

CV SUMMARY – ANNA BERNASCONI

1. GENERAL INFORMATION

DATE OF BIRTH 08/07/1990

EDUCATION

- 2021 **PhD Degree** (*cum Laude*) in Information Technology, Politecnico di Milano, Italy
- 2016 **Master of Science Degree** in Computer Science, University of Illinois at Chicago, USA
(Final Grade 3.85/4.00)
- 2015 **Master of Science Degree** in Computer Engineering, Politecnico di Milano, Italy
(Final Grade 110/110)
- 2012 **Bachelor Degree** in Mathematical Engineering, Politecnico di Milano, Italy (Final Grade 100/110)

CAREER

- Oct 2022 – present **Research Fellow – RTDA** with 0.4 PO (university-funded units) at Politecnico di Milano, Italy
- Jan 2022 – June 2022 **Visiting Researcher** at Universitat Politècnica de València, Spain (6 months)
- Apr 2021 – Dec 2021 **Visiting Researcher** (remote status) at Universitat Politècnica de València, Spain (9 months)
- Nov 2020 – Sept 2022 **Post-Doctoral Researcher** (“assegnista di ricerca”), Politecnico di Milano, Italy
- Nov 2016 – Oct 2020 **PhD Student**, Politecnico di Milano, Italy
- Nov 2014 – Feb 2016 **Junior Business Intelligence Consultant and Data Warehouse Architect**, Quanyca S.R.L., Monza, Italy

SCIENTIFIC NATIONAL HABILITATION

Type of habilitation	Country	SSD (if Italian habilitation) or topic area	Date of achievement
Associate Professor (II Fascia)	Italy	09/H1 (SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI)	15/06/2023
Associate Professor (II Fascia)	Italy	01/B1 (INFORMATICA)	12/12/2023

RESEARCH INTERESTS

- **Research area 1: Databases**, conceptual modeling, data integration, knowledge management.
- **Research area 2: Bioinformatics**, human genomics, viral genomics, integrative data analysis.
- **Research area 3: Data science**, data mining, big data analysis, natural language processing.

2. QUALITY OF SCIENTIFIC AND/OR PROJECT PRODUCTION

PRODUCTIVITY AND IMPACT METRICS

- **Scientific Productivity: 70** publications (**58** entries and **3,628*** co-authors according to Scopus):
 - Author/Co-author of **29** international journal papers (**18** top ranked **Q1** journal papers based on WoS and Scimago), **1** Editorial article, **2** Consortia papers, and **1** National journal article.
 - Author/Co-author of **19** scientific publications on peer-reviewed conferences, including **1** top-level **A+ (Class 1)** conference paper (according to GII-GRIN-SCIE conf. ranking), **9 A (Class 2)** conferences papers (according to CORE), and **11** workshops papers/posters.
- **Publication Impact (on December 12th, 2023):**
 - Google Scholar: h-index **16** citations **1,310**
 - Scopus: h-index **12** citations **817**

*About 3,500 co-authors on Scopus are due to participation to consortia papers with the COVID-19 Host Genetics Initiative

AWARDS AND RECOGNITIONS

- 2023 Recipient of the **CAiSE PhD Award 2023**, 35th International Conference on Advanced Information Systems Engineering, granted to an outstanding Ph.D. thesis in the field of Information Systems Engineering
- 2022 **‘Runner Up Paper’ Award**, 41st International Conference on Conceptual Modeling, **ER 2022**
- 2022 Recipient of the **Award for the Best Ph.D. Thesis on Big Data & Data Science 2022, CINI Lab on Big Data**, granted to excellent PhD theses based on originality and potential impact on the Big Data e Data Science community
- 2021 Recipient of **Dimitris N. Chorafas Foundation 2021 Prize**, from Weizmann Institute of Science and Feinberg Graduate School, in recognition of results obtained during the research activity pursued during my PhD – the foundation awards prizes in engineering/medicine/natural sciences where results can have a significant impact on future research and practical applications
- 2021 Recipient of the Springer award, from Politecnico di Milano IT PhD Board of professors

TALKS AND SEMINARS

- 2023 **Lightning Talk.** *Building an Integrated Repository for Genomic Data.* For the Research Data Alliance collaborative meeting on FAIRification of Genomic Tracks (Nov 2023, online)
- 2023 **Interview.** *Building a metadata model for COVID-19.* For CSC – IT Center for Science Ltd, administered by the Finnish Ministry of Education and Culture (Oct 2023, online)
- 2022 **Invited seminar.** *Data-driven SARS-CoV-2 understanding and hunting.* At Information Sciences Institute - University of Southern California (Sep 2022, online)
- 2022 **Invited seminar.** *Data analysis for unveiling the SARS-CoV-2 evolution.* At the First International Workshop on Data Analysis for Life Science (DALs@ECML-PKDD2022) (Sep 2022, online)
- 2022 **Invited seminar.** *Data-bases-driven research on viral genomes and SARS-CoV-2.* At Universidade Eduardo Mondlane, Maputo, Mozambique (July 2022, online)
- 2022 **Seminar.** *Data-driven SARS-CoV-2 understanding and hunting (searching for the new Omicron).* At Politecnico di Milano and Universitat Politècnica de València (Jan 2022, UPV)
- 2021 **Invited Talk.** *Modeling, integrating, and searching processed genomic datasets.* At “Integrating Diverse Datasets Webinar Series” by Front Line Genomics (Nov 2021, online)
- 2019 **Invited Talk.** *Metadata Integration Framework for Genomic Datasets.* At “Challenges in Data-Driven Genomic Computing” Workshop, Como, Villa del Grumello, Italy (Mar 2019)
- 2019 **Seminar.** *Metadata Integration Framework for Genomic Datasets.* At EPFL, Lausanne, Switzerland (Feb 2019)
- 2017 **Seminar.** *Data-Driven Genomic Computing: Making Sense of the Signals from the Genome through systematic application design and open data repository.* At ETH Zurich and IBM Zurich, Switzerland (Jul 2019)

Additionally, I gave **2 tutorials** on “*Conceptual models and databases for searching the genome*” at the ER 2021 and the EDBT 2022 conferences, and **14 paper presentations** (between 2017 and 2023) at the conferences ER 2017, SWAT4HCLS 2018, ER 2019, ER 2020, REFSQ 2021, ER 2021, ICC 2022, SWAT4HCLS 2022, ER 2022, CAiSE 2023, IEEE SOSE 2023, ER 2023, and at the workshops DTMBio@CIKM2018, EmpER@ER2019.

INSTITUTIONAL RESPONSIBILITIES

- 2022 – present Member of the Interuniversity Teaching Board for the Master “Bioinformatics for Computational Genomics” as Professor, Politecnico di Milano & University of Milan
- 2023 – present Member of the Scientific Committee of the Honours Program, Computer Science and Engineering Track (Data, web, and society), Politecnico di Milano

ORGANIZATION OF SCIENTIFIC MEETINGS

- 2020–2023 **Chair and co-organizer**, International Workshop on Conceptual Modeling for Life Sciences (CMLS) for **four editions**, co-located with the ER conference (online 2020–2022 and Lisbon, Portugal in 2023)
- 2022–2023 **Chair and co-organizer**, International Workshop on Web Applications for Life Sciences (WALS) for **two editions**, co-located with the ICWE conference (Bari, Italy in 2022 and Alicante, Spain in 2023)
- 2023 **Publicity Chair** of the 42nd International Conference on Conceptual Modeling (ER 2023), November 2023, Lisbon, Portugal
- 2023 **Program Chair**, 9th IEEE International Conference on Big Data Computing Service and Machine Learning Applications (IEEE BigDataService 2023) (July 2023, Athens, Greece)
- 2022 **Track Chair** (Big Data and Analytics for Healthcare) and **Proceedings Chair**, 8th IEEE International Conference on Big Data Computing Service and Machine Learning Applications (IEEE BigDataService 2022) (August 2022, San Francisco Bay Area, USA)
- 2019 **Organization Member**, “Challenges in Data-Driven Genomic Computing” ERC AdG GeCo Workshop (March 2019, Como, Italy)

Additionally, I served (or currently serving) as **Program Committee Member** for CAiSE 2024, Demo Track SIGMOD/PODS 2024, FNA@BIBM 2023, KDIR 2023, KEOD 2023, ER 2023, ER Forum 2023, EmpER@ER 2023, CIBB 2023, HiDA@ADBIS 2023, IWWBIO 2023, NPAHI@IEEE ICHI 2023, CAiSE 2023 Project Exhibitions, FNA@BIBM 2022, DAL@ECML-PKDD 2022, IWBBIO 2022, CIBB 2021.

PARTICIPATION IN EDITORIAL BOARDS

Academic Editor for Plos One (ISSN 1932-6203, Public Library of Science) – since Feb 2023

Editorial Board Member of BMC Bioinformatics (ISSN 1471-2105, Springer Nature) – since May 2022

- 2023 **Guest Editor** for the BMC Bioinformatics Collection “Big data management in biological domains”
- 2023 **Guest Editor** for the Journal Future Generation Computer Systems (ISSN: 0167-739X, Publisher: Elsevier), for the Special Issue “Special Issue on Big Data Computing Service and Machine Learning Applications”
- 2023 **Supplement Editor** for the BMC Bioinformatics/BMC Medical Informatics and Decision Making (ISSN: 1472-6947) Journals, for the joint supplement “Selected articles from the CMLS/WALS 2022 workshops”
- 2022 **Guest Associate Editor** for Research Topic “Identification of phenotypically important genomic variants” on Frontiers in Bioinformatics (ISSN: 2673-7647, Publisher: Frontiers Media S.A.)
- 2022 **Volume Editor** for “Current Trends in Web Engineering - ICWE 2022 International Workshops, BECS, SWEET and WALS, Bari, Italy, July 5–8, 2022, Revised Selected Papers”. Conference Proceedings Book in Springer CCIS, volume 1668
- 2022 **Volume Editor** for “Computational Intelligence Methods for Bioinformatics and Biostatistics”, 17th International Meeting, CIBB 2021, Virtual Event, November 15–17, 2021, Revised Selected Papers. Conference Proceedings Book in Springer LNBI, volume 13483
- 2022 **Guest Editor** for the BioTech Journal (EISSN 2673-6284, Publisher: MDPI), Special Issue “Bioinformatics: Present and Future challenges”
- 2022 **Supplement Editor** for the BMC Bioinformatics/BMC Medical Informatics and Decision Making Journals, for the joint supplement “Selected articles from the CIBB 2021 conference”
- 2022 **Supplement Editor** for the BMC Bioinformatics / BMC Medical Informatics and Decision Making / Genomic Data Journals, for the joint supplement “Selected articles on Conceptual Modeling for Life Sciences (CMLS/ER 2021)”
- 2021 **Supplement Editor** for the BMC Bioinformatics Journal, for the supplement “Selected articles from the 1st International Workshop on Conceptual Modeling for Life Sciences (CMLS 2020)”

Overall, I acted as **editor for more than 30 journal manuscripts** and as a **reviewer for more than 60 journal manuscripts from 25 different journals** (including Applied Clinical Informatics Journal; Briefings in Bioinformatics; Genomics, Proteomics & Bioinformatics; GigaScience; ACM Journal of Data and Information Quality; International Journal of Medical Informatics; Machine Learning; npj Digital Medicine, and Scientific Data).

3. TEACHING ACTIVITIES

COURSES WITH A PRIMARY RESPONSIBILITY

Institution name	Course name	Credits	Role	Reference Study Course	Time period	Students Evaluation
Politecnico di Milano	Data management and analysis for computational biology	2 (of 5)	Lecturer	Doctoral Program in Information Technology	AA 2022-2023	N/A
Politecnico di Milano	Programming	6	Lecturer	Bioinf. for Computational Genomics – Master Level	AA 2022-2023 AA 2023-2024	N/A (interuniversity master program)

OTHER TEACHING ACTIVITY

Institution name	Course name	Credits	Role	Reference Study Course	Time period	Students Evaluation
Politecnico di Milano	Informatica A	10	Teaching assistant	Mathematical Eng. – Bachelor Level	AA 2016-2017 AA 2017-2018 AA 2021-2022	High 3.4/4.0 3.4/4.0
Politecnico di Milano	Basi di Dati 1	5	Teaching assistant	Computer Science Eng – Bachelor Level	AA 2021-2022 AA 2022-2023 AA 2023-2024	3.3/4.0 2.9/4.0 N/A
Politecnico di Milano	Data Bases	6	Teaching assistant	Bioinf. for Computational Genomics – Master Level	AA 2021-2022 AA 2022-2023 AA 2023-2024	N/A (interuniversity master program)
Universitat Politècnica de Valencia	Guest Lectures on Modeling Languages	Guest lectures (10h+5h)	Guest lecturer	Computer Science courses (Bachelor and Master's levels)	April 2022	N/A

SUPERVISION OF MASTER, DOCTORAL STUDENTS

2017 – present **Co-advisor** of **8** successfully defended Master Theses (two students started the PhD in the group) and **4** bachelor/post-graduate projects in Computer Engineering degree, Politecnico di Milano, Italy

2021 – present **Co-supervisor** of **4** (currently) Doctoral Students in Information Technology, Politecnico di Milano, Italy

4. PARTICIPATION/RESPONSIBILITY FOR FUNDED PROJECTS**PARTICIPATION IN COMPETITIVE RESEARCH PROJECTS**

Project Acronym	Time Period	Funding Institution	Funding Scheme	Role of the applicant	Budget for applicant's institution
SENSIBLE (Small-data Early warning System for viral pathogens In public Health), <i>project n. P2022CNN2J</i>	Dec 2023 – Nov 2025	MUR	PRIN PNRR 2022	Principal Investigator	186K Euros
BETTER (Better real-world health-data distributed analytics Research platform), <i>grant n. 101136262</i>	Dec 2023 – May 2027	EU	Horizon Europe Framework Programme	Task Leader in WP4: “Health datasets FAIRification and preprocessing”	639K Euros

TETYS (Topics Evolution That You See), <i>project n. SEARCH OC2_18</i>	Sep 2023 – Aug 2024	EU	NGI Initiative (NGI Search Project)	Principal Investigator	150K Euros
Future Artificial Intelligence Research (FAIR)	Jan 2023 – Dec 2025	EU	NextGenerationEU program, PNRR-PE-AI scheme	Task Leader: “Graph-based ML for Network Medicine”	-
Innovation Activity ‘Virusurf’, <i>project n. 20663</i>	Jul 2021– Dec 2021	EIT	EIT Digital, “Data against COVID-19” program	Task leader: “User experience and interfaces”	172K Euros
Data-Driven Genomic Computing (GeCo), ERC Advanced grant <i>grant n. 693174</i>	Sep 2016 – Aug 2021	EU	H2020	WP leader: “WP4: Integrated access to large data sources”	2.5 million Euros

5. TECHNOLOGY TRANSFER

DEVELOPMENT OF PRODUCTS / OPEN-SOURCE TOOLS / APPLICATIONS / SYSTEMS / SERVICES

Contribution to open-source tools, which aim to facilitate the work of cancer genomics researchers, clinicians, geneticists, molecular biologists:

- OpenGDC (<http://www.bioinformatics.deib.polimi.it/opengdc/>): extraction and integration tool for cancer genomics data;
- GenoSurf (<http://gmql.eu/genosurf>): multi-ontology semantic search system providing access to a consolidated collection of human genome datasets for bioinformatic tertiary analysis;
- GeMI (<http://gmql.eu/gemi>): interactive interface to annotate genomic samples exploiting a deep-learning-based prediction model, accelerating genomic metadata extraction into structured formats;
- VarSum (<http://gmql.eu/popstudy>): API computing summary statistics on user-defined populations and their variants, using a large genomic data repository;
- META-BASE (<https://github.com/DEIB-GECO/Metadata-Manager>): tool that integrates and enriches heterogeneous biological and clinical data sources, exploiting several configurations to build a repository.

Contribution to open-source tools, which aim to facilitate the work of experts in virology, phylogenetics, epidemiology, pathogenesis, immunology, focusing on SARS-CoV-2 and other viral species:

- Virusurf (<http://gmql.eu/virusurf>): search system for public viral sequences described by their metadata and exhibiting nucleotide and amino acid mutational patterns;
- EpiSurf (<http://gmql.eu/epivirusurf/>): search system for evaluating the impact of viral mutational patterns on epitopes, i.e., areas of the virus recognized by the human immune system;
- VirusViz (<http://gmql.eu/virusviz>): visualization tool for the comparison of the mutational distributions of user-selected viral populations;
- ViruClust (<http://gmql.eu/viruclost>): data analysis tool to perform temporal and spatial comparison of mutational patterns of SARS-CoV-2 sequences;
- VariantHunter (http://gmql.eu/variant_hunter): tool for fast detection of emerging SARS-CoV-2 variants and sub-variants;
- CoV2K API (<http://gmql.eu/cov2k/api/>): knowledge base of SARS-CoV-2 sequence mutations, variants, and their effects;
- CoVEffect (<http://gmql.eu/coveffect>): interactive web interface for deep learning-assisted scientific text annotation with effects of SARS-CoV-2 mutations and variants.

Contribution to open-source tools for the exploration of COVID-19-related research literature:

- GRAPH-SEARCH (<http://gmql.eu/graph-search>): explorer of COVID-19-related literature using graph-

based queries with shortest paths and sophisticated ranking;

- CORToViz (<http://gmql.eu/cortoviz>): user-friendly visualizer for trends of topics extracted from COVID-19 scientific research abstracts, with easy-to-drive statistical testing. *This is the first prototype developed in the context of the NGI Search Program (TETYS project), aiming to build the next-generation open-source Web topics explorer (applicable to many other text document corpora).*

PARTICIPATION OR LEADERSHIP IN INDUSTRY-RELEVANT EXPLOITATION OR STANDARDIZATION INITIATIVES

Participation in the **COVID-19 Host Genetics Initiative** for designing the COVID-19 clinical data dictionary for describing COVID-19 patients' phenotype; our work was used in the publication <https://doi.org/10.1038/s41586-021-03767-x> on Nature and adopted within ELIXIR-CONVERGE, as deemed appropriate as standardization guidelines (<https://doi.org/10.5281/zenodo.4893222>)

Participation in the **FAIRification of Genomic Annotations RDA Working Group**, for the standardization of Genomic Annotations and FAIRification data processing pipelines (<https://www.rd-alliance.org/fairification-genomic-tracks-data-driven-life-science-through-granular-discovery-biological-sequence>).