



UNIVERSIDAD
DE MURCIA

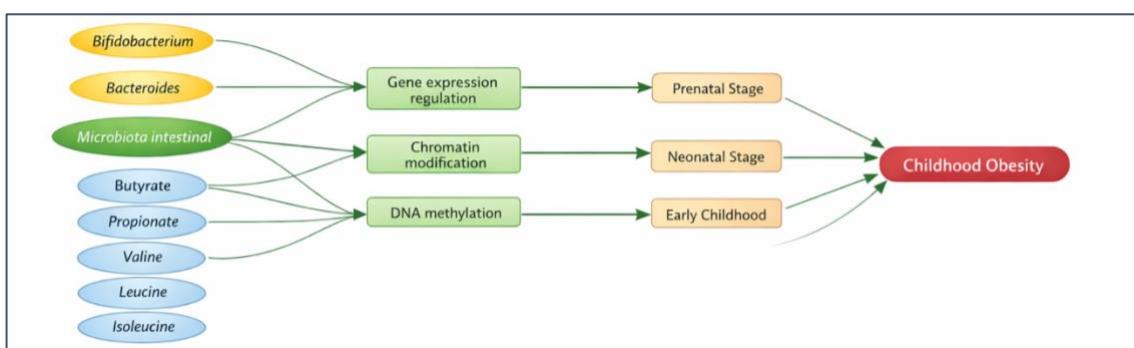


Facultad de Biología
Universidad de Murcia

TRABAJO FINAL EXPLOTACIÓN SEMÁNTICA DE DATOS

Máster en **Bioinformática**
Facultad de Biología – Universidad de Murcia

**Modelado y explotación semántica del eje
microbiota-epigenética-adiposidad temprana
mediante grafos RDF**



Autor: Irene Xue-Feng Lope Mateo

Curso 2025-2026

Tabla de contenido

1.	INTRODUCCIÓN.....	3
1.1.	Obesidad infantil como problema de salud pública	3
1.2.	Programación metabólica temprana y factores maternos.....	3
1.3.	Papel de la epigenética en la adiposidad temprana	3
1.4.	Microbiota intestinal y metabolitos en etapas tempranas.....	4
1.5.	Eje microbiota-metabolitos-epigenoma en el eje madre-infante	4
1.6.	Necesidad de enfoques integradores y motivación del trabajo	4
2.	OBJETIVOS.....	4
3.	MATERIALES Y MÉTODOS.....	5
3.1.	Fuentes de conocimiento.....	5
3.2.	Definición de prefijos y espacios de nombres.....	5
3.3.	Definición y modelado de las entidades del dominio.....	5
3.4.	Definición y justificación de propiedades y relaciones semánticas	5
3.5.	Modelado de las tripletas biológicas.....	7
3.6.	Validación sintáctica y comprobación del dataset RDF	8
3.7.	Carga del dataset RDF y configuración del entorno Blazegraph	9
3.8.	Generación y ejecución de consultas SPARQL desde R	10
4.	Resultados y discusión.....	10
4.1.	Resultados de la validación del dataset RDF.....	10
4.2.	Comprobación de la carga correcta del dataset en Blazegraph	11
4.3.	Resultados de las consultas SPARQL.....	12
	Consulta 1: Producción metabólica asociada a la microbiota intestinal	12
	Consulta 2: Metabolitos que regulan procesos epigenéticos	12
	Consulta 3: Procesos epigenéticos activos en etapas tempranas del desarrollo.....	13
	Consulta 4: Transmisión madre-infante de influencias biológicas	13
	Consulta 5: Rutas completas desde metabolitos microbianos hasta obesidad.....	14
5.	Conclusiones.....	14
6.	Bibliografía.....	15

1. INTRODUCCIÓN

1.1. Obesidad infantil como problema de salud pública

La obesidad infantil constituye en la actualidad uno de los principales problemas de salud pública a nivel mundial, con una prevalencia creciente tanto en países desarrollados como en regiones en vías de desarrollo. De acuerdo con la Organización Mundial de la Salud, decenas de millones de niños menores de cinco años presentan sobrepeso u obesidad, lo cual se asocia con un mayor riesgo de desarrollar patologías metabólicas crónicas en etapas posteriores de la vida, incluyendo diabetes mellitus tipo 2, hipertensión arterial, dislipidemia y enfermedad hepática metabólica. Esto conlleva un notable impacto clínico y socioeconómico a largo plazo (World Health Organization, 2020; Jebeile *et al.*, 2022; Hamilton *et al.*, 2018).

1.2. Programación metabólica temprana y factores maternos

El desarrollo de la obesidad infantil responde a una etiología compleja y multifactorial, en la que interactúan factores genéticos, ambientales, nutricionales y del desarrollo. Numerosos estudios han demostrado que características tempranas como la adiposidad neonatal, el tamaño corporal al nacimiento y la velocidad de crecimiento durante los primeros meses de vida están estrechamente relacionadas con un mayor riesgo de obesidad y de alteraciones metabólicas en la infancia y la edad adulta. Asimismo, se ha observado que los hijos de mujeres con obesidad pregestacional o con un aumento excesivo de peso durante el embarazo presentan una mayor probabilidad de desarrollar adiposidad temprana, lo que pone de manifiesto la relevancia del entorno materno en la programación metabólica del descendiente (Koontz *et al.*, 2014; Moore *et al.*, 2020; Carlsen *et al.*, 2014; Renault *et al.*, 2015).

1.3. Papel de la epigenética en la adiposidad temprana

En este contexto, el concepto de programación metabólica temprana ha adquirido una importancia creciente. Las exposiciones ambientales críticas durante períodos sensibles del desarrollo, como la gestación y los primeros años de vida, pueden inducir cambios biológicos persistentes que condicionan la salud metabólica futura. Entre los mecanismos propuestos para explicar este fenómeno, los procesos epigenéticos destacan como un elemento central. Las modificaciones epigenéticas, especialmente los cambios en los patrones de metilación del ADN, permiten regular la expresión génica sin alterar la secuencia del genoma y actúan como un nexo funcional entre el entorno y el fenotipo. Diversos estudios han relacionado alteraciones epigenéticas tempranas con un mayor riesgo de ganancia rápida de peso y obesidad a largo plazo, subrayando la influencia de factores maternos y ambientales durante el embarazo y la primera infancia sobre el epigenoma del niño (Robinson *et al.*, 2021; Lesseur *et al.*, 2014; Vahamiko *et al.*, 2019; Jonsson *et al.*, 2021).

1.4. Microbiota intestinal y metabolitos en etapas tempranas

La microbiota intestinal desempeña un papel clave en la regulación metabólica desde etapas muy tempranas de la vida, y alteraciones en su composición durante el primer año pueden influir en el riesgo de obesidad futura. Gran parte de su efecto se ejerce a través de la producción de metabolitos bioactivos, como ácidos grasos de cadena corta, aminoácidos de cadena ramificada y ácidos biliares, que modulan rutas metabólicas y epigenéticas especialmente relevantes durante el desarrollo temprano (Milani et al., 2017; Bäckhed et al., 2015; Dogra et al., 2015; Kumar et al., 2014; Cuevas-Sierra et al., 2019; Wahlström et al., 2016; Li et al., 2024; Li et al., 2025).

1.5. Eje microbiota-metabolitos-epigenoma en el eje madre-infante

El estudio de estas interacciones adquiere una dimensión adicional en el contexto del eje madre-infante. Evidencias recientes indican que la microbiota, el metaboloma y el epigenoma maternos pueden influir de forma directa o indirecta en los patrones biológicos del recién nacido, contribuyendo a la transmisión intergeneracional del riesgo metabólico. La integración de información procedente de la madre y del hijo a lo largo de distintas etapas del desarrollo temprano permite abordar el fenómeno de la adiposidad infantil desde una perspectiva longitudinal y sistémica, tal y como propone el estudio prospectivo de la cohorte española NEMO (Suárez-Cortés et al., 2025; Franzago et al., 2024; Pauwels et al., 2019).

1.6. Necesidad de enfoques integradores y motivación del trabajo

La integración de información procedente de múltiples dominios biológicos plantea importantes retos debido a la heterogeneidad de terminologías, que dificulta su análisis mediante enfoques tradicionales. En este contexto, las tecnologías de la Web Semántica y los grafos de conocimiento basados en RDF, apoyados en ontologías biomédicas estandarizadas, ofrecen un marco adecuado para representar de forma interoperable las relaciones entre microbiota, metabolitos, procesos epigenéticos y fenotipos de adiposidad temprana.

2. OBJETIVOS

El objetivo de este trabajo es construir y explotar un grafo de conocimiento basado en tecnologías de la Web Semántica que permita representar de forma estructurada las relaciones entre la microbiota intestinal, los metabolitos microbianos, los procesos epigenéticos y los fenotipos de adiposidad temprana en el eje madre-infante, tomando como referencia el estudio de la cohorte española NEMO. Para ello, se pretende reutilizar ontologías biomédicas estandarizadas, modelar relaciones causales y temporales relevantes para el desarrollo temprano, y demostrar el potencial de las consultas SPARQL para integrar y analizar conocimiento procedente de distintos niveles biológicos, facilitando así la explotación semántica de datos complejos en el ámbito de la investigación biomédica.

3. MATERIALES Y MÉTODOS

3.1. Fuentes de conocimiento

Para el desarrollo de este trabajo se ha seguido una metodología basada en la construcción y explotación de un grafo de conocimiento mediante tecnologías de la Web Semántica, con el objetivo de modelar de forma formal las relaciones biológicas descritas en el estudio de referencia sobre el eje madre-infante y la adiposidad temprana. Como fuente principal de conocimiento se ha utilizado el protocolo del estudio prospectivo de la cohorte española NEMO, a partir del cual se han identificado las entidades biológicas relevantes (microbiota, metabolitos, procesos epigenéticos, fenotipos y etapas del desarrollo) y las relaciones funcionales entre ellas.

3.2. Definición de prefijos y espacios de nombres

Para la construcción del dataset RDF se ha definido, en primer lugar, un conjunto de prefijos que permiten abbreviar las URIs utilizadas en el grafo y mejorar la legibilidad del modelo. Este enfoque facilita tanto la creación manual del archivo Turtle como la posterior formulación de consultas SPARQL, además de favorecer la interoperabilidad con otros recursos semánticos.

Se han incluido los prefijos estándar `rdf` y `rdfs`, correspondientes a los vocabularios básicos de RDF y RDF Schema, que permiten definir clases, propiedades y anotaciones semánticas como etiquetas (`rdfs:label`) y relaciones jerárquicas.

Con el objetivo de reutilizar vocabularios biomédicos consolidados, se han incorporado prefijos asociados a ontologías ampliamente aceptadas en el ámbito de las ciencias de la vida. En particular, se ha utilizado el prefijo `ncbi` para referenciar taxones bacterianos mediante la NCBI Taxonomy, `chebi` para la identificación de metabolitos y clases químicas, `go` para procesos biológicos y epigenéticos definidos en Gene Ontology, y `hp` para la representación de fenotipos clínicos relacionados con obesidad a través de la Human Phenotype Ontology. Asimismo, se han incluido los prefijos `hsapdv` y `fma` para modelar etapas del desarrollo humano y conceptos anatómicos relevantes. Para la definición de relaciones entre entidades se ha empleado el prefijo `ro`, correspondiente a la Relations Ontology.

Finalmente, se han definido dos espacios de nombres propios: `ma_r`, destinado a los recursos específicos del dominio del estudio (entidades como microbiota, metabolitos, procesos epigenéticos y fenotipos), y `ma_o`, reservado para posibles extensiones ontológicas.

3.3. Definición y modelado de las entidades del dominio

La creación del dataset RDF (`maternal_infant_adiposity.txt`) se basó en la identificación de las principales entidades biológicas implicadas en el eje madre-infante y su relación con la adiposidad temprana. Para ello, se definió un modelo

conceptual sencillo con un conjunto reducido de clases propias que permiten organizar entidades procedentes de distintas ontologías biomédicas.

En concreto, se definieron las clases **Microbiota**, **Metabolite**, **EpigeneticProcess**, **Phenotype** y **LifeStage**, que proporcionan una estructura semántica común para integrar información de distintos niveles biológicos. La microbiota intestinal se representó mediante instancias de la clase *Microbiota*, mientras que los metabolitos se modelaron combinando subclases químicas generales y metabolitos concretos enlazados a ChEBI.

Los procesos epigenéticos y los fenotipos metabólicos se representaron como instancias de las clases *EpigeneticProcess* y *Phenotype*, reutilizando términos de Gene Ontology y de la Human Phenotype Ontology, respectivamente. Finalmente, las etapas del desarrollo temprano se modelaron mediante la clase *LifeStage*, incorporando períodos críticos como la etapa prenatal, neonatal y la infancia, lo que permite contextualizar temporalmente los procesos biológicos representados en el grafo.

3.4. Definición y justificación de propiedades y relaciones semánticas

La representación de las interacciones biológicas en el dataset RDF se ha basado en la utilización de propiedades semánticas estandarizadas, con el objetivo de describir de forma coherente las relaciones existentes entre microbiota, metabolitos, procesos epigenéticos, fenotipos y etapas del desarrollo. Se ha priorizado la reutilización de propiedades definidas en la **Relations Ontology (RO)**, evitando la creación de relaciones específicas del dominio siempre que ha sido posible.

Para modelar la producción metabólica de la microbiota intestinal, se ha empleado la propiedad `ro:0003000 (produces)`, que permite establecer una relación directa entre taxones bacterianos o comunidades microbianas y los metabolitos que generan. Esta propiedad se ha utilizado para representar la producción de ácidos grasos de cadena corta, aminoácidos de cadena ramificada y ácidos biliares por parte de la microbiota intestinal, reflejando el papel funcional de la microbiota como fuente de metabolitos bioactivos.

La composición de la microbiota intestinal se ha representado mediante la propiedad `ro:0000052`, que describe una relación de pertenencia o característica. A través de esta relación, se indica que determinados taxones bacterianos forman parte de la microbiota intestinal como comunidad biológica, enfatizando que el interés del modelo no sólo reside en bacterias aisladas, sino en la microbiota como ecosistema funcional.

El origen microbiano de los metabolitos se ha expresado mediante la propiedad `ro:0001000 (derives from)`, que conecta metabolitos concretos con la microbiota intestinal de la que proceden.

Para representar las interacciones entre metabolitos, procesos epigenéticos y fenotipos, se han utilizado distintas propiedades de regulación definidas en la Relations Ontology. En particular, se ha empleado `ro:0002213 (positively regulates)` para indicar efectos reguladores positivos, y `ro:0002212 (negatively regulates)` para reflejar efectos inhibitorios. Estas relaciones se han aplicado tanto a la regulación de procesos epigenéticos por metabolitos como a la influencia de procesos epigenéticos sobre fenotipos metabólicos intermedios y finales.

Adicionalmente, se ha incluido la propiedad `ro:0002211 (regulates)` como relación genérica de regulación, reservada para aquellos casos en los que no se especifica la direccionalidad del efecto.

Las asociaciones causales generales entre entidades se han modelado mediante la propiedad `ro:0002410 (causally related to)`. Esta relación se ha utilizado, por ejemplo, para describir la influencia global de la microbiota o del epigenoma materno sobre sus contrapartes infantiles, así como para conectar fenotipos finales con alteraciones tisulares.

Finalmente, la dimensión temporal del desarrollo temprano se ha incorporado mediante la propiedad `ro:0002092 (happens during)`, que conecta procesos biológicos y entidades relevantes con etapas concretas del desarrollo humano. Esta relación permite representar explícitamente las ventanas críticas durante las cuales la microbiota, los metabolitos y los procesos epigenéticos pueden ejercer una influencia duradera sobre la salud metabólica futura.

3.5. Modelado de las tripletas biológicas

Una vez definidas las entidades y las propiedades semánticas, el dataset RDF se ha completado mediante la incorporación de un conjunto de tripletas que describen las relaciones biológicas relevantes entre microbiota, metabolitos, procesos epigenéticos, fenotipos y etapas del desarrollo. Estas tripletas constituyen el núcleo del grafo de conocimiento y permiten representar de forma estructurada la cadena de eventos biológicos propuesta en el estudio de referencia.

En primer lugar, se han incluido tripletas destinadas a describir la **composición de la microbiota intestinal**. Mediante la relación de pertenencia `ro:0000052`, se ha establecido que determinados taxones bacterianos forman parte de la microbiota intestinal como comunidad biológica.

A continuación, se han modelado las **relaciones de producción metabólica** entre la microbiota y los metabolitos bioactivos. Utilizando la propiedad `ro:0003000`, se han conectado taxones bacterianos concretos y la microbiota intestinal en su conjunto con metabolitos relevantes, como ácidos grasos de cadena corta, aminoácidos de cadena ramificada y ácidos biliares.

De forma complementaria, se han incorporado tripletas que describen el **origen microbiano de los metabolitos**, empleando la relación `ro:0001000`. Este bloque permite expresar explícitamente que determinados metabolitos derivan de la microbiota intestinal.

El siguiente conjunto de tripletas representa las **interacciones entre metabolitos y procesos epigenéticos**. Mediante el uso de relaciones de regulación positiva y negativa (`ro:0002213` y `ro:0002212`), se han modelado los efectos moduladores de distintos metabolitos sobre procesos epigenéticos relevantes, como la regulación de la expresión génica, la organización de la cromatina y la regulación del metabolismo lipídico.

Posteriormente, se han definido tripletas que relacionan los **procesos epigenéticos con fenotipos metabólicos intermedios**. A través de relaciones de regulación positiva, se ha representado la influencia de procesos epigenéticos sobre características como el aumento del índice de masa corporal, la resistencia a la insulina y la hiperinsulinemia.

Las tripletas correspondientes a los **fenotipos clínicos finales** describen la transición desde fenotipos intermedios hacia condiciones clínicas como la obesidad y la obesidad infantil. Asimismo, se ha incorporado una relación causal general entre la obesidad y alteraciones tisulares, representando la asociación con cambios en la morfología del tejido adiposo.

El eje madre-infante se ha modelado mediante tripletas que establecen relaciones causales generales entre la microbiota y el epigenoma maternos y sus equivalentes infantiles.

Finalmente, se ha incorporado explícitamente la **dimensión temporal del desarrollo temprano** mediante tripletas que conectan entidades y procesos biológicos con etapas concretas del desarrollo humano, como la etapa prenatal, neonatal y la infancia.

3.6. Validación sintáctica y comprobación del dataset RDF

Con el fin de verificar la corrección sintáctica del dataset RDF antes de su carga en el triplestore, se realizó un proceso de validación previo utilizando herramientas externas especializadas. En primer lugar, el archivo RDF serializado en formato Turtle fue convertido a formato RDF/XML mediante la herramienta **Easy RDF Converter**, lo que permitió detectar posibles errores de serialización y asegurar la compatibilidad del dataset con distintos formatos RDF.

Posteriormente, el archivo RDF/XML generado se validó empleando el **validador oficial del W3C**, comprobando que la sintaxis RDF cumplía con los estándares establecidos. Este proceso permitió confirmar la validez formal del dataset, así como

obtener información adicional sobre el número total de tripletas incluidas en el grafo.

Adicionalmente, las herramientas de validación utilizadas permitieron generar una representación gráfica del grafo RDF, ofreciendo una visualización global de las entidades y relaciones definidas. Esta visualización resultó útil como comprobación cualitativa del modelo, facilitando la identificación de posibles inconsistencias estructurales y proporcionando una visión intuitiva de la conectividad del grafo antes de su explotación mediante consultas SPARQL.

3.7. Carga del dataset RDF y configuración del entorno Blazegraph

La carga y gestión del dataset RDF se ha realizado utilizando el triplestore **Blazegraph**, desplegado en un entorno servidor remoto. Para ello, se accedió al servidor `dayhoff.inf.um.es` mediante conexión SSH, donde se creó un directorio de trabajo específico (`trabajo_esd`) destinado a contener los archivos necesarios para la ejecución del triplestore. En dicho directorio se almacenó el archivo ejecutable `blazegraph.jar`, que permite iniciar Blazegraph como servidor SPARQL independiente.

El triplestore se lanzó desde la terminal del servidor mediante la ejecución del comando `java -server -Xmx2g -jar blazegraph.jar`. Este comando inicia Blazegraph en modo servidor, exponiendo por defecto su interfaz web y su endpoint SPARQL en el puerto local 9999 del servidor remoto.

Para acceder a la interfaz web de Blazegraph desde el entorno local, se estableció un túnel SSH mediante redirección de puertos. Concretamente, se utilizó el comando `ssh -N -L 3040:localhost:9999 alumno11@dayhoff.inf.um.es`, que permite mapear el puerto 9999 del servidor remoto al puerto 3040 de la máquina local. Este procedimiento garantiza un acceso seguro al triplestore sin necesidad de exponer directamente el puerto del servidor.

Una vez establecido el túnel, se accedió a la interfaz web de Blazegraph desde el navegador local mediante la URL `http://localhost:3040/blazegraph/#splash`, utilizando una sesión de navegación privada para evitar interferencias con sesiones previas. A través de esta interfaz se procedió a la gestión del dataset RDF.

En la sección de administración de namespaces de Blazegraph se creó un nuevo espacio de nombres denominado “`ma`”, destinado a albergar el grafo correspondiente al eje madre-infante. Tras la creación del namespace, se utilizó la opción *Update* para cargar el contenido del archivo TTL generado, copiando directamente el dataset RDF en el editor proporcionado por la interfaz web. Este proceso permitió la inserción de todas las tripletas en el triplestore de forma inmediata.

Finalmente, una vez completada la carga del dataset, se seleccionó la opción *Use* asociada al namespace `ma`, estableciéndolo como espacio de trabajo activo. De este

modo, todas las consultas SPARQL ejecutadas a partir de ese momento se realizaron sobre el grafo RDF correspondiente al dataset construido.

3.8. Generación y ejecución de consultas SPARQL desde R

Una vez cargado el grafo RDF en el triplestore Blazegraph, se procedió a su explotación mediante la ejecución de consultas SPARQL desde el entorno de programación **RStudio** (“*queries_script.R*”). Este enfoque permite interactuar de forma programática con el grafo de conocimiento, automatizando la recuperación de información y facilitando su posterior análisis e integración con otras herramientas disponibles en R.

Para ello, se emplearon las librerías `httr` y `jsonlite`, que permiten enviar peticiones HTTP a un endpoint SPARQL remoto y procesar las respuestas devueltas en formato JSON. En primer lugar, se definió el endpoint correspondiente al servicio Blazegraph en el que se encontraba cargado el dataset RDF, apuntando al namespace creado para este trabajo. En concreto, el endpoint utilizado fue:

```
http://localhost:3040/blazegraph/namespace/ma/sparql
```

Con el fin de simplificar y unificar la ejecución de las consultas, se implementó una función auxiliar en R que permite enviar consultas SPARQL al endpoint especificado, comprobar la correcta respuesta del servidor y transformar los resultados obtenidos en estructuras de datos tabulares (*data frames*). Esta función automatiza la gestión de las peticiones y facilita la reutilización del código para múltiples consultas.

Las consultas SPARQL se construyeron como cadenas de texto siguiendo la sintaxis estándar del lenguaje. Cada consulta incluye la declaración explícita de los prefijos necesarios, correspondientes tanto al espacio de nombres propio del proyecto como a las ontologías reutilizadas, seguida de una cláusula `SELECT` que especifica las variables a recuperar y un bloque `WHERE` que define los patrones de tripletas empleados para interrogar el grafo RDF. El conjunto de consultas se diseño de forma progresiva, abarcando desde consultas descriptivas básicas hasta consultas complejas de tipo multisalto, con el objetivo de analizar las relaciones entre microbiota, metabolitos, procesos epigenéticos y fenotipos de adiposidad temprana.

Los resultados devueltos por cada consulta se almacenaron como *data frames* en R, lo que permitió su inspección directa mediante funciones de visualización como `View()` y facilitó una exploración ordenada de la información obtenida. Finalmente, los resultados de cada consulta se exportaron a archivos en formato CSV, garantizando la reproducibilidad del análisis y permitiendo su reutilización posterior.

4. Resultados y discusión

4.1. Resultados de la validación del dataset RDF

Como paso previo a la explotación del grafo RDF, se realizó una validación sintáctica del dataset utilizando el **servicio de validación RDF del W3C**, tras convertir el archivo original en formato Turtle a RDF/XML (“*maternal_infant_adiposity_xmlformat.txt*”). El resultado de dicha validación (“*validacion_w3c.png*”) confirma que el documento RDF es **sintácticamente correcto**, indicando explícitamente que el modelo ha sido validado con éxito y que cumple con los estándares RDF establecidos por el W3C.

En total, el proceso de validación permitió confirmar que el dataset RDF está compuesto por **166 triplets**, lo que proporciona una medida objetiva del tamaño y la complejidad del grafo construido. Esta representación tabular permite comprobar que las clases definidas (como *Microbiota*, *Metabolite*, *EpigeneticProcess* o *Phenotype*) están correctamente tipadas mediante `rdf:type` y que las anotaciones semánticas, como las etiquetas (`rdfs:label`), se han aplicado de forma coherente. La ausencia de errores o advertencias durante este proceso indica que la definición de prefijos, clases, propiedades y relaciones se ha realizado de manera adecuada.

Asimismo, el servicio de validación permite obtener una **visualización gráfica del grafo RDF** (“*grafo.png*”), en la que se representan de forma explícita las entidades y las relaciones entre ellas. En esta representación se observa una estructura altamente conectada, en la que la microbiota intestinal ocupa una posición central, enlazando con metabolitos, procesos epigenéticos y fenotipos metabólicos.

En definitiva, los resultados de la validación confirman que el dataset RDF construido no solo es formalmente correcto, sino que presenta una estructura semántica consistente con el dominio biológico representado.

4.2. Comprobación de la carga correcta del dataset en Blazegraph



Figura 1. Consulta SPARQL de comprobación de la carga del dataset RDF en Blazegraph

Tras la carga del dataset RDF en el triplestore Blazegraph, se realizó una comprobación adicional para verificar que el grafo había sido importado correctamente.

El resultado de esta consulta, mostrado en la *Figura 1*, devolvió un total de **166 tripletas**, coincidiendo exactamente con el número obtenido previamente durante la validación sintáctica del dataset mediante el servicio del W3C. Esta concordancia confirma que la carga del archivo RDF en Blazegraph se realizó de forma completa y sin pérdidas de información, garantizando que el grafo almacenado en el triplestore refleja fielmente el modelo semántico validado.

4.3. Resultados de las consultas SPARQL

Consulta 1: Producción metabólica asociada a la microbiota intestinal

Los resultados (*Figura 2*) muestran asociaciones coherentes entre taxones bacterianos y metabolitos clave. *Bifidobacterium* se vincula con la producción de butirato y lactato, metabolitos implicados en la homeostasis energética y la regulación epigenética, mientras que *Bacteroides* se asocia con propionato y ácido desoxicólico. La microbiota intestinal como conjunto aparece relacionada con la producción de acetato y aminoácidos de cadena ramificada, reflejando su papel central como fuente metabólica con impacto potencial en el metabolismo del huésped.

consulta1_microbiota_metabolitos		
	microbe	metabolite
1	http://maternal-infant-adiposity.um.es/resource/Bacteroides	http://maternal-infant-adiposity.um.es/resource/DeoxycholicAcid
2	http://maternal-infant-adiposity.um.es/resource/Bacteroides	http://maternal-infant-adiposity.um.es/resource/Propionate
3	http://maternal-infant-adiposity.um.es/resource/Bifidobacterium	http://maternal-infant-adiposity.um.es/resource/Butyrate
4	http://maternal-infant-adiposity.um.es/resource/Bifidobacterium	http://maternal-infant-adiposity.um.es/resource/Lactate
5	http://maternal-infant-adiposity.um.es/resource/GutMicrobiota	http://maternal-infant-adiposity.um.es/resource/Acetate
6	http://maternal-infant-adiposity.um.es/resource/GutMicrobiota	http://maternal-infant-adiposity.um.es/resource/Isoleucine
7	http://maternal-infant-adiposity.um.es/resource/GutMicrobiota	http://maternal-infant-adiposity.um.es/resource/Leucine
8	http://maternal-infant-adiposity.um.es/resource/GutMicrobiota	http://maternal-infant-adiposity.um.es/resource/Valine
9		

Figura 2. Archivo CSV resultante de la Consulta 1 sobre producción metabólica de la microbiota

Consulta 2: Metabolitos que regulan procesos epigenéticos

Los resultados (*Figura 3*) revelan que distintos tipos de metabolitos presentan efectos reguladores sobre procesos epigenéticos específicos. Los ácidos grasos de cadena corta se asocian con procesos de organización y remodelado de la cromatina, mientras que los aminoácidos de cadena ramificada se relacionan con la regulación positiva del metabolismo lipídico.

	metabolite	metaboliteClass	epigeneticProcess
1	http://maternal-infant-adiposity.um.es/resource/DeoxycholicAcid	http://maternal-infant-adiposity.um.es/resource/BileAcid	http://maternal-infant-adiposity.um.es/resource/EpigeneticRegulationOfGeneExpression
2	http://maternal-infant-adiposity.um.es/resource/Isoleucine	http://maternal-infant-adiposity.um.es/resource/BranchedChainAminoAcid	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess
3	http://maternal-infant-adiposity.um.es/resource/Leucine	http://maternal-infant-adiposity.um.es/resource/BranchedChainAminoAcid	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess
4	http://maternal-infant-adiposity.um.es/resource/Valine	http://maternal-infant-adiposity.um.es/resource/BranchedChainAminoAcid	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess
5	http://maternal-infant-adiposity.um.es/resource/Butyrate	http://maternal-infant-adiposity.um.es/resource/ShortChainFattyAcid	http://maternal-infant-adiposity.um.es/resource/ChromatinRemodeling
6	http://maternal-infant-adiposity.um.es/resource/Butyrate	http://maternal-infant-adiposity.um.es/resource/ShortChainFattyAcid	http://maternal-infant-adiposity.um.es/resource/EpigeneticRegulationOfGeneExpression
7	http://maternal-infant-adiposity.um.es/resource/Propionate	http://maternal-infant-adiposity.um.es/resource/ShortChainFattyAcid	http://maternal-infant-adiposity.um.es/resource/ChromatinOrganization
8	http://maternal-infant-adiposity.um.es/resource/Propionate	http://maternal-infant-adiposity.um.es/resource/ShortChainFattyAcid	http://maternal-infant-adiposity.um.es/resource/RegulationOfChromatinOrganization
9			

Figura 3. Archivo CSV resultante de la Consulta 2 sobre metabolitos que regulan procesos epigenéticos

Consulta 3: Procesos epigenéticos activos en etapas tempranas del desarrollo

Los resultados (*Figura 4*) indican una distribución temporal diferenciada de los procesos epigenéticos. La regulación de la expresión génica se asocia principalmente a la etapa prenatal, el remodelado de la cromatina a la etapa neonatal y la organización de la cromatina a la infancia.

	epigeneticProcess	lifeStage
1	http://maternal-infant-adiposity.um.es/resource/ChromatinOrganization	http://maternal-infant-adiposity.um.es/resource/Infancy
2	http://maternal-infant-adiposity.um.es/resource/ChromatinRemodeling	http://maternal-infant-adiposity.um.es/resource/NeonatalStage
3	http://maternal-infant-adiposity.um.es/resource/EpigeneticRegulationOfGeneExpression	http://maternal-infant-adiposity.um.es/resource/PrenatalStage
4		

Figura 4. Archivo CSV resultante de la Consulta 3 sobre procesos epigenéticos asociados a etapas del desarrollo temprano

Consulta 4: Transmisión madre-infante de influencias biológicas

Los resultados (*Figura 5*) muestran relaciones causales generales entre entidades maternas e infantiles, incluyendo la microbiota y el epigenoma. La asociación de la microbiota materna con la microbiota infantil durante la etapa prenatal refleja la influencia materna temprana, mientras que las relaciones entre epigenoma y fenotipos tisulares ilustran posibles vías de transmisión intergeneracional del riesgo metabólico.

	maternalEntity	relation	infantEntity	lifeStage
1	http://maternal-infant-adiposity.um.es/resource/MaternalEpigenome	http://purl.obolibrary.org/obo/RO_0002410	http://maternal-infant-adiposity.um.es/resource/InfantEpigenome	NA
2	http://maternal-infant-adiposity.um.es/resource/Oesity	http://purl.obolibrary.org/obo/RO_0002410	http://maternal-infant-adiposity.um.es/resource/AbnormalAdiposeTissueMorphology	NA
3	http://maternal-infant-adiposity.um.es/resource/MaternalMicrobiota	http://purl.obolibrary.org/obo/RO_0002410	http://maternal-infant-adiposity.um.es/resource/InfantMicrobiota	http://maternal-infant-adiposity.um.es/resource/PrenatalStage
4				

Figura 5. Archivo CSV resultante de la Consulta 4 sobre la transmisión madre-infante de influencias biológicas

Consulta 5: Rutas completas desde metabolitos microbianos hasta obesidad

Los resultados (*Figura 6*) permiten reconstruir rutas coherentes que conectan metabolitos microbianos con obesidad y obesidad infantil a través de procesos epigenéticos y fenotipos intermedios como el aumento del índice de masa corporal, la resistencia a la insulina o la hiperinsulinemia. Estas rutas sintetizan el eje microbiota-metabolitos-epigenoma-fenotipo y evidencian el valor del enfoque semántico para integrar múltiples niveles biológicos.

	A	B	C	D
consultas5_rutas_completas				
1	metabolite	epigeneticProcess	intermediatePhenotype	finalPhenotype
2	http://maternal-infant-adiposity.um.es/resource/DeoxythalicAcid	http://maternal-infant-adiposity.um.es/resource/EpigenticRegulationOfGeneExpression	http://maternal-infant-adiposity.um.es/resource/increasedBMI	http://maternal-infant-adiposity.um.es/resource/ChildhoodObesity
3	http://maternal-infant-adiposity.um.es/resource/Butyrate	http://maternal-infant-adiposity.um.es/resource/EpigenticRegulationOfGeneExpression	http://maternal-infant-adiposity.um.es/resource/increasedBMI	http://maternal-infant-adiposity.um.es/resource/ChildhoodObesity
4	http://maternal-infant-adiposity.um.es/resource/Isoleucine	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess	http://maternal-infant-adiposity.um.es/resource/increasedBMI	http://maternal-infant-adiposity.um.es/resource/ChildhoodObesity
5	http://maternal-infant-adiposity.um.es/resource/Leucine	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess	http://maternal-infant-adiposity.um.es/resource/increasedBMI	http://maternal-infant-adiposity.um.es/resource/ChildhoodObesity
6	http://maternal-infant-adiposity.um.es/resource/Valine	http://maternal-infant-adiposity.um.es/resource/PositiveRegulationOfLipidMetabolicProcess	http://maternal-infant-adiposity.um.es/resource/increasedBMI	http://maternal-infant-adiposity.um.es/resource/ChildhoodObesity
7	http://maternal-infant-adiposity.um.es/resource/Propionate	http://maternal-infant-adiposity.um.es/resource/ChromatinOrganization	http://maternal-infant-adiposity.um.es/resource/hyperinsulinemia	http://maternal-infant-adiposity.um.es/resource/Obesity
8	http://maternal-infant-adiposity.um.es/resource/Butyrate	http://maternal-infant-adiposity.um.es/resource/ChromatinRemodeling	http://maternal-infant-adiposity.um.es/resource/insulinResistance	http://maternal-infant-adiposity.um.es/resource/Obesity

Figura 6. Archivo CSV resultante de la Consulta 5 sobre rutas biológicas desde metabolitos microbianos hasta obesidad

5. Conclusiones

- En este trabajo se ha construido un dataset RDF que representa de forma estructurada y semánticamente coherente las relaciones entre microbiota intestinal, metabolitos microbianos, procesos epigenéticos y fenotipos de adiposidad temprana en el eje madre-infante.
- La reutilización de ontologías biomédicas estandarizadas y de propiedades definidas en la Relations Ontology ha permitido garantizar la interoperabilidad del modelo y evitar la definición de vocabularios ad hoc, siguiendo buenas prácticas de la Web Semántica.
- La validación sintáctica del dataset mediante herramientas del W3C y la comprobación de su carga correcta en el triplestore Blazegraph han confirmado la calidad formal del grafo, compuesto por un total de 166 triplets.
- La explotación del grafo mediante consultas SPARQL ha demostrado la capacidad del modelo para integrar múltiples niveles de información y reconstruir rutas biológicas complejas, desde la producción metabólica por la microbiota hasta la aparición de fenotipos clínicos finales.
- En conjunto, el enfoque basado en grafos de conocimiento pone de manifiesto el potencial de las tecnologías semánticas para apoyar el análisis integrado de procesos biológicos complejos y constituye una base sólida para futuras ampliaciones del modelo y su aplicación en investigación biomédica.

6. Bibliografía

- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host & Microbe*, 17(5), 690–703. <https://doi.org/10.1016/j.chom.2015.04.004>
- Carlsen, E. M., Renault, K. M., Nørgaard, K., Nilas, L., Jensen, J. E., Hyldstrup, L., Michaelsen, K. F., Cortes, D., & Pryds, O. (2014). Newborn regional body composition is influenced by maternal obesity, gestational weight gain, and birthweight standard score. *Acta Paediatrica*, 103(9), 939–945. <https://doi.org/10.1111/apa.12674>
- Cuevas-Sierra, A., Ramos-Lopez, O., Riezu-Boj, J. I., Milagro, F. I., & Martínez, J. A. (2019). Diet, gut microbiota, and obesity: Links with host genetics and epigenetics and potential applications. *Advances in Nutrition*, 10(Suppl. 1), S17–S30. <https://doi.org/10.1093/advances/nmy078>
- Dogra, S., Sakwinska, O., Soh, S. E., Ngom-Bru, C., Brück, W. M., Berger, B., Brüssow, H., Lee, Y. S., Yap, F., Chong, Y. S., et al. (2015). Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *mBio*, 6(1), e02419–14. <https://doi.org/10.1128/mBio.02419-14>
- Franzago, M., Borrelli, P., Di Nicola, M., Cavallo, P., D'Adamo, E., Di Tizio, L., Gazzolo, D., Stuppia, L., & Vitacolonna, E. (2024). From mother to child: Epigenetic signatures of hyperglycemia and obesity during pregnancy. *Nutrients*, 16(21), 3502. <https://doi.org/10.3390/nu16213502>
- Hamilton, D., Dee, A., & Perry, I. J. (2018). The lifetime costs of overweight and obesity in childhood and adolescence: A systematic review. *Obesity Reviews*, 19(4), 452–463. <https://doi.org/10.1111/obr.12649>
- Jebeile, H., Kelly, A. S., O'Malley, G., & Baur, L. A. (2022). Obesity in children and adolescents: Epidemiology, causes, assessment, and management. *The Lancet Diabetes & Endocrinology*, 10(5), 351–365. [https://doi.org/10.1016/S2213-8587\(22\)00047-0](https://doi.org/10.1016/S2213-8587(22)00047-0)
- Jonsson, J., Renault, K. M., Garcia-Calzon, S., Perfiliev, A., Estampador, A. C., Nørgaard, K., Lind, M. V., Vaag, A., Hjort, L., Michaelsen, K. F., et al. (2021). Lifestyle intervention in pregnant women with obesity impacts cord blood DNA methylation, which associates with body composition in the offspring. *Diabetes*, 70(4), 854–866. <https://doi.org/10.2337/db20-0858>
- Koontz, M. B., Gunzler, D. D., Presley, L., & Catalano, P. M. (2014). Longitudinal changes in infant body composition: Association with childhood obesity. *Pediatric Obesity*, 9(6), e141–e144. <https://doi.org/10.1111/j.2047-6310.2014.00248.x>
- Kumar, H., Lund, R., Laiho, A., Lundelin, K., Ley, R. E., Isolauri, E., & Salminen, S. (2014). Gut microbiota as an epigenetic regulator: Pilot study based on whole-genome methylation analysis. *mBio*, 5(6), e02113–14. <https://doi.org/10.1128/mBio.02113-14>
- Li, S., Ma, X., Mei, H., Chang, X., He, P., Sun, L., Xiao, H., Wang, S., & Li, R. (2025). Association between gut microbiota and short-chain fatty acids in children with obesity. *Scientific Reports*, 15, 483. <https://doi.org/10.1038/s41598-025-00483-0>
- Li, X. M., Lv, Q., Chen, Y. J., Yan, L. B., & Xiong, X. (2024). Association between childhood obesity and gut microbiota: 16S rRNA gene sequencing-based cohort study. *World Journal of Gastroenterology*, 30(16), 2249–2257. <https://doi.org/10.3748/wjg.v30.i16.2249>
- Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J., Belzer, C., Delgado Palacio, S., Arboleya Montes, S., Mancabelli, L., et al. (2017). The first microbial colonizers of the human gut: Composition, activities, and health implications of the infant gut microbiota. *Microbiology and Molecular Biology Reviews*, 81(4), e00036-17. <https://doi.org/10.1128/MMBR.00036-17>
- Moore, B. F., Harrall, K. K., Sauder, K. A., Glueck, D. H., & Dabelea, D. (2020). Neonatal adiposity and childhood obesity. *Pediatrics*, 146(3), e20200737. <https://doi.org/10.1542/peds.2020-0737>
- Pauwels, S., Symons, L., Vanautgaerden, E. L., Ghosh, M., Duca, R. C., Bekaert, B., Freson, K., Huybrechts, I., Langie, S. A. S., Koppen, G., et al. (2019). The influence of the duration of breastfeeding on the infant's metabolic epigenome. *Nutrients*, 11(6), 1408. <https://doi.org/10.3390/nu11061408>
- Renault, K. M., Carlsen, E. M., Nørgaard, K., Nilas, L., Pryds, O., Secher, N. J., Cortes, D., Jensen, J. E., Olsen, S. F., & Halldorsson, T. I. (2015). Intake of carbohydrates during pregnancy in obese women is associated with fat mass in the newborn offspring. *American Journal of Clinical Nutrition*, 102(6), 1475–1481. <https://doi.org/10.3945/ajcn.115.110098>
- Robinson, N., Brown, H., Antoun, E., Godfrey, K. M., Hanson, M. A., Lillycrop, K. A., Crozier, S. R., Murray, R., Pearce, M. S., Relton, C. L., & the SWS Study Group. (2021). Childhood DNA methylation as a marker of early life rapid weight gain and subsequent overweight. *Clinical Epigenetics*, 13, 8. <https://doi.org/10.1186/s13148-020-00980-9>
- Suárez-Cortés, M., Juan-Pérez, A., Molina-Rodríguez, A., Araújo de Castro, J., Castaño-Molina, M. Á., Fernández-Ruiz, V. E., Jiménez-Méndez, A., Martínez Pérez-Munar, P., Rico-Chazarra, S., Ramos-Molina, B., et al. (2025). Dynamics of the epigenome, microbiome, and metabolome in relation to early adiposity in the maternal–infant axis: Protocol for a prospective, observational pilot study in the Spanish NEMO cohort. *Journal of Clinical Medicine*, 14(19), 6694. <https://doi.org/10.3390/jcm14196694>
- Wahlström, A., Sayin, S. I., Marschall, H. U., & Bäckhed, F. (2016). Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metabolism*, 24(1), 41–50. <https://doi.org/10.1016/j.cmet.2016.05.005>
- World Health Organization. (2020). *Obesity and overweight*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>