



Universidad de Valladolid

Escuela de Ingeniería Informática

Grado en Ingeniería Informática

Mención en Tecnologías de la Información

TRABAJO DE FIN DE GRADO

SemCyLRAT: Registro de Actividades de
Tratamiento de Datos Personales en CyL con
información semántica

ALUMNO:

Crhistian de la Puerta Verdejo

TUTORA:

Dra. M^a Mercedes Martínez González

Agradecimientos

Este proyecto va dedicado a mis padres por el apoyo en toda mi educación. A mis tíos y primos por estar ahí siempre que les necesité.

A Mamahos por ser el grupo más divertido que jamás he conocido por mucho tiempo que pase, a Galácticos por saber menos de fútbol que de informática. Y por esas barbacoas en Alaejos tan especiales como necesarias.

A todos ellos,

Muchas gracias.

Resumen

Los RAT (*Registros de Actividades de Tratamiento*) son los registros de las actividades donde están involucradas datos personales que las instituciones y empresas deben hacer. Es obligatorio que estos registros sean públicos, y se suelen publicar en la web en formato de hoja de cálculo.

En este TFG se realizará la transformación de estos registros a un modelo de datos semántico y su publicación como datos abiertos enlazados, aplicado a RAT de Castilla y León.

Palabras clave

RAT, RDF, URI, SPARQL, grafo, tripleta, ontología, OWL, Dublin Core, SemRAT, SemCyLRAT, prefijo, Web semántica, sujeto, predicado, objeto, literal, OpenRefine, AllegroGraph.

Abstract

The RAT (*Records of Treatment Activities*) are the records of the activities where personal data is involved that institutions and companies must do. These records are required to be public, and are typically published on the web in spreadsheet format.

In this TFG, the transformation of these records to a semantic data model and their publication as linked open data will be carried out, applied to RAT of Castilla y León.

Keywords

RAT, RDF, URI, SPARQL, Graph, Triplet, Ontology, OWL, Dublin Core, SemRAT, SemCyLRAT, Prefix, Semantic Web, Subject, Predicate, Object, Literal, OpenRefine, AllegroGraph.

ÍNDICE

Agradecimientos	3
Resumen	5
1. Introducción.....	11
1.1 Motivación	11
1.2 Objetivos	11
1.3 Entorno	12
1.4 Estructura Memoria.....	12
2. Planificación	13
2.1 Alcance, propósitos y objetivos	13
2.2 Plan de Proceso	13
2.3 Planificación del proyecto.....	15
2.3.1 Identificación de tareas concretas	15
2.3.2 Calendario de tareas: Diagrama de Gantt	17
3. Tecnologías utilizadas	18
3.1 Web Semántica.....	18
3.2 RDF	19
3.3 Ontologías usadas.....	23
3.3.1 OWL	23
3.3.2 Dublin Core.....	24
3.3.3 Ontologías propias	27
3.4 SPARQL.....	30
3.5 Los Registros de las Actividades de Tratamiento (RAT) original y transformado.....	33
3.6 Openrefine.....	39
3.7 AllegroGraph.....	41
3.7.1 Instalación del entorno.....	42
4. Desarrollo	46
4.1 Pipeline.....	46
4.2 OpenRefine.....	47
4.3 Comprobación funcionamiento	57
4.4 Generación de grafos.....	61
5. Conclusiones.....	66
5.1 Objetivos Alcanzados.....	67
5.2 Proyecto futuro	67
5.2.1 Cloud Computing.....	68
5.2.2 Creación del dominio de las ontologías.....	68

6.	Bibliografía.....	70
7.	ANEXO I: Instalación OpenRefine.....	72
8.	ANEXO II: Instalación AllegroGraph.....	78
9.	ANEXO III: Instalación VirtualMachine	98
10.	ANEXO IV: Tripletas y Grafos RAT	111
10.1	Servicios Sociales	111

ÍNDICE DE FIGURAS

Figura 1. Fases PMBOK.....	13
Figura 2. Acciones en cada fase.....	14
Figura 3. Método seguido	15
Figura 4. Grafo RDF simple	22
Figura 5. Grafo RDF compuesto	22
Figura 6. Ejemplo de consulta en SPARQL	31
Figura 7. Visualización de la Base de Datos	31
Figura 8. Consulta cualquier artista holandés.....	31
Figura 9. Visualización de la Base de Datos	32
Figura 10. Diseño grafo general transformado a RDF	35
Figura 11. Ejemplo de la primera tripleta.....	36
Figura 12. Ejemplo de la segunda tripleta	37
Figura 13. Pipeline.....	46

ÍNDICE DE IMÁGENES

Imagen 1. Arquitectura de la Web Semántica	19
Imagen 2. Grafo RDF diseñado por Eric Miller	20
Imagen 3. RDF/XML de Eric Miller	21
Imagen 4. Interfaz de OpenRefine.....	39
Imagen 5. Extensión RDF en OpenRefine	40
Imagen 6. Creación del proyecto en OpenRefine.....	40
Imagen 7. Instalación de AllegroGraph 7.1.0 Virtual Machine	43
Imagen 8. Escritorio Ubuntu 20.04.....	44
Imagen 9. Consola de comandos de Shell	45
Imagen 10. Interfaz AllegroGraph.....	45
Imagen 11. Creación del proyecto	47
Imagen 12. Añadir una conciliación basada en virtuoso	48
Imagen 13. Servicios de conciliación añadidos	49
Imagen 14. Conciliación del recurso Conserjería.....	50
Imagen 15. Definición del servicio de conciliación	50
Imagen 16. Recurso tratamiento	51
Imagen 17. URI y prefijos necesarios.....	51
Imagen 18. Lista de prefijos para el proyecto.....	52
Imagen 19. Nodo raíz	53
Imagen 20. Añadir propiedad	53
Imagen 21. Añadir el recurso tratamiento	54
Imagen 22. Esqueleto RDF	55
Imagen 23. Código RDF Presidencia	55
Imagen 24. RAT presidencia exportado como XML	56
Imagen 25. RAT presidencia exportado como TTL.....	57
Imagen 26. Validación archivo Presidencia.ttl.....	58
Imagen 27. Interfaz W3C Validator	59
Imagen 28. Validación y tripletas del documento Presidencia.xml.....	60
Imagen 29. Parte del grafo Presidencia.xml	60
Imagen 30. Importación del archivo Presidencia.xml	61
Imagen 31. Interfaz AllegroGraph.....	62
Imagen 32. Generación de tripletas	63
Imagen 33. Generar el grafo Presidencia en AllegroGraph.....	63

Imagen 34. Muestra de la leyenda del Grafo Presidencia.....	64
Imagen 35. Selección de todos los predicados del RAT Presidencia	64
Imagen 36. Grafo Presidencia completo y organizado	65
Imagen 37. Inicio sesión AllegroGraph.....	84
Imagen 38. Seleccionar archivos	85
Imagen 39. Crear repositorio	86
Imagen 40. Información del repositorio	87
Imagen 41. Importar archivos.....	88
Imagen 42. kennedy.ntriples.....	89
Imagen 43. copia de seguridad	90
Imagen 44. Consulta AllegroGraph.....	92
Imagen 45. Ejemplo 1 de consulta.....	93
Imagen 46. Ejemplo 2 de consulta.....	94
Imagen 47. Resultado de la consulta	95
Imagen 48. Información de la consulta.....	96
Imagen 49. Warnings de la consulta.....	97
Imagen 50. Interfaz VMware VirtualMachine	99
Imagen 51. Archivo AllegroGraph.....	99
Imagen 52. Inicio sesión Ubuntu	100
Imagen 53. Escritorio Ubuntu.....	101
Imagen 54. Consola Ubuntu	102
Imagen 55. Iniciando AllegroGraph.....	103
Imagen 56. Parando AllegroGraph.....	104
Imagen 57. Inicio AllegroGraph.....	105
Imagen 58. Usuario y Password	105
Imagen 59. Archivo de configuración AllegroGraph.....	106
Imagen 60. Start Gruff.....	107
Imagen 61. VirtualMachine Settings	108
Imagen 62. Configuración Máquina Virtual.....	109
Imagen 63. Problemas de compatibilidad.....	110
Imagen 64. Grafo Servicios Sociales.....	111
Imagen 65. Grafo AllegroGraph Servicios Sociales.....	112

ÍNDICE DE TABLAS

Tabla 1. Cronograma	17
Tabla 2. Resultado de la consulta	32
Tabla 3. Excel de ejemplo RAT Economía y Hacienda	34
Tabla 4. Primeras tripletas de RAT economía y hacienda.....	36
Tabla 5. Ejemplo de tripletas de RDF complejo RAT economía y hacienda.....	38
Tabla 6. Tripletas Servicios Sociales.....	138

1. Introducción

1.1 Motivación

Una de las asignaturas que más me ha marcado y más me ha gustado realizar es aquella enfocada a las bases de datos, esto es debido a que en la Informática actual hay ingentes cantidades de datos, muchos de ellos de gran relevancia, que debe ser almacenada para que en los momentos que sea necesario se consulte la información.

Esta consulta puede ir desde un alumno que quiere consultar los libros de la biblioteca para realizar una reserva, hasta un banco que necesita tener una base de datos segura y accesible con todos los clientes sus cuentas, prestamos, deudas, descuentos y beneficios que deben de ser actualizables, optimizables y accesibles en cualquier momento tanto por los usuarios como por la entidad si así lo necesitaran.

En vista de la importancia de una base de datos, quise centrar mi trabajo en intentar mejorar el sistema de almacenamiento que usa la web oficial de la Junta de Castilla y León.

La web almacena varios **RAT** *Registros de Actividades de Tratamiento de Datos Personales*:

- Presidencia
- Transparencia
- Economía y hacienda
- Empleo e industria
- Fomento y Medio Ambiente
- Agricultura y ganadería
- Sanidad
- Familia e igualdad de oportunidades
- Educación
- Cultura y turismo

1.2 Objetivos

El objetivo principal es mejorar el sistema de almacenamiento de datos de la junta de Castilla y León.

- Convertir los Registros de Actividades de SQL (*Structured Query Language*) a RDF (*Resource Description Framework*) lo que convertiría los datos en semánticos mejorando así la integridad de los datos y en su interoperabilidad.
- Crear una nube basada en RDF, que destaque por la facilidad para habilitar el procesamiento automatizado de los recursos Web para almacenar los archivos convertidos.

- Realizar consultas a los archivos para comprobar la eficiencia de RDF debido a la conversión de tablas a tripletas donde tenemos: sujeto, predicado y objeto.
- Proporcionar documentación que describe la estructura de la Base de Datos, sus procedimientos y funciones ofrecidas.
- La Base de Datos deberá cumplir con los requerimientos identificados siguientes:
 - R1: Almacenamiento.

La Base de Datos deberá funcionar como un sistema de almacenamiento empleado por la Junta de Castilla y León para el almacenamiento y gestión de los archivos denominados *Registros de Actividades de Tratamiento de Datos Personales*.
 - R2: Consulta.

Cualquier usuario que acceda a la nube podrá hacer las consultas necesarias a la Base de Datos, sobre alguna tripleta u objeto específico, o consultar el grafo de alguno de los archivos.
 - R3: Gestión de la Base de Datos.

Registro y modificación de datos sobre los *Registros de Actividades de Tratamiento de Datos Personales*.
 - R4: Escalabilidad.

Capacidad de la Base de Datos para adaptarse a los cambios.

1.3 Entorno

El entorno de este proyecto está formado por todos los *Registros de Actividades de Tratamiento de Datos Personales de la Junta de Castilla y León*, puesto que todos ellos comparten el mismo formato, y la intención es realizar una transformación de todos esos ficheros a RDF para su posterior almacenamiento y consulta en Nube.

1.4 Estructura Memoria

Este proyecto se va a diferenciar en varias ramas. Para cada una de estas ramas se va a dedicar una explicación en el orden indicado a continuación.

- *Capítulo 1: Introducción.*

En primer lugar se realiza una breve introducción del trabajo realizado, donde se detallan las motivaciones y objetivos del proyecto.

- *Capítulo 2: Planificación.*

Se hablará sobre la planificación del proyecto y de la metodología utilizada para su realización.

- *Capítulo 3: Tecnologías usadas.*

Se describirán las tecnologías utilizadas en el proyecto. Qué es una web semántica y su relación con RDF, para qué sirven las ontologías y las razones para las que se usan en este proyecto OWL, Dublin Core, SemRAT y SemCyLRAT. Y por último los entornos de trabajo usados OpenRefine y AllegroGraph

- **Capítulo 4: Desarrollo.**

Indicar cual ha sido la metodología de trabajo que se ha seguido en el desarrollo del proyecto, explicando cada una de las etapas que se han realizado.

- **Capítulo 5: Conclusiones y proyecto futuro.**

Se pretende establecer una serie de conclusiones a las que llegaremos una vez realizado el proyecto y aportar algunas ideas que sirvan para posteriores trabajos que ayuden a mejorar la base de datos.

El documento que estamos realizando también posee una serie de secciones que se comentaran después de las principales, como son:

- *Bibliografía*

Contiene los enlaces a las fuentes usadas para el desarrollo de este proyecto.

- *Anexos I – III*

Se incluyen las guías de instalación de las distintas herramientas para la comprensión y utilización de las mismas.

- *Anexos IV*

Ejemplo de grafo y tripletas resultantes de la transformación de los archivos a RDF.

2. Planificación

2.1 Alcance, propósitos y objetivos

El objetivo de este proyecto es realizar una mejora en el almacenamiento y consulta de los Registros de Actividades de Tratamiento de Datos Personales de la Junta de Castilla y León, reubicándolos todos ellos en una nube accesible y consultable.

2.2 Plan de Proceso

Para la elaboración del proyecto se ha seguido según la definición PMBOK:

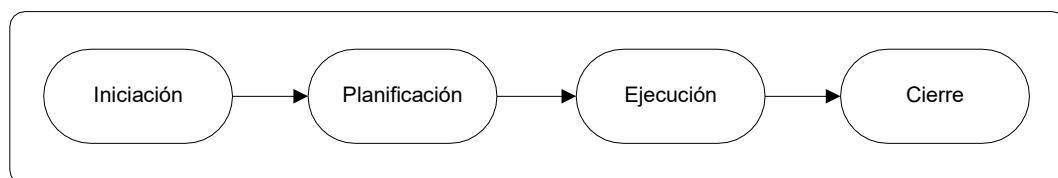


Figura 1. Fases PMBOK

Una vez dentro de cada fase, se definen las siguientes acciones para su realización:

A continuación, se describe las diferentes fases y acciones:

- **Iniciación:** Definición de los objetivos y el alcance inicial (véase el apartado *Alcance, propósitos y objetivos*).
- **Planificación:** Identificación de tareas, estimación de esfuerzo para cada tarea y cronograma.
- **Análisis:** Definición del Alcance Detallado, y Modelo Conceptual de la Base de Datos.
- **Diseño:** Definición del Modelo Lógico de la Base de Datos e identificación de las estructuras necesarias.
- **Construcción:** Desarrollo de los archivos a RDF y la construcción de las tripletas, así como de las estructuras de la Base de Datos para su posterior subida a la nube.
- **Pruebas:** Redacción y ejecución de los guiones de pruebas. Gestión de incidencias.
- **Cierre:** Entrega del proyecto finalizado.

Para el desarrollo de las fases se ha seguido una mezcla entre la metodología Clásica (en Cascada) y la de ciclos de vida iterativos.

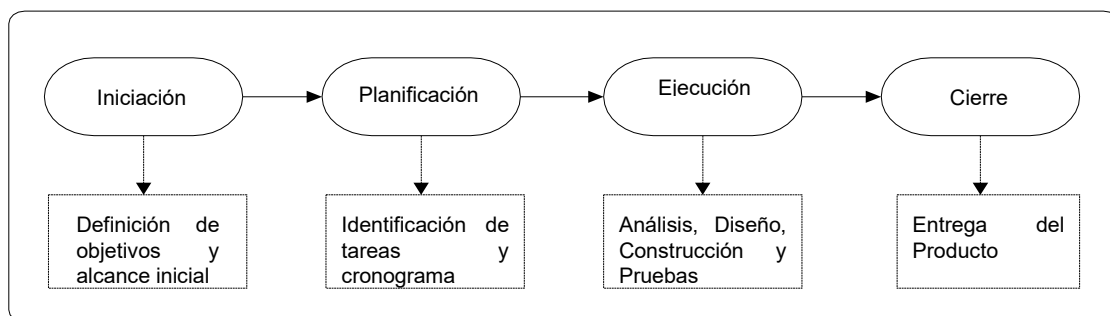


Figura 2. Acciones en cada fase

Se siguió la metodología Clásica para el desarrollo de las fases del proyecto. Dentro de la fase Ejecución, se ha seguido la metodología del Proceso Iterativa por Bloques para las acciones de transformación de los archivos a RDF, generación de grafos, almacenamiento en la Nube y la realización de consultas.

El diagrama siguiente describe el método seguido es el siguiente:

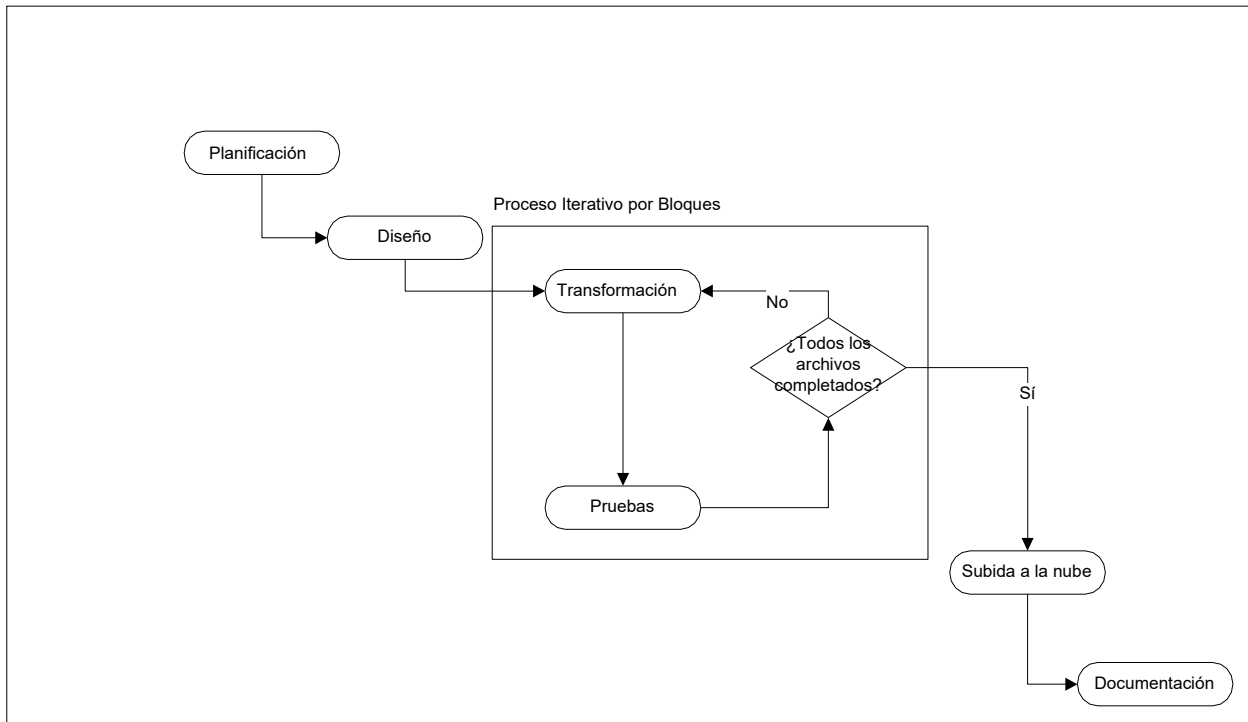


Figura 3. Método seguido

2.3 Planificación del proyecto

En la fase de Iniciación del proyecto, entre otros, se identificaron los objetivos del proyecto y el alcance inicial (véase el apartado *Objetivos*).

Para la fase de Planificación, se identificaron las tareas concretas que deben ser realizadas durante el proyecto, su duración aproximada y el intervalo de fechas en que deban realizarse. Se ha definido un diagrama de Gantt para mostrar la planificación temporal de las tareas.

2.3.1 Identificación de tareas concretas

A continuación, se muestra la lista de tareas identificadas para la realización del proyecto. Las tareas están clasificadas en los grupos generales siguientes: Preparación del proyecto, Diseño, Transformación de archivos, Subida de archivos, Generación de consultas y Documentación final. Se incluye una descripción de cada tarea.

Preparación del Proyecto

- **Tarea: Instalación de Openrefine en estación de trabajo.**
Instalación de Openrefine para la transformación de archivos a RDF. Se instalará versión *Openrefine-win-3.4.1* junto con la extensión *rdf-extension-1.3.0*

Diseño

- **Tarea: Decisiones de Diseño.**
Documentación de las decisiones de diseño que se han tomado.
- **Tarea: Modelos Lógico de la Base de Datos.**
Diagrama de tripletas para describir el modelo lógico de la *Base de Datos*.
- **Tarea: Selección del vocabulario**
Selección o creación de un vocabulario específico para las tripletas en función del archivo que se esté transformando.

Transformación de XML a RDF

- **Tarea: Transformación de los archivos *Registros de Actividades de Tratamiento de Datos Personales a RDF*.**
Utilizando Openrefine introducimos los archivos a RDF utilizando la ontología OWL.
- **Tarea: Creación de tripletas en la Base de Datos.**
Componer un esquema resumen del pipeline de la transformación.
- **Tarea: Creación de Grafos de la Base de Datos.**
Comprobar que el script creado sea correcto y generar un grafo para previsualizar el diseño de la base de datos.

Subida de archivos

- **Tarea: Búsqueda de nube.**
Buscar un servicio en la nube con soporte RDF con el fin de almacenar los archivos creados.
- **Tarea: Subida a la nube.**
Subir las tripletas de los archivos para su almacenamiento y poder desarrollar y realizar las consultas

Generación de consultas

- **Tarea: Desarrollo de consultas.**
Realizar un desarrollo de consultas para la comprobación del funcionamiento de la nube
- **Tarea: Ejecución de los guiones de pruebas.**
Ejecutar los guiones de prueba para comprobar que las consultas y la construcción de la nube sea correcta.
- **Tarea: Gestión de incidencias.**
Arreglar las incidencias detectadas durante la ejecución de los guiones de prueba.

Documentación final

- **Tarea: Documentación.**
Redactar la documentación final del proyecto.

2.3.2 Calendario de tareas: Diagrama de Gantt

A continuación se muestra el cronograma para la realización de las tareas del apartado anterior:

ID	Tarea	Inicio	Final	Duración	Febrero 2021				Marzo 2021					Abril 2021				Mayo 2021					Junio 2021	
					05/02	12/02	17/02	28/02	01/03	15/03	22/03	26/03	31/03	05/04	19/04	26/04	30/04	08/05	17/05	17/05	25/05	30/05		
1	Instalación de Openrefine en la estación de trabajo	05/02/2021	08/02/2021	4d																				
2	Instalación de extensiones RDF en la aplicación y lectura de <i>Manual de Uso</i>	05/02/2021	08/02/2021	4d																				
3	Creación de un Vocabulario específico para la ontología OWL	Se crea en función de cada archivo		--																				
4	Transformación de archivo <i>Presidencia</i> a RDF	12/02/2021	16/02/2021	5d																				
5	Instalación máquina virtual VMware Workstation 16 Player	17/02/2021	19/02/2021	2d																				
6	Instalación del almacén AllegroGraph	17/02/2021	19/02/2021	2d																				
7	Diseño de grafo	23/02/2021	26/02/2021	4d																				
8	Documentación y planificación del proyecto	01/03/2021	11/03/2021	10d																				
9	Transformación de archivo <i>Transparencia</i> a RDF	15/03/2021	17/03/2021	2d																				
10	Transformación de archivo <i>Economía y Hacienda</i> a RDF	17/03/2021	19/03/2021	2d																				
11	Transformación de archivo <i>Empleo e Industria</i> a RDF	22/03/2021	24/03/2021	2d																				
12	Transformación de archivo <i>Fomento y Medio</i> a RDF	24/03/2021	26/03/2021	2d																				
13	Transformación de archivo <i>Agricultura y Ganadería</i> a RDF	29/03/2021	31/03/2021	2d																				
14	Transformación de archivo <i>Transparencia</i> a RDF	02/04/2021	05/04/2021	3d																				
15	Transformación de archivo <i>Sanidad</i> a RDF	07/04/2021	09/04/2021	2d																				
16	Transformación de archivo <i>Familia e Igualdad</i> a RDF	11/04/2021	13/04/2021	2d																				
17	Transformación de archivo <i>Educación</i> a RDF	14/04/2021	16/04/2021	2d																				
18	Transformación de archivo <i>Cultura y Turismo</i> a RDF	17/04/2021	19/04/2021	2d																				
19	Comprobación de la transformación de los archivos a RDF	Siempre tras cada transformación		--																				
20	Desarrollo de grafos	19/04/2021	26/04/2021	8d																				
21	Desarrollo de consultas	27/04/2021	29/04/2021	2d																				
22	Búsqueda de una nube que soporte RDF y subida de los archivos al almacén	30/04/2021	08/05/2021	8d																				
23	Comprobación de almacenamiento y funcionamiento de las consultas	08/05/2021	10/05/2021	2d																				
24	Gestión de incidencias	11/05/2021	17/05/2021	6d																				
25	Documentación Final	18/05/2021	25/05/2021	8d																				

Tabla 1. Cronograma

Para cada tarea se muestra una descripción corta, la fecha en que debe comenzar, la fecha en que debe haberse finalizado, la duración en días entre la fecha inicio y la fecha fin y una visualización gráfica de la extensión de cada tarea a lo largo del calendario.

3. Tecnologías utilizadas

3.1 Web Semántica

La Web Semántica ha sido impulsada por Tim Berners-Lee, y otras personas relacionadas con el W3C (*World Wide Web Consortium*). El primer avance en este sentido fue la publicación en septiembre de 1998, por parte de Berners-Lee de 2 documentos denominados *Semantic Web Road Map* y *What the Semantic Web can represent*.

Basada en documentos y enlaces de hipertexto, cambió por completo la forma en la que podemos buscar, almacenar, gestionar y acceder a un sinfín de información. Fue diseñada para la lectura humana y no para que la información que contiene pudiera procesarse de forma automática. Si hacemos una búsqueda de documentos, por ejemplo, del término "aula", la Web no distingue entre los distintos significados o contextos en los que aparece este término (definición de aula, información docente, universidades que anuncian su web "*aula virtual*", etc.). La Web actual tampoco permite automatizar procesos, como hacer la reserva de plaza, consultar los medios de transporte disponibles hasta la ciudad donde se celebre el evento, reservar billete o conseguir un plano de dicha ciudad. Aun utilizando un potente buscador, se pierden muchas horas navegando por los resultados obtenidos tras la consulta para acceder a la información de forma manual; cuando esto lo podría hacer un programa o agente inteligente.

La Web Semántica vendría a ser una extensión de la Web actual dotada de significado, esto es, un espacio donde la información tendría un significado bien definido de manera que pudiera ser interpretada tanto por agentes humanos como por agentes computarizados.

En el año 2000, Berners-Lee ofreció una conferencia en el marco del W3C donde propuso: "La nueva información debe ser reunida de forma que un buscador pueda "comprender" en lugar de ponerla simplemente en una "lista". La Web semántica sería una red de documentos "más inteligentes" que permitan, a su vez, búsquedas más inteligentes. La idea sería aumentar la inteligencia de los contenidos de las páginas web dotándolas de contenido semántico. La Web actual posee una gran capacidad para almacenar datos y puede leer y visualizar los contenidos, pero no es capaz de pensar ni de entender todo lo que contiene. Se precisa, por lo tanto, de una nueva Web -*la Web semántica*- que hará posible no sólo almacenar los datos, sino entender e interpretar el sentido de esta información. De esta forma, Berners-Lee presenta la nueva arquitectura en que se basará la Web Semántica, no entendida como una nueva Web, sino como una extensión de la Web existente.

De acuerdo con Berners-Lee, la arquitectura de la Web Semántica se podría representar de la siguiente forma:

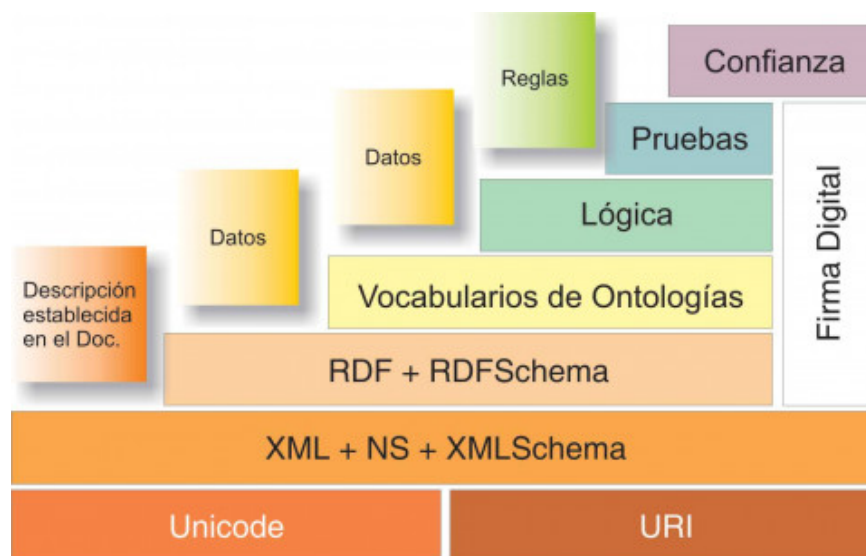


Imagen 1. Arquitectura de la Web Semántica

Según la W3C la definición de Web Semántica es "Una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida". Esto se consigue, como decía Berners-Lee, dotando a la Web de más significado y, por lo tanto, de más semántica, lo que permite obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

3.2 RDF

El Resource Description Framework (**RDF**) es un estándar, aunque siendo estrictos en la definición se debe entender como una recomendación de la W3C, para la descripción de recursos. En torno a los recursos existe aún un debate acerca de qué es considerado un recurso y qué no. En el ámbito de esta investigación se considera un recurso cualquier entidad que se pueda identificar, desde una persona o una página web hasta un número. Por lo tanto, podemos definir **RDF** como un lenguaje para especificar metadatos que permite la interoperabilidad entre aplicaciones que intercambian información comprensible por la página web, para proporcionar una infraestructura que soporte actividades de metadatos.

RDF está diseñado para situaciones en las que esta información debe ser procesada por aplicaciones en lugar de mostrarse solo a las personas y proporciona un marco común para expresar esta información, de modo que pueda intercambiarse entre aplicaciones sin perder su significado.

Dado que es un marco común, los diseñadores de aplicaciones pueden aprovechar la disponibilidad de herramientas de procesamiento y analizadores RDF comunes. La capacidad de intercambiar información entre diferentes aplicaciones significa que la información puede estar disponible para aplicaciones distintas de aquellas para las que se creó originalmente. RDF se basa en la idea de identificar cosas mediante identificadores web, llamados identificadores uniformes de recursos o URI, y describir los recursos en términos de propiedades simples y valores de propiedad. Esto permite que RDF represente declaraciones simples sobre recursos como un gráfico de nodos y arcos que representan los recursos, sus propiedades y valores.

Una **URI Uniform Resource Identifier** (Identificador Uniforme del Recurso) Es un conjunto genérico de todos los nombres y direcciones que se refieren a un recurso.

En el siguiente ejemplo mostramos un grafo diseñado por Eric Miller en el que se declara "hay una Persona identificada por un nombre que es Eric Miller, cuya dirección de correo electrónico es em@w3.org, y cuyo título es Dr."

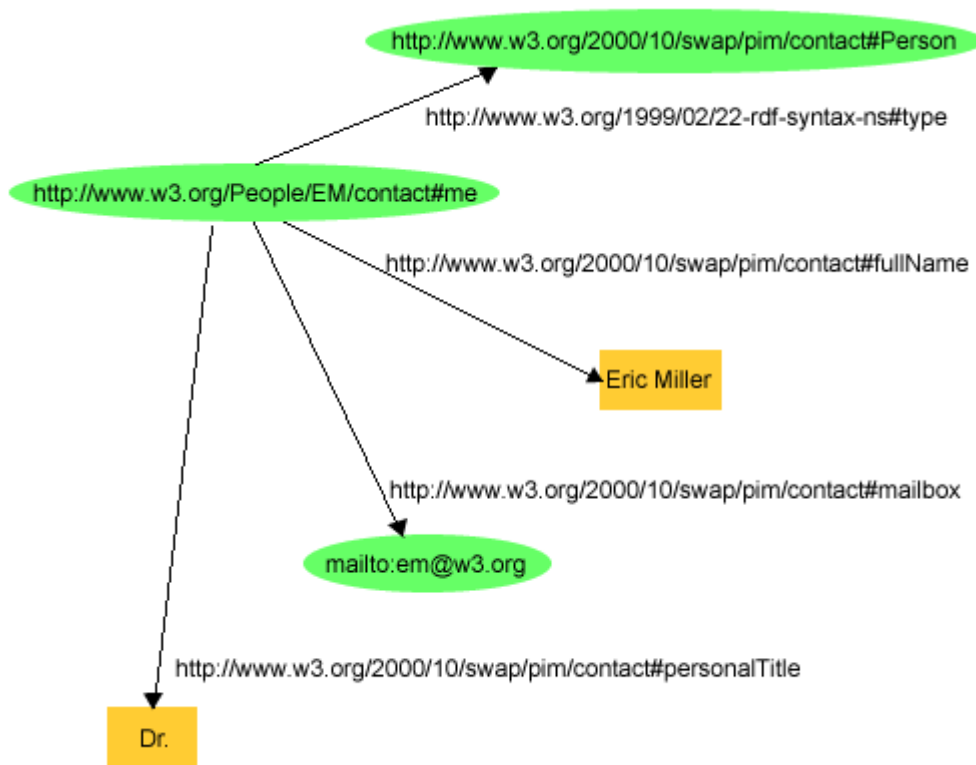


Imagen 2. Grafo RDF diseñado por Eric Miller

RDF nos proporciona una sintaxis basada en XML, denominada RDF/XML para poder editar estos grafos, en este caso el ejemplo proporcionado por Eric Miller sería:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

Imagen 3. RDF/XML de Eric Miller

RDF/XML es procesable por máquina y, usando URIs, se pueden enlazar las piezas de información a lo largo de la Web. Sin embargo, las URIs de RDF pueden hacer referencia a cualquier cosa identificable, incluyendo cosas que pueden no ser directamente recuperables en la Web. El resultado es que, para describir tales cosas como páginas web, RDF puede también describir coches, negocios, gente, noticias, eventos etc. E incluso, las propiedades RDF que tienen los URIs, para precisamente identificar las relaciones que existen entre los términos enlazados. En RDF tanto los sujetos, como las propiedades y los objetos, son recursos.

El fundamento de RDF es un modelo para representar propiedades designadas y valores de propiedades. El modelo RDF se basa en principios perfectamente establecidos de varias comunidades de representación de datos. Las propiedades RDF pueden recordar a atributos de recursos y en este sentido corresponden con los tradicionales pares de atributo-valor. Las propiedades RDF representan también la relación entre recursos y, por lo tanto, un modelo RDF puede parecer un diagrama entidad-relación. En la terminología del diseño orientado a objetos, los recursos corresponden con objetos y las propiedades corresponden con objetos específicos y variables de una categoría.

El modelo de datos básico consiste en tres tipos de objetos:

1. **Recursos.** Todo aquello descrito por expresiones RDF se denominan recursos. Los recursos se designan siempre por URIs más identificadores de anclas opcionales. Cualquier cosa puede tener un URI; la extensibilidad de URIs permite la introducción de identificadores para cualquier entidad imaginable.
2. **Propiedades.** Una propiedad es un aspecto específico, característica, atributo, o relación utilizada para describir un recurso. Cada propiedad tiene un significado específico, define sus valores permitidos, los tipos de recursos que puede describir y sus relaciones con otras propiedades.
3. **Sentencias.** Un recurso específico junto con una propiedad denominada más el valor de dicha propiedad para ese recurso es una sentencia RDF. Así pues, una declaración o sentencia está compuesta por 3 partes individuales: sujeto, predicado y objeto.

Como vemos en el ejemplo propuesto por Eric Miller nos encontramos con estas tres partes individuales de una sentencia denominadas, respectivamente, sujeto, predicado y objeto.

- **Sujeto.** Es el recurso desde el que parte el arco en el grafo formado.
- **Predicado.** Es la propiedad que etiqueta el arco en el grafo.
- **Objeto.** Es el recurso o literal al que llega el arco en el grafo. Puede ser otro recurso o puede ser un literal; es decir, un recurso (especificado por un URI), una cadena simple de caracteres u otros tipos de datos primitivos definidos por XML.

Eso se representa de forma gráfica en nodos y flechas donde representamos los recursos en forma de óvalos, las propiedades son las flechas que unen los recursos con los objetos y, por último, los objetos se representan en dos formas: de óvalo si es un recurso o en forma de rectángulo si es un literal.

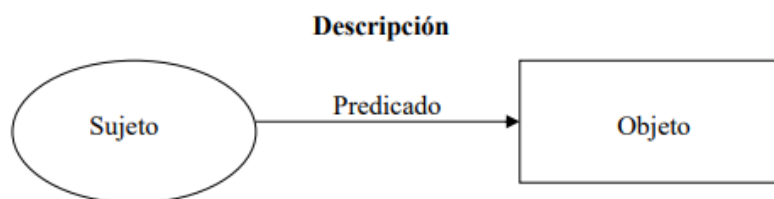


Figura 4. Grafo RDF simple

A este grafo RDF que muestra la descripción lo denominamos **tripleta**. Pero un grafo puede ir creciendo cuando se le añadan más elementos a la descripción del recurso.

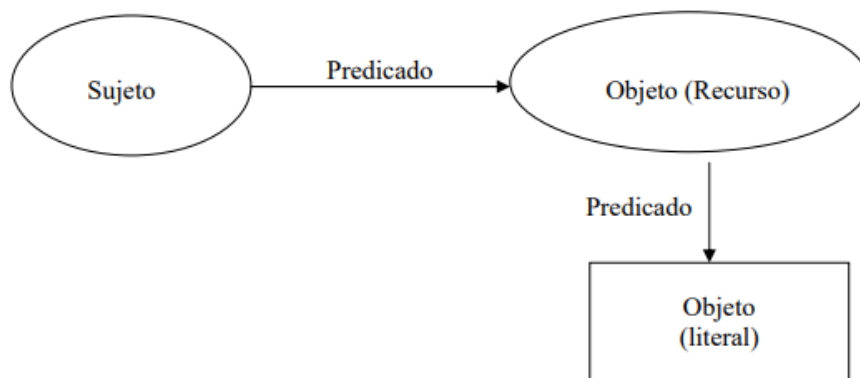


Figura 5. Grafo RDF compuesto

Como podemos ver, este grafo está compuesto en lugar de por una tripleta simple, por dos tripletas. Una formada por sujeto y predicado donde el Objeto es a su vez otro Recurso; y es este último recurso el que, al contener una propiedad, genera otra tripleta, donde éste es el sujeto -de la segunda tripleta- generando, a su vez, un predicado para unirlo a su objeto (*este formado por un literal*).

La forma que hay de relacionar y obtener las propiedades de un grafo se realiza mediante el acceso a las URIs correspondientes donde se identifican los recursos y propiedades.

3.3 Ontologías usadas

La principal tecnología de representación del conocimiento usada en la Web Semántica es la **ontología**: que facilita tanto la búsqueda de contenidos e información como mejorar el rastreo en la Web gracias a que proporciona una marca común que permite la integración, compartición y reutilización de los datos desde múltiples fuentes.

Se pueden encontrar un gran número de definiciones de ontología. Una de las más relevantes es la de que afirma que "una ontología es una especificación formal y explícita de una conceptualización compartida". Esto se refiere a la necesidad de que las ontologías sean comprensibles por la máquina. Esta definición enfatiza la necesidad de alcanzar un acuerdo para establecer la conceptualización compartida. Las ontologías proporcionan un vocabulario común de un área y define, con diferentes niveles de formalidad, el significado de términos y las relaciones que se establecen entre ellos.

Las ontologías pretenden describir el conocimiento (conceptos, atributos y relaciones) de un dominio dado. Fueron diseñadas para poderse compartir, de forma que distintas aplicaciones manejen los mismos conocimientos y puedan colaborar entre sí.

Para este proyecto han sido necesarias el uso de cuatro ontologías que permitieran realizar una correcta transformación de los Registros de Actividades de Tratamiento a RDF las cuales son:

- OWL
- Dublin Core
- SemCyLRAT
- SemRAT

En los siguientes puntos se definirá concretamente para qué son necesarias dichas ontologías y su uso en el proyecto.

3.3.1 OWL

Existen diferentes lenguajes para la creación de ontologías, aunque la W3C recomienda el uso de OWL

El Lenguaje de Ontologías Web (OWL) está diseñado para ser usado en aplicaciones que necesitan procesar el contenido de la información en lugar de únicamente representar información para los humanos. OWL facilita un mejor mecanismo de interpretabilidad de contenido Web que los mecanismos admitidos por XML, RDF, y esquema RDF (RDF-S) proporcionando vocabulario adicional junto con una semántica formal. En realidad, OWL es una extensión del lenguaje RDF y emplea las tripletas de RDF, aunque es un lenguaje con más poder expresivo ya que éste posee más funcionalidades para expresar el significado y semántica que XML, RDF, y RDFS; pero OWL va más allá que estos lenguajes pues ofrece la posibilidad de representar contenido de la Web interpretable por máquina. OWL es una revisión del lenguaje de ontologías web DAML+OIL que incorpora lecciones aprendidas desde el diseño y aplicaciones de DAML+OIL.

OWL tiene tres sublenguajes, con un nivel de expresividad creciente: OWL Lite, OWL DL, y OWL Full. A continuación, se proporciona una breve descripción de cada uno de ellos:

- **OWL Lite** está diseñado para aquellos usuarios que necesitan principalmente una clasificación jerárquica y restricciones simples. Por ello, a la vez que admite restricciones de cardinalidad, sólo permite establecer valores cardinales de 0 ó 1. OWL Lite proporciona además una ruta rápida de migración para tesauros y otras taxonomías. Tiene también una menor complejidad formal que OWL DL.
- **OWL DL** está diseñado para aquellos usuarios que quieren la máxima expresividad conservando completitud computacional, y resolubilidad. OWL DL incluye todas las construcciones del lenguaje de OWL, pero sólo pueden ser usados bajo ciertas restricciones. Por ejemplo, mientras una clase puede ser una subclase de otras muchas clases, una clase no puede ser una instancia de otra.
- **OWL Full** está dirigido a usuarios que quieren máxima expresividad y libertad sintáctica de RDF sin garantías computacionales. Por ejemplo, en OWL Full una clase puede ser considerada simultáneamente como una colección de clases individuales y como una clase individual propiamente dicha. OWL Full permite una ontología para aumentar el significado del vocabulario preestablecido (OWL o RDF). Es poco probable que cualquier software de razonamiento sea capaz de obtener un razonamiento completo para cada característica de OWL Full.

3.3.2 Dublin Core

La iniciativa de metadatos Dublin Core apoya la innovación en el diseño de metadatos y las mejores prácticas. Dublin Core es un sistema de 15 definiciones semánticas descriptivas con las que poder transmitir su mismo significado semántico.

Estas definiciones:

- Son opcionales
- Se pueden repetir
- Pueden aparecer en cualquier orden

Este sistema de definiciones fue diseñado específicamente para proporcionar un vocabulario de características básicas capaces de proporcionar la información descriptiva sobre cualquier recurso, sin que importe el formato de origen, acerca del área de especialización o el origen cultural.

En general, podemos clasificar estos elementos en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos:

- Elementos relacionados principalmente con el contenido del recurso.
- Elementos relacionados principalmente con el recurso cuando es visto como una propiedad intelectual.
- Elementos relacionados principalmente con la instanciación del recurso.

Dentro de cada clasificación definiremos los elementos que podemos encontrar. A pesar de que de Dublin Core solo usaremos un elemento en particular, viene bien conocer todos ya que se pueden ampliar el número de recursos y entonces deberíamos utilizar alguno de los siguientes elementos.

a) Contenido del Recurso

DC.Title

- Titulo: el nombre dado a un recurso, habitualmente por el autor.

DC.Subject

- Claves: los temas del recurso. Típicamente, Subject expresará las claves o frases que describen el título o el contenido del recurso. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación formales.

DC.Description

- Descripción: una descripción textual del recurso. Puede ser un resumen en el caso de un documento o una descripción del contenido en el caso de un documento visual.

DC.Source

- Fuente: secuencia de caracteres usados para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.

DC.Type

- Tipo del Recurso: la categoría del recurso. Por ejemplo, página personal, romance, poema, diccionario, etc.

DC.Relation

- Relación: es un identificador de un segundo recurso y su relación con el recurso actual. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos.

DC.Coverage

- Cobertura: es la característica de cobertura espacial y/o temporal del contenido intelectual del recurso. La cobertura espacial se refiere a una región física, utilizando por ejemplo coordenadas. La cobertura temporal se refiere al contenido del recurso, no a cuándo fue creado (que ya lo encontramos en el elemento Date).

b) Propiedad intelectual

DC.Creator

- Autor o Creador: la persona u organización responsable de la creación del contenido intelectual del recurso. Por ejemplo, los autores en el caso de documentos escritos; artistas, fotógrafos e ilustradores en el caso de recursos visuales.

DC.Publisher

- Editor: la entidad responsable de hacer que el recurso se encuentre disponible en la red en su formato actual.

DC.Contributor

- Otros Colaboradores: una persona u organización que haya tenido una contribución intelectual significativa, pero que esta sea secundaria en comparación con las de las personas u organizaciones especificadas en el elemento Creator. (por ejemplo: editor, ilustrador y traductor).

DC.Rights

- Derechos: son una referencia (por ejemplo, una URL) para una nota sobre derechos de autor, para un servicio de gestión de derechos o para un servicio que dará información sobre términos y condiciones de acceso a un recurso.

c) Instanciación

DC.Date

- Fecha: una fecha en la cual el recurso se puso a disposición del usuario en su forma actual. Esta fecha no se tiene que confundir con la que pertenece al elemento Coverage, que estaría asociada con el recurso en la medida que el contenido intelectual está de alguna manera relacionado con aquella fecha.

DC.Format

- Formato: es el formato de datos de un recurso, usado para identificar el software y, posiblemente, el hardware que se necesitaría para mostrar el recurso.

DC.Identifier

- Identificador del Recurso: secuencia de caracteres utilizados para identificar unívocamente un recurso. Ejemplos para recursos en línea pueden ser URLs y URNs. Para otros recursos pueden ser usados otros formatos de identificadores, como por ejemplo ISBN ("International Standard Book Number").

DC.Language

- Idioma: Idioma del contenido intelectual del recurso.

Cualquier persona puede utilizar los metadatos de Dublin Core para describir los recursos de un sistema de información. Las páginas Web son uno de los tipos más comunes de recursos que utilizan las descripciones de Dublin Core. Los metadatos de Dublin Core están siendo utilizados como la base para los sistemas descriptivo:

- Organizaciones educativas.
- Bibliotecas.
- Instituciones del gobierno.
- Sector científico de la investigación.
- Autores de páginas Web.
- Negocios que requieren lugares más investigables.
- Corporaciones con sistemas de gerencia extensos en conocimiento.

Para este proyecto, al tratarse de una institución del gobierno como es la Junta de Castilla y León, utilizaremos de Dublin Core el elemento **DC.Source** de Contenido del recurso ya que vamos a enlazar en él la URL de la que hemos sacado cada uno de los Registros de Actividades de Tratamiento, para que redirijan al usuario que este consultando en ese momento.

3.3.3 Ontologías propias

Una vez usadas las ontologías recomendadas por W3C y existentes para recursos específicos del archivo se nos plantea un problema, que era cómo realizar las tripletas, si no hay recursos descritos previamente en otras ontologías. La solución a este problema consiste en crear dos ontologías basadas tanto en los recursos proporcionados por los archivos de *Registro de Actividades de Tratamiento*, como las descritas en el *Artículo 30 del Reglamento General de Protección de Datos (RPGD)*.

Posteriormente plasmaré al completo dicho *Artículo 30*, para poder explicar concretamente cuales son los elementos que se pueden obtener de este artículo para los recursos de los Registros de Actividad de Tratamiento.

Artículo 30 del Reglamento General de Protección de Datos (RPGD).

“Registro de las actividades de tratamiento

1. *Cada responsable y, en su caso, su representante llevará un registro de las actividades de tratamiento efectuadas bajo su responsabilidad. Dicho registro deberá contener toda la información indicada a continuación:*
 - a) *el nombre y los datos de contacto del responsable y, en su caso, del corresponsable, del representante del responsable, y del delegado de protección de datos;*
 - b) *los fines del tratamiento;*
 - c) *una descripción de las categorías de interesados y de las categorías de datos personales;*
 - d) *las categorías de destinatarios a quienes se comunicaron o comunicarán los datos personales, incluidos los destinatarios en terceros países u organizaciones internacionales;*
 - e) *en su caso, las transferencias de datos personales a un tercer país o una organización internacional, incluida la identificación de dicho tercer país u organización internacional y, en el caso de las transferencias indicadas en el artículo 49, apartado 1, párrafo segundo, la documentación de garantías adecuadas;*
 - f) *cuando sea posible, los plazos previstos para la supresión de las diferentes categorías de datos;*
 - g) *cuando sea posible, una descripción general de las medidas técnicas y organizativas de seguridad a que se refiere el artículo 32, apartado 1.*
2. *Cada encargado y, en su caso, el representante del encargado llevará un registro de todas las categorías de actividades de tratamiento efectuadas por cuenta de un responsable que contenga:*

- a) *el nombre y los datos de contacto del encargado o encargados y de cada responsable por cuenta del cual actúe el encargado, y, en su caso, del representante del responsable o del encargado, y del delegado de protección de datos;*
 - b) *las categorías de tratamientos efectuados por cuenta de cada responsable;*
 - c) *en su caso, las transferencias de datos personales a un tercer país u organización internacional, incluida la identificación de dicho tercer país u organización internacional y, en el caso de las transferencias indicadas en el artículo 49, apartado 1, párrafo segundo, la documentación de garantías adecuadas;*
 - d) *cuando sea posible, una descripción general de las medidas técnicas y organizativas de seguridad a que se refiere el artículo 30, apartado 1.*
3. *Los registros a que se refieren los apartados 1 y 2 constarán por escrito, inclusive en formato electrónico.*
 4. *El responsable o el encargado del tratamiento y, en su caso, el representante del responsable o del encargado pondrán el registro a disposición de la autoridad de control que lo solicite.*
 5. *Las obligaciones indicadas en los apartados 1 y 2 no se aplicarán a ninguna empresa ni organización que emplee a menos de 250 personas, a menos que el tratamiento que realice pueda entrañar un riesgo para los derechos y libertades de los interesados, no sea ocasional, o incluya categorías especiales de datos personales indicadas en el artículo 9, apartado 1, o datos personales relativos a condenas e infracciones penales a que se refiere el artículo 10.”*

De este texto podemos sacar los siguientes elementos para las ontologías creadas SemCyLRAT y SemRAT:

- *semCyLRAT:tratamiento*
- *semCyLRAT:consejeria*
- *semCyLRAT:centro*
- *semRAT:responsable*
- *semRAT:delegado*
- *semRAT:base_juridica*
- *semRAT:finas*
- *semRAT:categoria_interesado*
- *semRAT:categoria_datos_personales*
- *semRAT:cesiones_datos*
- *semRAT:transferencias_internacionales*
- *semRAT:plazo_supresion*
- *semRAT:medidas_seguridad*
- *semRAT:encargado*
- *semRAT:categoria_tratamientos*
- *semRAT:categoria_destinatarios*

En los siguientes puntos pasaré a definir para que se usó y qué función tiene cada uno de los elementos.

3.3.3.2 *SemCyLRAT*

Para la creación de los elementos de los recursos que se usan en los Registros de Actividades de Tratamiento hemos creado SemCyLRAT con el fin de poder dar significado semántico a los recursos de los RAT que no vienen expresados en el Artículo 30 del RPDG pero que si constan en dichos RAT.

Para ello ha sido necesario la creación de un dominio que no estuviera en uso en este caso usamos el nombre del proyecto “<http://semcykrat.edu/>” al que podamos redirigir la cantidad de elementos encontrados que son:

semCyLRAT:tratamiento:

- Tratamiento: En este recurso incluiremos de los archivos RAT los nombres de los departamentos de Procesos, Gestión, Registro, etc. Que tiene la junta de Castilla y León en función de lo que busque el usuario.

semCyLRAT:consejeria:

- Consejería: Referencia a la consejería de la que forma parte el tratamiento al que hacemos referencia.

semCyLRAT:centro:

- Centro: Centro directivo que gestiona los tratamientos.

3.3.3.3 *SemRAT*

En esta ontología los recursos de los Registros de Actividades de Tratamiento sí que vienen expresadas en el Artículo 30 del RPDG. Por tanto, al igual que en el caso anterior, tendremos que crear un dominio que no pertenezca a ninguna ontología existente. No podrá llamarse igual ya que el anterior no forma parte de ninguno de los puntos del Artículo 30, pese a que la Junta de Castilla y León tenga en consideración que sí son necesarios; por lo tanto, he tomado la decisión de asignarle a la ontología un dominio similar que será: “<http://semcykrat.edu/>”.

semRAT:responsable

- Responsable: Se estipula la ubicación y las maneras de contacto tanto telefónicamente como por correo del responsable que se hace cargo del tratamiento.

semRAT:delegado

- delegado: Ubicación y contacto, teléfono y correo, del delegado de protección de datos.

semRAT:base_juridica

- Base jurídica: Artículos del RPDG y decretos ley en los que se basa el tratamiento.

semRAT:fines

- Fines: Finalidad del tratamiento de acuerdo con los Artículos del RPDG.

semRAT:categoria_interesado

- Categorías de interesados: Personas que puedan estar interesadas o a las que afecta el tratamiento.

semRAT:categoria_datos_personales

- Datos personales: Documentación que debe aportar la persona interesada a la institución.

semRAT:cesiones_datos

- Cesión de datos: Los datos que tiene la obligación de proporcionar a algunas de las instituciones gubernamentales competentes.

semRAT:transferencias_internacionales

- Transferencias internacionales: Informa a la persona interesada si están previstas o no, y qué tipo de datos se van a transferir internacionalmente.

semRAT:plazo_supresion

- Plazo de supresión: Tiempo máximo por el cuál la institución conservará los datos proporcionados según normativa.

semRAT:medidas_seguridad

- Medidas de seguridad: Medidas que se tomarán para proteger los datos proporcionados de manera segura.

semRAT:encargado

- Encargado: Persona que llevará un registro de todas las categorías efectuadas.

semRAT:categoria_tratamientos

- Categoría de tratamientos: Tipo de tratamiento que se está tratando.

semRAT:categoria_destinatarios

- Categoría de destinatarios: A quienes se comunicarán los datos personales, incluidos los destinatarios en terceros países u organizaciones internacionales.

3.4 SPARQL

El aspecto principal del uso de ontologías para el problema que se desea resolver es el de la posibilidad de utilizar SPARQL (*Protocol and RDF Query Language*) como lenguaje de consultas, ya que está basado en RDF. SPARQL es una tecnología clave en el desarrollo de la Web Semántica que permite extraer información precisa mediante consultas lógicas, reduciendo así la complejidad de la propia consulta.

SPARQL es un lenguaje de consultas para RDF que está diseñado para cumplir con los casos de uso y necesidades identificadas por el Grupo de Trabajo de Acceso a Datos RDF. SPARQL nos permite traducir datos en grafo, intensamente enlazados, en datos normalizados en formato tabular, esto es, distribuidos en filas y columnas. Resulta útil pensar las consultas SPARQL como un *Mad Lib* -un conjunto de oraciones con espacios en blanco-. La base de datos tomará esta consulta y encontrará cada conjunto de oraciones

que encaje correctamente en estos espacios en blanco, devolviéndonos los valores coincidentes como una tabla.

```
SELECT ?pintura
WHERE {
  ?pintura <utiliza la técnica de> <óleo sobre lienzo> .
}
```

Figura 6. Ejemplo de consulta en SPARQL

En esta imagen podemos ver un ejemplo de una consulta realizada en SPARQL dónde *?pintura* representa el nodo que la base de datos nos devolverá. Una vez recibida la consulta, la base de datos buscará todos los valores para *?pintura* que adecuadamente complete la declaración RDF *<utiliza la técnica de> <óleo sobre lienzo>*.

En esta Base de Datos de ejemplo podemos ver como los recursos que cumplen con dicha consulta son “*La ronda de noche*” y “*La tostadora de perlas*”



Figura 7. Visualización de la Base de Datos

Ahora bien, lo que hace a RDF y a SPARQL herramientas tan potentes es su habilidad para crear consultas complejas que referencian múltiples variables al mismo tiempo. Por ejemplo, podríamos buscar en nuestra base de datos RDF pinturas creadas por cualquier artista que fuese holandés. Cuyo ejemplo de consulta sería:

```
SELECT ?artista ?pintura
WHERE {
  ?artista <es de nacionalidad> <holandesa> .
  ?pintura <utiliza la técnica de> <óleo sobre lienzo> .
}
```

Figura 8. Consulta cualquier artista holandés

En este ejemplo hemos introducido una segunda variable: *?artista*. Por lo tanto la base de datos RDF devolverá todas las combinaciones que coincidan con *?artista* holandés y *?pintura* que utilice óleo sobre lienzo.

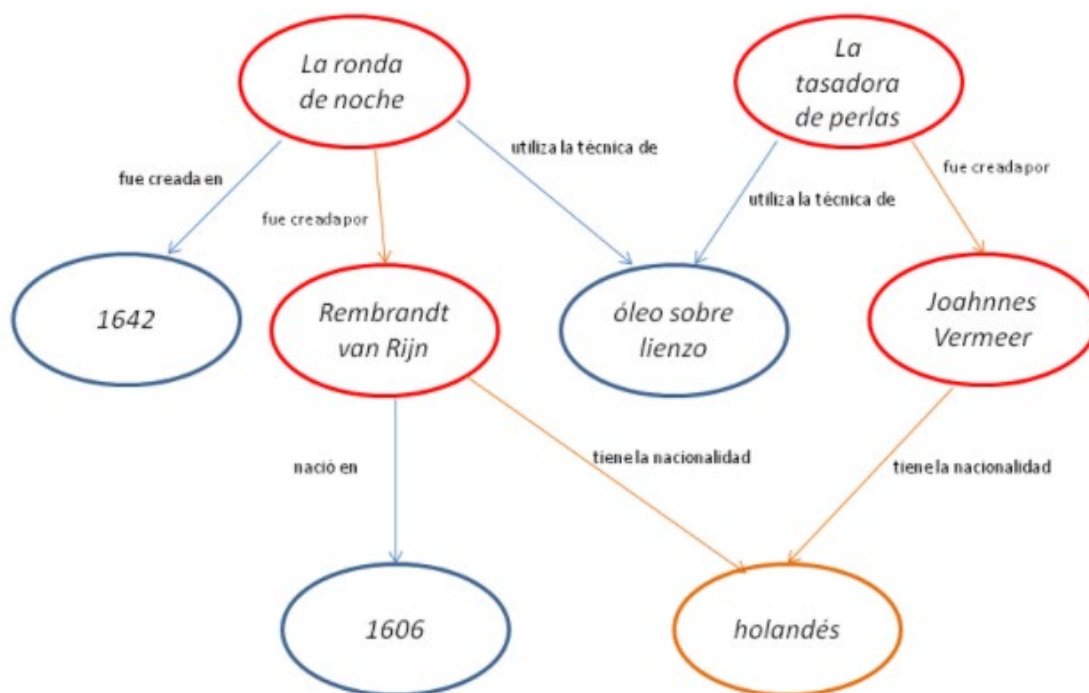


Figura 9. Visualización de la Base de Datos

En esta segunda consulta podemos ver como nuestro resultado sería:

Artistas	Pinturas
Rembrandt van Rijn	La ronda de noche
Johannes Vermeer	La tasadora de perlas

Tabla 2. Resultado de la consulta

Este ejemplo es muy útil para comprobar el funcionamiento de la base de datos y las consultas en RDF. Sin embargo, RDF se almacena principalmente en formato URI que separa las entidades conceptuales de sus etiquetas lingüísticas.

Por lo tanto, la consulta `?pintura <creadas por> <Rembrandt>` .

Sería más parecida a `?pintura <http://purl.org/dc/terms/creator> <http://dbpedia.org/resource/Rembrandt>` .

Como se puede observar, a diferencia de las URI que en esta consulta están enmarcadas por los signos `<>`, los objetos son cadenas de texto entrecomilladas. Esto es lo que se conoce como *literales*. Los literales representan valores, mientras que las URI representan referencias. Por ejemplo, `<http://dbpedia.org/resources/Rembrandt>` representa una entidad que puede referenciar y ser referenciada por muchas otras declaraciones, mientras que la cadena de texto "Rembrandt van Rijn" solo se representa a sí misma.

Las URIs pueden llegar a ser difíciles de manejar cuando se componen consultas SPARQL. Para simplificar este proceso se utilizan los *prefijos*. Los prefijos son atajos que nos liberan de tener que escribir toda la larga cadena de caracteres que constituye una

URI. Por ejemplo, el predicado para recuperar el título de *La ronda de noche*, <http://purl.org/dc/terms/title>. Con los siguientes prefijos, solo necesitamos escribir *dc:title* cuando queramos utilizar un predicado *purl.org*. *dc:* representa la cadena completa <http://purl.org/dc/terms>, y *title* simplemente se agrega al final de este enlace.

3.5 Los Registros de las Actividades de Tratamiento (RAT) original y transformado

En este apartado explico el contenido generalizado de cada uno de los archivos Excel obtenidos de la web de la Junta de Castilla y León. En caso de querer comprobar particularmente cada uno de los contenidos de los archivos, se podrán analizar en el apartado *Anexo I*.

Como ya se ha expresado anteriormente, la motivación de este proyecto se basa en querer mejorar los archivos proporcionados por la Junta de Castilla y León para su consulta en la web. Dichos archivos denominados Registros de Actividades de Tratamiento de Datos Personales son:

- Presidencia
- Transparencia
- Economía y hacienda
- Empleo e industria
- Fomento y Medio Ambiente
- Agricultura y ganadería
- Sanidad
- Familia e igualdad de oportunidades
- Educación
- Cultura y turismo

Cada uno de estos archivos proporcionado por la Junta de Castilla y León consiste en un documento Excel que consta de trece columnas, siendo cada columna la información específica que se debe proporcionar en función del tratamiento en particular.

Cada una de estas columnas será identificada en SemRAT y SemCyLRAT que tocó crear con anterioridad para poder relacionar el sujeto proporcionado en este archivo con el objeto al que está referenciando. Por tanto las columnas del archivo son las siguientes:

- *Tratamiento*
- *Consejería*
- *Centro*
- *Responsable*
- *Delegado*
- *Base jurídica*
- *Fines del Tratamiento*
- *Categoría del interesado*
- *Categoría de datos personales*
- *Cesiones de datos*
- *Transferencias internacionales*
- *Plazo supresión*
- *Medidas de seguridad*

El contenido de cada una de estas columnas viene explicado en el apartado donde se definen las ontologías utilizadas, en este caso las ontologías propias, creadas para SemRAT y SemCyLRAT. Aquí podemos ver un ejemplo de cómo está compuesto un Excel en este caso el de Economía y Hacienda.

	A	B	C	D	E	F	G	H
1	TRATAMIENTO	CONSEJERÍA	CENTRO DIRECTIVO	RESPONSABLE DEL TRATAMIENTO	DELEGADO DE PROTECCIÓN DE DATOS	BASE JURÍDICA	FINES DEL TRATAMIENTO	CATEGORÍAS DE INTERESADOS
2	CONSULTA DE LA DOCUMENTACIÓN DEL ARCHIVO CENTRAL DE LA CONSEJERÍA DE LA PRESIDENCIA DE LA JUNTA DE CASTILLA Y LEÓN	Consejería de la Presidencia	Secretaría General	SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcyll.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyll.es.	Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1.c) del RGPD cumplimiento de una obligación legal. Ley 6/1931, de 19 de abril, de Archivo y del Patrimonio Documental de Castilla y León. Ley 39/2015, de 1 de octubre, del procedimiento administrativo común de las Administraciones públicas. Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público. Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español.	Atender la consulta libre y gratuita de los documentos integrantes del Patrimonio Documental de Castilla y León con fines de estudio e investigación o de información para la defensa de sus derechos o el conocimiento de sus obligaciones. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1b) y 89.1 del RGPD	Personas físicas incluidos los representantes de personas jurídicas que soliciten información relativa al contenido de los documentos que custodia el Archivo Central de la Consejería de la Presidencia, por vía telefónica, correo electrónico u ordinario) y de forma presencial en la sede de la calle Santiago Alba, 1 de Valladolid.
3	GESTIÓN ECONÓMICA Y CONTABLE REALIZADA POR LA SECRETARÍA GENERAL DE LA CONSEJERÍA DE LA PRESIDENCIA Y A LA TRAMITACIÓN DE LOS EXPEDIENTES DE CONTRATACIÓN PÚBLICA REALIZADA POR LA SECRETARÍA GENERAL DE LA CONSEJERÍA DE	Consejería de la Presidencia	Secretaría General	SECRETARÍA GENERAL. C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyll.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyll.es.	Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1.c) del RGPD cumplimiento de una obligación legal. Ley 2/2006, de la Hacienda y del Sector Público de la Comunidad de Castilla y León. Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público. Ley 39/2003, de 17 de noviembre, General de Subvenciones. Ley 5/2008, de 25 de septiembre, de Subvenciones de Castilla y León. Ley 7/2005, de 24 de mayo, de la Función Pública de Castilla y León. Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas	La tramitación de expedientes de gasto y de documentos contables en materias propias de la Consejería de la Presidencia. La tramitación de la contratación pública en materias propias de la Consejería de la Presidencia. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1b) y 89.1 del RGPD	Personas físicas incluidos los representantes de las personas jurídicas que participan, y en su caso resu adjudicatarios, en los procedimientos de contratación gestionados por la Consejería de la Presidencia. Personas físicas y representantes de personas jurídicas, acreedores de la Consejería de la Presidencia.
4	REGISTRO DE ACTIVIDADES RELATIVAS A LA GESTIÓN DE RECURSOS HUMANOS DE LA CONSEJERÍA DE LA PRESIDENCIA DE LA JUNTA DE CASTILLA Y LEÓN	Consejería de la Presidencia	Secretaría General	SECRETARÍA GENERAL. C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyll.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyll.es.	Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 3.2.b) del RGPD por cumplimiento de obligaciones y el ejercicio derechos en materia laboral Real Decreto Legislativo 5/2015, de 30 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto Básico del Empleado Público. Ley 7/2005, de 24 de mayo, de la Función Pública de Castilla y León. Real Decreto Legislativo 2/2015, de 23 de octubre, por el que se aprueba el texto refundido de la Ley del Estatuto de los Trabajadores.	Gestión de los recursos humanos de la Consejería de la Presidencia de la Junta de Castilla y León. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1b) y 89.1 del RGPD.	Personas físicas al servicio de la Administración
5	GESTIÓN DE SUBVENCIONES, AYUDAS, BECAS, PREMIOS Y RECONOCIMIENTOS GESTIONADOS POR SECRETARÍA GENERAL DE LA CONSEJERÍA DE LA PRESIDENCIA DE LA JUNTA DE CASTILLA Y LEÓN	Consejería de la Presidencia	Secretaría General	SECRETARÍA GENERAL. C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyll.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyll.es.	Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1.c) del RGPD cumplimiento de una obligación legal. Artículo 3.2.g) del RGPD por razones de un interés público esencial. Ley 39/2003, de 17 de noviembre, General de Subvenciones. Ley 5/2008, de 25 de septiembre, de Subvenciones de Castilla y León. Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas. Decreto Legislativo 1/2003, de 18 de junio, por el que se aprueba el texto refundido de las normas vigentes sobre aportaciones económicas distintas a las subvenciones: Estatuto de Autonomía de Castilla y León (artículo 16.8).	Gestión de subvenciones públicas, ayudas públicas, becas, premios y reconocimientos. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1b) y 89.1 del RGPD	Personas físicas, incluidos los representantes de las personas jurídicas, que participen en las diferentes subvenciones, ayudas, becas y premios, que se gestionan por la Consejería de la Presidencia.
6	CONTROLES DE ACCESO Y VIDEOVIGILANCIA DEL CAMPUS SANTIAGO ALBA N	Consejería de la Presidencia	Secretaría General	SECRETARÍA GENERAL. C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyll.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyll.es.	Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Ley 5/2014 de 4 de abril, de Seguridad Privada.	Garantizar la seguridad de personas, bienes e instalaciones en las dependencias situadas en la calle Santiago Alba, 1 de Valladolid así como el registro y control de acceso de las personas al edificio.	Personas físicas que accedan a las dependencias situadas en la calle Santiago Alba, 1 de Valladolid.
	EJERCICIO DE LA COMPETENCIA			SECRETARÍA GENERAL. C/ Santiago Alba,		Artículo 6.1.e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1.c) del RGPD cumplimiento de una obligación legal.	Gestión de solicitudes de información y de reconocimiento honorífico de Víctimas del Terrorismo.	Personas físicas, incluidos los representantes de las personas jurídicas y personas jurídicas

Tabla 3. Excel de ejemplo RAT Economía y Hacienda

Para realizar su transformación a RDF cada una de las filas que hacen referencia a los tratamientos consistirá en un sujeto de una de estas tripletas nuevas. Cada una de las columnas será un objeto al que hace referencia este sujeto, que tendrá que estar unido a él por un predicado. Este predicado se forma llamando a prefijos existentes de alguna ontología existente, como es el caso de DC.Source, o bien generando un prefijo nuevo que haga referencia particularmente a este recurso nuevo.

Como en nuestro caso no hay ninguna ontología que haga referencia a estos recursos, es necesario crear estos prefijos explicados en el apartado ontologías propias, albergados dentro de las ontologías SemRAT y SemCyLRAT.

Ejemplo de cómo quedaría en forma de grafo el diseño generalizado de cualquiera de los archivos una vez transformado a RDF:

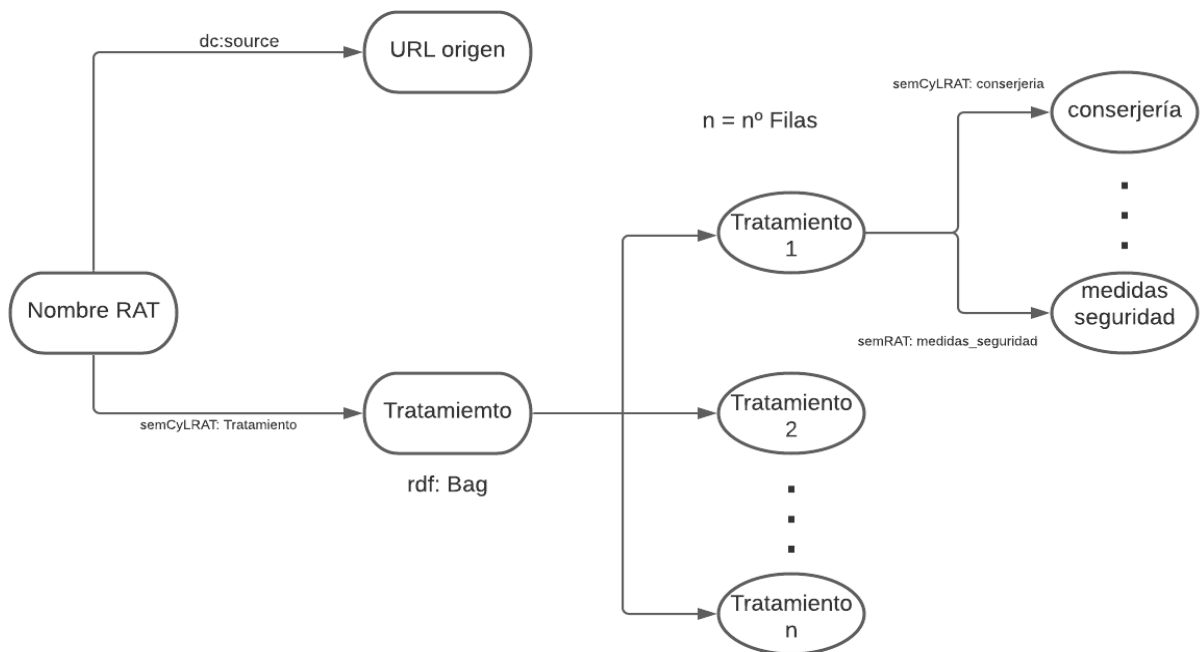


Figura 10. Diseño grafo general transformado a RDF

En este grafo podemos observar que la primera tripleta que formaremos consistirá en un sujeto que es el nombre del Registros de Actividades de Tratamiento de Datos Personales, unido a dos recursos.

El primero de ellos será la página web de la que obtenemos el archivo original y el segundo será la lista de tratamientos que irán almacenadas en un prefijo rdf:Bag ya que los organizaremos de todos juntos de manera no ordenada. Este sujeto irá relacionado a estos recursos mediante los predicados dc:source para especificar la URL origen, y SemCyLRAT:tratamiento para especificar el recurso de tipo tratamiento que estamos almacenando.

Esto lo podemos apreciar en el siguiente diagrama:

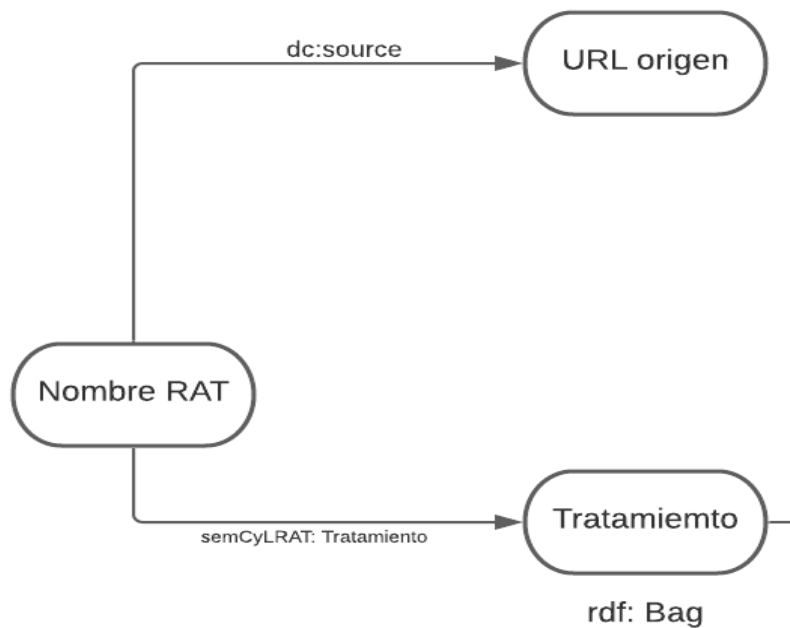


Figura 11. Ejemplo de la primera tripleta

Por lo tanto, al generar las tripletas se puede comprobar que las primeras representan a la URL original de la que obtenemos el archivo y el bag referenciando a la lista de tratamientos no ordenados, como podemos observar en la siguiente imagen.

Subject	Predicate	Object
http://semcytrat/economiayhacienda/RAT_EyH	http://purl.org/dc/elements/1.1/source	"https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-economia.html"
http://semcytrat/economiayhacienda/registro-base-de-datos-de-licitadores-y-contratistas-de-la-plataforma-electronica-duero-de-la-junta-de-castilla-y-leon	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag

Tabla 4. Primeras tripletas de RAT economía y hacienda

Una vez estipulados los tratamientos se puede observar que habrá una tripleta por fila, es decir, tendremos tantos sujetos como número de tratamientos haya en el archivo. Esto generará una segunda elaboración de tripletas, lo cual demuestra que nuestra transformación a RDF consistirá en un RDF de tipo complejo.

En la siguiente elaboración de tripletas tendremos que relacionar el sujeto (que en este caso consistirá en uno de los tratamientos), con cada uno de sus recursos objeto que serán cada una de las columnas que forman la tabla. Como es el caso, debemos generar los predicados correspondientes a las columnas restantes que vendrán expresados por los prefijos:

- *semCyLRAT:consejeria*
- *semCyLRAT:centro*
- *semRAT:responsable*
- *semRAT:delegado*
- *semRAT:base_juridica*
- *semRAT:fines*
- *semRAT:categoria_interesado*
- *semRAT:categoria_datos_personales*
- *semRAT:cesiones_datos*
- *semRAT:transferencias_internacionales*
- *semRAT:plazo_supresion*
- *semRAT:medidas_seguridad*
- *semRAT:encargado*
- *semRAT:categoria_tratamientos*
- *semRAT:categoria_destinatarios*

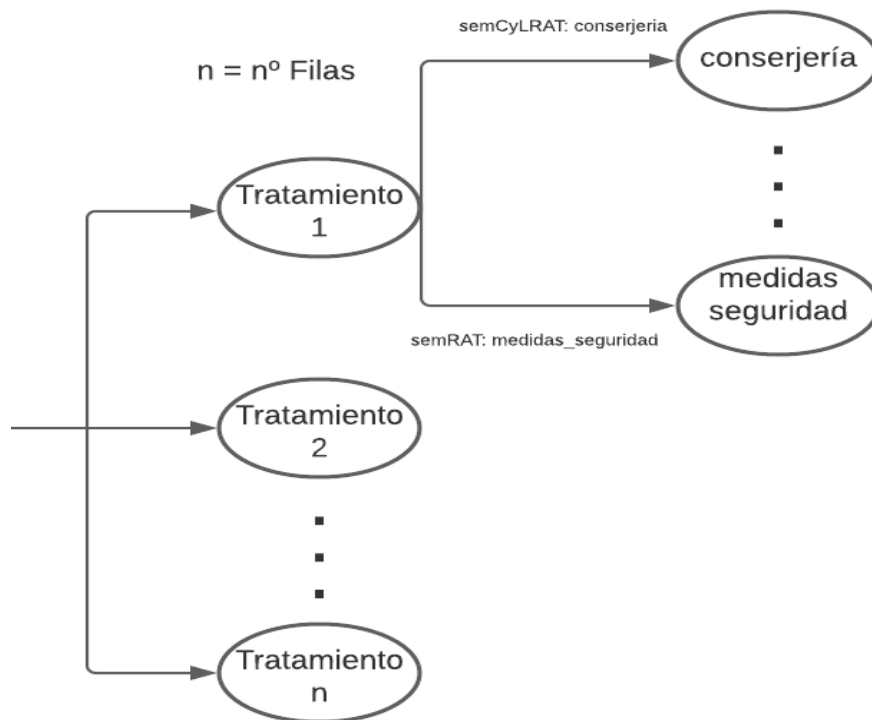


Figura 12. Ejemplo de la segunda tripleta

Se comprobar cómo se generan las tripletas en la siguiente tabla.

Observamos que para un tratamiento, en este caso el *registro base de datos de licitadores*, que forma parte de un bag de tratamientos como dice la siguiente tripleta, se generan tras ella el resto de recursos objetos como conserjería, centro, etc...

Subject	Predicate	Object
http://semcytrat/economiayahacienda/RAT_EyH	http://semcytrat.edu/tratamiento	http://semcytrat/economiayahacienda/registro-base-de-datos-de-licitadores-y-contratistas-de-la-plataforma-electrnica-duero-de-la-junta-de-castilla-y-leon
http://semcytrat/economiayahacienda/-registro-de-colegios-profesionales-y-consejos-de-colegios-de-castilla-y-leon	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat/economiayahacienda/-registro-de-colegios-profesionales-y-consejos-de-colegios-de-castilla-y-len	http://semcytrat.edu/consejeria	"Consejería de Economía y Hacienda"
http://semcytrat/economiayahacienda/-registro-de-colegios-profesionales-y-consejos-de-colegios-de-castilla-y-leon	http://semcytrat.edu/centro	"Secretaria General"
http://semcytrat/economiayahacienda/-registro-de-colegios-profesionales-y-consejos-de-colegios-de-castilla-y-leon	http://semrat.edu/responsable	"Secretaria General C/ José Cantalapiedra,2.47014 Valladolid. Teléfono 983 414000. Correo electrónico :sgeconomiayahacienda@jcyl.es"
http://semcytrat/economiayahacienda/-registro-de-colegios-profesionales-y-consejos-de-colegios-de-castilla-y-leon	http://semrat.edu/delegado	"Consejería de Economía y Hacienda.C/ José Cantalapiedra,2.47014 Valladolid.Teléfono 983 414000. Correo electrónico:dpd.economiayahacienda@jcyl.es"

Tabla 5. Ejemplo de tripletas de RDF complejo RAT economía y hacienda

El proceso de creación y transformación de los registros de actividades de tratamiento así como los registros de actividades de tratamiento originales y los grafos vendrán definidos en el siguiente apartado Desarrollo y en la documentación ANEXO IV.

3.6 Openrefine

Una de las herramientas más importantes para el proceso de transformación de los registros de actividades de tratamientos de la Junta de Castilla y León es OpenRefine.

OpenRefine (anteriormente Google Refine) es una herramienta para trabajar con datos desordenados: limpiarlos; transformándolo de un formato a otro; y ampliándolo con servicios web y datos externos. Esta herramienta es útil para el manejo de bases de datos donde podemos importar archivos para trabajar con las extensiones TSV, CSV, XML, JSON, XLS, entre otros. Lo que nos permite transformar archivos de cualquiera de estos formatos a otro.

OpenRefine RAT_Presidencia.xlsx Permalink

Facet / Filter Undo / Redo 271 / 271

26 rows

Show as: rows records Show: 5 10 25 50 rows

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

All	TRATAMIENTO	CONSEJERÍA	HTTP	CENTRO DIREC	RESPONSABLE DEL TRATAMIENTO	DELEGADO DE PROTECCI	BASE JURÍDICA
1	CONSULTA DE LA DOCUMENTACIÓN DEL ARCHIVO CENTRAL DE LA CONSEJERÍA DE LA PRESIDENCIA DE LA JUNTA DE CASTILLA Y LEÓN	Consejería de la Presidencia de la Junta de Castilla y León	https://gobierno.jcyl.es/web/ies/consejerias/registro-actividades-tratamiento-presidencia.html	Secretaría General	SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcyl.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyl.es	Artículo 6.1 e) del RGPD cumplimiento de una obligación legal. Ley 6/1991, de 19 de abril, de Archivo y del Patrimonio Documental de Castilla y León. Ley 39/2015, de 1 de octubre, de régimen jurídico del sector público. Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español.
2	GESTIÓN ECONÓMICA Y CONTABLE REALIZADA POR LA SECRETARÍA GENERAL DE LA CONSEJERÍA DE LA PRESIDENCIA Y A LA TRAMITACIÓN DE LOS EXPEDIENTES DE CONTRATACIÓN PÚBLICA REALIZADA POR LA SECRETARÍA GENERAL DE LA CONSEJERÍA DE LA	Consejería de la Presidencia de la Junta de Castilla y León		Secretaría General	SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyl.es	Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyl.es	Artículo 6.1 e) del RGPD cumplimiento de una obligación legal. Ley 2/2006, de la Hacienda y del Sector Público de la Comunidad de Castilla y León. Ley

Imagen 4. Interfaz de OpenRefine

La versión utilizada the OpenRefine para este proyecto es la versión OpenRefine 3.4.1, lanzada el 24 de septiembre de 2020. Con la que trabajaré en Windows 10, con lo que se necesitará tener instalado Java para su posterior ejecución.

Para este proyecto se usa OpenRefine para la transformación de los archivos XLS a RDF, para ello se necesita la instalación de una extensión que se debe añadir a OpenRefine para poder gestionar archivos RDF. La extensión RDF se denomina RDF Refine, y consiste en una bifurcación mejorada para exportar RDF que agrega una interfaz gráfica de usuario (GUI) para exportar datos en formato RDF.

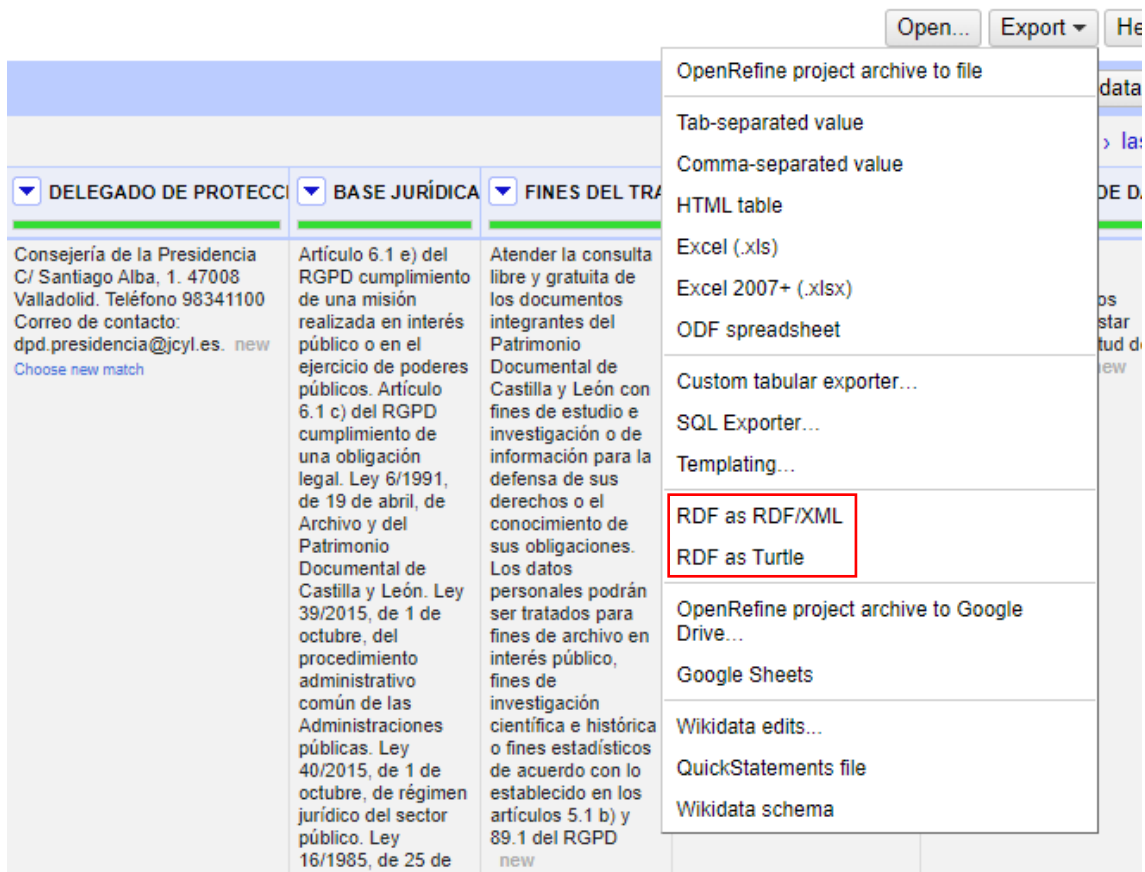


Imagen 5. Extensión RDF en OpenRefine

Una vez instalado tanto OpenRefine como su extensión RDF, podremos ejecutar OpenRefine y se abrirá una página con URL <http://localhost:3333>. Esta dirección la podemos cambiar por cualquier otra dirección con el nombre que queramos dar a nuestro proyecto en OpenRefine.

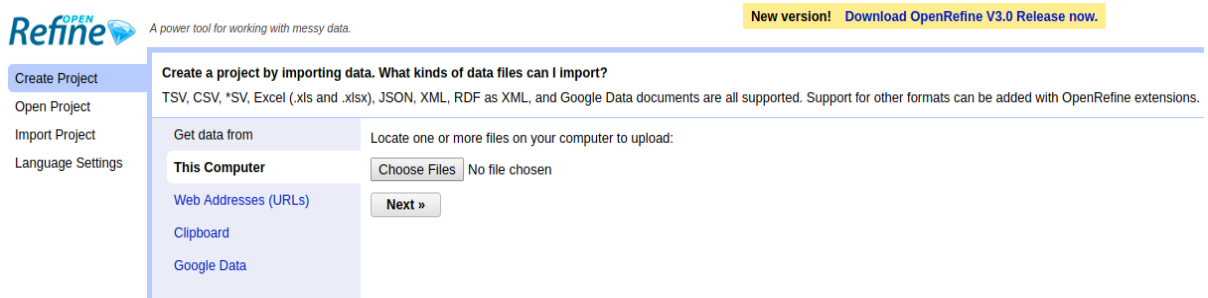


Imagen 6. Creación del proyecto en OpenRefine

Para ello seleccionaremos el archivo, XLS que queramos, le damos a “next”, y generará la tabla sobre la que tenemos que trabajar para su transformación a RDF.

3.7 AllegroGraph

Para este proyecto es necesario AllegroGraph para la gestión, creación y modelado de grafos obtenidos a través de la transformación a RDF, debido a la alta cantidad de nodos que no pueden ser gestionados por herramientas gráficas más básicas.

AllegroGraph es una tecnología de gráfico de conocimiento de eventos y entidades distribuida horizontalmente, multimodelo (documento y gráfico) que permite extraer información de decisiones sofisticada y análisis predictivo de datos altamente complejos y distribuidos que no pueden responderse con bases de datos convencionales. AllegroGraph es una base de datos de documentos y gráficos con fragmentación horizontal, de alto rendimiento y persistente, construida especialmente para este propósito. AllegroGraph utiliza una gestión de memoria eficiente en combinación con el almacenamiento en disco, lo que le permite escalar a miles de millones de documentos mientras mantiene su rendimiento. AllegroGraph cumple con los estándares W3C / ISO y admite JSON, JSON-LD, SPARQL 1.1, OWL Reasoning, SHACL y Prolog.

En este caso he utilizado la versión AllegroGraph 7.1.0 en una máquina Ubuntu versión 20.04, las características de esta versión de AllegroGraph son:

- FedShard acelera las consultas complejas: función de federación y fragmentación patentada.
- Modelado de gráficos de conocimiento de eventos de entidad.
- Documento multimodelo (JSON, JSON-LD) y base de datos gráfica (RDF, OWL).
- Gruff - Explorador de gráficos de conocimiento integrado directamente y basado en navegador.
- Cloud Native: Amazon Marketplace y Amazon EC2: AMI.
- AllegroGraph es 100% ACID y admite transacciones: compromiso, reversión y puntos de control.
- Seguridad: la base de datos de gráficos más segura de la industria con atributos triples.
- Replicación multimaestro para la coordinación de centros de datos distribuidos globalmente.
- Copias de seguridad en línea, recuperación en un momento determinado, replicación, espera en caliente.
- Indexación dinámica y automática: todos los triples comprometidos siempre están indexados (7 índices).
- Indexación de texto avanzada: indexación de texto por predicado.
- En dos fases - SOLR y MongoDB Integración.
- Soporte SHACL y SPIN.
- Todos los clientes basados en protocolo REST - RDF4J, Java Jena, Python, C#, Clojure, Perl, Rubí, Scala, y Lisp clientes.

- Completamente basado en multiprocesamiento (SMP): administración automática de recursos para todos los procesadores y discos, y uso optimizado de la memoria. Compresión de índices basada en columnas: paginación reducida, mejor rendimiento.
- Procedimientos almacenados del lado del servidor utilizando la API de JavaScript.
- Interfaz basada en JavaScript (JIG) para recorrido general de grafos.
- Compatibilidad con Soundex: permite la indexación de texto libre basada en la pronunciación fonética.
- Índices definidos por el usuario: totalmente controlados por el administrador del sistema.
- Búsqueda de texto incorporada.

3.7.1 *Instalación del entorno*

Para la utilización de esta herramienta, por problemas de capacidad y características de mi ordenador, he optado por usar una máquina virtual con sistema operativo Ubuntu 20.04, que es la opción recomendada por Franz.inc para la instalación de AllegroGraph.

Para comenzar instalaremos la máquina virtual en la que almacenaremos el entorno de AllegroGraph. En este caso la máquina virtual elegida será VMware Workstation 16 Player, es la que mejor se adapta a las características de mi ordenador con sistema operativo Windows 10. Y descargamos de la web oficial de Franz.inc <https://franz.com/agraph/downloads/> el archivo comprimido *agraph-7.1.0-ubuntu-20.04-20210204.zip* el cuál contendrá el sistema operativo mencionado antes, Ubuntu 20.04 y la herramienta AllegroGraph, lo que nos ahorrará tiempo en la instalación. Este comprimido recién descargado lo guardamos en una dirección accesible y ejecutamos la Máquina Virtual VMware Workstation 16, donde seleccionaremos la opción *Open a Virtual Machine* y seleccionaremos la ubicación en la que se encuentra el archivo de extensión VMX y nos generará una Máquina virtual a la que podremos acceder directamente con el nombre AllegroGraph 7.1.0 Virtual Machine, tal y cómo podemos ver en la imagen.

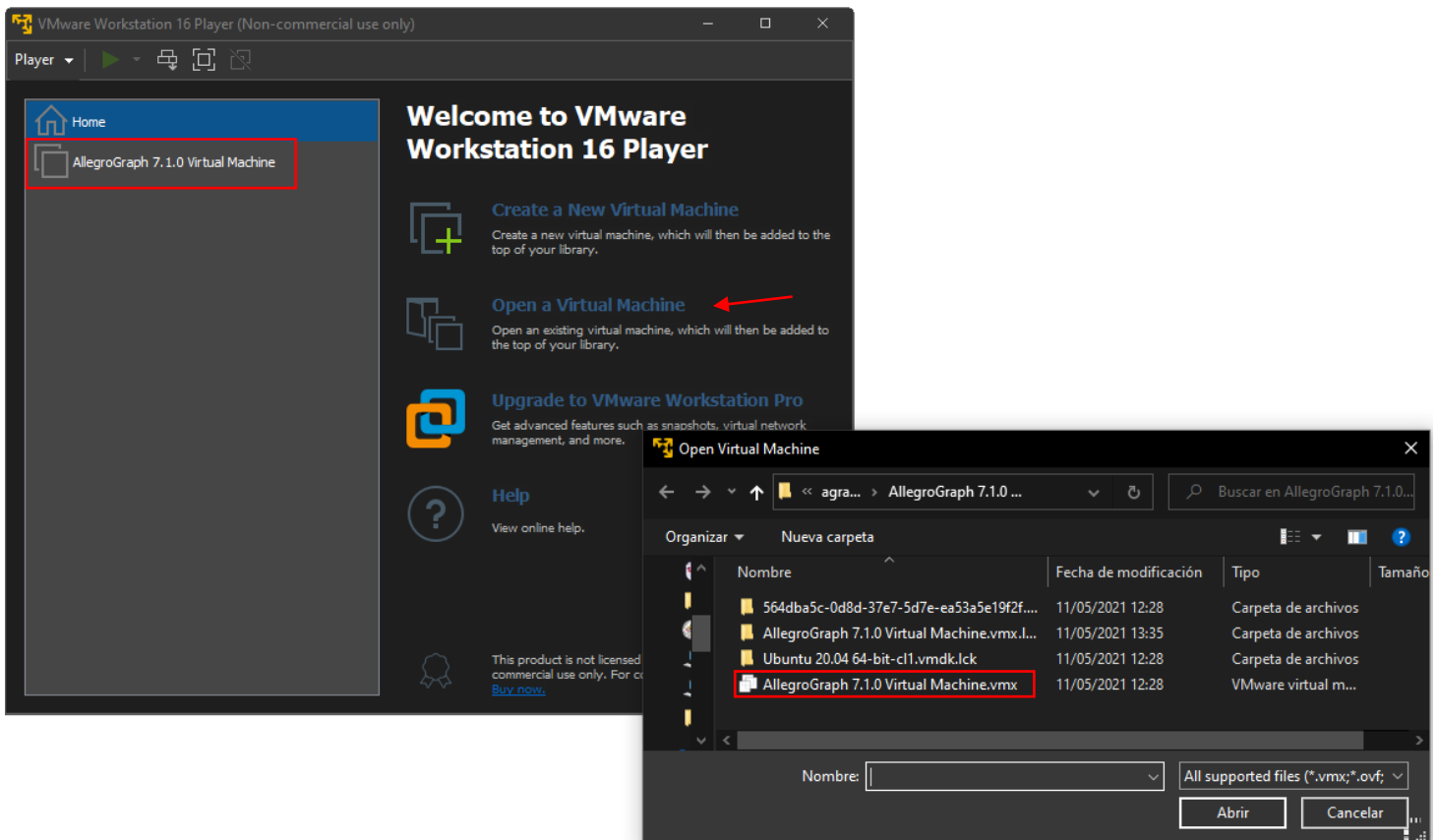


Imagen 7. Instalación de AllegroGraph 7.1.0 Virtual Machine

Una vez instalada la Máquina Virtual, seguimos los pasos estipulados en la guía de instalación de AllegroGraph 7.1.0 Virtual Machine y accederemos a nuestro sistema operativo Ubuntu con las credenciales, dadas por Franz.inc

Nombre de usuario de Ubuntu: Franz

Contraseña: allegrograph

Y ya tendremos acceso a nuestro escritorio Ubuntu en el cual se muestra una barra de iconos a la izquierda del escritorio que son:

- Navegador **Firefox**.
- **Terminal**: Muestra un shell de UNIX. Escribe comandos en el indicador para iniciar y detener AllegroGraph.
- **Ventana de archivo**: Muestra archivos en el sistema de archivos local.
- **Software de Ubuntu**: Muestra enlaces al software de Ubuntu disponible.
- Programa de **Ayuda** de Ubuntu

En el área principal del escritorio podemos observar

- Acceso directo al directorio **Franz**: Contiene los scripts para lanzar la herramienta AllegroGraph.
- **Papelera**.
- Carpeta **ag-triple-data**: Contiene varios conjuntos de datos de ejemplo.

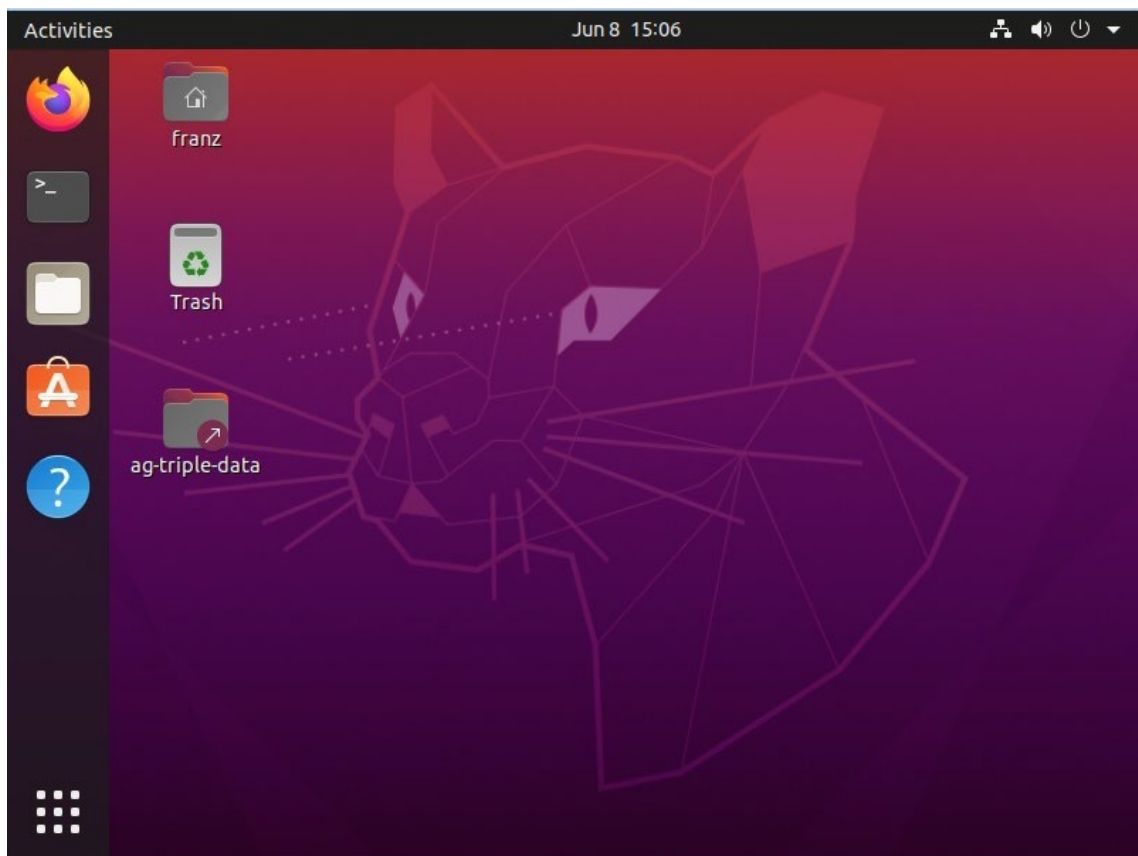


Imagen 8. Escritorio Ubuntu 20.04

Ahora que ya tenemos acceso a nuestro Sistema Operativo deberemos instalar la herramienta AllegroGraph para poder editar, gestionar y exportar las consultas, tripletas y grafos de cada uno de los Registros de Actividades de Tratamiento. Para ello deberemos seguir como hasta ahora la guía de instalación facilitada por Franz.inc

Para comenzar el lanzamiento de AllegroGraph deberemos abrir la consola de comandos de shell para iniciar y detener AllegroGraph, para determinar la dirección IP de la VM y para ejecutar programas como el editor vi.

La ilustración muestra la consola de comandos de shell en la que se ven tres comandos predefinidos como mensaje de inicio:

- Escriba '**agstart**' para iniciar AllegroGraph
- Escriba '**agstop**' para detener AllegroGraph
- Escriba '**hostname -I**' para buscar el ipaddr externo de la máquina virtual invitada. Utilice esta dirección para conectarse a la máquina virtual desde otras computadoras.

Para lanzar AllegroGraph deberemos lanzar la opción *start* dónde se encuentra el ejecutable *agraph.cfg* el cuál podemos encontrar en la ubicación *home/Franz/lib*. Al ejecutar este comando se nos abrirá una ventana en Firefox con la dirección *http://127.0.0.1:10035*

```
franz@ubuntu: ~  
AG start/stop scripts are in /home/franz/franz/  
..  
Type 'agstart' to start Allegrograph  
Type 'agstop' to stop Allegrograph  
Type 'hostname -I' to find the external ipaddr of the Guest VM. Use this address  
to connect to the VM from other computers.  
franz@ubuntu:~$ /home/franz/tmp/ag7.1.0/bin/agraph-control --config /home/franz/  
ag/lib/agraph.cfg start  
  
AllegroGraph Server Edition 7.1.0 (rc6), built on January 26, 2021 06:50:12 GMT-  
0800  
Copyright (c) 2005-2021 Franz Inc. All Rights Reserved.  
AllegroGraph contains patented and patent-pending technologies.  
AllegroGraph build cb50af5c9  
  
Current time: Tuesday, May 11, 2021 03:29:01 AM  
  
Daemonizing...  
Server started normally:  
Running with free license of 5,000,000 triples; no-expiration.  
Access AGWebView at http://127.0.0.1:10035  
franz@ubuntu:~$
```

Imagen 9. Consola de comandos de Shell

Esta dirección nos brindará acceso a usar la herramienta AllegroGraph desde la que ya sí importaremos nuestros documentos de Registro de Actividades de Tratamiento transformados.

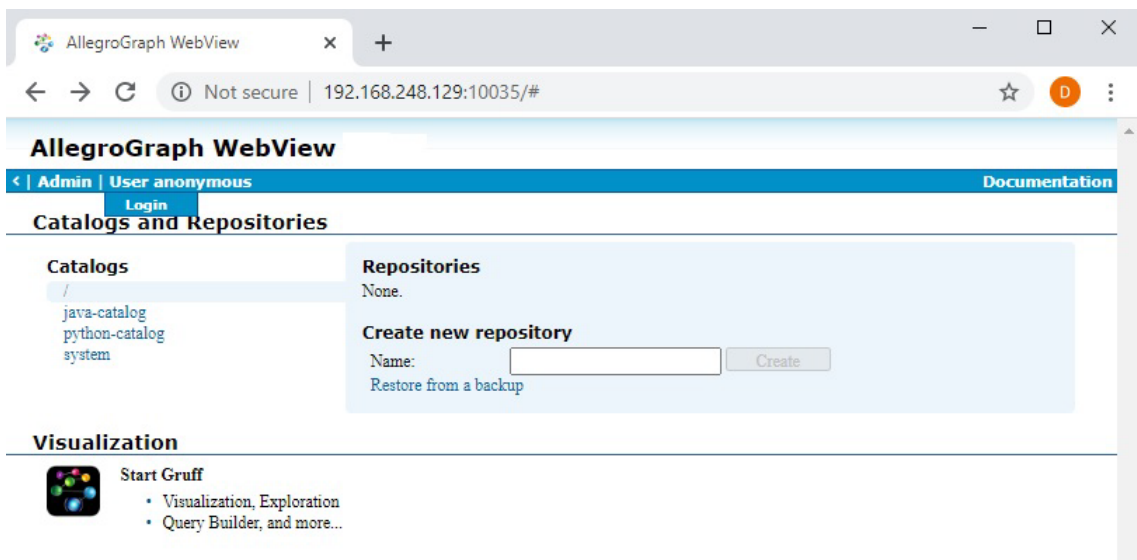


Imagen 10. Interfaz AllegroGraph

El proceso de importación consulta y generación de grafo será explicado más adelante en el apartado de desarrollo.

4. Desarrollo

En este apartado se va a explicar el proceso de transformación de cada uno de los ficheros de los registros de actividades de tratamiento de la Junta de Castilla y León desde que obtenemos su XLS hasta su transformación a RDF, creando sus ontologías propias y obteniendo el grafo y las tripletas resultantes tras su transformación.

4.1 Pipeline

En este primer apartado podemos ver un diagrama paso a paso del proceso que conlleva la transformación de un archivo RAT como muestra la siguiente imagen:

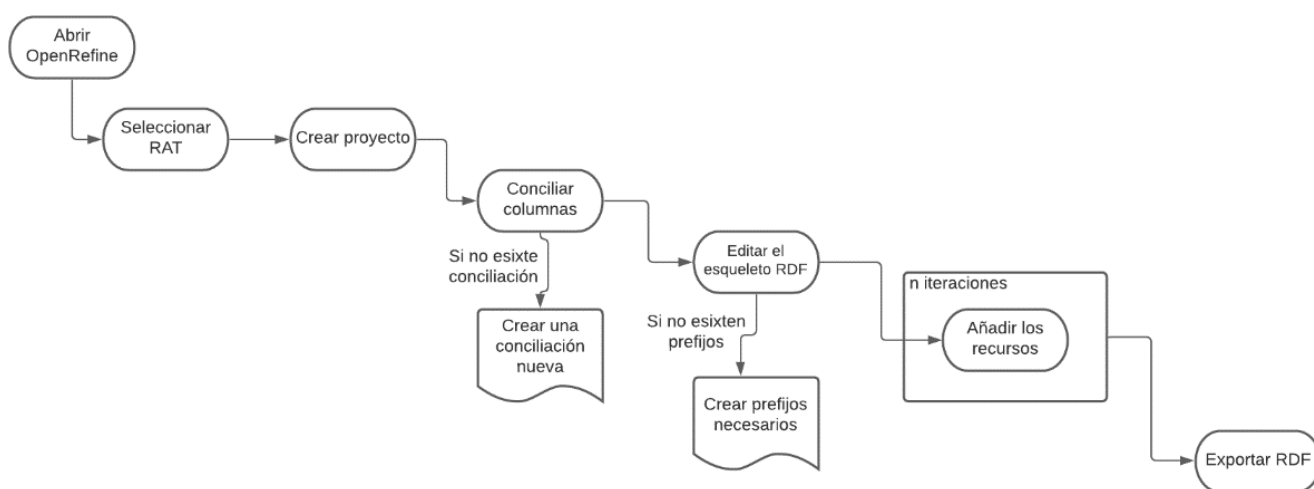


Figura 13. Pipeline

De esta imagen podemos distinguir 7 pasos:

1. Abrimos nuestra herramienta OpenRefine y seleccionamos el archivo XLS que queramos transformar a RDF.
2. Crearemos un proyecto nuevo, con el nombre del archivo que estemos transformando y dirección <http://semcytrat.edu/> y de extensión el nombre del archivo de registro de actividad sin espacios.
3. Una vez creado el proyecto deberemos conciliar las columnas, relacionar nuestro recurso con uno igual en una base de datos existente, con los recursos semejantes almacenados en web semánticas, en caso de no existir ninguna conciliación, crear una nueva.
4. Tras haber conciliado todas las columnas, deberemos organizar la estructura de nuestras tripletas en un código, esqueleto RDF, ya se trate de una estructura RDF simple, o compleja como es nuestro caso, para ello añadiremos las ontologías necesarias, como son las que necesitaremos en este proyecto Dublín Core y nuestras ontologías propias SemRAT y SemCyLRAT, por lo cual deberemos crear un prefijo para cada columna.

5. Una vez estructurado el código se van añadiendo las columnas conciliadas a los nuevos prefijos creados para que puedan ser almacenadas correctamente en las tripletas.
6. Por último, comprobamos la estructura del código y ya lo podremos exportar a RDF.

En los siguientes apartados se detallará en profundidad el proceso de transformación.

4.2 OpenRefine

Como ya se ha explicado anteriormente, OpenRefine es una de las herramientas más importantes que se han usado en este proyecto, ya que es necesaria para la transformación de los registros de actividades de tratamiento a RDF.

Para el proceso de transformación vamos a utilizar como base el registro de actividades de tratamiento el archivo Presidencia.xls ya que el proceso de transformación es el mismo para cada uno de los archivos obtenidos de la Junta de Castilla y León.

Para ello deberemos de crear un proyecto nuevo, en el que ordenemos las columnas y borremos si las hay, aquellas columnas que estén vacías, lo cual nos deberá quedar de una forma similar a eso.

The screenshot shows the OpenRefine web interface. At the top, the project name is 'RAT_Presidencia.xlsx'. Below the navigation bar, there is a table with 10 columns: TRATAMIENTO, CONSEJERÍA, CENTRO DIRECTIVO, RESPONSABLE DEL TRATAMIENTO, DELEGADO DE PROTECCIÓN DE DATOS, BASE JURÍDICA, FINES DEL TRATAMIENTO, CATEGORÍAS DE INTERESADOS, CATEGORÍAS DE DATOS PERSONALES, and CESIONES DE DATOS. The first row contains detailed information for a specific activity, such as 'CONSULTA DE LA DOCUMENTACIÓN DEL ARCHIVO CENTRAL DE LA CONSEJERÍA DE LA PRESIDENCIA DE LA JUNTA DE CASTILLA Y LEÓN'. Below the table, the 'Parse data as' panel is open, showing 'Excel files' selected and 'RAT_Presidencia.xlsx#PRE' as the source. The 'Update Preview' button is visible on the right.

Imagen 11. Creación del proyecto

Esto nos creará un proyecto con una tabla igual a la que tenemos en el archivo xls, sobre la cuál tendremos que ir añadiendo las modificaciones para convertirla correctamente en un archivo RDF.

Lo primero que tendremos que comenzar a añadir serán servicios de conciliación basados en SPARQL, esto es necesario para enlazar los recursos con Web Semánticas para que a la hora de hacer consultas poder relacionar los nodos.

Como podemos ver en esta imagen para añadir un servicio de conciliación debemos darles un nombre que nos sirva para identificar el servicio añadido, le especificamos la URL en la que el servicio hace las consultas basadas en virtuoso, y especificamos el tipo de sintaxis en el que se realizan las consultas, como ya hemos dicho todas estarán basadas en Virtuoso.

Add SPARQL-based reconciliation service

Name:
A human readable name

Endpoint details

Endpoint URL:

Graph URI:
Leave empty to use the default graph

Type: ▼
This determines syntax that will be used for search

Label properties

Select properties that are used to label resources in the endpoint. These properties will be used to match resources:

rdfs:label skos:prefLabel dcterms:title dc:title
 foaf:name
 Other...

OK Cancel

Imagen 12. Añadir una conciliación basada en virtuoso

Este proceso debe repetirse para añadir todas las bibliotecas digitales basadas en Webs Semánticas que nos interesen, para este proyecto he añadido las más comunes y mejores para archivos gubernamentales como es el caso de la junta de Castilla y León, que como se puede ver en la siguiente imagen son:

- **Wikidata:** Es una base de conocimiento abierta y gratuita que tanto humanos como máquinas pueden leer y editar. Wikidata viene por defecto en la herramienta OpenRefine para la conciliación de recursos y actúa como almacenamiento central para los datos estructurados de sus proyectos hermanos de Wikimedia, incluidos Wikipedia, Wikivoyage, Wiktionary, Wikisource y otros. Podemos utilizarla añadiendo el Endpoint <https://query.wikidata.org/>
- **DBpedia:** El proyecto DBpedia ha generado durante mucho tiempo información semántica a partir de la Wikipedia inglesa. Desde junio de 2011 el proceso de generación de información extrae información de Wikipedia en 15 de sus versiones (idiomas). Uno de ellos es el español. El comité de internacionalización de DBpedia ha asignado un sitio web y un SPARQL Endpoint para cada uno de estos idiomas. Como vemos en la imagen anterior se usa añadiendo el Endpoint <https://dbpedia.org/sparql>

- **Europeana:** Es la biblioteca digital europea, de acceso libre, cuyo prototipo comenzó a funcionar el 20 de noviembre de 2008 que reúne contribuciones ya digitalizadas de reconocidas instituciones culturales de los 27 estados miembros de la Unión Europea. Sus fondos incluyen libros, películas, pinturas, periódicos, archivos sonoros, mapas, manuscritos y otros archivos. Su Endpoint basado en virtuoso es <http://sparql.europeana.eu/>
- **Datos.gob:** Es una biblioteca digital que almacena vocabularios de datos que reutiliza la información pública del gobierno de España. <https://datos.gob.es/es/sparql>
- **Getty vocabularies Reconciliation service:** Los usuarios experimentados de la herramienta pueden conciliar sus datos cuando estén en OpenRefine con la URL: <https://services.getty.edu/vocab/reconcile/>

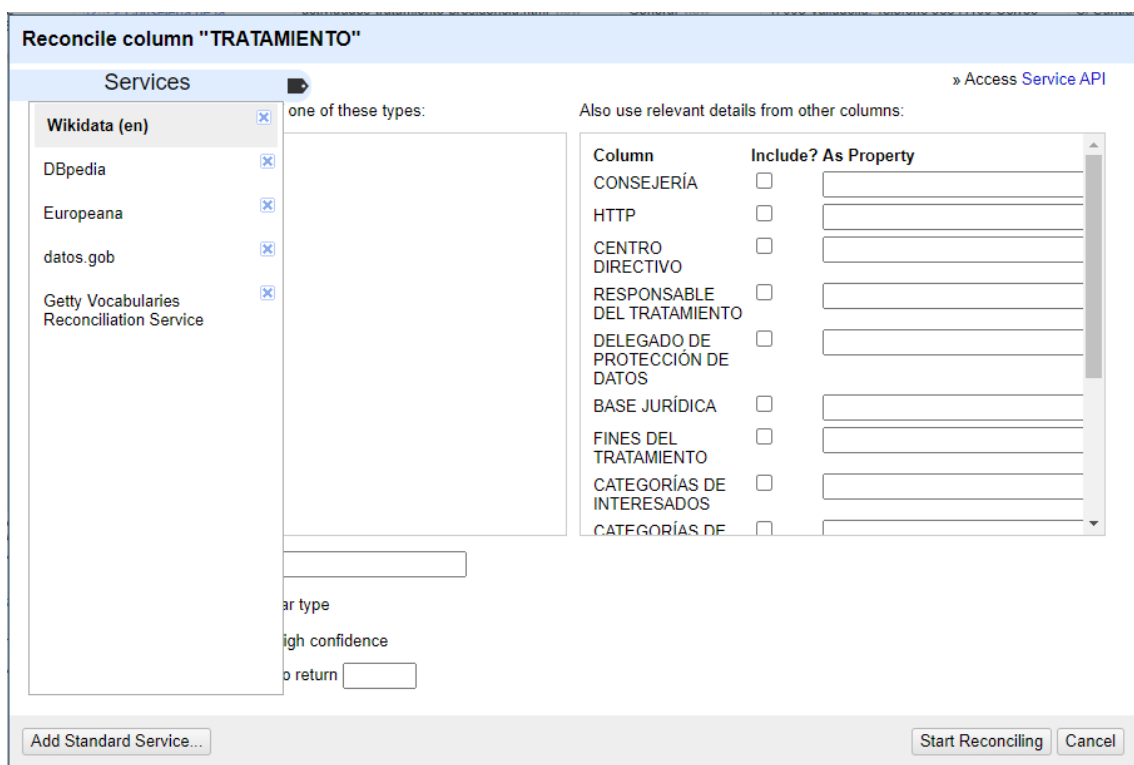


Imagen 13. Servicios de conciliación añadidos

Una vez elegido el servicio sobre el que queremos conciliar los recursos deberemos comprobar si existe un ítem que exprese a la perfección la conciliación con dicho recurso. Es decir, deberemos comprobar uno a uno todos los servicios con cada uno de los recursos para ver con cuál de ellos se ajusta más a su definición. En este caso podemos comprobar como el recurso Conserjería nos da dos opciones de conciliación [Consejería de la Presidencia de la Junta de Castilla y León\(100\)](#) y [Castilla y León \(Spain\). Consejería de Presidencia y Administración Territorial\(64\)](#)

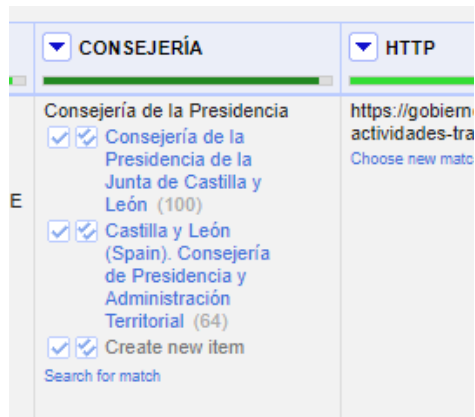


Imagen 14. Conciliación del recurso Conserjería

Para comprobar cuál de los dos ítems se asemeja más a nuestra definición deberemos comprobarlo, yendo al servicio de conciliación y leyendo la definición en este caso el servicio usado es Wikidata



Imagen 15. Definición del servicio de conciliación

En este caso la primera opción corresponde a la perfección con el recurso definido así que se le da el doble check, “Match all identical cells”, que tiene la opción que nos interesa para que concilie todos los recursos que tengan el mismo nombre con esta misma definición.

Pero no en todos los recursos se encuentra la definición que nos interesa, ya que este proyecto tiene recursos muy específicos que no están definidos en ninguna de las bibliotecas digitales que están añadidas, como podemos ver en el siguiente ejemplo de tratamiento.

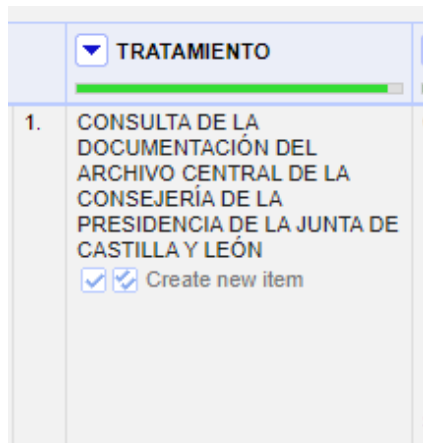


Imagen 16. Recurso tratamiento

En este caso cada uno de los recursos de tratamiento no tienen su semejante en ninguna de las bibliotecas anteriores. Por tanto, la solución que tendremos que realizar será crear un nuevo ítem específico para este proyecto y así poder conciliar los recursos con las ontologías.

Una vez conciliadas todos los recursos ya podremos editar el esqueleto RDF, que es editar el código transformado a RDF para la gestión y organización de las tripletas. Para ello deberemos cambiar la URI asignada por defecto a una con el nombre del archivo que estemos transformando, en este caso utilizaremos la URI `http://semcytrat/presidencia/` y además deberemos añadir los prefijos que nos hagan falta en función de las ontologías que sean necesarias para nuestro proyecto, en este caso deberemos añadir 3 prefijo. El primero será el prefijo Dublin Core, como ya hemos explicado anteriormente es útil para especificar en el predicado la URL de origen de la que obtenemos el archivo, este prefijo al ser una ontología existente simplemente añadimos `dc` y la URL se añadirá automáticamente. Como podemos observar en la imagen.

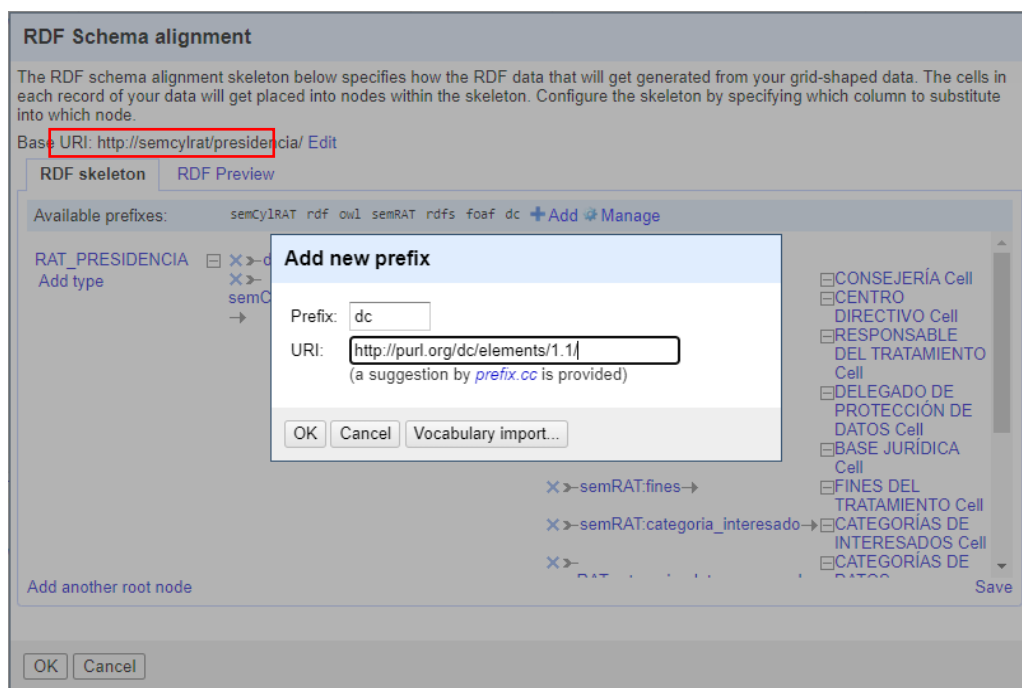


Imagen 17. URI y prefijos necesarios

Cómo explicábamos en el apartado ontologías propias, es necesario para este proyecto añadir dos prefijos creadas por nosotros mismos, el prefijo SemCyLRAT y SemRAT, para estipular el predicado de las tripletas a qué recurso se está haciendo referencia. El problema está en que estas ontologías no están relacionadas con ninguna URI existente, por tanto, deberemos gestionarlo como un prefijo exclusivo de este proyecto. Para ello añadiremos el nombre que le queremos dar al prefijo en este caso *SemCyLRAT* y *SemRAT* y como URI serán <http://semcykrat.edu/> y <http://semrat.edu/> respectivamente. En la siguiente imagen podemos ver la lista de prefijos por defecto más aquellos que hemos añadido para este proyecto.

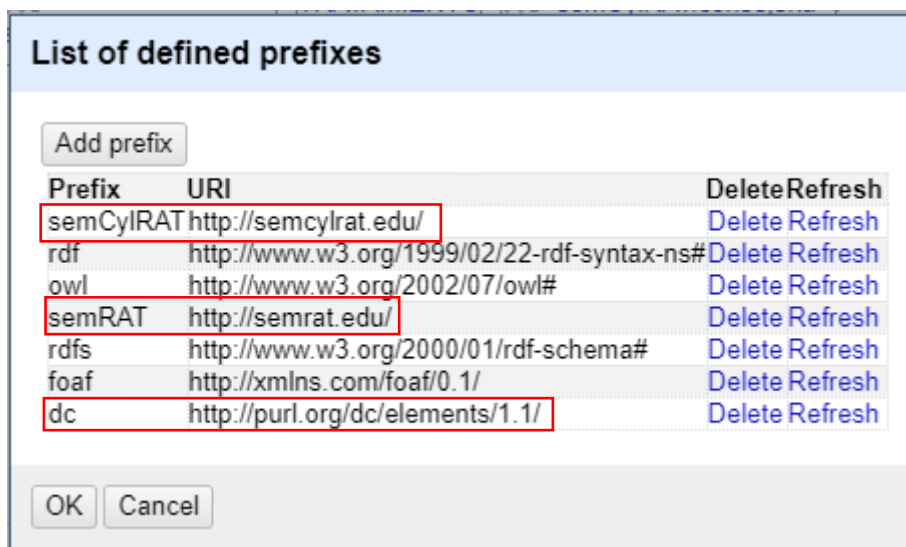


Imagen 18. Lista de prefijos para el proyecto

Ahora que ya tenemos las ontologías añadidas deberemos de organizar el diagrama de nuestra estructura RDF, este diagrama será la base en la que se sustentará la formación de las tripletas y los nodos del grafo. El diagrama comenzará siempre con un valor constante que hará referencia al archivo que estamos transformando seguido por dos nodos uno será el nodo que nos proporciona información de la ubicación de la que obtuvimos el archivo y el otro el nodo de cada uno de los tratamientos que irán ubicados en una lista no ordenada. De este último nodo TRATAMIENTO añadiremos un nodo por cada una de las columnas del archivo que nos proporcionarán información de la consejería, el centro directivo, el responsable, etc...

La construcción de este diagrama de nodos como se comenta anteriormente comenzará con un nodo raíz que consistirá en una constante con el nombre del archivo a transformar, es importante que la tipología de este nodo raíz sea una URI ya que nos permitirá añadir más nodos de manera secuencial que es lo que generará las tripletas.

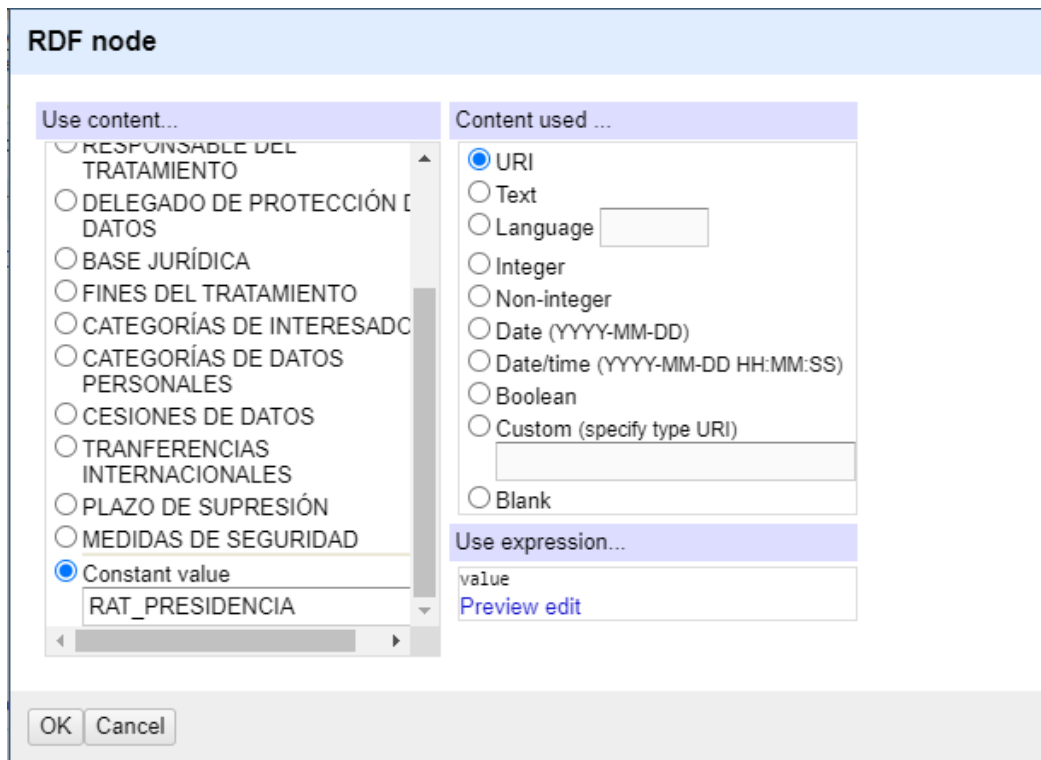


Imagen 19. Nodo raíz

Una vez añadido el nodo raíz deberemos añadir los prefijos de los recursos a los que haremos referencia en nuestra tripleta. En este caso serán las propiedades source del prefijo dc perteneciente a Dublin Core y tratamiento del prefijo SemCyLRAT.

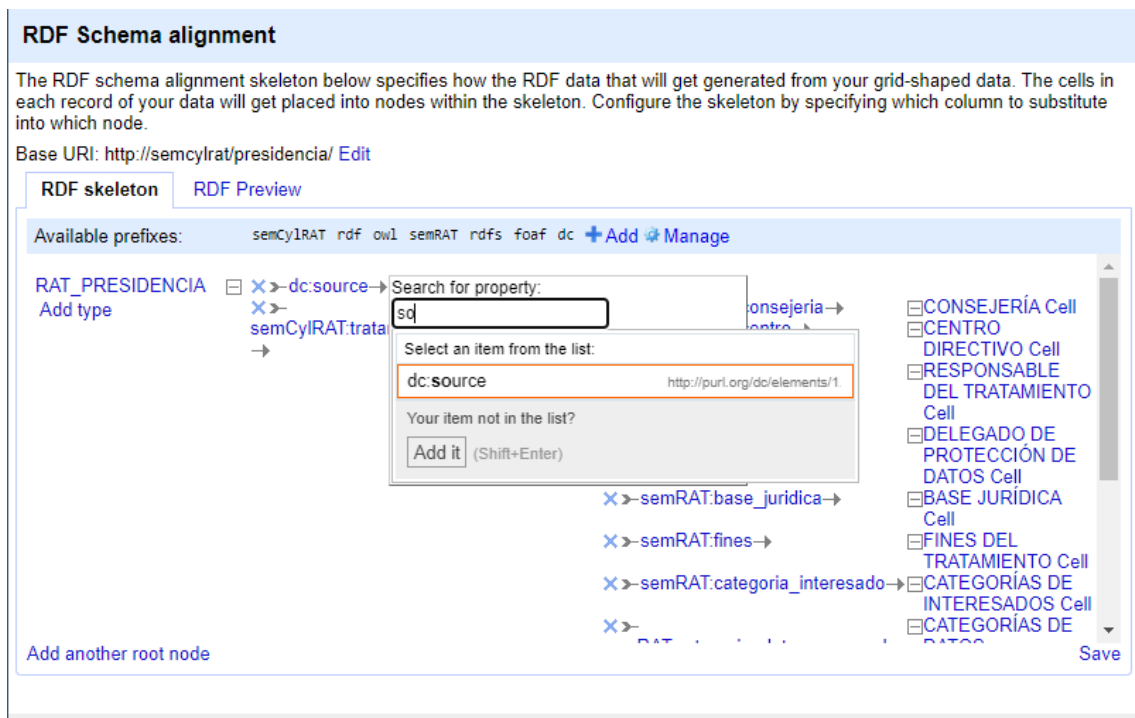


Imagen 20. Añadir propiedad

Tras definir la propiedad que vamos a añadir, añadimos el recurso que corresponda al prefijo indicado en este caso a dc:source añadiremos el recurso HTTP que contendrá la URL de origen de la que hemos obtenido el recurso, en este caso de tipo texto, ya que no añadiremos ningún otro recurso a este enlace. Y al prefijo SemCyLRAT:tratamiento añadiremos el recurso tratamiento, que si será de tipo URI, ya que a este objeto sí que tenemos que añadir más recursos, además deberemos estipular la tipología de este recurso que en este será rdf:bag porque nos interesa organizar cada tratamiento en un listado, para su posterior análisis de grafo y tripletas

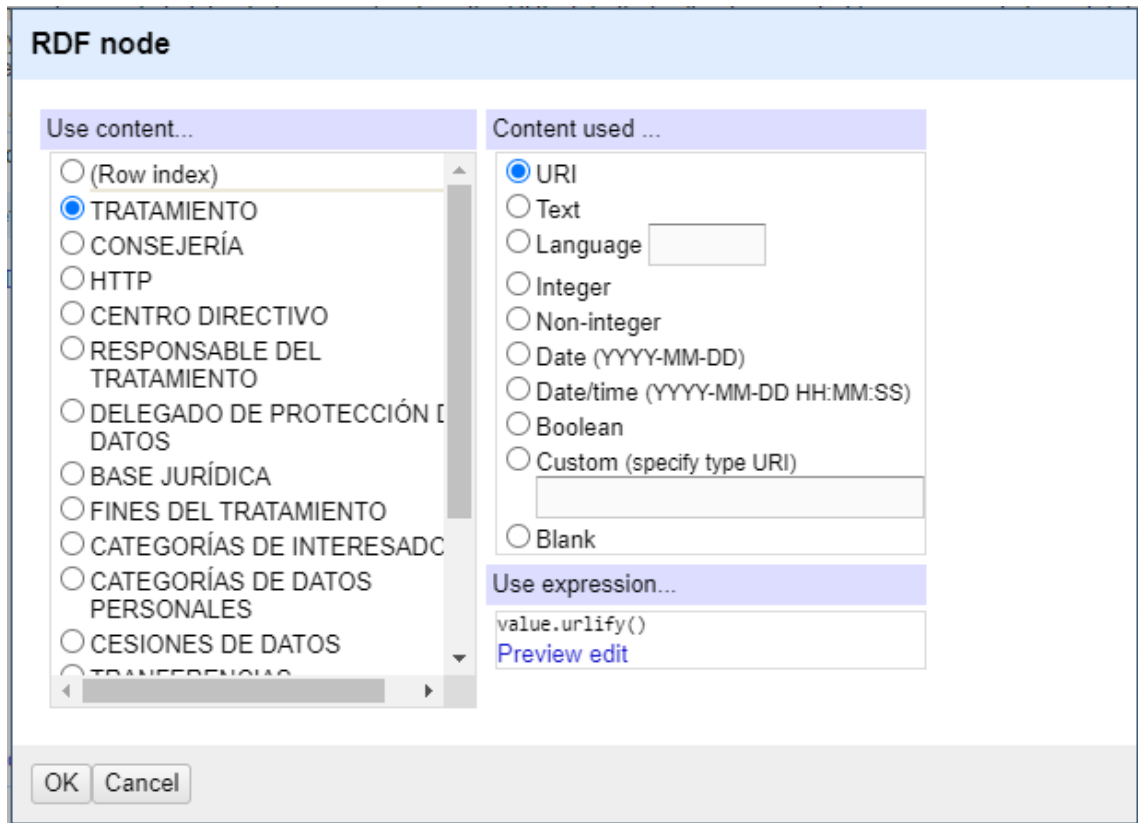


Imagen 21. Añadir el recurso tratamiento

Al ser el recurso TRATAMIENTO un nodo URI, tiene el mismo funcionamiento que el nodo raíz inicial por lo tanto a este nodo le podremos añadir el resto de los recursos que nos proporcionan información sobre cada TRATAMIENTO en particular, añadiendo para cada caso su prefijo y el recurso correspondiente, como podemos ver en la siguiente imagen.

Cada prefijo correspondiente a SemRAT y SemCyLRAT viene explicado en el apartado Ontologías propias.

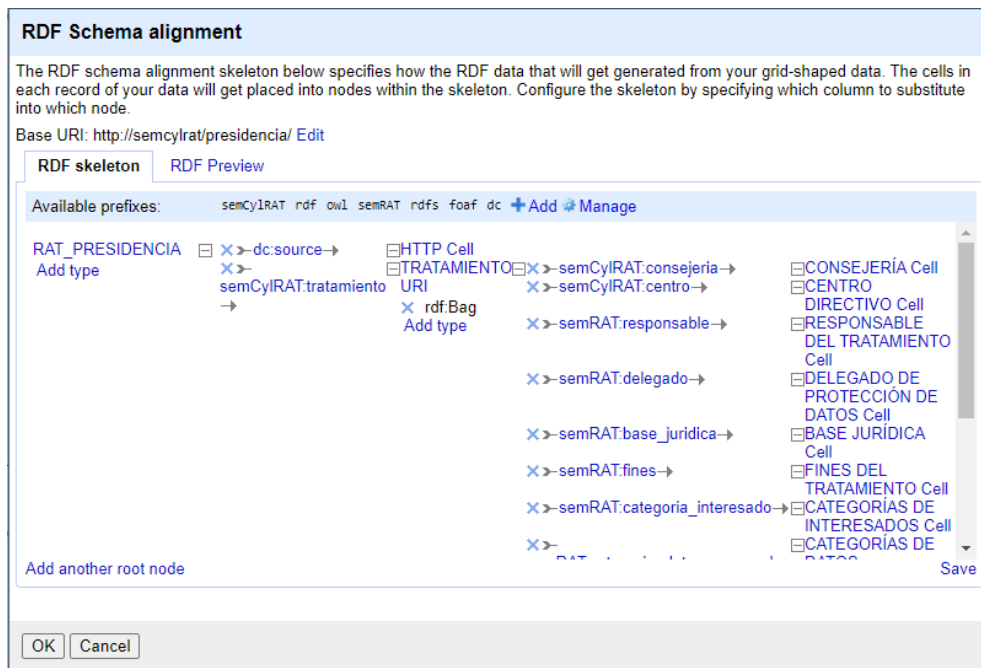


Imagen 22. Esqueleto RDF

Este proceso nos va generando automáticamente un código en lenguaje RDF que deberemos ir comprobando, si es el correcto antes de exportar el archivo final transformado. Para comprobarlo simplemente accedemos a la pestaña RDF Preview y ahí nos aparecerá por orden, los prefijos usados en este proyecto con la denominación @prefix y su URL, después el nodo raíz seguido de los prefijos source y tratamiento y los objetos a los que hace referencia, y luego cada uno de los objetos a los que hace referencia cada uno los tratamientos que completan este RDF complejo. Como se muestra en esta imagen.

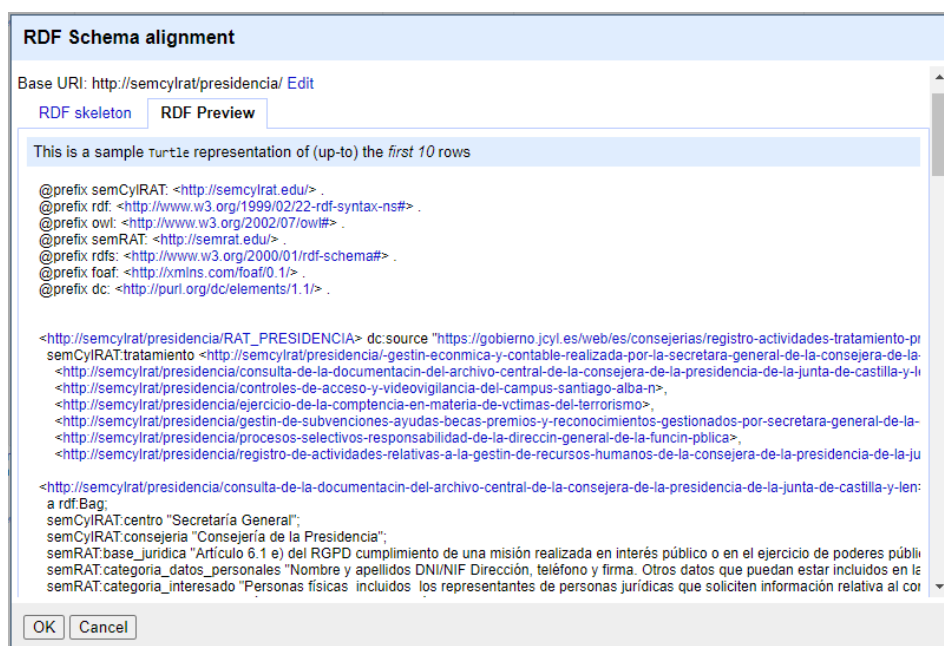


Imagen 23. Código RDF Presidencia

Una vez terminado este proceso de creación del diagrama y comprobación ya podremos exportar nuestro archivo a RDF. Podremos elegir si deseamos exportarlo como RDF/XML o como TTL, en este caso la conversión se hará a ambos formatos ya que no hay diferencias significativas, salvo el que TTL sea más fácil de leer y editar manualmente que XML.

Podemos ver en las siguientes imágenes la parte inicial del código transformado del Registro de Actividad de Tratamiento denominado presidencia (siendo el primero el referente al formato XML y el siguiente al formato TTL). Cabe destacar que en ambos códigos se comienza enumerando los prefijos usados, después la generación de la primera tripleta que hará referencia al nodo raíz y sus objetos URL origen y tratamientos, y después se listan los tratamientos de este RAT con cada uno de los recursos que proporcionan información al tratamiento.

```

RAT_Presidencia.rdf X
C:\Users> crhis > Desktop > UVA > 5º Prácticas y TFG > TFG > bbdd > Presidencia > RAT_Presidencia.rdf
1 <?xml version="1.0" encoding="UTF-8"?>
2 <rdf:RDF
3   xmlns:semCylRAT="http://semcylrat.edu/"
4   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
5   xmlns:owl="http://www.w3.org/2002/07/owl#"
6   xmlns:semRAT="http://semrat.edu/"
7   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
8   xmlns:foaf="http://xmlns.com/foaf/0.1/"
9   xmlns:dc="http://purl.org/dc/elements/1.1/"
10
11 <rdf:Description rdf:about="http://semcylrat/presidencia/RAT PRESIDENCIA">
12   <dc:source>https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-presidencia.html</dc:source>
13 </rdf:Description>
14
15 <rdf:Description rdf:about="http://semcylrat/presidencia/consulta-de-la-documentacin-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len">
16   <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag"/>
17   <semCylRAT:consejeria>Consejería de la Presidencia</semCylRAT:consejeria>
18   <semCylRAT:centro>Secretaría General</semCylRAT:centro>
19   <semRAT:responsable>SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcyl.es</semRAT:res
20   <semRAT:delegado>Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyl.es.</semRAT:delegado>
21   <semRAT:base_juridica>Artículo 6.1 e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1 c) del RGPD
22   <semRAT:finas>Atender la consulta libre y gratuita de los documentos integrantes del Patrimonio Documental de Castilla y León con fines de estudio e investigación o d
23   <semRAT:categoría_interesado>Personas físicas incluidos los representantes de personas jurídicas que soliciten información relativa al contenido de los documentos q
24   <semRAT:categoría_datos_personales>Nombre y apellidos DNI/NIF Dirección, teléfono y firma. Otros datos que puedan estar incluidos en la solicitud de consulta o acceso
25   <semRAT:cesiones_datos>No están previstas, salvo obligación legal.</semRAT:cesiones_datos>
26   <semRAT:transferencias_internacionales>No están previstas.</semRAT:transferencias_internacionales>
27   <semRAT:plazo_supresion>Los datos se conservarán durante el tiempo necesario para cumplir con la finalidad para la que se recabaron y determinar las posibles responsa
28   <semRAT:medidas_seguridad>Las medidas de seguridad implantadas se corresponden con las previstas en el Anexo II (Medidas de Seguridad) del Real Decreto 3/2010, de 8 d
29 </rdf:Description>
30
31 <rdf:Description rdf:about="http://semcylrat/presidencia/RAT PRESIDENCIA">
32   <semCylRAT:tratamiento rdf:resource="http://semcylrat/presidencia/consulta-de-la-documentacin-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-cas
33 </rdf:Description>
34
35 <rdf:Description rdf:about="http://semcylrat/presidencia/-gestin-economica-y-contable-realizada-por-la-secretaria-general-de-la-consejera-de-la-presidencia-y-a-la-tramitaci
36   <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag"/>
37   <semCylRAT:consejeria>Consejería de la Presidencia</semCylRAT:consejeria>
38   <semCylRAT:centro>Secretaría General</semCylRAT:centro>

```

Imagen 24. RAT presidencia exportado como XML


```
RAT_Presidencia.rdf  RAT_Presidencia.ttl X
C: > Users > chris > Desktop > UVa > 5º Prácticas y TFG > TFG > bbdd > Presidencia > RAT_Presidencia.ttl
1 @prefix dc11: <http://purl.org/dc/elements/1.1/> .
2 @prefix ns0: <http://semcyrat.edu/> .
3 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
4 @prefix ns1: <http://semrat.edu/> .
5
6 <http://semcyrat/presidencia/RAT_PRESIDENCIA>
7   dc11:source "https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-presidencia.html" ;
8   ns0:tratamiento <http://semcyrat/presidencia/consulta-de-la-documentacin-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>, <http://
9
10  <http://semcyrat/presidencia/consulta-de-la-documentacin-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>
11   a rdf:Bag ;
12   ns0:consejeria "Consejería de la Presidencia" ;
13   ns0:centro "Secretaría General" ;
14   ns1:responsable "SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcyl.es" ;
15   ns1:delegado "Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyl.es." ;
16   ns1:base_juridica "Artículo 6.1 e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1 c) del RGPD cum
17   ns1:finas "Atender la consulta libre y gratuita de los documentos integrantes del Patrimonio Documental de Castilla y León con fines de estudio e investigación o de inf
18   ns1:categoria_interesado "Personas físicas incluidos los representantes de personas jurídicas que soliciten información relativa al contenido de los documentos que cu
19   ns1:categoria_datos_personales "Nombre y apellidos DNI/NIF Dirección, teléfono y firma. Otros datos que puedan estar incluidos en la solicitud de consulta o acceso" ;
20   ns1:cesiones_datos "No están previstas, salvo obligación legal." ;
21   ns1:transferencias_internacionales "No están previstas. " ;
22   ns1:plazo_supresion "Los datos se conservarán durante el tiempo necesario para cumplir con la finalidad para la que se recabaron y determinar las posibles responsabilid
23   ns1:medidas_seguridad "Las medidas de seguridad implantadas se corresponden con las previstas en el Anexo II (Medidas de Seguridad) del Real Decreto 3/2010, de 8 de ene
24
25  <http://semcyrat/presidencia/-gestin-economica-y-contable-realizada-por-la-secretara-general-de-la-consejera-de-la-presidencia-y-a-la-tramitacin-de-los-expedientes-de-con
26   a rdf:Bag ;
27   ns0:consejeria "Consejería de la Presidencia" ;
28   ns0:centro "Secretaría General" ;
29   ns1:responsable "SECRETARÍA GENERAL. C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100. Correo electrónico: protecciondatos.sg.presidencia@jcyl.es" ;
30   ns1:delegado "Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcyl.es." ;
31   ns1:base_juridica "Artículo 6.1 e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1 c) del RGPD cum
32   ns1:finas "La tramitación de expedientes de gasto y de documentos contables en materias propias de la Consejería de la Presidencia. La tramitación de la contratación pú
33   ns1:categoria_interesado "Personas físicas incluidos los representantes de las personas jurídicas que participan, y en su caso resultan adjudicatarios, en los procedim
34   ns1:categoria_datos_personales "Relativos a la gestión económica y contable: Nombre y apellidos |DNI/NIF|Domicilio|Cuenta corriente|Firma electrónica. Relativos a la tr
35   ns1:cesiones_datos "Los datos básicos del adjudicatario se publicarán en: Diarios oficiales. Plataforma de contratación del sector público. Plataforma Duero. Portal de
36   ns1:transferencias_internacionales "No están previstas. " ;
37   ns1:plazo_supresion "Los datos se conservarán durante el tiempo necesario para cumplir con la finalidad para la que se recabaron y determinar las posibles responsabilid
38   ns1:medidas_seguridad "Las medidas de seguridad implantadas se corresponden con las previstas en el Anexo II (Medidas de Seguridad) del Real Decreto 3/2010, de 8 de ene
39
40  16.1 col.1 - Enlaces 2 - UTE-8 - CRIE - TUB
```

Imagen 25. RAT presidencia exportado como TTL

4.3 Comprobación funcionamiento

Una de las partes más importantes del proceso de transformación es la comprobación de que en ambos códigos, tanto en el que tenga la extensión XML cómo la extensión TTL, el lenguaje sea el correcto, porque si no podremos comprometer el resultado de nuestras tripletas o de nuestro grafo.

Debido a la cantidad ingente de información que obtenemos en nuestro código sería muy tedioso y prácticamente imposible validar manualmente que el código generado sea correcto. Para ello W3C nos proporciona dos páginas en las que podremos validar de forma online nuestro código:

La primera de ellas se trata de <https://www.w3.org/2015/03/ShExValidata/> es una página de validación online proporcionada por W3C para la sintaxis de archivos de extensión TTL, TriG, N-Triples or N-Quads. Simplemente importamos nuestro archivo TTL y le damos a validar, la propia página se encargará de comprobar el esquema, los datos y validará si nuestro archivo es válido proporcionando información de la cantidad de errores, *warnings* y *matches*. Si alguno de los dos primeros da un resultado distinto de 0 el código no será válido. Tal y como se ve en la imagen.

The screenshot displays the W3C RDF Validator interface. On the left, a sidebar contains a numbered list of steps: 1. Select Schema, 2. Input Data, 3. Configure Options, and 4. Validation Results. Below this is a 'Load Demo Data' section with a 'ChEMBL 2015 Demo' button. The main content area is divided into several sections: 'Upload Data File' where 'RAT_Presidencia.ttl' is selected; 'Directly Input Data' showing a code snippet; 'Configure Options' where the resource is set to 'http://rdf.ebi.ac.uk/chembl/chembl' and the shape to '<SummaryLevelShape>', with a 'Closed World validation' checkbox checked; and 'Validation Results' which reports 'Errors: 0', 'Warnings: 0', and 'Matches: 23'. A red box highlights the 'Validation Results' section.

Imagen 26. Validación archivo Presidencia.ttl

Para comprobar el archivo transformado en XML, también W3C nos proporciona una página más completa que la anterior que es <https://www.w3.org/RDF/Validator/>. Es una página más completa y bastante más conocida por el entorno RAT ya que nos proporciona tanto como la validación de que nuestro código es correcto, tanto como las tripletas y su grafo, siempre que este último no exceda un tamaño concreto debida a la capacidad de memoria de la propia herramienta.

Para su uso simplemente deberemos copiar nuestro código XML, en esta comprobación utilizamos el código transformado Presidencia.xml, lo incluimos en el validador y elegimos si queremos obtener tripletas y grafo o simplemente una de las dos, como se muestra en la imagen.

Validation Service

[Skip Navigation](#) [Home](#)
[Documentation](#)
[Feedback](#)

Check and Visualize your RDF documents

[oldie serviet](#)

Enter a URI or paste an RDF/XML document into the text field above. A 3-tuple (triple) representation of the corresponding data model as well as an optional graphical visualization of the data model will be displayed.

Check by Direct Input

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:sencylrat="http://sencylrat.edu/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:semrat="http://semrat.edu/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Display Result Options:
Triples and/or Graph:
Graph format:

Paste an RDF/XML document into the following text field to have it checked. More options are available in the [Extended interface](#).

Check by URI

Display Result Options:
Triples and/or Graph:
Graph format:

Enter the URI for the RDF/XML document you would like to check. More options are available in the [Extended interface](#).

Last modified \$Date: 2006/02/28 16:30:38 \$
[Eric Prud'hommeaux](#)

Imagen 27. Interfaz W3C Validator

Una vez pegado el texto le podemos dar a “*Parse RDF*” y la herramienta nos dará la información sobre si el documento es correcto y mostrará un listado de las tripletas generadas y su grafo. En las siguientes imágenes se puede comprobar como el archivo Presidencia sí que es correcto y se verá una pequeña parte de las tripletas y su grafo (ya que es demasiado grande para incluir en estas imágenes).

Jump To:

- Source
- Triples
- Messages
- Graph
- Feedback
- Back to Validator Input

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://purl.org/dc/elements/1.1/source	"https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-presidencia.html"
2	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
3	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/consejeria	"Consejería de la Presidencia"
4	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/centro	"Secretaría General"
5	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/responsable	"SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcy.es"
6	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/delgado	"Consejería de la Presidencia C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo de contacto: dpd.presidencia@jcy.es."
7	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/base_juridica	"Artículo 6.1 e) del RGPD cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos. Artículo 6.1 c) del RGPD cumplimiento de una obligación legal. Ley 6/1991, de 19 de abril, de Archivo y del Patrimonio Documental de Castilla y León. Ley 39/2015, de 1 de octubre, del procedimiento administrativo común de las Administraciones públicas. Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público. Ley 16/1985, de 25 de junio, del Patrimonio Histórico Español."
8	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/finis	"Atender la consulta libre y gratuita de los documentos integrantes del Patrimonio Documental de Castilla y León con fines de estudio e investigación o de información para la defensa de sus derechos o el conocimiento de sus obligaciones. Los datos personales podrán ser tratados para fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos de acuerdo con lo establecido en los artículos 5.1 b) y 5.1 del RGPD"
9	http://semovirat.presidencia.es/registro-actividades-tratamiento-presidencia.html	http://semovirat.edu/categoria_interesado	"Personas físicas incluidos los representantes de personas jurídicas que solicitan información relativa al contenido de los documentos que custodia el Archivo Central de la Consejería de la Presidencia, por vía telefónica, correo (electrónico u ordinario) y de forma presencial en la sede de la calle Santiago Alba, 1 de Valladolid."

Imagen 28. Validación y tripletas del documento Presidencia.xml

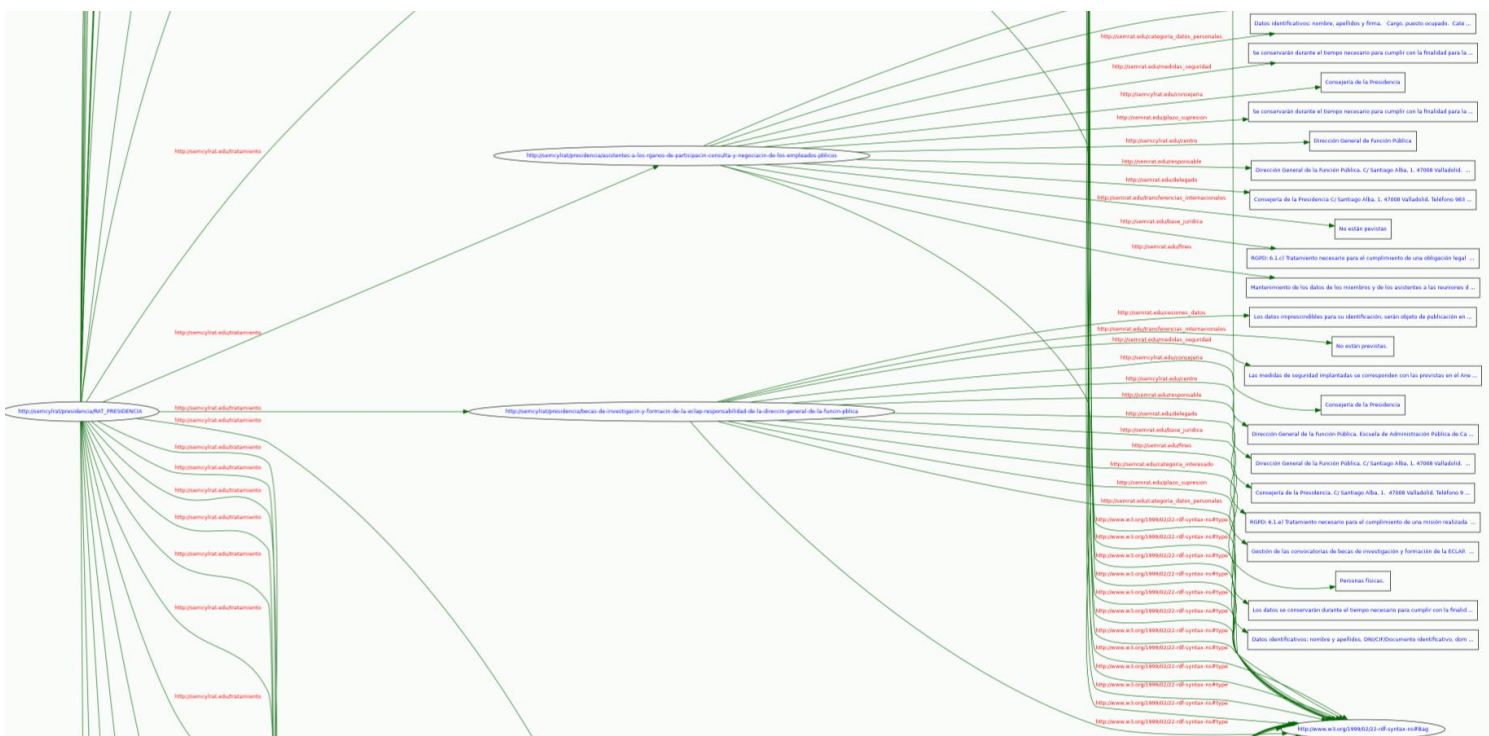


Imagen 29. Parte del grafo Presidencia.xml

Este proceso deberemos repetirlo para cada uno de los documentos generados para ver si no hemos tenido ningún problema a la hora de la transformación a RDF. Una vez ya comprobada la validación podremos comenzar a generar un archivo que nos muestre las tripletas y otro que nos genere el grafo mediante AllegroGraph, explicado en el siguiente punto. Cada uno de los archivos generados, así como las tripletas y sus grafos vendrán completas en los archivos anexos a la documentación.

4.4 Generación de grafos

Como ya hemos explicado anteriormente AllegroGraph es una poderosa herramienta para el análisis de archivos RDF, ya que nos permite realizar consultas en lenguaje SPARQL, generación de tripletas y generación y edición del grafo.

Una vez ya hemos comprobado que el código es válido (como se explica en el apartado anterior), deberemos de generar los grafos y las tripletas correspondientes a los archivos transformados para poder ver su estructura y generar las consultas.

Para ello, lo primero que deberemos generar en la herramienta AllegroGraph será un repositorio con el nombre del archivo transformado en el que importaremos dicho archivo, se podrá importar tanto el documento de extensión XML como de extensión TTL. En este caso, importaremos el archivo XML, ya que tanto las tripletas como el grafo resultante que se generará será exactamente el mismo, especificamos la base URI en la que estamos trabajando, tal y como hemos especificado en la herramienta OpenRefine y aceptamos como se ve en la imagen.

The screenshot shows the AllegroGraph WebView 7.1.0 interface. At the top, it says 'repository Presidencia'. Below that is a navigation bar with 'Repository | Queries | Admin | User anonymous'. The main area is titled 'Select one or more local files to import'. It contains several fields and options: 'File:' with a 'Browse...' button and the text 'presidencia'; 'Format:' with a dropdown menu set to 'RDF/XML' and a note about auto-detecting formats; 'Context:' with an empty text box and '(leave blank to use the default graph)'; 'Base URI:' with the text 'http://semcylrat/presidencia/'; 'Attributes:' with an empty text box; 'Relax syntax:' with an unchecked checkbox; 'External references:' with an unchecked checkbox and 'External reference timeout:' with a spinner; 'Store source:' with an unchecked checkbox and 'Suppress duplicate embedded triples (for Turtle*/TriG*):' with a checked checkbox; 'File loading mode:' with 'Transactional' selected (radio button) and 'Using agload' as an option; 'Agload threads:' with a spinner set to 2 and '(blank or 0 = default)'; 'Error handling:' with 'Cancel load' selected (radio button) and 'Ignore errors' as an option; and 'Use bulk load mode:' with an unchecked checkbox. At the bottom, there is a note: 'Uploading files larger than 100 MB using a browser may cause performance issues. We suggest using server-side file loading mode to efficiently import larger datasets.'

Imagen 30. Importación del archivo Presidencia.xml

Todo ello nos generara que nuestro repositorio Presidencia se completara con los statements que contenga este archivo. En este caso, este documento consta transformado de 474 *staments*, y desde aquí podremos añadir o eliminar algún statement que nos interese o que nos falte, explorar dicho repositorio así como ver sus tripletas, generar consultas y explorar y organizar el grafo.

AllegroGraph WebView 7.1.0 repository Presidencia

< | Repository | Queries | Admin | User anonymous

Repository Presidencia — 474 statements

[edit description]

Load and Delete Data

- Add a statement
- Delete statements
- Import RDF:
 - from an uploaded file
 - from a server-side file
 - from a text area input

Explore the Repository

- View triples
- View quads
- View repository's classes
- View repository's predicates
- View repository's named graphs
- Explore repository in Gruff

Replication

- Multi-master: convert store to a replication instance ?
- Warm standby: control replication ?

Repository Control

- Export repository as
- Start a session — support transactions and Prolog functors
- Warmup store
- Back-up this repository
- Export duplicate statements
- Delete duplicate statements
- Suppress duplicate statements false

Tools

- Explore Presidencia using Gruff

Reports

- Storage report
- Triple indices
- String table
- Full list of reports ...

Imagen 31. Interfaz AllegroGraph

Para ver las tripletas que nos genera este documento sabemos que son 474 statements de los cuales obtenemos la información que son 365 statements de tripletas más las 111 que añaden de información de manera automática en cualquier archivo AllegroGraph. Nosotros nos centraremos exclusivamente en las 365 que hemos creado nosotros en la transformación de nuestro archivo RAT Presidencia. Con este fin, borraremos estas tripletas sobrantes mediante *delete statements*. Una vez borrados nos dirigimos a la opción *view triples* y realizamos la siguiente consulta en SPARQL

```
SELECT ?s ?p ?o { ?s ?p ?o . }
```

De esta manera obtendremos todas las tripletas, independientemente de cuál sea su sujeto, predicado u objeto.

Edit query

```

1 # View triples
2 SELECT ?s ?p ?o { ?s ?p ?o . }
    
```

Language: SPARQL ▾
 Limit to 1000 results
 Reasoning
 Long parts
 Cancel on warnings
 Use MJQE ?

[Permalink to query](#)

365 Results in 5.595 ms

s	p	o
<http://semcytrat.presidencia/RAT PRESIDENCIA>	<http://purl.org/dc/elements/1.1/source>	"https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-presidencia.html"
<http://semcytrat.presidencia/consulta-de-la-documentacion-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>	<http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag>
<http://semcytrat.presidencia/consulta-de-la-documentacion-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>	<http://semcytrat.edu/consejeria>	"Consejería de la Presidencia"
<http://semcytrat.presidencia/consulta-de-la-documentacion-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>	<http://semcytrat.edu/centro>	"Secretaría General"
<http://semcytrat.presidencia/consulta-de-la-documentacion-del-archivo-central-de-la-consejera-de-la-presidencia-de-la-junta-de-castilla-y-len>	<http://semrat.edu/responsable>	"SECRETARÍA GENERAL C/ Santiago Alba, 1. 47008 Valladolid. Teléfono 98341100 Correo electrónico: protecciondatos.sg.presidencia@jcyl.es"

Imagen 32. Generación de tripletas

Una vez generadas las tripletas y comprobado el funcionamiento de las consultas, revisaremos el grafo resultante que nos genera AllegroGraph. Lo realizaremos pulsando en cualquiera de las opciones Gruff para la visualización del grafo, lo cual nos habilitará una pestaña nueva con una interfaz vacía. Para poder visualizar el grafo habilitamos la opción *Display Some Sample Triples*

Gruff 8.0.4 on AllegroGraph 7.1.0 Presidencia read / write 365 triples server 127.0.0.1:10035 (main port)

File	View	Text Search	Display	Link	Remove	Layout	Select	Edit	Global Options	Visual Graph Options	Help	
			Display a Node by URI or Literal ...							Shift+U		
			Display a Node by Label ...							Shift+L		
			Display a Recently Selected Node ...							Shift+R		
			Display an Instance Node by Type ...							Shift+J		
			Display a Sample Instance Node by Type ...							Ctrl+J		
			Display a Type from All Types ...							Ctrl+Shift+J		
			Display an Instance Node by Class Hierarchy ...							Shift+K	d session with its own port was requested when -store, but it could not be used. Requests are oxiex through the main port of the AllegroGraph it affect performance. See "A Note About Dedic Firewalls" in the Gruff documentation.)	
			Display a Sample Instance Node by Class Hierarchy ...							Ctrl+K		
			Display a Class Node by Class Hierarchy ...							Ctrl+Shift+K		
			Display Some Sample Triples							Ctrl+Alt+A		
			Display Triples of One Graph ...							Alt+Shift+A	ph view, use a command on the Display menu (c the Text Search menu) to display a starter node. Then displav nod	

Imagen 33. Generar el grafo Presidencia en AllegroGraph

Lo que nos generará algunos nodos de nuestro documento, pero no el grafo al completo. Solo los más importantes, que nos sirve para visualizar parte del grafo y comprender la leyenda que usa AllegroGraph, donde distinguiremos los tipos de nodos generados y los enlaces entre nodos que harán referencia a lo que en las tripletas se denomina predicado.

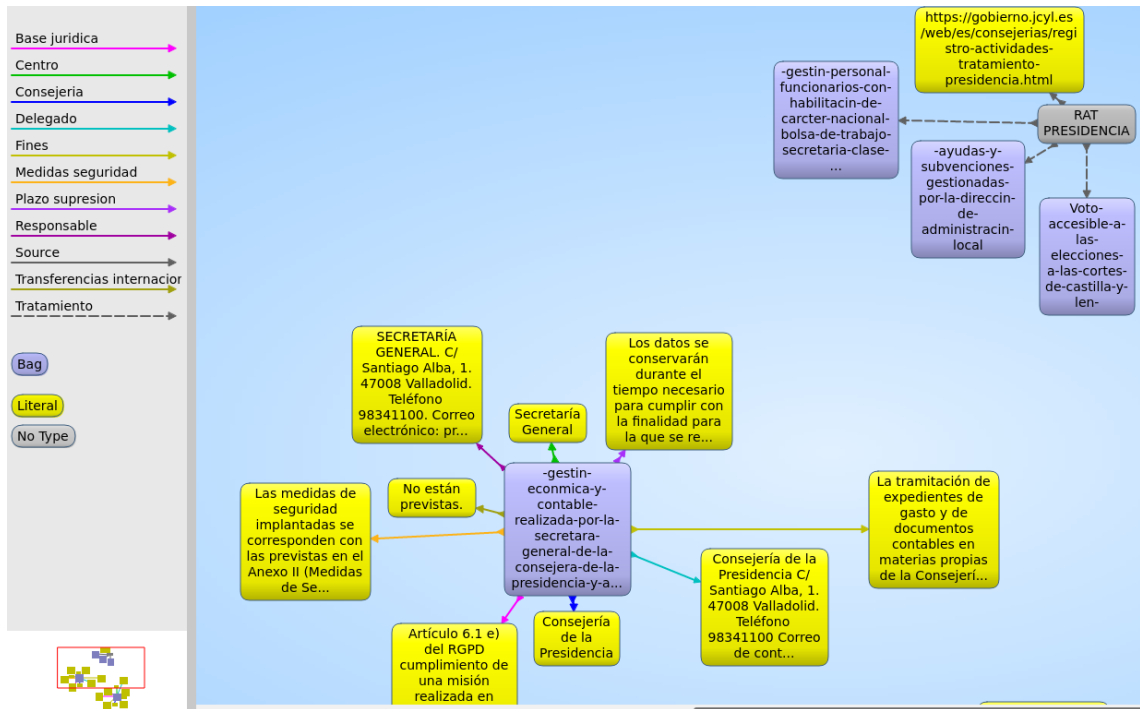


Imagen 34. Muestra de la leyenda del Grafo Presidencia

Una vez que hayamos comprobado que la leyenda es correcta lo que tendremos que generar el grafo al completo, para ello tendremos que seleccionar el Nodo Raíz, que en nuestro caso será *RAT PRESIDENCIA* y seleccionar todos los predicados que nacen de dicho nodo, esto nos generara nuestro grafo con todos los nodos y sus predicados, tal y como se estipula en las tripletas.

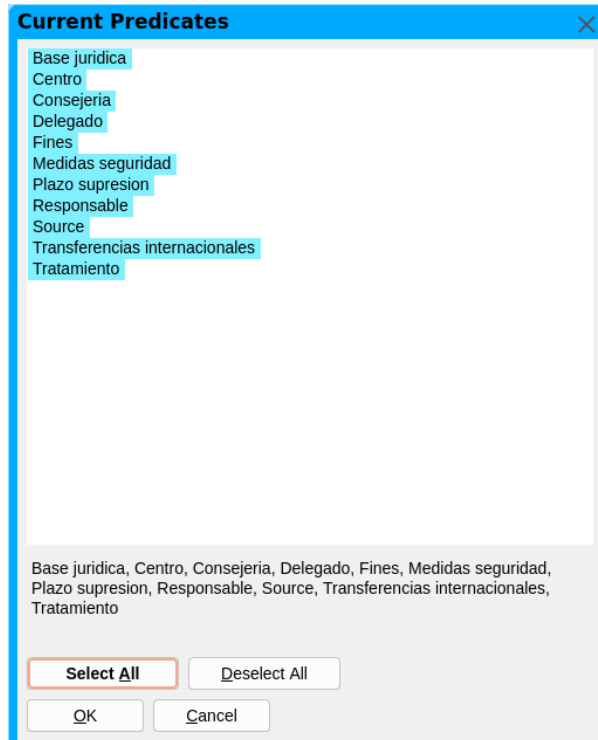


Imagen 35. Selección de todos los predicados del RAT Presidencia

Una vez generados todos los nodos en nuestro grafo, le aplicamos un “layout” para establecer una jerarquía que haga nuestro grafo más legible. De esta manera, yo he preferido que lo primero que se genere sea el Nodo Raíz *RAT PRESIDENCIA* del que salen el nodo URL origen, y cada uno de los nodos tratamiento que organizamos anteriormente en un bag, listado no ordenado, y para su visualización, de uno de estos tratamientos salen los nodos que nos proporcionan información con respecto al centro la base jurídica, etc...

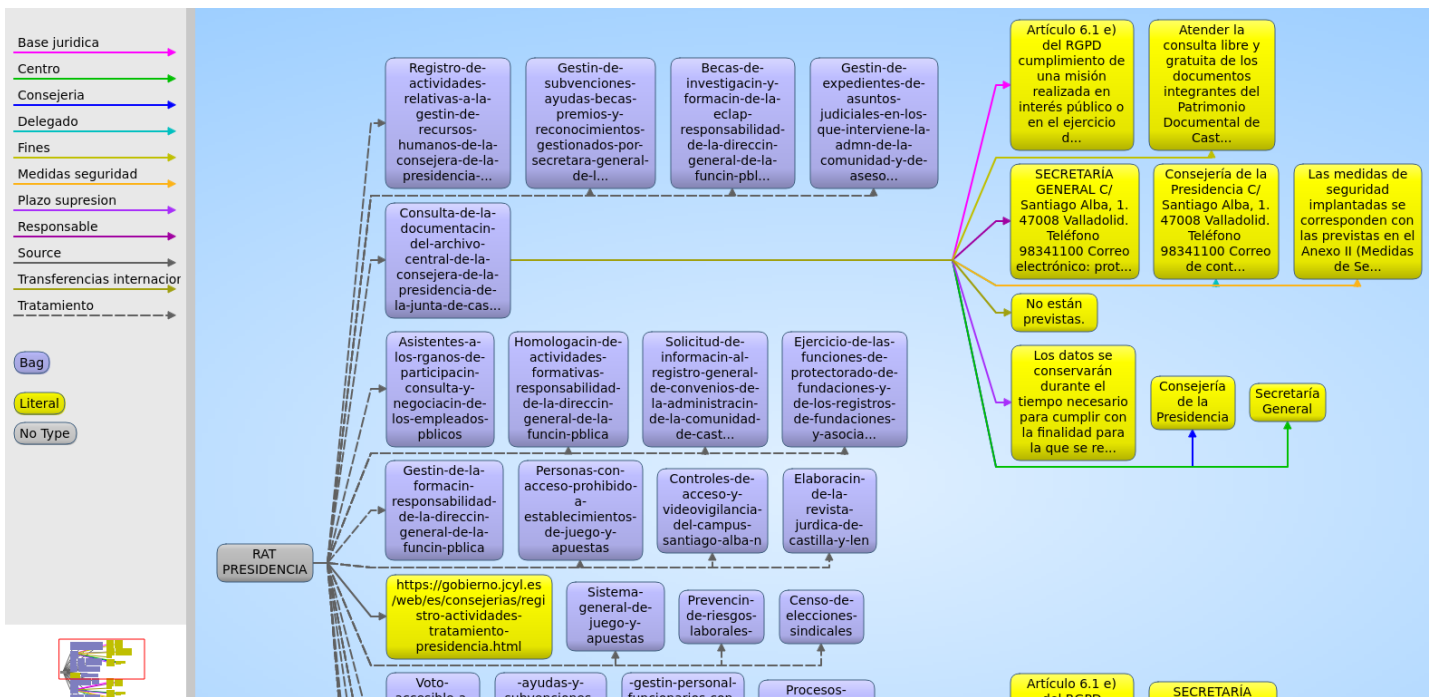


Imagen 36. Grafo Presidencia completo y organizado

Una vez generados y exportados estos documentos, de cada uno de los Registros de Actividades de Tratamiento proporcionados por la Junta de Castilla y León, tanto los archivos transformados, como las tripletas y el grafo, se podrá dar por concluido el proyecto. Los grafos completos se mostrarán en los archivos anexos a la documentación.

5. Conclusiones

Como punto de partida para este trabajo se intenta aproximar al lector al dominio en el que se centra este Trabajo de fin de grado: familiarización con el entorno RDF, ventajas de almacenar datos en RDF en lugar de las bases de datos convencionales y transformación de los archivos.

A lo largo de todo el documento que acredita y expone la investigación realizada, se ha profundizado en las distintas áreas que se abarcan para dotar al trabajo de una base sólida y fundamentada. De esta forma se han podido definir unas hipótesis adecuadas e interesantes que han guiado la totalidad del proceso hacia la obtención de unos resultados relevantes.

El apartado de *Tecnologías usadas* ha clarificado diferentes aspectos sobre las tecnologías y conceptos utilizados en la investigación, lo que ha permitido fijar un punto de partida sobre el que iniciar las aportaciones del trabajo.

En el desarrollo de este proyecto he podido poner en práctica conocimientos técnicos del funcionamiento de las bases de datos aprendidos en diferentes asignaturas cursadas, como son TDBD (*Tecnología y Diseño de Bases de Datos*), ABD (*Análisis y Diseños de Bases de Datos*) y PGPI (*Planificación y Gestión de Plataformas Informáticas*).

A parte de poner en práctica dichos conocimientos, ha sido necesario realizar un proceso de aprendizaje de técnicas y herramientas utilizadas en el proyecto, debido a que en ninguna de esas asignaturas se enseña RDF. Este aprendizaje se desarrolló mediante el libro "*A Developer's Guide to the Semantic web*" artículos en castellano proporcionados por Mercedes, mi tutora, y otros encontrados en internet, en los que incluyo videos explicativos de plataformas como YouTube, Xakata y Udemey.

Otro de los conocimientos a poner a prueba fue la gestión del proyecto y ser consciente de la importancia de una buena planificación para poder realizar todas las tareas dentro de una situación de fechas previamente establecida con entrega final. Teniendo en cuenta el enfoque que se le iba a dar (*Cascada*), como se explica en el apartado Planificación, se han establecido las pertinentes etapas en las que se iba a subdividir el proyecto.

La solución planteada para la mejora de los datos almacenados por la Junta de Castilla y León se basa en la utilización de tecnologías de web semánticas, ya que resultan más adecuadas para aprovechar al máximo la información recogida de usuarios y la intención de almacenarlo en aplicaciones de entorno de Cloud Computing. Dentro de la Web Semántica, las ontologías cobran especial importancia para el desarrollo del trabajo. Gracias a estas herramientas usadas en la transformación de los documentos a RDF se ha podido almacenar la información procedente de los archivos y establecer una estructura en ellos que han hecho que esos datos se transformen en recursos utilizables. Estas ontologías sirven como bases del lenguaje para identificar los recursos dentro de la web semántica.

Cabe destacar que la intención de este proyecto es crear un sistema de consulta de documentos de Registro de Actividades de Tratamiento para la Junta de Castilla y León, ya que no supondría un elevado coste de desarrollo porque simplemente tendrían que conseguir una nube que soporte datos en RDF, como es el caso de Azure. Y esto es algo que el Gobierno de España ya realiza en páginas como <https://datos.gob.es/> usada como iniciativa de datos abiertos del Gobierno de España para proyectos realizados por

personas o comunidades. Y gracias al diseño planteado con sus propias ontologías se potencia la usabilidad y escalabilidad, ya que no se depende de prefijos ya creados para hacer referencia a los recursos, lo que supondría un problema a la hora de modificar o añadir nuevos, con las ontologías propias se crean nuevos prefijos que hagan referencia a estos nuevos recursos a añadir y se incluyen al archivo RDF ya creado.

Para finalizar, este trabajo de fin de grado me ha supuesto un aprendizaje al modelo de almacenamiento de metadatos RDF, un modelo totalmente desconocido para mí, pero con una cantidad inmensa de funcionalidades para proyectos con grandes cantidades de datos. He puesto en práctica los conocimientos teóricos adquiridos en las diferentes asignaturas que he cursado y he aprendido nuevos sistemas de almacenamiento y consulta en lo que a bases de datos de refiere.

5.1 Objetivos Alcanzados

Tras la finalización de este proyecto, se puede asegurar que se han cumplidos los siguientes objetivos propuestos:

- Estudio pormenorizado de RDF y sus ventajas frente a las bases de datos convencionales.
- Transformación de los archivos obtenidos de la Junta de Castilla y León de formato Excel a RDF para su posterior almacenamiento en nubes que soporten bases de datos basadas en RDF.
- Creación de ontologías propias con sus prefijos específicos para la identificación de los recursos.
- Realización, gestión y comprobación de las tripletas generadas basadas en una estructura RDF compleja, tal y como se explica en los puntos anteriores.
- Realización de grafos de cada uno de los archivos para comprobar y consultar la estructura de los documentos en RDF y poder comprobar los recursos que lo componen.
- Realizar consultas a cada uno de los documentos transformados a RDF y comprobar que los resultados sean correctos.

5.2 Proyecto futuro

Uno de los enfoques con los que se comenzó este proyecto consistía en realizar la transformación de los Registros de Actividad de Tratamiento de la Junta de Castilla y León, con la intención de crear una nube, Cloud Computing, basada en RDF para generar una base de datos consultable por los usuarios. Esto no ha sido posible debido a que fue imposible encontrar una nube gratuita que soportará RDF, ya que todas las encontradas eran de pago.

Por tanto en los siguientes puntos se exponen dentro de un marco teórico los pasos a seguir para que crear una Cloud Computing y como almacenar la información de las ontologías en bibliotecas digitales específicas.

5.2.1 Cloud Computing

La computación en la nube es un término general para la prestación de servicios alojados a través de Internet, y permite a las empresas consumir recursos informáticos como una utilidad en lugar de tener que construir y mantener infraestructuras de computación en tu casa o tus oficinas.

Podríamos catalogar tres clases de computación en la nube, donde los diferentes servicios son proporcionados para una empresa. Hay que tener en cuenta que existe cierta superposición entre ellos.

- *IaaS (Infraestructura como Servicio)*
Un alojamiento web ordinario es un ejemplo de IaaS, en el cual pagas una suscripción mensual o una tarifa por gigabytes consumidos. Esto significa que estás comprando el acceso al hardware de computación en bruto través de la red, tales como servidores o almacenamiento.
- *SaaS (Software como Servicio)*
Significa que se utiliza una aplicación completa que se ejecuta en el sistema de otra persona, estos pueden ser, por ejemplo: el correo electrónico basado en web y documentos de Google.
- *PaaS (Plataforma como servicio)*
Significa poder desarrollar aplicaciones haciendo uso de herramientas basadas en la Web para que se ejecuten en sistemas de software y hardware proporcionado por otra compañía. Así, por ejemplo, podrías desarrollar tu propio sitio web de comercio electrónico, pero tienes todo, incluyendo el carrito de la compra, pago, envío y el mecanismo de pago ejecutándose en el servidor de un comerciante.

Para los objetivos que se querían alcanzar en el proyecto se decidió un enfoque combinado a través de la web semántica y las plataformas de Cloud Computing. De esta forma se podrían utilizar infraestructuras de computación en nube para construir sistemas de bases de datos consultables en forma de PaaS semántico mediante la cual nosotros obtenemos las licencias de una de las plataformas de Cloud Computing basadas en RDF, como es el caso de Azure de Microsoft y de esta manera se almacenarían los documentos en los servidores de Azure. Esto nos permite que a la hora de consultar los archivos almacenados dentro del Cloud Computing podamos crear una interfaz de búsqueda similar a la que nos muestra la interfaz de AllegroGraph donde realizaremos las consultas pertinentes para buscar los recursos que deseamos mediante un método de consultas basado en SPARQL.

5.2.2 Creación del dominio de las ontologías

Como se explica en el punto anterior, el planteamiento de la creación de un Cloud Computing se basa en la utilización de tecnologías semánticas, ya que resultan más que adecuadas para aprovechar al máximo la información recogida de usuarios y aplicaciones en el entorno.

Dentro de la Web Semántica, las ontologías son de especial importancia para el desarrollo del trabajo. Gracias a estas herramientas se ha podido almacenar la información procedente de los documentos y establecer una serie de relaciones entre ellos que han hecho que esos datos se transformen en conocimiento utilizable. Es necesario un elemento capaz de utilizar el conocimiento almacenado en las ontologías y determinar las acciones a realizar en cada momento para relacionar los recursos con una librería digital en la que se puedan almacenar dichos datos.

Para ello es necesario en los dominios SemRAT y SemCyLRAT que se han usado para la transformación de los archivos, "<http://semcykrat.edu/>" y "<http://semcykrat.edu/>" respectivamente, crear un sistema de almacenamiento de información basado en Virtuoso usando SPARQL endpoint que es un protocolo HTTP en red que es capaz de recibir y procesar solicitudes del protocolo SPARQL. En sendos dominios se almacenarán los prefijos explicados previamente en el apartado *Ontologías usadas*. Esto es necesario para poder realizar las consultas SPARQL en la nube en la que se almacenarán los Registros de Actividades de Tratamientos para poder obtener un enlace a la web semántica de los recursos existentes para comprender su definición.

6. Bibliografía

ARANO Silvia. "Los tesauros y las ontologías en la Biblioteconomía y la Documentación". [En línea] // Hipertext.net. - 2003. - 2021. - <http://www.hipertexto.info/documentos/ontologias.htm>.

Avilleira Ana Beatriz Castelló WEB SEMÁNTICA: RDF Y SGBD QUE LO SOPORTAN [En línea]. - 2006. - <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/496/1/38211tfc.pdf>.

Beckett Dave TURTLE [En línea] // Turtle - Terse RDF Triple Language. - 2011. - <https://www.w3.org/TeamSubmission/turtle/>.

Code for Science & Society OpenRefine [En línea]. - <https://openrefine.org/>.

DBpedia Acceso global y unificado a los gráficos de conocimiento [En línea]. - 2021. - <https://www.dbpedia.org/>.

Deborah L McGuinness Frank Van Harmelen OWL Web ontology language: Overview [Libro]. - 2004.

DIENG-KUNTZ Rose. "Corporate Semantic Webs" [Libro]. - [s.l.] : ERCIM News, No. 51, 2002. - Vol. https://www.ercim.eu/publication/Ercim_News/enw51/EN51.pdf.

DING Ying. FENSEL, Dieter. "OntoWeb: The Thematic Network for the Semantic Web" [Libro]. - 2002. - Vol. Special Semantic Web. http://www.ercim.org/publication/Ercim_News/enw51.

Dublin Core Dublin Core™ Metadata Initiative [En línea]. - <https://dublincore.org/>.

Fensel D., Facca, F.M., Simperl, E., Toma, I. Semantic Web Services [Libro]. - 2011. - Vol. III.

FRANZ.INC AllegroGraph 7.1.0 [En línea] // Introducción. - <https://franz.com/agraph/support/documentation/current/agraph-introduction.html>.

GONZÁLEZ MARTÍNEZ MARÍA MERCEDES PROYECTO DOCENTE E INVESTIGADOR [En línea]. - 2019.

Graham Klyne Jeremy Carroll Resource Description Framework (RDF): Concepts and Abstract Syntax: W3C Recommendation [En línea]. - 10 de Febrero de 2004. - <https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>.

GUTIÉRREZ Claudio "La Web Semántica" [En línea]. - <https://users.dcc.uchile.cl/~cgutierrez/websemantica/websemantica.pdf>.

Hernández Margarita Base22. [En línea] // Digital Transformation Consulting.. - 18 de Diciembre de 2018. - <https://base22.com/es/blog-es/las-ventajas-del-resource-description-framework-rdf/>.

Hughes. Mike Cotterell Bob "Software Project Management". [Libro]. - : McGraw-Hill Education / Europe, Middle East & Africa, 2009. - 5th Edition : Vol. V.

Jorge Pérez Marcelo Arenas, and Claudio Gutierrez Semantics of SPARQL [En línea]. - 2006. - <https://arxiv.org/abs/cs/0605124>.

Junta de Castilla y León [En línea]// gobierno.jcyl. - 2021. - mayo de 2021. - <https://gobierno.jcyl.es/web/es/consejerias.html>.

Lincoln Matthew Uso de SPARQL para acceder a datos abiertos enlazados [En línea]. - 2015. - <https://programminghistorian.org/es/lecciones/retirada/sparql-datos-abiertos-enlazados>.

Liyang Yu "A Developer's Guide to the Semantic Web" [Libro]. - [s.l.] : Springer, 2011.

Martínez M. Mercedes Introducción a la Web Semántica [En línea]. - 2016. - <https://www2.infor.uva.es/~mercedes/websem/fileswebsemvalladolid/introduccion.pdf>.

Matute Jorge Hernández Un estudio comparativo entre los sistemas gestores RDF [En línea]. - 2009. - https://e-archivo.uc3m.es/bitstream/handle/10016/6566/PFC_Jorge_Hernandez.pdf?sequence=1&isAllowed=y.

Peter F. Patel-Schneider Patrick Hayes, Ian Horrocks, OWL-Semantics [En línea]// OWL Web Ontology Language Semantics and Abstract Syntax. - 2004. - <https://www.w3.org/TR/2004/REC-owl-semantics-20040210/>.

RDF SEMANTICS RDF Semantics [En línea]. - 2004. - <https://www.w3.org/TR/2004/REC-rdf-nt-20040210/>.

Ríos Manuel Open Refine [En línea]// qué es + tutorial. - 2013. - <https://es.schoolofdata.org/2014/06/30/openrefine/>.

SPARQL SPARQL Query Language for RDF [En línea]. - 2008. - <https://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>.

ZANG Jun. CARAGEA, Doina. HONAVAR, Vasant. "Learning Ontology -Aware Classifiers". [En línea] // Proceedings of the Eight International Conference on Discovery Science (DS'05),. - Octubre de 2005. - <http://www.cs.iastate.edu/~honavar/Papers/ds05.pdf>.

7. ANEXO I: Instalación OpenRefine

Instalación de OpenRefine

<https://docs.openrefine.org/manual/installing>

- **Requisitos del sistema**

OpenRefine no requiere acceso a Internet para ejecutar sus funciones básicas. Una vez que lo descarga e instala, se ejecuta como un pequeño servidor web en su propia computadora, y accede a ese servidor web local usando su navegador. Solo requiere una conexión a Internet para importar datos de la web, conciliar datos mediante un servicio web o exportar datos a la web.

OpenRefine requiere tres cosas en su computadora para funcionar:

- **Sistema operativo compatible**

OpenRefine está diseñado para funcionar con los sistemas operativos **Windows**, **Mac** y **Linux**. Nuestro equipo lanza paquetes para cada uno.

Java

Java debe estar instalado y configurado en su computadora para ejecutar OpenRefine. La versión para Mac de OpenRefine incluye Java; nuevo en OpenRefine 3.4, también hay un paquete de Windows con Java incluido.

Si instala e inicia OpenRefine en una computadora con Windows sin Java, se abrirá automáticamente una ventana del navegador a la página de descargas de Java, y simplemente puede seguir las instrucciones allí.

Le recomendamos que descargue e instale Java antes de continuar con la instalación de OpenRefine.

Navegador compatible

OpenRefine funciona mejor en navegadores basados en Webkit, como:

- Google Chrome
- Ópera
- Microsoft Edge

Somos conscientes de algunos problemas menores de renderizado y rendimiento en otros navegadores como Firefox. No admitimos Internet Explorer. Si tiene problemas para ejecutar OpenRefine, consulte la sección sobre ejecución.

Versiones de lanzamiento

OpenRefine siempre tiene una versión estable más reciente, así como algunos desarrollos más recientes disponibles en versiones beta, candidatas a versiones o instantáneas. Si está instalando por primera vez, le recomendamos la última versión estable.

Si desea utilizar una extensión que solo sea compatible con una versión anterior de OpenRefine y no requiera las funciones más recientes, puede encontrar que una versión estable anterior es la mejor para usted en nuestra lista de lanzamientos. Consulte las versiones posteriores para ver qué vulnerabilidades de seguridad se están solucionando, a fin de evaluar su propia tolerancia al riesgo para el uso de versiones anteriores. Busque versiones de "lanzamiento final" en lugar de versiones "beta" o "candidatas a lanzamiento".

Versiones inestables

Si necesita una función desarrollada recientemente y está dispuesto a arriesgar código no probado, puede mirar los elementos más recientes de la lista y ver qué cambios le atraen.

Las versiones "Beta" y "Release Candidate" pueden tener errores no reportados y son más adecuadas para las personas que están dispuestas a ayudarnos a solucionar estos problemas mediante la creación de informes de errores .

Para obtener las últimas actualizaciones de desarrollo absolutas, consulte las versiones instantáneas . Estos se crean con cada compromiso.

Qué ha cambiado

Nuestra última versión es OpenRefine 3.4.1 , lanzada el 24 de septiembre de 2020. Los principales cambios en esta versión se enumeran en la página de lanzamiento 3.4.1 con los paquetes descargables.

Puede encontrar información sobre todas las versiones de OpenRefine en la página de lanzamientos en Github .

OTRAS DISTRIBUCIONES

OpenRefine también puede funcionar en otros entornos, como Chromebooks, donde hay terminales Linux disponibles. Consulte nuestra lista de otras distribuciones en la página de descargas para conocer otras formas de ejecutar OpenRefine y consulte nuestra comunidad de colaboradores para ver nuevos entornos en desarrollo.

Instalando o actualizando

Haga una copia de seguridad de sus datos

Si está actualizando desde una versión anterior de OpenRefine y ya tiene proyectos en su computadora, debe crear copias de seguridad de esos proyectos antes de instalar una nueva versión.

Primero, ubique el directorio de su espacio de trabajo. Luego, copie todo lo que encuentre allí y péguelo en una carpeta en otro lugar de su computadora.

Para mayor seguridad, puede exportar sus proyectos OpenRefine existentes.

PRECAUCIÓN

Toma nota de las extensiones que tienes instaladas actualmente. Es posible que no sean compatibles con la versión mejorada de OpenRefine. Las instalaciones se pueden instalar en dos lugares, así que asegúrese de verificar tanto el directorio de su espacio de trabajo como el directorio de instalación existente.

Instalar o actualizar OpenRefine

Si está actualizando una instalación de OpenRefine existente, puede eliminar los archivos de programa antiguos e instalar los archivos nuevos en el mismo espacio. No sobrescriba los archivos, ya que algunos archivos obsoletos pueden quedar innecesariamente.

PRECAUCIÓN

Si tiene extensiones instaladas, no elimine la *webapp\extensions* carpeta donde las instaló. Es posible que desee instalar extensiones en el directorio del espacio de trabajo en lugar del directorio del programa. No hay garantía de que las extensiones sean compatibles con las nuevas versiones de OpenRefine y no mantenemos extensiones.

WINDOWS

Una vez que haya descargado el *.zip* archivo, extraígallo en una carpeta donde desee almacenar archivos de programa (como *D:\Program Files\OpenRefine*).

Puede hacer clic con el botón derecho en *openrefine.exe* o *refine.bat* y anclar uno de esos programas a su menú Inicio o crear accesos directos para facilitar el acceso.

Establecer dónde se almacenan los datos

OpenRefine almacena datos en dos lugares: archivos de programa en el directorio del programa, donde sea que lo haya instalado; y archivos de proyecto en lo que llamamos el "directorio del espacio de trabajo". Puede acceder a esta carpeta fácilmente desde OpenRefine yendo a la pantalla de inicio (en *http://127.0.0.1:3333/*) y haciendo clic en *Examinar el directorio del espacio de trabajo*.

Por defecto esto es:

Dependiendo de su versión de Windows, los datos se encuentran en uno de estos directorios:

- *%appdata%\OpenRefine*
- *%localappdata%\OpenRefine*
- *C:\Documents and Settings\ (user id) \Local Settings\Application Data\OpenRefine*
- *C:\Users\ (user id) \AppData\Roaming\OpenRefine*
- *C:\Users\ (user id) \AppData\Local\OpenRefine*
- *C:\Users\ (user id) \OpenRefine*

Para versiones anteriores de Google Refine, reemplace *OpenRefine* con *Google\Refine*. Puede cambiar esto agregando esta línea al archivo *openrefine.14j.ini* y especificando la unidad deseada y la ruta de la carpeta:

```
-Drefine.data_dir = D: \ MyDesiredFolder
```

Si la ruta de su carpeta tiene espacios, use comillas neutrales alrededor:

```
-Drefine.data_dir = "D: \ Mi carpeta deseada"
```

Si la carpeta no existe, OpenRefine la creará.

Registros

OpenRefine no genera actualmente un registro de errores, pero debido a que la ventana de la consola OpenRefine siempre está abierta (en Linux y Windows) mientras OpenRefine se ejecuta en su navegador, puede copiar información de la consola si ocurre un error. Con una Mac, puede ejecutar OpenRefine usando la terminal para capturar errores.

Aumento de la asignación de memoria

OpenRefine se basa en tener memoria de computadora disponible para funcionar de manera efectiva. Si planea trabajar con grandes conjuntos de datos, es posible que desee configurar OpenRefine para manejarlo desde el principio. Por "grande" generalmente nos referimos a uno de los siguientes indicadores:

- más de un millón de células en total
- un tamaño de archivo de entrada de más de 50 megabytes (MB)
- más de 50 filas por registro en modo registros

De forma predeterminada, OpenRefine está configurado para funcionar con 1 gigabyte (GB) de memoria (1024 MB). Si cree que OpenRefine se está ejecutando lentamente o está obteniendo errores de "memoria insuficiente" (por ejemplo, *java.lang.OutOfMemoryError*), puede intentar asignar más memoria.

Una buena práctica es comenzar con no más del 50% de la memoria que quede después del uso estimado de su sistema operativo, para dejar memoria para que su navegador se ejecute.

Todas las configuraciones siguientes utilizan un número de cuatro dígitos para especificar los megabytes (MB) utilizados (en realidad, mebibytes). El valor predeterminado suele ser 1024 MB, pero no es necesario que el nuevo valor sea múltiplo de 1024.

TRATAR CON GRANDES CONJUNTOS DE DATOS

Si su proyecto es lo suficientemente grande como para necesitar más que la cantidad predeterminada de memoria, considere apagar *Analizar el texto de la celda en números, fechas, ...* en la importación. Es conveniente, pero menos eficiente que convertir explícitamente cualquier columna que necesite como un tipo de datos que no sea el tipo de "cadena" predeterminado.

Usando openrefine.exe

Si lo ejecuta *openrefine.exe*, deberá editar el *openrefine.14j.ini* archivo que se encuentra en el directorio del programa y editar la línea

```
# tamaño máximo del montón de memoria de la memoria
```

```
# tamaño máximo del montón de memoria de la memoria
```

```
-Xmx1024M
```

La línea `"-Xmx1024M"` define la cantidad de memoria disponible en megabytes. Cambie el número "1024"; por ejemplo, edite la línea a `"-Xmx2048M"` para que haya 2048 MB [2 GB] de memoria disponible.

¿OPENREFINE.EXE NO SE ESTÁ EJECUTANDO?

Una vez que aumente la asignación de memoria, es posible que no pueda ejecutar *openrefine.exe*. En este caso, su computadora necesita una versión de 64 bits de Java (esto es diferente de Java JDK . Busque la descarga "Windows Offline (64-bit)" en la página de Descargas e instálela. Su sistema también debe estar configurado para utilice la versión de 64 bits de Java cambiando la configuración de Java .

Usando refine.bat

En Windows, OpenRefine también se puede ejecutar utilizando el archivo *refine.bat* en el directorio del programa. Si inicia OpenRefine usando *refine.bat*, la memoria disponible para OpenRefine se puede especificar a través de las opciones de la línea de comandos o mediante el *refine.ini* archivo.

Para establecer la cantidad máxima de memoria en la línea de comando cuando se usa *refine.bat*, *cd* en el directorio del programa, luego escriba

```
refine.bat /m 2048m
```

donde "2048" es la cantidad máxima de MB que desea que use OpenRefine.

Para cambiar el valor predeterminado que *refine.bat* usa, edite la *refine.ini* línea que dice

```
REFINE_MEMORY=1024M
```

Tenga en cuenta que este archivo solo se lee si usa *refine.bat*, no *openrefine.exe*.

Instalando extensiones

Nuestra comunidad de colaboradores ha creado extensiones para agregar funcionalidad o proporcionar accesos directos convenientes para usos comunes de OpenRefine. Enumeramos las extensiones que conocemos en nuestra página de descargas.

CONTRIBUYENDO EXTENSIONES

Si desea crear o modificar una extensión, consulte nuestra documentación para desarrolladores aquí . Si tiene un problema, use nuestra página de descargas para ir a la página de la extensión e informar el problema allí.

Dos formas de instalar extensiones

Puede instalar extensiones en uno de dos lugares :

- En la carpeta del programa OpenRefine, por lo que solo estarán disponibles para esa versión / instalación de OpenRefine (lo que significa que la extensión no se ejecutará si actualiza OpenRefine), o
- En su espacio de trabajo, donde se almacenan sus proyectos, para que estén disponibles sin importar qué versión de OpenRefine esté usando.

Ofrecemos estas opciones porque es posible que desee reinstalar una extensión determinada manualmente cada vez que actualice OpenRefine, para asegurarse de que funcione correctamente.

Encuentre el lugar adecuado para instalar

Si desea instalar la extensión en la carpeta del programa, vaya al directorio de su programa y luego vaya a `webapp\extensions` (o créelo si no existe).

Si desea instalar la extensión en su espacio de trabajo, puede:

- inicie OpenRefine y haga clic en *Proyecto abierto* en la barra lateral
- En la parte inferior de la pantalla, haga clic en *Examinar el directorio del espacio de trabajo*
- Se abrirá una ventana de buscador o explorador de archivos en su espacio de trabajo
- Cree una nueva carpeta llamada "extensiones" dentro del espacio de trabajo si no existe.

También puede encontrar su espacio de trabajo en cada sistema operativo siguiendo estas instrucciones .

Instale la extensión

Algunas extensiones tienen sus propias instrucciones: asegúrese de leer la documentación antes de comenzar a instalar.

Algunas extensiones pueden tener varias versiones, para coincidir con las versiones de OpenRefine, así que asegúrese de elegir la versión correcta para su instalación. Si tiene preguntas sobre compatibilidad o desea solicitar o expresar su apoyo para una actualización, use nuestra página de descargas para ir a la página de la extensión e informar el problema allí.

Generalmente, el proceso de instalación será:

- Descargue la extensión (generalmente como un archivo zip de GitHub)
- Extraiga el contenido del zip en el directorio de "extensiones", asegurándose de que todo el contenido vaya a una carpeta con el nombre de la extensión.
- Inicie (o reinicie) OpenRefine.

Para confirmar que la instalación fue un éxito, siga las instrucciones proporcionadas por la extensión. Cada extensión aparecerá a su manera dentro de la interfaz de OpenRefine. Asegúrese de leer su documentación para saber dónde aparecerá la funcionalidad, como en menús desplegables específicos.

8. ANEXO II: Instalación AllegroGraph

Introducción

Este documento lo guía para comenzar con AllegroGraph. AllegroGraph es un producto complejo y poderoso, por lo que tiene muchas opciones. Estas opciones se describen en detalle en documentos especializados que tratan aspectos específicos de AllegroGraph. En esta guía de inicio rápido, sugeriremos opciones estándar (a menudo predeterminadas) y enlaces a documentos más especializados. Eso debería permitirle estar listo y funcionando rápidamente.

Descarga de software

Descargue el servidor AllegroGraph desde la página de descargas de AllegroGraph . Hay varias opciones para los clientes de AllegroGraph. En este documento de inicio rápido, asumiremos que utilizará el cliente *AGWebView* (utilizando un navegador web como interfaz). *AGWebView* está integrado en el servidor AllegroGraph. El servidor se ejecuta de forma nativa en máquinas Linux x86-64 bits. Puede ejecutarlo en una máquina Mac o Windows usando una máquina virtual Linux, como se describe en la página de descarga de AllegroGraph . También puede ejecutar AllegroGraph dentro del Subsistema de Windows para Linux (WSL).

En el resto de este documento asumimos que se ha descargado el archivo de instalación del servidor.

WSL: descargando el archivo de instalación

Debido a que la estructura de directorios del Subsistema de Windows para Linux (WSL) Unix se puede ocultar en Windows, es más fácil descargar los archivos necesarios con curl o wget ejecutados en un shell de UNIX. Consulte la sección Descarga del servidor en el documento Instalación del servidor para obtener información sobre cómo obtener el enlace de descarga adecuado como argumento para curl o wget .

Instalación del servidor

La instalación del servidor se describe en detalle en Instalación del servidor . Hay dos métodos de instalación: usando un archivo RPM o usando un archivo tar con formato gzip. Describiremos la instalación desde un archivo tar.gz aquí.

Instalación desde el archivo tar.gz

Los comandos de esta sección no requieren acceso "root".

Coloque el archivo tar.gz descargado en un directorio de su elección. Luego extraiga los archivos de instalación con el siguiente comando:

```
$ tar xzf agraph-7.1.0-linuxamd64.64.tar.gz
```

Esto crea el subdirectorio "agraph-7.1.0" dentro del directorio de trabajo actual, donde reside el script de instalación. El script de instalación es "install-agraph". Debe proporcionar la ruta a un directorio grabable (que se creará si es necesario) donde desea instalar AllegroGraph.

```
$ agraph-7.1.0/install-agraph /home/joe/tmp/ag7.1.0
```

```
Installation complete.
```

```
Now running configure-agraph.
```

Ejecución del script de configuración

El script de instalación de AllegroGraph crea un archivo de configuración de AllegroGraph (agraph.cfg) con la configuración apropiada para su entorno. El script crea el archivo de configuración mínimo para comenzar. El archivo de configuración se puede editar más tarde.

El script de configuración se ejecuta automáticamente cuando se instala desde el archivo tar.gz. Hace varias preguntas. Las respuestas predeterminadas suelen ser adecuadas y se pueden reconfigurar más adelante si es necesario.

```
Welcome to the AllegroGraph configuration program. This script will
```

```
help you establish a baseline AllegroGraph configuration.
```

```
You will be prompted for a few settings. In most cases, you can hit return
```

```
to accept the default value.
```

```
Location of configuration file to create:
```

```
[/home/joe/tmp/ag7.1.0/lib/agraph.cfg]:
```

```
Directory to store data and settings:
```

```
[/home/joe/tmp/ag7.1.0/data]:
```

```
Directory to store log files:
```

```
[/home/joe/tmp/ag7.1.0/log]:
```

```
Location of file to write server process id:
```

```
[/home/joe/tmp/ag7.1.0/data/agraph.pid]:
```

Si está instalando como **root** (lo cual no es necesario), se le pedirá una cuenta de usuario no root para que se ejecute AllegroGraph. (Sería un riesgo para la seguridad dejar que AllegroGraph se ejecute como root.) Esta cuenta por defecto es "agraph" y está designada en agraph.cfg como "RunAs agraph". La cuenta "agraph" no tiene contraseña y solo la usa el servidor AllegroGraph. El sistema determinará que ha iniciado sesión como root y le hará las siguientes preguntas. Por lo general, puede tomar el valor predeterminado:

```
User to run as:
[agraph]:

User 'agraph' doesn't exist on this system.
Create agraph user:
[y]:
```

El script de configuración le pedirá un nombre de usuario y una contraseña para ser su superusuario de AllegroGraph. Esta cuenta es específica de AllegroGraph y no es la cuenta de inicio de sesión de su sistema operativo. El siguiente ejemplo muestra el usuario "prueba" con la contraseña "xyzy". Esta es la cuenta esperada por los ejemplos del tutorial, por lo que es conveniente crear esa cuenta ahora. La administración de cuentas de usuario es fácil de realizar a través de la interfaz AllegroGraph WebView , por lo que esta combinación inicial de usuario / contraseña se puede cambiar fácilmente más adelante.

```
Now you must set up an initial user account for AllegroGraph.
This
account will have "super user" privileges within AllegroGraph.

SuperUser account name:
[super]: test

SuperUser account password:
xyzy
```

Finalmente, se le pedirá por los segundos del tiempo de espera de la instancia. Este es el período de tiempo (en segundos) que una base de datos permanecerá abierta sin que se acceda a ella. El valor predeterminado es 604800, que es 60 x 60 x 24 x 7, es decir, una semana en segundos. Por el contrario, una hora es 3600.

```
Instance timeout seconds:
[604800]:
```

El valor se escribirá en el archivo `agraph.cfg` como `InstanceTimeout 604800`. Si elimina esa línea del archivo, el intervalo de tiempo de espera será de una hora (3600 segundos).

El script escribe el archivo `agraph.cfg` en el directorio `<yourInstallDirectory> / lib` (la ubicación estándar para la instalación TAR.GZ). Tome nota de dónde reside su archivo `agraph.cfg`.

```
/home/joe/tmp/ag7.1.0/lib/agraph.cfg has been created.

If desired, you may modify the configuration.
```


El script concluye con una visualización de los comandos de inicio y detención personalizados para su instalación. Tome nota de los comandos de inicio y detención de su instalación.

```
You can start AllegroGraph by running:

/home/joe/tmp/ag7.1.0/bin/agraph-control          --config
/home/joe/tmp/ag7.1.0/lib/agraph.cfg start

You can stop AllegroGraph by running:

/home/joe/tmp/ag7.1.0/bin/agraph-control          --config
/home/joe/tmp/ag7.1.0/lib/agraph.cfg stop
```

A menos que esté utilizando la versión gratuita (que permite hasta 5 millones de triples), debe agregar una clave de licencia al archivo de configuración. Consulte a continuación para obtener más información. Si está utilizando la versión gratuita, vaya a Iniciar el servidor a continuación .

Instalación de la clave de licencia

Cuando descarga la versión gratuita de AllegroGraph, no se proporciona ni se requiere ninguna clave de licencia. La versión gratuita te permite cargar un máximo de 5.000.000 triples y no tiene fecha de caducidad.

Franz emite una clave de licencia especial para los usuarios de la versión paga de AllegroGraph. La clave de licencia define un número máximo de triples y posiblemente una fecha de vencimiento para la licencia, y se emite a un individuo o empresa cliente. Su clave de licencia será emitida por el departamento de ventas de Franz y le llegará en su correo electrónico. Una clave de licencia típica se parece a este ejemplo (que no funciona):

```
Licensee Customer Name

LicenseLimit 1000000

LicenseExpires 2010-06-16

<LicenseCode>

829P93952R9E8P7X4E8W565Z7A2A4Z592E4W5E49693P9K325X296S3R945A7
89D4

S843K9455787K3R3A2K6Z6N8K8V6V658A3V3D293H8X5S7X3V9A4R6533979V
246V

6J8Y3T928MA69B928R7E4D8C7C4T6X9F28626Z448E986R6E9B6F4S9J9U6H8
A7Q8

S324C67783Y89495Q8W867B7S349Q6C6T5W42A59P8Q5M967R5B554C2K3Q42
4T

</LicenseCode>
```

La "Clave de licencia" consta de los cuatro campos: el Licenciario, LicenseLimit, LicenseExpires y LicenseCode. Todos deben estar presentes para que la Llave funcione. Los primeros tres valores reflejan los términos de su contrato con Franz. LicenseCode verifica los otros valores.

Para instalar su clave de licencia, copie todo el bloque de texto (Licensee, LicenseLimit, LicenseExpires y LicenseCode) y péguelo en su archivo agraph.cfg. (En el ejemplo de instalación de tar.gz anterior, el archivo agraph.cfg se encontraba en /home/joe/tmp/ag7.1.0/lib/agraph.cfg.) La clave de licencia puede ir a cualquier parte del archivo siempre que está fuera de una definición de catálogo. Guarda el archivo. La licencia será efectiva cuando inicie o reinicie el servidor AllegroGraph.

Si hay algún problema con su clave de licencia, AllegroGraph imprimirá un mensaje de error y el servidor no podrá iniciarse. Póngase en contacto con sales@franz.com si tiene problemas con su clave de licencia. Mientras tanto, siempre puede ejecutar la versión gratuita (cinco millones de límite triple) eliminando o comentando la parte de la licencia del archivo de configuración.

Iniciar el servidor

El servidor se instala en el directorio especificado en el comando de **instalación** (ver más arriba). En el ejemplo anterior era /home/joe/tmp/ag7.1.0/ pero probablemente eligió un directorio diferente. Llamaremos a ese directorio **[ag-dir]** en los ejemplos siguientes.

Como se indicó anteriormente, cuando completa la instalación, se imprime un mensaje que indica los comandos para iniciar y detener el servidor AllegroGraph. El comando es **un control gráfico** en el subdirectorio *bin / del* directorio **[ag-dir]**. Inicie el servidor con:

```
[ag-dir]/bin/agraph-control --config [ag-dir]/lib/agraph.cfg
start
```

Un mensaje similar a

```
AllegroGraph Server
Copyright (c) Franz Inc. All Rights Reserved.
AllegroGraph contains patented technology.

Current time: [day, date, and time]

Daemonizing...

Server started normally: Running with free license of 5,000,000
triples; no-expiration.

Access AGWebView at http://127.0.0.1:10035
```

se imprimirá (si instaló una clave de licencia, el mensaje de licencia será diferente). El servidor AllegroGraph ahora se está ejecutando.

El error 'AllegroGraph ya se está ejecutando'

Puede ejecutar varios servidores en una sola máquina, pero deben usar puertos diferentes y no deben usar el mismo directorio especificado por el `SettingsDirectory` parámetro de configuración. Consulte Configuración y control del servidor para obtener información sobre cómo especificar parámetros. Si ese directorio está en uso por otro

proceso de servidor, cuando intente iniciar un servidor que usará ese mismo directorio, el inicio del servidor fallará con un mensaje similar a (con valores para PID y PORT):

```
Daemonizing...
Starting server failed:

  There appears to already be an AllegroGraph server running
(pid PID) .

  If it is your intention to run another AllegroGraph server
  simultaneously,

  please make a separate configuration file which has different
values

  for the following parameters:

Port                PORT

SettingsDirectory  /disk1/allegrograph/settings/
```

Puede terminar ese servidor (después de asegurarse de que es lo que desea hacer y que no está incomodando a otro usuario) con

```
agraph-control --config [path to agraph.cfg file of the running
server] stop
```

Si no puede determinar dónde está el archivo agraph.cfg apropiado o si ese comando falla por alguna razón, puede (nuevamente, asegurándose de que no está interfiriendo con otros usuarios) hacer

```
kill PID
```

Si el puerto deseado está en uso (presumiblemente por un proceso que no es AllegroGraph pero quizás por otro servidor AllegroGraph que no usa el directorio mencionado anteriormente), el servidor no se iniciará y se imprimirá un mensaje como el siguiente:

```
Local socket address already in use" (errno 98) occurred while
creating a passive socket on any interface port PORT.
```

Para usar el socket deseado, debe encontrar y terminar el proceso usándolo. También puede modificar el archivo agraph.cfg para especificar otro puerto no utilizado. Consulte el `Port` parámetro en Configuración y control del servidor .

Conectando al servidor

Hay varios clientes potenciales que pueden conectarse al servidor de AllegroGraph, ya sea desde la misma máquina que ejecuta el servidor o desde una máquina diferente. Los distintos clientes y cómo se descargan (cuando es necesario) y cómo se instalan e inician se describen en Instalación del servidor , que tiene enlaces a otros documentos sobre clientes.

En este documento, describimos la conexión al servidor AllegroGraph con el cliente AGWebView . Este cliente utiliza un navegador web y no requiere (más)

descarga o instalación. La única información que necesita es el nombre de la máquina que ejecuta el servidor (que puede ser la misma máquina que ejecuta su navegador) y el número de puerto en el que está escuchando el servidor. El número de puerto predeterminado es 10035, y si lo instaló como se describe arriba, ese será el número de puerto. Si tiene alguna dificultad con ese número de puerto (como, por ejemplo, está siendo utilizado por otro programa), edite el archivo de configuración ([ruta] /ag7.1.0/lib/agraph.cfg) modificando la línea

```
Port 10035
```

para especificar un puerto disponible diferente. Si cambia esa línea, reemplace 10035 con el nuevo número en los ejemplos a continuación en este documento.

Usando AGWebView

AGWebView es parte del servidor AllegroGraph. Cuando el servidor se está ejecutando, AGWebView se puede abrir escribiendo el nombre de host de la red del servidor o la dirección IP y el número de puerto en el campo de dirección de un navegador web, como el siguiente:

```
http://localhost:10035
```

localhost se refiere a la máquina que ejecuta el navegador (por lo que tanto el servidor como el navegador se ejecutan en la misma máquina).

Iniciando sesión

Si hay un servidor AllegroGraph escuchando en ese puerto, AGWebView responderá inmediatamente con un formulario de inicio de sesión (en este ejemplo, el servidor se está ejecutando en la *incursión del host*):

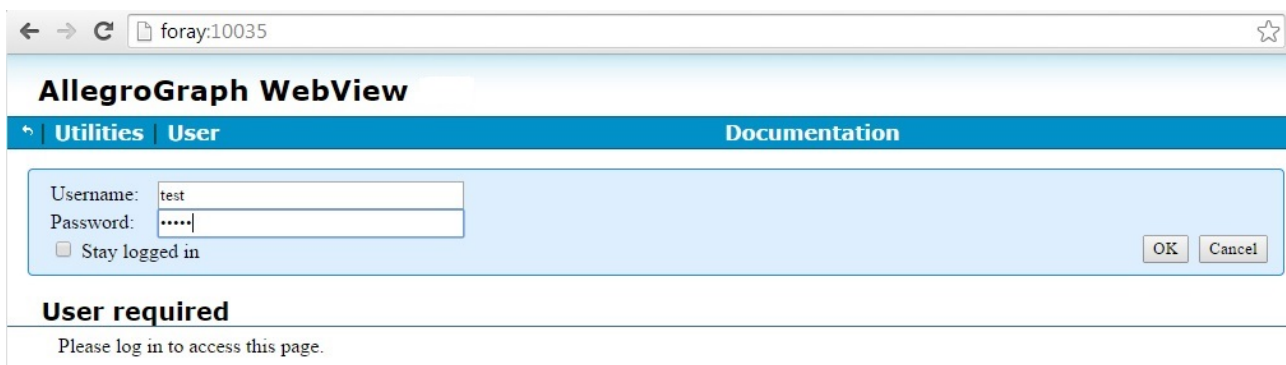


Imagen 37. Inicio sesión AllegroGraph

Para este sencillo tutorial, inicie sesión con el nombre y la contraseña del superusuario de AllegroGraph que creó durante la instalación del servidor. Si aceptó los valores predeterminados cuando instaló AllegroGraph arriba, el usuario es "prueba" con la contraseña "xyzyz" (como se muestra en la ilustración, aunque la contraseña está oculta).

Una vez que haya iniciado sesión, puede crear cuentas de usuario adicionales con diferentes tipos de acceso y privilegios. Consulte Gestión de usuarios para obtener más detalles.

El directorio de tutoriales

En el nivel superior de la instalación de AllegroGraph (junto con directorios como *bin* / y *lib* /), hay un directorio *tutorial* / . Este directorio contiene varios archivos de ejemplo usados en algunos de los ejemplos que se encuentran en la documentación, incluido el archivo *kennedy.ntriples* (usado a continuación en este documento) y el archivo *PhoneCalls.ttl.gz* usado en el ejemplo geoespacial n-dimensional en N-dimensional Guía y ejemplo de uso geoespacial . Cuando, en una página de repositorio de AGWebView, elige **Cargar y eliminar datos | Importar RDF | desde un archivo cargado** (indicado por la flecha roja en la ilustración, que muestra el ejemplo de Kennedy desarrollado a continuación), la **opción Seleccionar archivos** El cuadro de diálogo que se muestra comienza en el último directorio que se mostró. Pero la primera vez que se invoca, muestra el directorio *tutorial* / en el directorio de distribución del servidor (suponiendo que ese directorio esté en el lugar esperado; si hay un problema, el cuadro de diálogo mostrará el directorio raíz de distribución del servidor).

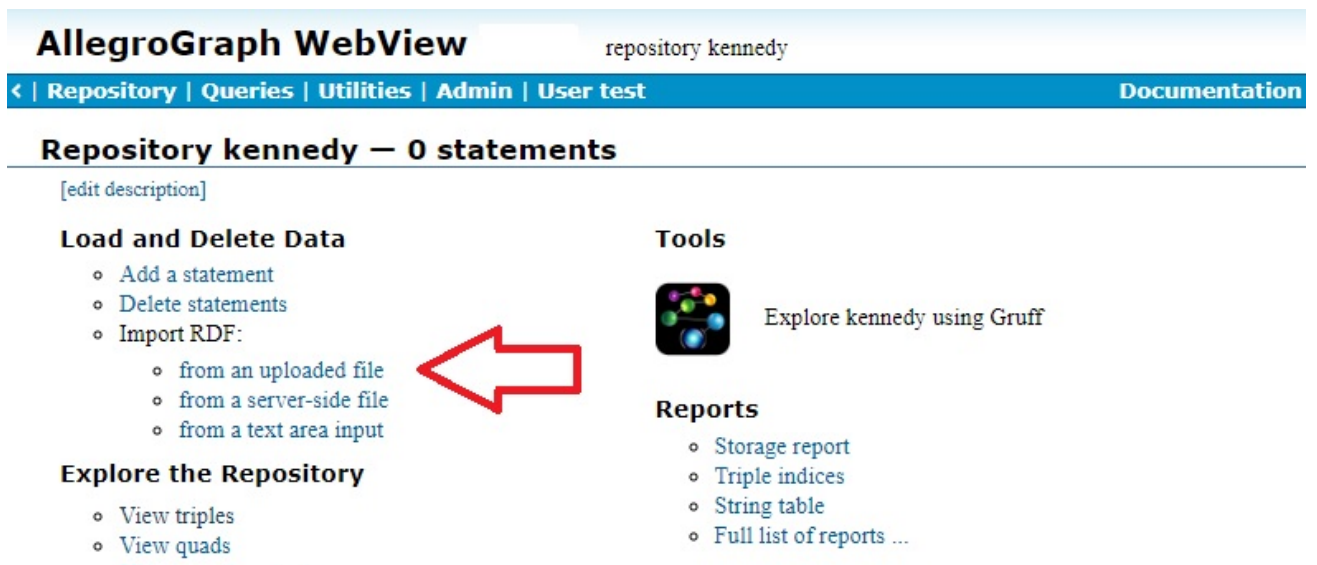


Imagen 38. Seleccionar archivos

Es posible que no tenga acceso a ese directorio por alguna razón. Si ese es el caso, pídale al administrador del sistema que le proporcione o ponga a disposición una copia del directorio.

Crear una base de datos

Una vez que haya iniciado sesión correctamente, el navegador se verá así:

Catalogs and Repositories

Catalogs

/
system

Repositories

None.

Create new repository

Name:

[Restore from a backup](#)

Visualization



Start Gruff

- Visualization, Exploration
- Query Builder, and more...

Imagen 39. Crear repositorio

Esta es la página Catálogo que enumera los catálogos y los repositorios del catálogo seleccionado. El catálogo raíz (indicado por /) se selecciona en la ilustración. El único otro catálogo es el catálogo del **sistema**, que es solo para uso interno.

No hay repositorios en el catálogo raíz ya que acabamos de abrir AllegroGraph por primera vez (si los datos se hubieran almacenado, veríamos uno o más repositorios que contienen los datos).

Vamos a crear un repositorio llamado *kennedy* y lo completaremos con los datos disponibles en un archivo de ejemplo incluido con la distribución. En el área **Crear nuevo repositorio**, en el campo **Nombre** :, ya *ingresamos kennedy*. Haga clic en el botón **Crear**. Se creará el repositorio de *Kennedy* y el navegador mostrará la página del repositorio de *Kennedy* :

Repository kennedy — 0 statements

[\[edit description\]](#)**Load and Delete Data**

- Add a statement
- Delete statements
- Import RDF:
 - from an uploaded file
 - from a server-side file
 - from a text area input

Explore the Repository

- View triples
- View quads
- View repository's classes
- View repository's predicates
- View repository's named graphs
- Explore repository in Gruff

Replication

- Multi-master: convert store to a replication instance [?](#)
- Warm standby: control replication [?](#)

Repository Control

- Export repository as ▼
- Start a session — support transactions and Prolog functors
- Warmup store
- Back-up this repository
- Export duplicate statements ▼
- Delete duplicate statements
- Suppress duplicate statements false
- View/manage active transactions
- Recognize geospatial datatypes automatically:
- Control durability (bulk-load mode)
- Manage triple attribute definitions and static filter
- View/manage triple indices
- Optimize the repository [?](#)
- Manage free-text indices
- Materialize Entailed Triples [?](#)
- Delete Materialized Triples
- Manage external Solr free-text indexer [?](#)
- Manage external MongoDB connection [?](#)
- Shutdown instance [?](#)

Tools

Explore kennedy using Gruff

Reports

- Storage report
- Triple indices
- String table
- Full list of reports ...

Imagen 40. Información del repositorio

Se muestra información sobre el repositorio (por ejemplo, no hay datos, *declaraciones 0*, hasta el momento) y varios comandos para administrar los datos. Queremos cargar datos del archivo *kennedy.ntriples*, que se encuentra en el directorio *tutorial /* de la instalación del servidor; consulte la sección del directorio del tutorial más arriba). AllegroGraph admite varios formatos de archivo de datos (consulte la carga de una herramienta). El formato ntriples tiene un triple (sujeto, predicado, objeto) por línea, y cada línea termina en un punto. Aquí están las primeras líneas:

```

<http://www.franz.com/simple#person1>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.franz.com/simple#person> .

<http://www.franz.com/simple#person1>
<http://www.franz.com/simple#first-name> "Joseph" .

<http://www.franz.com/simple#person1>
<http://www.franz.com/simple#middle-initial> "Patrick" .

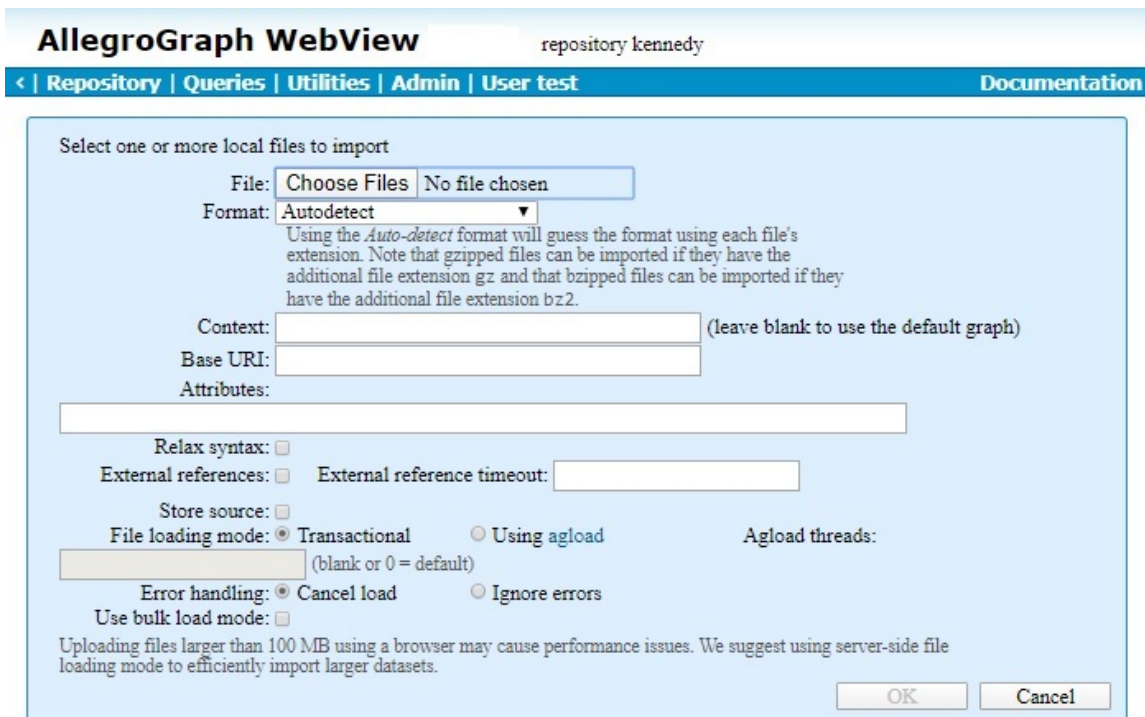
<http://www.franz.com/simple#person1>
<http://www.franz.com/simple#last-name> "Kennedy" .

<http://www.franz.com/simple#person1>
<http://www.franz.com/simple#suffix> "none" .

<http://www.franz.com/simple#person1>
<http://www.franz.com/simple#alma-mater>
<http://www.franz.com/simple#Harvard> .

```

Cargamos estos datos haciendo clic en **Cargar y eliminar datos | Importar RDF: | desde un archivo cargado** (indicado por la flecha roja en la ilustración). Aparece un panel de diálogo cerca de la parte superior de la página WebView que le permite especificar el archivo o archivos que se cargarán.



Repository kennedy – 0 statements

[\[edit description\]](#)

Load and Delete Data

- o [Add a statement](#)
- o [Delete statements](#)
- o [Import RDF:](#)
 - o [from an uploaded file](#)
 - o [from a server-side file](#)
 - o [from a text area input](#)



Explore the Repository

Tools



Explore kennedy using Gruff

Reports

- o [Storage report](#)
- o [Triple indices](#)

Imagen 41. Importar archivos

Este diálogo tiene muchas opciones. Consulte la descripción para obtener más detalles. Aquí solo especificaremos un archivo para cargar, dejando todas las opciones en sus valores predeterminados.

Haga clic en el botón **Elegir archivos** . Mostrará el último directorio que mostró, pero la primera vez que se invoca, muestra el directorio *tutorial* / en el directorio de instalación del servidor (asumiendo que el directorio está en el lugar esperado; si hay un problema, el cuadro de diálogo mostrará el servidor directorio raíz de distribución).

Especifique el archivo en el *tutorial* / directorio llamado *kennedy.ntriples* . Luego haga clic en el botón **Aceptar** en la parte inferior derecha del panel de diálogo. (Si el sistema no puede acceder al directorio *tutorial* / por algún motivo, por supuesto, tendrá que hacer que el archivo sea accesible de alguna otra manera).

Una vez que se cargan los *triples* en el archivo, vemos que el **Repositorio Kennedy** ahora tiene 1214 declaraciones (triples).

The screenshot shows the AllegroGraph WebView interface for the 'repository kennedy'. The top navigation bar includes links for Repository, Queries, Utilities, Admin, User test, and Documentation. The main heading is 'Repository kennedy — 1,214 statements' with an '[edit description]' link. The interface is divided into several sections:

- Load and Delete Data:**
 - Add a statement
 - Delete statements
 - Import RDF:
 - from an uploaded file
 - from a server-side file
 - from a text area input
- Explore the Repository:**
 - View triples
 - View quads
 - View repository's classes
 - View repository's predicates
 - View repository's named graphs
 - Explore repository in Gruff
- Tools:**
 - Explore kennedy using Gruff (with a Gruff icon)
- Reports:**
 - Storage report
 - Triple indices
 - String table
 - Full list of reports ...
- Replication:**
 - Multi-master: convert store to a replication instance ?
 - Warm standby: control replication ?
- Repository Control:**
 - Export repository as N-Triples (dropdown menu)
 - Start a session — support transactions and Prolog functors
 - Warmup store
 - Back-up this repository
 - Export duplicate statements (dropdown menu: Subject, Predicate, Object and Graph (spog))
 - Delete duplicate statements
 - Suppress duplicate statements false
 - View/manage active transactions
 - Recognize geospatial datatypes automatically:
 - Control durability (bulk-load mode)
 - Manage triple attribute definitions and static filter
 - View/manage triple indices
 - Optimize the repository ?
 - Manage free-text indices
 - Materialize Entailed Triples ?
 - Delete Materialized Triples
 - Manage external Solr free-text indexer ?
 - Manage external MongoDB connection ?
 - Shutdown instance ?

Imagen 42. *kennedy.ntriples*

Ahora hemos creado un repositorio y cargado los datos. Ahora deberíamos respaldarlo. Elija **Hacer una copia de seguridad de este repositorio** en las opciones de **Control del repositorio** (encerrado en un círculo rojo en la ilustración). Aparece el panel de diálogo que se muestra en la parte superior de la pantalla. Ingrese la ruta y el nombre del archivo (un nombre de ruta de directorio absoluto no deja ambigüedad sobre dónde se colocará el archivo; el archivo se guardará en la máquina del servidor) en el cuadro de texto y haga clic en **Aceptar**.

The screenshot shows the AllegroGraph WebView interface for a repository named 'kennedy'. The page title is 'AllegroGraph WebView repository kennedy'. The navigation bar includes links for 'Repository', 'Queries', 'Utilities', 'Admin', 'User test', and 'Documentation'. The main content area is titled 'Repository kennedy – 1,214 statements' and includes a link to '[edit description]'. The interface is divided into several sections:

- Load and Delete Data:**
 - Add a statement
 - Delete statements
 - Import RDF:
 - from an uploaded file
 - from a server-side file
 - from a text area input
- Explore the Repository:**
 - View triples
 - View quads
 - View repository's classes
 - View repository's predicates
 - View repository's named graphs
 - Explore repository in Gruff
- Replication:**
 - Multi-master: convert store to a replication instance
 - Warm standby: control replication
- Repository Control:**
 - Export repository as N-Triples
 - Start a session — support transactions and Prolog functors
 - Warmup store
 - Back-up this repository** (highlighted in red)
 - Export duplicate statements
 - Subject, Predicate, Object and Graph (spog)
 - Delete duplicate statements
 - Suppress duplicate statements false
 - View/manage active transactions
 - Recognize geospatial datatypes automatically:
 - Control durability (bulk-load mode)
 - Manage triple attribute definitions and static filter
 - View/manage triple indices
 - Optimize the repository
 - Manage free-text indices
 - Materialize Entailed Triples
 - Delete Materialized Triples
 - Manage external Solr free-text indexer
 - Manage external MongoDB connection
 - Shutdown instance
- Tools:**
 - Explore kennedy using Gruff
- Reports:**
 - Storage report
 - Triple indices
 - String table
 - Full list of reports ...

Imagen 43. copia de seguridad

Otras operaciones en el repositorio

Como muestra la ilustración, hay muchos comandos que puede invocar para manipular la base de datos y los datos que contiene. Estos se describen en la sección Página de descripción general del repositorio de AGWebView.

También hay un icono a la derecha de las opciones de **Explorar el repositorio de Gruff**. Al hacer clic en esto, se abrirá una ventana del navegador Gruff. Gruff es un generador de consultas y visualización de gráficos diseñado para trabajar con AllegroGraph. Consulte el documento Gruff para obtener más información y la sección Gruff en un navegador en el documento AGWebView . No hablaremos más de Gruff en este documento.

Consultas

Una vez que se cargan los datos, desea consultar la base de datos. El *Kennedy* base de datos contiene datos públicos sobre las relaciones entre los individuos y las características de la familia de Joseph Kennedy, padre del presidente de Estados Unidos John F. Kennedy. Las consultas pueden incluir lo siguiente:

- ¿Quiénes de José y sus descendientes son varones y fueron a la Universidad de Harvard y cuáles son sus profesiones?
- ¿Quién es el descendiente vivo más antiguo (al menos cuando se creó la base de datos)?

Las consultas se pueden formular utilizando SPARQL, un lenguaje de consulta adaptado a bases de datos RDF. Tomemos una consulta muy simple: las personas descritas en la base de datos de Kennedy están etiquetadas `<http://www.franz.com/simple#personN>y <http://www.franz.com/simple#person1>` es Joseph Patrick Kennedy, el patriarca de la familia. Puede ver esto examinando las primeras líneas del archivo de datos.

Entonces, ¿cuáles son los predicados y los objetos de todos los triples cuyo sujeto es `<http://www.franz.com/simple#person1>`? La consulta SPARQL

```
select ?p ?o WHERE {<http://www.franz.com/simple#person1> ?p ?o}
```

dirá, como se muestra en la siguiente ilustración (seleccionamos **Nuevo** en el menú **Consultas** en la barra de menú, y escribimos la consulta en el cuadro **Editar consulta** y hacemos clic en el botón **Ejecutar**).

AllegroGraph WebView repository kennedy

Repository | Queries | Utilities | Admin | User test Documentation

Edit query

```
1 select ?p ?o WHERE {<http://www.franz.com/simple#person1> ?p ?o}
```

Language: SPARQL
 Limit to 1000 results
 Reasoning
 Long parts
 Graphs
 Show namespaces
 Add a namespace
 Edit initfile
[Permalink to query](#)

Execute Log Query Show Plan Save as Add to repository

23 Results in 3.652 ms Information

p	o
rdf:type	person
first-name	"Joseph"
middle-initial	"Patrick"
last-name	"Kennedy"
suffix	"none"
alma-mater	Harvard
birth-year	"1888"
death-year	"1969"
sex	male
spouse	person2
has-child	person3
has-child	person4
has-child	person6

Download using format SPARQL JSON

Imagen 44. Consulta AllegroGraph

¿Quiénes son los varones entre Joseph Kennedy y sus descendientes que fueron a la Universidad de Harvard y cuáles son sus respectivas profesiones? Aquí está la consulta SPARQL:

```
PREFIX o: <http://www.franz.com/simple#>

SELECT ?x ?p {
  ?x o:sex o:male ;
    o:alma-mater o:Harvard ;
    o:profession ?p .
}

ORDER BY ?p
```

Y la respuesta es 13, como se muestra en la consulta mostrada (también mostramos sus profesiones):

Edit query

```

1 PREFIX o: <http://www.franz.com/simple#>
2 SELECT ?x ?p {
3   ?x o:sex o:male ;
4     o:alma-mater o:Harvard ;
5     o:profession ?p .
6 }
7 ORDER BY ?p

```

Language: SPARQL ▾

Limit to 1000 results

Reasoning

Long parts

Graphs

Show namespaces

Add a namespace

Edit initfile

[Permalink to query](#)

Execute Log Query Show Plan Save as Add to repository

13 Results in 40.599 ms Information

x	p
person1	ambassador
person17	army
person47	attorney
person1	banker
person47	broadcaster
person4	congressman
person3	navy
person4	navy
person4	president
person1	producer
person47	professor
person4	senator
person17	senator

Download using format SPARQL JSON ▾

Imagen 45. Ejemplo 1 de consulta

¿Quién es la persona viva de mayor edad entre Joseph Kennedy y sus descendientes (en el momento en que se creó la base de datos, hace varios años)? Aquí está la consulta SPARQL. Tenga en cuenta el uso de una cláusula OPCIONAL. Eso produce vinculaciones de resultados para las personas con un año de nacimiento, tengan o no también un año de muerte. La cláusula FILTER luego filtra los resultados que tienen un año de muerte.

```

PREFIX o: <http://www.franz.com/simple#>

SELECT ?person ?birthYear {
  ?person o:birth-year ?birthYear
  OPTIONAL {
    ?person o:death-year ?deathYear
  }
  FILTER( !bound( ?deathYear ) )
} order by ?birthYear

LIMIT 1

```

Como solo nos interesa el más antiguo, agregamos una `LIMIT 1` cláusula a la consulta SPARQL. Ejecutamos la consulta y encontramos que el resultado es `person10`:

The screenshot shows the AllegroGraph WebView interface. At the top, it says "AllegroGraph WebView" and "repository kennedy". Below that is a navigation bar with "Repository | Queries | Utilities | Admin | User test" and "Documentation". The main area is titled "Edit query" and contains a text editor with the following SPARQL query:

```

1 PREFIX o: <http://www.franz.com/simple#>
2 SELECT ?person ?birthYear {
3   ?person o:birth-year ?birthYear
4   OPTIONAL {
5     ?person o:death-year ?deathYear
6   }
7   FILTER( !bound( ?deathYear ) )
8   } order by ?birthYear
9 LIMIT 1

```

On the right side, there are settings for the query: "Language: SPARQL", "Limit to 1000 results" (checked), "Reasoning" (unchecked), "Long parts" (unchecked), and "Graphs" (unchecked). There are also buttons for "Show namespaces", "Add a namespace", "Edit initfile", and a "Permalink to query" link.

Below the query editor, there are buttons for "Execute", "Log Query", "Show Plan", "Save as", and "Add to repository".

The results section shows "1 Result in 4.248 ms" and a "Warnings" button. The results are displayed in a table:

person	birthYear
person10	"1915"

At the bottom, there is a "Download" button and a "using format" dropdown menu set to "SPARQL JSON".

Imagen 46. Ejemplo 2 de consulta

¿Y quién es `person10`? Podríamos usar la misma consulta que hicimos anteriormente para `person1`:

```
select ?p ?o WHERE {<http://www.franz.com/simple#person10> ?p
?o}
```

y obtenemos todos los triples donde `person10` es el sujeto, pero solo nos interesan los nombres (primero, segundo y último), por lo que restringimos la consulta a eso (el segundo nombre se muestra en el `middle-initial` campo con el nombre incorrecto):

```
PREFIX o: <http://www.franz.com/simple#>

SELECT ?fn ?mn ?ln

WHERE {o:person10 o:first-name ?fn ;
        o:middle-initial ?mn ;
        o:last-name ?ln .

```

```
}
}
```

AllegroGraph WebView repository kennedy

← | Repository | Queries | Utilities | Admin | User test Documentation

Edit query

```
1 PREFIX o: <http://www.franz.com/simple#>
2   SELECT ?fn ?mn ?ln
3   WHERE {
4     o:person10 o:first-name ?fn ;
5               o:middle-initial ?mn ;
6               o:last-name ?ln .
7 }
```

Language: SPARQL ▼
Planner: default ▼
Result limit: 100 ▼
Reasoning
Long parts
Graphs
Show namespaces
Add a namespace
Edit initfile
Copy link to query

Execute Log Query Show Plan Save as Add to repository

1 Result in 0.500 ms **Warnings**

fn	mn	ln
"Robert"	"Sargent"	"Shriver"

Imagen 47. Resultado de la consulta

person10 es Robert Sargent Shriver, comúnmente conocido como Sargent Shriver, primer director del Cuerpo de Paz. (El Sr. Shriver murió en 2011, pero eso fue después de que se construyó la base de datos).

Tenga en cuenta que hay una pestaña de **Advertencias** en lugar de una pestaña de **Información**. Las advertencias sobre la consulta se muestran allí si las hay (si no hay advertencias, la pestaña se etiqueta **Información**). Si miramos las advertencias vemos:

Edit query

```
1 PREFIX o: <http://www.franz.com/simple#>
2 SELECT ?fn ?mn ?ln
3   WHERE {o:person10 o:first-name ?fn ;
4           o:middle-initial ?mn ;
5           o:last-name ?ln .
6 }
```

Language: SPARQL ▼
 Limit to 1000 results
 Reasoning
 Long parts
Show namespaces
Add a namespace
Edit initfile
[Permalink to query](#)

Execute Log Query Show Plan Save as Add to repository

1 Result in 0.896 ms Warnings

Query Information
warn

joinCrossProduct	Plan has 2 JOIN cross-products. These joins were not mergable: 1: join LHS: ?fn, ?mn; RHS: ?ln 2: join LHS: ?fn; RHS: ?mn	
time		
query		0.000471
plan		0.000425
total		0.000895
memory		
maximumMap		0
maximumChunk		11600000
other		
verb		select

Download using format SPARQL JSON ▼

Imagen 48. Información de la consulta

Se le advierte que la consulta tiene un producto cruzado (lo que significa una combinación sin variables comunes, lo que puede llevar mucho tiempo y producir resultados voluminosos). El resto de la pantalla es la información.

Aunque no es obvio, técnicamente es cierto que la consulta tiene un producto cruzado, ya que el plan de la consulta, que se muestra en el botón **Mostrar plan**, es

SPARQL select query plan

Query had warnings

Variables

?fn ?mn ?ln

Query Information

warn	joinCrossProduct	Plan has 2 JOIN cross-products. These joins were not mergable: 1: join LHS: ?fn, ?mn; RHS: ?ln 2: join LHS: ?fn; RHS: ?mn
time	plan	0.002539
	total	0.002539
other	verb	select

Plan

```
(:join
 (:lhs
 (:join
 (:lhs
 (:graph current
 (:triple-query <http://www.franz.com/simple#person10 http://www.franz.com/simple#first-name ?fn>
 :estimate 1.0133333 :cursor-estimate 0.5066666)))
 (:rhs
 (:graph current
 (:triple-query <http://www.franz.com/simple#person10 http://www.franz.com/simple#middle-initial ?mn>
 :estimate 5.5066667 :cursor-estimate 5.5066667))))))
 (:rhs
 (:graph current
 (:triple-query <http://www.franz.com/simple#person10 http://www.franz.com/simple#last-name ?ln>
 :estimate 5.5066667 :cursor-estimate 5.5066667))))))
```

Imagen 49. Warnings de la consulta

De hecho, no hay variables compartidas. A pesar de que indica una advertencia, se puede argumentar que la advertencia debe suprimirse si el LHS o el RHS tienen solo una fila, ya que en ese caso no es probable que haya un problema. En esta impresión, se señala la advertencia. Eso podría cambiar en el futuro. Pero es útil comprender que pueden aparecer advertencias y se pueden ver, si las hay, haciendo clic en **Mostrar plan** antes de la ejecución.

Esta ha sido una breve introducción a SPARQL pero muestra su flexibilidad y poder. Hay más ejemplos en el Tutorial de SPARQL .

Otra documentación

Si ha seguido los pasos de este documento, debería tener AllegroGraph instalado, al menos una base de datos cargada, y debería haber preguntado y recibido respuestas sobre esa base de datos. Consulte el índice de documentación para obtener más información sobre AllegroGraph.

9. ANEXO III: Instalación VirtualMachine

Introduction

AllegroGraph Server runs natively on Linux x86-64. To run AllegroGraph on other operating systems (such as Windows or MacOS) we suggest you set up a Linux virtual machine, as described in this document, or use our Docker container image.

There are also cloud service options such as AllegroGraph on AWS Marketplace.

For the Linux virtual machine, we provide a VMware virtual machine image to help facilitate this installation or you can create one on your own. The AllegroGraph virtual machine can be downloaded from the AllegroGraph download page.

Performance in the virtual machine will be slower than running natively, so we encourage you to install AllegroGraph natively for performance evaluation.

Guide to the AllegroGraph Virtual Machine

Download the Virtual Machine Image

The AllegroGraph virtual machine image can be downloaded from the AllegroGraph download page. After download, unzip the file into your desired location on your computer.

NOTE: The native Windows unzip for many versions of Windows cannot unzip the downloaded folder due to its size. You will need to use a free program like 7-Zip.

Download and install the VMware Player

- 32/64-bit Windows: VMware Workstation Player
- 32/64-bit Mac OS X (Intel): VMware Fusion

Launch VMware Player and start the VM

- Launch VMware Player

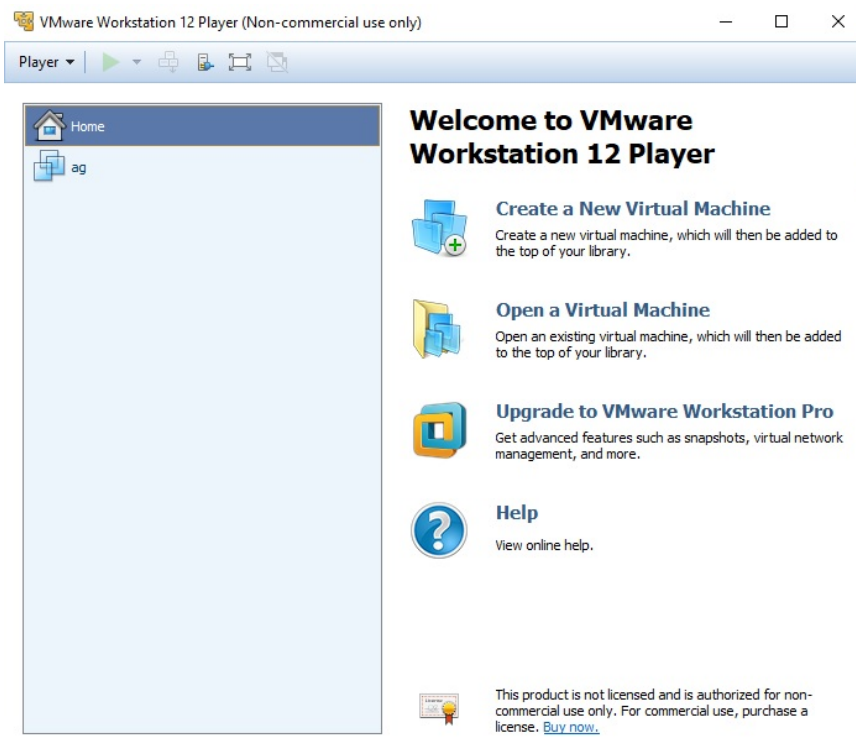


Imagen 50. Interfaz VMware VirtualMachine

Click on "Open a Virtual Machine"

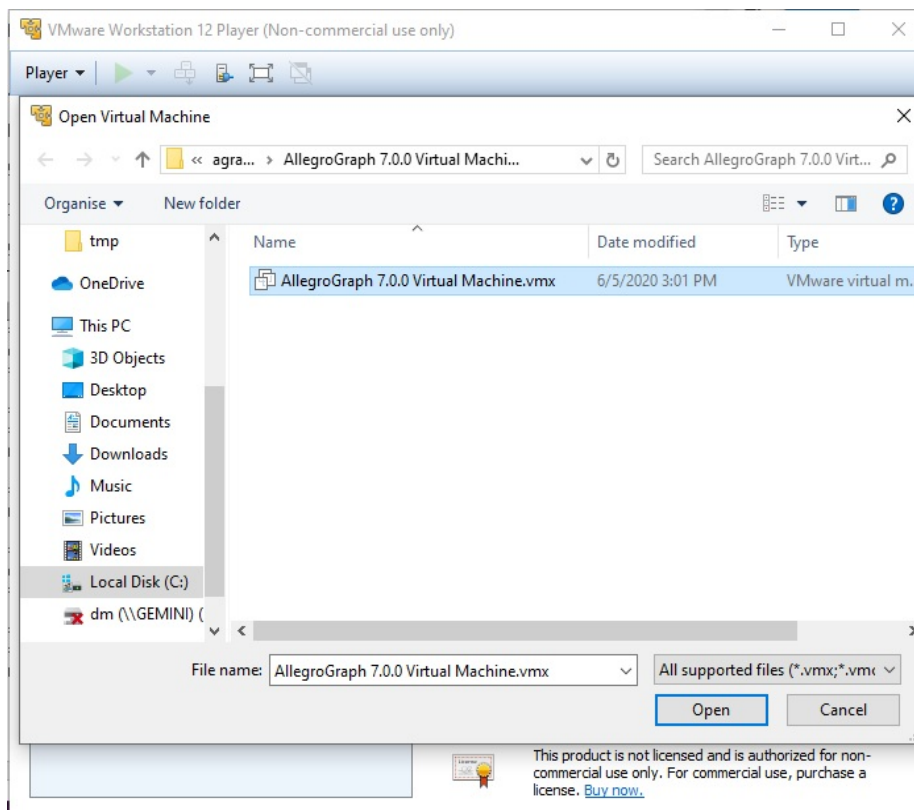


Imagen 51. Archivo AllegroGraph

- Navigate to the **AllegroGraph Virtual Machine.vmx** file in the directory where you unzipped the AllegroGraph download (it is shown in the file dialog in the illustration). Open this file. Take ownership if prompted.
- Click on "Play virtual machine". If on Windows you get a message saying (roughly) **VMware Player and Device/Credential Guard are not compatible**, please see Appendix 1: Device/Credential Guard on Windows below.
- If prompted for **Moved** or **Copied**, select **Copied**.

Once the VM starts playing, you should see this screen:

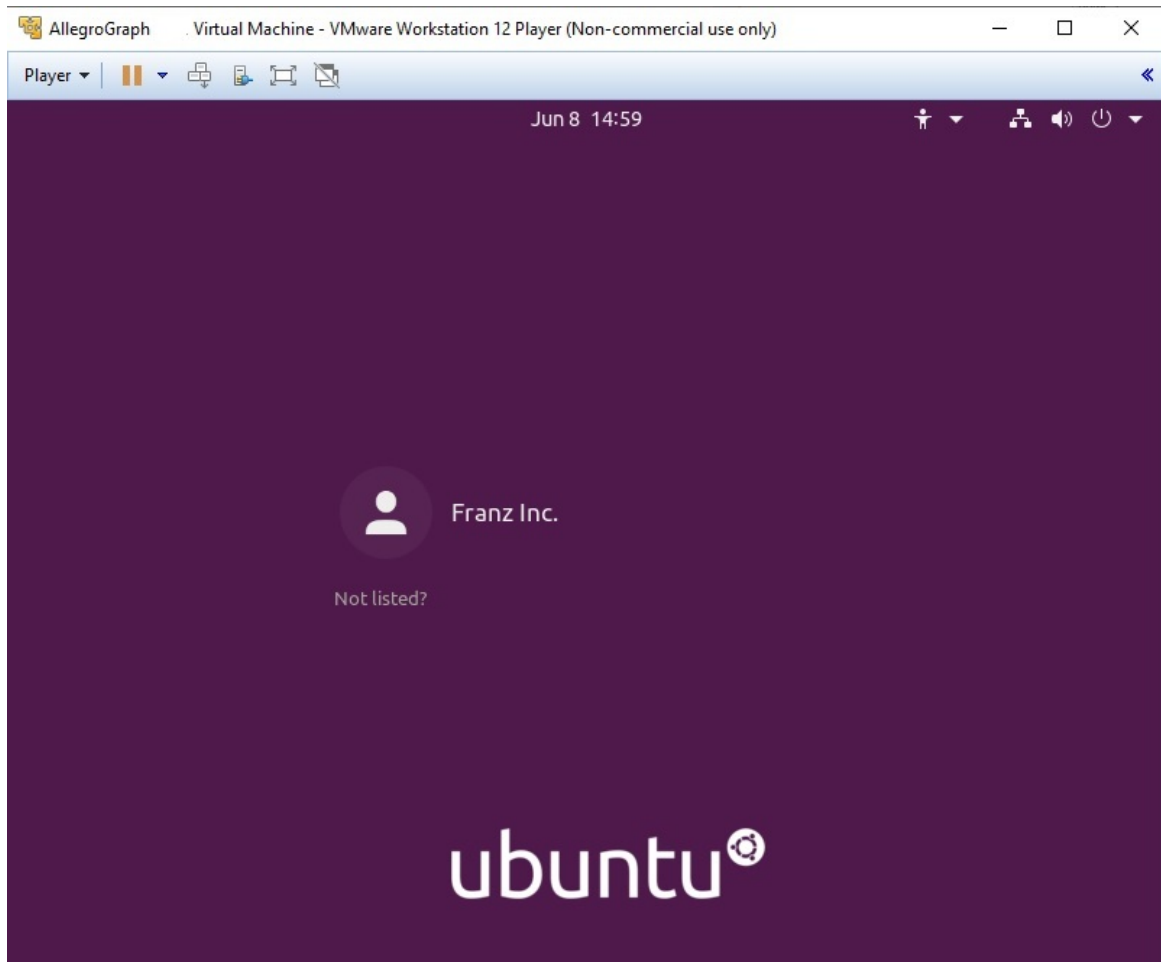


Imagen 52. Inicio sesión Ubuntu

Double-click on **Franz Inc.** and you will be prompted for a password. The password is **allegrograph**. Once you have entered the password, you will see this screen:

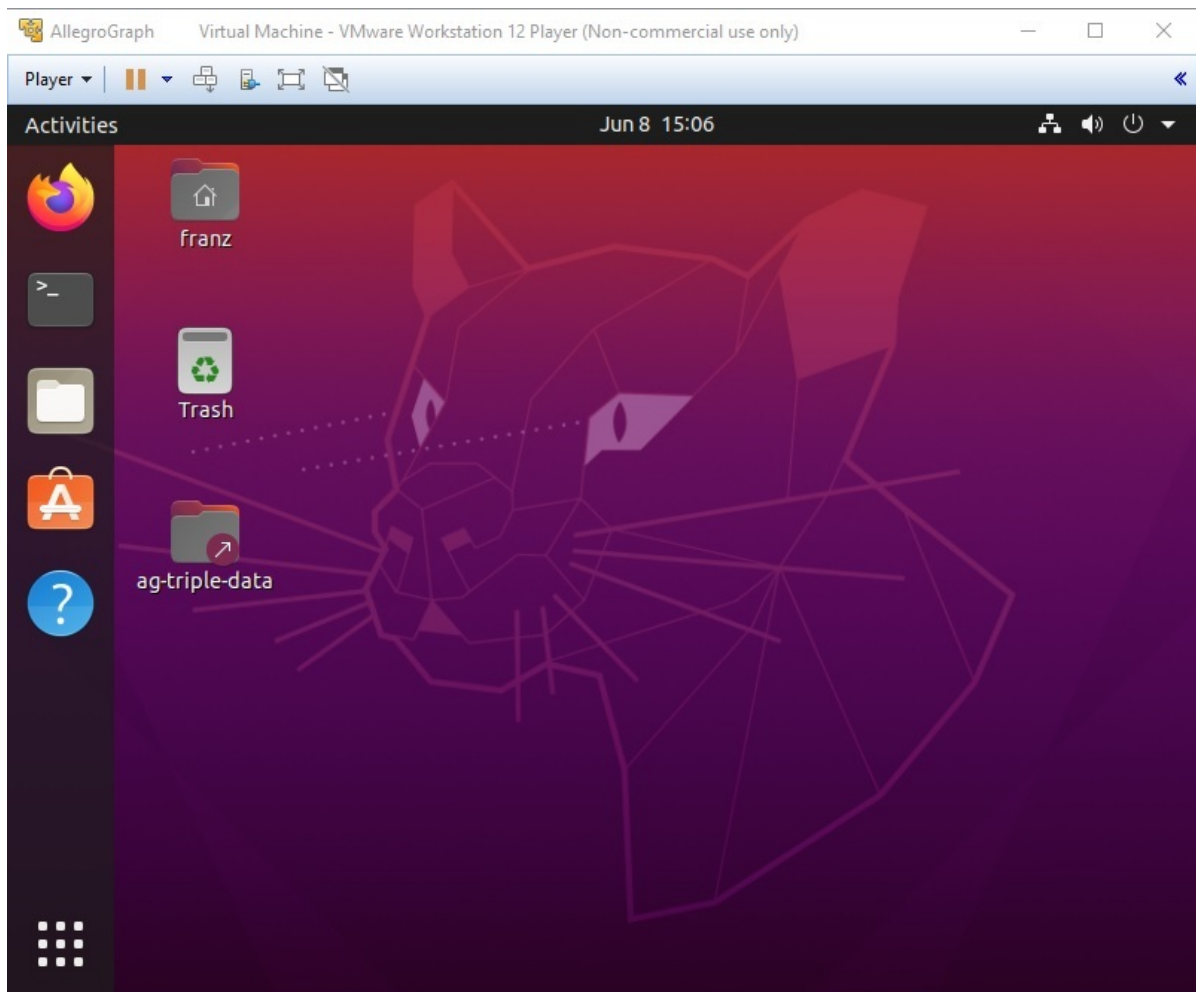


Imagen 53. Escritorio Ubuntu

There is an icon bar to the left and several icons in the main area. The icons in the icon bar are:

- **Firefox:** double-click to display a Firefox browser.
- **Terminal:** displays a UNIX shell. You type commands to the prompt to start and stop AllegroGraph. See Using the shell below for more information.
- **File window:** displays files in the local filesystem.
- **Ubuntu software:** displays links to available Ubuntu software.
- **Ubuntu help program**

In the main area, there are icons for the **Franz** directory file window, the **Trash** folder, and the **ag-triple-data** folder, which contains various example datasets.

Using the shell

You use the shell to start and stop AllegroGraph, to determine the VM's IP address, and to run programs like the **vi** editor. Display the shell by double-clicking on the terminal icon in the icon bar.

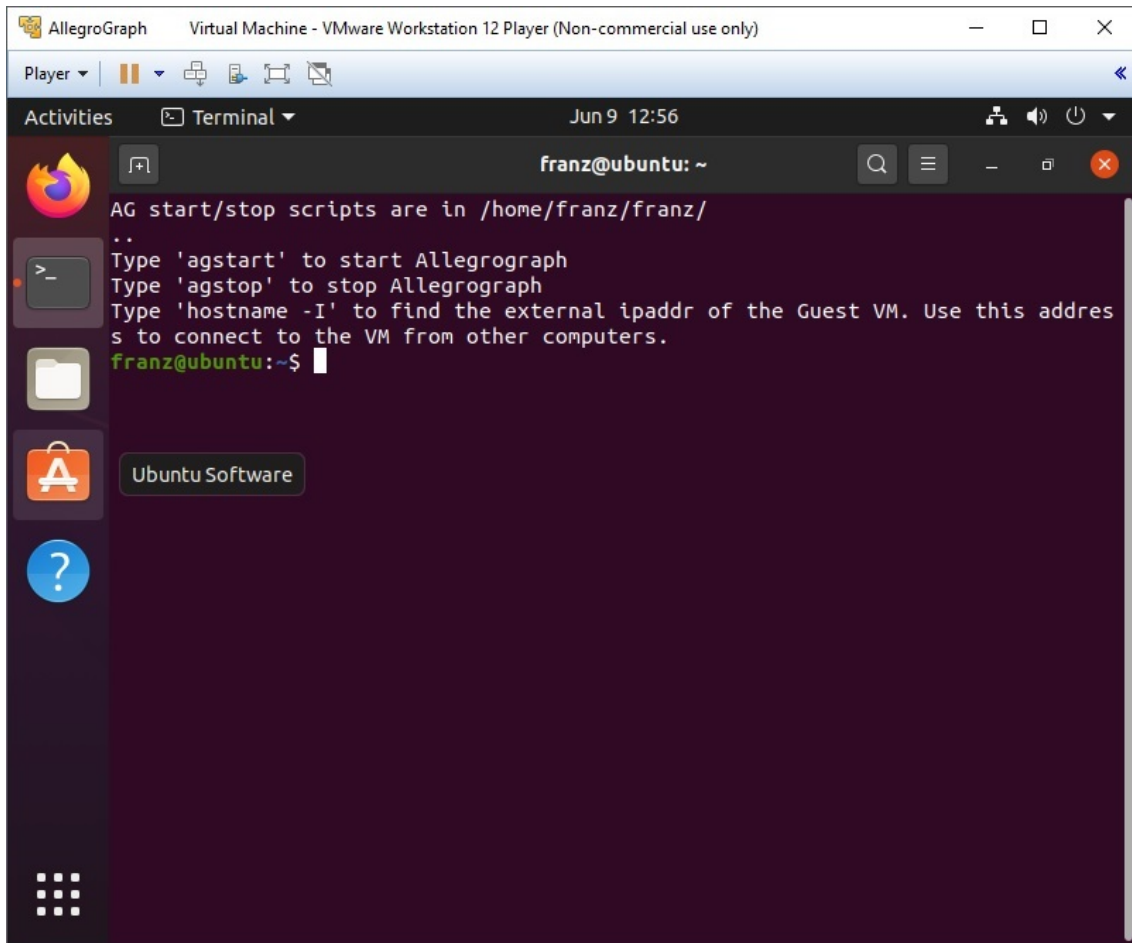


Imagen 54. Consola Ubuntu

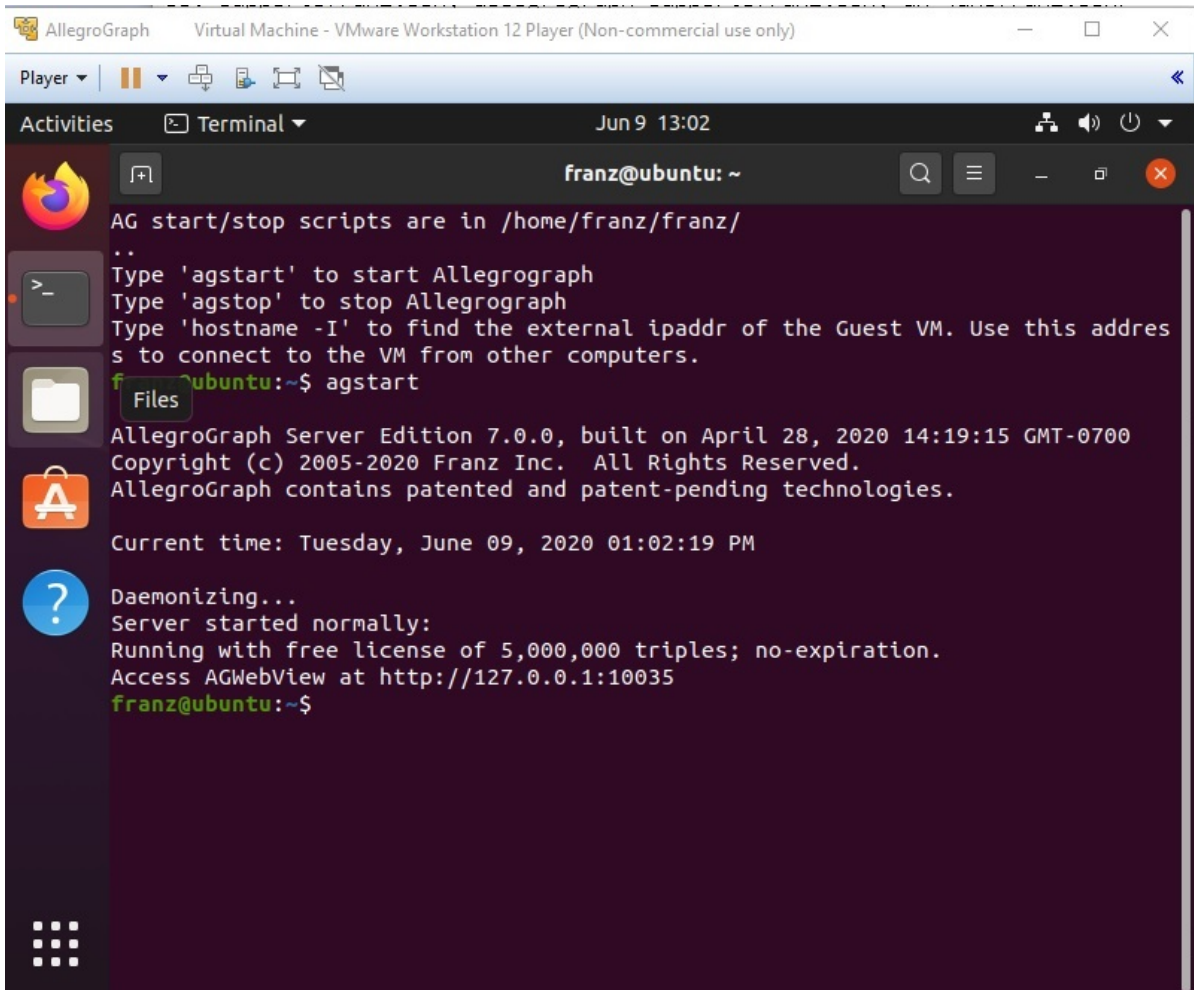
The illustration show the shell when it comes up. Three predefined commands are displayed as a startup message:

- Type 'agstart' to start Allegrograph
- Type 'agstop' to stop Allegrograph
- Type 'hostname -I' to find the external ipaddr of the Guest VM. Use this address to connect to the VM from other computers.

All these commands will work regardless of the current directory in the shell.

Starting AllegroGraph

Type **agstart** to the shell prompt:



```
AllegroGraph Virtual Machine - VMware Workstation 12 Player (Non-commercial use only)
Player | [Icons]
Activities Terminal Jun 9 13:02
franz@ubuntu: ~
AG start/stop scripts are in /home/franz/franz/
..
Type 'agstart' to start Allegrograph
Type 'agstop' to stop Allegrograph
Type 'hostname -I' to find the external ipaddr of the Guest VM. Use this address to connect to the VM from other computers.
franz@ubuntu:~$ agstart
AllegroGraph Server Edition 7.0.0, built on April 28, 2020 14:19:15 GMT-0700
Copyright (c) 2005-2020 Franz Inc. All Rights Reserved.
AllegroGraph contains patented and patent-pending technologies.

Current time: Tuesday, June 09, 2020 01:02:19 PM

Daemonizing...
Server started normally:
Running with free license of 5,000,000 triples; no-expiration.
Access AGWebView at http://127.0.0.1:10035
franz@ubuntu:~$
```

Imagen 55. Iniciando AllegroGraph

The server starts and runs in the background. Because we are using the default *agraph.cfg* which does not include a license, we are restricted to 5 million triples. The local port (10035) is reported. See [Using AGWebView](#) below for information on using [AGWebview](#) to control AllegroGraph.

The **agstart** command works regardless of the current directory of the shell.

Stopping AllegroGraph

You stop AllegroGraph with the **agstop** command. This illustration show **agstop** being called (at the bottom) just below the call to **hostname** described next.

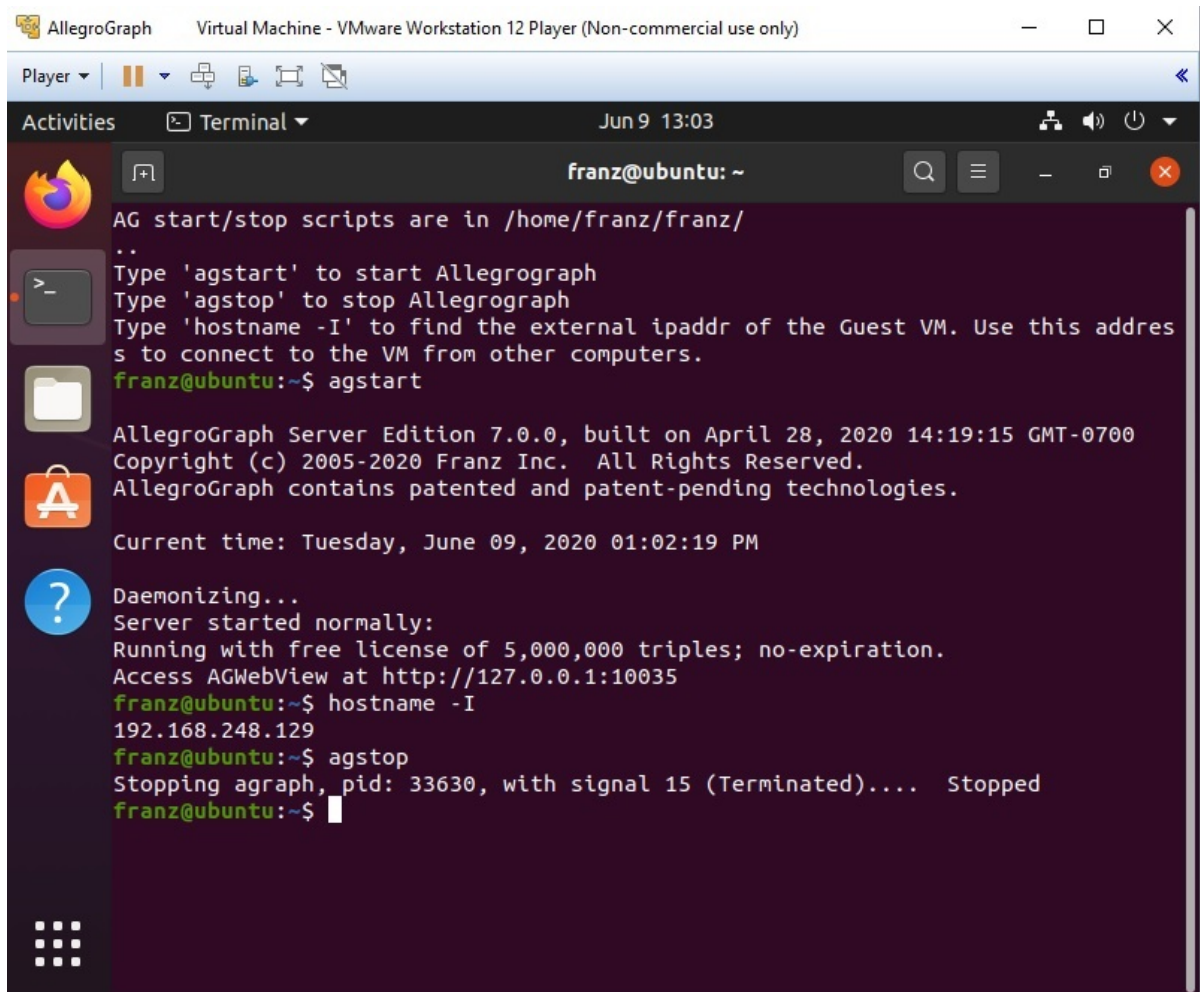


Imagen 56. Parando AllegroGraph

Determining the IP address

The third predefined command listed when the shell is opened is **hostname**. When called with the **-I** argument it returns the hostname which can be used by browsers outside the VM for connecting with the AllegroGraph server. See the Remote Connection to the VM for information on allowing connections to the VM just from the local machine (running the VM) or from any accessible machine. (That depends on using a **NAT** or a **Bridged** connection. The VM uses a **NAT** connection when initially started.)

Once you have the IP address, you can use AGWebview to control AllegroGraph.

Using AGWebView

Using **hostname -I** we determined that the IP address, usable at least on the machine running the VM, was 192.168.248.128 (you will likely get a different IP address). AllegroGraph by default listens on port 10035. So bringing up a browser on the local machine and entering 192.168.248.128:10035 will open AGWebView:

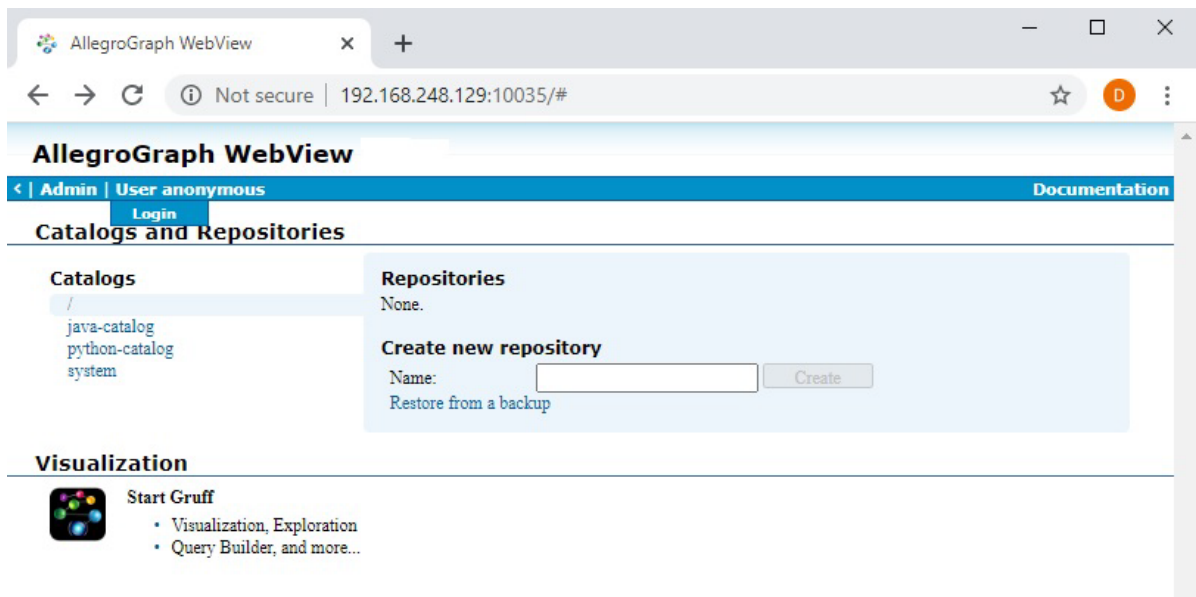


Imagen 57. Inicio AllegroGraph

Anonymous login is permitted by default in the VM AllegroGraph so AGWebView typically comes up with the Anonymous login. Clicking on **User Anonymous** in the menu bar displays a **login** choice (shown in the picture) and choosing that displays the usual AllegroGraph login prompt:

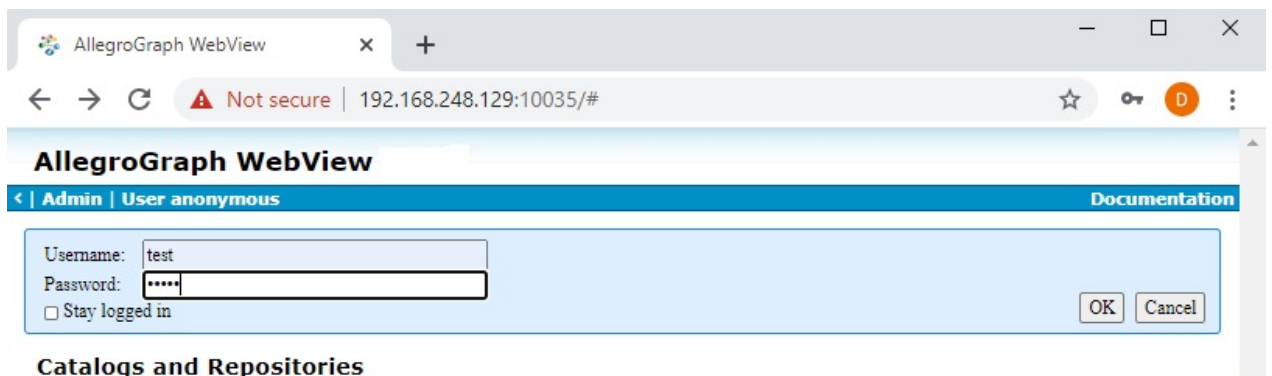


Imagen 58. Usuario y Password

See the AGWebview document for information on controlling AllegroGraph with AGWebView.

The agrapg.cfg file

AllegroGraph is customized by its configuration file *agraph.cfg*. That file is located in the */home/franz/ag/lib/* directory. It can be opened with the **vi** built in editor with the command **vi /home/franz/ag/lib/agraph.cfg** as we have done in the illustration:

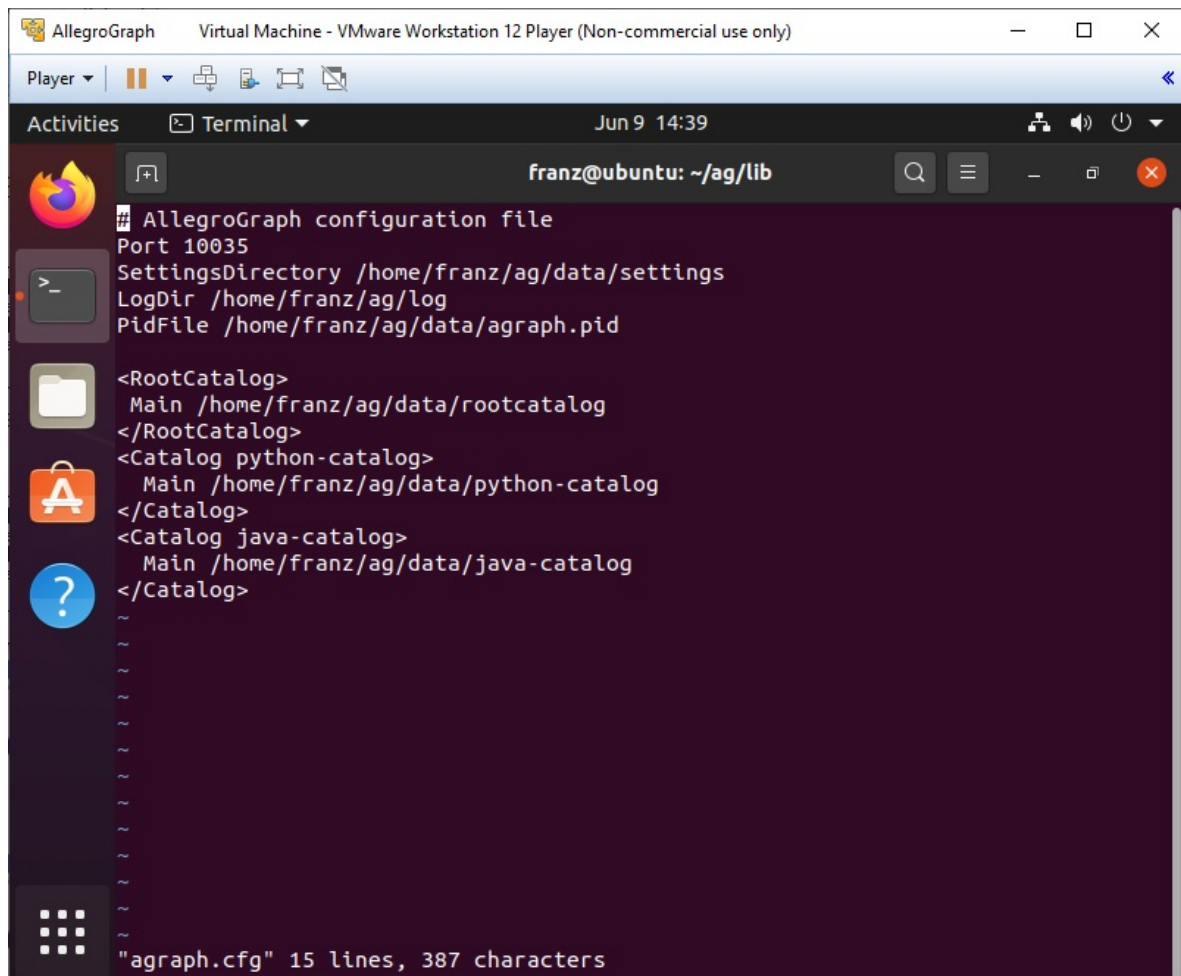


Imagen 59. Archivo de configuración AllegroGraph

See the Server Configuration and Control document for a list of configuration directives. If you have an AllegroGraph license, it goes at the end of the file.

agtool

agtool is a general command utility for AllegroGraph, see the agtool General Command Utility document. It is located in the `/home/franz/ag/bin/` directory.

Running Gruff in the VM

Gruff is a graph visualization and graphical query builder designed to work with AllegroGraph. GRUFF can be run in a browser. Open AGWebView (here we have the anonymous login shown above):

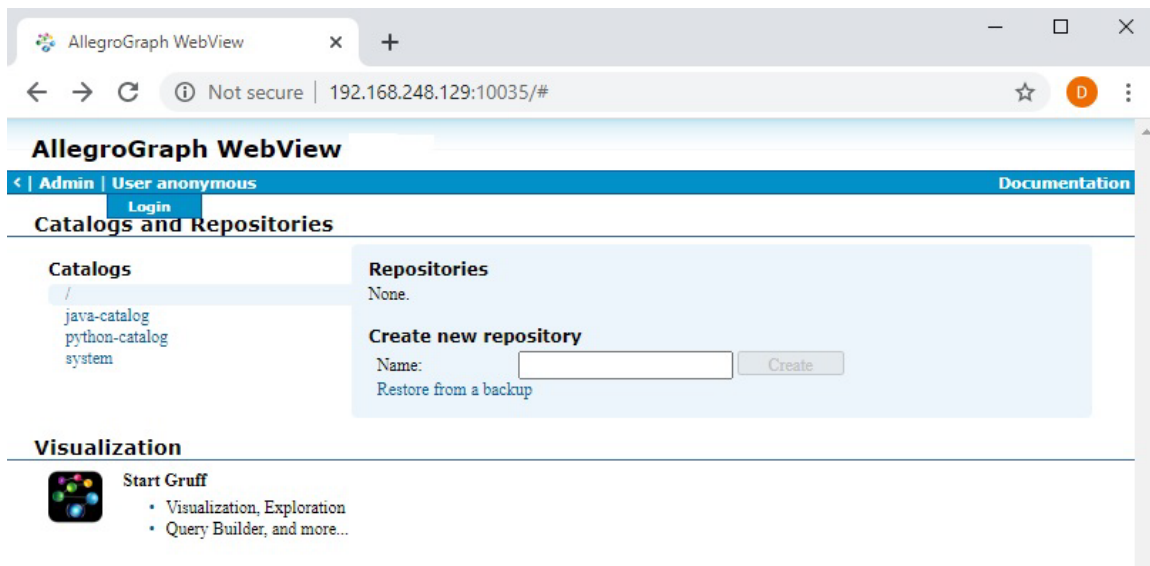


Imagen 60. Start Gruff

Clicking on the GRUFF icon on this page, and also on a repository page, will bring GRUFF up in a your browser. See the Gruff in AllegroGraph document.

The Wilburwine preloaded dataset

The AllegroGraph VM has various example data files located in the `/home/franz/franz/ag-triple-data/` directory. The file is named `wilburwine.ntriples.gz`. You can load this file into AllegroGraph with the command

```
• /home/franz/ag/bin/agtool load wilburwine  
/home/franz/franz/ag-triples-data/wilbutwine.ntriples.gz
```

The same directory also contains other example datasets.

Remote Connection to the VM

NAT vs. Bridged Network Connections

The AllegroGraph VM is configured with a network adaptor that uses NAT. In this configuration, the VM shares the IP address of the host server. Only the host machine (and possibly other VMs running on the same host) will be able to connect to the AllegroGraph VM. No other machines on the host network will be able to connect to the AllegroGraph VM.

In Bridged mode, the AllegroGraph VM and the host machine will share the host's network connection, but the VM will attempt to join the same network that the host belongs to. In this configuration, any machine on the same network as the host will also be able to connect to the AllegroGraph VM.

To change the type of network connection, make sure your AllegroGraph VM is selected and open the **Player | Manage** menu and select **Virtual Machine Settings**.

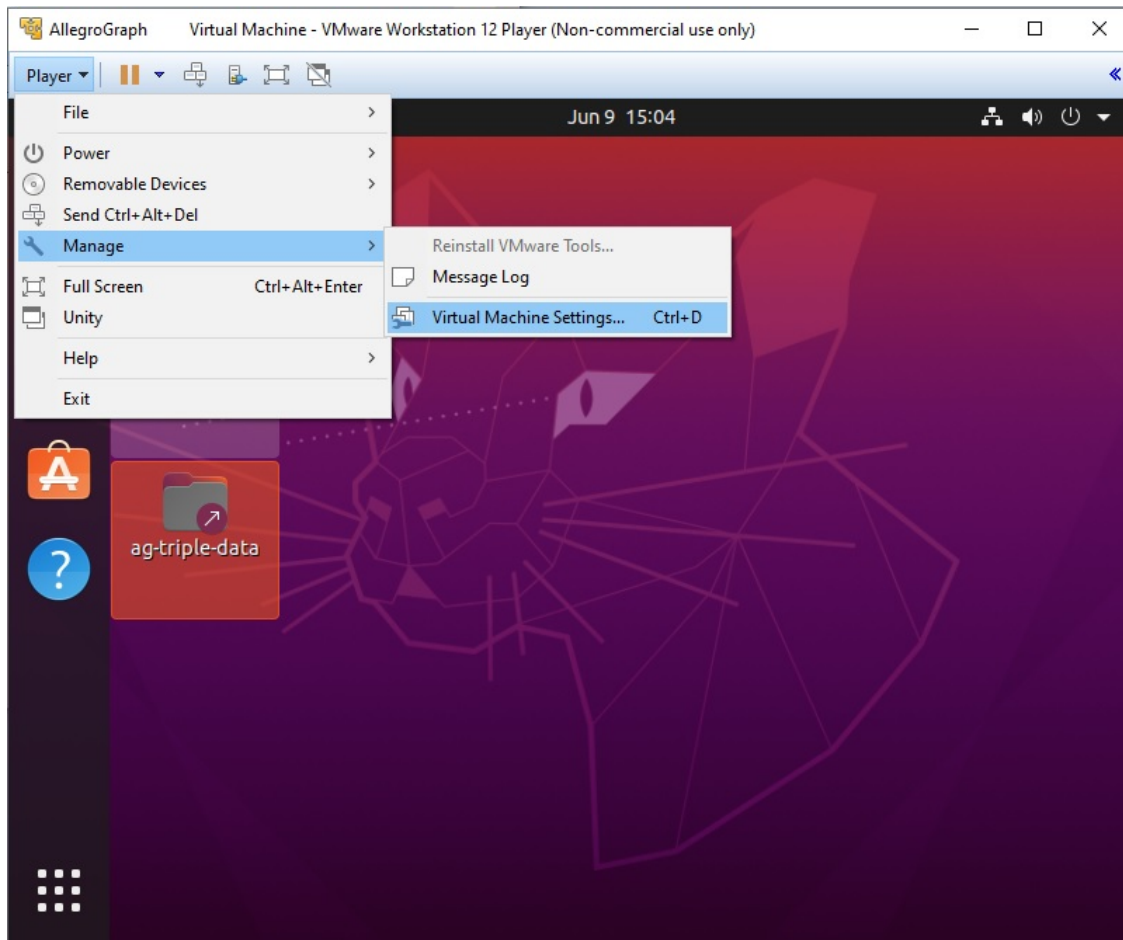


Imagen 61. VirtualMachine Settings

This displays the settings dialog. To connect from a remote machine, you want to turn off NAT and enable Bridged instead. (The illustrations are from a VM running on Windows.) We have selected the **Network Adaptor** setting in the illustration.

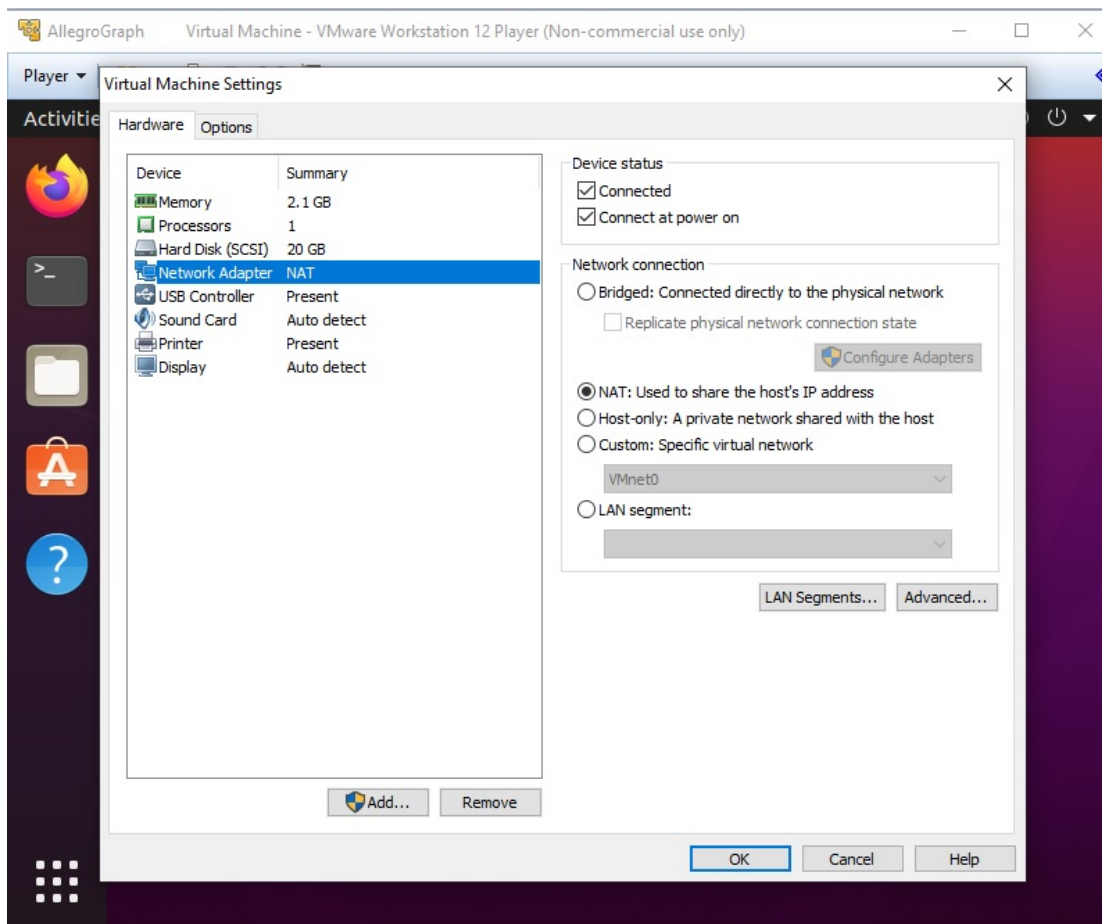


Imagen 62. Configuración Máquina Virtual

This dialog is from Windows. The Mac dialog is similar but has somewhat different wording.

Windows

Select "Bridged: Connect directly to the physical network." Check the option that says "Replicate the physical network connectivity state." Close the dialog box and reboot the virtual machine.

Getting back to the NAT setting

Select instead **NAT: Used to share the host's IP address**. Close the dialog box and reboot the virtual machine.

Appendix 1: Device/Credential Guard on Windows

Note that this information, which was correct at the time this document was written, discusses MicroSoft Windows, VMware, and Docker, none of which are products of Franz Inc. It is possible that details will have changed and the advice given here might be out of date. The problem is not related to Franz Inc. products.

A feature of Windows, *Windows Device Credential Guard*, is not compatible with VMware. Trying to run a virtual machine on WMware when that feature is enabled will fail with a message similar to the following:

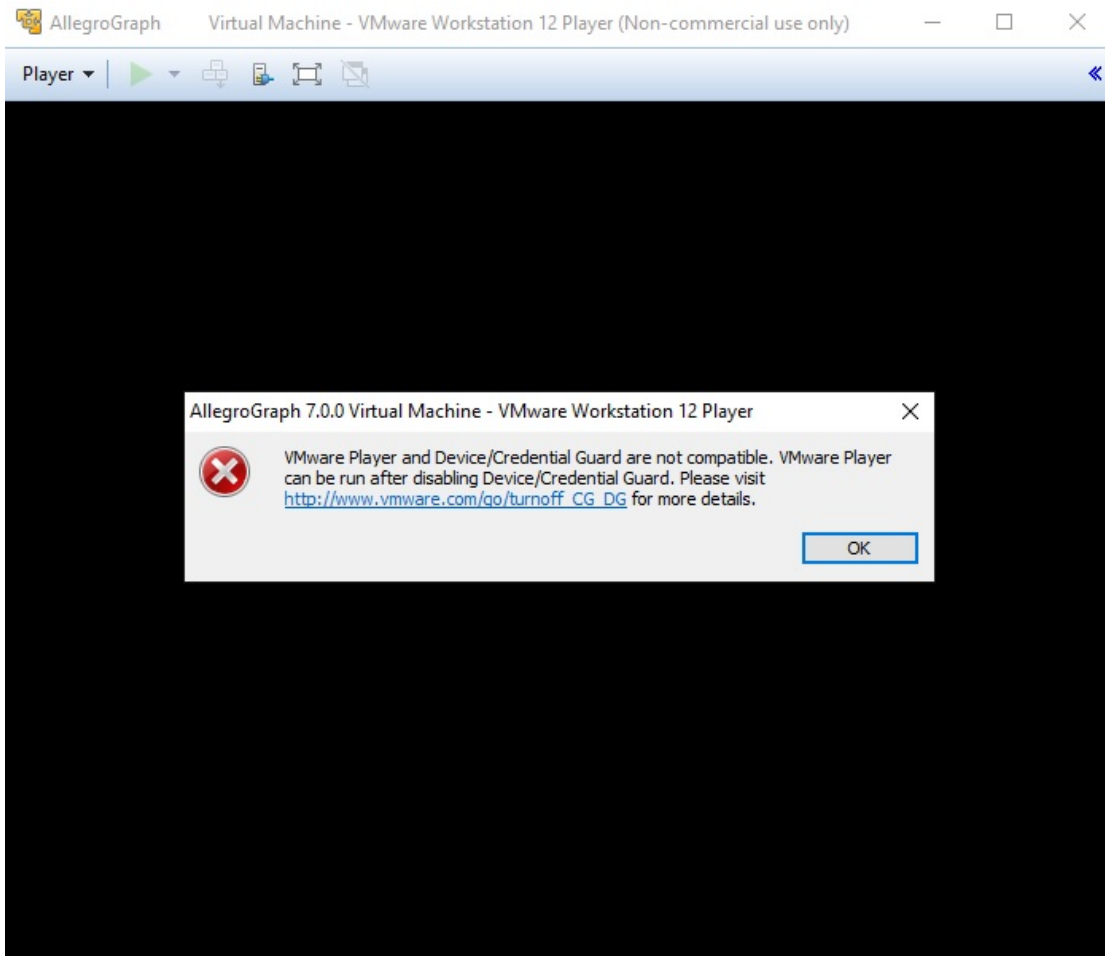


Imagen 63. Problemas de compatibilidad

The solution is to run, at an Administrator command prompt,

- `bcdedit /set hypervisorlaunchtype off`

and to restart your computer. Note that Docker requires that the Device Credential Guard be enabled, which means you apparently cannot run Docker and VMware at the same time. You re-enable the Device Credential Guard by running, at an Administrator command prompt,

- `bcdedit /set hypervisorlaunchtype auto`

and restarting your computer.

To repeat the note at the top of this appendix, this information is about products which are not products of Franz Inc., and while what is said here was tested and correct at the time of writing, things may have changed since. Users should contact the various vendors of the mentioned software products for more information.

10. ANEXO IV: Tripletas y Grafos RAT

10.1 Servicios Sociales

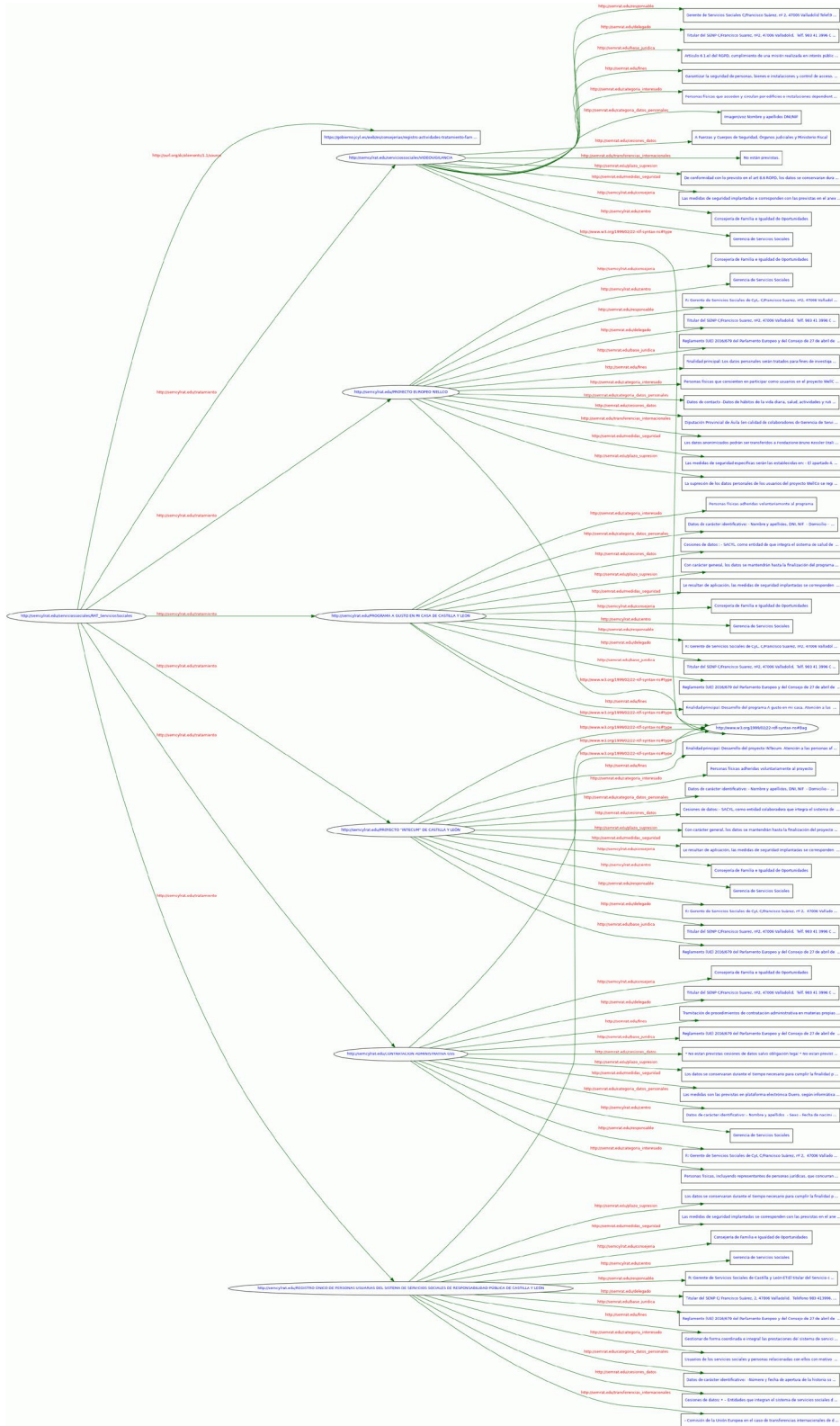


Imagen 64. Grafo Servicios Sociales

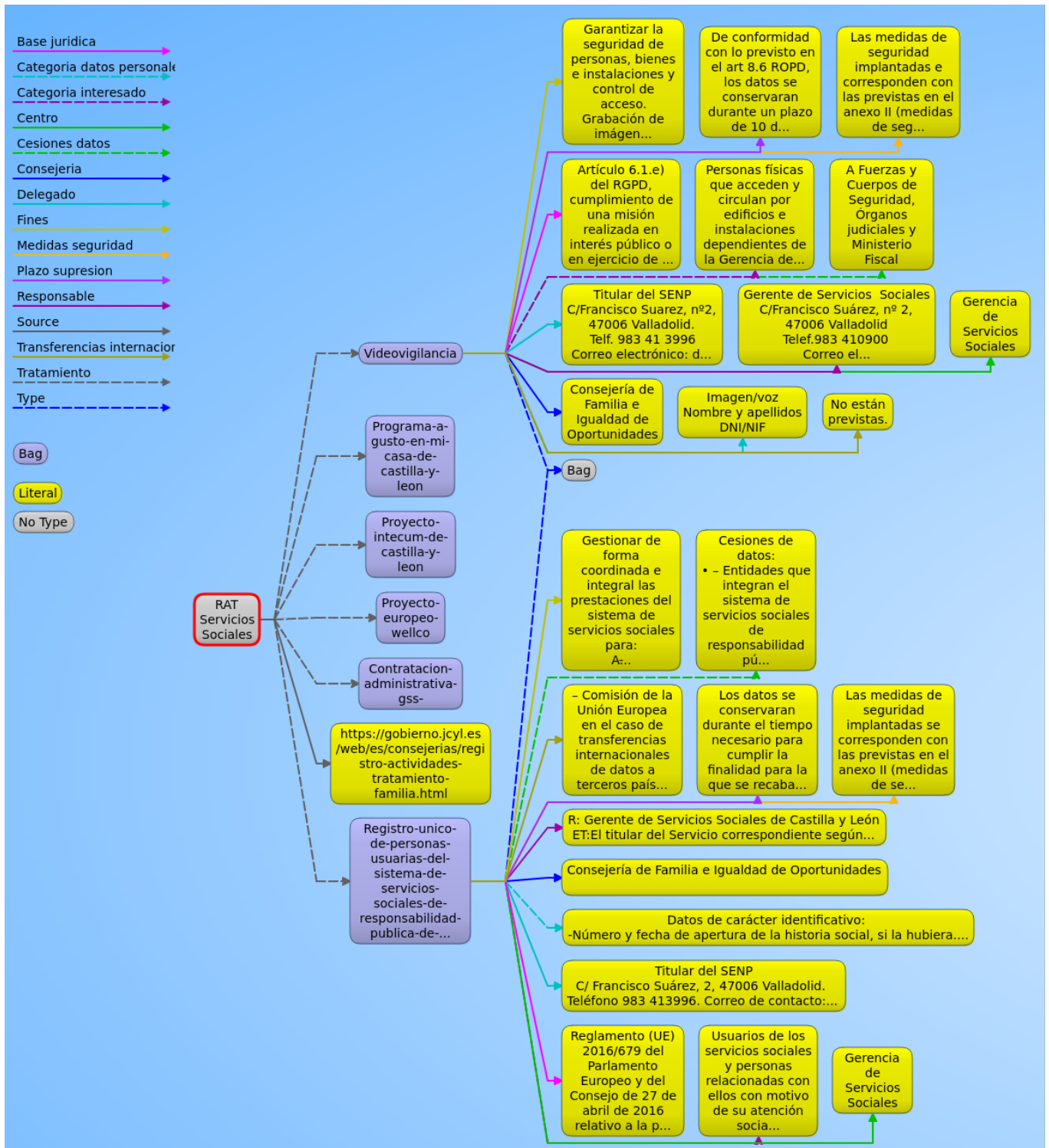


Imagen 65. Grafo AllegroGraph Servicios Sociales

Subject	Predicate	Object
http://semcytrat.edu/serviciosociales/RAT_ServiciosSociales	http://purl.org/dc/elements/1.1/source	"https://gobierno.jcyl.es/web/es/consejerias/registro-actividades-tratamiento-familia.html"
http://semcytrat.edu/REGISTRO_ÚNICO_DE_PERSONAS_USUARIAS_DEL_SISTEMA_DE_SERVICIOS_SOCIALES_DE_RESPONSABILIDAD_PÚBLICA_DE_CASTILLA_Y_LEÓN	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat.edu/REGISTRO_ÚNICO_DE_PERSONAS_USUARIAS_DEL_SISTEMA_DE_SERVICIOS_SOCIALES_DE_RESPONSABILIDAD_PÚBLICA_DE_CASTILLA_Y_LEÓN	http://semcytrat.edu/consejeria	"Consejería de Familia e Igualdad de Oportunidades"
http://semcytrat.edu/REGISTRO_ÚNICO_DE_PERSONAS_USUARIAS_DEL_SISTEMA_DE_SERVICIOS_SOCIALES_DE_RESPONSABILIDAD_PÚBLICA_DE_CASTILLA_Y_LEÓN	http://semcytrat.edu/centro	"Gerencia de Servicios Sociales"

<p>http://semcylrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/responsable</p>	<p>"R: Gerente de Servicios Sociales de Castilla y León ET:El titular del Servicio correspondiente según materia y/o entidad prestadora de la atención social correspondiente del sistema de servicios sociales de responsabilidad publica. Gerente de Servicios Sociales C/Francisco Suárez, nº 2, 47006 Valladolid Telef.983 410900 Correo electrónico: gerente.servicios.sociales@jcyL.es"</p>
<p>http://semcylrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/delegado</p>	<p>"Titular del SENP C/ Francisco Suárez, 2, 47006 Valladolid. Teléfono 983 413996. Correo de contacto: dpd.fio@jcyL.es"</p>

<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/base_juridica</p>	<p>"Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, en cumplimiento de una obligación legal y asimismo, en ejercicio de una misión realizada en interés público o en ejercicio de poderes públicos. Artículo 6.1.e) y c). Igualmente, en cumplimiento de obligaciones en el ámbito de la protección social y para la gestión del sistema de asistencia social. Artículo 9.2. b) y h) del RGPD. Ley 16/2010, de 20 de diciembre, de Servicios Sociales de Castilla y León. Artículos 34, 35, 42 y 43 Ley 4/2018, de 2 de julio, de ordenación y funcionamiento de la de la Red de protección e inclusión de las personas y familias en situación de mayor vulnerabilidad social o económica en Castilla y León. Artículo 30."</p>
--	--	---

<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/fines</p>	<p>"Gestionar de forma coordinada e integral las prestaciones del sistema de servicios sociales para: -Atención social integral. -Detección, acceso, provisión y seguimiento de las prestaciones sociales. -Gestión de peticiones de cita previa e información. -Coordinación con otras AA.PP. y entidades colaboradoras. - Elaboración de informes de carácter asistencial y social. - Fines estadísticos"</p>
<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/categoria_interesa do</p>	<p>"Usuarios de los servicios sociales y personas relacionadas con ellos con motivo de su atención social."</p>

<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/categoria_datos_personales</p>	<p>"Datos de carácter identificativo: -Número y fecha de apertura de la historia social, si la hubiera. -Nombre y apellidos, DNI, NIE, NIF, pasaporte u otros documentos de identidad; fecha y lugar de nacimiento; sexo; nacionalidad, estado civil y fecha de fallecimiento, en su caso. -Firma o huella. -Domicilio (nombre de vía, número, localidad, municipio, provincia, país, código postal). -Datos de contacto (teléfono fijo, teléfono móvil, correo electrónico, fax). - Número de seguridad social o mutualidad. - Imagen y voz. -Identificación de la zona de acción social y el ayuntamiento o la diputación provincial titular del centro de acción social correspondiente al domicilio. -Profesional de referencia de los servicios sociales que coordina la intervención. -Datos sobre guarda; tutela; profesión; situación laboral, escolar y económica, financieros, bancarios y seguros, nivel educativo, estado civil y de vida; condiciones básicas de salud; etnia; personas cuidadoras, personas de contacto del solicitante; datos relacionados con el uso, abuso y dependencia de tabaco, alcohol y otras drogas. - Datos relativos a la unidad de convivencia familiar, integrada por las personas que conviven con la persona interesada y guardan con ella algún tipo de parentesco o afinidad que se consideren relevantes para la intervención social. -Solicitudes de servicios sociales. -Datos sobre</p>
--	--	---

		<p>actuaciones de órganos jurisdiccionales, de salud, de empleo, de educación o de cuerpos y fuerzas de seguridad del estado. -Datos relativos al origen de las derivaciones asistenciales: Centro o servicio, profesional, teléfono, fecha de primer contacto y fecha de derivación. Datos de la atención: -Valoraciones técnicas e instrumentos de valoración, particularmente, en materia de infancia, menores infractores, familia, violencia de género, dependencia, discapacidad, exclusión social, drogodependencia y otras áreas de interés. -Objetivos de la asistencia. - Prestaciones recibidas del sistema de servicios sociales de responsabilidad pública y otras prestaciones y actuaciones complementarias. - Resultado de intervenciones y evaluación de la adecuación de las prestaciones a las necesidades de la persona usuaria."</p>
--	--	---

<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/cesiones_datos</p>	<p>"Cesiones de datos: • – Entidades que integran el sistema de servicios sociales de responsabilidad pública de Castilla y León. • – Otros órganos de la Administración Autonómica de Castilla y León y del sector público de Castilla y León. • – Administración General del Estado y el sector público institucional dependiente de aquella, Administraciones autonómicas y locales y su sector público institucional respectivo. • – Entidades privadas proveedoras de servicios sociales y entidades colaboradoras con el sistema de servicios sociales de responsabilidad pública."</p>
<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/transferecias_internacionales</p>	<p>"– Comisión de la Unión Europea en el caso de transferencias internacionales de datos a terceros países respecto a datos referidos a las víctimas de violencia de género."</p>
<p>http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/plazo_supresion</p>	<p>"Los datos se conservarán durante el tiempo necesario para cumplir la finalidad para la que se recabaron y para determinar las posibles responsabilidades que se pudieran derivar de dicha finalidad y del tratamiento de los datos, en los términos del art. 8.6 RGPD. Sera de aplicación el Decreto 79/2015, de 17 de diciembre, por el que se regula la Historia Social Única, y la normativa en materia de archivos y documentación."</p>

http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN	http://semrat.edu/medidas_seguridad	<p>"Las medidas de seguridad implantadas se corresponden con las previstas en el anexo II (medidas de seguridad) del Real Decreto 3/2010, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica y que se encuentran descritas en los documentos que conforman la política de protección de datos y política de seguridad de la información de la Junta de Castilla y León, según se prevé en la creación del fichero protegido."</p>
http://semcytrat.edu/serviciossociales/RAT_ServiciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/REGISTRO ÚNICO DE PERSONAS USUARIAS DEL SISTEMA DE SERVICIOS SOCIALES DE RESPONSABILIDAD PÚBLICA DE CASTILLA Y LEÓN
http://semcytrat.edu/serviciossociales/VIDEOVIGILANCIA	http://semcytrat.edu/consejeria	<p>"Consejería de Familia e Igualdad de Oportunidades"</p>
http://semcytrat.edu/serviciossociales/VIDEOVIGILANCIA	http://semcytrat.edu/centro	<p>"Gerencia de Servicios Sociales"</p>
http://semcytrat.edu/serviciossociales/VIDEOVIGILANCIA	http://semrat.edu/responsable	<p>"Gerente de Servicios Sociales C/Francisco Suárez, nº 2, 47006 Valladolid Telef.983 410900 Correo electrónico: gerente.servicios.sociales@jcy.es"</p>
http://semcytrat.edu/serviciossociales/VIDEOVIGILANCIA	http://semrat.edu/delegado	<p>"Titular del SENP C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 3996 Correo electrónico: dpd.fio@jcy.es"</p>

http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/base_juridica	"Artículo 6.1.e) del RGPD, cumplimiento de una misión realizada en interés público o en ejercicio de poderes públicos. Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales. Ley 5/2014, de 4 de abril, de Seguridad Privada. ."
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/fines	"Garantizar la seguridad de personas, bienes e instalaciones y control de acceso. Grabación de imágenes de las personas que acceden y transitan por edificios e instalaciones."
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/categoria_interesado	"Personas físicas que acceden y circulan por edificios e instalaciones dependientes de la Gerencia de SS"
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/categoria_datos_personales	"Imagen/voz Nombre y apellidos DNI/NIF"
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/cesiones_datos	"A Fuerzas y Cuerpos de Seguridad, Órganos judiciales y Ministerio Fiscal"
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/transferencias_internacionales	"No están previstas."

http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/plazo_supresion	<p>"De conformidad con lo previsto en el art 8.6 ROPD, los datos se conservarán durante un plazo de 10 días, a contar desde la fecha de su recogida para videovigilancia. Los datos relativos a los accesos se conservarán durante el tiempo necesario para cumplir con la finalidad para la que se recabaron y determinar las posibles responsabilidades que se pudieran derivar de dicha finalidad y del tratamiento de los datos."</p>
http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA	http://semrat.edu/medidas_seguridad	<p>"Las medidas de seguridad implantadas e corresponden con las previstas en el anexo II (medidas de seguridad) del Real Decreto 3/20101, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica y que se encuentran descritas en los documentos que conforman la política de protección de datos y política de seguridad de la información de la Junta de Castilla y León, según se prevé en la Orden FAM/ 1053/2009, de creación del fichero protegido."</p>
http://semcytrat.edu/serviciosociales/RAT_ServiciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/serviciosociales/VIDEOVIGILANCIA

http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/plazo_supresion	<p>"La supresión de los datos personales de los usuarios del proyecto WellCo se regirá de conformidad con: - El "Acuerdo de Subvención de la Comisión Europea: Grant Agreement number: 769765 — WellCO — H2020-SC1-2016-2017/H2020-SC1-2017-CNECT-1". - El Acuerdo entre socios "WellCo Consortium Agreement". - Los apartados 3 "Duración" y 4 "Obligaciones del encargado del tratamiento" de las "Cláusulas contractuales para el encargado del tratamiento de datos del proyecto WellCo – Acuerdo de Subvención del Programa H2020 nº 769765" firmado entre HI Iberia Ingeniería y Proyectos SL (que gestiona la base de datos del proyecto) y la Gerencia de Servicios Sociales de Castilla y León."</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semcytrat.edu/consejeria	<p>"Consejería de Familia e Igualdad de Oportunidades"</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semcytrat.edu/centro	<p>"Gerencia de Servicios Sociales"</p>

http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/responsable	<p>"R: Gerente de Servicios Sociales de CyL. C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 410900 Correo electrónico: gerente.servicios.sociales@jcy1.es ET: El titular del Servicio de Atención a Personas Mayores y Prevención de Dependencia y entidades colaboradoras en el desarrollo del proyecto. C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 413877 Correo electrónico: GSSmayores@jcy1.es"</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/delegado	<p>"Titular del SENP C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 3996 Correo electrónico: dpd.fio@jcy1.es"</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/base_juridica	<p>"Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Artículo 6.1.a). Ley 16/2010, de 20 de diciembre, de servicio sociales de Castilla y León"</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/fines	<p>"Finalidad principal: Los datos personales serán tratados para fines de investigación científica de acuerdo con lo establecido en los artículos 5.1.b) y 89.1 del RGPD. Finalidad específica: Participar en el co-diseño y prueba de los prototipos a desarrollar en el proyecto europeo WellCo."</p>

http://semcylrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/categoria_interesado	"Personas físicas que consienten en participar como usuarios en el proyecto WellCo."
http://semcylrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/categoria_datos_personales	"Datos de contacto -Datos de hábitos de la vida diaria, salud, actividades y rutinas, gustos, preferencias, etc que los usuarios deseen proporcionar a partir de cuestionarios. -Datos de actividad física y descanso recogidos a través de cuestionarios y dispositivos electrónicos (relojes y teléfonos inteligentes). -Datos procesados a partir de grabaciones de audio, vídeo e imagen. I6"

http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/cesiones_datos	<p>"Diputación Provincial de Ávila (en calidad de colaboradores de Gerencia de Servicios Sociales), Universidad de Valladolid (en calidad de colaboradores de Gerencia de Servicios Sociales), Bravo Solution S.A (en calidad de adjudicataria del contrato de Servicio de Apoyo Técnico del proyecto H2020 WellCo) y las entidades que forman parte del proyecto WellCo: HI Iberia Ingeniería y Proyectos SL (España), Fondazione Bruno Kessler (Italia), Kobenhavns Universitet (Dinamarca), Jožef Stefan Institute (Eslovenia), ConnectedCare Services, B.V. (Países Bajos), Monsenso Aps (Dinamarca), Syddansk Universitet (Dinamarca) y la Fundación Acción Social y Tutela de Castilla y León (en calidad de Tercera Parte Vinculada en el marco de la ejecución del proyecto WellCo)."</p>
http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/transferencias_internacionales	<p>"Los datos anonimizados podrán ser transferidos a Fondazione Bruno Kessler (Italia), Kobenhavns Universitet (Dinamarca), Jožef Stefan Institute (Eslovenia), ConnectedCare Services, B.V. (Países Bajos), Monsenso Aps (Dinamarca), Syddansk Universitet (Dinamarca), entidades socias del proyecto WellCo fuera de España"</p>

http://semcytrat.edu/PROYECTO EUROPEO WELLCO	http://semrat.edu/medidas_seguridad	<p>"Las medidas de seguridad específicas serán las establecidas en: - El apartado 4.o "Obligaciones del encargado del tratamiento" de las "Cláusulas contractuales para el encargado del tratamiento de datos del proyecto WellCo – Acuerdo de Subvención del Programa H2020 nº 769765", que establece los mecanismos a implantar a tal fin. - El apartado 2.3.5.4 "Security Safeguards" del documento entregable del proyecto D7.1 – POPD – Requirement no.2. En cualquier, caso de forma genérica, le es aplicación, las medidas de seguridad implantadas se corresponden con las previstas en el anexo II (medidas de seguridad) del Real Decreto 3/20101, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica y que se encuentran descritas en los documentos que conforman la política de protección de datos y política de seguridad de la información de la Junta de Castilla y León"</p>
http://semcytrat.edu/serviciosociales/RAT_ServiciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/PROYECTO EUROPEO WELLCO
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semcytrat.edu/consejeria	<p>"Consejería de Familia e Igualdad de Oportunidades"</p>

http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semcytrat.edu/centro	"Gerencia de Servicios Sociales"
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/responsable	"R: Gerente de Servicios Sociales de CyL. C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 410900 Correo electrónico: gerente.servicios.sociales@jcy1.es ET: El titular del Servicio de Atención a Personas Mayores y Prevención de Dependencia y entidades colaboradoras en el desarrollo del proyecto. C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 413877 Correo electrónico: GSSmayores@jcy1.es"
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/delegado	"Titular del SENP C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 3996 Correo electrónico: dpd.fio@jcy1.es"
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/base_juridica	"Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Art 6.1.a) Ley 16/2010, de 20 de diciembre, de servicio sociales de Castilla y León"
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/fines	"Finalidad principal: Desarrollo del programa A gusto en mi casa. Atención a las personas, en su domicilio, hasta el final Finalidades Adicionales: Evaluación del programa A gusto en mi casa"

http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/categoria_interesa do	<p>"Personas físicas adheridas voluntariamente al programa"</p>
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/categoria_datos_personales	<p>"Datos de carácter identificativo: - Nombre y apellidos, DNI, NIF - Domicilio - Datos de contacto Datos de la atención: - Valoraciones técnicas e instrumentos de valoración, particularmente, en materia de dependencia, personas mayores y otras áreas de interés. - Objetivos de la asistencia. - Prestaciones recibidas del sistema de servicios sociales de responsabilidad pública y otras prestaciones y actuaciones complementarias. - Resultado de intervenciones y evaluación de la adecuación de las prestaciones a las necesidades de la persona usuaria. - Hábitos de la vida diaria, salud, actividades y rutinas, gustos, preferencias, etc que los usuarios deseen proporcionar a partir de cuestionarios. - Datos sobre su historia de vida - Datos de los expedientes de dependencia - Datos relativos a la unidad de convivencia, familiares, amigos y otras personas del entorno cercano con las que se relaciona la persona usuaria. - Datos procesados a partir de grabaciones de audio, vídeo e imagen."</p>

<p>http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/cesiones_datos</p>	<p>"Cesiones de datos : - SACYL, como entidad de que integra el sistema de salud de responsabilidad pública de Castilla y León. - Diputación provincial de Ávila, entidad colaboradora, responsable de los servicios sociales de la zona en la que se desarrolla el proyecto piloto - Universidad de Valladolid, institución responsable de la evaluación del programa y de la calidad de vida de las personas usuarias - PRONISA, entidad privada proveedora de servicios - ASPRODES, entidad privada que realiza funciones de apoyo técnico al programa - Plena Inclusión, entidad privada que difundirá información sobre los resultados del programa - Cruz Roja, entidad privada proveedora de servicios de teleasistencia, voluntariado y otros productos tecnológicos"</p>
--	--	---

http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/plazo_supresion	<p>"Con carácter general, los datos se mantendrán hasta la finalización del programa, según los convenios suscritos con las diferentes entidades. No obstante, en el caso de personas usuarias que causen baja voluntaria en el programa, no se obtendrán nuevos datos, pero se podrán utilizar los datos obtenidos a efectos de evaluación de resultados. Así mismo, con independencia de la finalización de la experiencia piloto, según los convenios firmados, se mantendrá el registro de datos de las personas usuarias hasta que finalice la intervención y los apoyos que se estén prestando."</p>
http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN	http://semrat.edu/medidas_seguridad	<p>"Le resultan de aplicación, las medidas de seguridad implantadas se corresponden con las previstas en el anexo II (medidas de seguridad) del Real Decreto 3/20101, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica y que se encuentran descritas en los documentos que conforman la política de protección de datos y política de seguridad de la información de la Junta de Castilla y León."</p>
http://semcytrat.edu/serviciosociales/RAT_ServiciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/PROGRAMA A GUSTO EN MI CASA DE CASTILLA Y LEÓN
http://semcytrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag

http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semcytrat.edu/consejeria	"Consejería de Familia e Igualdad de Oportunidades"
http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semcytrat.edu/centro	"Gerencia de Servicios Sociales"
http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semrat.edu/responsable	"R: Gerente de Servicios Sociales de CyL C/Francisco Suárez, nº 2, 47006 Valladolid E-mail: gerente.servicios.sociales@jcy.es ET: Titular del Servicio de Concertación C/ Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 22 61 E-mail: intecum@jcy.es"
http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semrat.edu/delegado	"Titular del SENP C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 3996 Correo electrónico: dpd.fio@jcy.es"
http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semrat.edu/base_juridica	"Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, art 6.1.a). Ley 16/2010, de 20 de diciembre, de servicio sociales de Castilla y León"
http://semcytrat.edu/PROYECTO “INTECUM” DE CASTILLA Y LEÓN	http://semrat.edu/fines	"Finalidad principal: Desarrollo del proyecto INTecum. Atención a las personas afectadas por una enfermedad en fase avanzada/terminal, en su domicilio, hasta el final de la vida Finalidades Adicionales: Evaluación del proyecto INTecum"

http://semcylrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN	http://semrat.edu/categoria_interesado	"Personas físicas adheridas voluntariamente al proyecto"
http://semcylrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN	http://semrat.edu/categoria_datos_personales	"Datos de carácter identificativo: - Nombre y apellidos, DNI, NIF - Domicilio - Datos de contacto Datos de la atención: - Valoraciones técnicas e instrumentos de valoración, particularmente, en materia de dependencia, personas mayores o con discapacidad y otras áreas de interés. - Objetivos de la asistencia. - Prestaciones recibidas del sistema de servicios sociales de responsabilidad pública y otras prestaciones y actuaciones complementarias. - Resultado de intervenciones y evaluación de la adecuación de las prestaciones a las necesidades de la persona usuaria. - Hábitos de la vida diaria, salud, actividades y rutinas, gustos, preferencias, etc que los usuarios deseen proporcionar a partir de cuestionarios. - Datos sobre su historia de vida - Datos de los expedientes de dependencia - Datos relativos a la unidad de convivencia, familiares, amigos y otras personas del entorno cercano con las que se relaciona la persona usuaria. - Datos procesados a partir de grabaciones de audio, vídeo e imagen"

<p>http://semcytrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN</p>	<p>http://semrat.edu/cesiones_datos</p>	<p>"Cesiones de datos: - SACYL, como entidad colaboradora que integra el sistema de salud de responsabilidad pública de Castilla y León, y prestadora de los cuidados sanitarios. - Diputación Provincial de Palencia, entidad colaboradora, responsable de los servicios sociales de la zona en la que se desarrolla el proyecto piloto. - Ayuntamiento de Palencia, entidad colaboradora, responsable de los servicios sociales de la zona en la que se desarrolla el proyecto piloto. - Universidad de Valladolid, institución responsable de la evaluación y seguimiento del proyecto y de la calidad de vida de las personas usuarias - FUNDACIÓN SAN CEBRIÁN, entidad privada proveedora de servicios. - ASPRODES, entidad privada que realiza funciones de apoyo técnico al proyecto. - PLENA INCLUSIÓN, entidad privada que difundirá información sobre los resultados del proyecto. - Cruz Roja, entidad privada proveedora de servicios de teleasistencia, voluntariado y otros productos tecnológicos."</p>
--	--	--

http://semcytrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN	http://semrat.edu/plazo_supresion	<p>"Con carácter general, los datos se mantendrán hasta la finalización del proyecto, según los convenios suscritos con las diferentes entidades. No obstante, en el caso de personas usuarias que causen baja voluntaria en el proyecto, no se obtendrán nuevos datos, pero se podrán utilizar los datos obtenidos a efectos de evaluación de resultados. Así mismo, con independencia de la finalización de la experiencia piloto, según los convenios firmados, se mantendrá el registro de datos de las personas usuarias hasta que finalice la intervención y los apoyos que se estén prestando."</p>
http://semcytrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN	http://semrat.edu/medidas_seguridad	<p>"Le resultan de aplicación, las medidas de seguridad implantadas se corresponden con las previstas en el anexo II (medidas de seguridad) del Real Decreto 3/20101, de 8 de enero, por el que se regula el Esquema Nacional de Seguridad en el ámbito de la Administración Electrónica y que se encuentran descritas en los documentos que conforman la política de protección de datos y política de seguridad de la información de la Junta de Castilla y León."</p>
http://semcytrat.edu/serviciosociales/RAT_Ser viciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/PROYECTO "INTECUM" DE CASTILLA Y LEÓN

http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/categoria_datos_personales	"Datos de carácter identificativo: - Nombre y apellidos - Sexo - Fecha de nacimiento - No se tratan datos de categoría especial"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semcytrat.edu/centro	"Gerencia de Servicios Sociales"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/responsable	"R: Gerente de Servicios Sociales de CyL C/Francisco Suárez, nº 2, 47006 Valladolid Telf.983 410900 E-mail: gerente.servicios.sociales@jcy1.es ET: Titular del Servicio de Contratación Administrativa y entidades adjudicatarias de los contratos"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Bag
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/categoria_interesado	"Personas físicas, incluyendo representantes de personas jurídicas, que concurren al procedimiento de contratación pública"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semcytrat.edu/consejeria	"Consejería de Familia e Igualdad de Oportunidades"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/delegado	"Titular del SENP C/Francisco Suarez, nº2, 47006 Valladolid. Telf. 983 41 3996 Correo electrónico: dpd.fio@jcy1.es"

http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/fines	<p>"Tramitación de procedimientos de contratación administrativa en materias propias de la CFIO y de la GSS. Los datos personales podrán ser tratados para fines de archivo en interés público fines de investigación científica, histórica o fines estadísticos, de acuerdo con lo establecido en los artículos 5.1b) y 89.1 del RGPD"</p>
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/base_juridica	<p>"Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Artículos 6.1.c) y e) Ley 9/2017 de 8 de noviembre de Contratos del Sector Público. La aportación de los datos es requisito indispensable para realizar los trámites y ser adjudicatario de los contratos advos. Ley 39/2015 de 1 de octubre del Procedimiento Administrativo Común de las Administraciones Publicas"</p>
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/cesiones_datos	<p>"* No están previstas cesiones de datos salvo obligación legal * No están previstas transferencias de datos a terceros países"</p>

http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/plazo_supresion	"Los datos se conservarán durante el tiempo necesario para cumplir la finalidad para la que se recabaron y para determinar las posibles responsabilidades que se pudieran derivar de dicha finalidad y del tratamiento de los datos. Sera de aplicación lo dispuesto en la normativa de archivos y documentación, Orden CYT/27/2012, de 10 de enero"
http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS	http://semrat.edu/medidas_seguridad	"Las medidas son las previstas en plataforma electrónica Duero, según informática corporativa de la Admón. de la Comunidad."
http://semcytrat.edu/serviciosociales/RAT_ServiciosSociales	http://semcytrat.edu/tratamiento	http://semcytrat.edu/CONTRATACION ADMINISTRATIVA GSS

Tabla 6. Tripletas Servicios Sociales

Estas tres imágenes, tanto los dos grafos como la tabla con las tripletas, muestran los resultados de la transformación del Registro de Actividades de Tratamiento de la Junta de Castilla y León *Servicios Sociales* a RDF.

Por motivos de tamaño se muestra en este documento solamente estos resultados como ejemplo, el resto de los resultados de los *RAT* transformados se acompañarán como archivos anexos a esta documentación.