

UNIVERSITÀ DEGLI STUDI DI FIRENZE
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica



Tesi di Laurea in Informatica

ONTOLOGIE PER LE NEUROSCIENZE:
HUMAN BRAIN PROJECT

LORENZO MASSAI

Relatore: *Elena Barucci*

Co-relatore: *Paolo Nesi*

Anno Accademico 2011-2012

Lorenzo Massai: *Ontologie per le neuroscienze: Human Brain Project*,
Corso di Laurea in Informatica, © Anno Accademico 2011-2012

INDICE

1	LA MODELLAZIONE DELLA CONOSCENZA	1
1.1	Cos'è un'ontologia	1
1.2	Il modello RDF	2
1.3	Inferenza e query	4
1.4	Ontologie per il web semantico: OWL	6
1.4.1	Sintassi	6
1.4.2	Tipologie di OWL	10
1.5	Protégé	11
2	LO STATO DELL'ARTE NELLE NEUROSCIENZE	13
2.1	Principali architetture	14
2.1.1	Ontologie anatomico-funzionali: NIF	15
2.1.2	Ontologie genetiche: Gene Ontology	21
2.1.3	Ontologie istologiche: SAO (BFO)	26
2.2	Unificazione del lessico	30
2.2.1	Neurolex	30
2.2.2	NeuroNames	32
2.3	Ontologie	34
2.3.1	Ontologie anatomiche: ABA	34
2.3.2	Ontologie di indagine: BIRNlex	37
2.3.3	Ontologie funzionali: Senselab	39
2.3.4	Ontologie murine: MGI	41
2.3.5	Ontologie di patologie: MPATH	42
2.3.6	Ontologie fenotipiche: PATO	43
3	INTEGRAZIONE DEI LIVELLI	47
3.1	Moduli di coordinamento	49
3.2	Lessici	50
3.3	Conoscenza di dominio	51
4	UN ESEMPIO D'USO	55
4.1	Mouse Ontology	55
4.1.1	Proprietà	58
5	CONCLUSIONI E SVILUPPI SUCCESSIVI	65
A	APPENDICE: ESTRATTI DALLE ONTOLOGIE	69
	BIBLIOGRAFIA	77

ELENCO DEGLI ACRONIMI

ABA	Allen Brain Atlas	MGI	Mouse Genome Informatics
AGEA	Anatomic Gene Expression Atlas	MPATH	Mouse PATHology
BIRN	Biomedical Informatics Research Network	MRI	Magnetic Resonance Imaging
BFO	Basic Formal Ontology	NIF	Neuroscience Information Framework
CCDB	Cell Centered Database	NIFSTD	NIF STanDard ontology
CHEBI	Chemical Entities of Biological Interest	NINDS	National Institute of Neurological Disorders and Stroke
DTI	Diffusion Tensor Imaging	OBO	Open Biomedical Ontologies
EAV/CR	Entity Attribute Values, Classes and Relationships	OBO-RO	OBO-Relation Ontologies
EMAP	Edinburgh Mouse Atlas Project	ORDB	Olfactory Receptor Database
fMRI	functional Magnetic Resonance Imaging	PATO	Phenotypic And Trait Ontology
GO	Gene Ontology	RDF	Resource Description Framework
HBP	Human Brain Project	RDFS	RDF Schema
INCF	International Neuroinformatics Coordinating Facility	SAO	Subcellular Anatomy Ontology
IUPHAR	International Union of PHARmacology	OWL	Web Ontology Language
LENS	European Laboratory for Non-linear Spectroscopy	URI	Universal Resource Identifier
MA	Mouse Anatomy	UML	Unified Modeling Language
MeSH	Medical Subject Headings	XML	eXtensible Markup Language

INTRODUZIONE

Questo lavoro di tesi si colloca all'interno del progetto Human Brain (HBP), il cui obiettivo è la comprensione delle caratteristiche anatomiche e funzionali del cervello umano e di come esse siano interconnesse a formare i meccanismi cerebrali. La sezione *Projectome* del progetto HBP è volta alla realizzazione di un atlante interattivo tramite il quale sia possibile esplorare il cervello a più livelli di granularità, a partire dal livello macro-anatomico fino a quello neuronale. I progressi di *Projectome* ad oggi riguardano principalmente l'indagine microscopica a livello neuronale e si basano sugli esperimenti condotti al Laboratorio Europeo di Spettroscopia Non-lineare (L.E.N.S.), attraverso il quale si hanno a disposizione immagini di neuroni ottenute tramite il microscopio confocale *Conf-U* in seguito all'applicazione di contrasti specifici e alla colorazione dei tessuti afferenti a tutte le zone cerebrali di un topo adulto [17]; i metodi di indagine neuroscientifica attuale consentono inoltre l'esplorazione dell'aspetto funzionale del cervello attraverso risonanze magnetiche mirate (*fMRI*) basate sull'iniezione di contrasti specifici con cui è possibile evidenziare la reazione dei circuiti neuronali ed accedere a un tipo di conoscenza che trascende quella anatomica.

L'enorme quantità di informazioni derivanti dal L.E.N.S. e dall'indagine funzionale rende impossibile la definizione manuale di tutte le caratteristiche dei neuroni, la cui stima in un cervello umano è di $\sim 80 - 120 \times 10^9$ per $10^{14} - 10^{15}$ sinapsi, mentre si ha un numero di neuroni inferiore di almeno tre ordini di grandezza in un cervello di topo. Per gestire l'ingente quantità di informazioni è necessario studiare un approccio che non sia volto all'aspetto quantitativo dei dati, ma a quello qualitativo; si pone quindi il problema di definire le regole secondo le quali le parti si coordinano e di dare luogo a una struttura per i dati che dovrà essere così eterogenea da accogliere i diversi tipi di informazioni che costituiscono la contestualizzazione di un'entità anatomica.

Lo schema che meglio si adatta alla descrizione formale di un dominio di interesse è quello delle *ontologie*, che costituiscono una concettualizzazione gerarchica della struttura del dominio. Attraverso un

modello ontologico è quindi possibile sovrapporre i livelli di conoscenza associando alle immagini del *L.E.N.S.* i metadati di contesto (atlante anatomico, immagini di risonanza, conoscenza di dominio) ed esplicitando le relazioni tra le parti in modo da rendere la struttura esplorabile semanticamente.

In un contesto così ampio è necessario identificare i ruoli dei gruppi che partecipano all'astrazione della conoscenza di dominio e il modo in cui essi si coordinano per garantire una ricerca inter-operante tra i livelli di conoscenza; si vanno dunque ad analizzare gli assetti secondo i quali i principali enti di ricerca organizzano le informazioni e gestiscono il rapporto tra i domini specifici.

Essendo le ontologie un sistema che si presta all'estensione e alla modularizzazione, è possibile fare uso di strutture specifiche di più domini di interesse per coprire più aspetti della conoscenza associata a ogni neurone; sono volte a tale scopo la maggior parte delle ontologie neuroscientifiche, delle quali in questo elaborato si studia la struttura, il funzionamento e il contributo agli ambiti della conoscenza di dominio più vicini agli scopi del progetto. Tali ambiti riguardano tutte le qualità che caratterizzano una cellula (neurone) e la sua contestualizzazione all'interno dell'organismo; sono quindi analizzate quelle strutture che ricoprono l'aspetto anatomico, funzionale, dei metodi sperimentali, delle patologie, dei fenotipi e della genetica.

Per consentire la collaborazione tra le ontologie è necessario che i moduli indipendenti che fanno riferimento alle stesse caratteristiche siano messi in relazione; a tale scopo una sezione della tesi è volta allo studio degli strumenti e delle tecniche utilizzate dalle principali ontologie per sopperire alla mancanza di un lessico unificato all'interno della letteratura neuroscientifica.

Dopo aver appurato quali sono i modelli di riferimento, gli obiettivi che essi si propongono e come sono gestiti i livelli di conoscenza, si ha una visione d'insieme che consente di definire quali ambiti possano costituire un'integrazione delle risorse di cui il progetto dispone e quali sono le esigenze in relazione agli obiettivi del progetto.

Avendo individuato e definito il tipo di conoscenza ricoperto dalle ontologie di contesto, si propone un modello ontologico iniziale che integra le conoscenze settoriali coperte dalle ontologie esposte nella tesi ed estende quelle relative alle tecniche di indagine con i metadati relativi al microscopio *Conf-U* e al protocollo di preparazione del campione seguito al *L.E.N.S.* nel corso degli esperimenti.

L'elaborato è suddiviso in 5 capitoli:

primo capitolo: Analisi degli strumenti per la modellazione della conoscenza. Concetti fondamentali di ontologia, metadati e inferenza, definizione di relazioni logiche tra entità, linguaggi per l'implementazione di un'ontologia;

secondo capitolo: Contestualizzazione dell'approccio ontologico all'interno del progetto HBP, valutazione dello stato dell'arte nelle ontologie neuroscientifiche, struttura e funzionamento delle principali ontologie, strategie per l'unificazione del lessico, analisi delle ontologie il cui dominio di conoscenza riflette gli obiettivi del progetto;

terzo capitolo: Integrazione dei moduli ontologici, prospetti, diagrammi delle competenze, coordinazione delle ontologie, copertura della conoscenza di contesto;

quarto capitolo: Proposta di un modello iniziale che integra gli aspetti inerenti a Human Brain Project in base agli esperimenti svolti al L.E.N.S.

quinto capitolo: Conclusioni e sviluppi successivi;

appendice: Estratti dalle ontologie analizzate nella tesi.

LA MODELLAZIONE DELLA CONOSCENZA

1.1 COS'È UN'ONTOLOGIA

L'ontologia è quella parte della metafisica che tratta dell'essere come tale, considerato in se' stesso, indipendentemente dai suoi modi di manifestarsi.

— Accademia della Crusca [1]

In filosofia l'ontologia è la scienza che studia l'essenza (dal greco ὄντος - ente, λόγος - discorso) ovvero ciò che contraddistingue un essere da un altro. La sua definizione in ambito informatico non si discosta particolarmente dal significato etimologico: in informatica ci si riferisce infatti a un'ontologia come a una descrizione esplicita e formale dei concetti in un dominio di rappresentazione, le proprietà che ne descrivono gli attributi e i vincoli su di esse; attraverso questi strumenti è possibile fornire una descrizione schematica delle entità che costituiscono l'ambito di rappresentazione e delle relazioni che intercorrono tra esse. Lo scopo di una modellazione ontologica non si limita alla mera collezione di dati, ma si occupa della loro descrizione tramite la definizione di un linguaggio formale attraverso il quale saranno successivamente definiti i dati; attraverso questa infrastruttura si definisce un prototipo che consente:

- la condivisione delle informazioni
- il riuso della conoscenza di dominio
- la rappresentazione *strutturale* della conoscenza
- la distinzione tra la conoscenza di dominio e quella relazionale
- la navigazione della conoscenza di dominio

La *condivisione delle informazioni* è facilitata dall'uso di un modello comune. Informazioni provenienti da ambiti differenti fanno riferimento alla stessa struttura e sono quindi organizzate allo stesso modo.

Il *riuso della conoscenza di dominio* consiste nella possibilità di estrarre da un'ontologia già esistente parti della struttura che si adattino ad un dominio differente. Allo stesso modo ontologie che fanno riferimento a domini distinti possono essere integrate per dare luogo ad una ontologia più generale.

La *rappresentazione strutturale della conoscenza* consente di effettuare cambiamenti alla conoscenza di dominio senza alterare la struttura dell'ontologia. La separazione tra i due livelli permette di operare cambiamenti sui dati mantenendo la stessa architettura.

La *distinzione tra la conoscenza di dominio e quella relazionale* esplicita la distinzione tra i livelli di conoscenza. Descrivere un modello per rappresentare le entità e relazioni coinvolte rende possibile l'integrazione con strumenti dedicati alla definizione dei dati nel dominio.

La *navigazione della conoscenza di dominio* è consentita solo in seguito alla specifica dei dati nell'ontologia. E' possibile eseguire una navigazione tra le entità attraverso le proprietà *esplicite* (definite) e quelle *implicite*, facendo inferenza sulle relazioni esplicite tra le classi.

La necessità principale che si presenta in una descrizione tipo ontologico è quindi quella di poter definire le parti che contribuiscono alla descrizione del contesto da rappresentare e il modo in cui sono correlate; dal punto di vista informatico non è quindi sufficiente limitarsi alle primitive messe a disposizione dai linguaggi di programmazione, ma è necessario utilizzare un *metalinguaggio*, ovvero uno strumento che consenta di definire in modo specifico la tipologia di ogni figura che compare nell'ontologia e il suo ruolo in relazione con le altre.

1.2 IL MODELLO RDF

Lo strumento che meglio si adatta alla modellazione della conoscenza è il *Resource Description Framework (RDF)*, che si prepone gli stessi obiettivi di una descrizione ontologica astratta, fornendo strumenti computazionali analoghi a quelli di una schematizzazione ontologica per l'organizzazione dei metadati [2]; in *RDF* le entità sono rappresentate da *risorse* legate tra loro da *relazioni* che ne esprimono la struttura gerarchica. Un'informazione rappresentata tramite *RDF* è tipicamente una tripla soggetto-predicato-oggetto in cui il soggetto è rappresentato da una risorsa, il predicato è una relazione e l'oggetto è un valore di tipo primitivo (intero, stringa, ecc.) oppure un'altra

risorsa; ogni oggetto puo' costituire un riferimento ad un'altra risorsa, assegnando come predicato della tripla un puntatore (ad esempio una stringa) ad un'altra entità. Il modello *RDF* si adatta alla condivisione di informazioni in rete; in questo caso ogni risorsa corrisponde ad un *URI* (Universal Resource Identifier), ovvero ad un puntatore alla posizione di quella risorsa nella rete. Lo scopo del modello è di definire la struttura secondo la quale le informazioni sono correlate, ma la specifica funzione che le relazioni hanno all'interno dell'ontologia è ancora incognita a questo livello di astrazione, in quanto *RDF* è indipendente dal dominio; a tale scopo è infatti volto *RDF Schema (RDFS)*, tramite il quale è possibile istanziare le classi e definire la funzione delle relazioni all'interno della struttura.

Possiamo rappresentare un esempio di asserzione secondo lo schema *RDF* tramite un grafo orientato in cui i nodi rappresentano il soggetto e l'oggetto e gli archi orientati rappresentano i predicati:

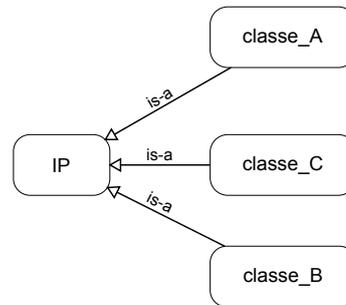


Figura 1.1: Grafo di esempio per rappresentare classi e le relazioni tra esse

Utilizzando i link *URI* l'informazione sulla posizione della risorsa nella rete è un metadata associato all'entità ed è indicato come una proprietà che la descrive; in figura 2.2 gli *URI* compaiono nel grafo come nomi di entità a scopo esemplificativo.

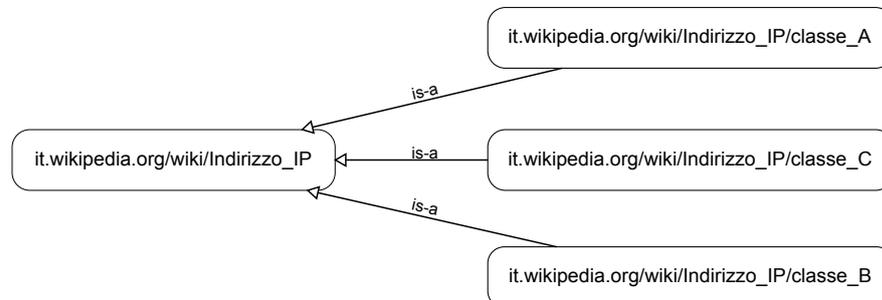


Figura 1.2: Grafo di esempio con URI

I limiti principali a livello espressivo di *RDF* e *RDFS* riguardano:

relazioni specifiche - l'unica relazione consentita in *RDFS* è quella di sottoclasse, quindi non è possibile definire ad esempio il rapporto che lega due sottoclassi di una classe.

combinazione logica di classi - non è possibile definire classi come composizione o partizione di altre.

restrizioni globali - le restrizioni applicabili alle relazioni in *RDF* riguardano tutte le classi del codominio e non è possibile definirle per un sottoinsieme di esse.

restrizioni sulla quantità di istanze - non è possibile limitare il numero di istanze che deve avere una classe.

proprietà funzionali delle relazioni - non è possibile definire una relazione biunivoca tra due o più classi, ad esempio la relazione tra due classi non può essere transitiva o inversa.

1.3 INFERENZA E QUERY

La differenza principale tra un'ontologia e un database si può individuare nell'obiettivo che si prepongono: un database è volto alla descrizione delle istanze di un ambito di rappresentazione e la sua modellazione è strettamente legata ad esse; un'ontologia è invece orientata alla descrizione delle entità astratte, quindi alla coordinazione tra di esse all'interno della struttura. L'approccio ontologico permette di fare inferenza sulle relazioni che legano le entità, quindi di estrarre *nuova conoscenza* da quella "dichiarata"; da un punto di vista computazionale questa possibilità si ha applicando a un'ontologia un *ragionatore*, tramite il quale è possibile valutare la consistenza delle relazioni tra le classi e mettere in relazione entità che non sono esplicitamente correlate in base alla definizione di *regole di inferenza* che rispecchiano la visione di chi plasma l'ontologia. Successivamente al popolamento dell'ontologia con istanze e dopo la definizione delle relative regole di inferenza si ha una *base di conoscenza*.

Il motivo per cui si sceglie di dare una struttura di tipo ontologico alla conoscenza è dare la possibilità di interrogare la struttura sulle relazioni inferite, ad esempio su:

appartenenza di classe - Se una classe B ha una istanza x e tale classe è sottoclasse di una classe A , si può inferire che x sia un'istanza di A , anche se non è esplicitamente dichiarato.

equivalenza di classi - Tra classi vale la proprietà transitiva: Se una classe A è dichiarata equivalente ad una classe B e la classe B è dichiarata equivalente alla classe C , possiamo inferire l'equivalenza tra le classi A e C .

consistenza - Dichiarando le classi A, B, C, D tali che A è sottoclasse di $B \cap C$ e allo stesso tempo di D , se x è un'istanza di A e $B \cap D = \emptyset$, si ha un errore di consistenza, ovvero un errore nell'ontologia, poiché la classe A non dovrebbe avere alcuna istanza.

classificazione - Se una classe A è univocamente determinata dalla presenza di una o più relazioni, possiamo inferire che qualsiasi classe o istanza che presenti tali relazioni sia afferente alla classe A (Figura 1.3).

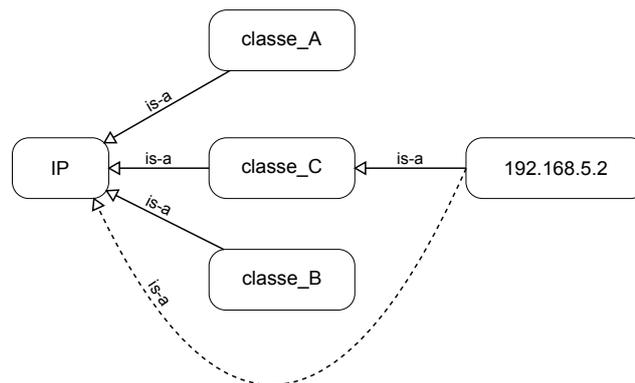


Figura 1.3: Grafo di esempio per l'inferenza sull'appartenenza di classe

Nel grafo in figura 1.3 l'istanza 192.168.5.2 dell'entità classe_C non è esplicitamente dichiarata come istanza dell'entità IP, ma tramite un ragionatore questa relazione diventa esplicita (arco tratteggiato); allo stesso modo se l'istanza 192.168.5.2 fosse un'istanza di IP potremmo dedurre la classe di IP a cui appartiene definendo come restrizione di classe_C lo spettro di valori 192 – 223 per il primo otetto [3], condizione soddisfatta dall'istanza e quindi sufficiente per inferirne l'appartenenza a classe_C (classificazione).

Tale organizzazione delle informazioni è navigabile tramite la logica dei predicati del primo ordine, che consente di individuare tutte le

componenti che l'ontologo ha modellato (classi, relazioni e vincoli) e le relazioni inferite attraverso il ragionatore.

La limitata capacità espressiva di *RDF* ha il vantaggio di consentire un'analisi efficiente dal punto di vista del ragionatore e quindi dell'inferenza, ma ha il forte svantaggio di avere un vocabolario estremamente contenuto per descrivere le proprietà che legano le entità; a tale scopo è volto *Web Ontology Language (OWL)*, un'estensione di *RDF* che a discrezione delle necessità espressive o computazionali dell'ontologo consente di utilizzare un vocabolario più o meno elaborato.

1.4 ONTOLOGIE PER IL WEB SEMANTICO: OWL

Il *Web Ontology Language (OWL)* è un metalinguaggio che aggiunge costrutti a quelli già presenti in *RDF* e ne aumenta la capacità descrittiva. L'utilizzo di *OWL* consente una maggiore e più accurata espressività rispetto alle strategie precedenti, integrandole a formare lo standard *de facto* per la modellazione ontologica.

1.4.1 Sintassi

La sintassi di *OWL* è basata sullo standard *XML* e può essere costruita a partire dalle regole definite per la descrizione con *RDF* oppure tramite un grafo orientato che segue lo schema *UML* per i linguaggi a oggetti con cui è immediato rappresentare le entità e le istanze come nodi e le relazioni e restrizioni tramite archi annotati. Ogni espressione è della forma: `vocabolario:tipo`, quindi utilizzando ad esempio in un documento *OWL* l'espressione `rdfs:label` si indica che il vocabolario a cui fa riferimento il tipo dichiarato è quello di *RDFS*.

1.4.1.1 Classi

In *OWL* ogni entità è dichiarata tramite `owl:Class` ed estende la generica classe *Thing* (`owl:Thing`) dalla quale vengono generate tutte le entità dell'ontologia; per ogni entità è inoltre presente una generica sottoclasse *Nothing* (`owl:Nothing`) che la estende; la relazione di inclusione tra queste entità è espressa da `rdfs:subClassOf` e, come in *RDFS*, la sottoclasse eredita tutte le proprietà della sopraclasse. La dichiarazione esplicita di una risorsa come classe (lo stesso vale per le

proprietà e le istanze) si ottiene tramite `rdf:type`, seguito dal tipo di risorsa.

1.4.1.2 Proprietà

A differenza di *RDF*, in cui una proprietà generica è espressa da `rdf:Property`, in *OWL* è possibile definire due tipi di proprietà tra classi: le *proprietà degli oggetti* (`owl:ObjectProperty`) e le *proprietà dei dati* (`owl:DatatypeProperty`) che mettono in relazione rispettivamente risorse con altre risorse e risorse con valori di tipo primitivo; il dominio di una proprietà è espresso da `rdfs:domain` e il codominio da `rdfs:range`. Come per le entità, anche per le proprietà è possibile definire una struttura gerarchica dichiarando sottoproprietà (`rdfs:subPropertyOf`) e relazioni di equivalenza tra esse.

1.4.1.3 Proprietà di classi

OWL consente di individuare classi che non sono esplicitamente definite da un'entità, ma sono univocamente determinate dalla combinazione booleana di altre già esistenti; avendo quindi definito ad esempio le classi `classe_A`, `classe_B` e `classe_C` come tutte e sole le classi di indirizzi IP, l'unione delle tre classi dà luogo alla classe IP. Nello specifico possiamo considerare:

unione (`owl:unionOf`):

$$\text{IP} = \text{classe_A} \cup \text{classe_B} \cup \text{classe_C}$$

intersezione (`owl:intersectionOf`):

$$\text{classe_C} = \text{IP} \cap \neg(\text{classe_A} \cup \text{classe_B})$$

complemento (`owl:complementOf`):

$$\neg(\text{classe_A} \cup \text{classe_B} \cup \text{classe_C}) \cap \text{IP} = \emptyset$$

1.4.1.4 Proprietà di relazioni

Per ogni relazione tra istanze si può specificare il rapporto tra le entità correlate, quindi definire le proprietà di una relazione *R* sull'insieme *I* delle istanze:

proprietà simmetrica (`owl:SymmetricProperty`):

$$\forall x, y \in I, x R y \Rightarrow y R x$$

proprietà transitiva (`owl:TransitiveProperty`):

$$\forall x, y, z \in I, x R y \wedge y R z \Rightarrow x R z$$

proprietà riflessiva (owl:ReflexiveProperty):

$$\forall x \in I, x R x$$

proprietà antisimmetrica (owl:AsymmetricProperty):

$$\forall x, y \in I, x R y \Rightarrow \neg(y R x)$$

proprietà antiriflessiva (owl:IrreflexiveProperty):

$$\forall x, y \in I, x R y \Rightarrow x \neq y$$

proprietà funzionale (owl:FunctionalProperty):

$$\forall x \in I, y, z \in I \cup \text{Letterali}, x R y \wedge x R z \Rightarrow y = z$$

proprietà inversa (owl:InverseFunctionalProperty):

$$\forall x, y, z \in I, x R z \wedge y R z \Rightarrow x = y$$

1.4.1.5 Restrizioni sulle proprietà

Il cambiamento più rilevante rispetto a *RDF* è quello di poter definire *restrizioni* specifiche sulle proprietà, ovvero di considerare solo determinati valori del dominio di una relazione; si ottiene questo risultato considerando la relazione esistente tra due classi come se fosse divisa in due relazioni: la prima tra il dominio originale e una sua sottoclasse e la seconda tra la nuova sottoclasse e il codominio originale. Ad esempio se un indirizzo IP è un numero intero e un indirizzo IP di classe *C* è un numero intero le cui prime tre cifre sono comprese tra 192 e 223, la relazione tra *classe_C* e il suo valore può essere definita come una restrizione sulla relazione tra IP e un suo valore che abbia le prime tre cifre tra 192 e 223; possiamo quindi applicare una prima funzione (la restrizione) per le prime tre cifre tra il dominio originale (\mathbb{Z}) e una sua sottoclasse (192, ..., 223) e applicare una seconda funzione (la relazione) tra l'insieme (192, ..., 223) e il codominio originale.

Le restrizioni si dichiarano tramite owl:Restriction su proprietà (owl:onProperty) o classi (owl:onClass) e le più significative sono quelle che consentono di selezionare una parte del dominio in base a quantificatori o alla cardinalità:

restrizione universale (owl:allValuesFrom): limita il codominio della relazione solo a quello specificato.

restrizione esistenziale (owl:someValuesFrom): esprime che la classe specificata fa parte del codominio della relazione.

restrizione di appartenenza (`owl:hasValue`): esprime il valore specifico che assumono gli elementi del dominio a cui sia applicata la relazione.

restrizioni booleane (`owl:unionOf`, `owl:intersectionOf`, `owl:complementOf`): esprime un rapporto booleano tra classi, del tipo $A \vee B$, $A \wedge B$, $\neg A$.

restrizioni di cardinalità (`owl:minCardinality`, `owl:maxCardinality`, `owl:cardinality`): esprime il numero minimo, massimo o esatto di elementi del codominio che gli elementi del dominio a cui sia applicata la relazione devono assumere.

enumerazione (`owl:oneOf`): consente di definire una classe elencando i suoi elementi.

disgiunzione (`owl:disjointWith`): l'esplicita dichiarazione della disgiunzione tra due classi consente a un ragionatore di individuare le inconsistenze.

1.4.1.6 Istanze

L'istanza di una classe è dichiarata tramite `rdf:type` e ogni istanza è unica, a meno che non sia specificata l'uguaglianza con altre tramite `owl:sameAs`; due istanze con la stessa etichetta risultano quindi distinte in seguito a un'interrogazione dell'ontologia; l'attributo che identifica univocamente una risorsa rispetto a un'altra è la *chiave*, il cui valore può essere specificato tramite `owl:hasKey` e deve essere unico. La distinzione tra le classi può essere resa esplicita per precauzione tramite `owl:differentFrom`.

1.4.1.7 Annotazioni

Un tipo particolare di proprietà sono le *annotazioni*, ovvero metadati associati a classi, proprietà e istanze disgiunte. Le annotazioni si dichiarano con l'espressione `owl:AnnotationProperty` e consentono di specificare:

etichetta (`rdfs:label`): associa alla risorsa una stringa di testo che ne identifica il nome.

commento (`rdfs:comment`): associa alla risorsa una stringa di testo.

definizione (`rdfs:isDefinedBy`): associa alla risorsa la sua descrizione, che può essere locale oppure un URI.

definizione separata (`rdfs:seeAlso`): indica che la risorsa è descritta in un'altra sezione dell'ontologia

1.4.2 Tipologie di OWL

L'estensione di *RDF* con i costrutti di *OWL* garantisce un'alta capacità espressiva, ma lo scopo per cui un'ontologia viene sviluppata è principalmente quello di poterla analizzare tramite il ragionatore per ottenere nuova conoscenza da quella esplicita e questo compito può risultare oneroso dal punto di vista computazionale se il "vocabolario" del linguaggio è troppo articolato; per questo motivo la modellazione di un'ontologia tramite *OWL* può avvenire attraverso tre sotto-linguaggi, ognuno dei quali limita il precedente, ma garantisce una maggiore efficienza.

Il rapporto di inclusione dei tre linguaggi è schematizzato come segue:

$$\text{OWL Lite} \subseteq \text{OWL DL} \subseteq \text{OWL Full}$$

1.4.2.1 OWL Lite

La forma più semplificata di *OWL* è *OWL Lite*, che garantisce una complessità inferiore rispetto alle varianti *DL* e *Full*, ma impone limiti a livello delle relazioni e quindi dell'espressività del linguaggio. La struttura di *OWL Lite* è volta principalmente alla classificazione gerarchica delle informazioni e limita il vocabolario standard in quanto:

- Non sono permesse le restrizioni di appartenenza, enumerazione e booleane.
- Tutte le restrizioni di cardinalità non possono avere valori diversi da 0 e 1.
- L'equivalenza di classi (`owl:equivalentClass`) è permessa solo tra identificatori di classe.

1.4.2.2 *OWL DL*

Il linguaggio più utilizzato per la descrizione formale di ontologie è *OWL DL* (Description Logic) in quanto consente un'alta espressività con costi computazionali accettabili (tutti i risultati di inferenza sono computabili e la complessità è finita). Questa variante include tutto il vocabolario *OWL* ma presenta dei limiti a livello di risorse, proprietà e relazioni:

- Ogni risorsa deve essere esclusivamente una tra: classe, dati, proprietà di dati, proprietà di classe, istanza, valore.
- La distinzione e la definizione di ogni risorsa deve essere esplicita, come anche il suo ruolo nell'ontologia.
- Le proprietà di relazioni non sono applicabili ai dati, perchè le proprietà di classe sono distinte dalle proprietà di dati.
- Non sono consentite restrizioni di cardinalità per le proprietà definite come transitive.
- Le uniche classi a cui è consentito non avere una definizione sono quelle derivanti dall'equivalenza, dalla disgiunzione e quelle che sono il codominio di una relazione di sottoclasse.

1.4.2.3 *OWL Full*

Il linguaggio con la maggiore espressività è *OWL Full*, ma la varietà dei costrutti che utilizza non consente un'implementazione completa quindi anche la possibilità di fare inferenza è limitata. *OWL Full* può essere visto come un'estensione di *RDF* in quanto permette l'espansione del vocabolario *OWL*.

1.5 PROTÉGÉ

Protégé è un progetto open source dell'Università di Stanford che fornisce un'interfaccia per la modellazione di ontologie tramite i linguaggi *RDF*, *RDFS*, *OWL* e *XML-Schema*; implementa tutti i sotto-linguaggi di *OWL*, ma è possibile applicare il ragionatore solo alle ontologie che rientrano nei vincoli di *OWL DL* e Lite. Protégé può essere integrato con numerose estensioni per facilitare la progettazione e l'analisi di un'ontologia: l'estensione *OWL code generation* ad

esempio fornisce la generazione automatica di codice a partire da un'ontologia modellata sotto forma di grafo orientato, *SPARQL* è uno strumento per effettuare query sulle ontologie, *Ontograf* permette la visualizzazione di un'ontologia tramite un grafo orientato e *Hermit* è un ragionatore che permette di ampliare la struttura definita in base alle regole di inferenza)

L'eterogeneità consentita da una modellazione di tipo ontologico evidenzia la necessità di costruire un modello che permetta non solo di essere esteso, ma anche di costituire un'estensione di modelli preesistenti e che non tralasci la possibilità di un'integrazione con altre ontologie; ad esempio un'ontologia che descrive le relazioni gerarchiche tra le parti anatomiche di un organismo è estendibile con una che ne descrive il livello funzionale; il risultato è un modello unico interrogabile sia dal punto di vista anatomico che funzionale in cui i due livelli descrittivi sono correlati.

Questa sezione descrive le maggiori ontologie volte all'espressione delle relazioni cerebrali e di come esse si coordinino per fornire una conoscenza il più possibile accurata; nella maggior parte dei modelli il tipo di relazione presente tra le entità è di tipo gerarchico, quindi la descrizione delle parti riguarda principalmente la formalizzazione di un atlante di riferimento e l'annotazione con metadati derivanti dalla letteratura neuroscientifica.

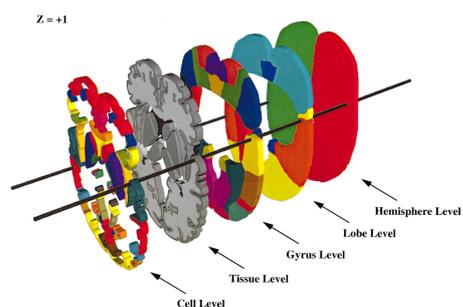


Figura 2.1a: Livello anatomico, macroscopico

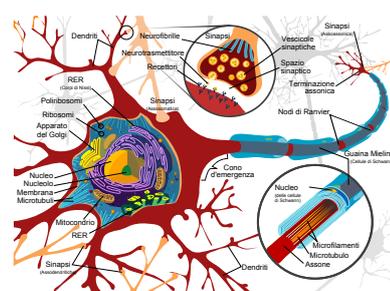


Figura 2.1b: Livello anatomico, microscopico

La profondità della gerarchia a cui i modelli anatomici descritti fanno riferimento è basato sui progressi dell'indagine microscopica attuale, che può arrivare a livelli molto alti, ma limitatamente alla zona del cervello presa in esame in ogni esperimento (ad esempio in figura 2.2b [6]), quindi la quantità di annotazioni di alto livello sarà fortemente legata al numero di osservazioni eseguite.

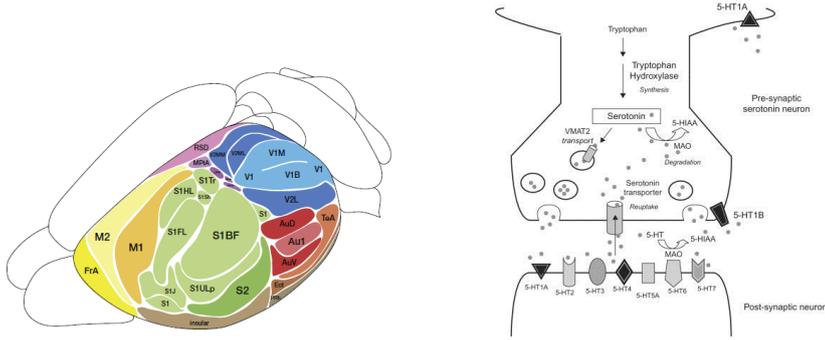


Figura 2.2a: Livello funzionale, macroscopico Figura 2.2b: Livello funzionale, microscopico

La principale innovazione che deriva dagli esperimenti condotti al *L.E.N.S.* è quella di consentire un altissimo livello di microscopia su *tutto* il cervello (figura 2.3¹), quindi da un punto di vista ontologico di consentire l'annotazione dei neuroni in ogni zona del cervello da un punto di vista sia anatomico (figura 2.1a [4], figura 2.1b) che funzionale (figura 2.2a [5], figura 2.2b).

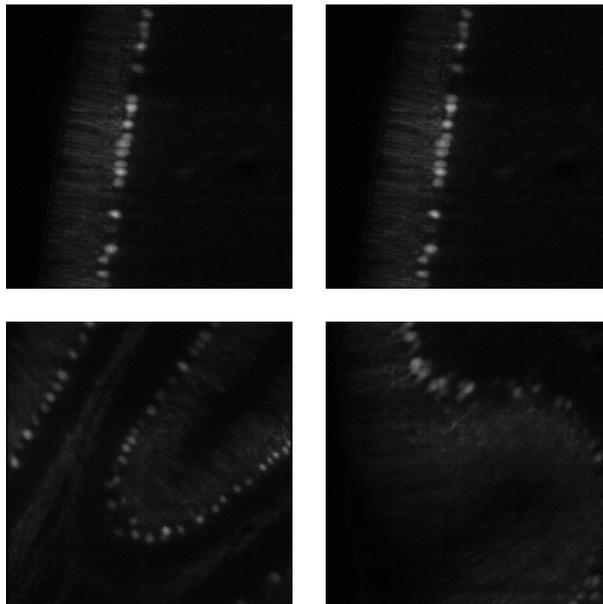


Figura 2.3: Immagini di neuroni raccolte al L.E.N.S.

2.1 PRINCIPALI ARCHITETTURE

Questa sezione è volta all'esplorazione delle strategie che sono state seguite per gestire la frammentazione delle categorie e delle strutture

¹ per gentile concessione di Ludovico Silvestri

ontologiche nei settori neuroscientifici. Per garantire l'interoperabilità dei gruppi la cui conoscenza settoriale riguarda aspetti diversi dello stesso organismo l'organizzazione delle ontologie è basata su moduli di coordinamento il cui scopo, oltre a quello di fornire un contributo alla conoscenza di dominio, è quello di proporre una struttura comune che definisca i vincoli delle ontologie partecipanti, come ad esempio i tipi di relazione consentiti, il tipo di ereditarietà, o la massima profondità della gerarchia.

2.1.1 Ontologie anatomico-funzionali: NIF

Il *Neuroscience Information Framework*² (NIF) è un'interfaccia per le risorse neuroscientifiche che consente la ricerca semantica dei dati attraverso un motore di ricerca ontologico, tramite il quale è possibile accedere contemporaneamente a diversi tipi di conoscenza, organizzata in base alle categorie di indagine. L'architettura NIF fa riferimento a quattro settori principali (figura 2.4 [10]): Neurolex, Textpresso, Ontoquest e Bioportal.

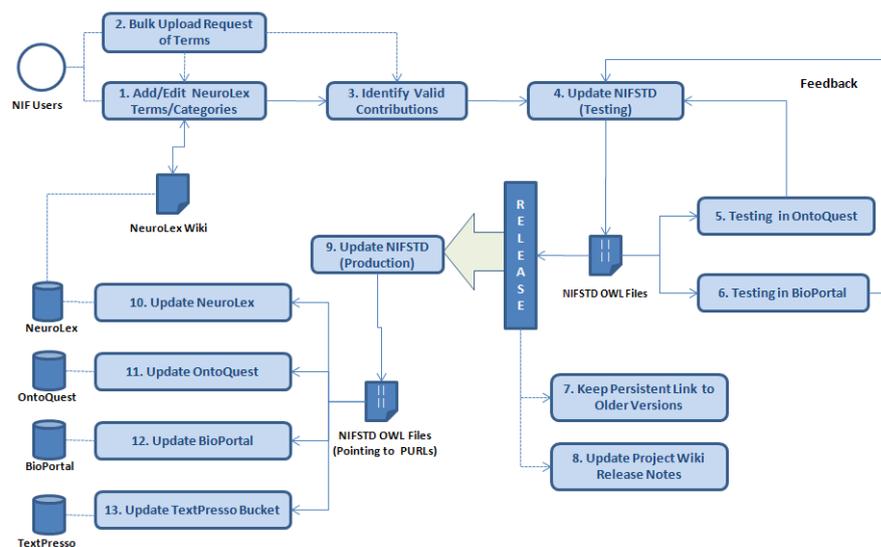


Figura 2.4: Schema dell'architettura NIF

Neurolex è la sezione relativa al lessico e contiene i termini che vengono utilizzati per costruire le asserzioni; tali asserzioni sono in parte derivate da strumenti di estrazione automatica del testo come *Text-Presso*, uno strumento che a partire dal vocabolario Neurolex estrae

² <http://www.neuinfo.org/>

dalla letteratura neuroscientifica i dati e i metadati navigabili tramite le ontologie; le ontologie sono accessibili tramite Ontoquest, che rappresenta il *Data Mediator* di NIF, ovvero un database che contiene i riferimenti a tutte le ontologie afferenti a NIF e che le rende interrogabili come se fossero un'unica struttura. L'ontologia del Neuroscience Information Framework è NIFSTD, tramite la quale si accede a tutti i moduli OWL che fanno parte di Ontoquest (figura 2.5)

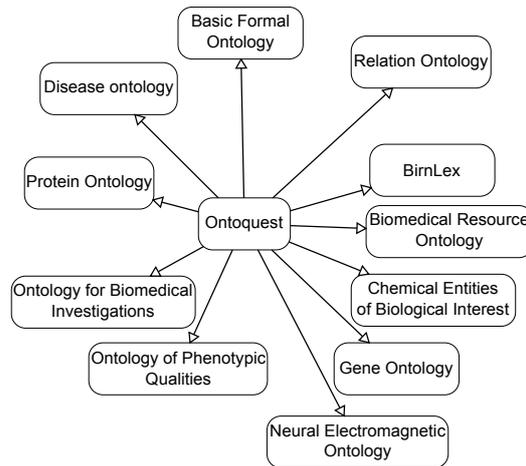


Figura 2.5: Ontologie accessibili da Ontoquest

2.1.1.1 NIFSTD

La struttura della *NIF Standard Ontology (NIFSTD)*, costruita a partire dal vocabolario Neurolex secondo gli standard *OBO-RO*, è attualmente suddivisa in moduli gerarchici che rappresentano i maggiori domini di interesse neuroscientifico:

- *Tassonomia organica*: modulo NIF-Organism (i)

Esprime le relazioni che intercorrono tra le entità anatomiche e una tassonomia standard.

- *Cellule*: modulo NIF-Cell (ii)

La conoscenza del modulo cellulare deriva da NIF, che ha integrato tramite strumenti software semi-automatici l'ontologia minimale della fondazione OBO, quella *SenseLab* contenente la fisiologia neuronale, il modello morfologico *Neuromorpho*, il *Cell-Centered Database* e parte dell'ontologia *Sub-Cellular Anatomy*.

- *Molecole*: modulo NIF-Molecule (iii)

L'ontologia che si occupa della descrizione molecolare deriva dal reperimento semi-automatico delle informazioni dai database dell' *International Union of Pharmacology*, contenenti la conoscenza relativa ai canali ionici e i recettori proteici dei neuroni, per la quale le righe e colonne sono state convertite in classi, annotazioni e relazioni OWL. Le informazioni reperibili sui geni sono relative al nome del canale di trasmissione, la categoria di appartenenza, il numero identificativo di *GENBANK*, i sinonimi, la posizione cromosomica, le sottostrutture conosciute e le sequenze di amminoacidi analizzate.

- *Strato sub-cellulare*: modulo NIF-Subcellular (iv)

Deriva completamente dall'ontologia *Sub-Cellular Anatomy*, che fornisce un modello anatomico in base alle tecniche di indagine microscopica presenti nel *Cell-Centered Database*. La conoscenza *SAO* è integrata nella gerarchia *Gene Ontology* e le relazioni al suo interno seguono lo standard *OBO-RO*, come tutte le altre relazioni presenti in *NIFSTD*, quindi l'unico tipo di ereditarietà tra le classi è quella singola.

- *Anatomia macroscopica*: modulo NIF-GrossAnatomy (v)

L'organizzazione anatomica è importata da *BIRNlex*, che per lo sviluppo del modulo ha sfruttato la paronimia *NeuroNames*, che fornisce una nomenclatura standard sopperendo alle disomogeneità presenti nel lessico neuroanatomico con un'alta quantità di sinonimi per facilitare l'interrogazione della struttura da parte degli esperti del settore. Le relazioni tra le entità anatomiche all'interno del modulo sono di tipo *part of*, in modo da non confondere la relazione di sottoclasse *is a* con quelle di inclusione fisica.

- *Aspetto funzionale del sistema nervoso*: modulo NIF-Function (vi)

Questo modulo provvede all'apporto della conoscenza relativa all'aspetto sensoriale, comportamentale e cognitivo derivante dall'esperienza *BIRN* e *BrainMap* e integra nell'ontologia i dati dei repository delle *fMRI* (risonanze magnetiche funzionali) raccolte in stati di specifiche funzioni cerebrali.

- *Disfunzioni del sistema nervoso*: modulo NIF-Dysfunction (vii)

Le informazioni presenti in questo modulo derivano in parte da *Medical Subject Headings (MeSH)* e riguardano le disfunzioni nervose, muscolari e visive. In questa sezione sono inoltre presenti strutture relative alle disfunzioni multisistemiche, neuromuscolari e alle sindromi sintomatiche. L'annotazione delle disfunzioni comprende collegamenti al *National Institute of Neurological Disorders and Stroke (NINDS)*, in modo tale da rendere possibile una correlazione tra le conoscenze relative alle disfunzioni anatomiche, le patologie e i sintomi.

- *Qualità fenotipiche*: modulo NIF-Quality (viii)

Deriva completamente dall'ontologia delle qualità fenotipiche della fondazione *OBO (PATO)*. Per fenotipo si intende un cambiamento osservabile nell'organismo, quindi questa sezione riguarda ad esempio la relazione tra patologie e cambiamenti anatomici, o la degenerazione cellulare.

- *Processi biologici*: modulo NIF-BioProcess (ix)

Esprime la correlazione tra le attività cerebrali e i processi biologici.

- *Paradigmi cognitivi*: modulo NIF-Investigation (x)

Esprime la correlazione tra il livello funzionale e quello fenotipico.

Oltre alla descrizione anatomica e funzionale la struttura è predisposta per la descrizione dei dati relativi alle tecniche sperimentali e strumentali utilizzate durante le osservazioni, come:

- *Strumenti, protocolli, strategie, reagenti*: modulo NIF-Investigation (xi)

Il modulo relativo alle tecniche di indagine deriva completamente dalla conoscenza dell'*Ontology of Biomedical Investigation (OBI)* che a partire dalla precedente ontologia *BIRNlex* raccoglie le metodologie utilizzate negli esperimenti a più livelli di microscopia.

- *Dati relativi al campione*: modulo NIF-Resource (xii)

La sezione relativa ai dati riguardanti l'organismo analizzato raccoglie dati di contesto, ovvero le condizioni in cui si presenta il campione al momento dell'esperimento.

La varietà delle ontologie presenti all'interno del Neuroscience Information Framework è dovuta al contributo della comunità neuroscientifica³, alla quale è consentita l'integrazione tra le strutture già presenti e quelle derivanti dalla propria conoscenza attraverso gli strumenti messi a disposizione dal NIF. L'inserimento di nuovi dati e l'annotazione di quelli già presenti all'interno dell'ontologia è di tipo collaborativo, ovvero tramite il Wiki ogni utente può sottoporre a una valutazione esperta informazioni sempre più specifiche che entreranno a far parte della conoscenza Neurolex sotto forma di documenti OWL.

L'ontologia NIFSTD fornisce un tipo di conoscenza modulare in cui ogni modulo corrisponde a un dominio di conoscenza e la correlazione tra di essi e tra le entità che li compongono rispetta gli standard OBO-RO (figura A.1). Lo sviluppo di ognuno dei domini deriva dall'apporto della conoscenza di settore, rappresentata da un'ontologia in formato OWL-DL e organizzata all'interno della struttura secondo lo schema in figura 2.6; essendo OWL-DL un linguaggio decidibile, è possibile applicare all'ontologia un ragionatore per esplicitare le relazioni implicite (figura 2.7a, b, c), inoltre per favorire i costi computazionali l'ereditarietà tra le classi non può essere multipla se non per le classi che derivano da combinazioni logiche.

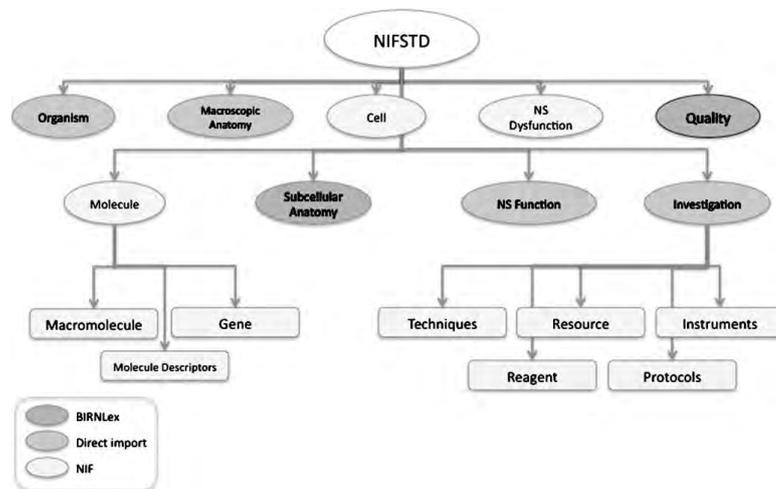


Figura 2.6: Conoscenza di dominio coperta da NIFSTD

³ partecipanti: **i:** NCBI, GBIF, ITIS, IMSR, Jackson Labs; **ii:** CCDB, NeuronDB, NeuroMorpho, OBO; **iii:** IUPHAR, SO, ChEBI, PRO, NCBI, NIDA; **iv:** SAO, GO; **v:** NeuroNames, BIRN, SumsDB, BrainMap; **vi:** NIF, BIRN, BrainMap, MeSH, NINDS, DO; **vii:** MeSH, NINDS, DO; **viii:** PATO, OBO; **ix:** GO; **x:** CogPO; **xi:** OBI, RefSeq; **xii:** NIF, OBI, NITRC, BRO.

L'inserimento di un modulo ontologico in *NIFSTD* varia in base allo stato in cui si presenta: se la fonte è in formato *OWL* è sufficiente importarla (`owl:import`), ma è necessario che essa rispetti la terminologia utilizzata da tutte le altre; se così non fosse è necessario creare un modulo di collegamento che definisca i riferimenti tra l'ontologia importata e il vocabolario standard (ad esempio una proprietà definita come `soma_located_in` tra un neurone e una parte del cervello viene convertita in:

```
Neuron X has_part
    some ('Soma' and (part_of some 'Brain region Y'))
```

all'interno di un modulo di collegamento). Il processo di integrazione può essere parzialmente automatizzato in base a un'analisi della struttura in cui viene fornito all'ontologia (come nel caso della terminologia *IUPHAR* per il modulo *NIF-Molecule*).

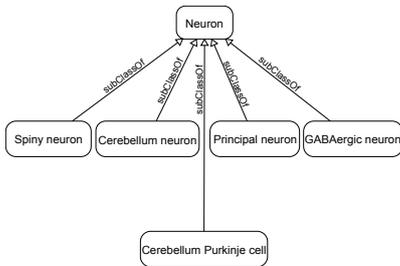


Figura 2.7a: *NIFSTD*, modello asserito

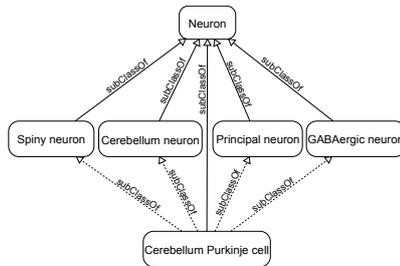


Figura 2.7b: *NIFSTD*, modello inferito

Nome della classe	Espressione asserita
Cerebellum neuron	Is a 'Neuron' whose soma lies in any part of the 'Cerebellum' or 'Cerebellar cortex'
Principal neuron	Is a 'Neuron' which has 'Projection neuron role', i.e., a neuron whose axon projects out of the brain region in which its soma lies
GABAergic neuron	Is a 'Neuron' that uses 'GABA' as a neurotransmitter
Nome della classe	Regola di inferenza
Cerebellum Purkinje cell	1. Is a 'Neuron' 2. Its soma lies within 'Purkinje cell layer of cerebellar cortex' 3. It has 'Projection neuron role' 4. It has 'GABA' as a neurotransmitter 5. It has 'Spiny dendrite quality'

Figura 2.7c: Regole di inferenza

Applicando il ragionatore al modello le regole di inferenza diventano esplicite per tutti i dati ed è possibile effettuare ricerche in base alla classificazione dei dati, ad esempio un neurone può essere individuato tramite la posizione anatomica, i suoi neurotrasmettitori, la sua funzione, la sua morfologia o le sue molecole [10].

Le annotazioni meno specifiche legate a *tutte* le entità dell'ontologia *NIFSTD* sono:

preferred label - indica il nome della classe come è stato inserito

synonym - un insieme di termini alternativi per identificare la classe (per sopperire alle incongruenze nella terminologia neuroscientifica). La presenza di questa annotazione consente di reperire la stessa informazione tramite ricerche distinte (ad esempio "Malattia di Huntington" e "Corea di Huntington").

definition - una breve descrizione della classe

defining citation - contiene la definizione dell'entità attraverso link *URI* o database in rete

curator - indica l'autore della definizione

external source id - definisce una fonte esterna all'ontologia in cui il termine sia definito (come i vocabolari *UMLS*, *MeSHID*, *NeuroNamesID*)

curation status - indica il livello di affidabilità delle informazioni

date - indica la data di inserimento dei dati nell'ontologia

obsolete properties - questa annotazione è assegnata alle classi sostituite, ma per le quali è utile mantenere traccia della precedente definizione

Nello specifico le classi relative ai neuroni contengono le ulteriori annotazioni: *Soma location*, *Dendrite location*, *Origin of axon*, *Localization of distant axon arborization*, *Neurotransmitter released*, *Molecular constituents*, utili alle ricerche di tipo semantico.

La parte meno curata dell'ontologia è quella relativa allo strato sub-neuronale, annotata con informazioni relative al rilascio di molecole in base alla stimolazione cellulare tramite i canali ionici e i recettori; gran parte di questa conoscenza deriva dall'estrazione automatica del database *IUPHAR* e la quantità di dati reattiva ai circa 750 geni annotati è nell'ordine di 3000-4000 considerando i geni, i trascrittori *mRNA* e i recettori peptidergici e tutte le relazioni che intercorrono tra di essi [11].

2.1.2 Ontologie genetiche: Gene Ontology

Il progetto *Gene Ontology* (*GO*) ha il proposito di rendere accessibili tramite una struttura ontologica tutte le informazioni relative ai processi biologici a cui i geni partecipano, alla *posizione della cellula*

in cui i geni sono attivi e alle *funzioni molecolari* (attività biochimiche) dei geni della specie presa in esame. *GO* contiene dati relativi ai genomi di 40 specie differenti, mantenendo un'organizzazione delle informazioni che consente l'integrazione di progetti esterni per favorire l'estensione dell'ontologia principale con una conoscenza di dominio settoriale. L'ontologia che ricopre la sezione relativa ai dati anatomico-funzionali del *mus musculus* è quella del progetto Mouse Genome Informatics (*MGI*) e per capirne il funzionamento è necessario descrivere la struttura della Gene Ontology.

2.1.2.1 Struttura

Gli aspetti ricoperti da *GO* riguardano la funzione dei geni in relazione a:

cellule	interazione con altri organismi
membrane indipendenti e costituenti	relazione con altri geni
proteine	risposta agli stimoli
ciclo cellulare	percezione sensoriale
crescita	comunicazione molecolare
processi metabolici	localizzazione

Tabella 2.1: Conoscenza di dominio coperta da *GO*

Rappresentando la struttura dell'ontologia come un grafo orientato, ogni nodo prende il nome di *GO term* e gli archi che rappresentano le relazioni tra di essi costruiscono la gerarchia *GO*, in cui ogni nodo che abbia uno o più padri riflette una maggiore specializzazione rispetto ad essi.

La Gene Ontology è divisa in tre principali domini ontologici scorrelati tra loro (*cellular component* (figura A.2), *biological process*, *molecular function*), ognuno dei quali è un *GO term* che identifica il settore di pertinenza di tutti i nodi discendenti; essendo i domini scorrelati, nessun termine accessibile tramite uno dei tre nodi padre avrà nodi condivisi con gli altri.

Ogni *GO term* è necessariamente annotato con un nome, un identificativo numerico del tipo *GO:0123456*, un dominio di appartenenza, una definizione testuale, e i tipi di relazione a cui il termine è associato (tabella 2.2); inoltre ad ogni *GO term* è possibile associare anno-

tazioni ausiliari quali identificativi ausiliari, sinonimi, riferimenti ad altre fonti, commenti, tipologia, attualità del termine.

<i>Accession</i>	GO:0009987		
<i>Ontology</i>	Biological Process		
<i>Synonyms</i>	narrow: cell growth and/or maintenance exact: cell physiology exact: cellular physiological process alt_id: GO:0008151 alt_id: GO:0050875		
<i>Definition</i>	Any process that is carried out at the cellular level, but not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level. Source: GOC:go_curators, GOC:isa_complete		
<i>Comment</i>	None		
<i>Subset</i>	PIR GO slim Plant GO slim Prokaryotic GO subset		
<i>Relazioni</i>	cellular process	is_a	biological process
	microtubule-based process	is_a	cellular process
	negative regulation of cellular process	negatively _regulates	cellular process
	positive regulation of cellular process	positively _regulates	cellular process
	regulation of cellular process	regulates	cellular process

Tabella 2.2: Risultato della ricerca⁴ di cellular process

⁴ <http://amigo.geneontology.org/cgi-bin/amigo/go.cgi>

2.1.2.2 Relazioni

Le relazioni che esprimono i legami tra le entità nella Gene Ontology sono sia di tipo gerarchico che funzionale:

Relazioni gerarchiche		Relazioni funzionali	
is-a	regulates	positively-regulates	
part-of		negatively-regulates	

Tabella 2.3: Tipi di relazione nella Gene Ontology

La semantica delle relazioni gerarchiche è immediata: un *GO term* in relazione *is-a* con un altro *GO term* ne indica la tipologia, mentre la relazione *part-of* ne esprime l'appartenenza; la relazione funzionale *regulates* è invece definita come un processo che modula la frequenza, la quantità o l'estensione di un processo biologico, di una qualità o di una funzione (GO:0065007). Le relazioni *positively-regulates* e *negatively-regulates* sono sotto-relazioni di *regulates* e quindi ne ereditano la semantica, aggiungendo una specializzazione. L'inferenza derivante dalla composizione della relazione *is-a* con *regulates* è sempre possibile, mentre la composizione con la relazione *part-of* consente di fare inferenza solo nel caso in cui la classe meno specifica sia la stessa a cui è associata la relazione *regulates* o sue sotto-relazioni (figura 2.8).

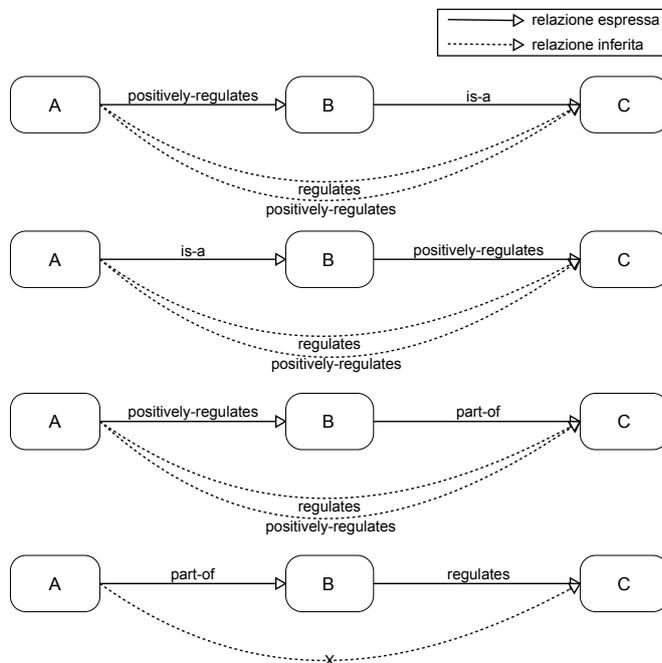


Figura 2.8: Tipi di inferenza sulla composizione delle relazioni

2.1.2.3 *Annotazioni e inferenza*

Definendo delle relazioni logiche tra i termini di *GO* ed associandole alle entità, si ottengono asserzioni nella forma *genus-differentia*, in cui la prima parte (*genus*) è un *GO term* e la seconda parte (*differentia*) è la relazione esistente tra esso e un altro *GO term*; tutte le asserzioni di questo tipo saranno quindi nella forma

GO TERM is *GO term* that **relation** *GO term*

Ad esempio dalla definizione *OBO* del termine *osteoblast development*:

```
[Term]
id: GO:0002076
name: osteoblast development
def:"The process whose specific outcome is the progression of an
osteoblast over time, from its formation to the mature
structure. Osteoblast development does not include the steps
involved in committing a cranial neural crest cell or an
osteoprogenitor cell to an osteoblast fate. An osteoblast is
a cell that gives rise to bone." [source: GOC:dph]
intersection_of: GO:0048468 ! cell development
intersection_of: OBO_REL:results_in_complete_development_of CL
:0000062 ! osteoblast
```

si ottiene la relazione:

"OSTEOBLAST DEVELOPMENT (GO:0002076) is *cell development* (GO:0048468)
that **results in the** *complete development of an osteoblast* (CL:0000062)"

Questo tipo di approccio consente la generazione automatica di relazioni inferite a partire sia da quelle interne a *GO* che dalla cooperazione delle ontologie a cui *GO* fa riferimento. Ogni annotazione deve essere attribuita ad una fonte, che può provenire dalla letteratura, da un altro database o da strumenti automatici (inferenza); in questo caso i tipi di relazioni inferite sono associati a un codice che ne indica la provenienza:

<i>Inferenza: esperimenti</i>	<i>Inferenza: strumenti computazionali</i>	<i>Inferenza: autore</i>
EXP: Experiment	ISS: Sequence or Structural Similarity	TAS: Traceable Author Statement
IDA: Direct Assay	ISO: Sequence Orthology	NAS: Non-traceable Author Statement
IPI: Physical Interaction	ISA: Sequence Alignment	
IMP: Mutant Phenotype	ISM: Sequence Model	<i>Inferenza: curatore</i>
IGI: Genetic Interaction	IGC: Genomic Context	IC: Inferred by Curator
IEP: Expression Pattern	IBA: Biological aspect of Ancestor	ND: No biological Data available
	IBD: Biological aspect of Descendant	
	IKR: Key Residues	<i>Inferenza: automatica</i>
	IRD: Rapid Divergence	IEA: Electronic Annotation
	RCA: Reviewed Computational Analysis	

Tabella 2.4: Codici associati alle relazioni inferite

2.1.3 *Ontologie istologiche: SAO (BFO)*

L'ontologia *SAO (Subcellular Anatomy Ontology)* costituisce il livello intermedio tra la descrizione anatomica e quella sub-cellulare; tramite *SAO* è possibile la navigazione semantica dei dati del *Cell Centered Database (CCDB)*, contenente immagini derivanti da analisi di microscopia ottica ed elettronica. Attraverso l'ontologia le strutture cellulari e sub-cellulari come i neuroni e le sinapsi possono essere annotate con metadati sia tassonomici che funzionali. Il linguaggio in cui è sviluppata *SAO* è il *Web Ontology Language*, in modo tale da fornire un'estensione alle ontologie anatomiche già presenti con un livello di microscopia più alto e la possibilità di fare inferenza.

L'approccio *SAO* è dal basso verso l'alto, ovvero definisce le macrostrutture a partire dalla descrizione delle loro componenti, a differenza delle ontologie anatomiche il cui maggiore raffinamento consiste in un più alto livello microscopico.

L'architettura dell'ontologia *SAO* è basata su quella della *Basic Formal Ontology (BFO)* ed è principalmente divisa nei due settori: *continuant* e *occurrent* (figura A.3), che esprimono rispettivamente l'aspetto strutturale (figura 2.1-b) e funzionale (fasi cellulari, motilità, secre-

zione di sostanze) degli organismi descritti. La parte strutturale è suddivisa nelle classi `dependent_continuant` e `independent_continuant`, che esprimono entità che persistono in funzione della struttura di appartenenza o indipendentemente da essa (ad esempio “colore” non è strettamente correlato con l’apparato cellulare, mentre “numero di dendriti” è una caratteristica che hanno solo le cellule neuronali).

2.1.3.1 *Classi*

Le classi principali della sezione `independent_continuant` riguardano:

- *Cellule*: classe `Cell`

Include la conoscenza relativa a neuroni, cellule gliari, e tutte le strutture derivanti dall’analisi strutturale del sistema nervoso, come le cellule vascolari, endoteliali e muscolari. Le informazioni reperibili riguardano criteri di classificazione relativi a domini differenti, come la morfologia (ad esempio “Neurone piramidale”) o la tassonomia (ad esempio “Neurone di Purkinje”). I nomi assegnati alle entità hanno il solo scopo di rendere la struttura comprensibile per un’esplorazione manuale, poiché ad ogni classe è assegnata una stringa univoca che la identifica (ad esempio a `Nerve Cell` corrisponde la stringa `sao1224657022`), come definito dalle specifiche della fondazione *OBO*.

- *Sezioni sub-cellulari*: classi `Regional Part of Cell`, `Cell Component`

Le classi riguardanti la descrizione sub-cellulare sono divise in `regional part` e `component part`, in continuità con la struttura della *Basic Formal Ontology*. La parte *regionale* intende descrivere quelle strutture che costituiscono una sezione tassonomicamente distinta da quelle adiacenti, ma che non presentano una distinzione fisica da essa (ad esempio un assone nasce dal corpo cellulare del neurone e ne fa parte, pur essendo un concetto distinto da esso); il livello di specializzazione delle classi regionali varia in base alla profondità dell’indagine, ad esempio al livello dendritico c’è una distinzione tra la parte legata al corpo cellulare del neurone, le diramazioni e le specializzazioni terminali. La parte *componente* deriva maggiormente dalla *Gene On-*

tology (GO) e riguarda tutte le sotto-strutture che costituiscono l'apparato cellulare, come le membrane e i mitocondri.

- *Struttura extra-cellulare*: classe Extracellular Structure

Le entità che distinguono il dominio sopracellulare (le micro-strutture esterne al corpo cellulare) sono principalmente le classi *site* e *object_aggregate* che indicano rispettivamente una caratteristica morfologica in relazione alle entità circostanti e un insieme di elementi scorrelati il cui unico legame è rappresentato dalla vicinanza spaziale; ad esempio una sinapsi è il *luogo* in cui si scambiano i neurotrasmettitori (l'aspetto funzionale non è esplicito, ma è inferito dalla presenza dei neurotrasmettitori) e una diramazione dendritica è un *aggregato* di un neurone, cioè è un'entità "vicina" ad ogni cellula neuronale. Alcune entità sono modellate in entrambe le classi, ad esempio le sinapsi (figura 2.9 [12]) hanno una descrizione in *Synaptic site* (Pre-synaptic site, Post-synaptic site, Junctional site) e in *Synapse* (Chemical Synapse, Electrotonic Synapse, Mixed Synapse), in modo da distinguere le qualità spaziali da quelle morfologiche.

- *Molecole*: classe Molecule

La classe molecolare è composta dalle parti che costituiscono una molecola, ovvero amminoacidi, acetilcolina, carboidrati, neuropeptidi, acidi nucleici (DNA, RNA) e proteine; tramite questa suddivisione è possibile identificare un'entità molecolare a partire dall'individuazione di caratteristiche specifiche (ad esempio la presenza o l'assenza di una sostanza in una molecola può essere indice di una particolare patologia).

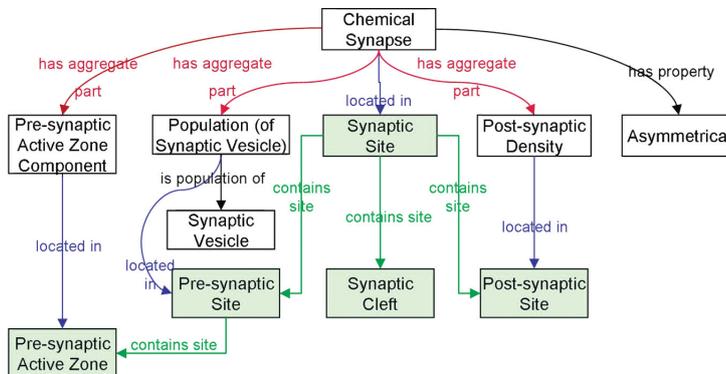


Figura 2.9: Sinapsi chimica: luoghi e aggregati

2.1.3.2 Proprietà

Le proprietà dell'ontologia possono essere divise in tre maggiori categorie:

- *Relazioni spaziali*

Le relazioni spaziali tra le entità sono espresse tramite la proprietà `continuous_with`, che esprime l'adiacenza di un'entità cellulare ad un'altra: ad esempio un assone sarà in relazione di continuità con il corpo cellulare e una sinapsi avrà la stessa relazione con un dendrite (figura 2.1b). La proprietà `anatomical_location` esprime la posizione anatomica della struttura a cui si riferisce ed è associata ad ogni parte della cellula piuttosto che alla cellula stessa, in modo tale da consentire la localizzazione di un apparato a partire dalle sue componenti (ad esempio un particolare neurone può essere individuato dalla presenza di un numero noto di dendriti e di un assone); tale proprietà si specializza in `general_anatomical`, `specific_anatomical` e `atlas_anatomical`, al variare del livello di profondità che si intende esprimere o del riferimento alle coordinate di un atlante.

- *Relazioni di appartenenza*

Le relazioni di appartenenza (`part_of`) derivano dall'analisi microscopica attraverso tecniche che evidenziano lo strato sub-cellulare e riguardano principalmente il livello molecolare; la proprietà `has_molecular_constituent` ad esempio serve per definire la presenza di un qualsiasi tipo di molecola all'interno di un corpo cellulare, ad eccezione dei neurotrasmettitori per i quali è presente una proprietà dedicata (`has_neurotransmitter`) al livello cellulare.

- *Relazioni morfologiche*

Le relazioni morfologiche comuni alla maggior parte delle entità sono importate dalla *Phenotype Trait Ontology (PTO)* e riguardano il livello più alto della descrizione istologica; i tipi di assone, di nucleo, dei vescicoli e di tutte le altre strutture sub-cellulari sono invece state introdotte da *SAO*. Le proprietà morfologiche sono divise tra gli insiemi `dependent_continuant` e `independent_continuant` in base all'attinenza a una particolare struttura: le qualità più specifiche di un'entità sono quelle

dipendenti da essa, mentre quelle più frequenti e comuni alla maggior parte delle entità sono quelle *indipendenti*.

La possibilità di fare inferenza riguarda i livelli di conoscenza tramite i quali sono organizzate le informazioni: ad esempio i neuroni nell'ontologia *SAO* sono organizzati per neurotrasmettitore, proprietà morfologiche, posizione anatomica e circuito di appartenenza, quindi a partire da dati empirici che evidenziano una di queste proprietà (immagini ottenute dalla microscopia elettronica in seguito all'applicazione di un contrasto) si può fare inferenza su tutte le classi correlate; un'informazione esplicita come: "*Il reticolo endoplasmatico dei dendriti delle cellule di Purkinje utilizza il recettore IP₃*" consente di inferire, oltre alla proprietà dichiarata, che le cellule di Purkinje utilizzano il recettore IP₃, che la corteccia cerebrale utilizza il recettore IP₃, e che il cervelletto utilizza il recettore IP₃ (inclusione dal basso verso l'alto).

2.2 UNIFICAZIONE DEL LESSICO

Questa sezione è volta all'analisi dei tipi di lessico utilizzabili in un contesto neuroscientifico per la modellazione di un'ontologia. Si esaminano due differenti strategie di unificazione del lessico, in modo da evidenziare lo strumento più adatto all'annotazione e all'inquadramento di nuovi dati nel contesto neuro-biologico.

2.2.1 *Neurolex*

Una risorsa fondamentale per l'indagine neuroscientifica è rappresentata dal *Neurolex Lexicon* (precedentemente *BIRNlex*), parte integrante del progetto *NIF*, che consiste in un vocabolario di migliaia di termini neuroscientifici tra cui neuroni (attualmente sono presenti 254 neuroni annotati⁵), parti del cervello e componenti del sistema nervoso. Il progetto *Neurolex* è gestito da *INCF* e fornisce una terminologia standard con la quale caratterizzare tutte le componenti utili all'inserimento e alla ricerca di risorse neuroscientifiche. Il vocabolario *Neurolex* è disponibile tramite una pagina internet di tipo *Wiki*⁶ ed è navigabile tramite la struttura *NIF* sia da un punto di vista formale

⁵ fonte: http://neurolex.org/wiki/Cells_Hierarchy

⁶ Un *Wiki* è una pagina che viene aggiornata dai suoi utilizzatori e i cui contenuti sono sviluppati in collaborazione da tutti coloro che vi hanno accesso. [cfr. <http://it.wikipedia.org/wiki/Wiki>]

GO term: Cerebellum Purkinje cell:

Name:	Cerebellum Purkinje cell
Description:	Principal neuron (projection neuron) of the cerebellar cortex; cell bodies arranged in a single layer; characterized by a pear-shaped cell body, 1 (rarely 2) primary dendrites and an elaborate dendritic tree heavily invested with dendritic spines.
Synonym(s):	Purkinje neuron, Purkinje's corpuscles, Cerebellar Purkinje neuron, Purkinje Cell, Purkyne cell, Corpuscles of Purkinje
Has role:	Principal neuron
Super-category:	Neuron
Id:	sa0471801888
Defining criteria:	Morphology; Connectivity; Biochemical
Organism:	Mammal
Soma Specific Properties:	
Cell Soma Shape:	Spherical
Cell Soma Size:	Large soma
Soma location:	Cerebellum
Dendrite Specific Properties:	
Spine density on dendrites:	spiny high density
Branching type:	unipolar
Axon Specific Properties:	
Axon myelination	myelinated
Axon projection laterality:	ipsilateral
Origin of axon:	soma
Location of distant axon arborization:	Deep cerebellar nuclear complex
Location of local axon arborization:	Purkinje cell layer of cerebellar cortex
Intrinsic Properties:	
Neurotransmitter released:	GABA
Neurotransmitter receptors:	AMPA-type glutamate-gated cationic channel, GABA-gated anionic channel, ga
Firing patterns:	Regular firing and bursting
Spontaneous firing patterns:	Regular firing and bursting
Spontaneous firing rate:	Rapid firing

Tabella 2.5: Esempio di ricerca di un termine Neurolex

che semantico, possibilità derivante dalla presenza dell'ontologia *NIFSTD* sottostante. L'inserimento dei dati nell'ontologia tramite il *Wiki* di Neurolex è quindi dal basso verso l'alto, ovvero le informazioni vengono prima incluse nel vocabolario dagli utenti (o da un'ontologia) e poi assegnate alla relativa sezione; da un punto di vista informatico è un metodo meno efficace dell'approccio *top-down* poiché richiede un'analisi aggiuntiva all'inserimento, ma l'approccio collaborativo è l'unico che rende l'inserimento fruibile a chi non si occupa di ontologie. Gli esperti dei settori neuro-biologici inseriscono i dati in base alla propria conoscenza e gli ontologi del *NIF* si occupano di rendere il lessico uniforme a quello specificato dallo standard *OBO-RO*, utilizzato da tutte le ontologie *NIFSTD*.

2.2.2 *NeuroNames*

Neuronames è un lessico standard volto a uniformare i vocabolari neuroscientifici e consentire quindi un'interrogazione della struttura a partire da nomenclature differenti. La gestione delle differenti terminologie si ottiene tramite un largo uso di *sinonimi*, uno dei quali sarà il *default name*, il sinonimo principale; ogni termine è fornito inoltre di un'*abbreviazione* che lo identifica univocamente. Il *dominio spaziale* di NeuroNames è costituito dalle strutture primarie (*primary structures*) e composte (*superstructures*), ognuna delle quali è associata a una lista di sinonimi che costituiscono il *dominio simbolico*. Le strutture presenti riguardano i livelli di indagine macroscopico e istologico e si distinguono in *volumetriche* e *superficiali*, ovvero componenti anatomiche interne e superfici aderenti ad esse; la struttura superficiale è a sua volta divisa in superficie topografica (non volumetrica) e appendice (volumetrica), mentre quella volumetrica è suddivisa in *parenchimale* (materia grigia/bianca) e *fluida*.

<i>Sinonimi</i>	<i>Strutture primarie</i>	<i>Soprastrutture</i>	<i>Termini ausiliari</i>
identificativo univoco	identificativo univoco	identificativo univoco	identificativo univoco
lingua (inglese, latino, ...)	nome di default (inglese)	nome di default (inglese)	nome di default
fonte bibliografica	nome di default (latino)	nome di default (latino)	specie
specie di riferimento	abbreviazione	abbreviazione	definizione gerarchica
classificazione nell'atlante	entità volumetrica o superficiale	entità volumetrica o superficiale	definizione di default
nome di default	composizione (materia grigia/bian- ca/fluida)	antenati nella gerarchia	fonte della definizione
riferimento PubMed	antenati nella gerarchia		

Tabella 2.6: Annotazioni dei termini in NeuroNames

Le entità esterne alla gerarchia sono i *termini ausiliari*, che riguardano i criteri secondo i quali le strutture possono essere organizzate: l'aspetto strutturale, connettivo, evolutivo, neurotrasmissivo e funzionale (ad esempio le aree di Brodmann, figura 2.2a).

La gerarchia delle strutture primarie è composta da nove livelli di inclusione (figura A.4) in cui le strutture costituiscono una descrizione dei volumi cerebrali per prossimità, ovvero le entità di dimensione maggiore sono formate dall'aggregazione di strutture più piccole in esse contenute. Un gruppo di sezioni anatomiche dà luogo a una *soprastruttura*, anch'essa dotata delle annotazioni comuni ai termini NeuroNames (sinonimi, default name, abbreviazioni). Tutti i termini che non fanno parte della gerarchia sono i termini ausiliari (*ancillary terms*).

La nomenclatura Neuronames è catalogata in basi di dati che riguardano le disposizioni anatomiche di *umani* (Crosby, Roberts, Carpenter - Sutin, Paxinos, Ono, Mai), *primati* (Walker, Bonin-Bailey, Krieg, Szabo-Kowan) e *topi* (Bowden-Martin, Paxinos-Watson, Swanson) [13], che possono essere consultate tramite *BrainInfo*, un client web tramite

il quale è possibile la navigazione della gerarchia. Il modello gerarchico da cui derivano le classi delle strutture primarie è quello esposto in *Nomina Anatomica*, il testo di riferimento per la nomenclatura anatomica in latino.

Lo scopo di NeuroNames non è dunque quello di compilare una lista di termini neuroscientifici, bensì di trovare un compromesso lessicale nella varietà di nomenclature che compaiono nella terminologia neuroanatomica; a tale scopo la terminologia NeuroNames include per ogni elemento la denominazione inglese, latina e i sinonimi.

2.3 ONTOLOGIE

2.3.1 *Ontologie anatomiche: ABA*

Una delle associazioni più accreditate per l'esplorazione virtuale del cervello è l'*Allen Institute for Brain Science*, che dedica le proprie risorse all'ampliamento e all'integrazione delle conoscenze informatiche nelle neuroscienze. I mezzi sia a livello umano che strumentale di cui la compagnia dispone sono ingenti e una sezione della ricerca dell'istituto è dedicata all'aspetto ontologico dell'analisi cerebrale.

Le ontologie che raccoglie il gruppo Allen riguardano la descrizione gerarchica del cervello umano e del *mus musculus*, il topo comune; per entrambe è presente un'ontologia completa che descrive un cervello adulto e un'altra, *AGEA*, divisa in sette stadi di evoluzione del cervello di topo in base all'età (figura 2.10, Allen). La descrizione del cervello di un topo tramite l'ontologia è suddivisa in 13 livelli gerarchici (figura A.5-a) che rappresentano la specializzazione delle sezioni macroscopiche in base alla crescita [7]; i livelli descrittivi più bassi si riferiscono a stadi giovanili del campione, che con la crescita modifica e specializza le proprie strutture cerebrali, descritte nei livelli più alti. L'atlante stereotassico di riferimento per la nomenclatura delle ontologie riguardanti il *mus musculus* è quello di Franklin e Paxinos, *The Mouse Brain in Stereotaxic Coordinates*, sul quale si basano tutte le gerarchie presenti nel progetto (figura A.5-b [8]).

livelli 04 e 05 - Al quarto e quinto livello dell'ontologia si evidenziano le regioni dorsale e ventrale, in particolare al quinto livello si distingue tra sezione superiore e inferiore di ogni entità di livello superiore (*roof plate* e *floor plate*).

livelli 06, 07 e 08 - In questi livelli la suddivisione topologica è maggiore e interessa principalmente il telencefalo, ad esempio si identifica il *pallium* e tutte le parti che lo compongono: quella ventrale, laterale, mediale e dorsale.

livelli 09 e 10 - Il nono livello descrive le zone *ventricolari* e *mantelari* delle aree superiori, mentre il decimo ne è una specializzazione che considera lo strato periventricolare, intermedio e superficiale delle zone descritte al nono livello.

livelli 11, 12 e 13 - Il livello 11 è attualmente il più specializzato; gli ultimi due livelli dell'ontologia sono necessari perché da analisi empiriche è risultato che limitare la descrizione ai livelli inferiori non consente di esprimere tutte le differenze tra i cervelli adulti. Il nucleo ipotalamico paraventricolare, ad esempio, compare al livello 11 e le sue sezioni al livello 12; se fosse necessaria un'ulteriore specializzazione, ad esempio sullo strato cellulare, essa sarebbe contemplata nel livello 13. Al livello attuale dell'ontologia il dodicesimo livello è quello con il minor numero di entità e il livello 13 non è definito, ovvero la gerarchia cerebrale non include ne' il livello neuronale ne' quello sinaptico.

2.3.1.1 EMAP

Il progetto EMAP (*Edinburgh Mouse Atlas Project*), parte integrante di INCF, rende contemporaneamente accessibili i livelli di conoscenza ricoperti dall'*Allen Institute for Brain Science*, dall'atlante di riferimento *Waxolm Space (WHS)* e dall'atlante delle espressioni geniche EMAGE tramite un modello sia bidimensionale che tridimensionale che mette in relazione:

- relazioni tra le entità macroscopiche (ontologia Allen, figura A.5-a)
- disposizione tassonomica (atlante Paxinos, figura A.5-b)
- immagini derivanti da tecniche di indagine differenti (fluorescenza cellulare, immunoistochimica, iniezione virale,...)

- espressione dei geni in base allo stadio di sviluppo (*EMAGE*)

Le immagini fornite dal gruppo Allen (relative a ~ 300 zone cerebrali [20]) sono esplorabili a più livelli di granularità e sono volte a mettere in evidenza particolari zone del cervello (sezione *Mouse Connectivity* di Allen) annotate con le informazioni relative al tipo di esperimento e ai dati del campione.

2.3.2 Ontologie di indagine: *BIRNlex*

Il progetto *Biomedical Informatics Research Network (BIRNlex)* si propone un'indagine biologica che integri più livelli di conoscenza. Le discipline che il progetto raccoglie sono riassunte in figura 2.11 e riguardano principalmente dati di risonanze magnetiche (*MRI, MRM*), trasparenze specifiche (*contrasto Nissl, APP*), trattografie ottenute tramite *DTI (Diffusion Tensor Imaging)*, database anatomici e cellulari (*CCDB*), informazioni su specifiche zone del cervello e atlanti di organismi sia sani (neo e post-natali) che patologici (morbo di Parkinson, sclerosi multipla, . . .). Lo scopo finale è quello di consentire un accesso simultaneo alle risorse relative ai vari settori disciplinari a discrezione della necessità di chi accede ai dati navigando l'ontologia.

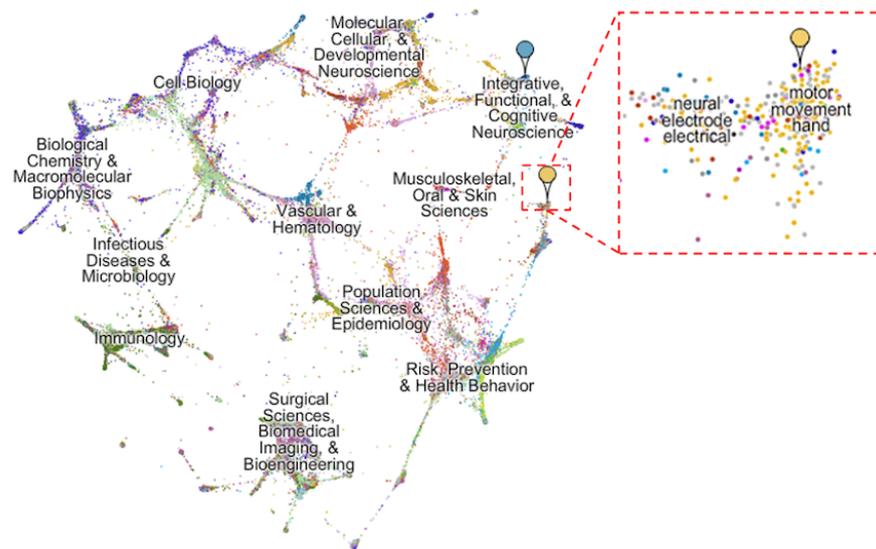


Figura 2.11: Settori disciplinari di *BIRNlex*

Dal punto di vista ontologico il progetto *BIRNlex* ha sviluppato la propria *ontologia BIRNlex* (figura A.6) fino alla completa fusione con il

lessico Neurolex; il contributo di *BIRNlex* al vocabolario Neurolex riguarda prevalentemente la parte dei metadati di tipo non anatomico, come quelli strumentali, di contesto, di trattamento e di indagine.

L'ontologia *BIRNlex* esprime principalmente relazioni di inclusione e fornisce l'integrazione di metadati riguardanti:

- Condizioni degli esperimenti
(Confocal imaging protocol, Fluorescence resonance energy transfer imaging,...)
- Software e risorse usate per l'analisi dei dati dopo l'estrazione
(Alignment software, Statistical parametric mapping,...)
- Tecniche di indagine
(MRI imaging protocol, Golgi staining protocol,...)
- Patologie o mutazioni inducibili sul campione
(Isaacs syndrome, Glaucomatous retinal degeneration)
- Età e caratteristiche fisiche e funzionali
(Prenatal→Early prenatal→First trimester,
Cognitive state→Emotion→Fear)
- Classificazione all'interno della specie

Le annotazioni presenti nell'ontologia riguardano prevalentemente definizioni statiche e locali (ad esempio l'annotazione di:

Phenotype

- Behavioral phenotype
- Behavioral laterality
 - Handedness
 - Right handedness

è "*A behavioral phenotype where the subject preferentially uses the right hand for tasks requiring the use of a single hand or a dominant hand.*") e il livello di attendibilità delle informazioni è in base alla data di immissione delle annotazioni.

L'ontologia è predisposta per l'integrazione con altre, infatti esplorando i livelli troviamo all'interno di: Information entity→ Algorithm entità come Physiological model, Sequence-based model e Structural model, prive di sottosezioni; queste entità, per il momento annotate con la sola definizione della struttura, saranno il punto di

contatto tra l'ontologia *BIRNlex* e altri tipi di conoscenza (anatomica, istologica, subcellulare, genetica, funzionale, . . .).

2.3.3 *Ontologie funzionali: Senselab*

Il progetto Senselab è volto a fornire la competenza relativa all'aspetto funzionale di alcune classi di neuroni, in particolare alla sezione relativa alla percezione degli odori e le zone del cervello coinvolte in essa. I database Senselab sono divisi in 8 categorie: *NeuronDB*, *CellPropDB*, *ModelDB*, *ORDB* (Olfactory Receptor Database), *OdorDB*, *OdorMapDB*, *BrainPharm* e *MicrocircuitDB* e per gran parte dei dati analizzati è presente un modello neuronale e sub-neuronale che ne descrive le proprietà anatomiche e funzionali. La struttura dei database è unificata: si basa su *EAV/CR* (*Entity Attribute Values with Classes and Relationships*), un modello analogo a *RDF* sia per la possibilità di annotare i dati con metadati, sia per la struttura tramite la quale sono accessibili i dati: le informazioni sono registrate come entità descritte da attributi che possono assumere valori e tali valori possono essere sia dati che entità, in modo da consentire una descrizione più articolata delle relazioni e delle funzioni descritte. Per garantire l'interoperabilità tra le fonti (il progetto è accessibile attraverso *NIF*) alcune sezioni di Senselab sono state convertite nei formati standard per la modellazione di ontologie, ad esempio *NeuronDB* è stato convertito in *OWL* e *BrainPharm* in *RDF* (figura 2.12, riadattato da [14]).

Le strutture principali di Senselab sono:

- *CellPropDB*, *NeuronDB*

I due database anatomici di Senselab riguardano le analisi delle principali proprietà fisiologiche dei neuroni: il primo riguarda le proprietà che caratterizzano prevalentemente i neuroni coinvolti nell'attività olfattiva: i canali di conduzione, i neurotrasmettitori e i recettori sinaptici; il secondo mette in relazione le proprietà con le qualità morfologiche di tali neuroni, consentendo quindi una ricerca "per comparti" [15] che presentano determinate proprietà. La suddivisione in due database riflette la granularità delle informazioni reperibili: tramite *CellPropDB* è possibile ottenere la definizione di un neurone come entità singola in base alle proprietà che lo caratterizzano, mentre tramite *NeuronDB* si ha accesso alle sue componenti.

La lista delle cellule annotate (dati) comprende, oltre a 11 inter-neuroni, 32 neuroni individualmente accessibili per regione cerebrale oppure tramite i loro attributi (canali, neurotrasmettitori, recettori).

- *ORDB, OdorDB*

I database riguardanti il sistema olfattivo forniscono informazioni genomiche, proteomiche e strutturali sui recettori chemosensoriali e olfattivi (*OR*); accedendo a *ORDB* è possibile reperire dati sugli apparati recettivi e feromonici di 60 specie tra cui mammiferi, organismi acquatici e anfibi, correlati con le sostanze che li stimolano o li inibiscono (gli odori) presenti in *OdorDB*, dove sono registrati i dati sulla composizione molecolare, l'identificativo chimico, il gruppo funzionale o il modello bidimensionale degli attivatori chimici.

- *OdorMapDB*

Lo scopo del database *Odormap*, costruito a partire dai dati in *ORDB* e *OdorDB*, è quello di definire la risposta del bulbo olfattivo (*OB*) agli stimoli e di formalizzarne i meccanismi. Il risultato di questo tipo di approccio è una risorsa che provvede alla conoscenza relativa alle proprietà funzionali del bulbo olfattivo in base agli odori. I dati di ogni esperimento derivano dall'analisi di risonanza del cervello in seguito a una stimolazione e comprendono i metadati che descrivono le proprietà fisiologiche del campione, la zona del cervello presa in esame, il tipo di odore, il tempo di esposizione e l'immagine della risonanza.

- *ModelDB, MicrocircuitDB*

I database dei modelli neuronali sono strettamente correlati con *NeuronDB* e contengono un registro contributivo di modelli coi quali è possibile simulare il comportamento dei neuroni a due livelli di profondità (cellulare e sub-cellulare). Questa sezione contiene i modelli di neuroni, sinapsi, assoni, canali di voltaggio, recettori, trasmettitori e geni, i quali sono accessibili per nome, autore del modello e regione cerebrale; per ognuno di essi è possibile osservarne il comportamento tramite il simulatore neuronale *NEURON*⁷.

⁷ <http://www.neuron.yale.edu/neuron/>

- *BrainPharm*

BrainPharm è un database contenente le informazioni relative ai risultati delle ricerche sull'effetto di sostanze chimiche sul cervello nel trattamento di disturbi neurologici. Gli agenti chimici sono descritti in base all'azione che hanno su organismi sani e malati; per ognuno di essi è individuabile la collocazione morfologica, gli agenti patologici intra/extra cellulari, i recettori dei neuroni coinvolti, e la reazione che la sostanza provoca in essi. L'accesso ai dati è possibile tramite una ricerca per neurone, meccanismo patologico e agente farmacologico.

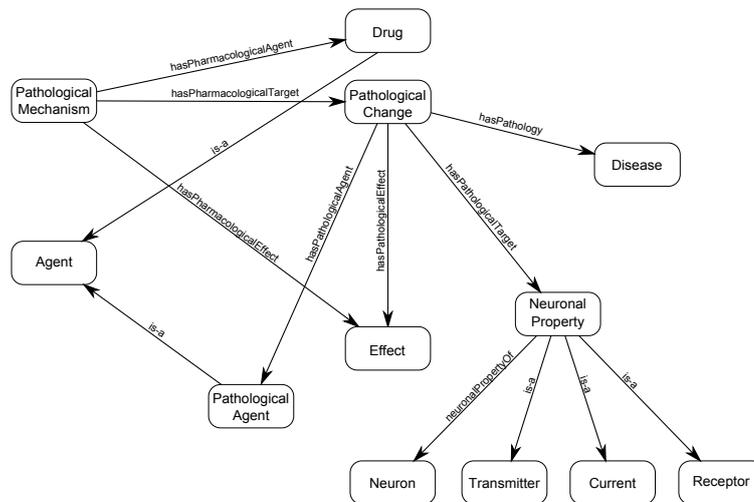


Figura 2.12: Senselab, sezione dell'ontologia BrainPharm

2.3.4 Ontologie murine: MGI

Mouse Genome Informatics (MGI) rappresenta la sezione della Gene Ontology che si occupa esclusivamente della specie *mus musculus*; lo scopo del progetto è studiare le sequenze genetiche e i relativi aspetti funzionali nel topo comune, poiché la riproposizione delle sequenze nell'una e nell'altra specie può essere indice del mantenimento dell'aspetto funzionale a cui tali sequenze sono legate, quindi se la conoscenza di dominio deriva da esperimenti condotti su una delle due specie, viene valutata la possibilità di associare tale conoscenza anche all'altra; le relazioni che derivano dalla conoscenza inferita

tramite strumenti automatici non vengono invece associate a specie diverse da quella presa in esame perché ritenute inaffidabili⁸.

La conoscenza ricoperta dall'ontologia *MGI* deriva dagli esperimenti condotti ai Jackson Labs, e riguarda malattie, geni, alleli, classi di marcatori genetici, tipi di mutazioni, procedure di analisi, stadi di evoluzione, tessuti, cloni e cellule staminali; l'organizzazione di queste informazioni rende tali domini accessibili in base a: gene, fenotipo, modello patologico, espressione genica, ricombinasi, funzione, percorso, polimorfismo nucleotidico, ortologia e presenza tumorale⁹.

I moduli ontologici gestiti da *MGI* per gestire le informazioni derivanti dagli esperimenti dei Jackson Labs sono l'ontologia anatomica del topo adulto (*MA*) e l'ontologia delle patologie murine (*MPATH*).

2.3.5 *Ontologie di patologie: MPATH*

L'ontologia *Mouse Pathology (MPATH)* ha lo scopo di costruire una struttura che consente di associare metadati alle immagini del progetto *Pathbase*, relative all'aspetto istologico delle patologie riscontrate o indotte nei topi, in modo tale da consentirne una navigazione semantica (figura 2.13). *MPATH* è collegata con le ontologie anatomiche (*Mouse Anatomy*), genomiche (*Gene Ontology*) e cellulari (*Cell Ontology*) in modo tale da coprire tutti gli aspetti fenotipici legati allo studio dell'istopatologia [16].

La struttura dell'ontologia è analoga a quella della *Gene Ontology* (figura A.7): le relazioni presenti sono di tipo *is-a* e *part-of* e ogni termine accessibile tramite *MPATH* è fornito di un identificativo, un nome, una definizione, i sinonimi e le relazioni logiche con altri termini; tutte le annotazioni morfologiche, le relazioni logiche tra le classi e le descrizioni derivano dai database *Pathbase*, nei quali sono registrate secondo lo standard *OBO*.

L'organizzazione della gerarchia è studiata in base al tipo di patologia analizzata: oltre alle patologie annotate ma non riferite a una tipologia particolare, le disfunzioni sono divise in entità anatomiche e processi patologici. Il primo gruppo riguarda le alterazioni al tessuto e alle cellule, i disordini circolatori, le deformità strutturali e di sviluppo, le immunodeficienze e i tumori; il secondo gruppo riguar-

⁸ Mouse Genome Informatics scientific curators, 2001. External accession(s): MGI:2154458, J:73065

⁹ <http://www.informatics.jax.org/>

da i processi di danneggiamento cellulare, i disordini di sviluppo e differenziazione cellulare, le immunopatologie e le neoplasie.

I metadati associati alle immagini si riferiscono, oltre al nome dell'immagine, alla sua provenienza e a una descrizione, alle informazioni che ne definiscono la localizzazione all'interno di altre gerarchie; sono quindi presenti i dati relativi al genere, la razza, la specie, lo stadio di sviluppo (in riferimento alle ontologie *EMAP*), il genotipo, la patologia (in riferimento all'ontologia *MPATH*), la manipolazione genetica e sperimentale, la localizzazione anatomica (in riferimento all'ontologia *MA*), e gli alleli; i riferimenti derivano dall'integrazione delle tre ontologie anatomiche, genomiche e cellulari con Pathbase.

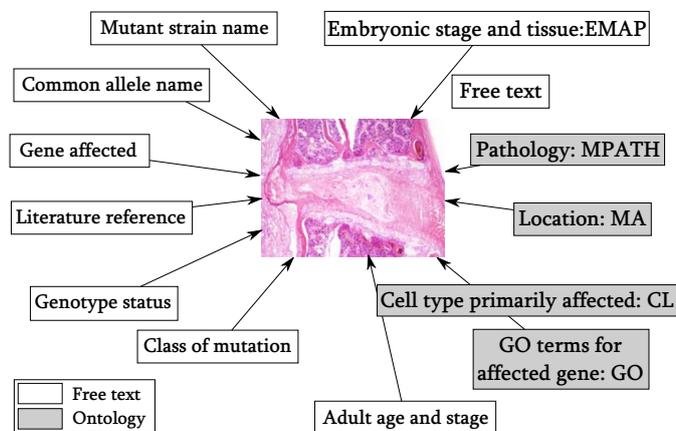
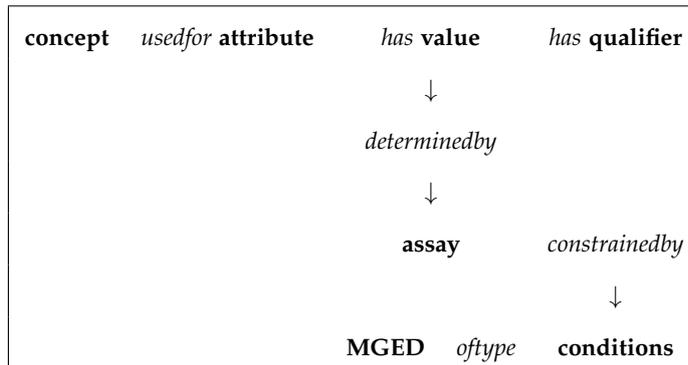


Figura 2.13: Metadati associati alle immagini Pathbase

2.3.6 Ontologie fenotipiche: PATO

La *Phenotypic And Trait Ontology (PATO)* è l'ontologia che ricopre la conoscenza relativa alla descrizione formale del fenotipo murino e si propone come un progetto di collegamento tra le ontologie anatomiche (*MGI*), di sviluppo (*EMAP*), di patologie (*MPATH*), genetiche (*GO*) e comportamentali. L'obiettivo che si propone PATO è quello di formalizzare la descrizione di un fenotipo sotto forma di asserzione in un'ontologia che costituisce un modulo di conoscenza all'interno di un organismo i cui aspetti sono curati da moduli indipendenti e interoperanti. Le asserzioni modellate dall'ontologia fenotipica sono

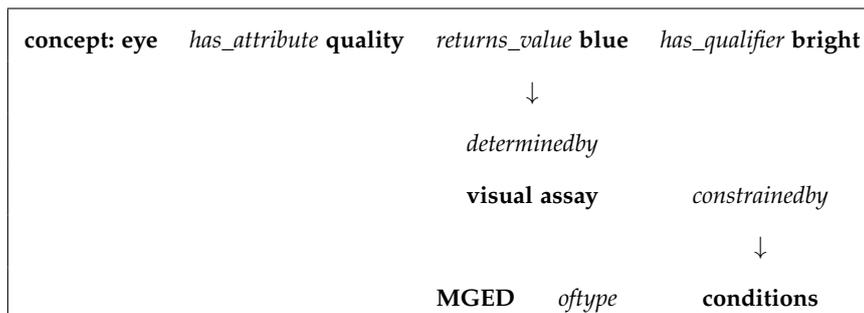
del tipo **entità - relazione - attributi** e sono schematizzate come segue [18]:



Ad esempio si può modellare l'asserzione:

“eye has color blue that is bright. Blue is determined by visual assay”

attraverso:



Le sezioni in cui è suddivisa l'ontologia sono:

organism contiene le informazioni dell'organismo da cui provengono le caratteristiche fenotipiche

entity contiene le classi delle ontologie di contesto importate in *PATO*

attribute contiene la gerarchia dell'ontologia *PATO* (la cui classe principale è *quality*), riguardanti le strutture fenotipiche

assay contiene i metadati relativi al fenotipo osservato; la relazione presente tra *assay* (tipi di dati) e *values* (dati) varia in base al tipo di esperimento

values contiene l'interpretazione che viene data alle caratteristiche fenotipiche in relazione al tipo di esperimento

A partire da un'asserzione che descrive un particolare fenotipo si ha quindi lo schema (riadattato da [19]):

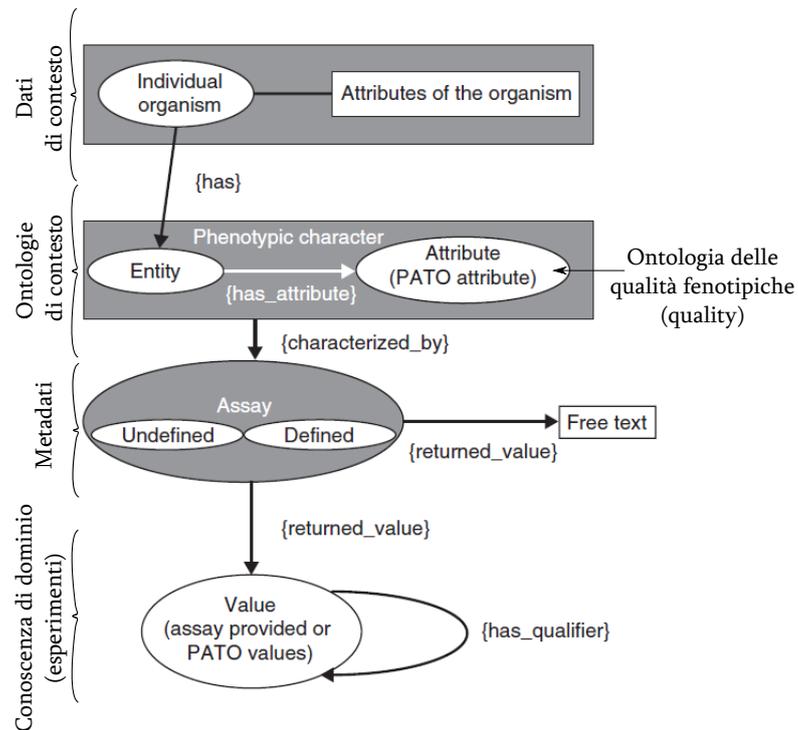


Figura 2.14: Collocazione dell'ontologia fenotipica in *PATO*

L'ontologia che riguarda le qualità fenotipiche e che costituisce il contributo della gerarchia *PATO* alla modellazione ontologica di un organismo è collocata nella classe *attribute*, nella quale si sviluppa l'entità *quality* (figura A.8), tramite la quale è possibile navigare semanticamente la struttura in relazione alle ontologie di contesto ed accedere ai metadati che caratterizzano gli altri aspetti dell'organismo.

INTEGRAZIONE DEI LIVELLI

Per poter contestualizzare le ontologie analizzate è necessario identificare quali sono i loro moduli di coordinamento, ovvero quelle ontologie che costituiscono il punto di contatto tra le categorie neuroscientifiche e le ontologie settoriali, consentendo l'inserimento di una struttura circoscritta ad un aspetto particolare in un ambito più ampio (figura 3.1). E' inoltre necessario fare uso del lessico che meglio si adatta al tipo di dati che costituiranno la conoscenza dell'ontologia e degli aspetti che le ontologie di contesto ricoprono, in modo da avere una visione d'insieme su quali sono gli ambiti la cui conoscenza di dominio è ricoperta e dove emerge la necessità di un contributo.

La prima sezione contiene la tassonomia dei moduli di coordinamento, ovvero le categorie di cui le ontologie analizzate fanno parte, in modo da consentire la contestualizzazione delle nuove strutture all'interno dell'ambito di indagine.

La seconda sezione riguarda i tipi di lessico utilizzati e il loro contributo al vocabolario utilizzato dalle ontologie neuroscientifiche; si evidenziano la profondità del livello di microscopia che tali lessici raggiungono e il metodo con cui essi vengono costruiti in modo da adattarsi alle esigenze e alle risorse disponibili in fase di costruzione di un'ontologia.

La terza sezione riguarda le strutture relative alla conoscenza di dominio ricoperta dalle ontologie analizzate. Sono evidenziati i punti di contatto tra i moduli e le ontologie, tutti i metadati di cui ogni ontologia fornisce il contributo e la quantità di ambiti ricoperti da ogni modulo di coordinamento in base all'integrazione con altre componenti.

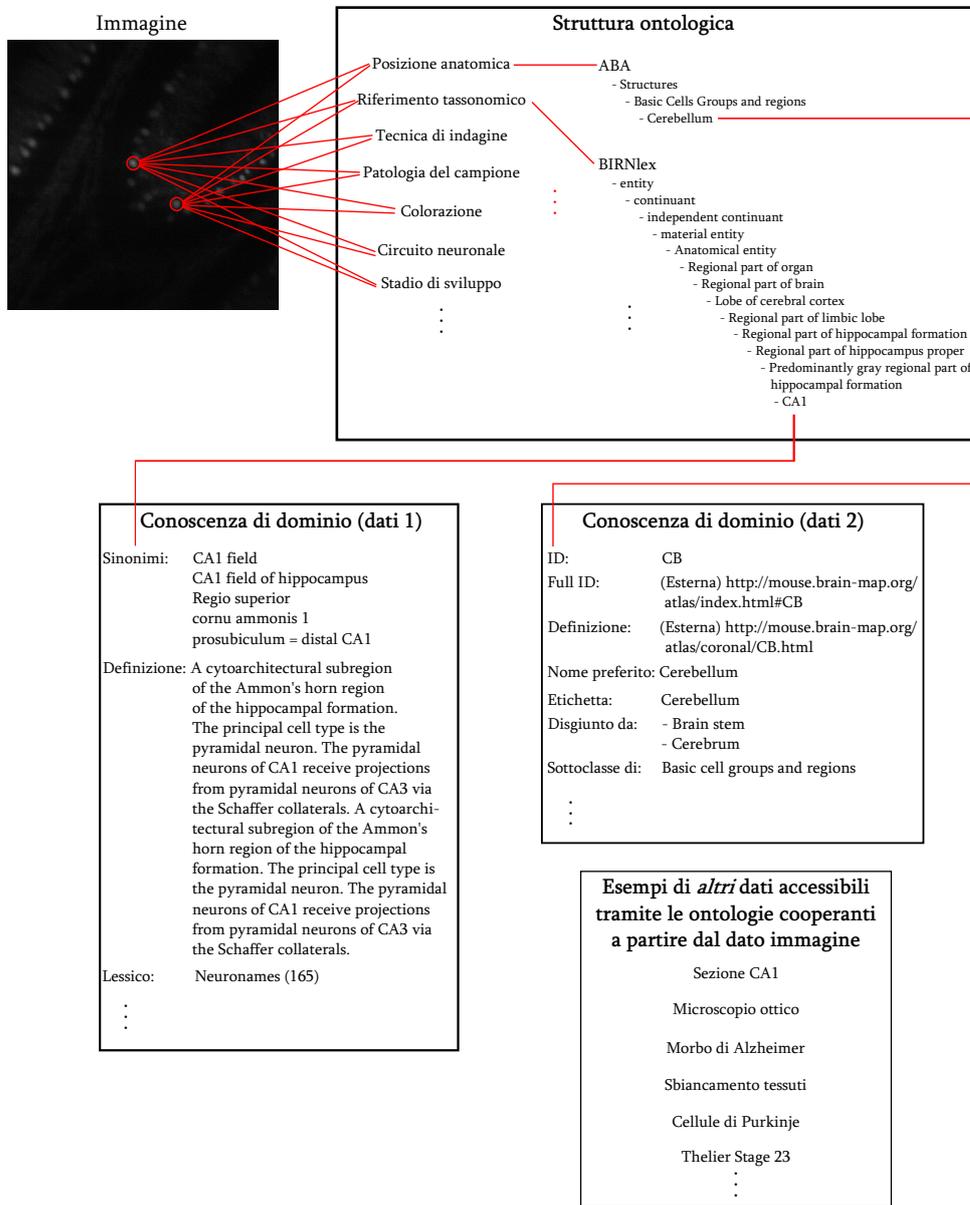


Figura 3.1: Correlazione tra le immagini e i dati

3.1 MODULI DI COORDINAMENTO

	ABA	NIFSTD	GO	BIRNlex	SAO	Senselab	EMAP	MGI
Ontologie anatomiche	X	X				X	X	X
Ontologie funzionali	X	X				X		
Ontologie genetiche			X				X	X
Ontologie istologiche		X			X	X		
Ontologie di indagine		X		X				
Ontologie di patologie		X				X		X
Ontologie fenotipiche		X				X		X

Tabella 3.1: Tassonomia dei moduli di coordinamento

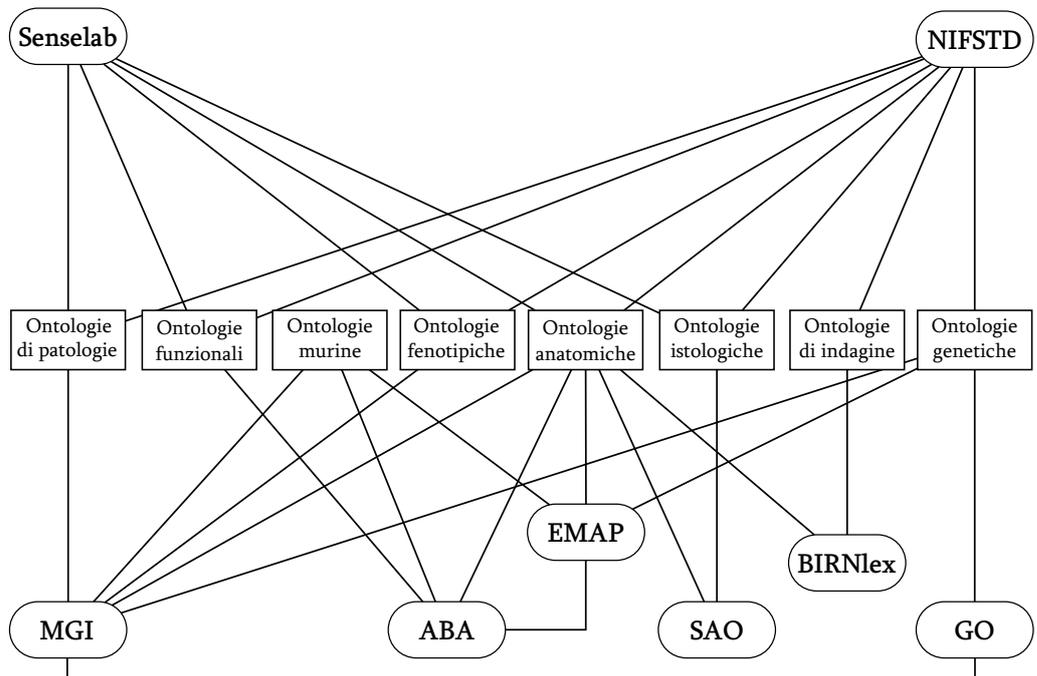


Figura 3.2: Dominii di conoscenza e ontologie afferenti

3.2 LESSICI

	Neurolex	Neuronames
Sinonimi	X	X
Annotazione contributiva	X	
Metodo top-down		X
Metodo bottom-up	X	
Livello anatomico	X	X
Livello cellulare	X	
Livello neuronale	X	
Tassonomia		X
Localizzazione spaziale		X
Organismo: uomo	X	X
Organismo: topo		X

Tabella 3.2: Ambiti ricoperti dai vocabolari utilizzati

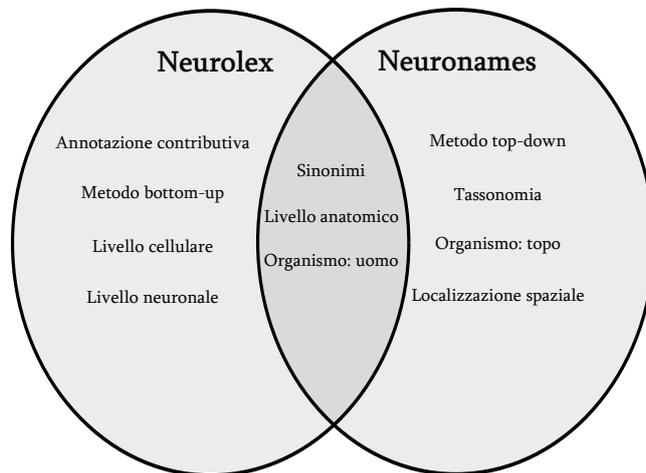


Figura 3.3: Convergenze lessicali tra i vocabolari

3.3 CONOSCENZA DI DOMINIO

	ABA	NIFSTD	GO	BIRNlex	SAO	MGI	Senselab	EMAP	MPATH	PATO
X	modulo individuale									
*	modulo importato									
Tassonomia anatomica	X	X	X	X		X		*		
Sviluppo	X	*	X					*		*
Anatomia macroscopica	X	*	*		X	X		*		*
Processo biologico		X	X	X		X	X	*		
Tecnica di indagine		*		X						
Cellule		X	X	X	X		X			
Anatomia cellulare		X			X		X			
Chimica		*					X			
Patologia		X	*	X		X	X	*	X	*
Comportamento		*					X			X
Condizione sperimentale		*		X						
Genomica e proteomica	X	*	X			X		*		*
Immagini	X		*					*	X	
Immagini di risonanza	X	*	*				X	*	X	
Immunologia		*					X			
Malattia neurologica		X					X			
Ciclo cellulare		X	X							
Molecole		X	X							
Anatomia del topo	X	*	*			X	X	*		
Disordine neurologico		X					X			
Fenotipo		*	*	X		X	X	*		X
Processo psicochimico		*	X				X			
Proteine		*	X							
Anatomia sub-cellulare		X			X		X			
Paradigma cognitivo		*					X			
Interazioni farmacologiche		*					X			
Circuiti neuronali		*			X		X			

Tabella 3.3: Tassonomia delle ontologie

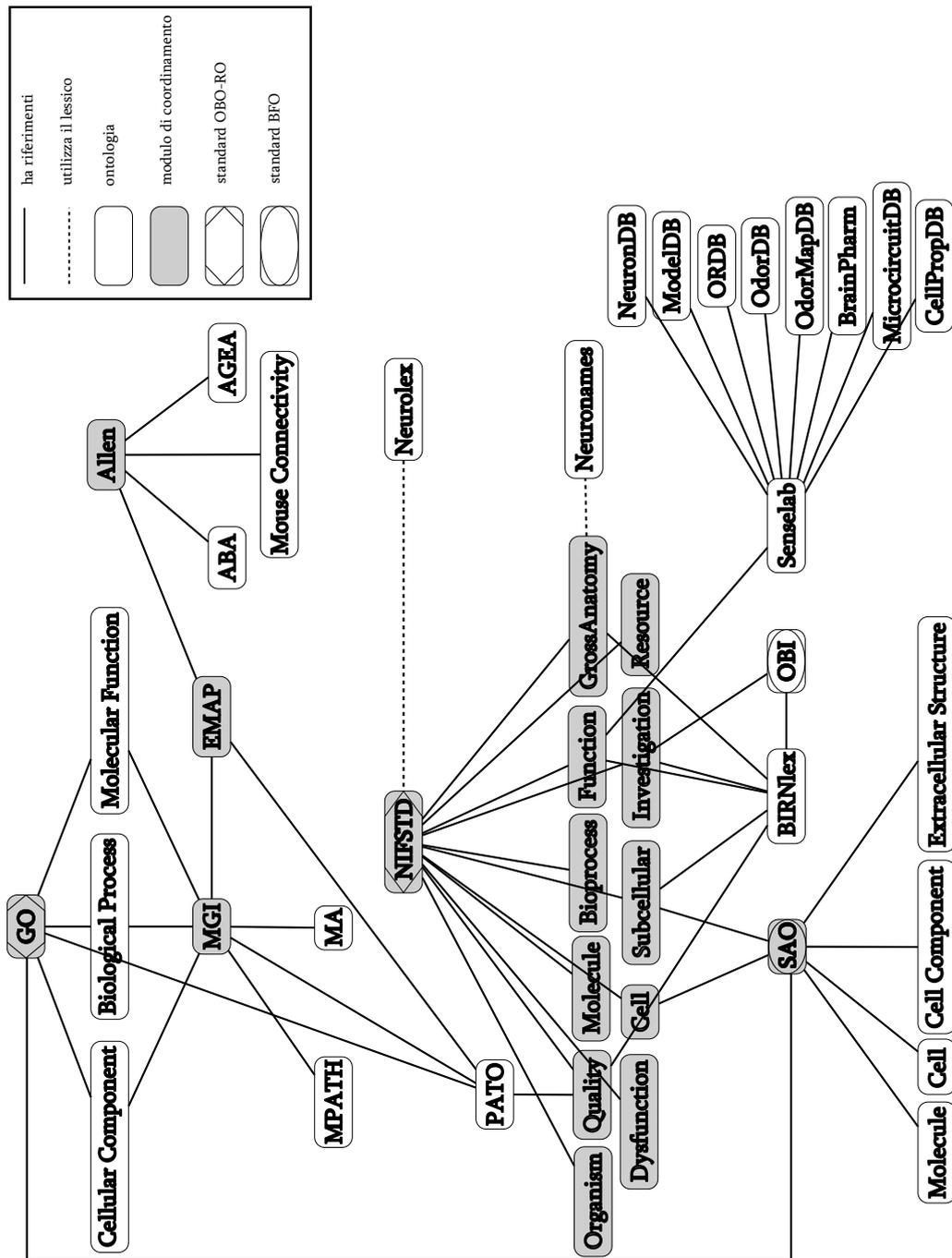


Figura 3.4: Coordinazione delle ontologie

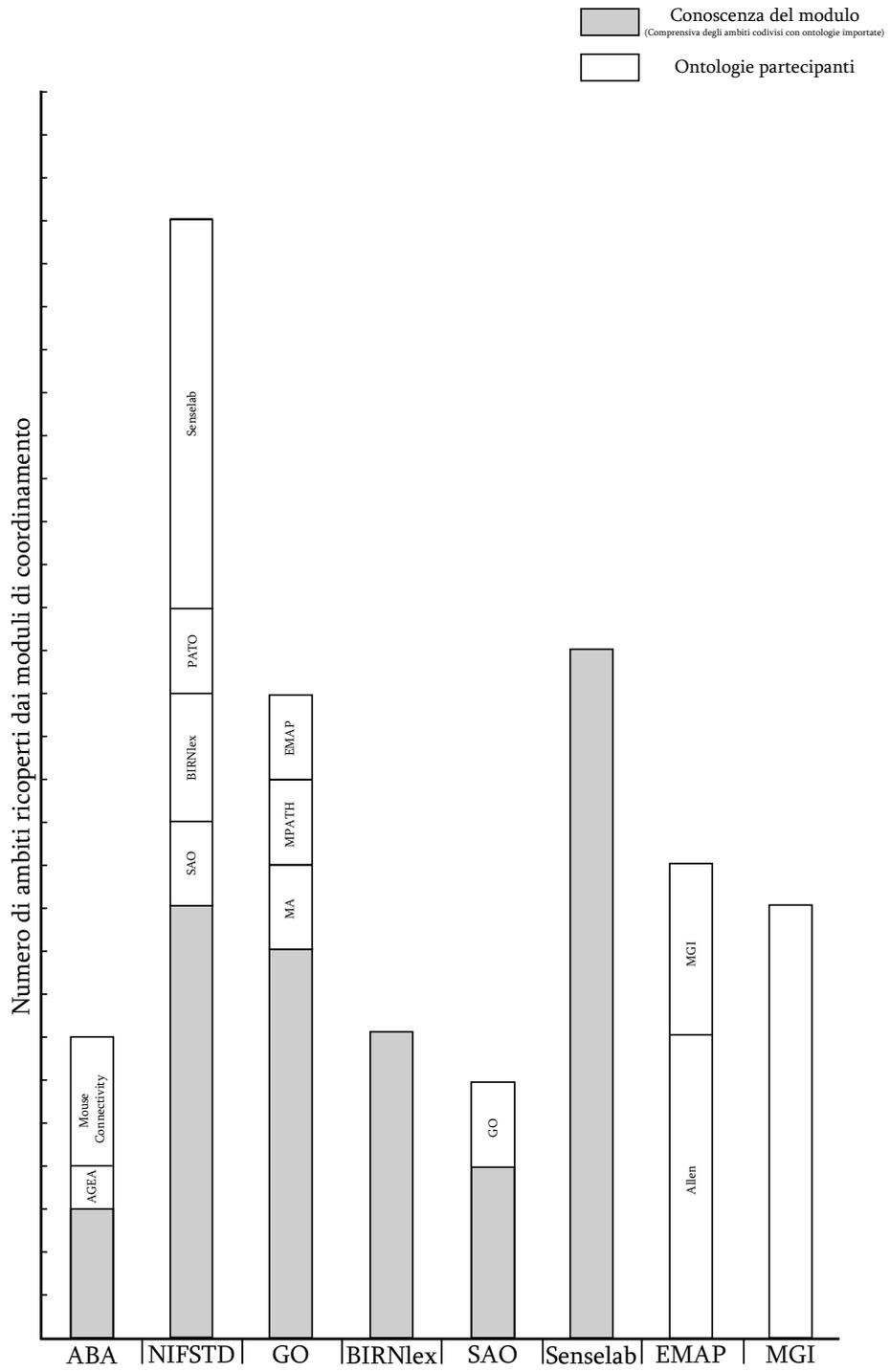


Figura 3.5: Copertura della conoscenza di dominio delle ontologie

UN ESEMPIO D'USO

Si propone di seguito un esempio d'uso delle tecniche ontologiche nel contesto neuroscientifico. A partire da un modello in cui i dati siano stati inseriti all'interno della struttura ontologica sotto forma di immagini, se ne predispone una navigazione di tipo semantico che consente di accedere a tutti i metadati di contesto in relazione alle ontologie analizzate nella tesi, inoltre si presenta una struttura iniziale per l'annotazione con i metadati strumentali relativi agli esperimenti condotti al *L.E.N.S.* tramite il microscopio *Conf-U*, che data l'originalità della tecnica di indagine non sono presenti nelle ontologie esposte e sono pertanto collocati nel contesto delle ontologie di indagine.

4.1 MOUSE ONTOLOGY

L'ontologia modella la conoscenza relativa al *mus musculus*, considerandone gli aspetti anatomici, istologici, sub-cellulari, di sviluppo, di indagine, fenotipici, funzionali, genetici e patologici; tali aspetti

Sezione dell'ontologia	Modulo importato
Gross_Anatomy	ABA
Histology	CL
Subcellular_Anatomy	SAO
Development	EMAP
Investigation	BIRN
Phenotypic_Quality	PATO
Function	Senselab
Genetics	GO
Pathology	MPATH

Tabella 4.1: Ontologie coinvolte nel modello

```

<Import>http://purl.obolibrary.org/obo/cl.owl
</Import>
<Import>http://purl.obolibrary.org/obo/mpath.owl
</Import>
<Import>http://ccdb.ucsd.edu/SA0/1.2
</Import>
<Import>http://purl.obolibrary.org/obo/emap.owl
</Import>
<Import>http://ontology.neuinfo.org/NIF/Function/NIF-Function.owl
</Import>
<Import>http://bioontology.org/projects/ontologies/birnlex
</Import>
<Import>http://purl.obolibrary.org/obo/pato.owl
</Import>
<Import>http://purl.obolibrary.org/obo/go.owl
</Import>
<Import>http://mouse.brain-map.org/atlas/index.html
</Import>

```

Figura 4.1: Riferimenti ai file OWL

sono gestiti attraverso le ontologie esposte nella tesi importandone i moduli *OWL* attraverso Protégé (figura 4.1); la struttura delle informazioni consente quindi un accesso intuitivo ai metadati di contesto (tabella A.1) contenuti nei moduli importati.

La sezione relativa alle tecniche di indagine quali condizioni sperimentali, protocolli, trattamento del campione è gestita dal modulo *BIRN* ed è all'interno di questa che modelliamo le sottoclassi delle entità Laser scanning confocal microscope (figura 4.2-a) e Laser scanning confocal imaging protocol (figura 4.2-b), riguardanti rispettivamente il tipo di microscopio confocale e il protocollo di preparazione del campione in un esperimento in cui viene utilizzato tale strumento; i moduli che costituiscono la specializzazione delle due classi sono *Conf-U.owl* (figura 4.7) e *PrepProtocol.owl* (figura 4.8).

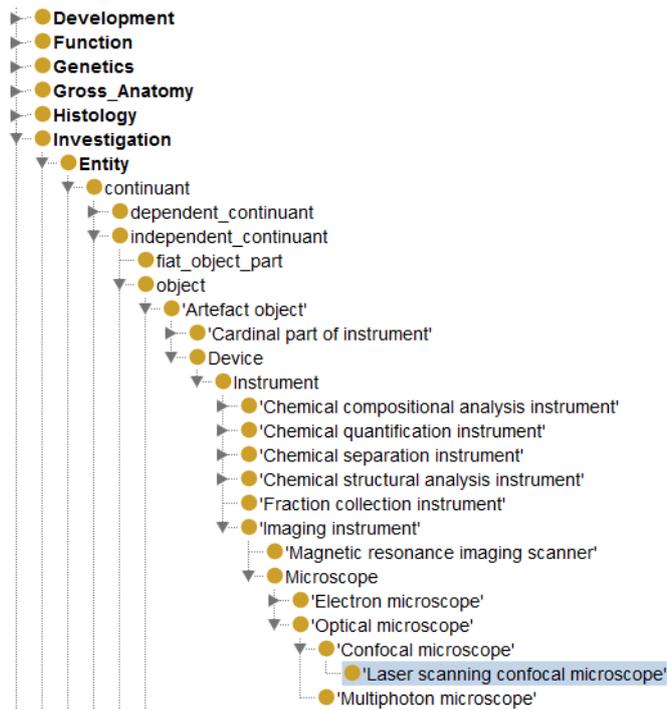


Figura 4.2-a: Collocazione delle entità da estendere (microscopio)

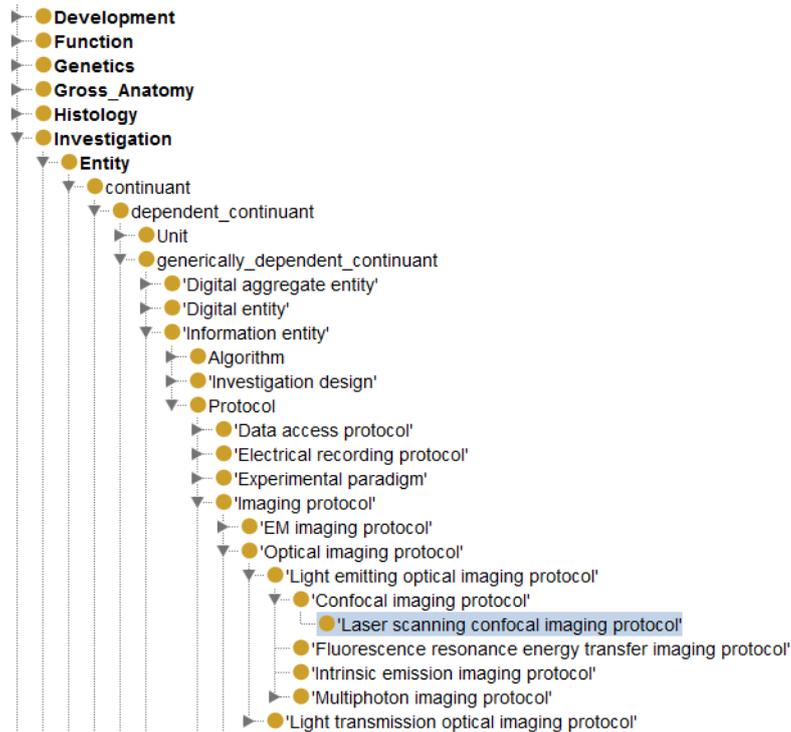


Figura 4.2-b: Collocazione delle entità da estendere (protocollo)

4.1.1 Proprietà

4.1.1.1 Annotation properties

Le Annotation properties definite nel modello riguardano principalmente la definizione dei termini tramite la proprietà `rdfs:isDefinedBy`; per garantire l'accuratezza delle informazioni si è legato ad ogni nuovo metadato la definizione in riferimento al lessico Neurolex (se presente) o a Wikipedia¹.

```

<AnnotationAssertion>
  <AnnotationProperty abbreviatedIRI="rdfs:isDefinedBy"/>
    <IRI>
      http://www.semanticweb.org/ontologies
        /2012/2/Mouse_Ontology.owl#Conf-U
    </IRI>
    <Literal
      datatypeIRI="&xsd:anyURI">http://neurolex
        .org/wiki/Category:Laser_scanning_
        confocal_microscope
    </Literal>
</AnnotationAssertion>

<AnnotationAssertion>
  <AnnotationProperty abbreviatedIRI="rdfs:isDefinedBy"/>
    <IRI>
      http://www.semanticweb.org/ontologies
        /2012/2/Mouse_Ontology.owl#ConfU_
        Voxel
    </IRI>
    <Literal
      datatypeIRI="&xsd:anyURI">http://neurolex
        .org/wiki/Category:Voxel
    </Literal>
</AnnotationAssertion>

<AnnotationAssertion>
  <AnnotationProperty abbreviatedIRI="rdfs:isDefinedBy"/>
    <IRI>
      http://www.semanticweb.org/ontologies
        /2012/2/Mouse_Ontology.owl#ConfU_
        labeling
    </IRI>

```

¹ <http://www.wikipedia.org/>

```

    <Literal
      datatypeIRI="&xsd:anyURI">http://neurolex
      .org/wiki/Category:Fluorescence_
      labeling_protocol
    </Literal>
  </AnnotationAssertion>

```

Figura 4.3: Esempi di Annotation properties definite nell'ontologia

4.1.1.2 Data properties

Le Data properties riguardano i metadati da associare alle entità che compongono il modello e derivano dal dialogo con Ludovico Silvestri (autore di *Conf-U*) che ci ha fornito i tipi di informazioni alle quali un esperto di dominio potrebbe essere interessato navigando l'ontologia.

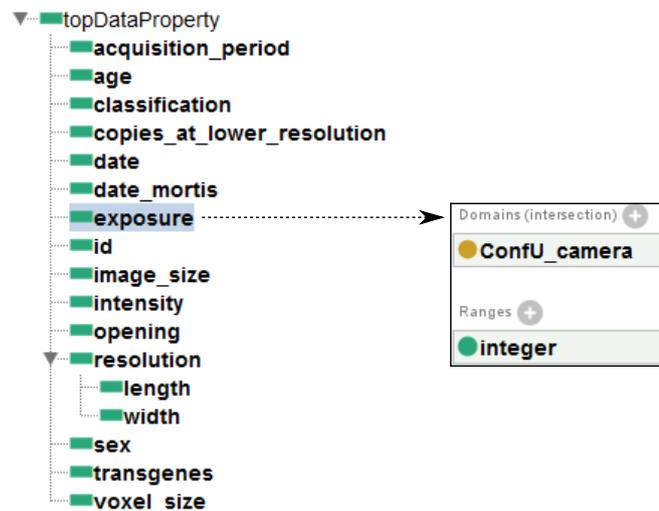


Figura 4.4: Data properties di *Conf-U*

4.1.1.3 Object properties

Le *Object properties* incluse nell'ontologia sono volte principalmente all'espressione della proprietà di inclusione tra le entità che costituiscono l'apparato *Conf-U* e riguardano le relazioni di:

component_of

Dominio: Sample_Chamber, Filter_Wheel, Galvo_Mirror, Expansion_Telescope, ConfU_laser, ConfU_slit, AOTF, Lens, EM-CCD

Codominio: ConfU

has

Dominio: Sample dissection

Codominio: ConfU_labeling

includes

Dominio: ConfU_optical_microscopy

Codominio: ConfU_laser

part_of

Dominio: ConfU_Voxel

Codominio: ConfU_image

Object properties importate

- topObjectProperty
 - adjacent_to
 - 'anatomical Location'
 - 'associated With'
 - 'bearer of'
 - belongsTo
 - 'boundary Of'
 - capable_of
 - component_of**
 - 'contains Site'
 - correlates_with
 - derives_from
 - develops_from
 - 'different_in_magnitude_relative_to'
 - directly_associated_with
 - has
 - 'has Boundary'
 - 'has Location Of'
 - 'has Part'
 - 'has Property'
 - hasMany
 - has_completed
 - has_cross_section
 - has_dividend_entity
 - has_divisor_entity
 - has_high_plasma_membrane_amount
 - has_low_plasma_membrane_amount
 - has_modification

Object properties definite

The figure shows two screenshots of a software interface. The top screenshot displays a 'Domains (intersection)' list with one item: 'Sample dissection'. Below it, a 'Ranges (intersection)' list contains one item: 'ConfU_labeling'. The bottom screenshot displays a 'Domains (intersection)' list with nine items: 'Sample Chamber', 'Filter Wheel', 'Galvo Mirror', 'Expansion Telescope', 'ConfU_laser', 'ConfU_slit', 'AOTF', 'Lens', and 'EM-CCD'. Below it, a 'Ranges (intersection)' list contains one item: 'Conf-U'. Dotted arrows from the 'component_of' property in the left list point to these two screenshots.

Figura 4.5: Object properties di *Conf-U*

4.1.1.4 Relazioni logiche

Le relazioni logiche tra le entità provengono dalla conoscenza di dominio [17] e riguardano l'utilizzo delle classi all'interno dell'ontologia.

Ad esempio la relazione tra i tipi di lenti utilizzate durante un esperimento è di disgiunzione, mentre la relazione di sottoclasse è stata utilizzata per collegare l'ontologia *ConfU* a quelle di contesto (figura 4.6).

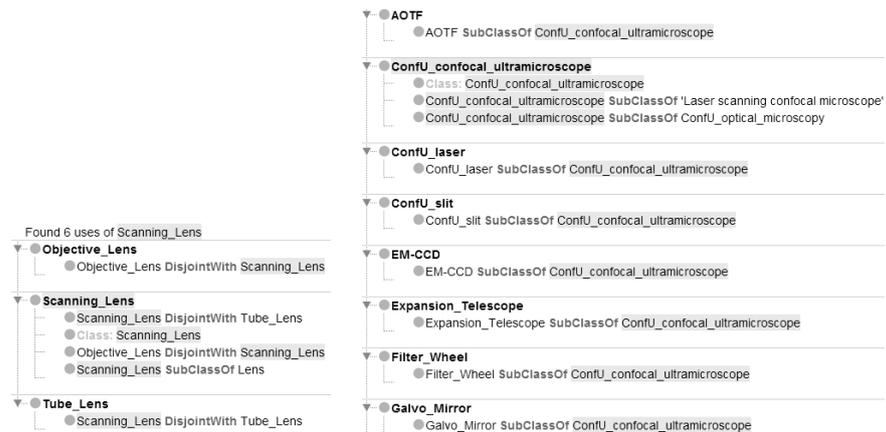


Figura 4.6: Relazioni logiche tra le classi

CONCLUSIONI E SVILUPPI SUCCESSIVI

Dallo studio sullo stato dell'arte nella modellazione ontologica nel campo delle neuroscienze emerge che il panorama delle ontologie neuroscientifiche è estremamente frammentato, coerentemente con l'eterogeneità delle informazioni che una struttura comune si propone di unificare.

Un primo passo in questo senso è costituito dalla centralizzazione del lessico: per consentire la cooperazione di più strutture che facciano riferimento ad aspetti differenti di uno stesso organismo è necessario che esse utilizzino un lessico comune, così come più esperti di dominio che facciano riferimento alla stessa entità anatomica necessitano di una flessibilità tale da potersi accordare sulla terminologia da utilizzare (ad esempio *rombencefalo* e *lobo occipitale* fanno riferimento alla stessa zona del cervello). A tale scopo si sono analizzati due tipi di lessico: Neuronames e Neurolex; il primo consiste in una definizione statica dell'anatomia cerebrale basata su *Nomina Anatomica*, il cui contributo è limitato alla definizione di uno standard tassonomico circoscritto all'anatomia macroscopica, mentre il secondo è un lessico contributivo, pertanto estendibile a prescindere dalla quantità e dalla tipologia delle informazioni derivanti dalla conoscenza di dominio che vengono immesse nel lessico dagli utenti; l'utilizzo di un lessico contributivo consente di ricoprire ambiti diversi e di fornire accesso alle definizioni e ai sinonimi sia di entità anatomiche (ad es. *cervelletto*) che funzionali (ad es. *malattia di Huntington*) e strumentali (ad es. *microscopio ottico*),... ed è pertanto l'alternativa più vicina agli scopi del progetto Human Brain, che si propone una contestualizzazione interdisciplinare delle cellule neuronali.

Per fornire un modello ontologico dei neuroni visibili nelle immagini ottenute attraverso *Conf-U* che sia ben contestualizzato è necessario riferirsi alle ontologie analizzate nello stato dell'arte, le quali si basano su tre ontologie che ricoprono gli ambiti principali della conoscenza di dominio:

L'ontologia più estesa, sia per apporto individuale alla conoscenza che per la maggiore integrazione con le ontologie settoriali, è l'ontologia del progetto *NIF* (figura 3.5). *NIFSTD* è suddivisa in moduli che gestiscono le strutture relative a: tassonomia organica, cellule (*SAO*, *Senselab*), molecole e strato sub-cellulare (*SAO*), anatomia macroscopica (*BIRN*), aspetto funzionale del sistema nervoso (*Senselab*), disfunzioni del sistema nervoso, qualità fenotipiche (*PATO*), processi biologici (*GO*), paradigmi cognitivi, strumenti - protocolli - reagenti (*BIRN*, *OBI*), e dati relativi al campione; l'unificazione della conoscenza trattata dai moduli rende la navigazione dei modelli trasparente per l'utente che servendosi di *un'unica* ontologia accede a informazioni di ambiti correlati in base al tipo di dati immessi nell'ontologia. Dal punto di vista del lessico l'ontologia *NIFSTD* integra entrambi i vocabolari Neurolex e Neuronames, destinando il secondo all'anatomia macroscopica e ampliando il primo con i termini derivanti dai settori coperti dalle ontologie specifiche.

Uno dei settori più vasti nella contestualizzazione cellulare è sicuramente quello della genetica, la cui conoscenza è trattata dalla *Gene Ontology*, che formalizza le relazioni esistenti nella genomica di organismi di tipo diverso e consente l'annotazione con metadati non forniti esplicitamente; la sezione di *GO* utile agli scopi del progetto Human Brain è quella relativa al *mus musculus*, la cui gestione è curata da *Mouse Genome Informatics (MGI)*, che copre la conoscenza relativa all'aspetto delle patologie, della genetica, del trattamento e delle procedure di analisi utilizzate negli esperimenti condotti ai *Jackson Labs*.

La sezione relativa all'aspetto istologico è coperta in gran parte dalla *Subcellular Anatomy Ontology (SAO)*, la cui struttura è basata su quella della *Basic Formal Ontology (BFO)*, che divide le entità descritte in base a relazioni spaziali (continuant, occurrent) e di attinenza (dependent, independent).

Il passo successivo a partire dallo studio delle architetture esposte è la progettazione di un modello interoperante che accolga le immagini del *L.E.N.S.* e ne consenta la navigazione semantica garantendo l'accesso a tutti i metadati di domini differenti e la relativa inferenza; ad esempio definendo come regola di inferenza la relazione tra una determinata conformazione istologica (accesso tramite strutture *SAO*) e una particolare patologia (accesso tramite strutture *MPATH*)

è possibile, dopo l'inserimento dei dati nella struttura, fare inferenza sulla presenza della patologia nell'organismo immesso sotto forma di immagini in base alla morfologia delle cellule.

A partire dallo stato dell'arte è quindi necessario:

- definire una struttura indipendente dai dati nella quale più tipi di dati e le relazioni tra di essi possano essere inseriti a partire dal dialogo con gli esperti di dominio e dalla letteratura neuroscientifica;
- rendere accessibile il modello ontologico agli esperti di dominio sia per la navigazione delle immagini che per l'estensione della conoscenza settoriale tramite l'annotazione dei dati (ad esempio sovrapponendo il livello delle immagini a quello tassonomico attraverso il riferimento a un atlante stereotassico);
- definire nuove relazioni tra aspetti anatomici e funzionali (ad esempio tra malformazioni e patologie) in base all'interazione con esperti di dominio, in modo da inferire nuova conoscenza a partire dalle relazioni esplicite e dalle analogie tra i campioni sperimentali (immagini);
- definire i tipi di domanda a cui trovare risposte attraverso l'ontologia (regole di inferenza);
- estendere il modello proposto contenente i riferimenti alle ontologie di contesto analizzate e i metadati relativi agli esperimenti condotti al *L.E.N.S.* (condizioni sperimentali, trattamento del campione, caratteristiche del microscopio confocale,...)

Per conseguire gli obiettivi preposti è essenziale il dialogo con gli esperti di dominio, in quanto una modellazione di tipo ontologico consiste principalmente nella formalizzazione della conoscenza dei settori analizzati, che in parte deriva dalla letteratura e in parte dall'esperienza degli specialisti, che attraverso strumenti affini alla propria competenza sono grado di esaminare e di estendere la struttura generando conoscenza a partire dall'annotazione dei dati e dalla definizione di nuove relazioni tra le entità che compongono il modello.

APPENDICE: ESTRATTI DALLE ONTOLOGIE

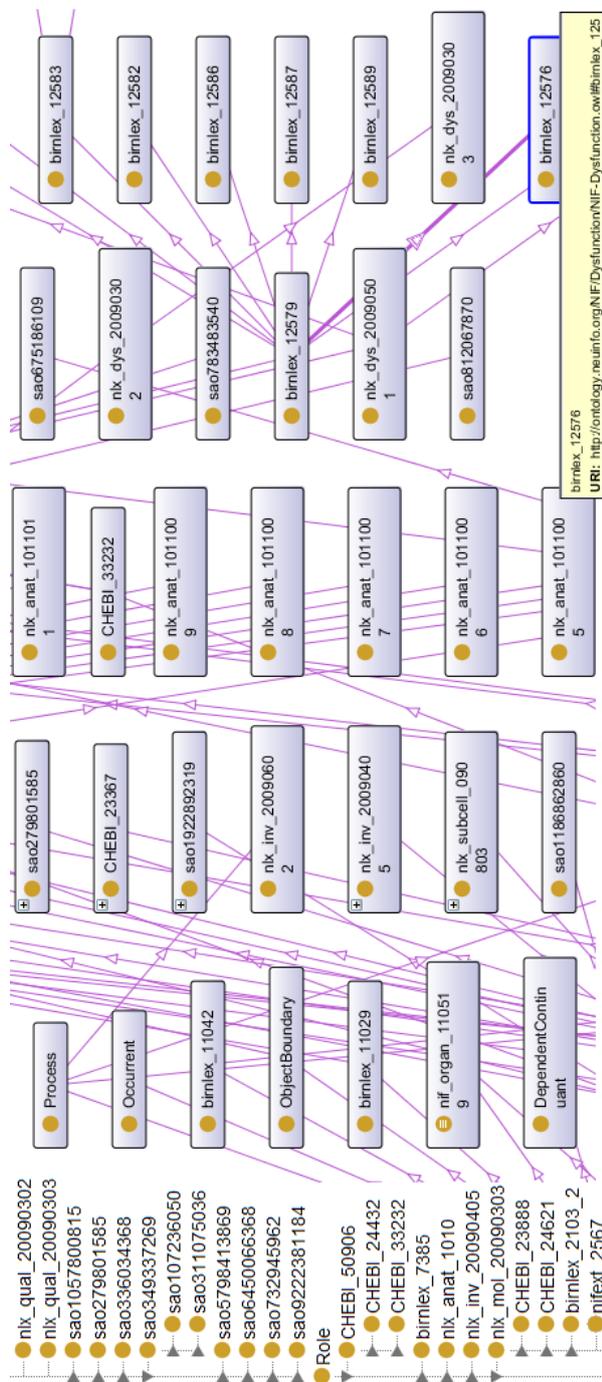


Figura A.1: Estratto dall'ontologia NIFSTD

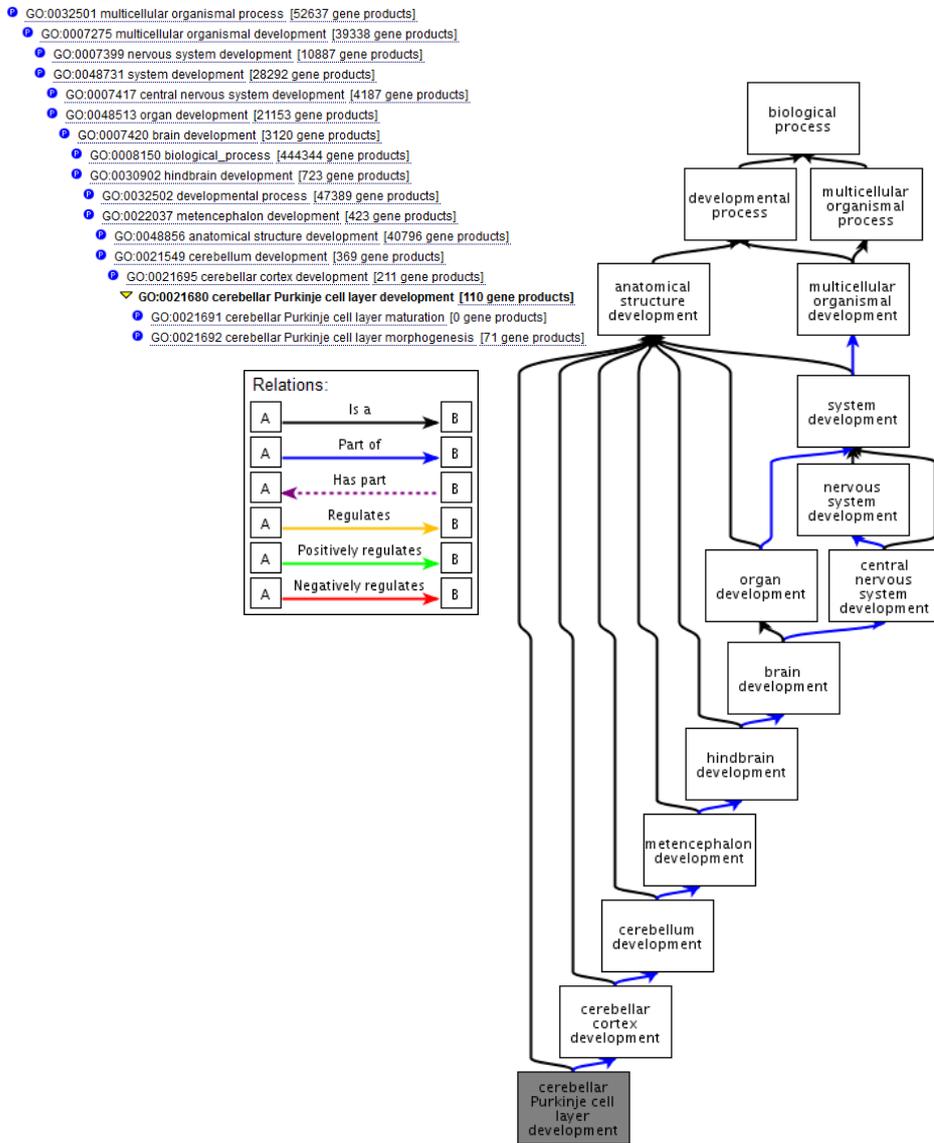


Figura A.2: Estratto dall'ontologia GO

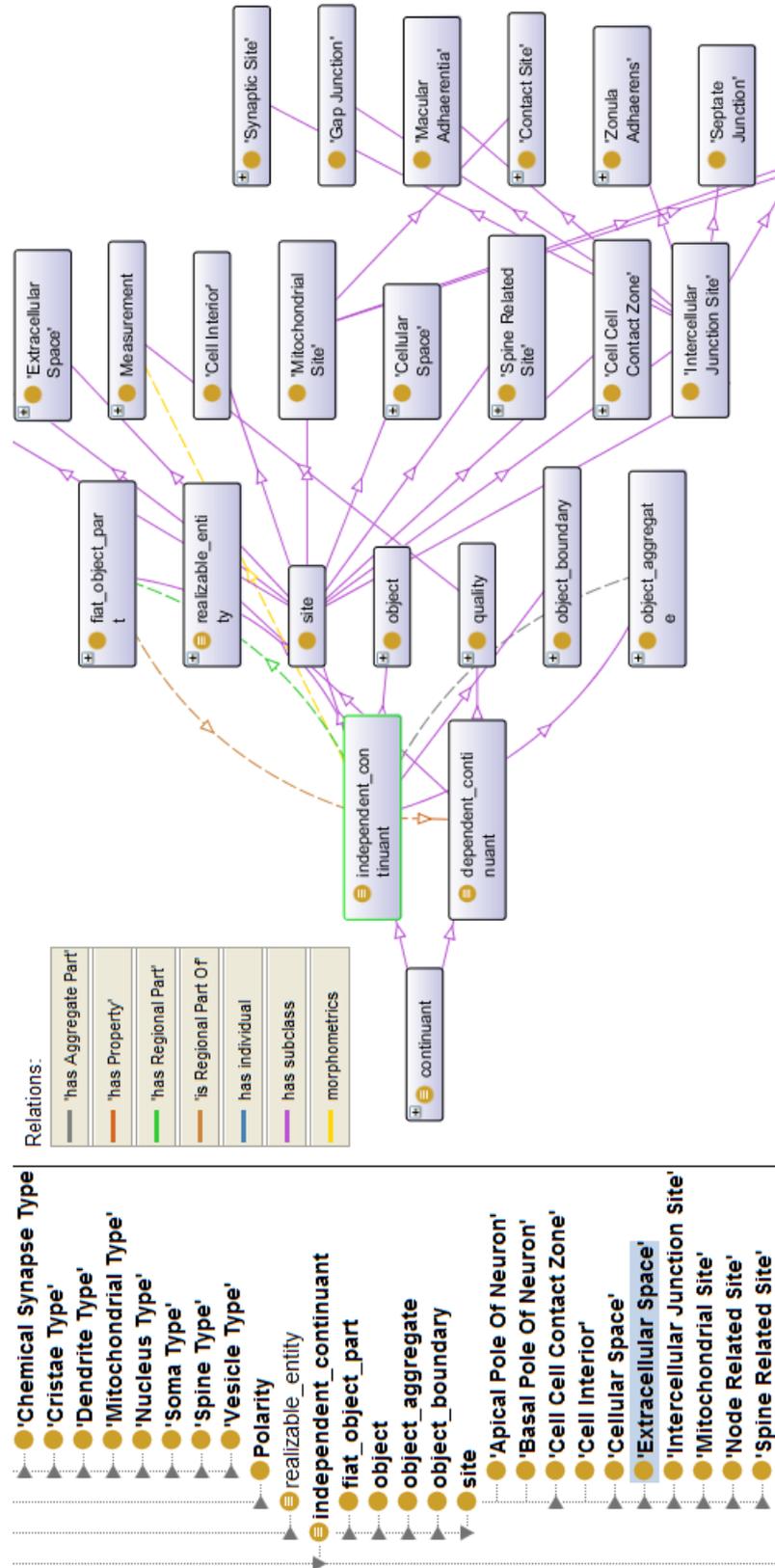


Figura A.3: Estratto dall'ontologia SAO

NEURONAMES BRAIN HIERARCHY

TABLE 1 — *Continued*

1	2	3	4	5	6	7	8	9	Abbreviation	
									lateral tuberal nuclei	LTu
									dorsal hypothalamic area	DH
									postcommissural fornix	pofx
									dorsal longitudinal fasciculus of hypothalamus	dlfh
									medial forebrain bundle	mfbb
									SUBTHALAMUS	
									subthalamic nucleus	STh
									nucleus of ansa lenticularis	AL
									nucleus of field H	FH
									zona incerta	ZI
									field H1	H1
									field H	H
									field H2	H2
									capsule of subthalamic nucleus	csTh
									pallidohabenular bundle	phb
									ansa lenticularis	al
									ansa peduncularis	ap
									THIRD VENTRICLE	3V
									<i>interventricular foramen</i>	<i>ivf</i>
									<i>tenia of thalamus</i>	<i>tth</i>
									<i>hypothalamic sulcus</i>	<i>hys</i>
									<i>habenular sulcus</i>	<i>hbs</i>
									velum interpositum	vi
									choroid plexus of third ventricle	cp3v
									tela choroidea of third ventricle	tc3v
									subfornical organ	SFO
									suprapineal recess	SPiRe
									pineal recess	PiRe
									optic recess	ORe
									infundibular recess	IRe
									optic chiasm	ox
									optic tract	opt
									habenulo-interpeduncular tract	hip
									MIDBRAIN	MB
									<i>dorsolateral fissure of midbrain</i>	<i>dlfmb</i>
									TECTUM	Tec
									<i>trochlear nerve</i>	<i>4n</i>
									PRETECTAL REGION	PTec
									nucleus of pretectal area	PTA
									principal pretectal nucleus	PPt
									nucleus of optic tract	OT
									sublentiform nucleus	SLn
									olivary pretectal nucleus	OPt
									SUPERIOR COLLICULUS	SC
									zonal layer of superior colliculus	Zo
									superficial gray layer of superior colliculus	SuG
									superficial white layer of superior colliculus	SuWh
									STRATUM LEMNISCII OF SUPERIOR COLLICULUS	StL

Figura A.4: Estratto dalla nomenclatura NeuroNames

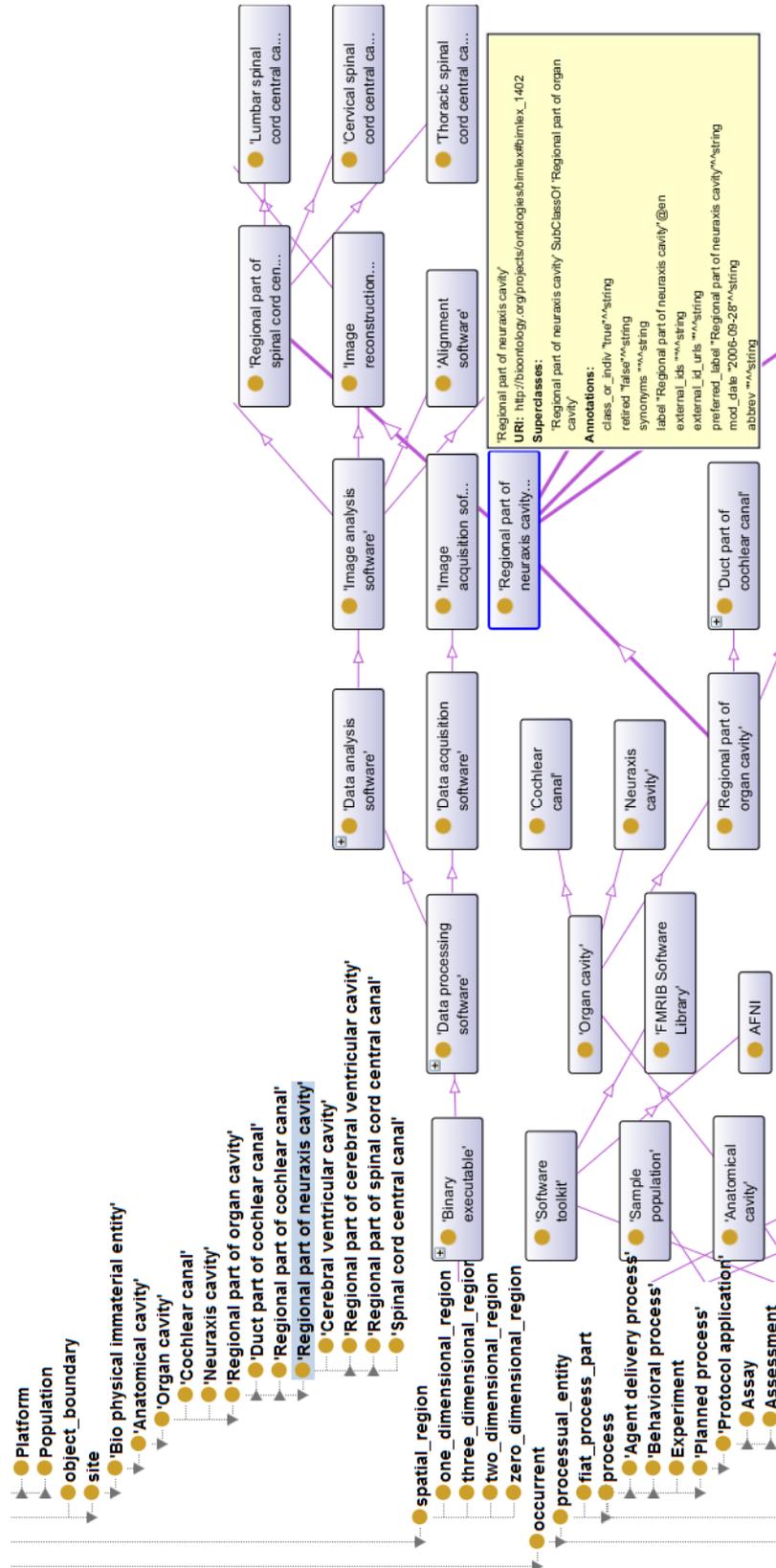


Figura A.6: Estratto dall'ontologia *BIRNlex*

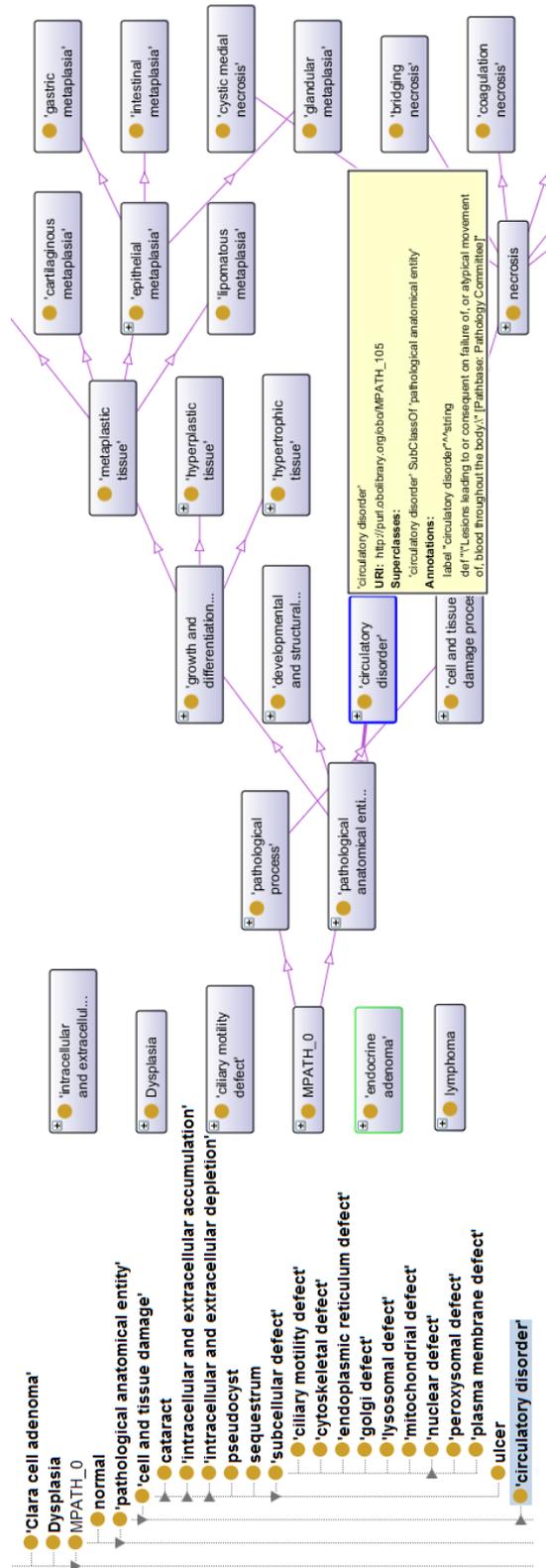


Figura A.7: Estratto dall'ontologia MPATH

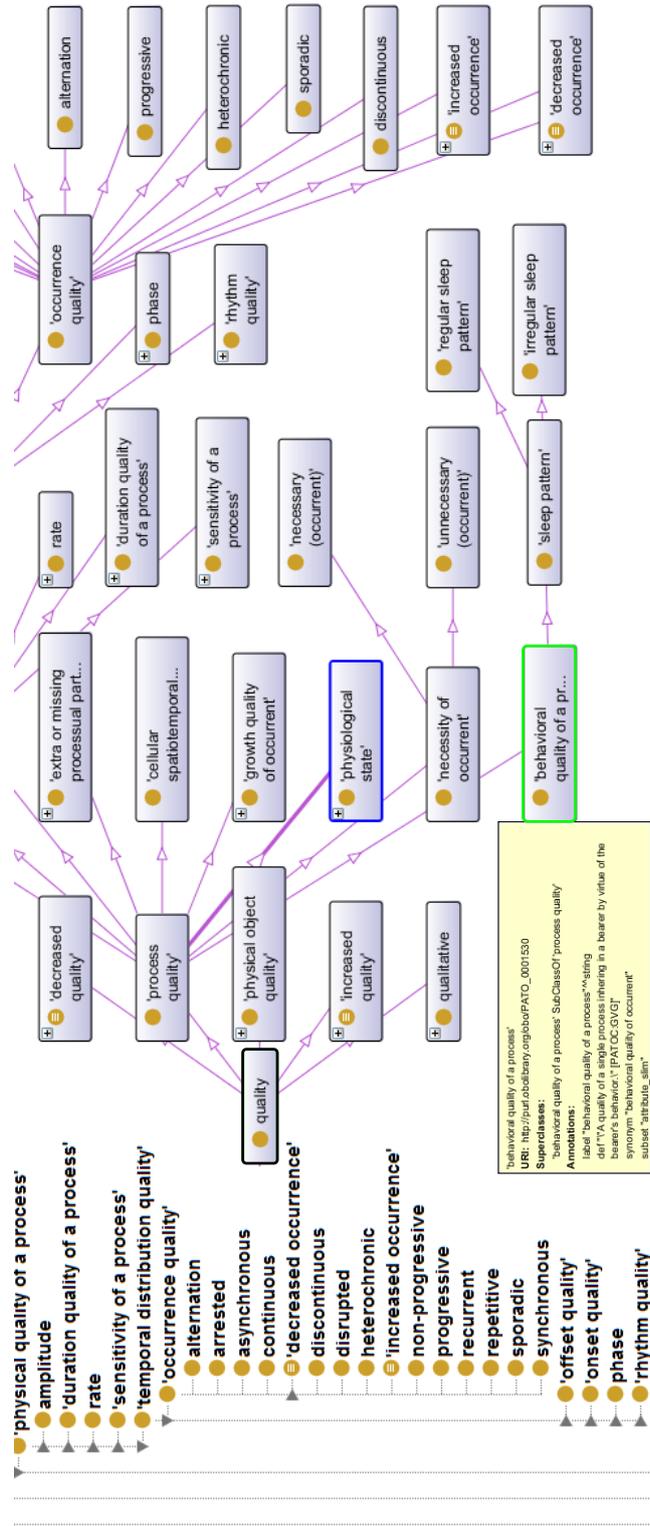


Figura A.8: Estratto dall'ontologia PATO

BIBLIOGRAFIA

- [1] Accademia della Crusca. *Vocabolario degli accademici* vol. 11, V edizione, 1923
- [2] Dean Allemang, James Hendler, *Semantic web for the working ontologist: modeling in RDF, RDFS and OWL*, II edizione, 2008
- [3] Larry L. Peterson, Bruce S. Davie, *Reti di calcolatori*, III edizione, 2004
- [4] Jack L. Lancaster et al, *Automated Talairach Atlas Labels For Functional Brain Mapping*, Research Imaging Center, University of Texas Health Science Center at San Antonio, 2000
- [5] Charles Watson, George Paxinos, Luis Puelles, *The Mouse Nervous System*, Academic Press, 2011
- [6] Olivia F. O'Leary, John F. Cryan, *The Behavioral Genetics of Serotonin: Relevance to Anxiety and Depression*, Handbook of Behavioral Neuroscience vol.21, 2010
- [7] Lein, E.S. et al, *Genome-Wide Atlas of Gene Expression in the Adult Mouse Brain*, Nature 445, 2007
- [8] Paul G. Allen et al, *Allen Mouse Brain Atlas* [Internet], Seattle (WA): Allen Institute for Brain Science ©2009
- [9] Keith B.J. Franklin, George Paxinos, *The Mouse Brain in Stereotaxic Coordinates*, III edizione, 2007
- [10] Fahim T. Imam, *NIFSTD and NeuroLex: A Comprehensive Neuroscience Ontology Development Based on Multiple Biomedical Ontologies and Community Involvement*, Working with Multiple Biomedical Ontologies Workshop, 2011
- [11] William J. Bug et al, *The NIFSTD and BIRNlex Vocabularies: Building Comprehensive Ontologies for Neuroscience*, Neuroinformatics, 2008

- [12] Stephen D. Larson et al, *A formal ontology of subcellular neuroanatomy*, Frontiers in Neuroinformatics, 2007
- [13] Douglas M. Bowden, Mark F. Dubach, *Neuronames 2002*, Neuroinformatics vol. 1 ©2003
- [14] Chiquito J. Crasto et al, *Senselab: New developments in disseminating neuroscience information*, Brief Bioinform, 2007
- [15] Perry L. Miller et al, *A Pilot Model of the Human Brain Project Approach*, Journal of the American Medical Informatics Association, 2001
- [16] P. N. Schofield et al, *Pathbase and the MPATH Ontology: Community Resources for Mouse Histopathology*, Veterinary Pathology vol. 47, 2010
- [17] Ludovico Silvestri, *Confocal ultramicroscopy: micron-scale neuroanatomy of the entire mouse brain*, PhD Thesis, 2012
- [18] Georgios V. Gkoutos et al, *Building Mouse Phenotype Ontologies*, Pacific Symposium on Biocomputing, 2004
- [19] Georgios V. Gkoutos et al, *Using ontologies to describe mouse phenotypes*, Genome Biology, 2004
- [20] Paul G. Allen et al, *Allen Mouse Brain Connectivity Atlas* [Internet], Seattle (WA): Allen Institute for Brain Science ©2012