

# Relazione Finale Progetti MITD

## **TEAM WINX:**

Tania Brusadin - Davide Del Castello  
Giada Maurizzi - Minh Davide Ragagni

Luglio 2024

Alma Mater Studiorum - Università di Bologna

Insegnamento di *Metodi informatici per la trasformazione digitale*

Facoltà di *Governance e Politiche dell'Innovazione Digitale* (GEPID)

A.A. 2023/2024

# Indice

<b>Introduzione</b>	<b>2</b>
Team e ruoli assunti . . . . .	2
Metodologia di lavoro e collaborazione . . . . .	3
Diario Sprint . . . . .	3
Sprint 1 . . . . .	4
Sprint 2 . . . . .	4
Sprint 3 . . . . .	5
Sprint 4 . . . . .	6
Resoconto della retrospettiva finale e Criticità . . . . .	7
<b>1 Progetto 1: Protezione dei dati personali nell’ambito della sperimentazione clinica</b>	<b>8</b>
<b>2 Progetto 2: Trasformazione dei Dati Finanziari PNRR in formato aperto</b>	<b>10</b>
<b>3 Progetto 3: Storytelling dell’analisi dei dati delle strutture di vendita della regione Lombardia</b>	<b>12</b>
<b>4 Progetto 4: Creazione di un sistema software ex-novo</b>	<b>14</b>

# Introduzione

Scopo del seguente elaborato è di descrivere il processo che ha portato il Team Winx alla realizzazione dei quattro progetti d'esame dell'insegnamento *Metodi Informatici per la Trasformazione Digitale* dell'Università di Bologna. In particolare, di seguito si analizzeranno i ruoli rivestiti da ciascun partecipante nello svolgimento del progetto, i mezzi e gli strumenti che hanno permesso lo scambio di informazioni tra i soggetti, la metodologia lavorativa adottata, il tipo di collaborazione, le retrospettive ed, infine, le caratteristiche di ogni sprint. A seguire la sezione introduttiva è la descrizione del processo di lavoro, cioè l'analisi nel dettaglio dei risultati ottenuti al termine di ciascun progetto e le modalità di lavoro utilizzate, sottolineandone i ruoli, l'oggetto, lo svolgimento e le criticità.

## Team e ruoli assunti

Ciascun partecipante ha svolto un ruolo specifico nella realizzazione del progetto finale:

Nome e Cognome	Numero di Matricola	Ruolo
Tania Brusadin	1118143	Product Owner
Davide Del Castello	1118138	Scrum Master
Giada Maurizzi	1129249	Team
Minh Davide Ragagni	1132794	Team

Il Product Owner è responsabile della definizione della visione del prodotto e della gestione del backlog del progetto. Lavorando a stretto contatto con i membri del team, il Product Owner raccoglie i requisiti, stabilisce le priorità delle funzionalità e garantisce che il progetto soddisfi le aspettative degli ipotetici utenti finali e delle indicazioni dei Professori. Lo Scrum Master, invece, facilita l'adozione delle pratiche Agile e Scrum del parte del team, rimuovendo gli ostacoli che potrebbero impedire il progresso del lavoro. Lo Scrum Master organizza e guida le riunioni Scrum come le Sprint Planning, assicurandosi che il team segua correttamente i processi Scrum e promuovendo una comunicazione efficace all'interno del gruppo. I membri del Team sono coloro che concretamente progettano, sviluppano e consegnano le funzionalità del progetto. Collaborano tra loro e a stretto contatto con il Product Owner e lo Scrum Master per completare il lavoro in modo efficiente e rispettare gli standard di qualità. Partecipano attivamente alle riunioni, risolvono problemi, testano le funzionalità sviluppate e lavorano insieme per raggiungere gli obiettivi e le scadenze stabilite.

Per la definizione dei ruoli di ciascun progetto, invece, il Team ha deciso di alternarsi al fine di sviluppare le competenze trasversali di ciascuna mansione, di mettersi alla prova e di valorizzare le singole inclinazioni e gli interessi personali.

Studente	Progetto I	Progetto II	Progetto III	Progetto IV
Tania Brusadin	Dipendente dell'ISTC (Titolare e Responsabile del trattamento)	Membro Team Open Data	Product Owner	Team Creazione Software
Davide Del Castello	Product Owner	Membro Team Open Data	Team tecnico e di comunicazione	Scrum Master
Giada Maurizzi	Dipendente l'AIRETT (Titolare e Responsabile del trattamento)	Membro Team Open Data	Team tecnico e di comunicazione	Team Creazione Software
Minh Davide Ragagni	DPO	Product Owner	Pubblico generale	Product Owner

## Metodologia di lavoro e collaborazione

Subito dopo la creazione del team, abbiamo creato un gruppo Whatsapp con cui coordinare le nostre attività. Durante il primo incontro abbiamo svolto un'analisi generale e collettiva dei progetti, stabilito i ruoli, le priorità e le scadenze. In questo contesto, lo SM ha definito le sessioni di incontro e il PO ha determinato il numero degli Sprint. Il team ha lavorato in modalità mista: gli incontri generali sono avvenuti sempre da remoto (via Teams) mentre per i singoli progetti le coppie hanno lavorato anche in presenza. Per l'organizzazione del lavoro il team ha utilizzato **Taiga**, dove sono state raccolte le user stories e le task relative a ciascun progetto. Complessivamente, il lavoro ha richiesto quattro sprint di una settimana ciascuno, nei quali il Team diviso in coppie, differenti per ogni lavoro, ha portato avanti parallelamente i quattro progetti. Il team ha gestito lo stato di avanzamento dei progetti tramite la repository su **Github**. Tra gli strumenti utilizzati, inoltre, si ricordano:

Organizzazione del lavoro	Coordinamento tra team	Realizzazione dei progetti
Taiga	Whatsapp	Google Doc
Trello	Teams	Software PIA
Github	Google Calendar	Word
Google Drive		Blocco Note
		Colab
		Chatgpt
		Matey
		LaTeX

## Diario degli Sprint

Dedichiamo questa sezione ad un conciso riassunto dell'evoluzione degli sprint per ciascun progetto, informando che per maggiori dettagli, come le user stories e le task, si può fare riferimento al **Taiga**.

## Sprint 1

L'obiettivo del primo sprint ha riguardato principalmente la divisione e l'organizzazione del lavoro. Per ciascun progetto sono stati stabiliti i seguenti obiettivi e le US ad essi correlate:

- **Progetto 1**

**Obiettivo:** Comprensione dell'ambito di analisi e delle informazioni necessaria per la compilazione della documentazione finale.

**User Stories:** Definizione dei contratti di nomina, Registro delle attività e Relazione del DPO.

**Resoconto della retrospettiva:** Le attività da svolgere erano di prassi e di ricerca, quindi, in definitiva non hanno fatto insorgere difficoltà all'intero del Team. I risultati sono emersi nel secondo sprint.

- **Progetto 2:**

**Obiettivo:** Trascrizione dei dati presenti nel documento d'esame nel formato .jpeg in formato .csv e .json

**User Stories:** Conversione dei dati in formato aperto.

**Resoconto della retrospettiva:** Il team ha svolto le task senza difficoltà.

- **Progetto 3:**

**Obiettivo:** Download dei dataset e prima analisi dei dati a disposizione (elaborazione e pulizia).

**User Stories:** Analisi dei dati n. 1, 2, 3, 4, ossia fasi preliminari di analisi del dataset per poter svolgere le indagini successive.

**Resoconto della retrospettiva:** Il team ha svolto le task senza particolari difficoltà se non quelle riguardanti lo strumento Colab.

- **Progetto 4:**

**Obiettivo:** Backlog di prodotto, definizione di epiche e relative user stories. Definizione della struttura per il sistema di autenticazione degli utenti.

**User Stories:** Autenticazione degli utenti n. 1, 2, 3.

**Resoconto della retrospettiva:** Il team non ha trovato difficoltà.

## Sprint 2

L'obiettivo del secondo sprint ha riguardato la realizzazione dei primi elementi dei vari progetti. Per ciascun progetto sono stati stabiliti i seguenti obiettivi e le US ad essi correlate:

- **Progetto 1**

**Obiettivo:** Individuazione dei dati necessari per la stesura della DPIA e dell'Informativa, valutazione dei rischi e delle misure tecnico-organizzative.

**User Stories:** Informativa Privacy e modulo di consenso informato.

**Resoconto della retrospettiva:** Le attività da svolgere richiedevano attenzione e tempo per la ricerca delle informazioni, perciò si è ritenuto necessario prorogare la realizzazione delle task al terzo sprint.

- **Progetto 2:**

**Obiettivo:** Creazione del file RDF Turtle sulla base dell'ontologia fornita nel documento di esame.

**User Stories:** Definizione della politica degli URI, Creazione dei Linked Open Data, Definizione delle regole YARRML.

**Resoconto della retrospettiva:** Mancavano degli identificativi che permettevano la redazione dell'RDF Turtle per cui è stato necessario un confronto con la Professoressa Lodi.

- **Progetto 3:**

**Obiettivo:** Prime analisi comparative e realizzazione dei grafici.

**User Stories:** Analisi dei dati n. 5, 6, 8 con le prospettive di: un ricercatore di comportamenti di consumo, analista del mercato retail, dirigente di catena di retail.

**Resoconto della retrospettiva:** Il team ha svolto le analisi ma le user stories 5 e 8 non sono state completate a causa della mancanza di dati e dell'impossibilità di sviluppare modelli predittivi.

- **Progetto 4:**

**Obiettivo:** Prioritizzazione del backlog di prodotto e definizione dei requisiti specifici per la gestione dei vocabolari controllati.

**User Stories:** Gestione dei vocabolari controllati n. 1, 2.

**Resoconto della retrospettiva:** Data la mancata considerazione della realizzazione parallela degli altri progetti, si è verificato un sovraccarico del lavoro. Il problema è stato risolto grazie alla ridefinizione delle priorità e alla ridistribuzione delle user stories su tutti gli sprint, operate congiuntamente dal PO e dello SM.

Verso la fine del secondo sprint, il team ha richiesto un colloquio con i Professori al fine di confrontarsi sulle problematiche riscontrate nello svolgimento dei progetti e sulla loro risoluzione. L'incontro ha portato alla ridefinizione di alcune priorità e il team ha compreso come affrontare gli ostacoli.

## Sprint 3

Nel terzo sprint sono stati definiti i seguenti obiettivi e US:

- **Progetto 1**

**Obiettivo:** Compilazione dei campi guida per la stesura della DPIA sul software PIA.

**User Stories:** Utilizzo del software PIA.

**Resoconto della retrospettiva:** Il team ha avuto difficoltà nel lavorare contemporaneamente sul software, non adatto al lavoro di gruppo simultaneo. Il problema è stato risolto lavorando separatamente e inviando agli altri membri l'esportazione in formato .json.

- **Progetto 2:**

**Obiettivo:** Realizzazione della metadatazione.

**User Stories:** Produzione di metadatazione DCAT-AP.IT.

**Resoconto della retrospettiva:** Il team ha trovato difficoltà non avendo dimestichezza con il linguaggio e sono sorti dubbi sulla correttezza del risultato.

- **Progetto 3:**

**Obiettivo:** Realizzazione delle dashboard mediante le analisi grafiche dei dati.

**User Stories:** Analisi n. 7 e Produzione della documentazione (Colab con le dashboard)

**Resoconto della retrospettiva:** Il team, dopo un'iniziale difficoltà nella creazione delle dashboard, è poi riuscito a produrre il materiale grazie all'ausilio di tutorial e Chatgpt. Tuttavia, non è stato possibile realizzare dashboard in formato orizzontale ma solo in modalità "scrolling", quindi in verticale.

- **Progetto 4:**

**Obiettivo:** Definizione dei casi d'uso con la collaborazione degli altri membri del team e dei requisiti specifici per la funzionalità di ricerca dei beni culturali già catalogati.

**User Stories:** Definizione dei beni culturali.

**Resoconto della retrospettiva:** Il team non ha riscontrato difficoltà, ma si riconosce il limite del risultato dovuto alla mancata produzione di un software.

## Sprint 4

- **Progetto 1**

**Obiettivo:** Stesura della DPIA e revisione dell'informativa.

**User Stories:** Produzione della documentazione finale.

**Resoconto della retrospettiva:** Il software PIA non ha prodotto un file .doc facilmente leggibile di conseguenza il team ha dovuto elaborare un altro documento sulla base delle informazioni inserite nel software. Inoltre, l'informativa è stata aggiornata con le icone, in modo da renderla più user-friendly.

- **Progetto 2:**

**Obiettivo:** Produzione di un codice Python in grado di produrre automaticamente i file richiesti.

**User Stories:** Creazione del Colab.

**Resoconto della retrospettiva:** Il team ha riscontrato notevoli problematiche non avendo mai affrontato la produzione di un codice simile. Ciò lo ha portato ad affidarsi a risorse online e a Chatgpt, che, però, non hanno prodotto nell'immediato il risultato desiderato.

- **Progetto 3:**

**Obiettivo:** Creazione dello Storytelling in formato .doc e rilascio della Dashboard tramite download del Colab in .html.

**User Stories:** Storytelling dei dati e Rilascio della dashboard.

**Resoconto della retrospettiva:** Per la produzione dello storytelling è stato necessario aggiungere un comando per il download in .png dei grafici, in quanto il formato .html non era supportato da Google Doc e non permetteva di esporre la storia dei dati. Dopo il confronto con l'utente, si sono applicate alcune modifiche riguardo il formato e i colori delle dashboard per rendere lo storytelling più user-friendly.

- **Progetto 4:**

**Obiettivo:** Controllo ex-post per verificare le funzionalità implementate, identificare eventuali lacune e apportare le ultime modifiche.

**User Stories:** Testing funzionalità implementate.

**Resoconto della retrospettiva:** Il lavoro si è svolto coerentemente al programma stabilito, ma ha portato alla sola definizione del processo di realizzazione e non alla realizzazione stessa del prodotto, a causa della mancanza di competenze tecniche.

## Resoconto della retrospettiva finale e Criticità

Il team è riuscito a portare a termine tutti e quattro i progetti, nelle tempistiche e modalità previste. Non sono sorte criticità interne al team e l'ambiente di lavoro è sempre stato collaborativo, propositivo ed efficiente. Le criticità riscontrate riguardano principalmente la difficoltà nell'interpretare le direttive e i requisiti dei progetti, come il ruolo dell'aggiornamento di Taiga. Seppur il team abbia compilato le varie user stories, definito le specifiche task e si sia incontrato abitualmente per aggiornarsi sullo stato di avanzamento, non è stata posta sufficiente attenzione all'aggiornamento costante di Taiga. Gli aggiornamenti sono stati registrati con altre modalità, come l'interazione diretta, producendo un burndown poco rappresentativo dell'effettivo stato di avanzamento del progetto e dell'esecuzione delle fasi previste. Per le singole criticità relative a ciascun progetto, fare riferimento alle sezioni successive.



# Capitolo1

## Progetto 1: Protezione dei dati personali nell'ambito della sperimentazione clinica

### Ruoli

Ruolo	Nostro Nome
Product Owner	Davide del Castello
DPO	Minh Davide Ragagni
Dipendente dell'ISTC (Titolare e Responsabile del trattamento)	Tania Brusadin
Dipendente l'Associazione Italiana Sindrome di Rett (Titolare e Responsabile del trattamento)	Giada Maurizzi

### Oggetto

L'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR e l'Associazione Italiana Sindrome di Rett stanno collaborando per una ricerca che prevede la sperimentazione di giochi interattivi su soggetti con Sindrome di Rett (RTT) con lo scopo di valutare se questi possano stimolare le competenze senso-motorie e sociali nei soggetti RTT, di età compresa tra 3 e 30 anni. La sperimentazione, della durata di 36 mesi, prevede la raccolta e conservazione dei dati personali anche oltre la fine del progetto. In questo contesto, il team Winx deve produrre la documentazione necessaria per garantire la protezione dei dati personali ai fini della sperimentazione. In particolare, gestire il backlog del prodotto, confrontarsi con il DPO per una consulenza sul trattamento dei dati personali previsto dal progetto ed infine redigere la documentazione necessaria sul trattamento dei dati personali.

### Svolgimento

Il backlog del prodotto ha evidenziato la necessità di informarsi sulla procedura di protezione dei dati personali e di conseguenza di focalizzarsi su alcune specifiche questioni: determinazione dei dati raccolti, individuazione del tempo di raccolta e conservazione dei dati (5 anni), definizione dei ruoli e loro formazione, definire le finalità del trattamento, le basi giuridiche e le misure tecnico organizzative da implementare per rispondere ai rischi.

A seguito di questa ricerca, si è redatto un *documento che contenesse le informazioni relative al trattamento e delle proposte per il materiale da produrre* per la protezione dei dati personali degli individui soggetti alla sperimentazione.

Dopo un primo confronto con il DPO, le incaricate hanno avuto conferma dell'analisi e dei documenti da produrre. Inoltre, il colloquio con la professoressa Lodi ha confermato la necessità di produzione di DPIA e informativa privacy con modulo del consenso al trattamento; così le due incaricate hanno prodotto tali documenti. La realizzazione del lavoro ha richiesto 4 sprint da 1 settimana ciascuno, di cui le ultime due sono state interamente dedicate alla stesura dei documenti (per un'analisi più precisa fare riferimento ai diari degli sprint). Il gruppo si è riunito a volte online (tramite Teams) e a volte in presenza.

# Output

I documenti sono disponibili nella repository **Github** del progetto. Si trovano:

- Il **file contenente il backlog del prodotto e l'analisi del trattamento dei dati personali** con le informazioni necessarie per la redazione dei documenti.
- La DPIA (in **.json**, in **.docx** e in **.pdf**) prodotta con il software PIA e la relativa rappresentazione del ciclo di vita dei dati.
- L'informativa privacy (in **.docx** e in **.pdf**) con i relativi moduli di consenso alla partecipazione e la dichiarazione del medico o professionista sanitario che ha fornito i dati personali e particolari del paziente. Le icone sono state prese dai vincitori del contest "Informative privacy più chiare grazie alle icone? È possibile" indetto dal Garante per la protezione dei dati personali.

# Criticità

Il processo di produzione ha visto più momenti di retrospettiva, durante i quali non sono emerse criticità organizzative e comunicative, nonostante una modalità di lavoro mista. Tali momenti erano dedicati al confronto su come meglio affrontare la stesura dei documenti per garantire al meglio la protezione dei dati personali dei soggetti. Le principali questioni sorte ruotavano attorno all'individuazione del tempo di conservazione dei dati, alle misure tecniche ed organizzative da individuare vista la natura particolare dei dati clinici e al trattamento di dati relativi anche a minori. Le responsabili hanno lavorato ad entrambe le documentazioni in modo che vi fosse un continuo scambio e supporto. Dal confronto è emersa la necessità di rendere i documenti più "facili" e chiari per gli utenti, vista la materia. Il problema è stato risolto tramite l'utilizzo di grafiche per l'informativa privacy e una riduzione del contenuto della DPIA, congiunta ad una grafica del ciclo di vita dei dati, per rendere la documentazione più fruibile.

# Capitolo2

## Progetto 2: Trasformazione dei Dati Finanziari PNRR in formato aperto

### Ruoli

Ruolo	Nostro Nome
Product Owner	Minh Davide Ragagni
Membro Team Open Data	Tania Brusadin
Membro Team Open Data	Giada Maurizzi
Membro Team Open Data	Davide del Castello

### Oggetto

Il progetto richiede la trasformazione di alcuni dati, riguardanti i finanziamenti PNRR ricevuti dagli enti italiani, da un formato tabellare JPEG a formati standard più flessibili ed interoperabili come CSV, JSON e RDF/Turtle. Al Team è, inoltre, richiesto di produrre la relativa metadatazione obbligatoria DCAT-AP\_IT del dataset, disponibile sia come CSV, che come JSON e RDF/Turtle, utilizzando la licenza Creative Commons Attribuzione 4.0 (CC-BY 4.0). A tali fini, è previsto l'utilizzo degli strumenti e delle tecniche apprese durante il corso *"Metodi Informatici per la Trasformazione Digitale"* e di cimentarsi nella scrittura di un codice Python in grado di generare automaticamente i risultati richiesti, tramite l'utilizzo della piattaforma Colab utilizzata durante il corso *"Laboratorio di Programmazione"*.

### Svolgimento

Il progetto è stato gestito sotto la supervisione di un Product Owner (PO) e con la collaborazione di un team dedicato alla gestione e trasformazione degli Open Data, risultante da una preliminare definizione dei ruoli: il PO (Minh Davide Ragagni) è stato incaricato di gestire il backlog del prodotto, le priorità sulla produzione dei dati e della documentazione richiesta; gli altri tre membri del team (Giada Maurizzi, Tania Brusadin e Davide Del Castello) hanno formato il team open data incaricato di collaborare con il PO per la comprensione del dato e la redazione dell'elaborato. Si è utilizzato un approccio strutturato e metodologico per la definizione degli obiettivi e la conseguente produzione degli output. Più specificatamente, il progetto ha preso avvio con l'estrazione dei dati dal formato JPEG, avvenuta attraverso la loro trascrizione direttamente in formato CSV; a tale fine si è utilizzato il programma predefinito "Blocco Note" del computer, adatto per creare e modificare file di testo semplici, prestando attenzione a mantenere una codifica UTF-8 per la compatibilità universale. In seguito, si è convertito manualmente il CSV in un formato più strutturato e facilmente manipolabile (il formato JSON), seguendo gli esempi e le istruzioni presenti tra i materiali del corso. Il passaggio successivo ha previsto la mappatura manuale del CSV in RDF/Turtle sulla base dell'ontologia fornita al Team nel documento d'esame; a ciò è seguita l'integrazione semantica dei dati, cioè l'adozione del linguaggio YARRRML manualmente prodotto per mappare il CSV in RDF/Turtle

tramite l'utilizzo della piattaforma web **Matey**. Infine, il Team ha prodotto la metadatazione obbligatoria secondo lo standard DCAT-AP\_IT; questo ha incluso alcune informazioni cruciali specificate nel documento d'esame, come la licenza Creative Commons Attribuzione 4.0 e il tema di classificazione dei dati (Scienza e Tecnologia). Come ultimo passaggio, il Team si è servito della piattaforma Colab per scrivere un codice Python in grado di svolgere interamente il compito d'esame e, quindi, di restituire tutti i file richiesti. Questo passaggio è risultato fondamentale per verificare la correttezza degli output. Il completamento del lavoro ha richiesto 4 sprint da 1 settimana ciascuno (per un'analisi più precisa fare riferimento ai diari degli sprint). Il gruppo si è riunito a volte online e a volte in presenza.

## Output

I documenti sono disponibili nella repository **Github** del progetto. Si trovano:

- Il file **CSV** contenente i dati finanziari PNRR estratti dalla tabella in formato JPEG presente sul documento d'esame;
- Il file **JSON** ottenuto dalla trasformazione del CSV;
- Il file **RDF/Turtle** prodotto utilizzando YARRRML per la mappatura del CSV in grafi RDF;
- Il file **YARRRML** utilizzato per la mappatura del CSV in grafi RDF;
- La **Metadatazione DCAT-AP\_IT** in CSV, JSON e RDF/Turtle;
- Il file **IPYNB** contenente il codice Python

## Criticità

Durante il corso dello svolgimento del progetto sono emerse alcune criticità: prima di tutto, la mappatura manuale del CSV in RDF/Turtle ha costituito per i membri del Team una fase relativamente complessa a causa della non domestichezza nell'interpretazione dell'ontologia e, quindi, il rischio di commettere errori. La produzione della metadatazione secondo lo standard DCAT-AP\_IT ha implicato la necessità di rispettare dettagliate specifiche normative, inclusa l'accuratezza delle informazioni come la licenza e il tema di classificazione, di cui i membri del Team non hanno avuto immediata e facile comprensione. Le problematiche maggiori sono, però, emerse nella stesura del codice Python: non avendo mai affrontato nulla del genere prima, il Team si è completamente affidato alle risorse web, non sempre risultate adeguate.

# Capitolo3

## Progetto 3: Storytelling dell'analisi dei dati delle strutture di vendita della regione Lombardia

### Ruoli

Ruolo	Nostro Nome
Product Owner	Tania Brusadin
Team tecnico e di comunicazione	Giada Maurizzi
Team tecnico e di comunicazione	Davide Del Castello
Pubblico generale	Minh Davide Ragagni

### Oggetto

Il progetto prevede la creazione di una dashboard relativa al commercio al dettaglio della Regione Lombardia, utilizzando tre dataset aperti disponibili online:

- **Serie Storica Esercizi di Vicinato:** Contiene dati relativi agli esercizi di vicinato nella Regione Lombardia.
- **Serie Storica Medie Strutture di Vendita:** Include dati sulle medie strutture di vendita presenti nella Regione Lombardia.
- **Serie Storica Grandi Strutture di Vendita:** Contiene dati sulle grandi strutture di vendita nella Regione Lombardia.

L'analisi prevede, utilizzando gli strumenti e le tecniche apprese durante il corso "Laboratorio di Programmazione", la creazione di una dashboard con l'obiettivo di identificare trend e pattern nel commercio al dettaglio della Lombardia, di creare rappresentazioni grafiche efficaci (grafici a torta, boxplot, histplot ecc.) per supportare lo storytelling e mostrare i dati ad un pubblico generale.

### Svolgimento

Il progetto è stato realizzato seguendo un approccio strutturato e metodologico. Si è cominciato dall'analisi dei desiderata e dalla definizione degli obiettivi. Il team ha concentrato l'attenzione sull'analisi descrittiva dei tre dataset identificando le variabili di interesse e i possibili trend da evidenziare.

Successivamente, è stato formato il team e sono stati assegnati i ruoli: il PO (Tania Brusadin) è stato incaricato di gestire il backlog del prodotto e le priorità sull'analisi dei dati. Due membri del team hanno formato il gruppo tecnico e di comunicazione (Giada Maurizzi e Davide Del Castello), incaricato di eseguire l'analisi dei dati, creare visualizzazioni grafiche e sviluppare la storia basata sui dati. Un quarto membro (Minh Davide Ragagni) ha svolto il ruolo del pubblico, fornendo feedback sulla presentazione dello storytelling.

Dopo l'upload dei dati su Colab sono stati effettuati passaggi di pulizia e pre-elaborazione per gestire dati mancanti o anomali, garantendo la coerenza e l'integrità degli stessi.

Durante il primo sprint, utilizzando tecniche di analisi dati e librerie Python come Pandas, NumPy e Matplotlib/Seaborn per la visualizzazione, il team ha esplorato trend, correlazioni e pattern all'interno dei dataset. Sono stati creati grafici, diagrammi a barre, istogrammi e scatter plot per rappresentare chiaramente le informazioni estratte dai dati. Nella seconda e terza settimana (rispettivamente secondo e terzo sprint), il team ha sviluppato diverse dashboard interattive con l'ausilio di Colab, utilizzando le visualizzazioni precedentemente create e focalizzandosi sulle analisi di ciascun dataset per la creazione di una dashboard totale con i confronti tra i tre. Le dashboard sono state progettate per fornire una panoramica intuitiva e informativa sul commercio al dettaglio nella Regione Lombardia, consentendo agli utenti di esplorare i dati in modo dinamico. A seguito del completamento dell'analisi dei dati e della redazione del documento dello storytelling, il lavoro è stato presentato al membro del gruppo che rappresenta il pubblico per raccogliere feedback al fine di identificare aree di miglioramento e possibili sviluppi futuri, considerando implementazioni aggiuntive per arricchire ulteriormente la dashboard e migliorare l'esperienza utente. Alcuni dei consigli sono stati implementati nell'ultima settimana (ultimo sprint).

## Output

Il team ha analizzato i dati tramite Colab e sviluppato diverse dashboard interattive. I risultati sono presenti nella cartella **Github**. All'interno del **Colab** sono contenuti i passaggi di analisi e i principali risultati che prevedono la produzione di 5 dashboard: una per ciascun dataset, una "dashboard totale" che le contiene tutte, e una "dashboard unica" che riassume i confronti più esemplificativi facendo riferimento anche ad un caso specifico: la Provincia di Milano, per arricchire ulteriormente l'analisi svolta. Infine è stato redatto un **documento contenente lo storytelling** che presenta le tendenze e le caratteristiche del commercio al dettaglio nella Lombardia per un pubblico generale.

## Criticità

In ultimo, è doveroso evidenziare alcune criticità riscontrate durante la realizzazione del progetto. In primo luogo, si sono riscontrate delle problematiche con lo strumento Colab: non ha permesso il lavoro contemporaneo da remoto del team tecnico, infatti, se ci si collega contemporaneamente da due dispositivi diversi il sistema presenta alcuni bug. Lo strumento, in aggiunta, non sembrerebbe indicato per creare delle dashboard compatte, permettendo unicamente la realizzazione di dashboard estese da "scrollare". In secondo luogo, invece, bisogna portare all'attenzione due questioni riguardanti i dataset: la numerosità delle osservazioni presenti nei tre è molto sbilanciata, rendendo più difficile la possibilità di confronti: nello specifico, il dataset delle grandi strutture ha 8990 osservazioni, quello delle medie 117482 e il dataset degli esercizi di vicinato 28535. Nei dataset delle grandi e medie strutture, inoltre, la categoria *Settore merc non alim* presenta un elevato numero di valori nulli impedendo l'analisi della stessa, che altrimenti avrebbe condotto a risultati fuorvianti.

Rispetto al processo di realizzazione della dashboard e al lavoro nel team, non si sono riscontrate particolari criticità. Solo nella fase di feedback è stato consigliato dal membro che rappresentava il pubblico di cambiare i colori dei grafici, predisponendo scale di colori diverse per i rispettivi dataset, così da facilitare il confronto. A tal proposito, il team tecnico si è mobilitato, ritenendo la richiesta conforme al fine ultimo del progetto stesso, e ha modificato i colori, che poi sono stati mostrati in un secondo incontro e approvati.

# Capitolo4

## Progetto 4: Creazione di un sistema software ex-novo

### Ruoli

Ruolo	Nostro Nome
Product Owner	Minh Davide Ragagni
Scrum Master	Davide Del Castello
Team Creazione Software	Tania Brusadin
Team Creazione Software	Giada Maurizzi

### Oggetto

La consegna richiede di creare ex novo il sistema software NSIGEC per il catalogo dei beni culturali italiani, basato su tecnologie web. Il sistema deve permettere ai catalogatori di varie organizzazioni di ricerca di autenticarsi tramite SPID, selezionare e approvare l'accesso alle loro organizzazioni e cercare nonché inserire dati sui beni culturali. Gli utenti ICCD, inoltre, possono gestire vocabolari controllati. Il progetto segue il paradigma Agile, con ruoli di Product Owner e Scrum Master definiti, un backlog prioritizzato delle funzionalità da sviluppare, epiche, user story e casi d'uso. Le prime due release sono previste per giugno e luglio 2024.

### Svolgimento

Per l'implementazione del nuovo sistema di catalogazione dei beni culturali italiano, denominato NSIGEC, l'Istituto Centrale di Catalogazione e Documentazione (ICCD) del Ministero della Cultura ha deciso di adottare il paradigma della programmazione Agile. Questo approccio ci ha permesso di gestire il progetto in modo flessibile e iterativo, rispondendo rapidamente ai cambiamenti e alle esigenze specifiche del sistema.

Il nostro team si è inizialmente organizzato dividendo i ruoli fondamentali per la gestione del progetto. Minh Davide Ragagni ha assunto il ruolo di Product Owner (PO), responsabile della definizione delle funzionalità del prodotto e della prioritizzazione del backlog. Davide Del Castello ha svolto il ruolo di Scrum Master (SM), facilitando il processo Agile e assicurando che le pratiche Scrum fossero seguite correttamente.

All'inizio del progetto, il PO ha definito le principali funzionalità richieste dal sistema NSIGEC, suddividendole in tre epiche principali, ognuna delle quali è stata poi suddivisa in user stories e tasks specifici. Queste epiche includevano:

- Autenticazione e autorizzazione dei catalogatori e utenti ICCD
- Ricerca e gestione dei beni culturali
- Gestione dei vocabolari controllati

Durante la prima fase, abbiamo lavorato alla definizione dettagliata delle epiche e delle relative user stories. Il PO ha identificato i requisiti per la funzionalità di login tramite credenziali

SPID per i catalogatori e utenti ICCD, la selezione dell'organizzazione di appartenenza e il processo di approvazione delle richieste di accesso da parte dei referenti delle organizzazioni.

Nella seconda fase, il backlog è stato prioritizzato assegnando le attività ai vari sprint. Il PO ha collaborato con lo SM per distribuire le user stories in modo equilibrato, tenendo conto delle risorse disponibili e della necessità di lavorare su altri progetti in parallelo. Questo ha richiesto una continua revisione e adattamento delle priorità per evitare sovraccarichi di lavoro, soprattutto quando ci siamo resi conto che il carico di lavoro era eccessivo. Grazie all'intervento dello SM, il backlog è stato riorganizzato per distribuire le attività in modo più omogeneo.

Durante la terza fase, abbiamo concentrato gli sforzi sulla realizzazione delle user stories prioritarie. Sulla base delle user stories realizzate, abbiamo definito i casi d'uso con l'aiuto degli altri due membri del team (Giada Maurizzi e Tania Brusadin), che hanno fornito un prezioso contributo esterno per affinare i requisiti e garantire una visione completa del progetto.

Nella fase finale, è stato effettuato un controllo ex post per verificare che tutte le funzionalità richieste dagli stakeholders fossero state implementate correttamente. Questo controllo ha permesso di identificare eventuali lacune e di apportare le ultime modifiche prima del rilascio del prodotto. Il sistema NSIGEC è stato rilasciato in due diverse release, una a giugno 2024 e una a luglio 2024, rispettando le scadenze previste.

## Output

Per il progetto 4, è stata prodotta una **bacheca su Trello** che esplicita il backlog prioritizzato, quindi le epiche e le user stories individuate, nonché i casi d'uso.

## Criticità

In un primo momento, si sono riscontrate alcune criticità. La principale è stata la mancata considerazione iniziale della realizzazione parallela di altri progetti, che ha portato a un sovraccarico di lavoro. Questo problema è stato risolto grazie alla ridefinizione delle priorità e alla ridistribuzione delle user stories su tutti gli sprint, operata congiuntamente dal PO e dallo SM.