

Stéphane Ghozzi

I am a data scientist at the World Health Organization. I help conceive, set up and run data-science projects, including: domain research; mathematics, statistics and machine learning; data processing, exploration and visualization. I facilitate and deliver reports and concept notes, code notebooks and packages, interactive graphics.

Graduate from the École normale supérieure, physics PhD, professional work in theoretical physics, infectious-disease surveillance and epidemiology. Fluent in English, French and German. He/him.

stephaneghozzi.com

github.com/stephaneghozzi · mastodon.social/@stephaneghozzi · linkedin.com/in/stephaneghozzi · orcid.org/0000-0002-3911-9573 · scholar.google.com/citations?user=uGVLwREAAAAJ



Experience

Data scientist World Health Organization (WHO)

25 May 2023 – present · 2 yrs 1 mo

Berlin, Germany

WHO Hub for Pandemic and Epidemic Intelligence

pandemichub.who.int

International Pathogen Surveillance Network (IPSN)

who.int/initiatives/international-pathogen-surveillance-network

- member of the team leading global efforts towards making pathogen genomic surveillance data infrastructure, analysis, and sharing more accessible, equitable, and sustainable
- stakeholder engagement, meeting preparation, output synthesis, technological advancements and research monitoring, to formulate a work plan for the implementation of WHO's Global Genomic Surveillance Strategy

Epidemic Intelligence from Open Sources (EIOS)

who.int/initiatives/eios

- development and implementation of two machine learning and generative AI software and infrastructure projects
- supervision of multidisciplinary teams of developers and business analysts who designed application architectures and built strategies for data querying, data processing and machine learning pipelines (Python), implementation of Large Language Models (OpenAI), web applications (Streamlit, Flask) to interact with the data analysis
- one project was cloud native, with DevOps pipelines automating the code compilation, component deployment, and connection to the necessary cloud services

Prototype for the Surveillance Portfolio Assistant (SPA):

- design, development, implementation, and deployment of a web application for recommending surveillance approaches based on country context and needs
- R package that allows the ingestion of relevant datasets, implements a scoring algorithm, and includes the code for a Shiny web application to interactively set processing parameters, and see and download the results

AI for Public Health Intelligence strategy:

- conception of position papers, engagement with internal leadership and external partners, organisation of internal seminars

Various project lead activities, such as designing work plans, reporting on progress and budget, engaging with stakeholders.

Data scientist Independent developer and consultant

1 November 2022 – 24 May 2023 · 7 mos

Berlin, Germany

stephaneghozzi.com

Research, statistics, machine learning, data visualization.

**Research
associate**

Helmholtz-Zentrum für Infektionsforschung (HZI)

1 June 2020 – 31 May 2022 · 2 yrs

Brunswick (Braunschweig), Germany

helmholtz-hzi.de/en/research/research-topics/bacterial-and-viral-pathogens/epidemiology/our-research/ · sormas.org

Senior scientist, statistical methods and tools for public health, in the Epidemiology department (Krause group).

Surveillance Outbreak Response Management and Analysis System

(SORMAS): Automated support for decision making by SORMAS users.

Strategies to bring computations to data so as to allow for advanced analytics without compromising data privacy.

Effect of climate change on infectious disease spread: Statistical modelling of the impact of short and long-term environmental and demographic changes on Lyme disease spread in Germany.

Machine learning for multiplex serology.

Management and organization:

- contact person for topics of automated analyses and processes;
- student support;
- writing of grant applications.

**Research
associate**

Robert-Koch Institut (RKI)

15 April 2016 – 31 May 2020 · 4 yrs 2 mos

Berlin, Germany

rki.de/signale-project

Machine learning, informatics, statistics and visualizations, Signale team, Department of Infectious-Disease Epidemiology.

Research and development:

- development of outbreak-detection algorithms and modelling of infection dynamics;
- performance evaluation and parameter optimization of the algorithms;
- natural language processing of online articles to support international infectious-disease surveillance;
- interactive visualizations of the results and other data for public-health professionals.

Management and organization:

- speaker for the team (four data scientists, two web developers, an average of two students);
- coordination of the international Topic Group Outbreaks under the Focus Group AI for Health of ITU and WHO: itu.int/en/ITU-T/focusgroups/ai4h/Pages/tg.aspx
- supervision of master theses;
- organization of workshops and hackathons;
- writing of grant applications.

World Health Organization (WHO)

Seconded from RKI

1 May 2019 – 31 October 2019 · 6 mos

Geneva, Switzerland

who.int/eios · who.int/emergencies/outbreak-toolkit

Machine learning and web-application development for epidemic intelligence and investigation of outbreaks of unknown origins.

In the Detection, Verification and Risk Assessment (DVA) and the Health Operations Monitoring and Data Collection (MDC) units of the Health Emergency Information and Risk Assessment (HIM) department within the WHO Health Emergencies (WHE) program of WHO.

Visual artist

Self employed

1 March 2012 – 14 April 2016 · 4 yrs 1 mo

Berlin, Germany

Portfolio as of 2014 (pdf): stephaneghozzi.com/docs/portfolio-ghozzi.pdf

Drawing, generative animation, photography, video, computer animation.

Selected projects and collaborations:

- 2014: videos exhibited during the *backup* festival, E-Werk, Weimar;
- 2005: twelve illustrations for *Trace.project*, a compilation album of original electronic music;
- 2005: videography for the dance piece *Entre-Deux* by Mirjam Fruttiger, Paris and Rome (including one-week invitation at the Villa Médicis);
- 2004: videography on the documentary *Manchay Tiempo* by Florence Blum and María Pía Medina-Luna (four-weeks filming in Peru);
- 2002: drawings and photographs published in the magazine *R de réel*.

Postdoctoral researcher

Institut für Theoretische Physik (THP), Universität zu Köln

1 March 2010 – 29 February 2012 · 2 yrs

Cologne, Germany

www.thp.uni-koeln.de/~lassig

Statistical and mechanical models of biological evolution, in the Lässig group.

Mathematical model and analysis of bacterial growth and gene expression, interpretation of experimental results.

Signatures of selection in DNA sequences and comparison with population genetics models:

- long-term influenza evolution;
- transcription-binding-site motifs in yeast.

Co-evolution signatures in protein sequences.

**Teaching
assistant**

Universität zu Köln

1 September 2010 – 30 January 2012 · 1 yr 5 mos

Cologne, Germany

Mathematics and statistical physics for Bachelor students.

UPMC Sorbonne Universités

1 October 2005 – 31 August 2009 · 3 yrs 11 mos

Paris, France

Thermodynamics, optics and waves, mathematical methods for Bachelor students.

**Doctoral
researcher**

Laboratoire de Physique Statistique (LPS), École normale supérieure

1 September 2005 – 31 December 2009 · 4 yrs 4 mos

Paris, France

www.labos.upmc.fr/ljp/?article7 · www.lps.ens.fr

Theoretical and experimental biophysics, in the Chatenay group: Dynamics of gene regulatory networks.

Time series of gene expression levels, via fluorescent microscopy, of the lysis-lysogeny decision network of the bacteriophage Lambda:

- molecular biology (extraction of viral genes, insertion of genes coding for fluorescent protein, modification of bacterial genomes);
- microbiology (bacterial and viral cultures);
- automation and fluorescent microscopy;
- image analysis.

Mathematical analysis of noise statistics of bacterial gene expression.

Computer simulations of the dynamics and evolution of gene regulatory networks.

Won a 60 k grant to fund the experimental project (over 3 years, used to buy apparatus and consumables): program “Interface physique, biologie et chimie : soutien à la prise de risque 2007-2009” of the CNRS.

**Volunteering
Founding
member**

Celsius

1 August 2007 – 31 August 2009 · 2 yrs 1 mo

Paris, France

Celsius was a think tank with the goal of developing the European project.

- Elaboration of background documents on technical themes;
- preparation, organization of, and follow-up on, two-days meetings in Madrid, Brussels, and Paris, each with more than 30 participants;
- development of publication strategies, print and online.

Internships	Laboratoire de l'Accélérateur Linéaire (LAL), Université Paris-Sud 1 February 2005 – 31 March 2005 · 2 mos Orsay, France Theoretical particle physics, in the Davier group: Computation of fundamental physical quantities for high energy physics from particle collision experiments data.
	Deutsches Elektronen-Synchrotron (DESY), Humboldt-Universität zu Berlin 17 January 2003 – 31 August 2003 · 7 mos Zeuthen, Germany Theoretical particle physics, in the Jegerlehner group: Analysis of experimental data with ad hoc and first-principle mathematical models of elementary particles.
	Laboratoire Kastler Brossel (LKB), École normale supérieure 1 June 2002 – 31 August 2002 · 3 mos Paris, France Experimental quantum physics, in the Grynberg group: Construction of an optical atom trap.

Education

Certifications	Coursera 1 January 2016 – 31 March 2016 · 3 mos Practical Predictive Analytics: Models and Methods Machine Learning Data Manipulation at Scale: Systems and Algorithms
Ph.D.	École normale supérieure 1 September 2005 – 31 December 2009 · 4 yrs 4 mos Paris, France Theoretical and experimental biophysics: Dynamics of gene regulatory networks. See “Experience” above for details.
B.Sc. & M.Sc.	École normale supérieure 1 September 2001 – 31 August 2005 · 4 yrs Paris, France Theoretical and mathematical physics: specialization in particle physics and statistical physics.
	Competitive fellowship.
	Activities and societies: President of the Photography and Video-Making student associations in 2002 and 2003: <ul style="list-style-type: none"> · presenting, defending, and managing budgets; · initiation and support for members.

M.A. **École nationale supérieure des arts décoratifs**
1 September 2003 – 31 August 2004 · 1 yr
Paris, France
CGI, 3d animation, post-production.

Skills

Programming	R, Python, LaTeX, Hugo, HTML/CSS/Bootstrap, Mathematica, Matlab, Processing (Java), SQL, Perl, C++, Shell
Tools	Git, Jira, Confluence, Team Foundation Server (agile management: Scrum, Kanban), SQL Server Management Studio, Photoshop, Illustrator, After Effects, Microsoft Office, 3ds Max, Blender
Languages	French (native) German (professional) English (professional) Japanese (basics)

Selected projects

outbreak detection evaluation

a new general approach to evaluating outbreak-detection algorithms implemented in the Python package epi-quark (main developer Auss Abbood)

preprint: medrxiv.org/content/10.1101/2022.03.16.22272469

code: github.com/aauss/epi-quark

eios anomaly

Python notebooks for natural language processing and machine learning to identify epidemiologically relevant texts and support epidemic intelligence

code & motivation: github.com/stephaneghozzi/eios-anomaly

slides (pdf): github.com/stephaneghozzi/eios-anomaly/blob/master/eios_anomaly-seoul-20191113.pdf

sormasdatagen

an R package for generating synthetic data for the outbreak-management software SORMAS

code & motivation: github.com/stephaneghozzi/sormasdatagen

world health organization outbreak toolkit virtual assistant

a prototypical interactive R Shiny web application with analyses and visualizations for the investigation of outbreaks of unknown origins

code & motivation: github.com/stephaneghozzi/wotva

slides (pdf): github.com/stephaneghozzi/wotva/blob/master/doc/outbreak_toolkit-assistant-presentation-20191016-long.pdf

live demo: stephaneghozzi.shinyapps.io/wotva

asmodee trendbreaker evaluation

an R script for evaluating a new outbreak-detection algorithm on simulated and observed Covid-19 time series

paper: doi.org/10.1098/rstb.2020.0266

code: github.com/stephaneghozzi/asmodee-trendbreaker-evaluation

signale logo

Processing scripts to generate patterns used for the graphic identity of a data-science team

code, motivation & output: github.com/stephaneghozzi/signale-logo

Publications

- [1] Martín Lotto Batista, Bruno Moreira Carvalho, Rory Gibb, Balakrishnan Solaraju-Murali, Stefan Flasche, Stefanie Castell, Stéphane Ghazzi, and Rachel Lowe, “Modelling the impact of climate and the environment on the spatiotemporal dynamics of Lyme borreliosis in Germany,” *eBioMedicine*, 10.1016/j.ebiom.2025.105701, 2025. [https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964\(25\)00145-8/fulltext](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(25)00145-8/fulltext)
- [2] Daniela Gorny, Martina Scharlach, Jona Theodor Stahmeyer, Anika Großhennig, Sebastian Häckl, Abhishek Bakuli, Stéphane Ghazzi, Carolina Klett-Tammen, Wilfried Bautsch, Gérard Krause, and Stefanie Castell, “Who chooses to learn? Characteristics of (non-)attendees of trainings on rational oral antibiotic therapy,” *submitted*, 2023.
- [3] Martín Lotto Batista, Eleanor M. Rees, Andrea Gómez, Soledad López, Stefanie Castell, Adam J. Kucharski, Stéphane Ghazzi, Gabriela V. Müller, and Rachel Lowe, “Towards a leptospirosis early warning system in northeastern Argentina,” *Journal of the Royal Society Interface*, 10.1098/rsif.2023.0069, 2023. <https://doi.org/10.1098/rsif.2023.0069>
- [4] Sangeeta Bhatia, Natsuko Imai, Oliver J Watson, Auss Abbood, Philip Abdelmalik, Thijs Cornelissen, Stéphane Ghazzi, Britta Lassmann, Radhika Nagesh, Manon L Ragonnet-Cronin, Johannes Christof Schnitzler, Moritz UG Kraemer, Simon Cauchemez, Pierre Nouvellet, and Anne Cori, “Lessons from COVID-19 for rescalable data collection,” *The Lancet Infectious Diseases*, 10.1016/S1473-3099(23)00121-4, 2023. [https://doi.org/10.1016/S1473-3099\(23\)00121-4](https://doi.org/10.1016/S1473-3099(23)00121-4)
- [5] Manas K Akmatov, Jakob Holstiege, Lotte Dammertz, Joachim Heuer, Claudia Kohring, Martin Lotto-Batista, Friedrich Boeing, Stéphane Ghazzi, Stefanie Castell, and Jörg Bätzing, “Epidemiology of Lyme borreliosis based on outpatient claims data of all people with statutory health insurance, Germany, 2019,” *Eurosurveillance*, 10.2807/1560-7917.ES.2022.27.32.2101193, 2022. <https://doi.org/10.2807/1560-7917.ES.2022.27.32.2101193>
- [6] Bernard C Silenou, Carolin Verset, Basil B Kaburi, Olivier Leuci, Stéphane Ghazzi, Cédric Duboudin, and Gérard Krause, “A Novel Tool for Real-time Estimation of Epidemiological Parameters of Communicable Diseases Using Contact-Tracing Data: Development and Deployment,” *JMIR Public Health Surveillance*, 10.2196/34438, 2022. <https://publichealth.jmir.org/2022/5/e34438>
- [7] Auss Abbood and Stéphane Ghazzi, “General Framework for Evaluating Outbreak Prediction, Detection, and Annotation Algorithms,” *medRxiv*, 10.1101/2022.03.16.22272469, 2022. <https://www.medrxiv.org/content/10.1101/2022.03.16.22272469v2>
- [8] Thibaut Jombart*, Stéphane Ghazzi*, Dirk Schumacher, Timothy J. Taylor, Quentin J. Leclerc, Mark Jit, Stefan Flasche, Felix Greaves, Tom Ward, Rosalind M. Eggo, Emily Nightingale, Sophie Meakin, Oliver J. Brady, Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Graham F. Medley, Michael Höhle, and W. John

Edmunds, “Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, 10.1098/rstb.2020.0266, 2021. <https://doi.org/10.1098/rstb.2020.0266>

- [9] Matthias Becker, Monika Strengert, Daniel Junker, Philipp D. Kaiser, Tobias Kerrinnes, Bjoern Traenkle, Heiko Dinter, Julia Häring, Stéphane Ghazzi, Anne Zeck, Frank Weise, Andreas Peter, Sebastian Hörber, Simon Fink, Felix Ruoff, Alex Dulovic, Tamam Bakchoul, Armin Baillot, Stefan Lohse, Markus Cornberg, Thomas Illig, Jens Gottlieb, Sigrun Smola, André Karch, Klaus Berger, Hans-Georg Rammensee, Katja Schenke-Layland, Annika Nelde, Melanie Märklin, Jonas S. Heitmann, Juliane S. Walz, Markus Templin, Thomas O. Joos, Ulrich Rothbauer, Gérard Krause, and Nicole Schneiderhan-Marra, “Exploring beyond clinical routine SARS-CoV-2 serology using MultiCoV-Ab to evaluate endemic coronavirus cross-reactivity,” *Nature Communications*, 10.1038/s41467-021-20973-3, 2021. <https://doi.org/10.1038/s41467-021-20973-3>
- [10] Auss Abbood, Alexander Ullrich, Rüdiger Busche, and Stéphane Ghazzi, “EventEpi A natural language processing framework for event-based surveillance,” *PLOS Computational Biology*, 10.1371/journal.pcbi.1008277, 2020. <https://doi.org/10.1371/journal.pcbi.1008277>
- [11] Olivera Stojanovi, Johannes Leugering, Gordon Pipa, Stéphane Ghazzi, and Alexander Ullrich, “A Bayesian Monte Carlo approach for predicting the spread of infectious diseases,” *PLOS ONE*, 10.1371/journal.pone.0225838, 2019. <https://doi.org/10.1371/journal.pone.0225838>
- [12] Benedikt Zacher, Alexander Ullrich, and Stéphane Ghazzi, “Supervised Learning for Automated Infectious-Disease-Outbreak Detection,” *Online Journal of Public Health Informatics*, 10.5210/ojphi.v11i1.9770, 2019. <https://doi.org/10.5210/ojphi.v11i1.9770>
- [13] Fabian Eckelmann, Stéphane Ghazzi, and Alexander Ullrich, “Dashboards as strategy to integrate multiple data streams for real time surveillance,” *Online Journal of Public Health Informatics*, 10.5210/ojphi.v11i1.9701, 2019. <https://doi.org/10.5210/ojphi.v11i1.9701>
- [14] Navina Sarma, Alexander Ullrich, Hendrik Wilking, Stéphane Ghazzi, Andreas K. Lindner, Christoph Weber, Alexandra Holzer, Andreas Jansen, Klaus Stark, and Sabine Vygen-Bonnet, “Surveillance on speed: Being aware of infectious diseases in migrants mass accommodations - an easy and flexible toolkit for field application of syndromic surveillance, Germany, 2016 to 2017,” *Eurosurveillance*, 10.2807/1560-7917.ES.2018.23.40.1700430, 2018. <https://doi.org/10.2807/1560-7917.ES.2018.23.40.1700430>
- [15] Lilia Perfeito, Stéphane Ghazzi, Johannes Berg, Karin Schnetz, and Michael Lässig, “Nonlinear Fitness Landscape of a Molecular Pathway,” *PLOS Genetics*, 10.1371/journal.pgen.1002160, 2011. <https://doi.org/10.1371/journal.pgen.1002160>
- [16] Stéphane Ghazzi, Jérôme Wong Ng, Didier Chatenay, and Jérôme Robert, “Inference of plasmid-copy-number mean and noise from single-cell gene expression data,” *Phys. Rev. E*, 10.1103/PhysRevE.82.051916, 2010. <https://doi.org/10.1103/PhysRevE.82.051916>
- [17] Stéphane Ghazzi, “Expression Dynamics of a Genetic Regulatory Network: the Lysis/Lysogeny Decision of Bacteriophage Lambda,” Theses, Université Pierre et Marie Curie - Paris VI, Dec. 2009. <https://tel.archives-ouvertes.fr/tel-00515109>
- [18] Stéphane Ghazzi and Fred Jegerlehner, “Isospin violating effects in e^+e^- vs. τ measurements of the pion form factor $|F_\pi|^2(s)$,” *Physics Letters B*, 10.1016/j.physletb.2004.01.021, 2004. <https://doi.org/10.1016/j.physletb.2004.01.021>