

# Joana Marques Barros

BSc BIOLOGY, MSc BIOINFORMATICS, PHD COMPUTER SCIENCE

 joanambarros |  Joana M. Barros

## Summary

---

R programmer and interdisciplinary data scientist with experience in the pharmaceutical industry, including clinical data standards and GxP compliance. Skilled at translating complex, cross-disciplinary insights into actionable results. Known for a curious, organized, and dedicated approach, I believe the AI revolution starts with bringing clarity and structure to our data.

## Work Experience

---

### Idorsia Pharmaceuticals

*Allschwil, Switzerland*

CLINICAL DATA SCIENTIST

*June, 2022 - Current*

- Supported Biostatistics, Data Management, and Statistical Programming teams on multiple Phase 3 studies within the Biometry unit
- Led an initiative to teach R programming to over 15 coders and non-coders, transitioning from a supporting role to lead instructor
- Developed R packages for internal Shiny applications and general internal use
- Streamlined statistical programming processes by creating R packages for tasks like releases and QC on datasets, while contributing to the initiative to modernize and increase efficiency through the adoption of internal R packages, version control, and CI/CD automation
- Collaborated with IT to test and troubleshoot internal R-based infrastructure
- Worked cross-departmentally in Global Clinical Development with Biometry, Clinical Operations, and Clinical Science to develop an application for Risk-Based Quality Management
- Brought automation to routine tasks in a >5000 patient study leading to significant efficiency gains
- Managed a 5-week project with one direct report to standardize execution and storage of internal analysis results
- Cultivated trust with colleagues within the Biometrics unit taking a central role as the go-to R support in Data Science

### Novartis

*Basel, Switzerland*

AI4LIFE RESIDENT

*May, 2021 - April 29 2022*

- Worked on a project to rethink the analysis and reporting of clinical trials in a transparent, traceable, interoperable, and easy-to-use way
- Developed concepts and a framework motivating the importance of storing analysis results as data
- Established and implemented a high-level grammar for data analysis designed to be an end-to-end approach to handle clinical analysis results efficiently and enable automation
- Developed working prototypes using R, SQL, and other technologies as well as data standards such as CDISC
- Tested the use of internal software
- Established successful communications with stakeholders within the global management
- Participated in an internal challenge on subgroup analysis, ranking in the top 3

### National University of Ireland, Galway

*Galway, Ireland*

PHD RESEARCHER

*January, 2016 - September, 2020*

- Led extensive research and data analysis of unstructured text and time-series data
- Designed and built pipelines leveraging biomedical semantics, time-series modeling, and machine learning
- Established productive collaborations and managed projects delivering results on agreed deadlines
- Authored and edited over 5 scientific publications and presentations as well as reviewed multiple articles

### National University of Ireland, Galway

*Galway, Ireland*

TEACHING ASSISTANT - PROGRAMMING FOR DATA ANALYTICS (WITH R)

*September, 2017 - December, 2018*

- Coordinated and conducted postgraduate lab classes
- Performed weekly student performance evaluation

### Lasige - Large-Scale Informatics Systems Laboratory

*Lisbon, Portugal*

JUNIOR RESEARCHER

*August, 2014 - November, 2015*

- Designed and implemented quantitative structure-activity relationship (QSAR) models
- Developed machine learning predictive models for drug interactions with the human heart

### CIIMAR - Interdisciplinary Centre of Marine and Environmental Research

*Porto, Portugal*

RESEARCH INTERN

*September, 2012 - June, 2013*

- Outlined and implemented field experiments replicating real-world macro-algae stressors
- Generated statistical analysis on complex data from sensors and algae growth indicators

## Education

---

### National University of Ireland Galway

PHD IN COMPUTER SCIENCE

Ireland  
January, 2016 - September, 2020

- Thesis: Applying Informal Health Reports and Search Queries for Public Health Monitoring: An Evaluation of Characteristics, Potentials, and Requirements of Online Self-Reporting, Discussions, and Search Behaviour
- Advisors: Prof. Dietrich Rebholz-Schuhmann, Prof. Jim Duggan & Dr. Paul Buitelaar

### University of Lisbon

MASTER IN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY, GRADE: 16/20

Portugal  
September, 2013 - November, 2015

- Thesis: Mining Cardiac Side Effects of Known Drugs
- Advisor: Prof. André Falcão

### University of Porto

BACHELOR IN BIOLOGY, GRADE: 14/20

Portugal  
September, 2010 - July, 2013

- Internship Report: The Consequences of Global Climate Change in the Portuguese Coastal Communities
- Advisor: Prof. Isabel Sousa Pinto

## Communications

---

### GENERAL PUBLIC (SELECTED 2/5)

- 2023 **A Case-Study Driven Motivation of Analysis Results Data**, R in Pharma *Online Workshop*
- 2019 **Irish scientists using AI to help predict national suicide rates in new study**, Irish Examiner *Online Newspaper*

### ACADEMIC (SELECTED 3/5)

- 2019 **Unsupervised Classification of Health Content on Reddit**, Digital Public Health Conference *Marseille, France*
- 2017 **Text Mining from Social Media for Public Health Applications**, International Conference on Digital Health *London, UK*

## Skills

---

<b>Technical Skills</b>	Code Management, Code Development, Cloud Computing, Data Modeling, Data Management, Data Visualization, Deep Learning, Machine Learning, Natural Language Processing, Research Engineering, Statistical Analysis, Text Mining, Time-Series Analysis
<b>Programming Skills</b>	Python, R, HTML
<b>Tools</b>	Git, Visual Studio Code, Keras/Tensorflow, MetaMap, MongoDB, MySQL, Shiny, SQLite, NLTK, Numpy, Pandas, Posit Workbench, Posit Connect, RStudio, Scikit-Learn
<b>Compliance &amp; Regulatory</b>	Research Integrity Training, EU GDPR and Health Research Regulations, Good Clinical Practice
<b>Languages</b>	Portuguese (Native), English (Proficient), Spanish (Intermediate), German (Intermediate)

## Extra Activities

---

- 2024-2025 **Local Organizing Committee**, 46th Annual Conference of the International Society for Clinical Biostatistics *Basel, Switzerland*
- 2022- **Committee Member**, Basel Biometrics Society - NextGen *Basel, Switzerland*
- 2018 **Research Integrity Training**, National University of Ireland Galway *Galway, Ireland*