Herminio García González, PhD

Data integration specialist

Kazerne Dossin
Goswin de Stassartstraat 153
2800 Mechelen
Belgium

🌐 https://herminiogarcia.com
@ herminio.garciagonzalez@kazernedossin.eu
📞 +32 (0) 15 28 86 52

Work experience

2021 – Present	**Data integration specialist** *Kazerne Dossin* Inside Kazerne Dossin (KD), I mainly work for the European Research Holocaust Infrastructure (EHRI) project in two parallel projects: the EHRI-3 taking care that new data from different partners and associated institutions is integrated seamlessly in the EHRI portal and EHRI-IP securing and advancing EHRI's technological stack for its future phase as an established European Research Infrastructure Consortium (ERIC). I also collaborate with KD to ensure that their digital operations run as expected and are future-proof. In addition, I pursue my scientific interests to give newer solutions that could enhance EHRI's and KD's technological ecosystems in the field of the Digital Humanities.
2020 – 2021	**Software Analyst** *IT and Communications Service, University of Oviedo* I started to work in the service from a competitive exam call. In the service we develop and maintain the whole academic system that supports the teaching and administrative procedures in the University of Oviedo. We give service to around 21000 students and 2000 professors. Here, I acquired a more industrial vision of my profession in a production and high-demanding environment. Therefore, I got to know and understand what the industry needs and how the scientific world can contribute to it.
2016 – 2020	**PhD Student** *Department of Computer Science, University of Oviedo* As part of my PhD student journey I prepared and applied for competitive calls in order to get fund for my PhD project. This allowed me to get knowledge on scientific project proposal and I obtained two grants to perform my thesis. Also during my PhD

thesis period I dedicated time for teaching in the Software Architecture course and the supervision of a final degree project.

2014 – 2020	**Developer**

University of Oviedo Foundation, WESO research group

I started to work as a developer in the WESO research group in many knowledge transference projects in which we delivered semantic solutions for industry necessities. Here, I got a vision of semantic technologies and how they can solve industrial problems.

Education

2016 – 2021	**PhD in Computer Science**

University of Oviedo

I focused my PhD thesis in the Semantic Web field and how it can contribute to the integration of heterogenous data sources. I also studied how this integration can be performed not only for data but also for schemas, providing a full integration environment in both levels. During this time, I also applied these experiences to the Digital Humanities and e-Learning fields.

2014 – 2016	**MSc in Web Engineering**

University of Oviedo

During my studies at the Master level, I had the opportunity to acquire a whole set of skills and technological knowledge of the most recent tendencies in web development. Besides, I also started to learn about the scientific process which culminated in my master thesis published as a journal paper "Enhancing e-learning content by using semantic web technologies".

2010 – 2014	**Bachelor's degree in Software Engineering**

University of Oviedo

In my Bachelor's 4-year studies I acquired all the basis for technological and engineering mind-set. During this epoch, I performed many works in groups which developed in me a collaborative spirit. In the last year, I started to collaborate with WESO research group in knowledge transference projects with the Semantic Web as the main topic.

Languages

Language	Reading	Speaking	Writing
Spanish	C2 (Mother tongue)	C2 (Mother tongue)	C2 (Mother tongue)
English	C1	C1	C1
French	B1	B1	B1

Publications

Journals

- Ruiz-Saavedra, S., García-González, H., Arboleya, S., Salazar, N., Labra-Gayo, J. E., Díaz, I., ... & de Los Reyes-Gavilán, C. G. (2021). Intestinal microbiota alterations

by dietary exposure to chemicals from food cooking and processing. Application of data science for risk prediction. *Computational and Structural Biotechnology Journal*, 19, 1081.

- García-González, H., Boneva, I., Staworko, S., Labra-Gayo, J. E., & Lovelle, J. M. C. (2020). ShExML: improving the usability of heterogeneous data mapping languages for first-time users. *PeerJ Computer Science*, *6*, e318.
- Garcia-Gonzalez, H., & Labra-Gayo, J. E. (2020). XMLSchema2ShEx: Converting XML validation to RDF validation. *Semantic Web*, 11(2), 235-253.
- García-González, H., Fernández-Álvarez, D., Labra-Gayo, J. E., & Ordóñez de Pablos, P. (2019). Applying big data and stream processing to the real estate domain. *Behaviour & Information Technology*, *38*(9), 950-958.
- García-González, H., Gayo, J. E. L., & Paule-Ruiz, M. (2016). Enhancing e-learning content by using semantic web technologies. IEEE Transactions on Learning Technologies, 10(4), 544-550.

Conferences

- Dermentzi, M., Bryant, M., Rovigo, F., García-González, H. (2024, June). Multilingual Automated Subject Indexing: a comparative study of LLMs vs alternative approaches in the context of the EHRI project. DH Benelux 2024, June 5-7, 2024, Leuven, Belgium
- García-González, H., Bryant, M. (2023, November). The Holocaust Archival Material Knowledge Graph. ISWC 2023: The 22nd International Semantic Web Conference, In-Use Track, (pp. 362-379), November 6-10, 2023, Athens, Greece. Cham: Springer Nature Switzerland.
- García-González, H., Dimou, A. (2022, September). Why to tie to a single data mapping language? Enabling a transformation from ShExML to RML. In Proceedings of Poster and Demo Track and Workshop Track of the 18th International Conference on Semantic Systems co-located with 18th International Conference on Semantic Systems (SEMANTiCS 2022), Vienna, Austria (Vol. 3235, pp. paper-11).
- García-González, H. (2021, June). A ShExML perspective on mapping challenges: already solved ones, language modifications and future required actions. In Proceedings of the 2nd International Workshop on Knowledge Graph Construction co-located with 18th Extended Semantic Web Conference (ESWC 2021), Online (paper 1).
- García-González, H. (2021, June). Everything for the users, nothing by the users: Lessons learnt from a heterogeneous data mapping languages user study. In Proceedings of the 2nd International Workshop on Knowledge Graph Construction co-located with 18th Extended Semantic Web Conference (ESWC 2021), Online (paper 2).
- García-González, H., Albarrán-Fernández, E., & Emilio, J. (2020, June). Converting Asturian Notaries Public deeds to Linked Data using TEI and ShExML. In Proceedings of the Third Workshop on Humanities in the Semantic Web (WHiSe 2020) co-located with 15th Extended Semantic Web Conference (ESWC 2020), Heraklion, Greece (pp. 41-46).
- Sanchez-Doria, M., Gayo, J. E. L., Fernandez-Alvarez, D., & Garcia-Gonzalez, H. (2019, June). Testing the usability of SNOMED CT terms suggestion in medical

report. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 672-677). IEEE.

- Garcia-Gonzalez, H., Fernandez-Alvarez, D., & Gayo, J. E. L. (2018). ShExML: An Heterogeneous Data Mapping Language based on ShEx. In *EKAW (Posters & Demos)* (pp. 9-12).
- Fernández-Álvarez, D., García-González, H., Frey, J., Hellmann, S., & Gayo, J. E. L. (2018, October). Inference of Latent Shape Expressions Associated to DBpedia Ontology. In *International Semantic Web Conference (P&D/Industry/BlueSky)*.
- Labra Gayo, J. E., Fernández-Álvarez, D., & García-González, H. (2018, October). RDFShape: An RDF playground based on Shapes. In *International Semantic Web Conference (P&D/Industry/BlueSky)*.
- Fernández-Alvarez, D., Labra-Gayo, J. E., & García-González, H. (2016). Inference and serialization of latent graph schemata using ShEx.

Book chapters

- Labra-Gayo, J. E., García-González, H., Fernández-Alvarez, D., & Prud'hommeaux, E. (2019). Challenges in RDF validation. In *Current Trends in Semantic Web Technologies: Theory and Practice* (pp. 121-151). Springer, Cham.

Thesis

- González, H. G. (2020). Semantic integration of heterogeneous big data sources. PhD Thesis dissertation, University of Oviedo, March 2021. Highest mark and cum laude mention.

Grants

- **Severo Ochoa PhD grant (2018)**
  Grant awarded by the Principality of Asturias government under a competitive call to perform a doctoral thesis project with a duration of 4 years.
- **University of Oviedo PhD bridge grant (2017)**
  Grant awarded by the Vice-rectorate for Research of the University of Oviedo under a competitive call to start a PhD thesis project and to support researcher activities while applying to a full-time PhD grant.

Awards

- **University of Oviedo's Extraordinary Doctorate Award**
  Prize awarded to most innovative PhD thesis during an academic year. They are awarded on the basis of having the unanimous nomination of the PhD VIVA committee and on a competitive basis, awarding a maximum of 1 thesis for every 5 defended thesis in the university.

Teaching experience

Teaching

- **Software Architecture course in the Bachelor's Degree in Software Engineering (120h) (2017-2018) - *University of Oviedo***

During two consecutive academic years I was in charge of some practical lessons. In these practical lessons we taught students about groundbreaking technologies and architectural patterns and guide them through the construction of a software architecture in which they have to work in groups to achieve a functional architecture that not only holds in theory but also in practice.

Supervision of final year projects

- **Improving Research Access for Holocaust Scholars: Metadata Integration in the European Holocaust Research Infrastructure (EHRI) Portal by Suzanne Swartz (2024) Mark: 17 (out of 20) -** *KU Leuven*
  I was the mentor for the internship that Suzanne carried out in Kazerne Dossin in which she contributed to the data integration in the EHRI Portal and developed a test case for the integration of archival descriptions from the FranceArchives site.
- **Using Linked Data to Support Holocaust Research Building a Knowledge Graph for Metadata in the Kazerne Dossin Collections by Tereza Podhajská (2024) Mark: 17 (out of 20) -** *KU Leuven*
  I was the mentor for the internship that Tereza carried out in Kazerne Dossin in which she developed a conversion of the Holocaust victim's data, spread in different Access databases, to a cohesive Knowledge Graph in RDF.
- **Web portal about the Spanish constitution by Santiago Marqués Riviera (2018) Mark: 8,5 (out of 10) -** *University of Oviedo*
  I was the co-supervisor of Santiago's final degree project in which he developed a portal to access the contents and history of the Spanish constitution in the year of its 40th anniversary.

Selected Open Source Projects

- **ShExML**
  https://github.com/herminiogg/ShExML
  ShExML is a language based on ShEx to map and merge heterogeneous data sources. It is designed with usability in mind trying to make the script creation easier to the users.
- **DMAOG**
  https://github.com/herminiogg/dmaog
  DMAOG is a library that allows to use data mapping rules or RDF data to automatically generate code to access the generated data generated, decreasing the amount of code written by developers.
- **DataIntegrationLabResources**
  https://github.com/EHRI/DataIntegrationLabResources
  This repository holds the resources used for integrating the data from different archives in the EHRI Portal during the duration of the EHRI-3 project.
- **EHRI2LOD**
  https://github.com/herminiogg/EHRI2LOD
  EHRI2LOD contains all the mapping rules and the orchestration scripts used to generate the EHRI-KG which is described in the publication "The Holocaust Archival Material Knowledge Graph".

- **XMLSchema2ShEx**
  https://github.com/herminiogg/XMLSchema2ShEx
  XMLSchema2ShEx is a library that allows to convert a subset of XMLSchema specification elements to Shape Expressions (ShEx). The implementation of mentioned mappings is based on the homonymous publication.

You can find more repositories in which I am working or I have been working in my Github profile: https://github.com/herminiogg/

International activity

Invited talks

- **Intro and practical session on Linked Open Data**
  EHRI API & Innovation Workshop, London, 15th March 2023
- **The EHRI portal, from data aggregator to central hub: how it works now and how it should evolve to cope with the maturity challenges**
  International Memorial's Database Conference, Mechelen, March 2022
- **The EHRI (Mobile) Data Integration Lab: what it is, what it solves and which the challenges are**
  Sustainable Publishing of Metadata – DARIAH Working Group Open Meetings 2021 (https://metadata.hypotheses.org/80)

Collaborative groups

- **W3C Knowledge Graph Construction Community Group**
  https://www.w3.org/community/kg-construct/
  The group goal is to provide better ways to construct Knowledge Graphs studying and discussing existing tools and use cases. The main community work output is about new methods, guidelines, resources, and better practises around the Knowledge Graphs building processes.

Research stays

- **INRIA Lille Nord Europe, Villeneuve d'Ascq, Nord – Pas-de-Calais, France**
  I performed a research stay of 3 months as a part of a collaboration with Iovka Boneva and Sławek Staworko with the aim to develop the semantics of the ShExML language.

Reviewer tasks and Membership of Program Committees

- PeerJ Computer Science in 2023 and 2024
- Program Committee member of the Posters & Demos track in SEMANTiCS 2022, 2023 and 2024
- Program Committee member of the CBMS 2023 and 2024
- Program Committee member of 4th International Workshop on Knowledge Graph Construction, co-located with ESWC in 2021, 2022, 2023 and 2024
- Reviewer of the Semantic Web Journal in 2022, 2021, 2020 and 2019
- Reviewer of the Poster & Demo track in ESWC 2021
- Sub-reviewer in ISWC 2020
- Reviewer of the Behaviour & Information Technology journal in 2018

Projects participation

- **European Research Holocaust Infrastructure Implementation Phase (EHRI-IP)**
  14 partners from 13 countries – from February 2024 until present
  Funded by the European Commission under the call HORIZON-INFRA-2023-DEV-01, with grant agreement ID 101129732.
- **European Research Holocaust Infrastructure Phase 3 (EHRI-3)**
  26 partners from 17 countries – from September 2021 until present
  Funded by the European Commission under the call H2020-INFRAIA-2018-2020, with grant agreement ID 871111 and DOI 10.3030/871111.
- **Implementing the Research Software MetaData (RSMD) guidelines**
  From May 2024 until October 2024
  Funded by the FAIR-IMPACT project through a cascading grant.
- **FAIR-enabling repositories**
  From May 2024 until October 2024
  Funded by the FAIR-IMPACT project through a cascading grant.
- **SIGINFO: Information for the assets sale**
  Izertis S.L. and WESO research group – 1 year starting on June 2016
- **LODLearning:Enriching the learning experience in eLearning**
  Izertis S.L. and WESO research group – 6 months starting on April 2015
- **Data treatment for the A4AI (Alliance for Affordable Internet)**
  The World Wide Foundation and WESO research group – 2 months starting on February 2015
- **Rebuilding IFAD's land portal**
  SBC consulting and WESO research group – 6 months starting January 2014

Dissemination

- ***"If this Portal could talk..."| EHRI's Veerle and Herminio Talk About Data Identification and Integration***
  Interview on the EHRI project website (https://www.ehri-project.eu/veerle-and-herminio-talk-about-EHRI-Portal)
- ***What can I do with this messy spreadsheet?* Converting from Excel sheets to fully compliant EAD-XML files**
  EHRI Document Blog (https://blog.ehri-project.eu/2022/04/25/converting-from-excel-to-ead-xml/)
- ***Creating a To Do list app with React and Solid***
  Personal webpage (https://herminiogarcia.com/blog/2020/02/05/todo-list-react-solid/)

Courses

- **Management and treatment of data for research: Introduction to the statistical analysis (Level 2)**
  University of Oviedo – 20 hours – January 2017
- **Management and treatment of data for research: Introduction to the statistical analysis (Level 1)**
  University of Oviedo – 20 hours – January 2017

- **Workshop about the elaboration and evaluation of research projects in the IT and digital communications field**
  University of Oviedo – 40 hours – November 2016
- **Parallel Programming in Scala**
  Coursera Inc. – June 2016
- **Machine Learning**
  Coursera Inc. – Verified certificate – May 2016
- **Principles of Reactive Programming**
  Coursera Inc. – Verified certificate with distinction – May 2015
- **Functional Programming Principles in Scala**
  Coursera Inc. – Verified certificate with distinction – June 2014