



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Plataforma de visualización y comparador
dinámico de contrataciones. El caso de
Generalitat Valenciana y Generalitat Cataluña.

Trabajo Fin de Máster
Máster Universitario en Gestión de la Información

Autor: Germán Andrés Boris
Tutor: Antonia Ferrer Sapena
José Manuel Calabuig Rodríguez

2017-2018



Esta obra est a sujeta a la licencia Reconocimiento 4.0 Internacional de Creative Commons. Para ver una copia de esta licencia, visite.

<http://creativecommons.org/licenses/by/4.0/>

Resumen

La información relacionada con las licitaciones y contratación pública se encuentra cada vez más accesible a través de las distintas plataformas autonómicas y la plataforma nacional. Aunque la información se encuentre accesible, su consulta de manera comparada es difícil. De esta manera, el objetivo de este proyecto es realizar una plataforma web que de forma automática cargue los datos de contratos publicados en los portales abiertos como “GVA Oberta” y “Portal Nacional” representándolos en forma de tablas, gráficos y grafos. Haciendo especial énfasis en los grafos, ya que es una forma innovadora y simple de presentar la información, de esta forma el ciudadano podrá interactuar con los diferentes datos y además entender de mejor manera la información ya publicada.

La plataforma web además de presentar la información de contratos contará con un comparador con el resto de comunidades autónomas, de esta manera el ciudadano podrá identificar cual es el gasto de la Generalitat Valenciana en comparación con otras. De inicio y para este trabajo de fin de master se realizará el comparador únicamente con la Generalitat de Catalunya.

Palabras clave: contratos, licitaciones, grafos, gobierno abierto, datos abiertos, web, R, Shiny.

Abstract

The aim of this project is to create a web platform that automatically loads the data of contracts which were published in open portals such as "GVA Oberta" and "Portal Nacional", representing them in the form of tables, charts and graphs. Since the graphs are an innovative and simple way of presenting information, there will be emphasis on them. In this way, the citizens will be able to interact with the different data and will understand better the published information.

In addition to presenting contract information, the web platform provides a comparator, which enables the citizens to compare the different autonomous communities. So the citizen can see the expenditure of the Generalitat Valenciana in comparison with others. At the beginning, as well as for this master's thesis, the comparator will only be carried out with the Generalitat de Catalunya.

Keywords: contracts, tenders, graphs, Open Government, open data, web, R, Shiny.

Tabla de contenidos

CAPÍTULO 1 : INTRODUCCIÓN	9
1.1 ANTECEDENTES	9
1.2 PROBLEMA DE INVESTIGACIÓN	10
1.3 JUSTIFICACIÓN	10
1.4 OBJETIVOS	11
1.4.1 PRINCIPALES	11
1.4.2 ESPECÍFICOS	11
1.5 DISEÑO METODOLÓGICO	11
1.5.1 METODOLOGÍA AGILE ANALYTICS	11
CAPÍTULO 2 : MARCO TEÓRICO	17
2.1 GOBIERNO ABIERTO	17
2.2 DATOS ABIERTOS	19
2.2.1 PROCESO DE APERTURA DE DATOS	20
2.3 PORTALES ABIERTOS	25
2.3.1 CKAN	25
2.3.2 DADES OBERTES	27
2.3.3 GOVERN OBERT	28
2.4 NORMATIVA	29
2.5 CONTRATACIÓN PÚBLICA	31
2.5.1 LEYES	31
2.5.2 TIPO DE CONTRATOS	32
2.5.3 PROCEDIMIENTOS	35
2.5.4 RACIONALIZACIÓN TÉCNICA DE LA CONTRATACIÓN	37
CAPÍTULO 3 : TECNOLOGÍAS	39
3.1 API's	39
3.1.1 API CKAN	39
3.1.2 API SOCRATA	44
3.2 R-STUDIO	49
3.3 DOCKER	51
3.4 SHINY	52
3.5 NEO4J	52
3.5.1 CYPHER	54
CAPÍTULO 4 :DESARROLLO DE PLATAFORMA	55
4.1 DESARROLLO EN SHINY R	55
4.1.1 EXTRACCIÓN	55
4.1.2 LIMPIEZA DE DATOS	60
4.2 PLATAFORMA	61
4.3 ELABORACIÓN DE GRÁFICOS	63
4.3.1 IMPORTACIÓN POR TIPO DE CONTRATO	63
4.3.2 EMPRESAS CON MÁS ADJUDICACIONES	64

4.3.3	TIPOS DE CONTRATO ANUAL	65
4.3.4	MONTO DE ADJUDICACIÓN POR COMUNIDAD	66
4.4	ELABORACIÓN DE GRAFOS	67
4.4.1	CONEXIÓN Y CARGA EN R	68
4.4.2	VISUALIZACIÓN DE GRAFOS	68
4.4.3	CONTRATOS POR COMUNIDAD AUTÓNOMA	69
4.4.4	INFORMACIÓN DE CONTRATOS POR COMUNIDAD AUTÓNOMA	69
<u>CAPÍTULO 5 :IMPLEMENTACIÓN E IMPLANTACIÓN</u>		71
5.1	PUBLICACIÓN DE LA PLATAFORMA	71
5.1.1	ARQUITECTURA	71
5.1.2	DOCKER	72
5.1.3	NEO4J	72
5.1.4	SHINY-SERVER	73
5.2	LIMITACIONES	73
<u>CONCLUSIONES</u>		74
<u>BIBLIOGRAFÍA</u>		75
<u>RECURSOS ELECTRÓNICOS</u>		76



Glosario de t rminos

APM	Agile Project Managment
BI	Business Intelligent
GVA	Generalitat Valenciana
ORG	Organization Open Government Partnership
RDF	Resource Description Framework
CMS	Content Management System
RPC	Remote Procedure Call
API	Application Programming Interface
SOAP	Simple Object Access Protocol
REST	Representational State Transfer
HTTP	Hypertext Transfer Protocol
SQL	Structure Query Language
SoSQL	Socrata SQL
IDE	Integrated Delevopment Enviroment
CRUD	Create, Update, Delete
OLTP	on-line transaction processing
BDOG	base de datos orientada a grafos
RAM	Random Access Memory

Índice de ilustraciones

<i>Ilustración 1: Ciclo de APM</i>	13
<i>Ilustración 2: Iteraciones Iniciales</i>	15
<i>Ilustración 3: Metodología APM</i>	16
<i>Ilustración 4: Gobierno Abierto</i>	18
<i>Ilustración 5: Presentación de metadatos</i>	24
<i>Ilustración 6: Portal de Datos abiertos comunidad de Valenciana</i>	28
<i>Ilustración 7: Portal de Datos Abiertos Cataluña</i>	29
<i>Ilustración 8: Tipos de contratos</i>	32
<i>Ilustración 9: Contratos administrativo y privados</i>	34
<i>Ilustración 10: Procedimientos de adjudicación</i>	35
<i>Ilustración 11: API Cataluña</i>	45
<i>Ilustración 12: Modelo R Studio</i>	50
<i>Ilustración 13: Comparación entre contenedores y máquinas virtuales</i>	51
<i>Ilustración 14: Docker Enterprise Edition</i>	52
<i>Ilustración 15: Estructura del grafo</i>	52
<i>Ilustración 16: Plataforma Neo4j</i>	54
<i>Ilustración 17: Menú Principal</i>	62
<i>Ilustración 18: Glosario de términos</i>	62
<i>Ilustración 19: Importe por tipo de contrato</i>	63
<i>Ilustración 20: Empresa con mayor importe</i>	64
<i>Ilustración 21: Tipo de contrato anual</i>	65
<i>Ilustración 22: Importe anual por comunidad</i>	66
<i>Ilustración 23: Empresas por comunidad</i>	69
<i>Ilustración 24: Contratos por comunidad</i>	70
<i>Ilustración 25: Arquitectura docker</i>	71

Índice de Tablas

<i>Tabla 1: 5 estrellas para Datos Abiertos</i>	21
<i>Tabla 2: Leyes Involucradas</i>	29
<i>Tabla 3: Normativas y Directivas</i>	30
<i>Tabla 4: Leyes relacionadas</i>	30
<i>Tabla 5: Contratos sujetos a regulación armonizada</i>	33
<i>Tabla 6: API Acciones de consulta</i>	40
<i>Tabla 7: API Acciones de inserción</i>	41
<i>Tabla 8: API Acciones de actualización</i>	41
<i>Tabla 9: API Acciones de borrado</i>	41
<i>Tabla 10: API Acciones de búsqueda</i>	42
<i>Tabla 11: API Métodos de búsqueda</i>	42
<i>Tabla 12: API Formatos de búsqueda</i>	43
<i>Tabla 13: API Parámetros de recurso</i>	43
<i>Tabla 14: Parámetros de revisión</i>	43
<i>Tabla 15: Cuadro SQL</i>	47
<i>Tabla 16: Funciones Socrata</i>	47
<i>Tabla 17: Campos Cataluña</i>	55
<i>Tabla 18: Campos Comunidad Valenciana</i>	57
<i>Tabla 19: Campos de la plataforma</i>	59

Capítulo 1 : Introducción

1.1 Antecedentes

A lo largo de la historia los ciudadanos han exigido a sus gobernantes una justificación a las decisiones que toman. Ante este antecedente, sumado al avance de la tecnología y otras problemáticas sociales nace el termino Gobierno Abierto. Término que se asimila a la transparencia de la administración pública a través de portales electrónicos que deben contar con los principios básicos de: colaboración, participación y transparencia.

Los gobiernos de Europa y diferentes países alrededor del mundo han elaborado leyes para la transparencia de los datos de la administración pública, optando por el paradigma del gobierno abierto, de esta realidad no está exenta España, quien en el año 2013, crea la ley de Transparencia, Acceso a la Información Pública y Buen Gobierno que tiene por objeto ampliar y reforzar la transparencia de la administración pública, regular y garantizar el derecho de acceso a la información relativa a aquella actividad y establecer las obligaciones de buen gobierno que debe cumplir los responsable públicos¹.

Una vez adoptado el concepto de gobierno abierto en España y con una ley que respalda la transparencia de la información del gobierno hacia sus ciudadanos, se comienzan a desarrollar diferentes portales de administración pública como es el caso del portal del Consejo Transparencia y Buen Gobierno² (CTBG), el Portal de datos abiertos³, Portal de Transparencia, etc. A esta iniciativa se le suma las comunidades Autónomas con sus diferentes portales de transparencia para el ciudadano.

En el caso de la Comunidad Valenciana rige la Ley 2/2016, de 4 de marzo, de la Generalitat, de modificación de la Ley 2/2015, de 2 de abril, de Transparencia, Buen Gobierno y Participación Ciudadana de la Comunitat Valenciana ⁴ la misma que regula diversos aspectos relacionadas con la transparencia, el buen gobierno, los datos abiertos y la participación ciudadana en el ámbito de la Comunidad Valenciana. Bajo la ley 2/2015, de 2 de abril de Transparencia, Buen Gobierno y Participación Ciudadana de la Comunitat Valenciana se crea el Consejo de Transparencia, Acceso a la Información Pública y Buen Gobierno el cuál tiene como fin el garantizar los derechos de acceso a la información, velar por el cumplimiento de las obligaciones de publicidad y garantizar y promover la observancia de las disposiciones de buen gobierno.

Para el caso de la contratación pública la ley más reciente es la 9/2017⁵, de 8 de noviembre, de Contratos del Sector Público, la misma transponen al ordenamiento jurídico español las

¹ Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.

² http://www.consejodetransparencia.es/ct_Home/index.html

³ <http://datos.gob.es/es>

⁴ [Ley 2/2016, de 4 de marzo, de la Generalitat, de modificación de la Ley 2/2015, de 2 de abril, de Transparencia, Buen Gobierno y Participación Ciudadana de la Comunitat Valenciana](#)

⁵ Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público

Directivas del Parlamento Europeo y del Consejo 2014/23/UE y 2014/24/UE, de 26 de febrero de 2014. Con esta nueva ley se busca una mejor transparencia en la contratación pública, obligando el uso de los portales electrónicos. La contratación electrónica presta una mayor seguridad y por lo tanto un mayor freno a las malas prácticas en materia de contratación pública.

Por otra parte, tenemos los datos abiertos que básicamente son datos generados o recabados por personas, organizaciones públicas o privadas los cuales se pueden utilizar, reutilizar y redistribuir, con el único requisito de atribución de su fuente o reconocimiento de su autoría dependiendo de la licencia a usarse. Se detalla con mayor detalle en el capítulo 2 del presente trabajo.

El portal de datos abiertos ⁶ de la Comunidad Valenciana, cuenta con un gran número de conjunto de datos de diferentes temas como lo son medioambiente, economía, salud, educación etc. Estos datos están publicados de forma abierta para el ciudadano y en diferentes formatos para que puedan ser consumidos y tratados según convenga.

1.2 Problema de investigación

Según el informe de reutilización de datos realizado por COTEC, existen alrededor de 103 portales, en los cuales tras realizar un filtrado y eliminar los datos federados para evitar duplicidad de la información, se han podido identificar 20.026 juegos de datos. Un claro ejemplo, el portal datos.gob.es tiene más de 14.000 juegos de datos listados, claramente se puede ver que el problema no es la falta de datos, sino la calidad, relevancia e importancia de la información publicada, la información que se puede analizar no es fácil de comprender para el ciudadano, cuentan con un gran número de campos e información, muchas veces irrelevante y difícil de interpretar. Otra de las problemáticas que existen es que la información no se encuentra estandarizada bajo normas generales para todos los portales. La misma información dependiendo del portal cuenta con una mayor o menor cantidad de campos en diferentes nomenclaturas.

Otro de los problemas a los que nos enfrentamos al momento de publicar nuestros datos o reutilizar datos públicos son las leyes que rigen este tipo de actividades, por ejemplo, los portales de datos abiertos deben regirse entre las principales leyes a la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público estatal que a manera general permite la reutilización de los documentos públicos para fines comerciales y no comerciales siempre que dicho uso no constituya una actividad administrativa pública, además también se encuentra la Directiva 2013/37/EU que sustituye a la Directiva 2003/98/CE e incorpora la necesidad de establecer una revisión de la iniciativa sobre accesibilidad y reutilización de la información pública.

1.3 Justificación

Como parte de mi formación profesional y con el afán de conseguir mayor conocimiento y experiencia realizo prácticas en la Catedra de transparencia y datos abiertos de la UPV, donde pude conocer de antemano la administración electrónica, gobierno abierto y especialmente los

⁶ <http://www.gvaoberta.gva.es/>

datos abiertos y sus diferentes portales, haciendo un especial énfasis en el portal de datos abiertos de la Comunidad Valenciana.

Trabajando en algunos proyectos pude conocer ciertas necesidades que el ciudadano puede requerir de la administración pública. Entre estas necesidades, creo que es fundamental que el ciudadano o usuario final pueda visualizar la información publicada de una manera fácil de interpretar y lo más grafica posible, por tanto, propongo este trabajo de fin de master enfocado a suplir estas necesidades. Ya que a pesar de existir muchos datos en los portales de cara al ciudadano no aporta ninguna información si esta no está representada de forma simple.

1.4 Objetivos

1.4.1 Principales

El objetivo principal es realizar una plataforma web que de forma automática y gráfica represente los datos existentes en los portales de datos abiertos tanto de la comunidad Valenciana como de Cataluña realizando comparaciones entre estas dos comunidades.

1.4.2 Específicos

- Conocer los diferentes portales de datos abiertos, nacional y autonómicos.
- Analizar los datasets en los diferentes portales de contratación, nacional y autonómicos.
- Comparar los datos obtenidos de los portales.
- Conocer la manera de conexión a los diferentes portales a través de API's.
- Desarrollar un portal web, en el cual se puedan visualizar los datos de manera gráfica, haciendo un especial énfasis en los grafos.
- Realizar comparaciones graficas de los diferentes portales analizados.

1.5 Diseño Metodológico

Para este trabajo fin de master se usará una metodología de desarrollo de aplicaciones que maneja amplia cantidad de datos. Se eligió esta metodología ya que los proyectos que involucran grandes cantidades de datos no tienen el mismo ciclo de vida que otro tipo de proyectos mucho más interactivos.

1.5.1 Metodología Agile Analytics

Este capítulo está basado en la metodología Agile Analytics de Ken W. Collier. Este nos da a conocer que, Agile Analytics es una metodología reciente que nace por una nueva tendencia llamada Big Data que ha ido tomando fuerza en estos últimos años, los proyectos relacionados con Big Data, Analytics y BI, no se pueden manejar de la manera tradicional por la gran cantidad de información que involucran, por tanto, esta metodología nos da una guía de cómo manejar nuestro proceso de desarrollo dentro de estas ramas. Hemos de tener en cuenta que es



únicamente una guía ya que todos los proyectos varían dependiendo sus objetivos y alcance, por este motivo debemos adaptar la metodología a nuestro proyecto.

1.5.1.1 Características de la metodología

Iterativo, incremental, evolutivo. Se trabaja en iteraciones cortas máximo de cuatro semanas. El sistema se va construyendo en pequeños paquetes de acuerdo a las necesidades del usuario. Y evolucionamos el sistema de trabajo adaptándonos a los comentarios o requerimientos de los usuarios. Las iteraciones cortas con frecuentes revisiones de los usuarios ayudan a garantizar que nunca estemos muy lejos de nuestro desarrollo. Para nuestro proyecto es muy cómodo trabajar por paquetes por la tecnología que se usa.

Desarrollo de valor. El objetivo de cada iteración de desarrollo es un producto valorado por el usuario. Si bien para los desarrolladores/equipo de trabajo es de gran dificultad diseñar las arquitecturas de datos, modelos de datos, secuencias de comandos, procesos etc. A los usuarios en general les importa menos estas cosas. Lo que les importa a los usuarios de los sistemas es la presentación, visualización, facilidad de uso y el acceso a la información que les ayuda a resolver un problema. Cada iteración debe producir al menos una nueva característica valorada por el usuario a pesar del hecho de que las características del usuario no reflejan todo el trabajo y desarrollo que existe para obtener ese resultado.

Calidad del producto: Cada característica desarrollada recientemente debe probarse y depurarse completamente durante la iteración de desarrollo. El desarrollo ágil no se trata de construir prototipos incompletos; se trata de evolucionar gradualmente a la solución correcta con los mejores fundamentos tecnológicos. Los desarrolladores deben planificar e incluir pruebas rigurosas en su proceso de desarrollo.

Procesos mínimos. Dado que nuestro objetivo principal es la producción de sistemas de alta calidad y alto valor, debemos ser capaces de minimizar la cantidad de procesos requeridos para otras actividades, como son actas de reunión, autorizaciones a cambios etc. Este tipo de procesos en las metodologías clásicas suelen ser un dolor de cabeza y toman mucho tiempo. Para nuestro caso este punto ha sido básicamente las revisiones realizadas por los tutores y/o representantes de la conselleria, pero en un proyecto con clientes este puede ser un punto que nos dé más de un dolor de cabeza, por el tiempo que conlleva esta actividad.

Automatización. La única manera de ser realmente ágil es automatizar tantos procesos rutinarios como sea posible, Agile Analytics busca automatizar cualquier proceso que se realice más de una vez. Esto nos permite ahorrar tiempo innecesario y concentrarnos en lo que en realidad importa, las necesidades del cliente.

Colaboración. Con demasiada frecuencia en los proyectos tradicionales, el equipo de desarrollo solo asume la responsabilidad de garantizar que se cumplan los plazos, se entrega el alcance completo, se gestionan los presupuestos y se garantiza la calidad. La inteligencia de negocios ágil reconoce que hay un proyecto más amplio. Si bien es este trabajo de fin de master no se tiene un equipo de desarrollo, es más bien un desarrollo individual, si existió un equipo de trabajo en este caso serían los tutores y la conselleria de transparencia.

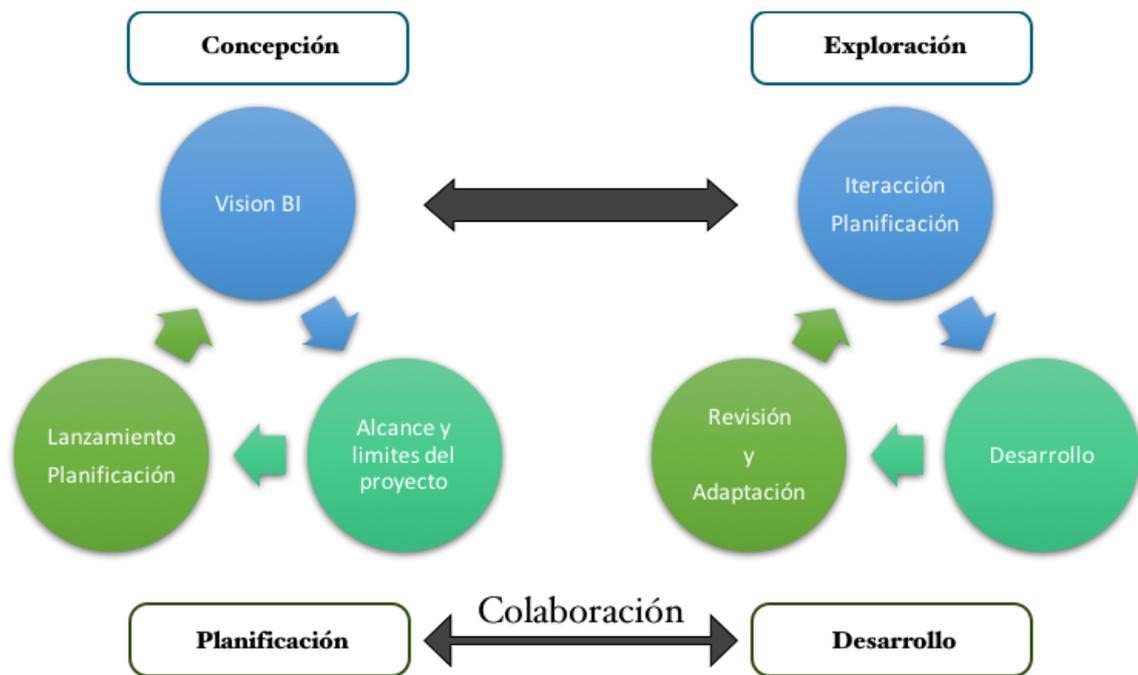
1.5.1.2 Ciclo de vida

Agile Analytics está marcado por un enfoque altamente iterativo con un alto grado de colaboración entre desarrolladores, usuarios y partes interesadas.

El marco APM de Highsmith se basa en un ciclo de acuerdo como se muestra (ver Figura 1.1). La importancia de este cambio de paradigma es que reconoce que los proyectos están sujetos a incertidumbre y cambio, y los buenos equipos de proyecto buscan adaptarse a ese cambio e incertidumbre. El proceso APM es altamente colaborativo y fomenta la interacción frecuente entre desarrolladores, usuarios comerciales y partes interesadas a lo largo del ciclo del proyecto.

En este proceso existen dos ciclos internos que son fundamentales para la aplicación del proceso APM, el primero es la concepción del proyecto y el segundo la exploración (ver Figura 1.2).

Ilustración 1: Ciclo de APM



Fuente: Adaptado de Agile Analytics, Ken W. Collier.

Fase de concepción

En la fase de concepción se ha de averiguar qué se va a hacer en el proyecto y cómo. Se define el alcance, límites, equipo de trabajo, limitaciones del proyecto y la estrategia comercial a seguir. Como resultado de la fase se debe tener el resultado visionado del proyecto y luego especular acerca de cómo lograr esa meta de manera incremental. En este proyecto de omitió la parte comercial que puede conllevar una gran dificultad de cara al proyecto, además de que no era uno de los objetivos planteados dado que es un proyecto directamente soportado por la Cátedra de Transparencia y Gestión de datos.

En base a la fase de concepción para nuestro proyecto se planteó lo siguiente:

Visión Datos

Como parte del análisis de datos se desea conectar de manera automática a los portales abiertos de las comunidades autónomas a analizarse para realizar una comparativa en base a los datos que se tenga de contrataciones. La comparativa se realizará con visualizaciones dinámicas.

Alcance

Desarrollar una plataforma web que de forma automática represente los datos existentes en los portales de datos abiertos de forma gráfica y pueda realizar comparaciones entre las distintas comunidades autónomas.

Limites

- Los datos de las comunidades autónomas que se analizarán serán únicamente de contratos.
- Se realizará la carga, análisis y visualización de datos únicamente de dos comunidades autónomas (Valenciana y Cataluña)
- Para el caso que no se cuente con API para la conexión automática, se la realizara de manera manual, descargando los CSV.

Características

- Conexión de datos mediante API portal Dades Obertes (GVA)
- Conexión de datos mediante API portal Dades Obertes (CAT)
- Carga de datos en plataforma R- Studio
- Limpieza de datos en R-Studio
- Implementación en plataforma Shiny
- Presentación de datos en plataforma Shiny
- Creación de grafos en Neo4j
- Conexión entre Neo4j y Shiny
- Visualización de datos y gráficos en Shiny
- Creación de filtros y métodos de búsqueda para todas las visualizaciones.
- Pruebas e implementación del portal.
- Publicación del portal Shiny

Fase de exploración

El desarrollo tradicional se basa en tareas o actividades (requisitos, codificación). El desarrollo ágil se basa en los productos (pequeñas historias de usuarios). Se seguirán las actividades/características que se han detallado en la fase concepción con ellas se irán realizando pequeñas iteraciones, hasta completar con toda la funcionalidad.

El ciclo de exploración se denomina así porque nos brinda la oportunidad de explorar, experimentar, probar ideas, evaluar y finalmente establecer lo correcto para construir y la forma

correcta de construirlo. Una de las mejores cosas de trabajar en estas iteraciones cortas es que el desarrollo no se aleja demasiado de los requerimientos en caso de estar erróneo.

No vamos a construir todo el sistema correctamente ni completo en una iteración de este ciclo. Debemos repetir este proceso varias veces, así el producto evoluciona hasta obtener un producto completo y de calidad. En cada vuelta del ciclo es fundamental buscar retroalimentación y aceptación, para el caso de este TFM esta evaluación la realizaran los tutores y compañeros de la Cátedra.

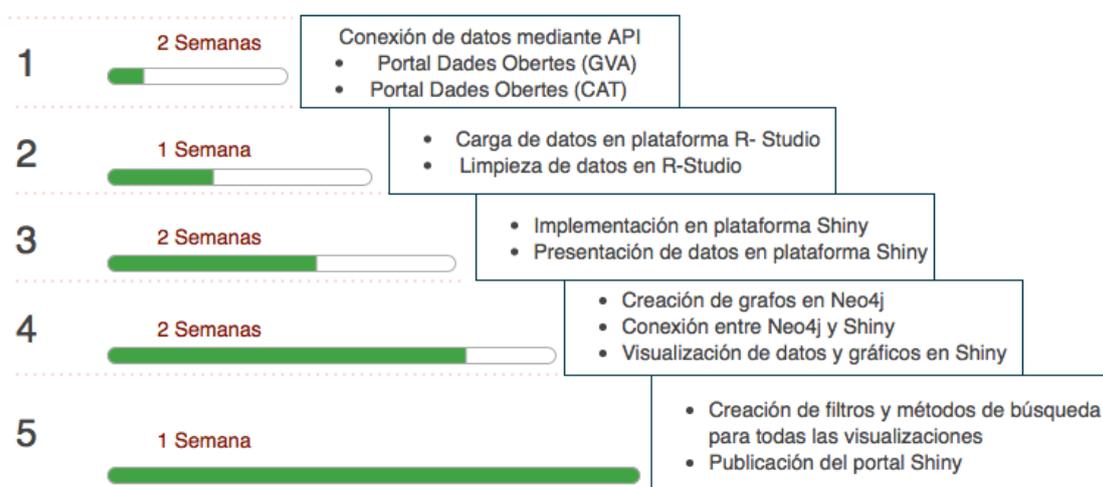
Normalmente cada una de las iteraciones toma de 1 a 2 semanas dependiendo del avance en el desarrollo, en cada una de ellas se incrementará al menos una característica del sistema. Esto dependerá del ritmo del desarrollador y la dificultad que conlleva llevar a cabo las funcionalidades del sistema. En algunas de las iteraciones se desarrollará más de una característica a la vez.

Cada iteración a través del ciclo de exploración comienza con una sesión de planificación de iteración. Dependiendo de qué cómo se mantenga la acumulación de historias, la planificación de la iteración se realizará semanalmente y no se tardará más de 2 horas en definir todos los puntos. Durante este tiempo, el equipo de desarrollo realiza la estimación del esfuerzo de la historia del usuario (funcionalidad), mueve una o más historias de usuarios desde la parte superior de la acumulación al plan de iteración, se compromete a completarlas y define las tareas subyacentes requeridas para completar cada historia. En esencia, el equipo realizará todo lo necesario para comenzar a trabajar.

Cada iteración termina en una revisión de una característica o funcionalidad, y una retrospectiva o reflexión. La revisión de características es una sesión de colaboración con usuarios finales para demostrar nuevas funciones, como un componente de informe o panel de control. El objetivo de la revisión es obtener la aceptación de la (s) nueva (s) característica (s) para satisfacer las necesidades de los usuarios.

Iteraciones Iniciales

Ilustración 2: Iteraciones Iniciales

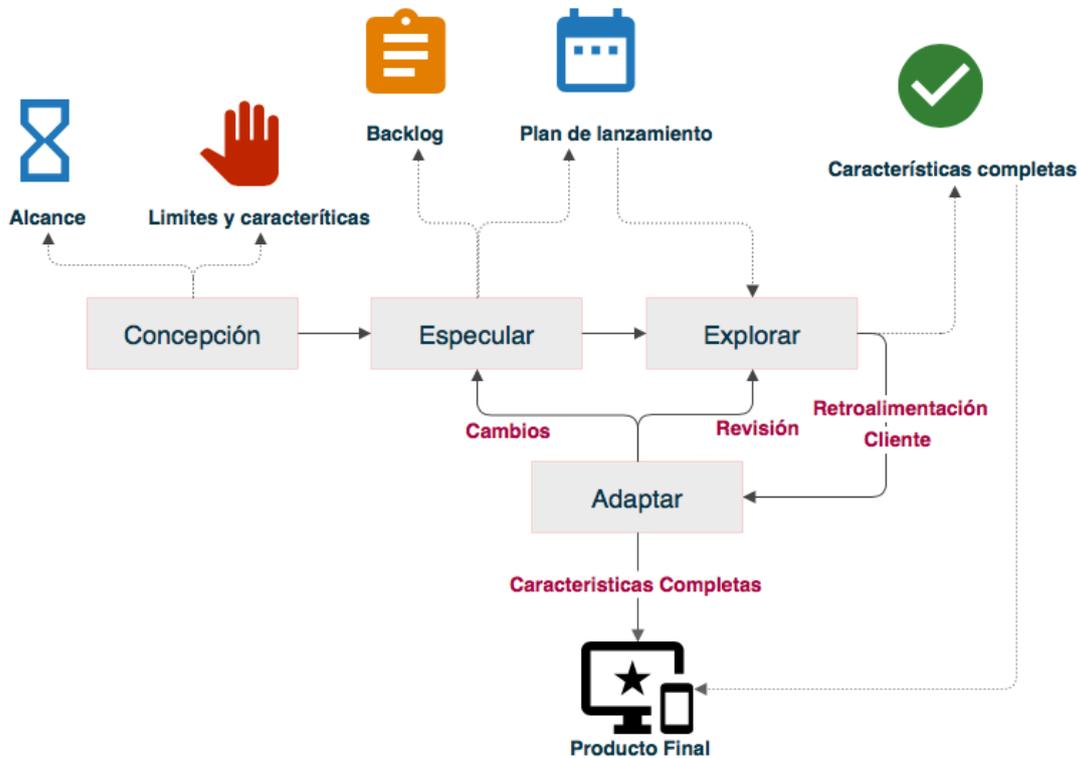


Fuente: Elaboración propia

1.5.1.3 Diagrama de proceso de APM

En el siguiente diagrama se muestra el flujo completo de APM, incluido sus dos fases primordiales como son la concepción y la exploración.

Ilustración 3: Metodología APM



Fuente: Adaptado de Agile Analytics, Ken W. Collier.

En resumen, el proceso de la metodología consta de 4 fases primordiales:

- **Concepción:** Se definirá el alcance, requerimientos funcionales, requerimientos no funcionales, límites y limitantes para el proyecto.
- **Especulación:** En esta fase por una parte se realiza el plan de lanzamiento inicial y se van registrando los cambios realizados conforme las propuestas vayan siendo aceptadas o rechazadas. Es una fase en la cual se registran toda la actividad antes de pasar a la fase de exploración.
- **Exploración:** Es la fase en la cual se realizan los prototipos o se desarrollan las características planteadas en las fases anteriores. Una vez finalizada esta fase, se aprobará el desarrollo por partes del usuario, caso contrario se lo revisará o se cambiará en su totalidad dependiendo de la valoración del usuario.
- **Adaptación:** Una vez los desarrollos son aprobados en la fase de exploración se van integrando al sistema, cuando todas las características son aprobadas se tendrá el producto final completo.

Capítulo 2 : Marco Teórico

En este capítulo se explicará los diferentes conceptos involucrados en el proyecto. En algunos casos el nivel de detalle será a manera general, existen temas que contiene una gran cantidad de información por lo que no se puede redactar todo su contenido en un solo proyecto mucho menos conocerlo a profundidad, se abarcaran únicamente los conceptos generales para entender el desarrollo y el propósito del proyecto.

2.1 Gobierno abierto

Se encuentran varias definiciones de gobierno abierto, las que mejor se adaptan al contexto que aquí se aborda son las de:

Ramirez Alujas (2011), “El Gobierno Abierto surge como un nuevo paradigma y modelo de relación entre los gobernantes, las administraciones y la sociedad: transparente, multidireccional, colaborativo y orientado a la participación de los ciudadanos tanto en el seguimiento como en la toma de decisiones públicas, a partir de cuya plataforma o espacio de acción es posible catalizar, articular y crear valor público desde y más allá de las fronteras de las burocracias estatales.”

Obama (2011), se comprometió a “Crear un nivel de apertura en el gobierno sin precedentes y un sistema de transparencia, participación pública y colaboración que reforzara la democracia, asegurara la confianza pública y promoviera la eficacia y la eficiencia gubernamental”

Si algo tienen en común las definiciones mencionadas y la mayoría de definiciones sobre gobierno abierto es que se basan en tres pilares fundamentales: transparencia, participación y colaboración. Bajo estos pilares los ciudadanos tendrán una interacción directa con el gobierno, aportando su opinión y teniendo una relación más participativa en las acciones tomadas por sus gobernantes además de la posibilidad de exigir la rendición de cuentas.

En la siguiente grafica se podrá visualizar de mejor manera los 3 pilares básicos de gobierno abierto y sus principales características.



Ilustración 4: Gobierno Abierto



Fuente: Adaptado de Armel Le Coz & Cyril Lage

Entorno al gobierno abierto se creó una organización Open Government Partnership (OGP), creada el 20 de septiembre de 2011. Inicialmente estuvo conformada por 8 países (Brasil, Indonesia, México, Noruega, Filipinas, South África, Reino Unido y Estados Unidos). Esta organización busca que los gobiernos que forman parte, cumplan con sus objetivos los cuáles se encuentran relacionados con la transparencia, la colaboración y la participación con el fin de luchar contra la corrupción aprovechando las nuevas tecnologías y así promover la transparencia del gobierno. OGP está supervisado y dirigido por un comité directivo, el cual incluye representantes de los países miembros y organizaciones civiles.

Para ser miembro de la OGP, los países participantes deben respaldar una Declaración de Gobierno Abierto, entregar un plan de acción de país desarrollado con consulta pública, Y se comprometen a informar de manera independiente sobre su progreso.

España forma parte de esta organización desde el año de su creación. En la que cuenta en la actualidad con 20 compromisos adquiridos, de los cuales cabe recalcar:

- e-Salud (1/1/2012)
- Ley de Transparencia y Acceso a la Información (3/1/2012)
- Transparencia y rendición de cuentas en la Ayuda Oficial al Desarrollo (1/1/2012)
- Mayor control y transparencia en las subvenciones públicas (1/1/2014)

Con esta alianza multilateral, se plantea cumplir con el Plan de Acción de Gobierno Abierto que los gobiernos deben presentarla de manera bianual. OGP evalúa a fin del periodo si los miembros de OGP cumplen con lo propuesto y en qué grado.

2.2 Datos abiertos

Son datos recabados por organizaciones públicas o privadas que cualquiera puede utilizar, reutilizar y redistribuir, con el único requisito de atribución de su fuente o reconocimiento de su autoría dependiendo de la licencia a usarse.

En una reunión de diciembre 2007 de Open Government Working Group Meeting ⁷ en la que participaron treinta expertos propulsores de la open data, se definieron unas pautas de referencia, las mismas fueron sometidas posteriormente a consulta a los diferentes gobiernos (open gov data, 2007). De esta reunión se obtuvieron los siguientes principios básicos que debe seguir los datos abiertos:

- **Completos**

Los datos públicos son aquellos que no están sujetos a restricciones de privacidad o seguridad de acuerdo al derecho español vigente. Artículo 3 de la Ley 18/2015.

- **Primarios**

Los datos deberán estar disponibles en las fuentes primarias, sin haber sido modificados o procesados de ningún modo.

- **Periódicos**

Se procurará que la información puesta a disposición se actualice en un tiempo razonable que permita el uso adecuado de dicha información, con una frecuencia análoga con la que actualicen dicha información internamente, así como su disponibilidad, incluida la temporal, completitud e integridad de acuerdo con el marco normativo aplicable en cada caso. Artículo 2.3 del Real Decreto 1495/2011 por el que se desarrolla la Ley 37/2007.

- **Accesibles**

Los datos deben estar disponibles de tal manera que cualquiera sin discriminación pueda usarlos, modificarlos y compartirlos para cualquier fin.

- **Procesables**

Los datos deben estar estructurados de tal forma que puedan ser procesados de forma automática.

- **No discriminatorios**

Los datos deberán estar disponibles para el público en general sin ningún tipo de restricción o registro de cualquier tipo

- **No propietarios**

⁷ <https://opengovdata.org/>

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Los datos deben de presentarse en un formato de cogido abierto como lo son (HTML, XML, RDF, JSON, etc.), evitando los formatos propietarios (XLS, DOC, etc.).

- **Licencia Libre**

Los datos públicos no estarán sujetos a derechos de autor, licencias o patentes

2.2.1 Proceso de apertura de datos

Siguiendo la guía publicada por el gobierno español “Cómo publicar datos abiertos de manera rápida y sencilla (con CKAN)” en su portal datos.gov.es y la guía de catalogación de datos abiertos realizado por la comunidad de Aragón, el proceso de apertura de datos sigue la siguiente dinámica:

Estas directrices de publicación de datos se enfocan únicamente para instituciones públicas, aunque también se podría adaptar para instituciones privadas.

2.2.1.1 Localización de datos

Como punto inicial se debe conocer el alcance y la extensión aproximada de los datos. Dado que en un inicio no se conoce la cantidad de datos, el número de ellos que se tomará será a modo de referencia. Los datos a incluir se irán ingresando de forma progresiva. Este proceso será continuo a lo largo del proyecto, debido a que la información se estará actualizando continuamente.

Existen factores a lo largo de este proceso que dan lugar a la generación o actualización masiva de los datos como puede ser cambios legislativos, nueva información, peticiones de datos etc., debemos por tanto tener en cuenta que existiría la posibilidad de cambios dentro de la plataforma para incluir modificaciones futuras.

Se deberá plantear, además, un proceso de digitalización de la información que no se encuentre en formato digital.

Si la información ha de ser reutilizable los datos deben tener en cuenta los siguientes puntos:

- La Información debe ser de calidad y fiable.
- El contenido de la información, incluyendo sus metadatos, no debe ser alterado.
- Hay que citar la fuente.
- Mencionar fecha de última actualización.
- Cuando la información contenga datos de carácter personal, se debe anonimizar los datos.

2.2.1.2 Elección condiciones reutilización

Aplicables dentro del marco establecido por la Ley 37/2007. Los organismos y entidades del sector público estatal estarán sujetos a lo establecido por el Real Decreto 1495/2011.

- Disposición de los recursos de información reutilizable sin sujeción a condiciones específicas o reutilización en condiciones generales, siendo esta la modalidad general por defecto e incluye:
 - Condiciones generales para la reutilización de datos.
 - Exclusión de responsabilidad al publicador de cualquier uso posterior de los datos.
 - Responsabilidad del agente re utilizador en el uso de los datos.
 - Reflejar la atribución de la fuente original.
- Reutilización con sujeción a condiciones establecidas en licencias-tipo. Si, se optase por la aplicación de un licenciamiento alternativo para la publicación de los datos, es recomendable que lo haga exclusivamente a través de alguna de las licencias-tipo compatibles con los principios de los datos abiertos.
- Solicitud a administración pública para iniciar proceso de reutilización de datos.

2.2.1.3 Catalogación

La catalogación de la información se genera en base a la información y los metadatos, estos serán publicados en un único punto de acceso común (catálogo de datos) para mejorar la reutilización, lo que permitirá a los desarrolladores o usuarios una mejor búsqueda y manejo de la información.

Dentro de la catalogación de los datos se debe elegir el formato que se considere más adecuado para su posterior tratamiento. Se puede tomar como guía la Norma Técnica de Interoperabilidad del Catálogo de estándares⁸ siempre que sea posible, en particular en lo relativo a formatos de ficheros, gestión documental, archivística y semántica.

Selección de Formatos

Según Tim Berners-Lee, el inventor de la Web e iniciador de los Datos Enlazados (Linked Data), sugirió un esquema de desarrollo de 5 estrellas para Datos Abiertos (5stardata, 2015).

Tabla 1: 5 estrellas para Datos Abiertos

Nivel	Descripción	Formato
1 Estrella	Publicar los datos en la web en cualquier formato bajo licencia abierta	PDF
2 Estrellas	Se presenta los datos de una forma estructurada, se pueden procesar con algún tipo de software de forma automática.	XLS
3 Estrellas	Usa formatos no propietarios, los formatos publicados son abiertos ejemplo (CSV)	CSV

⁸ Resolución de 3 de octubre de 2012, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Catálogo de estándares: <http://www.boe.es/boe/dias/2012/10/31/pdfs/BOE-A-2012-13501.pdf>

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

4 Estrellas	Los datos (más importantes) tienen una URI y puede ser compartido. Una forma de representar los datos es utilizar RDF, sin embargo, existen otros formatos como atom.	RDF
5 Estrellas	Enlazar datos a otros datos para dar contexto a la información.	LOD

Fuente: Elaboración propia

Tomando en cuenta la selección de formatos del gráfico anterior, la elección de formato dependerá del nivel en el cual queramos presentar la información.

1 Estrella: Formato PDF e Imágenes, no es una buena manera de publicar los datos, no pueden ser manejados de forma automática, mucho menos realizar operaciones estadísticas.

2 Estrellas: Formato XLS, si bien es un formato en el cual la información podría tratarse, generalmente suelen estar asociadas a licencias privadas. No es lo más recomendable se podría incurrir en gastos innecesarios

3 Estrellas: Formato CSV, JSON y XML, son formatos altamente recomendados, con ellos se puede tratar la información, realizar procesos estadísticos y además son formatos abiertos lo que facilita el uso en diferentes programas informáticos.

4 Estrellas: Formato RDF, se usan URL's para identificar los datos, lo que brinda una manera conveniente de interconectar iniciativas de open data existentes. Los datos RDF pueden ser almacenados en XML y JSON.

5 Estrellas: Linked Open Data: Se pueden ver los datos relacionados, con esto tenemos un diagrama de cómo están interconectados los datos.

El objetivo ideal en cuanto a la elección de los formatos sería publicar en varios formatos donde al menos uno de ellos se encuentre en formato abierto, estándar, estructurado y legible por las máquinas.

Selección de los Datos

Se deberá seleccionar los datos más relevantes de cara a la reutilización ya que se dispone de unos recursos limitados, aunque el objetivo final es categorizar la mayor cantidad de datos posibles.

El criterio a seguir para la selección de datos según lo que nos indica las guías del portal abierto del gobierno de España podría ser el siguiente. Esto solo debe ser tomado como referencia ya que en cada proyecto varía su alcance.

- Exigencias de las normativas existentes
- Relevancia de los datos para la ciudadanía
- Relevancia de los datos para la economía
- Actualidad de los datos
- Volumen de datos
- Nivel de detalle de los datos

Aunque como norma general se debería publicar todo, se debe tener en cuenta que existen datos que tienen ciertas limitaciones como son los siguientes casos:

- Datos personales protegidos por ley
- Sometidos a secreto estadístico
- Afectación a seguridad nacional
- Existencia de derecho intelectual

2.2.1.4 Preparación de datos

El mayor reto que se tiene al exponer los datos de carácter abierto es que estos puedan ser procesados de manera masiva y automática. Esto se logra únicamente si la información se encuentra en formatos electrónicos adecuados, la información esta correctamente estructurada y organizada por categorías entendibles.

Detalle de los datos

El nivel de detalle dependerá de la cantidad de información que se tenga respecto al tema a publicarse, mientras más información se tenga mejor. Es importante respetar el nivel de detalle de origen de los datos. Se incluirán los datos originales tal y como se generaron en origen.

Otros aspectos a tomar en cuenta es el formato, confidencialidad, privacidad y la seguridad como se ha comentado en puntos anteriores de este trabajo. Siempre manteniendo la misma semántica, lo que ayudara al programador con el tratamiento de los datos.

Calidad y fiabilidad de los datos

Este punto depende en gran medida de la empresa u organización publica que los publica. Ellos son los responsables de si los datos publicados son fiables. En cuanto a la presentación se debe mantener un estándar por cada campo en el dataset, esto nos ayuda a evitar posibles depuraciones en un futuro. El realizar estándares es un punto importante para tener datos de calidad.

Metadatos

El último paso y uno de los más importantes antes de publicación es la preparación de los metadatos, fundamentales para la localización, clasificación y reutilización de la información. El objetivo ideal es proporcionar la mayor cantidad de metadatos posibles, pero de una manera correcta por lo que se tiene las siguientes pautas:

- **Obligatorios:** Aquellos que, por su relevancia son indispensables o en cumplimiento de la ley vigente
- **Recomendados:** Aquellos que, legalmente no son obligatorios, pero son relevantes publicarlos.
- **Opcionales:** Aquellos que, no siempre se encuentran disponibles, aun siendo igualmente recomendados



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

El metadato no solo ayudará a localizar y clasificar la información, también ayudará a describir un conjunto de datos y los diferentes recursos. Un ejemplo de metadatos a presentarse podría ser el siguiente:

Ilustración 5: Presentación de metadatos

Obligatorios

- Nombre
- Descripcion
- Tematica
- Organismo / Empresa / Departamento

Recomendados

- Etiqueta
- Licencia

Opcionales

- Fecha de creación
- Fecha ultima actualización
- Idioma
- Locación
- Vigencia
- Recursos asociados

Fuente: Elaboración Propia

2.2.1.5 Publicación

Para la publicación de la información se publica tanto los conjuntos de datos seleccionados, como cualquier otra información adicional en forma de metadatos que pueda ser útil para su localización, clasificación y reutilización con el objetivo de que sean accesibles para todo el mundo. Los puntos más importantes a tener en cuenta sobre la publicación son: infraestructura tecnológica, sistemas de información, enlace con otros catálogos.

Infraestructura Tecnológica

Herramientas propias

Este sería el método más largo y difícil, deberíamos desarrollar nuestra propia plataforma de administración de datos, con las opciones que actualmente existen en el mercado no sería la opción más adecuada.

Herramientas de terceros

Existen muchas herramientas que nos ayudan con la publicación de catálogos de datos, estos podrían ser comparados con los CMS tradicionales, pero en lugar de contenidos lo que administra son datos, se los llama Open Data Portals (Portales de datos abiertos). Como sucede con muchos sistemas de información muchos son de código abierto y gratuitos, otros son licenciados y de pago. Entre los más conocidos se encuentran CKAN, Socrata, Junar, ArcGIS Open Data, DKAN etc.

Sistemas de información

Como sistemas de información deberían contar con al menos las siguientes características:

- **Documentación:** Guías o manuales del uso del sistema
- **Esquemas URIs:** Ofrecen la capacidad de identificación de forma unívoca de cualquier recurso de información a través de la Web, lo que permite establecer identificadores universales para cualquier pieza de información que se desee exponer.
- **Taxonomía:** Es un esquema independiente de conceptos, se la pueda estructurar de la forma que se desee.
- **Actualización:** Se debe poder realizar actualizaciones a los datos dependiendo de su disponibilidad
- **Estabilidad del formato:** El sistema debe mantener el formato o al menos poder exportar el contenido en otro formato, siendo este quien realice la conversión.
- **Seguridad:** Debe contar con los estándares de seguridad básicos.

2.3 Portales abiertos

Los portales de datos abiertos suelen confundirse con un catálogo de datos, pero es mucho más que un catálogo de datos. Un portal suele proveer servicios de búsqueda más avanzados que un catálogo de datos convencional, además puede tener visualizadores de datos dinámicos. Por ejemplo, puede haber un servicio de búsqueda de texto sobre los metadatos que etiqueten a un conjunto de datos, o la posibilidad de pre visualizar, visualizar y hasta explorar los datos desde el mismo portal sin necesidad de descargar ningún archivo.

Existen varios tipos de portales abiertos, la mayoría están desarrollados con software libre como es el caso de CKAN y Socrata. El análisis de estos dos portales se debe a que tanto GVA Oberta (portal Comunidad Valenciana) y Govern Obert (portal Comunidad Cataluña) están implementados en estas dos tecnologías.

2.3.1 CKAN

CKAN es una herramienta de código abierto cuyo principal objetivo es la administración de datos, haciendo que los datos sean accesibles, además provee funcionalidades que permiten agilizar la publicación, el intercambio, la búsqueda y el uso de datos. Al ser un software de código abierto se lo puede usar sin ninguna tarifa de licencia y conservar todos los derechos de los datos y metadatos.

CKAN dispone de una interfaz capaz de mostrar de forma simplificada los datos. A cada conjunto de datos se le puede asignar su propia página con diferentes características, recursos, amplia colección de metadatos, convirtiéndose en un catálogo de datos muy potente capaz de catalogar y presentar la información de forma ordenada, facilitando así la búsqueda.



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Si las características que trae por defecto CKAN no cumplen con las necesidades del proyecto, se tiene la posibilidad de añadir más características disponibles en su comunidad. En la actualidad existen más de 200 extensiones, para casos más específicos se deberá desarrollar una característica adicional.

2.3.1.1 Conjuntos de datos y recursos

En CKAN, se define los datos que se publican en unidades llamadas "conjuntos de datos"; por ejemplo, podría ser las licitaciones de las diferentes instituciones públicas, gastos de las comunidades, estadísticas de criminalidad, variación de la meteorología de la región, etc. Cuando los usuarios buscan datos, los resultados de búsqueda que ven serán conjuntos de datos individuales.

Los conjuntos de datos tienen dos elementos fundamentales

- Información o "metadatos" sobre los datos. Por ejemplo, el título, creador, la fecha, formatos disponibles, licencia entre muchos más.
- Una cantidad de "recursos", que contienen los datos en sí. A CKAN no le importa en qué formato están los datos. Un conjunto de datos puede contener cualquier cantidad de recursos. Por ejemplo, las contrataciones pueden contener datos de diferentes años o con diferentes formatos.

2.3.1.2 Características CKAN

Las características que se mencionaran en esta sección son las que trae CKAN por defecto, no son desarrollos externos por parte de la comunidad.

API: CKAN cuenta con una potente API al estilo RPC expone todas las funcionalidades de la solución al cliente, se puede conectar de manera externa haciendo una llamada a la API

Data Store: Es una base de datos ad hoc para el almacenamiento de los datos estructurados. Los datos pueden importarse al Data Store desde los recursos. La ventaja de usar este tipo de almacenamiento es tener pre visualizaciones automáticas de datos en la página del recurso, usando la extensión Data Explorer y contar con la API nativa de Data Store, la cual posee mayores funcionalidades.

Federación: Nos ayuda a realizar una recolección de datos y metadatos de otros portales, se puede usar para crear una red federada de portales de datos que comparten datos entre sí. Esto es de gran ayuda si, por ejemplo, un portal nacional desea agregar información de instancias CKAN de una comunidad autónoma o gobierno local.

File Store: permite a los usuarios cargar archivos de datos a los recursos y cargar imágenes de logotipos para grupos y organizaciones.

Geoespaciales: Tiene características avanzadas de visualización, búsqueda y descubrimiento.

Cuando datos estructurados con información de geo posición se cargan, se puede visualizar un mapa interactivo con los datos.

Puede comprender una ubicación asociada a un conjunto de datos, y usar esto para ofrecer capacidades de búsqueda geoespacial a través de la interfaz web y la API. Un usuario que busca

conjuntos de datos puede filtrar los resultados por ubicación geográfica, especificando un cuadro delimitador para limitar el área que le interesa.

Metadatos: Posee un amplio conjunto de metadatos por defecto son: Título, Identificador único, Grupos, Descripción, Vista previa de datos, Historial de revisiones, Campos adicionales, Licencias, Tags, Formatos múltiples, Clave API. Pudiendo crear otros de manera personalizada.

Manejo de datos: Mediante su aplicación web los editores pueden registrar, actualizar y refinar los datos fácilmente, pudiendo distribuir las actividades a diferentes colaboradores.

El ingreso de los Datos se lo puede realizar directamente por la interfaz web, mediante la API o a través de hojas de cálculo personalizadas. Además, se puede realizar una recolección de los siguientes repositorios: servidores geoespaciales CSW, catálogos web existentes, páginas de índice HTML simples o carpetas accesibles en la web, ArcGIS, servidores de geoportales y bases de datos Z39.5, otras instancias de CKAN.

Panel de administración, se pueden administrar conjuntos de datos, definir si los datos vana a ser públicos o privados, administrar las fuentes entre otras acciones.

Los conjuntos de datos pueden ser públicos o privados. Si son privados, solo son visibles para los miembros que tengan los permisos necesarios para poder ver esta información. Los administradores pueden aprobar conjuntos de datos para su publicación con una herramienta de edición masiva disponible en CKAN, que le permite buscar, seleccionar facetas y elegir conjuntos de datos para que sean públicos o privados.

Búsqueda: La búsqueda es al estilo Google, se puede usar filtros por cada categoría que se tenga e incluso por los metadatos ingresados.

Temas: Dispone de múltiples temas que pueden ser usados, también se puede desarrollar un tema propio con es el caso de portales como Datos.gob.

Pre visualización: Herramienta que permite pre visualizar los datos cargados en forma de tablas, gráficos, mapas, imágenes o crear uno propio.

2.3.2 Dades Obertes⁹

Es el portal de transparencia de la Generalitat Valenciana, en este se encuentran publicados todos los catálogos de datos disponibles en la comunidad Valenciana, la información está catalogada por temas, etiquetas y formatos.

El portal se encuentra implementado sobre CKAN por lo que cuenta con toda la funcionalidad que nos brinda CKAN, se puede destacar que nos indica el tipo de formatos disponibles para su descarga, además de tener la posibilidad de visualizar los datos antes de descargarlos.

⁹ <http://www.dadesobertes.gva.es>

Ilustración 6: Portal de Datos abiertos comunidad de Valenciana

The screenshot shows the 'dades Obertes' portal interface. At the top, there is a navigation bar with the logo 'dades Obertes' and the text 'PORTAL DE TRANSPARÈNCIA DE LA GENERALITAT VALENCIANA'. The main navigation menu includes 'CONJUNTS DE DADES', 'TEMES', 'PARTICIPA', 'QUÈ SÓN LES DADES OBERTES?', and 'GVA OBERTA'. The current page is 'CONJUNTS DE DADES'. On the left, there are filters for 'Temes' (Ocupació (3)), 'Etiquetes' (treball (3), estadístiques (3), contractes (3)), and 'Formats' (XML (3), JSON (3), CSV (3), PDF (2)). The main content area shows a search for 'Contractació' with 3 results. It lists 'Estadístiques de contractació 2016', 'Estadístiques de contractació 2015', and 'Estadístiques de contractació 2017', each with download options for CSV, JSON, PDF, and XML. The source URL is provided as 'http://www.dadesobertes.gva.es/va/'.

Fuente: <http://www.dadesobertes.gva.es/va/>

El portal de la comunidad Valenciana está bajo la misma tecnología del portal nacional la diferencia radica en la composición de sus datos y el número de ellos.

2.3.3 Govern Obert

Portal de datos abiertos de la comunidad de Cataluña, cuenta con todos los catálogos de datos publicados, está catalogado principalmente por categorías, tipo de datos, tipos de actores y etiquetas. Como cualquier otro portal cuenta con un buscador al estilo del google, facilitando la búsqueda de sus contenidos. Actualmente el portal se encuentra únicamente en el idioma catalán.

Este portal de datos abiertos esta implementado bajo el sistema Socrata. Una de las ventajas que ofrece Socrata es su interfaz simple y limpia, y permite la posibilidad de visualizar y exportar datos directamente del portal. Además, cuenta con la ventaja que representa la facilidad con la que se puede usar el API de conexión, permitiéndonos desarrollar sistemas de mayor calidad.

Una de sus desventajas es que es un sistema propietario, aunque sus datos son públicos la tecnología usada no lo es.

Fuente: <https://analisi.transparenciacatalunya.cat>

2.4 Normativa

Existen diversas leyes, normas y decretos sobre la transparencia, protección de datos, contratos públicos y manejo de los portales de administración electrónica, en las siguientes tablas se menciona las leyes y normas más importantes donde se detallarán las más relevantes para el proyecto actual.

La primera tabla nos nombra las leyes más importantes para este proyecto que rigen aspectos como la reutilización de información en el sector público, leyes de contratos en sector público entre otras.

Tabla 2: Leyes Involucradas

Ley	Descripción
Noviembre 2007 Gobierno de España	LEY 37/2007 sobre Reutilización de la Información del Sector Público. Regula la directiva 2003/98/CE. Última modificación julio 2015.
Octubre 2011 Gobierno de España	Real Decreto 1495/2011 por el que se desarrolla la Ley 37/2007 sobre reutilización de la información del sector público, para el ámbito del sector público estatal.
Noviembre 2011	Real Decreto Legislativo 3/2011, de 14 de noviembre, por el que se aprueba el texto refundido de la Ley de Contratos del Sector Público.
Junio de 2013 Parlamento Europeo y del Consejo	DIRECTIVA 2013/37/UE relativa a la reutilización de la información del sector público. Modifica la Directiva 2003/98/CE
Julio 2015 Gobierno de España	LEY 18/2015 sobre reutilización de la información del sector público. Regula la directiva 2013/37/UE. Modifica y consolida la Ley 37/2007
Noviembre 2017	Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público, por la que se transponen al ordenamiento jurídico español las Directivas del Parlamento Europeo y del Consejo 2014/23/UE y 2014/24/UE, de 26 de febrero de 2014.

Fuente: Elaboración propia

Normativa y Directiva

Son normativas o directivas dictadas por el gobierno español o el parlamento europeo donde se topan temas como la reutilización de información, administración electrónica e interoperabilidad.

Tabla 3: Normativas y Directivas

Normativa/Directiva	Descripción
Noviembre 2003 Parlamento Europeo y del Consejo	DIRECTIVA 2003/98/CE relativa a la reutilización de la información del sector público
Enero 2010 Gobierno de España	Real Decreto 4/2010, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica. Última modificación noviembre 2011
Febrero de 2013 Secretaría de Estado de Administraciones	Resolución por la que se aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información.
Junio de 2013 Parlamento Europeo y del Consejo	DIRECTIVA 2013/37/UE relativa a la reutilización de la información del sector público. Modifica la Directiva 2003/98/CE
Enero 2015 Secretaría de Estado de Administraciones Publicas	Norma UNE 178301 Ciudades Inteligentes y Datos Abiertos Última modificación julio 2015

Fuente: Elaboración propia

Leyes relacionadas

Estas leyes intervienen de manera indirecta al proyecto actual, deben ser tomadas en cuenta, pero no es el foco fundamental del proyecto.

Tabla 4: Leyes relacionadas

Ley	Descripción
Diciembre 1999 Gobierno de España	LEY Orgánica 15-1999 de Protección de Datos (LOPD) Última modificación marzo 2011
Junio 2007 Gobierno de España	LEY 11/2007 de acceso electrónico de los ciudadanos al Sector Público (ENI)
Diciembre 2007 Gobierno España	Real Decreto 1720-2007 Reglamento de desarrollo LOPD – Última modificación marzo 2013
Diciembre 2013 Gobierno España	LEY 19/2013 de Transparencia Acceso a la Información pública y Buen Gobierno (LTAIBG)
Abril 2015 Generalitat Valenciana	LEY 2/2015 de la Generalitat de Transparencia Buen Gobierno y Participación Ciudadana de la CV
Octubre 2015 Gobierno de España	LEY 39/2015 del Procedimiento Administrativo Común de las AAPP. Deroga Ley 11/2007

Fuente: Elaboración propia

2.5 Contratación pública

La contratación pública es el procedimiento por el cual una administración adjudica a una persona física o jurídica la ejecución de una obra, prestación de un servicio o suministro de un bien, para satisfacer una finalidad pública. Al ser los contratos los datos a analizarse en este proyecto se explicarán de manera general la estructura de los contratos y sus diferentes características, al igual que las leyes que rigen a la contratación pública.

2.5.1 Leyes

La ley más actual de contratación pública es la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público, por la que se transponen al ordenamiento jurídico español las Directivas del Parlamento Europeo y del Consejo 2014/23/UE y 2014/24/UE, de 26 de febrero de 2014. Además, deroga el Real Decreto Legislativo 3/2011 de 14 de noviembre. Dentro de la ley actual se busca mejorar y simplificar los procesos haciendo uso de las nuevas tecnologías, así como establecer medidas para evitar la corrupción, entre ellas se encuentra la supresión del procedimiento negociado sin publicidad por razón de la cuantía, la creación de dos tipos de procedimientos abiertos simplificados o la formación de un órgano independiente de supervisión de los contratos públicos. Todo esto bajo la obligatoriedad del uso de portales electrónicos para todas las administraciones públicas.

Se pueden destacar los siguientes puntos de la nueva normativa:

- Se introduce un mejor acceso para las pymes y se impone una menor burocracia en los trámites para los licitadores, además de exigir un mayor control en responsabilidad social, ambiental y hacia los trabajadores.
- El funcionario será el responsable del contrato con la prohibición de las indemnizaciones por expectativas de mercado no cumplidas, por tanto, la administración quedará exenta de responsabilidades en caso de incumplimiento del contrato.
- Se crea un mayor control en el proceso de adjudicación indicando el responsable del contrato – funcionario o empleado público, que se encargara del seguimiento y cumplimiento del contrato.
- Reducción para que las administraciones públicas para adjudicar contratos menores.
- La ley establece una nueva adjudicación de contratos no sujetos a regulación armonizada, cambiando los valores de adjudicación.
- Se reducen hasta 40.000 y 15.000 euros los límites vigentes de 50.000 euros para contratos de obras y 18.000 euros para el resto (IVA excluido) y tampoco podrán adjudicarse contratos sin publicidad.
- La ley obliga a fraccionar en diferentes lotes los contratos grandes, flexibiliza algunos requisitos de solvencia a las características y circunstancias de las empresas y no se exigirá a las sociedades con menos de cinco años la acreditación de trabajos previos, como sucedía hasta ahora.

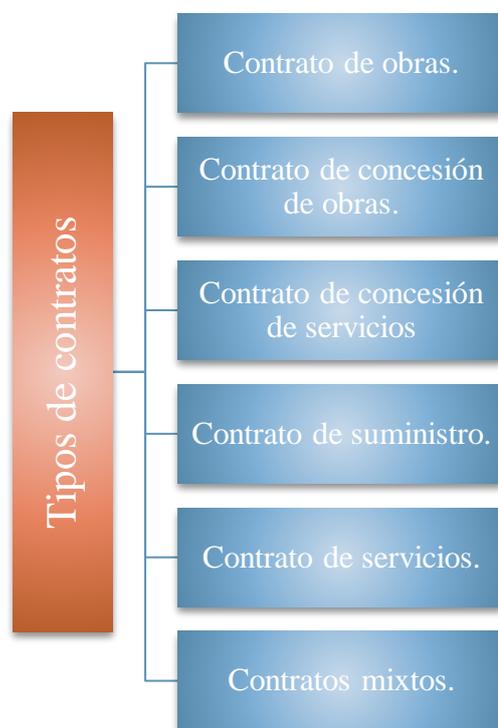


- Mejoras en proyectos sociales dándoles una cierta ventaja si se emplea personas con discapacidad, así como la no admisión de empresas licitantes que estando obligadas no acrediten el cumplimiento de la reserva legal de empleo del 2% de trabajadores con discapacidad.

2.5.2 Tipo de contratos

En esta sección se detallará los tipos contractuales para el sector público. Es importante conocer esta clasificación y sus diferentes definiciones ya que en nuestro proyecto se presentarán algunas graficas filtrando por tipo de contratos. A continuación, realizaremos un pequeño cuadro resumen. La información mostrada en esta sección fue tomada de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público.

Ilustración 8:Tipos de contratos



Fuente: Elaboración propia

Contrato de obras.

Se entiende por contrato de obras a la realización, por cualquier medio, de una obra que cumpla los requisitos fijados por la entidad del sector público contratante que ejerza una influencia decisiva en el tipo o el proyecto de la obra. Por «obra» se entenderá el resultado de un conjunto de trabajos de construcción o de ingeniería civil, destinado a cumplir por sí mismo una función económica o técnica, que tenga por objeto un bien inmueble. (Ley 9/19, Art 13, 2017)

Contrato de concesión de obras.

La concesión de obras es un contrato que tiene por objeto la realización por el concesionario de algunas de las prestaciones a que se refiere el contratos de obras, incluidas las de restauración y reparación de construcciones existentes, así como la conservación y

mantenimiento de los elementos construidos, y en el que la contraprestación a favor de aquel consiste, o bien únicamente en el derecho a explotar la obra en el sentido del apartado cuarto siguiente, o bien en dicho derecho acompañado del de percibir un precio. (Ley 9/19, Art 14, 2017)

Contrato de concesión de servicios

El contrato de concesión de servicios es aquel en cuya virtud uno o varios poderes adjudicadores encomiendan a título oneroso a una o varias personas, naturales o jurídicas, la gestión de un servicio cuya prestación sea de su titularidad o competencia, y cuya contrapartida venga constituida bien por el derecho a explotar los servicios objeto del contrato o bien por dicho derecho acompañado del de percibir un precio. (Ley 9/19, Art 15, 2017)

Contrato de suministro.

Son contratos de suministro los que tienen por objeto la adquisición, el arrendamiento financiero, o el arrendamiento, con o sin opción de compra, de productos o bienes muebles. (Ley 9/19, Art 16, 2017)

Contrato de servicios.

Son contratos de servicios aquellos cuyo objeto son prestaciones de hacer consistentes en el desarrollo de una actividad o dirigidas a la obtención de un resultado distinto de una obra o suministro, incluyendo aquellos en que el adjudicatario se obligue a ejecutar el servicio de forma sucesiva y por precio unitario. (Ley 9/19, Art 16, 2017)

Contratos mixtos.

Se entenderá por contrato mixto aquel que contenga prestaciones correspondientes a otro u otros de distinta clase. (Ley 9/19, Art 18, 2017)

Contratos sujetos a regulación armonizada

Son contratos sujetos a una regulación armonizada los contratos de obras, los de concesión de obras, los de concesión de servicios, los de suministro, y los de servicios, cuyo valor estimado, calculado conforme a las reglas que se establecen en el artículo 101, sea igual o superior a las cuantías que se indican en los artículos, siempre que la entidad contratante tenga el carácter de poder adjudicador. El siguiente cuadro se resumen los contratos sujetos a regulación armonizada de acuerdo a los montos establecidos. (Ley 9/19, Art 19, 2017)

Tabla 5: Contratos sujetos a regulación armonizada

Tipo de contrato	Descripción	
Contrato de obras.	5.225.000,00 €	
Contrato de concesión de obras.		
Contrato de concesión de servicios		



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

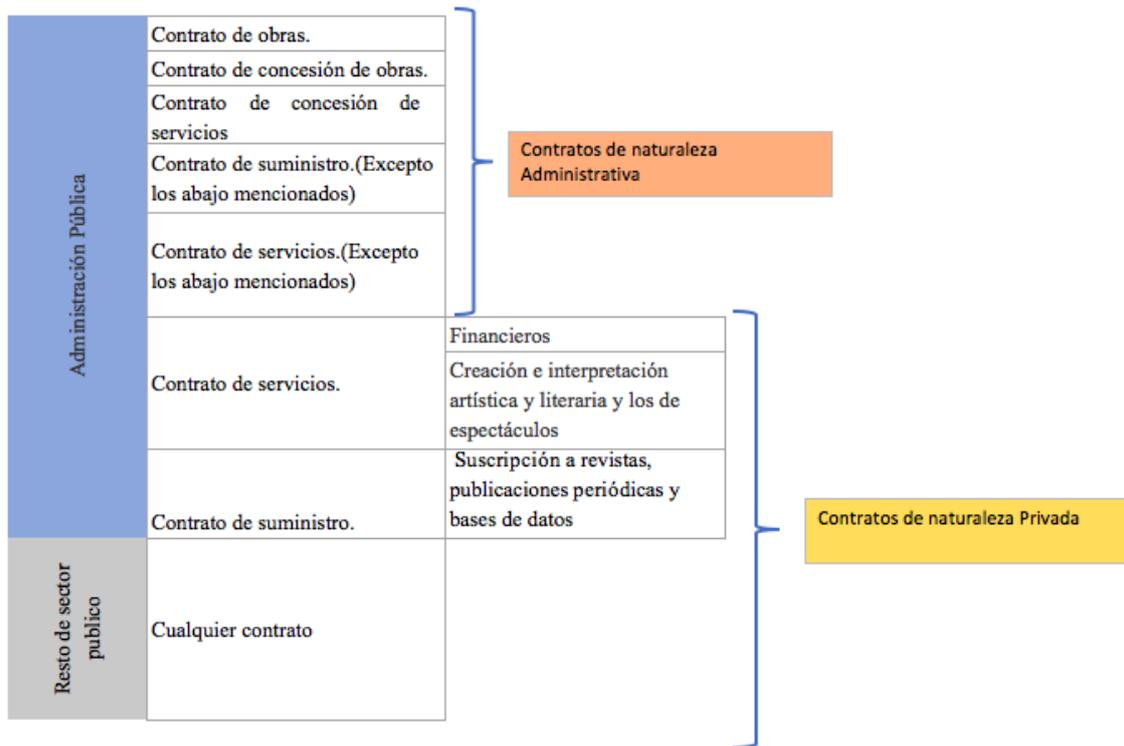
Contrato de suministro.	135.000,00 €	Contratos adjudicados por la Administración General del Estado, sus Organismos Autónomos, o las Entidades Gestoras y Servicios Comunes de la Seguridad Social.
	209.000,00 €	Cuando se trate de contratos de suministro distintos
Contrato de servicios.	135.000,00 €	Contratos hayan de ser adjudicados por la Administración General del Estado, sus Organismos Autónomos, o las Entidades Gestoras y Servicios Comunes de la Seguridad Social
	209.000,00 €	Contratos hayan de adjudicarse por entidades del sector público distintas a la Administración General del Estado, sus Organismos Autónomos o las Entidades Gestoras y Servicios Comunes de la Seguridad Social
	750.000,00 €	Contratos que tengan por objeto los servicios sociales y otros servicios específicos
Contratos mixtos.		

Fuente: Elaboración propia

Contratos administrativos y contratos privados

Los contratos del sector público podrán estar sometidos a un régimen jurídico de derecho administrativo o de derecho privado, dependiendo del tipo de contrato se registrará de diferente manera. (Ley 9/19, Art 25,26, 2017)

Ilustración 9: Contratos administrativo y privados



Fuente: Elaboración propia

2.5.3 Procedimientos

Los procedimientos de adjudicación son un conjunto de trámites legales necesarios para la selección de la persona, física o jurídica, para garantizar no solo la transparencia en la adjudicación del contrato, sino también un mayor control de la administración pública al momento de seleccionar una oferta ventajosa, en beneficio mutuo.

Ilustración 10: Procedimientos de adjudicación



Fuente: Elaboración propia



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Según la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público define los procedimientos de la siguiente manera:

2.5.3.1 Procedimiento Abierto

En el procedimiento abierto todo empresario interesado podrá presentar una proposición, quedando excluida toda negociación de los términos del contrato con los licitadores. (Ley 9/19, Art 156, 2017)

Procedimiento abierto Simplificado

Los órganos de contratación podrán acordar la utilización de un procedimiento abierto simplificado en los contratos de obras, suministro y servicios cuando se cumplan las dos condiciones siguientes

- a) Que su valor estimado sea igual o inferior a 2.000.000 de euros en el caso de contratos de obras, y en el caso de contratos de suministro y de servicios, que su valor estimado sea igual o inferior a 100.000 euros.
- b) Que entre los criterios de adjudicación previstos en el pliego no haya ninguno evaluable mediante juicio de valor o, de haberlos, su ponderación no supere el veinticinco por ciento del total, salvo en el caso de que el contrato tenga por objeto prestaciones de carácter intelectual, como los servicios de ingeniería y arquitectura, en que su ponderación no podrá superar el cuarenta y cinco por ciento del total.

(Ley 9/19, Art 159, 2017)

2.5.3.2 Procedimientos restringidos

Solo podrán presentar proposiciones aquellos empresarios que, a su solicitud y en atención a su solvencia, sean seleccionados por el órgano de contratación. (Ley 9/19, 160, 2017)

2.5.3.3 Procedimientos con negociación

En los procedimientos con negociación la adjudicación recaerá en el licitador justificadamente elegido por el órgano de contratación, tras negociar las condiciones del contrato con uno o varios candidatos.

(Ley 9/19, Art 166, 2017)

2.5.3.4 Dialogo competitivo

En el diálogo competitivo, la mesa especial de diálogo competitivo dirige un diálogo con los candidatos seleccionados, previa solicitud de los mismos, a fin de desarrollar una o varias soluciones susceptibles de satisfacer sus necesidades y que servirán de base para que los candidatos elegidos presenten una oferta.

(Ley 9/19, Art 172, 2017)

2.5.3.5 Procedimiento de asociación para la innovación

La asociación para la innovación es un procedimiento que tiene como finalidad el desarrollo de productos, servicios u obras innovadores y la compra ulterior de los suministros, servicios

u obras resultantes, siempre que correspondan a los niveles de rendimiento y a los costes máximos acordados entre los órganos de contratación y los participantes.

(Ley 9/19, Art 177, 2017)

2.5.4 Racionalización técnica de la contratación

2.5.4.1 Acuerdo Marco

Consiste en preestablecer una serie de términos y condiciones que serán comunes a todos los contratos basados en este acuerdo. Trata de simplificar la gestión de contratos públicos y así facilitar la adhesión de organismos o entidades públicas.

- Podrán celebrarse con uno o varias empresas.
- Los acuerdos marcos pueden ser con todos los términos establecidos o sin todos los acuerdos establecidos, además.
- Se los puede licitar a varios o un solo empresarios. La duración máxima será de 4 años salvo excepciones que deberán ser debidamente justificadas.
- La duración de los contratos basados en un acuerdo marco será independiente de la duración del acuerdo marco.

(Ley 9/19, Art 219, 2017)

2.5.4.2 Basados en un acuerdo marco.

Solo podrán celebrarse contratos basados en un acuerdo marco entre las empresas y los órganos de contratación que hayan sido originariamente partes en el mismo

Cuando el acuerdo marco se hubiese concluido con una única empresa, los contratos basados en aquel se adjudicarán con arreglo a los términos en él establecidos. Los órganos de contratación podrán consultar por escrito al empresario, pidiéndole, si fuere necesario, que complete su oferta. (Ley 9/19, Art 221, 2017)

2.5.4.3 Sistema dinámico

Proceso de adquisición de obras, servicios y suministros de uso corriente, totalmente electrónico y abierto durante la vigencia a las empresas que cumplan los criterios de selección.

- Proceso totalmente electrónico
- Duración limitada fijada y delimitada en los pliegos
- Abierto durante toda su duración período de vigencia a cualquier empresa interesada que cumpla los criterios de selección
- Contratación mediante procedimiento restringido

(Ley 9/19, Art 223, 2017)

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

2.5.4.4 Contrato centralizado

Incluye prestaciones destinadas a varias entidades del sector público como:

- Servicios especializados a los que las entidades del sector público atribuyen la contratación centralizada de obras, servicios y suministros
- Adquieren suministros y servicios para otros entes del sector público: Mayorista (Compra, almacena, revende)
- Adjudica contratos, acuerdos marco y sistemas dinámicos para otros entes: Intermediaria (no paga, ni decepciona, ni vende)

Capítulo 3 : Tecnologías

En este capítulo describiremos las tecnologías a utilizar para el desarrollo de nuestro portal. La explicación de las tecnologías será teórica, ya que en el siguiente capítulo se detallará el desarrollo realizado usando las diferentes herramientas. Además, daremos una breve revisión a las API's de los portales. Ésto nos ayudará a entender de mejor manera como se extrajeron los datos de forma automática.

3.1 API's

Una API es el medio de comunicación entre dos softwares, este puede contener procedimientos, funciones o solo información. Son usadas generalmente en las bibliotecas de programación y los lenguajes más populares para su conexión suelen ser XML y JSON. Esta es la manera en la cual podemos automatizar nuestros procesos, minimizando tiempo y aumentando la productividad.

API de servicios web: Nos permite conexión e intercambio de información entre un servicio web y una aplicación. Normalmente el intercambio se realiza a través del protocolo http o https. Los formatos más usados para esta transferencia de información suelen ser: XML o JSON.

Entre las API de servicios web destacan dos: SOAP, un protocolo de intercambio de información que usa XML para la transferencia de datos y REST que usa el protocolo http. Las funciones más importantes en cualquier servicio REST son: POST (crear), GET (leer y consultar), PUT (editar) y DELETE (eliminar).

API basadas en bibliotecas: este tipo de APIs son las que permiten que una aplicación importe una biblioteca de otro software para hacer el intercambio de información. La gran mayoría de estas bibliotecas están desarrolladas en lenguajes java o javascript.

API basadas en clases: este tipo de APIs nos permite la conexión con las clases o funciones que generalmente se las llama paquetes, como es común en programación orientada a objetos Java.

En nuestro proyecto usaremos APIs de servicios web. Los diferentes portales nos expondrán una URL, la cual nos dará información dependiendo de la consulta que realicemos.

3.1.1 API CKAN

Esta sección mostrará un resumen de la API CKAN, para desarrolladores que deseen incorporar datos o simplemente realizar consultas mediante la API. En el portal de GVA dades obertes, podemos usar la API versión 3 para interactuar.

La API de acción de CKAN es una potente API de estilo RPC que expone todas las funciones principales de CKAN.

La API contará con un enlace principal que en nuestro caso será <http://www.dadesobertes.gva.es/api/3/action/> {logic_action}



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Ejemplo. Para obtener listas con formato JSON de conjuntos de datos, grupos u otros objetos CKAN de un sitio escribimos:

http://www.dadesobertes.gva.es/api/3/action/package_list

http://www.dadesobertes.gva.es/api/3/action/group_list

http://www.dadesobertes.gva.es/api/3/action/tag_list

Las acciones que **consulta** que podemos usar son las siguientes:

Tabla 6: API Acciones de consulta

Acción lógica	Parámetros
site_read	(none)
package_list	(none)
current_package_list_with_resources	limit
revision_list	(none)
package_revision_list	id
group_list	all_fields
group_list_authz	(none)
group_list_available	(none)
group_revision_list	id
licence_list	(none)
tag_list	q, all_fields, limit, offset, return_objects
user_list	q, order_by
package_relationships_list	id, id2, rel
package_show	id
revision_show	id
group_show	id
tag_show	id
user_show	id
package_show_rest	id
group_show_rest	id
tag_show_rest	id
package_autocomplete	q
tag_autocomplete	q, limit
format_autocomplete	q, limit
user_autocomplete	q, limit
package_search	q, fields, facet_by, limit, offset

Las siguientes acciones no podrán ejecutarse al no tener permiso por parte del administrador del portal.

Acciones de **inserción**:

Tabla 7: API Acciones de inserción

Acción lógica	Parámetros
package_create	(package keys)
package_create_validate	(package keys)
resource_create	(resource keys)
package_relationship_create	id, id2, rel, comment
group_create	(group keys)
rating_create	package, rating
user_create	(user keys)
package_create_rest	(package keys)
group_create_rest	(group keys)

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Acciones de **actualización**:

Tabla 8: API Acciones de actualización

Acción lógica	Parámetros
make_latest_pending_package_active	id
resource_update	(resource keys)
package_update	(package keys)
package_update_validate	(package keys)
package_relationship_update	id, id2, rel, comment
group_update	(group keys)
user_update	(user keys), reset_key
package_update_rest	(package keys)
group_update_rest	(group keys)

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Acciones de **borrado**:

Tabla 9: API Acciones de borrado

Acción lógica	Parámetros
package_delete	id
package_relationship_delete	id, id2, rel



group_delete	id
--------------	----

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

API de búsqueda

Los recursos están publicados y disponibles con una variedad de formatos de datos, por cada recurso existe una ubicación. Cada una de las ubicaciones tiene métodos específicos.

Estos son los recursos publicados de la API de búsqueda.

Tabla 10: API Acciones de búsqueda

Recurso	Ubicación
Dataset Search	/search/dataset
Resource Search	/search/resource
Revision Search	/search/revision
Tag Counts	/tag_counts

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Ejemplo: <http://www.dadesobertes.gva.es/api/search/dataset>

Métodos de búsqueda

Dependen del recurso que se necesite.

Tabla 11: API Métodos de búsqueda

Recurso	Método	Solicitud	Respuesta
Dataset Search	POST	Dataset-Search-Params	Dataset-Search-Response
Resource Search	POST	Resource-Search-Params	Resource-Search-Response
Revision Search	POST	Revision-Search-Params	Revision-List
Tag Counts	GET		Tag-Count-List

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Ejemplo: [http://www.dadesobertes.gva.es/api/search/dataset?q=contratos%](http://www.dadesobertes.gva.es/api/search/dataset?q=contratos%25)

Formatos de búsqueda

Al realizar una búsqueda podemos usar diferentes parámetros dependiendo el recurso que se use, a continuación, se enumeran los parámetros por recurso.

Parámetros del dataset

Tabla 12: API Formatos de búsqueda

Parámetro	Valor del parámetro	Ejemplo
q	Search-String	q=geodata
		q=government%20sweden
		q=%22drug%20abuse%22
		q=title:census
		q=tags:maps&tags:country-uk
qjson	JSON encoded options	['q': 'geodata']
fl	list of fields	fl=name
		fl=name,title
		fl=*
sort	field name, asc / dec	sort=name asc
		sort=metadata_modified asc
start, rows	result-int (defaults: start=0, rows=20)	start=40&rows=20
all_fields	0 (default) or 1	all_fields=1

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Ejemplo: <http://www.dadesobertes.gva.es/api/search/dataset?q=contratos;sort=name%20asc>

Parámetros del recurso

Tabla 13: API Parámetros de recurso

Parámetro	Valor del parámetro	Ejemplo
url, format, description	Search-String	url=statistics.org
		format=xls
		description=Research+Institute
qjson	JSON encoded options	['url': 'www.statistics.org']
hash	Search-String	hash=b0d7c260-35d4-42ab-9e3d-c1f4db9bc2f0
all_fields	0 (default) or 1	all_fields=1
offset, limit	result-int (defaults: offset=0, limit=20)	offset=40&limit=20

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Ejemplo: <http://www.dadesobertes.gva.es/api/search/resource?format=csv>

Parámetros de revisión

Tabla 14: Parámetros de revisión

Parámetro	Valor del parámetro	Ejemplo
-----------	---------------------	---------

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

since_time	Date-Time	since_time=2010-05-05T19:42:45.854533
since_id	Uuid	since_id=6c9f32ef-1f93-4b2f-891b-fd01924ebe08

Fuente: <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Otros ejemplos para el portal GVA Dades Obertes

Como hemos visto antes para obtener listas con formato JSON de conjuntos de datos, grupos u otros objetos CKAN de un sitio escribimos:

http://www.dadesobertes.gva.es/api/3/action/package_list

http://www.dadesobertes.gva.es/api/3/action/group_list

http://www.dadesobertes.gva.es/api/3/action/tag_list

Para obtener una representación JSON completa de un conjunto de datos, recursos u otro objeto pondríamos:

http://www.dadesobertes.gva.es/api/3/action/package_show?id=tra-ocu-contratos-2015

http://www.dadesobertes.gva.es/api/3/action/tag_show?id=alumnat

http://www.dadesobertes.gva.es/api/3/action/group_show?id=economia

Para buscar paquetes o recursos que coincidan con una consulta usamos:

http://www.dadesobertes.gva.es/api/3/action/package_search?q=tra-ocu-contratos-2017

http://www.dadesobertes.gva.es/api/3/action/resource_search?query=name:Contra

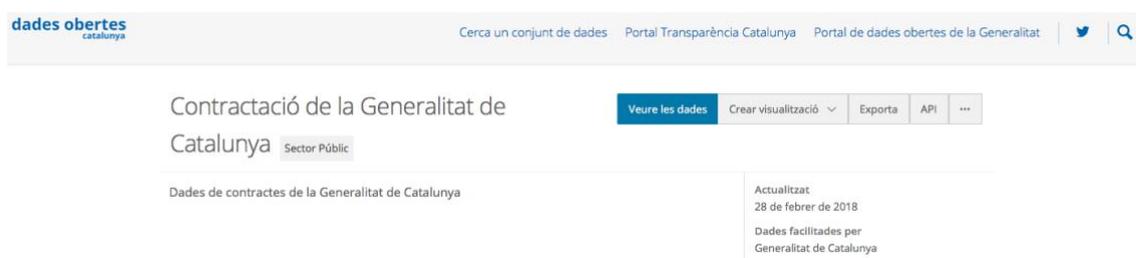
3.1.2 API Socrata

La comunidad de Cataluña alberga sus datos y los gestiona en la plataforma Socrata, una plataforma de código propietario, aunque su API ofrece una versión de código abierto. En ésta podremos encontrar catálogos de datos de diferentes gobiernos, ayuntamientos, ONG y diferentes entidades que tengan su portal de datos abiertos con Socrata. Esta API, al igual que CKAN, nos proporciona toda la funcionalidad que posee Socrata. Sin embargo, en esta sección únicamente nos enfocaremos en la extracción de datos, no en toda la funcionalidad que ofrece el API de Socrata.

La API de Socrata es muy fácil de usar. En primer lugar, debemos identificar el sitio web. En este trabajo se tomará como referencia el portal de dades obertes Catalunya, de ahí se extraerá el conjunto de datos necesario para su análisis.

Todos los conjuntos de datos abiertos de Socrata tienen una API de SODA incorporada. Dependiendo como se ha implementado en el sistema tendremos variaciones en el API Endpoint y en el link para la documentación de del API.

En cada conjunto de datos se encontrará un botón llamado API en la esquina superior derecha, en ella encontraremos los detalles sobre la API Endpoint y un enlace a la documentación de la API.



Fuente: <https://analisi.transparenciacatalunya.cat/Sector-P-blic/Contractaci-de-la-Generalitat-de-Catalunya/hb6v-jcbf>

La ubicaci n del bot n API puede variar dependiendo el portal al que se est  accediendo. Si accedemos al link donde se encuentra toda la documentaci n de la API podremos encontrar las siguientes secciones:

3.1.2.1 Descripci n general

En esta secci n se detalla la manera en la que la API puede conectarse y los formatos en los cuales se pueden exportar los datos. Tambi n podemos acceder al Endpoint del conjunto de datos, siendo  sta una direcci n com n que apunta al dataset.

La informaci n m s relevante de esta secci n podr a ser la informaci n general del dataset del conjunto de datos al cual accedimos. Por ejemplo, ingresando el conjunto de datos ‘‘Dades de contractes de la Generalitat de Catalunya’’ encontraremos la siguiente informaci n del dataset:

- **Dataset Identifier:** ig8n-drgz
- **Total Rows:** 205974
- **Source Domain:** analisi.transparenciacatalunya.cat
- **Created:** 6/30/2016, 1:12:31 PM
- **Last Updated:** 2/28/2018, 9:03:46 AM
- **Category:** Sector P blic
- **Attribution:** Generalitat de Catalunya
- **License:**
- **Owner:** Generalitat de Catalunya
- **Endpoint Version:** 2.1

3.1.2.2 APP Tokens

Las aplicaciones que accedan a los datos mediante la API de desarrollo deber n usar un token de aplicaci n para cada solicitud que realicen. De esta manera se podr  identificar a la aplicaci n que est  solicitando los datos. Se pueden hacer solicitudes sin token de aplicaci n, pero est n sujetas a disponibilidad y la velocidad de respuesta puede variar dependiendo de la disponibilidad, teniendo en cuenta que su prioridad es baja. Con un token de conexi n la aplicaci n solicitante tiene garantizado su acceso a la informaci n.

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Para solicitar un token de aplicación únicamente se tiene que registrar como una aplicación que va a realizar solicitudes a ese portal o siendo más específico a ese conjunto de datos. Para el portal de dades obertes de Catalunya se tiene que realizar el registro solicitando un permiso especial a la Generalitat de Catalunya, por tanto, para este trabajo realizaremos las solicitudes sin registro.

3.1.2.3 Campos

Describen los campos que conforman el dataset, así como el tipo de formato y una breve descripción de cada uno de ellos. Usando filtros y consultas SoQL, puede buscar registros, limitar sus resultados y cambiar la forma en que se emiten los datos. Para afinar el nivel de búsqueda se puede combinar el filtrado mediante URL, donde se añaden parámetros en la URL o mediante el uso de consultas SoQL.

Ejemplo de descripción de campos

Ilustración 12: Descripción de campos

+	situaci_contractual	text	SITUACIÓ CONTRACTUAL
+	exercici	number	EXERCICI
+	subjecte_ambit	text	SUBJECTE/AMBIT
+	id_agrupacio_organisme	number	ID AGRUPACIO ORGANISME
+	agrupacio_organisme	text	AGRUPACIO ORGANISME
+	id_organisme_contractant	number	ID ORGANISME CONTRACTANT
+	organisme_contractant	text	ORGANISME CONTRACTANT
+	codi_expedient	text	CODI EXPEDIENT
+	procediment_adjudicacio	text	PROCEDIMENT ADJUDICACIO
+	tipus_contracte	text	TIPUS CONTRACTE

Fuente: <https://dev.socrata.com/foundry/analisi.transparenciacatalunya.cat/ig8n-drgz>

Filtros Simples

Los filtros simples son la manera más sencilla de filtrar contenido. Únicamente se tiene que colocar el nombre de la columna con el valor que se desee filtrar.

Ejemplo: https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?agrupacio_organisme=DEPARTAMENT%20DE%20SALUT

Se puede aumentar en número de filtros con el operador lógico *and* (&) para hacer consultas un poco más detalladas.

Ejemplo: https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?agrupacio_organisme=DEPARTAMENT%20DE%20JUSTÍCIA&tipus_contracte=SERVEIS

Consultas SoQL

SoSQL¹⁰, nos permite realizar consultas avanzadas sobre el dataset. Como en su nombre se puede evidenciar tiene una estructura similar al lenguaje SQL, por lo que usaremos la misma estructura y parámetros a través de la url o usando un lenguaje externo.

Cuadro de parámetros

Tabla 15: Cuadro SQL

Parámetros	Descripción	Default
\$select	Similar a SELECT en SQL	Todas las columnas, \$select=*
\$where	Similar a WHERE en SQL	Sin filtro
\$order	Similar a ORDER BY en SQL	Sin orden
\$group	Similar a GROUP BY en SQL	Sin agrupación
\$having	Similar a HAVING en SQL	Sin filtro
\$limit	Similar a LIMIT en SQL	1000 (v2.0: Maximo de 50,000;v 2.1: unlimited »)
\$offset	Similar a OFFSET en SQL	0
\$q	Realiza una búsqueda de texto completo para un valor.	Sin busqueda
\$query	Una cadena de consulta SoQL completa, todo como un parámetro	N/A
\$\$bom	Prepara una marca de orden de bytes UTF-8 al comienzo de la salida CSV	FALSO

Fuente: <https://dev.socrata.com/docs/queries/>

Con estos parámetros podríamos construir una consulta avanzada como la siguiente:

Ejemplo: [https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?agrupacio_organisme=DEPARTAMENT%20DE%20SALUT&\\$where=import_adjudicacio%3E0&\\$order=import_adjudicacio%20desc](https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?agrupacio_organisme=DEPARTAMENT%20DE%20SALUT&$where=import_adjudicacio%3E0&$order=import_adjudicacio%20desc)

Cuadro de funciones

Al igual que los parámetros las funciones que se mencionan a continuación se pueden usar en una consulta SoSQL.

Tabla 16: Funciones Socrata

Versión	Descripción
avg(...)	Devuelve el promedio de un conjunto dado de números
between ... and ...	Devuelve TRUE para valores en un rango dado

¹⁰ Lenguaje de consulta Socrata

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

case(...)	Devuelve diferentes valores basados en la evaluación de comparaciones booleanas
convex_hull(...)	Devuelve la geometría convexa mínima que encierra todas las geometrías dentro de un conjunto
count(...)	Devuelve un recuento de un conjunto determinado de registros
date_extract_d(...)	Extrae el día de la fecha como un número entero.
date_extract_dow(...)	Extrae el día de la semana como un número entero entre 0 y 6 (inclusive).
date_extract_hh(...)	Extrae la hora del día como un número entero entre 0 y 23 (inclusive).
date_extract_m(...)	Extrae el mes como un entero.
date_extract_mm(...)	Extrae el minuto del tiempo como un entero.
date_extract_ss(...)	Extrae el segundo del tiempo como un entero.
date_extract_woy(...)	Extrae la semana del año como un número entero entre 0 y 51 (inclusive).
date_extract_y(...)	Extrae el año como un número entero.
date_trunc_y(...)	Trunca una fecha del calendario en el umbral del año
date_trunc_yr(...)	Trunca una fecha del calendario en el umbral de año / mes
date_trunc_yrmd(...)	Trunca una fecha del calendario en el umbral año / mes / fecha
distance_in_meters(...)	Devuelve la distancia entre dos puntos en metros
extent(...)	Devuelve un cuadro delimitador que incluye un conjunto de geometrías
in(...)	Coincide con los valores en un conjunto dado de opciones
intersects(...)	Le permite comparar dos tipos geoespaciales para ver si se cruzan o se superponen entre sí
like '...'	Permite búsquedas de subcadenas en cadenas de texto
lower(...)	Devuelve el equivalente en minúsculas de una cadena de texto
max(...)	Devuelve el máximo de un conjunto dado de números
min(...)	Devuelve el mínimo de un conjunto dado de números
not between ... and ...	Devuelve TRUE para valores que no están en un rango dado
not in(...)	Coincide con los valores que no están en un conjunto dado de opciones
not like '...'	Permite la coincidencia de campos de texto que no contienen una subcadena
num_points(...)	Devuelve la cantidad de vértices en un registro de datos geoespaciales
simplify(...)	Reduce el número de vértices en una línea o polígono
simplify_preserve_topology(...)	Reduce el número de vértices en una línea o polígono, preservando la topología

starts_with(...)	Coincide con cadenas de texto que comienzan con una subcadena dada
stddev_pop(...)	Devuelve la desviación estándar de población de un conjunto dado de números
stddev_samp(...)	Devuelve una desviación estándar muestrada de un conjunto dado de números
sum(...)	Devuelve la suma de un conjunto dado de números
upper(...)	Devuelve el equivalente en mayúsculas de una cadena de texto
within_box(...)	Devuelve las filas que tienen geodatos dentro del cuadro especificado, definido por latitud y longitud
within_circle(...)	Devuelve las filas que tienen ubicaciones dentro de un círculo específico, medido en metros
within_polygon(...)	Devuelve las filas que tienen ubicaciones dentro del cuadro especificado, definidas por latitud y longitud
within_polygon(...)	Devuelve las filas que tienen ubicaciones dentro del cuadro especificado, definidas por latitud y longitud

Fuente: <https://dev.socrata.com/docs/functions>

Ejemplo: [https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?\\$select=count\(agrupacio_organisme\)as%20cuenta&\\$where=%20agrupacio_organisme=%22DEPARTAMENT%20DE%20SALUT%22](https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json?$select=count(agrupacio_organisme)as%20cuenta&$where=%20agrupacio_organisme=%22DEPARTAMENT%20DE%20SALUT%22)

Nota: Cabe destacar que el explorador sustituye los espacios por %20 y las comillas por %22, por lo que la sentencia podría escribirse con espacios y comillas.

3.1.2.4 Fragmentos de código

En esta sección se exponen diferentes ejemplos en lenguajes como jQuery, Python, PowerShell, R, SAS y SODA, este código es para acceder al dataset de forma automática.

En nuestro caso accederemos mediante R-Socrata. Donde se usaría el siguiente código:

```
## Install the required package with:
## install.packages("RSocrata")

library("RSocrata")

df <- read.socrata(
  "https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json"
)
```

Si tuviéramos el token de la aplicación deberíamos aumentar los campos de app_token, email y password. De esta manera nos aseguraríamos que nuestra solicitud va a ser aceptada, como se describe en el punto 3.1.2.2.

3.2 R-Studio

Es un entorno de desarrollo integrado (IDE) especializado para el lenguaje R. Cuenta con dos versiones una de código abierto y la otra comercial en caso de requerir características



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

adicionales y una licencia especial. Además cuenta con dos modos de despliegue la de escritorio y de servidor, generalmente en un entorno comercial se usan las versiones de pago con un despliegue sobre servidores para mejorar el rendimiento en el momento de desarrollar.

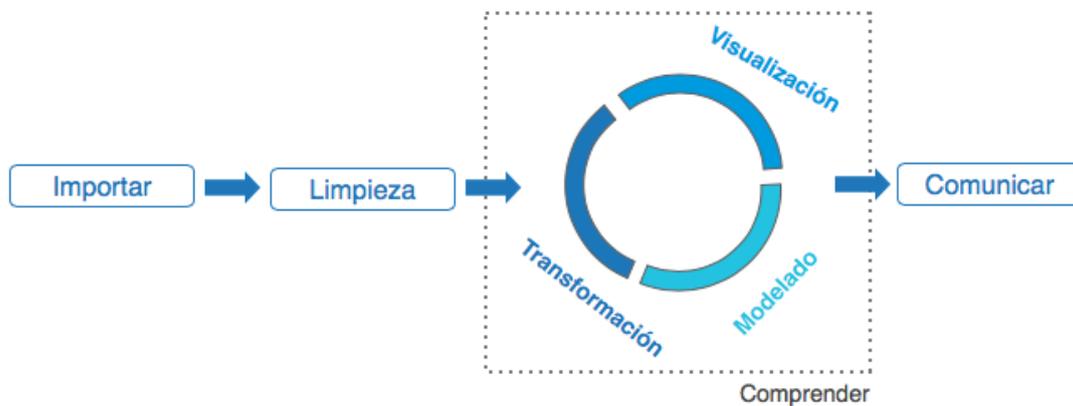
La diferencia principal entre la versión de escritorio de código abierto y la licencia comercial es el nivel de soporte que ofrece la de pago, ya que en la licencia de código abierto únicamente tenemos soporte por parte de la comunidad.

En el caso de la versión servidor si se encuentran grandes diferencias, se puede encontrar todas sus diferencias en la siguiente página web: <https://www.rstudio.com/products/RStudio/>

Una de las ventajas más reconocidas de R-Studio es la gran facilidad de integrar funcionalidades creadas por desarrolladores externos. Esto se realiza en forma de paquetes que cuentan con la correspondiente documentación y que permiten implementar diferentes funciones. La facilidad con la que se instalan las nuevas características brindan una ventaja de cara al desarrollador, siendo capaz de manejar múltiples acciones en un solo entorno de desarrollo.

Para este trabajo se usarán múltiples paquetes que nos ayudarán a tener un modelo como se presenta en la siguiente figura.

Ilustración 13: Modelo R Studio



Fuente: Adaptado de R for data science, Wickham, H., & Golemund, 2017.

Más concretamente usaremos los siguientes paquetes:

Importar: API's, Rsocrata

Limpieza: Tidyverse

Visualización: Shiny, ggplot 2

Modelado: modelr

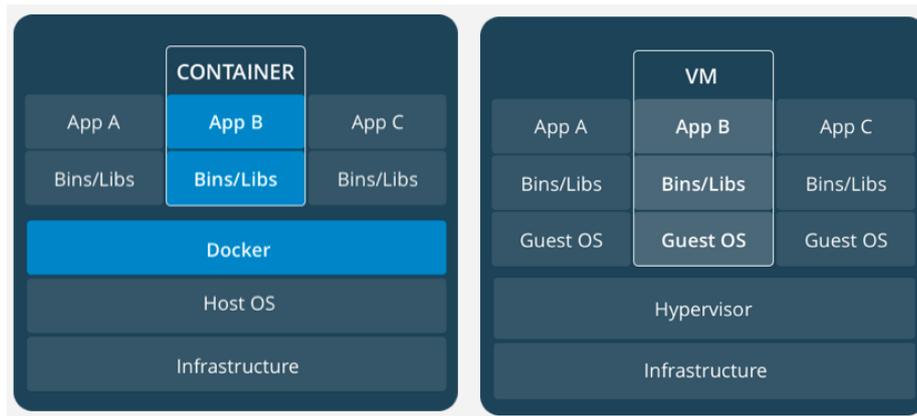
Transformación: dplyr

Comunicar o graficación: ggplot 2, billboarder, visNetwork, igrph, sunburstR

3.3 Docker

Es la tecnología de virtualización de contenedores. Una imagen de contenedor es un paquete ligero, autónomo y ejecutable de una pieza de software que incluye todo lo necesario para ejecutarlo: código, tiempo de ejecución, herramientas del sistema, bibliotecas del sistema, configuraciones. Disponible para aplicaciones basadas en Linux y Windows, el software en contenedor siempre funcionará igual, independientemente del entorno. Los contenedores aíslan el software de su entorno.

Ilustración 14: Comparación entre contenedores y máquinas virtuales



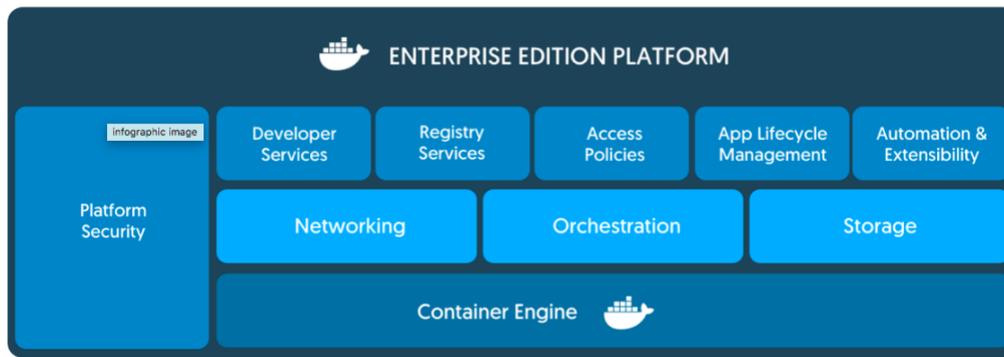
Fuente: <https://www.docker.com/what-container>

Entonces, Docker es como las máquinas virtuales, pero mucho más ligeras. Esto es posible gracias al uso de contenedores sin tener que cargar todo el sistema operativo para cada máquina virtual. Se revuelve este problema corriendo los contenedores como un subproceso de sistema operativo, por lo tanto, tendremos múltiples contenedores en un mismo sistema este puede ser virtual o físico. Con esto tendremos una verdadera independencia entre las aplicaciones, la infraestructura, los desarrolladores y las operaciones de TI. Además de construir contenedores, se facilita el desarrollo, ayudando a los programadores a construir contenedores y aplicaciones dentro de contenedores para luego compartirlos entre sus compañeros de equipo.

Para un ambiente empresarial se debería usar la versión de Docker Enterprise Edition, ya que cuenta con prestaciones importantes para la empresa como lo son: flexibilidad de uso de varias nubes o infraestructuras, seguridad, actualizaciones, procesos unificados y sobre todo el soporte.

La arquitectura de Docker Enterprise Edition vendría a ser la siguiente (la versión Community es la misma, pero sin tener algunos beneficios).

Ilustración 15: Docker Enterprise Edition



Fuente: <https://www.docker.com/enterprise-edition>

3.4 Shiny

Es un paquete de R que permite crear aplicaciones web dinámicas directamente desde R Studio. Se puede crear las aplicaciones en lenguaje R de forma nativa, con la ventaja de contar con todas las características que contaríamos en un desarrollo web tradicional, incluso temas CSS, JavaScript y WidgetsHTML.

Adicional a este se usó “shinydashboard”, que cuenta con una funcionalidad mayor en cuanto a interfaz gráfica, trae por defecto configuraciones predeterminadas entre la que cabe destacar que todo el contenido se adapta a cualquier contenido sin importar la resolución de la pantalla además de tener temas CSS pre-definidos.

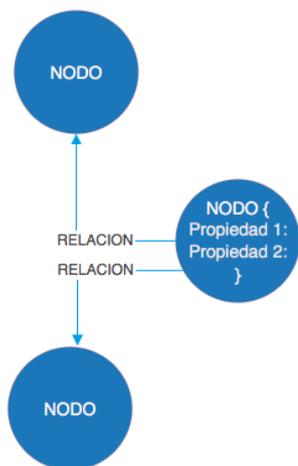
La ventaja que nos proporciona usar Shiny es la incrustación del código de manera fácil sin tener que hacer llamados innecesarios a componentes externos, además cuenta con la posibilidad de alojar la web de manera local o en los servidores de Shiny. En ambos casos se puede optar por la versión gratuita o comercial, siendo la versión comercial más apta para empresas que deseen tener mayor disponibilidad, seguridad y mantenimiento.

3.5 Neo4j

Es una base de datos orientada a grafos(BDOG), usa grafos para representar datos y las relaciones entre ellos. Un grafo es un conjunto de nodos y relaciones que se conectan entre si representado la relación que existe entre los datos.

Elementos de un grafo

Ilustración 16: Estructura del grafo



Nodo: Es la parte fundamental de un grafo, si hacemos relación con un modelo tradicional sería la entidad de un BDD.

Propiedades: Se compone de una clave y un valor donde se describe al nodo o relaciones.

Relación: Conecta dos nodos donde se especifica el sentido puede ser unidireccional o bidireccional.

Etiquetas: Asocia un nombre común al conjunto de nodos o relaciones. Un nodo o relación puede tener una o más etiquetas.

Fuente: elaboración propia

En una BDOG se permite realizar operaciones como crear, leer, actualizar, eliminar (CRUD) sobre un modelo de grafos. A diferencia de las bases de datos tradicionales, las relaciones tienen prioridad en las bases de datos de grafos. Esto significa que su aplicación no tiene que inferir conexiones de datos usando claves externas o procesamiento innecesario, como MapReduce. Esto representa una gran ventaja hablando del procesamiento de las operaciones, mientras que en una BDD relacional se tiene múltiples operaciones de JOIN para enlazar tablas, en Neo4j se mantienen relaciones entre nodos facilitando así la búsqueda de la información.

La característica principal de las BDOG es ayudar a dar sentido y relacionar la gran variedad de datos que poseemos para poder analizar la información, de esta manera poder generar valor de los datos que poseemos. Para ello se puede crear una variedad de grafos:

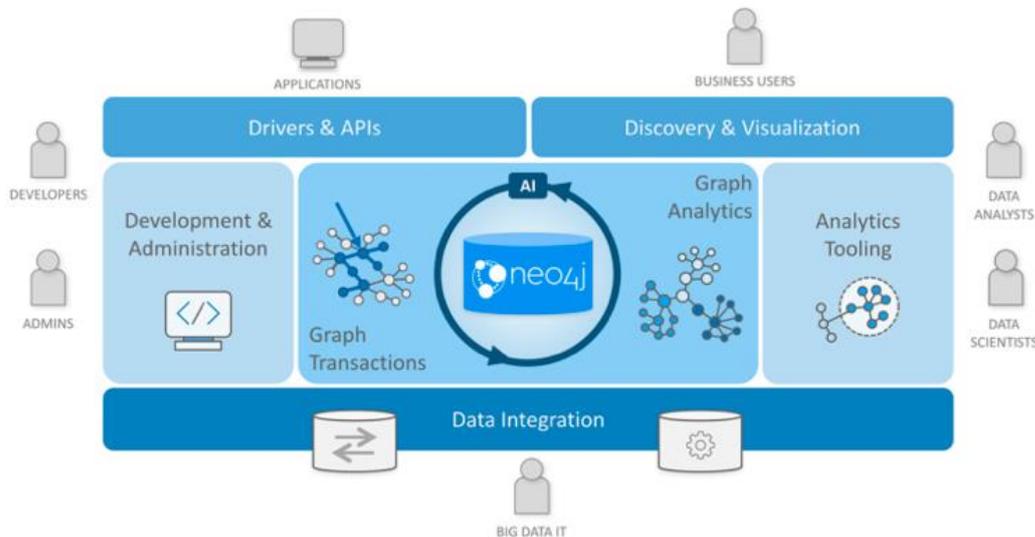
- Grafos no dirigidos: La relación entre los nodos se puede interpretar en cualquier sentido, no necesariamente en una dirección.
- Grafos dirigidos: La relación que existe entre los nodos es dirigida no bidireccional, un ejemplo de ellos es las solicitudes en Instagram, se puede seguir a una persona sin que ella te siga a ti.
- Grafos con peso: Son grafos que dentro de su relación tienen un valor numérico que en futuro se usará para realizar operaciones. Ejemplo: Un contrato hacia una empresa cuenta con un importe, ese sería el peso asignado.
- Grafos con etiquetas: Se etiquetan los grafos para tener un mejor entendimiento de la relación que existe entre los nodos. En Facebook podríamos tener nodos definidos por términos como 'amigo' o 'compañero de trabajo' y las relaciones como 'amigo de' o 'socio de'.
- Grafos de propiedad: Son grafos que cuentan con pesos, etiquetas y además propiedades definidas para cada nodo. Es el más complejo de los grafos, pero a su vez el más usado.



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Neo4j no solo se ha convertido en una base de datos de grafos, sino que ha creado una plataforma de grafos, que cuenta entre otras funcionalidades con: herramientas de análisis de grafos, APIs para conectarse con Neo4j de manera automática, ETL (Extract, Transform and Load) PONER LO QUE ES de grafos, visualización de grafos e integración de descubrimiento y una nueva edición de Enterprise para desarrolladores

Ilustración 17: Plataforma Neo4j



Fuente: <https://neo4j.com/whats-new-in-neo4j/>

3.5.1 Cypher

Es el lenguaje estándar para la tecnología basada en grafos. Como en las bases de datos tradicionales se usa SQL para realizar consultas, inserción, actualización y eliminación de datos, en las bases de datos orientadas a grafos se cuenta con el lenguaje Cypher que es un lenguaje que nos permite realizar operaciones sin tener que programar exactamente lo que desea realizar.

Neo4j cuenta con un compilador Cypher que desempeña un papel prominente, lo que le permite utilizar este lenguaje de consulta simple pero poderoso para integrar su trabajo con todas las diversas herramientas y facetas de Graph Platform.

Los principios básicos de Cypher son los siguientes:

- Buscar los patrones de los nodos y las relaciones en el grafo, para extraer información o modificar los datos.
- Tiene el concepto de identificadores que denotan nombres, límite de elementos y parámetros.
- Puede realizar operaciones CRUD sobre nodos, relaciones, etiquetas y propiedades.
- Maneja índices y restricciones.

Capítulo 4 : Desarrollo de plataforma

4.1 Desarrollo en Shiny R

Usualmente el desarrollo en Shiny se compone de dos tipos de contenidos UI (que se refiere a toda la interfaz gráfica) y Server (que es lo que permite ejecutar los procesos). Cabe destacar que dichos contenidos no son los únicos pudiendo haber otros. Así, en este proyecto se incluyó otro tipo de componentes como son hojas de estilo (css) y elementos javascript. En la sección del server se encuentran las funciones de limpieza, transformación y extracción de datos. En la sección de UI se encuentran las funciones de gráficos, grafos y hojas de estilo, aunque en cierta manera al ser funciones interactivas algunas operaciones de procesamiento se las debe incluir necesariamente en el UI (por ejemplo, al cargar las diferentes comunidades autónomas debemos extraer los valores de la base de datos de grafos para poder llenar los campos dinámicamente).

4.1.1 Extracción

Los datos han sido extraídos mediante las API's de cada uno de los portales de datos abiertos de ambas comunidades, se los puede descargar en varios formatos, la elección para este proyecto fue CSV por la facilidad de inserción de los datos en datasets de R.

4.1.1.1 Cataluña

Los datos de contratación pública en el portal de datos abiertos de Cataluña cuentan con los siguientes campos y opciones para cada uno de los campos.

Tabla 17: Campos Cataluña

CATALUÑA		
CAMPO	TIPO DE CAMPO	OPCIONES
SITUACIÓ CONTRACTUAL	Texto	Prorroga
		Adjudicació
		Menor
EXERCICI	Numérico	1998 - 2018, 2035
SUBJECTE/AMBIT	Texto	
ID AGRUPACIO ORGANISME	Numérico	
AGRUPACIO ORGANISME	Texto	
ID ORGANISME CONTRACTANT	Numérico	
ORGANISME CONTRACTANT	Texto	
CODI EXPEDIENT	Texto	
PROCEDIMENT ADJUDICACIO	Texto	Obert
		Negociat
		Restringit
		Altres adjudicacions segons instruccions internes aprovades
		Derivat d'acord marc
		Contractes d'emergències
		Diàleg Competitiu
		Específic de sistema dinàmic de contractació
Menor		

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

		Negociat amb publicitat
		Negociat sense publicitat
		Concurs de projectes
		Obres
		Gestió de servei públic
		Subministraments
		Serveis
		Administratiu especial
		Colaboracion Publica Privada
		Concesión de obra publica
		Consultoría Asistencia
TIPUS CONTRACTE	Texto	
DESCRIPCIO EXPEDIENT	Texto	
NUMERO LOT	Numérico	0 - 99
CODI CPV	Texto	
ADJUDICATARI	Texto	
IMPORT ADJUDICACIO	Numérico	
DATA ADJUDICACIO	Fecha	
CONTRACTE	Texto	
LOT DESERT	Texto	S / N
DIES DURADA	Numérico	
MESOS DURADA	Numérico	
ANYS DURADA	Numérico	
NUMERO PRORROGA	Numérico	
DATA INICI PRORROGA	Fecha	
DATA FI PRORROGA	Fecha	

Fuente: Elaboración propia

Para la extracción de datos de Cataluña se usó la API de Socrata, ejecutada desde nuestro programa R.

Últimos registros

```
actualizar_Online<-function(){
tf_cat<-read.csv(file = "https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.csv",fill =
TRUE,flush = TRUE,encoding = "UTF-8",stringsAsFactors = FALSE)
return(tf_cat%>%
select(exercici,procediment_adjudicacio,tipus_contracte,organisme_contractant,adjudicatari,import
t_adjudicacio,dies_durada,mesos_durada,anys_durada))
}
```

Toda la información

```
## Install the required package with:
## install.packages("RSocrata")

library("RSocrata")

df <- read.socrata(
  "https://analisi.transparenciacatalunya.cat/resource/ig8n-drgz.json"
)
```

4.1.1.2 Valencia



Los datos de contratación pública en el portal de datos abiertos de la comunidad Valenciana cuentan con los siguientes campos y opciones para cada uno de los campos.

Tabla 18: Campos Comunidad Valenciana

COMUNIDAD VALENCIANA		
CAMPO	TIPO DE CAMPO	DESCRIPCIÓN
EJERCICIO (AÑOS)	Numérico	2013 a 2018
NUMERO DE REGISTRO	Numérico	
EXPEDIENTE	Texto	
CLASE CONTRATO	Texto	Menor
		No menor
NUM. LOTES	Numérico	1 al 34
CONSELLERIA/ENTIDAD ADJUDICADORA	Texto	
UNIDAD ADMINISTRATIVA	Texto	
OBJETO (DESCRIPCION)	Texto	
LUGAR DE EJECUCIÓN	Texto	
TIPO	Texto	Administrativo Especial
		Concesión de obras
		Concesión de Servicios
		Gestión de Servicios Públicos
		Obras
		Otros Privados
		Seguros
		Servicios
		Servicios bancarios y de inversiones
		Servicios de creación e interpretación artística o literaria
		Servicios de Espectáculos
TIPO TRAMITACIÓN	Texto	Suministros
		Ordinaria
		Emergencia
ACUERDO/CONTRATO MARCO	Texto	Urgente
PLAZO EJECUCIÓN	Numérico	S / N
UNIDADES PLAZO EJECUCIÓN	Texto	0 al 924
		Meses
		Años
PLAZO MAXIMO CONTRATO INCLUIDAS PRÓRROGAS	Numérico	Días
UNIDADES PLAZO MÁXIMO CONTRATO INCLUIDAS PRÓRROGAS	Numérico	Meses
		Años
		Días
NUMERO DE ANUALIDADES PREVISTAS	Numérico	
FECHA PREVISTA DE FINALIZACIÓN	Fecha	
IMPORTE/VALOR LICITACIÓN SIN IVA	Numérico	
IMPORTE IVA PREVISTO	Numérico	
IMPORTE TOTAL LICITACIÓN	Numérico	
SISTEMA DE PRECIOS	Texto	
FECHA PUBLICACIÓN DE LA LICITACIÓN DOCV	Fecha	
FECHA PUBLICACIÓN DE LA LICITACIÓN BOE	Fecha	

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

FECHA PUBLICACIÓN DE LA LICITACIÓN DOUE	Fecha	
FECHA PUBLICACIÓN DE LA LICITACIÓN PERFIL	Fecha	
PROCEDIMIENTO	Texto	Abierto
		Adjudicación directa
		Adjudicación directa previa consulta a una empresa
		Adjudicación directa previa consulta a varias empresas
		Basado en acuerdo marco celebrado con varios empresarios
		Basado en acuerdo marco concluido con un único empresario
		Negociado
		Restringido
		Según instrucciones
FORMA DE ADJUDICACIÓN	Texto	Subasta Concurso
NUMERO DE EMPRESAS INVITADAS	Numérico	
NÚMERO DE LICITADORES	Numérico	
NACIONALIDAD DEL ADJUDICATARIO	Texto	
FECHA DE ADJUDICACIÓN	Fecha	
FECHA DE FORMALIZACIÓN	Fecha	
NIF/CIF DEL ADJUDICATARIO	Texto	
NOMBRE (O RAZÓN SOCIAL)	Texto	
IMPORTE ADJUDICACIÓN SIN IVA	Numérico	
IMPORTE IVA ESTIMADO	Numérico	
IMPORTE TOTAL ADJUDICACIÓN	Numérico	
TIENE INCIDENCIAS DE VARIACIÓN DE DATOS IDENTIFICATIVOS DEL ADJUDICATARIO (IDE)	Booleano	
FECHA INCIDENCIA IDE	Fecha	
TIENE INCIDENCIAS DE TIENE INCIDENCIAS DE VARIACIÓN DE PLAZOS Y/O IMPORTES (MPI)	Booleano	
FECHA INCIDENCIA MPI	Fecha	
FECHA RECEPCIÓN/CONFORMIDAD (EXTINCIÓN POR CUMPLIMIENTO)	Fecha	
IMPORTE TOTAL FINAL (EXTINCIÓN POR CUMPLIMIENTO)	Numérico	
FECHA DE RESOLUCIÓN (EXTINCIÓN POR RESOLUCIÓN)	Fecha	
CAUSA DE RESOLUCIÓN	Texto	
IMPORTE FINAL DEL CONTRATO (EXTINCIÓN POR RESOLUCIÓN)	Numérico	

Fuente: Elaboración propia

Para la extracción de datos de la comunidad Valenciana no se pudo usar la API de CKAN, pese a que CKAN tiene una gran funcionalidad mediante su API, el CSV que cuenta con la

información de los contratos no se encuentra publicado en el catálogo de datos, se lo debe descargar desde otro sitio web. Por tal motivo se descargó manualmente el archivo CSV y se lo cargo mediante la plataforma en R.

```
datos_offline_Val<-function(){
  tf_val<-read.csv(file = "data/dataset_val.csv",sep = ";",dec = ".",fill = TRUE,flush = TRUE,string
sAsFactors = FALSE,header = TRUE,encoding = "latin1",nrows=10000)#,nrows=10000
  return(tratamiento_val(tf_val%>%select("Ejercicio","Procedimiento","Tipo","Conselleria.Entida
d.adjudicadora","Nombre..o.razón.social.,""Importe.total.adjudicación","Plazo.ejecución","Unidad
es.plazo.ejecución"))
}
```

Como podemos observar los datos de ambas comunidades tienen muchos campos, pero no todos son datos comunes. Por este motivo se tuvo que homogenizar y encontrar una concordancia entre ellos. Tras un análisis exhaustivo y una limpieza de los campos y datos los datos que se consideraron relevantes fueron los siguientes:

Tabla 19: Campos de la plataforma

Campo	Tipo de Campo	Opciones
Ejercicio	Numérico	
Procedimiento	Texto	Abierto
		Negociado
		Restringido
		Según instrucciones
		Basado en acuerdo marco celebrado con varios empresarios
		Basado en acuerdo marco concluido con un único empresario
		Concurso de proyectos
		Contratos de emergencias
		Derivado de acuerdo marco
		Diálogo Competitivo
		Específico de sistema dinámico de contratación
		Menor
		Negociado con publicidad
		Negociado sin publicidad
Tipo	Texto	Obras
		Servicios
		Suministros
		Gestión de Servicios Públicos
		Administrativo Especial

		Colaboración público - privado
		Concesión de obras publicas
		Consultoría y Asistencia
		Servicios bancarios y de inversiones
		Servicios de creación e interpretación artística o literaria
		Servicios de Espectáculos
		Concesión de Servicios
		Seguros
Organismo Contratante	Texto	
Adjudicado	Texto	
Importe	Numérico	
Día	Numérico	
Mes	Numérico	
Año	Numérico	

Fuente: Elaboración propia

4.1.2 Limpieza de datos

Este proceso es indispensable y de los más importantes al momento de tratar información, ya que al momento de analizarla pueden existir inconsistencias y la interpretación de la información puede variar de manera considerable. Los datos proporcionados en muchas ocasiones presentan inconsistencias, ya sea por el formato o el tipo de lenguaje que se esté usando, por ello hay que establecer estándares bajo los cuales se presenta la información. Para este proyecto se han establecido los siguientes estándares:

- Se normalizaron las cabeceras bajo etiquetas comunes para las dos comunidades autónomas y se definió el tipo de dato para cada uno de los campos, las cabeceras definidas fueron las siguientes:
 - Ejercicio: numérico
 - Procedimiento: texto
 - Tipo: texto
 - Organismo Contratante: texto
 - Adjudicado: texto
 - Importe: numérico decimal
 - Fecha: texto

Esta estandarización nos facilita el momento de realizar un filtrado y análisis de los datos.

- Se eliminó o reemplazó los espacios vacíos en todos los campos. Este es un problema común que en el momento de realizar los gráficos genera errores.
- Se estableció el punto como separador decimal y la coma como separador de miles. No siempre se realiza este tipo de especificaciones, pero es mejor tenerlas definidas para evitarnos problemas con la configuración del servidor.
- Los datos en común como tipo de contrato y procedimiento fueron traducidos a un mismo idioma (español) y bajo un mismo nombre, por ejemplo, existían varias denominaciones para un mismo tipo de contrato: serveis, 5. Serveis, servicios, etc., para este caso se usó un mismo nombre “servicios” y así con cada tipo o procedimiento donde se identificó este problema. Si no realizamos esta acción, el sistema lo interpretará como si fueran términos diferentes lo que nos dará como resultado datos inconsistentes.
- Se quitaron caracteres especiales como “*:-” y los acentos en los dos idiomas catalán y español.
- En el campo adjudicado, se identificó que la misma empresa estaba escrita de maneras diferentes. Por ejemplo: empresa S.A, empresa SA, empresa SA., empresaSA, a pesar de ser la misma empresa y que difiere únicamente en uno o dos caracteres en el momento de analizar los datos nos dará una inconsistencia por lo que se creó una función que corrija este tipo de inconsistencias.

La limpieza de datos se realiza de forma automática al cargar los datos a la plataforma, las librerías a usarse fueron Tidyverse y dplyr. Uno de los problemas de la limpieza de datos es que no se puede limpiar la totalidad de ellos. Los datos ingresados por errores humanos (como es el caso de escribir mal el nombre de una empresa) son imposibles de identificar de manera automática ya que no existen patrones para la corrección de este tipo de errores. Sin embargo, se debe limpiar la mayor cantidad de datos en los cuales si se pueda identificar un patrón de fallo.

Otro de los puntos a tener en cuenta es el procesamiento al momento de la limpieza de los datos. Este proceso generalmente conlleva un alto uso de memoria. Esto debido a que se lee toda la información y se realiza un reemplazo masivo en toda la información.

4.2 Plataforma

En nuestro menú principal tenemos algunos datos resumen como son:

- Número de comunidades sincronizadas en el portal.
- Número total de registro de contratos.
- Progreso de carga de los archivos online (no se encuentra implementado).
- Incertidumbre en la limpieza y fiabilidad de los datos.
- Grafico interactivo filtrado por el tipo de contrato teniendo en cuenta los importes de las comunidades autónomas analizadas (Valenciana y Cataluña). Es interactivo



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

por lo que al pasar el ratón sobre los colores nos indicara el nodo hijo y el porcentaje de cada uno de ellos

En el siguiente grafico se muestra la interfaz principal desarrollada en Shiny y que cuenta con un tema personalizado basada en “Shinydashboard”

Ilustración 18: Menú Principal



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

También contamos con una sección de glosario de términos para el usuario que no esté familiarizado con los términos que contiene este proyecto. Como tópicos principales se tiene clases de contratos, tipos de contratos y procedimientos de adjudicación. Cada tópico contiene el significado y los términos que se usaron durante este proyecto.

Ilustración 19: Glosario de términos



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

4.3 Elaboración de gráficos

La elaboración de gráficos se hizo mediante los paquetes de billboarder y dygraphs en R, usando los datos que arroja las consultas de Neo4j y visualizándolos en shiny. Los gráficos que se elaboraron cumplen diferentes problemáticas que se puede plantear el ciudadano o la administración pública. Todos ellos son interactivos para realizar filtrado dependiendo de lo que se desee conocer.

4.3.1 Importación por tipo de contrato

Nuestro primer gráfico es el importe por tipo de contrato. Este gráfico representa la suma del importe de contratación anual filtrado por comunidad. Los componentes del gráfico son los siguientes:

- Eje x: Comunidades
- Eje y: Suma del importe de contratación anual
- Filtrado: Tipo de contrato, Año

El gráfico es interactivo y cuenta con opciones como: el cambio de año, quitar o agregar tipo de contrato con hacer *click* en su leyenda. Además, al pasar el ratón por las barras nos mostrara los datos exactos de cada una de las columnas.

Ilustración 20: Importe por tipo de contrato



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

Con este gráfico podemos identificar de manera fácil el gasto total de cada comunidad y los tipos de contratos en los cuales se ha invertido a lo largo del año seleccionado.

4.3.2 Empresas con más adjudicaciones

Esta gráfica nos muestra la suma del importe de adjudicación a lo largo de los años por empresa. Se encuentra filtrado por comunidades y se tiene la posibilidad de manera dinámica de aumentar o quitar el número de empresas.

Los componentes del grafico son los siguientes:

- Eje x: Empresas
- Eje y: Suma de los importes adjudicados en todos los años
- Filtrado: Comunidad, Número de Empresas

De igual manera que en los gráficos anteriores, al deslizar el ratón sobre una barra nos dará el nombre de la empresa y el importe total a lo largo de los años.

Ilustración 21: Empresa con mayor importe



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

En este gráfico podemos ver claramente las empresas que más monto por contratación han tenido a lo largo de todos los años por comunidad.

4.3.3 Tipos de contrato anual

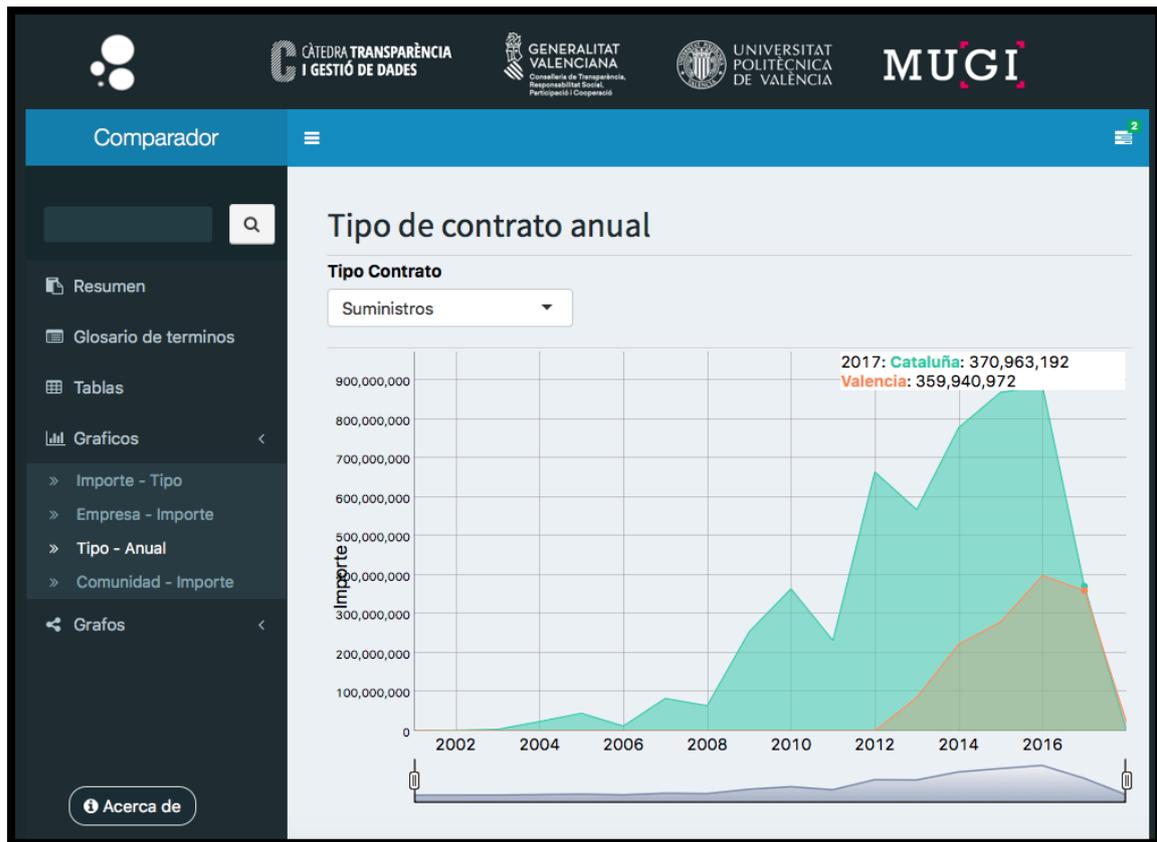
Esta gráfica nos muestra los contratos adjudicados por tipo a lo largo de los años filtrados por comunidad autónoma.

Los componentes del gráfico son los siguientes:

- Eje x: Años
- Eje y: Suma de los importes por tipo
- Filtrado: Comunidad, tipo de contrato

Al ser un gráfico interactivo al deslizar el ratón por cualquier año, nos da el importe adjudicado anual por comunidad, para mayor detalle existe la posibilidad de restringir los años de selección.

Ilustración 22: Tipo de contrato anual



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

La gráfica nos muestra los importes anuales por tipo de contrato de las dos comunidades autónomas, se puede identificar la progresión de los gastos por tipo de contrato.

4.3.4 Monto de adjudicación por comunidad

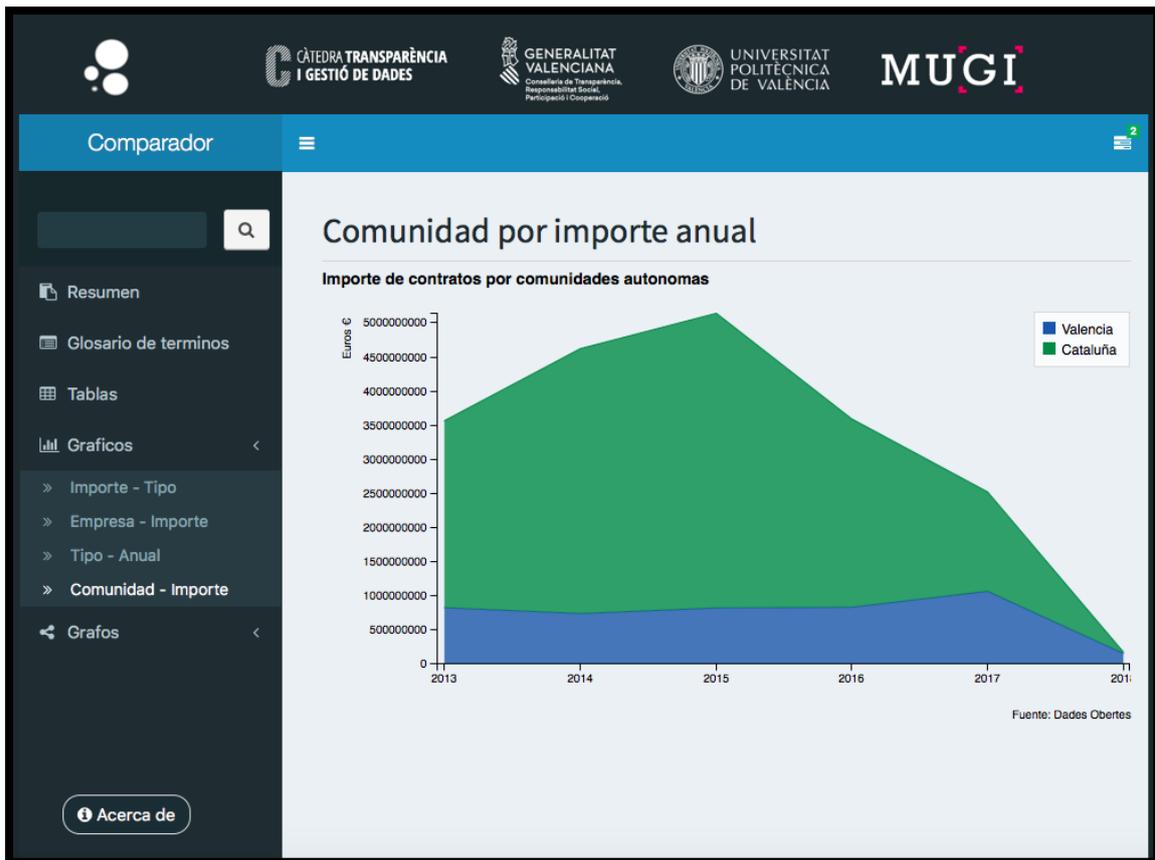
Este gráfico nos muestra en una gráfica comparativa el monto total de contrataciones por comunidad autónoma a lo largo de los años.

Los componentes del grafico son los siguientes:

- Eje x: Años
- Eje y: Suma de los importes totales
- Filtrado: Comunidad

Se puede evidenciar claramente los resultados en una gráfica por áreas.

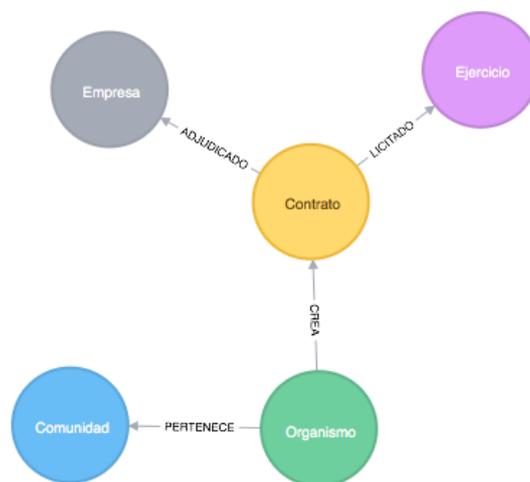
Ilustración 23: Importe anual por comunidad



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

4.4 Elaboración de grafos

Para elaborar los grafos se usó la tecnología Neo4j detallada en el capítulo 3. Lo primero que se hizo es ver la relación que existen entre los datos para así crear los nodos y sus relaciones. Después de acabar con el análisis de las relaciones se concluyó que la mejor estructura para los grafos sería la siguiente:



Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

Los grafos cuentan con nodos (empresas, ejercicio, contrato comunidad y organismo) y relaciones (adjudicado, licitado, crea, pertenece), cada nodo puede contar con propiedades en nuestro caso son las siguientes:

- Contrato: tipo, procedimiento, duración e importe
- Ejercicio: año
- Organismo: nombre
- Comunidad: nombre
- Empresa: adjudicado

Las relaciones al igual que los nodos pueden contener propiedades. Generalmente se usa estas propiedades para dar un peso a la relación, en nuestro diagrama tenemos la relación entre contrato y empresa, esta relación tiene la propiedad importe y es la que determina el peso de la relación.

4.4.1 Conexión y carga en R

En nuestra plataforma en R debemos cargar la librería RNeo4j para poder conectarnos a nuestro Neo4j, para nuestro caso se instaló Neo4j en docker en el puerto 7494.

Una vez establecida la conexión procedemos a crear la BDD de grafos con los datos almacenados de las comunidades autónomas, esta acción la realizamos con Cypher (lenguaje que se usa en Neo4j para realizar consultas). Aquí se crean los nodos y relaciones con sus respectivas propiedades, después se genera un bucle en el cual se carga la información contenida en los datasets de las comunidades autónomas, la estructura debe seguir la de los nodos y propiedades. Este proceso suele demorar bastante tiempo básicamente porque existen gran cantidad de datos.

4.4.2 Visualización de grafos

Para la visualización de grafos en R se necesitó otro paquete llamado visNetwork. Este paquete nos ayuda en la representación de los grafos en R. La parte clave en la graficación de los grafos en Neo4j son las consultas a realizarse, las consultas determinaran los resultados a presentarse. En nuestro primer grafo tenemos las siguientes consultas:

```
MATCH (o)-[:CREA]->(c)-[ad:ADJUDICADO]->(e) RETURN o.comunidad as from,e.adjudica  
do as to, ad.importe as value order by value desc limit {limit}
```

```
MATCH (o)-[:CREA]->(c)-[ad:ADJUDICADO]->(e) where o.comunidad={comunidad} RETU  
RN o.comunidad as from,e.adjudicado as to, ad.importe as value order by value desc limit {limit}
```

Como podemos apreciar en la consulta se puede evidenciar la relación que existe entre organismo, contrato y comunidad donde estamos consultando las empresas con los mayores importes por comunidad autónoma. Una vez tenemos los datos de la consulta realizada tenemos que presentarlo gráficamente por lo que se usó visNetwork.

En visNetwork se ingresan los datos de los nodos, relaciones y propiedades respectivamente para obtener una gráfica similar a la que se obtendría nativamente en Neo4j. Si analizamos el



proceso de graficación utilizado podemos observar que al realizar una consulta a Neo4j éste nos devuelve una tabla con la información (nodos, relaciones y propiedades) en contra de lo que se podría llegar a pensar, que nos devuelve un objeto tipo gráfico.

4.4.3 Contratos por comunidad Autónoma

Como podemos evidenciar en la gráfica tenemos la posibilidad de controlar ciertos parámetros como son la comunidad y número de nodos a presentarse. Esta capacidad de poder interactuar de manera dinámica con el gráfico se debe a la potencia que tiene Shiny con sus funciones reactivas (las mismas que al modificar un componente carga automáticamente la información).

Ilustración 24: Empresas por comunidad



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

4.4.4 Información de contratos por comunidad autónoma

Los grafos se los puede presentar de varias maneras. Por este motivo en este proyecto se abarcó otro tipo de presentación a parte de la convencional, para ello se usó otro tipo de paquetes. El uso de Maptree, nos permite mapear de manera sencilla la relación entre los datos de forma jerárquica, visual e interactiva de cara al usuario.

En el siguiente gráfico se ha estructurado la información por niveles:

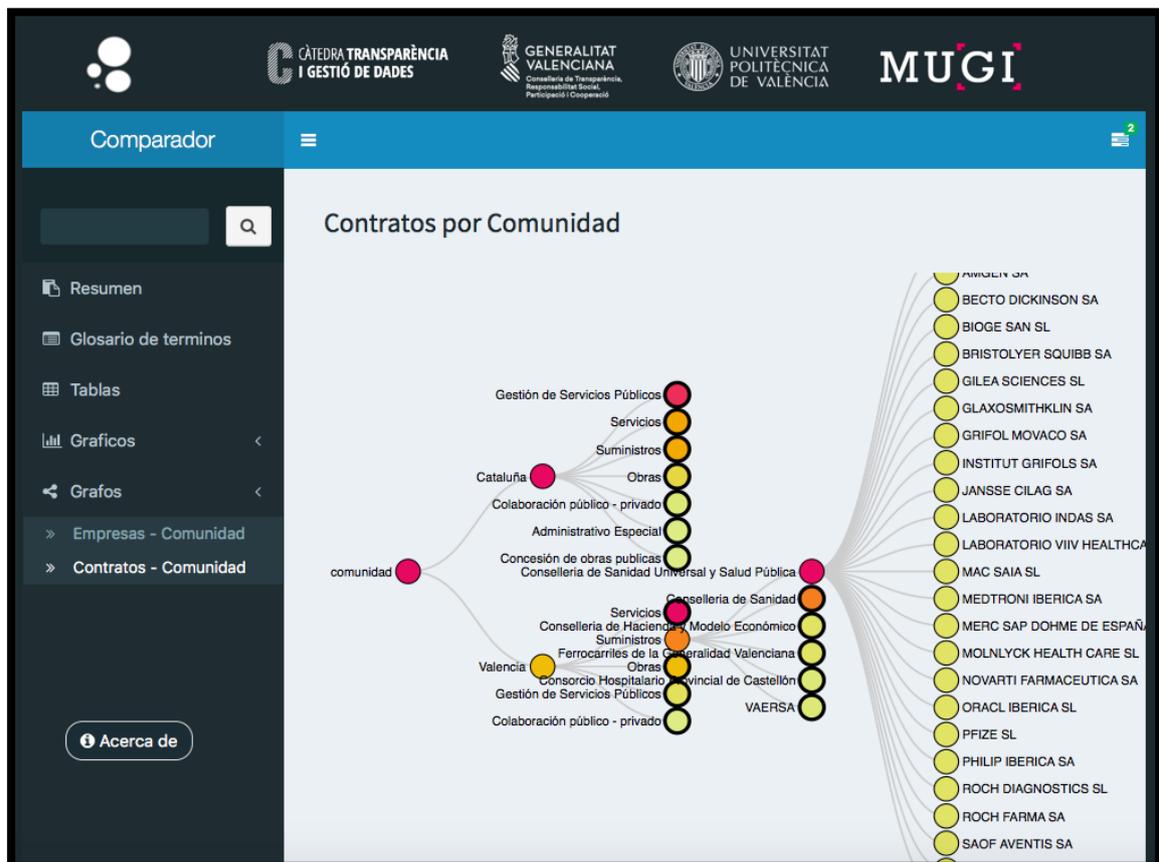
Nivel 1. Comunidades autónomas

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

- Nivel 2. Tipos de contrato
- Nivel 3. Organización adjudicadora
- Nivel 4. Empresa

Todos los niveles cuentan con colores, estos colores representan la cantidad de datos, siendo el color más rojizo el que más datos tiene y los cercanos al blanco los que menos datos poseen. Por ejemplo, podemos ver que Cataluña respecto a la comunidad Valenciana tiene mayor cantidad de contratos y así con cada nivel.

Ilustración 25: Contratos por comunidad



Elaboración propia: <http://madphy.mat.upv.es:3838/Comparador/>

Únicamente se mostrarán los 1000 últimos contratos del más actual al menos actual. Ésto debido a que es imposible representar en este tipo de gráfica un número mayor al estipulado por el proyecto.

Capítulo 5 : Implementación e implantación

Este capítulo está enfocado a la publicación de la plataforma en la web, se mostrará la arquitectura que tiene la plataforma en los servidores y la manera en que las diferentes tecnologías interactúan entre sí. Además, se comentará las complicaciones limitaciones y restricciones que se tuvo en el momento de publicar la plataforma en la web.

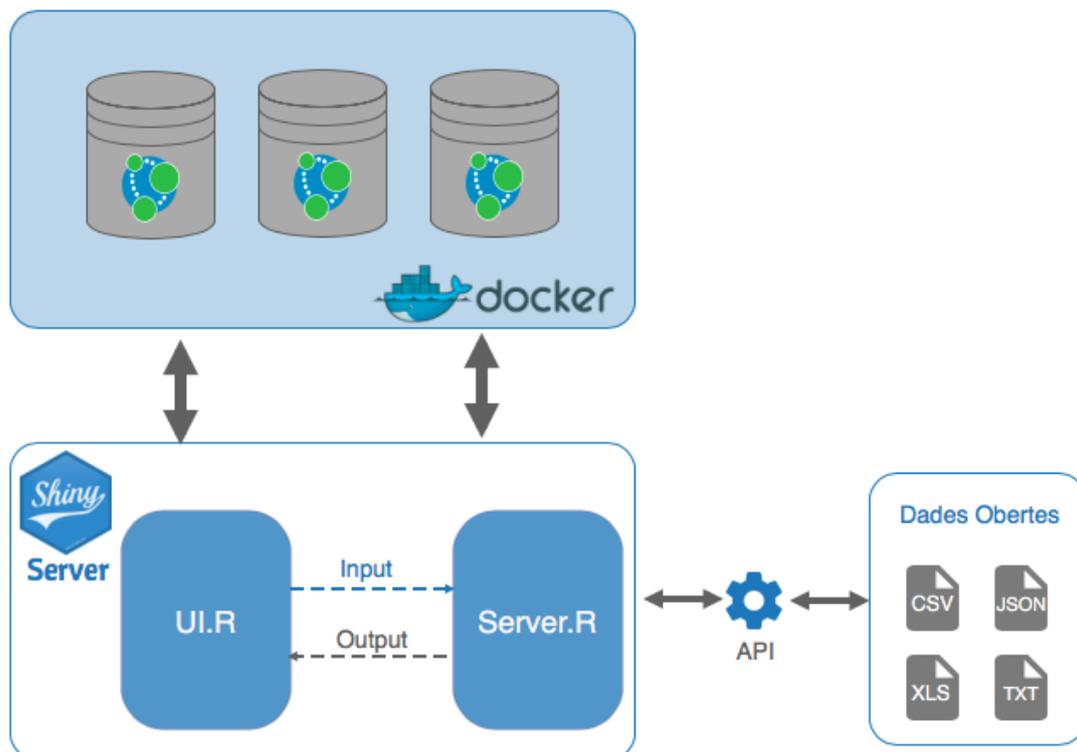
5.1 Publicación de la plataforma

La plataforma estará publicada en un servidor Linux del grupo MADPHY perteneciente al Departamento de Matemática Aplicada de la Universitat Politècnica de València en la dirección <http://madphy.mat.upv.es:3838/Comparador/>.

5.1.1 Arquitectura

La plataforma cuenta con la arquitectura tecnológica, dispuesta en tres grandes componentes BDOG, Servidor de Aplicaciones y componentes externos, se muestra de manera gráfica en la siguiente figura:

Ilustración 26: Arquitectura docker



Fuente: Elaboración propia

Como podemos observar contamos con tres componentes clave:

Plataforma de visualización y comparador dinámico de contrataciones. El caso de Generalitat Valenciana y Generalitat Cataluña.

1. La base de datos de grafos (BDOG) se encuentra en docker bajo diferentes contenedores, lo que nos permite tener diferentes BDOG en un mismo entorno sin cambiar de servidor. Se deberá realizar configuraciones sobre todo en los puertos de conexión con las diferentes BDOG.
2. Shiny server: sería el equivalente a un servidor de aplicaciones en una arquitectura tradicional. Shiny server tendrá interacción tanto con nuestra BDOG como con la API de extracción de datos. Será la interfaz de interacción de cara a nuestro usuario final.
3. API de los diferentes portales de datos abiertos para la extracción o consulta de datos en diferentes formatos. En nuestro proyecto únicamente usamos los datos en CSV, básicamente por la facilidad que representa su manipulación.

5.1.2 Docker

El primer paso es instalar docker en nuestro servidor y configurar el número de cores a asignar, la memoria RAM y la locación donde se encontrará la imagen de docker. La ubicación de nuestra imagen docker dentro de nuestro servidor es importante, la necesitaremos si deseamos copiar la imagen del docker en otro servidor y realizar replicas.

Se crearon dos contenedores en docker: uno con Neo4j y el otro con Shiny Server. Las dos imágenes se encuentran disponibles (gratuitamente) en la tienda de docker. Hay que tener en cuenta la configuración de los puertos para ambos casos, ya que estos dos componentes interactuarán entre sí. La instalación de los contenedores es muy sencilla y se las realiza por terminal.

5.1.3 Neo4j

Neo4j para docker se encuentra oficialmente en la tienda de docker, únicamente se debe correr el código donde se especifica la versión de Neo4j que se desea ejecutar, el puerto donde se ejecutara y la ruta donde estarán las carpetas. Por defecto se corren sobre los puertos 7474 para http, 7473 para HTTPS, 7687 para Bolt, hay que tener especial cuidado en los puertos ya que si se tiene más de una aplicación se deben cambiar los puertos, tanto en docker como en nuestro código en shiny.

Para nuestro caso se creó un ejecutable que al momento que se inicie el servidor este corra automáticamente los comandos para levantar nuestro docker con Neo4j y Shiny server.

```
docker run
--publish=7474:7474 --publish=7687:7687
--volume=$HOME/neo4j/data:/data
--volume=$HOME/neo4j/config:/conf
--volume=$HOME/neo4j/plugins:/plugins
--volume=$HOME/neo4j/import:/import
--volume=$HOME/neo4j/run:/run
--volume=$HOME/neo4j/certificates:/certificates
--volume=$HOME/neo4j/logs:/logs
--env NEO4J_dbms_memory_heap_maxSize=2g
neo4j
```

5.1.4 Shiny-Server

Shiny server corre únicamente en versiones Linux, por lo tanto, nuestro servidor es una distribución ligera de Linux. En el servidor se tiene que compilar los paquetes y el código desarrollado en R, el comando de instalación “*sudo R install.packages("paquete")*”, normalmente es muy fácil realizar el despliegue en el servidor. Los problemas suelen darse principalmente al no tener instalados las librerías utilizadas en el desarrollo. Para ello debemos cerciorarnos que todas las librerías estén instaladas antes de copiar los archivos.

Shiny Server aloja todos nuestros desarrollos R-Shiny, así como los complementos usados (archivos HTML, archivos CSS, etc.) usando la configuración de `site_dir` de forma predeterminada, Shiny Server usará un `site_dir` para hacer que las aplicaciones y los activos estén almacenados y disponibles en `/srv/shiny-server/`. Es posible poner más de una aplicación en este directorio y luego hacer referencia a ellas en su servidor. Por ejemplo, una aplicación Shiny almacenada en `/srv/shiny-server/comparador`, donde `comparador` sería el nombre de nuestra aplicación, estaría disponible en <http://madphy.mat.upv.es:3838/Comparador/>, (donde `madphy.mat.upv.es` es el nombre del servidor). También se puede colocar archivos HTML en este directorio para que estén disponibles el servidor.

Se puede configurar roles y usuarios que puedan acceder a nuestra aplicación, para nuestro caso no se tienen ninguna restricción de este tipo, ya que se necesitaría el uso de Shiny Pro para extender la funcionalidad de nuestro sistema.

Shiny server al igual que Neo4j es ejecutado automáticamente al iniciar o reiniciar el servidor, esto en caso de cualquier fallo o reinicio en el servidor.

5.2 Limitaciones

Las limitaciones como en todo proyecto en el que se maneja gran cantidad de datos es básicamente el procesamiento de la información. Los servidores deben ser lo suficientemente robustos para soportan el procesamiento, sobre todo la limpieza de datos y la visualización de gráficas que son los procesos que más recursos consumen.

Otra limitación que se tiene al interactuar con los portales de datos abiertos es la disponibilidad de la información y la manera en la que se encuentra publicada. Muchas veces se tiene la información publicada pero imposible de ser analizada por su falta de estandarización e incompatibilidad con otros portales.

La actualización automática de la información puede representar un problema de cara a la limpieza de los datos, si se cambia el formato del archivo o se aumenta o disminuye campo, seguramente la plataforma presentará errores.



Conclusiones

La utilización de una metodología orientada al análisis de datos permitió tener mejores resultados de los esperados, se pudo cargar la información en su totalidad y además realizar la actualización automática de los datos que se publiquen recientemente. Las iteraciones que la metodología recomienda ayudaron en gran medida a suplir todas las necesidades de cara al usuario final, permitiendo así cumplir con el objetivo principal “realizar una plataforma web que de forma automática y gráfica represente los datos existentes en los portales de datos abiertos tanto de la comunidad Valenciana como de Cataluña realizando comparaciones entre estas dos comunidades”.

Si bien se pudo realizar la conexión automática con los dos portales de las comunidades autónomas sólo en el caso de Cataluña se pudo obtener la información de los contratos de manera automática, en el caso de la Comunidad Valenciana específicamente el dataset de contratación no estaba disponible a través de la API por lo que no se pudo realizar la actualización automática para esta comunidad.

Al analizar los datos de los diferentes portales se pudo evidenciar la falta de estandarización que tienen los datasets. Si bien todos deben regirse a las leyes de transparencia de datos y contratación pública (entre otras), en dichas leyes no se especifica claramente el formato de los campos, ni el tipo de datos y mucho menos el texto permitido. Es por esto que la estandarización y homologación de campos resulta difícil (lo que podría producir alguna distorsión en la comparativa).

Al realizar el programa se tornó casi imposible encontrar un método automático que abarque todos los portales de las diferentes comunidades autónomas, esto debido a los siguientes problemas: falta de estandarización del acceso a los datos, varía su tecnología (CKAN, DKAN, Socrata, etc), varía la manera de buscar la información y sus campos, por lo que prácticamente sería realizar un programa particular para cada comunidad autónoma, dependiendo de la personalización que han realizado en cada uno de los portales.

La limpieza de los datos fue una de las etapas más problemáticas del proyecto debido a que los dataset de una misma comunidad no mantenía una concordancia entre sus datos esto por diferentes motivos: los tipos de formatos diferentes, errores en códigos, errores humanos al ingresar datos etc. Todo esto influyó en que el proceso fuese largo y consuma una gran cantidad de recursos en cuanto a procesamiento. En el proceso de limpieza no se pudo asegurar una limpieza del 100% debido a que existen errores que no se pueden detectar de forma automática, como es el caso del mal ingreso de un campo por una escritura errónea.

Las comparativas de los gráficos realizados nos muestra que de manera general Cataluña gasta mucho más en contratos públicos de manera general que la comunidad Valenciana, las comparaciones en algunos casos solo se pudieron hacer de los últimos años ya que la comunidad Valenciana solo cuenta con datos posteriores a 2012. En ciertos gráficos se puede tener ambigüedades por el cambio de las leyes de contratación pública a lo largo de los años.

El uso de base de datos orientada a grafos de Neo4j permitió que el proyecto tuviera la versatilidad desde un punto de vista de manejo de los datos, ganando velocidad de

procesamiento al realizar consultas y manteniendo la relación entre los datos de manera automática. Sin embargo, la carga de la información llevó mucho tiempo.

Bibliografía

Collier, Ken. (2012). Agile analytics: A value-driven approach to business intelligence and data warehousing.

Fernández Alarcón, V. (2006). Desarrollo De Sistemas De Información: Una Metodología Basada En El Modelado.

Wickham, H., & Grolemund, G. (2017). R for data science: Import, tidy, transform, visualize, and model data.

Ian Robinson, Jim Webber, and Emil Eifrem. (2015). Graph Databases. United States of America: O'Reilly Media.

Criado, J., & Ruvalcaba Gómez, E. (2016). ¿Qué es y qué se entiende por gobierno abierto? (1st ed.). NovaGob.

Observatorio de Sociedad, Gobierno y Tecnologías de Información. Guía de buenas prácticas en gobierno abierto. Bogotá: Universidad Externado de Colombia.

Neo4j Team. (2017). The Neo4j Developer Manual v3.2. ND: Neo Technology.

AMETIC. Recomendaciones de buenas prácticas en la contratación pública de consultoría y servicios TI, pp. 30.

Gobierno de España, 2013. Boletín oficial del estado, pp. 11370-11421. ISSN 1133-2654. DOI BOE-A-2012-5403.

Gobierno de España, 2013. Boletín Oficial del Estado, pp. 97922-97952. ISSN 0212-033X DOI BOE-A-2013-12887

Anderson, C. (2015). Docker [software engineering]. *IEEE Software*, 32(3), 102-c3.

Arellano, P. (2018). RACIONALIZACIÓN TÉCNICA DE LA CONTRATACIÓN.

Recursos electrónicos

Guía de aplicación del Real Decreto 1495/2011, Retrieved from <https://datos.gob.es/es/documentacion/guia-de-aplicacion-del-real-decreto-14952011>

Ley 2/2015, de 2 de abril, de Transparencia, Buen Gobierno y Participación Ciudadana de la Comunidad Valenciana. (2015). Retrieved from <https://boe.es/buscar/pdf/2015/BOE-A-2015-4547-consolidado.pdf>

Norma técnica de interoperabilidad de reutilización de recursos de información. (2013). Retrieved from <http://www.boe.es/boe/dias/2013/03/04/pdfs/BOE-A-2013-2380.pdf>

Informe de reutilización de datos (2016).
Retrieve from: http://informecotec.es/media/INFORME_REUTILIZACION-DE-DATOS.pdf

User guide — CKAN 2.9.0a documentation. Docs.ckan.org. Retrieved from <http://docs.ckan.org/en/latest/user-guide.html>

Version 3 — CKAN (Comprehensive Knowledge Archive Network) v1.5.1 Administration Guide. Docs.ckan.org, from <http://docs.ckan.org/en/ckan-1.5.1/apiv3.html>

Conselleria de Transparencia, Responsabilidad social, Participación y Cooperación - Generalitat Valenciana. (2018). Transparencia.gva.es. Retrieved 16 March 2018, from <http://www.transparencia.gva.es/inicio>

Datos Abiertos - Generalitat Valenciana. Dadesobertes.gva.es. Retrieved 10 April 2018, from <http://www.dadesobertes.gva.es/es/about>

Catalunya, P. (2018). Socrata. Transparència Catalunya. Retrieved 10 April 2018, from <https://analisi.transparenciacatalunya.cat>

RStudio. RStudio. Retrieved 6 February 2018, from <https://www.rstudio.com/products/RStudio/>

Guía de catalogación de datos abiertos. (2014) (1st ed.). Retrieved from http://opendata.aragon.es/public/documentos/Guia_Gestion_v1.0.pdf

Opengovdata.org. (n.d.). The 8 Principles of Open Government Data (OpenGovData.org). [online] Available at: <https://opengovdata.org>.

Datos.gob.es. (2018). Guía para publicar datos abiertos de manera rápida y sencilla (con CKAN) | datos.gob.es. [online] Available at: <http://datos.gob.es/es/documentacion/guia-para-publicar-datos-abiertos-de-manera-rapida-y-sencilla-con-ckan>.

Elconsultor.laley.es. (2018). Cuadro de las novedades más importantes que incorpora la Ley 9/2017, de 8 de noviembre, de contratos del sector público. [online] Available at: <http://elconsultor.laley.es/Content/Documento.aspx?params=H4sIAAAAEAMtMSbF1>

[CTEAAiMzE1Mjc7Wy1KLizPw8WyMDQwsDYwNDkEBmWqVLfnJIZUGqbVpiTnEqAHrL_K81AAAAWKE](https://doi.org/10.1016/j.ijl.2018.08.001)

Jurídicas, N. (2018). La nueva Ley de Contratos del Sector Público impone la obligatoriedad de la contratación electrónica · Noticias Jurídicas. [online] Noticias Jurídicas. Available at: <http://noticias.juridicas.com/actualidad/noticias/12775-la-nueva-ley-de-contratos-del-sector-publico-impone-la-obligatoriedad-de-la-contratacion-electronica/>

Lawandtrends.com. (2018). Tabla Práctica para entender la Nueva Ley de Contratos del Sector Público | Administrativo | LawAndTrends. [online] Available at: <http://www.lawandtrends.com/noticias/administrativo/tabla-practica-para-entender-la-nueva-ley-de-contratos-del-sector-publico-1.html>

P, V. (2018). Rpubs - billboarder examples. [online] Rpubs.com. Available at: http://rpubs.com/Victorp/billboarder_examples.

studio.github.io. (2018). Using in Shiny Applications. [online] Available at: <https://rstudio.github.io/dygraphs/shiny.html>

