Emanuel Dima

Email: emanueldima@gmail.com   Mobile: +49 176 70463270
Info: emanueldima.github.io   GitHub: github.com/emanueldima

Address: 72555 Metzingen, Germany

Full stack developer with 15+ years of professional development experience in Python, Go, Java, C++ backend APIs and CLI apps and mainly JS/React web frontends; 3 years of experience as a team leader. Good general knowledge of many diverse technologies. Calm, open and friendly.

Skills

Python related: Python 2/3; Flask, Pyramid, FastAPI, Pydantic, SQLAlchemy; basic asyncio; Jinja templates;

Java related: Java 8/11/17; JDBC; Maven; Dropwizard; Apache Tomcat

Web technologies: HTML, CSS; JavaScript, React, Redux, Webpack

Web servers: Apache, Nginx; **Analytics**: Piwik/Matomo; **Payments**: Stripe

Other languages: Bash (scripting), C/C++, Go, Rust

Databases: SQL (PostgreSQL, MySQL, SQLite)

Data/configuration formats: JSON, YAML, TOML; **Xml technologies**: XML, XSLT, XPath; **Specification formats**: JSON Schema, RelaxNG, XML Schema

Version control: Git, Svn; **Project management**: GitHub, Jira

Virtualization & DevOps: Docker and Docker Swarm, BuildKit, Nomad; Vagrant; basic AWS and GCP with Terraform

Basic knowledge of: Swift, Clojure, Scala, C#, Erlang, Haskell, Lisp, Matlab, Pascal, Perl, Prolog; LaTeX; Node.js; XACML; Mustache templates; ML

Design Patterns; UML; REST; SOA/Microservices; Linux, networking

Education

Master degree in Distributed Systems (master thesis grade: 10), Computer Science Faculty, "Al. I. Cuza" University, Iași (2007 - 2009)

Bachelor of Science (bachelor paper grade: 10; average ECTS grade: 9.49) Computer Science Faculty, "Al. I. Cuza" University, Iași (2003 – 2007)
   - Erasmus scholarship, Granada University, Spain (oct. 2006 – feb. 2007)

Experience

SYSTEM ENGINEER, EXPLOSION.AI, REMOTE, 2023

- **Orbyt** (Machine Learning experiment execution framework), 2023, github.com/explosion/orbyt: I developed this project from scratch in Go as a CLI app. It packs an ML experiment into a Docker image and runs it multiple times with varying parameters either locally or remotely in a Nomad cluster with GPU nodes.

- I also developed a revenue generation site prototype for an open source project using Stripe, a local DB and a pypi compatible endpoint, deployed on AWS with Terraform. The code was forked from a similar project and adapted to new requirements.

SENIOR SOFTWARE DEVELOPER, FIGSHARE / DIGITAL SCIENCE REMOTE, 2022-2023

- **HGV** (public web registry of human genome variations), 2022, hgv.figshare.com: I rewrote the project in a SPA style using new technologies due to rising maintainability costs of the old version. The new backend was Python/Pyramid/SQLAlchemy while the frontend was JS with React.

- **FigShare** – other internal projects, including contributing to the dockerization of operations and updating the front-end of a Kibana installation.

ACADEMIC RESEARCHER / SOFTWARE DEVELOPER, UNIVERSITÄT TÜBINGEN; TÜBINGEN, 2010-2021

- **WebLicht** (web app for automatic linguistic annotations), 2013-2021, weblicht.sfs.uni-tuebingen.de: WebLicht is a tool for automatic annotations of text corpora, relying on an ecosystem of external services. I worked on maintenance and new features: design and implementation of a service chaining language, support of new linguistic data formats (TCF versions 4 and 5), service testing, coordination of changes across multiple institutions, Shibboleth configuration, etc.

- **Switchboard**, (web app for finding tools appropriate to given research data sets), 2019-2021, switchboard.clarin.eu, github.com/clarin-eric/switchboard: in 2019 I became the main developer, rewriting key modules for improved performance and ease of administration. I also added many new features: automatic recognition of file types, content inspection of data archives, seamless integration with other data repositories, UI improvements, a new tool specification format, etc.

- **Stratus** (on-premises hosting solution), 2019-2021: a system for deploying web services, based on Docker Swarm. I designed it as a simple transition path from an existing Rancher hosting solution with focus on simplicity and ease of maintenance.

- **Tündra** (web app for browsing text corpora), 2018-2020, weblicht.sfs.uni-tuebingen.de/Tundra: at first only involved in the UI design, I later became the project maintainer, fixing bugs and adding small features.

- **B2SHARE** (web app for publication of scientific data), 2014-2018, b2share.eudat.eu, github.com/EUDAT-B2SHARE/b2share: Technologies: Python, Flask web framework, Jinja templates, using the CERN's Invenio framework as backend; HTML/JS/React in the frontend. I started as developer and became team coordinator in 2015. I was involved in the design and implementation of the UI, authentication (using OAuth), the input of rich metadata (DublinCore, Marc21) with custom schemas, enforcing publication embargoes, allocation of unique identifiers (PIDs, DOIs), data versioning. As a team coordinator I was responsible for defining requirements, project management, roadmap, QA, deployments, as well as interactions with the upper management.

- **FCS Aggregator** (search interface for distributed text corpora), 2014-2015, contentsearch.clarin.eu: I implemented a new version of the required search protocol (based on SRU protocol), redesigned the UI and maintained the project for a short while.

- **Erdo** (web frontend of data repository), 2011-2015: a simple HTML/Javascript frontend with a Java backend communicating to a Fedora Commons server. It was used as the department's institutional repository.

- **service prototypes: GEF** (Generic Execution Framework), 2013-2018, github.com/EUDAT-GEF/GEF: Environment for execution of lightweight scientific workflows, provided as docker containers, at the site of the data. Useful for data filtering and data statistics. Technologies: Go, gorilla/mux, sqlite; **AAI delegation** (2019) delegation of access to a protected resource; **DEX** (2017-2018), a rule engine prototype based on Drools.

During this period I was actively involved in the following German, European and transatlantic research projects: EUDAT, EUDAT2020, CLARIN, CLARIAH, LAPPS-CLARIN.

SENIOR SOFTWARE DEVELOPER, BITDEFENDER; IAȘI, 2008-2010

- **Bitdefender 2010, Bitdefender 2011** (antivirus software): I worked on a custom user interface (a modular, plugin based system) and on an internal communication system that reflects the state of different modules

in different processes to the main user interface (Windows/C++/WinAPI, MFC, COM, HTMLayout)

RESEARCH ASSISTANT, "AL. I. CUZA" UNIVERSITY; IAŞI, 2008–2009

- **ALEAR** (European research project): I worked on a simulation concerning the emergence of anaphoric phenomena in artificial languages developed by software agents (Java)

- **InterOb** (European research project): I worked on the physical simulation and rendering (Java/Java3D) of a muscle-based facial animation engine.

SOFTWARE DEVELOPER, EMBARCADERO TECHNOLOGIES; IAŞI, 2007–2008

- **DSAuditor** (a network-based database auditing solution): worked on project maintenance, Linux/C, SQL

- **Performance Center** (a database performance monitoring solution): project maintenance, Windows/C++, SQL/WInAPI, MFC, COM

Publications

Claus Zinn, Emanuel Dima: The CLARIN Language Resource Switchboard: Current state, impact, and future roadmap. In Darja Fišer & Andreas Witt (eds.), CLARIN. The infrastructure for language resources. Berlin, deGruyter, 2022

Claus Zinn, Wei Qui, Marie Hinrichs, Emanuel Dima, Alexandr Chernov: Handling Big Data and Sensitive Data Using EUDAT's Generic Execution Framework and the WebLicht Workflow Engine. LREC 2018

Claus Zinn, Thorsten Trippel, Steve Kaminski, Emanuel Dima: Crosswalking from CMDI to Dublin Core and MARC 21. LREC 2016

Sarah Berenji Ardestani, Carl Johan Hakansson, Erwin Laure, Ilja Livenson, Pavel Stranák, Emanuel Dima, Dennis Blommesteijn, Mark van de Sanden: B2SHARE: An Open eScience Data Sharing Platform. e-Science 2015: 448-453

Chris Culy, Corina Dima, Emanuel Dima: Through the Looking Glass: Two Approaches to Visualizing Linguistic Syntax Trees. IV 2012: 214-219

Emanuel Dima, Verena Henrich, Erhard W. Hinrichs, Marie Hinrichs, Christina Hoppermann, Thorsten Trippel, Thomas Zastrow, Claus Zinn: A Repository for the Sustainable Management of Research Data. LREC 2012: 3586-3592

Emanuel Dima, Christina Hoppermann, Erhard W. Hinrichs, Thorsten Trippel, Claus Zinn: A Metadata Editor to Support the Description of Linguistic Resources. LREC 2012: 1061-1066

Emanuel Dima, Erhard Hinrichs, Marie Hinrichs, Alexander Kislev, Thorsten Trippel, Thomas Zastrow: Integration of Weblicht into the Clarin infrastructure." In Proceedings service-oriented architectures (SOAs) for the humanities: Solutions and impacts. Joint CLARIN-D/DARIAH Workshop at Digital Humanities Conference. 2012.

Dan Cristea, Emanuel Dima, Corina Dima: Why Would a Robot Make Use of Pronouns? An Evolutionary Investigation of the Emergence of Pronominal Anaphora. DAARC 2009: 1-14

Language Skills

Fluent: **English, Romanian** (native)

Basic: **German** (A2/B1 level)

References - on request