RUST TURAKULOV

Data Analysis | Machine Learning | Bioinformatics and Genomics

Wantirna South, VIC, 3152 | 0405 174 704 | rust.turakulov@gmail.com | https://rustt.app | LinkedIn

ABOUT ME

I am data specialist with biomedical background. Proficient in Perl, R, SQL, Python, Linux, and cloud configuration. Delivered data-driven insights for institutions like AGRF, ANU and NIH. Proven expertise in machine learning applications, data science, and business intelligence, poised for senior roles in data and business analytics. Recognised for driving impactful solutions and actionable insights through advanced technical and analytical skills. Check my portfolio on https://rustt.app

MAJOR QUALIFICATIONS

▪ PhD in Molecular Genetics – Lomonosov Moscow State University	1998
▪ Honorary Lecturer – College of Science & Engineering, Flinders University, Adelaide, SA	In Progress

PROFESSIONAL DEVELOPMENT

▪ VIC Driver's Licence	Current
▪ AWS Cloud Practitioner (certified)	2024
▪ DevOps Engineering on AWS	2024
▪ Bioinformatics & Programming Courses	2001

SOFTWARE AND TECHNOLOGY

- Aurora DB
- AWS IAM
- AWS LightSail
- EC2
- CKAN
- Cloud VM
- Docker
- HPC Slurm
- LINUX
- MS OLE
- MySQL/SQL
- Palantir Foundry
- Perl
- Python
- R
- S3
- Shell
- Shiny Web Server

PROFESSIONAL EXPERIENCE

Senior Developer | Australian Genome Research Facility (AGRF), Melbourne, VIC **09/2023 - Present**

- Developed applications to analyse bacterial profiles and in-house data assets, enabling streamlined insights for agricultural research studies.
- Integrated microbial dataset with over 10k+ entries with 2 Australian governments and maintained geographical databases with over 10m+ records.
- Designed website application with over 40+ predictive models and classification algorithms for sorting soil bacterial profiles uploaded to the site in real-time and providing interactive analysis features.
- Developing a soil bacteria profile report to assist decision-making for environmental monitoring of rehabilitation sites in the mining industry. This report product potentially could have a huge economic impact on multimillion-dollar businesses. Work in progress.
- Developed an R Shiny dashboard to monitor the clinical data for laboratory quality control metrics, facilitating longitudinal tracking and seamless navigation of over 5+ historical and current clinical samples.
- Established an open data registry platform to foster clinical research collaboration, supporting transparent and accessible data sharing for clinical datasets complying with the Privacy Act, CLIA and HIPAA.
- Implemented MVP (Minimal Viable Product) for secure clinical data ledger application through AWS Aurora database and Marquez platform, enabling detailed data lineage tracking files for up to 10k records.
- Developed the CRC TiME Data Browser for advanced data exploration, supporting collaborative research and the project-wide dataset exploration in interactive mode on R Shiny application with multidimensional reduction of hundreds of soil samples and over 5k+ variables; prototype launched.

- Integrated continent-wide datasets into an interactive application with an intuitive dashboard via the R-Shiny platform, allowing potential clients to compare their data with a reference dataset and open-sourced metadata. Queried over 10k spatial samples and integrated 3k samples.
- Led a developer team from consulting company Infoobjects to develop AGRF's first AWS-based web service (MVP), enabling global access to data applications with 20+ daily users. The service featured 4 Shiny apps and an AWS Serverless Aurora DB-powered data ledger.
- Developed the data share platform based on the CKAN (MVP) project purposed for efficient data storage and content search, enhancing clinical research accessibility and facilitating collaboration.
- Managed all project stages, from quote collection and price negotiation through all development stages, including architecture and testing.
- Secured funding as a Genovite23 grants idea winner, obtaining nearly $150k across 2 awarded grants.
- Supported 3 research projects, including 3 conference abstracts published within one year.

Laboratory Information System Developer, Staff Scientist | Laboratory of Pathology, National Cancer Institute, National Institute of Health, Bethesda, USA **11/2020 - 08/2023**

- Designed a Laboratory Information System (LIS) data workflow with the Palantir enterprise system.
- Managed routine public domain dataset scanning and integration with in-house curated clinical datasets, supporting research through large-scale dataset sharing and analysis.
- Collaborated with a dozen clinicians from 7 groups to optimise R-Shiny and in-house systems while designing a pipeline for methylation data on the NIH HPC cluster.
- Implemented the Laboratory of Pathology LIS for 10k+ records, 3k+ historical data points, and weekly processing of 50-100 new samples to impact cancer diagnostics.
- Launched the Methylscape portal housing a curated collection of 100k+ tumour samples, including metadata from LP and public domain samples utilised by 200+ molecular pathologists globally.
- Collaborated on high-profile publications in Nature Medicine, advancing 3 research studies on applying machine learning algorithms for kidney, haematological, and pan-cancer tumour diagnostics.
- Led tumour classification research on brain and hematological tumours, testing machine learning algorithms, optimising datasets, and conducting exploratory analysis on methylation profiling clusters.

Bioinformatician | Centre for Biodiversity Analysis, Research School of Biology, Australian National University, Canberra, ACT **04/2019 - 10/2020**

- Created a Docker container for a variant calling pipeline with 200+ downloads and developed a Target Capture Cloud Platform pipeline with 100+ downloads.
- Migrated the Laboratory Dataset from MySQL to a new cloud provider, leveraging a robust transition strategy, impacting over 10k+ data entries.
- Developed 3 analytical sequencing pipelines across in-house servers and national supercomputing facilities, including the Australian Supercomputer Facility, PAWSEY, and NECTAR cloud machines.
- Supported High-Performance Computing (HPC) workflows, boosting computational efficiency reducing computational timeframes from ~3 months to approx. 1 week.
- Provided over 5 workshops and group training sessions for students and researchers, including interstate trips.

Data Scientist (Contractor) | SmartDNA, Clayton, VIC **04/2016 - 04/2019**

- Enhanced the efficiency of bacterial analysis by developing a new workflow and maintaining an in-house sequence database, processing over 100+ samples monthly.
- Developed a patented diagnostic method for identifying human medical conditions based on bacterial profiles, securing patent AU 2019203763 ("Method of Diagnosing a Dysbiosis") on May 29, 2019.
- Developed a Perl-coded pipeline for 16S sequence analysis and the smartHIT™ microbiome database, creating an automated workflow that processes data, merges patient data with a reference group, and generates a client-ready PDF report using a random forest model.
- Leveraged Perl, R (with various libraries), PDF::API2, and Linux bash commands to develop a robust pipeline, delivering precision data and insights for medical practitioners.
- Developed an advanced diagnostic prediction model for irritable bowel syndrome (IBS), supporting treating 500–1k patients annually.
- Leveraged smartGUT™ reporting with IBS scoring to enhance clinical decision-making and guide therapy selection for optimal patient outcomes.
- Improved client communication by developing automated reports for delivering results, reducing client feedback time from over 2 weeks to under 3 days.
- Maintained the sequence database and scripting report generation from spreadsheet data, including creating a script to convert multi-client Excel files into client-specific reports in Word and PDF formats.
- Managed operational processes by handling routine batches of 10 to 100 clinical samples weekly, ensuring annual timely analysis for over 3k clinical tests.
- Secured intellectual property rights in Australia with the IVD IBS diagnostic test patent, driving potential market access for 9–15% of the Australian population or 2.2 to 3.7 million people.
- Expanded market presence by developing SmartGut, SmartSpectrum, and SmartIBS direct-to-consumer products with fully automated pipelines, increasing sales revenue by 30%.
- Strengthened organisational data resources by establishing an in-house database of cases and bacterial collections, enhancing research capabilities for 2 ongoing projects, including 1 clinical trial of a commercial probiotic strain.

Senior Scientist/Bioinformatician | AGRF, Melbourne, VIC **April/2004 - April/2016**

- Delivered cutting-edge bioinformatic analyses by overseeing data processing and interpretation for up to 100 genome research projects per year across Melbourne, enhancing data accuracy and throughput.
- Developed a Perl-coded variant calling pipeline for a high-performance cluster, integrating it with the in-house Laboratory Information System (LIMS) to streamline sample tracking and data processing.
- Played a pivotal role in data integration, extensive client documentation, and internal accreditation processes, enhancing service quality and driving a high client return rate.
- Led over 10 training sessions and workshops, contributing to company-wide knowledge sharing and skill development.
- Instrumental in launching new company services and refining existing ones, including the 16S quality control system, MLST testing protocol and analysis script, BVDV test, GoldenGate assay, and GBS pipeline.
- Developed R scripts to automate microarray data analysis, improving process efficiency and switching analysis from licensed software Partek to open-source R libraries, saving the company over $10k+yearly on licence fees.
- Created and maintained three core production analysis pipelines, established quality control metrics, and integrated these into the LIMS, ensuring precise and reliable data management.
- Conducted statistical analysis for routine internal and client experiments, provided bioinformatics support, and advanced collaborative projects, driving high-quality research outcomes daily.

REFERENCES

Available on request