

# LUCÍA PRIETO SANTAMARÍA

Universidad Politécnica de Madrid  
lucia.prieto.santamaria@upm.es  
Madrid, Spain  
Last updated: January 2023



## EDUCATION

---

<b>PhD candidate</b>	Software, Systems and Computing (Industrial doctor distinction) Ezeris Networks Global Services S.L. Universidad Politécnica de Madrid	Ongoing (started in October 2019)
<b>MSc</b>	Computational Biology Data Science itinerary Universidad Politécnica de Madrid	July 2019
<b>BSc</b>	Biotechnology Computational biotechnology itinerary Universidad Politécnica de Madrid	July 2018

## MAIN RESEARCH INTERESTS

---

- Biomedical informatics and computational biology
- Artificial intelligence and machine learning applied to biomedical domains
- Human disease networks and drug repurposing
- Knowledge representation and semantic web
- Health-related social media analysis

## RESEARCH EXPERIENCE

---

### Ezeris Networks Global Services S.L

February 2020 – Present

PhD thesis (Supervisors: Alejandro Rodríguez González and Yuliana Pérez Gallardo)

- Creation, integration and analysis of disease networks towards a better understanding in the context of drug repurposing

### MEDAL (<https://medal.ctb.upm.es/>)

September 2017 – Present

Medical Data Analytics Laboratory

Centre for Biomedical Technology (CTB)  
Polytechnic University of Madrid (UPM)

MSc and BSc theses and internships (under the supervision of Alejandro Rodríguez González)

- Master thesis and internship. “Biological and phenotypical data analysis to generate new nosological models of diseases”, July 2019.
- Bachelor thesis and internship. “Extraction, similarity computing and analysis of biological features towards the creation of complex human disease networks”, July 2018.

Active involvement in research **projects** and main responsibilities and contributions:

- **DISNET** (“Drug repositioning and disease understanding through complex networks creation and analysis”, <http://disnet.ctb.upm.es/>, RTI2018-094576-A-I00 from the Spanish Ministry of Science, Innovation and Universities).
  - Human disease complex networks to gain a better disease understanding and drug repurposing. Extraction and integration of biomedical data structured in 3 levels: the phenotypical, the biological and pharmacological.
  - Intensive analysis of the information and generation of drug repurposing hypotheses. Employment of machine learning techniques both supervised and unsupervised.
  - Management tasks. Scientific writing and communication, as well as dissemination to the general public.
- **CUREX** (“Secure and private health data exchange”, <https://curex-project.eu/>, European Commission, Programme Horizon 2020, G.A. n 826404).
  - Creation of a basic NLP module for texts related to computer network topologies.
  - Development of an ontology for the semantic representation of data regarding asset discovery in distributed networks for healthcare environments.
- **P4-LUCAT** (“Personalized medicine for lung cancer treatment”, <https://p4-lucat.eu/>, ERA PerMed of Joint Transnational Call 2019 No. 163).
  - Design and implementation of a relational database for the storage of lung cancer patient information.
  - Drafting of deliverables and assistance in the project management tasks in its initial stages.
- **MAVIS** (“Study on sentiments expressed on social media during the period 2015-2018 in posts related to vaccines”, funded by MSD, Spain, VEAP ID: 7789).
  - Analysis, visualization and understanding of data collected from Twitter posts and their polarity.
  - Scientific writing and dissemination in the form of journal papers.
- **3DR-GNN** (“Data-driven drug repositioning applying graph neural networks”, PID2021-122659OB-I00 from the Spanish Ministry of Science, Innovation and Universities).
  - Writing of the project proposal. Project planification and management.
- **GRENADA** (“Drug repurposing hypotheses through a data-driven approach”, PDC2022-133173-I00 from the Spanish Ministry of Science, Innovation and Universities).
  - Writing of the project proposal. Project planification and management.

**Barabasi’s Lab** (<https://www.barabasilab.com>)

January 2023 – June 2023

Visiting PhD student (5-month research stay, funded by UPM)

Center for Complex Networks Research (CCNR)

Network Science Institute (NetSI)

Northeastern University (NEU)

## PUBLICATIONS

---

### *JCR-indexed journals (13)*

1. A.J. Diaz-Honrubia, A. Blázquez Herranz, **L. Prieto Santamaría**, E. Menasalvas Ruiz, A. Rodríguez-González, G. Gonzalez-Granadillo, R. Diaz, E. Panaousis, and C. Xenakis, «A Trusted Platform Module-based, Pre-emptive and Dynamic Asset Discovery Tool», *J. Inf. Secur. Appl.*, vol. 71, p. 103350, Dec. 2022, doi: 10.1016/j.jisa.2022.103350.

2. B. Otero-Carrasco, **L. Prieto Santamaría**, E. Ugarte Carro, J. P. Caraça-Valente Hernández, and A. Rodríguez-González, «Repositioning Drugs for Rare Diseases Based on Biological Features and Computational Approaches», *Healthcare*, vol. 10, no. 9, Art. no. 9, Sep. 2022, doi: 10.3390/healthcare10091784.
3. **L. Prieto Santamaría\***, M. Díaz Uzquiano\*, E. Ugarte Carro, N. Ortiz-Roldán, Y. Pérez Gallardo, and A. Rodríguez-González, «Integrating heterogeneous data to facilitate COVID-19 drug repurposing», *Drug Discov. Today*, vol. 27, no. 2, pp. 558–566, Feb. 2022, doi: 10.1016/j.drudis.2021.10.002.
4. **L. Prieto Santamaría\***, E. Ugarte Carro\*, M. Díaz Uzquiano, E. Menasalvas Ruiz, Y. Pérez Gallardo, and A. Rodríguez-González, «A data-driven methodology towards evaluating the potential of drug repurposing hypotheses», *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 4559–4573, 2021, doi: 10.1016/j.csbj.2021.08.003.
5. **L. Prieto Santamaría**, E. P. García del Valle, M. Zanin, G. S. Hernández Chan, Y. Pérez Gallardo, and A. Rodríguez-González, «Classifying diseases by using biological features to identify potential nosological models», *Sci. Rep.*, vol. 11, no. 1, p. 21096, Oct. 2021, doi: 10.1038/s41598-021-00554-6.
6. **L. Prieto Santamaría**, J. M. Tuñas, D. Fernández Peces-Barba, A. Jaramillo, M. Cotarelo, E. Menasalvas Ruiz, A. J. Conejo Fernández, A. Arce, A. Gil de Miguel, and A. Rodríguez-González, «Influenza and Measles-MMR: two case study of the trend and impact of vaccine-related Twitter posts in Spanish during 2015-2018», *Hum. Vaccines Immunother.*, pp. 1-15, Mar. 2021, doi: 10.1080/21645515.2021.1877597.
7. **L. Prieto Santamaría**, D. Fernández Lobón, A. J. Díaz-Honrubia, E. Menasalvas Ruiz, S. Nifakos, and A. Rodríguez-González, «Towards the representation of network assets in health care environments using ontologies», *Methods Inf. Med.*, vol. 60, no. S 02, pp. e89–e102, Dec. 2021, doi: 10.1055/s-0041-1735621.
8. E. P. García del Valle, G. Lagunes García, **L. Prieto Santamaría**, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez-González, «DisMaNET: A network-based tool to cross map disease vocabularies», *Comput. Methods Programs Biomed.*, p. 106233, Jun. 2021, doi: 10.1016/j.cmpb.2021.106233.
9. E. P. García del Valle, G. Lagunes García, **L. Prieto Santamaría**, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez-González, «Leveraging network analysis to evaluate biomedical named entity recognition tools», *Sci. Rep.*, vol. 11, no. 1, Jun. 2021, doi: 10.1038/s41598-021-93018-w.
10. A. Rodríguez-González, J. M. Tuñas, **L. Prieto Santamaría**, D. Fernández Peces-Barba, E. Menasalvas Ruiz, A. Jaramillo, M. Cotarelo, A. J. Conejo Fernández, A. Arce, and A. Gil de Miguel, «Identifying Polarity in Tweets from an Imbalanced Dataset about Diseases and Vaccines Using a Meta-Model Based on Machine Learning Techniques», *Appl. Sci.*, vol. 10, no. 24, p. 9019, Dec. 2020, doi: 10.3390/app10249019.
11. G. Lagunes García, A. Rodríguez-González, **L. Prieto Santamaría**, E. P. García del Valle, M. Zanin, and E. Menasalvas Ruiz, «How does Wikipedia disease information evolve over time? An analysis of disease-based articles changes», *Inf. Process. Manag.*, vol. 57, no. 3, p. 102225, May 2020, doi: 10.1016/j.ipm.2020.102225.
12. G. Lagunes García, A. Rodríguez González, **L. Prieto Santamaría**, E. P. García del Valle, M. Zanin, and E. Menasalvas Ruiz, «DISNET: a framework for extracting phenotypic disease information from public sources», *PeerJ*, vol. 8, p. e8580, Feb. 2020, doi: 10.7717/peerj.8580.
13. E. P. García del Valle, G. Lagunes García, **L. Prieto Santamaría**, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez-González, «Disease networks and their contribution to disease understanding: A review of their evolution, techniques and data sources», *J. Biomed. Inform.*, vol. 94, p. 103206, Jun. 2019, doi: 10.1016/j.jbi.2019.103206.

**Conference proceedings (9)**

1. A. Álvarez Pérez, A. Iglesias-Molina, **L. Prieto Santamaría**, M. Poveda-Villalón, C. Badenes-Olmedo, and A. Rodríguez-González, ‘EBOCA: Evidences for BiOmedical Concepts Association Ontology’, in Knowledge Engineering and Knowledge Management, Cham, 2022, pp. 152–166. doi: 10.1007/978-3-031-17105-5\_11.
2. A. Ayuso Muñoz, E. Ugarte Carro, **L. Prieto Santamaría**, B. Otero-Carrasco, E. Menasalvas Ruiz, Y. Pérez Gallardo, and A. Rodríguez-González, “REDIRECTION: Generating drug repurposing hypotheses using link prediction with DISNET data”, in 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), Jul. 2022, pp. 7-12, doi: 10.1109/CBMS55023.2022.00009.
3. B. Otero-Carrasco, A. Pérez Pérez, E. Mensalvas Ruiz, J. P. Caraça-Valente Hernández, **L. Prieto Santamaría**, and A. Rodríguez-González, “Drug repositioning with gender perspective focused on Adverse Drug Reactions” in 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), Jul. 2022, pp. 435-440, doi: 10.1109/CBMS55023.2022.00084.
4. B. Otero-Carrasco, **L. Prieto Santamaría**, E. Ugarte Carro, J.P. Caraça-Valente Hernández, A. Rodríguez González, ‘A computational drug repositioning method for rare diseases’, in: Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence. IWINAC 2022. Lecture Notes in Computer Science, vol 13259, pp. 551-561. Springer, Cham. doi: 10.1007/978-3-031-06527-9\_55.
5. E. P. García Del Valle, **L. Prieto Santamaría**, G. Lagunes García, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez-Gonzalez, ‘A meta-path-based prediction method for disease comorbidities’, in 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2021, pp. 219–224. doi: 10.1109/CBMS52027.2021.00022.
6. **L. Prieto Santamaría**, E. P. García Del Valle, G. Lagunes Garcia, M. Zanin, A. Rodríguez-Gonzalez, E. Menasalvas Ruiz, Y. Pérez Gallardo, G. S. Hernández Chan. ‘Analysis of new nosological models from disease similarities using clustering’, in 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), Jul. 2020, pp. 183–188. doi: 10.1109/CBMS49503.2020.00042.
7. E. P. García del Valle, G. Lagunes García, E. Menasalvas Ruiz, **L. Prieto Santamaría**, M. Zanin, and A. Rodríguez-González, ‘Completing missing MeSH code mappings in UMLS through alternative expert-curated sources’, in 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2019, pp. 174–179. doi: 10.1109/CBMS.2019.00044.
8. G. Lagunes García, **L. Prieto Santamaría**, E. P. García del Valle, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez González, ‘Wikipedia disease articles: an analysis of their content and evolution’, in 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2019, pp. 664–671. doi: 10.1109/CBMS.2019.00136.
9. E. P. Garcia del Valle, G. Lagunes García, **L. Prieto Santamaría**, M. Zanin, E. Menasalvas Ruiz, and A. Rodríguez González, ‘Evaluating wikipedia as a source of information for disease understanding’, in 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), Jun. 2018, pp. 399–404. doi: 10.1109/CBMS.2018.00076.

**CONGRESSES AND CONFERENCES**

**EKA2022** (23rd International Conference on Knowledge Engineering and Knowledge Management), Bolzano (Italy), September 2022.

**CBMS2022** (2022 IEEE 35th International Symposium on Computer-Based Medical Systems), Online, June 2022.

**CamAIDD** (1st Cambridge AI in Drug Discovery Conference), Online, February 2022.

**CDW21** (Connected Data World 2021), Online, December 2021.

**CAEPIA20/21** (XIX Conference of the Spanish Association for Artificial Intelligence), Málaga (Spain), September 2021.

**CBMS2021** (2021 IEEE 34th International Symposium on Computer-Based Medical Systems), Online, June 2021.

**CBMS2020** (2020 IEEE 33rd International Symposium on Computer-Based Medical Systems), Online, July 2020.

**Drug repurposing meeting**, Biochemical Society Event, Birmingham (UK), November 2019.

**CBMS2019** (2019 IEEE 32nd International Symposium on Computer-Based Medical Systems), Córdoba (Spain), June 2019.

## TALKS

---

**Talk**, “¿How can we study and combat diseases with computers?”, Week of the Science and Innovation in Madrid 2022, November 2022.

**Seminar**, “Disease understanding: Dealing with complex and unstructured big data in biomedical domain”, MSc in ICT Innovation: Data Science (EIT Digital – UPM), February 2022.

**Talk**, “¿How can we study and combat diseases with computers?”, Week of the Science and Innovation in Madrid 2021, November 2021.

**Doctoral symposium**, “My Thesis in a Nutshell: Disease networks towards drug repurposing”, UPM, June 2021. Finalist award.

**Seminar**, “From data to treatments: Can we use public data to have a better understanding of diseases and reuse treatments?”, MSc in Computational Biology (UPM), Online, May 2021.

**Workshop**, “Healthcare security ontologies for semantic representation of data / CUREX Asset Discovery”, MSc in Data and Web Science (Aristotle University of Thessaloniki, Greece), Online, March 2021.

**Seminar**, “Disease understanding: Dealing with complex and unstructured big data in biomedical domain”, MSc in Data Science (UPM), Online, February 2021.

**Talk**, “¿How can we study and combat diseases with computers?”, Week of the Science and Innovation in Madrid 2020, Online, November 2020. (<https://short.upm.es/vaxv3>)

**Oral Communication**, “DISNET: Drug repurposing and disease understanding through complex networks creation”, Drug repurposing meeting, Biochemical Society Event, Birmingham (UK), November 2019.

**Seminar**, “Disease understanding: Dealing with complex and unstructured big data in biomedical domain”, MSc in Data Science (UPM), February 2019.

**PEER REVIEWING**

---

***Journals***

- IJMEDI - International Journal of Medical Informatics (2023)
- Frontiers in Digital Health (2022)
- BMC Medical Informatics (2022)
- KHVI - Human Vaccines and Immunotherapies (2022, 2021)
- JMIR - Journal of Medical Internet Research (2022, 2021)
- Frontiers in Medical Technology (2021)
- JBI - Journal for Biomedical Informatics (2020)
- IPM - Journal of Information Processing and Management (2019)

***Conference member committee/reviewer***

- EURECA-PRO Conference - 2022
- 2022 IEEE 35<sup>th</sup> International Symposium on Computer-Based Medical Systems (CBMS2022)
- 2021 IEEE 34<sup>th</sup> International Symposium on Computer-Based Medical Systems (CBMS2021)
- 2020 IEEE 33<sup>rd</sup> International Symposium on Computer-Based Medical Systems (CBMS2020)
- Knowledge Discovery in Databases 2020 (KDD2020)
- SIAM International Conference on Data Mining (SDM2020)
- 2019 12<sup>th</sup> International Conference on Health Informatics (HEALTHINF2019)
- 2019 IEEE 32<sup>nd</sup> International Symposium on Computer-Based Medical Systems (CBMS2019)

**FELLOWSHIPS**

---

Competitive call for mobility research fellowships in North America funded by Universidad Politécnica de Madrid (January – June 2023).

**IND2019/TIC-17159** “Support for research and innovation programme (Industrial doctorates)”, Community of Madrid, Spain (2020 – 2023).

“Academic excellence scholarship”, Community of Madrid, Spain (2014 – 2015).

**TEACHING EXPERIENCE**

---

***Universidad Politécnica de Madrid***

Course 2022/23

**Teaching assistant**, Computer Languages and Systems and Software Engineering Department (“Escuela Técnica Superior de Ingenieros Informáticos”)

Subjects:

- Data Processes (EIT Digital MSc in Digital Innovation) (4h)
- Data Processing Engineering (MSc in Computer Engineering) (4h)
- Complex Data in Health (EIT Health MSc Health and Medical Data Analytics) (12h)
- Processes in Data Science (MSc in Data Science) (4h)
- Databases (BSc in Biotechnology) (10h)

***Universidad Politécnica de Madrid***

Course 2021/22

**Teaching assistant**, Computer Languages and Systems and Software Engineering Department (“Escuela Técnica Superior de Ingenieros Informáticos”)

Subjects:

- Data Processes (EIT Digital MSc in Digital Innovation) (1.32h)
- Data Processing Engineering (MSc in Computer Engineering) (1.32h)
- Complex Data in Health (EIT Health MSc Health and Medical Data Analytics) (8h)
- Data Analytics (BSc in Mathematics and Computer Science) (8h)
- Processes in Data Science (MSc in Data Science) (1.32h)
- Databases I (BSc in Data Science and Artificial Intelligence) (8h)
- Databases (BSc in Mathematics and Computer Science) (8h)
- Databases (BSc in Biotechnology) (9h)

**Co-supervision of Master Theses**

- Andrea Álvarez Pérez, “On giving disease networks data a semantic context: generating an ontology reusing existing ones”, MSc in Computational Biology (UPM), July 2022.

**Co-supervision of Bachelor Theses**

- Natalia García Sánchez, “Biological sequence analysis in the context of drug repurposing and human disease complex networks”, BSc in Biotechnology (UPM), July 2022.
- Adrián Ayuso Muñoz, “Análisis y desarrollo de modelos de predicción de ejes en redes mediante técnicas de Deep Learning”, BSc in Computer Engineering (UPM), June 2022. [With honours]

***Universidad Politécnica de Madrid***

Course 2020/21

**Teaching assistant**, Computer Languages and Systems and Software Engineering Department (“Escuela Técnica Superior de Ingenieros Informáticos”)

Subjects:

- Complex Data in Health (EIT Digital MSc in ICT Innovation: Data Science – Health Itinerary) (7.5h)
- Data Analytics (BSc in Mathematics and Computer Science) (9h)
- Databases (BSc in Biotechnology) (7h)

**Co-supervision of Master Theses**

- Pablo Soto García, “Study and integration of new biological features data from public sources in the context of human complex disease networks”, MSc in Computational Biology (UPM), July 2021.

**Co-supervision of Bachelor Theses**

- Ignacio Montero Callejas, “Desarrollo de procesos para el análisis y visualización de datos fenotípicos y biológicos en el contexto de las redes de enfermedades humanas”, BSc in Biomedical Engineering (UPM), July 2021.

- Esther Ugarte Carro, “Drug repurposing approaches and validation through DISNET data”, BSc in Biotechnology (UPM), July 2021. [With honours]

**Universidad Politécnica de Madrid**

Course 2019/20

**Co-supervision of Master Theses**

- Sara Jaramillo Cárdenas, “Analysis of public data sources and drugs information extraction towards the elaboration of human complex disease networks”, MSc in Computational Biology (UPM), September 2020.
- Rizwan Saleem, “Towards drug repositioning: analysis of relevant drug features and computational methods”, MSc in Computational Biology (UPM), September 2020.
- Youness Elkhailil, “Analysis of phenotypic data on diseases and their relationships for drug repositioning”, MSc in Computational Biology (UPM), July 2020.

**TAUGHT COURSES**

---

<b>Unsupervised machine learning (5h)</b> Banco Santander Internal Educational Programme, Madrid, Spain	October 2020
<b>Data mining, solutions and horizontal applications (4h)</b> University of Military Forces, Ecuador	October 2020
<b>Unsupervised machine learning (5h)</b> Banco Santander Internal Educational Programme, Madrid, Spain	February 2020
<b>Introduction to databases for Data Science (5h)</b> Banco Santander Internal Educational Programme, Madrid, Spain	January 2020
<b>Unsupervised machine learning (5h)</b> Banco Santander Internal Educational Programme, Madrid, Spain	December 2019



## TECHNICAL SKILLS

---

### Information extraction and storage

- Programmatic access to data sources by means of APIs and web services.
- Strong knowledge of relational databases, with emphasis on MySQL database manager system.
- Strong knowledge of Python scripting and basic knowledge of R.
- Familiar with NoSQL databases, in particular with graph databases. Notions on Neo4J.
- Familiar with semantic technologies, biomedical ontologies, RDF language and SPARQL queries.
- Familiar with NLP pipelines and tools.

### Data processing, manipulation and analysis

- Python (pandas, numpy, scipy, scikit-learn) and R.

### Data visualization

- Python (matplotlib, seaborn, plotly) and R (ggplot).
- Graph visualization (gephi, cytoscape).

### Machine learning

- Strong knowledge in unsupervised and supervised learning.
- Python (scikit-learn, PyTorch, DGL, pyclustering).
- Basic knowledge of Weka and Knime.

### Bioinformatic tools

- Sequence alignment (BLAST, ClustalW) and phylogenetic studies.
- Basic knowledge of genomic data analysis.

### Others

- Anaconda, JupyterLab
- Unix/Linux systems
- Version control with Git
- Notions of Perl, Ruby and Matlab
- Docker, VMware, Oracle VMVirtualBox
- Microsoft Word, LaTeX

## LANGUAGES

---

- Spanish: native
- English: C1 (Cambridge Assessment English – Certificate in Advanced English, March 2022)
- French: Basic

## COURSES

---

- Data visualization courses (Domestika), online, ongoing.
  - “Introduction to data visualization”
  - “Data visualization and information design: creating a visual model”
  - “Data visualization for editorial projects”
- KNIME Analytics Platform for Data Scientist: Basics, online, October 2021.
- SNOMED-CT E-Learning, online, March 2021.

**PROFESSIONAL AFFILIATIONS**

---

Since September 2017      ETS Ingenieros Informáticos, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, España

Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, España

Since February 2020                      Ezeris Networks Global Services S.L., 28028 Madrid, España

**OTHERS**

---

- Title of Rhythmic Gymnastics Coach (1<sup>st</sup> Level), RFEG (Royal Spanish Gymnastics Federation), July 2014, coaching in several schools and clubs (Madrid, Spain), 2014 – 2019.
- Driving license (Madrid, Spain), May 2015.
- Private tutor for high school students in subjects such as mathematics, physics and chemistry (Madrid, Spain), 2012 – 2014.
- YMCA Camp Weona counselor volunteer (Buffalo, USA), July 2012 and July – August 2013.