuesta favorable a este pedido, que busca el fortalecer el trabajo interinstitucional en temas tan importantes para el desarrollo nacional, como es la investigación educativa, me despido.

Atentamente,


Dra. María Augusta Espin Estévez
RECTORA (s)



FACULTAD DE CIEN
ADMINISTRATI
RECEPCION DE DOCUM

copia - Dra. María Mercedes Gavilánez, VICERRECTORA DE INVESTIGACIÓN, DOCTORADOS E INNOVACIÓN
- Ing. Susana Cadena, DOCENTE, FACULTAD DE CIENCIAS ADMINISTRATIVAS

15 JUL 2019

DECANATO

Ciudad Universitaria: 53010000 - 593 22 505 000 - 593 22 501 500 - Telefax: 593 22 505 890

10.5 Anexo 5: Certificado de aportación en la co-creación de la Política Nacional de Datos Abiertos.

PRESIDENCIA DE LA REPÚBLICA · MINISTERIO DE TELECOMUNICACIONES
Y DE LA SOCIEDAD DE LA INFORMACIÓN · SECRETARÍA NACIONAL DE
PLANIFICACIÓN Y DESARROLLO



Confieren el presente reconocimiento a:

Susana Cadena

Por su valioso aporte en las mesas de diálogo para la
“**Co-creación de la Política Nacional de Datos Abiertos**”
realizadas en la ciudad de Quito, 10 y 11 de diciembre de 2018.

Quito, D.M., 27 de diciembre de 2018

Firmado Digitalmente por: RAUL JAVIER JARA
IÑIGUEZ
Hora local: 02/01/2019 17:13

Mgs. Javier Jara Iñiguez
Subsecretario de Gobierno Electrónico

Universitat d'Alacant
Universidad de Alicante