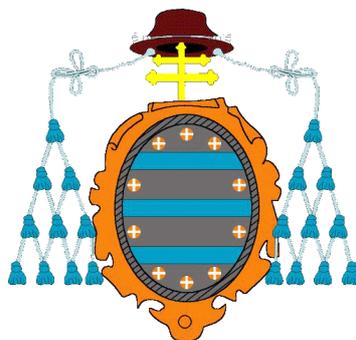


UNIVERSIDAD DE OVIEDO
Departamento de Informática



TESIS DOCTORAL

“METAMODELO PARA INTEGRACIÓN DE DATOS
ABIERTOS APLICADO A INTELIGENCIA DE
NEGOCIOS”

Autor

Luz Andrea Rodríguez Rojas

Directores

Doctor D. Juan Manuel Cueva Lovelle

Doctor D. Carlos Enrique Montenegro Marín

Oviedo, 2017

Agradecimientos

Primero quiero darle las gracias a Dios por mi vida, mi salud y por las oportunidades que ha puesto en mi camino para crecer como persona y profesional.

A mis Directores, el Dr. D. Juan Manuel Cueva Lovelle y el Dr. D. Carlos Enrique Montenegro, quienes con su apoyo y paciencia me acompañaron y contribuyeron al feliz término de este proceso de formación. Al Dr. Ruben Gonzalez Crespo por su apoyo incondicional y oportuno.

Al grupo de investigación MDA Research Group, de la Universidad de Oviedo, especialmente a los compañeros Guillermo, Cristian, Edward, Jordan y Vicente que estuvieron siempre dispuestos a colaborar y proporcionar un ambiente amable y cálido.

A mis padres Antonio y Carmenza que con su ejemplo me enseñaron el valor del trabajo, la responsabilidad y la persistencia, por tantos años de sacrificios para que sus hijos hoy en día sean personas de bien. A mis hermanos Claudia, Lucia, Lilia y Miguel, sin su apoyo incondicional hoy no sería posible culminar este proceso. Estare eternamente agradecida por todo lo que han hecho por mí.

A mi esposo Giovanni por alentarme y no dejarme desfallecer, por su ejemplo y apoyo, por darme la mayor bendición de mi vida: Abigail. A ella por cambiar mi mundo, por enseñarme a amar a alguien más que a mi misma y ser el motivo que me impulsa cada día para ser mejor persona.

A todas aquellas personas que con sus consejos, paciencia y colaboración aportaron al logro de esta meta, especialmente a Doña Olga Maldonado a quien considero de mi familia y agradezco por sus consejos y sabiduría, a la Dra. Jenny Danna por su sincera amistad y aprecio y a Monica Fernanda Sanchez por su tiempo y colaboración.

Resumen

El sector público, en el cumplimiento de sus funciones, genera, gestiona, sufraga y conserva una gran cantidad de información de calidad. Dicha información es potencialmente reutilizable tanto por los ciudadanos como por las empresas. Esa información que obra en poder de las instituciones y administraciones públicas presenta un potencial económico y social enorme a la par que contribuye a la transparencia en la gestión pública y a la generación de confianza por parte de la ciudadanía.

La finalidad de las iniciativas open data consiste en hacer públicos los datos en poder de las administraciones públicas y el sector privado para que puedan ser posteriormente reutilizados con la finalidad de generar nuevos conocimientos y aplicaciones que en muchos casos, añadirán un valor al dato primario y contribuirán a potenciar la innovación. Este reciente paradigma provee a las empresas de enorme oportunidades para sacar provecho de diversos tipos de datos que contribuyan a la toma de decisiones acertadas y por ende mayores posibilidades de éxito en un entorno tan competitivo.

La apertura de los datos conlleva beneficios para todos los que participan de ella: para los ciudadanos, para las empresas y para la propia administración. Para las empresas porque a partir de estos datos pueden diseñar y crear nuevos servicios y productos con los que se mejora su productividad y su competitividad; para los ciudadanos porque les permite estar mejor informados respecto a la gestión pública lo que contribuye a que tomen mejor sus decisiones; a la propia administración pública porque el hecho de proporcionar una mayor cantidad de información a sus ciudadanos permite a estos la búsqueda de soluciones a problemáticas para las cuales las instituciones no tienen tiempo ni recursos.

El reto está en procesar cantidades enormes de datos y convertirlas en decisiones inteligentes y oportunas, por eso se hace necesario aprovechar esta importante oportunidad de crear valor al combinar y utilizar con eficacia conjuntos de datos de diversas fuentes.

La Ingeniería Dirigida por Modelos es una propuesta que se ha venido trabajando desde hace varios años por diversos autores que plantean el uso de modelos como eje fundamental en todo el ciclo de vida de un proyecto de software que reduce el tiempo y esfuerzo en el desarrollo. Este es el camino propuesto para la presentación de un metamodelo, materializado en un lenguaje de dominio específico, que a partir de una fuente de datos abiertos captura, transporta, transforma y analiza datos, y finalmente permite visualizar información que apoya la toma de decisiones.

Palabras Claves: Datos abiertos gubernamentales, Toma de decisiones, Lenguaje de dominio específico, Metamodelo, Ingeniería dirigida por modelos.

Abstract

The public sector, in the fulfillment of its functions, generates, manages, pays and retains a great quantity of quality information. Such information is potentially reusable by both citizens and businesses. This information held by public institutions and administrations has enormous economic and social potential and contributes to transparency in public management and the generation of trust by citizens.

The purpose of open data initiatives is to make public data held by public administrations and the private sector so that they can be later reused with the purpose of generating new knowledge and applications that in many cases, add value to the primary data and contribute to innovation. This recent paradigm provides companies with huge opportunities to take advantage of various types of data that contribute to successful decision making and thus greater chances of success in such a competitive environment.

The opening of the data brings benefits to all who participate in it: for the citizens, for the companies and for the own administration. For the companies because from this data they can design and create new services and products with which it improves its productivity and its competitiveness; For the citizens because it allows them to be better informed about the public management that contributes to that they make better their decisions; To the public administration itself because providing more information to its citizens allows them to find solutions to problems for which institutions do not have the time or resources.

The challenge is to process huge amounts of data and make them intelligent and timely decisions, so it is necessary to take advantage of this important opportunity to create value by combining and effectively using data sets from various sources.

The Model-Driven Engineering is a proposal that has been working for several years by various authors who propose the use of models as the fundamental axis throughout the life cycle of a software project that reduces time and effort in development. This is the proposed way to present a metamodel, materialized in a specific domain language, that from an open data source captures, transports, transforms and analyzes data, and finally allows to visualize information that supports decision making.

Keywords: Open government data, Decision making, Domain specific language, Metamodel, Model-Driven Engineering.

TABLA DE CONTENIDO

PARTE I PLATEAMIENTO DEL PROBLEMA	
CAPÍTULO 1 -INTRODUCCIÓN.....	12
1.1 PLANTEAMIENTO DEL PROBLEMA	13
1.2 HIPÓTESIS	15
1.3 OBJETIVOS	16
1.3.1 <i>Objetivos Específicos</i>	16
1.4 APORTACIONES Y BENEFICIOS	17
CAPÍTULO 2 -METODOLOGÍA Y DESARROLLO DE LA INVESTIGACIÓN.....	18
2.1 FASE 1: RECOLECCIÓN DE INFORMACIÓN	19
2.2 FASE II: SELECCIÓN DE HERRAMIENTAS.....	20
2.3 FASE III: ESTUDIO DE PLATAFORMAS Y DATASETS	20
2.4 FASE IV: DESARROLLO DEL METAMODELO, PRUEBAS Y DIFUSIÓN DE RESULTADOS	21
2.4.1 <i>Desarrollo metamodelo</i>	21
2.4.2 <i>Pruebas</i>	21
2.4.3 <i>Difusion de resultados</i>	21
2.5 ORGANIZACIÓN DE ESTA TESIS	21
PARTE II MARCO TEORICO	
CAPÍTULO 3 – DATOS ABIERTOS	25
3.1 DEFINICIÓN	25
3.2 TIPOS DE DATOS ABIERTOS.....	27
3.3 ACTORES INVOLUCRADOS	28
3.4 FORMATOS DE DATOS ABIERTOS	31
3.5 BENEFICIOS DE LOS DATOS ABIERTOS	36
3.6 BARRERAS DE LOS DATOS ABIERTOS	39
CAPÍTULO 4 – INTELIGENCIA DE NEGOCIOS Y BIG DATA.....	45
4.1 INTELIGENCIA DE NEGOCIOS	45
4.1.1 <i>Arquitectura típica</i>	45
4.1.2 <i>Cuadrante magico BI Gartner 2017</i>	47
4.1.3 <i>Beneficios de la BI</i>	51
4.2 BIG DATA	52
4.2.1 <i>Cadena de valor</i>	53
4.2.2 <i>Big data y creacion de valor</i>	61
4.2.3 <i>Retos del big data</i>	63
CAPITULO 5 – INGENIERIA DIRIGIDA POR MODELOS (MDE).....	66
5.1 INGENIERIA DIRIGIDA POR MODELOS MDE.....	66
5.1.2 <i>Beneficios de MDE</i>	72
5.2 ARQUITECTURA DIRIGIDA POR MODELOS (MDA).....	73
5.2.1 <i>Definición</i>	73
5.2.2 <i>Ciclo de vida del desarrollo con MDA</i>	74
5.2.3 <i>Niveles de abstración</i>	76
5.2.4 <i>Ventajas de MDA</i>	78
5.3 LENGUAJES DE DOMINIO ESPECÍFICO (DSL).....	80
5.3.1 <i>Definición</i>	80
5.3.2 <i>Tipos de Lenguajes de Dominio Específico (DSL)</i>	81
5.3.3 <i>Características de un lenguaje de dominio específico (DSL)</i>	82
5.3.4 <i>Ventajas de un lenguaje de dominio específico (DSL)</i>	83
PARTE III ESTADO DEL ARTE	
CAPITULO 6: INICIATIVAS DE DATOS ABIERTOS	87

6.1 ANTECEDENTES	89
6.2 LATINOAMERICA Y CARIBE	91
CAPITULO 7- TOMA DE DECISIONES EN AGROINDUSTRIA	102
7.1 DATOS ABIERTOS Y AGROINDUSTRIA.....	105
7.1.1 <i>Toma de decisiones más eficiente y eficaz</i>	105
7.1.2 <i>Fomentar la innovación</i>	108
7.1.3 <i>Impulsar el cambio a través de la transparencia</i>	110
PARTE IV DESARROLLO DEL PROTOTIPO	
CAPITULO 8 - DESCRIPCION DEL PROTOTIPO	114
8.1 CONSTRUCCION DEL DSL.....	117
CAPITULO 9 - VALIDACIÓN Y PRUEBAS	121
9.1 PRUEBAS DE USABILIDAD.....	121
9.1.1 <i>Evaluación Heurística</i>	121
9.1.2 <i>Test de Usabilidad</i>	123
9.1.3 <i>Entrevista</i>	123
9.2 VALIDACIÓN ESTADÍSTICA DE SUPUESTOS E HIPÓTESIS	124
PARTE V CONCLUSIONES	
CAPÍTULO 10 - CONCLUSIONES Y TRABAJO FUTURO	129
10.1 VERIFICACIÓN Y CONTRATE DE LA HIPÓTESIS	130
10.2 VERIFICACIÓN, CONTRASTE Y EVALUACIÓN DE OBJETIVOS	130
10.3 ESCRUTINIO PÚBLICO	132
10.3.1 <i>Libros</i>	132
10.3.2 <i>Capítulo de libro</i>	132
10.3.3 <i>Artículos Publicados (no JCR)</i>	132
10.3.4 <i>Artículos Publicados (JCR)</i>	133
10.3.5 <i>Congresos</i>	134
10.4 LÍNEAS DE INVESTIGACIÓN FUTURAS	134
REFERENCIAS BIBLIOGRAFICAS.....	136
<hr/>	
ANEXOS	151

Índice de Figuras

Figura 1. Metodología de la tesis	19
Figura 2 Cadena de valor de los datos abiertos	27
Figura 3. Tipos de datos abiertos.	28
Figura 4. Beneficios de los datos abiertos.	36
Figura 5. Barreras ámbito institucional.	40
Figura 6. Barreras ámbito complejidad de las tareas.	40
Figura 7. Barreras ámbito uso y participación.	41
Figura 8. Barreras ámbito legislación.	42
Figura 9. Barreras ámbito calidad de la información.	42
Figura 10. Barreras ámbito técnico.	43
Figura 11. Arquitectura típica de inteligencia de negocios.	46
Figura 12. Cuadrante mágico para plataformas analíticas y BI.	48
Figura 13. Power BI service.	49
Figura 14. Arquitectura Qlik Sense.	51
Figura 15. Ejemplo de nube de etiquetas	60
Figura 16. Arquitectura de los conceptos de MDE.	68
Figura 17. Proceso tradicional y proceso con MDA	75
Figura 18. Distribución regional de datos abiertos.	88
Figura 19. Top diez barómetro de datos abiertos tercera edición.	88
Figura 20. ¿Cómo funciona una clínica de plantas?	106
Figura 21. Rendimientos potenciales y análisis para cultivar arroz en marzo 2017	107
Figura 22. Búsqueda de documentos en AGRIS	109
Figura 23. Arquitectura general de la solución	114
Figura 24. Diagrama de capas aplicación web	116
Figura 25. Modelo de datos de la Aplicación Web	117
Figura 26. Lenguaje de Dominio específico para el prototipo	118
Figura 27. Modelo del prototipo	119
Figura 28. Vista de la creación de fuentes de datos en la aplicación web	120
Figura 29. Ejemplo de Dashboard usando portal de datos abiertos datos.gov.co	120
Figura 30. Resultados Evaluación Heurística	122

Índice de Tablas

Tabla 1. Búsqueda en bases de datos	19
Tabla 2. Clasificación de los conjuntos de datos	34
Tabla 3. Ranking datos abiertos Latinoamérica	91
Tabla 4. Componentes / tecnología de la arquitectura base	115
Tabla 5. Resultados test de usabilidad	123
Tabla 6. Aceptación y rechazo de los supuestos planteados	125
Tabla 7. Prueba de hipótesis supuestos del metamodelo	125

“La diferencia entre lo que hacemos y somos capaces de hacer, resolvería la mayoría de problemas del mundo”

Mahatma Gandhi.

PARTE I PLANTEAMIENTO DEL PROBLEMA

CAPÍTULO 1 -INTRODUCCIÓN

En los últimos años se ha iniciado una revolución en materia de acceso a la información pública por parte de los ciudadanos. A esta revolución se le ha denominado “Open Data” o Datos Abiertos [1], se trata de una nueva industria en torno a la reutilización de la información.

Durante mucho tiempo las entidades públicas han almacenado y protegido grandes cantidades de datos relativos a la actividad de gobierno. La liberación de esa información está ganando cada vez más atención debido a su potencial para estimular el crecimiento económico, apoyar el buen gobierno y facilitar la innovación social. Motivados por estos beneficios, varios gobiernos del mundo han comenzado a incluir temas de datos abiertos en su estrategias de administración electrónica y han puesto en práctica iniciativas que han dado lugar a la aparición del llamado movimiento de datos abiertos gubernamentales (OGD) [2].

Para la mayoría de los gobiernos los objetivos han estado relacionados con cuatro aspectos: el crecimiento económico, (innovación empresarial, la creación de empresas y puestos de trabajo); la participación ciudadana; aumento de la transparencia y rendición de cuentas; y la mejora en la eficiencia de las operaciones y servicios (mejor toma de decisiones mediante el acceso a la datos de otros organismos)[3].

Añadiendo valor a los datos originales proporcionados por gobiernos se puede facilitar la creación de nuevos servicios personalizados, nuevos mercados y nuevas empresas [4]. En los países pioneros en la apertura y utilización de datos abiertos ya hay evidencia de empresas que han explotado su potencial [5].

Sin embargo a pesar de los esfuerzos de los gobiernos para promover el movimiento OGD todavía hay una débil participación de los agentes sociales en la reutilización de los datos. Al mismo tiempo, se requiere que universidades, empresas y entidades con capacidad de promover la innovación se involucren activamente para aprovechar el potencial previsto [6]. Stephenson et al., encontraron cuatro obstáculos principales para el uso de información: (i) disponibilidad de los datos, (ii) voluntad para el uso, (iii) accesibilidad, usabilidad y calidad de los datos, y (iv) capacidad de

recolección, manejo y uso de datos [7]. Consideran clave mejorar la coordinación nacional e internacional y la colaboración intersectorial para la gestión de datos; producir y utilizar más productos derivados de los datos; trabajo conjunto de Gobiernos, sociedad civil y organismos académicos para romper las barreras al acceso a los datos y compartir e integrar la información para la toma de decisiones.

La Ingeniería Dirigida por Modelos es una propuesta que se ha venido trabajando desde hace varios años por diversos autores [8]–[14], los cuales plantean el uso de modelos como eje fundamental en todo el ciclo de vida de un proyecto de software que reduce el tiempo y esfuerzo en el desarrollo. Este es el camino propuesto para la presentación de un metamodelo, materializado en un lenguaje de dominio específico, que a partir de una fuente de datos abiertos captura, transporta, transforma, analiza datos y finalmente permite visualizar información que apoya la toma de decisiones.

1.1 Planteamiento del problema

La calidad y cantidad del suministro mundial de alimentos es crucial para el bienestar de todos los seres humanos. Esta situación constituye un reto pues se requiere que el sector agroindustrial sea eficiente y sostenible pues se estima que en el año 2050 la demanda mundial de alimentos se duplicará, impulsada por el crecimiento previsto de la población y de la economía que sacará de la pobreza a los consumidores de bajos ingresos[15]. Para alcanzar tales metas se requiere información, herramientas, oportunidades e incentivos para tomar decisiones que mantengan y mejoren la calidad de los recursos naturales y la calidad de vida humana.

Los "objetivos de Desarrollo Sostenible" - un marco para la próxima agenda global de desarrollo, dirigida por las Naciones Unidas - han destacado "lograr la seguridad alimentaria y mejorar la nutrición y promover la agricultura sostenible" como una prioridad global para los próximos 15 años[16]. Esta no es una tarea sencilla. Esto significará un aumento sustancial de la productividad agrícola, al mismo tiempo que se establecerán mercados más eficientes y equitativos. El progreso se impulsará principalmente proporcionando un mejor acceso a información precisa y oportuna para los pequeños agricultores, las empresas y los responsables políticos. Los datos abiertos pueden y deben ser parte de la solución.

Las tecnologías de la información y comunicaciones han permitido, en la agroindustria, la reducción de costos para obtener información, han facilitado la aparición de nuevos mercados y negocios y han permitido la creación de redes de colaboración y alianzas empresariales. Las agroindustrias solo serán sostenibles si son competitivas en términos de costos, precios, eficiencia operativa, oferta de productos y otros parámetros asociados [17]. Por ello información de calidad para la toma de decisiones constituye un factor clave en el crecimiento de este sector.

Los pequeños agricultores operan dentro de un contexto incierto y riesgoso. Además de la variabilidad del clima, el cambio climático, las dinámicas sociales, ambientales, institucionales y relacionadas con el mercado agrícola afectan sus decisiones y capacidad para hacer frente y adaptarse. Por ello requieren de información útil, por ejemplo, las enfermedades de las plantas y cómo protegerlas. Otros actores como los investigadores también requieren información de investigaciones previas para hacer seguimiento a problemáticas específicas; la pequeña y mediana empresa requiere información para conocer el estado de los negocios y las entidades gubernamentales necesitan un portal con información ejecutiva, en tiempo real, con alertas o seguimiento a eventos [18].

La creciente utilización de la agricultura de precisión y las tecnologías móviles, junto con mejoras en el software de gestión de datos genera oportunidades para una infraestructura de datos integrada que enlazadas a herramientas analíticas pueden contribuir a la toma de decisiones en el sector agroindustrial. Adicionalmente con el desarrollo del Internet de las cosas hay una tendencia al aumento del uso de dispositivos que permiten obtener información valiosa en todos los eslabones de la cadena productiva (proveedores, cultivo, postcosecha y comercialización). Ejemplo de ello son las redes inalámbricas de sensores que tienen diversas aplicaciones para la agricultura como lo son medición de las condiciones climáticas, irrigación inteligente, monitorización y control de plagas, agricultura de precisión, entre otras[19]–[21].

Gran parte de este creciente volumen de nuevos datos es privado, por ejemplo, información sobre dónde y cuándo ocurren las operaciones agrícolas, y sus consecuencias. También hay una creciente cantidad de datos públicos, tales como imágenes de satélite y datos meteorológicos y de suelos, rendimientos y datos económicos.

De esta manera se propone un prototipo, desarrollado con ingeniería dirigida por modelos, que permita usar los datos abiertos dispuestos por los gobiernos, relacionados con la agroindustria, tratarlos, analizarlos y visualizarlos con el ánimo de apoyar la toma de decisiones. Así se contribuye a la reutilización de la información dando lugar a aplicaciones mucho más ricas que aportan mayor conocimiento o que complementan la información ya existente. El reto está en procesar cantidades enormes de datos y convertirlas en decisiones inteligentes y oportunas [22].

1.2 Hipótesis

Hoy en día la información constituye un recurso esencial para desarrollar cualquier tipo de actividad en la sociedad. Vivimos en la economía del conocimiento, una economía caracterizada por utilizar la información como elemento fundamental para generar valor y riqueza. En este contexto, las nuevas tecnologías han modificado nuestra relación con la información facilitándonos su acceso y las vías para reutilizarla. De allí el incremento de iniciativas relacionadas con la apertura de datos gubernamentales y el creciente interés de la administración pública en garantizar su utilización.

Dentro de los impedimentos para el proceso de apertura de datos abiertos se encuentran la disponibilidad y el acceso, la capacidad, la facilidad de uso, la calidad, la vinculación y la combinación de datos, los metadatos, la falta de legislación y estandarización, la brecha digital, entre otros [23]–[25].

El límite potencial es la participación ciudadana real. Cuando se abren los datos, el problema se convierte en que todos los usen, con el fin de garantizar realmente el Gobierno Abierto e involucrar activamente a los ciudadanos en la toma de decisiones.

Si bien la participación de los ciudadanos es esencial, especialmente en tiempos de crisis social y económica, lograrlo a gran escala no será fácil. La frustración y la falta de confianza en las instituciones en muchos países son altas, por lo que no es ninguna sorpresa cuando la gente expresa dudas de que la apertura de los datos del gobierno no ayudará mucho. Si los ciudadanos no pueden encontrar y reconocer el significado real y relevante y el valor práctico en los datos, así como la forma de usarlos para hacer que el

cambio ocurra, no habrá ningún beneficio generalizado y duradero de la apertura [26].

El beneficio del uso de los datos abiertos impacta no sólo a los ciudadanos sino también a las empresas y al mismo gobierno, por ello se formula la siguiente hipótesis:

La incorporación de los beneficios del nuevo paradigma de datos abiertos, mediante un metamodelo que lo integre a soluciones de Inteligencia de negocios, crea valor para organizaciones públicas y privadas mediante la combinación de datos, aprovechando la información disponible en portales web y ofreciendo herramientas enriquecidas para la toma de decisiones

1.3 Objetivos

A partir de la hipótesis generada se propone como objetivo general de la tesis doctoral:

Proponer y diseñar un metamodelo que integre los beneficios del paradigma de datos abiertos a las soluciones de Inteligencia de negocios para crear valor en organizaciones públicas y privadas, mediante la combinación de datos y de esta manera ofrecer herramientas enriquecidas para la toma de decisiones

1.3.1 Objetivos Específicos

A fin de alcanzar el objetivo planteado, se proponen los siguientes objetivos específicos:

1. Diagnosticar la situación actual de datos abiertos. Se identificarán las iniciativas gubernamentales entorno a este reciente paradigma y los avances en implementación y uso. De igual manera se establecerá un estado de arte con referencia a la evolución de las soluciones de inteligencia de negocios y tratamiento de datos.
2. Especificar las herramientas tecnológicas aplicables y los requerimientos para la propuesta de un metamodelo que permitan

integrar los datos abiertos a soluciones de Inteligencia de negocios generando herramientas enriquecidas y viables para la toma de decisiones.

3. Estudiar las plataformas web y los conjuntos de datos abiertos disponibles para definir cuales datos se van a utilizar en el diseño y desarrollo del metamodelo.
4. Analizar, diseñar, desarrollar e implementar un prototipo de metamodelo que aproveche los beneficios del nuevo paradigma de datos abiertos.
5. Proponer y aplicar pruebas de validación para la propuesta.

1.4 Aportaciones y Beneficios

Se presentan las siguientes aportaciones y beneficios originales de esta tesis:

- Se desarrolla una revisión de las iniciativas gubernamentales de datos abiertos, avances en implementación y uso.
- Se identifican los beneficios y barreras de las iniciativas de datos abiertos
- Se realiza un estudio de portales de datos abiertos y conjuntos de datos abiertos para seleccionar aquellos que servirán de insumo al prototipo.
- Se desarrolla un DSL, que materializa el metamodelo, que permite la definición de la sintaxis abstracta y concreta.
- Se presenta el desarrollo una aplicación web que hace operable las funcionalidades propuestas en el Metamodelo.
- Se valida mediante pruebas la hipótesis inicialmente planteada.

CAPÍTULO 2 -METODOLOGÍA Y DESARROLLO DE LA INVESTIGACIÓN

Para el desarrollo de la investigación se ha optado por la aplicación de un enfoque **iterativo e incremental**. Cada iteración inicia estableciendo un modelo teórico de un conjunto de aspectos concretos que continua con la creación de una serie de prototipos para realizar experimentos sobre el modelo planteado. Los resultados parciales de estos experimentos e ideas eran contrastados y discutidos con expertos. La presentación de resultados ante la comunidad científica generó nuevas ideas que permitieron la mejora continua y el enriquecimiento del proceso investigativo.

Para desarrollar esta tesis se optó por una metodología de desarrollo ágil basada en el método científico (en el anexo 1 se describe detalladamente la metodología de trabajo SCRUM). Los resultados fueron enviados a congresos y revistas del índice Journal Citation Reports (JCR). La retroalimentación obtenida en estos envíos sirvió para mejorar diferentes aspectos de la propuesta.

El proceso investigativo consistió en una fase inicial de recolección de información y establecimiento del estado del arte de la problemática, una segunda fase en la que se definieron los requerimientos del metamodelo, una tercera fase en la que se estudiaron los portales de datos abiertos y los datasets disponibles para determinar aquellos más adecuados para utilizar en el prototipo y la última fase en la que se desarrolla la propuesta, se adelantan las pruebas y validación y se hace la difusión de resultados. La metodología seguida para la elaboración de esta tesis, se describe en la figura 1.

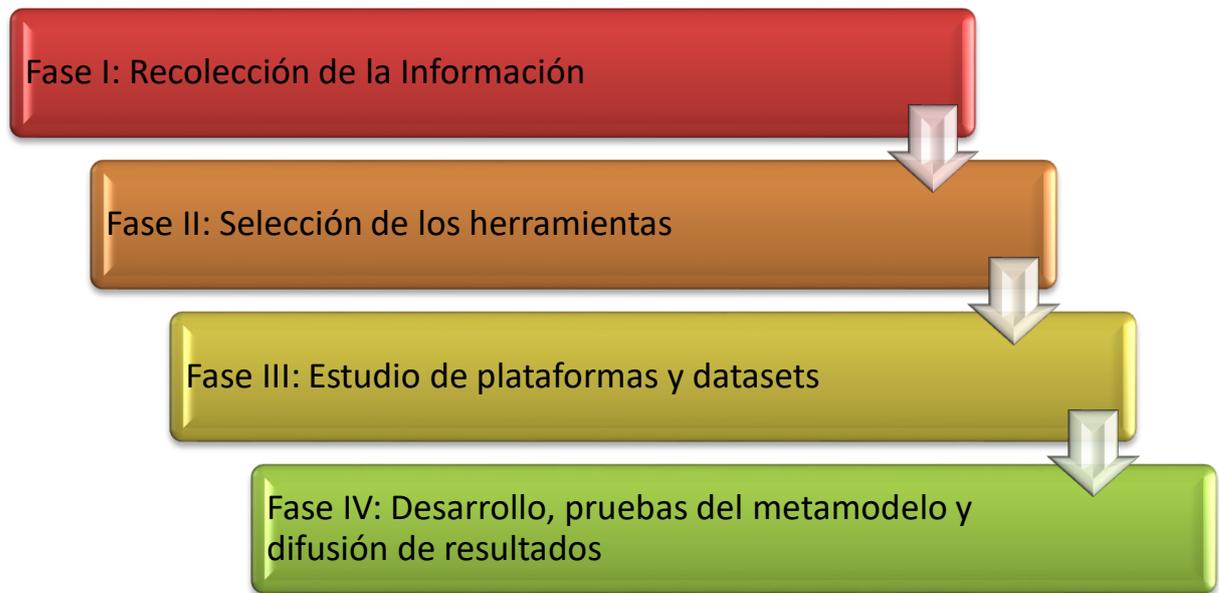


Figura 1. Metodología de la tesis

2.1 Fase 1: Recolección de información

La recolección de información se realizó mediante la búsqueda de artículos de revistas, artículos de conferencias, libros, informes gubernamentales y no gubernamentales y otra información en diversas bases de datos, incluyendo Science Direct, web of Science, ACM Digital Library, IEEE Explore, Google Scholar y Google. Las palabras clave que se utilizaron durante esta búsqueda fueron combinaciones de los términos datos abiertos, datos públicos abiertos, información del sector público, arquitectura + datos abiertos, metamodelos + datos abiertos. En la tabla 1 se muestran los resultados obtenidos.

Tabla 1. Búsqueda en bases de datos

Fuente	Encontrado	Relevante	No repetido
Science@direct	80	40	35
Web of Science	34	18	16
IEEE Explore	36	15	15
ACM Digital Library	70	35	33
Total	220	108	99

En total, se encontraron 220 documentos. Estos fueron filtrados a partir de las palabras claves y los resúmenes, así se determinaron los relevantes para la investigación. Por último se revisaron y eliminaron las publicaciones que aparecieron en más de una base de datos, de esta manera se seleccionaron 99 publicaciones que fueron relevantes y a partir de ellas se generó el estado de arte de la problemática a tratar.

2.2 Fase II: Selección de herramientas

En esta fase se realizó un estudio de herramientas y tecnología utilizadas para la generación, adquisición, almacenamiento y análisis de datos tanto en inteligencia de negocios como en Big data. De esta manera se determinó utilizar para el prototipo Microsoft Azure Machine learning para análisis de los datos y para la aplicación web SQL server 2016. En la capa de presentación se usa Google Charts para generar los elementos gráficos del dashboard.

2.3 Fase III: Estudio de plataformas y datasets

Se realizó un estudio de los portales de datos abiertos. Se incluyeron el portal de Estados Unidos, Gran Bretaña y Unión europea por ser los referentes a nivel internacional. Posteriormente se delimitó la búsqueda a los portales de países Latinoamericanos y del Caribe por mostrar un desempeño relativamente positivo comparado con el promedio global. Cuatro países (México, Brasil, Uruguay y Colombia) se encuentran en el top 20 del barómetro de datos abiertos 2016 debido a que han asumido un rol de liderazgo en Latinoamérica, y demuestran que las políticas de datos abiertos que equilibran esfuerzos entre capacidades e implementación pueden generar un impacto positivo. Se decidió usar para el prototipo conjuntos de datos del portal colombiano debido a que es la realidad más cercana para el investigador y dada la conveniencia de contar posteriormente con expertos que permitieron validar la propuesta.

2.4 Fase IV: Desarrollo del metamodelo, pruebas y difusión de resultados

2.4.1 Desarrollo metamodelo

Una vez realizada la base documental y seleccionadas las herramientas y conjuntos de datos, se procedió al desarrollo y elaboración de la propuesta en el siguiente orden:

- Arquitectura general de la propuesta
- Generación del lenguaje de dominio Específico
- Implementación de la propuesta

2.4.2 Pruebas

Para la validación de la propuesta se realizaron pruebas de usabilidad y validación estadística de supuestos e hipótesis con la ayuda de 11 expertos en el ámbito agroindustrial.

2.4.3 Difusion de resultados

La fase de difusión de resultados se realiza de manera continua durante el desarrollo de la tesis mediante la participación en congresos y eventos académicos y la publicación de artículos y libros. Estos resultados de escrutinio público se presentan en la cuarta parte de esta tesis.

2.5 Organización de esta Tesis

El presente documento está estructurado en secciones y capítulos. En este apartado se presenta en forma resumida la organización de esta tesis para facilitar su lectura posterior. La **parte I** titulada “**Planteamiento del Problema**” incluye los capítulos introductorios donde se presentan el problema, la hipótesis y su contexto. En el **capítulo 1**, “**Introducción**”, se presentan la justificación y planteamiento del problema a resolver en esta tesis. Luego se presenta la hipótesis y se establecen los objetivos de la investigación. Finalmente se enumeran brevemente las principales aportaciones y beneficios.

En el **capítulo 2** titulado “**Metodología y Desarrollo de la Investigación**” se presenta información metodológica sobre el desarrollo de esta tesis

incluyendo la metodología y la estructura de esta memoria. En la **parte II** titulada “**Marco teórico**” se presentan los conceptos y generalidades sobre los que se fundamenta la investigación realizada en esta tesis doctoral. Los primeros capítulos de esta sección se centran en el estudio de los datos abiertos, la inteligencia de negocios, big data, la ingeniería dirigida por modelos MDE, la arquitectura dirigida por modelos (MDA), los lenguajes de dominio específico, entre otros conceptos esenciales.

En la **parte III** titulada “**Estado de arte**” se presenta el estudio de las plataformas de datos abiertos e investigaciones relacionadas con la toma de decisiones en el ámbito de la agroindustria. En la **parte IV** “**Solución Propuesta**” se presenta la solución propuesta en esta tesis para el problema planteado al inicio en el capítulo 1. En el **capítulo 8** “**Arquitectura General de la Solución**” se presenta la solución a nivel conceptual y de arquitectura. Este capítulo ofrece una visión concreta de la solución desarrollada en esta tesis. En el **capítulo 9** se presentan las pruebas de validación de la propuesta.

En la **parte V** titulada “**Conclusiones**” se presentan las conclusiones generales que pueden extraerse de esta tesis doctoral (**capítulo 10**). Estas conclusiones incluyen un análisis general de los resultados obtenidos, la verificación del cumplimiento de objetivos, el escrutinio público del trabajo realizado. Posteriormente se presentan las futuras líneas de investigación derivadas del trabajo realizado. En la **parte VI** titulada “**Anexos**” se presentan apéndices varios que enriquecen los contenidos presentados a lo largo de la disertación.

“Aquel que le gusta la practica sin la teoría, es como el marino que navega barcos sin timón ni brújula y nunca sabe

dónde anclar”

Leonardo da Vinci

PARTE II

MARCO TEÓRICO

CAPÍTULO 3 – DATOS ABIERTOS

3.1 Definición

Desde principios de la década de los años noventa los gobiernos han tenido presencia en la web proveyendo distintos servicios para que el ciudadano pueda agilizar sus trámites y así obtener una administración más eficiente, tanto en tiempo como en uso de recursos, esta situación ha derivado a lo que se denomina como e-government o “administración electrónica”, en la cual la información de los servicios y datos que obran en poder de la Administración pública resultan de fácil acceso a los ciudadanos a través de los principales canales de comunicación, como páginas web, dispositivos móviles, etc., creando de esta forma una administración que fija como objetivo principal la transparencia y la objetividad, materializadas a través de distintas políticas.

El gobierno electrónico trae consigo una serie de beneficios especialmente para la ciudadanía, dentro de los que cabe resaltar la eliminación de las barreras de tiempo y espacio, la facilidad en las comunicaciones, el acceso igualitario a la información, el aumento en el intercambio colaborativo entre distintas reparticiones y hasta incremento en la producción de bienes y servicios de valor agregado. En general, la implementación de tecnología de esta índole contribuye a la modernización del estado, ayuda al control interno y externo, aportando a la transparencia del sector público, disminución de costos al compartir recursos, reduce la brecha entre estado y ciudadanía y finalmente promueve la participación de la ciudadanía en los diferentes procesos de toma de decisiones, entre muchos otros [1].

Aquellos gobiernos que divulgan la información del sector público reciben el nombre de Open Government o Gobierno Abierto, los cuales dejan disponible información en internet para ser reutilizada por otros [27]. Estos practican y promueven la transparencia en su gestión, potencian la colaboración y la participación en la puesta en marcha de políticas públicas, además de compartir recursos informativos y datos o registros de información que obran en poder de la Administración Pública y que han sido recabados con distintos objetivos en los distintos procesos administrativos como los referentes a: salud, cartografía, datos meteorológicos, educación, datos bibliográficos, contratación pública, legislación, etc. [28].

Algunas de las principales motivaciones son que el acceso abierto a los datos financiados con fondos públicos proporciona mayores retornos de la inversión pública, puede generar riqueza a través del uso de los productos hacia abajo, proporciona a los responsables de la formulación de políticas los datos necesarios para abordar problemas complejos y puede ayudar a involucrar a la ciudadanía en el análisis de grandes cantidades de conjuntos de datos.

En este sentido, gobierno abierto y gobierno electrónico son conceptos interrelacionados, pero no necesariamente uno sucede junto con el otro. De hecho, existen intensos debates acerca de si el gobierno abierto es resultado de las políticas de gobierno electrónico o si es un proceso anterior a la implementación de las tecnologías de información y comunicación TIC [30].

Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen [31].

La definición de apertura significa:

***Disponibilidad y acceso:** la información debe estar disponible como un todo y a un costo razonable de reproducción, preferiblemente descargándola de internet. Además, la información debe estar disponible en una forma conveniente y modificable.

***Reutilización y redistribución:** los datos deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.

***Participación universal:** todos deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. Restricciones “no comerciales” que prevendrían el uso comercial de los datos; o restricciones de uso para ciertos propósitos (por ejemplo sólo para educación) no son permitidos.

Por otro lado la cadena de valor alrededor de los datos abiertos (figura 2) inicia con la fuente de datos, estos pueden ser originados en organizaciones públicas o privadas. Cuando los datos provienen del gobierno son denominados Open Government Data. El siguiente eslabón en la cadena son los mecanismos técnicos que permiten la publicación y disposición de los datos al alcance de todos. De esta manera los infomediarios crean productos y servicios a partir de dichos datos. De allí surgen aplicaciones para

dispositivos móviles, portales, iniciativas que promueven transparencia, participación, desarrollo e innovación, entre otras. El último eslabón son los usuarios que pueden ser ciudadanos o empresas que utilizan los productos o servicios ofrecidos por los infomediarios.

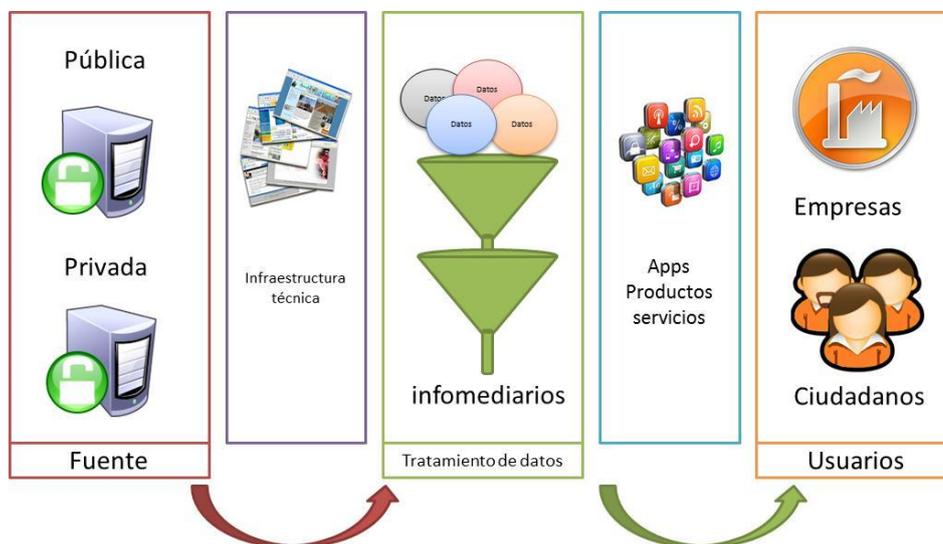


Figura 2. Cadena de valor de los datos abiertos

3.2 Tipos de Datos abiertos

Entre los tipos de datos abiertos, figura 3, se destacan los datos gubernamentales o la Información del Sector Público (ISP), que incluye una amplia gama de datos recopilados o financiados por los gobiernos nacionales, regionales y locales y agencias gubernamentales que pueden recopilarse a propósito (por ejemplo, estadísticas nacionales, meteorológicas, otros datos espaciales) o surgir como una parte integral de la función gubernamental (por ejemplo, registro de empresas, registros judiciales).

Otra área importante es la investigación o los datos científicos, especialmente los que se derivan de la investigación financiada con fondos públicos. También hay una enorme gama de datos del sector privado que podrían estar abiertos para beneficio público y privado (por ejemplo, información de seguimiento de vehículos para la gestión del tráfico y el diseño y desarrollo de la infraestructura, datos de ventas de códigos de barras para la gestión económica como la estimación del índice de precios al consumidor).

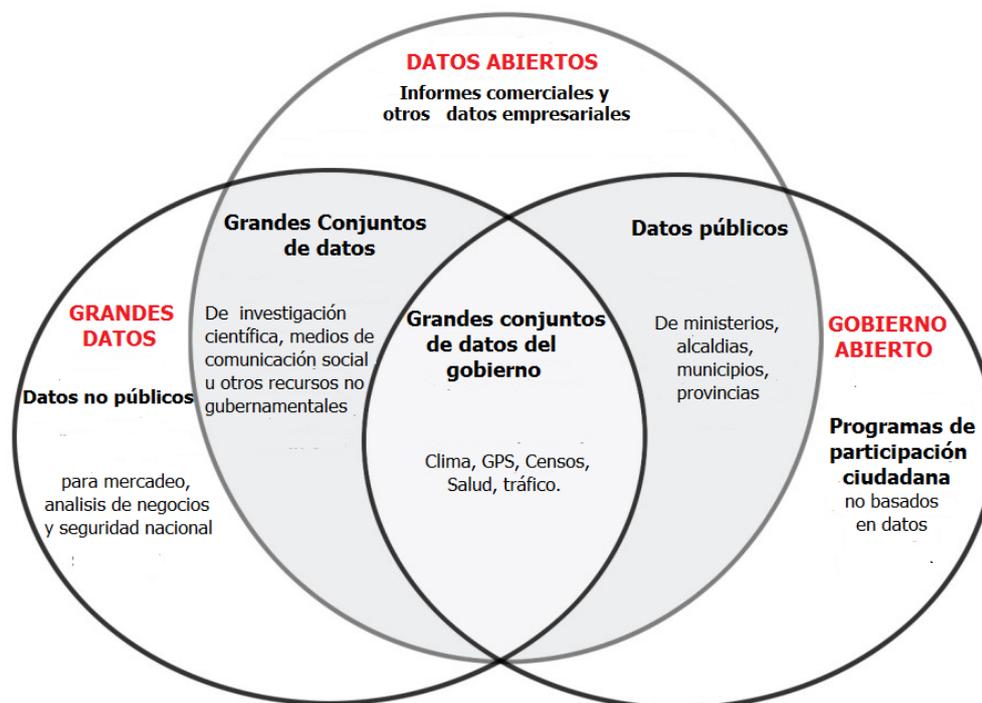


Figura 3. Tipos de datos abiertos.
Fuente [32]

3.3 Actores involucrados

Los gobiernos, las empresas y los individuos necesitan comprender cómo aprovechar los datos abiertos. Todas las partes interesadas gobiernos, organizaciones sin fines de lucro, empresas, individuos (como consumidores y ciudadanos) tienen un papel que desempeñar para maximizar los beneficios de los datos abiertos. Obtener información valiosa de datos abiertos requerirá nuevas reglas y procedimientos y nuevas actitudes, así como inversiones en tecnología y capacidades.

Los **gobiernos** pueden definir el éxito de las iniciativas de datos abiertos dentro de una sociedad, tanto mediante la liberación de datos como la configuración del entorno normativo. Un primer paso importante es establecer prioridades para la liberación de datos basadas en el valor potencial, en lugar de facilitar la "apertura" de los datos para compartirlos. Pueden establecer reglas claras que definan el tipo de datos que deben o no deben ser liberados, con especial énfasis en seguridad, seguridad nacional, privacidad, responsabilidad, derechos de propiedad intelectual y confidencialidad. Como

fuentes primarias de datos abiertos, los gobiernos de todos los niveles pueden ser líderes en el desarrollo de plataformas fáciles de usar para acceder a datos abiertos. Los líderes también pueden buscar asociaciones público-privadas o colaboraciones para apoyar las actividades de datos abiertos.

El gobierno puede dar forma al ambiente legal y económico que maximiza el valor social potencial del uso de datos abiertos, mientras que trata las preocupaciones legítimas de privacidad y propiedad intelectual de individuos y organizaciones. Esto puede incluir políticas sobre quién puede acceder a la información, el tipo de información que se puede recopilar o utilizar en determinadas funciones y protocolos para notificar a las personas cuando se dispone de información sobre ellas. Los reguladores también pueden crear políticas para alentar a las empresas a recopilar y liberar datos (con las protecciones adecuadas).

Los datos abiertos crean oportunidades para que los **empresarios** creen nuevas fuentes de ingresos y aumenten la productividad de sus operaciones. Otros tipos de datos abiertos disponibles por gobiernos, empresas y terceros (a menudo combinados con otros datos) pueden utilizarse para crear productos y servicios innovadores. Un valor particular puede residir en reunir información de sectores dispares de la economía y combinar esa información con datos de propiedad. Las empresas que buscan aprovechar las oportunidades de datos abiertos pueden buscar e inventariar fuentes de datos, desarrollar estrategias para influenciar a otros para que los datos valiosos estén más disponibles, invertir en el desarrollo de herramientas que puedan transformar datos abiertos en productos y servicios y dedicar personal a desarrollar ideas y diseñar productos utilizando datos abiertos.

Los datos abiertos pueden aportar nuevas ideas sobre cómo operan las **compañías** y ayudar a la administración a identificar variaciones innecesarias y otras barreras a la productividad. Las empresas pueden optar por compartir datos de propiedad para crear puntos de referencia que pueden mejorar el rendimiento general de la industria. Las empresas pueden refinar los requisitos del producto y crear nuevos productos y servicios. Las empresas también tendrán que tener estrategias sobre cómo, cuándo y en qué circunstancias abrir sus datos, teniendo en cuenta el impacto potencial de la liberación de sus datos podría tener. Las empresas pueden elegir plataformas para la liberación de datos, participar en la creación de un ecosistema de usuarios de datos y considerar formas de monetizar el valor de los datos.

Además, pueden participar en el establecimiento de normas, incluidos los metadatos, y aportar información sobre los marcos jurídicos emergentes que rigen los datos.

Las empresas también deben evaluar los riesgos potenciales que plantean los datos abiertos, deben ser conscientes de los tipos de datos abiertos que podrían causar daño a la reputación, perjudicar su posición competitiva o perturbar su industria. Ante estos riesgos, las empresas deben participar en los diálogos que establecen normas, desarrollan marcos legales y políticas e informan al público en general sobre temas como propiedad intelectual, privacidad y confidencialidad.

Los **consumidores** pueden capturar una gran parte del valor potencial de usar datos abiertos. Los individuos pueden buscar aplicaciones que usan datos abiertos y proporcionar retroalimentación para mejorar estas herramientas. Si bien los datos abiertos tienen numerosos beneficios potenciales, los consumidores pueden ayudar a salvaguardar sus intereses mediante el monitoreo de políticas y prácticas de privacidad para asegurar que los datos no se usan de tal manera que causan daño social o financiero. Los ciudadanos pueden trabajar con el gobierno para guiar la política de recolección y uso de datos.

Las organizaciones no gubernamentales (ONG) y otras **organizaciones sin fines de lucro** pueden incorporar información como ubicación de los recursos -escuelas, hospitales, caminos- y la calidad de los sistemas de salud, educación y economía para identificar las áreas de mayor necesidad. El hecho de presionar a gobiernos y organizaciones privadas para que divulguen información donde existan vacíos puede conducir a avances críticos.

ONG y organizaciones sin fines de lucro también pueden organizar voluntarios con habilidades relacionadas con los datos que pueden colaborar para crear herramientas útiles. La base de datos OpenStreetMap que se utilizó en Haití, por ejemplo, fue producida por voluntarios en varios continentes. Las organizaciones sin fines de lucro también pueden servir como una fuerza organizadora neutral para alinear las normas internacionales de datos, como la normalización de formatos para metadatos (datos sobre datos) y otros elementos de datos.

3.4 Formatos de Datos Abiertos

Los estándares de datos abiertos son el conjunto de especificaciones, modelos, formatos, protocolos, sintaxis e interfaces de programación que se usan para la publicación y reutilización automática de los datos.

En muchos sitios WEB se encuentra información estadística de interés o datos relevantes disponibles en diversos formatos y catalogados en función de su naturaleza. Estos datos, en el mejor de los casos, son datos dinámicos generados en tiempo real o en su defecto es un equipo de personas quien de forma periódica genera y ofrece esos datos a los ciudadanos (datos estáticos).

Cuanto mejor estructurados y enriquecidos estén los datos, más fácil será reutilizarlos y construir aplicaciones que los traten de manera automática. Por esto hay formatos que se consideran de menos valor que otros. En la tabla 2 se presentan algunos de los formatos estándares usados para datos abiertos.

La información de la web puede presentarse en diferentes formatos, los cuales permitirán o no reutilizar los datos con mayor facilidad para construir productos y servicios. Entre los formatos de datos más adecuados para la reutilización están XML, JSON y el modelo RDF (que permite la consulta de datos utilizando el lenguaje SPARQL), los cuales permiten disfrutar la posibilidad real de creación de servicios de alto valor añadido [33].

El formato JSON o JavaScript Object Notation es un formato ligero para el intercambio de datos basado en la notación literal de objetos de JavaScript, su sintaxis es muy simple, lo que lo hace fácil de leer por cualquier lenguaje de programación, esto significa que los computadores lo pueden procesar más fácilmente a otros formatos como el XML (eXtensible Markup Language), sin embargo, este último es muy usado en el intercambio de datos, ya que brinda grandes oportunidades para mantener la estructura de los datos y la forma cómo son construidos, además de permitir a los desarrolladores escribir partes de la documentación sin interferir en su lectura [33].

SGML son las siglas de Standard Generalized Markup Language o “Estándar de Lenguaje de Marcado Generalizado”, el cual consiste en un sistema para la organización y etiquetado de documentos, normalizado por la Organización Internacional de Estándares (ISO). Sirve para especificar las reglas de etiquetado de documentos y no impone en sí ningún conjunto de

etiquetas en especial. El lenguaje HTML está definido en términos del SGML [34].

El formato XML fue desarrollado por el World Wide Web Consortium (W3C), un consorcio internacional de reconocido prestigio donde las organizaciones miembro, el personal a tiempo completo y el público en general, trabajan conjuntamente para desarrollar estándares Web. Este formato XML es un estándar de creación posterior al SGML, que incorpora un subconjunto de su funcionalidad (suficiente para las necesidades comunes), y resulta más sencillo de implementar pues evita algunas características avanzadas de SGML. Se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas, ya que se puede usar en bases de datos, editores de texto, hojas de cálculo, etc [33], [34].

Really Simple Syndication RSS es un tipo de formato XML para la distribución de contenidos de páginas web. Facilita la publicación de información actualizada a los usuarios suscritos a la fuente RSS sin necesidad de usar un navegador, utilizando un software especializado en este formato

También se encuentran los formatos XLS o Microsoft Office Excel, formato propietario de Microsoft, que muestra la información en celdas organizadas en filas y columnas, y cada celda contiene datos o fórmulas, con referencias relativas o absolutas a otras celdas

SPARQL (Protocol and RDF Query Language), es una tecnología de consulta de información desde bases de datos y otros orígenes de datos en sus estados primitivos a través de la Web. Se compone de un lenguaje de consulta estandarizado y de un protocolo con el que ofrecer un servicio Web estándar, permitiendo realizar consultas a diversas fuentes de datos que almacenan los mismos nativamente en RDF o los presentan como tal.

El formato Resource Description Framework (RDF), es una infraestructura para describir semánticamente recursos, es decir, dotar de sentido a lo que representamos para que las máquinas lo comprendan, se puede representar en distintos formatos y hace posible representar datos que provengan de diversas fuentes; fomenta el uso de URLs (localizadores uniformes de recursos) como identificadores de tripletas (sujeto-propiedad-objeto), lo cual provee una forma conveniente de interconectar directamente iniciativas datos abiertos en la Web y los datos de este tipo pueden ser almacenados en XML, JSON, entre otros [33].

Otro tipo de documentos en formato abierto, son los llamados ficheros CVS o Comma-separated values, los cuales representan de manera sencilla, datos en formato de tabla, separando las columnas por comas (o punto y coma) y las filas por saltos de línea; los PDF o Portable Document Format, son otro tipo de especificación abierta y multiplataforma (se puede presentar en los principales sistemas operativos: Windows, Mac o Unix/Linux). Fue desarrollado por Adobe Systems y permite almacenar documentos de tipo compuesto (imagen vectorial, mapa de bits y texto), es decir, fue ideado para documentos susceptibles de ser impresos, ya que especifica toda la información necesaria para la presentación final del documento, determinando todos los detalles de cómo va a quedar, no requiriéndose procesos anteriores de ajuste ni de maquetación; para éste se han desarrollado herramientas de software libre como la suite ofimática de OpenOffice.org y el procesador de textos de LaTeX, que permiten crear, visualizar o modificar documentos .

En cuanto a los formatos utilizados para presentar información geográfica se encuentran los SHP Shapefile que es un formato propietario estándar de datos espaciales, desarrollado por la compañía ESRI, que almacena tanto la geometría como la información alfanumérica y en general puede ser reutilizable por Sistemas de Información Geográfica (SIG). Los formatos .sbn y .sbx almacenan el índice espacial de las entidades en las aplicaciones SIG, los .fbn y .fbx almacenan el índice espacial de las entidades para los shapefiles que son inalterables (solo lectura) en las aplicaciones SIG; los .prj guardan la información referida al sistema de coordenadas en las aplicaciones SIG, entre otros [35].

Para mapas interactivos, se utilizan los servicios wms (Web Map Service), el cual produce mapas a partir de información geográfica vectorial presentando la información como imágenes digitales susceptibles de ser visualizadas en pantalla en formatos PNG (Portable Network Graphics), GIF o JPEG y ocasionalmente, se representan como información vectorial en formato Scalable Vector Graphics (SVG), es decir en gráficos vectoriales bidimensionales tanto estáticos como animados, o Web Computer Graphics Metafile (WebCGM), óptimo para diagramas complejos en que se requiere hacer zoom o visualizar por capas. Los mapas visualizados pueden superponerse unos a otros, siempre y cuando los parámetros geográficos y el tamaño de salida sean los mismos.

Para bases de datos, se encuentran los tipos de formatos mdb, accdb, dbf, los cuales son legibles por aplicaciones de bases de datos, dbf es además el tipo de archivo que almacena la información de los atributos de los objetos en aplicaciones SIG.

Aunque oficialmente no existe una manera de clasificar los conjuntos de datos dispuestos en las soluciones WEB, si existe una propuesta formulada por Tim Berners-Lee [36], que clasifica a los conjuntos de datos puntuándolos con un número de estrellas entre 1 y 5, la tabla 2 muestra algunas de las ventajas y desventajas.

Tabla 2. Clasificación de los conjuntos de datos

Estrellas	Formato	Ventajas	Desventajas
★	PDF, texto, imagenes	No requiere ningún tipo de transformación previa, pueden ser publicados de una manera sencilla.	Los formatos son muy difíciles de manipular y reutilizar.
★★	Ficheros Excel, XLS		Para acceder a los datos se requiere de herramientas no abiertas.
★★★	CSV, XML o RDF	El acceso y la manipulación de los datos pueden hacerse de manera sencilla.	Se requiere la transformación previa.
★★★★	URLs	Facilita la labor de reutilización de los datos y actualización de los mismos.	Mayor esfuerzo de estructuración, separación de los datos y asignación de las URLs.
★★★★★	RDF, basado en XML Linked data	Enriquecimiento semántico de los conjuntos de datos y estandarización del vocabulario utilizado en los diferentes contextos.	

La esencia de los datos (o códigos) compartidos es que una parte del material abierto pueda a partir de ahí ser mezclado con otro material abierto. Esta interoperabilidad es absolutamente fundamental para entender los principales beneficios prácticos de la apertura: el incremento dramático de la habilidad de combinar distintas bases de datos o conjuntos de datos y así desarrollar más y mejores productos y servicios [31]

3.5 Catálogos de datos abiertos

Un catálogo de datos es la entrada organizada a los datasets o conjuntos de datos publicados, está enfocado en facilitar el acceso simple, abierto, y sin restricciones de uso de los datos. Se puede definir como la organización de los metadatos de los datos del repositorio, más metadatos y enlaces a otros

datos relacionados relevantes para los usuarios (entendiendo los metadatos como datos que describen otros datos, permitiendo organizarlos, clasificarlos y relacionarlos) [37].

Estos catálogos de datos, se han realizado en plataformas como Junar, Ckan, Dkan, entre otras, las cuales permiten a los desarrolladores crear sus páginas web donde los usuarios puedan buscar, publicar y reutilizar los datos que deseen; fomentando la creación de nuevas aplicaciones útiles para la sociedad.

La plataforma Junar ha sido utilizada por los gobiernos de Chile, Sacramento USA, Palo Alto USA, Lima Perú, entre otros; y CKAN, por su parte, ha sido utilizada como plataforma de Datos Abiertos del Reino Unido (data.gov.uk), de la Unión europea publicdata.eu, del portal de Helsinki Finlandia, entre otros.

Ckan es un software gratis y de licencia abierta que permite mantener y publicar colecciones de datos, usado por gobiernos nacionales y locales, instituciones de investigación y otras organizaciones que coleccionan gran cantidad de datos [31]. Con esta herramienta, los usuarios pueden buscar y encontrar los datos que necesitan y previsualizarlos usando mapas, gráficos y tablas.

Este software fue desarrollado por la Open Knowledge Foundation, una Iniciativa europea para generar nuevo conocimiento mediante el tratamiento de datos liberados o disponibles en la web, y ha sido muy útil para países que han desarrollado sus portales de datos abiertos como parte de los compromisos establecidos en la Alianza para el Gobierno Abierto (Open Government Partnership OGP), ya que esta alianza hace esfuerzos para promover la adopción de políticas de participación ciudadana, lucha contra la corrupción, empoderamiento ciudadano y gobierno electrónico para cada uno de sus Estados miembros. La OGP lidera varios proyectos aparte de CKAN como “The open definition”, que sienta las bases teóricas o “Data commons” que ofrece soluciones legales [38], [39].

En cuanto a DKAN, es una plataforma basada en Drupal, un sistema para manejo de datos (Data Management System -DMS-), contiene un conjunto completo de funciones de catalogación de datos, edición y visualización, la cual puede ser personalizada y extendida fácilmente por los usuarios, quienes pueden explorar datos, buscarlos y describirlos. El modelamiento de datos de DKAN es muy parecido al de CKAN, permitiendo describir los

datasets, las fuentes, crear grupos y etiquetas. Además, esta plataforma es totalmente abierta, lo cual permite acelerar el desarrollo, reducir costos y eliminar dependencia de proveedores y empresas [40].

3.5 Beneficios de los Datos abiertos

Los datos gubernamentales y los datos científicos desempeñan un papel importante en las economías basadas en el conocimiento y, con muchos de los beneficios aún por alcanzar, los datos más abiertos podrían contribuir al crecimiento económico. El principal desafío es que los datos abiertos no tienen valor en sí mismos; sólo se convierten en valiosos cuando se utilizan. Uno de los principales beneficios de la apertura de un sistema es la capacidad de aprovechar la inteligencia colectiva del público. La idea clave es que, en las circunstancias adecuadas, los grupos pueden generar mejores alternativas y tomar mejores decisiones que las personas más inteligentes pueden hacer por sí solas.

Los beneficios se agruparon en (1) beneficios políticos y sociales, (2) económicos y (3) operacionales y técnicos, se muestran en la figura 4.



Figura 4. Beneficios de los datos abiertos.
Fuente[41]-[43]

Desde otro punto de vista se conocen algunos estudios que exploran el valor económico de los datos abiertos, se han centrado en diferentes dimensiones de valor y han adoptado diferentes enfoques y métodos económicos. La mayoría se centran en la información del sector público, aunque algunos exploran datos de investigación y datos privados.

Adoptando un enfoque de arriba hacia abajo, [44] combinó medidas del costo de inversión (es decir, el monto gastado en la recolección / generación de la Información del Sector Público –ISP-) y el gasto por usuarios y reutilizadores. Luego, para los usuarios finales, el valor estimado como gasto o, cuando la ISP estaba disponible libremente, como el costo de inversión de su recolección / generación; y para los intermediarios que agregan valor a los datos, atribuyen una proporción de su valor agregado al ISP utilizado. De esta manera se estimó la inversión en unos 9.500 millones de euros al año en 1999 y el valor económico de la ISP, en la Unión Europea, a unos 68.000 millones de euros, aproximadamente el 1,4% del PIB representando un rendimiento de la inversión de siete veces.

Por otro lado utilizando una gran encuesta de productores y usuarios de ISP, [45] intentó estimar el tamaño del mercado ISP en Europa. En el estudio MEPSIR, la demanda y el desempeño económico se midieron en una extensa encuesta solicitando directamente a los titulares de contenido público y a los reutilizadores datos económicos clave, como el volumen de negocios total, el número total de funcionarios frente al número del personal dedicado a la manipulación de la ISP, y las estimaciones del mercado interno para un tipo particular de ISP. Sobre la base de las estimaciones de los reutilizadores, ponen el mercado total de la ISP en la UE más Noruega en unos 27.000 millones de euros (aproximadamente el 0,25% del PIB agregado).

DotEcon produjo un informe en el que desarrollaron un enfoque de base para estimar el valor económico de los productos y servicios de la ISP en el Reino Unido, viendo el valor económico neto de ISP como la disposición a pagar por ISP menos el costo de su suministro. También examinaron los costos (perjuicio) de los obstáculos al uso, entre ellos: precios indebidamente altos, distorsión de la competencia y falta de explotación de la ISP [46]. Los resultados indican que el valor neto de la ISP en el Reino Unido se situaba en torno a 590 millones de libras esterlinas al año en 2005. Los costes de los tres

tipos de perjuicios se estimaron en 20 millones de libras esterlinas, 140 millones de libras esterlinas por restricción de la competencia y 360 millones de libras esterlinas por no explotar ISP, lo que sugiere que el valor de la ISP podría duplicarse resolviendo los problemas identificados

Al revisar la literatura sobre el tamaño y el crecimiento del mercado europeo de la ISP para la Union Europea, en [5] se estimó que el mercado relacionado con la ISP directa de la UE era del orden de 28.000 millones de euros en 2008. Todos los estudios muestran el crecimiento relativamente rápido de los mercados relacionados con la ISP.

Basándose en el enfoque ascendente desarrollado por DotEcon, [41] estimó que el valor de la información del sector público para los consumidores, las empresas y el sector público era aproximadamente de 1.800 millones de libras esterlinas en los precios de 2011, pero sugirió que el uso y la reutilización de la información del sector público tiene impactos mucho más profundos que afectan a todas las áreas de la sociedad más allá del cliente directo. Sobre la base de suposiciones conservadoras, estimaron que podría superar los 5.000 millones de libras esterlinas a precios de 2011.

El estudio [47] sugiere que los datos abiertos tienen el potencial de desbloquear aproximadamente USD 3.2 trillones en valor adicional anualmente a través de los siete dominios de la economía que el estudio examinó (educación, transporte, electricidad, petróleo y gas, salud, productos de consumo y productos financieros). El valor puede surgir de varias maneras, incluyendo equipar a los trabajadores con las habilidades para aumentar la productividad, permitiendo a los vendedores poblaciones de micro-segmentos con más éxito, e impulsando el desempeño compartiendo puntos de referencia, datos de mercado e información de mejores prácticas. Los consumidores pueden ganar ahorrando dinero a través de una mayor transparencia de precios y usando más información para tomar decisiones.

En [48] se propone un marco para evaluar el impacto de la liberación de datos abiertos mediante la aplicación del enfoque de retorno social de la inversión. Allí se evidencian posibles impactos sociales como mayor calidad educativa, mas igualdad de oportunidades para niños, más intercambio de know-how entre hospitales y médicos, más ciudadanos que utilizan transporte, menos corrupción y mejor comprensión financiera de los ciudadanos, entre otras.

Específicamente para el sector privado [49], [50] encontraron como beneficios del uso de datos abiertos acciones colaborativas y mayor dialogo y participacion de los clientes. Las empresas también pueden aprovechar el conocimiento de los actores privados y las comunidades y aumentar la transparencia en su entorno de trabajo y procesos. Las pequeñas y medianas empresas (PYME), los empresarios y otros negocios, no sólo pueden encontrar un beneficio económico mediante el uso de datos gubernamentales, sino que gracias a servicios innovadores pueden repercutir indirectamente en la creación de empleo. Aunque el solo el acceso a los datos no proporciona ventaja competitiva, las empresas pueden innovar sobre los datos disponibles para crear servicios de valor añadido [51].

El uso y reutilización de la información del sector público tiene impactos mucho más profundos que afectan a todas las áreas de la sociedad. La información del sector público y privado puede actuar como catalizador en el proceso de generar innovación, optimizar los costos y ayudar a los funcionarios, líderes empresariales y ciudadanos comunes a tomar mejores decisiones gracias a información de calidad.

3.6 Barreras de los Datos abiertos

Cada vez más datos gubernamentales se están poniendo a disposición del público, pero no siempre en un formato abierto, dado los frecuentes problemas con las licencias, los formatos, las descargas a granel o la disponibilidad gratuita. Desafortunadamente muchos gobiernos con buenas intenciones piensan que están haciendo los datos abiertos pero no cumplen los requisitos clave para la verdadera facilidad de acceso y reutilización de los datos [52].

Si bien es cierto que gran parte de los resultados de los datos abiertos prometen ser benéficos en escenarios diversos, también es cierto que depende en gran medida de la estructura política de cada estado, es decir, es inminente que puedan presentarse problemas o barreras frente a la apertura de datos, estas pueden ser clasificadas en institucionales, de complejidad de las tareras, de uso y participacion, legislación, de calidad y técnicas. En la figura 5 se presentan las barreras a nivel institucional identificadas en la literatura.



Figura 5. Barreras ámbito institucional.
Fuente: autor a partir de [1], [23], [24], [43], [53]–[58]

La ISP se deposita en varias fuentes fragmentadas, por lo que los usuarios tienen dificultades para encontrar los datos, se requiere la transformación de las organizaciones del sector público para garantizar que no abusen de su poder de monopolio como únicos productores de determinados tipos de datos y prioricen las necesidades de los ciudadanos y no las propias. Se requiere identificar un terreno común donde las partes implicadas puedan centrarse en la construcción de información de valor que beneficie a todos.



Figura 6. Barreras ámbito complejidad de las tareas.
Fuente: autor a partir de [1], [23], [24], [43], [53]–[58]

En cuanto al ámbito de la complejidad de las tareas, figura 6, un tema subestimado parece ser la disponibilidad de todo tipo de capacidades y niveles de conocimiento de los usuarios para el uso de datos complejos y sofisticados, no sirve de nada disponer de conjuntos de datos en portales si los usuarios que podrían beneficiarse de su reutilización no cuentan con las herramientas para sacarles provecho. Adicionalmente en muchas ocasiones los datos no son fáciles de descubrir, están duplicados y son poco interoperables. Todo esto lleva a los usuarios a limpiar y tratar los conjuntos de datos antes de poder reutilizarlos.



Figura 7. Barreras ámbito uso y participación.
Fuente: autor a partir de [1], [23], [24], [43], [53]-[58]

En la figura 7, se presentan las barreras en cuanto al uso y participación. Los usuarios de datos encuentran difícil interpretarlos, porque no están familiarizados con las definiciones y categorías que se utilizan, por lo tanto, es difícil la comparación, el enlace y la reutilización de datos. Como no todos los usuarios están en capacidad de usar los datos abiertos se corre el riesgo de limitarlos a ciertos grupos y de esta manera lejos de aprovechar su valor se estaría contribuyendo a la brecha digital.



Figura 8. Barreras ámbito legislación.
Fuente: autor a partir de [1], [23], [24], [43], [53]-[58]

En muchos países ya existen iniciativas de datos abiertos pero carecen de un marco legal que reglamente su uso. Por ello la confiabilidad de la fuente, la procedencia de los datos y los aspectos legales de su consumo son aspectos preocupantes que no han sido abordados. Las partes interesadas están expuestas a la violación de la privacidad y a ser legalmente responsables cuando los datos abiertos sean mal utilizados. Otras barreras legales se muestran en la figura 8.



Figura 9. Barreras ámbito calidad de la información.
Fuente: autor a partir de [1], [23], [24], [43], [53]-[58]

En cuanto a la calidad de los datos, figura 9, se encuentra la carencia de metadatos que permitan conocer su calidad y la forma en que fueron recopilados. En muchos casos se requiere una carga de trabajo sustancial para limpiar los datos y para hacerlos comprensibles. Los portales estan llenos de conjuntos de datos obsoletos o con información poco relevante. Esto es una gran barrera pues disponer demasiada informacion no necesariamente resulta en decisiones mejores, más democráticas o más racionales. Más información puede resultar en menos comprensión, más confusión y menos confianza [59].



Figura 10. Barreras ámbito técnico.
Fuente: autor a partir de [1], [23], [24], [43], [53]-[58]

Por último en el ámbito técnico, figura 10, se encuentran barreras como la falta de soporte para los usuarios inexpertos y las falencias en infraestructura tecnológica para el almacenamiento, descarga, disposición y uso de los datos. Ser capaz de utilizar datos y encontrar patrones y tendencias en grandes cantidades de datos también sigue siendo un reto importante [55] pues esta capacidad es necesaria, al igual que el contenido automatizado y el análisis de temas. Los algoritmos de aprendizaje automático a menudo pueden analizar un conjunto de datos e inferir reglas para clasificar y agrupar elementos de datos y de esta manera se puede sacar provecho a los datos. La

automatización y clasificación de datos puede ayudar a simplificar búsquedas, analizar relaciones y extrapolar tendencias [43].

Los gobiernos y la sociedad civil deberían trabajar juntos para identificar las prioridades sociales más apremiantes y las necesidades de datos relacionadas con éstas. El éxito de los sistemas de datos abiertos implica tener en cuenta que se requiere la mejora de la calidad de la información gubernamental, la creación e institucionalización de una cultura de gobierno abierto y la provisión de herramientas e instrumentos para utilizar los datos. Es necesaria una infraestructura que ayude a los usuarios a darle sentido a los datos, y son necesarias medidas institucionales para garantizar la participación pública. En estas condiciones, los datos abiertos pueden ir más allá del nivel actual de participación ciudadana y podrían dar lugar a un diálogo continuo entre los gobiernos y sus electores, aprovechando la inteligencia colectiva del público [43].

El movimiento de datos abiertos está en un punto de inflexión. Los Objetivos de Desarrollo Sostenible de las Naciones Unidas han creado impulso para una revolución de datos, y los compromisos e iniciativas de políticas de datos abiertos se han extendido rápidamente. Sin embargo, la implementación y el impacto están rezagados, creando el riesgo de que el movimiento de datos abiertos se desvanezca en una ciudad fantasma de portales abandonados y aplicaciones olvidadas [52].

Además, es posible encontrar rechazo y resistencia en todos los niveles de la sociedad y al interior de entes gubernamentales debido a los cambios que implica la puesta en marcha de datos abiertos, la forma de pensar, cambios en los sistemas de información y no menos importante, sensación de pérdida de poder por parte del gobierno, ya que al manipular información que resulta inasequible al ciudadano del común se puede tener la falsa percepción de pérdida de poder en el caso de gobiernos menos versátiles [1].

Las iniciativas de datos abiertos prometen un potencial con múltiples beneficios, por ello los entes públicos y privados han mostrado su interés en promover su disponibilidad al público sin embargo no se deben dejar de lado las barreras que limitan su uso y de esta manera garantizar una verdadera reutilización de la información.

CAPÍTULO 4 – INTELIGENCIA DE NEGOCIOS Y BIG DATA

Aunque la Inteligencia de negocios (Business Intelligence- BI) y los grandes datos (Big Data) son dos tecnologías que sirven para analizar datos y así ayudar a las empresas en el proceso de toma de decisiones, difieren tanto en el modo en el que lo hacen, como en el tipo de datos que analizan. A continuación se presentan las generalidades de estas dos tecnologías.

4.1 Inteligencia de negocios

La inteligencia de negocios es un enfoque estratégico para orientar sistemáticamente el seguimiento, la comunicación y la transformación relacionada al débil conocimiento de la información procesable en la cual se basa la toma de decisiones [60]. Es una disciplina que integra información proveniente de diversas fuentes u orígenes facilitando su almacenamiento, abordaje, selección, y el tratamiento de datos históricos.

Con el fin de entregar la información útil para la toma de decisiones, la inteligencia de negocios es la tecnología clave para que los usuarios extraigan eficientemente información valiosa desde los océanos de datos. El concepto de BI fue introducido en primer lugar por Garter Group que lo define como un término general que incluye las aplicaciones, la infraestructura, las herramientas y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento [61].

4.1.1 Arquitectura típica

Una arquitectura típica de BI se muestra en la Figura 11. Los datos sobre los que se realizan las tareas de BI a menudo provienen de diversas fuentes, normalmente de múltiples bases de datos operacionales entre departamentos dentro de la organización, así como proveedores externos. Diferentes fuentes contienen datos de calidad variable, utilizan representaciones inconsistentes, códigos y formatos, que tienen que ser conciliados. Por lo tanto, los problemas de integrar, limpiar y estandarizar los datos en preparación para las tareas de BI pueden ser bastante desafiantes.

Estas tecnologías de back-end para preparar los datos para BI se denominan herramientas Extract-Transform-Load (ETL). Cada vez más, es necesario apoyar las tareas de BI en tiempo casi real, es decir, tomar decisiones empresariales basadas en los propios datos operativos. Motores especializados llamados motores de procesamiento de eventos complejos (CEP) han surgido para apoyar estos escenarios [62].

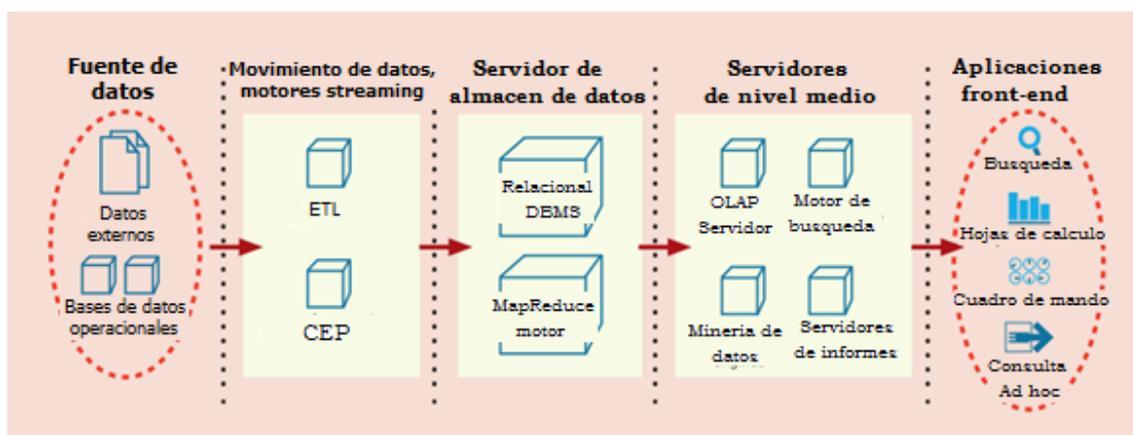


Figura 11. Arquitectura típica de inteligencia de negocios.
Adaptado de [63]

Los datos sobre los que se realizan las tareas de BI normalmente se cargan en un repositorio denominado almacén de datos (data warehouse) que es administrado por uno o más servidores. Una elección popular de motores para almacenar y consultar datos de almacén son los sistemas de gestión de base de datos relacional (RDBMS). Se han desarrollado varias estructuras de datos, optimizaciones y técnicas de procesamiento de consultas, principalmente para ejecutar consultas SQL complejas sobre grandes volúmenes de datos.

A medida que la cantidad de datos aumenta, se hacen necesarias plataformas de bajo costo que pueden soportar mucho más volumen de datos que el tradicionalmente manejado por RDBMS [64]. Por ello los motores basados en el paradigma MapReduce ahora se están apuntando para el análisis de la empresa.

Los servidores de almacén de datos se complementan con un conjunto de servidores de nivel medio que proporcionan funcionalidad especializada para diferentes escenarios de BI. Los servidores de procesamiento analítico en línea (OLAP) exponen eficientemente la vista multidimensional de los datos a aplicaciones o usuarios y permiten las operaciones de BI comunes, como el filtrado, la agregación, la clasificación y el pivoteo [65]. Los servidores de

informes permiten la definición, la ejecución eficiente y la prestación de informes, por ejemplo, reportar las ventas totales por región para un año y compararlas con las ventas del año anterior.

Los motores de búsqueda de empresas soportan el paradigma de búsqueda de palabras clave sobre el texto y los datos estructurados en el almacén (por ejemplo, buscan mensajes de correo electrónico). Los motores de minería de datos permiten un análisis en profundidad de los datos que va mucho más allá de lo que ofrece OLAP o servidores de informes, y proporciona la capacidad de construir modelos predictivos para ayudar a responder a preguntas específicas [66].

Existen varias aplicaciones de front-end populares a través de las cuales los usuarios realizan tareas de BI: hojas de cálculo, portales empresariales para búsqueda, aplicaciones de gestión de desempeño que permiten a los responsables de tomar decisiones, identificar los indicadores clave de rendimiento de la empresa utilizando paneles visuales, plantear consultas ad hoc, visores para modelos de minería de datos, entre otros. La visualización rápida y ad hoc de datos puede permitir la exploración dinámica de patrones, valores atípicos y ayudar a cubrir hechos relevantes para BI.

Además, existen otras tecnologías de BI como la analítica web [63], que permite comprender cómo interactúan los visitantes del sitio Web de una empresa con las páginas y la gestión de relaciones con el cliente (CRM). CRM puede proporcionar funcionalidad para segmentar a los clientes en los más probables y menos propensos a recomprar un producto en particular[67]. Otra área incipiente pero importante es el BI móvil [68] que presenta oportunidades para permitir nuevas y ricas aplicaciones de BI para trabajadores del conocimiento en dispositivos móviles.

4.1.2 Cuadrante magico BI Gartner 2017

Un mercado apretado con nuevos competidores, la evolución rápida y la innovación constante crea un ambiente difícil para que los vendedores de soluciones de BI diferencien sus ofertas de la competencia. La lista de proveedores se extiende desde los grandes actores de la tecnología, tanto los nuevos al espacio como los jugadores de larga duración que tratan de reinventarse para recuperar relevancia, a las startups respaldadas por enormes cantidades de capital de riesgo de empresas de capital privado.

El cuadrante Mágico de Gartner proporciona un posicionamiento gráfico competitivo de cuatro tipos de proveedores de tecnología, en mercados donde el crecimiento es alto y la diferenciación de proveedores es distinta:

Los líderes ejecutan bien contra su visión actual y están bien posicionados para mañana.

Los visionarios entienden dónde va el mercado o tienen una visión para cambiar las reglas del mercado, pero aún no ejecutan bien.

Los jugadores de nicho se enfocan con éxito en un segmento pequeño, o no se focalizan y no innovan o superan a otros.

Los desafiantes ejecutan bien hoy o pueden dominar un segmento grande, pero no demuestran una comprensión de la dirección del mercado.

Para 2017, los líderes tradicionales son relegados al cuadrante de “visionarios”. Ni SAP BO, ni IBM Cognos, ni Microstrategy, ni SAS, ni Oracle OBI están en el cuadrante de líderes. Los nuevos líderes son Tableau, Qlik y Microsoft Power BI [69]. A continuación se describen brevemente.



Figura 12. Cuadrante mágico para plataformas analíticas y BI.

Fuente [69]

Microsoft

Microsoft ofrece una amplia gama de capacidades de BI y análisis, tanto locales como en la nube de Microsoft Azure. Microsoft Power BI es el foco de este Cuadrante Mágico ofreciendo BI basada en la nube con una nueva interfaz de escritorio. Power BI ofrece preparación de datos, descubrimiento de datos y tableros interactivos a través de una sola herramienta de diseño. Power BI consta de una aplicación de escritorio de Windows denominada Power BI Desktop, un servicio SaaS (software como servicio) en línea denominado servicio Power BI, y aplicaciones móviles de Power BI disponibles para teléfonos y tabletas Windows, así como para dispositivos iOS y Android [70]

Microsoft ha bajado sustancialmente el precio de Power BI, lo que la convierte en una de las soluciones de menor precio en el mercado hoy en día, particularmente de los proveedores más grandes. El precio más bajo, además de sustanciales mejoras en los productos, explica la fuerte aceptación. El flujo de trabajo habitual de Power BI comienza en Power BI Desktop, figura 13, donde se integran los datos y se crea un informe, luego, ese informe se publica en el servicio Power BI donde se crean nuevas visualizaciones y después se comparte para que los usuarios de las aplicaciones de Power BI Mobile puedan compartir y usar la información.

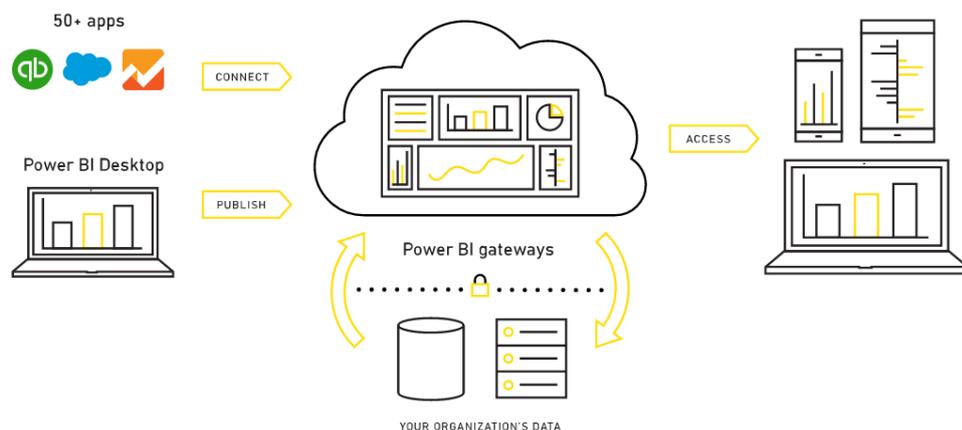


Figura 13. Power BI service.

Fuente [71]

El modelo de entrega basado en la nube de Microsoft y los bajos precios por usuario, además de la facilidad de uso para los usuarios empresariales y

la disponibilidad de recursos especializados fueron los motivos por los cuales los clientes lo destacan.

Tableau

Tableau ofrece productos de descubrimiento de datos altamente interactivos e intuitivos que permiten a los usuarios de negocios acceder fácilmente, preparar y analizar sus datos sin necesidad de codificación. Desde su creación, Tableau se ha centrado en mejorar la experiencia analítica de flujo de trabajo para los usuarios, siendo la facilidad de uso el principal objetivo de gran parte de sus esfuerzos de desarrollo de productos. Una de las virtudes principales de Tableau es su versatilidad, en términos de opciones de despliegue a través de la nube y en el local.

Tableau ofrece una amplia gama de opciones de aprendizaje -incluyendo tutoriales en línea, seminarios en línea y capacitación práctica en el aula- para educar y capacitar a sus usuarios[72]. Los puntos fuertes del producto de Tableau continúan siendo su diversa gama de conectividad de origen de datos, la cual está en constante expansión, así como sus capacidades de visualización y exploración interactivas[73]. Esta combinación cumple la misión de Tableau de ayudar a las personas a ver y comprender sus datos, permitiendo un acceso rápido a prácticamente cualquier fuente de datos, a la que los usuarios no técnicos pueden comenzar a interactuar de inmediato con una interfaz visual intuitiva.

Qlik

Qlik ofrece información y análisis de datos gobernados a través de sus dos productos principales: QlikView y Qlik Sense. Su motor de memoria y análisis asociativo permiten a los usuarios ver patrones en los datos. Qlik está altamente calificado para la facilidad de uso, la complejidad del análisis y los beneficios del negocio [74].

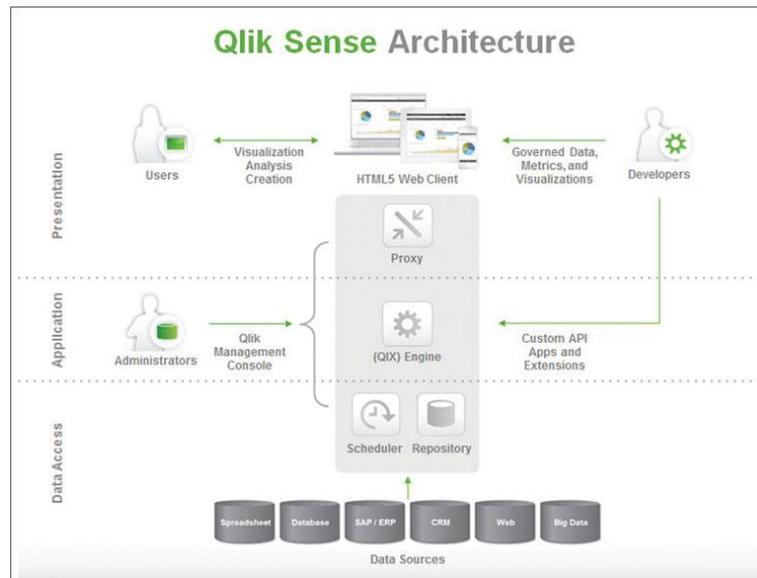


Figura 14. Arquitectura Qlik Sense.
Fuente [74]

En comparación con sus principales competidores, Tableau y Microsoft, Qlik obtiene puntuaciones significativamente más altas en la complejidad del análisis, esto debido a su mayor capacidad para soportar múltiples fuentes de datos, un motor de cálculo robusto y filtrado y búsqueda asociativos[75].

4.1.3 Beneficios de la BI

En la literatura se encuentra estudios que argumentan que el uso de los sistemas de inteligencia de negocios pueden traer numerosos beneficios. La BI puede ofrecer ciertas ventajas competitivas a las empresas, ya que generalmente proporciona una mayor funcionalidad en términos de acceso y análisis en comparación con los sistemas de planificación de recursos empresariales (ERP)[76]. En cuanto a la gestión de la relación con los clientes (CRM), el procesamiento analítico de la información sobre mismos y su comportamiento permite optimizar la gestión de las relaciones con los clientes maximizando su satisfacción y aumentando su lealtad y rentabilidad.

Otros beneficios son la reducción de costos; aumento de los ingresos; mejora de la satisfacción del cliente; y mejora la comunicación dentro de la empresa [77]–[80]. La información se considera a menudo como el segundo recurso mas importante que una empresa tiene junto a su gente. Así que cuando la empresa puede tomar decisiones basadas en nformación precisa puede mejorar su rendimiento.

En [81] se mencionan otros beneficios de la adopción de soluciones BI.

- Las empresas pueden identificar los clientes rentables y las razones de la lealtad de estos clientes, así como identificar futuros clientes potenciales.

- Analizar datos de flujo de clics para mejorar las estrategias de comercio electrónico.

- Determinar qué combinaciones de productos y líneas de servicio comprar y cuándo.

- Detectar y disuadir conductas fraudulentas, como los picos de uso cuando las tarjetas de crédito son robadas.

- Mejorar el proceso de creación de informes para datos sofisticados sin la necesidad de Tecnologías analíticas. Además, el acceso remoto a cualquier dato deseado a través de un teléfono inteligente[82].

- Detección temprana de riesgos para identificar oportunidades de negocio antes que los competidores[83].

La información es la clave y las organizaciones saben muy bien que la información vital para la toma de decisiones está en sus bases de datos. Por lo tanto, la gestión eficiente de la información, así como, la inteligencia de negocios permite ampliar la visión estratégica, reducir el riesgo y la incertidumbre en la toma de decisiones empresariales y construir ventajas competitivas de largo plazo.

4.2 Big Data

El crecimiento en el volumen de datos generados por diferentes sistemas y actividades cotidianas en la sociedad ha creado la necesidad de modificar, optimizar y generar métodos y modelos de almacenamiento y tratamiento de datos que suplan las falencias que presentan las bases de datos y los sistemas de gestión de datos tradicionales. Así "Datos grandes" (Big data) se refiere a conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas típicas de software de base de datos para capturar, almacenar, administrar y analizar [84].

4.2.1 Cadena de valor

La cadena de valor del big data puede ser dividida en cuatro fases: generación de datos, adquisición de datos, almacenamiento de datos y análisis de datos [85]. La primera fase se entiende como un proceso propio de diversas actividades de la sociedad, en estas se genera una cantidad inmensa de datos, que, según su naturaleza, puede estar almacenada y estructurada o puede corresponder a datos sin ninguna estructura pero con características de gran valor. En la segunda fase, se incluye la recopilación, transmisión y pre-procesamiento de los datos. Muchos conjuntos de datos presentan redundancia o datos inútiles y si no se tratan pueden incrementar el espacio de almacenamiento innecesariamente y afectar los resultados de la fase de análisis.

La fase de almacenamiento de Big Data ha creado la necesidad de generar nuevas estrategias que permitan afrontar los tipos de datos que no se pueden gestionar con un sistema de gestión de bases de datos relacionales. Surgen entonces, tecnologías de almacenamiento de datos masivos como almacenamiento con conexión directa y el almacenamiento en red, también diferentes motores NoSQL. Finalmente, la fase de análisis debe atender a la necesidad de extraer rápidamente información desde los datos masivos para poder generar valor en las organizaciones y facilitar procesos de toma de decisiones, se requiere de tecnologías que faciliten incluso el análisis en tiempo real.

Generación de datos

El origen de los datos esta estrechamente relacionado con la vida cotidiana de las personas. Por ejemplo los datos de Internet pueden no tener valor individualmente, pero a través de la explotación de grandes datos acumulados se puede identificar información útil como hábitos y aficiones de los usuarios. En la actualidad, las principales fuentes de grandes datos son la información operativa y comercial en las empresas, la información logística y la generada por internet de las cosas (IOT), la información sobre la interacción humana y los datos generados en la investigación científica. El IOT es una fuente importante de datos. La adquisición de los datos consiste principalmente en redes de sensores, luego la información se transmite y procesa y finalmente llega a una capa de aplicación que soporta aplicaciones específicas de IoT.

Adquisición de datos

La recopilación de datos consiste en utilizar técnicas especiales de recopilación de datos para adquirir datos brutos de un entorno específico de generación de datos. Los archivos de registro, los sensores y el rastreador web son ejemplos de dichas técnicas. Además de estos existen otros métodos, por ejemplo, en experimentos científicos, se pueden usar muchas herramientas especiales para recolectar datos experimentales como espectrómetros magnéticos y radiotelescopios.

Una vez completada la recopilación de datos en bruto, los datos se transfieren a una infraestructura de almacenamiento de datos para su procesamiento y análisis. Los datos grandes se almacenan principalmente en un centro de datos. El diseño de los datos debe ajustarse para mejorar la eficiencia de la computación o facilitar el mantenimiento del hardware.

Debido a la gran variedad de fuentes de datos, los conjuntos de datos recolectados varían con respecto al ruido, la redundancia y la consistencia por ello se requiere un preprocesamiento que incluya la integración, limpieza y eliminación de redundancia pues es indudablemente un desperdicio almacenar datos sin sentido.

Almacenamiento de datos

Este se refiere al almacenamiento y manejo de conjuntos de datos a gran escala, al tiempo que se logra confiabilidad y disponibilidad de acceso a los datos. Se encuentran sistemas de almacenamiento masivo, sistemas de almacenamiento distribuido y mecanismos de almacenamiento de grandes datos. Las tecnologías de almacenamiento masivo existentes pueden clasificarse como DAS (Direct Attached Storage) y almacenamiento en red, y a su vez el almacenamiento en red puede clasificarse en NAS (Network Attached Storage) y Storage Area Network (SAN). DAS se utiliza principalmente en computadoras personales y servidores de pequeño tamaño y el almacenamiento en red consiste en utilizar la red para proporcionar a los usuarios una interfaz de unión para el acceso y el intercambio de datos.

Por otro lado para utilizar un sistema distribuido para almacenar datos masivos, deben tenerse en cuenta factores como consistencia (asegurar que las copias múltiples de los mismos datos sean idénticas), disponibilidad (que el sistema no se vea gravemente afectado para satisfacer las peticiones del

cliente en términos de lectura y escritura) y tolerancia de partición (que el almacenamiento distribuido siga funcionando bien cuando la red esté particionada). Dado que la coherencia, la disponibilidad y la tolerancia de las particiones no pueden lograrse simultáneamente, se puede tener un sistema de CA (Consistency- Availability) ignorando la tolerancia de partición, un sistema de CP (Consistency- Partition) ignorando la disponibilidad y una AP (Availability- Partition) que ignora la consistencia, de acuerdo con diferentes objetivos de diseño.

Finalmente los mecanismos de almacenamiento para grandes datos pueden clasificarse en tres niveles: i) sistemas de archivos, ii) bases de datos y iii) modelos de programación. Los sistemas de archivos son la base de las aplicaciones en los niveles superiores. Por ejemplo Microsoft desarrolló Cosmos [86] para apoyar su negocio de búsqueda y publicidad y Facebook utiliza Haystack [87] para almacenar gran cantidad de fotos de pequeño tamaño. Por otro lado las bases de datos NoSQL (bases de datos relacionales no tradicionales) se están convirtiendo en la tecnología principal de los grandes datos. Se destacan tres bases de datos NoSQL: bases de datos de valores clave, bases de datos orientadas a columnas y bases de datos orientadas a documentos, cada una basada en ciertos modelos de datos. En las primeras encontramos Dynamo de Amazon [88] y Voldemort utilizado por LinkedIn. En las segundas encontramos BigTable de Google, Cassandra [89], HBase [90] y por último encontramos tres representantes importantes de sistemas de almacenamiento de documentos como MongoDB [91], SimpleDB [92] y CouchDB [93].

Algunos modelos de programación paralelos pueden mejorar el rendimiento de NoSQL y reducir la brecha de rendimiento a las bases de datos relacionales. Por lo tanto, estos modelos se han convertido en la piedra angular para el análisis de datos masivos. Dentro de estos podemos encontrar MapReduce [94], Dryad [95], All-Pairs [96] y Pregel [97].

Análisis de datos

Muchos métodos tradicionales de análisis de datos todavía se pueden utilizar para el análisis de datos de gran tamaño. A continuación presentamos algunas técnicas aplicables en una amplia gama de industrias.

Pruebas A / B es una técnica en la que un grupo de control se compara con una variedad de grupos de prueba con el fin de determinar qué tratamientos (es decir, cambios) mejorarán una variable objetivo dada. Un ejemplo de aplicación es determinar diseños, imágenes o colores que podrían mejorar un página web [98]. Los datos grandes permiten ejecutar y analizar un gran número de pruebas, asegurando que los grupos son de tamaño suficiente para detectar diferencias estadísticamente significativas entre los grupos control y tratamiento.

Fusión de datos e integración de datos. Conjunto de técnicas que integran y analizan datos de múltiples fuentes con el fin de desarrollar conocimientos de manera más eficiente y potencialmente más precisa que si se desarrollaran analizando una sola fuente de datos [99]. Los datos de las redes sociales, analizados por el procesamiento del lenguaje natural, se pueden combinar con datos de ventas en tiempo real, para determinar qué efecto tiene una campaña de marketing en el cliente y el comportamiento de compra.

Minería de datos. Es un proceso para extraer información o conocimiento ocultos, desconocidos, pero potencialmente útiles de datos masivos, incompletos, ruidosos, difusos y aleatorios [100]. Estas técnicas incluyen reglas de asociación, análisis de conglomerados, clasificación y regresión.

Reglas de asociación: estas técnicas consisten en una variedad de algoritmos para generar y probar posibles reglas. Una aplicación es el análisis de la cesta de mercado, en el cual se puede determinar qué productos son frecuentemente comprados juntos y usar esta información para la comercialización (muchos consumidores que compran pañales también tienden a comprar cerveza).

Análisis de conglomerados. Es un método estadístico para agrupar objetos, y específicamente, clasificar objetos según algunas características. El análisis de clúster se utiliza para diferenciar objetos con características particulares y dividirlos en algunas categorías (clusters) de acuerdo con estas características, de manera que los objetos de la misma categoría tendrán una alta homogeneidad mientras que diferentes categorías tendrán una alta heterogeneidad [101]. El análisis de clúster es un método de estudio sin supervisión sin datos de entrenamiento. Un ejemplo de análisis de clúster es la segmentación de los consumidores en grupos auto-similares para el marketing dirigido.

Muchas herramientas para la minería y análisis de datos están disponibles, incluyendo software profesional y amateur, software comercial costoso y software de código abierto. De acuerdo con la encuesta de “What Analytics, Data mining, Big Data software that you used in the past 12 months for a real project?” [102] los cinco programas más utilizados en 2014 fueron:

RapidMiner (44.2%): Rapidminer es un software de código abierto que se utiliza para la minería de datos, el aprendizaje automático y el análisis predictivo. Los programas de minería de datos y aprendizaje de máquinas proporcionados por RapidMiner incluyen Extract, Transform and Load (ETL), preprocesamiento y visualización de datos, modelado, evaluación e implementación [103]. Rapid Miner está escrito en Java. Integra el aprendizaje y el método de evaluación de Weka, y trabaja con R.

R (38.5%): R, un lenguaje de programación de código abierto y un entorno de software, está diseñado para la minería de datos / análisis y visualización [104]. Paso de 37, 4% en 2013 a 38,5% en 2014.

Excel (25,8%): Excel, un componente básico de Microsoft Office, ofrece potentes capacidades de procesamiento de datos y análisis estadístico. Cuando Excel está instalado, algunos complementos avanzados, como ToolPak de análisis y complemento Solver, con poderosas funciones para el análisis de datos se integran inicialmente, pero dichos complementos sólo pueden utilizarse si los usuarios los habilitan. Excel es también el único software comercial entre los cinco.

SQL (25.3%) Microsoft SQL Server es una plataforma de base de datos que se utiliza en el procesamiento de transacciones en línea (OLTP) a gran escala, es también una plataforma de Business Intelligence para soluciones de integración, análisis y creación de informes de datos [105]. En 2014 aparece en los cinco primeros lugares pero en la encuesta 2013 no se encontraba en el top 10.

Python, (19.5%). Python es un lenguaje multiplataforma de propósito general cuyo desarrollo se enfoca en proveer herramientas fundamentales para resolver cualquier tipo de problema en el campo de la computación, sin sacrificar la legibilidad y simplicidad del código [106]. Python paso del 13,3% en 2013 a 19,5 % en 2014.

Optimización. Una cartera de técnicas numéricas utilizadas para rediseñar sistemas y procesos complejos para mejorar su desempeño de

acuerdo con una o más medidas objetivas (por ejemplo, costo, velocidad o confiabilidad). Los algoritmos genéticos son un ejemplo de una técnica de optimización.

Algoritmos genéticos. Técnica utilizada para la optimización que se inspira en el proceso de evolución natural o "supervivencia del más apto". En esta técnica, las soluciones potenciales se codifican como "cromosomas" que pueden combinarse y mutarse. Estos cromosomas individuales se seleccionan para sobrevivir dentro de un "ambiente" modelado que determina la aptitud o el rendimiento de cada individuo en la población [107]. A menudo descrito como un tipo de "algoritmo evolutivo", estos algoritmos son adecuados para resolver problemas no lineales. Ejemplos de aplicaciones incluyen mejorar la programación de tareas en la fabricación y optimizar el rendimiento de una cartera de inversiones.

Aprendizaje automático. Una subespecialidad de la informática (dentro de un campo que históricamente se llama "inteligencia artificial") se refiere al diseño y desarrollo de algoritmos que permiten a las computadoras evolucionar comportamientos basados en datos empíricos[108]. Un foco importante de la investigación de aprendizaje de máquina es aprender automáticamente a reconocer patrones complejos y tomar decisiones inteligentes basadas en datos. El procesamiento del lenguaje natural es un ejemplo de aprendizaje automático.

Procesamiento del lenguaje natural (PNL). Un conjunto de técnicas de una subespecialidad de la informática (dentro de un campo llamado históricamente "inteligencia artificial") y lingüística que utiliza algoritmos informáticos para analizar el lenguaje humano (natural) [109]. Muchas técnicas de PNL son tipos de aprendizaje automático. Una aplicación de PNL está utilizando el análisis de sentimiento en las redes sociales para determinar cómo los clientes potenciales están reaccionando a una campaña de marca.

Redes neuronales. Modelos computacionales, inspirados en la estructura y el funcionamiento de las redes neuronales biológicas (es decir, las células y las conexiones dentro de un cerebro), que encuentran patrones en los datos [110]. Las redes neuronales son adecuadas para encontrar patrones no lineales. Pueden utilizarse para el reconocimiento y optimización de patrones. Algunas aplicaciones de redes neuronales implican aprendizaje supervisado y otras implican aprendizaje no supervisado. Ejemplos de aplicaciones incluyen la identificación de clientes de alto valor que están en riesgo de dejar una

empresa en particular y la identificación de reclamaciones de seguros fraudulentas.

Análisis de red. Conjunto de técnicas utilizadas para caracterizar relaciones entre nodos discretos en un grafo o una red. En el análisis de redes sociales, se analizan las conexiones entre individuos en una comunidad u organización, por ejemplo, cómo circula la información o quién tiene la mayor influencia sobre quién. Ejemplos de aplicaciones incluyen la identificación de los principales líderes de opinión para la comercialización y la identificación de cuellos de botella en los flujos de información empresarial.

Modelado predictivo. Un conjunto de técnicas en las que se crea o se elige un modelo matemático para predecir mejor la probabilidad de un resultado. Un ejemplo de una aplicación en la gestión de la relación con el cliente es el uso de modelos predictivos para estimar la probabilidad de que un cliente "churn" (es decir, cambiar de proveedor) o la probabilidad de que un cliente pueda comprar otro producto. La regresión es un ejemplo de las muchas técnicas de modelización predictiva.

Regresión. Un conjunto de técnicas estadísticas para determinar cómo cambia el valor de la variable dependiente cuando se modifica una o más variables independientes[111]. A menudo se utiliza para pronóstico o predicción. Ejemplos de aplicaciones incluyen la previsión de volúmenes de ventas basados en diversas variables de mercado y económicas o determinar qué parámetros de fabricación medibles influyen más en la satisfacción del cliente.

Análisis espacial. Un conjunto de técnicas, algunas aplicadas a partir de estadísticas, que analizan las propiedades topológicas, geométricas o geográficas codificadas en un conjunto de datos. A menudo los datos para el análisis espacial provienen de sistemas de información geográfica (SIG) que captan datos que incluyen información de localización, por ejemplo, direcciones o coordenadas de latitud / longitud [112]. Ejemplos de aplicaciones incluyen la incorporación de datos espaciales en regresiones espaciales (por ejemplo, ¿cómo está la disposición del consumidor a comprar un producto correlacionado con la ubicación?) O simulaciones (por ejemplo, ¿cómo funcionaría una red de cadena de suministro de fabricación con sitios en diferentes ubicaciones?)

Análisis de series de tiempo. Conjunto de técnicas tanto de estadística como de procesamiento de señales para analizar secuencias de puntos de

Clustergram: Es una técnica de visualización utilizada para el análisis de clúster que muestra cómo los miembros individuales de un conjunto de datos se asignan a los clústeres a medida que aumenta el número de clusters. La elección del número de clusters es un parámetro importante en el cluster análisis [114]. Esta técnica permite al analista alcanzar una mejor comprensión de cómo los resultados del agrupamiento varían con diferentes números de grupos.

En la actualidad, los principales métodos de procesamiento de grandes datos se muestran de la siguiente forma:

Bloom Filter: Bloom Filter consiste en una serie de funciones Hash. El principio de Bloom Filter es almacenar valores de Hash de datos distintos de los propios datos mediante la utilización de una matriz de bits, que es en esencia un índice de mapa de bits que utiliza funciones de Hash para llevar a cabo el almacenamiento con pérdida de compresión de datos [115]. Tiene ventajas tales como alta eficiencia espacial y alta velocidad de consulta, pero también tiene algunas desventajas en la falta de reconocimiento y eliminación.

Hashing: es un método que transforma esencialmente los datos en valores numéricos de longitud fija más cortos o valores de índice[116]. Hashing tiene ventajas tales como lectura rápida, escritura y alta velocidad de consulta, pero es difícil encontrar una función de Hash de sonido.

Computación Paralela: comparada con la computación en serie tradicional, la computación paralela se refiere a la utilización simultánea de varios recursos computacionales para completar una tarea de cálculo[117]. Su idea básica es descomponer un problema y asignarlos a varios procesos separados para ser completados independientemente, para lograr el coprocesamiento. Actualmente, algunos modelos clásicos de computación paralela incluyen MPI (Message Passing Interface), MapReduce y Dryad.

4.2.2 Big data y creación de valor

Hay cinco maneras ampliamente aplicables de aprovechar el potencial de los grandes datos para crear valor:

Creación de transparencia: En el sector público, por ejemplo, hacer que los datos pertinentes sean más fácilmente accesibles, a través de departamentos separados, puede reducir drásticamente el tiempo de

búsqueda y procesamiento, incrementar el descubrimiento fortuito y la redefinición de los datos y mejorar la rendición de cuentas[118], [119].

Habilitar la experimentación para descubrir las necesidades y mejorar el rendimiento: Las TI permiten a las organizaciones instrumentar procesos y establecer experimentos controlados. El uso de datos para analizar la variabilidad surgida naturalmente o generada por experimentos controlados y para comprender sus causas fundamentales, que permiten a los líderes administrar el rendimiento a niveles más altos[120], [121].

Segmentación de las poblaciones para personalizar las acciones: Los grandes datos permiten a las organizaciones adaptar productos y servicios precisamente a las necesidades de los segmentos de Mercado [122]. Familiarizado en marketing y gestión de riesgos, este enfoque puede ser revolucionario en otros lugares - por ejemplo, en el sector público. Incluso los bienes de consumo y las empresas de servicios que han utilizado la segmentación durante muchos años están comenzando a desplegar en tiempo real la micro-segmentación de los clientes para orientar las promociones.

Algoritmos para apoyo a las decisiones: La analítica sofisticada puede mejorar sustancialmente la toma de decisiones, minimizar los riesgos y desenterrar conocimientos valiosos que de otra manera permanecerían ocultos. Dichos análisis pueden desplegarse para triar los riesgos más altos de regulación y cumplimiento, afinar los inventarios y los precios en respuesta a las ventas en tiempo real en las tiendas y en línea. Las decisiones pueden basarse en el comportamiento de poblaciones enteras en lugar de muestras. La toma de decisiones nunca puede ser la misma; Algunas organizaciones ya están tomando mejores decisiones [123]–[125] analizando conjuntos de datos completos de clientes, empleados o incluso sensores incorporados en los productos.

Nuevos modelos, productos y servicios innovadores: Los grandes datos permiten a las empresas crear nuevos productos y servicios[126] [127], mejorar los ya existentes e inventar nuevos modelos de negocio. Los fabricantes están utilizando los datos obtenidos del uso de productos reales para mejorar el desarrollo de la próxima generación de productos y crear innovadoras ofertas de servicio posventa.

4.2.3 Retos del big data

El diluvio de datos en la era de los grandes datos trae consigo enormes desafíos en la adquisición, almacenamiento, administración y análisis de datos. Los sistemas tradicionales de gestión y análisis de datos se basan en el sistema de gestión de bases de datos relacionales y a su vez estos se aplican solo a datos estructurados, distintos de los datos semiestructurados o no estructurados. Algunos de los retos claves son:

Representación de datos: muchos conjuntos de datos tienen ciertos niveles de heterogeneidad en el tipo, la estructura, la semántica, la organización, la granularidad y la accesibilidad [128]. La representación de datos tiene como objetivo hacer que los datos sean más significativos para el análisis por computadora y la interpretación del usuario. Sin embargo, una representación inadecuada de los datos reducirá el valor de los datos originales e incluso puede obstruir el análisis efectivo de los datos. La representación eficiente de los datos deberá reflejar la estructura de los datos, la clase y el tipo, así como las tecnologías integradas a fin de permitir operaciones eficientes en los conjuntos de datos.

Reducción de redundancia y compresión de datos: en general, existe un alto nivel de redundancia en los conjuntos de datos. La reducción de la redundancia y la compresión de datos son eficaces para reducir el coste indirecto de todo el sistema sobre la base de que los valores potenciales de los datos no se ven afectados.

Gestión del ciclo de vida de los datos: en comparación con los avances relativamente lentos de los sistemas de almacenamiento, la detección omnipresente y la informática están generando datos a velocidades y escalas sin precedentes[129]. Uno de los retos mas apremiantes es que el sistema de almacenamiento actual no podía soportar datos masivos. En general, los valores ocultos en los datos grandes dependen de la frescura de los datos. Por lo tanto, debe desarrollarse un principio de importancia de los datos relacionado con el valor analítico para decidir qué datos se almacenarán y qué datos se descartarán.

Mecanismo analítico: el sistema analítico de grandes datos procesará masas de datos heterogéneos dentro de un tiempo limitado. Sin embargo, los RDBMS tradicionales están estrictamente diseñados con una falta de

escalabilidad y capacidad de expansión, que no podía satisfacer los requisitos de rendimiento[130]. Bases de datos no relacionales han demostrado sus ventajas únicas en el procesamiento de datos no estructurados y comenzó a convertirse en corriente principal en el análisis de datos grandes. Aún así, todavía existen algunos problemas de bases de datos no relacionales en su rendimiento y aplicaciones particulares [131].

Confidencialidad de los datos: la mayoría de los grandes proveedores de servicios de datos o propietarios en la actualidad no podrían mantener y analizar de manera efectiva tales enormes conjuntos de datos debido a su limitada capacidad. Deben contar con profesionales o herramientas para analizar tales datos, lo que aumenta los riesgos potenciales de seguridad [132]. Por ejemplo, el conjunto de datos transaccional generalmente incluye un conjunto de datos operativos completos para impulsar procesos empresariales clave. Dichos datos contienen detalles de la granularidad más baja y cierta información sensible como números de tarjetas de crédito. Por lo tanto, el análisis de grandes datos puede ser entregado a un tercero para el procesamiento sólo cuando se toman las medidas preventivas adecuadas para proteger esos datos sensibles y así garantizar su seguridad.

Gestión de energía: El consumo de energía de los sistemas informáticos principales ha atraído mucha atención desde la perspectiva de la economía y del medio ambiente. Con el aumento del volumen de datos y las demandas analíticas, el procesamiento, almacenamiento y transmisión de grandes datos consumirá inevitablemente más energía eléctrica [133]. Por ello esta realidad constituye un reto pues se requiere el control de consumo de energía a nivel de sistema garantizando el mínimo de costos y de impacto al medio ambiente.

Cooperación: el análisis de grandes datos es una investigación interdisciplinaria, que requiere que expertos en diferentes campos cooperen para aprovechar el potencial de los grandes datos. Debe establecerse una arquitectura amplia de redes de datos para ayudar a científicos e ingenieros en diversos campos a acceder a diferentes tipos de datos ya utilizar plenamente sus conocimientos técnicos, a fin de cooperar para completar los objetivos analíticos[134]

La revolución de los datos ha generado más ventajas y beneficios en sectores como los negocios, la salud, la ciencia y el gobierno y gracias a ello se ha contribuido a la solución de problemáticas específicas. Sin embargo,

también ha traído nuevos desafíos que no se contemplaban en los métodos tradicionales, que van desde la captura y almacenamiento de los datos, hasta su análisis e interpretación. No se puede dejar de lado la cultura en las organizaciones pues de esta depende el éxito o fracaso al emprender este tipo de iniciativas.

CAPITULO 5 – INGENIERIA DIRIGIDA POR MODELOS (MDE)

En este capítulo se abordará la ingeniería dirigida por modelos (MDE), la arquitectura dirigida por modelos (MDA) y los lenguajes de dominio específico (DSL).

5.1 Ingeniería Dirigida por Modelos MDE

Desde los inicios de la informática, los investigadores y desarrolladores de software han creado abstracciones que ayuden a programar en términos de su diseño y no pensando en el entorno computacional. La Ingeniería Dirigida por Modelos (Model-Driven Engineering -MDE) es el último paso en la tendencia histórica a elevar el nivel de abstracción, al que se diseña y construye el software. Los modelos son los artefactos claves de MDE. Un modelo es una descripción de un (parte de) sistema escrito en un lenguaje bien definido[135]. En este ámbito los modelos juegan un papel preponderante ya que no sólo son simples elementos de documentación del software que se diseña, sino que son los verdaderos directores de todo el proceso de desarrollo del software.

En un enfoque Dirigido por los modelos, el sistema de modelos tiene suficiente detalle como para permitir la generación de un sistema completo de aplicación de los modelos propios. De hecho, “el modelo es el código”, es decir, la atención se centra en el modelamiento y el código es generado mecánicamente a partir de modelos [136]. Bajo la premisa “todo es un modelo”, los investigadores y desarrolladores han descubierto que pueden expresar sus problemas en forma de modelos y aplicar técnicas de MDE para resolverlos ó simplificarlos aprovechando la automatización que proporcionan las herramientas disponibles. El ámbito de aplicación varía notablemente, abarcando desde dominios genéricos, como la Ingeniería Web la orientación a servicios, a dominios más específicos como los esquemas de base de datos o la domótica [137]

MDE propone comenzar el desarrollo con una definición formal de los conceptos y las reglas que caracterizarán a las representaciones de alto nivel (modelos), las cuales serán el artefacto principal de especificación de software. La obtención de representaciones de un menor nivel de abstracción, como por

ejemplo código fuente, es lograda a través de un proceso de derivación, el cual toma como entrada los modelos formales confeccionados de acuerdo a los conceptos y reglas definidos. La tecnología de MDE ofrece un enfoque prometedor para hacer frente a la incapacidad de los lenguajes de tercera generación para aliviar la complejidad de las plataformas y expresar conceptos de dominio de manera eficaz [138].

Ingeniería dirigida por modelos (MDE) se caracteriza por: a) poner al frente el nivel de abstracción ocultando los detalles específicos en la plataforma; b) aprovechar el uso de los modelos en todas las fases de desarrollo de software para mejorar la comprensión; c) desarrollar Framework y lenguajes específicos para lograr la comprensión del dominio; y d) obtener provecho de las transformaciones para automatizar trabajo repetitivo y mejorar la calidad del software[139].

La metodología propone comenzar el desarrollo con una definición formal de los conceptos y las reglas que caracterizarán a las representaciones de alto nivel (modelos), las cuales serán el artefacto principal de especificación de software. La obtención de representaciones de un menor nivel de abstracción, como por ejemplo código fuente, es lograda a través de un proceso de derivación, el cual toma como entrada los modelos formales confeccionados de acuerdo a los conceptos y reglas definidos[140].

Esta derivación es posible gracias a la cuota que se pone en la cantidad de aspectos considerados del problema de software a tratar, la cual conforma el “dominio específico” de los modelos y promueve un incremento en la semántica de los conceptos y reglas utilizadas en las descripciones. Tal incremento abre paso a una generación de código completa, en el sentido de que los desarrolladores no deben proveer más información que los mismos modelos a fin de obtener la implementación final, a costas de acotar las soluciones que se pueden obtener a un dominio particular.

Un dominio es definido como un área de interés en relación al desarrollo de software, y puede tratarse de situaciones comunes y acotadas a una problemática particular. Cuanto más se acote el dominio, más posibilidades habrán para generar aplicaciones funcionales directamente de las representaciones de alto nivel, dado que mayor será el contenido semántico de los componentes y las reglas que las conforman [140].

En contraste con otros procesos de desarrollo de software requiere, la creación de tres elementos principales:

- Un lenguaje de dominio específico formal a través del cual se definen las soluciones.
- Un generador de código para el lenguaje anteriormente especificado .
- Una infraestructura (framework) de dominio que sirva como base al código generado en el punto anterior, para que la traducción sea más sencilla

5.1.1 Terminología usada en MDE

En la figura 16 se ilustran una serie de términos relacionados con MDE los cuales permiten comprender el funcionamiento de dicha tecnología. A continuación se describirá cada uno de los términos.

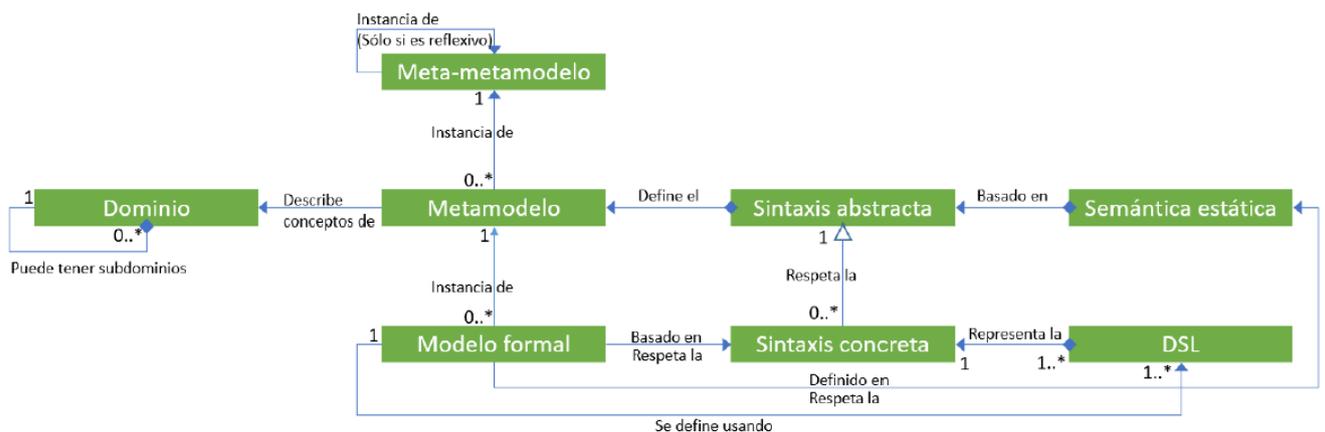


Figura 16. Arquitectura de los conceptos de MDE.
Fuente [141]

Sistema: conjunto de partes y de las relaciones entre estas partes que pueden ser organizados para lograr un propósito. Así, un sistema puede ser un sistema hardware o software, una compañía, procesos de negocio o la combinación de diferentes sistemas. Un sistema está formado por la plataforma y su aplicación [142].

Dominio: el dominio siempre es el punto inicial de MDE y delimita el campo de conocimiento. Así, el dominio es el área de conocimiento sobre la que se trabaja y sobre la que se quiere resolver el problema. Estos se dividen

en dominios tecnológicos y profesionales, donde los primeros hacen referencia a la tecnología de desarrollo de software y los segundos a los conceptos que manejará la aplicación. Los dominios a su vez pueden estar compuestos de varios subdominios[143]

Metamodelo: es una especificación del lenguaje de modelado que define las características del modelo y permite verificar el modelo expresado en ese lenguaje determinado de manera formal, es decir, es un modelo del lenguaje de modelado [144].

Metamodelo reflexivo: es cuando el metamodelo de un lenguaje de modelado usa el mismo lenguaje de modelado, es decir, se define el metamodelo utilizando el mismo lenguaje en el que el metamodelo está descrito. Dentro de estos, el metamodelo mínimo reflexivo es aquel que usa el mínimo número de elementos del lenguaje de modelado para los propósitos de ese metamodelo, luego, si se eliminase cualquier elemento sería imposible modelar o expresar cualquier estado esencial [141].

Meta-metamodelo: Es una especificación del metamodelo que define las características del metamodelo y permite verificar el metamodelo expresado en ese lenguaje determinado, es decir, en un modelo del metamodelo. El tener un meta-metamodelo, permite que exista un metamodelo para cada dominio del conocimiento a tratar, mientras se tiene un meta-metamodelo común a todos estos para así poder realizar operaciones sobre ellos, como pueden ser transformaciones automáticas, validaciones y búsquedas[141]

Lenguaje de dominio específico: un DSL está constituido por la estructura, los términos, notaciones, sintaxis, semántica y reglas de integridad que son usadas para expresar un modelo. Algunos ejemplos de lenguajes de modelado son UML, SQL Schema, Business Process Management and Notation (BPMN), E/R, Ontology Web Language (OWL) y XML Schema [142].

Punto de vista: un punto de vista de un sistema es una técnica de abstracción en la que se seleccionan una serie de conceptos y reglas estructurales de ese sistema con el fin de centrarse en las preocupaciones importantes de ese sistema. Es decir, se crea una abstracción para suprimir detalles irrelevantes y así obtener un modelo simplificado de una parte del sistema [145]. Cada punto de vista puede tener uno o más modelos, también conocidos como vistas.

Vista: una vista es la representación del sistema, desde una perspectiva elegida de un punto de vista. Por ejemplo, si tenemos un sistema para mostrar los datos de los usuarios de un videojuego online, una vista podría ser, la forma en que está la información estructurada, otra vista la que muestra la información que puede ver cada rol existente, otra vista podría ser, la que contiene los protocolos utilizados para transmitir la información y otra vista, para saber cómo se obtiene la información de los usuarios a partir de la información del sistema [141].

Capas de arquitectura del metamodelo[146]: Arquitectura basada en cuatro niveles de abstracción que van a permitir distinguir entre los distintos niveles conceptuales que intervienen en el modelado de un sistema. Esos niveles se les denominan comúnmente con las iniciales M0, M1, M2 y M3.

– El nivel M0 – Las instancias. El nivel M0 modela el sistema real, y sus elementos son las instancias que componen dicho sistema. Ejemplos de dichos elementos son el cliente “Juan” que vive en “Paseo de la Castellana, Madrid” y ha comprado el ejemplar número “123” del libro “Ulises”.

– El nivel M1 – El modelo del sistema. Los elementos del nivel M1 son los modelos de los sistemas concretos. En el nivel M1 es donde se definen, por ejemplo, los conceptos de “Cliente”, “Compra” y “Libro”, cada uno con sus correspondientes atributos (“dirección”, “numero de ejemplar”, “título”, etc.). Existe una relación muy estrecha entre los niveles M0 y M1: los conceptos del nivel M1 definen las clasificaciones de los elementos del nivel M0, mientras que los elementos del nivel M0 son las instancias de los elementos del nivel M1.

– El nivel M2 – El modelo del modelo (el metamodelo). Los elementos del nivel M2 son los lenguajes de modelado. El nivel M2 define los elementos que intervienen a la hora de definir un modelo del nivel M1. En el caso de un modelo UML de un sistema, los conceptos propios del nivel M2 son “Clase”, “Atributo”, o “Asociación”. Al igual que pasaba entre los niveles M0 y M1, aquí también existe una gran relación entre los conceptos de los niveles M1 y M2: los elementos del nivel superior definen las clases de elementos válidos en un determinado modelo de nivel M1, mientras que los elementos del nivel M1 pueden ser considerados como instancias de los elementos del nivel M2.

– El nivel M3 – El modelo de M2 (el meta-metamodelo). Finalmente, el nivel M3 define los elementos que constituyen los distintos lenguajes de modelado. De esta forma, el concepto de “clase” definido en UML (que pertenece al nivel

M2) puede verse como una instancia del correspondiente elemento del nivel M3, en donde se define de forma precisa ese concepto, así como sus características y las relaciones con otros elementos (p.e., una clase es un clasificador, y por tanto puede tener asociado un comportamiento, y además dispone de un conjunto de atributos y de operaciones).

Sintaxis abstracta: la sintaxis abstracta de un metamodelo es básicamente un modelo de datos de como almacenar o intercambiar datos en la estructura semántica del modelo a realizar, además de servir para definir las restricciones y estructura de los elementos del modelo[144].

Sintaxis concreta: Puede ser textual o gráfica. Bajo una misma sintaxis gráfica, se pueden tener diferentes sintaxis concretas, lo que nos permite tener diferentes tipos y vistas de esta, permitiendo diferentes tipos de forma para definirla, ya sea gráfica, textual o en forma de árbol, entre otros. Es decir, la sintaxis concreta define cómo se deben representar los modelos y que notaciones deben de utilizar los usuarios para ello [147].

Sintaxis estática: La semántica estática puede estar basada en términos de las cosas observadas en el mundo donde se describe, como puede ser el envío de mensajes, el estado de los objetos o el cambio entre estado, entre otras cosas, o bien, puede estar basada en traducciones de lenguajes de alto nivel construidos a partir de otras construcciones que tienen el significado bien definido[148].

Modelo: describe los elementos que pueden existir en un sistema, así como sus tipos. Por ejemplo, si definimos la clase “Persona” en un modelo, podremos utilizar instancias de dichas clases en nuestro sistema [146].

Infraestructura: conjunto de piezas software o hardware presentes cuando se desarrolla un artefacto software [148].

Plataforma: infraestructura de software que contiene una serie de recursos que permiten implementar o soportar un sistema determinado o implementado una tecnología específica [148]. La plataforma soporta la ejecución de la aplicación y, en conjunto, constituye el sistema. Así, la aplicación provee de la parte funcional del sistema descrito en el modelo.

Arquitectura: el propósito de la arquitectura es definir o mejorar los sistemas de manera que ayude a comprender el ámbito del sistema de interés y los requisitos [142].

Semantica del problema: esto hace referencia a que, cada vez que se incluye un artefacto en el modelo, lo que se está añadiendo es significado,

debido a que cada concepto capturado en el modelo tiene su propio significado. Así, cada concepto del lenguaje se mapea directamente a un concepto del dominio que se modela [147].

5.1.2 Beneficios de MDE

Dentro de los beneficios del MDE encontramos mejorar la productividad a corto plazo, debido a que se incrementa el valor de los artefactos de software primarios en términos de cuánta funcionalidad ofrecen y mejorar la productividad a largo plazo, debido a que incrementa el tiempo en el cual dichos artefactos se vuelven obsoletos, reduciendo la sensibilidad de los artefactos de software a: cambio de personal, requerimientos, plataformas de desarrollo, plataformas de despliegue [149].

Otros beneficios implican la reducción de la sensibilidad en cuatro formas:

Personal. El desarrollo del conocimiento vital no debe ser sólo almacenado en la mente de los desarrolladores, esta información se perderá en el caso de las fluctuaciones de personal. Por lo tanto, esta información debe ser de fácil acceso para cualquier persona, que no sean los creadores del artefacto software.

Requerimientos. Los cambios de los requerimientos son un problema conocido en la ingeniería de software. En el actual entorno empresarial aparecen nuevas funcionalidades más rápido que nunca. Todos estos cambios deben tener un bajo impacto sobre los sistemas existentes en términos de mantenimiento y no perturbar los sistemas en línea.

Plataformas de desarrollo. Están en un estado de constante evolución. Para disociar la vida de un artefacto de software de la herramienta de desarrollo utilizado para su creación inicial, es necesario disociar el objeto o el modelo que representa el objeto de la herramienta de desarrollo. Con MDE los modelos pueden ser independientes de la plataforma y así el peso de las modificaciones recae en las herramientas utilizadas para realizar las transformaciones.

El despliegue de plataformas. Para aumentar la vida útil de los objetos de software es necesario protegerlos contra los cambios en el despliegue de la plataforma.

Adicionalmente con MDE se puede generar todo el sistema a partir de los modelos y toda la documentación de alto nivel estaría en los modelos, reduciéndose así la cantidad de documentación a crear y mantener. El desarrollo del conocimiento vital no debe ser sólo almacenado en la mente de los desarrolladores, la información debe ser de fácil acceso para cualquier persona que no sean los creadores del artefacto software.

5.2 Arquitectura Dirigida por Modelos (MDA)

5.2.1 Definición

El reto actual es proponer esquemas de desarrollo en los cuales los modelos, antes que el código, son los actores centrales del proceso de desarrollo y donde se proveen mecanismos y herramientas que asisten al desarrollador en la construcción y transformación progresivas de modelos hasta llegar a la solución final. Esta corriente de trabajo, liderada por el Object Management Group (OMG) se conoce como arquitectura dirigida por modelos (MDA).

El OMG es una organización de compañías de sistemas de información creada en 1990, con el fin de potenciar el desarrollo de aplicaciones distribuidas orientadas a objetos. Esta organización ha definido estándares importantes como UML, CORBA, MOF, entre otros [150]. En los últimos años se ha tomado mayor interés e importancia al modelado en el desarrollo de cualquier tipo de software, debido a la facilidad que ofrece un buen diseño tanto a la hora de desarrollar como al hacer la integración y mantenimiento de sistemas de software.

La clave del MDA es la importancia de los modelos en el proceso de desarrollo de software. MDA propone la definición y uso de modelos a diferente nivel de abstracción, así como la posibilidad de la generación automática de código a partir de los modelos definidos y de las reglas de transformación entre dichos modelos.

El objetivo es separar la forma en que los sistemas de aplicación se definen a partir de la tecnología en que se ejecuten. El propósito de esa separación es asegurar que las inversiones realizadas en sistemas de construcción se pueden conservar incluso cuando las plataformas tecnológicas subyacentes cambian. Autores como [151], proponen que la esencia de la iniciativa MDA

es que el proceso de crear software debería ser dirigido por la formulación de modelos, en lugar de, por la escritura manual de código fuente.

MDA es quizá la más prometedora y ambiciosa propuesta de la OMG para abordar los retos de los actuales sistemas de software que se encuentran en constante cambio y con alto grado de interconexión. MDA se basa en los conceptos de MDE pero mediante la utilización de estándares de la OMG como son: Meta-Object Facility (MOF), XML Metadata Interchange (XMI), Object Constraint Language (OCL), Unified Modeling Language (UML), Common Warehouse Meta-model (CWM).

Con la combinación de estos lenguajes y el uso de mecanismos tales como la abstracción, refinamiento y vistas, MDA define un conjunto de modelos Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM) los cuales definirán el sistema desde diferentes perspectivas y niveles de abstracción [152].

La MDA rescata la importancia de los modelos como estrategia clave para entender y especificar una solución de software y progresivamente obtener la solución final. Aparece como respuesta a dos problemas fundamentales dentro de la industria informática: La diversidad de plataformas y tecnologías y la acelerada evolución tecnológica [153]. La estrategia para alcanzar beneficios fundamentales como productividad, interoperabilidad, portabilidad y facilidad de mantenimiento, se plantean en las ideas del manifiesto MDA [154].

5.2.2 Ciclo de vida del desarrollo con MDA

El ciclo de vida del desarrollo de software con MDA tiene las mismas seis etapas que el ciclo de vida del desarrollo de software tradicional (figura 17), pero con algunas diferencias[143]:

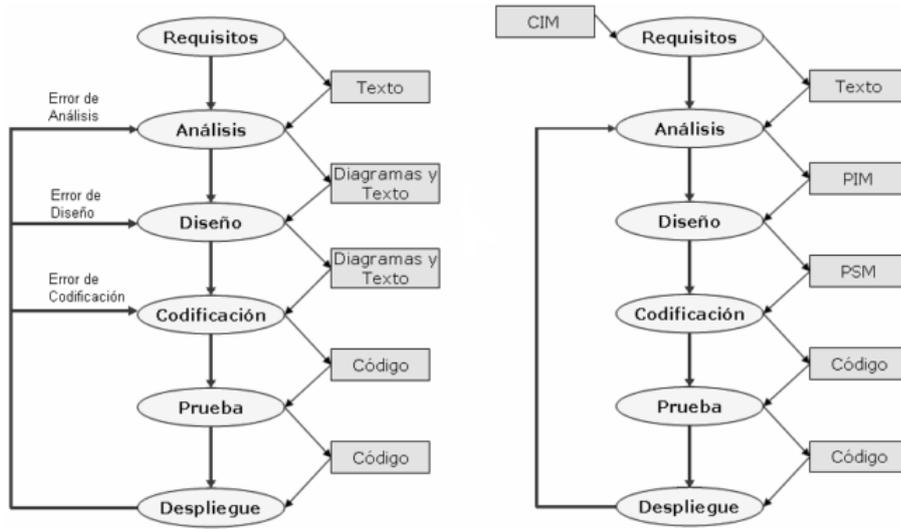


Figura 17. Proceso tradicional y proceso con MDA
Fuente [153]

Requisitos. La primera etapa es igual en ambos casos y corresponde con la captura de requisitos de forma textual en un papel. Utilizando la nomenclatura de MDA, esta etapa consiste en construir el Modelo independiente de cómputo (CIM).

Análisis. Con el desarrollo tradicional, los requisitos se convierten en diagramas de análisis que básicamente sirven como documentación y que casi nunca están sincronizados con el código debido a las continuas modificaciones que hay que hacer durante el desarrollo y durante el mantenimiento. Con MDA, en el análisis los requisitos se convierten en el Modelo Independiente de la Plataforma (PIM).

Diseño. En esta etapa, con el desarrollo tradicional lo que se hace es obtener de forma manual otros nuevos diagramas, mientras que con MDA lo que se hace es convertir el PIM al Modelo Específico de Plataforma (PSM) de la forma más automatizada posible.

Codificación. Mediante el desarrollo tradicional es necesario convertir los diagramas a código fuente de forma manual, mientras que con MDA la idea es que el código se genere de forma automática a partir del PSM.

Pruebas. Durante el desarrollo tradicional, es necesario emplear una cantidad muy grande de tiempo para realizar pruebas. En cambio, con MDA

muchas de las pruebas se pueden evitar o realizar de forma automática, lo que es otra ventaja adicional.

Despliegue. Con el desarrollo tradicional, el despliegue se suele realizar manualmente. La herramienta MDA podría automatizar parte o todo el despliegue.

5.2.3 Niveles de abstracción

La especificación MDA define cuatro puntos de vista o niveles de abstracción para analizar sistemas, de modo que la idea principal es empezar desde el nivel de abstracción más alto e ir haciendo transformaciones automática o semi-automáticamente hasta llegar a código ejecutable por una máquina física o virtual. Para cada punto de vista o nivel de abstracción se define un modelo del sistema:

CIM (Computational-Independent Model)

Corresponde con la captura de requisitos o modelo de dominio de la aplicación. El CIM se centra en los requerimientos y representa el nivel más alto del modelo de negocios. El CIM trasciende a los sistemas; cada proceso de negocio interactúa con trabajadores humanos y/o componente de máquina. Un objetivo fundamental del CIM, es que cualquiera que pueda entender el negocio y los procesos del mismo puede comprenderlo, ya que éste evita todo tipo de conocimiento especializado o de sistemas.

PIM (Platform Independent Model)

Cuando se realiza la transformación de CIM a PIM, se baja el nivel de abstracción de modo que se tiene un modelo que está pensado para ejecutarse en un ordenador pero sin tener en cuenta la plataforma. Con el fin de poner en práctica un PIM en una plataforma específica, se utiliza una herramienta para generar el PSM de la PIM. La herramienta comprende la tecnología de destino y sabe cómo traducir las construcciones lógicas del PIM en una forma adecuada para la plataforma elegida. Estas asignaciones pueden ser complejas, un elemento dado de la PIM puede ser asignado a múltiples elementos en el PSM.

Las características clave de PIM son las siguientes [155]–[158]:

1. Abstracción de alto nivel que es independiente de cualquier tecnología de implementación.

2. Describe un sistema de software que soporta algunos negocios y se especifica usando xUML.

3. El sistema es modelado desde el punto de vista que brinde mejor apoyo a la empresa.

4. Describe el comportamiento del sistema, independientemente del entorno informático y las tecnologías de implementación.

5. Los PIM pueden reutilizarse en múltiples plataformas.

PSM (Platform specific Model)

Es un modelo, pero ahora a nivel de diseño y reflejando la asignación a una plataforma de destino. La etapa final es generar a partir de la PSM el código y otros artefactos técnicos necesarios para implementar y ejecutar el sistema.

Las características clave de PSM son los siguientes [156]–[158]:

1. Especifica cómo la funcionalidad en un PIM se realiza en una plataforma informática.

2. Resulta del PIM mediante la adición de atributos específicos de la plataforma a los componentes estándar.

3. Es común tener múltiples PSMs con un solo PIM.

Una vez generado el PSM, el siguiente paso del proceso de desarrollo de software es la generación del código de PSM y desplegando el sistema en el entorno específico. Varios factores clave involucrados al pasar de PSM a código y despliegue son los siguientes [156]–[158]:

1. Generación de un modelo de objetos.

2. Repositorio de modelos que contiene toda la información necesaria para producir un tiempo de ejecución en una plataforma específica.

3. Traducción de abstracciones en los modelos a implementaciones específicas con la ayuda de compiladores de modelos.

4. PSM es un código ejecutable que se ejecuta en la arquitectura de destino.

Un punto central sobre el enfoque de MDA es que la automatización se utiliza para producir el PSM y Código a partir de los objetos de nivel superior. Esta automatización no sólo acelera el proceso de desarrollo, sino también

asegura que el código es una representación fiel de los modelos de más alto nivel, por lo que los modelos son una representación fiel del código.

5.2.4 Ventajas de MDA

El beneficio alegado por MDA es que el PIM, siendo independiente de la tecnología, es la mejor manera de capturar y preservar la inversión intelectual y financiera hecha en el desarrollo de una aplicación. Si la aplicación necesita ser desplegada en varias plataformas o emigrada de una plataforma a otra, el código requerido puede ser regenerado a partir del PIM. Esto, se dice, es más rápido y más barato que migrar el código desplegado.

Una de las principales ventajas que tendría un proceso de desarrollo dirigido por modelos es la independencia tecnológica. En la actualidad, el desarrollo de aplicaciones está fuertemente impregnado de las decisiones tecnológicas que se adopten. Éstas abarcan opciones middleware, frameworks arquitectónicos, servicios de persistencia y transaccionales, infraestructuras de seguridad o los lenguajes de programación utilizados. Los programas quedan inexorablemente ligados a las tecnologías que utilicen, y las tecnologías en el mundo de la informática son extremadamente volátiles [159].

La promesa de MDA es facilitar la creación de modelos legibles por máquina con un objetivo de flexibilidad a largo plazo en términos de [160]:

- **Obsolescencia tecnológica:** la nueva infraestructura de implementación puede ser fácilmente integrada y apoyada por diseños existentes.

Portabilidad: la funcionalidad existente se puede migrar más rápidamente a nuevos entornos y plataformas según lo dictado por las necesidades del negocio.

- **Productividad y tiempo de comercialización:** mediante la automatización de tareas de desarrollo arquitectos y desarrolladores se liberan para enfocar su atención en la lógica central del sistema.

- **Calidad:** la separación formal de las preocupaciones implicadas por este enfoque más la consistencia y fiabilidad de los artefactos producidos contribuyen a la calidad mejorada de todo el sistema.

- **Integración:** la producción de puentes de integración con sistemas heredados y/o externos.

- **Mantenimiento:** la disponibilidad del diseño en una forma legible por máquina ofrece a los analistas, desarrolladores y probadores acceso directo a la especificación para simplificar sus tareas de mantenimiento.

- **Pruebas y simulación:** los modelos pueden ser directamente validados así como probado contra varias infraestructuras.

- **Retorno de la inversión:** las empresas son capaces de extraer sus inversiones en herramientas.

En [161] se enumeran los siguientes beneficios para desarrolladores, dueños de negocio y expertos de dominio:

Los desarrolladores y/o diseñadores son los responsables de hacer realidad los requerimientos del sistema a desarrollar, algunos beneficios son:

- Un menor número de líneas de código escritas, ya que los niveles de abstracción fomentan el reuso del código y de los modelos.

- Alto nivel de abstracción para escribir aplicaciones y artefactos de software a través de la arquitectura de niveles del metamodelado y las capas de modelado de MDA. Este beneficio favorece diseñar una aplicación o artefactos de software partiendo de lo más general a lo más concreto, es decir, son independientes de la tecnología.

- Especificación de requisitos de usuario a varios niveles obteniendo un sistema flexible a los cambios. La flexibilidad de estos sistemas se observa si la funcionalidad que se desea agregar es posible implementarla a través de los modelos correspondientes y evidentemente en el código generado.

- Evitar la adopción de una única tecnología de hardware particular gracias a los niveles de abstracción presentes sin generar un vínculo particular, por tanto, los metamodelos se convierten en el eje central, porque representan el modelo del sistema de manera independiente a la tecnología.

Dueños del negocio: son los encargados de coordinar y/o financiar el proyecto de desarrollo e implementación del sistema dentro de la organización o institución, y los beneficios son:

- Desarrollo de componentes de software. Este beneficio se evidencia por las fases del MDA, que van del CIM a la generación de código, ayudando a los desarrolladores a crear diseños de artefactos correctos sin causar problemas en el sistema general con el fin de incrementar la productividad y calidad del mismo.

- Generación de código para plataformas previamente especificadas a través de la funcionalidad que ofrece el paradigma, ya que los artefactos de

software se generan en los PSM que luego se transformarán en el código deseado.

- Reducción de costes en el desarrollo de aplicaciones debido a la disminución del recurso humano requerido, de las horas hombre y del tiempo invertido en las diferentes actividades relacionadas.

- Documentación de todo el proceso de desarrollo de software, representado básicamente en el modelo PIM, documentación de alto nivel que se necesita para cualquier sistema de software.

Expertos del dominio: representan a los especialistas, los beneficios son:

- Permitir el desarrollo de modelos por parte de los distintos expertos del dominio, a diferencia de los desarrolladores que se concentran en los detalles técnicos a través de transformaciones de los modelos partiendo desde los PIM hasta la generación de código.

- Generación de lenguajes específicos del dominio. Este beneficio se obtiene en el desarrollo de modelos correspondientes a diferentes puntos de vista para las fases principales del paradigma, PIM y PSM, al mismo tiempo la posibilidad de implementar lenguajes específicos del dominio para esos puntos de vista garantizando el reuso de los conceptos.

- Interoperabilidad entre los distintos modelos PSM principalmente, ya que pueden pertenecer a distintas tecnologías (plataformas). Esta interoperabilidad se logra a través de puentes construidos por las herramientas de transformación de modelos garantizando los conceptos definidos.

5.3 Lenguajes de Dominio Específico (DSL)

5.3.1 Definición

La principal idea compartida por todos los paradigmas englobados dentro del Desarrollo de Software Dirigido por Modelos (DSDM) es la conveniencia de que los programadores empleen lenguajes de más alto nivel de abstracción que los lenguajes de programación, esto es, lenguajes que manejen conceptos más cercanos al dominio de la aplicación. Estos lenguajes que proporcionan mayor nivel de abstracción se denominan Lenguajes de Dominio Especifico (DSL).

A diferencia de lo que ocurre con los lenguajes de propósito general, gracias al uso de DSLs, se consigue que los conceptos de un lenguaje se mapeen directamente a conceptos del dominio que se modela, sin posibilidad de interpretaciones erróneas. En muchas situaciones del desarrollo de software se evidencia la repetitividad de los problemas en este caso para solucionarlos se puede utilizar un GPL (General Purpose Language o Lenguaje de Propósito General) como Java y C# o se puede recurrir a la utilización de un DSL.

Los expertos en un dominio crean los lenguajes específicos del dominio (DSL) y sus generadores de código, y los desarrolladores los usan para especificar una solución de alto nivel, de una forma más productiva que escribiendo el código en un lenguaje de programación.

5.3.2 Tipos de Lenguajes de Dominio Específico (DSL)

Se pueden distinguir tres clasificaciones para los DSL [143], desde el punto de vista de la construcción del lenguaje, desde el formato de lenguaje y desde el dominio del problema, a continuación se especifican cada uno.

Desde un punto de vista de la construcción del lenguaje:

Internos: Utilizan un determinado lenguaje anfitrión para darle la apariencia de otro lenguaje concreto. Un ejemplo claro son lo que actualmente se conoce como Fluent Interfaces.

Externos: Tiene su propia sintaxis y es necesario un parser para poder procesarlos. Un ejemplo claro de DSL externo es SQL (Structured Query Language)

Desde el punto de vista del formato del lenguaje:

Textuales: La mayoría de los lenguajes informáticos son textuales y están formados por un conjunto ordenado de sentencias. Un ejemplo muy conocido de DSL textual es SQL utilizado para realizar consultas a una base de datos. Una forma de crear DSLs textuales es mediante la creación de una determinada gramática (por ejemplo utilizando EBNF) y posteriormente crear o utilizar un parser para dicha gramática, para en etapas posteriores poder interpretar el DSL o generar código.

Gráficos: En los últimos años están ganando gran aceptación los lenguajes gráficos, podrían citarse como ejemplo UML. La creación de un lenguaje grafico es similar a la de un lenguaje textual, la única diferencia es que en lugar de usar texto para representar los conceptos, se utilizan conectores y figuras simples.

Desde el punto de vista del dominio del problema:

Horizontales: Los DSL horizontales son aquellos en los que el cliente que utilizará el lenguaje no pertenece a ningún dominio específico. Un ejemplo son los editores visuales de entornos de desarrollo que permiten generar interfaces de usuario automáticamente (por ejemplo Windows Forms de visual Studio).

Verticales: A diferencia de los DSL horizontales, el cliente que utilizará el lenguaje pertenece al mismo dominio que el lenguaje en sí. Como en el ejemplo anterior para un lenguaje de definición de encuestas, los usuarios finales serían los expertos en estadística encargados de definir dichas encuestas.

5.3.3 Características de un lenguaje de dominio específico (DSL)

Los DSL deben contar con algunas características, a continuación se relacionan [143], [162]:

Conformidad: los elementos del lenguaje deben de corresponderse con los conceptos del dominio.

Ortogonalidad: cada elemento del lenguaje es para representar un elemento del dominio.

Soporte: deben de existir herramientas para trabajar con el DSL.

Integrabilidad: el DSL y sus herramientas deben de poder interoperar con otras herramientas, así como deben de poder soportar la extensibilidad con otros elementos adicionales.

Longevidad: debe de haber una justificación temporal de cara a la creación al DSL, pues si este periodo de uso es demasiado corto, puede no ser rentable realizar el DSL.

Simplicidad: el DSL debe de poder representar de una manera simple los elementos del dominio.

Calidad: se deben proveer mecanismos para poder crear sistemas de calidad, por ejemplo, mediante la inclusión de precondiciones y postcondiciones.

Escalabilidad: este requisito no es necesario debido a que puede ser que se necesite un DSL para sistemas pequeños, pero a veces si es una característica deseable en algunos DSLs.

Usabilidad: el lenguaje debe de ser usable, no obstante, no siempre esta característica es posible debido a que a veces el propio lenguaje que se crear debido al dominio utilizar impide que este sea usable. No obstante, la simplicidad puede ayudar a que sea usable.

5.3.4 Ventajas de un lenguaje de dominio específico (DSL)

Existen multiples beneficios derivados del uso de DSLs [163], [164], entre los que se pueden destacar:

- Con un DSL es mucho menos probable cometer errores en la representación de un problema de un dominio que con un lenguaje de proposito general.
- Trabajar con los terminos de un dominio concreto facilita la comprension de los modelos que representan al software a personas no expertas en tecnologías de desarrollo de software.
- Cuando se trabaja con modelos expresados utilizando DSLs, dichos modelos pueden ser validados en el mismo nivel de abstraccion que el espacio del problema, lo cual implica que los errores seran detectados con mas antelacion.
- Los modelos podrian ser utilizados para simular salidas de las soluciones que se crearan.
- Cuando se captura conocimiento de un determinado dominio en un modelo, es más sencillo realizar migraciones entre diferentes tecnologías
- Los DSLs suelen proporcionar un API especifica de dominio para manipular sus modelos y así aumentar la productividad.

*Hay dos maneras de difundir la luz... ser la lámpara que
la emite o el espejo que la refleja*

– Lin Yutang

PARTE III ESTADO DEL ARTE

CAPITULO 6: INICIATIVAS DE DATOS ABIERTOS

Producido por la World Wide Web Foundation como un trabajo colaborativo de la red Open Data for Development (OD4D) y con el apoyo de la Red Omidyar, el Open Data Barometer (ODB) tiene como objetivo descubrir la verdadera prevalencia e impacto de las iniciativas de datos abiertos en torno al mundo. Analiza las tendencias mundiales y proporciona datos comparativos sobre países y regiones utilizando una metodología en profundidad que combina datos contextuales, evaluaciones técnicas e indicadores secundarios. Abarcando 92 países en el Barómetro se clasifican las naciones en:

- **Preparación** para iniciativas de datos abiertos
- **Implementación** de programas de datos abiertos
- **Impacto** que tienen los datos abiertos en los negocios, la política y la sociedad civil.

Para su realización se tiene en cuenta 3 tipos de información: Una encuesta de expertos con una serie de preguntas sobre contextos de datos abiertos, políticas, implementación e impactos y una encuesta detallada de datos para 15 tipos de datos en cada país, que aborda cuestiones de disponibilidad de datos, formato, concesión de licencias, puntualidad y descubrimiento. Una autoevaluación del gobierno con el mismo rango de contexto como una fuente adicional de información y por último datos secundarios seleccionados para complementar los datos de la encuesta de expertos (Foro Económico Mundial, Banco Mundial, Encuesta de Gobierno Electrónico de las Naciones Unidas y Freedom House).

La figura 14 muestra la distribución de las iniciativas de datos abiertos en las diferentes regiones del mundo. De esta manera se puede observar que Norteamérica, Europa y Asia Central lideran la preparación, implementación e impacto de la apertura de datos.

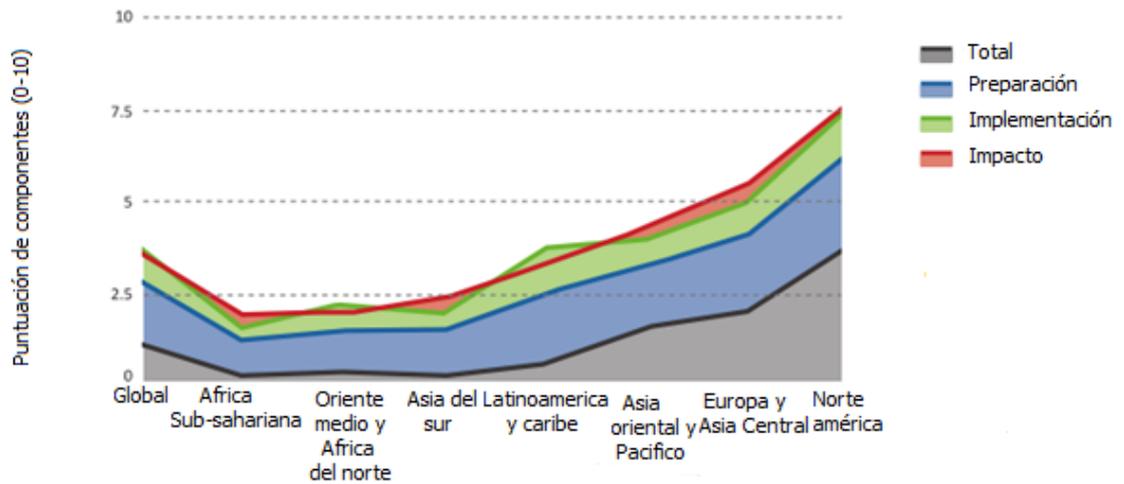


Figura 18. Distribución regional de datos abiertos.
Fuente [52]

Los países que ocupan el primer lugar en el ranking se caracterizan por su gran disposición. La implementación también es fuerte: el 46% de todos los conjuntos de datos abiertos se encuentran en estos 10 países. También se observa un impacto significativo en estos países.

Position	Country	Score	Readiness	Implementation	Impact
1	UK	100	100	100	100
2	USA	81.89	97	76	76
2	France	81.65	97	76	74
4	Canada	80.35	89	84	67
5	Denmark	76.62	77	77	78
6	New Zealand	76.35	87	62	87
7	Netherlands	75.13	90	69	70
8	Korea	71.19	95	64	58
9	Sweden	69.26	88	60	64
10	Australia	67.99	84	77	39
Average top 10		78.04	90.04	74.50	71.30

Figura 19. Top diez barómetro de datos abiertos tercera edición.
Fuente [52]

La mayoría de los países incluidos en la encuesta (55%) disponen de una iniciativa de datos abiertos y de un catálogo nacional de datos que ofrece acceso a los conjuntos de datos disponibles para su reutilización. Además, en

varios países, como Ecuador, Jamaica, Santa Lucía, Nepal, Tailandia, Botswana, Etiopía, Nigeria, Rwanda y Uganda, se están emprendiendo o se prometen nuevas iniciativas de datos abiertos.

Veintiséis de los treinta (30) primeros países del ranking son países de ingresos altos. La mitad de los datos abiertos se encuentran en sólo los diez (10) países de la OCDE, mientras que casi ninguno se encuentra en países africanos. Una nueva generación de líderes de datos abiertos, como Francia, Canadá, México, Uruguay, Corea del Sur y Filipinas, están empezando a desafiar a los clásicos líderes del ranking [52].

6.1 Antecedentes

A partir de 2009, países como Estados Unidos y el Reino Unido empezaron a poner en marcha sus iniciativas en materia de datos abiertos y de reutilización de la información. Algunas de las iniciativas más importantes en datos abiertos a nivel mundial son el portal Data.gov donde el gobierno Norteamericano a dispuesto a los ciudadanos cerca de 193,992 conjuntos de datos [165] para su reutilización. Mientras que la iniciativa de Reino Unido data.gov.uk dispone de 40,784 conjuntos de datos [166].

Cada vez son más los países en donde el concepto de gobierno abierto toma fuerza, ya sea como efecto de cambiar los sistemas tradicionales de gobernanza, poner en manifiesto su intención de procesos, acciones o datos más transparentes, colaborar interior y exteriormente con actores que aporten nuevas ideas en aras de solucionar problemas de carácter social o simplemente estimular la participación activa de ciudadanos en la formulación y ejecución de políticas.

El Reino Unido es uno de los pioneros en materia de gobiernos abiertos, una de sus motivaciones más relevantes es la de promover la toma de decisiones por parte de la ciudadanía, además de las sugerencias que puedan recibir. Allí la innovación se ha convertido en un pilar fundamental gracias a convocatorias que premian iniciativas relacionadas con datos abiertos gubernamentales y al desarrollar varios sistemas de información producto de alianzas estratégicas entre el gobierno y las compañías privadas locales. Hampshire Hub es un claro ejemplo de esta integración el cual consiste en un proyecto que asocia a proveedores de servicio público dentro y fuera de Hampshire (UK) y que a través de los datos abiertos se suplen de lo que

pueda ser ofrecido por el estado, personas naturales y hasta la propia competencia [167].

Por otro lado Europa ha encontrado en la revolución de datos abiertos otra razón para integrarse internamente, muestra de ello son las iniciativas que han ocasionado que los estados unifiquen en un solo portal web el contenido de los datos abiertos y la reutilización de los mismos (<https://data.europa.eu/euodp/es/data>) [168]. Este portal cuenta con 9383 datasets disponibles en 20 grupos diferentes como: trabajo y empleo, vida económica, industria, ciencia, medio ambiente, transporte, energía, unión europea, entre otras [169].

En el continente americano, sin lugar a dudas **Estados Unidos** se ha convertido por excelencia en el gran pionero de la reutilización de la información pública, y no solo porque posee portales especializados con un amplio catálogo de datos, sino porque también ostenta un robusto brazo legislativo que favorece la promulgación de información en el país; tal es el caso de la rendición de cuentas de fondos federales y la ley de rendición de cuentas (FFATA por sus siglas en inglés) promulgada en 2006, y que dentro de sus múltiples objetivos busca empoderar a los estadounidenses para que responsabilicen al gobierno del gasto público y en este orden de ideas se haga seguimiento a los recursos públicos. Es de esta manera como la legislación FFATA se materializa a través de un único portal de búsqueda denominado usaspending.gov [170].

El portal de datos Abiertos de Estados Unidos dispone de información y herramientas para aquellos desarrolladores externos que quieran trabajar en la construcción de aplicaciones y servicios, gracias a esto se han podido generar servicios con valor agregado que faciliten la rendición de cuentas.

Adicionalmente, **Canadá** ha mostrado significativos avances en materia de implementación de Datos Abiertos en su jurisdicción, muestra de ello es la creación del portal centralizado open.canada.ca, el cual posee poco más de 148085 datasets y posee un portal de transparencia (transparency.ca) que divulga datos de diversa índole que dan cuenta de su gestión en términos financieros. Es así como la ciudadanía ha adoptado un papel activo desde que el gobierno estatal en 2011 decidiera adquirir compromisos ante la alianza para el gobierno abierto (OGO) y que hasta la fecha ha cumplido a cabalidad [171].

6.2 Latinoamérica y Caribe

Los resultados del Barómetro 2015 indican que Latinoamérica muestra un desempeño relativamente positivo comparado con el promedio global. Por primera vez hay tres países (México, Brasil, Uruguay) de la región entre los top 20 y cuatro entre los top 30 (Colombia). En particular, los resultados destacan a México y Uruguay, tabla 3, como nuevos contendientes de los líderes tradicionales, con un significativo salto en el ranking gracias a nuevas políticas centradas en la generación de impacto para sus ciudadanos a través de mejores servicios públicos. Por estas razones se estudia con detenimiento esta región y también se justifica dada la conveniencia para el investigador al conocer el entorno y posibles expertos que validen la propuesta posteriormente.

Estos países han asumido un rol de liderazgo en Latinoamérica, y demuestran que las políticas de datos abiertos que equilibran esfuerzos entre capacidades e implementación pueden generar un impacto positivo. Por el contrario, la región también presenta dos de las caídas más significativas en el ranking global (tabla 3): Chile (-15) y Argentina (-16). Chile, un líder tradicional de datos abiertos en la cayó a la posición 30 debido a una preponderancia a realizar actividades a corto plazo por sobre políticas sustentables y de largo aliento en datos abiertos.

Tabla 3. Ranking datos abiertos Latinoamérica

POSICIÓN	PAÍS	INGRESO 2015*	PUNTAJE ODB
16	 México	Ingreso mediano alto	61.76
17	 Brasil	Ingreso mediano alto	61.16
19	 Uruguay	Ingreso alto	58.11
28	 Colombia	Ingreso mediano alto	45.39
30	 Chile	Ingreso alto	42.97
42	 Ecuador	Ingreso mediano alto	30.29
44	 Perú	Ingreso mediano alto	28.93
44	 Costa Rica	Ingreso mediano alto	28.52
52	 Argentina	High income	23.78
62	 Paraguay	Ingreso mediano alto	15.99
78	 Venezuela	Ingreso alto	6.79

Fuente: [52]

Uno de los países de la región que se esforzó por promulgar las primeras iniciativas de transparencia fiscal y rendición de cuentas, así como de acceso a la información pública fue **México**, la ley federal de transparencia y acceso a la información gubernamental de 2002 constituyó una base lo suficientemente sólida como para soportar futuras iniciativas en materia de regulación y preparó el país para la llegada de datos abiertos [172].

Otra ley crucial para que la apertura de datos dentro de México se llevara a cabo fue la ley federal de protección de datos personales firmada en 2011[173], también en este año se planteó un esquema inicial de interoperabilidad para publicar las bases, políticas, así como los principios para compartir servicios digitales dentro de las múltiples dependencias de la nación, esto facilitó el desarrollo de plataformas especializadas de datos abiertos en temas como rendición de cuentas (datamx.io) o participación ciudadana como lo es el sistema Infomex a través del cual puede solicitarse información pública del Gobierno Federal [174].

México actualmente cuenta con el Instituto Federal de Información y Protección de Datos (IFAI), dependencia encargada de garantizar el derecho de los ciudadanos a la información pública gubernamental y a la privacidad, la rendición de cuentas por parte del estado y fomentar el uso de la información pública, es decir, es una entidad promotora de la ley de transparencia en el estado Mexicano [175]

Brasil es uno de los países líderes en la región en la incorporación de la gestión pública a una era de transparencia, muestra de ello fue la cofundación de la Alianza para el gobierno abierto junto con Estados Unidos, además las iniciativas llevadas en la creación de repositorios de datos tanto para el gobierno federal (dados.gov.br) como para temas tan sensibles como lo es la rendición de cuentas en una plataforma centralizada de transparencia www.portaltransparencia.gov.br, la cual es objeto de mejoras permanentes por cuenta del estado; lo anterior soportado en un marco legal de acceso a la información pública que contempla leyes como la 12.527 de 2011 o la 9.507 de 1997 [176], [177].

Uruguay por su parte, en su primer Plan de Acción de Gobierno Abierto se enfocó en torno a tres grandes líneas de acción: aumentar la integridad pública, gestionar de forma más eficiente los recursos públicos y mejorar la prestación de servicios públicos al ciudadano [178].

Para gestionar de forma más eficiente los recursos públicos, se realizaron esfuerzos relacionados con la Agencia de Compras y Contrataciones del Estado (ACCE): en 2013 se implementó la Apertura Electrónica, que permite realizar en línea todo el proceso desde el llamado hasta la adjudicación y evita costos tanto a oferentes como al Estado al eliminar instancias presenciales y basadas en papel. Además se implementó el Registro Único de Proveedores del Estado (RUPE) que mejora y simplifica la relación entre el Estado y sus proveedores, y se creó un sistema profesional de reclutamiento y selección para la búsqueda de los mejores candidatos para trabajar con el Estado (www.uruguayconcurso.gub.uy) llamado Uruguay Concurso [178].

Para mejorar la prestación de servicios públicos al ciudadano, se aprobaron los Decretos de Simplificación de Trámites 177/013 y 178/013, y la reglamentación de algunos artículos de la ley 18.719, que versa sobre el intercambio de información entre distintos organismos públicos [178], [179].

Para fortalecer los Datos Abiertos de Gobierno, desde diciembre 2012 se dispone del Catálogo Nacional de Datos Abiertos, se desarrollaron guías para publicación de datos abiertos, se capacitaron funcionarios públicos y ciudadanos para acercar el tema a los diferentes involucrados. Se realizó el DateIdea.uy primer concurso de datos abiertos que cuenta con tres categorías: “aplicaciones existentes”, que reconoce a los pioneros, “prototipos”, que premia a aquellos que utilicen datos del catálogo con aplicaciones nuevas y por último una categoría de “ideas” originales que busquen desarrollarse. Además, se aprobó el Decreto de Simplificación de Trámites,; Uruguay fue sede de la Primera Conferencia Regional de Datos Abiertos para América Latina y el Caribe, donde hubo 600 asistentes presenciales, 1900 on-line, 106 expositores, 122 paneles temáticos, y se realizaron talleres dirigidos a periodistas como iniciativa para promover el periodismo de datos en Uruguay como el 2do. Databootcamp en Latinoamérica con la participación de periodistas, desarrolladores y expertos de Inglaterra, Francia, Alemania, España y Argentina [178].

Colombia cuenta con el Portal del Estado Colombiano (www.gobiernoenlinea.gov.co) a cargo del Ministerio de Tecnologías de la Información y las Comunicaciones, el cual es un punto integrado de acceso a la información, trámites y servicios que proveen las instituciones públicas a través de medios electrónicos. Esta plataforma ofrece servicios de acompañamiento a la población como el asesor virtual (chat disponible en el

Portal) y el centro de contacto al ciudadano (línea telefónica gratuita en todo el país). En cuanto a los Datos Abiertos, el Gobierno Nacional ha definido los lineamientos para que las entidades estructuren sus datos y los publiquen en sus diferentes sitios web y de forma unificada en el portal de datos abiertos www.datos.gov.co, cuya finalidad es convertirse en el mecanismo de coordinación para el acceso y uso de los diferentes datos publicados por las entidades [180].

En cuanto al gobierno de **Chile**, Sebastián Piñera, presidente de Chile (2010-2014), fue el primero que se comprometió al impulso del gobierno abierto y con la Open Government Partnership (OGP), de la cual Chile hace parte desde septiembre de 2011. Desarrolló planes de gobierno abierto para mejorar el acceso a los servicios públicos, e intentó incorporar todas las herramientas digitales para que el Estado se acercara a la gente, fuera más amistoso y grato [181].

Así, el gobierno de Chile impulsa lo que se denomina “agenda digital Imagina Chile 2013-2020”, que es la hoja de ruta del Gobierno en torno a las tecnologías de la información con el fin de impulsar el desarrollo digital y tecnológico del país. Establece cinco ejes estratégicos: Conectividad e Inclusión Digital, Innovación y Emprendimiento, Educación y Capacitación, Servicios y Aplicaciones y Entorno para el Desarrollo Digital. Cada uno cuenta con un plan de acción y ejecución además de metas e indicadores relevantes de manera que se monitorea su grado de implementación que impulsará la competitividad y la economía digital de Chile [182].

También se aprobó la Ley n° 20.285 sobre el acceso a la información pública, promulgada en agosto de 2008 por la Presidenta Michelle Bachelet, y entró en vigencia en abril de 2009, tiene como principal objetivo abrir y facilitar el acceso a la información de los organismos públicos a todas las personas, promoviendo un sistema estatal más cercano, generando nuevos espacios para la participación y mejorando las exigencias de rendición de cuentas de la gestión pública” [181].

Además, se han desarrollados varios portales para fortalecer el gobierno abierto como Chile Atiende, el cual reúne en un lugar distintos servicios y beneficios de diferentes instituciones públicas, el Portal de Gobierno Abierto de Chile <http://www.gobiernoabierto.cl/> un espacio virtual que centraliza todas las iniciativas de transparencia, participación y datos públicos de Chile, y el Portal “Acceso Inteligente”, desarrollado por la fundación Ciudadano

inteligente, que busca facilitar el acceso a la información por medio de la transparencia pasiva. Otros portales son: el portal de Linked Data de la Biblioteca del Congreso Nacional (BCN) que permite utilizar tecnologías innovadoras para mejorar procesos y servicios al Congreso y a los ciudadanos, el portal Open Data Consejo para la Transparencia es el catálogo de datos abiertos (Open data) del Consejo para la Transparencia y el Portal de Datos Abiertos www.datos.gob.cl [182].

En el Plan de Acción propuesto por **Argentina**, se busca fortalecer la interoperabilidad entre los distintos programas, fomentando el trabajo en conjunto y promoviendo nuevas soluciones para la ciudadanía. Cuentan con un Portal Web Ministerio de Desarrollo Social de la Nación que promueve una relación interactiva con la ciudadanía y dar acceso a la información pública a partir de múltiples canales de consulta, cumple con los principios de usabilidad y de accesibilidad web para personas con discapacidad y hace uso de un lenguaje amplio y sin tecnicismos, con el propósito de brindar la información necesaria para el acceso a las políticas públicas [183].

Otro de los esfuerzos realizados por el gobierno argentino fue la creación de la ley 26653 de accesibilidad de la información en las páginas web, con el objeto de garantizarles la igualdad real de oportunidades y trato, evitando así todo tipo de discriminación, y el decreto 1172/03 de Mejora de la Calidad de la Democracia y sus Instituciones, dictado en diciembre del año 2003 y aplicado en todo el ámbito del Poder Ejecutivo Nacional [184], [185].

En cuanto al Plan de acción de gobierno abierto de **Paraguay**, en 2012 se realizó el primer Plan de acción, en el cual se plantearon 15 compromisos, de los cuales se cumplieron 3: Un Sistema de intercambio de información entre instituciones del gobierno (SII) denominada “TEKOPORU”; el fortalecimiento del sistema nacional de contrataciones públicas, apuntando sobre todo a la simplificación y mayor transparencia de los procesos de compras y contrataciones y el Portal Único de Empleo Público “Paraguay Concurso” [186].

Entre los compromisos adquiridos en el Plan de Acción 2014-2016 se encuentran la promulgación e implementación de la Ley de libre acceso ciudadano a la información pública y transparencia gubernamental, expedida en el 2014, diseñar e implementar una política de datos abiertos y promoción de la generación de capacidades en la Sociedad Civil para su utilización, demostrar transparencia y dar acceso a la Información de

proyectos públicos emblemáticos, implementar canales de diálogo social y participación ciudadana para la consulta y el monitoreo de políticas públicas y crear un Sistema de Trámites Jurídicos Electrónicos en Contrataciones Públicas de acceso público [187].

El portal del gobierno de Paraguay se presentó oficialmente en 2014 (<http://www.paraguay.gov.py/>), en el cual se encuentra el portal oficial de datos abiertos del estado paraguayo o Catálogo de Datos Abiertos de Gobierno (<http://www.datos.gov.py/>), que se constituye como un único punto de acceso a los datos abiertos y presenta los conjuntos de datos (datasets) organizados por instituciones, agrupados por categorías y presentados en distintos formatos, para facilitar la libre descarga de los mismos; así como las licencias libres asociadas al uso, re-uso y distribución de cada set de datos.

En **Perú** como marco normativo, la Constitución (art. 2º, numeral 5), reconoce el derecho fundamental de acceso a la información pública y el Texto Único Ordenado de la ley 27806, Ley de Transparencia y Acceso a la Información Pública (DS 043-2003-PCM), desarrolla el derecho a la información [188].

En el Plan de acción de Gobierno Abierto Perú 2012 se propusieron 47 compromisos, los cuales fueron revisados y reorganizados en 12 compromisos nuevos para el Plan de Acción de 2014-2016 [189]. Perú se comprometió con la transparencia y acceso a la información pública, la participación ciudadana, rendición de cuentas y mejora de servicios públicos.

Sin embargo, Perú todavía no cuenta con un catálogo de datos abiertos oficial que brinde información al ciudadano y le permita reutilizar la información, a pesar de que un diagnóstico del Banco Mundial concluye que el Perú reúne las condiciones necesarias para implementar una política de datos abiertos.

En **Bolivia** no existe una cultura de acceso a la información que debe generar el Estado, sin embargo, el Decreto supremo 28168 de 2005 tiene por objeto garantizar el acceso a la información, como derecho fundamental de toda persona y la transparencia en la gestión del Poder Ejecutivo y el Decreto 0214 de 2009 aprueba la Política Nacional de Transparencia y Lucha contra la corrupción con el fin de contar con instrumentos orientados a la prevención, investigación, transparencia, acceso a la información y sanción de actos corruptivos [38], [190]. Actualmente Bolivia no participa de la Alianza

de Gobierno Abierto, cuenta con un portal de datos abiertos <https://datos.gob.bo/> donde apenas se dispone de 24 conjuntos de datos.

En **Venezuela** se identifica retraso, ya que el gobierno aún se encuentra en fases de aprobación y/o adopción de leyes que permitan su inserción en la revolución de Open data y consecuentemente, no cuentan con catálogo de datos abiertos; tampoco forma parte de la Alianza de Gobierno abierto y no sitúa los datos abiertos entre sus prioridades [37]. Actualmente Venezuela cuenta con un portal www.gobiernoenlinea.gov.ve, donde se encuentra información para facilitar los trámites del ciudadano, pero no contiene ningún tipo de información pública en formatos abiertos.

En **Ecuador**, La Ley Orgánica de Transparencia y Acceso a la Información Pública (LOTAIP) es la norma que regula el ejercicio del derecho a acceder a la información pública en Ecuador. Además, la Constitución Política de la República del Ecuador de 1998 es una de las constituciones latinoamericanas que desarrolla con mayor precisión el derecho de acceso a la información pública. Ecuador no pertenece a la Alianza de Gobierno Abierto y cuenta con un portal de Datos abiertos <http://catalogo.datosabiertos.gob.ec/> donde se encuentran 107 documentos que pueden ser consultados por los ciudadanos.

En **Bélice** cuentan con una ley que consagra de manera más o menos amplia el derecho de acceso a la información pública, es el “Freedom of Information Act, 2004”. Aunque cuenta con el portal oficial <http://www.belize.gov.bz/>, no cuenta con catálogos de Datos Abiertos ni pertenece a la Alianza para el Gobierno abierto (Open Government Partnership) [192].

En cuanto a **Cuba**, no existe ninguna garantía constitucional o jurídica con respecto al libre acceso a la información pública, no pertenece a la Alianza de gobierno abierto y no cuentan con portal de datos abiertos o de gobierno electrónico, sólo se puede observar el portal <http://www.cuba.cu/> pero no cuenta con información en formatos abiertos.

Con respecto a las Antillas, **Antigua y Barbuda** cuenta con ley de libertad de información del 2004 y sus órganos de aplicación de la ley son el Comisionado para la Información, y el Comisionado para la Integridad; en **Bahamas** tienen el Proyecto de ley sobre la libertad de información, **Barbados** cuenta con la Ley sobre la libertad de información; **Dominica** se encuentra en proceso de elaboración de la ley y **Granada** tiene el Proyecto de ley de

libertad de información aunque no cuenta con reglamentación sobre participación ciudadana. Ninguno cuenta con portales de datos abiertos [193].

Haití tampoco cuenta con ley de acceso a la información pública y aunque tampoco cuenta con portal oficial de datos abiertos, si cuenta con un portal sobre información geo-espacial disponible en <http://haitidata.org/>.

En **Guyana** sólo se cuenta con la ley de acceso a la información del 2011 [194], mientras que el órgano de aplicación de la ley de acceso de información en Jamaica, publicada en el 2003, es la Unidad de Acceso a la Información [195], sin embargo ninguno cuenta con reglamentación sobre participación ciudadana o portales de datos abiertos.

Desde hace poco más de diez años, el gobierno de la **Republica Dominicana** ha mostrado claros esfuerzos por promover la participación ciudadana en su territorio, y con el ánimo de concretar esto, formula un plan de acción con la intención de adherirse a la alianza gobierno abierto, en dicho plan se identifican compromisos como el de dar apertura a información, crear plataformas únicas entre ellas el sitio oficial de datos abiertos, entre otros, dichos compromisos se renuevan en su mayoría en el segundo plan de acción presentado ante la misma organización. Del mismo modo, Republica Dominicana cuenta con leyes orgánicas de acceso a la información pública y protección de datos personales [196], [197].

Puerto Rico en el presente cuenta con un sitio web de datos abiertos oficial denominado data.pr.gov, el cual divulga datos de acceso libre provenientes del estado y herramientas que facilitan su apropiación, no obstante, el país no posee legislación pertinente al acceso a la información pública ya que está en curso de ser aprobada, lo que ha propiciado que el gobierno no cuente con una plataforma unificada para la fiscalización de la gestión del gobierno, dicha labor ha sido adoptada por organizaciones sin fines de lucro en el país como el Centro de Investigación y Política Pública, cuyo proyecto ABRE tiene como propósito eliminar barreras para acceder a la información pública a través de una página que publica permanentemente datos en formatos abiertos [198].

Al igual que Brasil, **El Salvador** promulgó su ley de acceso a la información pública en 2011 así como la regulación de los servicios de información, de este modo, se abrió el paso para que desarrolladores digitales bajo la dirección del gobierno diera apertura a su plataforma oficial

datoselsalvador.org, la cual significa uno de los primeros resultados en materia de desarrollo de plataformas de esta índole en el país, no obstante y a pesar de los esfuerzos, El Salvador se encuentra en una etapa preliminar en la cual debe centrar esfuerzos para concebir planes estratégicos más estructurados, así como vincular múltiples dependencias en dicho proceso a fin de dar cumplimiento a la legislación vigente en materia de transparencia [199].

Guatemala por su parte se encuentra en una etapa conceptual, la cual se ha formalizado a través de los planes de acción nacional de gobierno abierto, el primero publicado en 2008 y el último contempla un horizonte de planeación hasta el 2016, dicho plan permite apreciar las estrategias que pretende emplear el gobierno para la implementación de datos abiertos en Guatemala que se sintetizan en cinco compromisos y veintidós acciones concretas a ser ejecutadas por una comisión de transparencia a cargo de la presidencia de la república. A pesar de lo anterior, Guatemala posee su propia ley de acceso a la información pública desde 2008 en donde además está consignado el derecho a la protección de datos personales, ya que no hay una ley independiente que desarrolle este derecho de forma más amplia e integral [200], [201].

Honduras durante los últimos tres años ha focalizado sus esfuerzos en la lucha contra la corrupción, y es este orden de ideas por hacer que la gestión estatal sea lo más transparente posible y propenda por la rendición de cuentas, de esta manera se crea el instituto de acceso a la información pública, cuyo objetivo es facilitar el acceso a la información pública por parte de los ciudadanos y supervisar los procedimientos de las instituciones obligadas se ciñan a la ley de transparencia y acceso a la información pública firmada en 2006 [202], [203].

El gobierno de Honduras adicionalmente, ha formulado catorce compromisos para los próximos cuatro años enmarcado dentro del segundo plan de acción de gobierno abierto de Honduras 2014-2018, este documento hace hincapié en estrategias anticorrupción en su mayoría, así como en el fomento de la transparencia fiscal. A pesar de los notables esfuerzos para la elaboración de estrategias por parte del estado en términos de acceso a la información, no hay acciones concretas en el desarrollo de plataformas de datos abiertos que soporten dichas estrategias, y en consecuencia no existe una plataforma oficial de datos abiertos en el país [202].

Nicaragua en 2007 aprueba la Ley de acceso a la información pública, más adelante, en 2012 publica la ley de protección de datos personales; a pesar de que cuenta con una base sólida en lo que respecta a normativa, Nicaragua se ha rezagado con respecto a países de la región al no contar con planes bien estructurados en aras de dar cumplimiento a cada una de las leyes, igualmente, hay ausencia de portales de transparencia o rendición de cuentas, participación ciudadana, finalmente el estado carece de una plataforma oficial de datos abiertos [204], [205].

Uno de los países con notorios avances en cuanto a datos abiertos se refiere en centro américa es **Costa Rica**, en donde a pesar de que no existe aún una ley exclusivamente dedicada al tema de acceso a la información pública, constitucionalmente el estado ha tratado de garantizar este derecho, además se ha promulgado la ley contra la corrupción y el enriquecimiento ilícito en la función pública, la cual establece pautas para prevenir, detectar y sancionar actos corruptos que soporta el derecho constitucional de garantizar el libre acceso a los departamentos administrativos con el objeto de que dicha información sea más asequible a la ciudadanía [206], [207].

Uno de los logros más trascendentales para Costa Rica en términos de plataformas digitales en el marco del gobierno abierto, es contar en la actualidad con un sitio oficial de datos abiertos datosabiertos.gob.cr en donde puede encontrarse información vinculada a ministerios y diversas dependencias del país. Adicionalmente, se cuenta con un plan de acción de gobierno abierto elaborado en colaboración con la sociedad civil, así como múltiples plataformas que sirven de canal entre el estado y la ciudadanía con la intención de mejorar los servicios públicos que faciliten la participación ciudadana y la colaboración por cuenta del nivel gubernamental [208], [209].

Finalmente **Panamá** en 2013 estructuró un plan de acción para el gobierno abierto, en el cual se consignan compromisos referentes a lucha contra la corrupción y transparencia, allí además, se menciona la importancia de crear plataformas oficiales tanto para garantizar el cumplimiento de la ley 6 de 2002 (Ley de transparencia), como para promover la participación ciudadana, asimismo, en este plan de acción propone crear el sitio oficial de datos abiertos para el país, ya que carece del mismo en la actualidad. Por su parte la autoridad nacional de transparencia y acceso a la información es quien ha ejercido la función de promover políticas de transparencia, cuenta con un

sitio web con normatividad de conformidad a su misión y un portal de datos abiertos con 40 conjuntos de datos abiertos [210], [211].

Aunque Latinoamérica destaca positivamente comparado a sus pares, el avance en iniciativas de datos abiertos está concentrado en sólo cuatro países. (México, Brasil, Uruguay y Colombia). Los resultados muestran que aunque algunos esfuerzos fueron hechos en el pasado, las iniciativas de datos abiertos presentan una falta de recursos y políticas sustentables y de largo plazo.

A la fecha, los datos abiertos no han sido totalmente adoptados en estos países y son aún vulnerables a los cambios programáticos y la voluntad política de los gobiernos en el corto plazo. Hay también un bajo nivel de compromiso o entendimiento de parte de los gobiernos sobre lo que significa implementar datos abiertos y sus beneficios, o una falta de cooperación con otros actores clave en proyectos de datos abiertos como intermediarios, usuarios, y donantes. Por el contrario, los resultados positivos vienen de gobiernos que han incorporado en sus proyectos políticas y acciones, liberación de datasets, e iniciativas orientadas a generar impacto, además de contar con mayores espacios de coordinación entre las agencias públicas a cargo de los datos abiertos y otros múltiples actores.

CAPITULO 7- TOMA DE DECISIONES EN AGROINDUSTRIA

La toma de decisiones en el ambito de la agroindustria ha sido estudiada desde diferentes perspectivas. Problematicas como el uso del suelo, las politicas de gestion de los recursos naturales y su uso racional han sido abordadas a partir de escenarios donde se incorpora la opinion de una amplio conjunto de partes interesadas para la toma de decisiones informadas [7], [212].

Tambien se han propuesto marcos para la toma de decisiones estrategias, tácticas y operativas en politica de administracion del agua y cadenas de suministro sostenibles [213], [214] evaluando pérdidas económicas agrícolas por la escases de agua y análisis de toma de decisiones bajo diferentes escenarios.

Por otro lado se ha estudiado la manera en la cual los pequeños agricultores toman decisiones encontrando que los principales impulsores son la propiedad de los recursos y el acceso [215], sus preocupaciones a largo plazo son principalmente el posicionamiento estratégico y la sostenibilidad de las granjas y a corto plazo temas de sostenibilidad detallados [216], mientras la presencia de los demás actores de la cadena alimentaria para abastecimiento y acceso a los mercados de sus productos, así como proporcionar información y servicios técnicos son primordiales [217]. La toma de decisiones teniendo en cuenta informacion cuantitativa, (suelo, materia orgánica, tasa de alfabetización, acceso al mercado, mejora de la fertilidad del suelo, agua y agroforestería), así como opiniones cualitativas de una amplia gama de partes interesadas sobre estos criterios tambien ha sido abordada [218].

Por otro lado a pesar de la existencia de multiples herramientas para la toma de decisiones en la agricultura su adopcion es baja. El uso de las herramientas disponibles puede mejorar la eficiencia de la producción, especialmente en la planificación agrícola regional y la gestión[219]. Sin embargo problematicas como la usabilidad, la falta de pertinencia, el alto costo y bajo desempeño son razones que limitan su uso [220].

Varias herramientas para la toma de decisiones a nivel de granja están actualmente disponibles. El Decision Support System for Agrotechnology

Transfer (DSSAT) es un programa de software que comprende modelos de simulación para más de 42 cultivos. Combina bases de datos de cultivos, suelo y clima para simular resultados y el manejo de cultivos. DSSAT integra los efectos del suelo, el fenotipo de la cosecha, el clima y las opciones de manejo. También incluye programas de aplicación mejorados para análisis estacionales, espaciales, secuenciales y de rotación de cultivos que evalúan los riesgos económicos y los impactos ambientales asociados al riego, manejo de fertilizantes y nutrientes, variabilidad climática, cambio climático, captura de carbono y manejo de precisión[221]

En [222] se presenta una herramienta analítica denominada AgBizLogic que integra datos, escenarios, calculadoras económicas y financieras, el clima y el medio ambiente. Genera estimaciones de las repercusiones económicas y resultados de las prácticas de gestión actuales y alternativas. AgBizLogic cuenta con los módulos AgBizProfi (rentabilidad potencial de una inversión dada), AgBizLease (arrendamientos alternativos de corto y largo plazo para cultivos, ganado y otros), AgBizFinance (toma de decisiones de inversión basadas en la liquidez financiera, la solvencia, la rentabilidad y la eficiencia de la granja), AgBizClimate (información sobre el cambio climático) y AgBizEnvironment (erosión, pérdida de suelos, detección del carbono del suelo y emisiones de gases de efecto invernadero GHG asociadas con los niveles de insumos y las prácticas administrativas). Los datos de herramientas como AgBizLogic necesitan ser integrados con los sistemas basados en la nube y con los otros datos públicos, pues el enfoque actual de llevar a cabo manualmente esta integración hace que este tipo de análisis sea costoso.

Por otro lado la plataforma inteligente para la gestión agraria (smart agri-management platform) fue desarrollada por el instituto para la información de la industria de Taiwan como un servicio integrado en la nube. Inspirado en los datos abiertos de los precios de los cultivos, se desarrolló un servicio de predicción de precios de cosechas que recupera los precios históricos del sitio web oficial y pronostica de precios, usando series de tiempo [223].

Grandes compañías como Monsanto y Dupont han desarrollado sus propias herramientas para facilitar la gestión de la cosecha y la toma de decisiones. Los Sistemas de Agricultura Integrados (Integrated Farming Systems (IFS)) son la plataforma de Monsanto que proporciona a los agricultores recomendaciones, incluyendo maneras de aumentar el

rendimiento, optimizar insumos y mejorar la sostenibilidad de la finca. Monsanto también ofrece tres ofertas de productos escalonados de Climate FieldView™ (Prime, Plus y Pro), con su oferta premium conecta los datos del campo a la nube. Esta plataforma digital permite a los agricultores combinar sus datos de suelo, clima y rendimientos de los cultivos con datos históricos y en tiempo real para administrar sus operaciones cotidianas, lo que permite al agricultor disminuir los riesgos que conlleva la agricultura moderna [224]. Por su lado DuPont cuenta con el software Pioneer Field360 Select, una herramienta basada en la web que combina datos con información agronómica y meteorológica en tiempo real para ayudar a los productores a tomar decisiones de gestión informadas. Este programa está respaldado por datos sobre el suelo, el clima y las precipitaciones que abarcan muchas décadas. También es apoyado por los estudios de datos de Pioneer Agronomy Sciences para determinar las etapas de crecimiento de los híbridos de maíz de la marca Pioneer[225].

En el Mercado también se encuentran disponibles aplicaciones para dispositivos móviles que ayudan a los productores agrícolas a tomar decisiones. Hay aplicaciones con el propósito de identificar mala hierba (Ag Weed ID, Alberta Weed Spotter), enfermedades (BASF disease ID, The Cereal Disease ID App), insectos (GRDC Insect ID, Pestbook by DuPont) y nutrientes (Crop Nutrient Deficiency Photo Library). Otras aplicaciones mantienen informados a los agricultores (DTN/The Progressive Farmer: Agriculture News, AgWeb News & Markets)), y proporcionan facilidad para toma de decisiones en cuanto al ganado (ThermalAid, iHerd) o la maquinaria (My Shed - Case IH, JDLink). Los tipos de datos requeridos varían durante el ciclo de vida agrícola, desde la planificación de la temporada de cultivo hasta la comercialización de los productos [226]

Dentro del movimiento de datos abiertos los gobiernos, los centros de investigación y las empresas trabajan para disponer conjuntos de datos que promuevan la reutilización de la información y la toma de decisiones. Global Open Data for Agriculture and Nutrition (GODAN, www.godan.info) es una iniciativa para la apertura de datos en el área de seguridad alimentaria. Es un grupo con más de 400 socios de gobiernos nacionales, organizaciones no gubernamentales, organizaciones internacionales y del sector privado que promueve la colaboración para aprovechar el creciente volumen de datos generados por las nuevas tecnologías para resolver problemas de larga data

y para beneficiar a los agricultores y la salud de los consumidores[227].

En cuanto al uso de la ingeniería dirigida por modelos en la agroindustria, en [228] proponen un metamodelo para estimar la disponibilidad de agua en el suelo de cultivos de uvas y [140] presentan una infraestructura para la construcción automática de sistemas de monitoreo de las condiciones internas de los granos almacenados en silos bolsa. Para ello definieron un lenguaje de dominio específico que permite configurar distintos sistemas de monitoreo, los cuales se construyen automáticamente a partir de dichas configuraciones.

7.1 Datos abiertos y Agroindustria

En [229] se presentan tres maneras específicas en que los datos abiertos pueden ayudar a resolver problemas prácticos en los sectores de agricultura y nutrición: i) hacer una toma de decisiones más eficiente y eficaz, ii) fomentar la innovación de la que todos pueden beneficiarse e iii) impulsar el cambio organizacional y sectorial a través de la transparencia. A continuación se mencionan algunos casos de éxito.

7.1.1 Toma de decisiones más eficiente y eficaz

Hemos llegado a un punto en el que se puede obtener información de una variedad de fuentes en línea cuando lo necesitamos. A medida que se ofrece más información en línea, esto nos ayuda a tomar decisiones más rápidas y más informadas. Una forma clave en la que los datos abiertos están transformando el sector agrícola es apoyando una variedad de herramientas para los propios agricultores. Se han construido una serie de aplicaciones y servicios utilizando datos abiertos para ayudar a los agricultores a mantenerse informados sobre diversos problemas que afectan su trabajo.

GroenMonitor

La productividad de la granja es a menudo afectada por daños a las cosechas causados por plagas. Los ratones y otras plagas son difíciles de detectar en granjas a través de solo la inspección manual. GroenMonitor es una herramienta que muestra un mapa de vegetación actual de los Países

Bajos, basado en imágenes de satélite y mapas puestos a disposición del público a través de la Agencia Espacial Europea (ESA)[230]. La herramienta está siendo explotada para otras aplicaciones, incluyendo la fenología de las plantas, identificación de cultivos y rendimiento, identificación de actividades agrícolas (por ejemplo, corte, arado y cosecha), manejo del agua y control de incendios.

Plantwise

Plantwise es una herramienta que ayuda a los pequeños agricultores de los países en desarrollo con problemas fitosanitarios. Su objetivo es aumentar la seguridad alimentaria y mejorar los medios de subsistencia rurales mediante la reducción de las pérdidas de cultivos de plagas y enfermedades. Lo hace combinando los datos de acceso abierto global y local de fuentes como las bases de datos del Centro Internacional de Agricultura y Biociencias (Centre for Agriculture and Biosciences International- CABI), publicaciones de investigación y datos gubernamentales. Hace los datos disponibles y fáciles de buscar a través de una plataforma en línea.

Trabajando en estrecha colaboración con los servicios nacionales de asesoramiento agrícola, se establecieron redes sostenibles de clínicas de plantas, dirigidas por médicos de plantas capacitados, donde los agricultores pueden encontrar consejos prácticos sobre salud de las mismas. Las clínicas de plantas trabajan de manera similar a las clínicas para la salud humana: los agricultores acuden con muestras de sus cultivos (figura 20) y los médicos diagnostican el problema y hacen recomendaciones científicas sobre las maneras de manejarlas.

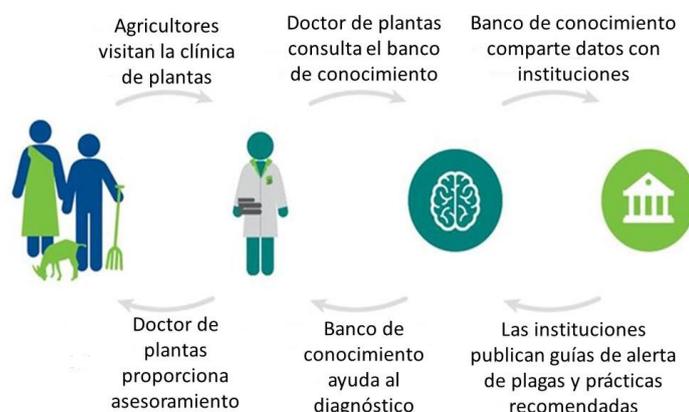


Figura 20. ¿Cómo funciona una clínica de plantas?

Fuente [231]

Las clínicas de plantas son reforzadas por el Plantwise Knowledge Bank, una puerta de acceso a información, en línea y fuera de línea, de la salud de la planta, incluyendo recurso de diagnóstico, asesoramiento de mejores prácticas en el manejo de plagas y análisis para la protección de cultivos.

Aclímate Colombia

Una reciente colaboración entre fuentes de datos públicas (abiertas) y privadas ayuda a los agricultores a tomar precauciones para evitar el daño causado por la sequía en Colombia. Entre 2007 y 2013, una asociación de agricultores (Fedearroz), el Centro Internacional de Agricultura Tropical (CIAT) y el Ministerio de Agricultura de Colombia se unieron para identificar los problemas que favorecen las reducciones anuales de los rendimientos de los cultivos de arroz.

Utilizando tanto datos abiertos como privados (los últimos obtenidos de empresas a través de acuerdos especiales), el CIAT analizó grandes conjuntos de datos de las encuestas anuales sobre el arroz, registros de cosecha, experimentos de campo y datos meteorológicos e identificó los problemas complejos y específicos de los rendimientos de la región. Esto llevó al desarrollo de una herramienta de toma de decisiones sobre la agricultura inteligente para los productores de arroz colombianos, la cual está abierta a todo el mundo. En la figura 21 se puede observar el tablero disponible con la fecha mas favorable para sembrar cierta variedad de arroz y su rendimiento potencial.

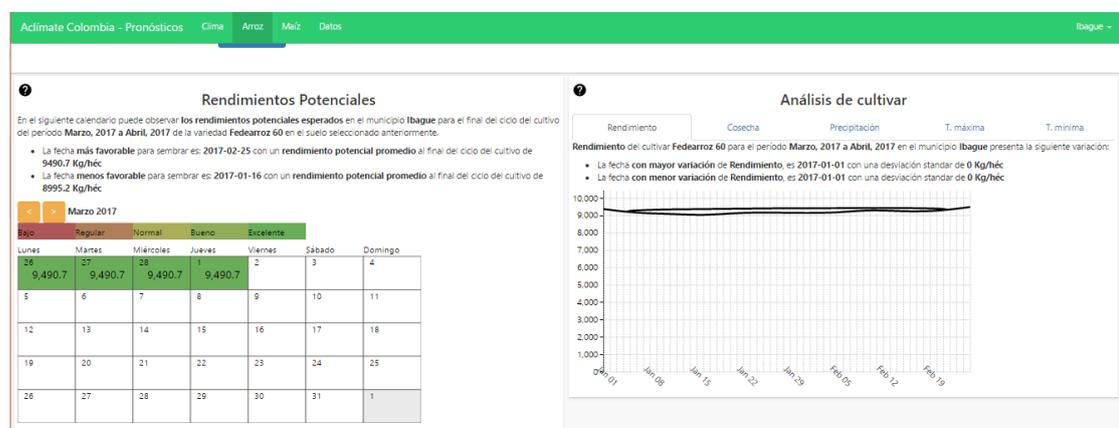


Figura 21. Rendimientos potenciales y análisis para cultivar arroz en marzo 2017
Fuente: [232]

7.1.2 Fomentar la innovación

Como materia prima para crear herramientas, servicios, conocimientos y aplicaciones, los datos abiertos hacen que sea barato y fácil crear nuevas innovaciones. Cuando los datos están abiertos, también se pueden realizar nuevas combinaciones con otros datos para obtener nuevos conocimientos. Beneficia a las empresas que aprenden y reaccionan ante la innovación en su sector, pudiendo invertir en estos nuevos productos y servicios, adquirir nuevos talentos y ajustar sus propias prácticas comerciales.

Climate Corporation

En el pasado, los agricultores lucharon con modelos climáticos predictivos que no tuvieron en cuenta las condiciones locales, lo que condujo a cálculos de riesgo ineficientes. Climate Corporation es un negocio abierto de datos que ofrece un seguro más preciso y un servicio de asesoramiento comercial para ayudar a los agricultores a manejar y adaptarse al cambio climático[233].

Lo hacen analizando volúmenes enormes de datos de fuentes abiertas y otras, para simular eventos meteorológicos y evaluar el riesgo para el rendimiento de cultivos específicos. La compañía utiliza datos abiertos de fuentes como la Administración Nacional Oceánica y Atmosférica (National Oceanic and Atmospheric Administration -NOOA), the Next Generation Radar (una red de 159 estaciones de radar Doppler operadas por el Servicio Meteorológico Nacional), así como mapas de terreno y tipos de suelos de estudios geológicos de EE.UU.

Portal AGRIS

Es una red internacional de instituciones de investigación y nodos de información que permite que la información sobre la investigación agrícola esté disponible a nivel mundial. Recopila y difunde información bibliográfica sobre diversas publicaciones alimentarias y agrícolas, de más de 150 proveedores de datos en 65 países diferentes.

AGRIS utiliza datos bibliográficos como un agregador para localizar contenido relacionado en línea y lo organiza a través de un repositorio de datos abierto (de más de 8 millones de registros) [234]. Una aplicación combina

registros con otros repositorios de datos abiertos y enlaces a otras fuentes de datos de calidad como el Banco Mundial, Nature y la Base de Datos de Germoplasma de China. AGRIS abarca una amplia gama de temas relacionados con la agricultura, incluyendo la silvicultura, la ganadería, las ciencias acuáticas y la pesca, la nutrición humana y la extensión. Su contenido incluye literatura gris única, como informes científicos y técnicos inéditos, tesis, artículos de conferencias, publicaciones gubernamentales y más.

Para buscar en AGRIS, simplemente se ingresa una palabra clave en el cuadro de búsqueda, figura 22, y el resultado de la búsqueda mostrará una página a la izquierda con los datos bibliográficos y a otra a la derecha con los recursos enlazados y un área de búsqueda avanzada.

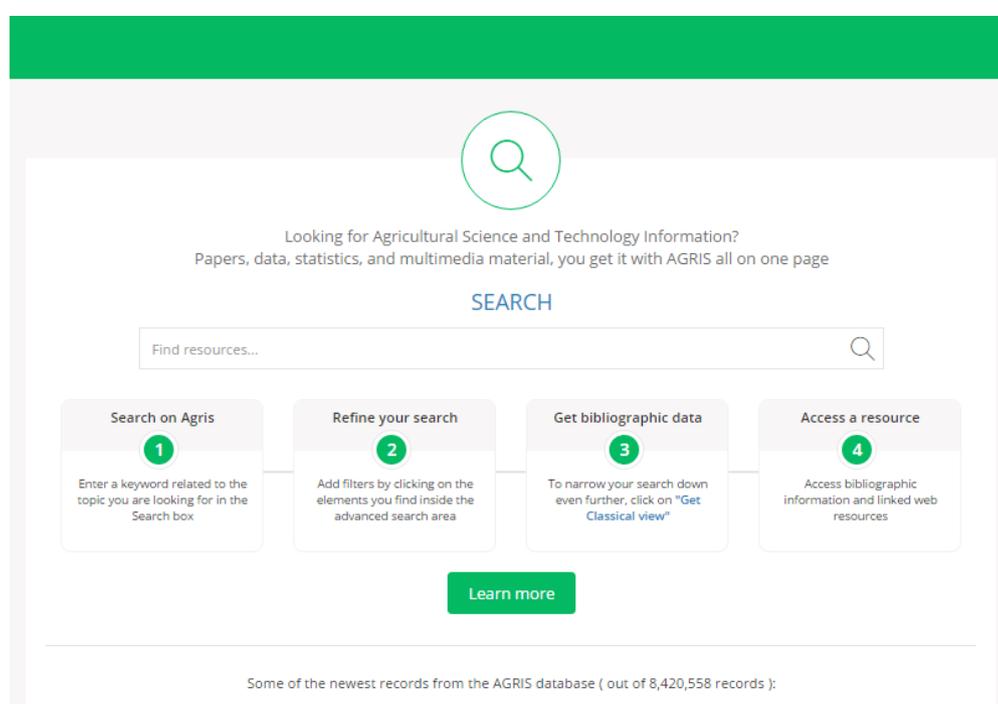


Figura 22. Búsqueda de documentos en AGRIS

Fuente [234]

La Metodología Linked Open Data se utilizó para vincular los conjuntos de datos bibliográficos de AGRIS - Registros, Seriales, Centros - a conjuntos de datos relacionados en la web.

The CIARD RING

A pesar de que ya hay una gran cantidad de información relacionada con datos abiertos (como conjuntos de datos, portales y estándares), la

recuperación de información relevante sigue siendo un tema importante que debe abordarse. En este contexto, el proyecto The Routemap to Information Nodes and Gateways (RING) implementado dentro de la iniciativa Coherence in Information for Agricultural Research for Development (CIARD) actúa como un registro global de información basada en la web, servicios y conjuntos de datos para el desarrollo de la investigación agrícola. El registro permite a los proveedores de datos registrar y categorizar sus servicios. También promueve y aprovecha los estándares, asegurando que todos los conjuntos de datos vayan acompañados de metadatos sobre los cuales se adopten las normas (como vocabularios, dimensiones y protocolos). Esto promueve la reutilización y descubrimiento de datos y permite una mayor automatización.

En la actualidad cuenta con más de 1000 servicios de información, de los cuales alrededor de un tercio son conjuntos de datos agroalimentarios[235]. Está diseñado principalmente para profesionales de la información agrícola y desarrolladores de sitios web que buscan fuentes de datos y servicios de información interoperables y científicos de datos que buscan conjuntos de datos que se pueden procesar.

7.1.3 Impulsar el cambio a través de la transparencia

Al exigir que las empresas, los departamentos gubernamentales y otras organizaciones publiquen conjuntos de datos clave - datos de rendimiento, datos de gasto o datos de la cadena de suministro - los gobiernos, los reguladores y las empresas pueden supervisar, analizar y responder a las tendencias en ese sector. Por otro lado al publicar datos sobre la distribución de los pagos agrícolas, como los subsidios, las agencias gubernamentales pueden hacer seguimiento para que llegue a quienes realmente la necesitan, ayudando así a fortalecer el sector en su conjunto. De allí la importancia del papel del uso de los datos abiertos para impulsar la transparencia.

Plataforma abierta y colaborativa Syngenta

Syngenta en 2013, anunció los compromisos de su Plan de crecimiento (The Good Growth Plan) que incluyen: mejorar la productividad de los cultivos, proteger los suelos y la biodiversidad, capacitar a los pequeños agricultores y garantizar las normas laborales, con metas a alcanzar para

2020 [236]. La iniciativa tiene como objetivo permitir a los agricultores aumentar los rendimientos de los cultivos de manera sostenible, a través de actividades de monitoreo como la aplicación de nutrientes y pesticidas y el uso de agua y combustible.

Se ha establecido un sistema de gestión de datos para realizar un seguimiento de estos productos e insumos agrícolas utilizando datos agrícolas y públicos, que son recopilados, validados y analizados por empresas independientes. A través de estas acciones, Syngenta, en colaboración con Open Data Institute (ODI), ofrecen una plataforma abierta y colaborativa para crear soluciones que minimicen el uso de los recursos necesarios para alimentar a una población creciente y preservar los hábitats para la biodiversidad.

FUNDAR Centro de Análisis e Investigación

PROCAMPO es el mayor programa federal de subsidios agrícolas en México para apoyar a los agricultores más pobres. Desde años atrás ha habido preocupaciones por garantizar que los subsidios fueran recibidos por aquellos que cumplen los requisitos y que en realidad necesitaran el apoyo. Por ello una ONG mexicana llamada FUNDAR pidió información relacionada con la distribución de subsidios de al Ministerio de Agricultura de México. Después de que las solicitudes iniciales dieron como resultado datos incompletos en formatos no legibles por máquina, la agencia encargada finalmente publicó los datos. El análisis mostró que el 57% de los beneficios se distribuyeron entre el 10% de los beneficiarios más ricos [237] confirmando los temores iniciales. De allí fue desarrollada una base de datos denominada Subsidios al Campo en México [238] donde se publica información sobre los subsidios agrícolas para asegurar una mayor transparencia sobre el proceso.

Base nacional de nutrientes de EE.UU

Aunque ya existe información básica sobre el envasado de los alimentos, una información más detallada sobre los nutrientes alimentarios podría permitir a las personas tomar mejores decisiones sobre la selección de alimentos según sus necesidades individuales. La Base Nacional de Nutrientes del EE.UU para la Referencia Estándar (SR25) es la fuente de datos sobre la composición de alimentos en los Estados Unidos y proporciona

fuentes de datos para la mayoría de las bases de datos de los sectores público y privado. El uso de estos datos no se limita a aplicaciones comerciales también proporciona la base para nuevos servicios como ChooseMyPlate.gov, una iniciativa lanzada por la primera dama estadounidense Michelle Obama para proporcionar información práctica a individuos, profesionales de la salud, educadores de nutrición y la industria alimentaria para ayudar a los consumidores a construir dietas saludables con recursos y herramientas para la evaluación dietética y otra información nutricional fácil de usar [239].

Los datos abiertos ayudan y pueden seguir ayudando al sector de la agricultura y la nutrición a enfrentar el reto de alimentar de manera sostenible al mundo en el contexto del crecimiento demográfico, el cambio climático y la volatilidad de los mercados. Los datos abiertos siguen siendo un concepto relativamente nuevo, y sigue habiendo un gran potencial inexplorado para innovar.

*“La mejor manera de empezar algo es dejar de hablar de
ello y empezar a hacerlo”*

Walt Disney

PARTE IV

DESARROLLO DEL PROTOTIPO

CAPITULO 8 - DESCRIPCION DEL PROTOTIPO

En este documento se propone una arquitectura que permite a un usuario escoger una fuente de datos abiertos disponible y a partir de una parametrización su transformación y análisis usando machine learning. Finalmente la visualización de la información se muestra en un dashboard de una aplicación web. Para este caso de estudio se escogió del portal de datos abiertos de Colombia datos.gov.co datasets relacionados con restitución de tierras y cantidad de créditos agropecuarios por departamento.

En el marco de la tecnología para los datos abiertos es pertinente tener en cuenta los elementos y/o dispositivos que producen o generan datos. Con la aparición de tecnologías como Big Data e Internet de las cosas se considera una arquitectura base que soporte la recolección y extracción de datos, tratamiento y análisis de los datos, y finalmente, la visualización de forma consolidada y agrupada para la toma de decisiones.

Esta Arquitectura se define en capas lógicas verticales y cada una de ellas realiza capacidades específicas dentro del esquema tecnológico que utiliza principalmente plataformas Cloud. En la figura 23 se muestra la arquitectura base y sus componentes y en la tabla 4 se explica con detalle cada capa lógica.

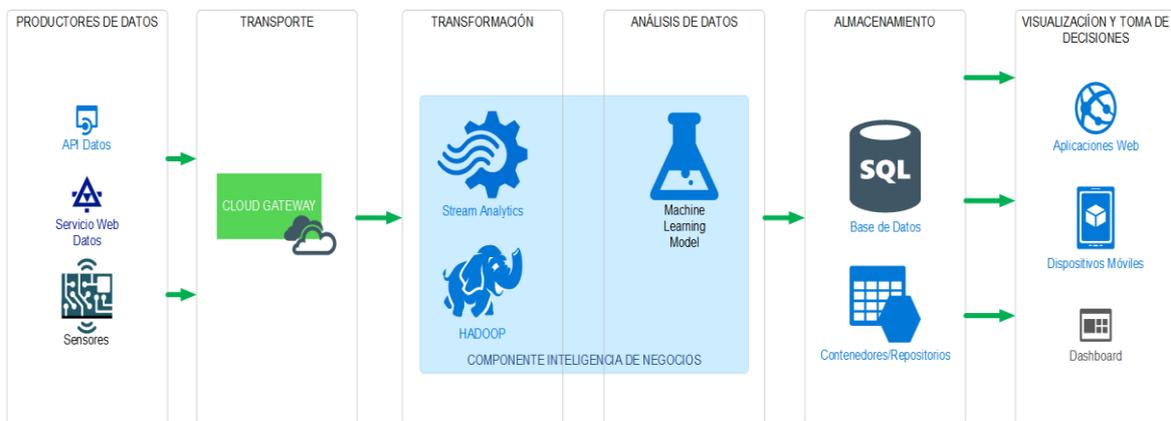


Figura 23. Arquitectura general de la solución

Tabla 4. Componentes / tecnología de la arquitectura base

Capa Lógica	Componente/Tecnología	Propósito
Productores de datos	API de datos	Permiten acceder a servicios de datos/datasets a través de REST
	Servicios Web de datos	Permiten exponer datasets en formato XML
	Sensores	Monitorean sucesos del mundo real, capturan datos y permiten enviarlos a través de Internet
Transporte	Cloud Gateway	Permite realizar la recolección de los datos que se capturan a través de eventos (eventos/segundo) producidos en los dispositivos y aplicaciones
Transformación	Stream Processors	Este componente permite procesar eventos provenientes desde los productores en tiempo real
	Hadoop	Este componente es la implementación que permite el procesamiento de grandes volúmenes de datos
Análisis de datos	Machine Learning Model	Este componente brinda capacidades de analítica de datos aplicando algoritmos y modelos matemáticos, además, puede publicar los resultados de los modelos a través de Web Services
Almacenamiento	Contenedores/Repositorios	Son repositorios permanentes de almacenamiento de los datos estructurados y no estructurados
	Base de Datos	
Visualización/Toma de decisiones	Web Applications	Componente utilizado para presentar la información con acceso a dispositivos como PCs y Laptops.
	Mobile Applications	Componente utilizado para presentar la información desde un dispositivo móvil

El componente Machine Learning dentro de la arquitectura propuesta se puede implementar utilizando soluciones SaaS (Software as a Service) en plataformas Cloud. El aprendizaje automático es una de las aplicaciones más importantes de la inteligencia artificial que evolucionó a partir del estudio del reconocimiento de patrones y de la teoría del aprendizaje computacional. El objetivo es crear y estudiar algoritmos que sean capaces de apoyarse en los datos y hacer predicciones sobre su base [240]. Para este caso, se utilizó Azure Machine Learning que tiene un entorno de desarrollo que permite crear modelos que son entrenados y ajustados con los algoritmos seleccionados con base a los criterios aplicados sobre los datos. Se trabajó con Microsoft Azure machine Learning por que es la plataforma Cloud con mayor tiempo en el mercado, cuenta con el mayor número de servicios para ambientes Cloud, es de fácil acceso, uso y consumo de servicios incluyendo la diferentes suscripciones y opciones de pago, permite una amplia flexibilidad y escalamiento de la infraestructura, arquitectura, aplicaciones y servicios,

satisface los estándares internacionales en seguridad y protección de los datos y contiene herramientas integradas para el desarrollo y gestión de aplicaciones empresariales, Web , móviles, IoT, Big Data y analítica de datos.

El resultado del modelo puede ser explotado con la opción de generar un Web Service (REST) y así, dicha información pueda ser consumida por una aplicación de negocio, aplicación móvil, entre otros.

En la figura 24 se muestra la arquitectura de la Aplicación Web, contiene las siguientes capas lógicas: capa de conexión a datos (Librería de clases), Capa de lógica de datos (Librería de clases), Capa de lógica de Negocio (Librería de Clases), Capa de Presentación (ASP.NET MVC), Capa de Utilidades (Librería de Clases) y Capa de Servicio (ASP.NET con XML Web Services)

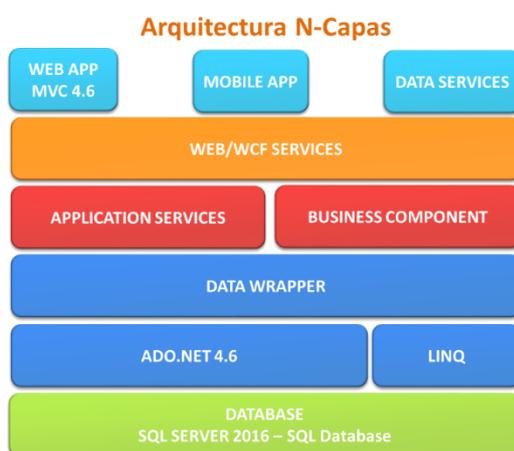


Figura 24. Diagrama de capas aplicación web

En la capa de presentación se usa Google Charts para generar los elementos gráficos del dashboard. En el prototipo, la carga de datos en el Dashboard se realiza a través de Stored (Capa acceso a datos) que para el ejemplo Restitución de Tierras se denomina Agro_SeleccionarAllRestitucionTierras y para Créditos por Departamento es Agro_CantidadCreditosPorDepartamento, y XML Web Services (asmx) que se denomina WSDatosAgro (Capa Servicios). La solución contempla la administración de usuarios, los tipos de fuente de datos y las fuentes de datos. En la base de datos, se registraron fuentes tomadas del portal de datos abiertos datos.gov.co

El modelo de datos de la Aplicación Web, desarrollado utilizando SQL Server 2016, se presenta en la figura 25:



Figura 25. Modelo de datos de la Aplicación Web

Las tablas resaltadas en el círculo rojo conforman el esquema de seguridad para acceso y control de los usuarios, la tabla resaltada en el círculo azul permite administrar los tipos de fuentes de datos del visor de datos mientras que la del círculo naranja permite registrar las fuentes de datos configuradas por el usuario. Por último las tablas del círculo verde permiten almacenar datos procesados a través de las distintas fuentes como Web Services, análisis de datos con Machine Learning y procesadores de flujo eventos y datos.

8. 1 Construcción del DSL

Para la definición del DSL se especificó su Meta-Modelo, el cual representa las construcciones del lenguaje a utilizar, junto con sus propiedades y las relaciones existentes entre las mismas. Integrando Modeling SDK for Visual Studio (MSDK) a Visual Studio 2015 se generó el DSL que permite especificar, de manera gráfica, las construcciones del lenguaje, las relaciones entre ellas y el aspecto gráfico que tendrán en un entorno integrado, cuyo aspecto general se muestra en la Figura 26 El Meta-Modelo es generado a través Domain-

Specific Language Designer lo que permite definir un esquema para la sintaxis abstracta.

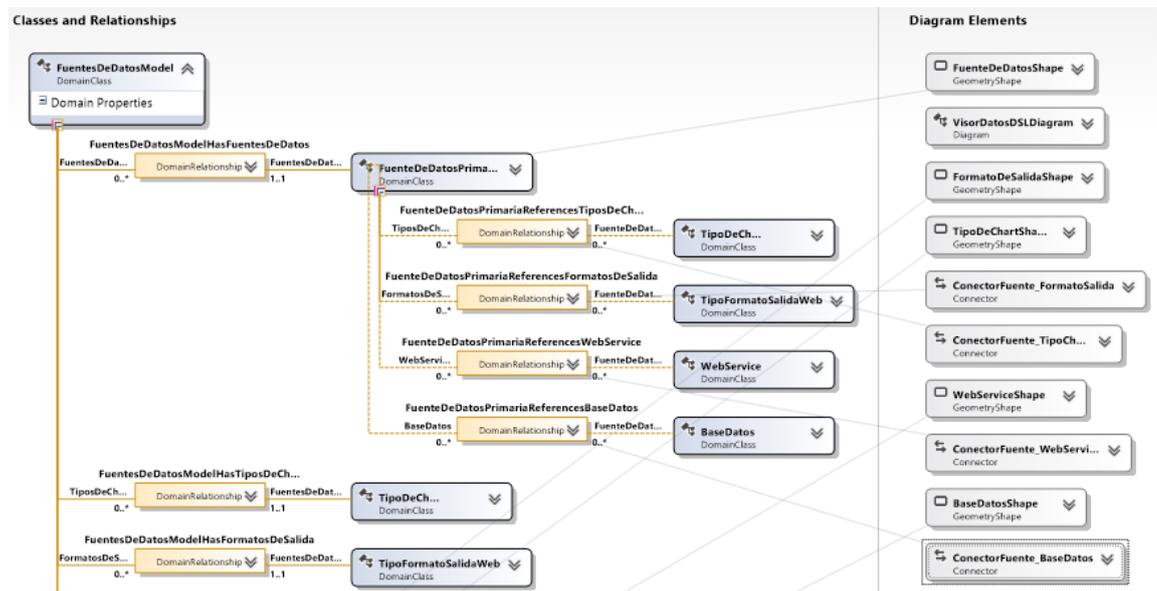


Figura 26. Lenguaje de Dominio específico para el prototipo

En el *Designer* observamos la construcción del modelo y clases de dominio y relaciones entre ellas. El modelamiento muestra las clases que representan las fuentes de datos relacionadas con los productores de datos, éstas a su vez, se especializan en *WebService* y *BaseDatos*. En el caso *WebService* permitiría que un sistema externo pueda consumir los datos expuestos en un tipo de formato como XML, JSON, entre otros, que se representa con la clase *TipoFormatoSalidaWeb*. En el caso *BaseDatos* permitiría extraer datos desde la conexión a una base de datos configurada en la arquitectura base. La visualización de la información en el prototipo también es modelada por la clase *TipoDeChart*, los tipos incluidos son tortas y Barras

Para este prototipo se establece que los cambios y/o modificaciones a las fuentes de datos primarias y las fuentes de salida establecidas se pueden automatizar y controlar a través de las características y capacidades incorporadas al modelo DSL apoyándonos en la herramienta de modelamiento sobre Visual Studio 2015. El backend de la arquitectura propuesta será administrado y controlado de forma manual a través de la configuración proporcionada por la plataforma Cloud Azure y los elementos relacionados con ésta se podrán automatizar a medida que los servicios de dicha plataforma así lo permitan.

El Modelo del DSL creado tiene la estructura que puede verse en la Figura 27:

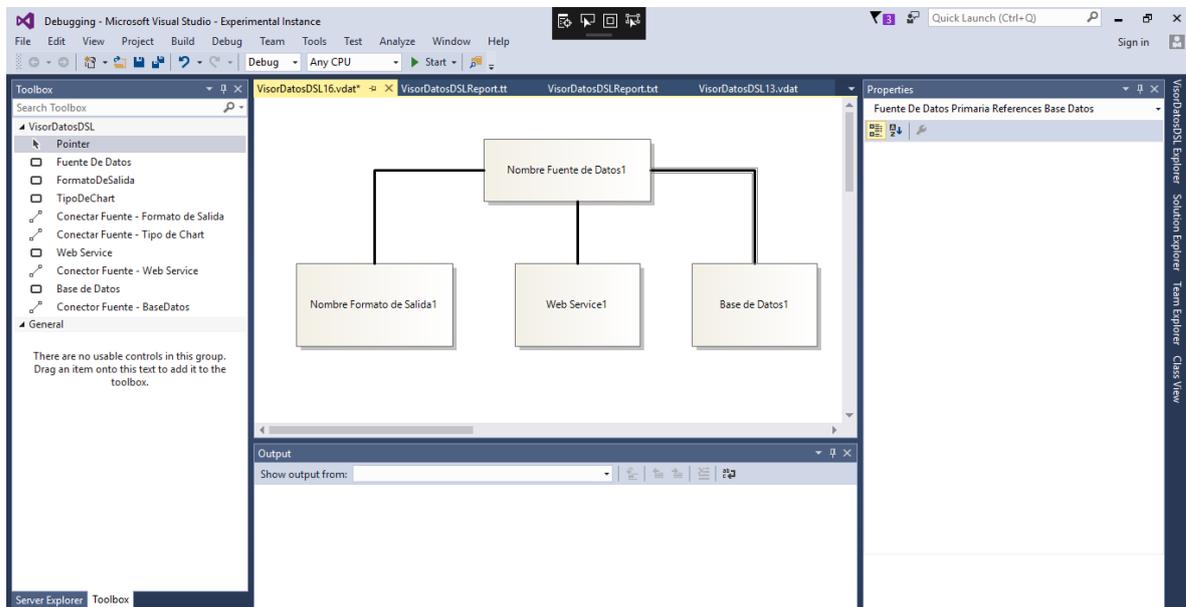


Figura 27. Modelo del prototipo

El modelo ajustado al contexto real permite finalmente crear herramientas de desarrollo basadas en el modelo que podrían contener vistas, generación de código, artefactos, comandos y la capacidad de interactuar con otros componentes u objetos que conformen la solución.

El DSL utiliza una aplicación web que permite la visualización de la información, en la figura 28 se muestra el visor de datos que contiene los parámetros, donde a su vez se crean las fuentes de datos. En la administración de la información del prototipo tenemos principalmente la gestión sobre las fuentes de datos que permite capturar, validar, visualizar, modificar y eliminar los registros creados por el usuario.



[Crear Fuente de Datos](#)

Tipo Fuente	Nombre	Descripción	Ruta	Método	
Web Service	Datos restitución tierras Web Service	Retorna los datos de restitución de tierras por medio de web service	http://localhost/VisorDatos.WSDatosAgro/DatosAgro.asmx	SeleccionarAllRestitucionTierras	Editar Detalle Borrar
Stored Procedure	Datos restitución tierras Stored Procedure	Retorna los datos de restitución de tierras por medio de procedimiento almacenado en la base de datos	Agro_SeleccionarAllRestitucionTierras		Editar Detalle Borrar
Web Service	Cantidad Créditos por Departamento Sector Agropecuario Web Service	Cantidad Créditos por Departamento Sector Agropecuario Web Service	http://localhost:8070/datosagro.asmx	SeleccionarCantidadCreditosPorDepartamento	Editar Detalle Borrar

[Volver al listado de parámetros](#)

© 2016 - Visor Datos

Figura 28. Vista de la creacion de fuentes de datos en la aplicacion web

En el Dashboard aparecen las diferentes opciones de servicios generados por el usuario con base a la parametrización previa. En la figura 29 se visualiza el dashboard con información relacionada con créditos para el sector agropecuario por departamento en Colombia. la fuente de datos consumida y parametrizada en la aplicación Web fue datos.gov.co



Figura 29. Ejemplo de Dashboard usando portal de datos abiertos datos.gov.co

CAPITULO 9 - VALIDACIÓN Y PRUEBAS

Con el propósito de validar el cumplimiento del objetivo de la solución propuesta se desarrollaron las siguientes pruebas:

- **Pruebas de Usabilidad:** Se desarrollo una evaluación heurística, se aplicó un test de usabilidad y una entrevista para conocer el desempeño del prototipo en diversos aspectos.
- **Validación estadística de supuestos e hipótesis:** Se convocó a 11 expertos de la cadena de valor agroindustrial para reflexionar y discutir la contribución de la propuesta al apoyo de toma de decisiones.

9.1 Pruebas de Usabilidad

La ISO define la usabilidad como la capacidad que tiene un producto para ser usado por determinados usuarios con el fin de alcanzar unos objetivos concretos con efectividad, eficiencia y satisfacción dentro de un contexto de uso específico [241]. La efectividad se refiere a la capacidad del sistema para ofrecer las funcionalidades para las que se ha diseñado; la eficiencia al esfuerzo necesario para conseguir realizar estas funcionalidades; y la satisfacción -el aspecto más subjetivo- a la sensación que el usuario tiene mientras lo usa y después de usarlo.

9.1.1 Evaluación Heurística

En las técnicas de evaluación de la usabilidad, se encuentra la evaluación heurística, que consiste en que determinados evaluadores revisan la interfaz siguiendo unos principios de usabilidad reconocidos (heurísticos). La revisión se realiza de manera individual asumiendo el papel de usuario. Hasta que no se realiza completamente la evaluación, no se le permite a los evaluadores comunicar los resultados y sintetizarlos.

La usabilidad se desarrolló formalmente a partir del trabajo de [242], surge, en el ámbito de estudio Interacción persona-ordenador, como una disciplina que busca que los usuarios se sientan cómodos al usar un software determinado. Desde el surgimiento del término usabilidad, autores y organizaciones prestigiosas han realizado aportaciones importantes al tema, y han brindado diferentes factores o atributos que permitan evaluarla, dependiendo cada enfoque de definición según se desea sea medida. El

término solo ha sido utilizado para analizar aquellos factores que contribuyen a que un sitio web resulte fácil de utilizar. El protagonista de la usabilidad es el usuario, a lo cual se le denomina diseño centrado en el usuario (user-centered design)[243].

Teniendo en cuenta [244] se seleccionaron tres revisores. Los principios heurísticos se definieron y la escala de valores con sus significados para cada una de las posibles respuestas. En la plantilla se tuvieron en cuenta aspectos como la navegación, la funcionalidad, el control por parte del usuario, el lenguaje y el contenido, la búsqueda de información, el acceso, la consistencia, la prevención y corrección de errores, la claridad visual y arquitectura del sistema, la flexibilidad y eficiencia. La escala de valores definida constaba de 4 opciones: totalmente en desacuerdo (1), en desacuerdo (2), de acuerdo (3), muy de acuerdo (4). Los resultados de la evaluación se convirtieron en porcentajes y se muestran en la figura 30.

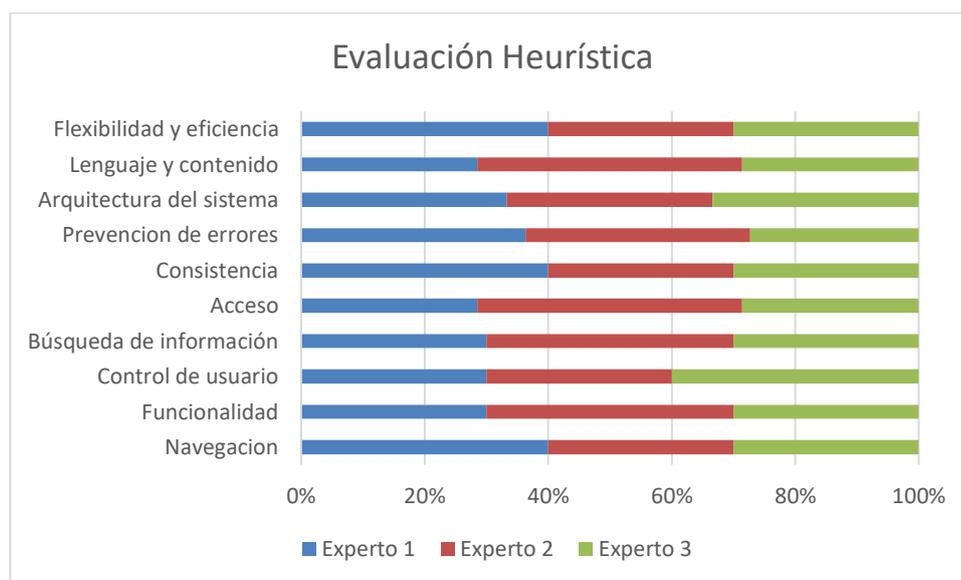


Figura 30. Resultados Evaluación Heurística

Los resultados muestran calificaciones aceptables en los parámetros evaluados, se encontraron problemas en la navegación, accesibilidad y lenguaje. Estos se han mejorado en una versión posterior.

9.1.2 Test de Usabilidad

Se realizó un test de usabilidad en el cual se planteó un conjunto de tareas concretas que el usuario debe realizar expresando en voz alta lo que hace y lo que piensa, mientras un observador toma nota. En este caso no se utilizaron medios de grabación de las sesiones. Se preparó una serie de 6 tareas y se realizó la prueba con 10 usuarios de manera individual.

Para realizar las pruebas se definieron dos perfiles. Por un lado desarrolladores que como parte de su trabajo gestionan sistemas de información y por el otro, usuarios concedores del ámbito de los datos abiertos y que conocen el funcionamiento básico de plataformas. En el caso de los desarrolladores su experiencia permite evaluar la efectividad y la eficiencia dado su conocimiento en las funcionalidades de las aplicaciones. En el caso de los usuarios su punto de vista resulta útil para evaluar si los análisis de datos generan valor para la toma de decisiones.

Se marcaron con un 0 las tareas no cumplidas, con un 1 aquellas conseguidas con dificultad y con un 2 las que se realizaron fácilmente. La tabla 5 muestra cuáles han sido las tareas que se han realizado más satisfactoriamente en el test de usuarios, sin hacer distinción entre desarrolladores y usuarios.

Tabla 5. Resultados test de usabilidad

Tarea	Tipo de Tarea	Porcentajes
1	Crear un usuario	87%
2	Crear fuentes de datos	85%
3	Editar fuentes de datos	90%
4	Administrar los tipos de fuentes de datos	85 %
5	Visualizar el servicio de información	80 %
6	Visualizar datos almacenados	75 %

9. 1.3 Entrevista

Una vez realizados los tests, los observadores mantuvieron una entrevista con cada uno de los usuarios. La entrevista ayudó a percibir cuestiones subjetivas y permitió conocer la sensación tras la prueba. En estas pruebas las preguntas que el evaluador planteaba iban encaminadas a afinar los resultados obtenidos en el test de usuario en aspectos relacionados con la

funcionalidad, facilidad de uso, rapidez, efectividad y satisfacción de la aplicación.

Algunas de las mejoras sugeridas se documentan a continuación:

1. Habilitar funcionalidad de la generación de informes y estadísticas de los procedimientos efectuados.
2. Habilitar comandos de teclado que disminuyan la dependencia del mouse
3. Revisar la función de búsqueda la cual requiere una validación redundante que retrasa el proceso.
4. Contar con un portafolio de fuentes de datos abiertos para seleccionar

Por tratarse de un prototipo se mejoraron aquellos aspectos contemplados en el alcance del proyecto. La solución propuesta tiene una arquitectura escalable que permitirá en futuras ampliaciones mejorar los aspectos uno (1), tres (3) y cuatro (4).

9.2 Validación estadística de supuestos e hipótesis

Se realizó una prueba de hipótesis de proporción con el objetivo de contrastar dos hipótesis estadísticas, aceptando o rechazando la hipótesis nula de aceptación del supuesto respecto a proporción de la población.

La proporción P_0 esperada para aceptar la hipótesis nula en todos los casos será de 10 respuestas positivas de la muestra conformada por 11 participantes. Los participantes fueron seleccionados a conveniencia pues por tratarse de una aplicación enfocada en el sector agroindustrial se requiere personal conocedor del contexto. Luego de una amplia convocatoria a actores involucrados en la cadena de valor de dicho sector, se consiguió una asistencia de 11 expertos. Se generó un espacio de discusión y reflexión alrededor de la problemáticas y el aporte que la propuesta podría hacer a la toma de decisiones. Se consideró un porcentaje de confianza del 95% y por tanto un valor de error o nivel de significancia (alfa) de 0,05. Este valor se debe considerar ya que existe la probabilidad de rechazar la hipótesis nula cuando esta sea verdadera.

Se consideraron las siguientes hipótesis nula H_0 y alternativa H_1 :

$$H_0 = 0,9090$$

$$H_1 < 0,9090$$

De acuerdo a la distribución normal se halló el valor del estadístico Z correspondiente al valor de alfa equivalente a $-1,64$, por medio de prueba de cola izquierda. Se calculó el estadístico de prueba Z_p (Ver ecuación 1).

$$Z_p = \frac{\frac{x}{n} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \quad (\text{Ecuación 1})$$

Donde:

P_0 =Proporción esperado

$\frac{x}{n}$ =Proporción de la muestra

n =Tamaño de la muestra

Z_p =estadístico de prueba a comparar

Z =estadístico z

Los valores Z_p y Z se compararon para tomar la decisión de aceptar o rechazar la hipótesis planteada, en la tabla 6 se detalla la aceptación y rechazo de cada uno de los supuestos planteados.

Tabla 6. Aceptación y rechazo de los supuestos planteados

$Z_p < Z$	Se rechaza la hipótesis nula
$Z_p > Z$	No se rechaza la hipótesis nula

El anterior proceso se realizó para evaluar cada uno de los supuestos planteados que se muestran en la tabla 7.

Tabla 7. Prueba de hipótesis supuestos del metamodelo

AFIRMACIÓN-SUPUESTO	RESPUESTA		CRITERIO
	Si	No	$Z_p > Z$
La disposición de información y su transformación a partir del DSL propuesto, así como su visualización soporta la toma de decisiones en la cadena agroindustrial	100%	0%	$Z_p = -1,049$ $Z = -3.46$
			No se rechaza la hipótesis nula
El modelo propuesto fomenta la disposición de datos públicos y abiertos	64%	36%	$Z_p = -1.64$ $Z = -3.46$

por parte de los involucrados en la cadena agroindustrial			Se rechaza la hipótesis nula
La acción de entidades gubernamentales propuesta en el modelo generará políticas de disposición de información pública y facilita la utilización de datos abiertos	100%	0%	$Z_p = -1.64$ $Z = 1,049$
			No se rechaza la hipótesis nula
La implementación y escalabilidad del DSL propuesto constituyen una alternativa tecnológica viable para la cadena agroindustrial.	91%	9%	$Z_p = -1.64$ $Z = 0$
			No se rechaza la hipótesis nula
La herramienta propuesta permite establecer relaciones transversales con la cadena del sector creando redes de información, con el objetivo de mejorar la toma de decisiones de las partes interesadas	91%	9%	$Z_p = -1.64$ $Z = 0$
			No se rechaza la hipótesis nula
El modelo propuesto implementado a nivel local permitirá crear nuevos nichos de mercado y potenciar la cadena agroindustrial	100%	0%	$Z_p = -1.64$ $Z = 1,049$
			No se rechaza la hipótesis nula.

De conformidad a los resultados obtenidos se puede evidenciar el cumplimiento de la hipótesis planteada. Con la herramienta descrita es posible facilitar el uso de los datos abiertos disponibles por entes públicos y privados con el ánimo de aprovechar su potencial. Una de las necesidades más críticas del sector agroindustrial es consolidar la información. Los expertos consideran que el modelo podría ser una oportunidad y herramienta de cambio siempre y cuando los actores interesados la conozcan y participen activamente en su mejoramiento a partir de su experiencia. Una mejora de esta propuesta es la integración de los datos abiertos con datos propios de tal manera que el usuario pueda usar datos históricos y tomar decisiones a partir de la realidad de su quehacer y la realidad del ambiente en el que se desempeña.

“Todos somos muy ignorantes. Lo que ocurre es que no todos ignoramos las mismas cosas”

Albert Einstein

PARTE V

CONCLUSIONES

CAPÍTULO 10 - CONCLUSIONES Y TRABAJO FUTURO

La arquitectura propuesta pretende contribuir al uso de los datos abiertos dispuestos por entes públicos o privados y proporcionar información valiosa para el apoyo de la toma de decisiones, sin embargo un aporte real dependerá de aspectos técnicos, semánticos y organizacionales.

Desde el punto de vista técnico, es necesario establecer parámetros comunes respecto a la información publicada, la calidad, el formato, las licencias, el procesamiento y la carga en cada sitio. Respecto a la semántica, el reto consiste en estandarizar los formatos y establecer un vocabulario de referencia estándar que las distintas administraciones pueden utilizar para facilitar la vinculación de los datos y ampliar el alcance a soluciones nacionales. Por último, desde el punto de vista organizativo, se requiere un fuerte liderazgo de la Administración pública para identificar cuáles son los datos que generan valor añadido, definir indicadores para medir y evaluar iniciativas y reconocer la importancia de los datos abiertos como motor de generación de beneficios económicos, la transparencia y la mejora de la interoperabilidad vertical y horizontal en las administraciones públicas.

La participación ciudadana real es esencial, la frustración y la falta de confianza en las instituciones en muchos países son altas, por lo que no es ninguna sorpresa cuando la gente expresa escepticismo respecto a que la apertura de los datos del gobierno contribuirá a solucionar problemas. Probablemente los problemas más difíciles de superar no son técnicos o semánticos, sino más bien aquellos relativos al entendimiento de que la información pertenece a todos y que el empoderamiento de las administraciones públicas es esencial para compartir su información con otros en este caso en particular actores de la cadena agroindustrial. En este sentido, la investigación sobre nuevos mecanismos que contribuyan a facilitar la búsqueda, extracción análisis y uso de datos abiertos es necesaria para aprovechar el potencial del Open Government Data.

Con la herramienta descrita es posible facilitar el uso de los datos abiertos disponibles por entes públicos y privados con el ánimo de aprovechar su potencial. Una de las necesidades más críticas del sector agroindustrial es consolidar la información. Los expertos consideran que el modelo podría ser una oportunidad y herramienta de cambio siempre y cuando los actores interesados la conozcan y participen activamente en su mejoramiento a partir de su experiencia. Una mejora de esta propuesta es la integración de los datos abiertos con datos propios de tal manera que el usuario pueda usar datos históricos y tomar decisiones a partir de la realidad de su quehacer y la realidad del ambiente en el que se desempeña.

En el capítulo 1 de este documento se plantearon un grupo de objetivos específicos, los cuales están estructurados con el fin de validar la hipótesis planteada, los mismos han sido cumplidos durante el desarrollo de esta tesis doctoral. Se presenta la contrastación de los mismos.

10.1 Verificación y contraste de la Hipótesis

La tesis doctoral planteo como hipótesis de partida, el siguiente enunciado ***“La incorporación de los beneficios del nuevo paradigma de datos abiertos, mediante un metamodelo que lo integre a soluciones de Inteligencia de negocios, crea valor para organizaciones públicas y privadas mediante la combinación de datos, aprovechando la información disponible en portales web y ofreciendo herramientas enriquecidas para la toma de decisiones”***. Como consecuencia de los resultados alcanzados en la investigación se puede inferir que la misma ha sido validada como positiva.

10.2 Verificación, contraste y evaluación de Objetivos

- ***Diagnosticar la situación actual de datos abiertos. Se identificarán las iniciativas gubernamentales entorno a este reciente paradigma y los avances en implementación y uso. De igual manera se establecerá un estado de arte con referencia a la evolución de las soluciones de inteligencia de negocios y tratamiento de datos.***

Se desarrolló un ejercicio de revisión de publicaciones relacionadas con los datos de abiertos gubernamentales, sus características, usos, formatos, catálogos, barreras y beneficios. También se realizó una revisión de las herramientas de inteligencia de negocios y big data

disponibles en el mercado, esto se evidencia en los capítulos 3, 4 y 6 de la memoria.

- ***Especificar las herramientas tecnológicas aplicables y los requerimientos para la propuesta de un metamodelo.***

A partir de los resultados del objetivo uno (1) se determinó utilizar para el prototipo Microsoft Azure Machine learning para análisis de los datos y para la aplicación web SQL server 2016. En la capa de presentación se usa Google Charts para generar los elementos gráficos del dashboard.

- ***Estudiar las plataformas web y los conjuntos de datos abiertos disponibles.***

Se realizó un estudio de los portales de datos abiertos. Se incluyeron el portal de Estados Unidos, Gran Bretaña y Unión europea por ser los referentes a nivel internacional. Posteriormente se delimitó la búsqueda a los portales de países Latinoamericanos y del caribe por mostrar un desempeño relativamente positivo comparado con el promedio global. Se decidió usar para el prototipo conjuntos de datos del portal colombiano debido a que es la realidad mas cercana para el investigador y dada la conveniencia de contar posteriormente con expertos que permitieron validar la propuesta.

- ***Analizar, diseñar, desarrollar e implementar un prototipo de metamodelo que aproveche los beneficios del nuevo paradigma de datos abiertos .***

Una vez realizada la base documental y seleccionadas las herramientas y conjuntos de datos, se procedió al desarrollo y elaboración de la propuesta en el siguiente orden:

- Arquitectura general de la propuesta
 - Generación del lenguaje de dominio Específico
 - Implementación de la propuesta
- ***Proponer y aplicar pruebas de validación para la propuesta.***

Para la validación de la propuesta se realizaron pruebas de usabilidad y validación estadística de supuestos e hipótesis con la ayuda de 11 expertos en el ámbito agroindustrial.

10.3 Escrutinio Público

10.3.1 Libros

Se publicó el libro resultado de investigación “**Gestión del Conocimiento en el Turismo Salud**” publicado como cuaderno de investigación de la colección de ingeniería informática, tres (3) autores incluido el doctorando con ISBN 84-695-8324 y depósito legal AS: 02507-2013[245]. El autor de esta tesis es el primer autor.

Resumen del Producto

El uso de las TIC`s, se constituye en una herramienta fundamental para lograr una actividad turística competitiva a nivel internacional y mejorar la toma de decisiones en el sector. La gestión del conocimiento permite dotar de valor al producto turístico como forma de adaptación y supervivencia, de diferenciación y competitividad del sector y como herramienta para la toma de decisiones.

10.3.2 Capítulo de libro

Se publicó el capítulo de libro resultado de investigación titulado “Model-driven engineering for Electronic Commerce” publicado en el libro Progressions and Innovations in Model-Driven Software Engineering, DOI: 10.4018/978-1-4666-4217-1.ch007. 2013. Páginas 196-208. Editorial IGI Global Disponible en <http://www.igi-global.com/chapter/model-driven-engineering-electronic-commerce/78213>. El autor de esta tesis es el segundo autor.

Se publicó el capítulo de libro resultado de investigación titulado “DSL TUNNOS Commerce: Model-Driven Architecture applied to E-Commerce Platforms” publicado en Handbook of Research on Innovations in Systems and Software Engineering, DOI: 10.4018/978-1-4666-6359-6.ch020. 2015. Páginas 500 – 509. Editorial IGI Global Disponible en <http://www.igi-global.com/chapter/dsl-tunnos-commerce/117938>. El autor de esta tesis es el tercer autor.

10.3.3 Artículos Publicados (no JCR)

“**Model innovation of process based on the standard e-commerce international GS1**” (2012) en The International Journal of Interactive

Multimedia and Artificial Intelligence– IJIMAI 1(7): 70-75 (2012) DOI: 10.9781/ijimai.2012.178. El autor de esta tesis es el primer autor.

"Open Data as a key factor for developing expert systems: a perspective from Spain" (2013) in International Journal Of Interactive Multimedia And Artificial Intelligence - Ijimai ISSN: 1989-1660, ed: v.2 fasc.2 p.51 - 55, (2013), DOI: 10.9781/ijimai.2013.226[246]. El autor de esta tesis es el primer autor.

"Open data and big data: a perspective from Colombia" (2014) en Lecture Notes In Business Information Processing, ISSN: 1865-1348, ed: Springer Verlag Berlin Heidelberg, v.185 p.35 – 41. El autor de esta tesis es el primer autor.

"State of the Art Construction Based on the J48 Classifier: Case Study of Internet of Things" en Lecture Notes In Business Information Processing, ISSN: 1865-1348, ed: Springer Verlag Berlin Heidelberg, v.224, p.463 - 473, 2015. El autor de esta tesis es el tercer autor.

"Datos abiertos: Oportunidades para empresas" en Revista Ventana Informática, ISSN: 0123-9678, ed: Universidad de Manizales, v.32, p.107 - 118, 2015. El autor de esta tesis es el primer autor.

10.3.4 Artículos Publicados (JCR)

"Reverse electronic auction web tool for B2B" (2014) en Computer in Industry in Special issue on New trends on E-Procurement applying Semantic Technologies. Impact Factor: 1.709. ISSN: 0166-3615. DOI 10.1016/j.compind.2013.12.008. El autor de esta tesis es el quinto autor de cinco autores.

"E-commerce metamodel based on MDA" en IEEE Transaction Latinoamerica Impact Factor: 0.34. ISSN: 1548-0992. El autor de esta tesis es el cuarto autor.

"Metamodel to support decision-making from open government data" en Journal of Ambient Intelligence and Humanized Computing. Impact factor 0,835 DOI 10.1007/s12652-016-0443-7. El autor de esta tesis es el primer autor

10.3.5 Congresos

Se publicó y presentó el paper titulado: **“Arquitectura de Datos Abiertos integrada a inteligencia de negocios para Vigilancia de Salud Laboral”** en el Simposio Doctoral de la CISTI'2013. El mismo se desarrolló del 19 al 22 de junio de 2013.

Se publicó y presentó el paper titulado: **“Open data and big data: a perspective from Colombia** en Ninth International Knowledge Management in Organizations Conference Knowledge Management to improve innovation and competitiveness through Big Data. El mismo se desarrolló del 2 al 5 de septiembre de 2014.

10.4 Líneas de Investigación Futuras

El desarrollo de la tesis motiva a continuar el trabajo en los siguientes campos de investigación como líneas futuras:

Calidad de los datos abiertos: Se han propuesto numerosas aplicaciones usando datos abiertos públicos y privados sin embargo se evidencian grandes falencias en la calidad de los datos publicados. Se requiere diseñar mecanismos que permitan que gobiernos, ciudadanos y empresas identifiquen la demanda real y así evitar la aparición de portales que disponen datos sin valor real. Investigaciones que contribuyan a la implementación de métricas de calidad para metadatos podrían ayudar a aprovechar el potencial de los datos abiertos.

Datos abiertos e internet de las cosas: Las ciudades están constantemente produciendo datos a través de sensores: las estaciones del metro, las cámaras de vigilancia, estaciones medioambientales, etc. Dichos datos, no son abiertos, van directamente a los departamentos y/o empresas responsables, quienes los custodian y quizá publican informes o estadísticas pero no el dato original. Tener información en tiempo real, permite ampliar las posibilidades de reutilización de los datos abiertos. Investigaciones que trabajen en la recepción e indexación de datos de diferentes fuentes, que permitan el análisis y la visualización en formatos fácilmente entendibles por los ciudadanos y máquinas en tiempo real garantiza estar más de la materialización de las denominadas Smartcities.

Datos abiertos e inteligencia de negocios: Los datos abiertos, normalmente se comparten como datos crudos (“raw data”) lo que dificulta al ciudadano medio su análisis para obtener conocimiento útil. Se necesita, entonces, mecanismos que permitan a los ciudadanos comprender y analizar los datos públicos de una manera amigable y sencilla. Investigaciones que propongan alternativas para guiar a usuarios inexpertos en el preprocesado y aplicación de algoritmos de minería, facilitando el análisis y visualización de datos abiertos y generando de manera sencilla informes, análisis multidimensional, cuadros de mando indudablemente aportaran al disminuir las barreras técnicas que hasta ahora aparecen referenciadas en la bibliografía disponible.

REFERENCIAS BIBLIOGRAFICAS

- [1] A. Naser and G. Concha, “Datos abiertos: Un nuevo desafío para los gobiernos de la región,” Chile, 2012.
- [2] T. Jetzek, M. Avital, and N. Bjorn-Andersen, “The Value of Open Government Data: A Strategic Analysis Framework,” *SIG eGovernment pre-ICIS Work. 2012*, no. 2002, 2012.
- [3] The World Bank, “Open data for economic growth,” 2014.
- [4] European Commission, “Open data: An engine for innovation, growth and transparent governance,” 2011.
- [5] G. Vickery, “Review of recent studies on PSI er-use and related market developments,” *Inf. Econ.*, 2011.
- [6] Á. Gomes and D. Soares, “Open government data initiatives in Europe: Northern versus southern countries analysis,” in *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance - ICEGOV '14*, 2014, pp. 342–350.
- [7] P. J. Stephenson, N. Bowles-Newark, E. Regan, D. Stanwell-Smith, M. Diagana, R. Höft, H. Abarchi, T. Abrahamse, C. Akello, H. Allison, O. Banki, B. Batieno, S. Dieme, A. Domingos, R. Galt, C. W. Githaiga, A. B. Guindo, D. L. N. Hafashimana, T. Hirsch, D. Hobern, J. Kaaya, R. Kaggwa, M. M. Kalemba, I. Linjouom, B. Manaka, Z. Mbwambo, M. Musasa, E. Okoree, A. Rwetsiba, A. B. Siam, and A. Thiombiano, “Unblocking the flow of biodiversity data for decision-making in Africa,” *Biol. Conserv.*, 2016.
- [8] J. F. López-Quintero, J. M. Cueva Lovelle, R. González Crespo, and V. García-Díaz, “A personal knowledge management metamodel based on semantic analysis and social information,” *Soft Comput.*, pp. 1–10, 2016.
- [9] E. R. Núñez-Valdez, V. García-Díaz, J. M. Cueva, Y. S. Achaerandio, and R. González-Crespo, “A model-driven approach to generate and deploy videogames on multiple platforms,” *J. Ambient Intell. Humaniz. Comput.*, pp. 1–13, 2016.
- [10] V. García-Díaz, H. Fernández, E. Palacios-González, P. G.-B. B. Cristina, O. Sanjuán-Martínez, and J. M. Cueva L., “TALISMAN MDE: Mixing MDE principles,” *J. Syst. Softw.*, vol. 83, pp. 1179–1191, 2010.
- [11] W. O. Rodrigues, F. Guyomarc’h, and J.-L. Dekeyser, “An MDE Approach for Automatic Code Generation from UML/MARTE to OpenCL,” *Comput. Sci. Eng.*, vol. 15, no. 1, pp. 46–55, Jan. 2013.
- [12] C. E. Montenegro-marín and J. M. Cueva-lovelle, “Domain specific language for the generation of learning management systems modules,” vol. 11, no. 1, pp. 23–50, 2012.
- [13] G. M. Tarazona and L. Rodriguez, “Model-Driven Engineering for Electronic Commerce,” in *Progressions and Innovations in ModelDriven Software Engineering*, vol. i, V. Garcia, J. M. Cueva-lovelle, C. Pelayo-Garcia, and O. Sanjuán, Eds. Hershey USA: Engineering Science Reference (an imprint of IGI Global), 2013, pp. 196–208.
- [14] J. Troya, A. Vallecillo, F. Durán, and S. Zschaler, “Model-driven performance analysis of rule-based domain specific visual models,” *Inf. Softw. Technol.*, vol. 55, no. 1, pp. 88–110, 2013.

- [15] C. Dennis, J. M. Aguilera, and M. Satin, "Tecnologías que dan forma al futuro," in *agroindustrias para el desarrollo*, 2013, pp. 103–135.
- [16] UN-DESA/DSD, "Open Working Group proposal for Sustainable Development Goals," 2014. [Online]. Available: <https://sustainabledevelopment.un.org/?page=view&nr=164&type=230&menu=2059>,.
- [17] Organización de las Naciones Unidas para la Alimentación y la Agricultura, *Agroindustrias para el desarrollo*. Roma, 2013.
- [18] A. Kawtrakul, "Ontology Engineering and Knowledge Services for Agriculture Domain," *J. Integr. Agric.*, vol. 11, no. 5, pp. 741–751, 2012.
- [19] Aqeel-ur-Rehman, A. Z. Abbasi, N. Islam, and Z. Shaikh, "A review of wireless sensors and networks' applications in agriculture," *Comput. Stand. Interfaces*, vol. 36, no. 2, p. 263,260, 2014.
- [20] I. Mohanraj, K. Ashokumar, and J. Naren, "Field Monitoring and Automation Using IOT in Agriculture Domain," *Procedia Comput. Sci.*, vol. 93, no. September, pp. 931–939, 2016.
- [21] N. Wang, N. Zhang, and M. Wang, "Wireless sensors in agriculture and food industry - Recent development and future perspective," *Comput. Electron. Agric.*, vol. 50, no. 1, pp. 1–14, 2006.
- [22] L. Duan and L. Da Xu, "Business intelligence for enterprise systems: a survey," *Ind. Informatics, IEEE Trans.*, vol. 8, no. 3, pp. 679–687, 2012.
- [23] A. Zuiderwijk, M. Janssen, and S. Choenni, "Socio-technical Impediments of Open Data," *Electron. J. eGovernment*, vol. 10, no. 2, pp. 156–172, 2013.
- [24] N. Huijboom and T. Van Den Broek, "Open data: an international comparison of strategies," *Eur. J. ePractice*, pp. 1–13, 2011.
- [25] M. Fioretti, "Open Data : Emerging trends , issues and best practices administration," 2011. [Online]. Available: http://www.lem.sssup.it/WPLem/odos/odos_report_2.pdf.
- [26] A. Abella, "Reutilización de información pública y privada en España," *Rooter Anal. S.L.*, 2011.
- [27] C. S. Albano, "Open government data : a value chain model proposal," in *Proceedings of the 14th Annual International Conference on Digital Government Research*, 2013, pp. 285–286.
- [28] Y. M. González, "Política europea de reutilización de la información del sector público. De la norma Jurídica al portal de datos abiertos.," *Rev. Univ. Eur.*, vol. 19, pp. 113–134, 2013.
- [30] D. Pando and N. Fernández, *El gobierno electrónico a nivel local. Experiencias, tendencias y reflexiones.*, CIPPEC. Buenos Aires, 2013.
- [31] Open Knowledge Foundation, "Open Data Handbook Documentation," 2012. .
- [32] J. Gurin, *Open data now: the secret to hot startups, smart investing, savvy marketing, and fast innovation.* 2013.
- [33] Open Knowledge Foundation, "Manual de los Datos Abiertos," 2012.
- [34] R. A. M. Calenti and F. S. Lorenzo, *Open Data y RISP generando valor social y económico ciudadanos gobierno abierto servicios.* 2011.
- [35] M. H. Lasanta, "Reutilización de datos del sector público: aplicación a la

- gestión de incendios forestales,” Universidad Complutense de Madrid, 2011.
- [36] R. Matheus and J. C. Vaz, “New Perspectives for Electronic Government in Brazil : The Adoption of Open Government Data in National and Subnational Governments of Brazil,” pp. 22–29, 2012.
- [37] Chile. Unidad de Modernización y Gobierno digital, “Norma Técnica para Publicación de Datos Abiertos en Chile.” pp. 1–28, 2013.
- [38] Alianza regional por la libre expresión e información, “Saber más IV Informe regional sobre acceso a la información pública y Alianza de gobierno abierto (Open Government Partnership),” 2012.
- [39] F. Peset, A. Ferrer-Sapena, and I. Subirats-Coll, “Open Data y Linked Open Data: Su Impacto en el Área de Bibliotecas y Documentación,” *El Prof. la Inf.*, vol. 20, no. 2, pp. 165–174, 2011.
- [40] A. Assaf and A. Senart, “HDL - Towards a Harmonized Dataset Model for Open Data Portals,” in *USEWOD-PROFILES@ESWC*, 2015.
- [41] Deloitte, “Market Assessment of Public Sector Information report to Department for Business Innovation and Skills,” 2013.
- [42] Omidyar Network, “Open for Business : How Open Data Can Help Achieve the G20 Growth,” 2014.
- [43] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, “Benefits, Adoption Barriers and Myths of Open Data and Open.pdf,” *Inf. Syst. Manag.*, vol. 29, pp. 258–268, 2012.
- [44] European Commission, “Commercial Exploitation of Europe’s Public Sector Information,” 2000.
- [45] M. Dekkers, F. Polman, R. te Velde, and M. de Vries, “Measuring European Public Sector Information Resources,” 2016.
- [46] DotEcon, “The commercial use of public information (CUPI),” 2006.
- [47] McKinsey Global Institute, “Open data : Unlocking innovation and performance with liquid information,” no. October, 2013.
- [48] M. Stuermer and M. M. Dapp, “Measuring the promise of open data : Development of the Impact Monitoring Framework,” in *Conference for E-Democracy and Open Government*, 2016, pp. 197–203.
- [49] A. Herala, E. Vanhala, J. Porrás, and T. Krri, “Experiences About Opening Data In Private Sector : A Systematic Literature Review,” in *SAI Computing Conference 2016*, 2016, pp. 715–724.
- [50] J. Lindman, T. Kinnari, and M. Rossi, “Business Roles in the Emerging Ecosystem,” *IEEE Softw.*, vol. 33, no. 5, 2016.
- [51] J. Attard, F. Orlandi, and S. Auer, “Value Creation on Open Government Data,” in *International Conference on System Sciences*, 2016, pp. 2605–2614.
- [52] World Wide Web Foundation, “Open Data Barometer Global Report Third Edition,” 2015.
- [53] Fundación COTEC Para la Innovación Tecnológica, “Reutilización de la información del sector público,” 2011.
- [54] R. G. Michener, “The Surrender of Secrecy: Explaining the Emergence of Strong Access to Information Laws In Latin America,” University of Texas,

2010.

- [55] J. Zurada and W. Karwowski, "Knowledge discovery through experiential learning from business and other contemporary data sources: A review and reappraisal," *Inf. Syst. Manag.*, vol. 28, no. 3, pp. 258–274, 2011.
- [56] M. M. Junior, M. S. E. Santos, M. C. R. Vidal, and P. V. R. de Carvalho, "Overcoming the blame game to learn from major accidents: A systemic analysis of an Anhydrous Ammonia leakage accident," *J. Loss Prev. Process Ind.*, vol. 25, no. 1, pp. 33–39, Jan. 2012.
- [57] M. Blakemore and M. Craglia, "Access to Public-Sector Information in Europe: Policy, Rights, and Obligations No Title," *Inf. Soc.*, vol. 22, no. 1, pp. 13–24, 2006.
- [58] R. McLaren and R. Waters, "Governing Location Information in the UK," *Cartogr. J.*, vol. 48, no. 3, pp. 172–178, 2011.
- [59] M. Strathern, "The tyranny of transparency," *Br. Educ. Res. J.*, vol. 26, no. 3, pp. 309–321, 2000.
- [60] K. Rouibah and S. Ould-ali, "PUZZLE: a concept and prototype for linking business intelligence to business strategy," *J. Strateg. Inf. Syst.*, vol. 11, no. 2, pp. 133–152, 2002.
- [61] Gartner, "IT Glossary." [Online]. Available: <http://www.gartner.com/it-glossary/business-intelligence-bi/>. [Accessed: 20-Jul-2002].
- [62] G. Hermosillo, L. Seinturier, and L. Duchien, "Using Complex Event Processing for Dynamic Business Process Adaptation," in *Services Computing (SCC), 2010 IEEE International Conference*, 2010, pp. 466–473.
- [63] S. Chaudhuri, U. Dayal, and V. Narasayya, "An overview of business intelligence technology," *Commun. ACM*, vol. 54, no. 8, pp. 88–98, Aug. 2011.
- [64] J. Duda, "Business intelligence and NoSQL Databases," *Inf. Syst. Manag.*, vol. 1, no. 1, pp. 25–37, 2012.
- [65] S. Mansmann, T. Neumuth, and M. H. Scholl, "OLAP Technology for Business Process Intelligence: Challenges and Solutions," in *International Conference on Data Warehousing and Knowledge Discovery*, 2007, pp. 11–122.
- [66] C. Vercellis, *Business intelligence: data mining and optimization for decision making*. John Wiley & Sons, 2011.
- [67] J. Ranjan and V. Bhatnagar, "Role of knowledge management and analytical CRM in business : data mining based framework," *Learn. Organ.*, vol. 18, no. 2, pp. 131–148, 2011.
- [68] D. Airinei and D. Homocianu, "The Mobile Business Intelligence Challenge," *Econ. Informatics*, vol. 10, no. 1, pp. 5–12, 2010.
- [69] J. Parenteau, R. L. Sallam, C. Howson, J. Tapadinhas, K. Schlegel, and T. W. Oestreich, "Magic Quadrant for Business Intelligence and Analytics Platforms," 2016.
- [70] S. Ccance, "Manual Power BI," 2016.
- [71] Microsoft, "¿Que es Power BI?," 2017. [Online]. Available: <https://powerbi.microsoft.com/es-es/what-is-power-bi/>.
- [72] Tableau, "Tableau aprendizaje," 2017. [Online]. Available: <https://www.tableau.com/es-es/learn/training>.

- [73] C. Murillo, A. Ardao, M. Giraldo, and G. Garay, “Estudio de las herramientas líderes de inteligencia de negocios basado en el cuadrante mágico de Gartner,” in *XVI convención y feria internacional Informática 2016*, 2016.
- [74] C. Ilacqua, H. Cronstrom, and J. Richardson, *Learning Qlik Sense®: The Official Guide*. Packt Publishing Ltd., 2015.
- [75] Qlik, “Why Qlik® is different,” 2017. [Online]. Available: <http://www.qlik.com/us/products/why-qlik-is-different>.
- [76] B. Hočevár and J. Jaklič, “Assessing benefits of business intelligence systems - a case study,” *Manag. J. Contemp. Manag. Issues*, vol. 13, no. 2, pp. 87–119, 2008.
- [77] L. T. Moss and S. Atre, *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley information Technology series, 2003.
- [78] A. Carver and M. Ritacco, “The Business Value of Business Intelligence. A Framework for Measuring the Benefits of Business Intelligence,” 2006.
- [79] H. J. Watson, “Tutorial: Business Intelligence – Past , Present , and Future,” *Commun. Assoc. Inf. Syst.*, vol. 25, 2009.
- [80] Ö. Işık, M. C. Jones, and A. Sidorova, “Business intelligence success: The roles of BI capabilities and decision environments,” *Inf. Manag.*, vol. 50, no. 1, pp. 13–23, Jan. 2013.
- [81] J. Ranjan, “Business Intelligence: Concepts , Components , Techniques and Benefits,” *J. Theor. Appl. Inf. Technol.*, pp. 60–70, 2009.
- [82] A. a. Mohamadina, M. R. B. Ghazali, M. R. B. Ibrahim, and M. a. Harbawi, “Business Intelligence: Concepts, Issues and Current Systems,” in *2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT)*, 2012, pp. 234–237.
- [83] F. Azma and M. A. Mostafapour, “Business intelligence as a key strategy for development organizations,” *Procedia Technol.*, vol. 1, pp. 102–106, Jan. 2012.
- [84] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung, “Big data: The next frontier for innovation, competition, and productivity,” *McKinsey Global Institute*, 2011. .
- [85] M. Chen, S. Mao, and Y. Liu, “Big Data: A Survey,” *Mob. Networks Appl.*, vol. 19, no. 2, pp. 171–209, Jan. 2014.
- [86] R. Chaiken, B. Jenkins, P. Å. Larson, B. Ramsey, D. Shakib, S. Weaver, and J. Zhou, “SCOPE: easy and efficient parallel processing of massive data sets.,” in *Proceedings of the VLDB Endowment*, 2008, pp. 1265–1276.
- [87] D. Beaver, S. Kumar, H. C. Li, J. Sobel, and P. Vajgel, “Finding a Needle in Haystack: Facebook’s Photo Storage,” in *In OSDI*, 2010, vol. 10, pp. 1–8.
- [88] G. DeCandia, D. Hastorun, M. Jampani, G. Kakulapati, A. Lakshman, A. Pilchin, and W. Vogels, “Dynamo: amazon’s highly available key-value store,” *ACM SIGOPS Oper. Syst. Rev.*, vol. 41, no. 6, pp. 205–220, 2007.
- [89] A. Lakshman and P. Malik, “Cassandra: structured storage system on a p2p network,” in *Proceedings of the 28th ACM symposium on Principles of distributed computing*, 2009.

- [90] L. George, *HBase: The Definitive Guide: Random Access to Your Planet-Size Data*. O'Reilly Media, 2011.
- [91] K. Chodorow, *MongoDB: the definitive guide*. O'Reilly Media, Inc., 2013.
- [92] J. Murty, *Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB*. O'Reilly Media, Inc., 2008.
- [93] J. C. Anderson, J. Lehnardt, and N. Slater, *CouchDB: the definitive guide*. O'Reilly Media, Inc., 2010.
- [94] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [95] M. Sard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly, "Dryad: distributed data-parallel programs from sequential building blocks," *ACM SIGOPS Oper. Syst. Rev.*, vol. 41, no. 3, pp. 59–72, 2007.
- [96] C. Moretti, J. Bulosan, D. Thain, and P. J. Flynn, "All-pairs: An abstraction for data-intensive cloud computing," in *Parallel and Distributed Processing IPDPS 2008*, 2008, pp. 1–11.
- [97] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski, "Pregel: a system for large-scale graph processing," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, 2010, pp. 135–146.
- [98] Unbounce, "The Ultimate Guide to A / B Testing." 2015.
- [99] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion : A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, 2013.
- [100] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier. 2011.
- [101] P. Berkhin, "A Survey of Clustering Data Mining Techniques," in *Grouping Multidimensional Data: Recent Advances in Clustering*, J. Kogan, C. Nicholas, and M. Teboulle, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 25–71.
- [102] KDnuggets, "What Analytics, Data Mining, Data Science software/tools you used in the past 12 months for a real project Poll," 2017. [Online]. Available: <http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-software-used.html>.
- [103] Rapidminer Inc., "Rapidminer," 2017. [Online]. Available: <https://rapidminer.com/>.
- [104] The R Foundation, "The R Project for Statistical Computing." [Online]. Available: <https://www.r-project.org/>.
- [105] Microsoft, "SQL server 2016," 2017. [Online]. Available: <https://www.microsoft.com/es-es/sql-server/sql-server-2016>.
- [106] Python Software Foundation, "Python." [Online]. Available: <https://www.python.org/>.
- [107] P. E. Valencia, "Optimizacion mediante algoritmos genéticos," *An. del Inst. Ing. Chile*, pp. 83–92, 1997.
- [108] M. Paluszek and S. Thomas, "An Overview of Machine Learning," in *MATLAB Machine Learning*, Berkeley, CA: Apress, 2017, pp. 3–15.
- [109] C. Augusto, H. Vega, and J. Pariona, "Procesamiento de lenguaje natural,"

- Rev. Ing. Sist. e informática*, vol. 6, no. 2, pp. 45–54, 2009.
- [110] N. K. Bose and P. Liang, *Neural Network Fundamentals with graphs, algorithms and applications*. Mc Graw-Hill, 1996.
- [111] S. Chatterjee and A. S. Hadi, *Regression analysis by example*, 5ta ed. Wiley Publishing, Inc., 2015.
- [112] S. Fotheringham and P. Rogerson, *Spatial analysis and GIS*. CRC Press, 2013.
- [113] F. A. Nava, *Procesamiento de series de tiempo*. Fondo de Cultura Económica., 2015.
- [114] M. Schonlau, “The clustergram : A graph for visualizing hierarchical and nonhierarchical cluster analyses,” *stata J.*, vol. 2, no. 4, pp. 391–402, 2002.
- [115] A. Broder and M. Mitzenmacher, “Network Applications of Bloom Filters: A Survey,” *Internet Math.*, vol. 1, no. 4, pp. 485–509, 2011.
- [116] J. D. Holt. and S. M. Chung, “Mining association rules using inverted hashing and pruning,” *Inf. Process. Lett.*, vol. 83, no. 4, pp. 211–220, 2002.
- [117] K. Asanovic, R. Bodik, J. Demmel, T. Keaveny, K. Keutzer, J. Kubiawicz, and D. Wessel, “A view of the parallel computing landscape,” *Commun. ACM*, vol. 5, no. 10, 2009.
- [118] G.-H. Kim, S. Trimi, and J.-H. Chung, “Big-data applications in the government sector,” *Commun. ACM*, vol. 57, no. 3, pp. 78–85, 2014.
- [119] B. G. Kim, S. Trimi, and J. Chung, “Big-Data Applications in the Government Sector.”
- [120] A. McAfee and E. Brynjolfsson, “Big Data: The management revolution,” *Harv. Bus. Rev.*, vol. 90, no. 10, pp. 61–67, 2012.
- [121] C. Yiu, *The Big Data Opportunity*. Policy Exchange, 2012.
- [122] N. Elgendy and A. Elragal, “Big Data Analytics: A Literature Review Paper,” in *Advances in Data Mining. Applications and Theoretical Aspects: 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014. Proceedings*, P. Perner, Ed. Cham: Springer International Publishing, 2014, pp. 214–227.
- [123] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare : promise and potential,” pp. 1–10, 2014.
- [124] H. Demirkan and D. Delen, “Leveraging the capabilities of service-oriented decision support systems : Putting analytics and big data in cloud,” *Decis. Support Syst.*, vol. 55, no. 1, pp. 412–421, 2013.
- [125] F. Provost and T. Fawcett, “Data science and its relationship data-driven decision making,” *Big data*, vol. 1, no. 1, pp. 51–59, 2013.
- [126] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, “The ‘ big data ’ revolution in healthcare accelerating value and innovation,” 2013.
- [127] N. Heerschap, S. Ortega, A. Priem, and M. Offermans, “Innovation of tourism statistics through the use of new big data sources,” *Stat. Netherlands*, 2014.
- [128] P. Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny)*, vol.

- 275, no. 10, pp. 314–347, 2014.
- [129] N. Khan, I. Yaqoob, I. Abaker, T. Hashem, Z. Inayat, W. Kamaleldin, M. Ali, M. Alam, M. Shiraz, and A. Gani, “Big Data : Survey, Technologies, Opportunities, and Challenges,” *Sci. World J.*, 2014.
 - [130] J. Fan, F. Han, and H. Liu, “Challenges of Big Data analysis,” *Natl. Sci. Rev.*, vol. 1, pp. 293–314, 2014.
 - [131] X. Chen, S. Member, and X. Lin, “Big Data Deep Learning : Challenges and Perspectives,” vol. 2, 2014.
 - [132] L. E. I. Xu, C. Jiang, and J. Wang, “Information Security in Big Data : Privacy and Data Mining,” *IEEE Access*, vol. 2, pp. 1149–1176, 2014.
 - [133] R. T. Kaushik and K. Nahrstedt, “T * : A Data-Centric Cooling Energy Costs Reduction Approach for Big Data Analytics Cloud,” in *T* : A data-centric cooling energy costs reduction approach for Big Data analytics cloud* Rini T. Kaushik; Klara Nahrstedt SC '12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, 2012, pp. 1–11.
 - [134] E. Ortoll, A. Canals, and J. Cobarsí, “Principales parámetros para el estudio de la colaboración científica en big science,” *Rev. Española Doc. Científica*, vol. 37, no. 4, pp. 1–11, 2014.
 - [135] J. Trejo and A. Robles, “Conceptos fundamentales de Ingeniería dirigida por Modelos y Modelos de Dominio Específico,” *Rev. Investig. Sist. e Informática*, vol. 7, no. 2, pp. 9–19, 2010.
 - [136] F. Oquendo, “A model-driven formal method for architecture-centric software engineering,” *SIGSOFT Softw. Eng. Notes*, vol. 31, no. 3, pp. 1–13, 2006.
 - [137] D. Granada, J. M. Vara, V. Andrikopoulos, and E. Marcos, “Aplicando la Ingeniería Dirigida por Modelos para soportar la evolución de servicios,” *Novática Rev. la Asoc. Técnicos Informática*, vol. 214, pp. 52–57, 2011.
 - [138] D. Schmidt, “Model-Driven Engineering,” *IEEE Comput. Soc.*, vol. 39, no. 2, pp. 25–31, 2006.
 - [139] J. Ledgard, T. Trejo, A. David, and E. Robles, “Conceptos fundamentales de Ingeniería dirigida por Modelos y Modelos de Dominio Específico,” vol. 7, no. 2, pp. 9–19, 2010.
 - [140] D. Calejari, C. Luna, M. Canabé, F. Sierra, N. Szasz, and C. Pons, “Ingeniería Dirigida por Modelos Aplicada al Control Automático del Almacenamiento en Silos Bolsa,” in *Memorias del II Congreso de agroinformática*, 2010, pp. 623–636.
 - [141] C. Gonzalez, “MIDGAR: interoperabilidad de objetos en el marco de Internet de las Cosas mediante el uso de Ingeniería Dirigida por Modelos,” Universidad de Oviedo, 2017.
 - [142] T. Stahl, M. Völter, J. Bettin, A. Haase, and S. Helsen, *Model-driven software development: technology, engineering, management*. John Wiley & Sons, 2013.
 - [143] V. Diaz, “MDCI: Model-Driven Continuous Integration,” Universidad de Oviedo, 2011.
 - [144] B. Selic, “Models, Software Models and UML,” in *UML for Real: Design of Embedded Real-Time Systems*, L. Lavagno, G. Martin, and B. Selic, Eds. Boston, MA: Springer US, 2003, pp. 1–16.

- [145] S. Kent, "Model driven engineering," in *International Conference on Integrated Formal Methods*, 2002, pp. 286–298.
- [146] L. Fuentes and A. Vallecillo, "Una Introducción a los Perfiles UML," *Novatica*, vol. 168, pp. 6–11, 2004.
- [147] J. Sutherland, "Business Objects in Corporate Information Systems," *ACM Comput. Surv.*, vol. 27, no. 2, pp. 274–276, 1995.
- [148] G. Poels and G. Dedene, "Complexity metrics for formally specified business requirements," in *Proceedings of the Annual Oregon Workshop on Software Metrics (AOWSM'97, 1997*, pp. 1–11.
- [149] C. Atkinson and T. Kühne, "Model-Driven Development: A Metamodeling Foundation," *IEEE Softw.*, vol. 20, no. 5, pp. 36–41, 2003.
- [150] Object Management Group, "About OMG," 2017. [Online]. Available: <http://www.omg.org/gettingstarted/gettingstartedindex.htm>.
- [151] J. Manrubia, "Desarrollo de Software Dirigido por Modelos," Universidad de Oviedo, 2006.
- [152] S. Meliá, "WebSA: un método de desarrollo dirigido por modelos de arquitectura para aplicaciones web," Universidad de Alicante, 2006.
- [153] J. Quintero and R. Anaya, "MDA y el papel de los modelos en el proceso de desarrollo de software," *Rev. EIA*, vol. 8, pp. 131–146, 2007.
- [154] G. Booch, A. Brown, S. Iyengar, J. Rumbaugh, and B. Selic, "An MDA manifesto," *IBM Ration. Softw.*, pp. 1–9, 2004.
- [155] Y. Liu and Y. Ma, "An Approach for MDA Model Transformation Based on JEE Platform," in *4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08.*, 2008.
- [156] S. Sendall, R. Hauser, J. Koehler, J. Küster, and M. Wahler, "Understanding model transformation by classification and formalization," in *Proc. Of Workshop on Software Transformation Systems*, 2004.
- [157] T. Mens and P. Van Gorp, "A Taxonomy of Model Transformation," *Electron. Notes Theor. Comput. Sci.*, vol. 152, pp. 125–142, 2006.
- [158] K. Czarnecki and S. Helsen, "Classification of Model Transformation Approaches," in *OOPSLA'03 Workshop on Generative Techniques in the Context of Model-Driven Architecture*, 2003, pp. 1–17.
- [159] J. Manrubia, "Desarrollo de Software Dirigido por Modelos," 2006.
- [160] F. Truyen, *The Fast Guide to Model Driven Architecture The Basics of Model Driven Architecture*. Cephas Consulting Corp, 2006.
- [161] J. Texier, M. De Giusti, N. Oviedo, G. Villarreal, and A. Lira, "Los beneficios del Desarrollo Dirigido por Modelos en los Repositorios Institucionales," in *II Conferencia Internacional Acceso Abierto, Comunicación Científica y Preservación Digital (BIREDIAL)*, 2012.
- [162] D. S. Kolovos, R. F. Paige, T. Kelly, and F. A. C. Polack, "Requirements for Domain-Specific Languages," in *ECOOP Workshop on Domain-Specific Program Development*, 2006, pp. 1–4.
- [163] S. Cook, G. Jones, S. Kent, and A. Wills, *Domain-specific development with visual studio dsl tools*. Addison-Wesley Professional, 2007.

- [164] A. Van Deursen and P. Klint, "Little Languages: Little Maintenance?," *J. Softw. maintenance*, vol. 10, no. 2, pp. 75–92, 1996.
- [165] U.S. General Services Administration, "data.gov," 2014. [Online]. Available: <http://www.data.gov/>.
- [166] HM Government, "data.gov.uk," 2013. [Online]. Available: <http://data.gov.uk/data>. [Accessed: 27-Feb-2013].
- [167] A. Ferrer-sapena and F. Peset, "Acceso a los datos públicos y su reutilización : open data y open government," pp. 4–5, 2011.
- [168] V. Gimenez, "La reutilización de la información en el sector público : aplicación en la Unión Europea," *e-ciencias la Inf.*, vol. 4, no. 1, pp. 0–11, 2014.
- [169] Unión Europea, "Portal de datos abiertos de la Unión Europea," 2016. [Online]. Available: <https://data.europa.eu/euodp/es/data>.
- [170] U.S. Government, "USAspending.gov," 2017. [Online]. Available: <https://www.usaspending.gov/Pages/Default.aspx>.
- [171] Government of Canada, "Canada's Action Plan on Open Government 2014–16," 2016. [Online]. Available: <http://open.canada.ca/en/content/canadas-action-plan-open-government-2014-16>.
- [172] Congreso Nacional de México, "Ley Federal de Transparencia y Acceso a la Información Pública." pp. 1–22, 2006.
- [173] Congreso Nacional de México, "Reglamento de la ley federal de protección de datos personales en posesión de los particulares." pp. 1–31, 2011.
- [174] Centro de administración y docencia Económica, "Acuerdo que establece el esquema de interoperabilidad y de datos abiertos de la administración pública federal." 2011.
- [175] Instituto Nacional de Transparencia acceso a la Información y Protección de Datos Personales, "Visión, Misión y Objetivos Estratégicos INAI," 2015. [Online]. Available: http://inicio.ifai.org.mx/nuevo/VISIO_MISION_OBJETIVOS_Acuerdo.pdf.
- [176] S. Fumega and F. Scrollini, "Primeros aportes para diseños de políticas de datos abiertos en América Latina," *Derecho Comp. la Inf.*, vol. 21, pp. 3–37, 2013.
- [177] República de Brasil. Presidencia de la República, *Ley 12.527 del 18 de noviembre de 2011. Ley general de acceso a la Información Pública*. 2011.
- [178] Presidencia República Oriental de Uruguay, "Plan de Acción Nacional de Gobierno Abierto 2014-2016," 2014. [Online]. Available: http://www.agesic.gub.uy/innovaportal/file/3801/1/plan_accion_uruguay_2014-2016_f.pdf.
- [179] República Oriental del Uruguay, *Ley 18381 del 7 de octubre de 2008. Ley de acceso a la información de Uruguay*. 2008, pp. 1–54.
- [180] República de Colombia, "Plan de Acción de Colombia - Alianza para el Gobierno Abierto," 2013. [Online]. Available: <http://www.secretariatransparencia.gov.co/estrategias/Documents/plan-de-accion-sep-30.pdf>.
- [181] Fundación CiGob, "Gobierno Abierto: experiencias de implementación. Los casos de EE.UU, Reino Unido, País Vasco y Chile," 2013. [Online].

- Available: <http://www.cigob.org.ar/media/investigaciones/doc-gobierno-abierto.pdf>.
- [182] Gobierno de Chile, “Plan de Acción del Gobierno de Chile 2013-2014,” 2013. [Online]. Available: http://portal.mma.gob.cl/wp-content/uploads/2014/08/9_Lineamientos_PlanAccion_Principio_10OGP.pdf.
- [183] Presidencia de la Republica de Argentina, “II Plan de Acción Nacional de Gobierno abierto de la República de Argentina,” 2013. [Online]. Available: http://www.opengovpartnership.org/sites/default/files/Argentina_NAP2_V.2.pdf.
- [184] República Argentina, “Reglamento general del acceso a la información pública para el poder ejecutivo nacional. Decreto 1172/2003.” Argentina, 2003.
- [185] Asociación de Derechos civiles, “Guia para usar el Decreto 1172-03. Acceso a la Información Pública.” Argentina, 2006.
- [186] Presidencia de Paraguay, “Plan de Acción Gobierno Abierto 2014 - 2016,” 2014. [Online]. Available: <http://www.stp.gov.py/v1/plan-de-accion-de-gobierno-abierto-20142016/>.
- [187] Presidencia de Paraguay, *Ley 5282 del libre acceso al ciudadano a la información pública y la transparencia gubernamental*. 2014.
- [188] Presidencia de Perú, *Plan de Acción Gobierno Abierto Perú 2014-2016*. 2014, pp. 1–19.
- [189] J. Casas, “Mecanismo de revisión independiente: El Perú Informe de avance 2012-13,” 2014.
- [190] República de Bolivia, *Decreto Supremo 28168 del 17 de mayo de 2005*. 2005, pp. 1–6.
- [191] C. J. L. Caicedo, “Datos abiertos gubernamentales Colombia vs Latinoamerica,” *Ambient. móviles*, vol. 66, pp. 21–24, 2014.
- [192] Government of Belize, *Freedom of Information Act - Cap. 13*. 2000.
- [193] Comisión Económica Para America Latina y el Caribe CEPAL, “Acceso a la Información, Participación y Justicia en Temas ambientales en América Latina y el Caribe: Situación Actual, perspectivas y ejemplos de buenas prácticas.” 2013.
- [194] Guyana. Gobierno de la Nación, *Access to Information Act*. Guyana, 2011.
- [195] Jamaica, *Access to Information Act. No. 21-2002*. 2002.
- [196] Presidencia República Dominicana, “Segundo plan de acción 2014-2016 República Dominicana,” 2014. [Online]. Available: http://www.opengovpartnership.org/sites/default/files/2do_Plan_de_Accion_Republica_Dominicana.pdf.
- [197] Republica Dominicana, *Ley General de Acceso a la Información Pública. Ley 200-04*. 2004.
- [198] Puerto Rico. Estado Libre Asociado, “Plan Estratégico de Manejo de Información, Tecnología y telecomunicaciones,” 2012.
- [199] El Salvador. Asamblea Legislativa, “Ley de Regulación de los Servicios de Información Sobre el Historial de Crédito de las Personas.” 2011.
- [200] Guatemala. Comisión Presidencial De Transparencia y Gobierno

- Electrónico, “Segundo Plan De Acción Nacional De Gobierno Abierto.” 2014.
- [201] República de Guatemala, *Decreto número 57-2008 Ley de Acceso a la Información Pública*. 2008, pp. 1–20.
- [202] Gobierno de Honduras, *II Plan de Acción de Gobierno Abierto Honduras 2014-2016*. 2014.
- [203] República de Honduras, *Ley de Transparencia y acceso a la información pública. Decreto No. 170/-2006*. 2006, pp. 1–19.
- [204] República de Nicaragua., *Ley de acceso a la información pública. Ley No. 621 del 16 de Mayo del 2007*. 2007.
- [205] Asamblea Nacional de la República de Nicaragua, *Ley de protección de datos personales. Ley No. 787 del 21 de Marzo del 2012*. 2012, pp. 1–19.
- [206] República de Costa Rica, *Constitución Política de Costa Rica*. 1949.
- [207] Presidencia de Costa Rica, *Ley N° 8432 Contra la Corrupción y el Enriquecimiento Ilícito en la Función Pública*. 2004.
- [208] Gobierno Nacional de Costa Rica, “Informe de Avances de Gobierno Digital en Costa Rica 2013,” 2013.
- [209] Costa Rica. Presidencia de la República, “Alianza para el Gobierno Abierto.” 2013.
- [210] Gobierno de la República de Panamá, “Segundo Plan de acción 2015-2017,” 2015.
- [211] República de Panamá, “Ley de transparencia de la gestión pública. Ley No.6 de 22 de 2002.” pp. 1–39, 2002.
- [212] S. Bech and P. Kristensen, “Land Use Policy Agriculture and landscape interaction — landowners ’ decision-making and drivers of land use change in rural Europe,” *Land use policy*, vol. 57, pp. 759–763, 2016.
- [213] E. Aivazidou, N. Tsolakis, E. Iakovou, and D. Vlachos, “The emerging role of water footprint in supply chain management: A critical literature synthesis and a hierarchical decision-making framework,” *J. Clean. Prod.*, vol. 137, pp. 1018–1037, 2016.
- [214] J. H. Kwakkel, M. Haasnoot, and W. E. Walker, “Comparing Robust Decision-Making and Dynamic Adaptive Policy Pathways for model-based decision support under deep uncertainty,” *Environ. Model. Softw.*, vol. 86, pp. 168–183, 2016.
- [215] C. Singh, P. Dorward, and H. Osbahr, “Developing a holistic approach to the analysis of farmer decision-making: implications for adaptation policy and practice in developing countries,” *Land use policy*, vol. 59, pp. 329–343, 2016.
- [216] M. Robert, A. Thomas, M. Sekhar, S. Badiger, L. Ruiz, H. Raynal, and J.-E. Bergez, “Adaptive and dynamic decision-making processes: A conceptual model of production systems on Indian farms,” *Agric. Syst.*, pp. 1–13, 2016.
- [217] M. Akimowicz, H. Cummings, and K. Landman, “Green lights in the Greenbelt? A qualitative analysis of farm investment decision-making in peri-urban Southern Ontario,” *Land use policy*, vol. 55, pp. 24–36, 2016.
- [218] P. Brandt, K. Butterbach-bahl, and M. C. Rufino, “How to target climate-smart agriculture? Concept and application of the consensus-driven

- decision support framework,” *Agric. Syst.*, no. June 2016, 2015.
- [219] G. wa Mbügwa, S. D. Prager, and J. M. Krall, “Utilization of spatial decision support systems decision-making in dryland agriculture: A Tifton burclover case study,” *Comput. Electron. Agric.*, vol. 118, pp. 215–224, 2015.
- [220] D. C. Rose, W. J. Sutherland, C. Parker, M. Lobley, M. Winter, C. Morris, S. Twining, C. Ffoulkes, T. Amano, and L. V. Dicks, “Decision support tools for agriculture: Towards effective design and delivery,” *Agric. Syst.*, vol. 149, pp. 165–174, 2016.
- [221] J. W. Jones, G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie, “DSSAT Cropping System Model,” *Eur. J. Agron.*, vol. 18, pp. 235–265, 2003.
- [222] S. M. Capalbo, J. M. Antle, and C. Seavert, “Next generation data systems and knowledge products to support agricultural producers and science-based policy decision making,” *Agric. Syst.*, 2016.
- [223] Y. Peng, C. Hsu, and P. Huang, “Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets,” in *Technologies and Applications of Artificial Intelligence (TAAI)*, 2015, pp. 172–175.
- [224] Monsanto, “Monsanto products,” 2016. [Online]. Available: <http://test.monsanto.com/products/pages/integrated-farming-systems.aspx>. [Accessed: 20-Jun-2011].
- [225] Dupont Pioneer, “Dupont Pioneer Products,” 2016. [Online]. Available: <https://www.pioneer.com/home/site/us/products/>. [Accessed: 20-Jun-2011].
- [226] S. Janssen, A. Porter, I. Moore, I. Athanasiadis, J. Foster, J. Jones, and J. Antle, “Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support,” 2015.
- [227] GODAN, “Global open data for agriculture & nutrition,” 2016. [Online]. Available: <http://www.godan.info/about>. [Accessed: 01-Nov-2016].
- [228] G. Barbier, V. Cucchi, and D. R. C. Hill, “Model-driven engineering applied to crop modeling,” *Ecol. Inform.*, vol. 26, no. P2, pp. 173–181, 2015.
- [229] GODAN & ODI, “How can we improve agriculture, food and nutrition with open data?,” 2015.
- [230] Groenmonitor, “Over de Groenmonitor,” 2017. [Online]. Available: <http://www.groenmonitor.nl/>.
- [231] CABI Plantwise, “About Plantwise,” 2017. [Online]. Available: <https://www.plantwise.org/about-plantwise/>.
- [232] CGIAR, “Clima y sector Agropecuario Colombiano adaptacion para la sostenibilidad productiva,” 2017. [Online]. Available: <http://www.aclimatecolombia.org/>.
- [233] The Climate Corporation, “About Us,” 2017. [Online]. Available: <https://www.climate.com/company/>.
- [234] Food and Agriculture Organization of the United Nations FAO, “AGRIS: International Information System for the Agricultural science and technology,” 2017. [Online]. Available: <http://agris.fao.org/content/about>.

- [235] Global Forum on Agricultural Research, “CIARDRING A directory of information services and datasets in agriculture,” 2017. [Online]. Available: <http://ring.ciard.net/about-ring>.
- [236] Syngenta, “The Good Growth Plan Progress Data,” 2017. [Online]. Available: <http://www4.syngenta.com/what-we-do/the-good-growth-plan/progress/progress-open-data>.
- [237] K. Pirker, “Subsidios al campo en México: Contraloría social, nuevas tecnologías e incidencia en políticas públicas,” 2010.
- [238] FUNDAR, “Subsidios al Campo en México,” 2010. [Online]. Available: <http://subsidiosalcampo.org.mx/analiza/informacion-entidades/>.
- [239] United States Department of Agriculture, “choosemyplate.gov,” 2017. [Online]. Available: <https://www.choosemyplate.gov/about-us>. [Accessed: 20-Jul-2003].
- [240] J. Carbonell, R. Michalski, and T. Mitchell, “An Overview of Machine Learning,” *Machine Learning SE - 1*. pp. 3–23, 1983.
- [241] ISO/IEC, “9241-11 Ergonomic requirements for office work with visual display terminals (VDT)s - Part 11 Guidance on usability.” 1998.
- [242] J. Nielsen, “Usability inspection methods,” in *Conference Companion on Human Factors in Computing Systems*, ACM, Ed. New York, New York, USA: ACM Press, 1995, pp. 377–378.
- [243] L. Perurena and M. Moráquez, “Usabilidad de los sitios Web , los métodos y las técnicas para la evaluación Usability of Web sites , methods and evaluation techniques,” *Rev. Cuba. Inf. en Ciencias la Salud*, vol. 24, no. 2, pp. 176–194, 2013.
- [244] J. Nielsen and T. K. Landauer, “A mathematical model of the finding of usability problems,” *Proceeding CHI '93 Proc. INTERACT '93 CHI '93 Conf. Hum. Factors Comput. Syst.*, pp. 206–213, 1993.
- [245] G. M. Tarazona, L. A. Rodriguez, and V. Medina, *Gestión del conocimiento en e l turismo de salud*. Oviedo, 2013.
- [246] L. A. Rodriguez-Rojas, J. M. Cueva-Lovelle, G. M. Tarazona-Bermudez, and C. E. Montenegro-Marin, “Open Data as a key factor for developing expert systems: a perspective from Spain,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 2, no. 2, p. 51, 2013.

*“La inspiración existe pero tiene que encontrarte
trabajando”*

Picasso

ANEXOS

ANEXO 1. Descripción de la metodología de trabajo (scrum)

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Proyecto
DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos

Descripción de la metodología de trabajo (scrum)

Versión 1.0

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Historial de Revisiones

Fecha	Versión	Descripción	Autor
20/9/2016	1.0	Primera versión con los apartados y contenidos básicos	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Juan Manuel Cueva Giovanny M. Tarazona
25/11/2016	2.0	Segunda Versión requerimientos	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Juan Manuel Cueva Giovanny M. Tarazona
20/01/2017	3.0	Tercera Versión requerimientos	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Juan Manuel Cueva Giovanny M. Tarazona
20/02/2017	4.0	Cuarta Versión estimaciones	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Giovanny M. Tarazona Benjamin Torres Alvarez
28/02/2017	4.5	Quinta versión	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Giovanny M. Tarazona Benjamin Torres Alvarez
10/03/2017	5.0	Sexta Versión	Luz Andrea Rodriguez Rojas Carlos E.Montenegro Giovanny M. Tarazona Benjamin Torres Alvarez

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Tabla de Contenido

1. Introducción.....	158
1.1 <i>Propósito de este documento.....</i>	<i>158</i>
1.2 <i>Alcance.....</i>	<i>158</i>
2. Descripción General de la Metodología	158
2.1 <i>Fundamentación.....</i>	<i>158</i>
2.2 <i>Valores de trabajo</i>	<i>159</i>
3. Personas y roles del proyecto.	160
4. Artefactos	¡Error! Marcador no definido.
4.1 <i>Pila de producto.....</i>	<i>161</i>
4.2 <i>Pila del sprint.....</i>	<i>162</i>
4.3 <i>Sprint.....</i>	<i>162</i>
4.4 <i>Incremento</i>	<i>162</i>
4.5 <i>Gráfica de producto (Burn Up)</i>	<i>162</i>
4.6 <i>Gráfica de avance (Burn Down).....</i>	<i>164</i>
4.7 <i>Reunión de inicio de sprint.....</i>	<i>166</i>
4.8 <i>Reunión técnica diaria.....</i>	<i>167</i>
4.9 <i>Reunión de cierre de sprint y entrega del incremento.</i>	<i>168</i>

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Descripción de la metodología de trabajo

1. Introducción

Este documento describe la implementación de la metodología de trabajo scrum como prototipo de desarrollo para validar la hipótesis de la tesis Doctoral METAMODELO PARA INTEGRACIÓN DE DATOS ABIERTOS A INTELIGENCIA DE NEGOCIOS elaborada por Luz Andrea Rodríguez Rojas y Co-dirigida por los Doctores Juan Manuel Cueva Lovalle y Carlos Enrique Montenegro con el apoyo del Dr. Giovanni Tarazona y el ingeniero Benjamín Torres Alvares del grupo de investigación Gicoecol de la Universidad Distrital Francisco José de Caldas en Bogotá Colombia en asocio con MDA Research Group de la Universidad de Oviedo en Asturias España para la gestión del desarrollo el proyecto Prototipo de Validación.

Incluye junto con la descripción de este ciclo de vida iterativo e incremental para el proyecto, los artefactos o documentos con los que se gestionan las tareas de adquisición y suministro: requisitos, monitorización y seguimiento del avance, así como las responsabilidades y compromisos de los participantes en el proyecto.

1.1 Propósito de este documento

Facilitar la información de referencia necesaria a las personas implicadas en el desarrollo del sistema Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos

1.2 Alcance

Personas y procedimientos implicados en el desarrollo del sistema Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos

2. Descripción General de la Metodología

2.1 Fundamentación

Las principales razones del uso de un ciclo de desarrollo iterativo e incremental de tipo scrum para la ejecución de este proyecto son:

- Sistema modular. Las características del sistema Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos permiten desarrollar una base funcional mínima y sobre ella ir incrementando las funcionalidades o modificando el comportamiento o apariencia de las ya implementadas.
- Entregas frecuentes y continuas al cliente de los módulos terminados, de forma que puede disponer de una funcionalidad básica en un tiempo mínimo y a partir de ahí un incremento y mejora continua del sistema.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

- Previsible inestabilidad de requisitos.

Es posible que el sistema incorpore más funcionalidades de las inicialmente identificadas.

Es posible que durante la ejecución del proyecto se altere el orden en el que se desean recibir los módulos o historias de usuario terminadas.

Para el cliente resulta difícil precisar cuál será la dimensión completa del sistema, y su crecimiento puede continuarse en el tiempo suspenderse o detenerse.



Fig. 1. Descripción general de la metodología

2.2 Valores de trabajo

Los valores que deben ser practicados por todos los miembros involucrados en el desarrollo y que hacen posible que la metodología Scrum tenga éxito son:

- Autonomía del equipo
- Respeto en el equipo
- Responsabilidad y auto-disciplina
- Foco en la tarea
- Información transparencia y visibilidad.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

3. Personas y roles del proyecto.

Persona	Contacto	Rol
Luz Andrea Rodríguez Rojas	Ing.rodriuezla@gmail.com	Scrum Manager
Juan Manuel Cueva Lovellev	cueva@uniovi.es	Director Proyecto
Carlos Enrique Montenegro	cemontenegrom@udistrital.edu.co	Co-directo Proyecto
Giovanny Mauricio Tarazona	gtarazona@udistrital.edu.co	Co-investigador
Benjamin Torres Alvarez	btorresal@gmail.com	Equipo Técnico

4. Artefactos

Documentos

- Pila de producto o Product Backlog
- Pila de sprint o Sprint Backlog

Sprint

Incremento

Gráficas para registro y seguimiento del avance.

- Gráfica de producto o Burn Up
- Gráfica de avance o Burn Down.

Comunicación y reporting directo.

- Reunión de inicio de sprint
- Reunión técnica diaria
- Reunión de cierre de sprint y entrega del incremento

El componente Machine Learning dentro de la arquitectura propuesta se puede implementar utilizando soluciones SaaS (Software as a Service) en plataformas Cloud. Para este caso, se utilizó Azure Machine Learning que tiene un entorno de desarrollo que permite crear modelos que son entrenados y ajustados con los algoritmos seleccionados con base a los criterios aplicados sobre los datos.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

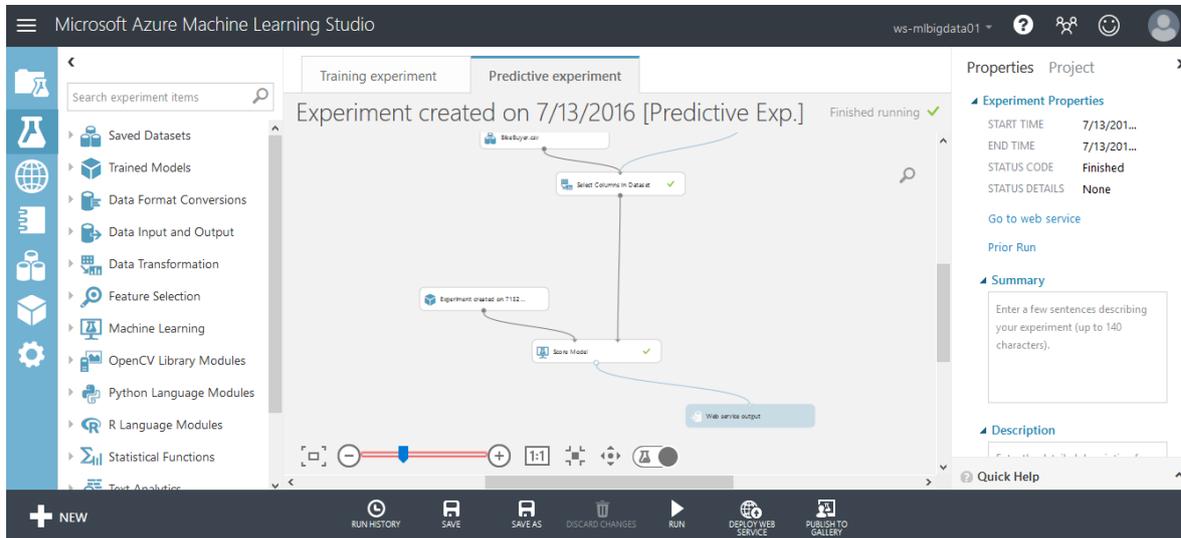


Fig. 1. Microsoft Azure Machine Learning

4.1 Pila de producto

Es el equivalente a los requisitos del sistema o del usuario (Con-Ops) en esta metodología.

El gestor de producto de su correcta gestión, durante todo el proyecto. El gestor de producto puede recabar las consultas y asesoramiento que pueda necesitar para su redacción y gestión durante el proyecto al Scrum Manager de este proyecto.

Responsabilidades del gestor de producto

- Registro en la lista de pila del producto de las historias de usuario que definen el sistema.
- Mantenimiento actualizado de la pila del producto en todo momento durante la ejecución del proyecto.
 - Orden en el que desea quiere recibir terminada cada historia de usuario.
 - Incorporación / eliminación / modificaciones de las historias o de su orden de prioridad.
 - Disponibilidad: Reunión diaria vía Skype, coordinación de trabajo mediante wasap, revisión de versiones utilizando correo electrónico.

Responsabilidades del Scrum Manager

- Supervisión de la pila de producto, y comunicación con el gestor del producto para pedirle aclaración de las dudas que pueda tener, o asesorarle para la subsanación de las deficiencias que observe.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Responsabilidades del equipo técnico

- Conocimiento y comprensión actualizado de la pila del producto.
- Resolución de dudas o comunicación de sugerencias Scrum
- Manager, codirectores y co-investigadores
- Responsabilidades del resto de implicados
- Conocimiento y comprensión actualizado de la pila del producto.
- Resolución de dudas o comunicación de sugerencias Scrum Manager

4.2 Pila del sprint

Es el documento de registro de los requisitos detallados o tareas que va a desarrollar el equipo técnico en la iteración (actual o que está preparándose para comenzar)

Responsabilidades del gestor de producto

- Presencia en las reuniones en las que el equipo elabora la pila del sprint. Resolución de dudas sobre las historias de usuario que se descomponen en la pila del sprint.

Responsabilidades del Scrum Manager

- Supervisión y asesoría en la elaboración de la pila de la pila del sprint.

Responsabilidades del equipo técnico

- Elaboración de la pila del sprint.
- Resolución de dudas o comunicación de sugerencias sobre las historias de usuario con el gestor del producto.

4.3 Sprint

Cada una de las iteraciones del ciclo de vida iterativo Scrum. La duración de cada sprint fue de cada veinte (20) días en promedio

4.4 Incremento

Parte o subsistema que se produce en un sprint y se entrega al gestor del producto completamente terminada y operativa.

4.5 Gráfica de producto (Burn Up)

Representación gráfica del plan de producto previsto por el gestor de producto. Es una gráfica que representa los temas o epics del sistema en el orden que se desean, y el tiempo en el que se prevé su ejecución.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Responsabilidades del gestor de producto

- Mantenimiento actualizado en todo momento durante la ejecución del proyecto
- Orden en el que desea disponer de los temas o “epics” del sistema, e hitos del producto (versiones).
- Incorporación / eliminación / modificaciones de los temas, de su orden de prioridad, estimaciones o hitos.
- Disponibilidad: Reunión diaria vía Skype, coordinación de trabajo mediante wasap, revisión de versiones utilizando correo electrónico.

Responsabilidades del Scrum Manager

- Supervisión del gráfico de producto, y comunicación con el gestor del producto para pedirle aclaración de las dudas que pueda tener, o asesorarle para la subsanación de las deficiencias que observe.

Responsabilidades del equipo técnico

- Conocimiento y comprensión actualizado del plan del producto.
- Resolución de dudas o comunicación de sugerencias con Scrum Manager, codirectores y co-investigadores
- Responsabilidades del resto de implicados
- Conocimiento y comprensión actualizado del plan de producto.

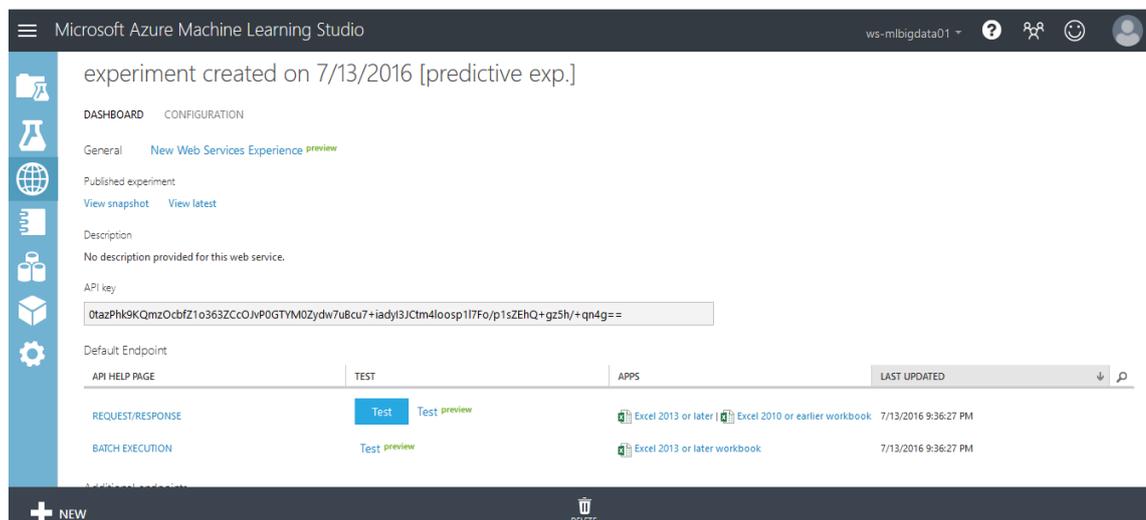


Fig. 2. Evidencia Primera versión

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

4.6 Gráfica de avance (Burn Down)

Gráfico que muestra el estado de avance del trabajo del sprint en curso.

Responsabilidades del gestor de producto

- Sin responsabilidades específicas, más allá de mantenerse regularmente informado del avance del sprint y disponible para atender decisiones para la resolución de opciones en sprints sobrevalorados o infravalorados (la gráfica de avance predice una entrega anterior o posterior a la fecha prevista)

Responsabilidades del Scrum Manager

- Supervisión de la actualización diaria por parte del equipo.

Responsabilidades del equipo técnico

- Actualización diaria del gráfico de avance

```

13  google.load("visualization", "1", { packages: ["corechart"] });
14  google.setOnLoadCallback(drawChart);
15
16  function drawChart() {
17    $.post("Url.Action("GetPieChartData", Model)", {},
18          function (data) {
19            var tdata = new google.visualization.DataTable();
20
21            for (var i = 0; i < 2; i++) {
22              if (i == 0) {
23                tdata.addColumn("string", 'Columna_Principal');
24              }
25              else {
26                tdata.addColumn("number", 'Columna' + i);
27              }
28            }
29
30            for (var i = 0; i < 5; i++) {
31              tdata.addRow([data[i].Departamento, data[i].Total_Solicitudes]);
32            }
33
34            var options = {
35              title: "Restitución de Tierras"
36            };
37
38            var chart = new google.visualization.PieChart(document.getElementById('divGrafico'));
39            chart.draw(tdata, options);
40
41          });
42    }
43  }
44  </script>
45  </head>
46  <div class="dashboardItem">
47    <div class="dashboardItemHeader">((VisorDatos.BE.BE_FuenteDatos)Model).Nombre </div>
48    <div class="dashboardItemBody">

```

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

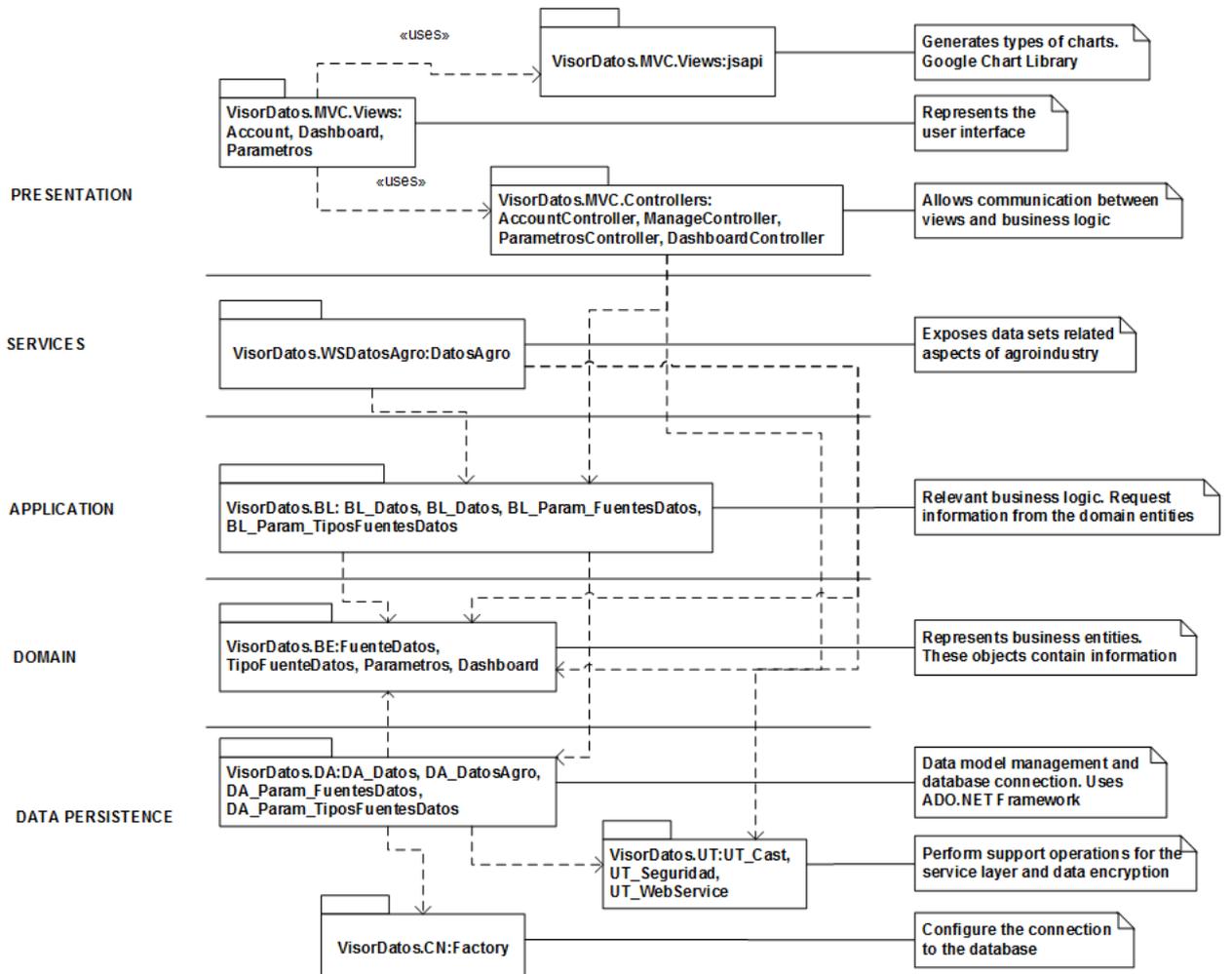


Fig. 3. Evidencia Tercer avance

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

4.7 Reunión de inicio de sprint

Reunión para determinar las funcionalidades o historias de usuario que se van a incluir en el próximo incremento.

Responsabilidades del gestor de producto

- Asistencia a la reunión.
- Exposición y explicación de las historias que necesita para la próxima iteración y posibles restricciones de fechas que pudiera tener.

Responsabilidades del Scrum Manager

- Moderación de la reunión

Responsabilidades del equipo técnico

- Confección de la pila del sprint.
- Auto-asignación del trabajo.

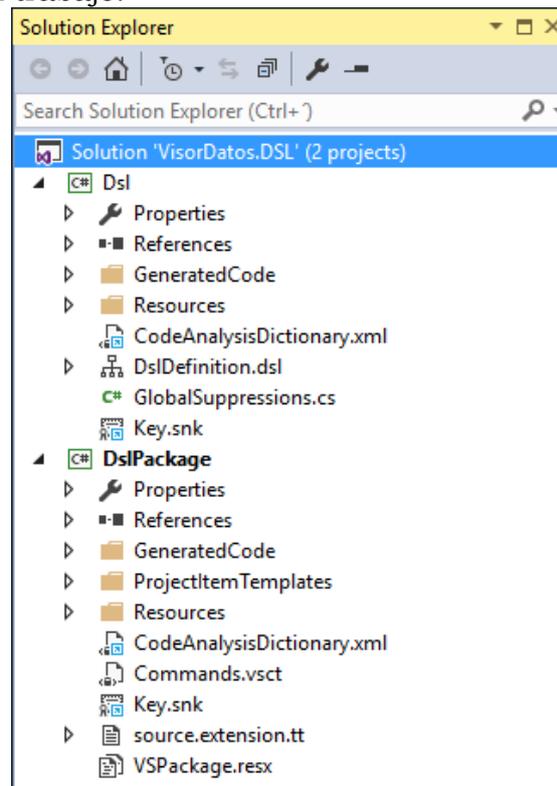


Fig. 5. Evidencia Cuarto avance

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

4.8 Reunión técnica diaria

Puesta en común diaria del equipo con presencia del Coordinador del proyecto o Scrum Manager de duración máxima de 10 minutos.

Responsabilidades del Scrum Manager

- Supervisión de la reunión y anotación de las necesidades o impedimentos que pueda detectar el equipo.
- Gestión para la solución de las necesidades o impedimentos detectados por el equipo.

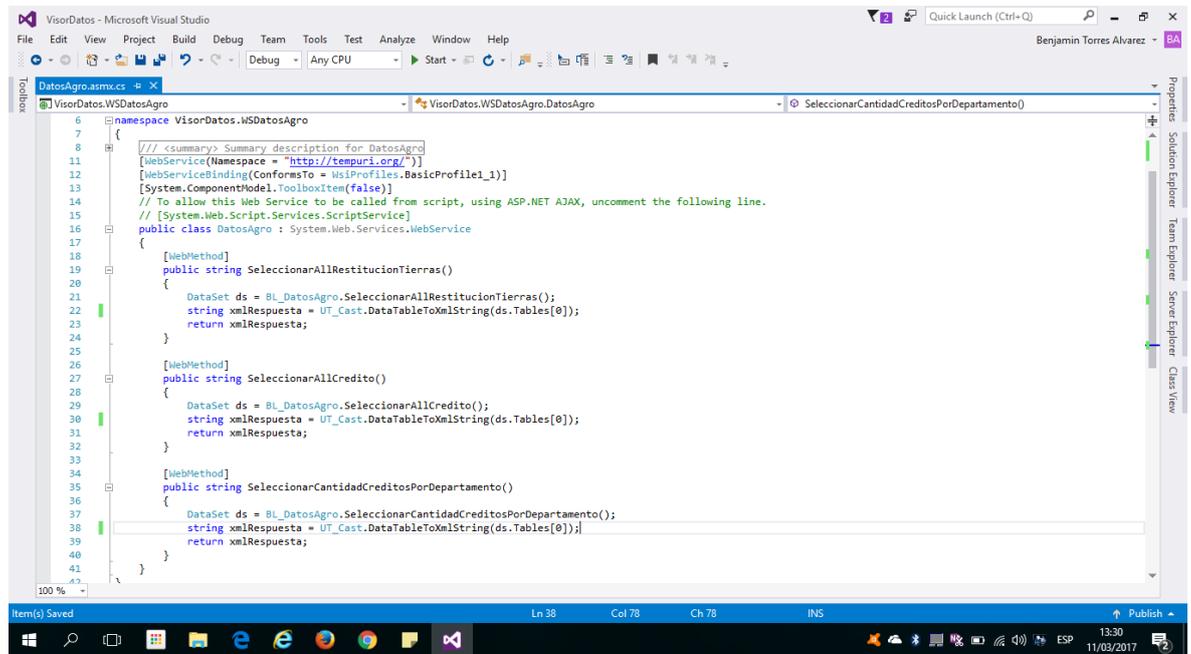


Fig. 6. Evidencia Quinto avance

Responsabilidades del equipo técnico

- Comunicación individual del trabajo realizado el día anterior y el previsto para día actual.
- Actualización individual del trabajo pendiente.
- Actualización del gráfico de avance para reflejar el estado de avance.
- Notificación de necesidades o impedimentos previstos u ocurridos para realizar las tareas asignadas.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

```

53
54 HttpRequest req = (HttpRequest)WebRequest.Create(urlServicio);
55 req.Headers.Add("SOAPAction", "\"http://tempuri.org/\" + Metodo + \"");
56 req.ContentType = "text/xml; charset=utf-8";
57 req.Accept = "text/xml";
58 req.Method = "POST";
59
60 using (Stream stm = req.GetRequestStream())
61 {
62     string postValues = "";
63     foreach (var param in Parametros)
64     {
65         if (encode)
66             postValues += string.Format("<{0}>{1}</{0}>", HttpUtility.UrlEncode(param.Key), HttpUtility.UrlEncode(param.Value));
67         else
68             postValues += string.Format("<{0}>{1}</{0}>", param.Key, param.Value);
69     }
70
71     soapStr = string.Format(soapStr, Metodo, postValues);
72     using (StreamWriter stmw = new StreamWriter(stm))
73     {
74         stmw.Write(soapStr);
75     }
76 }
77
78 using (WebResponse webResponse = req.GetResponse())
79 {
80     using (Stream stream = webResponse.GetResponseStream())
81     {
82         using (StreamReader responseReader = new StreamReader(stream))
83         {
84             string result = responseReader.ReadToEnd();
85             XmlDocument ResultXML = XmlDocument.Parse(result);
86             ResultString = HttpUtility.HtmlDecode(result);
87         }
88     }
89 }

```

Fig. 8. Evidencia séptima de avance

Vistas

Con la implementación del prototipo se generaron las siguientes vistas:

Registro: Captura de datos básicos del usuario para controlar el acceso a la aplicación Web

Login: Permíte iniciar sesión dentro de la aplicación Web

Index: Vista inicial que muestra la opciones brindadas por la aplicación Web

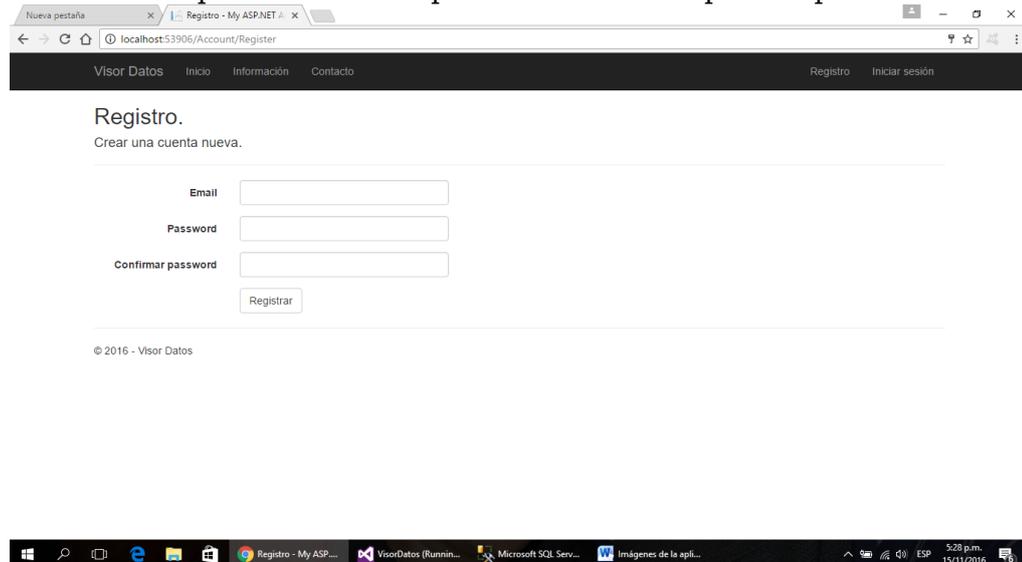


Fig. 9 vista 1.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

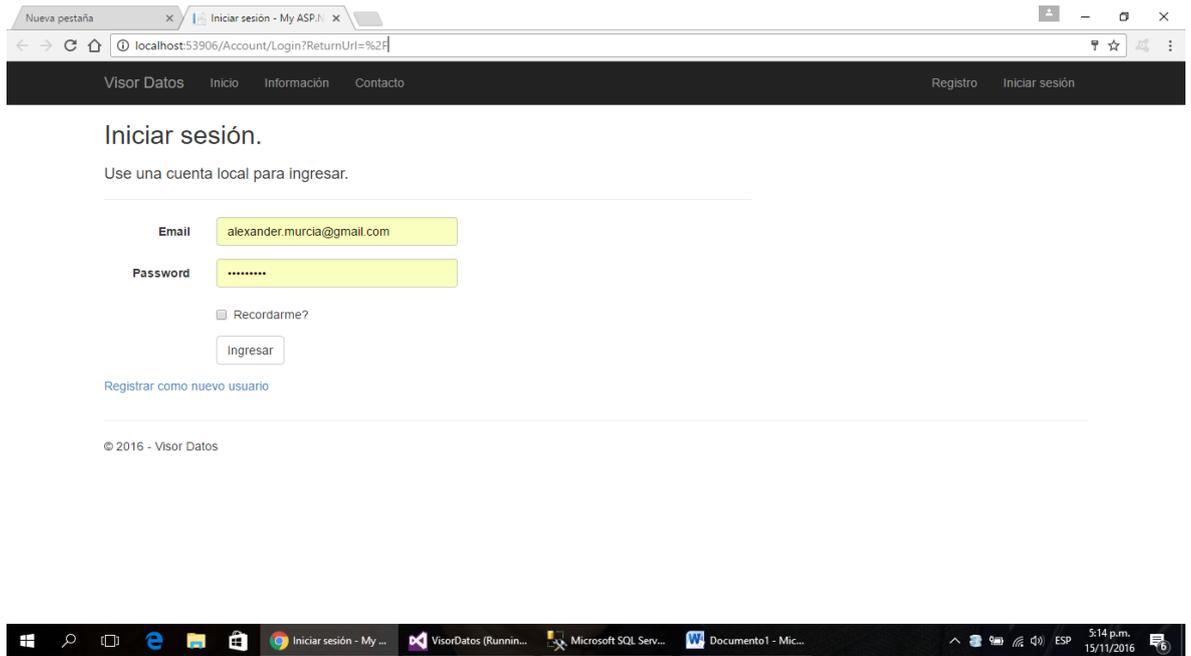


Fig. 10. vista 2.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

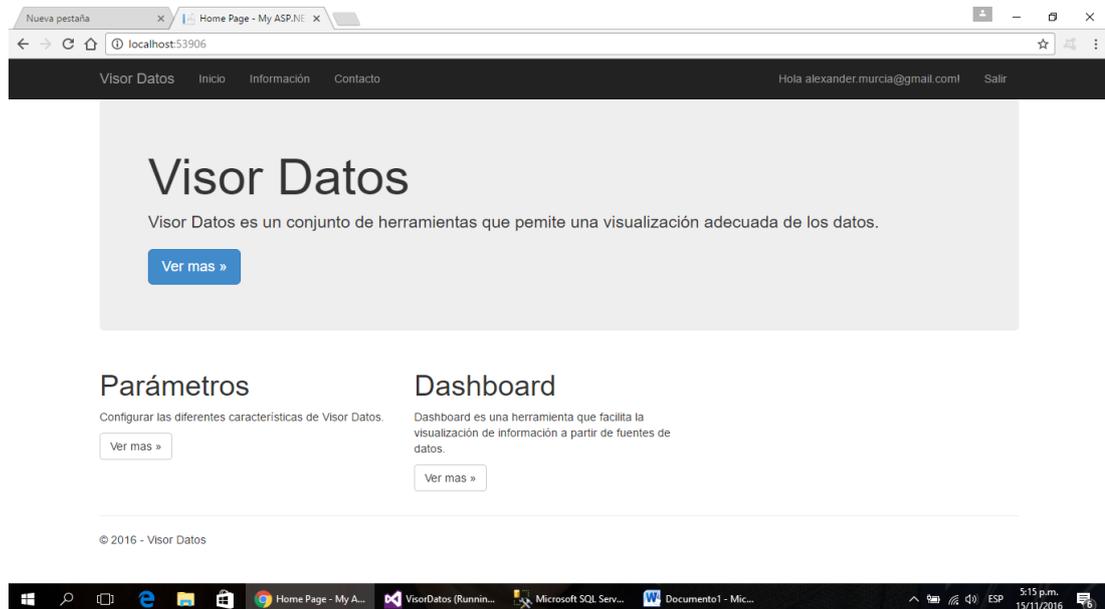
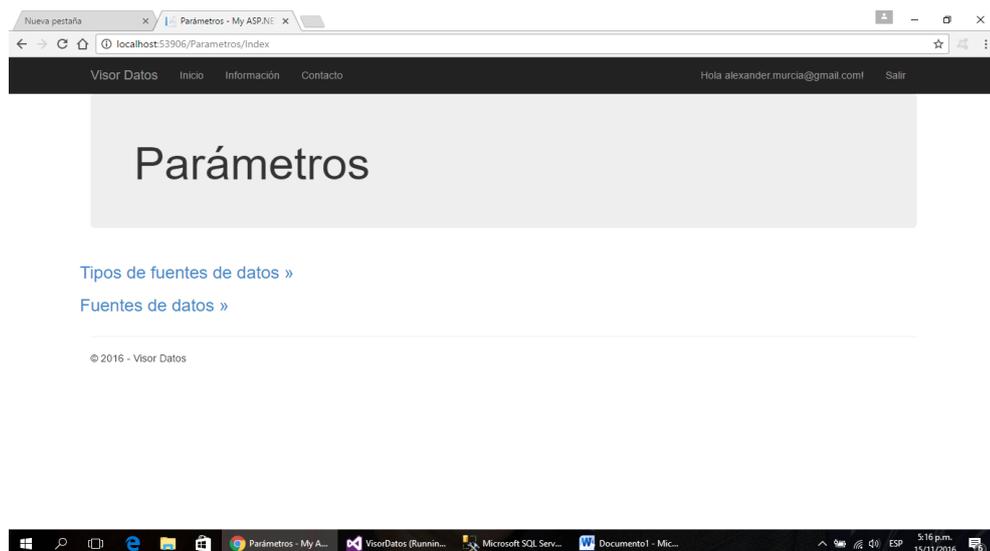


Fig. 11. vista 2.

En la Administración de la información del prototipo tenemos principalmente la gestión sobre las fuentes de datos primarias, la cual se agrupo en la sección *Parámetros*



Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Fig. 12. vista 3.

Administración Fuentes de Datos. Permite capturar, validar, visualizar, modificar y eliminar los registros creados por el usuario

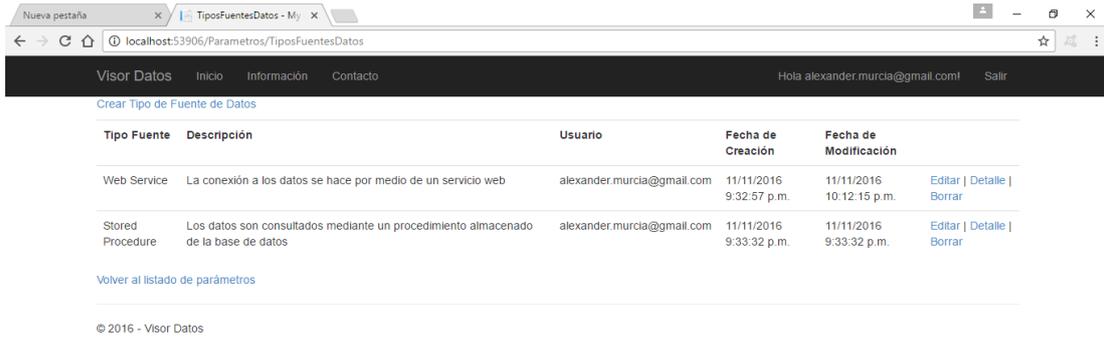
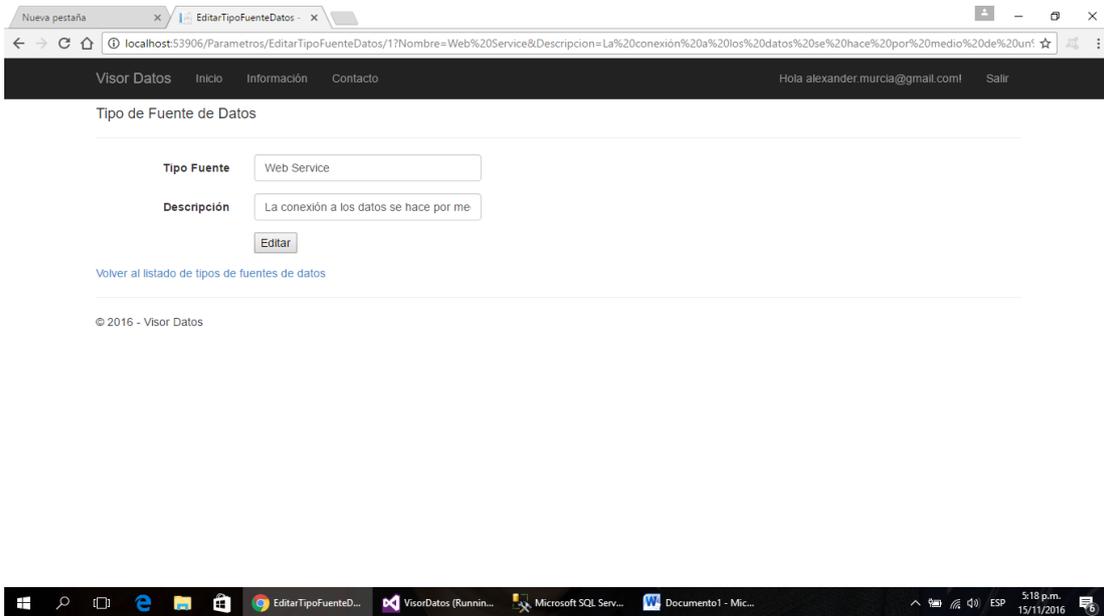


Fig. 13. vista 4.



Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

Fig. 14. vista 5.

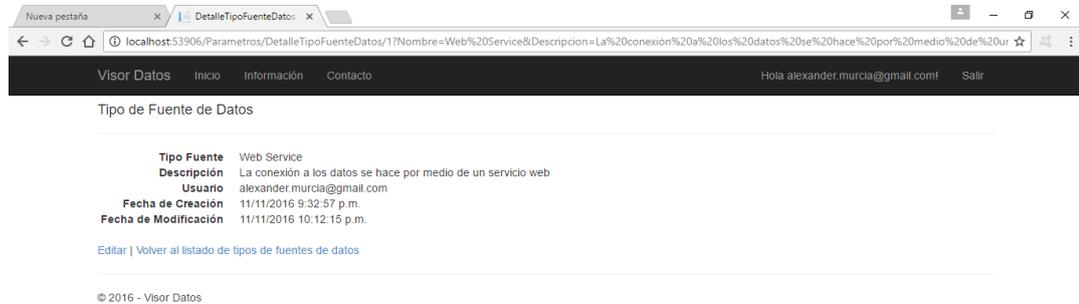


Fig. 15. vista 6.

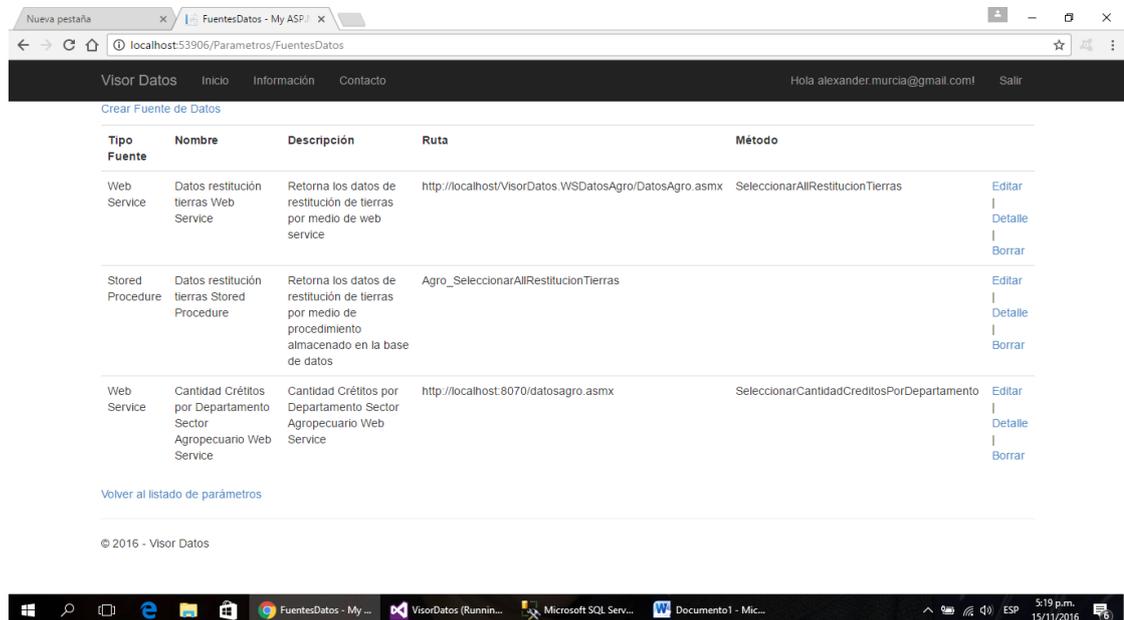


Fig. 16. vista 7.

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

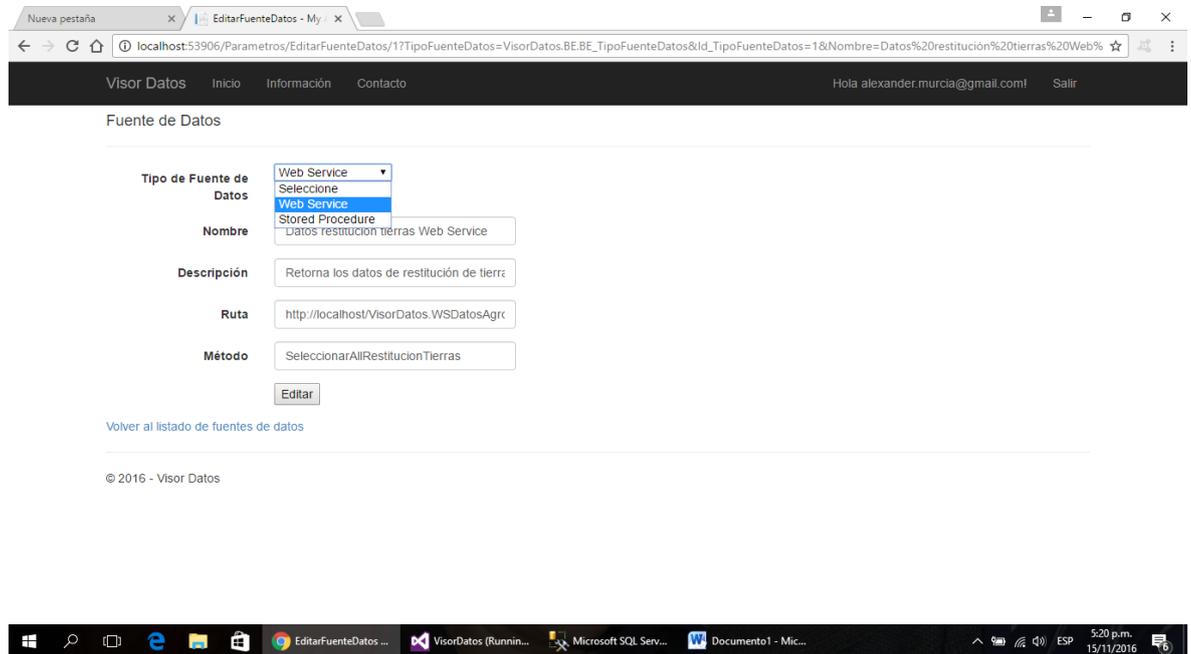


Fig. 17. vista 8.

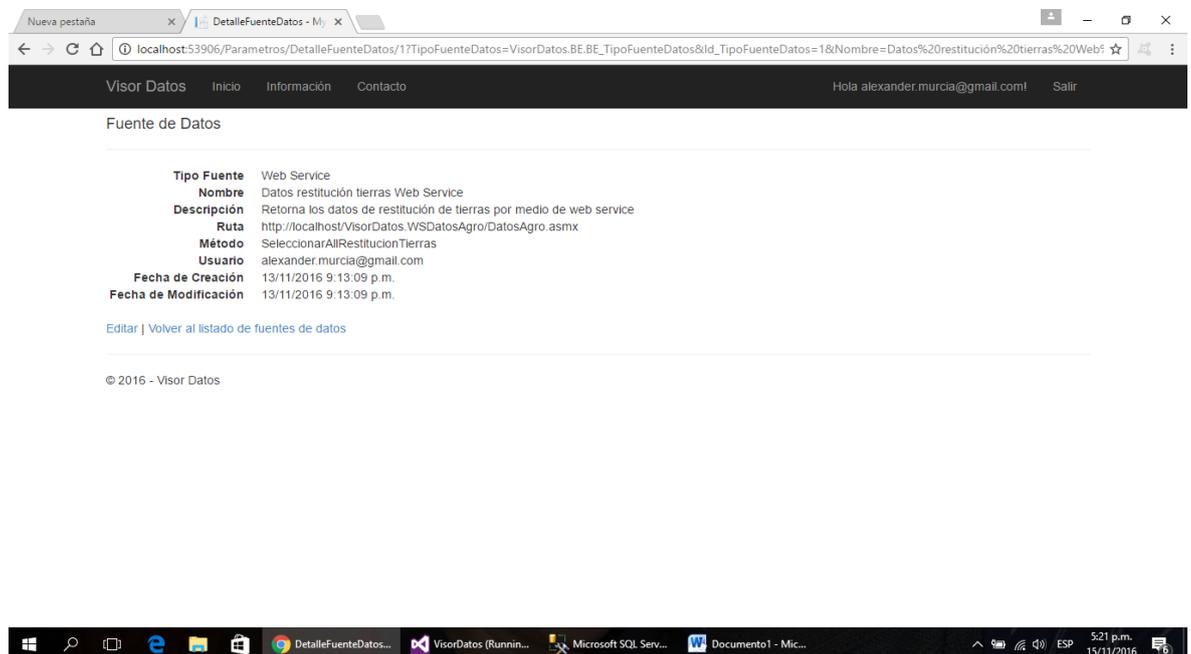


Fig. 18. vista 9.

Dashboard. En esta vista aparecen las diferentes opciones de servicios generados por el usuario con base a la parametrización previa

Desarrollo del Sistema: DSL Arquitectura orientada al procesamiento y aprovechamiento de datos abiertos	Versión: 5.0
Planificación	Fecha: 12/12/2017
Descripción de la metodología de trabajo	

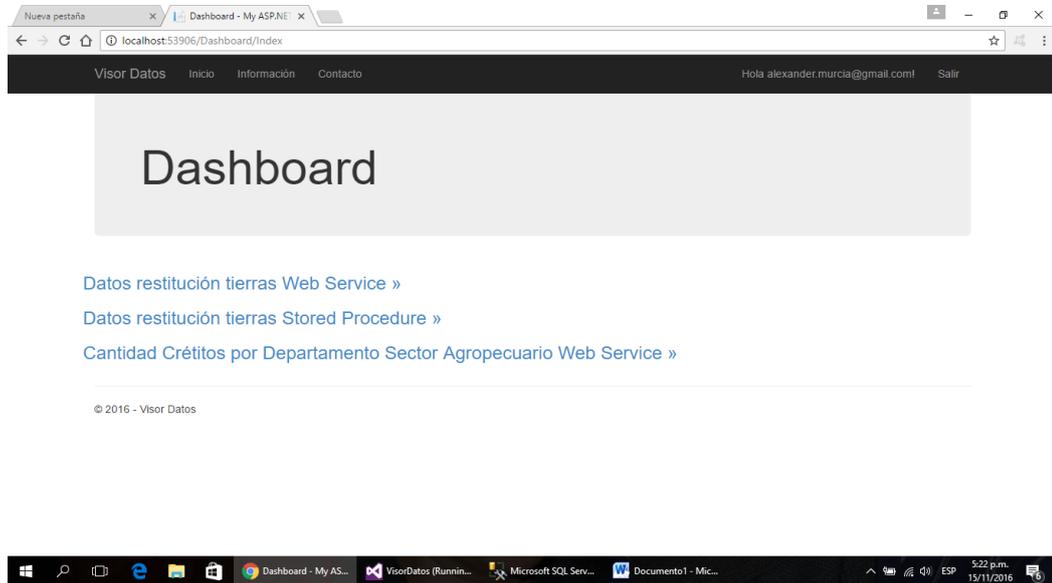


Fig. 19. vista 10.

En este caso, se visualiza en el dashboard la información relacionada con créditos para el sector agropecuario por departamento en Colombia con la fuente de datos consumida y parametrizada en la aplicación Web con el recurso datos.gov.co.



Fig. 20. vista 11.

