Vincent Emonet

Knowledge graph developer

📞 +33 6 75 72 59 52
✉ vincent.emonet@gmail.com
🖥 vemonet.github.io
⊙ vemonet
in vincent-emonet
13/11/1991 - Driving License

Experiences

Since July 2018

Data science developer,
Institute of Data Science, Maastricht University, the Netherlands.
- Developed a cross-platform toolkit to sign and publish RDF Nanopublications using Rust, with bindings to Python, and JavaScript (compiled to WebAssembly). *vemonet.github.io/nanopub-rs*
- Developed a python library and documentation to configure and deploy self-hosted open source Large Language Models offline with API and UI. *vemonet.github.io/libre-chat*
- Developed a python library to easily deploy SPARQL endpoints using RDFLib. Helped to integrate it to bioregistry.io. *github.com/vemonet/rdflib-endpoint*
- Developed a system with a web service to index standard knowledge representations, such as OWL ontologies and SHACL shapes, from public git repositories, alongside with generating metadata for active SPARQL endpoints (Python, React/TypeScript). Presented as a demo at the Web Conference 2021. *index.semanticscience.org*
- Developed a web service and a python library to define, publish, and execute extensible FAIR evaluations of resources online (Python, React/TypeScript). Presented at the International Conference on FAIR Digital Objects in 2022. *fair-enough.semanticscience.org* and *maastrichtu-ids.github.io/fair-test*
- Developed a web service with an interface to annotate biomedical text assisted by Natural Language Processing, using standard identifiers, and the BioLink model developed for the NIH NCATS Biomedical Data Translator project. Annotated statements are published as RDF Nanopublications. *collaboratory.semanticscience.org*
- Developed an API, to serve drug-disease predictions models, and expose the predictions using standards developed for the NIH NCATS Biomedical Data Translator project (Python). *openpredict.semanticscience.org*
- Developed a python library and documentation to assist with building knowledge graphs, by mapping structured data to RDF (Python, RML mappings). Published at the SWAT4LS conference. *d2s.semanticscience.org*
- Developed the Data Science Research Infrastructure at Maastricht University, an OpenShift cluster for academic researchers and students to deploy Data Science services and workspaces. Managed the cluster with its documentation, and defined the deployment of workspaces for data science on CPU and GPU, such as JupyterHub, RStudio, VSCode server (Kubernetes, Docker). *dsri.maastrichtuniversity.nl*
- Deployed and maintained numerous services in production, such as *bio2rdf.org* and *w3id.org/AmIFAIR* (Linux).
- Teaching assistant for the relational database, and semantic web course, at Bachelor level, and for the building and mining knowledge graph course at Master level, in the Department of Advanced Computing Sciences. Helped students in class, wrote and presented some parts of the courses and practical activities (e.g. SHACL, RML).

October 2017 to June 2018

Conversational design engineer,
Match Group, France.
- Involved in the development and tests of a Natural Language Understanding (NLU) tool to build chatbots.
- Designed conversational patterns using Google DialogFlow tool.
- Analyzed conversations flow to improve user engagement using tools such as Google Analytics.
- Managed the multilingual aspect and translations in 6 languages of the NLU tool developed.
- Won the French 2018 Digital Innovation Award with the chatbot developed.

January 2015 to July 2017

Computer science design engineer - **Ontologies & Semantic Web**,
CNRS - Laboratory of Informatics, Robotics and Microelectronics of Montpellier (LIRMM), France.
- Administration and new features development on the *agroportal.lirmm.fr* and *bioportal.lirmm.fr* platforms, used for ontology storage, exploration and text annotation (Ruby on Rails, Java). *github.com/sifrproject*
- Contributed to, and maintained compatibility with the *bioportal.bioontology.org* developed at Stanford University.
- Built a web platform to integrate and expose Natural Language Processing programs developed in the ADVANSE team (Java, AngularJS). *advanse.lirmm.fr*
- Enhanced and packaged the YAM++ ontologies matching tool (Java). *gite.lirmm.fr/opendata*
- Developed a web interface for ontologies mappings validation (JEE). *yamplusplus.lirmm.fr*
- Set up a Slurm Workload Manager on a supercomputer for research at the LIRMM.

March 2014 to September 2014 & June 2013 to September 2013

Bioinformatics design engineer - **Biomedical data processing**,
Université de Laval, Québec, Canada.
- Developed web services to access biomedical data in the RDF format for the *bio2rdf.org* project. *github.com/bio2rdf/bio2rdf-rest*
- Developed components for the Talend ETL for the Semantic Web (Java). *github.com/vemonet/talend4sw*
- Converted biomedical databases, such as PubMed and neXtProt, to RDF format using the Talend ETL.
- Set up a Hortonworks Hadoop cluster for Big Data on retired hospital servers.

June 2012 to August 2012	**Intern in bioinformatics in an ecology research laboratory**, *University of California*, Berkeley, USA.

- Developed DNA sequence analysis programs to build phylogenetic trees for the Xylella fastidiosa bacteria, using the neighbor-joining method with Python 2.7.
- Set up an Amazon Web Service machine image for bioinformatics research, based on Linux.

Education

2012 - 2014	**Master degree, Physiology, biotechnology and informatics**, *University of Poitiers, France*, Biotechnology.
2011 - 2012	**Bachelor's degree, Bioinformatics**, *University of Poitiers, France*, Biotechnology specialization.
2009 - 2011	**Technical College Diploma, Bioengineering**, *Toulon, France*, Biology and biochemistry analysis.

Publications

Platforms
- Google Scholar: scholar.google.fr/citations?user=G59f3woAAAAJ
- HAL: hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Vincent+Emonet
- dblp: dblp.uni-trier.de/pers/hd/e/Emonet:Vincent

IDS
- V. Emonet, A. Malic, A. Zaveri, A. Grigoriu and M. Dumontier, **Data2Services: enabling automated conversion of data to services**, 2018.

LIRMM
- C. Jonquet et al., **AgroPortal: a vocabulary and ontology repository for agronomy**, *Computers and Electronics in Agriculture, vol. 144, p. 126-143*, 2018.

Bio2RDF
- M. Dumontier, A. Callahan, J. Cruz-Toledo, P. Ansell, V. Emonet, F. Belleau, A. Droit, **Bio2RDF Release 3: A larger, more connected network of Linked Data for the Life Sciences**. *International Semantic Web Conference*, 2014.

Skills

Languages Python, TypeScript (React, NextJS, Lit, vanilla JS), Rust, Java, Ruby on Rails, Bash, SQL, LaTeX.

Web SPARQL, RDF, ontologies (OWL, SKOS, OBO, UMLS), data schemas (SHACL, ShEx, JSON Schema), RESTful API, GraphQL, JSON, XML, HTML, CSS.

Tools Git, GitHub, GitLab, Linux, Docker, Kubernetes, Talend ETL, Dialogflow, and Agile methodology.

Biology Notions in biotechnology, genetics, biochemistry and physiology.

Languages

French Mother tongue.

English TOEIC Score 2014: 935. Lived 1 year in North America, and 5 years in the Netherlands. Active participation in discussion, reading and writing documents.

Spanish Basic notions for communication.

Interests

Sports Mountain biking, cyclotourism, hiking.

Open source enthusiast I enjoy discovering and using open source softwares, which includes reporting issues, and improving the documentation or implementation when possible.

References

Maastricht University
Michel Dumontier,
Professor of Data Science
(michel.dumontier
@maastrichtuniversity.nl)

LIRMM lab
Computer science researchers
- Clément Jonquet (jonquet@lirmm.fr)
- Zohra Bellahsene (bella@lirmm.fr)
- Jérôme Azé (jerome.aze@lirmm.fr)

Bio2RDF project
François Belleau,
bio2rdf.org project co-founder
(francois.belleau@gmail.com)