

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

gitlab.com/stephaneghozzi · 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)**

May 25, 2023 – present · 10 mos

Berlin, Germany

pandemichub.who.int · who.int/initiatives/eios

Machine-learning operations, natural language processing in the Pandemic and Epidemic Intelligence Systems department of the Health Emergencies, Preparedness and Response programme.

Data scientist **Independent developer and consultant**

November 1, 2022 – May 24, 2023 · 7 mos

Berlin, Germany

stephaneghozzi.com

Research, statistics, machine learning, data visualization.

Research associate **Helmholtz-Zentrum für Infektionsforschung (HZI)**

June 1, 2020 – May 31, 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)

April 15, 2016 – May 31, 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)

May 1, 2019 – October 31, 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

March 1, 2012 – April 14, 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

March 1, 2010 – February 29, 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

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

Cologne, Germany

Mathematics and statistical physics for Bachelor students.

UPMC Sorbonne Universités

October 1, 2005 – August 31, 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

September 1, 2005 – December 31, 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

August 1, 2007 – August 31, 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

February 1, 2005 – March 31, 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

January 17, 2003 – August 31, 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

June 1, 2002 – August 31, 2002 · 3 mos

Paris, France

Experimental quantum physics, in the Grynberg group: Construction of an optical atom trap.

Education

Certifications

Coursera

January 1, 2016 – March 31, 2016 · 3 mos

Practical Predictive Analytics: Models and Methods

Machine Learning

Data Manipulation at Scale: Systems and Algorithms

Ph.D.

École normale supérieure

September 1, 2005 – December 31, 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 September 1, 2001 – August 31, 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 September 1, 2003 – August 31, 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: gitlab.com/stephaneghozzi/eios-anomaly

slides (pdf): gitlab.com/stephaneghozzi/eios-anomaly/-/raw/master/eios_anomaly-seoul-20191113.pdf?inline=false

sormasdatagen

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

code & motivation: gitlab.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: gitlab.com/stephaneghozzi/wotva

slides (pdf): gitlab.com/stephaneghozzi/wotva/-/raw/master/doc/outbreak_toolkit-assistant-presentation-20191016-long.pdf?inline=false

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: gitlab.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: gitlab.com/stephaneghozzi/signale-logo

Publications

- [1] M. Lotto Batista, B. Carvalho, R. Gibb, B. Solaraju-Murali, S. Flasche, S. Castell, S. Ghazzi, and R. Lowe, “Modelling the impact of climate and the environment on the spatiotemporal dynamics of Lyme borreliosis in Germany,” *to be submitted*, 2024.
- [2] D. Gornyk, M. Scharlach, J. T. Stahmeyer, A. Großhennig, S. Häckl, A. Bakuli, S. Ghazzi, C. Klett-Tammen, W. Bautsch, G. Krause, and S. Castell, “Who chooses to learn? Characteristics of (non-)attendees of trainings on rational oral antibiotic therapy,” *submitted*, 2023.
- [3] M. Lotto Batista, E. M. Rees, A. Gómez, S. López, S. Castell, A. J. Kucharski, S. Ghazzi, G. V. Müller, and R. 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] S. Bhatia, N. Imai, O. J. Watson, A. Abbood, P. Abdelmalik, T. Cornelissen, S. Ghazzi, B. Lassmann, R. Nagesh, M. L. Ragonnet-Cronin, J. C. Schnitzler, M. U. Kraemer, S. Cauchemez, P. Nouvellet, and A. 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] M. K. Akmatov, J. Holstiege, L. Dammertz, J. Heuer, C. Kohring, M. Lotto-Batista, F. Boeing, S. Ghazzi, S. Castell, and J. 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] B. C. Silenou, C. Verset, B. B. Kaburi, O. Leuci, S. Ghazzi, C. Duboudin, and G. 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] A. Abbood and S. 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] T. Jombart*, S. Ghazzi*, D. Schumacher, T. J. Taylor, Q. J. Leclerc, M. Jit, S. Flasche, F. Greaves, T. Ward, R. M. Eggo, E. Nightingale, S. Meakin, O. J. Brady, Centre for Mathematical Modelling of Infectious Diseases COVID-19 Working Group, G. F. Medley, M. Höhle, and W. J. 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] M. Becker, M. Strengert, D. Junker, P. D. Kaiser, T. Kerrinnes, B. Traenkle, H. Dinter, J. Häring, S. Ghazzi, A. Zeck, F. Weise, A. Peter, S. Hörber, S. Fink, F. Ruoff, A. Dulovic, T. Bakchoul, A. Baillet, S. Lohse, M. Cornberg, T. Illig, J. Gottlieb, S. Smola, A. Karch, K. Berger, H.-G. Rammensee, K. Schenke-Layland, A. Nelde, M. Märklin, J. S. Heitmann, J. S. Walz, M. Templin, T. O. Joos, U. Rothbauer, G. Krause, and N. 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] A. Abbood, A. Ullrich, R. Busche, and S. 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] O. Stojanovi, J. Leugering, G. Pipa, S. Ghazzi, and A. 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] B. Zacher, A. Ullrich, and S. 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] F. Eckelmann, S. Ghazzi, and A. 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] N. Sarma, A. Ullrich, H. Wilking, S. Ghazzi, A. K. Lindner, C. Weber, A. Holzer, A. Jansen, K. Stark, and S. 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] L. Perfeito, S. Ghazzi, J. Berg, K. Schnetz, and M. 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] S. Ghazzi, J. Wong Ng, D. Chatenay, and J. 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] S. 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] S. Ghozzi and F. 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>