
Javier Millán Acosta

Doctoral Researcher, Maastricht University, Department of Translational Genomics (TGX) · Maastricht, Netherlands

I work on software that makes biomedical data interoperable: schema and identifier reconciliation across databases and their releases to improve access to services and the tools we build with them.

javier.millanacosta@maastrichtuniversity.nl ORCID [GitHub](#) [Google Scholar](#) [LinkedIn](#) [PyPI](#)

EXPERIENCE

Doctoral Researcher

Maastricht University, Department of Translational Genomics (TGX) · Maastricht, Netherlands

Jul 2022 – present · PhD defense expected 2026

Thesis on the interoperability of life sciences data: how independently maintained RDF resources can be described, discovered, and queried together, and how the continual revision of the underlying databases affects integration and downstream usage.

Database description: I developed `rdfsolve` to mine schemas and paths from SPARQL endpoints and RDF dumps into descriptive (VoID) and prescriptive (SHACL) formats, generate typed Python models, and derive mappings across datasets. I maintain a registry of life sciences RDF resources recording what each exposes and how it links to the others.

Database evolution: I conceptualized a methodology to capture the historical events of identifier deprecation and label aliasing and supersession across ChEBI, Ensembl, HGNC, HMDB, NCBI, UniProt, and virtually any database with a release archive where deprecation events are explicit or implicit, and compare releases as the basis for measuring how these revisions propagate into downstream tools and resources (manuscript in preparation).

Researcher (internship)

FrieslandCampina · Wageningen, Netherlands

Jun 2021 – Oct 2021

Studied the wasting effect of dietary protein intake. Wrote scripts to extract, harmonize, and analyze food-systems and food-composition data from several countries; harmonized it and built dashboards with Power BI.

Researcher (internship)

Institut de Recerca Biomèdica de Lleida (IRBLleida), Clinical Neurosciences lab · Lleida, Spain

Jun 2018 – Sep 2018

Investigated conditioning in ischemic-stroke recovery using ELISA, RNA extraction, and preparation of mouse blood and brain-slice samples.

Projects and communities

Funded projects: VHP4Safety — NWA-ORC grant 1292.19.272, Dutch Research Council (NWO); SbD4Nano — H2020 grant 862195, European Commission; NanoSolveIT — H2020 grant 814572, European Commission; NanoCommons — H2020 grant 731032, European Commission.

Open communities: WikiPathways, eNanoMapper, BioDataFuse, BridgeDb, ELIXIR Toxicology Community, NFDI4Chem, PINK.

OPEN-SOURCE SOFTWARE

Libraries and tools

- **rdfsolve** — RDF schema discovery and conversion, typed Python clients, and SPARQL execution with batching, retries, and query provenance. · [docs](#) · [PyPI](#)
- **pysec2pri** — Reconciles retired identifiers and changed gene symbols against current database records, with historical mappings, ambiguity checks, and auditable resolution. · [docs](#) · [PyPI](#)
- **mapkgsutils** — Shared parsing, reconciliation, and release-comparison machinery for biomedical identifier mappings. · [PyPI](#)
- **pyaop** — Builds and enriches adverse outcome pathway networks from SPARQL sources, with gene, compound, and tissue associations for network visualization.

- **AOP-Suite** — web application for building adverse outcome pathway networks from SPARQL sources, with gene and compound queries and interactive Cytoscape.js views; built on pyAOP. · aopsuite.cloud.vhp4safety.nl · [tutorial](#)

Contributions to shared projects

- **BioDataFuse** — implemented RDF graph generation, AOP-Wiki and WikiPathways integrations, and GraphDB support; developed the corresponding API and Vue interface in `biodatafuseUI`; code maintenance. · [Zenodo](#) · [Bioinformatics 2026](#) · [CEUR Workshop Proceedings \(SWAT4HCLS 2025\)](#) 2025
- **VHP4Safety Virtual Human Platform** — implemented services and frontend, search and filtering, and glossary integration across Flask and JavaScript. Also developed the AOP-Suite (tutorial), for SPARQL-backed network construction, gene and compound queries, and interactive Cytoscape.js views, integrating the `pyAOP` library. · [Zenodo](#)
- **eNanoMapper** — developed ontology terms and imports for nanosafety; added integrity tests, corrected OWL profile violations, and maintained build and validation workflows. · [Zenodo](#)
- **BridgeDb** — automated Docker builds and versioned image releases across the webservice and data repositories; added webservice CI tests. · [Zenodo](#)
- **CNV Pathway Atlas** — built a phenotype-based browser with combined filters, ontology lookups, and an automated data-generation step. · [Scientific Data 2026](#)

Personal tooling: `seejobs`, a terminal interface for Slurm job monitoring, logs, and batch-script preparation over SSH.

Merged contributions · All repositories

PUBLICATIONS

Further research outputs are available on [ORCID](#) and [Google Scholar](#).

- **A dataset of rare copy number variants associated with neurodevelopmental and neuropsychiatric disorders**
A. Valeanu, Y. Duan, **J. Millán Acosta**, T.M. de Kok, T. van Amelsvoort, F. Ehrhart
Scientific Data · 2026 · Journal article · [10.1038/s41597-026-07345-6](https://doi.org/10.1038/s41597-026-07345-6)
- **pyBiodatafuse: extending interoperability of data using modular queries across biomedical resources**
Y. Gadiya, **J. Millán Acosta**, A. Ammar, A. Adriaque Lozano, D. Wetstede, D. Martinát, A.C. Sima, H. Mei, E. Willighagen, T. Abbassi-Daloi
Bioinformatics · 2026 · Journal article · [10.1093/bioinformatics/btag064](https://doi.org/10.1093/bioinformatics/btag064)
- **BioHackEU24 report: Expanding FAIR database integration through elucidation and transformation of underlying graph schemas**
J. Millán Acosta, S. Kawashima, T. Katayama, J. Bolleman, D. Martinat, H. Detering, J.E.L. Gayo, Y. Gadiya, T. Abbassi-Daloi
2025 · Preprint · [10.37044/osf.io/ptmg5_v1](https://doi.org/10.37044/osf.io/ptmg5_v1)
- **LP-63 Making the AOP-Wiki knowledge graph usable across disciplines in toxicological assessment**
M. Martens, S. Adil, **J. Millán Acosta**, E. Willighagen
Toxicology Letters · 2025 · Conference abstract · [10.1016/j.toxlet.2025.07.1074](https://doi.org/10.1016/j.toxlet.2025.07.1074)
- **Extended RDF support for Biomedical Knowledge Graphs in pyBioDataFuse: on-the-fly RDF graph generation and new resource annotators**
J. Millán Acosta, E. Willighagen, Y. Gadiya, T. Abbassi-Daloi
CEUR Workshop Proceedings (SWAT4HCLS 2025) · 2025 · Conference paper · [10.24406/publica-8969](https://doi.org/10.24406/publica-8969)
- **MCP server tools with RDF shapes**
J.E. Labra-Gayo, Y. Yamamoto, A.R. Kinjo, A. Waagmeester, **J. Millán Acosta**, S. Kawashima, Y. Okabeppu, J. Koblitz, S.B. Larriet
2025 · Preprint · [10.37044/osf.io/8qeh5_v1](https://doi.org/10.37044/osf.io/8qeh5_v1)
- **INTOXICOM Workshop Report: FAIRification of Toxicological Research Output: Leveraging ELIXIR Resources**
M. Martens, I. Lynch, T. Exner ... **J. Millán Acosta** ... E. Willighagen
2024 · Preprint · [10.37044/osf.io/un2rw](https://doi.org/10.37044/osf.io/un2rw)
- **Harmonising knowledge for safer materials via the “NanoCommons” Knowledge Base**
D. Maier, T.E. Exner, A.G. Papadiamantis ... **J. Millán Acosta** ... I. Lynch
Frontiers in Physics · 2023 · Journal article · [10.3389/fphy.2023.1271842](https://doi.org/10.3389/fphy.2023.1271842)
- **WikiPathways 2024: next generation pathway database**
A. Agrawal, H. Balci, K. Hanspers ... **J. Millán Acosta** ... A. Pico
Nucleic Acids Research · 2023 · Journal article · [10.1093/nar/gkad960](https://doi.org/10.1093/nar/gkad960)

Project deliverables

- **D1.5 Report on the final ontology releases and guidance**
NanoSolveIT, H2020 grant 814572, European Commission · 2023 · CORDIS
 - **D6.3 Case studies in toxicological profiling**
SbD4Nano, H2020 grant 862195, European Commission · 2023 · CORDIS
-

CONFERENCES & HACKATHONS

Conferences

- **SWAT4HCLS 2025 (Barcelona, Spain)** — Poster: on-the-fly RDF graph generation and new resource annotators in pyBioDataFuse, with a paper in CEUR Workshop Proceedings. · CEUR Workshop Proceedings (SWAT4HCLS 2025) 2025
- **NanoSolveIT Conference 2023 (Athens, Greece)** — Poster: “Evolving the eNanoMapper Ontology,” on enhancements to the ontology for nanoinformatics data management.
- **BioSB 2023 (Egmond aan Zee, Netherlands)** — Poster: the eNanoMapper ontology v9.0 and its move to automated, GitHub Actions-based builds. · 10.5281/zenodo.8032202

Workshops

- **ELIXIR Toxicology Community Workshop 2024 (Utrecht, Netherlands)** — co-authored the workshop report on FAIRification of toxicological research output. · Preprint 2024
- **Ontologies for Advanced Materials (Joint workshop of the ELIXIR Tox Community / PINK project) 2024 (Brussels, Belgium)** — Talk: “eNanoMapper Ontology: Updated curation and development workflows,” on harmonizing materials and biological ontologies for safe-and-sustainable-by-design data.
- **Ontologies4Chem Workshop 2023 (Online)** — Talk: the eNanoMapper ontology’s restructured curation, testing, and build workflows, and upcoming CHEMINF updates. · 10.5281/zenodo.10218107 · video

Hackathons

- **ELIXIR BioHackathon Europe 2025 (Berlin, Germany)** — Co-lead: shape-driven knowledge graph integration: a registry of RDF dataset schemas built with rdfsolve (Health-RI news). · Participation as project co-lead funded by ELIXIR
- **DBCLS BioHackathon 2025 (Mie, Japan)** — developed RDF schema tooling in rdfsolve.” · Preprint 2025 · Personally invited; participation funded by Database Center for Life Science (DBCLS)
- **ELIXIR BioHackathon Europe 2024 (Barcelona, Spain)** — worked on pyBioDataFuse and was first author of the report on graph-schema discovery and transformation for database integration, the starting point for rdfsolve. · Preprint 2025 · Invited; participation funded by ELIXIR

Other events

- **ELIXIR CZ–NL Staff Exchange 2026 (Brno, Czech Republic)** — Visiting researcher: visited RECETOX, Masaryk University, to bridge computational toxicology and metabolomics: AOP-framework technologies, RDF schema, ontology and identifier alignment, and molecular pathways in untargeted metabolomics. · ELIXIR Staff Exchange grant, funded by ELIXIR
 - **Virtual Human Platform launch 2026 (Utrecht, Netherlands)** — Demo: showcased the AOP-Suite at the official launch of the VHP4Safety Virtual Human Platform.
-

TEACHING

Teaching contributions at Maastricht University:

Programming in the Life Sciences

Maastricht Science Programme · PRA3006

Programming with life sciences: use web services connecting biological data and visualize results with JavaScript.

Programming I

BSc Brain Science · BRAIN1005

Python fundamentals, data visualization, and computational models, linking programming exercises to mathematics and neuroscience.

Biomedical Challenges

MSc Biomedical Sciences · MBS1001

Supervised student groups during the execution of a research plan around Biomedical Challenges.

Bioinformatics tutorials

BSc Biomedical Sciences · BBS

Practical tutorials on BLAST sequence searches, biostatistics with R and SPSS, and biological databases for biomedical science students.

Problem-based learning (PBL) tutoring

Facilitate small-group learning, supporting students as they explore problems, formulate learning questions, and discuss findings from independent study. UM's PBL approach.

Teaching certifications

- Introductory Course on the Principles of Problem-Based Learning (PBL) — Maastricht University, 2023.
- Small Group Teacher Training (tutoring and coaching) — Maastricht University, 2023.

EDUCATION

MSc Bioinformatics

Wageningen University & Research

Sep 2019 – Nov 2021

Thesis: proposed genes in de novo sesquiterpene synthesis in *Pieris brassicae*.

BSc Biotechnology

Universitat de Lleida

2015 – 2019

Included an Erasmus exchange at Wageningen University & Research.

TECHNICAL SKILLS

- **Knowledge graphs and ontologies:** RDF, OWL, and SPARQL; SHACL, VoID, and LinkML for schema extraction and validation; ontology development with ROBOT and ODK; RDFLib, GraphDB, and SPARQL endpoints; MCP tools over RDF shapes.
- **Data integration:** Identifier and label reconciliation, SSSOM mappings, release comparison, and provenance tracking; pandas and Polars.
- **Python:** Libraries and command-line tools with typed APIs and Pydantic models, published on PyPI with Sphinx documentation.
- **Web services and interfaces:** Flask and FastAPI backends; Vue, JavaScript, and TypeScript frontends; Cytoscape.js network views.
- **Testing and delivery:** pytest, tox, Ruff, and mypy; GitHub Actions, GitLab CI, and Jenkins for tests, data updates, and releases; Docker and Slurm.

LANGUAGES

Spanish (native) · Catalan (native) · English (professional) · Dutch (B1-, certified by Maastricht University, Oct 2023)