

4. CONSTRUCCIÓN DE LA ONTOLOGÍA M4L

4.1 INTRODUCCIÓN

El objetivo de construir una ontología en esta tesis es ofrecer una herramienta que simplifique y ayude a la recuperación e indexación de la información que se consigna en los cuadernos de laboratorio. La ontología diseñada en este estudio, provee términos relacionados con procesos de laboratorio comúnmente usados en biología molecular de plantas. La ontología llamada Metadata for Laboratory Notebook (M4L), reutiliza términos de otras ontologías e incorpora términos propios extraídos de cuadernos de laboratorio analizados. La labor de construcción de esta ontología fue facilitada por la elaboración de casos de uso representados a través de mapas conceptuales.

El proceso de desarrollo de una ontología, se refiere a las actividades que se deben llevar a cabo cuando se construyen ontologías (Fernández et al., 1997). Sin embargo, la comunidad científica aún no ha logrado un acuerdo que establezca qué metodología usar para la construcción de una ontología; y diferentes propuestas aparecen en la literatura, por ejemplo: TOVE (Gruninger and Fox, 1995), METHONTOLOGY (Fernández et al., 1997), UPON (De Nicola et al., 2009). Todas ellas comparten algunas características como: i) propósito de la ontología, ii) adquisición del conocimiento, iii) conceptualización y categorización de los términos y iv) formalización de la ontología.

Otras metodologías, además de las características antes mencionadas, coinciden en la importancia de soportar un ambiente colaborativo para el desarrollo de ontologías por parte de expertos de dominio. Algunos ejemplos de este tipo de

metodologías son: DILIGENT (Pinto et al., 2004), la metodología propuesta por García *et al* (Castro et al., 2006), y la presentada por Mainz *et al* (Mainz et al., 2008).

Para el desarrollo de M4L se reutilizaron actividades o procedimientos seguidos por las seis metodologías antes mencionadas; ellos son: i) propósito de la ontología, ii) conceptualización y categorización de términos, iii) captura del conocimiento que será representado, iv) elección de un lenguaje para el diseño de ontologías y v) selección de una herramienta para el diseño de ontologías.

Para explicar el proceso de desarrollo de M4L y las lecciones aprendidas, por parte de expertos de dominio en ciencias biológicas, el presente capítulo se presenta de la siguiente manera: en la primera sección se explica brevemente en que consiste cada uno de los cinco pasos sugeridos para la construcción de M4L; en la segunda sección se describe como se llevaron a cabo cada uno de los pasos y los resultados obtenidos en cada uno de ellos; y por ultimo, se presentan los comentarios finales.

4.2 PRESENTACIÓN DE LOS PASOS PARA MODELAR LA ONTOLOGÍA M4L

Existen varias metodologías generales para el desarrollo de ontologías, con diferente grado de dificultad y especificación en su aplicación. Todas ellas exponen los procedimientos y herramientas que pueden usarse para el desarrollo y validación de ontologías. Entre ellas se pueden destacar las siguientes: TOVE, METHONTOLOGY, UPON, DILIGENT, la metodología propuesta por García *et al.*, y la metodología presentada por Mainz *et al.* De las metodologías antes mencionadas, se identificaron aquellas actividades que soportaran el desarrollo, uso y mantenimiento de la ontología M4L. Las etapas o actividades que fueron reutilizadas de dichas metodologías, se presentan a continuación.

4.2.1 Paso 1: Identificación del Propósito de la Ontología. El propósito de una ontología es hacer explícito algún punto de vista. Esto es, definir el dominio que cubrirá la ontología, definir el uso, la aplicación final de dicha ontología, definir a que tipo de preguntas responderá la ontología y recolectar la información necesaria. Esta etapa puede ser facilitada a través de la elaboración de escenarios específicos o casos de uso.

4.2.2 Paso 2. Conceptualización y Categorización de los Términos. Puede lograrse por medio de: i) identificación de ontologías “modelo”, ii) reuso de ontologías y iii) de conceptos normalizados adoptados de diccionarios, tesauros, vocabularios controlados, etc. A continuación se presentan las principales fuentes de adquisición de términos no ambiguos para el desarrollo de la ontología M4L.

- Identificación de ontologías modelo. Se refiere a ontologías que hayan sido desarrolladas previamente con propósitos similares, para que sirvan de base a los expertos de dominio para construir su propia ontología.
- Considerar el reuso de ontologías. Se refiere a la importación de ontologías ya existentes que se adapten a las necesidades de la ontología que se desea desarrollar. En caso de poder aplicarse esta etapa, se recolectan términos y características propias de cada ontología reutilizada.
- Identificación de conceptos que podrían reusarse. Se refiere a términos provenientes de diccionarios, tesauros, vocabularios controlados u ontologías que tengan un significado similar al que se desea reutilizar en la ontología a desarrollar.

4.2.3 Paso 3: Capturar el Conocimiento que Será Representado. Es una etapa compleja, en este paso se deben escoger los mecanismos adecuados para lograr recolectar el conocimiento a representar. Como se describió en el paso uno, una manera de recolectar la información necesaria para la construcción de ontologías es a través de la presentación de escenarios específicos. Una herramienta que facilita la representación de estos escenarios son los mapas conceptuales (MC). Los MC, son instrumentos de representación del conocimiento sencillos y prácticos, que permiten transmitir con claridad mensajes conceptuales complejos y facilitar tanto el aprendizaje como la enseñanza. Los MC adoptan formas de grafos, donde los conceptos se representan como nodos mientras que los arcos representan las relaciones entre ellos (Novak and Cañas, 2008). La figura 42, ilustra la definición de un MC, mostrando los nodos y los arcos correspondientes.

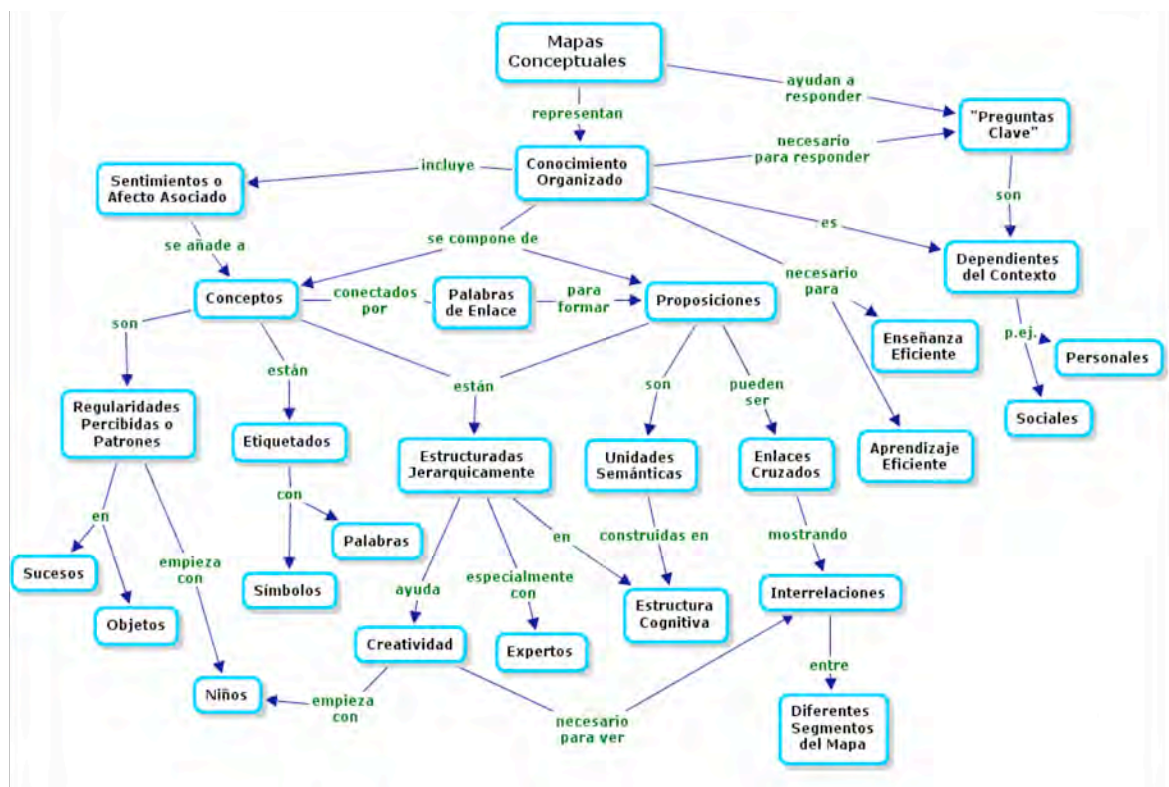


Figura 42. Definición esquemática de un mapa conceptual.

Instrumentos como los MC permiten:

1. una mejor comprensión del conocimiento que se captura en la ontología (Brilhante et al., 2006, Cuevas-Simon et al., 2008),
2. la representación gráfica, facilita visualizar conceptos claves y las relaciones dentro de un dominio de conocimiento (Novak and Cañas, 2008),
3. facilita el producir definiciones no ambiguas para los conceptos y relaciones (Simon et al., 2007); por ultimo,
4. promueve la colaboración, intercambio de conocimiento y evolución de la ontología entre expertos de dominio que incluso estén distribuidos geográficamente (Castro et al., 2006, Mainz et al., 2008).

4.2.4 Paso 4: Lenguaje para el Diseño de Ontologías. Existen multitud de lenguajes que permiten la representación de ontologías. Los primeros lenguajes diseñados en el campo de la inteligencia artificial para el trabajo con ontologías se crearon a principios de los 90. Entre esos lenguajes básicos están: CyCL, KIF, Ontolingua, LOOM, OCML y *Frame Logic* (Flogic) (Gómez-Pérez et al., 2004). Actualmente estos lenguajes han sido sustituidos por lenguajes de ontologías capaces de explotar las capacidades de la red. Entre los actuales lenguajes para ontologías se destacan: RDF, SHOE, XOL, OIL, DAM+OIL y *Ontology Web Language* (OWL) (Gómez-Pérez et al., 2004).

4.2.5 Paso 5. Herramientas para el Desarrollo de Ontologías. Las herramientas para el desarrollo de ontologías permiten la creación, modificación y gestión de modelos ontológicos, facilitando así, el trabajo del desarrollador. En Bioinformática

las herramientas mas usadas para el desarrollo de ontologías son Protégé-2000, Chimaera, OilEd y DAG-Edit (Lambrix et al., 2003).

De las herramientas antes mencionadas, la de uso mas extendido es Protégé-2000; desarrollada por el Stanford Center for BioMedical Informatics Research (BMIR) de la Universidad de Stanford. Su diseño ha perseguido dos objetivos: i) compatibilidad con otros sistemas de representación del conocimiento y, ii) ser configurable y fácil de usar (Lambrix et al., 2003). Las fortalezas de Protégé frente a otras herramientas para el desarrollo y edición de ontologías son:

1. Ser de libre acceso. Esta característica, ha hecho que Protégé cuente con un amplio número de seguidores, quienes generan nuevos plug-ins, dan soporte y comparten sus experiencias; por ello Protégé permanece en constante mantenimiento y evolución.
2. Los plug-ins de Protégé, ofrecen una funcionalidad importante. Así como los formatos que soporta (Frames, XML, Schema, RDF y OWL) para poder importar y exportar ontologías, y por ultimo,
3. su interfaz de usuario es simple y fácil de usar.

En la figura 43, se presentan los pasos escogidos de las seis metodologías analizadas en este estudio, para el desarrollo de la ontología M4L.



Figura 43. Pasos para el desarrollo de la ontología M4L.

4.3 RESULTADOS: DESARROLLO DE LA ONTOLOGÍA M4L PASO A PASO

4.3.1 Propósito de la Ontología. En la figura 44, se ilustra el primer paso desarrollado durante el proceso construcción de M4L, el propósito de la ontología.



Figura 44. Presentación del propósito de la ontología M4L.

Se inició el desarrollo de la ontología definiendo el dominio, alcance y granularidad de la misma; para ello se plantearon las siguientes preguntas significativas o de competencia (Gruninger and Fox, 1995).

- ¿Que dominio cubrirá la ontología?. Metadata for Laboratory Notebook (M4L), es una ontología diseñada dentro del dominio de la biotecnología de plantas. Se espera que M4L, proporcione un modelo para el diseño de una investigación, permita la selección de protocolos, de equipos empleados, reactivos utilizados y los datos que se generan de la misma investigación.
- ¿Para que se va a emplear la ontología?. Esta ontología facilitará la búsqueda, clasificación, recuperación e intercambio del conocimiento que se genere en el laboratorio.

Preguntas cuyas respuestas la ontología deberá soportar. Los usuarios esperan que la ontología M4L soporte las respuestas de preguntas como:

1. ¿Qué actividades experimentales deben llevarse a cabo en un proyecto de investigación específico?
2. ¿Qué protocolos son usados para realizar una determinada actividad experimental?
3. ¿Qué equipos, reactivos e insumos son requeridos al emplear un protocolo en particular?

- ¿Quien utilizará y mantendrá la ontología?. Los usuarios potenciales de esta ontología son estudiantes y profesionales que laboren en el laboratorio de biotecnología del Centro Internacional de Agricultura Tropical (CIAT). Se espera que la misma comunidad científica (que incluso se encuentre distribuida geográficamente), administre y actualice la ontología. La evolución de la ontología por parte de expertos de dominio puede facilitarse con el uso de herramientas como los MC, que permiten representar gráficamente el conocimiento expuesto en los casos de uso.

4.3.2 Conceptualización y Categorización de Términos. En la figura 45, se ilustra el segundo paso desarrollado en este estudio para la construcción de M4L, la conceptualización y categorización de términos. Los tres pasos que componen esta etapa pueden realizarse de manera simultánea, no se requiere un orden predeterminado para su ejecución.



Figura 45. Conceptualización y categorización de términos. Segunda etapa del desarrollo de M4L.

- Identificación de “ontologías modelo”. En este paso se revisaron ontologías de dominio biológico y biomédico, que han sido diseñadas con el propósito de estructurar la información derivada de procedimientos experimentales. La labor de búsqueda de ontologías fue facilitada por BioPortal²³, una aplicación Web desarrollada por el National Center for Biomedical Ontologies.

Entre los resultados de la búsqueda, se identificó una ontología diseñada para representar procedimientos experimentales, la Ontology for Biomedical Investigation (OBI) (Brinkman et al., 2010). OBI, fue diseñada con el propósito de soportar la descripción de investigaciones biológicas y clínicas. Esta ontología

²³ BioPortal, proporciona a la comunidad biológica y biomédica, un espacio que permite compartir y descubrir ontologías. Actualmente BioPortal es un repositorio de 204 ontologías y contiene 1,438,792 términos (<http://bioportal.bioontology.org/>).

representa todas las fases de los procesos experimentales y las entidades involucradas en la preparación, ejecución e interpretación de esos procesos (Brinkman et al., 2010). Debido al fundamento por el cual fue desarrollada OBI, esta se consideró como “ontología ideotipo”²⁴ u “ontología modelo”.

OBI, hace parte de la lista de ontologías candidatas dentro del proyecto Open Biological and Biomedical Ontologies (OBO) (Smith et al., 2007). Por lo tanto, la ontología cumple con los siguientes estándares: i) es expresada en Web Ontology Language^{25, 26} (OWL), ii) utiliza como base la ontología general BFO, por lo tanto todas las clases de OBI, son en realidad subclases de BFO, iii) usa la ontología de relaciones biológicas y biomédicas (OBO_REL) (Smith et al., 2005), finalmente vi) incluye un conjunto de metadatos para representar cada termino que hace parte de la ontología, ver figura 46.

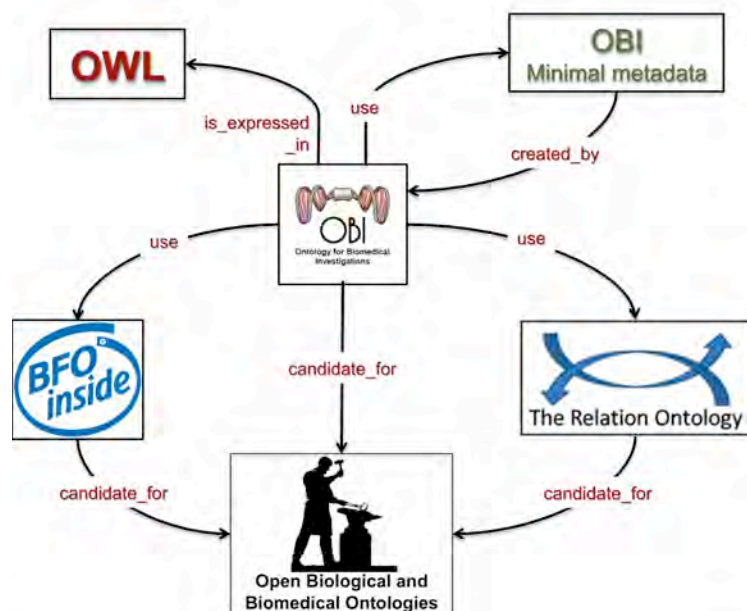


Figura 46. Mapa conceptual, que resume las características descritas sobre la ontología OBI.

²⁴ Ideotipo, en biología hace referencia a un individuo ideal, que contiene las características más favorables a un fin determinado.

²⁵ Web Ontology Language²⁵ (OWL), es un lenguaje que facilita expresar el sentido y la semántica de los términos y relaciones presentes en las ontologías.

²⁶ OWL Web Ontology Language, Overview. W3C Recommendation 10 February 2004. Available from: <http://www.w3.org/TR/owl-features/>.

- Identificación de ontologías que podrían reusarse. Al igual que OBI, la ontología M4L, reutiliza:

1. La ontología general Basic Formal Ontology (BFO), para seguir una estructura organizativa similar a las ontologías que hacen parte del proyecto OBO. En el anexo A, se presentan las clases de BFO a partir de las cuales se desarrolló M4L.
2. La ontología de relaciones biológicas y biomédicas (OBO_REL), para caracterizar los conceptos contenidos en ella. Las relaciones que provee OBO_REL, se presentan en el anexo B. Finalmente reutiliza,
3. la ontología de metadatos creada por OBI (ontology-metadata) (OBI), para representar cada término. En el anexo C, se presentan los metadatos que contiene ontology-metadata.

- identificación de conceptos que podrían reusarse. En los cuadernos de laboratorio analizados (capítulo 3), se observó como en ellos se registran todas las fases experimentales de distintos procedimientos de laboratorio, incluyendo las entidades involucradas en la preparación, ejecución e interpretación de esos procesos. Los procedimientos comúnmente registrados en los cuadernos son entre otros: la extracción de ADN, la técnica de PCR incluyendo algunas de sus variantes y la electroforesis en gel de agarosa y poliacrilamida. De acuerdo a esas observaciones, se procedió a buscar conceptos relacionados con los tres procedimientos experimentales antes mencionados.

La ontología M4L, contiene un total de 359 conceptos, de los cuales 145 son propios de este estudio y 214 fueron importados manualmente de 15 ontologías distintas disponibles en BioPortal, ver tabla 7. Los conceptos usados en la

ontología M4L y la proveniencia de los mismos, pueden ser visualizados en <http://www.biotea.ws>.

Tabla 7. Presentación del número de conceptos provenientes por ontología.

Ontología	No. de conceptos
Metadata for Laboratory Notebook (M4L)	145
Chemical Entities of Biological Interest (CHEBI) (Degtyarenko et al., 2008)	87
Ontology for Biomedical Investigation (OBI) (Brinkman et al., 2010)	59
Medical Subject Headings ontology (MSH) (Moerchen et al., 2008)	17
Gene Ontology (GO) (Ashburner et al., 2000)	14
Sample Processing and Separation Techniques (SEP) (http://psidev.info/index.php?q=node/312)	6
BIRN Project lexicon (BIRNLex) (Bug et al., 2008)	6
Gene Regulation Ontology (GRO) (Beisswanger et al., 2008)	5
National Cancer Institute thesaurus (NCIt) (Ceusters et al., 2005)	5
Plant Ontology Consortium (POC) (Jalswal et al., 2005)	5
SNOMED-CT (http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)	5
BioTop Ontology (Beisswanger et al., 2007)	1
Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003)	1
Ontology for Genetic Interval (OGI) (Lin et al., 2010)	1
Parasite Experiment Ontology (PEO) (http://wiki.knoesis.org/index.php/Parasite_Experiment_ontology)	1
Proteomics Data and Process Provenance (PDPP) (Sahoo et al., 2006)	1
Total de conceptos	359

4.3.3 Captura del Conocimiento que Será Representado en la Ontología M4L. En la figura 47 se ilustra el tercer paso desarrollado en este estudio para la construcción de M4L.



Figura 47. Captura del conocimiento que será representado en la ontología. Tercera etapa del desarrollo de M4L.

- Presentación de los casos de uso. Los casos de uso que se presentan en esta sección, contienen palabras o frases escritas entre comillas y en *letra cursiva* (itálica) que hace referencia a un termino específico adoptado de ontologías y/o vocabularios controlados. De igual manera, frases cortas unidas por un guión bajo (“_”), escritas en negrilla y letra cursiva representan las relaciones entre conceptos.

La ontología M4L toma como base la estructura de BFO y OBI. En estas ontologías se define “*investigación*”²⁷ como un “*proceso*”²⁸ que involucra varias partes; incluyendo la planificación general del “*diseño del estudio*”²⁹, ejecución del estudio diseñado y la documentación de los resultados. Una investigación,

²⁷ Investigation: ID: OBI_0000066. Tomado de Ontology Biomedical Investigation (OBI).

²⁸ process: Tomado de Basic Formal Ontology (BFO).

²⁹ study design: ID: OBI_0500000. Tomado de Ontology Biomedical Investigation (OBI).

típicamente incluye también la “*interpretación de datos*”³⁰ para sacar conclusiones.

Al igual que los procesos experimentales en ciencias biomédicas, los procesos relacionados con biotecnología de plantas involucran numerosos subprocesos. Estos a su vez incluyen materiales de experimentación; ya sean estos organismos completos (plantas), partes anatómicas (hojas, raíz, semillas, etc.), cultivos celulares o entidades moleculares (ácidos nucleicos).

OBI usa “*entidad material*”³¹ como la base para definir cosas físicas. Entidad material, **es_un** “*continuyente independiente*”³². Un “*continuyente*”³³, **es_una** entidad que existe como un todo en cualquier momento y persiste a través del tiempo manteniendo su identidad. Una entidad material es por ejemplo, un organismo, un tubo de ensayo, una centrífuga.

A su vez una entidad material puede tener “*roles*”³⁴ que se realizan dentro del marco de un proceso, por ejemplo, el “*rol de un objeto de estudio*”³⁵ el “*rol de un tampón*”³⁶, el “*rol de un cebador*”³⁷; y “*funciones*”³⁸, que dependen del diseño o estructura física de la entidad, por ejemplo, si cumple una “*función de medición*”³⁹, “*función de congelar*”⁴⁰ o una “*función mecánica*”⁴¹. La función, se considera

³⁰ interpreting data: ID: OBI_0000338. Tomado de Ontology Biomedical Investigation (OBI).

³¹ material_entity: Tomado de Basic Formal Ontology (BFO).

³² independent_continuant: Tomado de Basic Formal Ontology (BFO).

³³ continuant: Tomado de Basic Formal Ontology (BFO).

³⁴ role: Tomado de Basic Formal Ontology (BFO).

³⁵ study subject role: ID: OBI_0000097. Tomado de Ontology Biomedical Investigation (OBI).

³⁶ buffer role: ID: OBI_0302731. Tomado de Ontology Biomedical Investigation (OBI).

³⁷ primer role: ID: OBI_0000405. Tomado de Ontology Biomedical Investigation (OBI).

³⁸ function: Tomado de Basic Formal Ontology (BFO).

³⁹ measure function: ID: OBI_0000453. Tomado de Ontology Biomedical Investigation (OBI).

⁴⁰ freeze function: ID: OBI_0000375. Tomado de Ontology Biomedical Investigation (OBI).

inherente_a la entidad material y es ***realizada_por*** el rol que la entidad material juega en un proceso.

A continuación, se presentan tres casos de uso que describen e ilustran procedimientos de laboratorio o “*ensayos*”⁴², escogidos por ser ampliamente utilizados y por ende frecuentemente anotados en los cuadernos de laboratorio analizados en este estudio. El primer caso de uso es la extracción de ADN de alta calidad a pequeña escala en plantas. El segundo caso de uso hace referencia a los diferentes tipos de técnicas de PCR utilizados por los investigadores del laboratorio de Biotecnología del CIAT, el tercer caso de uso se enfoca en el proceso de electroforesis como medio de visualización de ácidos nucleicos. En los tres casos de uso, se muestran una serie de términos propios de este estudio y otros importados de ontologías externas o de vocabularios controlados.

Caso de uso 1: Extracción de ADN de alta calidad a pequeña escala en plantas.

La extracción de ADN a partir de tejidos vegetales es esencial para la aplicabilidad de varias técnicas empleadas en el análisis de un genoma. Es necesario obtener un ADN puro, con un alto rendimiento y de muy buena calidad. Un método de extracción de ADN que constituye una manera poco costosa, rápida, y fácil de obtenerlo, es el método de Dellaporta (Dellaporta et al., 1983). Este ha sido modificado para pruebas en el laboratorio de la Unidad de Agro Biodiversidad y Biotecnología del CIAT.

⁴¹ mechanical function: ID: OBI_0000379. Tomado de Ontology Biomedical Investigation (OBI).

⁴² assay: ID: OBI_0000070. Tomado de Ontology Biomedical Investigation (OBI).

Este ensayo de extracción de ADN involucra 3 procesos: i) cosecha del material vegetal, ii) proceso de obtención de tejido vegetal pulverizado a partir de hojas jóvenes y iii) el proceso de extracción y visualización del ADN.

i) Proceso 1: cosecha del material vegetal. El proceso “cosecha del material vegetal”⁴³, **es_un** ejemplo de “preparación de la muestra para un ensayo”⁴⁴, a su vez, **es_un** “proceso planeado”⁴⁵ adoptado de OBI.

1. Con la ayuda de una “tijera”⁴⁶, que cumple una *función mecánica*, se realiza el corte de “*hojas juveniles*”⁴⁷ sin zonas necróticas ni lesiones; también se pueden utilizar “*hojas adultas*”⁴⁸ que aún no llegan a la senescencia. Las “*hojas*”⁴⁹ cosechadas de las “*plantas*”⁵⁰ sembradas en el campo o en el invernadero, se presentan como *objeto de estudio*.

El concepto, *tijera*, importado de la ontología SNOMED-CT, **es_un** “*suministro de laboratorio*”⁵¹, clase creada en éste estudio. Los conceptos, *hojas*, *hojas juveniles*, *hojas adultas* y *plantas*, fueron importados de POC y usados según la definición lógica del ensayo.

⁴³ harvest of plant material: Tomado de Metadata for Laboratory Notebook (M4L)

⁴⁴ sample preparation for assay: ID: OBI_0000073. Tomado de Ontology Biomedical Investigation (OBI).

⁴⁵ planned process: ID: OBI_0000073. Tomado de Ontology Biomedical Investigation (OBI).

⁴⁶ scissors: ID: 64973003. Tomado de SNOMED-CT

⁴⁷ juvenile leaf: PO:0006339. Tomado de Plant Ontology Consortium (POC).

⁴⁸ adult leaf: PO:0006340. Tomado de Plant Ontology Consortium (POC).

⁴⁹ leaf: PO: 0009025. Tomado de Plant Ontology Consortium (POC).

⁵⁰ whole plant: PO: 0000003. Tomado de Plant Ontology Consortium (POC).

⁵¹ laboratory supply: tomado de Metadata for Laboratory Notebook (M4L).

2. Luego de cosechar las hojas, éstas se introducen en “*bolsas plásticas zip lock*”⁵², con la etiqueta que identifica la muestra. También se puede utilizar “*papel aluminio*”⁵³.

Los conceptos, *bolsas plásticas zip lock* (Withers, 2010) y *papel aluminio*^{54, 55}, son *suministros de laboratorio*; adicionalmente, ambos conceptos, hacen parte del conjunto de metadatos M4L, creados en este estudio.

3. Las hojas, se mantienen en una “*hielera portátil*”⁵⁶ u otro recipiente con hielo para mantener frías las hojas. Las bolsas plásticas zip lock cumplen con una función de almacenamiento y de evitar que se humedezcan las hojas. La *hielera portátil* tiene una “*función de enfriar*”⁵⁷. La razón de mantener las muestras a baja temperatura es evitar daños del ADN a causa de las “*enzimas DNasas*”⁵⁸ que empiezan a actuar cuando se produce el corte de la hoja.

La *enzima DNasa*, **es_un** ejemplo de “*biomacromolécula*”⁵⁹ clase importada de CHEBI. La, *hielera portátil*, **es_un** *suministro de laboratorio*. Las definiciones de ambos conceptos, *enzima DNasa* (Weir, 1994) y *hielera portátil* (Woods and Worsham, 1989) son propias de este estudio y adaptadas según las necesidades propias del ensayo.

⁵² reclosable bag: tomado de Metadata for Laboratory Notebook (M4L).

⁵³ aluminum foil: tomado de Metadata for Laboratory Notebook (M4L).

⁵⁴ European Aluminium Foil Association (EAFA); available from: http://www.alufoil.org/front_content.php

⁵⁵ The Aluminum Association, available from: http://www.alufoil.org/front_content.php

⁵⁶ portable ice chest: tomado de Metadata for Laboratory Notebook (M4L).

⁵⁷ cool function: ID: OBI_0000387. Tomado de Ontology Biomedical Investigation (OBI).

⁵⁸ deoxyribonuclease: tomado de Medical Subject Headings (MeSH).

⁵⁹ biomacromolecule: tomado de Chemical Entities of Biological Interest (CHEBI).

4. Una vez en el laboratorio, el material vegetal cosechado se almacena en un “congelador”⁶⁰ a -80°C. La “función de congelación”⁶¹ de esta entidad material, garantiza la conservación del material vegetal por largos periodos de tiempo. El concepto, *congelador*, fue importado de PEO y usado según la definición lógica del ensayo. De igual manera, *congelador es_un* “equipo de laboratorio”⁶², clase importada de SNOMED-CT.

En la figura 48, se presenta a manera de mapa conceptual el proceso de cosecha del material vegetal.

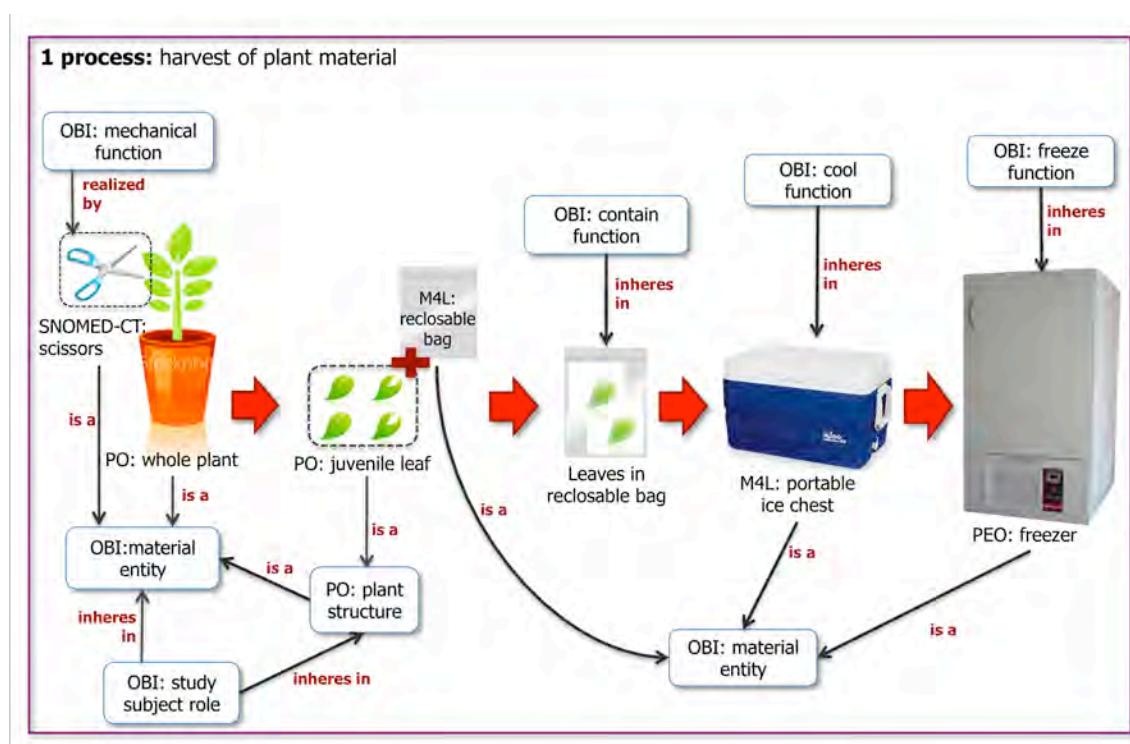


Figura 48. Mapa conceptual del proceso, cosecha del material vegetal.

⁶⁰ freezer: tomado de Parasite Experiment Ontology (PEO).

⁶¹ freeze function: ID: OBI_0000375. Tomado de Ontology Biomedical Investigation (OBI).

⁶² laboratory equipment: ID: 430853006. Tomado de SNOMED-CT.

ii) Proceso 2. Pulverización mecánica del material vegetal. El proceso “*pulverización del material vegetal*”⁶³ es fundamental para facilitar el rompimiento de las células y paredes celulares para permitir la liberación del ADN. El no lograr una óptima pulverización del tejido vegetal, reduce marcadamente la obtención final del ADN. El proceso pulverización mecánica del material vegetal, también **es_un** ejemplo de *preparación de la muestra para un ensayo*, que a su vez **es_un** *proceso planeado*, clase adoptada de OBI.

1. Esta etapa inicia sacando las hojas que se almacenaron en un *congelador* a -80°C al final del proceso 1. Las “*hojas congeladas*”⁶⁴, son un ejemplo de la clase “*cualidad de una entidad física única*”⁶⁵ representada en PATO.
2. El material vegetal *objeto de estudio*, al final de este proceso es convertido en un “*polvo fino*”⁶⁶. Este proceso de transformación se logra con la ayuda de “*nitrógeno liquido*”⁶⁷ cuya función es procurar una congelación rápida del material vegetal, para evitar el daño de estructuras y por ende del ADN; junto con la acción mecánica de un “*mazo y un mortero*”⁶⁸ (figura 49). El *polvo fino del material vegetal* es por lo tanto, un ejemplo de la clase “*material procesado*”⁶⁹ de OBI.

⁶³ mechanical pulverization of plant material: Tomado de Metadata for Laboratory Notebook (M4L).

⁶⁴ frozen leaf: Tomado de Metadata for Laboratory Notebook (M4L).

⁶⁵ quality of a single physical entity: ID: PATO_0001237. Tomado de Phenotypic Quality Ontology (PATO).

⁶⁶ fine powder of plant material: Tomado de Metadata for Laboratory Notebook (M4L).

⁶⁷ liquid nitrogen: ID: 52793004. Tomado de SNOMED-CT.

⁶⁸ mortar and pestle: Tomado de Metadata for Laboratory Notebook (M4L).

⁶⁹ processed material: ID: OBI_0000047. Tomado de Ontology for Biomedical Investigation (OBI).

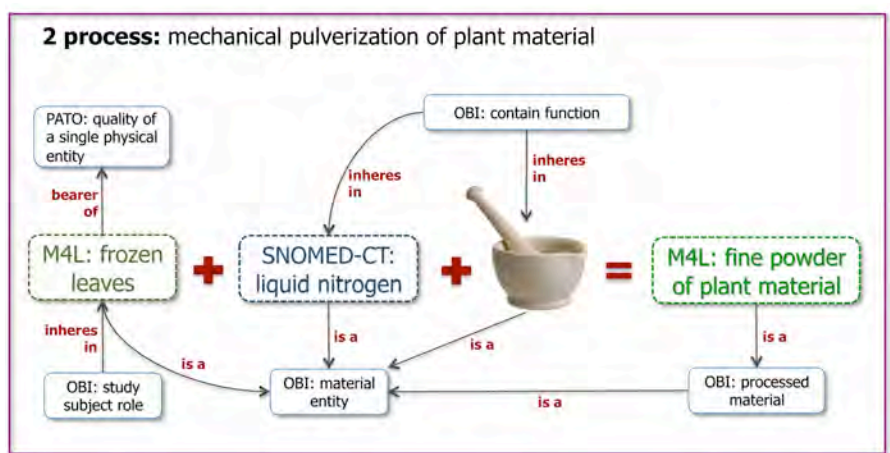


Figura 49. Mapa conceptual del proceso, pulverización mecánica del material vegetal.

El concepto, *nitrógeno líquido*, fue importado de la ontología SNOMED-CT, éste concepto, a su vez **es una** “entidad química”⁷⁰, clase importada de la ontología PDPP. El concepto *mazo y mortero*^{71, 72}, hace parte del conjunto de metadatos M4L y **es un** suministro de laboratorio.

iii) Proceso 3. Extracción de ADN (basado en el método Dellaporta, 1983). El proceso “extracción de ADN”⁷³, **es una** subclase de “extracción de ácidos nucleicos”⁷⁴ en OBI. Dado que la cantidad de ADN que se obtenga depende en buena parte del método utilizado, a continuación se presentan los pasos

⁷⁰ chemical entity: Tomado de Proteomics Data and Process Provenance (PDPP).

⁷¹ Mortar and Pestle. Wikipedia; available from: http://en.wikipedia.org/wiki/Mortar_and_pestle.

⁷² Illustrated Glossary of Antique Terms. Available from: http://www.ijohnsonappraisals.com/Illustrated_Glossary_of_Antique_Terms_Pages/Glossary_M.htm.

⁷³ DNA extraction: ID: OBI_0000257. Tomado de Ontology Biomedical Investigation (OBI).

⁷⁴ nucleic acid extraction: ID: OBI_0666667. Tomado de Ontology Biomedical Investigation (OBI).

fundamentales de uno de los “*protocolos*”⁷⁵ mas usados en el laboratorio de Biotecnología del CIAT para el aislamiento de ADN genómico.

1. Colocar 0,2 g del material vegetal pulverizado en un “*tubo de microcentrifuga*”⁷⁶. Posteriormente agregar el “*tampón de extracción*”⁷⁷, cuyo rol es disolver componentes de tejidos y membranas celulares, permitiendo la liberación del ADN. Incubar los tubos a 65°C en un “*baño de agua*”⁷⁸ por 30 min.

La entidad material *tubo de microcentrífuga*, **es_un** suministro de laboratorio. *tampón de extracción*, **es_un** “*tampón*”⁷⁹, clase adoptada de la ontología CHEBI. Ambos conceptos, *tubo de microcentrífuga*^{80,81} y *tampón de extracción* (Kawata et al., 2003), hacen parte del conjunto de metadatos M4L. La entidad material, *baño de agua*, fue importada de la ontología OBI.

2. Agregar 5M “*acetato de potasio*”⁸², cuya función es facilitar la remoción de proteínas y polisacáridos de los ácidos nucleicos. Mantener los tubos en hielo con agitación constante por 15 min.

El concepto, *acetato de potasio*, fue importado de la ontología CHEBI.

⁷⁵ protocol: ID: OBI_0000272. Tomado de Ontology Biomedical Investigation (OBI).

⁷⁶ microcentrifuge tube: Tomado de Metadata for Laboratory Notebook (M4L).

⁷⁷ DNA extraction buffer: Tomado de Metadata for Laboratory Notebook (M4L).

⁷⁸ water bath: ID: OBI_0400155. Tomado de Ontology Biomedical Investigation (OBI).

⁷⁹ buffer: CHEBI:35225. Tomado de Chemical Entities of Biological Interest (CHEBI).

⁸⁰ Microcentrifuge Tube. Wikipedia; available from: http://en.wikipedia.org/wiki/Microcentrifuge_tube#Microcentrifuge_tubes

⁸¹ *Sea Urchin Embryology Glossary*. Available from: <http://stanford.edu/group/Urchin/glossary.htm>.

⁸² potassium acetate: CHEBI:32029. Tomado de Chemical Entities of Biological Interest (CHEBI).

3. Colocar los tubos en una “*microcentrífuga*”⁸³, con el propósito de separar los restos de tejido vegetal, proteínas y polisacáridos (precipitados al fondo del tubo), del ADN (contenido en el sobrenadante).

La *microcentrífuga* (Marziali, 2001), **es_un** equipo de laboratorio.

4. Transferir el sobrenadante a un nuevo *tubo de microcentrífuga*. Agregar “*isopropanol*”⁸⁴ (2-propanol) frío y 3M “*acetato de sodio*”⁸⁵, ambas *entidades químicas* tienen el rol de precipitar el ADN. Mezclar por inversión para favorecer la precipitación de éste ácido nucleico. Se mantienen las muestras en un *congelador* a -20°C, por 2 h o durante toda la noche.

El *congelador* a -20° C **es_un** equipo de laboratorio.

5. Someter a centrifugación los tubos para terminar de precipitar y formar el botón de ADN. Descartar el sobrenadante y agregar “*etanol*”⁸⁶ al 70%, cuyo rol es lavar este ácido nucleico.

La *entidad material* “*ADN*”⁸⁷, **es_un** ejemplo de *biomacromolecula*, clase importada de CHEBI. EL *etanol* al 70% **es_un** “*alcohol*”⁸⁸, clase importada de la ontología CHEBI.

6. Descartar el *etanol* por inversión. Dejar evaporar el *etanol* a temperatura ambiente hasta que el botón de *ADN* se seque.

⁸³ microcentrifuge: Tomado de Metadata for Laboratory Notebook (M4L).

⁸⁴ propan-2-ol: CHEBI:17824. Tomado de Chemical Entities of Biological Interest (CHEBI).

⁸⁵ sodium acetate: CHEBI:32954. Tomado de Chemical Entities of Biological Interest (CHEBI).

⁸⁶ ethanol: CHEBI:16236. Tomado de Chemical Entities of Biological Interest (CHEBI).

⁸⁷ deoxyribonucleic acid: CHEBI:16991, Tomado de Chemical Entities of Biological Interest (CHEBI).

⁸⁸ alcohol: CHEBI:30879. Tomado de Chemical Entities of Biological Interest (CHEBI).

7. Agregar “*tampón TE*”⁸⁹ (o agua doblemente destilada y estéril) y “*enzima RNasa*”⁹⁰ e incubar por 30 min a 37°C. El propósito del *tampón TE* es proteger el ADN de la degradación y la *enzima RNasa* digiere o degrada moléculas de ARN que puedan interferir en los posteriores análisis del ADN.

El *Tampón TE* **es_un** ejemplo de *tampón*, clase importada de la ontología CHEBI. El concepto *enzima RNasa*, **es_un** ejemplo de *biomacromolecula*, clase importada de CHEBI. De igual manera las definiciones de los conceptos, *tampón TE*⁹¹ y *enzima RNasa* (Burrell, 1994), son propias de este estudio y adaptadas según las necesidades propias del ensayo.

8. Verificar la calidad y cantidad del ADN en un “*gel de agarosa*”⁹². Guardar las muestras a 4°C hasta utilizarlas; si no las va a utilizar en mucho tiempo, almacénelas a -20°C. La *entidad material, gel de agarosa*, **es_un** ejemplo de “*gel*”⁹³, clase importada de la ontología SEP.

Nota: Ver el caso de uso 3 (electroforesis basada en gel), donde se presentan los detalles acerca de la preparación del gel, las condiciones de corrida y la visualización del ADN.

En la figura 50, se presenta a manera de mapa conceptual el proceso de extracción del material vegetal.

⁸⁹ tris-EDTA buffer: Tomado de Metadata for Laboratory Notebook (M4L).

⁹⁰ ribonuclease: tomado de Metadata for Laboratory Notebook (M4L).

⁹¹ TE buffer. Wikipedia; available from: http://en.wikipedia.org/wiki/TE_buffer

⁹² agarose gel: ID: OBI_0000439. Tomado de Ontology Biomedical Investigation (OBI).

⁹³ gel: sep:00110. Tomado de Sample Processing and Separation Techniques (SEP).

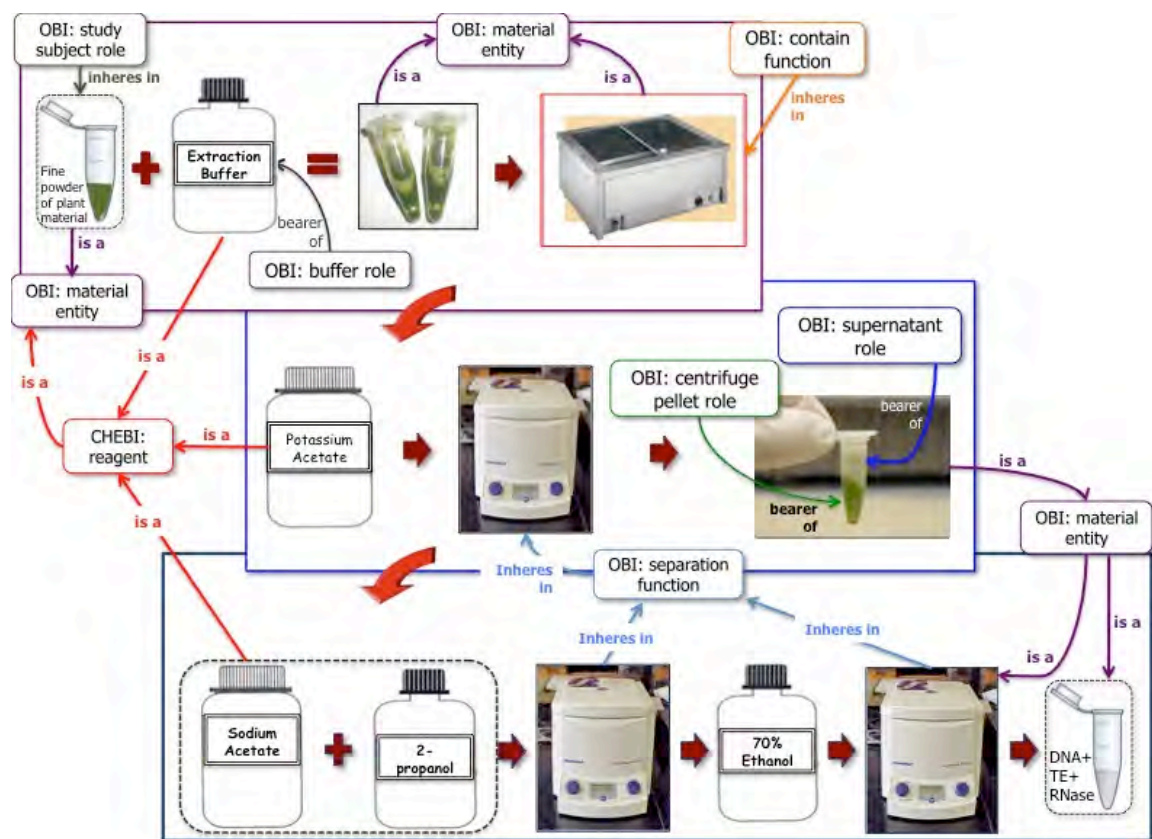


Figura 50. Mapa conceptual del proceso de extracción de ADN de alta calidad a pequeña escala en plantas.

Caso de uso 2: Amplificación de ácidos nucleicos *in Vitro*

El principal objetivo de las técnicas de “*amplificación de ácidos nucleicos in Vitro*”⁹⁴, es mejorar la sensibilidad de los ensayos o procedimientos de laboratorio basados en el uso de ácidos nucleicos, y hacerlos mas simples mediante la automatización y la incorporación de sistemas de detección no radioactivos.

En la actualidad, existen diferentes técnicas de amplificación de ácidos nucleicos y las mismas se mantienen en continua evolución con el avance de la ciencia. La

⁹⁴ in vitro nucleic acid amplification: Tomado de Metadata for Laboratory Notebook (M4L).

primera de estas técnicas es la “*Reacción en Cadena de la Polimerasa*”⁹⁵ (PCR), descrita por Kary Mullis en 1983 (Mullis, 1990b), la PCR se ha convertido en una técnica fundamental en la mayoría de laboratorios de biología molecular y a partir de su forma básica han surgido algunas variantes como la “*multiplex PCR*”⁹⁶, “*Real Time PCR*”⁹⁷, “*Whole Genome PCR*”⁹⁸, “*Reverse Transcriptase PCR*”⁹⁹, entre otras (Wang et al., 1993, Kinzler and Vogelstein, 1989, Gachon et al., 2004, Ciapina et al., 2004).

La técnica TAS “*Transcription-based Amplification System*”¹⁰⁰, fue el primer sistema de amplificación no basado en la PCR, descrito en 1989 por Kwoh *et al.* (Kwoh et al., 1989), y se basa en la amplificación mediante transcripción *in Vivo*. Una alternativa para solventar los inconvenientes de la TAS, es por medio de la técnica 3SR “*Self-sustaining Sequence Replication*”¹⁰¹, también conocida como NASBA (Nucleic Acid Sequence-based Amplification); ésta es una técnica de amplificación isotérmica (no requiere cambios en la temperatura de incubación), que se basa en las actividades enzimáticas simultáneas de tres enzimas: “*RNasa H de E. coli*”¹⁰², “*transcriptasa inversa del virus de la mieloblastosis aviar*”¹⁰³ y la “*RNA polimerasa del fago T7*”¹⁰⁴. La “*Reacción en Cadena de la Ligasa*”¹⁰⁵ (LCR), es otra técnica alternativa a las técnicas de amplificación dependientes de la

⁹⁵ polymerase chain reaction: ID: D016133, tomado de Medical Subject Headings (MeSH).

⁹⁶ multiplex polymerase chain reaction: tomado de Metadata for Laboratory Notebook (M4L).

⁹⁷ real time polymerase chain reaction: tomado de Metadata for Laboratory Notebook (M4L).

⁹⁸ whole genome amplification: tomado de NCI Thesaurus.

⁹⁹ reverse transcriptase polymerase chain reaction: ID: D020133, tomado de Medical Subject Headings (MeSH).

¹⁰⁰ transcripcion-based amplification system: tomado de Metadata for Laboratory Notebook (M4L).

¹⁰¹ self-sustaining sequence replication: ID: D021161, tomado de Medical Subject Headings (MeSH).

¹⁰² ribonuclease H: ID: D016914, tomado de Medical Subject Headings (MeSH).

¹⁰³ AMV reverse transcriptase: tomado de Metadata for Laboratory Notebook (M4L).

¹⁰⁴ bacteriophage T7 RNA polymerase: ID: C062210, tomado de Medical Subject Headings (MeSH).

¹⁰⁵ ligase chain reaction: ID: D022521, tomado de Medical Subject Headings (MeSH).

enzima polimerasa, fue también descrita en el año 1989 por Wu y Wallace (Wu and Wallace, 1989). El método, utiliza la DNA ligasa para unir dos pares de sondas complementarias después de que éstas se hayan unido a la secuencia molde *in Vitro*.

Dado que la técnica de amplificación de ácidos nucleicos *in Vitro* mas usada en biotecnología de plantas es la PCR, a continuación se amplía la descripción del método, considerándose: i) que es la PCR, ii) con que ingredientes se practica la reacción, y finalmente iii) como funciona la PCR.

i) Que es la PCR. La PCR (Reacción en Cadena de la Polimerasa) es un método *in Vitro* de síntesis de ADN. La amplificación de un segmento particular del ADN se logra al ser éste delimitado por un par de “cebadores”¹⁰⁶ o iniciadores que lo flanquean. Su copiado se logra en forma exponencial a través de repetidos ciclos de diferentes periodos de tiempo y temperaturas de incubación en presencia de una enzima, la “ADN polimerasa termoestable”¹⁰⁷. Así se obtienen en cuestión de horas millones de copias de la secuencia deseada de ADN (Mullis, 1990a).

ii) Ingredientes usados en la reacción de PCR. Para llevar a cabo una PCR se necesita mezclar en un tubo el ADN que contiene la secuencia a amplificar, es decir el ADN molde. Al mismo tubo se agregan ambos *cebadores* que se alinearán a las cadenas simples del ADN, la “mezcla de los dNTPs”¹⁰⁸ en cantidades suficientes, el “*tampón de PCR*”¹⁰⁹, la enzima *ADN polimerasa*

¹⁰⁶ forward primer, reverse primer: tomado de Metadata for Laboratory Notebook (M4L).

¹⁰⁷ thermostable DNA polymerase: tomado de Metadata for Laboratory Notebook (M4L).

¹⁰⁸ dNTP mix: tomado de Metadata for Laboratory Notebook (M4L).

¹⁰⁹ PCR buffer: tomado de Metadata for Laboratory Notebook (M4L).

termoestable y “agua grado PCR”¹¹⁰ para completar el volumen final de la reacción.

Los *cebadores*, son “*oligonucleótidos*”¹¹¹ químicamente sintetizados. La clase, *oligonucleótidos*, fue importada de la ontología CHEBI. Los conceptos *mezcla de dNTPs*^{112,113}, *tampón de PCR*¹¹⁴, *agua grado PCR*^{115,116} y *ADN polimerasa termoestable* (Abu Al-Soud and Radstrom, 1998, Davalieva and Efremov, 2010) hacen parte del grupo de metadatos M4L. En la figura 51, se presenta a manera de mapa conceptual los ingredientes usados en la reacción de PCR.

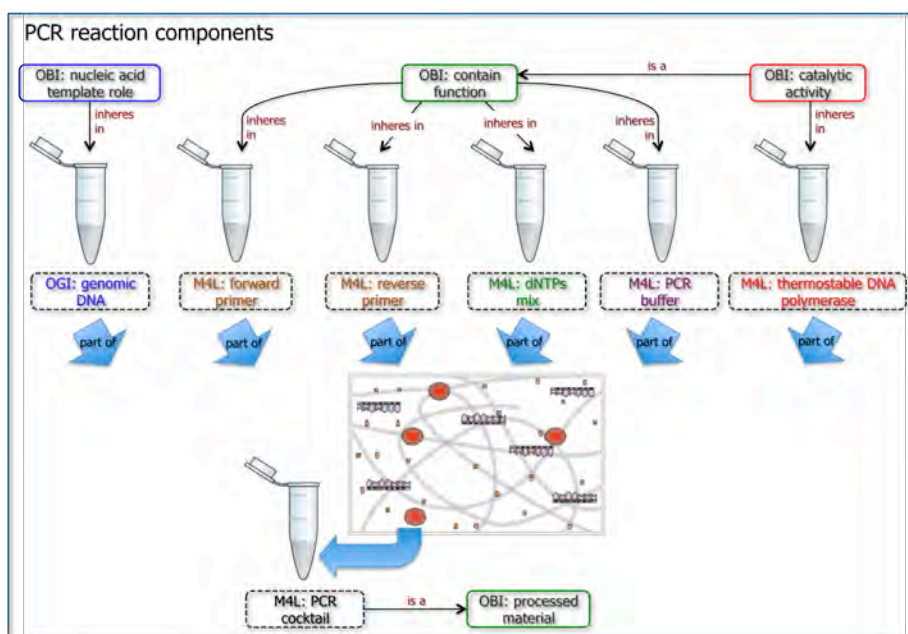


Figura 51. Mapa conceptual que ilustra los componentes de una reacción de PCR.

¹¹⁰ PCR grade water: tomado de Metadata for Laboratory Notebook (M4L).

¹¹¹ oligonucleotide: ID: CHEBI:7754, Tomado de Chemical Entities of Biological Interest (CHEBI).

¹¹² (2005) Deoxynucleoside Triphosphate Set, PCR Grade. Roche Applied Science.

¹¹³ *Protocols & Applications Guide: Nucleic Acid Amplification*. Available from: <http://www.promega.com/paguide/chap1.htm>.

¹¹⁴ PCR Glossary. PCRLinks.com. Available from: <http://www.pcrlinks.com/glossary/B.htm>.

¹¹⁵ (2007) Water, PCR Grade. IN ROCHE (Ed.). Roche Applied Science.

¹¹⁶ *PCR Grade Water*. Available from: <http://www.biocompare.com/Articles/ApplicationNote/778/PCR-Grade-Water.html>.

iii) Como funciona la PCR. La reacción consta, de ciclos repetitivos conformados cada uno de tres etapas: el primero consiste en la ruptura de los puentes de hidrogeno de ADN para desnaturalizarlo, para ello, se incuba a una temperatura alrededor de los 95°C. Esta etapa de “*desnaturalización*”¹¹⁷, expone las bases nitrogenadas del ADN molde. En la segunda etapa, ocurre la “*hibridación*”¹¹⁸ de las cadenas desnaturalizadas del ADN con los *cebadores*, a una temperatura que generalmente oscila entre 50 y 60°C. La tercera etapa se efectúa a 72°C, temperatura a la que la *enzima ADN polimerasa* “*extiende*”¹¹⁹ la longitud de los cebadores, añadiendo los diferentes nucleótidos libres en el orden que le va dictando la secuencia de nucleótidos que actúa como molde. La reacción de PCR se lleva a cabo en un “*termociclador*”¹²⁰, el cual es un *equipo de laboratorio* que puede programarse para cambiar de temperatura rápidamente en determinados tiempos. Gracias a estos equipos, el proceso de la PCR se realiza de manera automática.

En resumen, la *PCR*, es una “*amplificación enzimática*”¹²¹, clase importada de OBI. Donde al final del proceso se obtiene “*ADN amplificado*”¹²², *entidad material* que **es_un** ejemplo de *material procesado*, clase importada de OBI. La obtención del *ADN amplificado* se logra en forma exponencial a través de repetidos ciclos donde cada ciclo incluye una etapa de *desnaturalización*, *hibridación* y *extensión*. Los conceptos, *desnaturalización*, *hibridación* y *extensión* hacen parte del conjunto de metadatos M4L y definidos según las necesidades propias del ensayo. Finalmente, la reacción de PCR es realizada de manera automática en un

¹¹⁷ denaturing_activity: tomado de Metadata for Laboratory Notebook (M4L).

¹¹⁸ annealing activity: tomado de Metadata for Laboratory Notebook (M4L).

¹¹⁹ extension activity: tomado de Metadata for Laboratory Notebook (M4L).

¹²⁰ thermal cycler: ID: OBI_0400116, tomado de Ontology Biomedical Investigation (OBI).

¹²¹ enzymatic amplification: ID: OBI_0600058, tomado de Ontology Biomedical Investigation (OBI).

¹²² amplified DNA: OBI_0000809, tomado de Ontology Biomedical Investigation (OBI).

termociclador, concepto importado de OBI. En la figura 52, se presenta a manera de mapa conceptual las etapas de la PCR, obtención del producto amplificado y su visualización mediante electroforesis (los detalles del proceso de la electroforesis basada en gel, se muestran en el caso de uso 3).

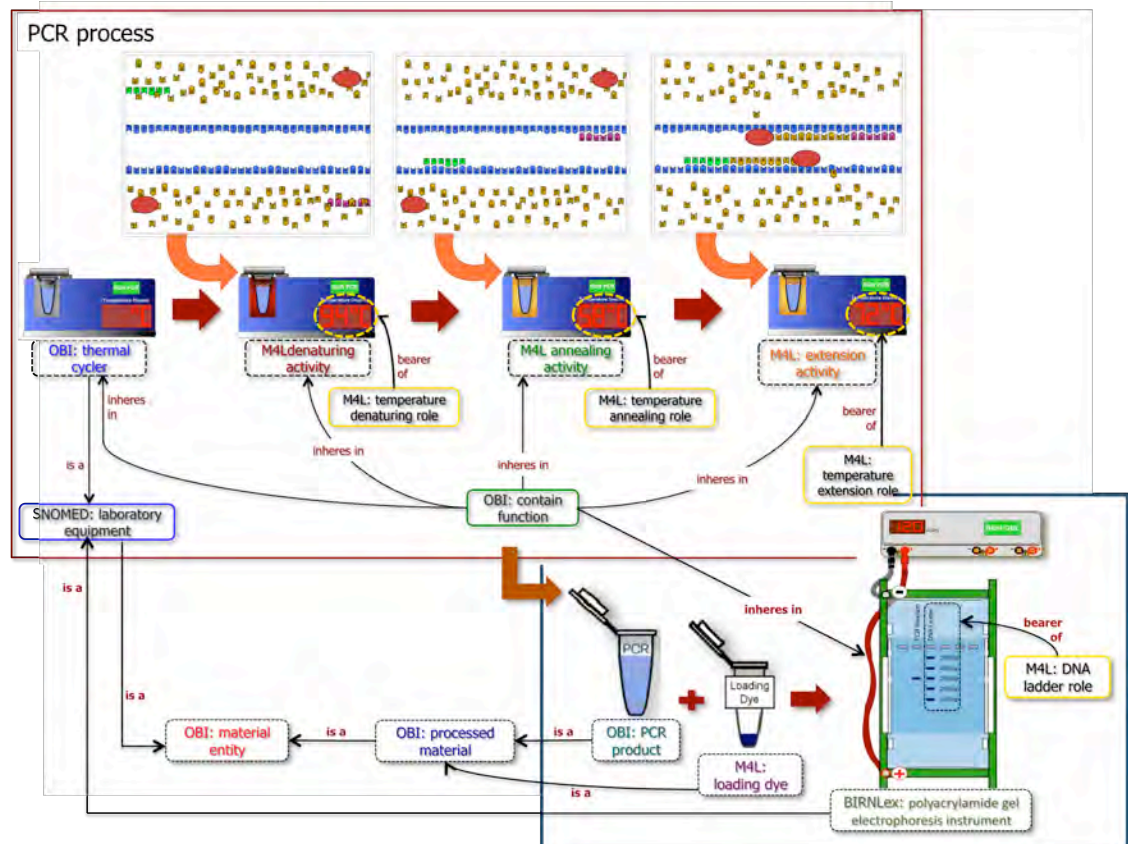


Figura 52. Mapa conceptual del proceso de la PCR. Imágenes tomadas de University Rochester Medical Center¹²³.

Los investigadores del proyecto Agro Biodiversidad y Biotecnología del CIAT, utilizan ampliamente la técnica de PCR por considerarse altamente específica, rápida, sensible y versátil. Por ejemplo, en el área de selección asistida por marcadores y mapeo genético, la PCR es la base de innumerables métodos

¹²³ Life Sciences Learning Center. Available from: <http://lifesciences.envmed.rochester.edu/animation.html>.

usados para la identificación de genes relacionados con diferentes tipos de estreses bióticos o abióticos (Blair et al., 2007, López et al., 2007), también es usada para facilitar la selección de plantas con características agronómicas de interés como el rendimiento y sus componentes (Gutiérrez et al., 2010, Aluko et al., 2004). De igual manera, la PCR ha servido para medir la expresión de genes (Lopez et al., 2005, Beltrán et al., 2008).

Caso de uso 3: Electroforesis basada en gel

La electroforesis es un método en el que se separan *biomacromoléculas* (*ADN*, “*ARN*”¹²⁴ o “*proteínas*”¹²⁵) bajo la influencia de un campo eléctrico; en dependencia de una combinación de su carga, peso molecular y estructura tridimensional. Los métodos electroforéticos son de alta sensibilidad, poder de resolución y versatilidad que consta de cinco componentes principales: la corriente eléctrica, la muestra (*ADN*, *ARN* o *proteínas*), el medio de soporte (celulosa, almidón, *gel de agarosa*, “*gel de poliacrilamida*”¹²⁶, etc.), el “*tampón de electroforesis*”¹²⁷ (“*TAE*”¹²⁸, “*TBE*”¹²⁹, etc.) y el sistema de detección (tinción).

El equipamiento requerido es simple (“*fuentes de poder*”¹³⁰ y “*cámara de electroforesis*”¹³¹ donde se coloca el medio de soporte). Con el desarrollo de la

¹²⁴ ribonucleic acid: ID: CHEBI:33697, tomado de Chemical Entities of Biological Interest (CHEBI)

¹²⁵ protein: ID: GRO:Protein, tomado de Gene Regulation Ontology (GRO)

¹²⁶ polyacrylamide gel: ID: OBI_0000432, tomado de Ontology for Biomedical Investigation (OBI)

¹²⁷ electrophoresis buffer: tomado de Metadata for Laboratory Notebook (M4L)

¹²⁸ TAE buffer: tomado de Metadata for Laboratory Notebook (M4L)

¹²⁹ TBE buffer: tomado de Metadata for Laboratory Notebook (M4L)

¹³⁰ laboratory power supply: tomado de Metadata for Laboratory Notebook (M4L)

¹³¹ electrophoresis instrument: ID: birnlex_2422, tomado de BIRNLex.

ciencia se han diseñado “*equipos automatizados de electroforesis*”¹³² que facilitan el análisis y secuenciación de macromoléculas biológicas como el “LI-COR 4200 sequencer”¹³³ (McArthur et al., 2000), “ABI PRISM 377 DNA Sequencer”¹³⁴ (Chartier, 2005, 1998) entre otros.

En el Proyecto Agro Biodiversidad y Biotecnología del CIAT, se usan métodos electroforéticos manuales y automatizados, basados en el uso de *gel de agarosa* y *poliacrilamida*. Entre los usos mas comunes de la electroforesis se destacan: medición del rendimiento y pureza de la extracción de ácidos nucleicos, calidad de productos amplificados y secuenciación automática del ADN. A continuación se describe un protocolo comúnmente usado de electroforesis manual utilizando como medio de soporte el *gel de agarosa*.

Electroforesis en gel de agarosa: La “*electroforesis en gel de agarosa*”¹³⁵ es un método utilizado en bioquímica y biología molecular para analizar y caracterizar “*ácidos nucleicos*”¹³⁶ de distintas procedencias. A continuación se describen los pasos de tres procesos fundamentales en la electroforesis: i) la preparación del *gel*, ii) carga de las muestras en el gel y iii) visualización.

¹³² automated DNA sequencer: tomado de Metadata for Laboratory Notebook (M4L)

¹³³ LI-COR 4200 sequencer: tomado de Metadata for Laboratory Notebook (M4L)

¹³⁴ ABI_PRISM_377_DNA_sequencer: tomado de Metadata for Laboratory Notebook (M4L)

¹³⁵ agarose gel electrophoresis: sep:00171, tomado de Sample Processing and Separation Techniques (SEP)

¹³⁶ nucleic acid: CHEBI:33696; tomado de Chemical Entities of Biological Interest (CHEBI)

i) Preparación del gel de agarosa.

1. determinar la concentración y el volumen del *gel* que se necesita. Por ejemplo si el *DNA* tiene un tamaño desconocido se debe utilizar un gel de agarosa al 1% (1g de agarosa/100ml de TBE 0.5X). Para visualizar “*productos de PCR*”¹³⁷ es necesario utilizar un *gel* al 2% (1g de agarosa/100ml de TBE 0.5X).
2. el volumen de gel a preparar depende del tamaño de la *cámara de electroforesis* que se disponga en el laboratorio. Por ejemplo, para preparar 40 ml de agarosa al 1%, se debe transferir a un frasco “*erlenmeyer*”¹³⁸:
 - 0.4g de “*agarosa*”¹³⁹
 - 40 ml de TBE 0.5X.
3. calentar en un “*horno microondas*”¹⁴⁰ o *baño maría* hasta que se disuelva la agarosa totalmente (la solución se torna transparente). **Nota:** es importante el uso de “*guantes resistentes al calor*”¹⁴¹ en los pasos 3, 4 y 5.
4. Enfriar la *agarosa* hasta alcanzar una temperatura aproximada de 60°C.
5. Servir la solución de agarosa por un extremo de la “*bandeja*”¹⁴² (previamente sellada con “*cinta de enmascarar*”¹⁴³), para evitar la formación de burbujas. Si

¹³⁷ PCR product: ID: OBI_0000406, tomado de Ontology for Biomedical Investigation (OBI)

¹³⁸ erlenmeyer flask: tomado de Metadata for Laboratory Notebook (M4L)

¹³⁹ agarose: CHEBI: 2511; tomado de Chemical Entities of Biological Interest (CHEBI)

¹⁴⁰ laboratory microwave oven: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴¹ heat resistant glove: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴² gel_casting_tray: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴³ masking tape: tomado de Metadata for Laboratory Notebook (M4L)

se forman burbujas, se pueden remover usando una “*punta para micropipeta*”¹⁴⁴.

6. Colocar el “*peine*”¹⁴⁵ en la *bandeja*, cuya función es formar los pozos en el gel. El *peine* y la *bandeja* son suministros de laboratorio y hacen ***parte_de*** la *cámara de electroforesis para geles de agarosa*.
7. Dejar solidificar el gel a temperatura ambiente (aproximadamente 30 minutos). El gel sólido tiene una apariencia opaca y ***es_un*** ejemplo de *material procesado*. Para que el gel quede de un espesor uniforme, la bandeja que contiene el gel debe estar en una superficie plana que esté nivelada.

Los conceptos usados en esta fase de la electroforesis provienen de OBI (productos de PCR), de CHEBI (agarosa) y de M4L (erlenmeyer, horno microondas, guantes resistentes al calor, bandeja, cinta de enmascarar, punta para micropipeta y peine). En la figura 53, se ilustra a manera de mapa conceptual el proceso de preparación del gel de agarosa.

¹⁴⁴ micropipette_tip: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴⁵ comb: tomado de Metadata for Laboratory Notebook (M4L)

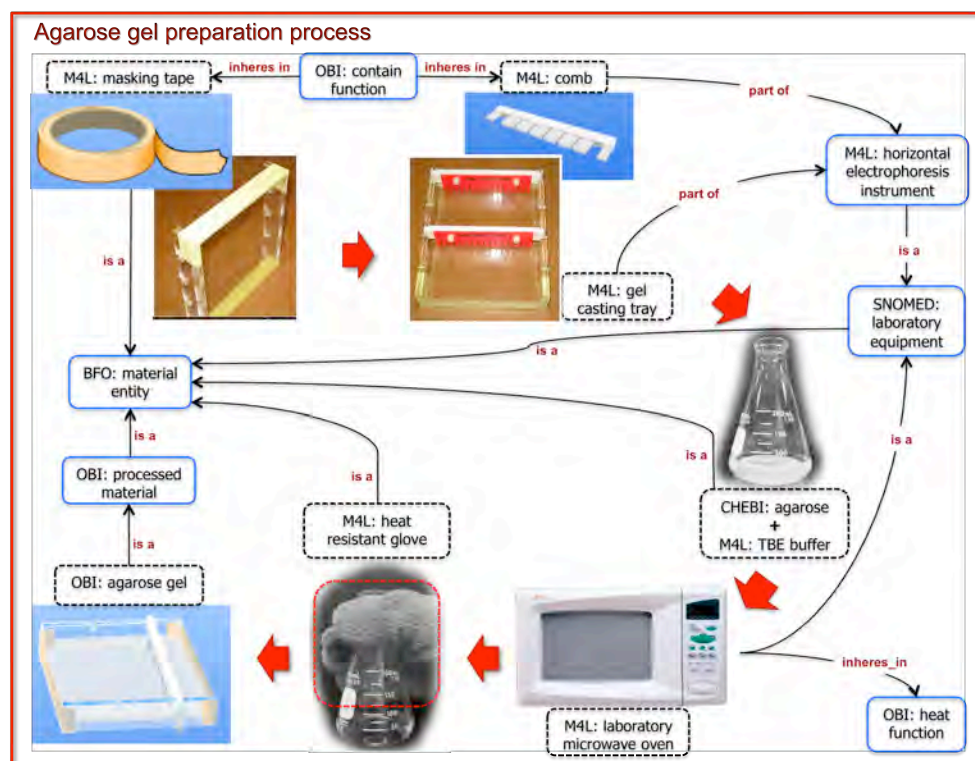


Figura 53. Mapa conceptual del proceso preparación del gel de agarosa.

ii) Carga y electroforesis de las muestras en un gel de agarosa

Nota: en esta etapa es recomendable el uso de “*guantes de latex*”¹⁴⁶.

1. Colocar la bandeja que contiene el *gel* sólido dentro de la *cámara de electroforesis*.
2. Añadir suficiente *TBE 0.5X* hasta cubrir completamente el *gel*. La principal función de este *tampón de electroforesis* es controlar el pH del gel, prevenir el daño y controlar el estado de ionización de las muestras en el gel.

¹⁴⁶ latex glove: tomado de Metadata for Laboratory Notebook (M4L)

3. En “*placas elisa*”¹⁴⁷ de 96 pozos, agregar 2ul de *ADN* y/o *producto de PCR*, posteriormente agregar a cada muestra 8ul de “*blue juice*”¹⁴⁸. El *blue juice*, facilita la carga y seguimiento de las muestras de *ADN* en el *gel de agarosa*.
4. Cargar 10 ul de muestra por pozo utilizando una “*micropipeta p-20*”¹⁴⁹ (2ul-20ul). Incluir en gel el “*marcador de peso molecular*”¹⁵⁰, cuyo rol es facilitar la estimación del tamaño molecular de *ácidos nucleicos* y *proteínas*.
5. Conectar el *equipo de electroforesis* a una *fuentes de poder*. Unir el electrodo positivo (rojo) y negativo (negro) a los terminales correspondientes.
6. Prender la *fuentes de poder*, seleccionar el voltaje deseado (aproximadamente 80 voltios). Si el equipo esta funcionando bien, se podrán observar burbujas saliendo de los electrodos. Debe asegurarse que las muestras están corriendo en la dirección correcta, hacia el electrodo positivo.
7. Correr la electroforesis hasta que las muestras hayan recorrido aproximadamente tres cuartos de la *bandeja* que contiene el *gel*.
8. Apagar el equipo.

Los conceptos descritos en esta etapa de la electroforesis (*guantes de latex*, *placas elisa*, *blue juice*, *micropipeta p-20*, *marcador de peso molecular*), son propios de este estudio. En la figura 54, se ilustra a manera de mapa conceptual el proceso de cargar las muestras en un gel de agarosa.

¹⁴⁷ elisa plate: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴⁸ blue juice, agarose gel loading buffer: tomado de Metadata for Laboratory Notebook (M4L)

¹⁴⁹ P-20 micropipette: tomado de Metadata for Laboratory Notebook (M4L)

¹⁵⁰ DNA ladder role: tomado de Metadata for Laboratory Notebook (M4L)

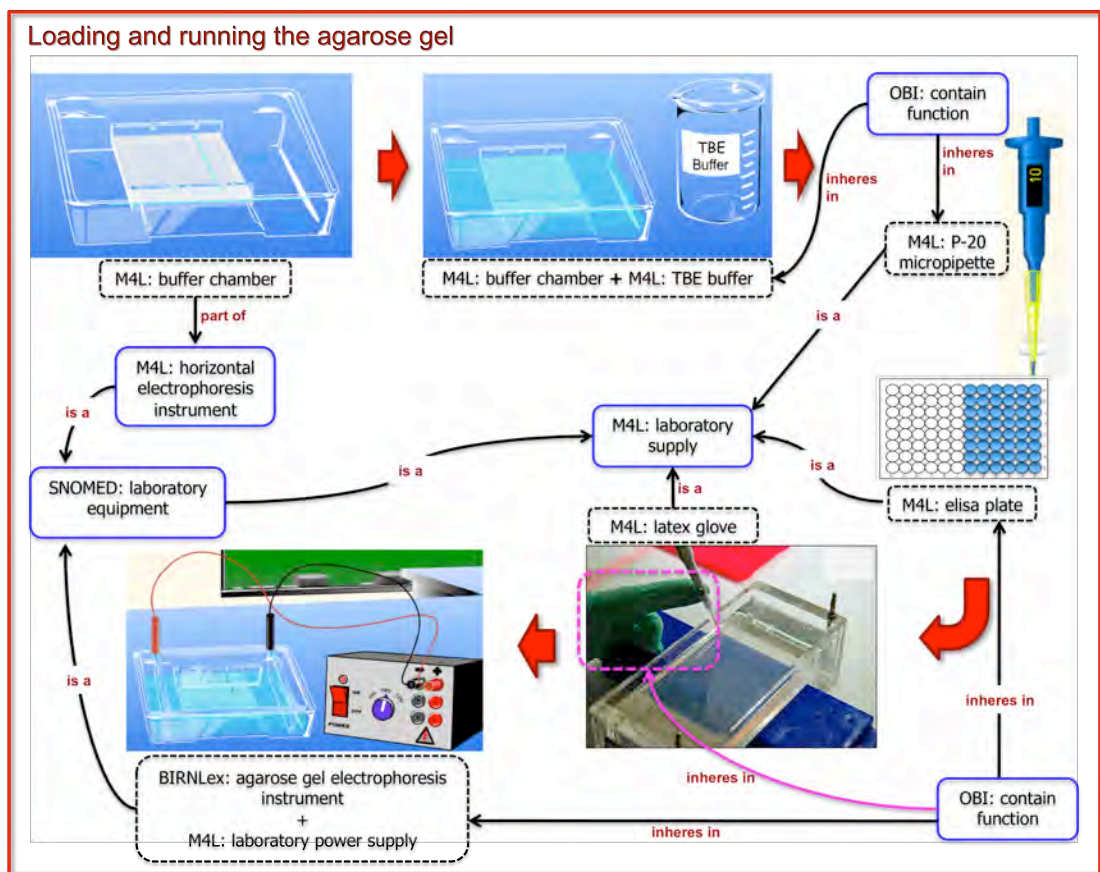


Figura 54. Mapa conceptual del proceso carga y corrida de las muestras en un gel de agarosa.

iii) Visualización de las muestras y análisis de resultados

- Visualizar el *ADN* o *producto de PCR* usando un “*transiluminador*”¹⁵¹.
- Tomar una fotografía del gel usando una “*cámara*”¹⁵² digital o equivalente.
- Finalmente hacer el análisis de resultados con el propósito de determinar el *tamaño molecular* de cada muestra presente en el *gel*.

¹⁵¹ transilluminator: tomado de Metadata for Laboratory Notebook (M4L)

¹⁵² camera: sep: 00099. Tomado de Sample Processing and Separation Techniques (SEP).

En la figura 55 se ilustra a manera de mapa conceptual el proceso de visualización y análisis de las muestras corridas durante una electroforesis.

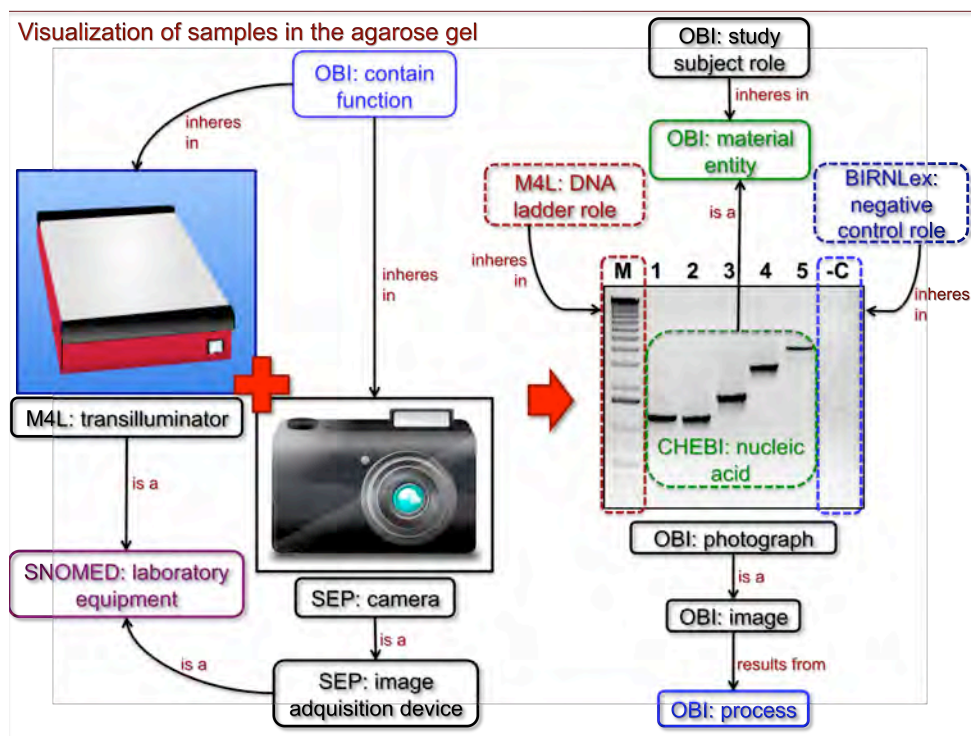


Figura 55. Mapa conceptual del proceso visualización de las muestras en un gel de agarosa y análisis de datos.

4.3.4 Elección de un Lenguaje para Representación de la Ontología M4L. En la figura 56 se presenta el cuarto paso del proceso de desarrollo de M4L. La selección de un lenguaje para el diseño de ontologías.



Figura 56. Selección de un lenguaje para el diseño de ontologías. Cuarto paso ejecutado para la construcción de M4L.

OWL fue el lenguaje usado en este trabajo de tesis para modelar el dominio de conocimiento que se desea representar en la ontología M4L. Las razones que se tuvieron en cuenta para la elección del lenguaje OWL se presentan a continuación:

- OWL permite representar los términos y las relaciones entre ellos de manera que sean fácilmente procesables por aplicaciones informáticas (Martínez-Fernández, 2010).
- Además, OWL es un lenguaje compatible con los actuales estándares utilizados en la Web, por lo tanto actualmente esta siendo muy empleado en el desarrollo de la Web semántica.
- OWL tiene tres sublenguajes: OWL Lite, OWL DL y OWL Full; todos ellos se diferencian en la riqueza de su expresividad, siendo el más simple OWL Lite y el lenguaje de máxima expresividad el OWL Full. En este estudio se usó OWL

DL, ya que por estar basado en lógica descriptiva es el mas usado para el desarrollo de ontologías.

4.3.5 Elección de una Herramienta para el Diseño de la Ontología M4L. En la figura 57, se ilustra el quinto paso llevado a cabo para el desarrollo de M4L, la selección de una herramienta para el diseño de ontologías.



Figura 57. Selección de una herramienta para el diseño de ontologías. Quinto paso del proceso de diseño de M4L.

Para el desarrollo de M4L se escogió Protégé-2000 por ser un sistema de adquisición de conocimiento de código abierto. La aplicación Protégé está escrita en Java y soporta dos formas principales para el modelado de ontologías:

- *Protégé-frames editor*, permite a los usuarios construir y editar ontologías que están basadas en marcos (frames, en ingles), de acuerdo con el protocolo Open Knowledge Base Connectivity (OKBC) (Dameron et al., 2005). La principal ventaja de un editor basado en marcos sobre uno basado en un lenguaje OWL, es que se apoya en el “supuesto del mundo cerrado” (closed world assumption): lo que no aparece explícitamente representado en un estado, es falso (Baader et al., 2003).

- *Protégé-OWL editor*, permite a los usuarios construir ontologías para la Web semántica, particularmente en lenguaje OWL. La ventaja de cualquier editor de ontologías basado en lenguaje OWL, es que tiene una semántica y sintaxis bien definida. Por el contrario, el editor Protégé basado en marcos no tiene una especificación sintáctica ni semántica (Dameron et al., 2005).

La ontología M4L fue desarrollada en un lenguaje OWL. En el anexo D, se presenta de forma detallada el proceso de construcción de la ontología usando la versión 3.4.4 de Protégé. Se espera que esta guía proporcione un punto de partida, una guía inicial que ayude a los nuevos diseñadores de ontologías a desarrollar ontologías de dominio en un lenguaje OWL.

4.4 EVALUACIÓN

Aunque existe gran cantidad de investigaciones sobre metodologías, algoritmos, herramientas de edición y lenguajes de construcción de ontologías en la Ingeniería de Software, un aspecto que no ha sido considerado con suficiente interés es la evaluación. Algunas razones que justifican el escaso desarrollo en esta área son las dificultades para establecer qué elementos evaluar y cuáles criterios considerar para determinar la calidad de las ontologías (Burton-Jones et al., 2005, Brewster et al., 2004).

En este estudio se consideró la actividad propuesta por Gomez-Perez, *et al* (Gomez-Perez et al., 1995) denominada evaluación de la terminología. El objetivo de la evaluación es determinar que define la ontología y la precisión de las definiciones. De acuerdo con esta metodología se deben tener en cuenta los siguientes criterios de evaluación:

- Consistencia: esto es, se asume que una definición es consistente si y solo si, no hay un conocimiento contradictorio que pueda ser inferido de otras definiciones o axiomas en la ontología.
- Completitud: esto es, se asume que las ontologías son en principio incompletas (Gomez-Perez et al., 1995, Gómez-Pérez et al., 2004). Sin embargo, la completitud puede ser evaluada dentro del contexto para lo cual la ontología va a ser usada. Una ontología se considera completa si y solo si, todo lo que se supone esta en la ontología se indica explícitamente, o puede ser inferida.
- Concisión: una ontología es concisa si la misma no almacena conocimiento innecesario, y la redundancia en el conjunto de definiciones ha sido removida apropiadamente.

La manera como se evaluó la ontología M4L, utilizando el esquema de evaluación propuesto fue la siguiente:

- Uso correcto del lenguaje, para evaluar la calidad de la ontología considerando la manera como está escrita. En este estudio, para codificar la ontología M4L se seleccionó el sub-lenguaje de OWL, OWL-DL, el cual permite máxima expresividad sin perder la completitud computacional.
- En cada fase del ciclo de desarrollo se utilizó el marco de chequeo que provee el editor Protégé-OWL. Esta funcionalidad permitió corregir inconsistencias sintácticas.

M4L es en realidad una extensión de la ontología OBI. Por lo tanto los términos y definiciones propuestos en este estudio también están siendo revisados por desarrolladores de OBI, con el propósito de identificación de inconsistencias, completitud de conceptos y existencia de redundancias en clases y relaciones para la posterior inclusión de términos en OBI. Dado que M4L es una extensión de OBI, esta hereda la estructura provista por BFO. En este sentido M4L es

compatible con OBI, tanto en su aspecto sintáctico como semántico. De igual manera, la estructura de metadatos usada en M4L es aquella recomendada por OBI en sus buenas practicas¹⁵³. Las definiciones, al igual que los casos de uso, se construyeron siguiendo los parámetros propuestos por OBI.

4.5 COMENTARIOS FINALES

Con el desarrollo de la bio-ontología M4L por parte de expertos de dominio en biotecnología de plantas, se logró comprender que cualquiera que sea el método empleado, el proceso de desarrollo de una ontología debe reconocerse como un proceso iterativo, compuesto de actividades sucesivas hasta llegar a una versión final de la misma.

Durante el proceso de aprendizaje para el modelamiento de una ontología, se pudo identificar tres puntos importantes a la hora de diseñar ontologías. Esos tres puntos podrían ser de utilidad para quienes desean desarrollar su propia ontología:

- No existe un único modo correcto de modelar una ontología de dominio; al contrario, siempre hay alternativas posibles. La mejor solución depende del propósito u objetivo final de la ontología, o de las aplicaciones en las que vaya a emplearse.
- El desarrollo de una ontología es necesariamente un proceso iterativo, permanece en constante mantenimiento y evolución.
- Los conceptos o clases de una ontología deberán referirse a objetos o entes (tanto físicos como lógicos) existentes en nuestro dominio de interés, al igual

¹⁵³ http://obi-ontology.org/page/OBI_Minimal_metadata

que las relaciones que caracterizan a las clases. Estos conceptos y relaciones podrán encontrarse en diferentes tipos de documentos (artículos, informes, cuadernos de laboratorio, etc.), vocabularios controlados, tesauros u ontologías que describen un dominio de interés.

Otros aspectos o lecciones aprendidas durante la etapa de diseño de una ontología de dominio, se describen a continuación:

Antes de comenzar la ejecución de la ontología, en este estudio era claro el propósito de la ontología y el tipo de conocimiento que se deseaba capturar en ella, pero no era claro como hacerlo. Por lo tanto el desafío en esta fase de la investigación, fue entender como representar semánticamente el conocimiento que se deseaba capturar en la ontología. La elaboración de casos de uso, representados por medio de mapas conceptuales (MC), facilitó esta labor de construcción de la ontología. A través de los MC, se consiguió que la información llegara a los investigadores de una manera sencilla de asimilar y con ello se facilitó la presentación de un conjunto de 359 términos que hacen parte de la primera versión de la ontología M4L.

Los conceptos que hacen parte de la ontología M4L, provienen de distintas fuentes como son: i) vocabularios controlados, ii) cuadernos de laboratorio de investigadores del proyecto Agro Biodiversidad y Biotecnología del CIAT, y de iii) ontologías que hacen parte de BioPortal. La incorporación de términos provenientes de otras ontologías se hizo manualmente, porque aunque existen estudios acerca de técnicas para mezclar ontologías, esta área de investigación aún es muy teórica ya que las ontologías de dominio representan conceptos muy específicos y en ocasiones se presentan problemas de incompatibilidad entre sí.

Esos problemas de incompatibilidad se deben a la diversidad de métodos empleados para el diseño de ontologías. La falta de métodos y normas estándar para la definición de una ontología tanto al interior de los equipos de trabajo, como entre ellos, obstaculiza el desarrollo de ontologías compartidas y

consensuadas, la extensión de ontologías desarrolladas por otros grupos y su reúso en otras ontologías y en aplicaciones finales (Soldatova and King, 2005, Bard and Rhee, 2004).

Debido a la amplia gama de posibilidades encontradas en la literatura para modelar una ontología, es importante precisar que en el mundo de las ontologías no hay verdades absolutas, ya que las ontologías son creadas artificialmente, no se descubren, por lo tanto solo especifican una forma de ver el mundo. Por lo tanto, es importante que una ontología sea diseñada con propósitos específicos para solucionar necesidades concretas.

4.6 BIBLIOGRAFIA

- ABU AL-SOUD, W. & RADSTROM, P. (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-Inhibiting samples. *Applied and Environmental Microbiology*, 64, 3748-3753.
- ADAM, M. (2004) Folksonomies - Cooperative Classification and Communication Through Shared Metadata. <http://www.adammathes.com>.
- ALAVI, M. & LEIDNER, D. E. (2001) *Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues*, Management Information Systems Research Center, University of Minnesota.
- ALUKO, G., MARTINEZ, C., TOHME, J., CASTANO, C., BERGMAN, C. & OARD, J. H. (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. *TAG Theoretical and Applied Genetics*, 109, 630-639.
- ANGELETOU, S. (2008) Semantic Enrichment of Folksonomy Tagspaces. *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.
- ANGUITA, A., CALLE, G., SILVA, A., MAOJO, V., WEILER, G., BROCHHAUSEN, M., KUMAR, A., DURAND, P., KONDYLAKIS, H. & DOERR, M. (2007)

Consolidated Requirements on ontological approaches for integration of multi-level biomedical information.

- ARANGUREN, M. E., ANTEZANA, E., KUIPER, M. & STEVENS, R. (2008) Ontology Design Patterns for bio-ontologies: a case study on the Cell Cycle Ontology. *BMC Bioinformatics*, 9, -.
- ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25, 25-9.
- BAADER, F., CALVANESE, D., MCGUINNESS, D., NARDI, D. & PATEL-SCHNEIDER, P. (2003) *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press
- BARD, J. B. L. & RHEE, S. Y. (2004) Ontologies in biology: Design, applications and future challenges. *Nature Reviews Genetics*, 5, 213-222.
- BAUMARD, P. & WAUCHOPE, S. (1999) ***Tacit knowledge in organizations***, London, SAGE.
- BEATTY, P., DICK, S. & MILLER, J. (2008) Is HTML in a race to the bottom? A large-scale survey and analysis of conformance to W3C standards. *Ieee Internet Computing*, 12, 76-80.
- BEISSWANGER, E., LEE, V., KIM, J. J., REBHOLZ-SCHUHMANN, D., SPLENDIANI, A., DAMERON, O., SCHULZ, S. & HAHN, U. (2008) Gene Regulation Ontology (GRO): design principles and use cases. *Stud Health Technol Inform*, 136, 9-14.
- BEISSWANGER, E., SCHULZ, S., STENZHORN, H. & HAHN, U. (2007) BIOTOP: An Upper Domain Ontology for the Life Sciences. *Applied Ontology*, 7.
- BELTRÁN, J., JAIMES, H., ECHEVERRY, M., LADINO, Y., LÓPEZ, D., DUQUE, M. C., CHAVARRIAGA, P. & TOHME, J. (2008) Quantitative analysis of transgenes in cassava plants using real-time PCR technology. *In Vitro Cellular & Developmental Biology - Plant*, 45, 48-56.
- BLAIR, M. W., FREGENE, M. A., BEEBE, S. E. & CEBALLOS, H. (2007) marker-assisted selection in common beans and cassava. IN GUIMARÃES, E. P., RUANE, J., SCHERF, B. D., SONNINO, A. & DARGIE, J. D. (Eds.) ***MARKER-ASSISTED SELECTION, Current status and future perspectives in crops, livestock, forestry and fish***.

- BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., SHERLOCK, G., SPELLMAN, P., STOECKERT, C., AACH, J., ANSORGE, W., BALL, C. A., CAUSTON, H. C., GAASTERLAND, T., GLENNISSON, P., HOLSTEGE, F. C. P., KIM, I. F., MARKOWITZ, V., MATESE, J. C., PARKINSON, H., ROBINSON, A., SARKANS, U., SCHULZE-KREMER, S., STEWART, J., TAYLOR, R., VILO, J. & VINGRONS, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *nature genetics* 29, 365-371.
- BREWSTER, C., ALANI, H., DASMAHAPATRA, S. & WILKS, Y. (2004) Data Driven Ontology Evaluation. *International Conference on Language Resources and Evaluation*. Lisbon, Portugal.
- BRILHANTE, V., MACEDO, G. T. & MACEDO, S. (2006) Heuristic Transformation of Well-Constructed Conceptual Maps into OWL Preliminary Domain Ontologies *Workshop on Ontologies and their Applications, WONTO*. Brazil.
- BRINKMAN, R. R., COURTOT, M., DEROM, D., FOSTEL, J. M., HE, Y., LORD, P., MALONE, J., PARKINSON, H., PETERS, B., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SOLDATOVA, L. N., STOECKERT JR., C. J., TURNER, J., ZHENG, J. & CONSORTIUM., T. O. (2010) Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*, 1, 11.
- BUG, W. J., ASCOLI, G. A., GRETHE, J. S., GUPTA, A., FENNEMA-NOTESTINE, C., LAIRD, A. R., LARSON, S. D., RUBIN, D., SHEPHERD, G. M., TURNER, J. A. & MARTONE, M. E. (2008) The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics*, 6, 175-94.
- BURRELL, M. M. (1994) RNase A (EC 3.1.27.5). IN BURRELL, M. M. (Ed.) *Enzymes of Molecular Biology*. New Jersey, Humana Press Inc.
- BURTON-JONES, A., STOREY, V. C., SUGUMARAN, V. & AHLUWALIA, P. (2005) A semiotic metrics suite for assessing the quality of ontologies. *Data & Knowledge Engineering*, 55, 84-102.
- BUTLER, D. (2005) Electronic notebooks: A new leaf. *Nature*, 436, 20-21.
- CANTARA, L. (2006) Encoding controlled vocabularies for the Semantic Web using SKOS Core. *OCLC Systems & Services: International digital library perspectives*, 22, 111-114.
- CASTRO, A. G., ROCCA-SERRA, P., STEVENS, R., TAYLOR, C., NASHAR, K., RAGAN, M. A. & SANSONE, S. A. (2006) The use of concept maps during

- knowledge elicitation in ontology development processes - the nutrigenomics use case. *BMC Bioinformatics*, 7, -.
- CEUSTERS, W., SMITH, B. & GOLDBERG, L. (2005) A terminological and ontological analysis of the NCI thesaurus. *Methods of Information in Medicine*, 44, 498-507.
- CHARTIER, J. (2005) Oragene™ and DNA sequencing with ABI PRISM®. Ottawa, ON., DNA Genotek Inc.
- CIAPINA, L. P., ALVES, L. M. C. & LEMOS, E. G. M. (2004) A nested-PCR assay for detection of *Xylella fastidiosa* in citrus plants and sharpshooter leafhoppers. *Journal of Applied Microbiology*, 96, 546-551.
- COURTOT, M., BUG, W., GIBSON, F., LISTER, A. L., MALONE, J., SCHOBBER, D., RYAN BRINKMAN, R. & RUTTENBERG, A. (2008) The OWL of Biomedical Investigations. *OWLED* 12.
- CUEVAS-SIMON, A., CECCARONI, L. & SUAREZ-ROSETE, A. (2008) An approach to formal modeling of environmental knowledge via concept maps and ontologies. *International Congress on Environmental Modelling and Software, iEMSs*.
- DAMERON, O., RUBIN, D. L. & MUSEN, M. A. (2005) Challenges in Converting Frame-Based Ontology into OWL: the Foundational Model of Anatomy Case-Study. *AMIA*, 2005.
- DASGUPTA, A. & RATH, D. S. (2009) Organizing Knowledge for Web Retrieval using SKOS: A Case Study in Human Protein Chain. *ICAL-Management Models and Framework*.
- DAVALIEVA, K. G. & EFREMOV, G. D. (2010) A new thermostable DNA polymerase mixture for efficient amplification of long DNA fragments. *Prikl Biokhim Mikrobiol*, 46, 248-52.
- DAVENPORT, T. H. & PRUSAK, L. (1998) *Working Knowledge: How Organizations Manage What They Know* Harvard Business School Press
- DE NICOLA, A., MISSIKOFF, M. & NAVIGLI, R. (2009) A software engineering approach to ontology building. *Information Systems*, 34, 258-275.
- DEGTYARENKO, K., DE MATOS, P., ENNIS, M., HASTINGS, J., ZBINDEN, M., MCNAUGHT, A., ALCANTARA, R., DARSOW, M., GUEDJ, M. & ASHBURNER, M. (2008) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36, D344-D350.

- DELLAPORTA, S. L., WOOD, J. & HICKS, J. B. (1983) A plant DNA miniprep: Version II *Plant Molecular Biology Reporter*, 1, 19-21.
- ELLIOTT, M. (2006) Electronic Study Management. New Tools for Improving the Efficiency of Pre-Clinical R&D. *Scientific Computing*, 7.
- ERCEGOVAC, Z. (1999) Special topic issue: Integrating multiple overlapping metadata standards - Introduction. *Journal of the American Society for Information Science*, 50, 1165-1168.
- FERNÁNDEZ, M., GOMEZ-PEREZ, A. & JURISTO, N. (1997) METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Stanford.
- GACHON, C., MINGAM, A. & CHARRIER, B. (2004) Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany*, 55, 1445-1454.
- GANGEMI, A., PISANELLI, D. M. & STEVE, G. (1998) Ontology Integration: Experiences with Medical Terminologies. *N. Guarino*, 163-178.
- GARCIA, A., O'NEILL, K., GARCIA, L. J., LORD, P., STEVENS, R., CORCHO, O. & GIBSON, F. (2010) Developing Ontologies within Decentralised Settings. IN CHEN, H. (Ed.) *Semantic e-Science*, In Press. Springer Science+Business Media.
- GARCIA-CASTRO, A. (2007) DEVELOPING ONTOLOGIES IN THE BIOLOGICAL DOMAIN *Institute for Molecular Bioscience*. Queensland, University of Queensland.
- GÓMEZ-PÉREZ, A., FERNÁNDEZ-LÓPEZ, M. & CORCHO, O. (2004) *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*, Springer.
- GOMEZ-PEREZ, A., JURISTO, N. & PAZOS, J. (1995) Evaluation and assessment of the Knowledge Sharing Technology. *Towards Very Large Knowledge Bases*, 289-296.
- GOOD, B., TRANFIELD, E. M., TAN, P. C., SHEHATA, M., SINGHERA, G. K., GOSSELINK, J., OKON, E. B. & WILKINSON, M. (2006) Fast, Cheap, and Out of Control: A Zero Curation Model for Ontology Development. *Pacific Symposium on Biocomputing*.
- GRAY, A. J. G., GRAY, N. & OUNIS, I. (2008) Finding Data Resources in a Virtual Observatory Using SKOS Vocabularies *Sharing Data, Information and Knowledge*. Springer Berlin / Heidelberg.

- GRENON, P., SMITH, B. & GOLDBERG, L. (2004) Biodynamic ontology: Applying BFO in the biomedical domain. *Ontologies in Medicine*, 102, 20-38.
- GRUBER, T. (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220.
- GRUBER, T. (2007) Collective Knowledge Systems. *Journal of Web Semantics*.
- GRUBER, T. R. (1995) Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43, 907-928.
- GRUNINGER, M. & FOX, M. S. (1995) Methodology for the Design and Evaluation of Ontologies. *Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal.
- GUTIÉRREZ, A. G., CARABALÍ, S. J., GIRALDO, O. X., MARTÍNEZ, C. P., CORREA, F., PRADO, G., TOHME, J. & LORIEUX, M. (2010) Identification of a Rice stripe necrosis virus resistance locus and yield component QTLs using *Oryza sativa* × *O. glaberrima* introgression lines. *BMC Plant Biology*, 10, 11.
- HAQUIN, S., OEUILLET, E., PAJON, A., HARRIS, M., JONES, A., TILBEURGH, H., MARKLEY, J., ZOLNAI, Z. & POUPON, A. (2008) *Data management in structural genomics: an overview.*, Humana Press.
- HARTMANN, J., PALMA, R., SURE, Y., SUAREZ-FIGUEROA, M. C., HAASE, P., GOMEZ-PEREZ, A. & STUDER, R. (2005) Ontology metadata vocabulary and applications. *On the Move to Meaningful Internet Systems 2005: Otm 2005 Workshops, Proceedings*, 3762, 906-915.
- HEERY, R. (1996) Review of metadata formats. *Program: electronic library and information systems*, 30, 345-373.
- HELFRICH, J. P. (2004) Knowledge management systems: coping with the data explosion. *DDT: BIOSILICO*, 2(1).
- HERRE, H., HELLER, B., BUREK, P., HOEHNDORF, R., LOEBE, F. & MICHALEK, H. (2006) General Formal Ontology (GFO), Part I: Basic Principles, Version 1.0. Leipzig, Institute of Medical Informatics, Statistics and Epidemiology (IMISE).
- HEYMANN, P. & GARCIA-MOLINA, H. (2008) Can Tagging Organize Human Knowledge?, Infolab Technical Report. CA, USA, Stanford University.

- HILLIARD, M. S., LARSON, D. L. & ROSENBERG, M. J. (2001) LIMS: A Suite of Database Tools for Laboratory Organization. *Computer-Based Medical Systems, 2001. CBMS 2001. Proceedings. 14th IEEE Symposium on*. Bethesda, MD, USA, <http://ieeexplore.ieee.org>.
- HUNTER, A., KAUFMAN, M. H., MCKAY, A., BALDOCK, R., SIMMEN, M. W. & BARD, J. B. L. (2003) An ontology of human developmental anatomy. *Journal of Anatomy*, 203, 347-355.
- JALSWAL, P., AVRAHAM, S., ILIC, K., KELLOGG, E. A., MCCOUCH, S., PUJAR, A., REISER, L., RHEE, S. Y., SACHS, M. M., SCHAEFFER, M., STEIN, L., STEVENS, P., VINCENT, L., WARE, D. & ZAPATA, F. (2005) Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages. *Comparative and Functional Genomics*, 6, 388-397.
- JAYASHREE, B., REDDY, T., LEELADEVI, Y., CROUCH, J., MAHALAKSHMI, V., BUHARIWALLA, H., ESHWAR, K., MACE, E., FOLKSTERMA, R., SENTHILVEL, S., VARSHNEY, R., SEETHA, K., RAJALAKSHMI, R., PRASANTH, V., CHANDRA, S., SWARUPA, L., SRIKALYANI, P. & HOISINGTON, D. (2006) Laboratory Information Management Software for genotyping workflows: applications in high throughput crop genotyping. *BMC Bioinformatics*, 7, 383.
- KAKALI, C., LOURDI, I., STASINOPOULOU, T., BOUNTOURI, L., PAPTAEODOROU, C., DOERR, M. & GERGATSOULIS, M. (2007) Integrating Dublin Core metadata for cultural heritage collections using ontologies. *Dublin Core and Metadata Applications*.
- KANARE, H. M. (1985) *Writing the Laboratory Notebook*, Washington D.C, American Chemical Society.
- KAWATA, M., MATSUMURA, Y., OIKAWA, T., KIMIZU, M., FUKUMOTO, F. & KURODA, S. (2003) Analysis of DNA extraction buffer components from plant tissue by polymerase chain reaction *Analytical Biochemistry*, 318, 314-317.
- KIMURA, F., TOBA, T., TEZUKA, T. & MAEDA, A. (2009) Federated Searching System for Humanities Databases Using Automatic Metadata Mapping. *Dublin Core and Metadata Applications*
- KINZLER, K. W. & VOGELSTEIN, B. (1989) Whole genome PCR: application to the identification of sequences bound by gene regulatory protein. *Nucleic Acids Research*, 17, 3645-3653.

- KÖHL, K., BASLER, G., LÜDEMANN, A., SELBIG, J. & WALTHER, D. (2008) A plant resource and experiment management system based on the Golm Plant Database as a basic tool for omics research. *Plant Methods*, 4, 1-11.
- KWOH, D. Y., DAVIS, G. R., WHITFIELD, K. M., CHAPPELLE, H. L., DIMICHELE, L. J. & GINGERAS, T. R. (1989) Transcription-Based Amplification System and Detection of Amplified Human Immunodeficiency Virus Type-1 with a Bead-Based Sandwich Hybridization Format. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 1173-1177.
- LAMBRIX, P., HABBOUCHE, M. & PEREZ, M. (2003) Evaluation of ontology development tools for bioinformatics. *Bioinformatics*, 19, 1564-1571.
- LIN, Y., TARUI, H. & SIMONS, P. (2010) From Ontology for Genetic Interval (OGI) to Sequence Assembly – Ontology Applying to Next Generation Sequencing. *Semantic Web Applications and Tools for Life Sciences (SWAT4LS)*.
- LOPEZ, C., SOTO, M., RESTREPO, S., PIEGU, B., COOKE, R., DELSENY, M., TOHME, J. & VERDIER, V. (2005) Gene expression profile in response to *Xanthomonas axonopodis* pv. *manihotis* infection in cassava using a cDNA microarray. *Plant Molecular Biology*, 57, 393-410.
- LÓPEZ, C. E., QUESADA-OCAMPO, L. M., BOHÓRQUEZ, A., DUQUE, M. C., VARGAS, J., TOHME, J. & VERDIER, V. (2007) Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. *Genome*, 50(12), 1078-1088.
- LYSAKOWSKI, R. (1997) Comparing Paper and Electronic Laboratory Notebooks. <http://www.censa.org>.
- MAHESH, K. (1996) Ontology Development for Machine Translation: Ideology and Methodology, MCCS-96-292. Computing Research Laboratory, New Mexico State University.
- MAINZ, D., PAULSEN, I., MAINZ, I., WELLER, K., KOHL, J. & VON HAESELER, A. (2008) Knowledge acquisition focused cooperative development of bio-ontologies - A case study with BIO2Me. *Bioinformatics Research and Development, Proceedings*, 13, 258-272.
- MARTINEZ-FERNANDEZ, J. L. (2010) Introduciendo la semántica en un proceso de desarrollo software a través de reglas de negocio *Escuela técnica superior de ingenieros de telecomunicación*. Madrid, Universidad Politécnica de Madrid.

- MARZIALI, A. (2001) Microcentrifuge. The Board of Trustees of the Leland Stanford Junior University, Palo Alto CA.
- MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., OLTRAMARI, A. & SCHNEIDER, L. (2003) WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies, Preliminary Report Padova, National Research Council, Institute of Cognitive Sciences and Technology.
- MATUSZEK, C., CABRAL, J., WITBROCK, M. & DEOLIVEIRA, J. (2006) An Introduction to the Syntax and Content of Cyc. *AAAI Spring Symposium, 2006*.
- MCARTHUR, A. G., MORRISON, H. G., NIXON, J. E., PASSAMANECK, N. Q., KIM, U., HINKLE, G., CROCKER, M. K., HOLDER, M. E., FARR, R., REICH, C. I., OLSEN, G. E., ALEY, S. B., ADAM, R. D., GILLIN, F. D. & SOGIN, M. L. (2000) The Giardia genome project database. *FEMS Microbiol Lett*, 189, 271-3.
- MIZOGUCHI, R., KOZAKI, K., SANO, T. & KITAMURA, Y. (2000) Construction and deployment of a plant ontology. *Knowledge Engineering and Knowledge Management, Proceedings, 1937*, 113-128.
- MOERCHEN, F., FRADKIN, D., DEJORI, M. & WACHMANN, B. (2008) Emerging trend prediction in biomedical literature. *AMIA Annu Symp Proc*, 485-9.
- MUELLER, J. D., PÜTZ, B. & HÖFLER, H. (1997) Self-sustained sequence replication (3SR): an alternative to PCR. *Histochemistry and Cell Biology*, 431-437.
- MULLIS, K. (1990a) The unusual origin of the polymerase chain reaction. *Scientific American*, 262(4), 56-61, 64-5.
- MULLIS, K. B. (1990b) The unusual origin of the polymerase chain reaction. *Sci Am*, 262, 56-61, 64-5.
- MURPHY, L. D. (1998) Digital Document Metadata in Organizations: Roles, Analytical Approaches, and Future Research Directions. *31st Hawaii International Conference on System Sciences*
- MYERS, J., FOX-DOBBS, C., LAIRD, J., LE, D., REICH, D. & CURTZ, T. (1996) Electronic Laboratory Notebooks for Collaborative Research. *IEEE Fifth Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*. Stanford University, California, <http://collaboratory.emsl.pnl.gov>.

- NATALE, D. A., ARIGHI, C. N., BARKER, W. C., BLAKE, J., CHANG, T. C., HU, Z. Z., LIU, H. F., SMITH, B. & WU, C. H. (2007) Framework for a Protein Ontology. *BMC Bioinformatics*, 8, -.
- NATIONAL AGRICULTURAL LIBRARY (NAL). & AGRICULTURAL RESEARCH SERVICE, U. S. D. O. A. (2010a) Glossary of Agricultural Terms. 2010 ed.
- NATIONAL AGRICULTURAL LIBRARY (NAL). & AGRICULTURAL RESEARCH SERVICE, U. S. D. O. A. (2010b) NAL Agricultural Thesaurus 2010 ed.
- NAUMAN, M., KHAN, S., AMIN, M. & HUSSAIN, F. (2008) Resolving Lexical Ambiguities in Folksonomy Based Search Systems through Common Sense and Personalization. *SemSearch 2008*. CEUR-WS.org/Vol-334/.
- NORRIS, C. (2008) Some Versions of Platonism: Mathematics and Ontology According to Badiou. *Philosophical frontiers*, 3, 28.
- NOVAK, J. D. & CAÑAS, A. J. (2008) The Theory Underlying Concept Maps and How to Construct and Use Them, Technical Report IHMC CmapTools 2006-01 Rev 01-2008. Florida, Institute for Human and Machine Cognition.
- OBI OBI Minimal metadata. science commons.
- ORGANIZATION, N. I. S. (2005) ANSI/NISO Z39.19 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. *National Information Standards Organization*. NISO Press.
- PAGANELLI, F., PETTENATI, M. C. & GIULI, D. (2006) A Metadata-Based Approach for Unstructured Document Management in Organizations. *Information Resources Management Journal*, 19, 22.
- PÄIVÄRINTA, T., TYRVÄINEN, P. & YLIMÄKI, T. (2002) Defining organizational document metadata: A case beyond standards. *Xth European Conference on Information Systems (ECIS)*. Gdańsk, Poland.
- PAJON, A., IONIDES, J., DIPROSE, J., FILLON, J., FOGH, R., ASHTON, A., BERMAN, H., BOUCHER, W., CYGLER, M., DELEURY, E., ESNOUF, R., JANNIN, J., KIM, R., KRIMM, I., LAWSON, C., OEUILLET, E., POUPON, A., RAYMOND, S., STEVENS, T., TILBEURGH, H., WESTBROOK, J., WOOD, P., ULRICH, E., VRANKEN, W., XUELI, L., LAUE, E., STUART, D. & HENRICK, K. (2005) Design of a Data Model for Developing Laboratory Information Management and Analysis Systems for Protein Production. *PROTEINS: Structure, Function, and Bioinformatics*, 58, 278-284.
- PASZKO, C., TURNER, E. & HINTON, M. (2002) *Laboratory Information Management Systems*, CRC Press.

- PEASE, A. & NILES, I. (2002) IEEE standard upper ontology: a progress report. *Knowledge Engineering Review*, 17, 65-70.
- PINTO, H. S., STAAB, S. & TEMPICH, C. (2004) DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of ontologies. *European Conference on Artificial Intelligence*. Valencia, Spain.
- POLONSKY, A., SIX ADRIEN., KOTELNIKOV MIKHAIL., POLONSKY VADIM., POLLY RENAUD. & PAUL., B. Semantic Laboratory Notebook. <http://knol.google.com>.
- REBBECK, J. (2005) What Makes a Great Science Lab Notebook? , <http://www.sciencebuddies.org>.
- ROSSE, C. & MEJINO, J. L., JR. (2003) A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*, 36, 478-500.
- SAHOO, S. S., THOMAS, C., SHETH, A., YORK, W. S. & TARTIR, S. (2006) Knowledge Modeling and its Application in Life Sciences: A Tale of two Ontologies. *World Wide Web Conference*. Edinburgh, Scotland.
- SANDARS, J. & HELLER, R. (2006) Improving the implementation of evidence-based practice: a knowledge management perspective. *Journal of Evaluation in Clinical Practice*, 12, 341-346.
- SARINI, M., BLANZIERI, E., GIORGINI, P. & MOSER, C. (2004) From actions to suggestions: supporting the work of biologists through laboratory notebooks. *Proceedings of 6th International Conference on the Design of Cooperative Systems (COOP2004)*. French Riviera, France, IOSPress.
- SCHOBBER, D., SMITH, B., LEWIS, S. E., KUSNIERCZYK, W., LOMAX, J., MUNGALL, C., TAYLOR, C. F., ROCCA-SERRA, P. & SANSONE, S. A. (2009) Survey-based naming conventions for use in OBO Foundry ontology development. *BMC Bioinformatics*, 10, -.
- SEOANE, C. (2007) La Web 2.0 y sus implicaciones en el mundo de la gestión de la información, Actas del VI Workshop CALSI. *Cal/si*.
- SHIRKY, C. (2005) Ontology is Overrated: Links, Tags and Post-hoc Metadata. *EMERGING TECHNOLOGY CONFERENCE*. San Diego, CA, <http://conferences.oreillynet.com>.
- SIMON, A., CECCARONI, L. & ROSETE, A. (2007) Generation of OWL Ontologies from Concept Maps in Shallow Domains. *Conference of the*

Spanish Association for Artificial Intelligence, CAEPIA. Salamanca, Spain, Springer-Verlag.

SMITH, B., ASHBURNER, M., ROSSE, C., BARD, J., BUG, W., CEUSTERS, W., GOLDBERG, L. J., EILBECK, K., IRELAND, A., MUNGALL, C. J., THE OBI CONSORTIUM, LEONTIS, N., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SCHEUERMANN, R. H., SHAH, N., WHETZEL, P. L. & LEWIS, S. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11), 1251 - 1255.

SMITH, B., CEUSTERS, W., KLAGGES, B., KOHLER, J., KUMAR, A., LOMAX, J., MUNGALL, C., NEUHAUS, F., RECTOR, A. L. & ROSSE, C. (2005) Relations in biomedical ontologies. *Genome Biology*, 6, -.

SMITH, B., WILLIAMS, J. & SCHULZE-KREMER, S. (2003) The ontology of the gene ontology. *AMIA Annual Symposium Proceedings*

SOCORRO, R., SIMÓN, A., VALDÉS, R., FERNÁNDEZ, F. O., ROSETE, A., MORENO, M., LEYVA, E. & PINA, J. Las ontologías en la representación del conocimiento. Centro de Estudios de Ingeniería de Sistemas (CEIS), Instituto Superior Politécnico “José Antonio Echeverría” (CUJAE).

SOLDATOVA, L. N. & KING, R. D. (2005) Are the current ontologies in biology good ontologies? *Nature Biotechnology*, 23, 1095-1098.

SPARROW, J. (1998) *Knowledge in Organizations: Access to Thinking at Work*, SAGE Publications.

SPITERI, L. F. (2007) Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*.

STEINACKER, A., GHAVAM, A. & STEINMETZ, R. (2001) Metadata Standards for Web-Based Resources. IN PEIYA LIU, S. C. R. (Ed.) *IEEE MultiMedia*.

STEVE, G., GANGEMI, A. & PISANELLI, D. M. (1998) Integrating Medical Terminologies with ONIONS Methodology.

STOECKERT JR, C. J. & PARKINSON, H. (2003) The MGED ontology: a framework for describing functional genomics experiments. *Comparative and Functional Genomics*, 4, 127-132.

UREN, V., CIMIANO, P., IRIA, J., HANDSCHUH, S., VARGAS-VERA, M., MOTTA, E. & CIRAVEGNA, F. (2005) Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1), 15.

- USCHOLD, M. & GRUNINGER, M. (1996) Ontologies: Principles, methods and applications. Knowledge Engineering Review
- VANDER-WAL, T. (2007) Folksonomy. vanderwal.net.
- VOELKERDING, K., DAMES, S. & DURTSCHI, J. (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, 55(4), 641-58.
- WANG, H., QI, M. Q. & CUTLER, A. J. (1993) A Simple Method of Preparing Plant-Samples for Pcr. *Nucleic Acids Research*, 21, 4153-4154.
- WASHINGTON, N. & LEWIS, S. (2008) Ontologies: Scientific Data Sharing Made Easy. *Scitable by nature education*.
- WEIBEL, S. (2000) The Dublin core metadata initiative - The Frankfurt focus and the year 2000. *Zeitschrift Fur Bibliothekswesen Und Bibliographie*, 47, 3-13.
- WEIGAND, H. & HOPPENBROUWERS. (1997) Experiences with a Multilingual Ontology-based Lexicon for News Filtering, the TREVI project. Infolab.
- WEIR, A. F. (1994) Nucleases. IN BURRELL, M. M. (Ed.) *Enzymes of Molecular Miology*. New Jersey, Humana Press Inc.
- WIEDMANN, M., WILSON, W., BARANY, F. & BATT, C. (2009) Ligase chain reaction (LCR)--overview and applications. *Genome Research*, 1994. 3, S51-S64.
- WITHERS, P. C. (2010) Reclosable plastic bags Australia, International Consolidated Bussiness Pty Ltd., Windsor (AU).
- WOODS, E. A. & WORSHAM, G. (1989) Quick- Chill Ice Chest. United States of America.
- WU, D. Y. & WALLACE, R. B. (1989) The Ligation Amplification Reaction (Lar) - Amplification of Specific DNA-Sequences Using Sequential Rounds of Template-Dependent Ligation. *Genomics*, 4, 560-569.
- ZHANG, Z., CHEUNG, K.-H. & TOWNSEND, J. (2008) Bringing Web 2.0 to bioinformatics. *BRIEFINGS IN BIOINFORMATICS*, 10(1), 1-10.

5. ANOTACIONES LIBRES + METADATOS + ONTOLOGIAS = TAGS4LABS

5.1 INTRODUCCIÓN

En el campo de las ciencias biomédicas, los laboratorios generan cantidades de datos a una velocidad cada vez mayor gracias a la disponibilidad de equipos de alta tecnología, capaces de procesar un gran número de muestras en el mismo periodo de tiempo que hace unas décadas hubiera parecido utópico. De igual manera, a través de la Web, el investigador tiene la posibilidad de acceder a todo tipo de información disponible en diferentes tipos de recursos que cada vez son mas interactivos y sociales.

El problema al que se enfrentan los laboratorios hoy en día no es la falta de acceso tecnológico, sino el tiempo que le toma al investigador en generar conocimiento –relacionar información. Ese tiempo dependerá de la cantidad de datos experimentales que deba organizar, clasificar y analizar para finalmente compartirlos con otros miembros de la comunidad científica. De igual manera el tiempo invertido en asimilar los volúmenes enormes de información almacenada tanto en medios electrónicos como en papel también influye en la generación de conocimiento.

El mantenimiento de información esparcida en diferentes medios, tanto físicos como electrónicos, hace cada vez mas difícil su consulta y recuperación. Lo anterior deja claro que parte del problema tiene que ver con el manejo de conocimiento (Knowledge Management).

A través del uso de conceptos que emergen del Web Semántico, anotaciones semánticas, constituyen un tejido para recuperar, clasificar y compartir el conocimiento. En este orden de ideas, la aproximación aquí presentada no solo usa técnicas y principios del Web Semántico y de folksonomías; esta aproximación se sustenta técnicamente en la orquestación de varias ontologías y contribuciones generadas por usuarios dentro de una estructura coherente.

Este capítulo presenta como dicha orquestación se hace realidad. A través de ejemplos extraídos de casos reales se ilustrará la estructura sobre la cual varias ontologías coexisten con anotaciones generadas por usuarios.

El presente capítulo está organizado como sigue: en la sección 4.2 se presenta la motivación para la construcción del prototipo Tags4Labs, en la sección 4.3 se presenta una estructura retórica de los cuadernos de laboratorio, producto del análisis de este tipo de documento hecho en este estudio, en la sección 4.4 se explica el medio usado para añadir semántica a las anotaciones libres, la Annotation Ontology (AO), en la sección 4.5 se muestra como fue pensado el diseño de Tags4labs, a continuación, en la sección 4.6 se presentan las funciones con que cuenta el prototipo actualmente, la sección 4.7 presenta la evaluación del prototipo y finalmente en la sección 4.8 se presentan los comentarios finales.

5.2 MOTIVACIÓN PARA LA CONSTRUCCIÓN DE TAGS4LABS

Actualmente existen muchas aplicaciones informáticas que tienen como objetivo común el almacenar, consultar y recuperar información biomédica (Rubacha et al., 2011, Topel et al., 2010, Stead et al., 2008, Marzolf and Troisch, 2006, Anderson et al., 2007).

Sin embargo las mismas no son ampliamente usadas por factores como:

1. Limitaciones en la captura fácil y rápida de información generada de manera espontánea a partir de una lluvia de ideas durante una reunión.
2. Baja difusión de la información, la cual se limita al uso de un correo electrónico o a través de conversaciones informales durante una pausa de actividades laborales. Este proceso de divulgación de la información no es eficiente debido a diferentes razones que se exponen a continuación:
 - Cuando se quiere difundir de manera abierta cierto tipo de información, en ocasiones no se involucra personal experto de dominio que podría tener conocimiento adicional para ofrecer.
 - Estas formas de divulgación de información toman tiempo y en ocasiones solo se logra difundir conocimiento de manera local.
 - Cuando personas involucradas en un proceso dejan el proyecto o la compañía, la información podría perderse debido a que el nuevo personal no tendría acceso a la información intercambiada previamente por no haber estado involucrado con anterioridad.
3. Incapacidad en la toma de datos en diferentes localidades (e.j. laboratorios, invernaderos, campos de cultivos, salas de conferencias, salas de reuniones, etc.).

Con el propósito de resolver estos problemas, en este estudio se plantea el diseño de un prototipo que permita una entrada rápida y flexible de la información, mejore la visibilidad y movilidad de la misma y los datos puedan ser capturados en cualquier sitio geográfico mediante el uso de dispositivos móviles como las tabletas. Dicho prototipo, llamado Tags4Labs (etiquetas para laboratorios), combina elementos del Web Social, como son las folksonomías, para facilitar la captura e intercambio de información; y elementos del Web Semántico, como las ontologías, para soportar la organización, clasificación y recuperación de la información allí consignada.

5.3 ESTRUCTURA RETÓRICA DE LOS CUADERNOS DE LABORATORIO

La teoría sobre la estructura retórica ofrece una explicación de la coherencia del documento. Entendiéndose coherencia como la ausencia de secuencias ilógicas y de lagunas. En otras palabras, cada parte de un documento coherente tiene una función –una razón verosímil o aceptable para su existencia- evidente a los lectores, y, además, produce la impresión de que no le falta nada (Green, 2010, Peng, 2009, Uzeda et al., 2008, Taboada and Mann, 2006a, Taboada and Mann, 2006b).

Con base en el análisis de los cuadernos de laboratorio y las ontologías desarrolladas; en este estudio se presenta una plantilla básica describiendo la estructura retórica de los cuadernos de laboratorio. La plantilla incluye un encabezado y un cuerpo. El encabezado está representado por 7 metadatos que auto-describen un cuaderno de laboratorio sin importar su contenido. El cuerpo, en primer lugar, contiene 10 metadatos que auto-describen una actividad experimental cualquiera y finalmente se presentan términos ontológicos relacionados con materiales & métodos y diseño experimental requeridos en diferentes procedimientos experimentales (figura 58). Los términos presentados sobre la organización retórica de los cuadernos de laboratorio provienen de diferentes normativas como DC y AgMes; tesauros como AGROVOC y NCIt; ontologías como OBI, CHEBI, PO, GO, SNOMED, entre otras.

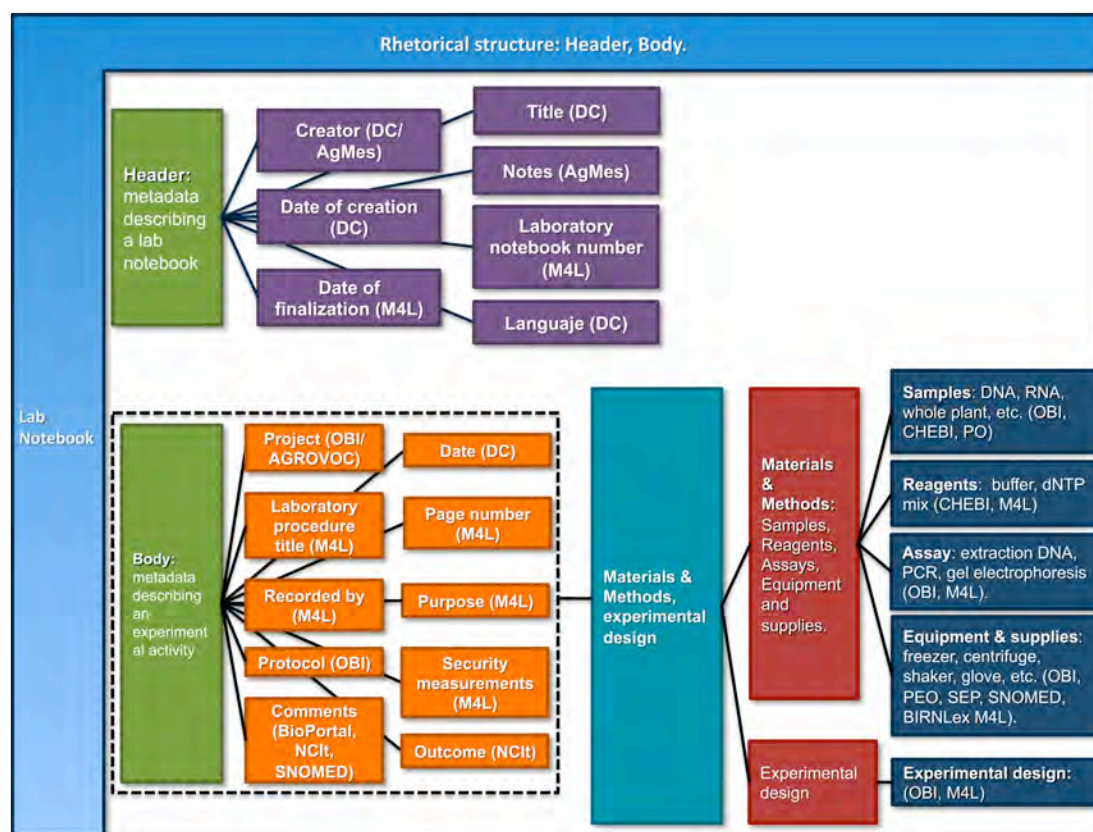


Figura 58. estructura retórica de los cuadernos de laboratorios.

El contar con términos normalizados para representar la organización retórica de los cuadernos de laboratorio facilita la auto-descripción de este tipo de documento y el tipo de información que puede contener. Es decir, en procesos automatizados, este tipo de documento debe ser capaz de “conocer su propio contenido” y “saber que hacer con el”. De esa manera se podrán soportar preguntas de competencia como:

1. Cuales cuadernos de laboratorio (Laboratory notebook number) fueron llevados por Ana Gonzalez (Creator) durante el 2008 y 2009 (Date of creation and finalization)? (figura 59)

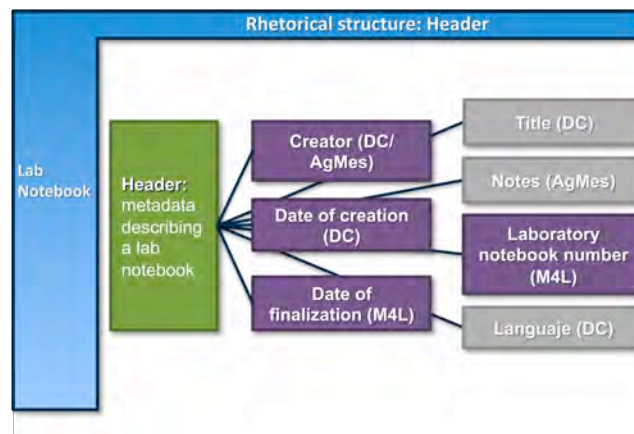


Figura 59. parte de la estructura retórica que soporta la pregunta de competencia 1.

2. En que fechas (Date) fue extraído el ADN (Assay) de los materiales de arroz (Sample) usados en el proyecto “identificación de marcadores moleculares asociados al contenido de Fe y Zn (Project)? (figura 60)

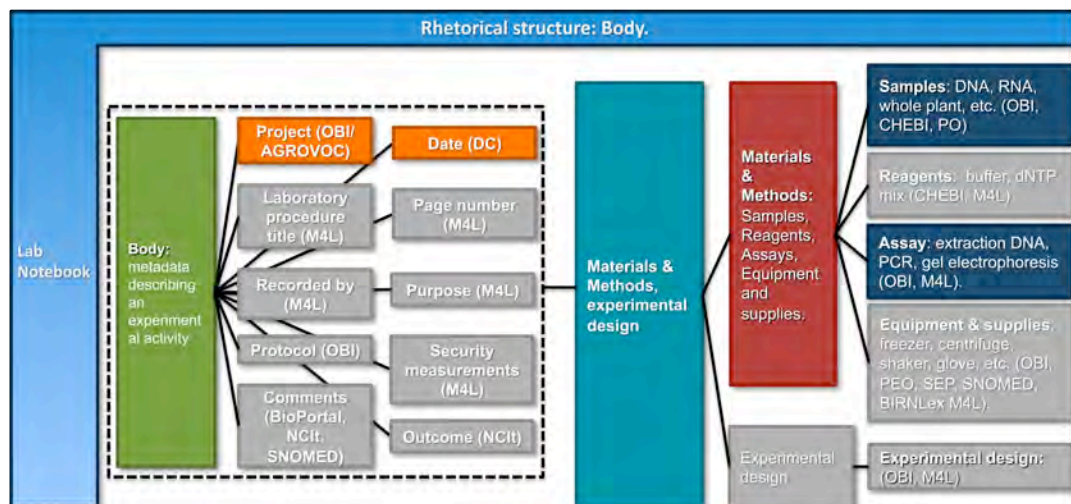


Figura 60. parte de la estructura retórica que soporta la pregunta de competencia 2.

3. En que paginas (Page number) fueron insertadas hojas de resultados provenientes de un LIMS (Outcome)? (figura 61)

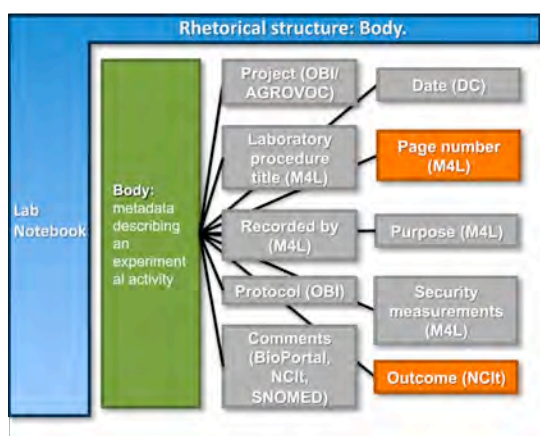


Figura 61. parte de la estructura retórica que soporta la pregunta de competencia 3.

4. Que medidas de seguridad (Security measurements) deben tenerse en cuenta al manipular equipos (Equipments) y reactivos (Reagents) utilizados en un ensayo (Assay) de electroforesis en geles de poliacrilamida? (figura 62)

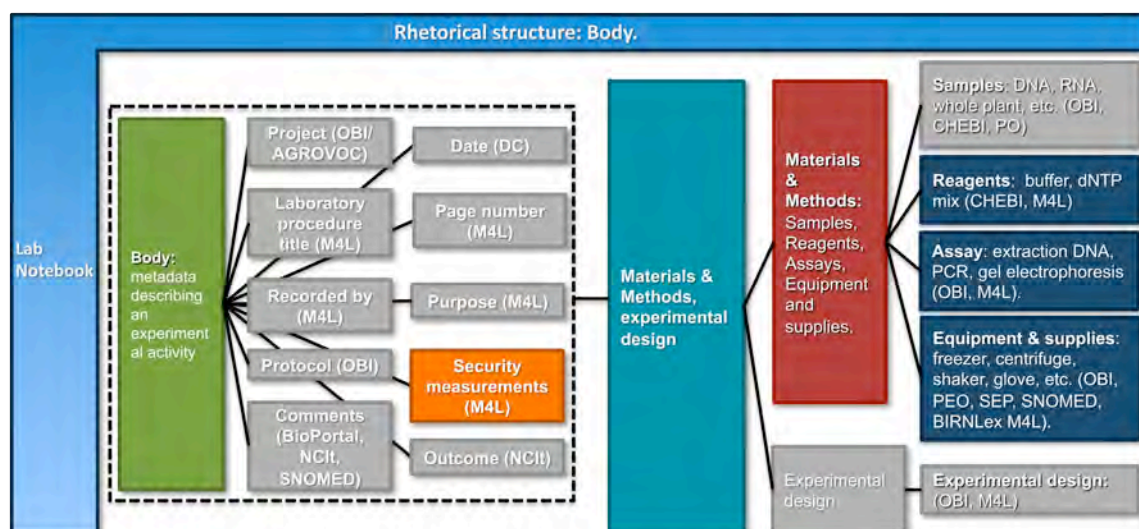


Figura 62. parte de la estructura retórica que soporta la pregunta de competencia 4.

5.4 ESTRUCTURANDO LAS ANOTACIONES LIBRES

Para estructurar el lenguaje natural usado por investigadores para describir actividades experimentales, en lo posible, a través de las ontologías aquí desarrolladas se decidió usar la Annotation Ontology (AO) (Ciccarese et al., 2011). La AO, es una ontología de naturaleza abierta, construida en un lenguaje OWL-DL que permite guardar datos etiquetados y publicarlos como linked data¹⁵⁴. Siguiendo esta vía, en este estudio se soportan preguntas complejas usando un lenguaje SPARQL que involucra varias ontologías por ejemplo, “recuperar de Tags4Labs las paginas etiquetadas por Tim Andrews o Lisa Watson con las etiquetas *arroz* y *hierro* para las cuales hay una entrada de datos proveniente de un LIMS” (figura 63).

```
SELECT ?Tags4Labs ?page

WHERE {

  ?annotation ann:annotates ?page .

  ?annotation pav:createdBy ?user . ?user foaf:name ?userName .

  FILTER(?userName = "lisa watson" || ?userName = "tim.andrews").

  ?annotation ao:hasTag ?tag .

  ?tag tags:name ?tagName .

  FILTER(?tagName = "rice" || ?tagName = "iron")

  ?eLabBook hasPage ?page.

  ?page hasLIMSDataEntry ?lims

}
```

¹⁵⁴ La propuesta linked data surge dentro del marco de la Web Semántica. El término linked data hace referencia al método con el que se pueden mostrar, intercambiar y conectar datos a través de URIs y RDF (<http://linkeddata.org/>).

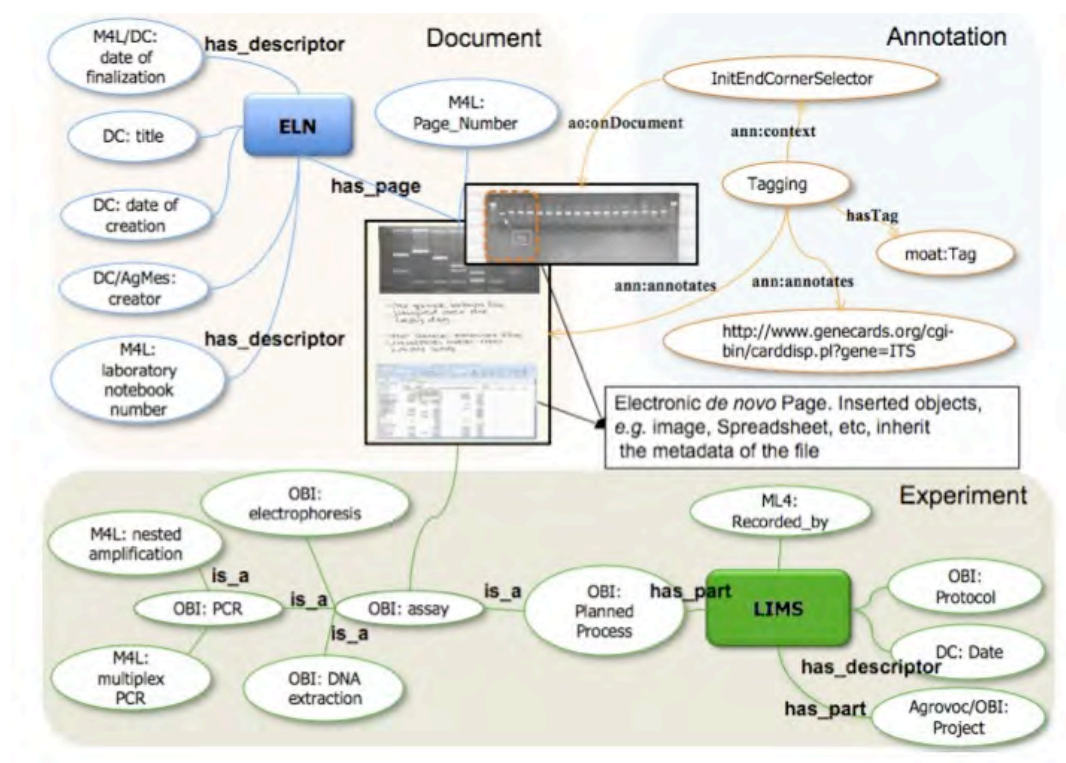


Figure 63. estructuración de la anotación semántica y las etiquetas generadas por usuarios.

La AO puede ser usada para agregar anotaciones o etiquetas a varios tipos de documentos como: paginas HTML, imágenes, videos y bases de datos. Las anotaciones pueden de igual manera identificar un documento completo o una porción del mismo (Ciccarese et al., 2011). Gracias a esta característica, Tags4Labs permite anotar paginas completas y a su vez seleccionar una porción de la misma gracias a un Selector provisto por la AO el cual permite identificar y delimitar la porción del recurso que se desea etiquetar (figura 64).

```
SELECT ?Tags4labs ?page ?file
```

```
WHERE {
```

```
    ?annotation ann:annotates ?page .
```

```
    ?annotation pav:createdBy <http://www.tags4lab.org/foaf.rdf#tim.andrews> .
```


psy promoter" con un término proveniente de un vocabulario controlado "GenBankAB005238". La ANNOT2, busca establecer una definición usando MOAT. Esta anotación se realizó sobre un documento de PubMed sin especificar ninguna porción dentro de él.

Gracias a esta forma de anotación, es posible enriquecer la información en Tags4Labs con términos ontológicos, texto libre, artículos científicos, imágenes, videos y todo aquello tenga un URI. Otras formas de anotación posibles son notas, definiciones, ejemplos y erratas (tabla 8).

Tabla 8. Tipos de anotación provistos por la AO.

Nota	Es un breve record de escritura asociado con un documento o parte de este. El cuerpo de una nota es usualmente texto libre.
Errata	Proporciona una corrección de un documento. El cuerpo de anotación es la versión corregida del contenido o una descripción del error.
Ejemplo	El cuerpo de anotación representa un caso típico de una clase o grupo.
Definición	El cuerpo de anotación contiene una explicación concisa del significado de una palabra, frase, símbolo o imagen que se está anotando.

Mediante el etiquetado en Tags4Labs los investigadores están generando nubes de etiquetas. Aunque el prototipo no cuenta con tablas de contenido, los usuarios identificaron las nubes de etiquetas como un recurso valioso para una rápida inspección de contenidos. Para facilitar la generación de etiquetas, combinando aquellas que vienen de ontologías con aquellas provistas por usuarios, en este estudio se soportan preguntas tales como “recuperar de Tags4Labs aquellas paginas que contienen un archivo de Excel y que ha sido etiquetado por Tim Andrews con la etiqueta *arroz* y opcionalmente con *PCR*”

5.5 TAGS4LABS PENSADO COMO UN MODELO DE MULTICAPAS

El diseño del prototipo, toma como principio los cuadernos de laboratorio. Con base en ello, se consideran tres capas de información:

1. La primera capa, es aquella que provee una pagina de cuaderno en papel, es decir una cuadrícula que facilita la escritura, un numero de pagina y el cuaderno al cual pertenece dicha pagina.
2. La segunda capa, la provee las anotaciones hechas por un usuario directamente sobre el papel y,
3. La tercera capa, es aquella que viene de medios externos por ejemplo, anotaciones hechas sobre post-it, tablas, figuras, fotos, etc.

Con base en este principio, los usuarios de Tags4Labs, podrán contar con dos funcionalidades. Una de ellas es poder conservar una copia en formato electrónico de la información contenida en los cuadernos en papel. La otra funcionalidad es la creación de paginas nuevas, para la toma de datos en tiempo real y en formato electrónico (figura 65). A continuación, en la sección 5.6, se presentan las funciones con que cuenta Tags4Labs actualmente.

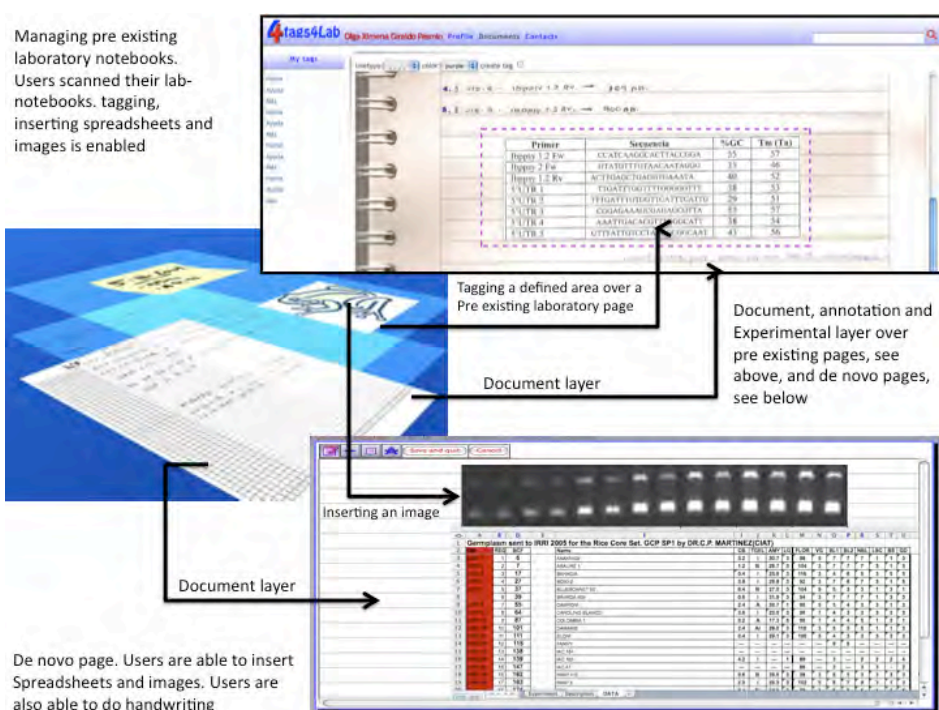


Figura 65. Capas de información de Tags4Labs. 1) la provista por el papel en blanco, 2) la provista por anotaciones manuscritas y/o digitales y 3) aquella importada de otros recursos.

5.6 TAGS4LABS EN ACCIÓN

Tags4Labs es un prototipo de libre acceso disponible en Biotea: <http://www.biotea.ws/node/5>. A continuación, se presentan las funciones con que actualmente cuenta el prototipo Tags4Labs con el propósito de facilitar la organización, recuperación e intercambio de información de información generada en un laboratorio. En primer lugar, para ingresar a Tags4labs, cada usuario deberá: 1) ingresar un nombre de usuario, 2) una clave y 3) iniciar sesión (figura 66).



Figura 66. iniciar sesión.

Tags4Labs permite manejar información pre-existente con el propósito de conservar ésta información en un ambiente donde se facilite la recuperación contextual de la información. El prototipo cuenta con una interfase desde la cual el usuario podrá importar desde una pagina o varias al tiempo en formato comprimido (.zip) (figura 67).



Figure 67. Tags4Labs permite conservar un histórico de los cuadernos en papel a manera de imágenes en formato jpg, gif o png.

De igual manera Tags4Labs facilita la recuperación de la información gracias a las anotaciones o etiquetas que se pueden agregar a cada actividad experimental allí consignada (figura 68).

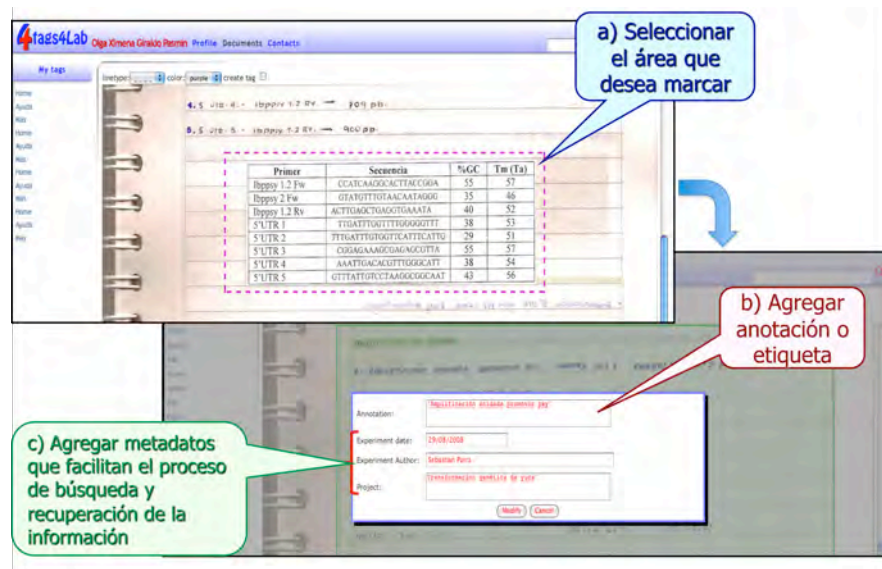


Figura 68. proceso de anotación en Tags4Labs.

Las etiquetas o anotaciones previamente creadas pueden ser editadas en cualquier momento (figura 69).

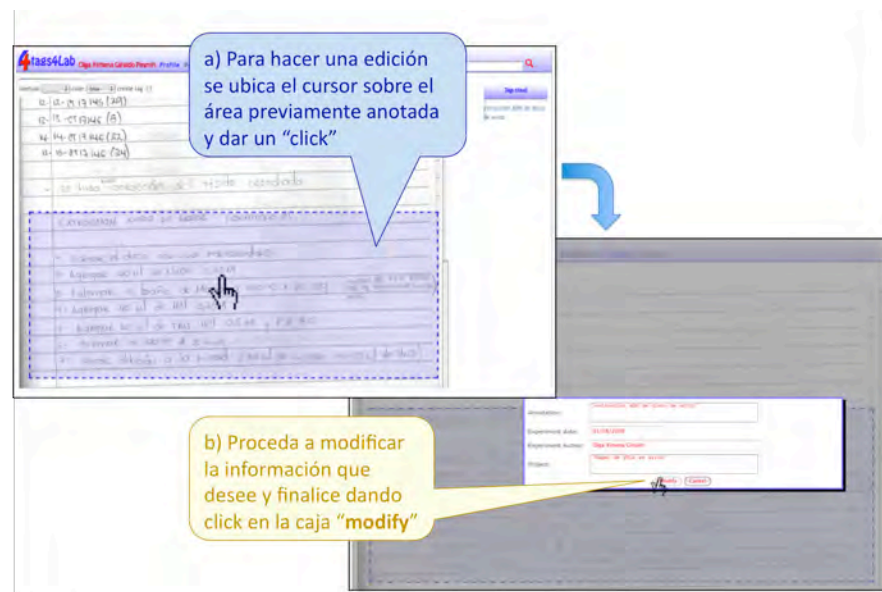


Figura 69. proceso de edición de las etiquetas en Tags4Labs.

como se vio en la sección 5.4, para facilitar la correcta indexación, recuperación, e intercambio de información, las anotaciones son estructuradas, en la medida de lo posible, sobre las ontologías (figura 70). De esta manera se podrán soportar preguntas de competencia como: recuperar las diferentes metodologías de extracción de ADN que un investigador ha consignado en Tags4Labs.

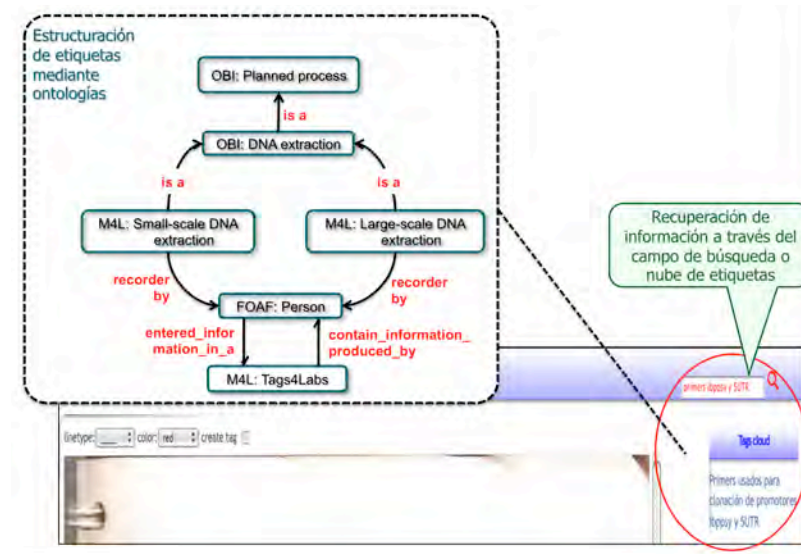


Figura 70. recuperación de información en Tags4Labs.

La figura 70 también ilustra que Tags4Labs funciona como una base de datos. Es decir, es un gran índice de nombres creados por cada usuario al anotar las paginas de su cuaderno de laboratorio. Por lo tanto, la recuperación de la información puede hacerse a través del campo de búsqueda o de la nube de etiquetas.

Finalmente, Tags4Labs permite la generación de páginas electrónicas sobre las cuales es posible insertar gráficas, hojas electrónicas, presentaciones, PDFs, plantillas de bases de datos, etc. Esto significa que permite la toma de datos en tiempo real y en formato electrónico de manera similar como se hace con un

cuaderno de papel. Cada elemento agregado a una nueva pagina podrá ser etiquetado (figura 71).

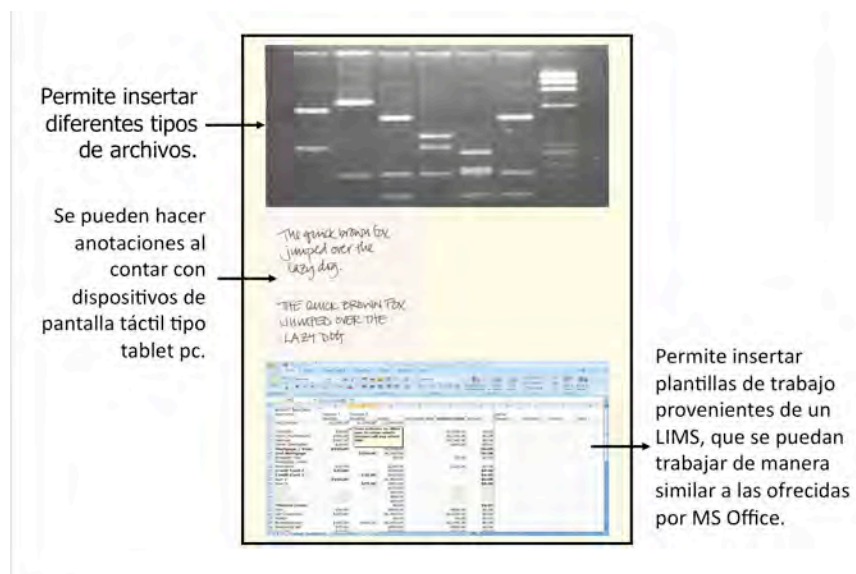


Figura 71. Tags4Labs permite la toma de datos en tiempo real.

5.7 EVALUACIÓN DE TAGS4LABS

El concepto de Tags4Labs es simple, la adecuada orquestación de ontologías facilita el marcaje de registros experimentales y la recuperación de información. La evaluación del mecanismo de orquestación de ontologías propuesto en esta tesis se llevó a cabo contra preguntas de competencia presentadas por los expertos de dominio que trabajan en el laboratorio de biotecnología del CIAT, además, se evaluó la ontología M4L siguiendo los parámetros de evaluación propuestos por el consorcio OBI. Esta evaluación esta disponible en (Brinkman et al., 2010).

Como trabajo futuro en relación a la información recuperada, se medirá como la respuesta es adecuada a la pregunta realizada; además se considera importante

el medir la satisfacción de quien ha hecho la demanda. Para ello se deberá medir la **relevancia** de un documento recuperado, es decir, si este sirve como respuesta a una pregunta (valor 1) o no (valor 0) (Hersh et al., 1994).

De igual manera se evaluará la recuperación de documentos relevantes. Los documentos pueden ser recuperados o rechazados al establecer la comparación entre la pregunta y la base de datos (Baeza-Yates. and Ribeiro-Neto., 1999). El conjunto de documentos recuperados se divide en dos grupos: documentos relevantes recuperados, es decir, aquellos que se han recuperado correctamente; y los no relevantes, recuperados erróneamente que provocan ruido en la salida. Los documentos no recuperados, que a su vez se dividen en los relevantes, rechazados por el sistema de manera errónea y los no relevantes, rechazados de manera correcta por el sistema (figura 72) (Baeza-Yates. and Ribeiro-Neto., 1999).

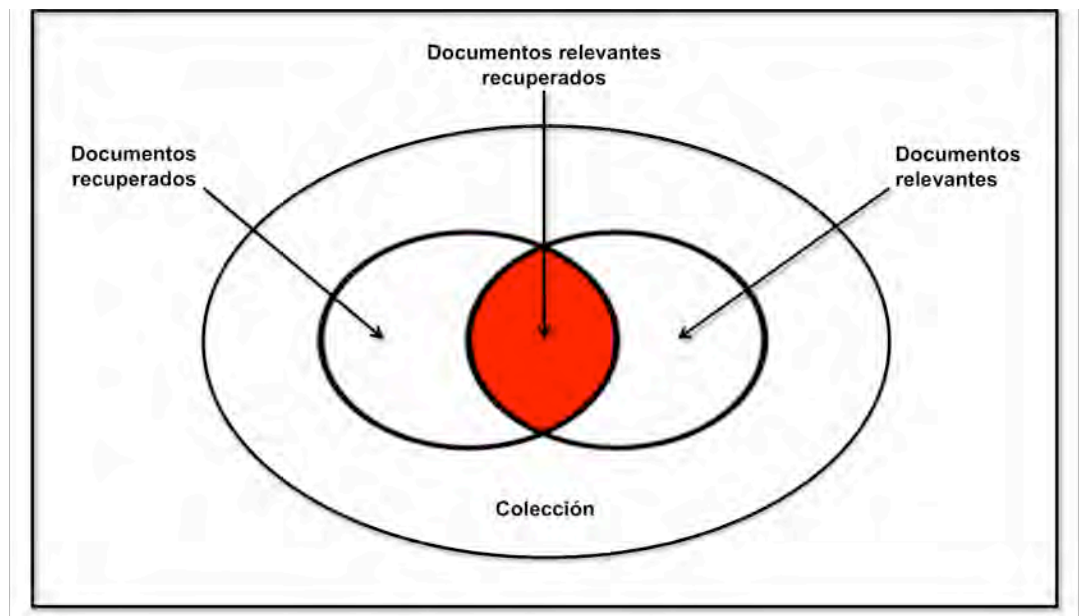


Figura 72. esquema recuperación de documentos. Adaptado de (Baeza-Yates. and Ribeiro-Neto., 1999).

5.8 COMENTARIOS FINALES

En este capítulo se mostró un prototipo que aprovecha la facilidad de uso que ofrecen anotaciones libres al mismo tiempo que considera el uso de ontologías para estructurar la información allí contenida.

Tags4Labs consideró el uso de la AO para estrechar la brecha entre el lenguaje natural usado por investigadores para describir su trabajo en el laboratorio con ontologías de dominio como M4N y M4L disponibles en la página Web Biotea (<http://www.biotea.ws/node/3>). Esta implementación que considera linked data, facilitaría el intercambio, visibilidad y relación de datos allí almacenados con otras fuentes de información. De igual manera este prototipo facilitaría la recuperación de la información y solucionaría problemas de ambigüedad típicos de las folksonomías al atar los datos anotados con términos que provienen de ontologías, tesauros, metadatos o vocabularios controlados.

De igual manera se quiere aprovechar el componente colaborativo y social de esta herramienta para que la misma comunidad de usuarios pueda soportar y evolucionar las ontologías.

En un futuro cercano se espera mejorar la usabilidad de Tags4Labs, de igual manera se planea liberar el software de manera abierta. Actualmente se continua mejorando las ontologías propuestas.

5.9 BIBLIOGRAFIA

ANDERSON, N. R., ASH, J. S. & TARCZY-HORNOCH, P. (2007) A qualitative study of the implementation of a bioinformatics tool in a biological research laboratory. *Int J Med Inform*, 76, 821-8.

- BAEZA-YATES. & RIBEIRO-NETO. (1999) *Modern Information Retrieval*, Boston, MA, USA, Addison-Wesley Longman Publishing Co., Inc.
- BRINKMAN, R. R., COURTOT, M., DEROM, D., FOSTEL, J. M., HE, Y., LORD, P., MALONE, J., PARKINSON, H., PETERS, B., ROCCA-SERRA, P., RUTTENBERG, A., SANSONE, S.-A., SOLDATOVA, L. N., STOECKERT JR., C. J., TURNER, J., ZHENG, J. & CONSORTIUM., T. O. (2010) Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*, 1, 11.
- CICCARRESE, P., OCANA, M., GARCIA CASTRO, L. J., DAS, S. & CLARK, T. (2011) An open annotation ontology for science on web 3.0. *J Biomed Semantics*, 2 Suppl 2, S4.
- GREEN, N. L. (2010) Representation of Argumentation in Text with Rhetorical Structure Theory. *Argumentation*, 24, 181-196.
- HERSH, W. R., ELLIOT, D. L., HICKAM, D. H., WOLF, S. L., MOLNAR, A. & LEICHTENSTIEN, C. (1994) Towards New Measures of Information-Retrieval Evaluation. *Journal of the American Medical Informatics Association*, 895-899.
- MARZOLF, B. & TROISCH, P. (2006) SLIMarray: lightweight software for microarray facility management. *Source Code Biol Med*, 1, 5.
- PENG, G. (2009) Using Rhetorical Structure Theory (RST) to describe the development of coherence in interpreting trainees. *Interpreting*, 11, 216-243.
- RUBACHA, M., RATTAN, A. K. & HOSSELET, S. C. (2011) A review of electronic laboratory notebooks available in the market today. *J Lab Autom*, 16, 90-8.
- STEAD, D. A., PATON, N. W., MISSIER, P., EMBURY, S. M., HEDELER, C., JIN, B., BROWN, A. J. & PREECE, A. (2008) Information quality in proteomics. *Brief Bioinform*, 9, 174-88.
- TABOADA, M. & MANN, W. C. (2006a) Applications of rhetorical structure theory. *Discourse Studies*, 8, 567-588.

- TABOADA, M. & MANN, W. C. (2006b) Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8, 423-459.
- TOPEL, T., SCHEIBLE, D., TREFZ, F. & HOFESTADT, R. (2010) RAMEDIS: a comprehensive information system for variations and corresponding phenotypes of rare metabolic diseases. *Hum Mutat*, 31, E1081-8.
- UZEDA, V. R., PARDO, T. A. S. & NUNES, M. D. V. (2008) Evaluation of Automatic Text Summarization Methods Based on Rhetorical Structure Theory. *Isda 2008: Eighth International Conference on Intelligent Systems Design and Applications, Vol 2, Proceedings*, 389-394.

6. DISCUSION, CONCLUSIONES Y TRABAJO FUTURO

Este capítulo hace un recorrido por los principales elementos descritos en este trabajo que contribuyen a responder la pregunta de investigación planteada. *¿Cómo organizar, clasificar y compartir la información consignada en los cuadernos de laboratorio?* Finalmente, se plantean algunas líneas de investigación futuras.

6.1 DISCUSION

El investigador ha visto la necesidad de robustecer los mecanismos para hacer mas eficiente la captura de datos en el laboratorio. Hoy en día existen varias alternativas para la administración de información producida en un laboratorio. Diferentes tipo de LIMS han sido diseñados de acuerdo a las necesidades industriales y académicas. Sin embargo, la capa semántica dentro de este tipo de herramientas necesita ser mejorada. Dicha capa requiere ser interoperable, contar con ontologías bien modularizadas para facilitar la recuperación de la información.

De acuerdo con lo anterior fueron varias las lecciones aprendidas durante el planteamiento de una metodología para mejorar la captura, clasificación, recuperación e intercambio de información producida en un laboratorio. En primer lugar, el uso de metadatos fue planteado como el primer paso de estandarización de datos digitales y no-digitales almacenados en cuadernos de laboratorio. Este primer paso de estandarización fue fundamental para:

1. describir e identificar recursos como el cuaderno de laboratorio y cualquier tipo de experimento
2. proporcionar un historial de los recursos, saber quien los creó y quien los modificó,
3. describir las características de los recursos,
4. proveer información de los derechos de uso, y
5. mejorar las capacidades de localización/recuperación.

Dado que los metadatos no son suficientes para soportar los diferentes tipos de procedimientos experimentales, en este estudio se consideró el uso de ontologías.

El encontrar en la literatura diversidad de métodos empleados para el diseño de ontologías refleja que en el mundo de las ontologías no hay verdades absolutas, estas son creadas artificialmente, no se descubren. Por lo tanto, una ontología es una forma de ver el mundo, ya que cada una de éstas, incorpora un punto de vista.

Sin embargo, la falta de métodos y normas estándar para la definición de una ontología tanto al interior de los equipos de trabajo, como entre ellos, obstaculiza el desarrollo de ontologías compartidas y consensuadas, la extensión de ontologías desarrolladas por otros grupos y su rehúso en otras ontologías y en aplicaciones finales(Soldatova and King, 2005, Bard and Rhee, 2004).

Dentro del proceso de aprendizaje en la construcción de ontologías, fue fundamental el entender que el desarrollo de una ontología debe ser planeado con anticipación a la ejecución misma del proceso. Particularmente, para el desarrollo de la ontología de procedimientos experimentales (M4L), era claro que tipo de conocimiento se deseaba capturar pero no la manera de como hacerlo. Por lo tanto el desafío en esta fase de la investigación, fue el entender como

representar semánticamente el conocimiento que se deseaba capturar en la ontología. La elaboración de casos de uso, representados por medio de mapas conceptuales, facilitó esta labor de construcción de M4L.

De igual manera la Annotation Ontology (AO), fue fundamental a la hora de estrechar el gap existente entre las ontologías de dominio y el uso de lenguaje natural proporcionado por la comunidad científica. Ahora el intercambio de información sucede sobre una base conceptual, por ejemplo, aquellos investigadores que realizan estudios sobre genes involucrados en el transporte de hierro, pueden compartir y ampliar información con aquellos quienes investigan sobre las causas relacionadas con la falta de ingesta y/o asimilación de hierro en poblaciones humanas. Tales conceptos basados en folksonomías fueron soportados en las ontologías aquí diseñadas. Dichas ontologías basadas en anotaciones libres facilitan la organización y recuperación de la información. Por lo tanto las anotaciones libres realizadas por investigadores que trabajan en proyectos similares constituyen un recurso valioso para proveer nuevos términos a las ontologías.

6.2 CONCLUSIONES

Como resultado de la investigación se propone un sistema que hace uso de varias ontologías; estas se orquestan de manera tal en que se permite la coexistencia de anotaciones libres (tags) y aquellas provenientes de ontologías. La orquestación de ontologías, al igual que la coexistencia de marcas provistas por usuarios y aquellas semánticamente validas, se soporta extensamente en el uso de la Annotation Ontology (AO). La coexistencia entre ontologías y anotación libre provisto por la comunidad probó, en este trabajo, ser útil. No solo para complementar las ontologías, también para apoyar la generación de mecanismos personales de organización sobre los cuadernos de laboratorio. El sistema

generado facilita la organización, recuperación, clasificación e intercambio de información generada en un laboratorio.

Además de haberse reusado ontologías, se generaron aquellas necesarias para soportar la integración e interoperabilidad requerida. Para el desarrollo de estas, se hizo especial énfasis en el aspecto metodológico. La metodología planteada en esta investigación retoma elementos previamente presentados por otros autores. Se siguen los estándares del proyecto OBO y buenas practicas sugeridas por la ontología OBI. Por ello M4L, ontología sobre procesos experimentales desarrollada en este estudio, no solo es compatible con OBI, también se considera una extensión de la misma. Un importante resultado de este trabajo lo constituye el marco metodológico generado. Este es replicable, se proveen detalles para cada paso sugerido y además se describe el proceso en términos de entrada (inputs) y salida (outputs); esto es útil para OBI y para aquellos grupos que deseen emprender desarrollos ontológicos que deban ser compatibles con OBI.

El extenso análisis sobre los cuadernos de laboratorio y su relación con la información generada, permitió definir una serie de tipos de datos (ver capítulo 3, sección 3.4), sobre los cuales se plantean ciertas funcionalidades. Todo lo anterior se hizo con la participación de los biólogos para quienes se estaba diseñando y desarrollando la tecnología. Este ejercicio participativo, *participatory design*, había sido planteado antes por otros autores, sin embargo en el presente trabajo se llevó a cabo a una escala mayor y con objetivos mas precisos en un ambiente real de laboratorio. La constante interacción con usuarios finales, su continuo involucramiento en el diseño de la tecnología facilitaron su compromiso y los hizo mas activos en todo el proceso.

Los resultados presentados en esta investigación, tales como: i) la “traducción” de tags o anotaciones libres a un vocabulario controlado provisto por el uso de metadatos y términos ontológicos tanto propios como de reuso, ii) la presentación

de la estructura retórica de los cuadernos de laboratorio, y iii) el concepto de cuaderno de laboratorio como un repositorio de información y facilitador de diseño, van mas allá de la tecnología con la cual se lleve a cabo la implementación. El aspecto tecnológico es importante; como también lo es la metodología que se defina para el manejo integral de información en un laboratorio. Este ultimo punto puede jugar un papel determinante en la adopción de la tecnología. Siendo el ambiente de laboratorio uno particularmente complejo, la tecnología sola puede no ser suficiente. Guías relativas a buenas practicas, metodología para manejar la información, haciendo énfasis en los beneficios para los investigadores, pueden facilitar la apropiación de la tecnología.

6.3 TRABAJO FUTURO

El prototipo Tags4Labs es el primer paso hacia una herramienta interactiva, portátil y de fácil uso para soportar el trabajo del investigador en el laboratorio. El estado actual del prototipo proporciona una plataforma para la futura investigación y desarrollo de tecnología para cuadernos de laboratorio electrónicos (ELN).

Se requiere mas estudios de usabilidad para avanzar el prototipo; para ello se debería trabajar con usuarios con el propósito de identificar una lista de funcionalidades prioritarias que deban ser incluidas en la primera versión del sistema. De igual manera es importante resaltar la importancia del diseño participativo en todo el proceso de desarrollo de software. La extensión de las ontologías y el desarrollo de nuevas puede ser soportado sobre la metodología aquí planteada. Se recomienda continuar con la estructuración de nuevos casos de uso que involucren otros procedimientos experimentales. De la misma manera es importante el establecer una mayor interacción con LIMS; además de estudiar mejor los aspectos semánticos en este tipo de sistemas.

ANEXOS

En esta sección se presentan: a) las clases de BFO a partir de las cuales se desarrolló M4L, b) las relaciones que provee OBO_REL, c) los metadatos que contiene ontology-metadata y d) el proceso de construcción de la ontología usando la versión 3.4.4 de Protégé.

Anexo A. Clases de la ontología general BFO, importadas a M4L

Tabla 9. Clases de la ontología general BFO, importadas a la ontología M4L.

ID	Nombre	Definición
BFO:0000001	entity	
BFO:0000002	continuant	An entity that exists in full at any time in which it exists at all, persists through time while maintaining its identity and has no temporal parts
BFO:0000003	occurrent	An entity that has temporal parts and that happens, unfolds or develops through time. Sometimes also called perdurants
BFO:0000004	independent continuant	A continuant that is a bearer of quality and realizable entity entities, in which other entities inhere and which itself cannot inhere in anything
BFO:0000005	dependent continuant	A continuant that is either dependent on one or other independent continuant bearers or inheres in or is borne by other entities
BFO:0000015	processual entity	An occurrent that exists in time by occurring or happening, has temporal parts and always involves and depends on some entity
BFO:0000017	realizable entity	A specifically dependent continuant that inheres in continuant entities and are not exhibited in full at every time in which it inheres in an entity or group of entities. The exhibition or actualization of a realizable entity is a particular manifestation, functioning or process that occurs under certain circumstances
BFO:0000019	quality	A specifically dependent continuant that is exhibited if it inheres in an entity or entities at all (a categorical property)

ID	Nombre	Definición
BFO:0000020	specifically dependent continuant	A continuant that inheres in or is borne by other entities. Every instance of A requires some specific instance of B which must always be the same
BFO:0000023	role	A realizable entity the manifestation of which brings about some result or end that is not essential to a continuant in virtue of the kind of thing that it is but that can be served or participated in by that kind of continuant in some kinds of natural, social or institutional contexts
BFO:0000031	generically dependent continuant	A continuant that is dependent on one or other independent continuant bearers. For every instance of A requires some instance of (an independent continuant type) B but which instance of B serves can change from time to time
BFO:0000034	function	A realizable entity the manifestation of which is an essentially end-directed activity of a continuant entity in virtue of that continuant entity being a specific kind of entity in the kind or kinds of contexts that it is made for
BFO:0000040	material entity	An independent continuant that is spatially extended whose identity is independent of that of other entities and can be maintained through time

Anexo B. Relaciones que hacen parte de OBO_REL, importadas A M4L

Tabla 10. Relaciones que hacen parte de OBO_REL, importadas a M4L.

ID	Nombre
BFO:0000050	part of
BFO:0000051	has part
BFO:0000052	inheres in
BFO:0000053	bearer of
BFO:0000054	realized by
BFO:0000055	realizes
BFO:0000056	participates in
BFO:0000057	has participant
BFO:0000058	concretization of
BFO:0000059	concretizes
BFO:0000060	immediately preceded by
BFO:0000061	immediately precedes
BFO:0000062	preceded by
BFO:0000063	precedes
BFO:0000064	course of
BFO:0000065	has course
BFO:0000066	occurs in
BFO:0000067	contains process

Anexo C. OBI minimal metadata, ontología de metadatos reusada en M4L

Tabla 11. Metadatos obligatorios

Nombre	Definición
editor preferred term	The concise, meaningful, and human-friendly name for a class or property preferred by the ontology developers that reflects community usage, or disambiguates the term if necessary.
definition	The official OBI definition, explaining the meaning of a class or property. Shall be Aristotelian, formalized and normalized. Can be augmented with informal definitions to further explain the meaning of the term.
definition editor	The name of the editor of the definition.
definition source	An unambiguous and traceable reference to the source of the definition. Examples: ISBN, URI plus date, MeSH Term, PUBMED ID, DOI.
curation status	Each class has a single curation status.

Tabla 12. Metadatos opcionales

Nombre	Definición
example of usage	A phrase describing how a term should be used. May also include other kinds of examples, such as widely known subclasses or instances of the class. Although essential for high level terms, examples for low level terms (e.g., Affymetrix HU133 array) are not.
alternative term	An alternative name for a class or property which means the same thing, i.e. semantically equivalent, as the preferred term.
editor note	A note containing points under consideration for further term development that may be included in released versions of the ontology. It should contain nothing embarrassing and something potentially useful for end users to understand the ontology. Editor notes should include the date of edit (YYYY/MM/DD) and the author.
curator note	An administrative note intended for the curator of the ontology. It will not be included in the released versions of the ontology, so it should contain nothing necessary for end users to understand the ontology. Curator notes should include the date of edit (YYYY/MM/DD) and the author.

Anexo D. CONSTRUCCIÓN DE UNA ONTOLOGÍA USANDO PROTÉGÉ

Como Instalar Protégé

1. Ir a la pagina principal de Protégé (<http://protege.stanford.edu/>) como se ilustra en la figura 73.

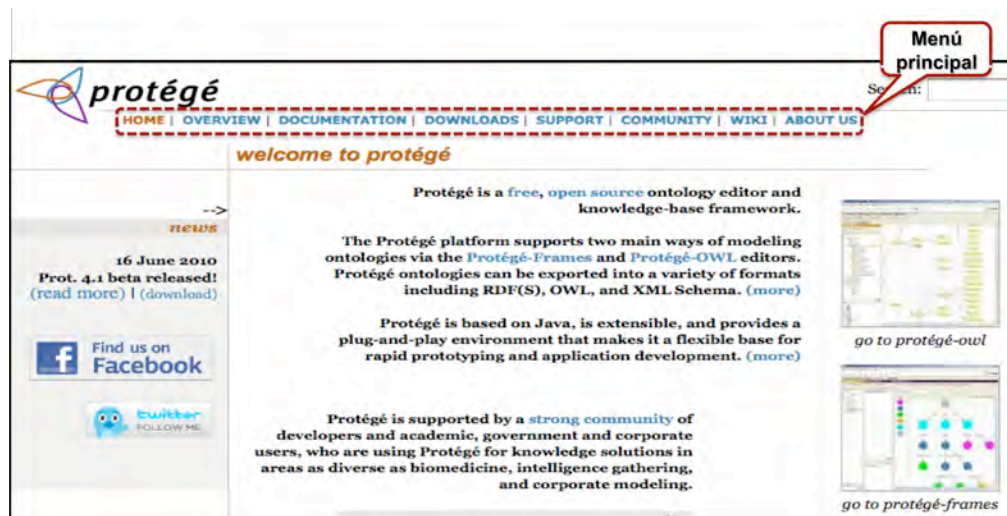


Figura 73. pasos para instalar Protégé. Primera parte.

2. Una vez en la página principal de Protégé, seleccione “download”, opción disponible en el menú principal, ver figura 39. Una vez se encuentre en la página **download de Protégé**, escoja la opción “download page”. A continuación se abrirá una nueva página, la cual presenta las diferentes versiones de Protégé disponibles. Para el caso particular, la ontología M4L se desarrolló usando la versión 3.4.4, ver figura 74.

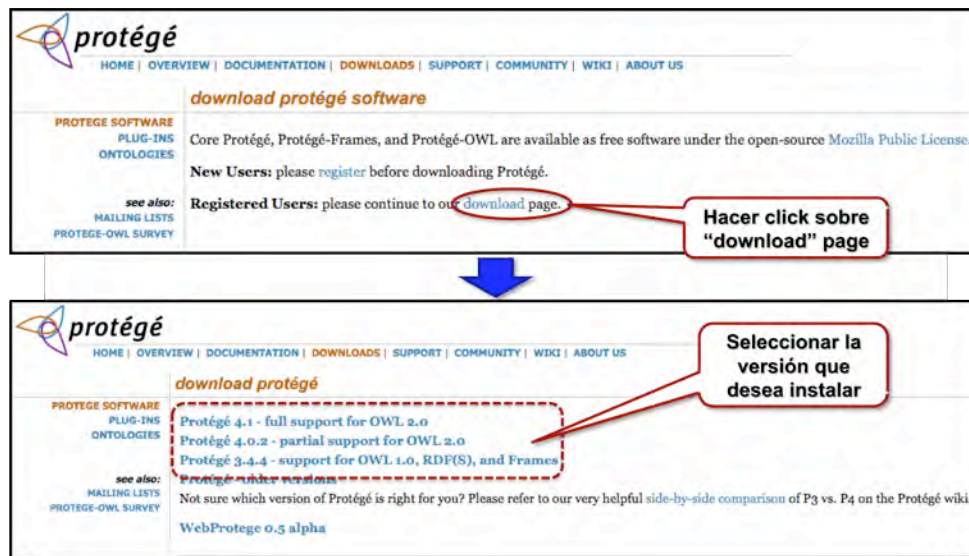


Figura 74. pasos para instalar Protégé. Segunda parte.

3. Seleccione la versión de Protégé que desea descargar y a continuación se abrirá una nueva ventana. En esa nueva ventana seleccione download "Protégé", ver figura 75.

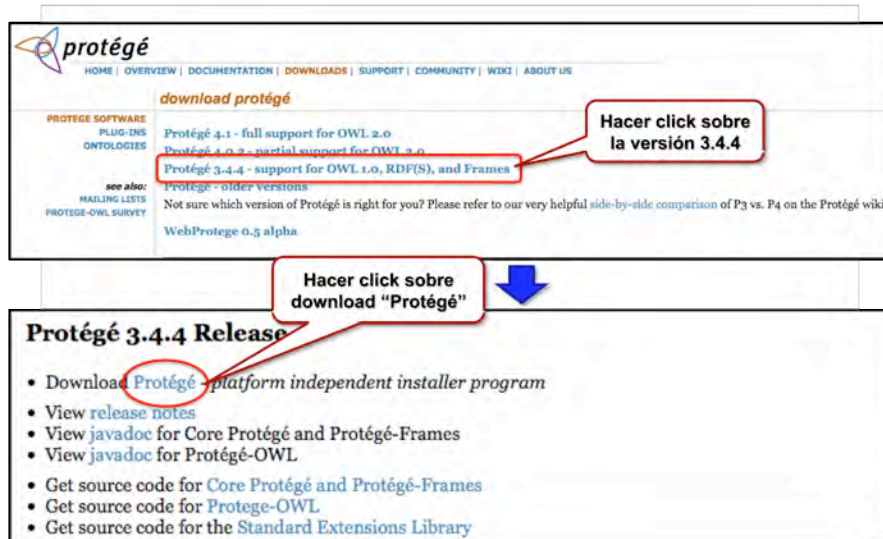


Figura 75. pasos para instalar Protégé. Tercera parte.

4. Requerimientos. Para instalar Protégé, primero se deberá descargar Java (lenguaje de programación orientado a objetos). Si se trabaja desde una plataforma Windows, se podrá descargar ambos programas de manera simultanea desde la página de Protégé. Si por el contrario se trabaja desde una plataforma MacOSX, deberá instalarlos por separado, ver figura 76.

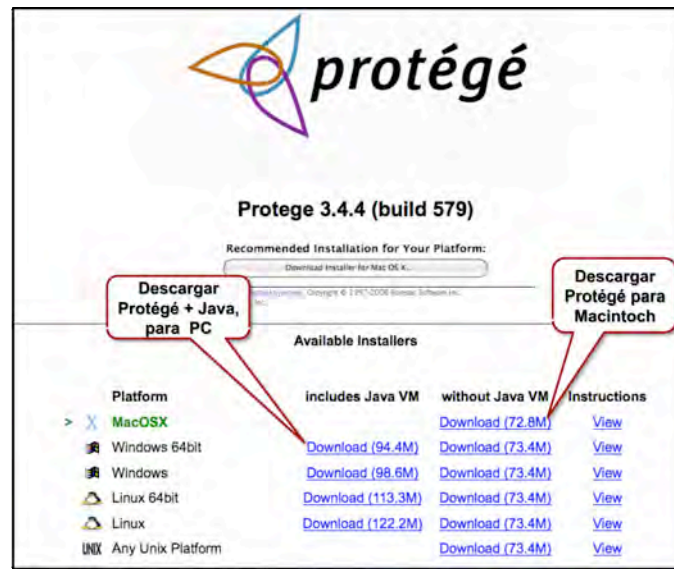


Figura 76. pasos para instalar Protégé. Cuarta parte.

Como Usar Protégé.

- Crear un nuevo proyecto basado en OWL.
1. Una vez se ha descargado exitosamente Protégé, estará listo para iniciar el desarrollo de una ontología. Para comenzar, abra Protégé dando click sobre el icono "*Protege.app*". A continuación se podrán seleccionar las características sobre las cuales desea desarrollar su ontología. Para iniciar el proceso de creación de una ontología, en el menú "*file*", seleccionar "*New Project*", ver figura 77.

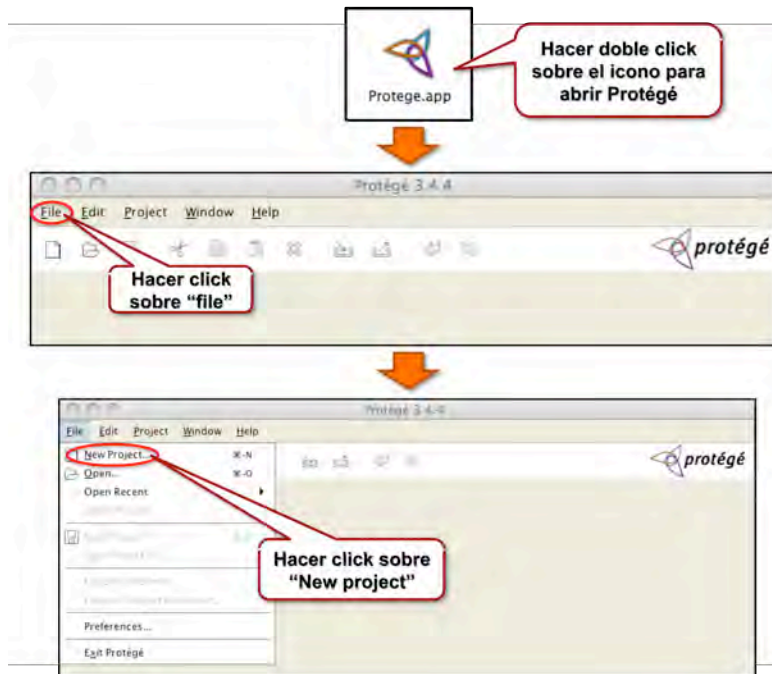


Figura 77. Pasos para crear una nueva ontología.

2. a continuación aparece una nueva ventana ofreciendo diferentes lenguajes para el desarrollo de ontologías. Seleccionar la opción "OWL/RDF files" y luego dar click en "next", ver figura 78.

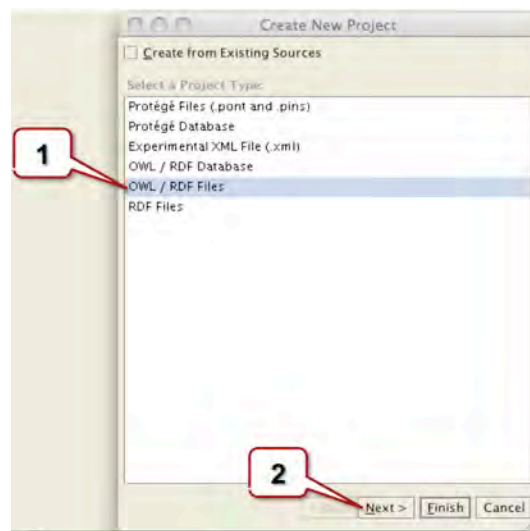


Figura 78. Selección del lenguaje para el diseño de una nueva ontología.

3. Especificar un URI único, que se convertirá en el identificador de su ontología. Por lo general, el URI debe representar la ubicación donde se publicará su ontología, pero esto no es un requisito obligatorio. El establecimiento de un URI único para su ontología puede ayudar a prevenir problemas cuando desee importar ontologías previamente diseñadas. Ver figura 79.

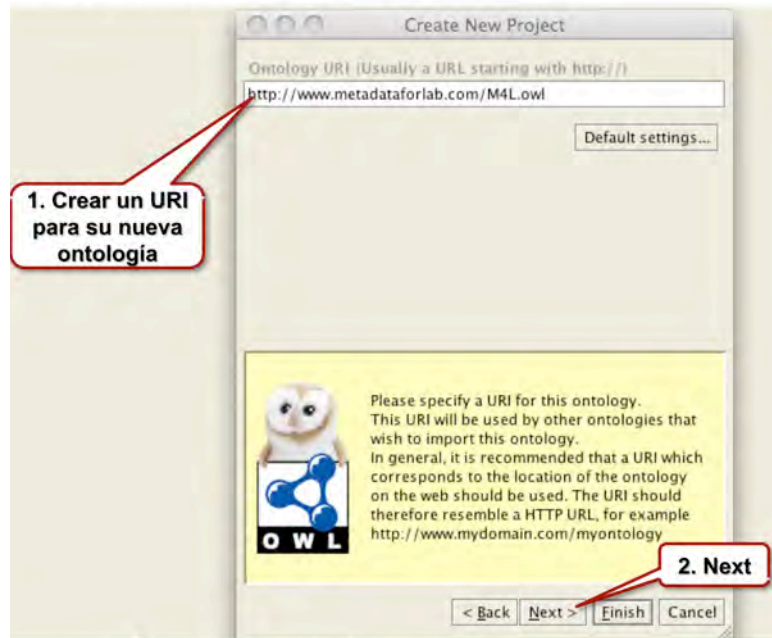


Figura 79. Creación de un URI para una nueva ontología.

4. Seleccione el lenguaje OWL/RDF que desea usar. La selección de un lenguaje particular, activa o desactiva ciertas características de la interfaz de usuario de Protégé-OWL. Para el caso particular, se eligió el lenguaje OWL DL para el desarrollo de M4L. En este punto se tienen dos opciones: i) escoger la opción “finish”, de este modo se crea un proyecto vacío, o ii) escoger la opción “next”, para pasar a la última página de asistencia y elegir la interfaz de usuario inicial, ver figura 80.

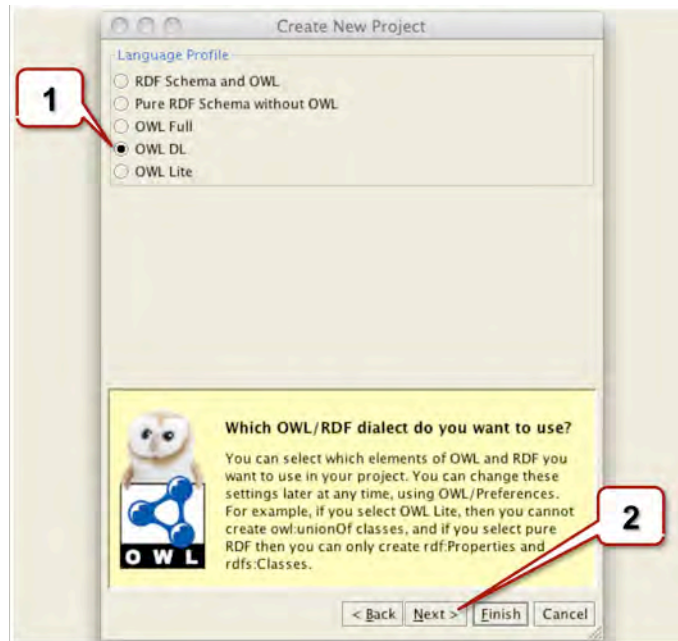


Figura 80. selección del lenguaje OWL/RDF que mejor se ajuste a la ontología a desarrollar.

5. Existen dos opciones para visualizar la interfaz de usuario. La primera opción es “logic view”, la cual es muy expresiva pero no muy intuitiva para los usuarios principiantes en OWL. La segunda opción “properties view”, proporciona una interfaz de usuario sencilla que es mas apropiada para los usuarios principiantes. Como se ilustra en la figura 81, se recomienda seleccionar “properties view” y luego “finish”. A partir de este punto un nuevo proyecto ha sido creado y usted estará listo para editar su propia ontología.

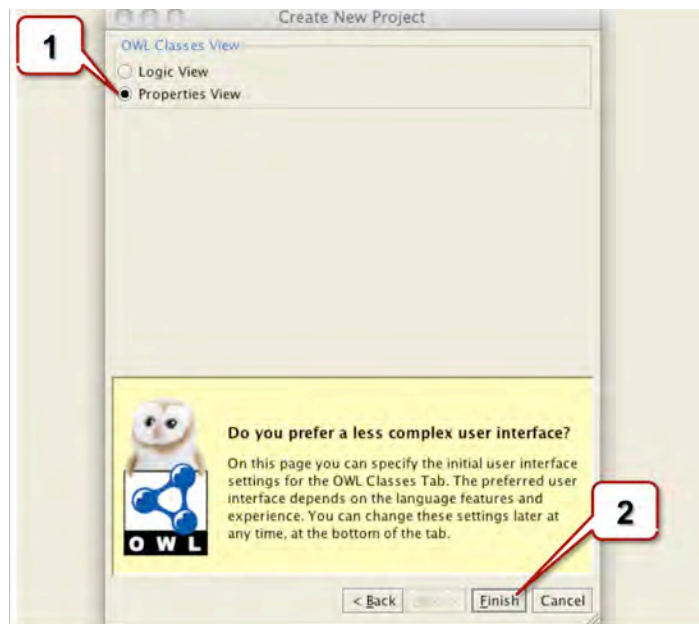


Figura 81. Escogencia de la interfaz de usuario que se ajuste a sus necesidades.

6. En la ventana principal de Protégé, deberá aparecer cinco pestañas: i) Metadata (Protégé se abre de manera predeterminada en esta pestaña), ii) OWL Classes, iii) Properties, iv) Individuals y v) Forms, ver figura 82.

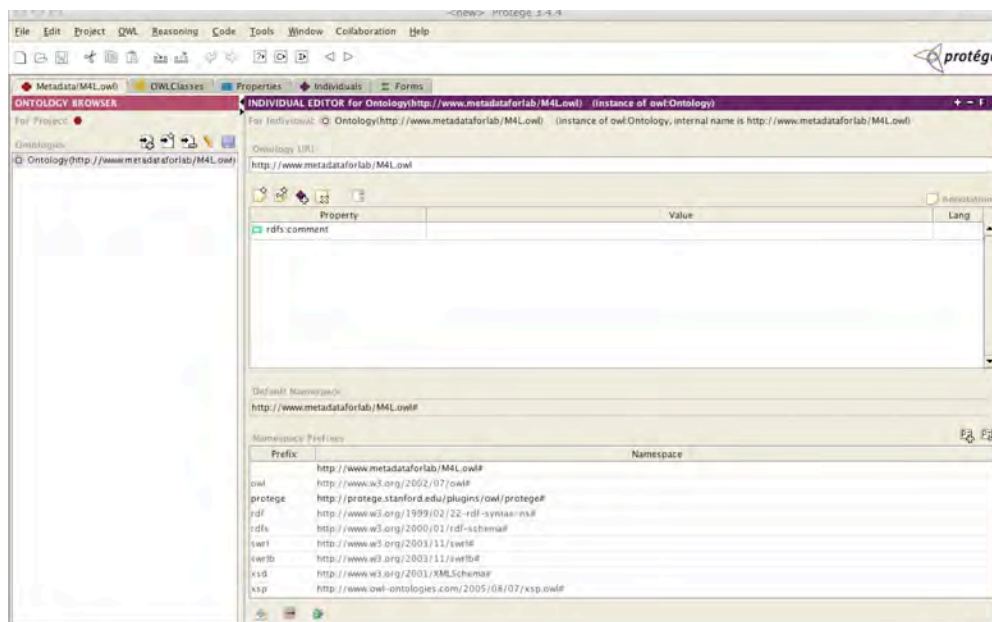


Figure 82. Página principal de Protégé.

- Como visualizar los términos de ontologías. Las ontologías tienen un número de identificación (ID) para cada uno de los términos, propiedades y relaciones que las componen, por lo tanto, para poder visualizar los términos de manera comprensible para las personas, en lugar de un código, se deberán seguir los siguientes pasos:
1. En Protégé, en el menú OWL, escoger “preferences”, allí se abrirá una nueva ventana, ver figura 83.

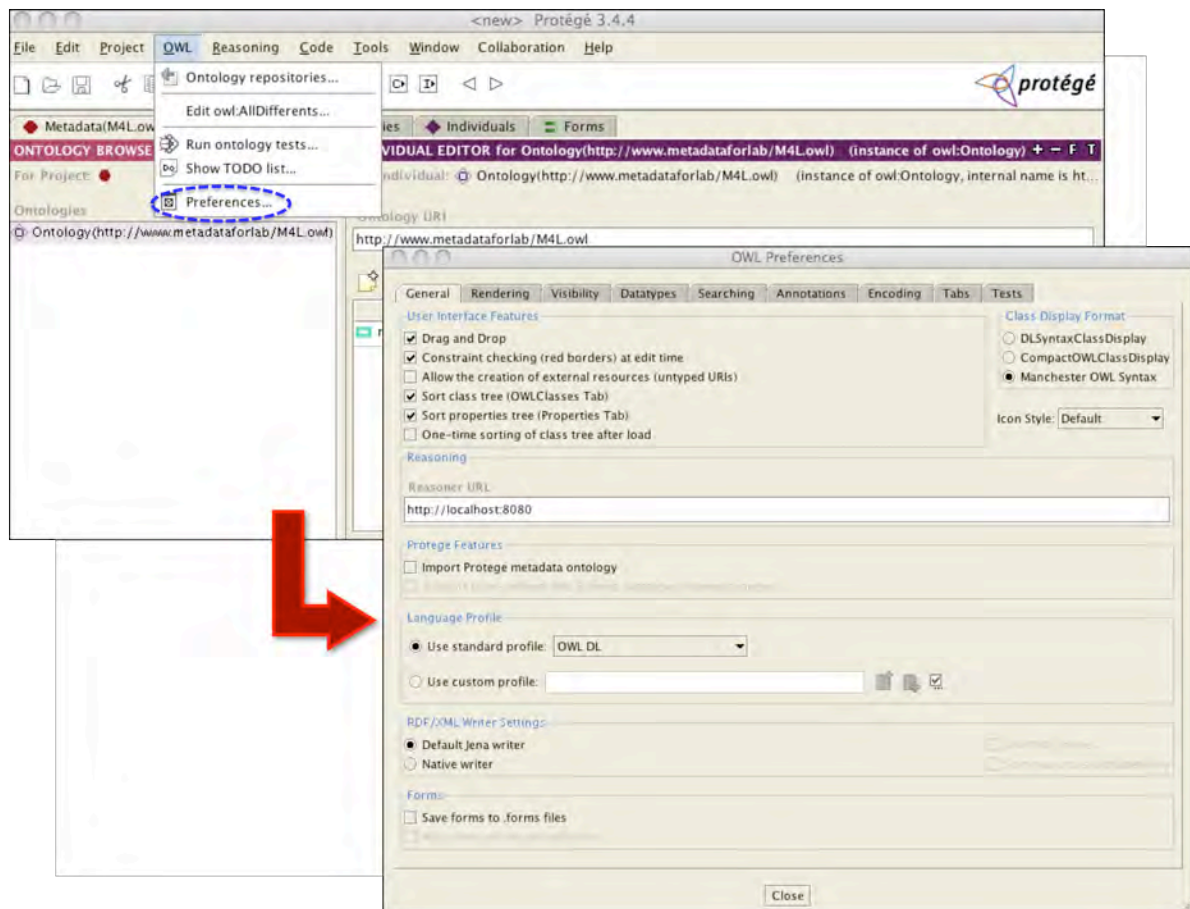


Figura 83. pasos a seguir para visualizar los términos de una ontología. Primera parte.

2. En la ventana “OWL preferences”, seleccionar la pestaña “Visibility”. La sección “metaclasses”, esta ubicada en la esquina superior izquierda de la pestaña, allí marque la casilla "All", a continuación se deberá marcar

automáticamente todas las demás casillas de la sección. Haga lo mismo en la sección de “RDF Properties” y “Anotation Properties”. Luego dar click en cerrar, ver figura 84.

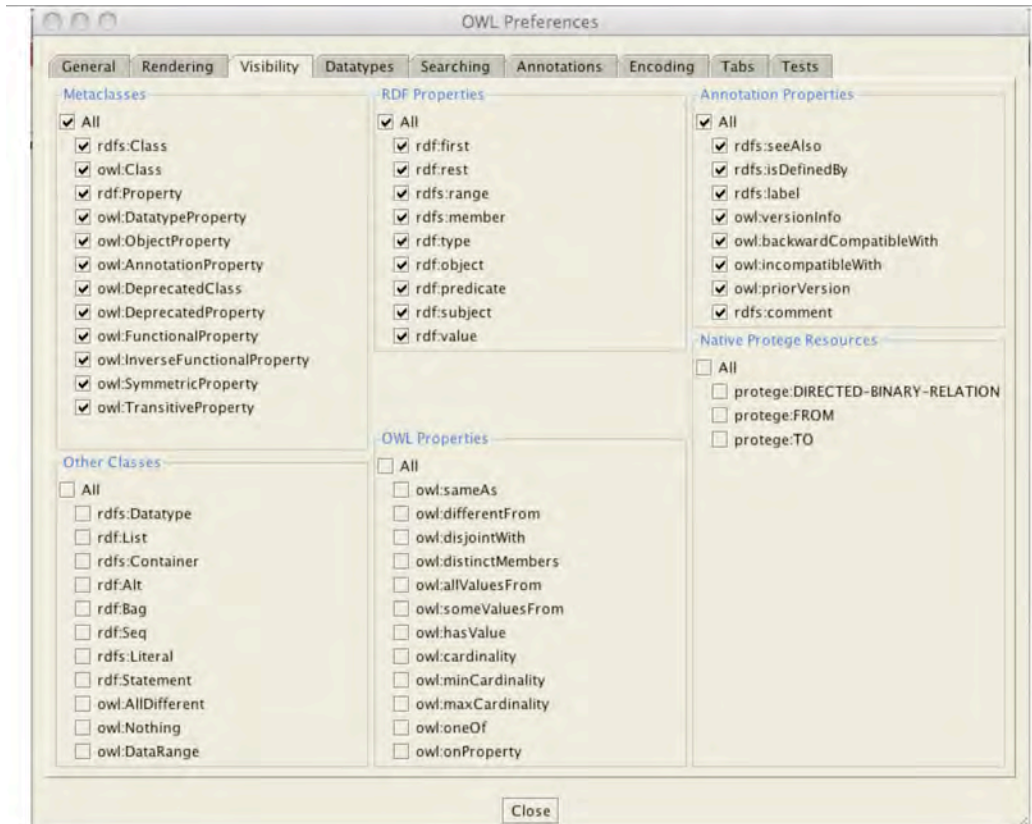


Figura 84. Pasos a seguir para visualizar los términos de una ontología. Segunda parte.

3. En la pagina principal de Protégé, seleccionar la pestaña “Forms”. En la parte izquierda se encuentra la sección “Form browser”. Dentro de esta sección, seleccionar “owl:Thing”, ver figura 85.

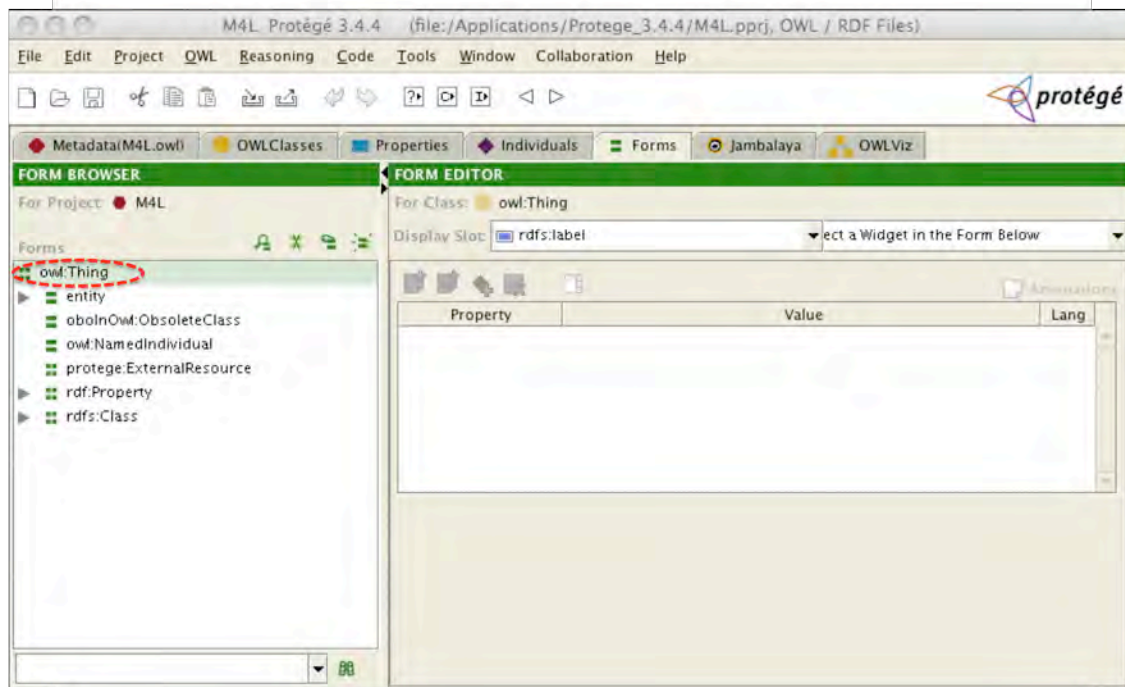


Figura 85. pasos a seguir para visualizar los términos de una ontología. Tercera parte.

4. En la parte derecha de la pestaña “Forms” se encuentra la sección “Form editor”, en esta sección desplegar el menú de opciones que ofrece “Display slot” y seleccionar la opción “rdfs:label”, ver figura 86.

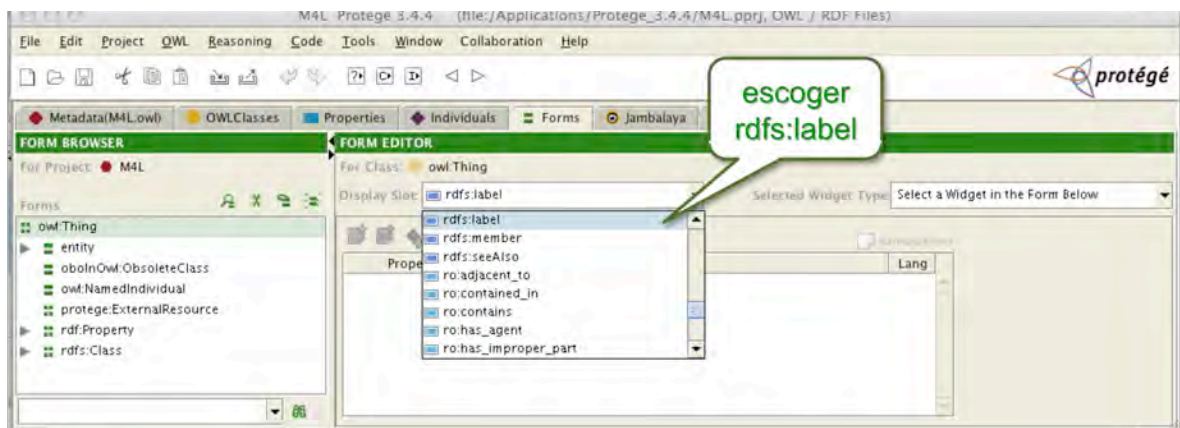


Figura 86. pasos a seguir para visualizar los términos de una ontología. Cuarta parte.

5. A continuación, despliegue las ramas del árbol “rdfs:Class”, haciendo click en el triangulo gris frente a ella. Luego seleccione “owl:Class”; por ultimo, al igual que en el paso 4, escoja la opción “rdfs:label” del menú “Display slot”, ver figura 87.

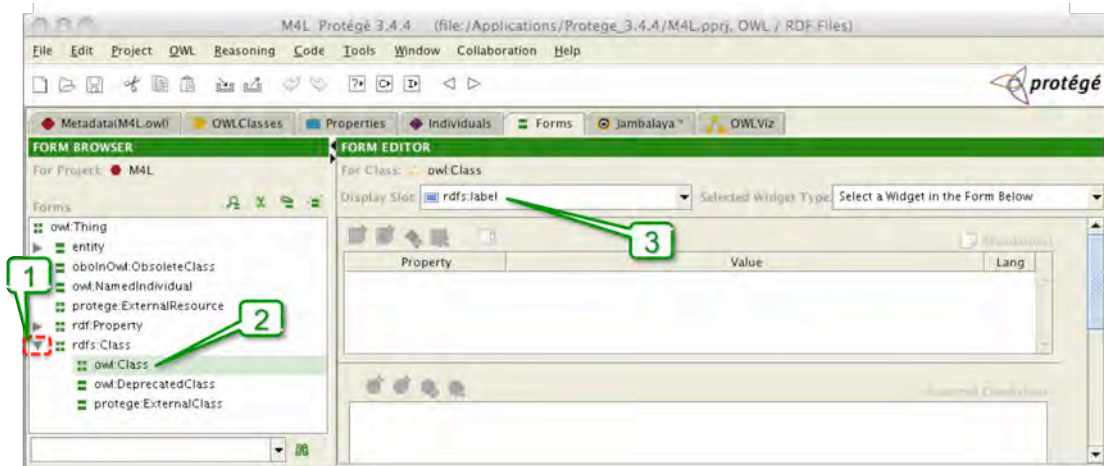


Figura 87. pasos a seguir para visualizar los términos de una ontología. Quinta parte.

6. Ahora despliegue las ramas del árbol “rdf:Property” como se hizo en el paso 5. A continuación, seleccione “rdf:Property” y finalmente escoja la opción “rdfs:label” del menú “Display slot”, ver figura 88.

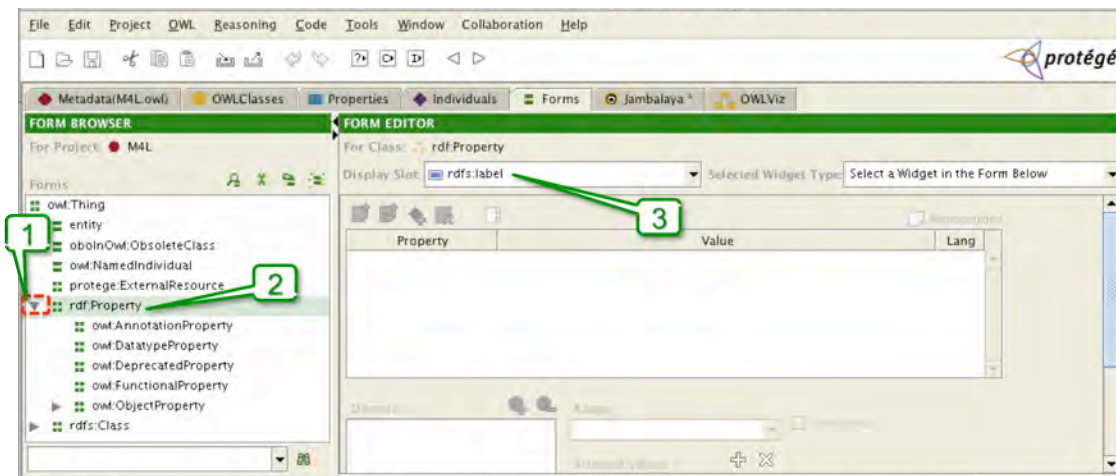


Figura 88. Pasos a seguir para visualizar los términos de una ontología. Sexta parte.

7. A partir de ahora podrá visualizar los términos de las ontologías que desee importar, en lugar de los identificadores numéricos.

- Como importar ontologías. Esta sección es una guía para importar ontologías usando Protégé 3.4.4 ya sea desde un lugar en internet o desde un archivo local.

Importando ontologías localizadas en internet.

1. Dentro de la pestaña “Metadata”, al lado izquierdo se encuentra la sección “ontology browser”; dentro de esta sección podrá ver el ícono “import ontology”. El primer paso para importar una ontología es dar click sobre el ícono “import ontology”. A continuación, se abrirá una nueva ventana, que muestra tres opciones de donde obtener la ontología que desea importar, seleccionar la opción “import an ontology from the web by specifying the http://...URL”. Finalmente dar click en “next”, ver figura 89.

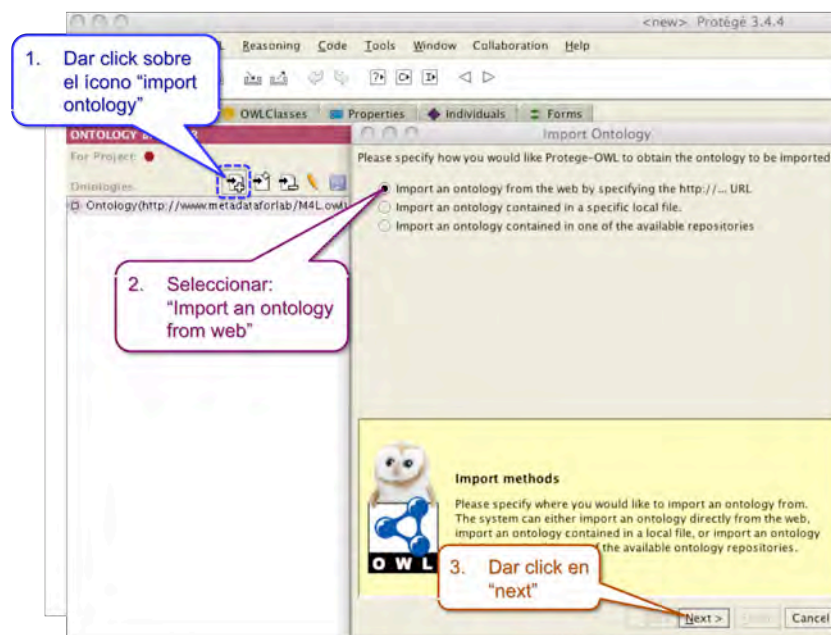


Figura 89. Pasos para importar ontologías disponibles en Internet. Primera parte.

2. Ahora deberá digitar el URL donde se encuentra la ontología que desea importar y luego seleccionar “next”, ver figura 90.

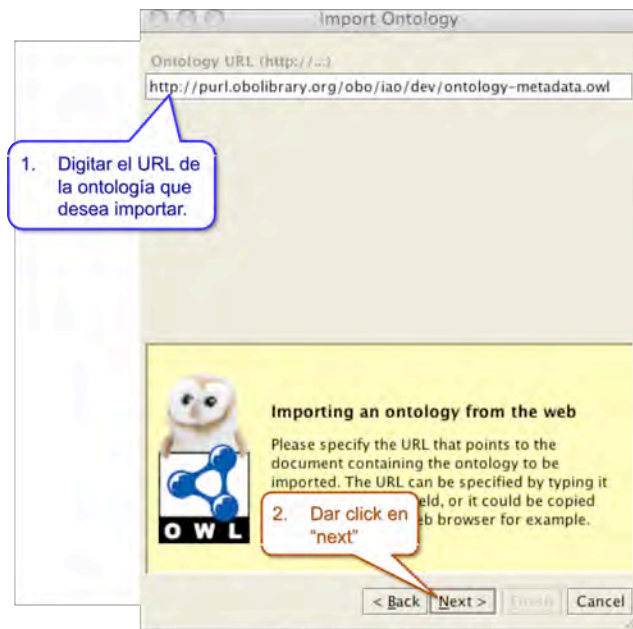


Figura 90. Pasos para importar ontologías disponibles en Internet. Segunda parte.

3. Luego, presionar “finish” para incorporar la ontología a Protégé, ver figura 91.

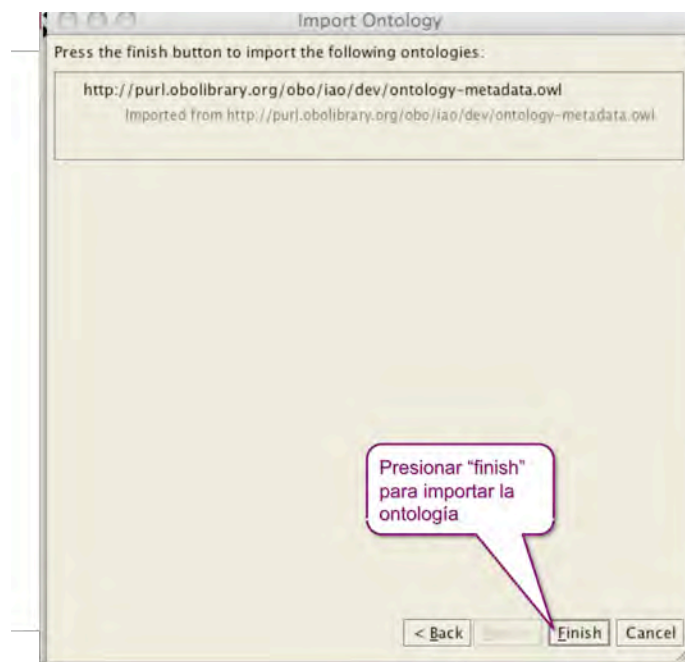


Figura 91. Pasos para importar ontologías disponibles en Internet. Tercera parte.

4. Automáticamente volverá a la pestaña “Metadata”, donde aparecerá la ontología recién importada, ver figura 92.

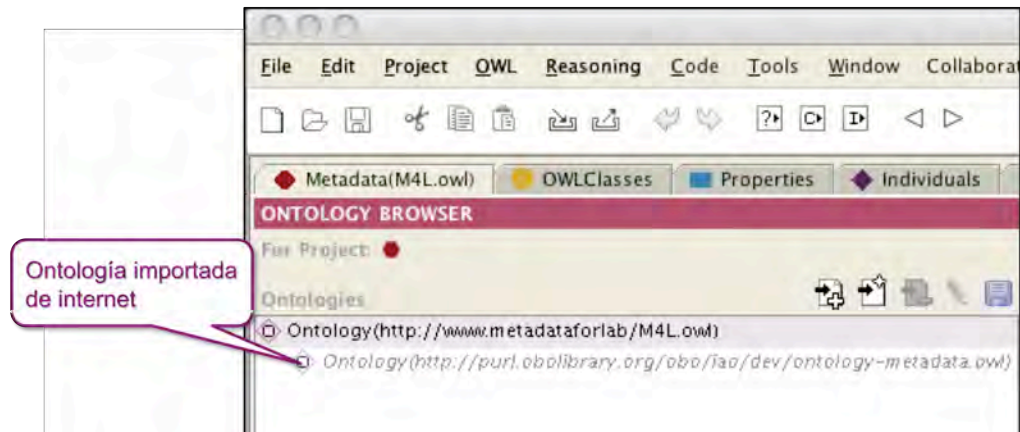


Figura 92. Pasos para importar ontologías disponibles en Internet. Cuarta parte.

Importando ontologías localizadas en un archivo local.

1. Al igual que en el paso 1 de la sección “importando ontologías localizadas en Internet”, dar click sobre el ícono “import ontology”. En la nueva ventana, seleccionar la opción “import an ontology contained in a specific local file”. Finalmente dar click en “next”, ver figura 93.

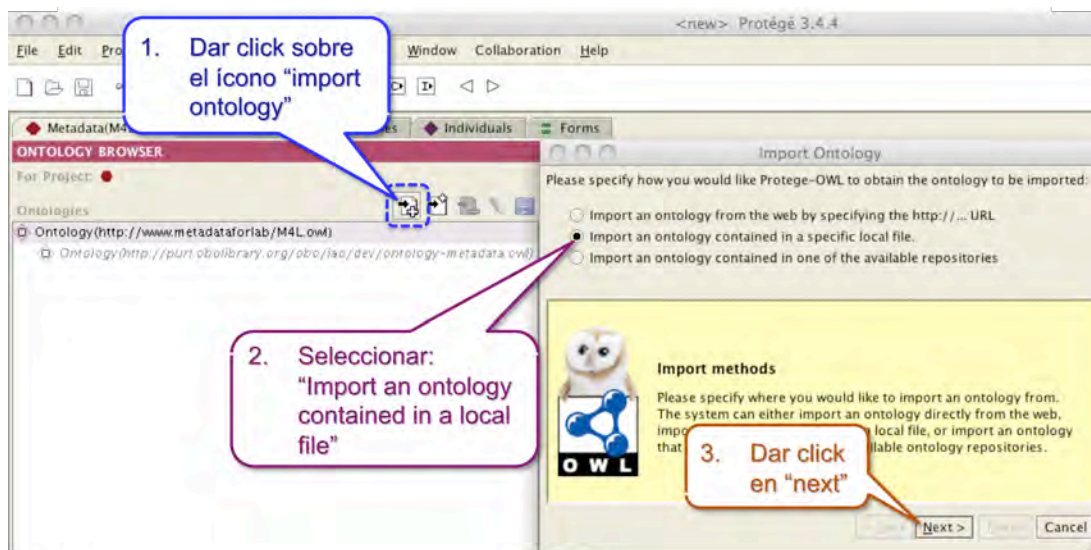


Figura 93. Pasos para importar ontologías desde un archivo local. Primera parte.

2. A continuación se abre una ventana que contiene el ícono “browse for file” (parte superior derecha de la ventana), que facilita la búsqueda de la ontología que se encuentra en un archivo local. Dar click sobre el ícono “browse for file” y seleccione el archivo OWL que desea importar, termine dando click en “select”, ver figura 94.

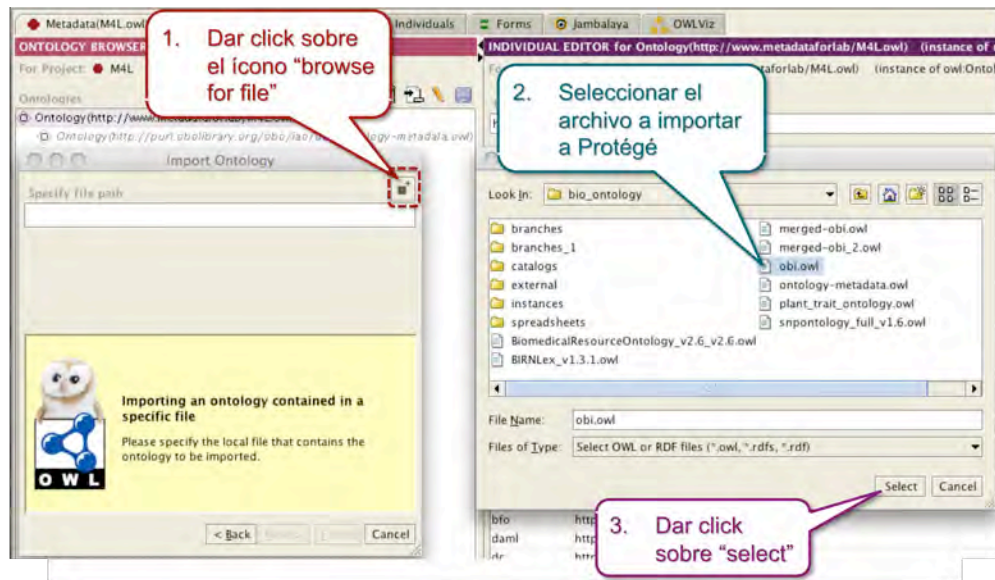


Figura 94. Pasos para importar ontologías desde un archivo local. Segunda parte.

3. De manera automática aparece la ubicación de la ontología a importar desde un archivo local. Si desea continuar con el proceso, dar click en “next”, ver figura 95.



Figura 95. Pasos para importar ontologías desde un archivo local. Tercera parte.

4. Finalmente, volverá a la pestaña “Metadata”, donde aparecerá la ontología recién importada, ver figura 96.

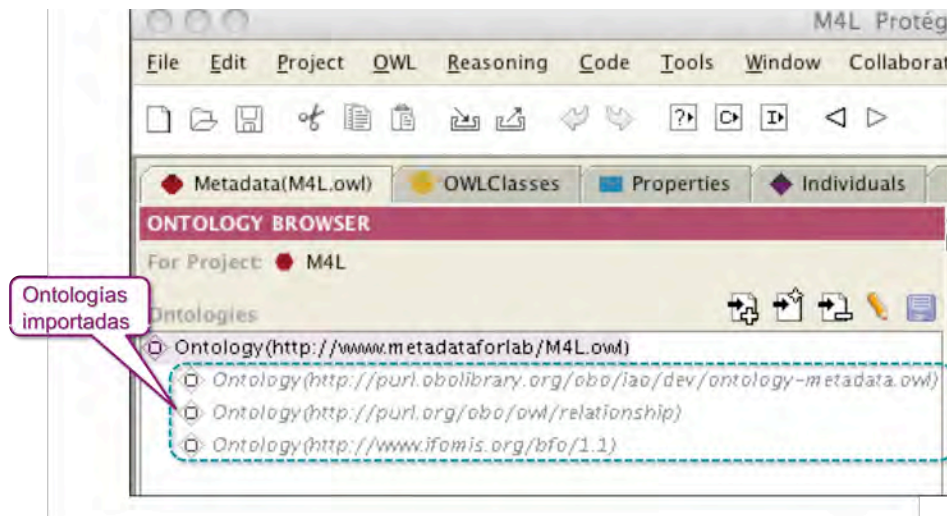


Figura 96. Pasos para importar ontologías desde un archivo local. Cuarta parte.

- Enumerar términos importantes para la ontología. Es útil escribir una lista con los términos que quisiéramos incluir en la ontología. ¿cuáles son los términos de los cuales quisiéramos hablar?, ¿Qué propiedades tienen esos términos?. Por ejemplo, términos importantes relativos a procedimientos experimentales como extracción de ácidos nucleicos incluirán: tipos de DNA, tipos de RNA, protocolos empleados, equipos y reactivos usados; la función o papel que cumplen tanto los ácidos nucleicos, como los equipos, reactivos y demás insumos que participan en este tipo de procedimiento experimental, etc.