

CASE STUDIES IN DATA SCIENCE: BIOMEDICAL RESEARCH

Jana Kristin Eckert, Postdoc/Data
scientist, Department of
Population Allergy, MCRI

BACKGROUND: A MARATHON, NOT A RACE

- Berlin: Uni, PhD: (wet) lab-based working on pathogen immune cell interaction, started wondering about stats (but didn't really know anything about it); MPH started but never finished (discovered data and what you could do with it)
- Heidelberg: Postdoc (DKFZ, first attempt at data science but wasn't allowed)
- Melbourne I: opportunity to work with and on data, becoming a data scientist at MCRI?
- Munich: work as clinical trial monitor and study coordinator
- Melbourne II: clinical trial manager, postdoc/data scientist

A DATA SCIENTIST

is a professional who uses data to help solve problems and make decisions. Their job is to extract insights from complex and often messy DATA.

OTHER DATA JOBS

Data engineering

builds the infrastructure that collects, stores, and processes data

Data analysis

focuses on interpreting existing data to find trends and support decision-making. A data analyst sits between business intelligence and data science. They provide vital information to business stakeholders.

Data governance & privacy

ensures data quality, compliance, and privacy.

Database administration & development

manages and optimizes databases.

Data visualization & reporting

creates dashboards and visual tools to communicate data insights.

Specialized/niche data roles

domain-specific or emerging roles that combine data expertise with another field.

Geospatial data analyst

Healthcare data analyst

Quantitative analyst – finance

WHAT A DATA SCIENTIST DOES

Collects Data

They gather data from various sources e.g. observational study, clinical trial, public data bases

Cleans and Prepares Data

Real-world data is often messy. Data scientists clean, format, and transform it to make it usable.

Explores and Analyzes Data

They look for patterns, trends, and relationships in the data using statistics and visualization tools.
(Overlap/cooperation with other data jobs)

Builds Models

Using machine learning or statistical models, they make predictions or classifications

Communicates Insights

They create reports, dashboards, or visualizations to explain their findings to decision-makers in a clear and actionable way.

Deploys Solutions

Sometimes, they help put their models into production, recommendations for guidelines etc..

CHALLENGE - COLLABORATION

- Different Backgrounds, Different Languages
- Conflicting Priorities (individual timeline and topics)
- Unclear objectives can result in irrelevant analyses or wrong solutions
- Data Access and Quality Issues

CHALLENGE - COLLABORATING WITH DATA SCIENTISTS IN THE MEDICAL FIELD

- Get involved as early as possible
- Establish data ownership and ways of communication
- Document everything, prioritize reproducibility
- Plan for data storage and distribution

CHALLENGE - COLLABORATING WITH DATA SCIENTISTS IN THE MEDICAL FIELD

- Clarify goals early and define the problem together
- Establish a timeline, even if you'll have to change it
- Embrace creativity, but Avoid 'scope creep'
- Authorship

Most important: Communication

- Learn the lingo
- Build bridges
- Encourage questions
- Be open

Q&A: open for discussion

HOW TO GET INVOLVED IN BIOMEDICAL DATA SCIENCE

- Volunteer
- Honor/Master projects
- PhD/Scholarships

Thank you

Q&A: open for discussion

Will I be replaced by AI?

Article in The Age: Learning to code was the golden ticket to a six-figure salary. Perhaps no longer, thanks to AI

Q&A: open for discussion

“AI will eventually absorb the tedious entry-level work that young employees traditionally cut their teeth on – summarising meetings, cleaning spreadsheets, drafting contracts or writing simple code. “

Q&A: open for discussion

“AI is automating the grunt coders, the ones who code without thinking as much. But AI can’t replace people who work to build things, synthesise stuff together.”

Q&A: open for discussion

“To make AI and robots, you need people.”