

Scheda di progetto per programmazione anno 2022

Titolo progetto: La ricerca scientifica nell'ambito delle Scienze Sociali: un Servizio Dati condiviso

Responsabili: Prof. Flavio Bazzana e Prof.ssa Gabriella Berloffia, Dipartimento di Economia e Management (DEM)

Descrizione obiettivi e risultati attesi

L'utilizzo di datasets pubblici e/o privati rappresenta una risorsa chiave per gran parte della produzione scientifica nell'ambito delle Scienze Sociali. Tipicamente, l'accesso a tali datasets avviene per iniziativa dei singoli ricercatori o gruppi di ricercatori che, oltre a reperire le eventuali risorse economiche necessarie, devono occuparsi delle procedure amministrative, fornire garanzie tecniche di sicurezza e protezione dei dati ed eseguire le operazioni preliminari necessarie per rendere fruibili i datasets per le analisi vere e proprie. I principali datasets utilizzati nelle Scienze Sociali si distinguono in due grandi tipologie: datasets di 'produzione pubblica' (contengono per lo più i risultati di indagini campionarie su famiglie e imprese condotte dagli Istituti di Statistica Nazionali dei diversi paesi e/o da Enti pubblici specifici; normalmente l'accesso ai microdati relativi a tali indagini è gratuito, ma richiede la presentazione di un progetto di ricerca e la fornitura di garanzie tecniche di sicurezza e protezione dei dati;) e datasets di 'produzione privata' (contengono per lo più dati di varia natura relativi a imprese e/o mercati specifici, raccolti da soggetti privati; l'accesso a tali dati è prevalentemente a pagamento).

Data la centralità delle banche dati e del loro utilizzo per lo sviluppo delle linee di ricerca dei docenti dell'Università di Trento afferenti a dipartimenti inclusi di Economia e Management (DEM, Dipartimento proponente), di Sociologia e Ricerca Sociale (DSRS) e di Facoltà di Giurisprudenza (DFG) e della Scuola di Studi Internazionali (SSI), oltre che delle scuole di dottorato, come la Scuola di Dottorato in Scienze Sociali (SDSS), l'Ateneo di Trento considera lo sviluppo di un Servizio Dati per le Scienze Sociali (SeDaSS), che possa divenire nel corso dei prossimi anni un punto di riferimento per la ricerca scientifica. Il SeDaSS dovrebbe rispondere alle seguenti esigenze:

- a) Raccogliere e unificare le conoscenze relative alle procedure di accesso ai principali datasets di produzione pubblica e privata, identificando un/una responsabile UNITN per i rapporti con i soggetti produttori dei dati e il personale di supporto per le procedure d'accesso;
- b) Diffondere le informazioni relative a nuove richieste di accesso a datasets pubblici e privati promosse da ricercatori UNITN in modo da raccogliere l'eventuale interesse di altri ricercatori per gli stessi dati, gestire in modo coordinato le varie richieste e favorire il reperimento degli eventuali fondi necessari;
- c) Disporre dell'hardware e del software necessario per lo stoccaggio dei dati e la fornitura delle necessarie garanzie di sicurezza e protezione dei dati;
- d) Gestire gli aspetti contrattuali per l'acquisto e l'utilizzo dei datasets;
- e) Eseguire le operazioni preliminari necessarie per rendere fruibili i datasets ai diversi utilizzatori finali. Da questo punto di vista è importante sottolineare che, nel caso di datasets di produzione pubblica, l'accesso ai dati è normalmente concesso per progetti di ricerca specifici e per periodi di tempo limitati. Tuttavia, è frequente il caso di utilizzo dello stesso dataset per progetti diversi, sia nello stesso intervallo temporale, sia in periodi successivi. Di conseguenza, l'accentramento delle operazioni di preparazione (in gergo, pulizia) preliminare costituirebbe un'importante razionalizzazione delle attività e un notevole risparmio di tempo per i ricercatori. La condivisione dei codici necessari a svolgere queste operazioni contribuirebbe inoltre a rendere più trasparente e inclusivo il lavoro di ricerca, in coerenza con l'impegno dell'Ateneo per l'*open access* e l'*open science*;
- f) Gestire, ove opportuno, banche dati costruite internamente, sia a partire dai dati dei provider, sia da dati pubblici. In questo caso dovranno essere definite delle politiche di utilizzo in accordo con i ricercatori.

Il presente progetto offrirebbe la possibilità di dare il via immediato allo sviluppo di SeDaSS, assicurando la messa a terra degli elementi fondamentali per questo progetto di ricerca trasversale e consentendo l'articolazione del piano di lavoro per il raggiungimento degli obiettivi a lungo termine.

Nell'ottica dello sviluppo di SeDaSS, il DEM e l'Ateneo lavoreranno per la valorizzazione del servizio e per la sua sostenibilità attraverso:

- La condivisione delle informazioni necessarie sia per quanto riguarda l'accesso ai vari datasets, sia per quanto riguarda il loro utilizzo;
- La condivisione di eventuali datasets già posseduti, il cui utilizzo si possa estendere ad altri ricercatori;
- La condivisione dei codici necessari a svolgere le operazioni preliminari di preparazione dei datasets, ove già disponibili;
- L'organizzazione di workshop in cui vengano presentate le attività di ricerca connesse ai datasets gestiti dal SeDaSS, al fine di promuovere nuove collaborazioni scientifiche, anche interdisciplinari;
- Il co-finanziamento dell'acquisto di altri datasets oltre a quelli indicati sopra;
- Il co-finanziamento futuro di tutti o alcuni dei datasets sopra indicati, sulla base delle valutazioni che verranno fatte alla fine del triennio.

L'obiettivo ultimo di questa proposta di progetto è di consentire l'analisi empirica di temi di pubblico interesse, anche in relazione ad aspetti culturali, sociali ed economici del territorio trentino.

Struttura, modalità di attuazione del progetto

Per sviluppare SeDaSS sono necessarie risorse, inclusi:

- uno spazio fisico, ad ingresso controllato, con alcune postazioni per l'interrogazione e l'eventuale download dei dati. A tal fine, il DEM può mettere a disposizione uno spazio all'interno del Palazzo di Economia;
- alcuni server da installare nel cluster di ateneo, con coda dedicata agli utenti del SeDaSS;
- una risorsa PTA (anche a tempo parziale) che si occupi dei servizi di cui ai punti a), b) e d) descritti sopra;
- una risorsa umana dedicata (tecnologo) che abbia competenze trasversali, non solo relative all'IT, ma anche nell'utilizzo delle banche dati nelle scienze sociali. Tale risorsa dovrebbe occuparsi dei servizi connessi ai punti c), e) e f) descritti sopra (gestione IT e verifica delle condizioni di sicurezza dei dati, operazioni di preparazione iniziale dei datasets e supporto ai ricercatori nella preparazione dei data set specifici per i vari progetti di ricerca; gestione banche dati costruite internamente);

È inoltre necessario il coordinamento dell'utilizzo e sviluppo del servizio, l'identificazione delle criticità *in itinere*, il confronto con altre realtà universitarie che svolgono attività di ricerca nell'ambito delle Scienze Sociali, l'individuazione ed il monitoraggio dei parametri di successo di SeDaSS su un arco temporale di medio-lungo termine.

Piano delle attività previste per il 2022

Per rendere SeDaSS operativo è importante poter procedere contestualmente all'acquisto di alcuni importanti datasets privati (o delle relative licenze di utilizzo), il cui costo è particolarmente elevato e di avvalersi di un tecnologo dedicato a questa prima fondamentale fase di sviluppo.

In particolare sono stati ad oggi identificate alcune banche dati rilevanti per DEM, DSRS, DFG, SSI, SDSS, che verranno acquisite nel corso del 2022. Esse includono:

ORBIS Full, Banca dati bilanci e informazioni su imprese (a livello mondiale);
Bankfocus, Banca dati mondiale bilanci e informazioni su intermediari finanziari;
Burning Glass, Banca dati mondiale su offerte di lavoro;
World Robotics, Banca dati su applicazioni di robot nelle imprese;
Add Health, Banca dati longitudinale US sulla salute.

È importante sottolineare che, nel caso di licenze di utilizzo, l'investimento sulle licenze per tre anni permetterebbe sia di valutare la bontà/qualità delle varie sezioni dei datasets, sia di verificare il numero

effettivo di ricercatori coinvolti, entrambi aspetti essenziali per definire la migliore strategia per il mantenimento e finanziamento futuro di tali datasets.

Impatto/ricadute sul territorio

In termini di impatto, il presente progetto permetterebbe di:

- Migliorare la conoscenza dei dati da parte dei ricercatori, con ricadute positive per l'attività di ricerca (in termini di numero e qualità delle pubblicazioni), favorendo altresì la partecipazione a bandi competitivi, nazionali e internazionali, da parte degli stessi;
- Migliorare l'attrattività di UNITN in fase di reclutamento di ricercatori/docenti nell'ambito delle Scienze Sociali;
- Migliorare la capacità di approfondire l'analisi empirica di temi di pubblico interesse, anche in relazione ad aspetti culturali, sociali ed economici del territorio trentino.

Come valore aggiunto, permetterebbe anche di:

- Migliorare la qualità del lavoro di tesi dei laureandi magistrali e delle attività didattiche di tipo laboratoriale per studenti triennali e magistrali (offrendo l'accesso a datasets complementari a quelli specificamente acquistati per le attività stesse);
- Offrire maggiori opportunità di ricerca ai dottorandi della Scuola di Scienze Sociali.