

Scheda di Programma

Per l'attivazione nell'ambito del Corso di Dottorato di ricerca in **Matematica e Scienze Computazionali** del seguente Programma di ricerca, a valere sulle risorse di cui al DM n. 351/2022, relativamente alla seguente Misura:

M4C1- Inv. 3.4 "Didattica e competenze universitarie avanzate" → **Dottorati dedicati alle transizioni digitali e ambientali.**

M4C1- Inv. 4.1 "Estensione del numero di dottorati di ricerca e dottorati innovativi per la pubblica amministrazione e il patrimonio culturale". In particolare:

Dottorati PNRR

Dottorati per la Pubblica Amministrazione

(selezionare l'area/le aree CUN di riferimento del programma tra quelle di seguito indicate)

- Area 09 – Ingegneria industriale e dell'informazione
- Area 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche
- Area 12 – Scienze giuridiche
- Area 13 – Scienze economiche e statistiche
- Area 14 – Scienze politiche e sociali

Dottorati per il patrimonio culturale

(selezionare l'area/le aree disciplinare/i e la tematica del programma tra quelle di seguito indicate)

- Area 01 – Scienze matematiche e informatiche **Tematica** – Informatica, patrimonio e beni culturali
- Area 02 – Scienze Fisiche **Tematica** – Fisica applicata al patrimonio culturale e ai beni culturali
- Area 03 – Scienze chimiche **Tematica** – Chimica, ambiente, patrimonio e beni culturali
- Area 04 Scienze della Terra **Tematica** – Georisorse minerarie per l'ambiente, il patrimonio e i beni culturali
- Area 05 Scienze Biologiche **Tematica** - Ecologia, patrimonio e beni culturali
- Area 08 – Ingegneria civile e Architettura **Tematiche** 1) Architettura, ambiente antropizzato, patrimonio e beni culturali 2) Architettura e paesaggio 3) storia dell'architettura; 4) Restauro; 5) Pianificazione e progettazione dell'ambiente antropizzato; 6) Design e progettazione tecnologica dell'architettura
- Area 10 Scienze dell'antichità, filologico-letterarie e storico -artistiche **Tematiche** 1) Archeologia; 2) Storia dell'arte; 3) Media, patrimonio e beni culturali
- Area 11 – Scienze storiche, filosofiche, pedagogiche, psicologiche **Tematiche** 1) Biblioteconomia; 2) Archivistica; 3) Storia del patrimonio e dei beni culturali 4) Paleografia; 5) Estetica; 6) Didattica dell'arte; 7) pedagogia dell'Arte
- Area 12 - Scienze giuridiche **Tematica** Diritto del patrimonio culturale
- Area 13 - Scienze Economiche e statistiche **Tematiche** 1) Economia della cultura e dell'arte 2) Economia e gestione delle imprese artistiche e culturali; 3) Statistica e Data Analytics per i beni culturali
- Area 14 Scienze Politiche e sociali **Tematiche** 1) Sociologia dei beni culturali 2) sociologia dell'ambiente e del territorio

Titolo del Programma di ricerca: Open Innovation per la competitività di Pubblica Amministrazione e PMI: Algoritmi su stringhe, Machine Learning e HPC per analisi e automazione massiva dei documenti gestionali

❖ **Descrizione** (MAX 5000 CARATTERI SPAZI ESCLUSI):

Obiettivi

La gestione di Big Data in ambito manageriale e amministrativo è di grande rilevanza per la Pubblica Amministrazione e le Imprese. Dati semi-strutturati come moduli e protocolli, fattura elettronica, fascicoli sanitari o schede documentali sono spesso generati, catalogati e conservati con standard disomogenei e tecnologie sub-ottimali, che impediscono l'implementazione di sistemi applicativi nonostante investimenti ingenti. Tecniche e protocolli informatici convenzionali sono spesso incompatibili con le necessità di post-processamento e interrogazione massiva. Ciò rende l'attività di information retrieval e system integration complessa e proibitiva dal punto di vista economico. Le mansioni amministrative e documentali spesso incidono in maniera decisiva sull'efficienza ed efficacia dei servizi offerti da PA e imprese, determinando talvolta l'impossibilità di garantire prestazioni e livelli assistenziali previsti per cittadini e aziende.

Il programma di ricerca si pone i seguenti obiettivi principali:

- A. migliorare significativamente framework, algoritmi, procedure e standard per il processamento e la gestione dei documenti disponibili per questi ambiti;
- B. rilasciare tecnologie, componenti e documentazioni Open Source che fungano da base di partenza accessibile e gratuita per i progetti di istituzioni pubbliche, PMI e Startup italiane.

Impatto

Molti investimenti ICT della Pubblica Amministrazione sono destinati a progetti su piattaforme, dati e analytics e digitalizzazione dei processi (Figura 1). Tuttavia, questi progetti ampi e complessi necessitano di tecnologie di base, API e standard per la gestione documentale e l'interoperabilità. Proprio questi potrebbero trarre benefici dal percorso di dottorato proposto, consentendo la creazione di strumenti e framework riutilizzabili in tutti gli interventi ICT della PA.



Figura 1 AGID, La Spesa ICT nella PA italiana 2020, Rep. 1/2021 (Fig. 35, pag.44)

Accelerazione di processi di trasformazione digitale

Attraverso la creazione di librerie software Open Source, algoritmi di Open Data Science e standard applicativi, sarà possibile fornire strumenti pronti all'uso per i progetti di innovazione della PA. La diffusione di standard aperti potrà portare alla creazione di implementazioni omogenee e potenzialmente interoperabili *by design*. Inoltre, gli strumenti elaborati saranno maggiormente accessibili anche per

Startup e PMI innovative in ambito ICT, con il risultato di rendere economico, veloce e aperto lo sviluppo di piattaforme e tecnologie innovative destinate alla PA.

Promozione dell'inclusione sociale

La diffusione di conoscenze scientifiche e di librerie Open Source per la relativa implementazione potrà consentire un miglioramento diffuso dell'accessibilità di tecnologie moderne e performanti, in grado di impattare positivamente le prestazioni e il livello assistenziale offerto ai cittadini. Ad esempio, strumenti predittivi e di estrazione di informazione da basi documentali non strutturate su larga scala, consentirebbe di rendere più economici, tempestivi ed accessibili gli strumenti di gestione di pratiche documentali (es. cartelle cliniche) e ridurre i tempi di attesa dovuti ai processi amministrativi e documentali per i servizi alla persona, con evidenti ricadute sulle fasce della popolazione più deboli, che non hanno accesso a strutture private o di consulenza per snellire tali procedure.

Miglioramento della sostenibilità ambientale

Attraverso strumenti di estrazione di entità geografiche dai documenti (es. Indirizzi, coordinate, etc.), sarà possibile geo-referenziare informazioni e renderle integrabili su scala massiva per attività quali la gestione del demanio, delle risorse naturali e per il monitoraggio di attività ad alto impatto ambientale. La disponibilità gratuita di strumenti e tecniche di base per la creazione di software e piattaforme che integrino tali tecnologie consentirà di rendere più accessibili e diffondere tali interventi.

Attività formativa e di ricerca

- *Indicizzazione compressa di collezioni di dati altamente ripetitivi*: nei contesti considerati, collezioni massive di dati altamente ripetitivi necessitano di essere memorizzate, analizzate e interrogate. Le strutture dati necessarie per elaborazioni complesse richiedono ulteriore spazio di ordine di grandezza superiore. La presente attività di ricerca prevede di definire e utilizzare nuove strutture dati di indicizzazione compressa capaci di memorizzare tutte le informazioni necessarie alla ricerca e localizzazione efficiente di pattern in uno spazio vicino alla taglia dei dati compressi. I dati non necessiteranno pertanto di essere decompressi ma potranno essere analizzati direttamente in forma compressa.
- *Classificazione e confronto di dati strutturati e semi-strutturati*: studio e utilizzo di strumenti di natura combinatoria per l'estrazione di informazioni, per la classificazione e il confronto tra testi. Gran parte dei dati con cui si intende operare sono strutturati in modo gerarchico. Tipicamente ciò accade per esempio nell'ambito Open Data. Si intende definire pertanto nuove metodologie per il confronto e la classificazione di dati strutturati in modo gerarchico, con lo scopo di localizzare similarità sotto-strutturali dei documenti analizzati.
- *Survival analysis e time-series analysis*: si prevede di utilizzare modelli basati su automi a stati finiti e algoritmi su grafi, integrati con tecniche di machine learning, per supportare survival analysis e time series analysis.

Attività di disseminazione e comunicazione dei risultati

Le istituzioni, in collaborazione con l'impresa BuildNN Srl, intendono:

- Pubblicare articoli scientifici e relazioni di ricerca su tecniche, framework e librerie;
- Creare librerie Open Source e strumenti per la prototipazione rapida in Cloud

❖ **PERIODO IN IMPRESA – CENTRI DI RICERCA – P.A.:**

Il Programma di ricerca sarà svolto in collaborazione con il seguente soggetto:

Ragione sociale: BuildNN Srl
Sede legale: Via Capri 44, 70022 Altamura (BA)
Rappresentante legale: Giacomo Barone

L'ente sopra citato ospiterà il dottorando beneficiario della borsa finanziata sulle risorse del DM 351/2022 per n. 12 mesi (**min 6 max 12**) nel corso del dottorato.

❖ **PERIODO ALL'ESTERO:**

Il Programma di ricerca prevede un periodo all'estero di n. 6 mesi (**min 6 max 18**) presso la seguente istituzione: Université Gustave Eiffel (Parigi, Francia)

Si dichiara inoltre che il presente programma è conforme al principio "di non arrecare un danno significativo" (DHS) ai sensi dell'art. 17 del regolamento (UE) 2020/852 in coerenza con gli orientamenti tecnici predisposti dalla Commissione Europea (Comunicazione della Commissione Europea 2021/C58/01) e garantisce il rispetto dei principi orizzontali del PNRR (contributo all'obiettivo climatico e digitale c.d. tagging, il principio della parità di genere e l'obbligo di protezione e valorizzazione dei giovani).

[REDACTED]

[REDACTED]

[REDACTED]