

Guida all'utilizzo di ToscanaOpenResearch

Qual è lo scopo di questa guida?	1
Cos'è il progetto ToscanaOpenResearch?	1
Qual è l'origine delle visualizzazioni TOR?	2
Quali dati sono disponibili su ToscanaOpenResearch?	4
Che dati possono essere esplorati su TOR?	4
Una visione semplificata dei dati tramite mind map	4
Come si interrogano i dati?	9
Cos'è lo SPARQL Endpoint e come funziona?	9
L'ontologia TOR	10
Come leggere il diagramma dell'ontologia?	10
Approfondimento - I vantaggi di un accesso ai dati basato su ontologie di dominio	11
Come scrivere le query per esplorare e scaricare dati?	12

Qual è lo scopo di questa guida?

Questo manuale di utilizzo offre una panoramica delle informazioni contenute nel portale [ToscanaOpenResearch](#) e si propone di presentare ad un pubblico senza competenze specifiche le modalità di consultazione e accesso ai dati sulla Ricerca, Innovazione e Alta Formazione in Toscana in esso reperibili.

Con il fine quindi di accrescere l'utilità e di estendere l'utilizzo di ToscanaOpenResearch, questo documento descrive il progetto che è alla base di questo portale, sottolineandone la ricchezza e varietà dei dati contenuti e illustrando il modo in cui è possibile interrogare il sistema per ottenere le informazioni desiderate.

Per qualsiasi commento o domanda riguardo ToscanaOpenResearch, si può mandare una mail al seguente indirizzo: staff@toscanaopenresearch.it.

Cos'è il progetto ToscanaOpenResearch?

Nel contesto del proprio Osservatorio Regionale della Ricerca e Innovazione, costituito con L.R. 20/2009 al fine di disporre di una base di conoscenza qualificata e di uno strumento informativo di supporto alle politiche, la Regione Toscana ha realizzato ToscanaOpenResearch (TOR) con una duplice finalità:

- Comunicare e valorizzare l'ecosistema dell'alta formazione e ricerca regionale, favorendo una governance sempre più trasparente e inclusiva;
- Analizzare l'ecosistema dell'alta formazione e ricerca regionale per avviare un

processo di orientamento delle politiche regionali in un'ottica data-driven.

TOR rappresenta quindi lo strumento informativo dell'Osservatorio, ed è costituito da differenti componenti:

- vari set di dati sulla ricerca, l'innovazione e l'alta formazione: i dati sono presentati in dettaglio nelle sezioni “[Quali dati sono disponibili su ToscanaOpenResearch?](#)” e “[Che dati possono essere esplorati su TOR?](#)”;
- un'ontologia di dominio, presentata nella sezione “[L'ontologia TOR](#)”;
- un endpoint SPARQL per esplorare e scaricare i dati, presentato nella sezione “[Cos'è lo SPARQL Endpoint e come funziona?](#)”;
- delle visualizzazioni interattive che rispondono a domande sull'[alta formazione](#), la [ricerca e l'innovazione](#) e le [collaborazioni](#) di ricerca in Toscana. Le visualizzazioni permettono di esplorare facilmente i dati e di scaricare direttamente i dati visualizzati in formato CSV, e sono presentate nella sezione “[Qual è l'origine delle visualizzazioni TOR?](#)”.

La componenti di ToscanaOpenResearch sono presentate schematicamente nella figura in basso:



Schema delle componenti di ToscanaOpenResearch.

Qual è l'origine delle visualizzazioni TOR?

Il dashboard TOR, ossia la sezione “[Esplora i dati](#)” che ospita le visualizzazioni, è costruito con **tecnologie open source**.

Le visualizzazioni sono basate su dati direttamente recuperati attraverso l'Endpoint SPARQL. Pertanto, ogni volta che i dati verranno aggiornati, **anche le visualizzazioni si aggiorneranno automaticamente**.

Le visualizzazioni che compaiono su ToscanaOpenResearch nella sezione “[Esplora i dati](#)”

sono state create per rispondere alle domande prodotte dalla **Conferenza Regionale per la Ricerca e l'Innovazione** tramite un [processo di co-design](#), sviluppatosi in 3 sessioni tra il 2017 e il 2018.

La Conferenza Regionale per la Ricerca e l'Innovazione è una struttura collegiale permanente con funzioni consultive, all'interno della quale siedono i rappresentanti di università, centri di ricerca, parchi scientifici e tecnologici, imprese e sindacati. A tal proposito, durante la fase di condivisione del Cruscotto Pilota, erano emersi diversi quesiti.

Tra i **quesiti** posti dalla Conferenza Regionale, a cui si è dato risposta tramite visualizzazioni interattive nel portale, troviamo:

1. Mappatura dei corsi di dottorato di ricerca

Una panoramica dei corsi di dottorato di ricerca in Toscana viene delineata da queste due visualizzazioni: [la prima](#) mostra il numero di iscritti e diplomati ai dottorati di ricerca in Toscana, per anno e per ateneo; [la seconda](#), fornisce informazioni sugli iscritti e i diplomati ai dottorati di ricerca, distinti per anno, ateneo, genere e titolo del dottorato.

2. Mappatura dei corsi di Master

Per rispondere a questa domanda, ToscanaOpenResearch include due visualizzazioni: [la prima](#) indica il numero di iscritti e diplomati ai master in Toscana distinti per anno e per università, tenendo in considerazione sia dei master di primo che di secondo livello; mentre [la seconda](#) riporta il numero di studenti di master in Toscana per ateneo, università, genere e titolo del Master.

3. Panoramica degli spin-off della ricerca pubblica in Toscana

[Questa visualizzazione](#) mostra una visione d'insieme degli spin-off della ricerca pubblica toscani, nella quale vengono riportate le seguenti informazioni: anno di fondazione; settore ATECO; capitale iniziale e lo stato dello spin-off.

4. Uno sguardo generale sulle reti di collaborazione che emergono dall'analisi dei progetti di ricerca e innovazione a finanziamento europeo

Per rispondere a questa domanda è stata creata un'[apposita visualizzazione](#), che mostra le collaborazioni degli attori toscani sui progetti di ricerca e innovazione a finanziamento europeo, sia con collaboratori italiani che con collaboratori internazionali. La mappa permette di filtrare per tipo di programma, per il paese collaboratore e per il tipo di attività.

Per ognuna delle visualizzazioni sopra menzionate, come per tutte quelle presenti all'interno del portale, è possibile scaricare direttamente i dati rappresentati all'interno della visualizzazione, in formato csv, cliccando sulla freccia rossa riportata subito sotto la rappresentazione.

Quali dati sono disponibili su ToscanaOpenResearch?

I dati aperti all'interno di ToscanaOpenResearch sono il risultato dell'integrazione di

numerose banche dati; protocolli di collaborazione e dati provenienti da database bibliometrici¹.

In particolare,

- sono disponibili dati provenienti da **banche dati open regionali, nazionali, europee e mondiali**;
- sono stati integrati una serie di ulteriori dati attraverso protocolli di collaborazione (es. Ministero dell'Istruzione, dell'Università e della Ricerca - MIUR, per l'integrazione dei dati inerenti ai Cluster Tecnologici Nazionali - CTN 2012 e 2016 e ai Progetti di ricerca di Rilevante Interesse Nazionale - PRIN 2012);
- per la parte relativa alle pubblicazioni, il sistema utilizza banche dati bibliometriche non open, ma è già predisposto per una facile integrazione con i dati di [CINECA-IRIS](#)².

Il sistema è attualmente **dedicato al perimetro Toscano**, però **per la maggior parte dei dati è già stato integrato il perimetro nazionale**, ed è pertanto possibile realizzare analisi di contesto sul piano nazionale o addirittura europeo (nel caso dei progetti di ricerca).

Che dati possono essere esplorati su TOR?

Le seguenti “mind map” offrono uno spaccato delle informazioni disponibili a seconda del soggetto o della tematica di interesse.

- **Soggetto/Tematica:**
 - Studenti
 - Personale accademico
 - Progetti
 - Europei
 - Italiani (CTN, PRIN 2012)
 - Erasmus+
 - CHAFEA
 - Spin-off accademici
 - Sperimentazioni cliniche
 - Dati sul diritto allo studio
 - Dati sul bilancio di genere

Una visione semplificata dei dati tramite mind map

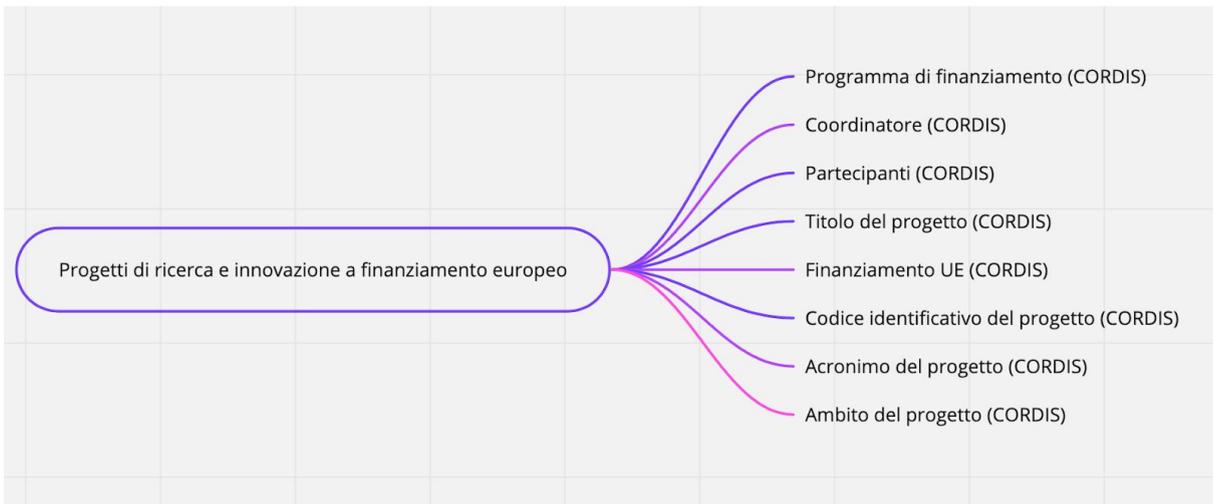
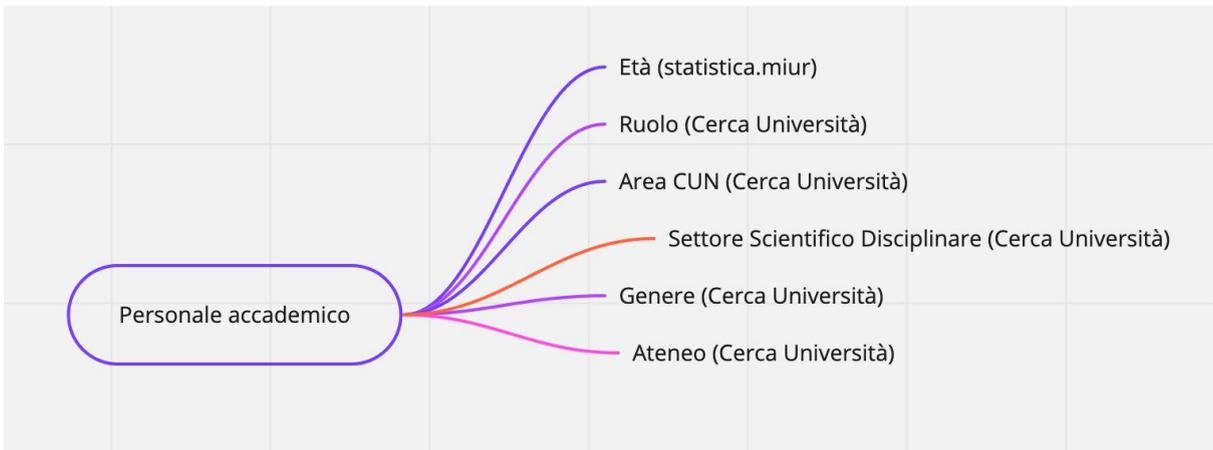
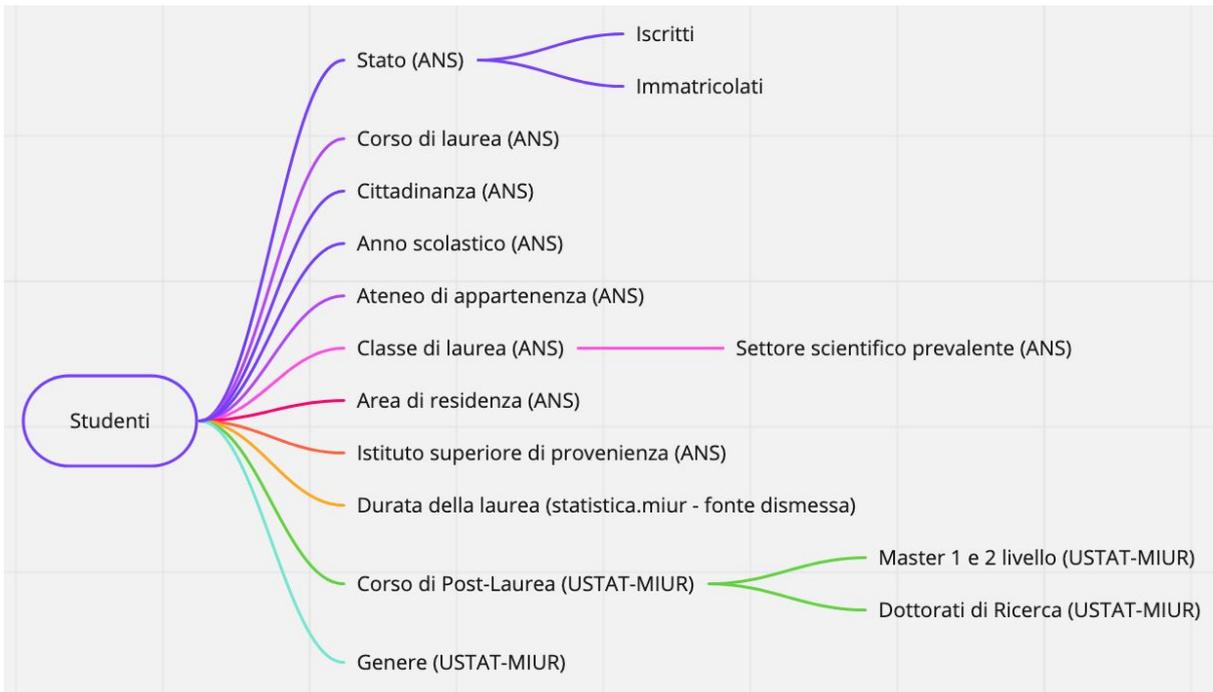
L'idea delle “mind map” sotto rappresentate è quella di delineare un **quadro generale dei dati disponibili su ToscanaOpenResearch**, suddividendoli per tematica.

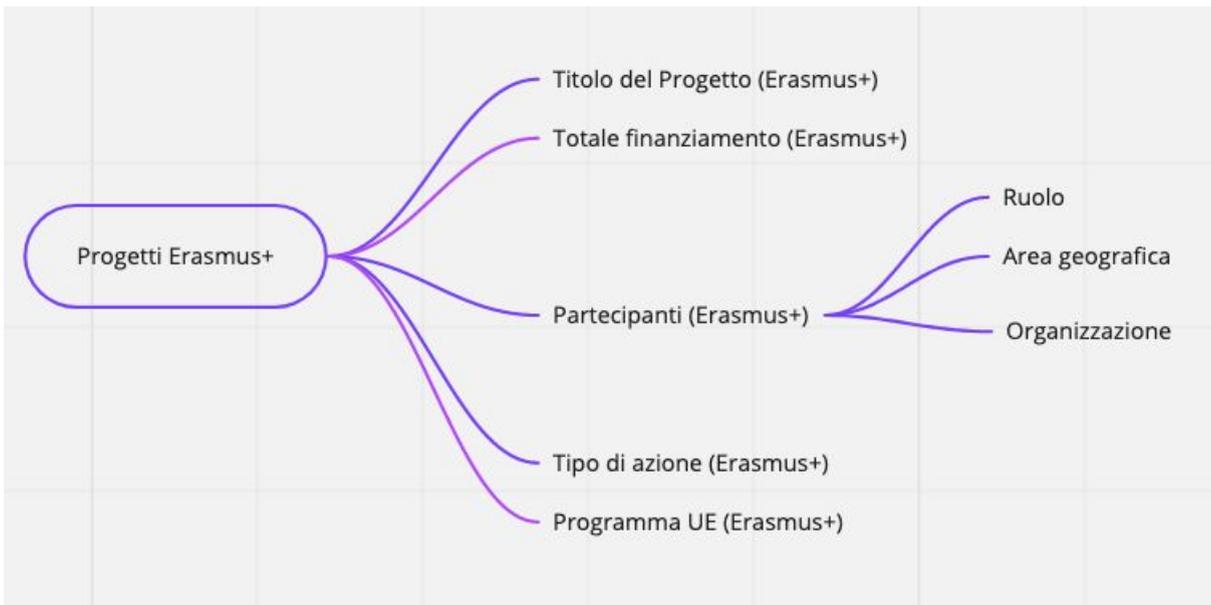
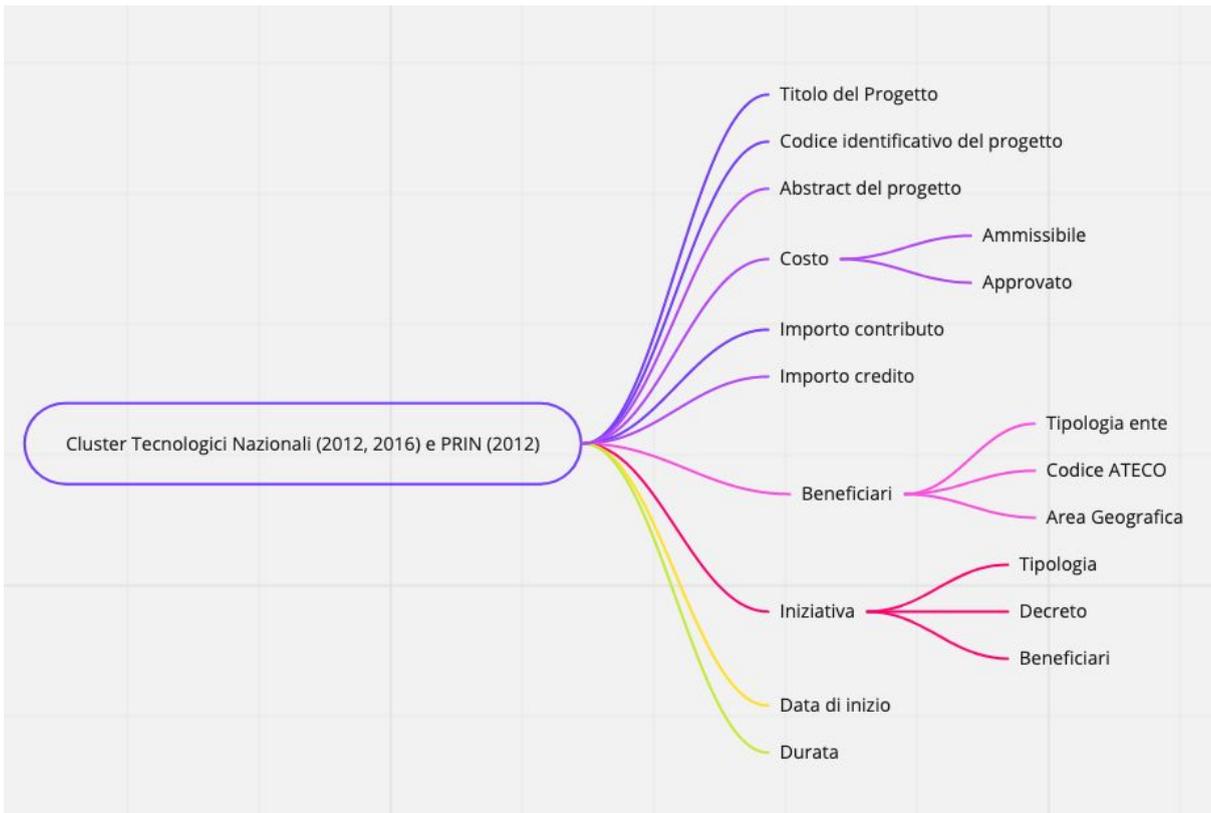
Infatti, la parola che si trova al centro di ogni map (all'interno della casella con il bordo arrotondato) rappresenta il soggetto di interesse, la tematica dalla quale è possibile ottenere le informazioni indicate da ogni freccia.

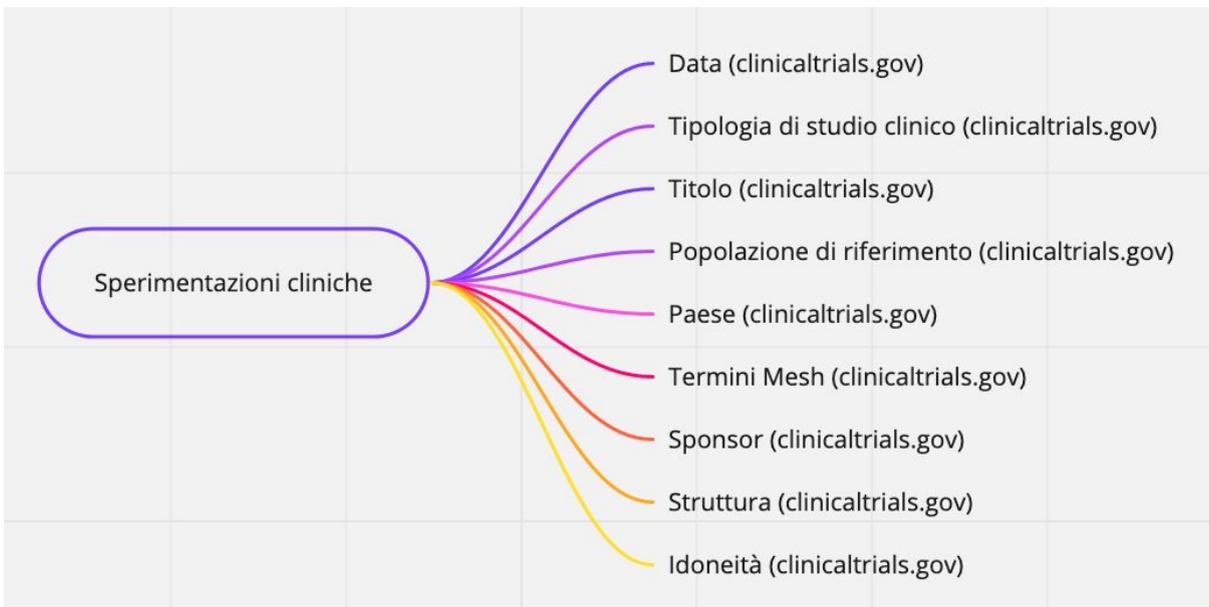
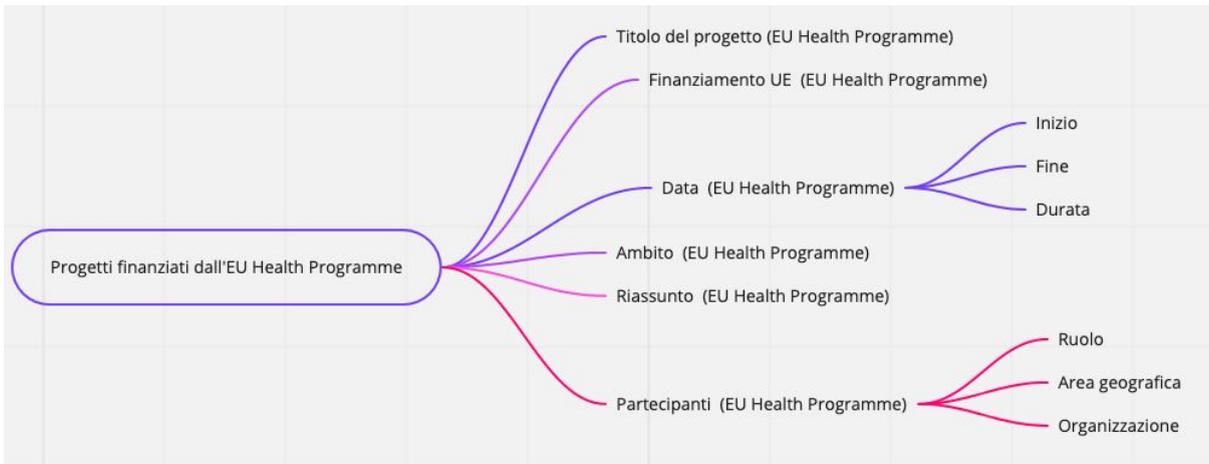
La fonte del dato viene riportata tra parentesi, ad eccezione dei dati CTN/PRIN 2012 che sono stati ottenuti tramite un protocollo di intesa tra il MIUR e la Regione Toscana.

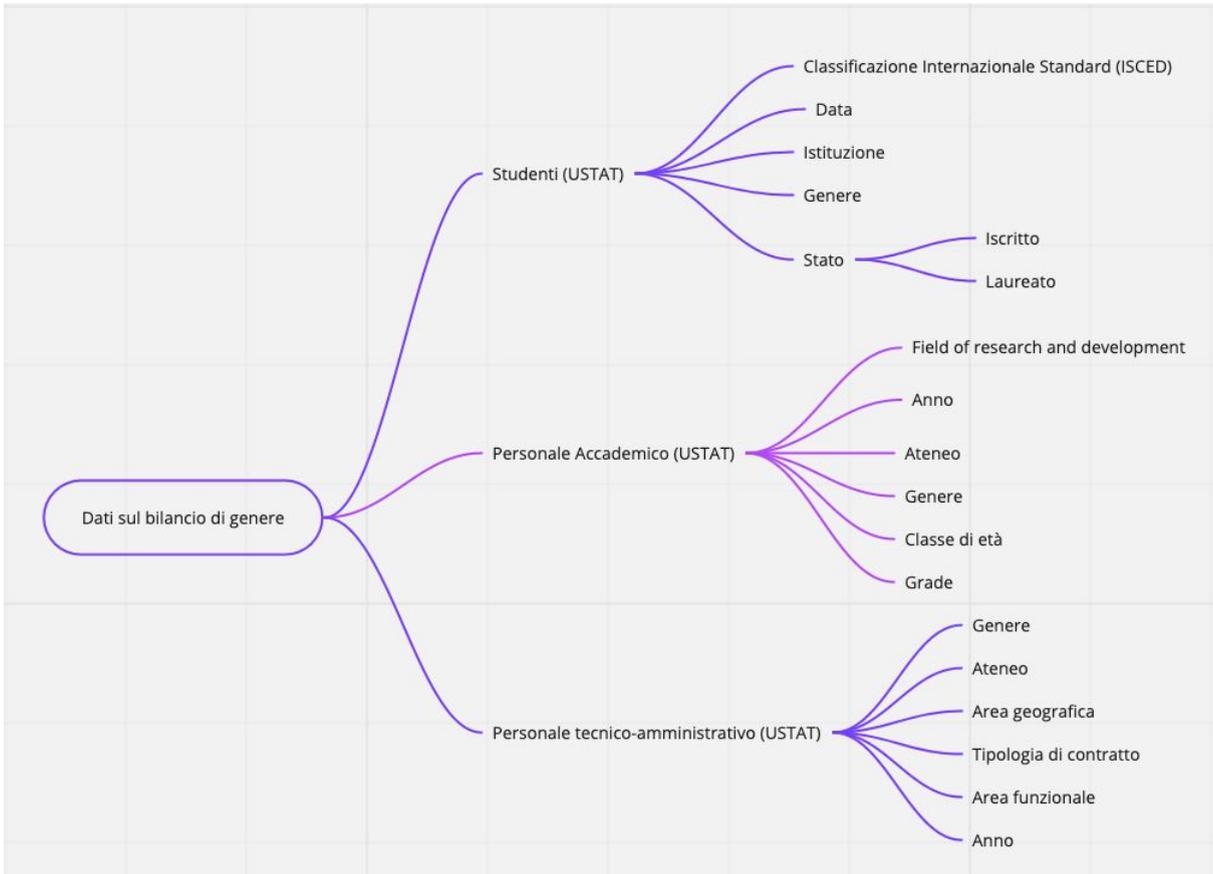
¹ Attualmente TOR si appoggia a una licenza di Web of Science, ma in futuro si prevede di integrare, laddove possibile, i dati provenienti dai repositories bibliografici delle singole università e centri di ricerca.

² Cineca è un Consorzio Interuniversitario senza scopo di lucro formato dal Ministero dell'Istruzione, il Ministero dell'Università e della Ricerca, 69 Università italiane e 11 Istituzioni pubbliche Nazionali (8 Enti di Ricerca, 3 Policlinici Universitari, 1 Agenzia).









Ad ogni modo, le mind map sopra rappresentante offrono solo un'idea generale dei dati a disposizione per ogni concetto e/o soggetto. Vedremo ora come analizzare in maggior dettaglio i dati disponibili e come consultarli tramite un'ontologia, ovvero una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse.

Come si interrogano i dati?

Questa sezione fornisce un'introduzione allo SPARQL Endpoint, la tecnologia utilizzata per interrogare i dati dell'Osservatorio, e ai diagrammi delle ontologie - delle rappresentazioni grafiche che permettono di descrivere delle entità (oggetti, concetti, ecc.) e le loro relazioni in un determinato dominio di conoscenza - con il fine di rendere il lettore autonomo nel poter interrogare i dati dell'Osservatorio.

Cos'è lo SPARQL Endpoint e come funziona?

Lo **SPARQL endpoint** permette agli utenti di recuperare i dati utilizzando il [linguaggio di interrogazione SPARQL](#), che è uno standard definito dal World Wide Web Consortium (W3C). Lo SPARQL endpoint può essere utilizzato sia dagli utenti umani che dalle macchine. Gli esseri umani interagiscono con esso attraverso un'interfaccia utente grafica, mentre le macchine (script, applicazioni, ecc.) comunicano con l'endpoint utilizzando il [protocollo standard SPARQL](#) definito dal W3C.

I dati recuperati da un endpoint SPARQL sono conformi all'approccio [Linked Data](#) per la pubblicazione di dati nel contesto del [Semantic Web](#). Ciò significa che il formato dei dati è **RDF (Resource Description Framework)**, in cui le informazioni sono rappresentate come risorse. Ogni risorsa è identificata da un **Uniform Resource Identifier (URI)**, che nella maggior parte dei casi assomiglia a un indirizzo web (ad es. <http://unics.cloud/ontology#Project-123>). Le risorse sono caratterizzate da proprietà. Alcune proprietà attribuiscono valori di dati a una risorsa, come l'acronimo di un progetto o il suo titolo, mentre altre collegano una risorsa con un'altra risorsa, ad esempio un progetto con uno dei suoi partecipanti. Le proprietà che attribuiscono valori di dati sono chiamate "**proprietà dato**", mentre le proprietà che collegano le risorse sono chiamate "**proprietà oggetto**". Le proprietà, come le risorse, sono identificate dagli URI (ad esempio <http://unics.cloud/ontology#acronym>³).

Concettualmente, **i dati formattati in RDF sono un grafo, in cui i nodi sono o risorse o valori di dati, e i collegamenti tra i nodi sono o proprietà dato o proprietà oggetto**. In termini pratici, tuttavia, i dati formattati in RDF sono normalmente rappresentati come una **lista di triple**. Una tripla è una frase sotto forma di "**soggetto predicato oggetto**". Queste triple sono usate per elencare i collegamenti nel grafo, cioè il soggetto di una tripla è sempre una risorsa, il predicato è una proprietà, e l'oggetto è o un valore di dati o un'altra risorsa. Come esempio, un set di dati contenente due progetti di uno stesso bando con il loro titolo e la loro sigla si presenterebbe in questo modo (gli URI sono spesso scritti tra parentesi angolari):

```
<http://unics.cloud/ontology#Project-1> <http://unics.cloud/ontology#acronym> "Acronym 1"  
<http://unics.cloud/ontology#Project-1> <http://unics.cloud/ontology#title> "Title 1"  
<http://unics.cloud/ontology#Project-1> <http://unics.cloud/ontology#call> <http://unics.cloud/ontology#Call-A>  
<http://unics.cloud/ontology#Project-2> <http://unics.cloud/ontology#acronym> "Acronym 2"  
<http://unics.cloud/ontology#Project-2> <http://unics.cloud/ontology#title> "Title 2"  
<http://unics.cloud/ontology#Project-2> <http://unics.cloud/ontology#call> <http://unics.cloud/ontology#Call-A>  
<http://unics.cloud/ontology#Call-A> <http://unics.cloud/ontology#name> "Name of call for projects A"
```

In RDF e SPARQL è comune utilizzare i prefissi per abbreviare i nomi delle risorse e delle proprietà, cioè gli URI. Come vediamo nell'esempio precedente, tutte le risorse e le proprietà di questo dataset hanno URI che iniziano con "<http://unics.cloud/ontology#>". Per evitare di

³ [UNICS](#) è una piattaforma Open Data, sviluppata da SIRIS Academic, basata su tecnologie semantiche che integra archivi di Open Data sull'istruzione superiore, la ricerca e l'innovazione in Europa. Questi dati integrati sono accessibili in formato open attraverso un endpoint SPARQL.

scrivere ogni volta questa parte comune, possiamo definire un'abbreviazione per questo prefisso, per esempio "tor:", e scrivere questa abbreviazione al posto del prefisso completo. L'esempio di cui sopra diventerebbe allora:

```
Prefix tor: <http://unics.cloud/ontology#>
tor:Project-1 tor:acronym "Acronym 1"
tor:Project-1 tor:title "Title 1"
tor:Project-1 tor:call tor:Call-A
tor:Project-2 tor:acronym "Acronym 2"
tor:Project-2 tor:title "Title 2"
tor:Project-1 tor:call tor:Call-A
tor:Call-A tor:name "Name of call for projects A"
```

In termini di sintassi, **l'abbreviazione di un prefisso deve essere in forma di "nome:"** (notare il ":" alla fine). L'abbreviazione più breve possibile è quella dove "nome" è vuoto, cioè l'abbreviazione è solo ":". Inoltre, è opportuno notare che quando le abbreviazioni sono usate per i prefissi, gli URI non sono più scritti tra parentesi angolari (questi sono usati solo per gli URI non abbreviati).

Per poter scrivere query su dati RDF utilizzando il linguaggio SPARQL, è necessario sapere quali proprietà possono essere applicate su quali insiemi di risorse. Qui entrano in gioco le ontologie. **Un'ontologia è essenzialmente una definizione della struttura dei dati formattati in formato RDF, dove risorse con proprietà simili sono raggruppate in concetti** (chiamati anche classi). In questo senso, un'ontologia definisce un insieme di concetti (ad esempio Progetto, Chiamata, Partecipante, Paese, ...) e un insieme di proprietà (sia proprietà dato che proprietà oggetto) che possono essere applicate a questi concetti. I concetti, proprio come le proprietà e le risorse, sono identificati dagli URI (che possono essere abbreviati come visto sopra).

L'ontologia TOR

Per l'**integrazione dei dati** il sistema utilizza l'approccio **Ontology-Based Data Access (OBDA)** che permette di integrare i dati attraverso un'ontologia di dominio e consente agli utenti di accedere ai dati per mezzo di "interrogazioni", eliminando la necessità di conoscere i termini tecnici legati all'organizzazione fisica dei database e alla struttura interna.

La documentazione dell'ontologia è disponibile sul sito ToscanaOpenResearch, nella sezione [Accesso ai Dati > Documentazione dell'ontologia](http://toscanaopenresearch.it/sparql-endpoint/it/docs/index.html), al link: <http://toscanaopenresearch.it/sparql-endpoint/it/docs/index.html>.

Come leggere il diagramma dell'ontologia?

Nei diagrammi, **le caselle con bordo arrotondato indicano i concetti**, con l'elenco dei nomi scritti all'interno di ogni casella che sono le **proprietà dato** applicabili a questo particolare concetto. Ad esempio, nella parte di ontologia TOR che riguarda i progetti CORDIS, il concetto "Organization" ha le proprietà dato "unicsId", "extendedName" ed "acronym". Le frecce con punta aperta tra le caselle rappresentano invece **proprietà oggetto**. Il nome di una proprietà oggetto viene indicato attraverso un'etichetta appostata sulla freccia, ed il verso di applicazione della proprietà è dato dal verso della freccia stessa. Ad esempio, la freccia etichettata con "principalInvestigator" che collega il concetto "EC-Project" con il concetto "Person", rappresenta una proprietà oggetto che associa ad un progetto europeo la persona che è il principal investigator di quel progetto.

Ad ogni proprietà dato e ad ogni proprietà oggetto è associato un **vincolo di cardinalità**, specificato attraverso una coppia "min..max", che stabilisce un limite al numero minimo e

massimo di valori che la proprietà in questione può avere. Ad esempio, il vincolo “0..1” sulla proprietà “principalInvestigator” indica che un oggetto in EC-Project (ovvero, un progetto europeo) può avere o meno un principalInvestigator (vincolo minimo 0), e in ogni caso può averne al più uno (vincolo massimo 1). In modo analogo, il vincolo “[0..1]” sulla proprietà dato “startingYear” di “EC-Project”, indica che un progetto europeo può avere o meno un singolo anno di inizio. In generale, la cardinalità minima può essere 0 (ovvero nessun vincolo minimo), oppure un intero positivo; quella massima può essere un intero positivo, oppure “n” (ovvero nessun vincolo massimo). Se nel diagramma manca la specifica del vincolo di cardinalità, viene assunto “1..1”, ovvero la proprietà è sia obbligatoria (ha sempre un valore), sia funzionale (ne ha uno solo).

Un altro elemento importante delle ontologie è il **meccanismo di generalizzazione**, che permette di **organizzare i concetti in una o più gerarchie**. Una generalizzazione (detta anche “is-a”), è indicata nel diagramma con una freccia che termina con un triangolo chiuso, che va da un concetto più specifico ad uno più generale. Un is-a indica che ogni oggetto che appartiene al concetto più specifico appartiene anche a quello più generale. Ad esempio, la freccia is-a da “ERC-Project” a “EC-Project” indica che ogni progetto ERC è anche un progetto europeo. In virtù di questo, la generalizzazione comporta l’*ereditarietà*, ovvero il fatto che un concetto più specifico eredita tutte le proprietà di un concetto più generale. Ad esempio, ogni progetto ERC, essendo anche un progetto europeo, ha tutte le proprietà dei progetti europei, ad esempio un titolo (proprietà dato “title”). Ma un progetto ERC ha anche ulteriori proprietà che i progetti europei in generale non hanno, ad esempio il fatto di avere un “ercGrant”.

Approfondimento - I vantaggi di un accesso ai dati basato su ontologie di dominio

L’approccio Ontology-Based Data Access (OBDA), detto anche *Virtual Knowledge Graph*, permette di accedere a database relazionali e a sorgenti di dati di diversa natura (ad es. file CSV, tabelle excel, dati JSON) mediando questi accessi attraverso un’**ontologia di dominio**. Tale ontologia fornisce un modello dei dati consolidato, condiviso tra diversi attori, ed espresso utilizzando una terminologia e concetti propri del dominio, che possono anche andare oltre quanto presente in specifiche sorgenti di dati. L’ontologia è poi collegata alle sorgenti utilizzando dei **“mapping” dichiarativi**. Questi rendono esplicito il modo in cui è effettuato il collegamento, e non lo “nascondono” in codice programmatico, come avviene nei sistemi di integrazione tradizionali. L’utente può quindi estrarre le informazioni di interesse, utilizzando i termini a lui familiari dell’ontologia. Può formulare con facilità interrogazioni anche complesse direttamente sull’ontologia, senza la necessità di conoscere i termini usati nei diversi database sorgente, spesso criptici e poco comprensibili, o come i dati sono ivi organizzati. L’interrogazione utente viene poi tradotta in opportuni accessi alle sorgenti di dati e le risposte vengono combinate per fornire la risposta all’utente, facendo uso dell’ontologia stessa e dei mapping. Il tutto avviene in modo completamente automatico e trasparente all’utente. L’intero approccio si basa su standard definiti dal World Wide Web Consortium (W3C) per l’ontologia, per i mapping, e

per il linguaggio di interrogazione.

È importante evidenziare che **l'ontologia fornisce una visione unificata dei dati come grafo** (insieme di elementi collegati) **di conoscenza**.

Da una parte, questo semplifica notevolmente l'integrazione e la riconciliazione dei dati provenienti da diverse sorgenti.

D'altra parte, l'ontologia può essere riutilizzata in contesti diversi in cui il dominio è analogo, semplicemente collegando le nuove sorgenti di dati.

Inoltre, i dati possono essere resi accessibili concentrandosi su chi poi li userà. In questa logica rientra anche la possibilità di tenere conto nel mapping dei dati solo dei dettagli specifici che sono di interesse per determinate tipologie di utente o in un determinato contesto, e rispettando criteri di privacy o confidenzialità.

In definitiva, un sistema **OBDA permette ai dati di essere fruiti da un pubblico più ampio e con una minore dimestichezza delle specifiche sorgenti**. L'impiego di sistemi OBDA tipicamente **non richiede modifiche ai database stessi**, permette dunque di mantenere piena operatività dei sistemi sottostanti. Riassumendo, l'approccio OBDA può essere considerato come un meccanismo per fornire **una API** (Application Programming Interface) **di alto livello per sorgenti di dati eterogenee**.

A livello tecnologico, il sistema OBDA di ToscanaOpenResearch utilizza [Ontop](#), una implementazione OBDA particolarmente performante e avanzata, come evidenziato anche dalla letteratura scientifica specifica. Questo software Open Source vanta una comunità di utenti eterogenea, che va da utenti che operano in un contesto industriale, fino a istituti finanziari e a gruppi di ricerca universitari. Il software è in continuo sviluppo, con aggiunta di funzionalità nuove oppure migliorate e la possibilità di un supporto professionale oltre a quello della community.

L'ontologia sviluppata all'interno di ToscanaOpenResearch è ora **in corso di allineamento con il progetto nazionale OntoPIA** (la rete di ontologie e vocabolari controllati della Pubblica Amministrazione), coordinato da AgID - Agenzia per l'Italia Digitale, per diventare **l'ontologia di riferimento nazionale su alta formazione e ricerca**: l'ontologia è disponibile a questo [link](#).

Come scrivere le query per esplorare e scaricare dati?

Dopo aver appreso le nozioni base sullo SPARQL Endpoint e sulla rappresentazione grafica dell'ontologia, è giunto il momento di interrogare in maniera autonoma il sistema e di sperimentare queries.

Una richiesta di informazioni ad un SPARQL endpoint tipicamente si compone dalle seguenti parole chiave:

- **PREFIX**: usato per specificare l'URL delle ontologie che si intende utilizzare nella query.
- **SELECT**: definisce i valori di output che saranno inclusi nella query. I suddetti valori devono essere presenti all'interno della clausola WHERE.
- **WHERE**: contiene le informazioni sotto forma di restrizione al fine di ottenere dei dati.

Di seguito un esempio di query che fornisce in output le università presenti nell'ontologia di

UNICS.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX unics: <http://unics.cloud/ontology#>
SELECT ?org
WHERE {
    ?org rdf:type unics:University .
}
```

L'istruzione PREFIX seguita da una parola (a scelta libera), introduce un'abbreviazione per URL potenzialmente lunghi. Riportando l'esempio delle università, nell'ontologia di UNICS le università sono disponibili attraverso l'URL "<http://unics.cloud/ontology/University>", usando l'istruzione PREFIX, è possibile riferirsi alle università usando "unics:University", poiché l'URL di unics è stato salvato nella parola unics.

Si nota che la query è divisa in almeno due parti.

Nella **prima parte, introdotta appunto da SELECT, si elencano le variabili che si vogliono avere come risultato**, separandole con spazi⁴.

La **seconda parte, introdotta da WHERE e racchiusa da parentesi graffe, specifica le regole sotto forma di triple RDF**. Una triple RDF è composta da:

- **Soggetto:** indica un risorsa disponibile nell'ontologia;
- **Predicato:** consente di creare una relazione tra soggetto e oggetto;
- **Oggetto:** può essere un risorsa identificabile tramite un univoco URI oppure un dato.

Nell'esempio sopra riportato troviamo:

- **?org:** soggetto della query
- **unics:University:** oggetto della query
- **Rdf:type:** predicato che permette di salvare nella variabile ?org il contenuto dell'oggetto unics:University.

Analizzando il contenuto della query, si nota che si sta chiedendo le entità che sono di tipo unics:University. Il termine ?org è una variabile, in cui è salvato il contenuto . Essa può prendere un nome arbitrario, purché inizi con un punto di domanda. Essa permette di usare tutti i valori che assume nelle triple "valide" nella tabella dei risultati. All'interno di una variabile è possibile salvare qualsiasi tipo di contenuto, che sia un dato o un URI che sarà usato successivamente per creare query più complesse.

In sostanza, WHERE seleziona dati da prendere in considerazione per costruire la tabella dei risultati.

È possibile inserire più di un triple pattern in una clausola WHERE:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX unics: <http://unics.cloud/ontology#>
SELECT *
WHERE {
    ?org rdf:type unics:University .
    ?org unics:extendedName "Università di Firenze" .
}
```

⁴ Similmente a SQL, l'asterisco è un modo comodo per elencare tutte le variabili definite più avanti.

All'esempio precedente viene aggiunta la condizione che con lo stesso soggetto del primo pattern e associato alla variabile ?org porti anche il nome di "Università di Firenze".

Si può intuire come le query possono crescere man mano in termini di complessità. In combinazione con la caratteristica di SPARQL di poter interrogare i metadati ("che proprietà può avere un'università?"), questo lo rende un linguaggio molto adatto all'esplorazione progressiva dei dati.

Vi sono poi vari modi più raffinati di selezionare ed elaborare dati, per i quali si rimanda alla letteratura più specifica e ai esempi contenuti sui portali TOR.

Una libreria di query predefinite, organizzate coerentemente con le sezioni principali dell'Osservatorio, è disponibile all'interno della sezione Accesso ai dati > SPARQL endpoint al seguente link: <http://toscanaopenresearch.it/sparql-endpoint/it/>.

Tramite questo SPARQL endpoint è inoltre possibile ottenere l'elaborazione visuale immediata dei dati ottenuti dall'esecuzione delle query, servendosi di strumenti quali "Pivot Table" e "Google Chart".

Tuttavia, una volta presa confidenza con lo SPARQL endpoint e il diagramma dell'ontologia, seguendo le istruzioni sopra riportate, sarà possibile cambiare le query e adattare in base alle esigenze, aggiungendo maggiori informazioni e/o cambiando il perimetro (sia temporale che geografico) di interesse.