



2023 Master internship at University of Cádiz



TITLE: Privacy-preserving and human-centered analytics on distributed ledger technologies

LAB & PEOPLE

- Name of the hosting lab: SPI&FM (Software Process Improvement and Formal Methods)
- General activities of the lab : Process and service engineering, Creative computing and learning assessment
- Website: <https://produccioncientifica.uca.es/grupos/7918/detalle>
- Number of staff / PhD: 17/ 15
- Supervisor name and contact: Juan Manuel Dodero (juanma.dodero@uca.es)

TOPIC OF THE INTERNSHIP

- **Scientific context of the internship (max 20 lines)** : The research focuses on the preservation of privacy in the analysis of data from distributed learning databases and health records. It addresses privacy-preserving data publishing as a key problem related to personal data. The research aims to develop techniques for combining multidimensional data from various sources, such as multimodal interaction devices, while preserving individuals' privacy. Privacy-preserving data publishing (PPDP) aims to make personal data indistinguishable from other data in a dataset while maintaining statistical utility. Semantic versions of PPDP techniques for nominal and categorical data will be explored. The research also aims to optimize existing PPDP techniques on applications in distributed environments, such as IoT, where computational costs and energy consumption are significant concerns. The research findings will be applied to datasets managed in Distributed Ledger Technologies (DLT) platforms such as IOTA, supporting sustainable and privacy-conscious data recording. The research also explores the use of Directed Acyclic Graphs in DLT platforms for implementing privacy-preserving data sharing systems. The scalable, open, decentralized, and interoperable technologies investigated in this project will improve human-centered learning analytics on distributed and multidimensional datasets. Privacy challenges arising from personal data in smart learning environments and health records need to be addressed to ensure ethical and privacy-preserving analytics.
- **Keywords** : Privacy by design, privacy-preserving data publishing & mining, distributed ledger technology, Human-centered learning analytics.

- **Tasks and duties entrusted to the student:** The student will be entrusted to study and use existing state-of-the-art techniques in the area of privacy-preserving data publishing for the processing of multimedia data obtained from learning experiences, health records and other open data sources. The data obtained from these techniques will be semantically characterised through knowledge graphs.
- **Skills to be acquired or developed:** Techniques for privacy-preserving data publishing, distributed ledger technology techniques.

Bibliography

Benjamin CM, M Fung, K Wang, R Chen, PS Yu. Privacy-preserving data publishing: a survey of recent developments. *ACM Computing Surveys*, 42 (4) (2010), pp. 1-53

Chamikara M, P Bertok, I Khalil, D Liu, S Camtepe. Privacy preserving distributed machine learning with federated learning. *Computer Communications*. 2021, 171, 112-125.

Dai W, C Dai, KKR Choo, C Cui, D Zou, H. Jin, SDTE: A Secure Blockchain-Based Data Trading Ecosystem, *IEEE Transactions on Information Forensics and Security* 15, 2020, pp. 725-737.

Gursoy ME, A Inan, ME Nergiz, Y Saygin. Privacy-Preserving Learning Analytics: Challenges and Techniques. *IEEE Transactions on Learning Technologies* 10(1), 2017, pp. 68-81.

Majeed A, S Lee. Anonymization techniques for privacy preserving data publishing: A comprehensive survey. *IEEE Access* 9, 2020, pp. 8512-8545.

Oviatt S, Ten Opportunities and Challenges for Advancing Student-Centered Multimodal Learning Analytics. *Proc. of ACM ICMI*, 2018, pp. 87-94.

Rodríguez-García M, MA Sicilia, JM Doderó. A privacy-preserving design for sharing demand-driven patient datasets over permissioned blockchains and P2P secure transfer. *PeerJ Computer Science*, 2021, article no. 568.

Sarwar K, S Yongchareon, J Yu, SU Rehman. A Survey on Privacy Preservation in Fog-Enabled Internet of Things. *ACM Computing Surveys* 55(1), 2023, Article 2.

Shum SB, R. Ferguson, R Martinez-Maldonado (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics* 6(2), 2019.

Zhao S, F Li, H Li, R Lu, S Ren, H Bao, JH Lin, S Han. Smart and Practical Privacy-Preserving Data Aggregation for Fog-Based Smart Grids, *IEEE Transactions on Information Forensics and Security* 16, 2021, pp. 521-536.

PROFILE OF THE DESIRED STUDENT

- Minimum level of study required: Level 6 - Bachelor's Degree
- Field(s) of study: Computer Science / Data Science
- Scientific skills: Data analysis, Maths & Statistics foundations, Technical writing, Strong programming skills
- Language skills required : English required (B2), Spanish optional

THE INTERNSHIP ASSIGNMENT:

Desired duration of the internship (in months): 9

Desired Starting date of the mission: *01/01/2024*

Indicative weekly schedule: *25h / week*

Remuneration : This internship will be funded by the Erasmus+ internships programme.

Internship agreement: *an internship agreement will be signed.*

To SEA-EU students:

*If you're interested please send your CV and letter of motivation to the scientist in charge,
juanma.dodero@uca.es before the date 30/09/2023.*