

DR. CARLOS VEGA

Research Engineer in Digital Health at the Luxembourg Institute of Health

vega.moreno.carlos@gmail.com (+34) 653 88-11-07 (+352) 621 65-48-20
http://research.cvega.es/ github.com/carlosvega Luxembourg

EXPERIENCE

Current position(s)

Research Engineer in Digital Health

Data Integration Center, Luxembourg Institute of Health

Oct. 2023 – Ongoing Strassen, Luxembourg

- API development in healthcare management software.
- Data wrangling solutions for clinical studies.
- Building data architecture for biomedical research.

Fellow

EU MIT Catalyst Europe

Sept. 2023 – Ongoing Catalyst Europe

- Co-leader of the project "Pacemaker interrogation made simple"
- Co-leader of the project "Period poverty related school absenteeism in Ghana"

Previous positions

Postdoctoral Researcher

Bioinformatics Core, Luxembourg Centre for Systems Biomedicine

Sep. 2018 – Sep. 2023 Belval, Luxembourg

- Web-based translational medicine knowledge management and visualization.
- R+D of a collaborative bio-curation web-app for browsing text-mined results.
- R+D of multi-appointment scheduling solutions for clinics and hospitals.
- R+D in neurodegeneration clinical studies: NCER-PD, PDP, MCI-BIOME, RBD, HeBA.
- COVID-19 · Literature curation, Text-Mining and Semantic Web Technologies.
- COVID-19 · Evolution Dashboard (Lux. COVID-19 Task Force WP0).
- COVID-19 · CON-VINCE study (Lux. COVID-19 TF WP01).

R+D Computer Science Engineer

Naudit HPCN 2015 – 2018 Madrid, Spain

- R+D of HPC tools for HTTP network traffic analysis and visualization.
- R+D of dashboards and anomaly detection tools for computer network traffic.
- EU H2020 project: Disaggregated Recursive Datacentre-in-a-Box (2016 - 2018).
- Spanish MINECO & ERDF: Traffic analysis for operational intelligence (2016 - 2019).

Technical and research staff

Universidad Autónoma de Madrid 2014 – 2015 Madrid, Spain

- 2015 · R+D for the 10G Trace Tester of the European project Fed4Fire.
- 2014 · R+D of a traffic repository for the European project EINS.
- 2012 - 2016 · Federation for FIRE. EU Project FP7-ICT-318389.
- 2012 - 2016 · EU FP7 project: Network of Excellence in Internet Science.

Scholarship assistant

Universidad Autónoma de Madrid 2012 – 2013 Madrid, Spain

- High Performance Computing and Networking group, Higher Polytechnic School.

Teaching Experience

University of Luxembourg

Master of Data Science

- 2021–Ongoing:
- Applied Philosophy of Science and Data Ethics (course director, conceptualisation, design and teaching)

Universidad Autónoma de Madrid

Bachelor's Degree in Telecommunication Technology and Service Engineering

- 2014-2015:
- Multimedia networking
- 2015-2016:
- Multimedia networking (joint course with CS Engineering)

Master's Degree in Telecommunication Engineering

- 2016–2017:
- Network planning
 - Network management

AIM STATEMENT

Computer Science and Engineering Ph.D.

My interests include data wrangling, engineering, science, and ethics. I research and develop solutions for data analysis, clinical study management, and AI bias in healthcare.

EDUCATION

Ph.D. Computer Science and Telecommunication

Universidad Autónoma de Madrid

2014 – 2017

Graduated *cum laude* with industrial mention

Exploring the collection, analysis and visualisation processes of computer networks traffic

My Ph.D. thesis tackles the ETL stages of high volume network traffic analysis, contributing on the data collection, data visualization and the intermediate stages of data processing and analysis.

M.Sc. in Computer Science Engineering

Universidad Autónoma de Madrid

2013 – 2014

My thesis (graded 10/10) addressed HTTP traffic dissection at high speeds (10 Gbps traffic rates). It has been used in different projects satisfying the needs of real enterprise problems in South America and Spain.

B.Sc. in Computer Science Engineering

Universidad Autónoma de Madrid

2009 – 2013

My thesis (10/10) evaluated web traffic performance, featuring a network traffic processing tool and a web interface for visualizing the analyzed data.

GRANTS

University of Gyor & MIT Catalyst Europe

- 2025 · As Co-leader · 50K€ · Pacemaker interrogation made simple
- 2025 · As Co-leader · 50K€ · Period Poverty Related School Absenteeism in Ghana

Luxembourg National Research Fund

- 2021 · As Grant co-author · 42K€ · PI: Prof. Reinhard Schneider · Covlit (14729738)

STRENGTHS AND SKILLS

</> Python JS C# bash awk MATLAB R C
PostgreSQL MySQL Elasticsearch solr Neo4j
Scikit-learn ONNX FastAI Pandas NumPy
Matplotlib Seaborn Plotly ggplot
Flask Django Vue.js React
Docker CI/CD Jenkins ArgoCD OpenShift
HPC Cloudera Hadoop
GDPR Data Ethics eCRFs REDCap
Problem Solving Teamwork Creativity

PUBLICATIONS

Journal Articles as first author

- Carlos Vega, Reinhard Schneider, and Venkata Satagopam (2023). "Analysis: Flawed Datasets of Monkeypox skin images". In: *Journal of Medical Systems*.
- Carlos Vega, Piotr Gawron, Jacek Lebioda, Valentin Grouès, Matyjaszczyk Piotr, Claire Pauly, Ewa Smula, Rejko Krüger, Reinhard Schneider, and Venkata Satagopam (2022). "SMASCH: multi-appointment scheduling system for longitudinal clinical research studies". In: *Jamia Open*.
- Carlos Vega (2021). "From Hume to Wuhan: an epistemological journey on the problem of induction in COVID-19 machine learning models and its impact upon medical research". In: *IEEE Access*.
- Carlos Vega, Valentin Grouès, Marek Ostaszewski, Reinhard Schneider, and Venkata Satagopam (2020). "BioKC: a collaborative platform for systems biology model curation and annotation". In: *bioRxiv*.
- Carlos Vega, Eduardo Miravalls, Guillermo Julián-Moreno, Jorge E López de Vergara, Eduardo Magaña, and Javier Aracil (2018). "On the design and performance evaluation of automatic traffic report generation systems with huge data volumes". In: *Int. Journal of Network Management*.
- Carlos Vega, Paula Roquero, and Javier Aracil (2017). "Multi-Gbps HTTP traffic analysis in commodity hardware based on local knowledge of TCP streams". In: *Computer Networks* 113, pp. 258–268.
- Carlos Vega, Paula Roquero, Rafael Leira, Ivan Gonzalez, and Javier Aracil (2017). "Loginson: a transform and load system for very large-scale log analysis in large IT infrastructures". In: *The Journal of Supercomputing* 73.9.

Conference Proceedings as first author

- Carlos Vega, Miroslav Kratochvil, Venkata Satagopam, and Reinhard Schneider (2022a). "Translational challenges of biomedical machine learning solutions in clinical and laboratory settings". In: *IWBBIO, June 2022*.
- Carlos Vega, Javier Aracil, and Eduardo Magaña (2018). "KISS methodologies for network management and anomaly detection". In: *26th Int. Conf. on Software, Telecommunications and Computer Networks (SoftCOM), Split – Supetar, Croatia*.
- Carlos Vega, Jose Fernando Zazo, Hugo Meyer, Ferad Zyulkyarov, Sergio Lopez-Buedo, and Javier Aracil (2017). "Diluting the Scalability Boundaries: Exploring the Use of Disaggregated Architectures for High-Level Network Data Analysis". In: *The 19th IEEE Int. Conf. on High Performance Computing and Communications, Bangkok, Thailand*.

Other Journal and Conference Articles

- Schröder, Valerie E, Amna Skrozic, Dorothee Erz, Anne Kaysen, Joëlle V Fritz, Joao M Loureiro, Deborah McIntyre, Laure Pauly, Jennifer Kemp, Sabine K Schmitz, et al. (2024). "Programme Dementia Prevention (pdp): A Nationwide Program for Personalized Prevention in Luxembourg". In: *Journal of Alzheimer's Disease Preprint*, pp. 1–14.
- Beatriz García Santa Cruz, Carlos Vega, and Frank Hertel (2021). "The need of standardised metadata to encode causal relationships: towards safer data-driven machine learning biological solutions". In: *17th Int. Conf. on Computational Intelligence Methods for Bioinformatics and Biostatistics*.
- Ostaszewski, Marek, Anna Niarakis, Alexander Mazein, Inna Kuperstein, Robert Phair, Aurelio Orta-Resendiz, Vidisha Singh, Sara Sadat Aghamiri, Marcio Luis Acencio, Enrico Glaab, et al. (2021). "COVID-19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms". In: *Molecular Systems Biology*.
- Daniel Perdices, Jorge E López de Vergara, Paula Roquero, Carlos Vega, and Javier Aracil (2018). "FlexiTop: a flexible and scalable network monitoring system for Over-The-Top services". In: *Network Protocols and Algorithms* 9.

Posters

- Vega, Carlos, Beatriz Garcia Santa Cruz, Frank Hertel, and Venkata Satagopam (2023). "Leveraging Client-side Computation for Biomedical Machine Learning Inference with Open Neural Network Exchange". In: *Northern Lights Deep Learning Conference Abstracts 2024*.
- Carlos Vega, Miroslav Kratochvil, Venkata Satagopam, and Reinhard Schneider (2022b). "Translational challenges of biomedical machine learning solutions in clinical and laboratory settings". In: *NLDL 2022*.
- Carlos Vega, Valentin Groues, Marek Ostaszewski, Venkata Satagopam, and Reinhard Schneider (2020). "BioKC: a platform for quality controlled curation and annotation of systems biology models". In: *19th ECCB 2020*.
- Welter, Danielle, Carlos Vega, Maria Biryukov, Valentin Groues, Soumyabrata Ghosh, Reinhard Schneider, and Venkata Satagopam (2020). "Supporting findability of COVID-19 research with large-scale text mining of scientific publications". In: *International FAIR Convergence Symposium*.
- Carlos Vega and Sergio López-Buedo (2015). "EINSrepo Distributed Traffic Repository". In: *2nd International Conference on Internet Science, Brussels, Belgium*.

LANGUAGES

Spanish (Mother tongue)
English (C1)
French (B1)



SELECTED REPOSITORIES

- Flawed Monkeypox datasets
- Applied Philosophy of Science and Data Ethics
- HTTPanalyzer
- Elasticsearch Importer

COURSES

University of Oxford

- Philosophy of Science

European infrastructure for translational medicine

- Translational Medicine Explained

University of Edinburgh

- Data Science in Stratified Healthcare and Precision Medicine

University of Michigan

- Applied Plotting, Charting & Data Representation in Python
- Applied Text Mining in Python
- Data Science Ethics

Big Data Specialization

University of California, San Diego

- Introduction to Big Data Analytics
- Hadoop Platform and Application Framework
- Graph Analytics for Big Data

MEMBER OF

- IEEE Senior Member since 2022

OCCASIONAL REVIEWER AT

- The Journal of Supercomputing
- Frontiers in Neurology
- Applied Artificial Intelligence
- Computers and Electrical Engineering
- IEEE Engineering in Medicine and Biology Society

PARTICIPANT AT

- Biohackathon Europe
- MIT Hacking Medicine

Talks

- *Web curation of systems biology models from text-mined literature facts* (2019). In: Current advances in text mining, University of Luxembourg.
- *KISS methodologies for network management and anomaly detection* (2018). In: SOFTCOM 2018. Split - Supetar, Croatia.
- *Multi-Gbps HTTP traffic analysis in commodity hardware based on local knowledge of TCP streams* (2018). In: RIPE 76 Congress in Marseille.
- *Diluting the Scalability Boundaries: Exploring the Use of Disaggregated Architectures for High-Level Network Data Analysis* (2017). In: HPCC 2017. Bangkok, Thailand.

As consortium author

HeBA consortium

The Healthy Brain Ageing (HeBA) consortium includes the University of Luxembourg, the Luxembourg Institute of Health, the Paracelsus-Elena-Klinik, the Göttingen University Medical Centre, the Innsbruck Medical University and the Hospital Clinic de Barcelona. This study is financed by the Michael J Fox Foundation. **Submitted abstracts:**

- Garrido, Alicia, Corinne GC Horlings, Carlos Vega, Sebastian Schade, Soumyabrata Ghosh, Kavita Rege, et al. (2023). *From the Healthy Brain Ageing (HeBA) Cohort: Frequency of key risk factors for Parkinson in a large European Cohort*. In: MDS 2023.
- Horlings, Corinne GC, Alicia Garrido, Carlos Vega, Sebastian Schade, Taina M Marques, Philipp Mahlke, et al. (2023). *The Healthy Brain Ageing (HeBA) PD risk survey: Study design and methods*. In: MDS 2023.
- Mahlke, Philipp, Taina M Marques, Carlos Vega, Corinne Horlings, Alicia Garrido, Sebastian Schade, et al. (2023). *From the Healthy Brain Ageing - HeBA - Cohort: Motivation for participating in the first online phase of HeBA*. In: MDS 2023.
- Sebastian Schade, Carlos Vega, Soumyabrata Ghosh, ..., Venkata Satagopam, Brit Mollenhauer, and the HEBA Consortium (2023). *Identifying prodromal non-motor-symptoms in a population-based recruitment strategy: Kassel Data of Healthy Brain Ageing*. In: MDS 2023.

NCER-PD consortium

The National Centre of Excellence in Research on Parkinson's disease conducts a long-term clinical study with PD patients from Luxembourg and neighbouring countries as well as healthy control subjects.

- Pauly, Claire, Enrico Glaab, Maxime Hansen, Camille Martin-Gallusiaux, Mirko Ledda, Taina M Marques, Paul Wilmes, Rejko Krüger, Nico J Diederich, and NCER-PD Consortium (2022). "Disease Progression, Resilience, and Inflammation Markers During the Coronavirus Disease 2019 Pandemic in Parkinson's Disease". In: *Movement Disorders*.
- Pavelka, L, A Rauschenberger, Z Landoulsi, S Pachchek, P May, E Glaab, and R Krüger (2022). "Age at onset as stratifier in idiopathic Parkinson's disease—effect of ageing and polygenic risk score on clinical phenotypes". In: *npj Parkinson's Disease* 8.1, pp. 1–10.
- Pavelka, Lukas, Armin Rauschenberger, Zied Landoulsi, Sinthuja Pachchek, Taina Marques, Clarissa PC Gomes, Enrico Glaab, Patrick May, and Rejko Krüger (2022). "Body-First Subtype of Parkinson's Disease with Probable REM-Sleep Behavior Disorder Is Associated with Non-Motor Dominant Phenotype". In: *Journal of Parkinson's Disease* Preprint, pp. 1–13.
- Hanff, Anne-Marie, Anja K Leist, Joëlle V Fritz, Claire Pauly, Rejko Krüger, Margareta Halek, NCER-PD Consortium, et al. (2021). "Determinants of Self-Stigma in People with Parkinson's Disease: A Mixed Methods Scoping Review". In: *Journal of Parkinson's Disease*.
- Hanff, Anne-Marie, Claire Pauly, Laure Pauly, Valerie E Schröder, Maxime Hansen, Guilherme Ramos Meyers, et al. (2021). "Unmet needs of people with Parkinson's disease and their caregivers during COVID-19-related confinement: an explorative secondary data analysis". In: *Frontiers in neurology*.
- Pauly, Claire, Fabiana Ribeiro, Valerie E Schröder, Laure Pauly, Rejko Krüger, Anja K Leist, CON-VINCE Consortium, et al. (2021). "The Moderating Role of Resilience in the Personality-Mental Health Relationship During the COVID-19 Pandemic". In: *Frontiers in Psychiatry*.
- Baldini, Federico, Johannes Hertel, Estelle Sandt, Cyrille C Thinnies, Lorieza Neuberger-Castillo, Lukas Pavelka, Fay Betsou, Rejko Krüger, Ines Thiele, and NCER-PD Consortium (2020). "Parkinson's disease-associated alterations of the gut microbiome predict disease-relevant changes in metabolic functions". In: *BMC biology*.
- Simons, Janine A, Michel Vaillant, Geraldine Hipp, Lukas Pavelka, Lara Stute, Claire Pauly, Rejko Krüger, and NCER-PD Consortium (2019). "Multilingual Validation of the First French Version of Munich Dysphagia Test—Parkinson's Disease (MDT-PD) in the Luxembourg Parkinson's Study". In: *Frontiers in neurology*.

CONVINCE consortium

Covid-19 National survey for assessing viral spread by non-affected carriers is a national project aiming to evaluate the dynamics of the spread of the COVID-19 disease within the Luxembourgish population.

- Holuka, Cyrielle, Chantal J Snoeck, Sophie B Mériaux, Markus Ollert, Rejko Krüger, Jonathan D Turner, CON-VINCE Consortium, et al. (2021). "Adverse Life Trajectories Are a Risk Factor for SARS-CoV-2 IgA Seropositivity". In: *Journal of clinical medicine*.
- Ribeiro, Fabiana, Valerie E Schröder, Rejko Krüger, Anja K Leist, CON-VINCE Consortium, et al. (2021). "The evolution and social determinants of mental health during the first wave of the COVID-19 outbreak in Luxembourg". In: *Psychiatry Research*.

Other consortiums

- Bergquist, Timothy, Thomas Schaffter, Yao Yan, Thomas Yu, Justin Prosser, Jifan Gao, et al. (2021). "Evaluation of crowdsourced mortality prediction models as a framework for assessing AI in medicine". In: *medRxiv*. And the Patient Mortality Prediction DREAM Challenge Consortium.