

## CV – ANNA BERNASCONI (UPDATED 03/05/2026)

Date of Birth 08/07/1990

Personal home page: <https://annabernasconi.faculty.polimi.it/>

Full CV: <https://annabernasconi.faculty.polimi.it/wp-content/uploads/FullCV.pdf>

### Summary.

*Anna Bernasconi is an Assistant Professor (Ricercatore Tenure-Track) with the Dipartimento di Elettronica, Informazione e Bioingegneria, the Data Science Lab, and the Bioinformatics Lab at Politecnico di Milano. She received her Ph.D. in Information Technology from Politecnico di Milano in February 2021, with a thesis on genomic data integration that won three prizes and was published in Springer's LNBIP. She was a visiting researcher at Universitat Politècnica de València, Jan-Jun 2022. Her research interests lie at the intersection of Bioinformatics, Databases, and Data Science Methods, focusing on building open-source tools and services.*

### Highlights.

- **Senior assistant professor** at DEIB, Politecnico di Milano, since May 2024.
- **National scientific qualification** as Italian associate professor (sectors 09/H1 and 01/B1), until 2035.
- Author/Co-author of 46 international journal papers, of which 37 articles in Q1 journals based on WoS and Scimago, including SciAdv, NatComm, NAR (2), SciData, GigaScience, Bioinformatics (2), Briefings in Bioinformatics (2), Information Processing & Management, Journal of Big Data, and Expert Systems with Applications. Author/Co-author of > 40 conferences/ workshops publications, including 4 in A++/A+ venues (SIGMOD, ICDE and PKDD-ECML) and 11 in A venue (according to GII-GRIN-SCIE).
- Recipient of the **Viruses Early Career Investigator Award 2026** (MDPI) and **three PhD thesis awards**: CAiSE PhD Award 2023, 35th Int. Conf. on Advanced Information Systems Engineering; Award for the Best Ph.D. Thesis on Big Data & Data Science 2022 – CINI Lab on Big Data; and N. Chorafas Foundation 2021 Prize from Weizmann Institute of Science and Feinberg Graduate School.
- Participation in 7 **competitive research projects**, funded by public institutions (2 National level, 5 European level), as **PI** (2), **WP-leader** (1), or **Task-leader** (4); Co-chair of Research Data Alliance FGA Working Group.
- **Invited speaker** (17 talks, including MMNGS'2025, CAiSE'2024, RDA Plenary meeting VP22, DAL@ECML'2022, ITADATA'2022, UEM, USC, UPV, EPFL, ETH, and IBM Zurich; 2 tutorials on “Conceptual models and databases for searching the genome” at the ER 2021 and the EDBT 2022; 19 conference paper presentations).
- **Academic editor** for Plos One, Scientific Data, BMC Bioinformatics, Viruses (organizer of 9 Supplements/Special issues); Editorial Review Board Member for ACM Journal of Data and Information Quality and PVLDB Volume 18; PC Member for CAiSE, RCIS, CIKM, ER, Demo Track SIGMOD/PODS.
- **Chair roles**: PC Chair (ER2026); Workshops Chair (ER2025); Forum Chair (RCIS2026). Founder of workshops on conceptual models (CMLS@ER), web applications (WALS@ICWE), and inf. sys. (iSAILS@CAiSE) for life sciences.
- **Lecturer** of the Doctoral courses “Data Management and Analysis for Computational Biology” (a.y. 2022-2023) and “Data Science for Bioinformatics and Healthcare” (a.y. 2025-2026); the course “Databases” (since a.y. 2024-2025, in the Computer Engineering Bachelor Degree), and the course “Programming” (since a.y. 2022-2023, in the Master on Bioinformatics for Computational Genomics).
- **(Co-)Supervisor** of 4 PostDocs, 6 Ph.D. students and 22 M.Sc. students at Politecnico di Milano (two at UPV).
- **Co-founder** of LegisRatio S.r.l. (in 2025), an innovative start-up in the LegalTech domain, accredited spin-off of Politecnico di Milano, supporting companies in navigating the legislative system

## Education.

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

## Career.

May 2024 – present **Tenured-track Researcher – RTT** at Politecnico di Milano, Italy  
Oct 2022 – Apr 2024 **Research Fellow (RTDA)**, 0.4 university-funded units, at Politecnico di Milano, Italy  
Jan 2022 – June 2022 **Visiting Researcher** at Universitat Politècnica de València, Spain (6-month)  
Apr 2021 – Dec 2021 **Visiting Researcher** (remote) at Universitat Politècnica de València, Spain (9m)  
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 **Jr. Business Intelligence Consultant and DataWH Architect**, Quanyca S.R.L.

**Scientific National Habilitation.** Associate Professor (II Fascia) Italy: 09/H1 (SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI); 01/B1 (INFORMATICA) [valid through 2023-2034]

## 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.

## Scientific Productivity: 90 publications

- Author/Co-author of **47** international journal papers (**37** top ranked **Q1** journal papers based on WoS and Scimago), **3** Editorial articles, **2** Consortia papers, and **1** National journal article.
- Author/Co-author of **26** scientific publications on peer-reviewed conferences, including **4** top-level **A++/A+ Class 1** conference papers (SIGMOD, ICDE and PKDD-ECML), **10 A (Class 2)** conferences papers (according to GII-GRIN-SCIE conf. ranking), and **14** workshops papers/posters.
- Google Scholar: h-index **21**; citations **2,323**; <https://scholar.google.com/citations?user=rCnI9L4AAAAJ>
- Scopus: h-index **17**; citations **1,476**; <https://www.scopus.com/authid/detail.uri?authorId=57195587336>
- WebOfScience: h-index **14**; cit. **1,140**; <https://www.webofscience.com/wos/author/record/AAF-2594-2019>

## Awards and Recognitions.

- Recipient: **Viruses Early Career Investigator Award 2026**, granted by the MDPI Viruses Journal, to acknowledge the achievements of early career investigators in the field of virology
- Top-3 champion on Italian selection for the 4th edition of the Frontiers Planet Prize with Paper "Lightweight multiscale early warning system for influenza A spillovers" published on Science Advances
- DEIB Young Researchers' Best Publications Award 2025 for the publication "An LLM-assisted ETL pipeline to build a high-quality knowledge graph of the Italian legislation" on Information Processing & Management, co-authored with the PhD student A. Colombo
- Member of the team **awarded by the Italian Chamber of Deputies**, XIX Legislature, for the LegisLab application”, July 2024
- Recipient: **CAiSE PhD Award 2023**, 35th Int. Conference on Advanced Information Systems Engineering, granted to an outstanding Ph.D. thesis in the field of Information Systems Engineering
- Recipient: **‘Runner Up Paper’ Award**, 41st International Conference on Conceptual Modeling, **ER 2022**
- Recipient: **Award for the Best Ph.D. Thesis on Big Data & Data Science 2022, CINI Lab on Big Data**, to excellent PhD theses based on originality and impact on the Big Data and Data Science community
- Recipient: **Dimitris N. Chorafas Foundation 2021 Prize**, from Weizmann Institute of Science and Feinberg Graduate School, in recognition of results of the research activity pursued during my PhD

## Talks and Seminars at National and International Events.

I gave several invited seminars and speeches (2017-2024), including at the CAiSE conference 2024, the RDA Plenary meeting VP 22, CSC – IT Center for Science Ltd, University of Southern California, ITADATA 2022 conference, DALs workshop, Universidade Eduardo Mondlane, UPV, EPFL, ETH, and IBM Zurich.

I presented **2 tutorials** on “*Conceptual models and databases for searching the genome*” at the ER 2021 and the EDBT 2022 conferences, and **20 paper presentations** (2017—2026) at the conferences CaiSE 2026, CAiSE 2024, SWAT4HCLS2024, SIGMOD 2024, ICDE 2024, ER 2023, IEEE SOSE 2023, CAiSE 2023, ER 2022, SWAT4HCLS 2022, ICC 2022, ER 2021, REFSQ 2021, ER 2020, ER 2019, SWAT4HCLS 2018, ER 2017, and at the workshops EmpER@ER2019, DTMBio@CIKM2018.

### Organization of Scientific Meetings.

I am founder and co-organizer of the 2<sup>nd</sup> edition (2026) of the Transforming Graph Data Workshop (**TGD**) [co-located with the EDBT conference, ]first six editions 2020-2025 of the International Workshop on Conceptual Modeling for Life Sciences (**CMLS**) [co-located with the ER conference], first three editions 2022-2024 of the International Workshop on Web Applications for Life Sciences (**WALS**) [co-located with the ICWE conference], and first edition 2025 of the International Workshop on Information Systems and AI for Life Sciences (**iSAILS**) [co-located with the CAiSE conference].

I will act as **Program Chair**, 45<sup>th</sup> edition (ER 2026), **Workshops co-chair**, 44<sup>th</sup> edition (ER 2025), **Publicity co-chair** during the 42<sup>nd</sup> edition (ER 2023) and 43<sup>rd</sup> edition (ER 2024) of the International Conference on Conceptual Modeling.

I acted as **Program Chair** in the 9th IEEE International Conference on Big Data Computing Service and Machine Learning Applications (IEEE BigDataService 2023), as **Track Chair** (Big Data and Analytics for Healthcare) and **Proceedings Chair**, 8th IEEE BigDataService 2022.

Additionally, I served as **Review Board Member** for PVLDB Volume 18 and **Program Committee Member** for CAiSE 2025, RCIS 2025, CIKM 2024, ER 2024, CAiSE 2024, Demo Track SIGMOD/PODS 2024 and 2025, FNA@BIBM 2023, KDIR 2023/2024, KEOD 2023/2024, 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 and Reviewing Tasks.

Since Apr 2026: **Editorial Board Member** for Viruses (ISSN 1999-4915, Publisher: MDPI)

Since Mar 2025: **Editorial Board Member** for Scientific Data (ISSN 2052-4463, Publisher: Springer Nature)

Since June 2024: **Associate Editor** for Bioinformatics and Biology Insights (ISSN1177-9322, Publisher: Sage)

Since Apr 2024: **Editorial Review Board Member** for ACM Journal of Data and Information Quality

Since Feb 2023: **Academic Editor** for Plos One (ISSN 1932-6203, Public Library of Science)

Since May 2022: **Editorial Board Member** of BMC Bioinformatics (ISSN 1471-2105, Springer Nature)

From 2021 to 2024 I acted as a **Guest Editor** for 9 Special Issues for the ACM Journal of Data and Information Quality, BMC Bioinformatics, Future Generation Computer, BMC Medical Informatics and Decision Making. Overall, I acted as **editor for 38 journal manuscripts** and as a **reviewer for more than 60 journal manuscripts from 30 different journals** – <https://www.webofscience.com/wos/author/rid/AAF-2594-2019>.

In 2024 I was **reviewer for funding proposals** for 1) the Research Infrastructure - National consortia (RI programme) by the Dutch Research Council (NWO); 2) the UK Research and Innovation (UKRI), Medical Research Council.

### Teaching.

- **Professor: Basi di Dati**, 5 CFU (Bachelor Degree, Engineering of Computing Systems, Politecnico di Milano) - **2024/25-present**
- **Professor: Programming**, 6 CFU (Master Degree, Bioinformatics for Computational Genomics, Politecnico di Milano & University of Milan) - **2022/23 – present**
- **Teaching assistant: Basi di Dati**, 5 CFU (Bachelor Degree, Engineering of Computing Systems, Politecnico di Milano - **2021/22 – 2023/24**
- **Teaching Assistant: Data Bases**, 6 CFU (Master Degree, Bioinformatics for Computational Genomics, Politecnico di Milano & University of Milan - **2021/22 – 2023/24**
- **Teaching Assistant: Informatica A**, 10 CFU (Bachelor Degree, Mathematical Engineering, Politecnico di Milano - **2016/17, 2017/18, 2021/22**
- **Professor: Data Science for Bioinformatics**, Ph.D. School of Information Technology (2 CFUs), Politecnico di Milano, Nov 2025
- **Professor: Data Management and Analysis for Computational Biology**, Ph.D. School of Information Technology (2 CFUs), Politecnico di Milano, Jan 2023
- **Lecturer: Ontological Unpacking: a Conceptual Model of the Human Genome**, Guest Lecturer for an intensive course of the Master 'Systèmes d'Information et de Connaissance' - The University Paris 1 Pantheon-Sorbonne (5 hours) - Universitat Politècnica de València, Jun 2022
- **Lecturer: UML and OntoUML modeling languages and ontological unpacking applications to the genomics domain**, Guest Lecturer in Bachelor/Master courses of Engineering of Computing Systems (10 hours), Universitat Politècnica de València, Apr 2022

## Supervision of Master, Doctoral Students.

**Co-advisor of 8, advisor of 12** successfully defended Master Theses (two students started their PhD under my supervision) and **7** bachelor/post-graduate projects in Computer Engineering degree, Politecnico di Milano, Italy. **Advisor of 6** ongoing Master Theses.

**Co-supervisor of 2** current, **3** former Doctoral Students in Information Technology (two started their PostDocs under my supervision), and **4** postdoctoral researchers at Politecnico di Milano; **1** Visiting PhD from UPV (at Polimi), Spain and **1** PhD at UPV, Spain.

## Institutional Responsibilities.

- Since 2023: Member of the Scientific Committee of the Honours Program, Computer Science and Engineering Track (Data, web, and society), Politecnico di Milano
- Since 2022: Member of the Interuniversity Teaching Board for the Master “Bioinformatics for Computational Genomics” as Professor, Politecnico di Milano & University of Milan
- Since 2022: Graduate committee member for 4 Master level sessions (PoliMi), 2 Bachelor level sessions (Polimi), and 5 Master level sessions (BCG interuniversity Master)

## Participation/responsibility in funded projects.

- Sep 2025 - Mar 2029: **Proposal co-writing and Task Leader.** EU4Health Programme (EU4H) Action Grant “UNICA: Unified Network for International Cancer Advancement”, grant agr. n.101218966, Funding PoliMi Unit: 602K Euros (total 4M Euros).
- Dec 2023 - Nov 2025: **Principal Investigator.** MUR PRIN PNRR 2022 project “SENSIBLE: Small-data Early warning System for viral pathogens In public Health”, project n. P2022CNN2J, Research Units: Politecnico di Milano and Univ. degli Studi di Milano; Funding PoliMi Unit: 186K Euros (Total: 240K Euros)
- Dec 2023 - May 2027: **Proposal co-writing and Task Leader.** Horizon Europe Framework Programme (HORIZON) project “BETTER: Better real-world health-data distributed analytics Research platform”, grant agreement n. 101136262, Funding PoliMi Unit: 639K Euros (total 9.8M Euros).
- Sep 2023 - Aug 2024: **Principal Investigator.** Next Generation Internet (NGI) Search, 2nd open call, project “TETHYS: Topics Evolution That You See”, agreement n. SEARCH OC2\_18, Funding: 150K Euros.
- Jan 2023 - Dec 2025: **Task Leader.** NextGenerationEU program project “FAIR: Future Artificial Intelligence Research”, within the PNRR-PE-AI scheme, Funding: 4.9M Euros.
- Jul 2020 - Dec 2020: **Task Leader.** EIT Digital “DATA against COVID-19” program, Innovation Activity ‘Virusurf’ n. 20663, Funding: 172K Euros.
- Nov 2016 - Aug 2021: **Work Package Leader.** Data-Driven Genomic Computing (GeCo), ERC Advanced Investigator Grant n. 693174, Funding: 2.5M Euros.

## Technology transfer.

- **Co-founder of LegisRatio S.r.l.**, Innovative start-up in the LegalTech domain, accredited **spin-off of Politecnico di Milano**. We support companies in navigating the legislative system, combining lightweight AI agents with structured, rich, and smart data modeled as easy explorable graphs (funded February 2025)
- Participation in the **COVID-19 Host Genetics Initiative** for designing the COVID-19 clinical data dictionary for COVID-19 patients’ phenotype (<https://doi.org/10.1038/s41586-021-03767-x> on Nature). (2020)
- **Co-chair** of the Research Data Alliance “**FAIRification of Genomic Annotations Working Group**”, for the standardization of Genomic Annotations and FAIRification data processing pipelines. (2024-2026)
- **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.
- **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/viruclust>): data analysis tool to perform temporal and spatial comparison of mutational patterns of SARS-CoV-2 sequences; VariantHunter ([http://gmql.eu/variant\\_hunter](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; OpenRecombinHunt (<http://gmql.eu/openrecombinhunt>): dashboard for automatic detection of recombination from publicly available viral sequences.

- **Contribution to open-source tools** for the exploration of 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);* TETYS (<http://gmql.eu/tetys>): Explorer of Sustainability Development Goals-related literature using a topic modeling pipeline and user-centered interfaces. *Mature product developed within NGI Search Program;* CLD-Explorer (<http://gmql.eu/cld-explorer>): tool for Causal Loop Diagrams exploration and analytics; LLMMap (<https://geco.deib.polimi.it/LLMap-demo/>): proof-of-concept of conceptual-modeling-driven exploration of LLM information; MLegis (<https://mlegis.eu/>): information system engineered to model, manage, and analyze legislative corpora through a property graph.

### Most relevant publications.

1. Canakoglu, A., **Bernasconi, A.**, Colombo, A., Masseroli, M., & Ceri, S. (2019). GenoSurf: metadata driven semantic search system for integrated genomic datasets. *Database*, 2019.
2. **Bernasconi, A.**, Canakoglu, A., Masseroli, M., & Ceri, S. (2021). The road towards data integration in human genomics: players, steps and interactions. *Briefings in Bioinformatics*, 22(1), 30-44.
3. **Bernasconi, A.**, Canakoglu, A., Masseroli, M., Pinoli, P., & Ceri, S. (2021). A review on viral data sources and search systems for perspective mitigation of COVID-19. *Briefings in Bioinformatics*, 22(2), 664-675.
4. Canakoglu, A., Pinoli, P., **Bernasconi, A.**, Alfonsi, T., Melidis, D. P., & Ceri, S. (2021). ViruSurf: an integrated database to investigate viral sequences. *Nucleic Acids Research*, 49(D1), D817-D824.
5. **Bernasconi, A.**, Gulino, A., Alfonsi, T., Canakoglu, A., Pinoli, P., Sandionigi, A., & Ceri, S. (2021) VirusViz: Comparative analysis and effective visualization of viral nucleotide and amino acid variants. *Nucleic Acids Research*, 49(15), e90.
6. **Bernasconi, A.**, Mari, L., Casagrandi, R., & Ceri, S. (2021). Data-driven analysis of amino acid change dynamics timely reveals SARS-CoV-2 variant emergence. *Scientific Reports*, 11(1), 21068.
7. Alfonsi, T., Al Khalaf, R., Ceri, S., & **Bernasconi, A.** (2022) CoV2K model, a comprehensive representation of SARS-CoV-2 knowledge and data interplay. *Scientific Data*, 9, 260.
8. Alfonsi, T., **Bernasconi, A.**, Chiara, M., Ceri, S. (2024). Data-driven recombination detection in viral genomes. *Nature Communications*, 15, 3313.
9. Invernici, F., **Bernasconi, A.**, Ceri, S. (2024) Exploring the evolution of research topics during the COVID-19 pandemic. *Expert Systems with Applications* 252, 124028.
10. Colombo, A., **Bernasconi, A.**, Ceri, S. (2025) An LLM-assisted End-to-End Pipeline to Build a High-Quality Knowledge Graph of the Italian Legislation. *Information Processing & Management*, 62(4), 104082.
11. Invernici, F., Curati, F., Jakimov, J., Samavi, A., **Bernasconi, A.** (2025) Capturing research literature attitude towards Sustainable Development Goals: an LLM-based topic modeling approach. *Journal of Big Data*, 12, 139.
12. Cambria, F., Invernici, F., **Bernasconi, A.**, Ceri, S. (2025) MINE GRAPH RULE: A New GQL Operator for Mining Association Rules in Property Graph Databases. *The VLDB Journal*, 34:54.
13. Alfonsi, T., Chiara, M., **Bernasconi, A.** (2025) A codon usage-based approach for the stratification of Influenza A. *Computational and Structural Biotechnology Journal*, 27, 2757-2771.
14. Alfonsi, T., **Bernasconi, A.**, Chiara, M., Ceri, S. (2025) Lightweight multi-scale early spillover warnings for influenza A viruses. *Science Advances*, 11, eadz7312.
15. Cambria, F., Invernici, F., **Bernasconi, A.**, Ceri, S. (2026) Data Analytics Algorithms in Property Graph Databases: A Survey. *Computer Science Review*, 62, 100981.