

# CURRICULUM VITAE SCIENTIFICO-PROFESSIONALE

## ORNELLA IRRERA

### Dati personali

Data di nascita: 9 settembre 1995  
Luogo di nascita: Padova (Italia)  
Residenza: via Loreto 7, Padova (Italia)  
Nazionalità: Italiana  
Codice fiscale: RRRRLL95P49G224L  
Cellulare: +39 348 602 89 27  
E-mail: ornella.irrera@unipd.it

### Formazione

- *Ottobre 2021 - in corso: Dottorato di ricerca in Ingegneria dell'Informazione*, Scuola di dottorato in Ingegneria dell'Informazione, indirizzo in "Scienza e Tecnologia dell'Informazione", Ciclo: XXXVII, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.  
Titolo della tesi: *Data search in practice: How to find scientific datasets and to link them to the literature*  
Supervisore: Gianmaria Silvello; Co-Supervisore: Paolo Manghi
- *Novembre 2020*: Esame di Stato per l'abilitazione all'esercizio della professione di Ingegnere dell'Informazione
- *Settembre 2018 - Ottobre 2020: Laurea Magistrale in Ingegneria Informatica*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.  
Titolo della tesi: *Background Linking: joining entity linking with learning to rank models*.  
Relatore: Prof. Gianmaria Silvello  
Voto: 107/110.
- *Settembre 2014 - Marzo 2018: Laurea Triennale in Ingegneria Elettronica*, Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova.  
Titolo della tesi: *Parametri di Scattering: definizioni e applicazioni*.  
Relatore: Prof. Andrea Galtarossa  
Voto: 100/110.
- *Giugno 2014*: Maturità scientifica conseguita presso il Liceo Scientifico Statale "Enrico Fermi" di Padova.  
Voto: 84/100.

### Assegni di Ricerca

*Novembre 2020 - Ottobre 2021*: Assegno di ricerca di tipo A, presso il Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Padova. *Definizione e sviluppo di metodi supervisionati per la*

creazione e revisione semi-automatica di nanopublications da letteratura medica. Finanziato dal progetto H2020 ExaMode.

Supervisore: Gianmaria Silvello

## Attività di didattica

Anno accademico 2022/2023: 20 ore di didattica integrativa per il corso di *Foundations of Databases* relativo ai corsi di laurea magistrale in Computer Engineering e ICT for Internet and Multimedia dell'Università degli Studi di Padova

## Attività formative e di ricerca a livello internazionale

- 25 luglio 2022 - 29 luglio 2022: FISCI 2022 - Collaboration for Action: Sharing Knowledge Across Boundaries (Online event)
- 18 luglio 2022 - 22 luglio 2022: ESSIR 2022 - The 13th European Summer School in Information Retrieval, Lisbona, Portogallo.

## Breve introduzione all'attività di ricerca

Dall'ottobre 2021 collaboro col CNR - Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" di Pisa all'interno del progetto Europeo OpenAIRE<sup>1</sup>. OpenAIRE realizza un'infrastruttura a sostegno della Scienza Aperta, il cui obiettivo è quello di fornire dei servizi e policies per monitorare e salvaguardare la ricerca in Europa e facilitare la comunicazione scientifica, promuovendo l'accesso e la pubblicazione di ogni elemento che è parte attiva della produzione dell'esperimento, non solo quindi pubblicazioni, dati e software, ma anche progetti europei, autori, enti e organizzazioni.

OpenAIRE aggrega i metadati riguardanti pubblicazioni, datasets, software, organizzazioni, autori, progetti e link tra tali prodotti, che sono salvati in data repositories, journals, e archivi e costruisce un grafo, l'OpenAIRE Research Graph, dove i metadati che descrivono gli outcome della ricerca, delle organizzazioni, degli autori e dei progetti sono tra loro interconnessi. I nodi di tale grafo sono gli outcomes descritti tramite i propri metadati, gli archi invece, hanno un'etichetta che rappresenta la semantica della relazione; tale semantica segue lo schema definito da DataCite<sup>2</sup>. La presenza di archi è fondamentale per dare visibilità ad outcome diversi, come ad esempio tutti dati collegati a una pubblicazione, oppure tutte le pubblicazioni che vengono citate. Al fine di scoprire nuovi archi e connettere nuovi outcome, sono applicati algoritmi di inferenza che si basano sui metadati salvati all'interno dei nodi. Attualmente, una visione ancora molto incentrata sulla pubblicazione del solo articolo testuale, fa in modo che i metadati relativi ad altri outcome come dati e software non siano definiti in dettaglio e che non siano definite le relazioni tra questi ed altri prodotti: questa impatta in modo negativo sulla possibilità di utilizzare i metadati per stabilire ed inferire nuove connessioni tra diversi outcome. Dati e software rimangono così isolati, portando quindi a una complessa reperibilità di questi e a problemi nell'attribuzione dei crediti ai relativi autori. Lo scopo del progetto di dottorato è quindi quello di sviluppare nuove tecniche e metodologie per stabilire nuove connessioni tra dati e pubblicazioni.

Durante il primo anno di progetto mi sono occupata della costruzione di una ground-truth a partire dal

---

<sup>1</sup><https://www.openaire.eu/>

<sup>2</sup><https://doi.org/10.14454/3w3z-sa82>

grafo originale di OpenAIRE. La ground-truth riguarda una precisa comunità scientifica di OpenAIRE, la *European Marine Science Community*, ed è stata curata in maniera semi-automatica. Su tale ground-truth ho effettuato un'analisi incentrata su come le pubblicazioni citano i dataset e i software all'interno della comunità elaborata. Al fine di visualizzare, analizzare e curare il grafo ho utilizzato *Neo4j*, un tool per la gestione dei property graphs.

Durante l'assegno di ricerca mi sono principalmente occupata dello sviluppo di MedTAG, un'applicazione web per l'annotazione manuale di documenti medici. Tale tool, il cui backend è stato sviluppato in *Django* e il frontend in *React.js*, è stato rilasciato come tool open-source installabile tramite Docker. Il tool è customizzabile, gestisce più di una lingua, implementa inter-annotator agreement tramite majority vote. I documenti possono essere annotati in quattro modi differenti: *label* annotation permette di associare una o più classi all'intero documento, *concepts* annotation permette di associare una serie di concetti al documento, *mentions* annotation permette di identificare parti di testo che possono essere associati a uno o più concetti e *linking* che consiste nell'associare i concetti alle mentions identificate.

## Publicazioni

### Riviste Internazionali

- Giachelle, F., Irrera, O., & Silvello, G. (2021). MedTAG: a portable and customizable annotation tool for biomedical documents. *BMC Medical Informatics and Decision Making*, 21.
- Marchesin, S., Giachelle, F., Silvello, G., Marini, N., Atzori, M., Boytcheva, S., Buttafuoco, G., Francesco Ciompi, F., Di Nunzio, G. M., Fraggetta, F., Irrera, O., Müller, H., Primov, T., & Vatrano, S. (2022). Empowering Digital Pathology Applications through Explainable Knowledge Extraction Tools. *Journal of Pathology Informatics*. Accepted for publication 2022.

### Atti di Conferenza

- Irrera O. (2022) Data Search in Practice: how to find scientific datasets and to link them to the literature. In *FDIA ESSIR 2022* (Accepted).
- Mannocci, A., Irrera, O., & Manghi, P. (2022). Open Science and Authorship of Supplementary Material. Evidence from a Research Community. arXiv preprint arXiv:2207.02775.
- Mannocci, A., Irrera, O., & Manghi, P. (2022). Will open science change authorship for good? Towards a quantitative analysis. arXiv preprint arXiv:2207.03121.
- Giachelle, F., Irrera, O., & Silvello, G. (2022). DocTAG: A Customizable Annotation Tool for Ground Truth Creation. In *European Conference on Information Retrieval* (pp. 288-293). Springer, Cham.
- Irrera, O., Giachelle, F., & Silvello, G. (2022). An Open-Source annotation tool for collaboratively annotating biomedical documents. In *Proceedings of the 18th Italian Research Conference on Digital Libraries*, volume 3160 of *CEUR Workshop Proceedings*, CEUR-WS.org. <http://ceur-ws.org/Vol-3160/short17.pdf>.
- Irrera, O., & Silvello, G. (2021). Background Linking: Joining Entity Linking with Learning to Rank Models. In *IRCDL* (pp. 64-77).

Le dichiarazioni rese nel presente curriculum sono da ritenersi rilasciate ai sensi degli artt. 46 e 47 del D.P.R. n. 445/2000.  
Il presente curriculum non contiene dati sensibili e dati giudiziari di cui all'art. 4, comma 1, lettere d) ed e) del D.Lgs.  
30.06.2003 n. 196.

Padova, 12 gennaio 2023