

Sanjay Kumar Srikakulam

Germany · [GitHub](#) · [Google Scholar](#) · [ORCID](#) · [ResearchGate](#) · [LinkedIn](#) · [Zenodo](#)

Postdoctoral researcher at Forschungszentrum Jülich, building small language models and agentic systems for mass spectrometry, including spectral annotation for lipidomics. Earlier work spans molecular dynamics and structure-based drug discovery, sequence indexing at petabyte scale, and three years running Galaxy Europe for over 170,000 researchers.

EDUCATION

PhD, Bioinformatics

2024 · Helmholtz Institute for Pharmaceutical Research Saarland and Saarland University, Saarbrücken, Germany

M.Sc., Bioinformatics

2018 · Saarland University, Saarbrücken, Germany

B.Tech., Bioinformatics

2014 · SRM University, Kancheepuram, India

EXPERIENCE

Postdoctoral Researcher

09/2025 to Present · Forschungszentrum Jülich GmbH, Jülich, Germany

- AI and agentic frameworks for biocuration of metabolomics and lipidomics literature: named entity recognition, ontologies, controlled vocabularies, and knowledge graphs.
- SeqSidecar, a columnar query engine for nucleotide and protein sequencing reads, supporting filtering, search, deduplication, and remote index-based querying of FASTA and FASTQ at scale.
- An OCI container registry with CVE scanning, SBOM generation, and image signing, run as a European service through de.NBI and ELIXIR.
- Novel algorithms and data structures for indexing and querying petabyte-scale sequencing datasets.
- Administration, strategic planning, and implementation of de.NBI infrastructure and services, focused on cloud infrastructure, AI, and agentic frameworks.
- Administration and strategic expansion of de.NBI and ELIXIR Germany infrastructure and services.
- Demonstrated EOSC EU Node use cases through the EOSC United project.
- Part of the EOSC Entrust project, establishing blueprints for secure data processing environments.
- Talks, posters, and workshops at national and international conferences.
- Represented de.NBI and ELIXIR Germany in national initiatives (NFDI) and European Commission projects.

Scientific Researcher (Postdoc)

08/2022 to 08/2025 · University of Freiburg, Freiburg, Germany

- Administration and management of Galaxy Europe, supporting over 170,000 users.
- Algorithms and optimisations for meta-scheduling, including reducing carbon footprint through efficient scheduling and distributed remote computing.
- Configuration, deployment, administration, and maintenance of the Galaxy HPC OpenStack cloud using Terraform, Ansible, Jenkins, Rsyslog, and the TIG stack.
- Application development, feature innovation, and research.
- Deliverables and reports for national and EU-funded initiatives.
- Research data management through Galaxy.
- Development and deployment of configurations through GitOps and Jenkins CI/CD.
- Scripting and automation.
- Infrastructure monitoring and performance analysis.
- Troubleshooting, logging, and log analysis.
- Support and documentation.
- Partnering with collaborators on national and international projects.
- Represented the organisation at conferences through exhibition booths, presentations, and scientific and industry discussions.

Bioinformatician

02/2021 to 02/2022 · Saarland University Hospital, Homburg, Germany

- Machine learning and survival analysis on longitudinal COPD patient data to identify biomarkers.
- OMICS data analysis.

System Administrator and DevOps, Bioinformatics IT Core Facility

10/2018 to 03/2023 · Helmholtz Institute for Pharmaceutical Research Saarland, Saarbrücken, Germany

- Setup, design, and establishment of the Bioinformatics IT core facility.
- Installation, configuration, and maintenance of CentOS and Red Hat Linux HPC clusters, servers, desktops, and workstations.
- Containerised web services (Gitea, MediaWiki, WordPress, Docker registry) and infrastructure monitoring through Icinga.
- Site reliability engineering processes.
- Budgeting, purchasing, and inventory.
- Bioinformatics application installation, configuration, deployment, and maintenance.
- Support and documentation for all researchers.
- On-demand compute provisioning, physical and virtual, with network and storage allocation.
- Storage provisioning and administration.
- MariaDB database server administration.
- Ansible and custom RPM packages for configuration, monitoring, and maintenance.
- Configuration of iDRAC/IPMI, DNS, TCP/IP, TLS/SSL, LDAP, Active Directory, SSH, Fail2ban, firewall, AutoFS, and printers.
- Troubleshooting and resolution of hardware and software issues.

System Administrator (Research Assistant)

09/2015 to 03/2022 · German Research Centre for Artificial Intelligence (DFKI), Saarbrücken, Germany

- Development and maintenance of an RPM-based distributed configuration management system.
- Development, configuration, and maintenance of web services including MediaWiki, WordPress, Gitea, and reverse proxies.
- Configuration and maintenance of Jenkins build automation and KVM virtual machines.
- Development and maintenance of installation and configuration routines.
- Configuration and maintenance of network monitoring through Icinga 2.
- Development and maintenance of containerised applications using Docker.
- Scripting and automation.
- Monitoring and maintenance of workstations and servers.
- Atlassian stack administration: Crowd, Jira, Confluence, and Bitbucket.
- Support and documentation for the DFKI-ASR user community and the team.

Project Assistant (Student)

05/2013 to 07/2014 · National Institute for Research in Tuberculosis, ICMR, Chennai, India

- Maintained the Mycobacterium tuberculosis structural database (MtbSD).
- Sequence and structural ensemble analysis of HIV and M. tuberculosis proteins.
- Molecular modelling, drug design, and docking.
- Statistical analysis.

PUBLICATIONS

The Galaxy Community. **Galaxy for accessible, reproducible, and collaborative data analyses: 2026 update**. Nucleic Acids Research, 2026. [doi:10.1093/nar/gkag469](https://doi.org/10.1093/nar/gkag469)

Azab A, Geest PD, Srikakulam SK, Vondrák T, Kuntz M, Grüning B. **Optimized Meta-Scheduling in Galaxy Using TPV Broker**. ICT for Intelligent Systems (ICTIS 2025), Smart Innovation, Systems and Technologies vol 325, Springer, 2026. [doi:10.1007/978-981-95-1365-9_42](https://doi.org/10.1007/978-981-95-1365-9_42)

Bacon WA, Batut B, Srikakulam SK, Zierep P, Bretaudeau A, et al.. **Ten common misconceptions about Galaxy (and why they are wrong!)**. PLOS Computational Biology, 2026. [doi:10.1371/journal.pcbi.1013869](https://doi.org/10.1371/journal.pcbi.1013869)

Twardziok S, Barysch S-V, Braun M, Buchhalter I, Grüning B, Kraft P, Krüger J, Kuchenbecker L, Srikakulam SK, Lawerenz C, Sczyrba A, Wagener H, Hoffmann N, Kohlbacher O. **Towards Federated, Certified Infrastructures for Sensitive Data Research in Germany and Europe with the de.NBI Cloud**. Zenodo, 2026. [doi:10.5281/zenodo.19882942](https://doi.org/10.5281/zenodo.19882942)

Data Snack: Galaxy, a research data powerhouse from analysis to FAIR management. Zenodo, 2025. [doi:10.5281/zenodo.15362960](https://doi.org/10.5281/zenodo.15362960)

Dynamic meta-scheduling in Galaxy with TPV Broker for smarter workload distribution. Zenodo, 2025. [doi:10.5281/zenodo.16312227](https://doi.org/10.5281/zenodo.16312227)

Interactive Galaxy Tools: combining real-time user interaction with scientific workflows. Zenodo, 2025. [doi:10.5281/zenodo.16735306](https://doi.org/10.5281/zenodo.16735306)

Publication on the smart jobs scheduler implementation. Zenodo, 2025. [doi:10.5281/zenodo.15827107](https://doi.org/10.5281/zenodo.15827107)

Tangaro MA, Nicotri S, Grüning B, Srikakulam SK, Dadras A, Zambelli F. **Scaling Scientific Workflows in Europe: Architecture and Deployment of the Galaxy-Pulsar Computational Network**. Zenodo, 2025. [doi:10.5281/zenodo.16761933](https://doi.org/10.5281/zenodo.16761933)

Bring Your Own Infrastructure. Zenodo, 2024. [doi:10.5281/zenodo.15729502](https://doi.org/10.5281/zenodo.15729502)

TECHNICAL SKILLS

AI and LLM stack: vLLM, PydanticAI, coding agents, pgvector, OpenAI API

Certificate management: Let's Encrypt

CI/CD: Jenkins

Cloud DNS: AWS Route 53, DigitalOcean

Cloud platforms and services: OpenStack, AWS

Containers and virtualisation: Docker, LXC, KVM

Databases: SQL, MySQL/MariaDB, PostgreSQL

High-throughput workload scheduling: HTCondor

Infrastructure as code and configuration management: Terraform, Ansible, RPM packaging, Pulp

Monitoring and observability: Icinga, Telegraf, InfluxDB, Grafana

Programming: Python, Shell scripting (Bash), R

Security: Firewall management, SSL/TLS certificates, Fail2ban

Storage: NAS, DAS, JBOD, NFS, S3, CVMFS

Version control: Git

Web servers: Apache, Nginx