

Curriculum Vitae
Helena Galhardas, PhD
March 2025

1. Private Data

Born 9/10/67 in Lisbon
Portuguese citizenship
Married
Currently living in Lisbon
<http://web.ist.utl.pt/helena.galhardas>

2. Education

2001 PhD in Informatics from University of Versailles.
1994 M.Sc. in Electrical and Computers Engineering from *Instituto Superior Técnico (IST)*, Technical University of Lisbon.
1990 BsC (5 years) in Electrical and Computers Engineering from *Instituto Superior Técnico*, Technical University of Lisbon.

3. Professional Activities

2018-now Associate Professor at *IST*, University of Lisbon (UL) and Researcher at INESC-ID.
2001-2018 Assistant Professor at *IST*, University of Lisbon (UL) and Researcher at INESC-ID.
1997-2001 On leave of absence from IST.
1995-96 Teaching Assistant at IST and Researcher at INESC - Software Engineering Group.
1991-94 Teaching assistant at IST and Researcher at INESC - Information Systems Group.
1991-94 Scholar of INESC - Communication Software Development Techniques Group.

4. Main research interests

- Data Transformation and Cleaning
- ETL (Extraction, Transformation and Loading)
- Data Integration
- Data Profiling
- Information Extraction

5. Publications

5.1. Journals

- J.L.M. Pereira, M. J. Fonseca, A. Lopes, H. Galhardas: Cleenex: Support for User Involvement during an Iterative Data Cleaning Process, *ACM J. Data Inf. Qual.* 16(1): 6:1-6:26, 2024.
- T. Bouganim, H. Galhardas and I. Manolescu, Efficiently Identifying Disguised Missing Values in Heterogeneous, Text-Rich Data, *Transactions on Large-Scale Data- and Knowledge-Centered Systems vol. 13410*, Oct. 2022, Lecture Notes in Computer Science, Springer.
- J.P.L. Pereira, J. Casanova, H. Galhardas and D. Shasha, AcX: system, techniques, and experiments for Acronym eXpansion, *Proceedings of the VLDB Endowment*, Sep. 2022, VLDB Endowment
- L. P. Oliveira e Silva, M. Resende, H. Galhardas, V. Manquinho and I. Lynce, DeepData: Machine Learning in the Marine Ecosystems, *Expert Systems With Applications*, Jun. 2022
- A.C. Anadiotis, O. Balalau, C. Conceição, H. Galhardas, M.Y. Haddad, I. Manolescu, T. Merabti and J. You, Graph integration of structured, semistructured and unstructured data for

- data journalism, *Information Systems* vol. 104, Feb. 2022, Elsevier
- O. Balalau, A.C. Anadiotis, F. Chimienti, T. Bouganim, H. Galhardas, M.Y. Haddad, S. Horel, I. Manolescu and Y. Youssef, Empowering Investigative Journalism with Graph-based Heterogeneous Data Management, *IEEE Data Engineering Bulletin* vol. 44 (3) pp. 12-26, Mar. 2021, IEEE.
 - P. Kathiravelu, A. Sharma, H. Galhardas, P. Van Roy and L. Veiga, On-demand big data integration, *Distributed and Parallel Databases* vol. 37 (2) pp. 273–295, Jun. 2019, Springer.
 - G. F. Simões, H. Galhardas and L. Gravano, When speed has a price: Fast information extraction using approximate algorithms, *Proceedings of VLDB Endowment (PVLDB)*, vol 6, n. 13, 2013.
 - A. Wichert and P. Teixeira and P. Santos and H. Galhardas, Subspace tree: High dimensional multimedia indexing with logarithmic temporal complexity, *Journal of Intelligent Information Systems* 35(3), pp. 495-516, Nov. 2010, Springer.
 - A. Wichert and P. Santos and H. Galhardas, Panoramix: a content-based health record system, *International Journal of Computer Assisted Radiology and Surgery*, 4(), pp. 152, Jun. 2009, Springer.
 - P. Carreira, H. Galhardas, A. Lopes and J. D. Pereira, One-to-many data transformations through data mappers, *Data & Knowledge Engineering*, 62(3), pp. 483-503, Sep. 2007, Elsevier.
 - J. Barateiro and H. Galhardas, A survey of data quality tools, *Datenbank-Spektrum*, (14), pp. 15-21, Aug. 2005, dpunkt.verlag.
 - E. Cardoso, H. Galhardas, M.J. Trigueiros, A. Rito Silva: A Decision Support System for IST Academic Information, (Best Award paper) *EUNIS 2003, July 2003 and Informatica (Int. Journal of Computing and Informatics)*, Volume 27, Number 3, 2003

5.2. Book Chapters

- V. Pequeno, H. Galhardas and V. Vidal, Using Complex Correspondences for Integrating Relational Data Sources, chapter in *Enterprise Information Systems (Revised and Extended Papers ICEIS 2014)*, Sep. 2015, LNBI, Springer
- H. Galhardas, Data Cleaning and Transformation Using the AJAX Framework, chapter in *Generative and Transformational Techniques in Software Engineering*, Nov 2006, Springer Berlin / Heidelberg.
- M. Bouzeghoub, F. Fabret, H. Galhardas, M. Matulovic, J. Pereira and E. Simon: Data Warehouse Refreshment, in *Fundamentals of Data Warehousing*, (eds.) M. Jarke, M. Lenzerini, Y. Vassiliou and P. Vassiliadis, Springer-Verlag, 1999.

5.3. Edited Proceedings

- M. Herschel, H. Galhardas, B. Reinwald, I. Fundulaki, C. Binnig and Z. Kaoudi: *Advances in Database Technology - 22nd International Conference on Extending Database Technology, EDBT 2019*, Lisbon, Portugal, March 26-29, 2019. OpenProceedings.org 2019, ISBN 978-3-89318-081-3
- H. Galhardas and E. Rahm Ed., *Data Integration in the Life Sciences, 10th International Conference (DILS 2014)*, Lecture Notes on Bioinformatics, subseries of LNCS, no. 8574, Jul 2014, Springer

5.3. International Conferences

- O. Balalau, S. Ebel, H. Galhardas, T. Galizzi, I. Manolescu, STaR: Space and Time-aware Statistic Query Answering. *CIKM*: 5190-5194, 2024.
- T. Tiam-Lee, R. Henriques, J. Costa, V. Manquinho and H. Galhardas, Consolidation of massive medical emergency events with heterogeneous situational context data sources, *International workshop on Data Analytics solutions for Real-Life Applications @ EDBT/ICDT 2022*, Mar. 2022.

- A.C. Anadiotis, O. Balalau, T. Bouganim, F. Chimienti, H. Galhardas, M-Y Haddad, S. Horel, I. Manolescu and Y. Youssef, Discovering Conflicts of Interest across Heterogeneous Data Sources with ConnectionLens (demo paper), presented at *ACM International Conference on Information and Knowledge Management (CIKM 2021)*, Nov. 2021
- L. P. Oliveira e Silva, H. Galhardas, V. Manquinho and R. Henriques, UNIANO: robust and efficient anomaly consensus in time series sensitive to cross-correlated anomaly profiles, presented at *SIAM International Conference on Data Mining*, Apr. 2021
- J. M. P. Moreira, H. Galhardas and M. Pardal, LeanBench: comparing software stacks for batch and query processing of IoT data, *International Conference on Ambient Systems, Networks and Technologies (ANT)*, May 2018.
- J. P. M. Pereira, H. Galhardas and B. Martins, A Benchmark for Relation Extraction Kernels, *East European Conference on Advances in Databases and Information Systems (ADBIS)*, Sep. 2015, pp. 184-197, Springer.
- P. Kathiravelu, H. Galhardas and L. Veiga, dudu Multi-Tenanted Framework: Distributed Near Duplicate Detection for Big Data, *23st International Conference on Cooperative Information Systems (CoopIS 2015)*, Sep. 2015, LNCS, Springer-Verlag
- P. Barrio, G. F. Simões, H. Galhardas, L. Gravano, Learning to Rank Adaptively for Scalable Information Extraction, *International Conference on Extending Database Technology (EDBT)*, Mar. 2015
- P. Barrio, G. F. Simões, H. Galhardas, L. Gravano, REEL: A Relation Extraction Learning Framework (poster paper), *Digital Libraries 2014*, Sep. 2014.
- R. Ferrão, H. Galhardas and L. Coheur, A Machine Learning based Natural Language Interface for a database of medicines (poster paper), *International Conference on Data Integration in the Life Sciences (DILS)*, Jul. 2014
- V. Pequeno, V. Vidal, M. Casanova, L. Neto and H. Galhardas, Specifying Complex Correspondences between Relational Schemas and RDF models for generating customized R2RML mappings, *International Database Engineering & Applications Symposium (IDEAS)*, Jul. 2014.
- V. Pequeno and H. Galhardas, Specifying Complex Correspondences Between Relational Schemas in a Data Integration Environment, *International Conference on Enterprise Information Systems (ICEIS)*, Apr. 2014.
- A. Thanopoulou, P. J. Carreira and H. Galhardas, Benchmarking with TPC-H on On-the-shelf Hardware (poster paper), *International Conference on Enterprise Information Systems (ICEIS)*, Jun. 2012.
- H. Galhardas and A. Lopes and E. Santos, Support for User Involvement in Data Cleaning, *International Conference on Data Warehousing and Knowledge Discovery (DAWAK)*, Aug. 2011
- H. Galhardas and L. Santos Torres and J. Damásio, Master Data Management: a proof of concept, *International Conference on Information Quality (ICIQ)*, Nov. 2010
- P. Carreira, H. Galhardas, J. D. Pereira, A. Wichert, One-to-many data transformation operations - Optimization and Execution on an RDBMS, *International Conference on Enterprise Information Systems*, Jun. 2007.
- A. Silva, H. Galhardas, J. Barateiro, H. Matos, J. Gonçalves, E. Portela, Data Analysis Features of the GESTBARRAGENS System, *International Conference on Innovative Views of NET Technologies (IVNET)*, Oct. 2006
- P. Carreira, H. Galhardas, A. Lopes, J. Pereira: Extending Relational Algebra to express one-to-many data transformations, *Simpósio Brasileiro de Banco de Dados (SBBd)*, October 2005, Uberlândia, Brasil.
- P. Carreira, H. Galhardas, A. Lopes, J. Pereira: Data Mapper: An operator for expressing one-to-many data transformations, *International Conference on Data Warehousing and Knowledge Discovery (DAWAK)*, May. 2005, Copenhagen.
- P. Carreira, H. Galhardas: Efficient Development of Data Migration Transformations, *SIGMOD 2004 demo paper, Paris, France, June 2004*,
- H. Galhardas, D. Florescu, D. Shasha, E. Simon and C. Saita: Declarative Data Cleaning: Language, Model and Algorithms, *VLDB 2001, Rome Italy, September 2001*.

- H. Galhardas, D. Florescu, D. Shasha and E. Simon: Declaratively cleaning your data using AJAX , *BDA 2000, Blois France, October 2000*.
- H. Galhardas, D. Florescu, D. Shasha and E. Simon: AJAX: An Extensible Data Cleaning Tool, *SIGMOD 2000 demonstration paper, Dallas, May 2000*.
- A. R. Silva, H. Galhardas, P. Sousa, J. Silva, P. M. Sousa, "Designing Distributed Databases from an Organisational Perspective", 4th European Conference on Information Systems (ECIS'96), Lisbon, July 1996.
- N. Duarte, H. Galhardas, R. Carapuça, M. Marques, "Towards an Integrated SDL/CHILL environment", 5th SDL Forum, Glasgow, 1991, ed. North Holland.
- M. Gomes, F. Ferreira, M. Gamito, H. Galhardas, N. Duarte, L. Silva, A. Freitas, J. Abreu, J. Martins, "Editores Gráficos Textuais", 3^o Encontro de Computação Gráfica, EUROGRAPHICS, Coimbra, Portugal, October 1990 (in portuguese).

5.4. International Workshops

- J.P.L. Pereira, H. Galhardas and D. Shasha, Acronym Expander at SDU@AAAI-21: an Acronym Disambiguation Module, presented at *Proceedings of the Workshop on Scientific Document Understanding co-located with 35th AAAI Conference on Artificial Intelligence (SDU@AAAI 2021)*, Feb. 2021
- P. Alves and M. J. Fonseca and J.D. Pereira and H. Galhardas, CDRGen: A Clinical Data Registry Generator, presented at *International Workshop on Data Management and Analytics for Medicine and Healthcare (in conj. with VLDB)*, Sep. 2020
- P. Kathiravelu, Y. Chen, A. Sharma, H. Galhardas, P. Van Roy and L. Veiga, On-Demand Service-Based Big Data Integration: Optimized for Research Collaboration, *VLDB Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH)*, Aug. 2017, Springer.
- N. F. Lages, B. S. Caetano, M. J. Fonseca, J. D. Pereira, H. Galhardas and R. Farinha, Umedicine: A System for Clinical Practice Support and Data Analysis, *VLDB Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH)*, Aug. 2017, Springer.
- B. Martins, H. Galhardas and N. Gonçalves, Using Random Forests Classifiers to Detect Duplicate Gazetteer Records, *Ibero-American Workshop on Data Quality (in conjunction with CISTI 2012)*, Jun. 2012.
- H. Galhardas, A. Lopes and E. Santos, User involvement in cleaning applications, *Ibero-American Workshop on Data Quality (in conjunction with CISTI 2011)*, Jun. 2011, pp. 737-742.
- J. P. Bastos, G. F. Simões, L. Coheur and H. Galhardas, Medicine.Ask: A Facility to assist Doctors on prescribing Medicines (poster/demo paper), *International Workshop on Data Integration in the Life Sciences (DILS)*, Jul. 2009.
- P. Carreira, H. Galhardas, J. D. Pereira, F. Martins, M. J. Silva, On the performance of one-to-many data transformations, *International Workshop on Quality of Databases (QDB)*, in conj. with VLDB'07, Vienna, Sep. 2007.
- H. Galhardas, D. Florescu, D. Shasha, E. Simon and C. Saita: Improving Data Quality using a Data Lineage Facility, *DMDW 2001 workshop paper, Interlaken Switzerland, June 2001*
- H. Galhardas, E. Simon and A. Tomasic, "A Framework for Classifying Scientific Metadata", AAAI'98 Workshop on AI and Information Integration, Madison, July 1998.
- H. Galhardas, M. Silva, R. Carapuça, "SOAP: A Next-Generation Database System Experiment", 3rd ERCIM Database Research Group Workshop, Pisa, September 1992.

5.5. Thesis

- H. Galhardas, Data Cleaning: Model, Declarative Language, and Algorithms, *PhD Thesis, University of Versailles*, September 2001.

6. Posters and Talks

- H. Galhardas, Clean your data if you care about its quality, DELix - Lisbon Winter School on Data Science and Engineering, Feb. 2017.
- H. Galhardas, Speeding up Information Extraction Programs, INRIA Saclay, France, March 2015.
- B. Martins, H. Galhardas, D. Florescu, M. Zacharioudakis, S. Nasoi, A library of data cleaning functions for XQuery, *XML Prague poster*, 2011.
- H. Galhardas, D. Florescu: XML Data Cleaning, *talk at Oracle*, CA, July 2010.
- H. Galhardas: Data cleaning and transformation using the AJAX framework, *Technology presentation at Summer School on Generative and Transformational Techniques in Software Engineering (GTTSE)*, Portugal, July 2005.
- H. Galhardas: AJAX - An Extensible Data Cleaning Tool, *talk at IBM Almaden*, February 2000.
- H. Galhardas, D. Florescu, D. Shasha and E. Simon: An Extensible Framework for Data Cleaning, *ICDE'2000 poster paper*, February 2000.
- H. Galhardas: A Framework for Data Cleaning, *talk at Dagstuhl Seminar No. 99261 on Foundations of Information Integration*, June 1999.

7. Funded Projects

- PICASSO: Pediatric Palliative Care Support System, funded by FCT, (2025-26) - Principal Investigator.
- LAIfBlood: Artificial Intelligence for Blood Management, funded by FCT (2020-23) - participation
- Data2Help: Data Science for Optimization of Emergency Medical Services, funded by FCT (2019-22) - participation
- DataStorm: Large Large-Scale Data Management in Cloud Environments, funded by FCT (2013-16) - participation
- METANET4U, funded by European Commission (2011-2013) - participation
- XML data cleaning, funded by FLWOR Foundation (2011-12) - Portuguese coordinator
- SInteliGIS: Services for Intelligent Geographical Information Systems, funded by FCT (2011-13) – participation.
- gestBarragens funded by IAPMEI (2004-05) – participation.
- XClean: XML data cleaning, with INRIA, funded by Program PESSOA GRICES (2005-06) – Portuguese coordinator

8. Professional service

8.1. Co-chair

- DILS 2014 - International Conference on Data Integration in the Life Sciences, with Erhard Rahm.
- InFORUM 2011, 2015, 2017 - *Simpósio de Informática, Lisboa*, Track: *Gestão de Informação e Conhecimento*
- InFORUM 2010 - *Simpósio de Informática, Lisboa*, Track: *Gestão e Tratamento de Informação*.
- InFORUM 2009 - *Simpósio de Informática, Lisboa*, Track: *Processamento de Dados, Informação e Linguagem*.
- IQIS 2006 - International Workshop on Information Quality in Information Systems (in conj. with SIGMOD 2006).

8.2. General Chair - Conferences

- EDBT 2019 - International Conference on Extending Database Technology.

8.3. Program Committee Member - Conferences

- BDR 2022 - Big Data Research.
- BDA 2021 - Bases de Données Avancées.
- VLDB 2020 - International Conference on Very Large Databases.
- EDBT 2017, 2016, 2009 - International Conference on Extending Database Technology.
- DILS 2017, 2015 - International Conference on Data Integration in the Life Sciences.
- CBMS 2017, 2016, 2015 - International Symposium on Computer-Based Medical Systems.
- SIGMOD 2012, 2008, 2004 - ACM Special Interest Group on Management of Data.
- ICDE 2011 - International Conference on Data Engineering.
- ER 2011, 2009 - International Conference on Conceptual Modeling.
- CIKM 2008, 2007 - ACM Conference on Information and Knowledge Management.
- QUATIC 2004 - Quality: the bridge to the future in ICT.

8.4. Program Committee Member - Workshops

- DILS 2017 - International Conference on data Integration in the Life Sciences.
- InFORUM 2022, 2016, 2014, 2013, 2012 - *Simpósio de Informática, Lisboa*, Track: *Gestão de Informação e Conhecimento*.
- QDB 2016, 2011, 2010, 2009, 2008, 2007 - International Workshop on Quality in Databases (in conj. with VLDB).
- PIKM 2007 - PhD Workshop in CIKM, (in conj. with CIKM), 2007, Steering Committee member.
- CleanDB 2006 - International Workshop on Clean Databases (in conj. with VLDB), 2006.
- IQIS 2006, 2005, 2004 International Workshop on Information Quality in Information Systems (in conj. with ACM SIGMOD)
- DOLAP 2008, 2006 - ACM International Workshop on Data Warehousing and OLAP (in conj. with ACM CIKM).
- DIWeb 2004 - International Workshop on Data Integration over the Web (in conj. with CAISE).
- DataGadgets 2004 - Workshop on Bringing Up Emerging Solutions for Data Warehousing Systems (in conj. with JISBD).
- EKDB 2003 - International Workshop on Extraction of Knowledge from Databases (in conj. with EPIA).

8.5. Journal Reviewer

- JODS - Journal on Data Semantics, Springer, 2020.
- VLDBJ -Very Large Databases Journal, 2019, 2007, 2002.
- DAPD - Distributed and Parallel Databases, Springer, 2017, 2014.
- TODS - Transactions on Database Systems, ACM, 2017.
- JDIQ - Journal of Data and Information Quality, ACM, 2016, 2014, 2013, 2008.
- Information Systems, Special Issue on Data Quality, Elsevier, 2012, 2003.
- DKE - Data and Knowledge Engineering, 2010.
- Internet Computing, IEEE, 2009.
- TKDE - Transactions on Knowledge and Data Engineering, IEEE, 2014.
- TOIS - Transactions on Information Systems, ACM, 2003.
- Journal on Information Sciences, 2003.

8.6. Evaluator of Scientific Programs

- June/2014 - Evaluator of project proposals H2020, Brussels.

- Fev/Mar 2012 - Evaluator of project proposals FP7, Brussels and Luxembourg.
- April/May 2008, Fev 2011/2012 - Evaluator of NSERC Strategic Network Grant, Canada.

9. Teaching

- Data Administration and Information Systems, MEIC (MSc Degree in Engineering Systems and Computer Engineering) and METI (MSc Degree in Telecommunications and Informatics), IST/UL, 2015/16 - 2020/21.
- Data Analysis and Integration, MEIC, IST/UL, 2015/16 - 2019/20.
- Data Quality, PhD Degree Information Systems and Computer Engineering, IST/UL, 2008/09 - 2016/17.
- Databases, LEIC (Degree in Engineering Systems and Computer Engineering) and LETI (Degree in Telecommunications and Informatics), IST/UL, 2003/04-2016/17
- Information Management and Processing, MEIC, IST/UL, 2008/09-2014/15
- Database Administration and Optimization, MEIC and METI, IST/UL, 2008/09 - 2014/15
- Database Technology, LEIC and MEIC, IST/UL, 2005/06 - 2006/07
- Support Decision Systems, LEIC and MEIC, IST/UL, 2003/04 - 2008/09.

10. University Management

- Executive Director INESC-ID 2025-.
- Director INESC-ID 2024.
- Vice-President for the Taguspark campus, Instituto Superior Técnico, Universidade Lisboa, 2020-2023.
- Coordinator of the Computer Science and Engineering Doctoral Program, Instituto Superior Técnico, Universidade Lisboa, 2018-2022.
- Member of the Management Committee Taguspark Campus, Instituto Superior Técnico, Universidade Lisboa, 2016-2019.
- Responsible for the Taguspark campus Library Instituto Superior Técnico, Universidade Lisboa, 2003-2009.
- Member of the Coordination Committee, Computer Science and Engineering Department, Instituto Superior Técnico, Universidade Lisboa, 2005-2006.

11. Foreign Languages

- Very good English (spoken and written).
- Very good French (spoken and written).
- Some knowledge of Spanish.
- Fluent Portuguese (mother tongue).