SIGNIFICANCE. STOP ILLICIT HERITAGE TRAFFICKING WITH ARTIFICIAL INTELLIGENCE

D. Abate[1*], M. Paolanti[2,3], R. Pierdicca[4], A. Lampropoulos[5], K. Toumbas[1], A. Agapiou[1], S. Vergis[5], E. Malinverni[4], K. Petrides[5], A. Felicetti[3], P. Zingaretti[3]

[1] The Cyprus Institute, Science and Technology in Archaeology and Culture Research Centre (STARC), 2121, Nicosia, Cyprus
[2] University of Macerata, Department of Political Sciences, Communication and International Relations, 62100 Macerata, Italy
[3] VRAI Lab - Università Politecnica delle Marche, Dipartimento di Ingegneria dell'Informazione, 60131 Ancona, Italy
[4] Università Politecnica delle Marche, Dipartimento di Ingegneria Civile, Edile e dell'Architettura, 60131 Ancona, Italy
[5] HYPERTECH S.A., 32 Perikleous St, 15232 Athens, Greece

Commission II, WG II/8

KEY WORDS: Artificial Intelligence, Social Media Intelligence, Dark Web, Illicit Traffic, Cultural Heritage.

ABSTRACT:

The inability to prevent or eliminate illicit trafficking of cultural goods is not limited to failed-state environments or any specific part of the globe. While the antiquities market denies that this illicit trade is a widespread phenomenon, the international community and Law Enforcement Agencies (LEAs) overwhelmingly recognize the problem indicating that organized crime is involved at all stages. Nowadays, web platforms play host to groups dedicated to illegal archaeological excavations and Illicit trade of cultural goods. Looters have the freedom to connect online with potential buyers around the world. At the same time, social media platform monitoring in search of criminal activities conducted by LEAs is poor due to the lack of expertise, efficient tools to scan the massive amounts of data, and funds. The COVID-19 crisis has compounded the problem by driving more and more dealers and buyers online – where they are discovering that by joining certain unmonitored groups, they can enter the illegal market with ease. The EU funded SIGNIFICANCE project (Stop Illicit heritaGe traffickiNg wIth artiFICiAl iNtelligenCE) has been designed to boost LEAs investigation capabilities in monitoring online illegal activities on social media platforms, the web and the dark web for the identification of cultural property crimes, exploiting Artificial Intelligence and Deep Learning algorithms for guaranteeing the successful prosecution of perpetrators unveiling criminal networks.

1. INTRODUCTION

The consequences of Cultural Property Crimes were first recognized in 1954 with 'The Hague' Convention for the Protection of Cultural Property in the Event of Armed Conflict (Toman, 2017). Numerous international instruments focused on addressing this crime have since been developed, such as the UNESCO Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property (1970); the UN Convention against Transnational Organised Crime (2000); the UNESCO Convention on the Protection of the Underwater Cultural Heritage (2001); and the Council of Europe Convention on Offences relating to Cultural Property (2017). During its 40th General Conference in 2019, UNESCO adopted the 14th of November as the International Day against Trafficking in Cultural Property with the aim of drawing more attention to this crime and the ways to combat it, highlighting the importance of international cooperation and proactive measures. However, despite the nearly 70 years of regulatory attempts through legislation and international efforts, cultural property crimes continue to damage heritage worldwide, supporting criminality through its illegal profits. This problem is by no means a new one, and it has grown more acute in recent decades. Political and security breakdowns in countries which were once the territory of ancient civilizations provide ideal conditions for these illegal activities. The richness of sites, remote locations, and modern communication tools further facilitate illicit business, taking it to

a global scale. The inability to prevent archaeological looting or to eliminate trafficking is not limited to a failed-state environment or any specific part of the globe. It is an issue that many EU Member States face. Moreover, after the outbreak of the COVID-19 pandemic the situation has been exponentially exacerbated. Thanks to an unprecedented technological innovation, many sectors now benefit from the availability of new tools, equipment, devices and specialised know-how. In the last few years Artificial Intelligence (AI) and Deep Learning (DL) have become disciplines widely used in several research domains with the potential or revolutionising the research capabilities in terms of speed and time needed for large volume of data analysis (Paolanti et al., 2020). Unlike simple artificial neural networks, DL algorithms are not only used for the mapping from representation to output but also to learn the representation itself (Granell et al., 2018). However, AI and DL, despite their potential, are rarely used in addressing the illicit trafficking of heritage, struggling to fully exploit their new possibilities. From these premises, the European project Stop Illicit heritaGe traffickiNg wIth artiFICiAl iNtelligenCE - SIGNIFICANCE (http://www.significance-project.eu) arises. It has been specifically designed to help and increase the responsiveness and effectiveness of the public authorities and police corps against the illicit trafficking of cultural goods through Internet channels (i.e., social platforms, web and dark web). The developed platform will allow relevant authorities to undertake proper actions for identifying, tracking,

---

*Corresponding author

STEP 1
Through web crawling algorithms on social platforms such as instagram (Instaloader) you can specifically query the platform to download the dataset of interest

STEP 2
Data annotation requires experts in the field and a cooperation with insiders and authorities to detect create the knowledge base

STEP 3
Our artificial intelligence algorithm automatically identifies illicit traffic of cultural goods, supported by High Performance computing infrastructures

STEP 4
The classification results will be available to the authorities through an alerting system

Figure 1. SIGNIFICANCE Strategy

and stopping illegal online actions and for guaranteeing the successful prosecution of perpetrators unveiling criminal networks. The level of ambition, innovation, and originality of SIGNIFICANCE is hence high, being the first platform tailored for the fight against online illicit trafficking of cultural assets. The main objective of SIGNIFICANCE is to develop a platform based on AI and DL algorithms, both interfaced with web, social media and dark web, which will improve the identification of artifacts (real or forgeries) sold and potentially link them to identify the criminal networks. By scanning forums and communications networks with an image- and text-based AI approaches (Paolanti et al., 2017) and combining methodologies of new cyber forensics, SIGNIFICANCE will flag suspicious activities to competent authorities which will be able to swiftly react and get a better understanding of the reach of online antiquities trafficking networks and combat them. SIGNIFICANCE, exploiting an automatic image- and text-based approach, aims at boosting intelligence-led investigations, including cross-border investigations between EU members and non- members state. It aims at increasing the identified amount of items sold or advertised online by an average rate of 10% to 15% on annual base at a national scale. In the long-term scenario, the developed tools will bring to a new level of efficiency and reliability in the detection and investigation methodologies, which will be tailored to the needs of different enforcement authorities. For the achievement of the project goals, several tasks, which demand serious computational resources, will be run on the CyClone HPC Infrastructure of The Cyprus Institute specifically configured for AI processing activities. This approach allows to optimize the procedures in terms of computing time, resources allocated and cost-efficiency. The two tasks which will mainly benefit from the use of the HPC infrastructure are: Algorithm Modelling and Training; Web Scraping and Data Annotation (Figure 1).

The remainder of the paper is structured as follows: Section 2 gives an overview of the background knowledge of related studies; Section 3 describes of the methods and the entire workflow together with the required materials; section 4 presents the SIGNIFICANCE platform accompanied by conclusions in section 5.

2. RELATED WORKS

Today public authorities and the scientific communities directly involved in the protection and preservation of cultural heritage are failing to conserve and to persuade others to conserve the world's archaeological heritage. The latter continues being destroyed at an undiminished pace both by natural disasters and human-made actions. Alarmingly a significant proportion of the ongoing lost is due by looters, acting mainly for commercial reasons, which are financed indirectly by private, and sometimes public, collectors of antiquities. Public Authorities, Law Enforcement Agencies (LEAs) and the Research community are gathering to maximize the efforts in the development of technological solutions.

Several databases of stolen objects are today available although not all of them are fully open to the public and the academia. One of the largest LEAs' database is LEONARDO (The Stolen Works of Art Database System), maintained by the Command for the Protection of Cultural Heritage of the Italian Carabinieri (Carabinieri for the Protection of Cultural Heritage- TPC, 2022). It was initially established in 1980, and today is composed by 1.285,765 stolen objects, 810.423 images and 65.970 theft cases. All the material is currently digitized in image and text formats. LEONARDO database is the reference point for the Italian and foreign LEAs. It allows to conduct a careful analysis of criminal phenomenon concerning the illicit trafficking of cultural property.

Recently, thanks to the SWOADS Project (Stolen Works Of Art Detection System) the software components of the LEONARDO database are being improved and expanded both in technological (i.e. big data, machine learning) and architectural terms (i.e. blockchain) (Carabinieri for the Protection of Cultural Heritage- TPC, 2020). Another digital tool developed by Carabinieri TPC is the mobile iTPC App (Carabinieri for the Protection of Cultural Heritage- TPC, 2017). It allows to (i) consult and download a bulletin of the stolen works of art which contain the most important artefacts stolen over time; (ii) perform a visual search in real time by comparing images with those contained in a special dedicated computer archive; (iii) create a record (Object ID) which allows an exhaustive and photo- graphic description of cultural assets, essential in the event of theft.

The INTERPOL Stolen Works of Art database combines descriptions and pictures of more than 52,000 items from 134 countries (INTERPOL Development Team 2022). It is the only database at the international level with certified police information on stolen and missing objects of art. Together with classic data fields, users can complement their search by uploading a picture of any object of art and checking it with the image-matching software. Anyone can apply to become an authorized user of the database, to check in real-time if an item is among the registered objects.

The ID-Art app instead, helps to identify stolen cultural property linking users to the INTERPOL Stolen Works of Art database through mobile devices (INTERPOL Development Team 2021). The app allows to (i) access the INTERPOL database of Stolen Works of Art to check if an object is registered as stolen; (ii) create an inventory of private art collections; (iii) report an item as stolen; (iv) report cultural sites potentially at risk or illicit excavations.

The EU-funded PREVISION project (Prediction and Visual Intelligence for Security Information) aimed to provide law enforcement agencies with advanced, almost-real-time, analytical support for multiple Big Data streams (PREVISION Development Team, 2020). The project outcomes intend to help investigators to better understand and address hybrid security threats (i.e. threats that combine physical and cyber-attacks). One specific task was focused on the illicit trafficking of cultural property.

3. SIGNIFICANCE WORKFLOW

SIGNIFICANCE project's final goal is to create a platform comprising of data and selected analytics on illicit traffic of cultural property on the internet and the dark web.

The system will automatically monitor and classify suspicious activities exploiting novel Artificial Intelligence algorithms, triggering an alarm to initiate further actions when a threat is identified.

Although the few excellences mentioned above (Databases and mobile Apps), LEAs' modus-operandi is still based on manual procedures for web-monitoring, and in some cases reluctant to adopt new digital technologies which may change the well-established investigation strategies.

Online investigations currently rely indeed on traditional web-search tools which are very time consuming compared with the amount of data to analyse. SIGNIFICANCE workflow is structured as follows. After the creation of different image-datasets and ontologies, the selected Neural Networks is trained exploiting HPC computing resources. Web Crawling techniques are used to download off-line targeted areas of the web and dark web on a dedicated storage system. Afterwards, the trained Neural Network searches the offline database to identify for possible risks. Information are stored and a warning to LEAs and relevant authorities is issued through the SIGNIFICANCE platform.

The system's design requires the implementation of state-of-the-art techniques in the fields of database creation, data semantic, AI, HPC programming, and web crawling.

In order to tailor the platform for the needs of Law Enforcement Agencies a questionnaire was structured with the final goal of providing a collection of user needs and statements to form the specification of the SIGNIFICANCE platform.

3.1 Training Datasets

Nowadays, the lack of image dataset created *ad-hoc* to train Convolutional Neural Networks (CNNs) for the identification on the web of illicit traffic of cultural property, represents a limit for the development of systems based on Artificial Intelligence and Deep Learning.

With the final goal of representing several artefact typologies, different object classes have been chosen to train the SIGNIFICANCE CNN, including ceramics, coins, icons, manuscripts, and fresco fragments.

Each database is divided into two groups respectively (Training set and Test set). The Training set teaches the network how to weigh different features, adjusting their parameters according to

their likelihood of minimizing errors in the results. The Testing set is used as a benchmark against the final random sampling. The results it produces will validate the accuracy of the trained network in identifying the images on the web or recognizing an 80% of them.

As part of the project implementation strategy the SIGNIFICANCE training database creation exploits a threefold approach based on (Table I):

- Traditional reality-based image data collection,
- Manual data collection from the web,
- web-scraping techniques.

| Name	N. of Files	Source
Ancient Coins (Greek)	213	INTERPOL DB
Ancient Coins (Roman)	89	INTERPOL DB
Cypriot Ceramic	39	INTERPOL DB
Ceramic Jugs	180	INTERPOL DB
Fresco Fragments	62	INTERPOL DB
Cypriot icons	54	INTERPOL DB
Icon Single Saint	466	INTERPOL DB
Illuminated Manuscript	6517	Web-Scraping
Greek Coins (I)	9995	Web-Scraping
Greek Coins (II)	12230	Web-Scraping
Greek Coins (III)	11494	Web-Scraping
Byzantine Icon Cyprus	850	Reality-Based
Table I. Training Datasets		
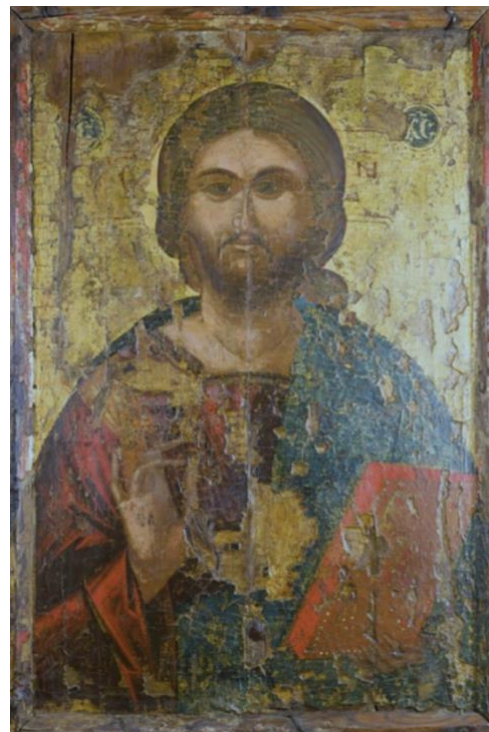
Figure 2. Training Dataset, Icon Samples

A different number of images and metadata has been collected with the final goal of testing the reliability of the trained neural network. An Image dataset have been newly created digitizing a large number of artefacts. This approach was chosen to demonstrate the sustainability of the platform, which can be expanded by different stakeholders training new Convolutional Neural Networks with new available datasets. A collection of Byzantine Icons produced in Cyprus and currently part of the collection of the Holy Bishopric of Limassol was digitized. The

database was created performing a photographic campaign conducted by the Cyprus Institute researchers. (Figure 2).

For specific classes of objects, the INTERPOL database has been consulted, and data have been downloaded and organized. Each stolen item featured an image coupled with metadata both concerning the artefact and the theft case. (Figure 3).

Figure 3. Training Dataset, Interpol Datasheet Sample

Few additional datasets containing an average of 5000 images of ancient coins and illuminated manuscripts were collected by scraping results from open online database. (Figure 4).

3.2 Web Crawling

One of the main components of the SIGNIFCANCE platform is the implementation of the Web Crawlers that will be used to retrieve raw data from the internet and the dark web. It will collect textual and visual information about heritage artefacts connected with illicit traffic activities for further analysis. Once data are stored, the trained Artificial Intelligence algorithms will be deployed and will analyse the new dataset in order to discover possible illegal activities.

The crawlers will scan well-known web applications, such as e-commerce platforms (i.e., e-bay), auction-houses websites, social networks (i.e., Facebook) and additional websites that have been pinpointed as relevant by domain experts during the user-requirements data collection (i.e., Catawiki).

Together with the web-resources mentioned above, it is well known that a large part of cultural property crimes is perpetrated in the dark and deep-web. To access those in- formation, a customised dark-Web crawler has been implemented targeting forums and e-stores where criminal activities are often hidden.

Figure 4. Training Dataset, Coin Samples

Crawling dark websites poses significant challenges, since dark web sites are built to be hardly discoverable, hidden in The Onion Router (TOR) network. TOR is a free, open-source software which enables anonymous communication among users. Existing implementations of dark Web Crawlers, such as the ACHE crawler, will be used and adapted for the purposes of the SIGNIFICANCE project. The data retrieved by the Web Crawlers will be stored to the SIGNIFICANCE database. As each web crawling channel represent the data of a transaction differently, a common, underlying way for representing online artifact transactions will be identified. More specifically, the following data models will be introduced. They will be common between SIGNIFICANCE components:

- Item Data Model (Table II),
- Seller Data Model,
- Buyer Data Model.

Attribute Name	Type	Required	Example
id	identifier	yes	'61e8518cb2cd54513bcb5fd1'
date_inserted	datetime	yes	2022-01-19T12:33:13.784+00:00
title	string	yes	'Nice German silver thaler'
description	string	yes	'A beautiful and very sharp German silver thaler, from Saxony or Sachsen 1638.'
price	float	yes	495.0
item_link	string	yes	'https://www.trocadero.com/stores/Senatus/items/1453814/Nice-German-silver-thaler-taler-

			Saxony-Johann-Georg-1638'
platform	string	yes	'trocadero'
seller_id	identifier	yes	'cb2cd54513bcb5fd161e8518'
buyer_id	identifier	no	'cb2cd5451361e8518bcb5fd1'
images	list of strings	no	[]
type	string	no	'icon'
technique	string	no	' '
material	string	no	'wood, tempera, handpainted'
shape	string	no	'circle'
dimensions	string	no	'32×27×2.5 cm'
signature	string	no	' '
artist	string	no	'unknown'
location	string	no	'Russia'
item_date	string	no	'20th century'
provenance	string	no	' '
certificate	string	no	' '
replica	string	no	'original'
dl_result	boolean	no	False
Table II. Fields for the Item Data Model			

Web Crawlers will share the collected data to other SIGNIFICANCE components through RESTful APIs, in order to facilitate component interconnection. This will facilitate the further analysis of the discovered data and their visualization.

The main REST Interface between the Web Crawlers and the Cultural Heritage Database will be the DataCollectionRestAPI. The interface will be implemented in Python programming language. Several REST Services will be exposed for inserting, updating, querying and deleting transaction data. Other SIGNIFICANCE components including the AI Algorithms and the SIGNIFICANCE Platform will have access to this interface.

3.3 Ontology-based Deep Learning

SIGNIFICANCE project has been focused on the development of effective investigation techniques tailored for as much classes of cultural objects as possible. The latter is strictly interlinked with the availability of training databases. However, beside developing and implementing dedicated database for keeping track of stolen artefacts, another issue needs to be assessed when dealing with social media and dark web scraping; this issue is related with the definition of data standards and ontologies to drive Artificial Intelligence (AI) based system to retrieve, automatically, the required information.

To design and develop an innovative knowledge-based investigation system dedicated to the illicit trafficking of cultural property, and specifically for web scraping a knowledge-based modelling approach is needed. The latter needs to consider several variables (activities, individuals, organizations, locations, black-markets and illegal trade, products and their interconnections and links), and to analyse heterogeneous online contents (i.e. text, audio, video, images). Advanced AI tools for the automatic extraction and analysis of the vast amount of multimodal and multimedia contents on web platform require structured methodologies for capturing, modelling, inferring, processing, and storing knowledge in a human understandable form. It includes a symbolic learning paradigm to extract knowledge from neural networks and store it in the form of ontologies (Figure 5).
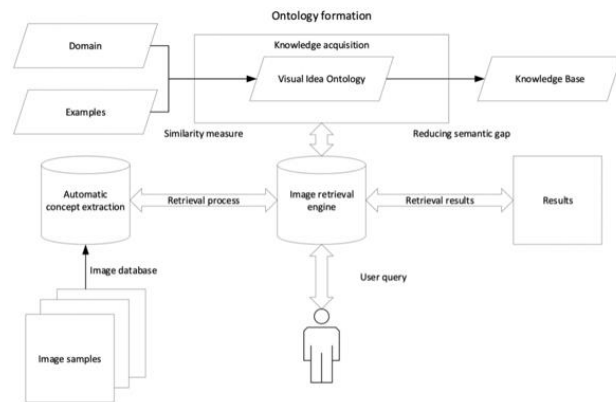

Figure 5. Workflow diagram for the creation and use of domain ontology.

The term "conceptualisation" is as a structured interpretation of a part of the world, used to think and communicate about it. The conceptualisation contains all the entities related to a particular area of interest (domain) and all the relationships between them. An ontology is a formal model that represents a domain of knowledge, based on specific requirements. It serves to describe the semantic of the data.

In the semantic web, ontologies are used to organise, formalise, publish, and retrieve information in an intelligent and efficient way.

The development of a series of functionalities, based on well-known ontologies, facilitates the transfer of domain knowledge/expertise to the expert users.

Any ontology can be described by its taxonomies, and its elements, included in the domain knowledge, can be represented in the adopted standard format:

- Entities are modelled as classes,
- Relationships are modelled as object properties,
- Attributes are modelled as data properties.

The OBJECT ID is an internationally recognized documentation standard conceived to identify and record cultural goods. It sets a standardised procedure to document and describe collections of archaeological, cultural, and artistic objects. By facilitating the identification of these objects, a standardised description can aid in their recovery in case of loss or theft (Thornes, 2000).

Object ID was developed in collaboration with the museum community, police and customs agencies, auction houses, and the insurance industry. It helps to combat the illicit trade of cultural heritage by encouraging the use of the standard and by bringing together organisations around the world that can encourage its implementation.

In case of theft, the information gathered and recorded using the Object ID norm can be checked against other databases of stolen artefacts, for example, the INTERPOL database of stolen works of art. Object ID was created as a practical tool for facilitating the recovery of stolen cultural goods and is now internationally recognised as a necessary and effective tool when inventorying a collection.

The Object ID standard defines nine categories of information as well as four steps to fulfil the procedure.

The categories are:

- Type of object,
- Materials and techniques,
- Measurements,
- Inscriptions and markings,
- Distinguishing features,
- Title,

- Subject,
- Date or period,
- Maker.

The four steps are divided as follows:
- Taking photographs of the object,
- Identifying the abovementioned categories,
- Writing a short description, including additional information,
- Keeping the constituted documentation in a secure place.

Documentation is indeed crucial for the protection of cultural objects, for police and customs officers can rarely recover and return objects that have not been photographed and adequately described. Police forces and customs administrations put into the custody of the museums and Ministries of Culture large numbers of objects that have been recovered during operations, but which cannot be returned to their rightful owners because there is no documentation that makes it possible to identify the victims.
The Object ID standard has reached a worldwide support. The majority of customs authorities of the member states of the European Union use it, and the Object ID checklist has already been translated into seventeen languages.

3.4 Ontology Definition

In order to define a cognitive based AI system, there is the need to identify, and develop, a specific ontology. As stated before, for the cultural heritage domain there are already existing ontologies, but the most complete and updated are the one chosen for the implementation of the SIGNIFICANCE project:
- ARCO (ARCO Development Team 2022),
- CULTURAL-ON (CULTURAL-ON Development Team 2022).

Given the specific typologies that have been identified for the case studies, classes and attributes that better fit with such ontologies have been identified (Figure 6).
In order to facilitate the visualization and the relations among these classes, Protègè (Protègè Development Team, 2022) has been used to visualize such elements. Few examples are defined in Figure 7.

3.5 Ontology-based Image classification

The definition of this ontology is essential for the implementation of the AI algorithms for the identification of illicit traffic. Indeed, the exploitation of domain specific ontologies for the image classification is a well know issue in the literature (Filali et al., 2020).
Due to lack of domain knowledge about the semantics of the image and image classification, the retrieval rate is usually unsatisfying. To improve this, image labels must be embedded in the dataset. Thanks to application of the aforementioned ontologies, it will be possible to extract relevant information from images, such as timeline, locations, visual features, and, at the same time, improve their interpretation.
Regarding the use of ontologies for image analysis, there has been a variety of recent studies in the field of cognitive computing (Ben et al., 2018). Moreover, with the rise of deep learning techniques it is worth to re-evaluate the benefits of applying them to existing object classification and recognition tasks.
One of the most important challenge when it comes to object recognition is the creation of image labelling system for the identification of ground truths. Some of the possibilities to create the ground truths involve image annotation by keywords, free text, or annotations based on ontologies which allows to add hierarchical structure to collection of keywords in order to produce a taxonomy.
Applying semantics greatly improve not only the overall performance of object recognition but also the performance and quality of individual tasks required for object recognition such as image segmentation. Furthermore, we have found that ontology can be used to substantially reduce semantic gap (i.e., the difference between the understanding of images by human and interpretation of images by machine), allowing for better automatization in training deep neural networks, as the dataset preparation can be offloaded to a ma- chine instead of being manufactured manually.

3.6 HPC Software and Hardware Configuration

High Performance Computing (HPC) architectures are definitely advantageous in the context of Artificial Intelligence and Deep Learning related activities. Graphical Processing Units (GPUs) and accelerated computing parallel programming tools (originally developed to benefit HPC applications) catalysed indeed the modern revolution in Artificial Intelligence and Deep Learning.
SIGNIFICANCE project will exploit the CyClone HPC Infrastructure of The Cyprus Institute for Artificial Intelligence and Deep Learning processing activities. It features 17 compute nodes equipped with 40 CPU cores each. An additionally 16 compute nodes come equipped with 40 CPU cores and 4 NVidia 32 GB V100 graphic cards each. Every node has a total of 192 GB RAM.
Based on the project required components, different type of resources from the CyClone HPC Infrastructure have been exploited to either enable the preparation and analysis of data or the hosting and execution of different parts of the platform To leverage the required resources, specific software deployment environments have been configured. Additionally, proper workflows to enable fast updates and bug fixing during the lifetime of the project have been designed and configured.
For any software being developed in the context of the SIGNIFICANCE project, the Git control- system version is used to allow to track-changes between code updates. It provides accountability and enable faster code development.
For the local "development environments", SIGNIFICANCE uses Docker.
Based on the nature of the SIGNIFICANCE components, two major types are identified based on their hardware resources needs:
- Computation Intensive Components (AI Modelling and Data Annotation),
- Memory Intensive Components (CH Database, Web Crawler, Data Collection, and the SIGNIFICANCE Platform).

A series of software environments (MongoDB, Python 3 Interpreter, CUDA + cuDDN support, Jupyter Notebooks) and libraries (TensorFlow 2, pyTorch, scikit-learn, matplotlib, OpenCV, Pandas, Numby) have been configured.
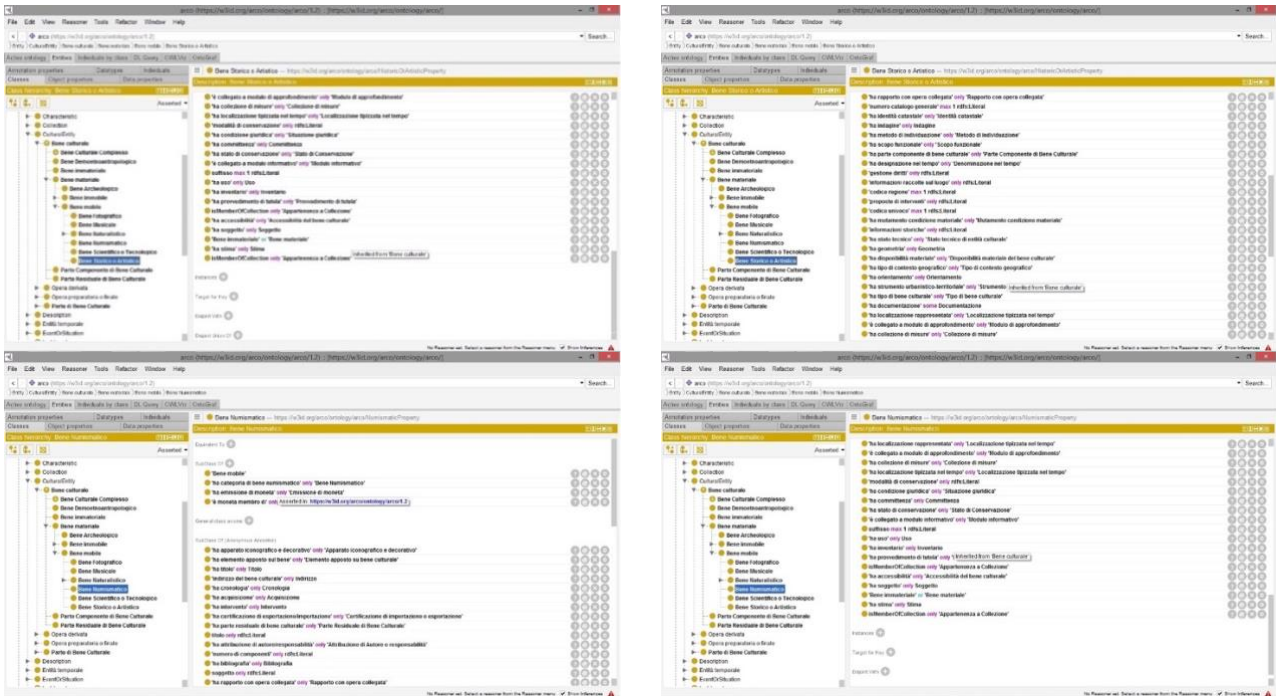
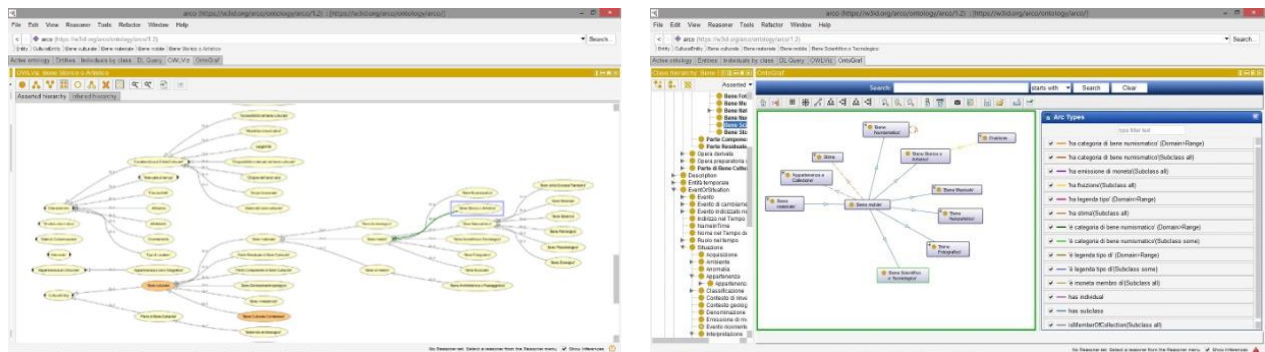Figure 6. Classes and attributes definition for the ontology.

Figure 7. Protegé visualization and the relations among these classes.

4. SIGNIFICANCE PLATFORM

SIGNIFICANCE platform will be an easily-accessible web application that will display metadata information on heritage objects illicitly traded on the internet. It will offer a web user interfaces for displaying all associated information exploitable for further investigations. This includes information of the object, images, activities or transactions, as well as additional details. More specifically, the platform will:

- feature an interface for CH Database (images and text),
- specify keywords and search criteria concerning surface and dark web crawlers,
- send notifications to authorities regarding new identified illegal transactions,
- provide the visualization of the collected information of the online artefact transactions,
- provide a User Interface (UI) with high usability and satisfaction for the end-users,
- be user friendly, comprehensive and easy to use.

SIGNIFICANCE Platform will be separated in two parts, the front-end and the back-end. The front end will act as the user interface of the whole system based on web building frameworks. The back end will be acting as the intermediate server between the UI and the Data Collection component, which is the database. The main visualization tool of the SIGNIFICANCE Platform will be the Dashboard View (Figure 8).

The user will be able to specify specific keywords (i.e., coin, Byzantine icon, etc.) and select the platform targets (i.e., ebay, catawiki, etc.) before the process of web crawling and AI analysis start. The retrieved activities will be listed including the name of the object, the seller, the link of the post and the estimation from the AI analysis (legal/illegal). The user will be able to notify the authorities for a single activity or several activities at once.
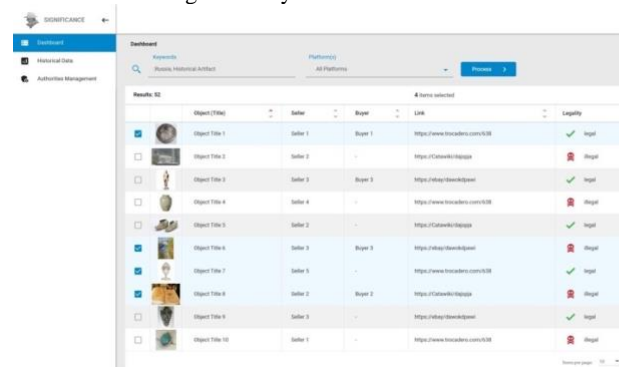

Figure 8. SIGNIFICANCE Platform: Dashboard mock-up.

The user will be able to visualize all the metadata regarding an identified online transaction and the related object (Figure 9). More specifically, the object metadata will be displayed as primary information. The latter will include:

- the title,
- the item description,
- the price,
- the link to the platform,
- the artistic/archaeological features (such as the technique, the material, the artist and the color),
- the data related to the originality and its provenance,
- the result of AI Modelling,
- the images of the object.

As far as the seller and buyer metadata are concerned, the platform will contain some basic information (examples can be name, mail and location) and some metrics related to the seller activity (such as the total number of the objects sold, the feedback received, etc.). When appropriate, the system will be notifying the authority for the identified illegal activity.

5. CONCLUSIONS

SIGNIFICANCE will contribute to the identification of illegally sold items on the internet and dark web using Artificial Intelligence (AI) and Deep Learning algorithms. The developed platform will allow LEAs and relevant authorities to undertake proper actions for identifying, tracking, and stopping illegal online auctions and unveiling criminal networks. By scanning forums and communication networks, SIGNIFICANCE will flag suspicious activities to competent authorities, which will be able to swiftly react and get a better understanding of the reach of online antiquities trafficking networks and combat them.
This will be held by using image and text AI methods as well as methodologies of new cyber forensics.
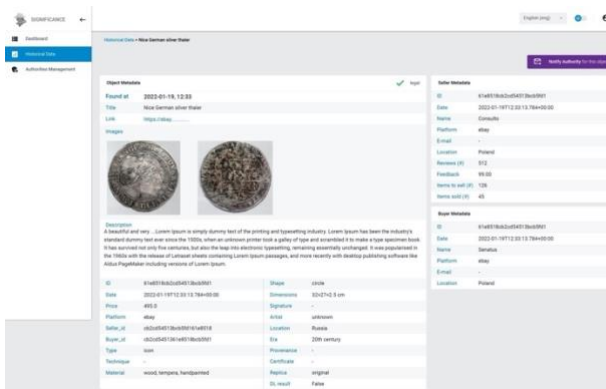

Figure 9. SIGNIFICANCE Platform: Object details mock-up

ACKNOWLEDGEMENTS

REFERENCES

ARCO Development Team 2022. http://wit.istc.cnr.it/arco (Last Access: 29/03/2022).

Ben Ellefi, M., Papini, O., Merad, D., Boi, J. M., Royer, J. P., Pasquet, J., ... & Drap, P. (2018, April). Cultural heritage resources profiling: Ontology-based approach. In Companion Proceedings of the The Web Conference 2018 (pp. 1489-1496).

Carabinieri for the Protection of Cultural Heritage- TPC, Development Team 2017.

http://tpcweb.carabinieri.it/SitoPubblico/psyche/generic?lang=EN (Last Access: 29/03/2022)

Carabinieri for the Protection of Cultural Heritage- TPC, Development Team 2020.

Carabinieri for the Protection of Cultural Heritage- TPC, Development Team 2022.

CULTURAL-ON Development Team 2022. https://dati.beniculturali.it/cultural-ON/ITA.html (Last Access: 29/03/2022).

Granell, E., Chammas, E., Likforman-Sulem, L., Martínez-Hinarejos, C. D., Mokbel, C., & Cîrstea, B. I. (2018). Transcription of spanish historical handwritten documents with deep neural networks. Journal of Imaging, 4(1), 15.

INTERPOL Development Team 2021. https://www.interpol.int/en/Crimes/Cultural-heritage-crime/ID-Art-mobile-app (Last Access: 29/03/2022)

INTERPOL Development Team 2022. https://www.interpol.int/en/Crimes/Cultural-heritage-crime/Stolen-Works-of-Art-Database (Last Access: 29/03/2022)

Paolanti, M., Kaiser, C., Schallner, R., Frontoni, E., & Zingaretti, P. (2017, September). Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks. In International conference on image analysis and processing (pp. 402-413). Springer, Cham.

Paolanti, M., & Frontoni, E. (2020). Multidisciplinary pattern recognition applications: a review. Computer Science Review, 37, 100276.

Protègè Developmnet Team 2022. https://protege.stanford.edu (Last Access: 29/03/2022).

PREVISION Development Team, 2020. http://www.prevision-h2020.eu (Last Access: 29/03/2022).

Toman, J. (2017). The protection of cultural property in the event of armed conflict: Commentary on the Convention for the Protection of Cultural Property in the Event of Armed Conflict and its Protocol, signed on 14 May 1954 in The Hague, and on other instruments of international law concerning such protection. Routledge.

Thornes, R., Dorrell, P.G., Lie, H. and Lie, P., 2000. Introduction to Object ID: guidelines for making records that describe art, antiques, and antiquities. Getty Publications.

Filali, J., Zghal, H. B., Martinet, J., 2020. Ontology-based image classification and annotation. International Journal of Pat-tern Recognition and Artificial Intelligence, 34(11), 2040002

APPENDIX

Image based training datasets are available among request for research purposes.