Remi Cura | PhD, Engineer

Remi.Cura@gmail.com | remicura@mit.edu
+1-(857)-413-1293

http://Remi.Cura.Info
ResearchGate   https://www.linkedin.com/in/remi-cura

SUMMARY

🚀Thrived in different fields (*databases*, *GIS*, *political science*), environments (*academia*, *industry*, *freelance*), and roles (*researcher*, *developer*, *project manager*, *supervisor*, *consultant*), bringing unconventional ideas and methods through communication. ⚙️ 7+ years of experience designing and leading development of prototypes for user-focused scalable solutions. 💡Wielded an exceptionally broad and deep scientific background (*information theory to social science*) to break new ground. Like to perform in cross-disciplinary team and share and discuss findings with other stakeholders. Experienced mentor and supervisor.

WORK EXPERIENCE

Post-doctorate Associate | Massachusetts Institute of Technology, Boston, USA   07/2018 – *present*

- **Architected and delivered complex and massive databases (DBs)** of money in politics in the USA (~*500 tables*, *100M rows*), which provide new insights into lobbying patterns, interest groups interactions with congress, lobbying and campaign financing coupling, and revolving door lobbyists. (**Data Model**) Invented a new design of interconnected DBs using a 3-layered approach for total transparency and interpretability. Incorporated advanced indexing to enable varied usages at scale (*composite keys*, *partial indexes*, *stats*). (**ETL**) Coded pipelines for dozens of data sources using classical (*scrapper*, *parser*, *API*) and advanced Postgres features (*Foreign Data Wrappers*, *xmltable*, *json path*).
  (**User access**) Tailored specific medium to access the data for different user profiles (*SQL, Python, R, Rest API, GraphQL API, Website, ready-made datasets*).

- **Disambiguated and linked data**. Authored state-of-the-art disambiguation pipelines to clean and link the tables and DBs. E.g., to automatically assess which lobbyists are the same person, generated a graph of lobbyists (nodes) connected by engineered similarity measures (edges). Used unsupervised clustering (*DBScan*) on this graph with fully integrated confidence from automated processing and user inputs to produce better consolidated data (*-10% entities*).

- **Unstructured text processing (NLP).** Starting from noisy free-form lobbyists declarations, extracted high quality lobbyist previous positions using old and new natural language processing methods (*regex*, *NLTK*), and advanced database logic (*nicknames*, *temporal distances*, *chamber*), producing new types of links (*4*).

- **Machine Learning projects. (Feature engineering on Lobbyists).** Engineered advanced features (based on text, statistics, graphs) to better qualify the effect of expertise and connection for lobbyists, strengthening the diversity (*x4*) and the premium effect estimation (*x3*).
  (**Automated ML**) Assembled a python library based on *scikit-learn* to quickly analyze and pick the appropriate supervised method for each linking problem, and assess data quality. Using caching and wrappers, expanded the library to transparently run outside and inside a database (*Pl/Python*).
  (**Labeling web app**). Directed a small development team to design a generic labeling app (*python + JS*, based on *plotly*) adapted to creating disambiguation ground truth. Completed *thousands of data points* collection.
  (**Lobbyists embedding**) Assembled a dataset of lobbyists colleagues (*7.8M*) to train a deep learning embedding (knowledge graph) to discover meaningful lobbyists distances and comparisons.

- **Advanced visualizations.** Factored the 80 original lobbying reports topics *to 18* with unsupervised network clustering (*successive K-means*). Initiated higher level data visualizations and analysis.

- **Project Management. (Financing)** Participated to 2 grants applications (obtained ~ *500k$, 2/2 success*). (**Supervising**) Recruited, interviewed, and mentored undergrade students (~*40*) from the MIT UROP program, went from low engagement levels to over *80%* of student fully involved, and *5 time increase* in applications.

(**Research**) Using the databases, generated many tailored datasets for *political scientists, economists, data scientists and physician*.

- **Web development. (Deployment)** Took ownership of an feature-rich website prototype (*search query engine, interactive visualizations, API with operators*), based on a complex stack (*ElasticSearch, Redis, Tornado, Angular, D3*). Rewrote data processing to solve performance issues, deployed on *AWS EC2 (Linux, command line)* from scratch. Won the IPES Best New Dataset Award. (**Redesign**) Guided an agile web development team to redesign from scratch the website, using a simplified and cutting-edge architecture (*Postgres, GraphQL and Rest API, React*) to attain a simplified and maintainable code base. Devised advanced authentication and role management system, and many advanced data processing methods for the visualizations, such as on-the-fly network data generation.

Freelance consulting and training | Oslandia, France   05/2017 – 01/2018

- Organized a *new training* on open source tools for point clouds processing and visualization (*3 days*). Customized the training to different clients (*energy, archeology*), with a consulting (*requirements, solution*) approach.

Post-doctorate associate, Paris School of Economics, Belle Epoque project | France   08/2016 – 02/2017

- **Designed a complete historical geocoding solution** as a geo-temporal recommendation engine *(<300ms)*, using fuzzy data modeling of time and addresses, tested on Paris (19th century, 10^6 *addresses*).

- **Engineered from scratch a full stack** (*Postgres, Geoserver, Rest API, Leaflet*) interactive web labeling app supporting collaborative editing with full handling of confidence and data quality control (*voting mechanism, ~5k edits*).

PhD candidate in industry | Thales TTS/IGN (MATIS, COGIT) | France   04/2013 – 09/2016

- **Authored a seminal work on** *street modeling and reconstruction* at city scale, seamlessly working in two research labs and a major defense company, **producing both high quality** *publications* (awarded) **and prototypes** (tested on real-life challenging data-sets).

- **Point Cloud feature engineering.** Engineered new low-level and robust geometric dimensionality feature based on octree occupation for visualization, unsupervised data exploration. Attained rough classification between vegetation (*94% prec., 92% recall*), ground, building and man-made objects.

- **Point Cloud Server.** Designed a full end-to-end point cloud management system. (**Storage**) Conceived a new data architecture based on storing compressed patches of points in a database (*pgpointcloud, 10^{10} points*). (**Level Of Details**) Invented a new LOD framework based on *octree occupation*, the *pseudo-random Halton sequence* (space filling curve), and a database. (**Visualization**) The LOD enabled visualization applications over 10^{10} *points*, such as by streaming the compressed point to a *WebGL* browser frontend, creating a lens for points in QGIS, and exporting rasters. (**Filtering**) Originated a complex indexing strategy (*R-Tree, GIN*) for rough filtering of patches of points to unlock *<100ms* advanced search (*spatio-temporal, semantic, road names, …*). (**Processing**) Performed object detections, image processing, and reconstruction methods directly within the database using *Pl/R* and *Pl/Python*, with easy re-use and parallelization.

- **Street Generation (Design).** Designed and implemented a full pipeline (*StreetGen*) to create coherent street models (*geometry, topology, semantic*) at scale in a robust way, using geometric boolean operations and Postgres to handle concurrency. (**Street models generation**). Demonstrated robustness and universality by generating a *full model of Paris in 20 minutes*, 3 other cities (*including in Africa*), a F1 racetrack and an *airport*. (**Traffic simulation**). Generated all the inputs (*~10k*) necessary for a hybrid *micro-meso traffic simulator* (*Simuvia*), including topologically correct street network and street lanes, trajectories in intersections and max cornering speeds, with *StreetGen* and *StreetUI*.

- **In-base interactions.** Invented a *world 1st* new UI framework (*StreetUI*) to handle interactions and processing within a database, allowing users to leverage any GIS software (*desktop, web*) to work. Spearheaded topology-preserving *GIS network editor*, visual collaborative cues for concurrent edit of GIS data enabling collaboration, tracking and quality analysis. Shaped a full UI for *Streetgen* that demonstrated new capabilities (*street model of 1 neighborhood in 1 hour*).

- **Street model optimization.** Programmed a software leveraging *Ceres C++ solver* to fit a generic street model to real world detections of urban feature (*markings*, *objects*, *sidewalks*) (*~million*). Found robust solutions in a scalable way (*~100km/min*) utilising custom regularization, jacobians, costs, and robust estimator to Seamlessly integrated confidence, uncertainty, and user inputs for optimal quality of results (*~0.1m*).

Lidar Engineer | IGN, France   10/2012 – 04/2013

- **Fine-tuned Lidar calibration**. Corrected global calibration and drafted fiber-by-fiber adjustments on *Velodyn* (*~5 cm)*. On *Riegl*, prototyped reflectance specularity compensation for street markings and qualified rare geometric and intensity errors happening on boundary of metal objects.

- **Increased point cloud integration with GIS**. Filtered quickly and conservatively outliers (using *trajectory* and *attributes*) for major reduction of octree volume and bounding box (*~4 levels*). Automated discovery of *points duplication* and *density* variation. Generated satellite-like images by fusing *Riegl* and *Velodyn* Lidar to limit occlusion.

- **Localization (Rough).** Using available traffic signs dataset, suggested an advanced indexing scheme for fast, scalable rough localization directly within a *Postgres* database. **(Precise)** Conceptualized scalable precise localization using roadway textures (*geometry* and *intensity gradients*) and image retrieval. **(Dual)** Demonstrated a pipeline to reconstruct traffic signs from individual raw signs detection.

Engineer-intern | Thales TTS, France   04/2012 – 10/2012

- **Qualified open source map layers for production (tree, building, signs...).** Performed analysis and discovery on *~20* layers and *~10^7* objects (coverage, quality, attributes). Delivered a normalized/cleaned version. Coded unsupervised ML tools to detect and correct hard-coded symbols (*R + SQL*, embedded in DB), and data mining (*association rules*) to check for patterns in the data.

- **Overhauled the backbone of a virtual world generation pipeline** to switch from file-based to DB-based. Reached majors increase in scalability (*10x faster*) and allowed distributed work (*internationally*). After development (*2 years)*, production integrated the prototype.

Engineer-intern | Ozone, France   07/2010 – 08/2010

- **Analyzed and formalized** the core business processes of a small company. Defined requirements, then prototyped a solution for Process automation by selecting and customizing an enterprise software (CiviCRM). The proposed solution was fully fleshed out and used in production.

SKILLS

Computer skills: SQL, PL/SQL, Python (numpy, scikit-learn), C/C++, Git, PostGIS
Soft skills: Supervising/mentoring (~40 students, 3 interns), international projects, organizer of social events at work
Languages: French (native), English (fluent), Spanish and German (basic conversations)

EDUCATION

Certified DBA: Postgres Associate and Postgres Professional (EDB)   09/2019
Certified Professional Cabinet Maker: CAP ebenisterie   10/2017
PhD in Computer Science (geographical information) – Universite Paris-Est, France   04/2013 – 09/2016
Master of Computer Science   (Engineer degree) - Telecom ParisTech   2007 – 2012

ACTIVITIES

Woodworking, *design* , trekking (*one year around the world 2011,* White Mountains (NH) AT section 06/2019) , mountaineering, Cooking.

PUBLICATIONS AND CONFERENCES

International Journal (peer reviewed):

> Cura, R., J. Perret, N. Paparoditis (2017): A scalable and multi-purpose point cloud server (PCS) for easier and faster point cloud data management and processing, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 127, pp.39-56
>
> Cura, R., B. Dumenieu, N. Abadie, B. Costes, J. Perret, M. Gribaudi (2018): Historical collaborative geocoding, MDPI IJGI (to be published).

International conferences (peer reviewed):

> (Best student paper award, session GeoBigData)
> Cura, R., J. Perret, N. Paparoditis (2015): Point Cloud Server (PCS): point clouds in-base management and processing, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W5, pp.531--539. ISPRS Geospatial Week 2015, La grande Motte, France.
>
> Cura, R., J. Perret, N. Paparoditis (2015): StreetGen: In-base procedural-based road generation, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W5, pp.409--416. ISPRS Geospatial Week 2015, La grande Motte, France.

Invited speaker:

> Cura, R. (2020) Postgres and Social Science, how we make it work. Postgres Vision, 2020, Boston.
> Cura, R. (2014) A PostgreSQL Server for Point Cloud Storage and Processing. PgDays 2014, Paris.

Professional exhibition:

> Stand Terra Mobilita, Future en Seine Festival. 2015, Paris

Open Access publication (not peer reviewed):

> Cura, R., J. Perret, N. Paparoditis (2016): Implicit LOD using points ordering for processing and visualisation in Point Cloud Servers, arxiv.org
> Cura, R., J. Perret, N. Paparoditis (2017): An octree cells occupancy geometric dimensionality descriptor for massive on-server point cloud visualisation and classification, arxiv.org
> Cura, R., J. Perret, N. Paparoditis (2017): StreetGen: In base city scale procedural generation of streets: road network, road surface and street objects, arxiv.org
> Cura, R., J. Perret, N. Paparoditis (2017): Interactive in-base street model edit: how common GIS software and a database can serve as a custom Graphical User Interface, arxiv.org
> Cura, R., J. Perret, N. Paparoditis (2017): User assisted and automatic inverse procedural road modelling at the city scale, arxiv.org
> Cura, R., J. Perret, N. Paparoditis (2017): A state of the art of urban reconstruction: street, street network, vegetation, urban feature, arxiv.org

PhD Thesis (peer reviewed): reviewers: Prof. Christian Heipke (L. U. Hannover), Prof. Peter Van Oosterom(T. U. Delft)

> Cura Remi (2016): Inverse procedural Street Modelling: from interactive to automatic reconstruction