# AICat: An AI Cataloguing Approach to Support the EU AI Act

Delaram Golpayegani[1,*], Harshvardhan J. Pandit[2] and Dave Lewis[1]

[1]ADAPT Centre, Trinity College Dublin, Dublin, Ireland
[2]ADAPT Centre, Dublin City University, Dublin, Ireland

**Abstract**

The European Union's Artificial Intelligence Act (AI Act) requires providers and deployers of high-risk AI applications to register their systems into the EU database, wherein the information should be represented and maintained in an easily-navigable and machine-readable manner. Given the uptake of open data and Semantic Web-based approaches for other EU repositories, in particular the use of the Data Catalogue vocabulary Application Profile (DCAT-AP), a similar solution for managing the EU database of high-risk AI systems is needed. This paper introduces **AICat**—an extension of DCAT for representing catalogues of AI systems that provides consistency, machine-readability, searchability, and interoperability in managing open metadata regarding AI systems. This open approach to cataloguing ensures transparency, traceability, and accountability in AI application markets beyond the immediate needs of high-risk AI compliance in the EU. AICat is available online at https://w3id.org/aicat under the CC-BY-4.0 license.

**Keywords**

AI Act, DCAT, AI catalogues, regulatory enforcement, trustworthy AI

## 1. Introduction

The European Union (EU) Artificial Intelligence Act (AI Act) [1], which entered into force on 1 August 2024, stands as a landmark legal regime for development and use of AI. Within the AI Act, there is a high demand for Regulatory Technology (RegTech) solutions to serve the foundational and technical backbone required for implementation and enforcement of the Act [2]. Drawing parallels with the EU digital regulations, notably the General Data Protection Regulation (GDPR) [3], and looking into the body of compliance and enforcement solutions proposed in regard to such regulations suggest adoption of Semantic Web for effective and scalable compliance and enforcement solutions.

In the context of the AI Act, one area where the Semantic Web is anticipated to be used is the implementation of the *EU database of high-risk AI systems*. The database, which is to be established and managed by the European Commission in collaboration with Member States, is intended to encompass information regarding high-risk AI systems as declared by their providers and deployers. From the technical perspective, the AI Act requires the information contained within the database to be *"easily navigable"* and *"machine-readable"* (Art. 71 (4)), with different levels of accessibility, i.e. publicly and non-publicly accessible.

To implement and manage the EU database, and any catalogues of AI-related resources, a layer of metadata is needed to facilitate cross-referencing, traceability, transparency, interoperability, and comparability. The current state of existing repositories of AI systems, models, and datasets shows that adoption of machine-readable metadata is limited (see section 2). In this paper, we address this gap in AI repositories as well as the pressing need for the EU to establish the high-risk AI database by proposing **AICat** as a cataloguing approach. AICat extends the Data Catalog Vocabulary (DCAT) [4], enabling describing AI systems and components, including AI models and datasets, in catalogues through a consistent, standardised, and interoperable mechanism. This leads to the contributions of this work as:

*AICS'24: 32nd Irish Conference on Artificial Intelligence and Cognitive Science, December 09–10, 2024, Dublin, Ireland*
*Corresponding author.
sgolpays@tcd.ie (D. Golpayegani); me@harshp.com (H. J. Pandit); delewis@tcd.ie (D. Lewis)
0000-0002-1208-186X (D. Golpayegani); 0000-0002-5068-3714 (H. J. Pandit); 0000-0002-3503-4644 (D. Lewis)

1. An in-depth analysis of the AI Act's registration requirements for providers and deployers of high-risk AI systems,
2. **AICat**, an extension of DCAT that provides a mechanism for cataloguing AI systems and their incorporating components in registries of AI systems, including but not limited to the EU database of high-risk AI systems.

## 2. Related Work

With the proliferation of AI models, systems, and use cases, open AI repositories and commercial marketplaces have been created to facilitate the discovery and sharing of resources [5]. This section investigates the literature to identify related studies that address registering and sharing metadata about AI systems and their risks, in particular within the EU database of high-risk AI systems, using Semantic Web technologies.

Currently, there are a few well-known repositories that leverage metadata for describing resources. The **Hugging Face Hub**[1] is a centralised repository of open-source models and datasets, wherein each model or dataset is accompanied with metadata describing them. This enables discovery, sharing, and filtering of resources available on Hugging Face's Model and Data Hubs through the use of open JSON-based metadata. The Hub contains a repository of Spaces, i.e. ML demo apps, which, unlike the Model and Data Hub, do not support the inclusion of documented information and structured metadata. Similarly, **Kaggle** provides repositories of Datasets[2] and Models[3], where datasets, models, and generative AI applications are indexed and documented using detailed Data and Model Cards. Using the Kaggle repository, datasets and models can be published, shared, tagged, searched, and sorted. Compared to Hugging Face Data Hub which supports indexing only open-source resources, Kaggle Datasets allows for sharing metadata about both proprietary and publicly available datasets. The **AI-on-Demand (AIoD) platform**[4] is a European-funded project that serves as a community-driven hub for cataloguing AI-related solutions and components that contribute to the European ecosystem of AI excellence and trust. AIoD's asset catalogue[5] covers a wide range of resources including datasets, libraries, ML models, AI services, tools, use cases, and even tutorials. AIoD also provides JSON-based metadata for describing resources[6].

**Croissant** [6] is a framework developed by MLCommons—a non-profit open AI engineering consortium that enables expressing metadata for datasets with a focus on information that is essential in machine learning workflows. The Croissant vocabulary[7] is an extension of `schema.org/Dataset` vocabulary for metadata of ML datasets, which is expressed in the JSON-LD format. The Croissant framework is supported by a user-friendly tool to assist non-technical users in creation and modification of metadata. Although it is not a dataset repository, it has been integrated with existing data repositories, including HuggingFace, adding a layer of metadata.

While the information in the aforementioned registries is mostly presented in semi-structured formats such as JSON, none of them follow standardised approaches for data sharing or cataloguing. In regard to standardised approaches, the Data Catalog Vocabulary (DCAT) [4]—the W3C's recommended vocabulary for publishing data catalogues—and particularly its application profile for data portals in Europe (DCAT-AP) [7] have been adopted by the European Commission to promote open, standardised, and interoperable data sharing, prominently in the European Data Portal (EDP)[8], which is the central point of access to open data provided by the EU's public agencies [8]. Recently, **MLDCAT-AP** [9] has been proposed as an extension of DCAT-AP for including information about machine learning models

[1] https://huggingface.co/docs/hub/index
[2] https://www.kaggle.com/datasets
[3] https://www.kaggle.com/models
[4] https://aiod.eu/
[5] https://www.ai4europe.eu/research/ai-catalog
[6] https://api.aiod.eu/redoc
[7] https://docs.mlcommons.org/croissant/docs/croissant-spec.html
[8] https://data.europa.eu/en

in data catalogues. One of the distinguishing features of MLDCAT-AP is inclusion of information about *risks* associated with ML models.

Of relevance to the contributions of this work is the **Data Processing Catalogue (DPCat)** [10], which is an extension of DCAT and DCAT-AP that enables representing, maintaining, and exchanging ROPA[9]-related information in the form of datasets and catalogues. DPCat further enables creating documentation to address the GDPR's ROPA requirements.

Table 1 shows a comparison of existing approaches for cataloguing AI, models, and datasets. Currently, providing metadata, typically in JSON format, regarding datasets and models is an established practice. However, there is little attention to cataloguing AI systems and consequently there is no standardised machine-readable vocabulary that supports cataloguing of AI systems as well as their incorporating components.

**Table 1**
Comparison of AI cataloguing approaches (a black circle (•) indicates the criterion is satisfied, while a blank circle (○) indicates that it is not)

| Work | Scope | Format of metadata | Use of standardised vocabularies |
|---|---|---|---|
| **Repositories** | | | |
| **Hugging Face Data/Model Hub** | Dataset/Model | JSON | ○ |
| **Kaggle dataset/model repository** | Dataset/Model | HTML | ○ |
| **AI-on-Demand (AIoD) platform** | AI assets (dataset, model, services) | JSON | ○ |
| **Approaches** | | | |
| **Croissant** [6] | Dataset | JSON-LD | ○ |
| **MLDCAT-AP** [9] | ML models | JSON-LD | • |
| **DPCat** [10] | GDPR's ROPA | Turtle | • |

## 3. Analysis of the AI Act's Registration Requirements

Under the AI Act, providers and deployers of Annex III high-risk AI systems and providers of non-high-risk Annex III systems, i.e. systems that meet the conditions of Annex III but are considered as non-high-risk by the provider, are required to register their systems into the EU database (Article 49). According to Article 71, the EU database should be set up and maintained by the European Commission, in collaboration with the Member States. It shall be *"accessible and publicly available"* (with some exceptions), provided in a *"user friendly manner"*, and should be *"easily navigable and machine-readable"*. The EU database aims to act as an instrument for the Commission and the Member States to facilitate monitoring the current uptake of Annex III AI systems—regardless of their associated risk category—within the EU market and to serve as a transparency measure for sharing information regarding such systems with the public (Article 71 and Recital 131). The EU database therefore is a key data interoperability point between the Commission, AI providers, AI deployers, and the public.

Table 2 provides a summary of the registration provisions specified in Article 49. As shown in Table, the list of information elements that should be registered and their level of openness, i.e. publicly accessible or not, depends on the role of the registrant and the type of the system. In this, notably, submitting information regarding incorporating AI models, whether they are general-purpose or not, is not needed. However, information about general-purpose AI models should be made available to downstream AI providers that intend to use the model within their systems (Article 53).

Annex VIII, wherein the information to be submitted upon the registration of high-risk AI systems is outlined, was analysed to identify the *general* information that should be provided when registering an AI system into the EU database. Detailed information, such as the system's logic, instructions for use,

[9]Register of Processing Activities

and summary of fundamental rights impact assessment are not included, due to their descriptive nature and the lack of guidelines. In addition, for the general description of the general-purpose AI model, the key information elements listed in Annex XII, Point 1, were included to enable representation of AI components. Table 3 shows the key information elements extracted from Annex VIII and XII.

**Table 2**
Registration requirements for high-risk AI systems under the EU AI Act

| AI Act Article | AI System | Where? | What Information? | Who? | When? |
|---|---|---|---|---|---|
| 49(1) | High-risk as per Annex III, P. 3, 4, 5, 8 | Public EU database | Annex VIII (A) | AI provider or authorised representative | Before placing on the market or putting into service |
| 49(1) & (4) | High-risk as per Annex III, P. 1, 6, and 7 | Non-public EU database | Annex VIII (A), points 1 to 10 (except 6, 8, and 9) | AI provider or authorised representative | Before placing on the market or putting into service |
| 49(2) | Meets Annex III, P. 2, 3, 4, 5, 8 conditions but non-high-risk as per assessment of the provider | Public EU database | Annex VIII (B) | AI provider or authorised representative | Before placing on the market or putting into service |
| 49(2) & (4) | Meets Annex III, P. 1, 6, & 7 conditions but non-high-risk as per assessment of the provider | Non-public EU database | Annex VIII (B), points 1 to 5 & points 8 & 9 | AI provider or authorised representative | Before placing on the market or putting into service |
| 49(3) | High-risk as per Annex III, P. 3, 4, 5, 8 | Public EU database | Annex VIII (C) | AI deployer (public authorities, Union institutions, bodies, offices, or agencies) | Before putting into service or using |
| 49(3) & (4) | High-risk as per Annex III, P. 1, 6, & 7 | Non-public EU database | Annex VIII (C), points 1 to 3 | AI deployer (public authorities, Union institutions, bodies, offices, or agencies) | Before putting into service or using |
| 49(5) | High-risk as per Annex III, P. 2 | Register at national level | Not mentioned | Not mentioned | Not mentioned |

**Table 3**
Key information elements to be registered into the EU database

| Annex | Clause | Requirement |
|---|---|---|
| Information about **operators**, including providers and deployers | | |
| VIII | A1, B1 | AI provider's name |
| | A1, B1 | AI provider's address |
| | A1, B1 | AI provider's contact details |
| | C1 | AI deployer's name |
| | C1 | AI deployer's address |
| | C1 | AI deployer's contact details |
| Information about **AI system** | | |
| VIII | A4, B4 | AI system's trade name |
| | A4, B4 | AI system's additional reference |
| | A5, B5 | AI system's intended purpose |
| | A7, B8 | AI system's market status |
| | A10, B9 | Countries where system is available |
| Information about **components**, i.e. datasets and models | | |
| VIII | A6 | Data used by the system |
| | A6 | Input data used by the system |
| | A5, B5 | Component's intended purpose |
| | – | AI models used within the system |
| XII | 1-1b | Model's use policy |
| | 1-1c | Model's date of release |
| | 1-1g | Model's input data |
| | 1-1g | Model's output data |
| | 1-1h | Model's license |

# 4. AICat - a DCAT Extension for Cataloguing AI Systems

AICat is an application profile for specifying catalogues of AI systems that avails a thin layer of metadata to enhance interoperability and cross-referencing within the AI value chain. Building upon DCAT version 3, which supports cataloguing varying resources, AICat enables providing metadata about AI systems, models, and datasets. In addition to the resources that are already used by DCAT, AICat reuses existing concepts from our previous work, including the Data Privacy Vocabulary (DPV)[10] [11], its technology extension[11], and the AI Risk Ontology (AIRO)[12] [12].

The key information elements identified from the AI Act's registration obligations, discussed in section 3, shape the functional requirements of AICat. These requirements, which are expressed in the form of competency questions following the methodology described in [13], are shown in Table 4.

## 4.1. AICat Overview

AICat extends DCAT version 3 [4], since this version of DCAT supports cataloguing resources beyond datasets. By extending DCAT, AICat aims to scale the cataloguing to include AI systems and models to address the needs of the EU database. Table 5 illustrates how the identified requirements are mapped into concepts from DCAT, AIRO, DPV, and DPV's TECH extension (for prefixes refer to Listing 2). As shown in the Table, the *intended purpose* of a system is represented as a policy modelled using the AI Use Policy (AIUP) profile [14], which is an extension of the Open Digital Rights Language (ODRL) [15], for expressing AI use offers, requests, and agreements between multiple parties across the AI value chain.

[10] https://w3id.org/dpv/
[11] https://w3id.org/dpv/tech
[12] https://w3id.org/airo

**Table 4**
AICat profile requirements specification

| **AICat Requirements Specification Document** | | |
|---|---|---|
| **1. Purpose** | | |
| The purpose of the AICat profile is to use DCAT and AIRO to describe catalogues of AI systems and their associated components, such as datasets and AI models. | | |
| **2. Scope** | | |
| The scope of AICat is limited to the *atomic* information that should be submitted upon the registration of high-risk AI systems into the EU database, outlined in Annex VIII. This means that descriptive information, for instance the system's logic and findings of the fundamental rights impact assessment, is not included in the scope. | | |
| **3. Implementation Language** | | |
| OWL, DCAT | | |
| **4. Key Uses** | | |
| USE 1. Maintaining and managing metadata about AI systems, datasets, and models in interoperable and standardised catalogues.<br>USE 2. Discovering and comparing AI solutions.<br>USE 3. Cataloguing and sharing information about AI systems with the public in a transparent manner. This includes the use by the European Commission for sharing metadata of the high-risk AI systems indexed in the EU database. | | |
| **5. Ontology Requirements** | | |
| **a. Non-Functional Requirements** | | |
| NFR 1. AICat shall be published online with standard documentation.<br>NFR 2. AICat shall reuse concepts and relations from existing ontologies, including AIRO, to the fullest extent possible. | | |
| **b. Functional Requirements: Groups of Competency Questions** | | |
| **CQG1. AI systems** | **CQG2. Datasets** | **CQG3. AI models** |
| CQ1-1. What is the name of the system?<br>CQ1-2. Who is the system's provider?<br>CQ1-3. Who is the system's deployer?<br>CQ1-4. What is the system's intended purpose?<br>CQ1-5. What is the system's market availability status?<br>CQ1-6. In which countries is the system made available?<br>CQ1-7. What are the additional references to the system? | CQ2-1. Which datasets are used by the system?<br>CQ2-2. What is the system's input data?<br>CQ2-3. What is the dataset's use policy? | CQ3-1. Which models are used by the system?<br>CQ3-2. What is the model's release data?<br>CQ3-3 What is the model's input data?<br>CQ3-4 What is the model's output data?<br>CQ3-5. What is the model's license?<br>CQ3-6. What is the model's use policy? |

Figure 1 depicts an overview of AICat's information model. As illustrated in the Figure, `aicat:Catalog` is a sub-class of `dcat:Catalog` that provides a curated collection of metadata about AI systems, models, and datasets. AICat extends DCAT by introducing `airo:AISystem` and `airo:AIModel` as sub-classes of `dcat:Resource`, enabling inclusion of their metadata in an `aicat:Catalog`. Given that `airo:Data` is a sub-class of `dcat:Dataset`, cataloguing data is also supported by AICat. While the inclusion of AI systems was directly linked to the scope of the EU database, whose aim is to index AI systems, the inclusion of models and datasets was driven by the existing focus in the state of the art on cataloguing these AI components, as reviewed in section 2.

`aicat:system`, `aicat:model`, and `dcat:dataset` are sub-properties of `dcat:resource` that allow linking the catalogue to the resources indexed therein. To enable modelling the relationships between the resources, for example to show which datasets used for training a model, `airo:hasTrainingData`, `airo:hasTestingData`, `airo:hasValidationData`, `airo:hasInput`, `airo:hasOutput`, and `airo:hasModel` are reused from AIRO.

AICat's documentation was generated using WIDOCO [16] and is available online at https://w3id.org/aicat under the CC-BY-4.0 license.

By following DCAT-AP [7], AICat can further distinguish between *mandatory*, *recommended*, *optional*, and *deprecated* elements based on the requirements of the AI Act. Even though implementing such normative profiles can easily be realised by defining the aforementioned property types for each of the information elements, in the context of the AI Act, identification of whether provision of an information element is mandatory, recommended, optional, or deprecated requires additional guidelines and codes of conduct.

**Table 5**
Specifications for representing AI systems and models in AICat

| CQ | AI Act Annex | Requirement | Metadata Field | Range |
|---|---|---|---|---|
| Information about **AI system** | | | | |
| 1-1 | VIII, A4 & B4 | AI system's trade name | `dct:title` | `rdfs:Literal` |
| 1-2 | VIII, A1 & B1 | Provider's information | `airo:isProvidedBy` | `airo:AIProvider` |
| 1-3 | VIII, C1 | Deployer's information | `airo:isDeployedBy` | `airo:AIDeployer` |
| 1-4 | VIII, A5 & B5 | AI system's intended purpose | `odrl:hasPolicy` | `aiup:UsePolicy` |
| 1-5 | VIII, A7 & B8 | AI system's market status | `tech:hasMarketAvailabilityStatus` | `tech:MarketAvailabilityStatus` |
| 1-6 | VIII, A10 & B9 | Countries where system is available | `dpv:hasCountry` | `dpv:Country` |
| 1-7 | VIII, A4 & B4 | AI system's additional reference | `dct:isReferencedBy` | `dcat:Resource` |
| Information about **components** | | | | |
| 2-3 | VIII, A5 & B5 | Component's intended purpose | `odrl:hasPolicy` | `aiup:UsePolicy` |
| Information about **datasets** | | | | |
| 2-1 | VIII, A6 | Data used by the system or model | `airo:hasTrainingData`, `airo:hasValidationData`, `airo:hasTestingData` | `airo:Data` |
| 2-2 | VIII, A6 | Input data used by the system | `airo:hasInput` | `airo:Data` |
| Information about **models** | | | | |
| 3-1 | – | AI models used within the system | `airo:hasModel` | `airo:AIModel` |
| 3-2 | XII, 1-1c | Model's date of release | `dct:issued` | `xsd:date` |
| 3-3 | XII, 1-1g | Model's input data | `airo:hasInput` | `airo:Data` |
| 3-4 | XII, 1-1g | Model's output data | `airo:hasOutput` | `airo:Data` |
| 3-5 | XII, 1-1h | Model's license | `airo:hasLicense` | `airo:License` |
| 3-6 | XII, 1-1b | Model's use policy | `odrl:hasPolicy` | `aiup:UsePolicy` |

AICat is introduced as a minimal extension of DCAT. This extension introduces the `aicat:Catalog` class and its relations with `airo:AISystem` and `airo:AIModel`, both added as new types of `dcat:Resource`. One of the key directions for improving AICat is using the Shapes Constraint Language (SHACL) [17] to specify the level of necessity for information elements—which can be mandatory, recommended, or optional. Listing 1 shows an example of a SHACL shape indicating that each AI system should have at least one provider. Currently, AICat does not define such a normative profile due to the absence of recommendations and guidelines in regard to the AI Act.

```
@prefix sh: <http://www.w3.org/ns/shacl#> .
@prefix airo: <https://w3id.org/airo#> .
:AIProviderShape a sh:NodeShape;
                                    sh:targetClass airo:AISystem ;
                                    sh:property [
                                      a sh:PropertyShape ;
                                      sh:path airo:isProvidedBy;
                                      sh:minCount 1 ] .
```

Listing 1: Example of a SHACL shape that specifies the requirement for presence of at least one provider for an AI system

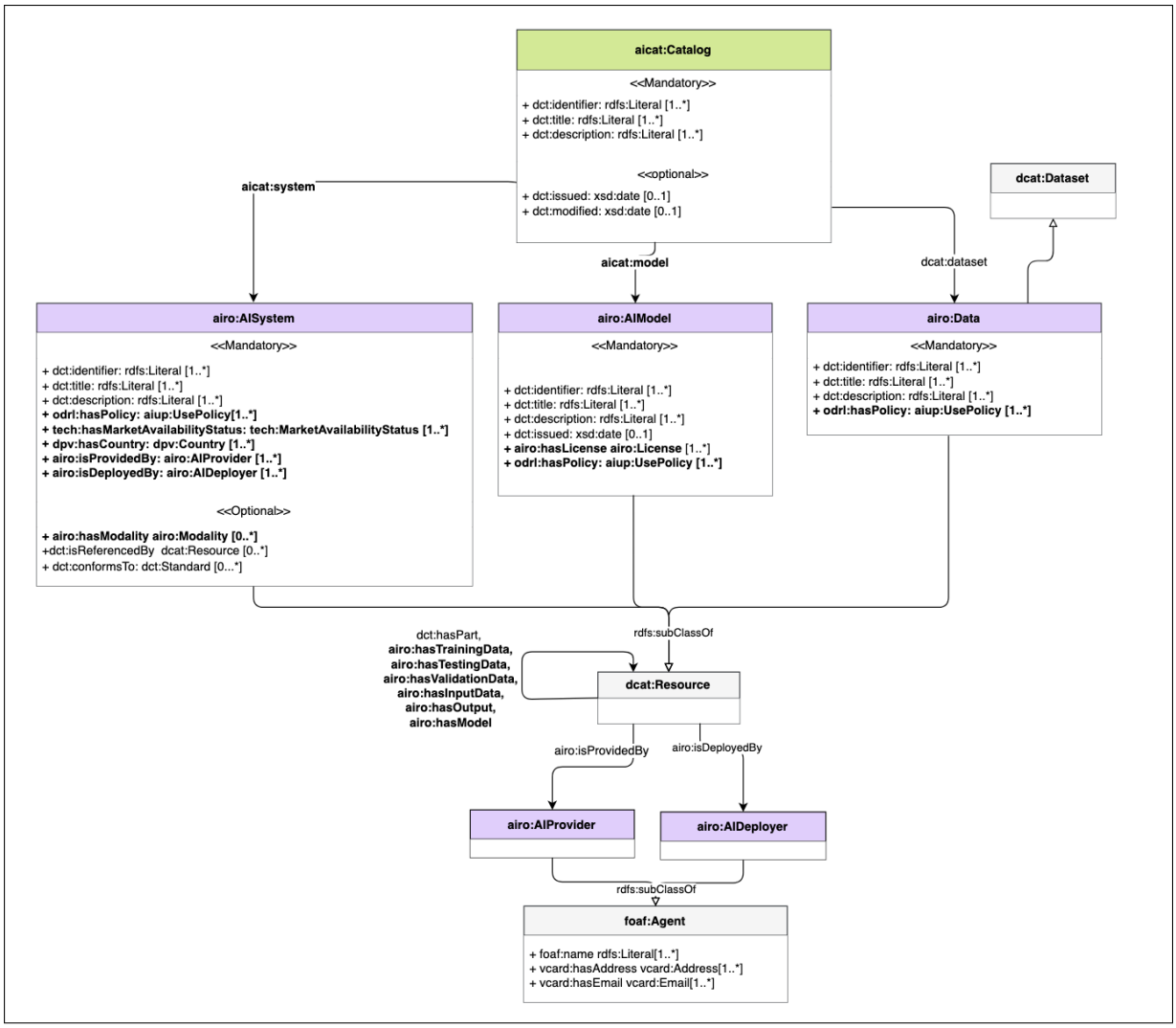


**Figure 1:** An overview of the AICat Profile

## 4.2. Proof-of-Concept Implementation

To illustrate an example of cataloguing, we use an example of an AI-based student proctoring system, described in [18, 19, 20]. The system, which is called *Proctify*, is provided by *AIEduX* and intended to detect suspicious behaviour during online exams by analysing facial behaviour. The system incorporates a facial analysis toolkit, provided by a third party, to analyse a person's facial information, including the head pose, gaze direction, and face landmarks' positions. This extracted information is then provided as an input to a model, *SusBehavedModel*, which has been trained in-house by the system's provider using *SusBehavedDataset*, to determine whether the student is displaying suspicious behaviour. Listing 2 presents a summarised version of an `aicat:Catalog` that contains metadata about *Proctify* and its components. As shown in the Listing, the policies for using the AI system and its incorporating components are expressed using the AIUP profile.

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <https://www.w3.org/TR/rdf12-schema/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dcat: <https://www.w3.org/TR/vocab-dcat-3/> .
@prefix dpv: <https://w3id.org/dpv#> .
@prefix tech: <https://w3id.org/dpv/tech#> .
@prefix airo: <https://w3id.org/airo#> .
@prefix aiup: <https://w3id.org/aiup#> .
@prefix aicat: <https://w3id.org/aicat#> .
@prefix ex: <http://example.com/proctify#> .

ex:aieduX-catalogue-01 a aicat:Catalog, dcat:Catalog  ;
    dct:identifier "aiedux-cat01"^^xsd:string ;
    dct:title "AIEduX catalogue"@en ;
    dct:description "AI systems and models provided by AIEduX"@en ;
    dct:created "2024-05-05"^^xsd:date ;
    dcat:dataset ex:susbehaved_dataset ;
    aicat:model ex:susbehaved_model;
    aicat:system  ex:proctify .

ex:susbehaved_dataset a dcat:Dataset, airo:Data ;
   dct:identifier "aiedux-d012"^^xsd:string ;
   dct:title "SusBehavedDataSet"@en ;
   dct:description ".. includes suspicious behaviour data.."@en ;
   odrl:hasPolicy ex:susbehaved_dataset_policy  .

ex:susbehaved_model a dcat:Resource, airo:AIModel ;
    dct:identifier "aiedux-m022"^^xsd:string ;
    dct:title "SusBehavedModel"@en ;
    dct:description ".. determines suspicious behaviour .."@en ;
    dct:issued "2024-02-15"^^xsd:date ;
    airo:hasTrainingData ex:susbehaved_dataset ;
    odrl:hasPolicy ex:susbehavedmodel_policy .

:proctify a dcat:Resource, airo:AISystem ;
    airo:isProvidedBy ex:aiedux ;
    dct:identifier "aiedux-ai031"^^xsd:string ;
    dct:title "Proctify"@en ;
    dct:description "An AI-based proctoring system..."@en ;
    tech:hasMarketAvailabilityStatus tech:MarketAvailable ;
    dpv:hasCountry <http://dbpedia.org/resource/Italy> ;
    dcat:contactPoint <http://example.org/aieduX-AI031/contact> ;
    airo:hasModel ex:susbehaved_model ;
    odrl:hasPolicy ex:proctify_use_policy .

 ex:susbehaved_dataset_policy a aiup:UseOffer  .
 ex:susbehavedmodel_policy a aiup:UseOffer .
 ex:proctify_use_policy a aiup:UseOffer .
```

Listing 2: An example of `aicat:Catalog` for describing a catalogue

## 5. Potential Benefits of AICat

In terms of potential benefits, through reusing widely-used W3C standardised vocabularies, the AICat enables expressing metadata regarding AI systems and AI components within catalogues, wherein common vocabularies and open linked data-based formats are used. Therefore, the AICat addresses the AI market needs for a consistent and interoperable mechanism for cataloguing AI solutions [21], in a way that enables federated search and comparison across AI, model, and data catalogues offered by

different vendors—a crucial feature often required in AI procurement processes. In relation to this, the European Commission's dataset of selected uses of AI in the public sector [22] is a prominent resource, whose interoperability and searchability can be enhanced through adoption of a cataloguing mechanism such as AICat.

At the organisational level, AICat could assist AI providers and deployers in providing structured catalogues of AI systems and components. At the European level, a similar approach to AICat is expected to be adopted for the implementation of the database of high-risk AI systems as required by Article 71 of the AI Act. Given that AICat ensures traceability while protecting privacy by providing metadata without revealing sensitive information within a database, it supports the implementation of the non-public section of the EU database and provides a structure for registration forms. AICat potentially addresses the gap in the European open data portal in providing FAIR (Findable, Accessible, Interoperable, and Reusable) information regarding existing AI systems and models provided or deployed by public organisations. AICat also has the potential to promote cross-border interoperability required by the recently-enforced Interoperable Europe Act [23], particularly in the implementation of the *Interoperable Europe portal*—the EU's single point of entry for information related to cross-border interoperability of trans-European digital public services (Interoperable Europe Act, Article 8). In this, AICat can be employed to facilitate sharing information and best practices to support interoperability in public procurement of AI-based solutions.

Compared to existing cataloguing approaches, reviewed in section 2, AICat expands the scope of cataloguing to AI systems. From this literature review, MLDCAT-AP [9] bears a close resemblance to AICat, especially in the use of DCAT. MLDCAT-AP has been supported by the European Commission's Semantic Interoperability Community (SEMIC), and therefore it might be a candidate to be adopted in the implementation of the EU database. However, prior to this, it needs to be extended to include specifications of AI systems in the catalogue in alignment with the requirements of the AI Act. This can be realised by the integration of MLDCAT-AP and AICat. Another key feature of MLDCAT-AP, in comparison with AICat, is the inclusion of risk information in the catalogue. While AICat can support DCAT-based documentation of risks by reusing `airo:hasRisk`, in its current form it does not go beyond the general, non-descriptive information elements of Annex VIII, mainly due to the absence of related official guidelines.

## 6. Conclusion and Further Work

In this paper, we proposed AICat as a novel technical solution for cataloguing AI systems in an open, machine-readable, and interoperable format based on the evolving requirements of the AI value chain, particularly the requirements of the EU AI Act. Using AICat facilitates discovery, integration, and sharing information associated with AI systems and components amongst the stakeholders involved in the AI value chain based on the existing proven mechanism of (open) data portals.

By demonstrating this solution, we hope that similar open and interoperable approaches will be adopted in the implementation of the AI Act, in particular the creation of the EU database of high-risk AI systems as per Article 71. Our work also contributes to trustworthy and responsible use of AI by enabling creation of scalable and interoperable AI catalogues on the internet by using a unified and coherent vocabulary.

## Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813497 (PROTECT ITN). The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant#13/RC/2106_P2.

## References


[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024. URL: http://data.europa.eu/eli/reg/2024/1689/oj.

[2] P. Fehlinger, Enabling the responsible use of technology at scale – Why Europe needs a regulatory technology innovation ecosystem, Technical Report, Sitra, 2023.

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj.

[4] R. Albertoni, D. Browning, S. J. D. Cox, A. G. Beltran, A. Perego, P. Winstanley, Data Catalog Vocabulary (DCAT) - version 3, 2024. URL: https://www.w3.org/TR/vocab-dcat-3/, w3C Recommendation.

[5] A. Kumar, B. Finley, T. Braud, S. Tarkoma, P. Hui, Sketching an AI marketplace: Tech, economic, and regulatory aspects, IEEE Access 9 (2021) 13761–13774. doi:10.1109/ACCESS.2021.3050929.

[6] M. Akhtar, O. Benjelloun, C. Conforti, P. Gijsbers, J. Giner-Miguelez, N. Jain, M. Kuchnik, Q. Lhoest, P. Marcenac, M. Maskey, P. Mattson, L. Oala, P. Ruyssen, R. Shinde, E. Simperl, G. Thomas, S. Tykhonov, J. Vanschoren, J. van der Velde, S. Vogler, C.-J. Wu, Croissant: A metadata format for ML-ready datasets, DEEM '24, Association for Computing Machinery, 2024, p. 1–6. doi:10.1145/3650203.3663326.

[7] B. V. Nuffelen, DCAT-AP 3.0, 2024. URL: https://semiceu.github.io/DCAT-AP/releases/3.0.0/.

[8] F. Kirstein, B. Dittwald, S. Dutkowski, Y. Glikman, S. Schimmler, M. Hauswirth, Linked data in the european data portal: A comprehensive platform for applying DCAT-AP, in: Electronic Government, Springer International Publishing, 2019, pp. 192–204.

[9] A. Schiltz, E. Stani, MLDCAT-AP, 2024. URL: https://semiceu.github.io/MLDCAT-AP/releases/2.0.0/.

[10] P. Ryan, R. Brennan, H. J. Pandit, DPCat: Specification for an interoperable and machine-readable data processing catalogue based on GDPR, Information 13 (2022).

[11] H. J. Pandit, B. Esteves, G. P. Krog, P. Ryan, D. Golpayegani, J. Flake, Data privacy vocabulary (DPV) – version 2.0, in: The Semantic Web – ISWC 2024, Springer Nature Switzerland, Cham, 2025, pp. 171–193.

[12] D. Golpayegani, H. J. Pandit, D. Lewis, AIRO: An ontology for representing AI risks based on the proposed EU AI Act and ISO risk management standards, in: Towards a Knowledge-Aware AI, volume 55, IOS Press, 2022, pp. 51–65.

[13] M. C. Suárez-Figueroa, A. Gómez-Pérez, B. Villazón-Terrazas, How to write and use the ontology requirements specification document, in: R. Meersman, T. Dillon, P. Herrero (Eds.), On the Move to Meaningful Internet Systems: OTM 2009, Springer Berlin Heidelberg, 2009, pp. 966–982.

[14] D. Golpayegani, B. Esteves, H. J. Pandit, D. Lewis, AIUP: an ODRL profile for expressing AI use policies to support the EU AI act, in: Joint Proceedings of Posters, Demos, Workshops, and Tutorials of the 20th International Conference on Semantic Systems co-located with 20th International Conference on Semantic Systems (SEMANTiCS 2024), 2024.

[15] R. Iannella, M. Steidl, S. Myles, V. Rodríguez-Doncel, ODRL version 2.2 ontology, 2017. URL: http://www.w3.org/ns/odrl/2/, w3C Recommendation.

[16] D. Garijo, WIDOCO: A wizard for documenting ontologies, in: The Semantic Web – ISWC 2017, Springer International Publishing, 2017, pp. 94–102.

[17] H. Knublauch, D. Kontokostas, Shapes constraint language (SHACL), 2017. URL: https://www.w3.org/TR/shacl/, w3C Recommendation.

[18] C. Panigutti, R. Hamon, I. Hupont, D. Fernandez Llorca, D. Fano Yela, H. Junklewitz, S. Scalzo, G. Mazzini, I. Sanchez, J. Soler Garrido, et al., The role of explainable AI in the context of the AI

act, in: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023, pp. 1139–1150.
[19] I. Hupont, D. Fernández-Llorca, S. Baldassarri, E. Gómez, Use case cards: A use case reporting framework inspired by the european AI act, Ethics and Information Technology 26 (2024).
[20] D. Golpayegani, I. Hupont, C. Panigutti, H. J. Pandit, S. Schade, D. O'Sullivan, D. Lewis, AI cards: Towards an applied framework for machine-readable AI and risk documentation inspired by the EU AI act, in: M. Jensen, C. Lauradoux, K. Rannenberg (Eds.), Privacy Technologies and Policy, Springer Nature Switzerland, 2024, pp. 48–72.
[21] European Commission, Joint Research Centre, M. Manzoni, R. Medaglia, L. Tangi, C. Van Noordt, L. Vaccari, D. Gattwinkel, AI Watch, road to the adoption of artificial intelligence by the public sector – A handbook for policymakers, public administrations and relevant stakeholders, Publications Office of the European Union, 2022. doi:10.2760/288757.
[22] European Commission, Joint Research Centre (JRC), Selected AI cases in the public sector (jrc129301), 2021. URL: http://data.europa.eu/89h/7342ea15-fd4f-4184-9603-98bd87d8239a, dataset.
[23] Regulation (EU) 2024/903 of the European Parliament and of the Council of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act), 2024. URL: http://data.europa.eu/eli/reg/2024/903/oj.