HORIZON-CL3-2021-FCT-01-08

RITHMS – Research, Intelligence and Technology for Heritage and Market Security

GA 101073932

Deliverable D8.3

OS final summary

WP 8 – Management and coordination

Authors: Michela De Bernardin, Sara Ferro (IIT)

Lead participant: IIT

Delivery date: 30 September 2025

Dissemination level: Public

Type: Report

Revision History

Author Name, Partner short name	Description	Date
Michela De Bernardin, Sara Ferro (IIT)	Draft deliverable	26/09/2025
Robert Vatasoiu (BEIA)	Contribution	28/09/2025
Riccardo Giovannelli (IIT), Tobias Mattes (HföD)	Revision and Security review	30/09/2025
Arianna Traviglia (IIT)	Final version	30/09/2025

Contents

Executive Summary

Deliverable D8.3 presents the RITHMS strategy for the management, dissemination, and publication of research outputs and data, ensuring full compliance with the Open Science (OS) framework and the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

Throughout the project's implementation, Open Science practices have been systematically integrated into research and dissemination activities. The main scientific outputs produced under RITHMS have been published under Open Access (OA) licences in peer-reviewed journals and also made publicly available through the project's official website (https://rithms.eu/results), in line with Horizon Europe's commitment to transparency and accessibility.

In addition, a selection of datasets collected or generated during the development of the RITHMS Platform has been cleaned, curated, and deposited in freely accessible online repositories as Open Data (OD). These datasets comply with the FAIR principles and are hosted in the IIT Dataverse, ensuring long-term preservation and usability by the wider research and policy community. The datasets can be accessed at URL: https://doi.org/10.48557/UCIKAW.

List of Abbreviations

AI - Artificial Intelligence
CDR - Call Detail Record
CPKU - Centar protiv krijumčarenja umjetninama
FAIR - Findable Accessible Interoperable Reusable
IIT - Italian Institute of Technology
LEA - Law Enforcement Agency
NGO - Non-governmental organisation
NLP - Natural Language Processing
OA - Open Access
OD - Open Data
OS - Open Science
SNA - Social Network Analysis
ROF - Romanian Obiecte Furate

List of Figures

List of Tables

1 Introduction

As anticipated in Deliverable D8.2 – *Data Management Plan*, the EU-funded project RITHMS (*Research, Intelligence and Technology for Heritage and Market Security*) embraces the principles of **Open Science** as a cornerstone of its research attitude, recognising openness, transparency, and inclusiveness as essential drivers of scientific progress and societal impact. Open Science represents an innovative and participatory approach that ensures that research results, data, methodologies, and publications are accessible to all. This model fosters collaboration, reproducibility, and collective learning across disciplines and institutions, thereby strengthening the credibility and effectiveness of scientific inquiry. Through the adoption of Open Science practices, RITHMS aims to extend the reach of its outcomes beyond the research community, encouraging active engagement by policymakers, law enforcement, cultural heritage professionals, and citizens.

In alignment with this vision, RITHMS and its Coordinator—the Italian Institute of Technology (IIT)—have placed a strong emphasis on the **FAIR principles** (Findable, Accessible, Interoperable, and Reusable) for the management and dissemination of research data. Ensuring that data and outputs adhere to FAIR standards is a fundamental requirement for the project's sustainability and long-term impact. By implementing rigorous data management procedures and interoperable infrastructures, the project guarantees that its datasets, tools, and results can be easily located, transparently accessed, and reused across diverse research and operational contexts. This approach not only facilitates verification and replication but also enhances the potential for cross-sector innovation and policy development in the field of cultural heritage protection.

1.1 Deliverable structure

Deliverable D8.3 is organised into four main sections. **Section 1** introduces the objectives and scope of the deliverable and provides an overview of its structure. **Section 2** presents the list of scientific papers produced within the framework of the RITHMS project, highlighting those already published under Open Access (OA) licences and those pending OA publication. **Section 3** summarises the dissemination of project results through posters presented at conferences and scientific events and available online. **Section 4** focuses on the datasets collected or generated during the development of the RITHMS Platform, describing in brief their content and publication processes.

2 Open Access Papers

Several publications have been produced during the project's lifetime, spanning from legal and ethical aspects related to the issue of illicit trafficking of cultural goods to research related to technical development of the RITHMS Platform's components (e.g., scrapers designed to collect data from diverse online sources and AI-based algorithms used to extract entities and relations from free-text).

The following list details journal articles published in Open Access, from the earliest to latest released. For the sake of completeness, a brief description of the content and corresponding practical classification labels have been included. The labels featured to classify the outputs are: 'LEGAL', 'ETHICS', 'COMPUTER SCIENCE', 'SOCIAL NETWORK ANALYSIS' and 'MACHINE LEARNING'.

Overall, as in the figure below, most of the publications fall under the category of 'COMPUTER SCIENCE'.

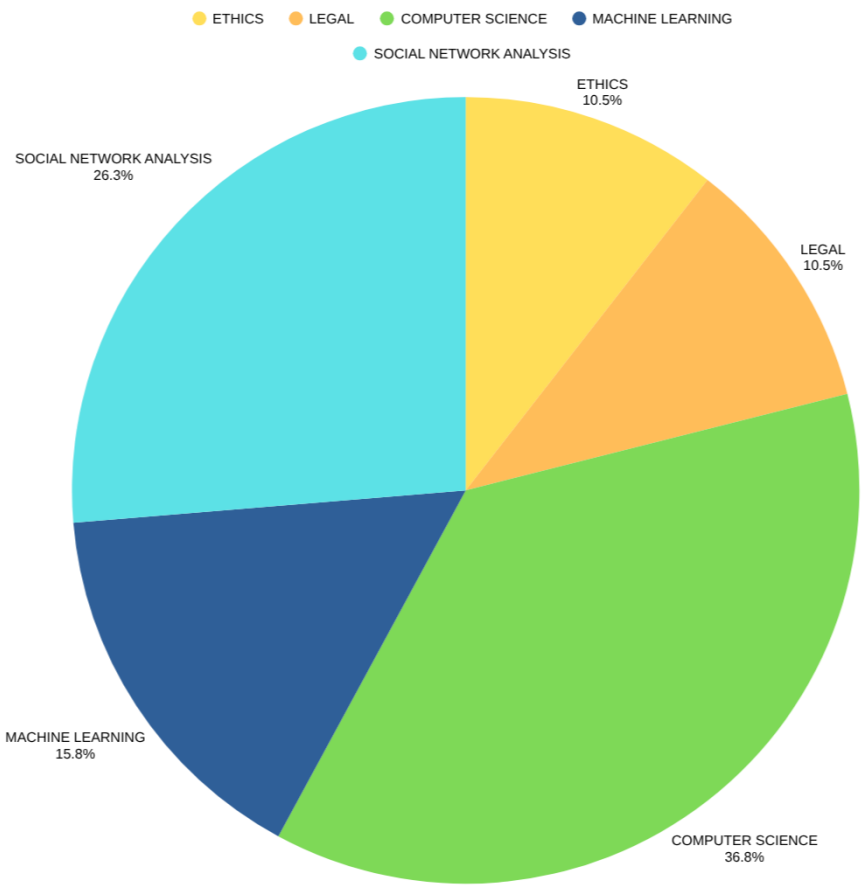

Figure 1. *Resulting OS products distribution, but OD*

1. **Ethical and Legal Risks of Algorithmic and AI Tools Developed to Fight Against Trafficking in Cultural Property in the RITHMS Project** (García Adán - Fuentes Loureiro, 2023). The paper examines the ethical and legal risks associated with the use of algorithmic technologies, big data, and artificial intelligence (AI) in combating the trafficking of cultural property. It focuses on the tools developed within the RITHMS Project, designed to identify criminal networks and monitor art markets, online auctions, and social media for suspicious activities. At the same time, these technologies raise significant ethical and legal challenges, requiring careful risk assessment and strict compliance with data protection, procedural, and fundamental rights standards. The ultimate goal is to ensure their responsible and lawful use by competent authorities. Classification label: 'ETHICS, 'LEGAL', 'MACHINE LEARNING'.

2. **Exploring Overlaps of Cultural Property Crime with Organised Crime in EU Policy Documents** (Faraldo Cabana, 2024). This article analyses how the European Union has addressed the links between the trafficking of cultural property and organised crime in its policies from 1993 to 2023. Based on a content analysis of 58 policy documents, it reveals conceptual and structural weaknesses that have limited policy effectiveness. Misconceptions about the organised nature of cultural property trafficking and its ties to terrorism financing risk leading to misguided measures, while recent steps—such as extending the EU anti-money laundering framework to art market actors—offer promising tools to counter illicit financial flows. Classification label: 'ETHICS', 'LEGAL'.

3. **Art Crime Does not pay: Multiplexed Social Network Analysis in Cultural Heritage Trafficking Forensics** (Salonen - Guarino, 2024). The paper examines how crimes against cultural heritage have become a key element of organised crime, often financing terrorism and international conflicts. To address this growing threat, the authors sketch an overview of the SNA-based methodology underpinning RITHMS project's proposal to support law enforcement agencies (LEAs) in tackling cultural property crime. Classification label: 'COMPUTER SCIENCE', 'SOCIAL NETWORK ANALYSIS'.

4. **War, art, and sanctions: social network analysis on the NACP's databases of sanctioned Russian individuals and art collectors** (Leeson et al., 2024). The study analyses data from Ukraine's War & Sanctions and War & Art databases (NACP) to map the social network of Russian elites sanctioned for supporting the war in Ukraine. Network analysis reveals a moderately centralised structure centred on key oligarchs and politicians, with art collectors playing particularly central roles. While efficient in information flow, the network's reliance on a few individuals creates vulnerabilities, and inconsistencies in sanction enforcement may

still allow the trade of cultural goods. Classification label: 'COMPUTER SCIENCE', 'SOCIAL NETWORK ANALYSIS'.

5. **Overcoming data siloes in cultural heritage crime research: a consolidated OSINT-derived dataset on art, antiquities, and the trade in cultural goods** (Leeson et al., 2025). The paper examines the challenges faced by *provenance* researchers, cultural heritage crime analysts, and law enforcement agencies, where data on cultural goods is often fragmented across multiple databases. The EU-funded RITHMS project addresses this by developing a platform that aggregates open, specialised, and police data to support investigations into illicit trafficking. A key phase involved creating 30 tailored web scrapers to collect data from databases of stolen, missing, protected, and unprovenanced items, resulting in the largest non-police dataset of its kind. The paper outlines the data collection and preprocessing process, demonstrates the value of consolidated data for intelligence-led investigations, and shows how mining this dataset can generate new scientific insights and actionable intelligence. Classification labels: 'COMPUTER SCIENCE'.

6. **A novel NLP-driven approach for enriching artefact descriptions, provenance, and entities in cultural heritage** (Ferro et al., 2025). The research tackles the challenge of structuring and normalising cultural heritage data, focusing on entity extraction from unstructured texts to support analysis and knowledge enrichment. It presents a procedure for creating domain-specific datasets to train NLP models and evaluates their performance across three RITHMS datasets: North American museum provenance data, stolen cultural goods in Romania, and unclassified WWII-looted Polish art. Results show that fine-tuned NLP models, particularly Transformers trained on provenance data, outperform non-fine-tuned models, while models trained on noisy descriptive datasets perform poorly. The study demonstrates the potential of automated entity extraction to build knowledge graphs and support advanced analyses, such as network analysis, in cultural heritage research. Classification labels: 'COMPUTER SCIENCE', 'MACHINE LEARNING'.

All the articles were published under the **Gold Open Access (GOA)** option, which presents the following properties[1]:

- *Timing*: Available immediately on publication;
- *Version*: Final published version of record, after any copyediting and typesetting;
- *Location and discoverability*: Article freely available and easily discoverable on publisher's platform, alongside other relevant content;

[1] Cfr. https://www.springernature.com/gp/open-science/about/green-or-gold-routes-to-oa.

- *Integrity of scientific record*: the article the is up-to-date and linked to any post-publication corrections;
- *Licensing*: Open license allow users to share onwards;
- *Path to OS*: Can be bi-directionally linked to OD sets and protocols, as well as included in open metrics, and complying with open standards.

All the abovementioned articles published under GOA option have **Creative Commons Attribution 4.0 International** licence.

2.1 Papers pending publications in OA

The following papers have been accepted for publication in journals/books with GOA option and will be released by the end of the year:

- **Faraldo Cabana P.,** *Exploring the EU Perspective on the Organised Nature of the Trafficking in Cultural Goods*, CEPOL European Law Enforcement Research Bulletin, forthcoming (2025).
- **Fuentes Loureiro, M.A./ Faraldo Cabana, P**., *How organised is the trafficking of cultural goods? Spain as case study*, forthcoming (2025).
- **Giovanelli, R., Leeson, M., De Bernardin, M., Ferro, S., and Traviglia, A**., *Social Network Analysis on the Proveana Database: Insights on the Circulation of Nazi-Looted Cultural Goods during and after WWII*, Social Network Analysis and Mining, forthcoming (2025).
- **Leeson, M., De Bernardin, M., Giovanelli, R., Ferro, S., and Traviglia, A**., *RITHMS Digital Platform: Social Network Analysis for Intelligence-Led Policing of Cultural Heritage Crime*. In Emiline Smith and Summer Austin (eds) *Researching a Rigged Game: Open-Source Data & the Trade in Cultural Heritage*. Cham, Switzerland: Springer, forthcoming (2025).

3 Posters

During the RITHMS project's lifespan, partners attended several conferences offering either oral presentations or research posters. The following list reports posters that have passed peer-review and have been made public during such conferences; they are all available on the RITHMS website:

1. **Art crime does not pay** (Salonen – Guarino, 2024). Classification labels: 'COMPUTER SCIENCE', 'SOCIAL NETWORK ANALYSIS'.

2. **RITHMS Project: How Digital Technology Can Help Protect Cultural Heritage from Illicit Trafficking** (De Bernardin et al., 2024). Classification labels: 'COMPUTER SCIENCE', 'MACHINE LEARNING', 'SOCIAL NETWORK ANALYSIS'.

3. **RITHMS Digital Platform. How Social Network Analysis Can Be Leveraged to Fight High-Risk Criminal Networks Trafficking in Cultural Goods** (De Bernardin et al., 2025). Classification labels: 'COMPUTER SCIENCE', 'SOCIAL NETWORK ANALYSIS'.

4 Datasets

The Coordinator received **written permission for publishing as Open Data** eight of the numerous datasets generated/collected during the research and development phase of the RITHMS project. These represent just the 16% of all data collected and used within the RITHMS Platform. However, valuable information is contained in those datasets and can be of support for interested researchers, scholars, and professionals.

The corpus (Fig. 2) mainly entails Romanian datasets on national cultural artefacts, organisations and locations (i.e., five CIMEC datasets, the ROF dataset, and the synthetic CDRs dataset), which represents the 87.5% of the OD. The seventh dataset released as OD by the RITHMS consortium concerns the stolen and missing works of art from Bosnia and Herzegovina (CPKU).

Figure 2. *RITHMS OD distribution within Europe*

The following sections provide a brief description of each dataset detailing their respective licences, while **Figure 3** displays a graphical representation of the distribution of the types of licence under which the Consortium has released its datasets.

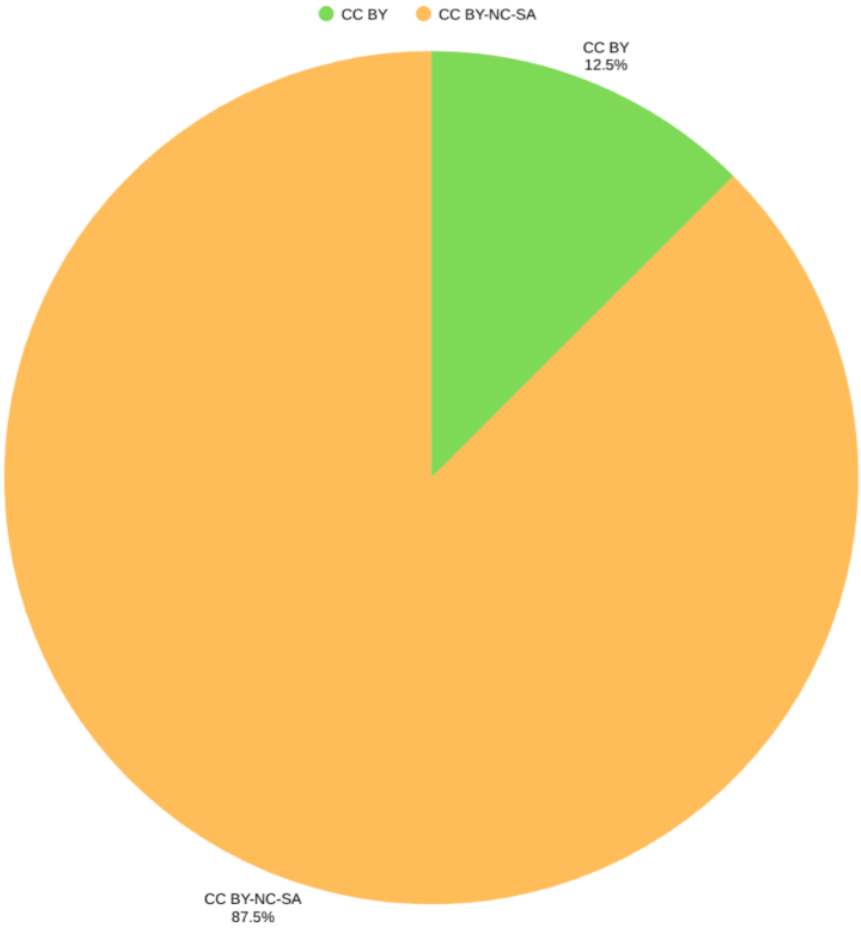

***Figure 3**. Types of licence distribution for the OD*

4.1 Scraped Datasets

4.1.1 CIMEC Dataset

The Institute of Cultural Memory in Romania (CIMEC), was founded in 1978 to create and manage multiple databases of Romanian cultural heritage and heritage experts. It is a database holding information on damaged, stolen, missing, illegally exported and protected cultural objects. It presents five different portions of data: one concerning the destroyed, stolen, missing, or illegally exported artefacts (which was named "CIMEC"), one about experts ("CIMEC_EXPERTS"), one about the Romanian museums ("CIMEC_MUSEUMS"), the database of Romanian places of worship "CIMEC_POW" and the "CIMEC_NAR" the dataset of the archaeological sites in Romania. Only "CIMEC_NAR" database was refined using Natural Language Processing (NLP) algorithms, performed on the "discoverer" field to detail whether it features a person or an organisation.
The target URLS of the original CIMEC databases are reported in **Table 1**:

	Scraped URLs
CIMEC	https://clasate.furate.ro/Lista_en.asp
	https://clasate.cimec.ro/Lista_en.asp
CIMEC_EXPERTS	http://acera.cimec.ro/RegistruArheologi.aspx
	https://old.cimec.ro/scripts/MuzeeSpecialisti/sel.asp?nr=1&NrSel=0&Lang=EN
	https://registru-conservatori-restauratori-r2.cimec.ro/sel.asp?nr=1&NrSel=0&Lang=EN
	https://registru-experti-r1.cimec.ro/sel.asp?nr=1&NrSel=0&Lang=EN
CIMEC_MUSEUMS	https://ghidulmuzeelor.cimec.ro/Muzee-Acreditate.asp
CIMEC_POW	https://lacasuridecult.cimec.ro/EN/Documente/ASP/culte.asp?
CIMEC_NAR	https://ran.cimec.ro/sel.asp?lang=EN

***Table 1**. Scraped URLs of the CIMEC databases*

The resulting RITHMS dataset (with scraped, cleaned, and organised data) is licenced under **CC BY-NC-SA**:

- *CC (Attribution)*: Necessary to give appropriate credit to the original creator.
- *NC (Non-Commercial)*: Denial of use of the dataset for commercial purposes.
- *SA (ShareAlike)*: Possibility to modify, adapt, or build upon the dataset, but any derivative works must be distributed under the same license as the original.

4.1.2 ROF Dataset

ROF (Romanian Obiecte Furate) database, managed by the Romanian police, records national stolen goods. Data entries are added based on the police reports of stolen goods belonging to both individuals and national institutions. The database entails a variety of artefact types, such as paintings,

books, coins, statues, jewellery, and ceramics. The scraped dataset has been further expanded for the RITHMS Platform's purposes by applying NLP algorithms to extract information on objects' details, like dimensions, weight, and material of the artefact.
The original database is available at the URL *https://www.politiaromana.ro/ro/obiecte-furate*.

The resulting RITHMS dataset (with scraped, cleaned, and organised data) is licenced under **CC BY-NC-SA**:

- *CC (Attribution)*: Necessary to give appropriate credit to the original creator.
- *NC (Non-Commercial)*: Denial of use of the dataset for commercial purposes.
- *SA (ShareAlike)*: Possibility to modify, adapt, or build upon the dataset, but any derivative works must be distributed under the same license as the original.

4.1.3 CPKU Dataset

CPKU stands for "Centar protiv krijumčarenja umjetninama" (Center Against Trafficking in Works of Art). The centre is the most active non-governmental organisation (NGO) working on the preservation of heritage and tracking the trafficking of cultural goods in Bosnia and Herzegovina. CPKU database presents information on stolen and missing works of art from Bosnia and Herzegovina.

While conducting a preliminary investigation on the CPKU open website's terms of use, the Coordinator discovered that these terms were not explicitly stated. Therefore, on September 18th, 2023, IIT submitted a form to obtain written permission from the CPKU managers to both scrape and re-publish CPKU data. Permission was granted; however, the website subsequently became accessible only through authentication. Access credentials were then provided to IIT researchers, enabling them to develop a new authentication-based scraper.
The original database is available at the URL *http://www.cpku.org/baza-nestalih-umjetnina-the-database-of-missing-art/*.

The resulting RITHMS dataset (with scraped, cleaned, and organised data) is licenced under **CC BY-NC-SA**:

- *CC (Attribution)*: Necessary to give appropriate credit to the original creator.
- *NC (Non-Commercial)*: Denial of use of the dataset for commercial purposes.
- *SA (ShareAlike)*: Possibility to modify, adapt, or build upon the dataset, but any derivative works must be distributed under the same license as the original.

4.2 Synthetic CDRs Dataset

This dataset was developed by the technical partner BEIA for the purposes of Pilot 3. It consists of synthetically generated Call Detail Records (CDRs) simulating Romanian mobile phone traffic.

Based on 300 real CDRs from Romania, 2,000 synthetic data entries were generated and linked to cell towers in areas previously identified as subject to looting. The **FAST ML synthesizer from the Synthetic Data Vault** (Python) library, optimised for modelling speed, was used for data generation. The resulting dataset closely preserves the properties of the original dataset.

This dataset is licenced under **CC BY** (**Creative Commons Attribution 4.0 International**)[2] and is hosted by Zenodo at the URL: *https://zenodo.org/records/17253518*. However, for the convenience of researchers, scholars, and interested individuals, a copy of the same dataset is saved and publicly available also within the IIT Dataverse together with all other databases released in the framework of the RITHMS Project.

[2] https://creativecommons.org/licenses/by/4.0/

5 Conclusions

The implementation of Open Science and FAIR data principles within RITHMS is not limited to compliance with Horizon Europe requirements but reflects a broader institutional commitment by the Coordinator-IIT and the project consortium to promote transparency, accountability, and long-term accessibility of research outputs, as also detailed in Deliverable D8.2 – *Data Management Plan*.

By embedding Open Science practices into its operational framework, RITHMS actively contributes to the creation of a sustainable research ecosystem where data sharing and knowledge exchange are integral components of innovation. The project's adherence to the **Horizon Europe Open Research Data policy** guarantees that its outcomes will remain accessible beyond the project's lifetime, maximising the scientific, societal, and policy impact of publicly funded research. Through this effort, RITHMS also sets a precedent for responsible and collaborative research in the domain of cultural heritage protection, demonstrating how openness and interoperability can serve as catalysts for both scientific excellence and the preservation.

References

1. **De Bernardin, M., Giovanelli, R., Leeson, M., and Traviglia, A.**, *RITHMS Project: How Digital Technology Can Help Protect Cultural Heritage from Illicit Trafficking*. **Poster** presented at the European Archaeologist Association Annual Event, Rome 2024. https://rithms.eu/images/publications/IIT_EAA_RITHMS%20Poster.pdf
2. **De Bernardin, M., Leeson, M., Giovanelli, R., Ferro, S., and Traviglia, A.**, *RITHMS Digital Platform. How Social Network Analysis Can Be Leveraged To Fight High-Risk Criminal Networks Trafficking In Cultural Goods*. **Poster** presented at the CEPOL Research and Science Conference 2024/2025. https://rithms.eu/images/publications/IIT_CEPOL%202025_Poster.pdf
3. **Faraldo Cabana, P.,** *Exploring overlaps of cultural property crime with organised crime in EU policy documents*, European Journal on Criminal Policy and Research (ISSN 0928-1371, e-ISSN 1572-9869), 2024, online first. https://doi.org/10.1007/s10610-024-09595-9
4. **Ferro, S., Giovanelli, R., Leeson, M., De Bernardin, M., and Traviglia, A.,** *A novel NLP-driven approach for enriching artefact descriptions, provenance, and entities in cultural heritage*. Neural Computing & Application 37, 2025. https://doi.org/10.1007/s00521-025-11449-2
5. **García Adán, A., Fuentes Loureiro, M.A.**, *Ethical and Legal Risks of Algorithmic and AI Tools Developed to Fight Against Trafficking in Cultural Property in the RITHMS Project*, en AA.VV., Proceedings XoveTIC 2023. Impulsando el talento científico. Servizo de Publicacións, A Coruña, 2023. https://xovetic.citic.udc.es/wp-content/uploads/2023/10/XoveTIC_2023_proceedings.pdf
6. **Leeson, M., Giovanelli, R., De Bernardin, M., and Traviglia, A.**, *War, Art, and Sanctions: Social Network Analysis on the NACP's Databases of Sanctioned Russian Individuals and Art Collectors,* International Journal of Digital Humanities, 2024. https://link.springer.com/article/10.1007/s42803-024-00089-y
7. **Leeson, M., Giovanelli, R., Ferro, S., De Bernardin, M., and Traviglia, A.,** *Overcoming Data Siloes in Cultural Heritage Crime Research: A Consolidated OSINT-Derived Dataset on Art, Antiquities, and the Trade in Cultural Goods*. Archival Science 25, 2025. https://doi.org/10.1007/s10502-025-09485-x
8. **Salonen, J., Guarino, A.,** *Art Crime Does not pay: Multiplexed Social Network Analysis in Cultural Heritage Trafficking Forensics*, Proceedings of the 19th International Conference on Cyber Warfare and Security, ICCWS 2024, pp. 617-620. https://doi.org/10.34190/iccws.19.1.2066
9. **Salonen, J., Guarino, A.,** *Art Crime Does not pay*. **Poster** presented at the 19th International Conference on Cyber Warfare and Security, ICCWS 2024. https://rithms.eu/images/publications/VTT-STAG_ICCWS24_Poster.pdf

Project Coordinator

Arianna Traviglia

arianna.traviglia@iit.it

Scientific Project Manager

Michela De Bernardin

michela.debernardin@iit.it