RESEARCH

Open Access

Toward a standard formal semantic representation of the model card report

Muhammad Tuan Amith, Licong Cui, Degui Zhi, Kirk Roberts, Xiaoqian Jiang, Fang Li, Evan Yu and Cui Tao*

*Correspondence:
cui.tao@uth.tmc.edu

School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, USA

Abstract

Background: Model card reports aim to provide informative and transparent description of machine learning models to stakeholders. This report document is of interest to the National Institutes of Health's Bridge2AI initiative to address the FAIR challenges with artificial intelligence-based machine learning models for biomedical research. We present our early undertaking in developing an ontology for capturing the conceptual-level information embedded in model card reports.

Results: Sourcing from existing ontologies and developing the core framework, we generated the Model Card Report Ontology. Our development efforts yielded an OWL2-based artifact that represents and formalizes model card report information. The current release of this ontology utilizes standard concepts and properties from OBO Foundry ontologies. Also, the software reasoner indicated no logical inconsistencies with the ontology. With sample model cards of machine learning models for bioinformatics research (HIV social networks and adverse outcome prediction for stent implantation), we showed the coverage and usefulness of our model in transforming static model card reports to a computable format for machine-based processing.

Conclusions: The benefit of our work is that it utilizes expansive and standard terminologies and scientific rigor promoted by biomedical ontologists, as well as, generating an avenue to make model cards machine-readable using semantic web technology. Our future goal is to assess the veracity of our model and later expand the model to include additional concepts to address terminological gaps. We discuss tools and software that will utilize our ontology for potential application services.

Keywords: Knowledge representation, Ontology, Model cards, FAIR, Standardization, Semantic web, Machine learning, Artificial intelligence

Amith *et al. BMC Bioinformatics*   (2022) 23:281

Page 2 of 17

Background

The FAIR principle [1] serves as a set of guidelines for researchers and data publishers for practices involving the release of digital research assets to be *Findable* (encoded with discoverable, clear metadata for indexing), *Accessible* (persistent availability of data through unique identifiers), *Interoperable* (integration through the utilization of shared terminologies for metadata), and *Reusable* (data resources to be well-documented to be effectively integrated with other environments).

The National Institutes of Health (NIH) is invested in the use of artificial intelligence (AI) and machine learning (ML) that address the need for data and tools to unite biomedical and behavioral researchers [2]. The principles outlined by the FAIR guidelines can assist in this initiative towards tools and data that are not only FAIR, but "credible, ethical, and generalizable" to enhance scientific endeavors and enterprises [2]. Model cards are one of the resources encouraged by NIH through their Bridge2AI program [3]. Model card reports, a one to two-page document, inform researchers, software developers, stakeholders, and impacted individuals (like patients) of what a specific ML approach is capable of performing, and potential technical or ethical issues [4]. For a lay description, model card reports are akin to nutritional or drug labels mandated by the Food and Drug Administration (FDA) to inform the consumer of food and drug products.

Nonetheless, as a static document they do limit opportunities to permit the information contained within the report to be computable. If the information in the model card were processable by machines, this could allow for sophisticated opportunities such as tools and analytical tasks that can help advance FAIR guidelines. One technological approach are ontologies which are part of the technology stack of Tim Berners-Lee's Semantic Web vision, a vision in which the web evolves from a document presentation layer to a web of linked computable electronic resources [5].

Ontologies are representational artifacts that model domain information. Within these artifacts are complex networks of concept terms that abstract information and knowledge of a specific domain. The life sciences community has been major contributors to the body of biomedical ontological resources and as evident on the Linked Open Data Cloud [6–8]. Many of these life sciences ontology resources, particularly those that are found in the Open Biological and Biomedical Ontology (OBO) Foundry, are well-documented, comprehensive knowledge bases, and have rigorous standards for design [9]. The majority of the life sciences ontologies, utilize the Basic Formal Ontology (BFO), which allows for a common framework for interoperability between various life science-related knowledge bases [10].

For background purposes, a concept is a unit of thought that binds a symbol or sign (i.e. term) with a real entity in the world. An ontology utilizes machine-level syntax to represent this concept that connects with some real and verifiable entity. These concepts are broken down into types (i.e. classes or categories) to indicate the variation of the concept, basically elucidating a taxonomy. While ontologies can be strictly a hierarchical taxonomy, some (and more semantically richer) ontologies can be polyhierarchical networks of concepts. This is accomplished through links, sometimes referred to as properties or predicates, that connect the various concepts to express some statement about the world or domain. For example, given two concepts from the Software

Amith *et al. BMC Bioinformatics*   (2022) 23:281

Page 3 of 17

Ontology, *algorithm* and *software*, that are linked together by a property of *implements*, the ontology representation of *software > implements > algorithm* expresses a schema elucidating some domain knowledge. Furthermore, this schema can serve as a scaffold for instance data to evoke meaning. For example, *Acme Software* is a *software*, *differential evolution (acme-variant)* is an *algorithm*, and with our sample schema, one can derive that *Acme Software > implements > differential evolution (acme-variant)*. These expressions (i.e. axioms) can be coded using OWL2 (Web Ontology Language) [11], RDF (Resource Description Framework) [12], or Turtle [13] into the ontology artifact and linked to other expressions to form a knowledge base. The logic-based consistency of the encoded expressions can be checked using a variety of available semantic reasoners, including inferring additional information. To illustrate, given *is implemented by* that is a mirrored, inverse of *implements*, an inferred axiom can be generated of *differential evolution (acme-variant) > is implemented by > Acme Software*. Lastly, the ontology can be linked to other ontologies or distributed heterogeneous resources to expand the meaning of the data. In addition, the linking of heterogeneous data resources provides functions to query the data using a query language like SPARQL [14] or SQWLR [15].

Objective

The objective of this study is to initiate standardization of the information embedded in model card reports using formal ontology methods and resources. We present our initial work on an ontology called the Model Card Report Ontology (MCRO) to represent information and metadata from model cards in a formal conceptual structure of its report. This ontology leverages verified resources like the Basic Formal Ontology (BFO) [10], Information Artifact Ontology (IAO) [16, 17], Provenance Ontology (PROV-O) [18, 19], and the Software Ontology (SWO) [20, 21] to help enrich the core model of this ontology.

By using an ontology to construct our formal representation of model cards, we can potentially standardize the nomenclature and representation of model card reports, essentially producing a computable version of the model card that can be shared, reused, and linked. Noted earlier, this proposed ontology will utilize BFO-based ontologies that have an open-sourced, community consensus on terms and terminological structure. Utilizing these existing ontologies provides rigorous scientific standards shared by experts. Piggybacking on this last point, the usage of existing standards can also help expand the metadata scope to be included in the model and further elaborate on reporting. With our ontology that covers the metadata of the model card report, there is also the potential to index these reports that can assist in searching and querying. Furthermore, the ability to index can furnish the capacity to collate the information for analysis and aggregation. Combined with the expressiveness afforded through OWL2 to code ontologies, there is the potential to enrich the querying facilities and provide machine-driven inferences. In addition to describing the development of this ontology, we will outline specific use cases to point toward future direction of this work.

With an ontology representing the documentation structure of model cards we can transform model cards to a computable artifact - one that can link information about machine learning models, leverage semantic reasoning to link the information and open up software tooling opportunities, aggregation of model card data, and querying of the

Amith *et al. BMC Bioinformatics*   (2022) 23:281

Page 4 of 17

information for biomedical machine learning research. We demonstrate the application of the ontology by using two bioinformatics-related model card samples - one pertaining to a machine learning model for adverse effects of stent implantation and the other for social networks of HIV patients. Also we provide insight to our proposed software publication engine that leverages ontology-based reasoners, like FaCT++ [22], to manage the informational pieces of the model card report and allow sophisticated querying functionality. Overall the implication of this work could supplant static textual model cards reports for a machine readable version that would support FAIR-based principles and improve clinical informatics research.

Results

The resulting ontology was encoded for OWL/XML (.owl) and Turtle syntax (.ttl) and is available at our GitHub link [23]. The current class count of the final OBO-aligned MCRO numbers at 954 with 145 object properties, 25 data properties, and 2147 logical axioms. The core foundation of MCRO was 38 classes, 11 data properties, and 79 logical axioms. Table 1 shows the breakdown of these numbers according to the imported ontologies (SWO, IAO, SKOS, PROV-O).

Essentially, an encoded ontology (in OWL 2) describes axioms (i.e. logical statements) adhering to the conceptual schema level, which in turn, formalizes other encoded axioms that describe instance data. The task of the reasoner is to ensure that axioms satisfy these logical constraints and express information that is sound. Using our earlier example, *software* and *algorithm* are two class-level concepts that are disjointed, meaning there is no conceptual overlapping of instantiated data. If we incorrectly encode an assertion *differential evolution (acme-variant)* is a *software* AND an *algorithm*, this would result in the ontology's representational model to be logically inconsistent, and the software reasoner would indicate this inconsistency. While the example stated is simple, reasoners have a crucial role in validating the logical soundness of an ontology, particularly large ontologies with complex axioms. Considering that our work imports existing OBO ontologies, passing logical checks also ensures alignment with these imported ontologies which signals interoperability to the body of standardized biomedical ontologies.

Within the Protégé environment, there exist built-in facilities to utilize software reasoners to check the logic of the ontology. For this work, we used the FaCT++ reasoner (v1.6.5) that is generally known to accommodate various axioms types and have one of the fast performance [24]. Activating the reasoner will load the ontology, and the reasoner will preform the logic-based checks of the axioms. The Protégé environment will flag and alert the user of any unsatisfiable inconsistencies that need to be rectified in order for the ontology to pass the validation process. The resulting check with the FaCT++ reasoner (v1.6.5) evaluated the logical consistency of the final OBO-aligned MCRO, and it revealed no logical inconsistencies thereby passing the logic checks of the coded axioms.

Figure 1 shows how the ontology would operationalize a sample model card report. Each textual information of a model card report would map to a concept from our ontology model. The textual information from the model card is expressed as instance data of the ontology and has metadata annotation associated with it. At our disposal,

Amith *et al. BMC Bioinformatics*   (2022) 23:281

Page 5 of 17

the imported SKOS and PROV-O framework provides an extensive array of annotations properties including from Dublin Core.

We produced two sample model card documents pertinent to the bioinformatics domain. One titled "Adverse endpoints prediction for patients undergoing PCI" details a recurrent neural network model for prediction of adverse outcomes from coronary artery stent implantation, and the other titled "Incorporating social network information to predict HIV status" details the use of graph convolutional network to predict HIV infection from social network data. Similar to what was described in above, each textual part of the two documents were encoded into the final ontology model as instance data and aligned to their corresponding concept. To show an example, Fig. 2 displays a general view from the Protégé environment of an encoded model card report sample. Also Listing 1 shows the underlying encoding of model card instance for one of the samples in Turtle syntax.

Listing 1 Sample Turtle syntax for one of the sample model cards.

```
1  ###  http://sbmi.uth.edu/ontology/mco#mc2_example_card
2  mco:mc2_example_card  rdf:type  owl:NamedIndividual,  mco:Model_Card_Report ;
3
4  <http://purl.obolibrary.org/obo/BFO_0000051>
5    mco:mc2_consideration_info  ,
6    mco:mc2_model_details  ,
7    mco:mc2_model_parameters  ,
8    mco:mc2_quantative_analysis ;
9
10 <http://purl.obolibrary.org/obo/IAO_0000136> mco:mc2_algrothim ;
11
12 <http://purl.obolibrary.org/obo/IAO_0000116>
13   "This is an example of model card report" ;
14
15 <http://purl.org/dc/elements/1.1/title>
16   "Model Card: Incorporating social network information
17   to predict HIV status" ;
18
19 <http://www.geneontology.org/formats/oboInOwl#created_by> "Evan Yu" ;
20
21 <http://www.geneontology.org/formats/oboInOwl#creation_date>
22   "January 26, 2022" .
```

Essentially, each major concept from the MCRO model (refer to Fig. 3 for details) is utilized as instance data and annotated with metadata and linked to text of the model
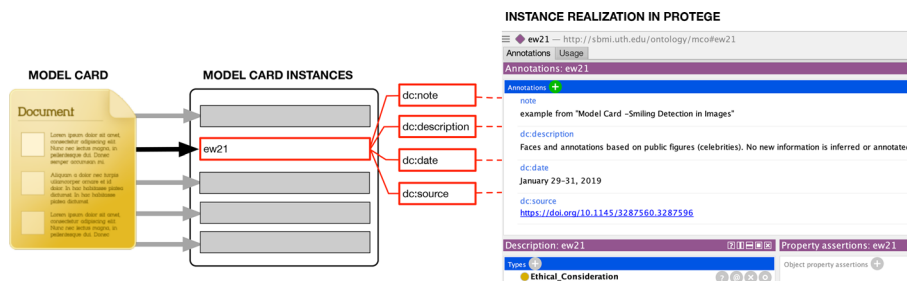

Fig. 1 An example showing how pieces of the textual information are mapped to instances of the Model Card Report Ontology concept. "ew21" is the identifier associated with a part of the model card that has links to data annotations (Dublin Core's *note*, *description*, *date*, *source*). In this example, we used the text from the Ethical Consideration section from the inception paper [4]
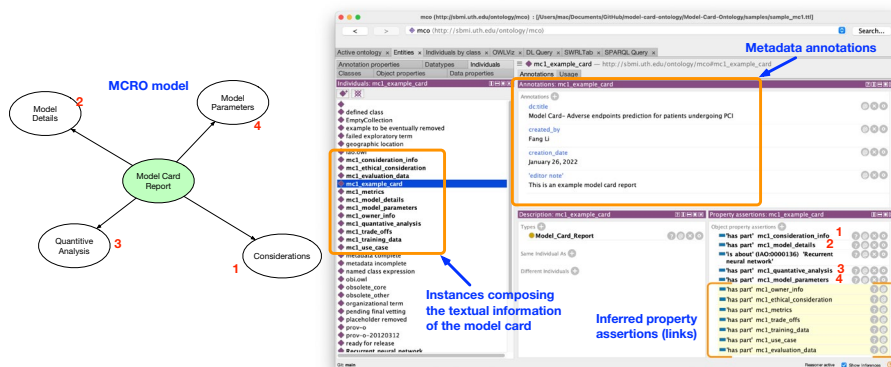
Fig. 2 A descriptive visualization showing the instances displayed through Protégé and their mapping to a segment of the MCRO model. The numbers from MCRO model (right) demonstrates how the model is manifested in the ontology artifact with matching numbers in the Protégé viewer. The yellow highlights shows the inferred assertions resulting from the OWL2 semantics and the FaCT++ reasoner

Table 1 Account of number of ontology features for the incorporated ontologies in MCRO

	Classes	Object properties	Data properties	Logical axioms
MCRO-Core	38	0	11	79
IAO	259	50	4	510
PROV	31	44	6	181
SWO	621	34	3	1335
SKOS	5	17	1	42
MCRO	954	145	25	2147

Italics indicate imported ontologies for MCRO

card (see Fig. 7). Both Figs. 7 and 8 show text of the model card represented as data attributes, e.g., in Fig. 7 as *documentation* for the *Trade off* and in Fig. 8 as the *overview* attribute for *Model Details.*

Also revealed in the figures are machine-based inferences generated by the semantics of the coding. In Fig. 2, due to the transitivity of the *has part* object property (and its inverse of *part of*), the FaCT++ reasoner produces all of the instances belonging to the model card sample beyond the first degree connection. For the sample in Fig. 7, the instance for *Consideration* information content coded that it has *Trade off* content part (See Listing 2, lines 6-7). *BFO_0000051* (*has part*) is defined as transitive property with an inverse equivalent (*part of*). This produces the inferred *part of* connection seen in Fig. 7.
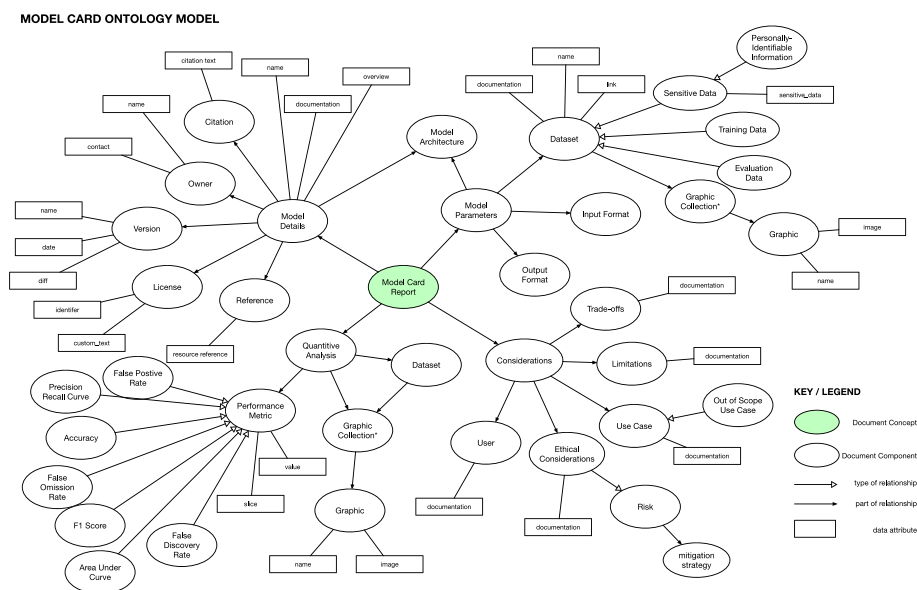
MODEL CARD ONTOLOGY MODEL

Fig. 3 Visualization of the abstraction for the Model Card Report Ontology

Listing 2 Turtle encoding for consideration instance for one of the sample model cards.

```
1  ###   http://sbmi.uth.edu/ontology/mco#mc1_consideration_info
2  mco:mc1_consideration_info rdf:type
3     owl:NamedIndividual ,
4     mco:Consideration_Information ;
5
6  <http://purl.obolibrary.org/obo/BFO_0000051> mco:mc1_ethical_consideration ,
7     mco:mc1_trade_offs ,
8     mco:mc1_use_case ;
9     mco:documentation
10      "The potentially relevant factors include age, gender, history
11      (e.g., prior bleeding, prior myocardial infarction),
12      comorbidities and risk factors (diabetes mellites, smoking,
13      congestive heart failure), as well as comedications
14      (e.g., statins, beta-blockers)"
```

This is also evident in the example in Fig. 8. The *Model Detail* instance has content part *User* information instance. Yet in this example, because of the transitive function of *BFO_0000051*, the reasoner also infers that *User* information instance (University of Texas Health Science Center, refer to "4" of Fig. 8) is also part of the instance of *Model Detail* and the referring model card. Both samples are available at the our GitHub repository to illustrate the overall coverage of the ontology model and the ability to translate document model cards to a computational format.

Discussion

Model card reports are one tool suggested by the National Institute of Health for their Bridge2AI initiative to advance behavioral and biological sciences impacted by research utilizing machine learning models. In this paper, we discussed our effort in developing the Model Card Report Ontology to represent and standardize the information in

a computable format to allow the reports to be machine-readable. This work is the first effort to develop an ontology model toward making model cards reports computable and serve scientific endeavors of FAIR for biomedical machine learning models.

To enforce some rigorous standard to our model, we utilized existing ontology resources from the OBO Foundry that adheres to strict principles for standardizing their ontology models. The benefit of using these community-agreed standards is the semantic precision that can be leveraged to evoke meaningful data and information. For example, we heavily used the partonomy relationship of *has part*. Partonomy relationships between concepts have varied and nuanced interpretations. A quick glance of this type of property [25] provide over 10 types that express very granular and nuanced meaning. Likewise, the same can be said of the entity categories that many bio-ontologies are aligned with in the BFO framework. Overall as noted by Obrst and colleagues [26], the more semantically rich the ontology, the more useful the artifact will be for sophisticated tools and applications. This is in effect would help formalize model card reports for application consumption and distribution.

Using our current version of the Model Card Report Ontology, we were able to transform the model card samples into ontology-based format. An issue with model card reports is that they are static documents. With an ontology-based representation we provide machine-readable artifact that can link to heterogeneous digital resources using a composition of formal ontologies. Also as ontology-based representational artifact we provide a formal structure for consistency and manage to derive coverage for the model card report.

While the ontology manages to cover structure and concepts of the model card report, we were limited by the incorporated ontologies. For example, specific types of neural network algorithms were not available from SWO, so we relied on the "generic" *neural network* for our sample. One of the benefits of using ontologies to encode model card reports is the possibility to further extend the representation with additional ontologies. While the OBO Foundry ontologies we used were sufficient to cover the model card report there could be other ontologies, like the Population Ontology or the Ontology for Biomedical Investigations [27] that can further enrich the model. Lastly, our two samples were limited to bioinformatics-focused model card reports, and therefore, other non-bioinformatics model card reports may have unique coverage requirements beyond the sufficient coverage we presented.

From an application and tooling perspective, the ontology can be harnessed by software application interface since the ontology, in of itself, is reusable software code that serves as a computable knowledge base. The work detailed, specifically the inferencing was localized in the Protégé authoring environment to show the potential feasibility of the ontology. Software APIs for the semantic web applications, like OWL-API [28], Jena [29], HermiT API [30], etc. enables the development of tailored software applications to utilize an ontology for useful real world purposes such as machine reasoning tasks outside of the Protégé environment.

For the sample model cards that we encoded into the ontology, the instance data, the instances' links, and their annotations were a manual process performed through Protégé. While Protégé provided a setting to experiment our approach, an ideal scenario is to have automated tools that will hide the arduous task of encoding the model card

manually where the author of the model card just provides the textual content and content will link automatically using the MCRO. One proposed tool we are currently developing relates to a MCRO-driven software application that can capture the textual inputs of the model card report from an author (see Fig. 9). The author identifies the concept-related section associated with their inputted text and the system will generate an output in the form of an ontology-based artifact (e.g., RDF, JSON-LD [31], nanopublication [32]). This export file can be distributed and shared with all of the necessary alignment with the ontology schema embedded in the export. Multiple model cards of this format that have the shared schema could be queried in aggregate for analysis or service some application tool.

So in addition, another future direction is to develop the natural language processing (NLP) pipeline that will align unstructured text from scientific literature and populate/link to the ontology with assistance of the publication engine. Figure 9 alludes to an natural language processing pipeline to extract unstructured text from published research involving ML models and linking to MCRO scaffold model. In this subsystem depicted in the figure, an automated method to assist in the authoring of ontology-based model card would involve NLP methods—named entity recognition, information and knowledge extraction, controlled natural language - to extract data from unstructured document sources (research publications, existing model card, web pages, etc.). For example, named entity recognition could identify concepts or phrases that match concepts in MCRO, or information extraction methods to parse subject-verb-object data from the text that can map to the ontology's concept level triples. Afterwards, the subsystem populates MCRO with the extracted data as instances that is coordinated by the publishing engine.

Another example of an applied use case is to utilize MCRO as an index component resource to support search services for querying model card documents and possibly aggregating their textual data and metadata for analysis. Anticipating the massive generation of datasets and tools resulting from AI and ML for biomedicine, this could assist potential biomedical and behavioral researchers toward the aforementioned goals.

Comparison to related works

There exist some related ontology-driven work towards adjacent goals but with a focus on niche areas (annotating the development of ML or a focus on neural nets). For example, Naja, et al. developed a duo of ontologies, The System Accountability Ontology (SAO) and Realising Accountable Intelligent Systems (RAInS) ontology, for auditing AI systems to assist in evaluating the risks of ML models [33]. Their ontologies annotate the development stages (i.e. software development lifecycle) for ML-based AI systems. Their work includes a proof of concept software application to show the utilization of their ontology, essentially an authoring tool. The FAIRnets Ontology defines the layers and aspects of neural networks for the purpose of accountability [34]. This ontology is limited to just neural networks with much of the classes (69) devoted to the types of layers. The resources are available for open source and we reviewed them for comparison purposes [35–37].

Our review of the resources highlighted some unique features of our work. One is we designed and encoded a semantic representation of a model card report that is

Amith *et al. BMC Bioinformatics*   (2022) 23:281

Page 10 of 17

computable and standardized through the use of community generated ontology models from biomedical ontologists. While some of these related studies relied on some existing standard resources like PROV-O and Dublin Core, they do not leverage the existing available ontologies that span over the relevant domains, like software (via SWO) or IAO which helped to structure the development our model to represent model cards. In addition, without standard models there is a risk of generating siloed resources that do not have an agreed consensus on the semantics of the concepts. This also enables our MCRO to continually evolve with emerging or other existing ontologies like the Ontology of Biomedical Investigation (OBI) or any other OBO Foundry ontology that are bioinformatics-related. Also the aforementioned resources are essentially used as knowledge graphs and do not leverage any semantic features of OWL2 that could enhance the application of their ontology models. Lastly, because our work utilizes existing standard resources ours is more comprehensive and is not focused on a specific niche.

Conclusion

The utilization of ML has advanced biomedical research in the last decade, contributing to the mass volume of complex data and tools. This has encouraged the National Institutes of Health's interest in making tools and data more FAIR and transparent for biomedical and behavioral researchers. We introduce the Model Card Report Ontology, a representational artifact encoded in OWL2 that can contribute toward FAIR data standards and transparency of machine learning models in biomedicine research. This initial version of the ontology can enable the linking of standardized metadata and encourage the development of tools and resources to search and aggregate machine learning models for health researchers. With two general bioinformatics-related model cards we were able to show how the model cards would translate to an ontology-based artifact and how it can rely on machine-based inferences to link text indexing and search. Future work will involve developing an engine that harnesses the ontology for application use, and continually improving the conceptual scope of the ontology to accommodate unique model card reports.

Methods

The basic steps in constructing the Model Card Report Ontology (MCRO) involved 1) identifying existing ontological resources that can be leveraged by MCRO, 2) developing the core framework of MCRO to synthesize the heterogeneous resources and fill any missing gaps to cover the scope of the model card report, and 3) testing the application of the ontology to link model card information and how integrative software usage. In Fig. 4, we outline the process in the development of the Model Card Report Ontology starting with identifying ontologies from the OBO Foundry that coincide with concepts with the domain, and accommodating conceptual gaps that are not supported by the existing ontology models by drafting a core abstraction with collaborators. Later stages involve encoding the core model and linking the core model with the identified existing ontologies. Our eventual goal is to publish this ontology toward making this a formal standard that aligns with other biomedical and health ontologies for reuse and sharing.
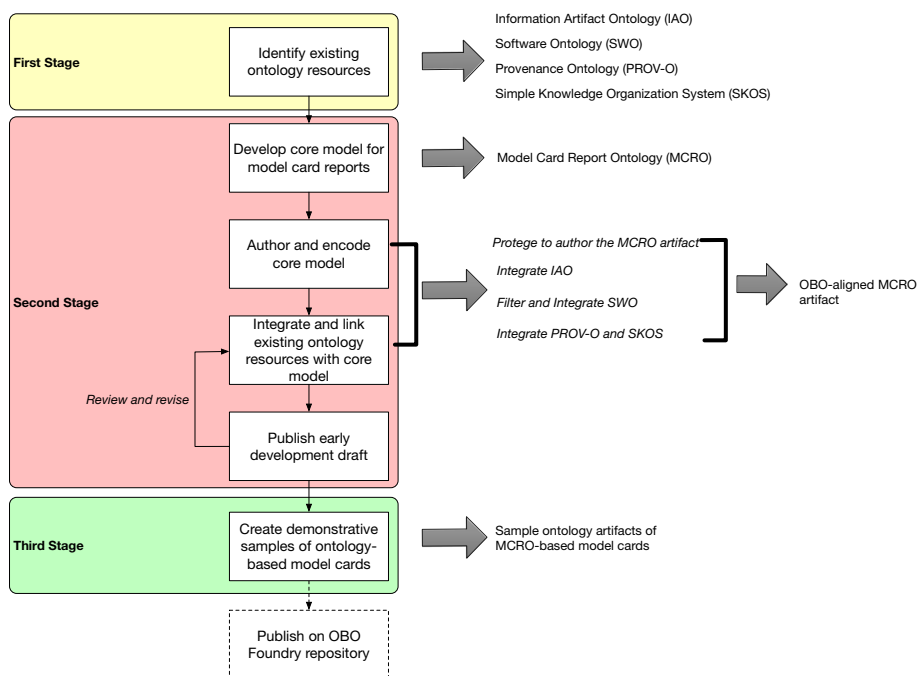
Fig. 4 Outline of the process in developing the Model Card Report Ontology

Integrating OBO foundry ontologies

The stated goal of the OBO Foundry is to "develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate" [9]. Following a prescribed approach to building an ontology [38], we leveraged existing OBO Foundry ontologies to enhance the terminological scope and also to promote reusability of existing ontology-based sources. Extending our core model to these ontologies will help align our work with a common architecture shared with a variety of consistent biomedical and life sciences ontologies.

We relied primarily on the Information Artifact Ontology (IAO), which represents a myriad of high-level information entities. Basically, a model card report is a report [4], which aligns to IAO's *report* concept (which is a type of *information content entity*). Each section of the model card is viewed as part of the document (i.e., IAO's *document part*) that composes the sectional and textual pieces of the model card. These sections are tied together using IAO's *has part* property (*BFO_0000051*), which defines a specific relationship in an ontology. Figure 5 outlines MCO's concept connection with the main classes of the IAO.

An important domain involved in machine learning models are software and periphery concepts involving software—application of the software, metadata specific to software, algorithms, etc. In addition to IAO, we utilized concepts and terminologies from the Software Ontology (SWO), particularlily, concepts such as *data, license, version, algorithm*, etc. These software concepts relate to the various concepts described in our core model to be discussed in the next subsection, and provide exhaustive list of software-related terminologies that can be integrated to our ontology model.

Since SWO relies on IAO, and to avoid redundancy, we extracted a subset SWO that covers target terms we need for model card. We used OBO's ROBOT to extract these essential concepts. ROBOT is a command line software tool that supports a variety of common development tasks when working with ontology files [39]. From SWO (swo.owl), we needed to extract the targeted terms (swo-seed.txt) and export them to a separate filtered version of SWO (swo_import_final.owl). Figure 6 shows the basic steps starting with the target seed terms and the SWO, followed by extraction and merging of the terms to its own ontology file. The resulting subset of SWO was uploaded to GitHub and then imported to furnish the MCRO model.

The Provenance Ontology is another OBO Foundry ontology that contains terminologies to annotate the origination of data and metadata . This ontology assists in providing coverage for authorship and metadata of resources, like a machine learning model resource (citation, dates, etc.). Lastly, we incorporated the Simple Knowledge Organization System (SKOS) ontology [40] for facilities to support annotations of similar concepts, alternative labels, etc. In the next subsection we discuss how these open ontologies are integrated to the core model.

Development of the core of model card report ontology

We examined the inception paper by [4] and the source code files for generating model card documents [41]. From our review of the sources, we drafted an abstraction that approximated the structure of the model card report to generate our ontology. The core concept of this representation is the *Model Card Report* concept which branches onto four main sections - *Model Parameters* (describes the assembly of the model, like algorithm, datasets for training and evaluation, the data format for input and output, etc.),
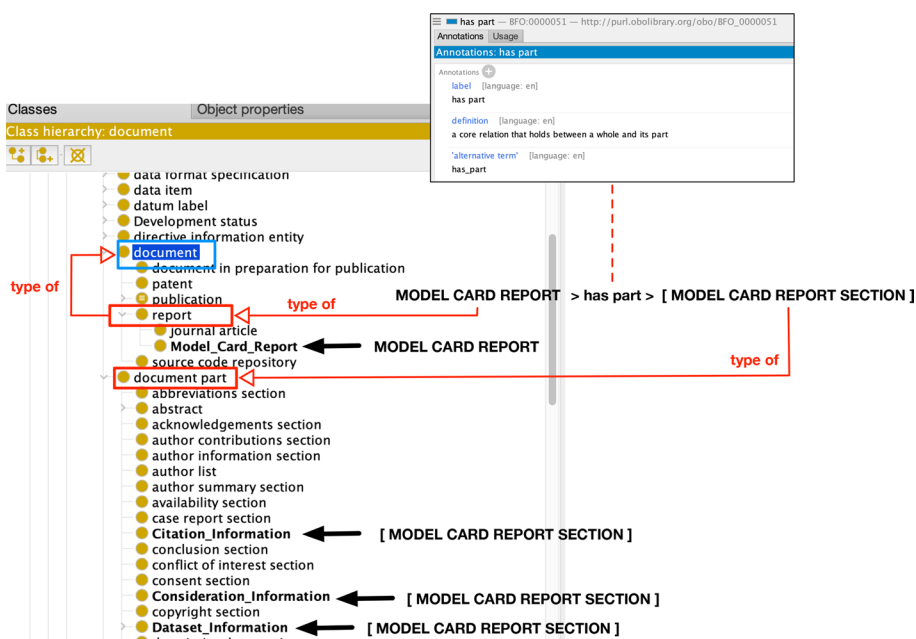

Fig. 5 Annotated Protégé screenshots of how the Information Artifact Ontology concepts of *report*, *document part*, and *has part* are aligned to the Model Card Report Ontology concepts (Model Card Report and its model card sections of Citation, Consideration, and Dataset Information)

Considerations (describes the limits, caveats, and impact on intended users), *Model Details* (general basic information of the model, like licensing, version, developers and owners, name, etc.), and *Quantitative Analysis* (describes the detailed data of model's performance). Essentially, the *Model Card Report* concept represents one model card report and its related textual sections. Figure 3 shows an overview of the representation of the abstraction.

Using the representation, we authored the Model Card Report Ontology using Protégé [42]. The Model Card Report Ontology relied on imports of the aforementioned ontologies discussed in *Integrating OBO Foundry Ontologies.* Alluded earlier, model cards are a type of report and the *Model Card Report* concept was subclassed as an IAO's *report*
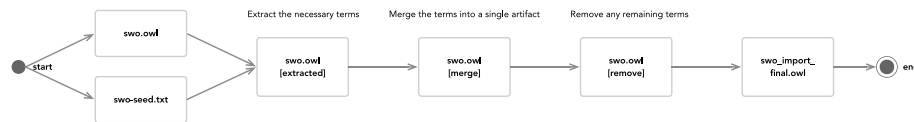
Fig. 6 An outline of the process of to generate a subset of the Software Ontology to import into the Model Card Report Ontology

Fig. 7 An example of some of the text from the sample model cards (*Trade off* information) mapped to the instances and the inferences generated in the Protégé environment (highlighted in yellow)

Fig. 8 An extensive visualization of the mapping of sample text as instance data of the Model Card Report Ontology, including displaying the transitive inference of one of the object properties. The dotted lines in red shows the inference generated from "4" which is an inference
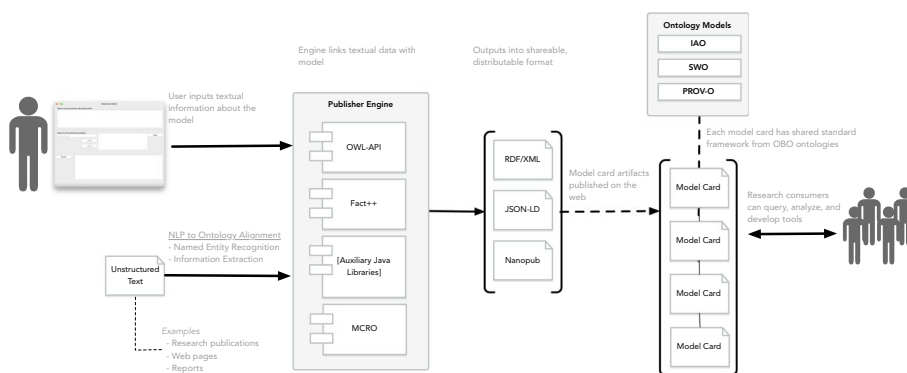
Fig. 9 An applied use-case of the Model Card Report Ontology for software publishing engine to generate ontology-based model card artifacts as RDF or JSON. This applied use case involves a manual authoring method or a proposed automated natural language processing (NLP) method

concept. The concept *Graphic Collection* that groups figures of quantitative data is subclassed as SKOS's *Collection* concept. Most of the links between concepts described in Fig. 3, utilizes BFO and IAO's *has part* of link the various information sections of the model card and *is about* was used to relate the model card card sections to other concepts found in SWO and IAO. Some of the concepts used alternative terms to describe certain sections, e.g., *Caveats* instead of *Trade offs*. This was facilitated with the *alternative label* annotation. Additional links and resources from the OBO Foundry ontologies are described in Table 2.

Demonstration use case for MCRO

To show proof of concept of this work, two bioinformatics researchers who are primary investigators on experimental machine learning endeavors, specifically for ML models on electronic health records (EHR) data to classify and predict adverse effects of stent operations and on social network survey data for HIV transmission among young MSM (Men who have sex with Men). We reviewed the format of the model card report and produced two sample model cards for the two experimental ML-based bioinformatics research projects. With the finalized MCRO, we authored two ontology-based

Table 2 Integration of OBO Concepts with Model Card concepts encoded in the ontology

Model card concept	OBO property link (URI)	OBO concept (URI)
Data set	Is about (★)	Data set (http://purl.obolibrary.org/obo/IAO_0000100)
Model architecture	Is about (★)	Algorithm (http://purl.obolibrary.org/obo/IAO_0000064)
Format information	Is about (★)	Data format specification (http://purl.obolibrary.org/obo/IAO_0000098)
Version information	Is about (★)	Version number (http://purl.obolibrary.org/obo/IAO_0000129)
Reference information	Equivalent to (≡)	Reference section (http://purl.obolibrary.org/obo/IAO_0000320)
Graphic	Equivalent to (≡)	Graph (http://purl.obolibrary.org/obo/IAO_0000038)
Citation information	Has part (∗)	Citation (http://purl.obolibrary.org/obo/IAO_0000301)
License information	Is about (★)	License (http://www.ebi.ac.uk/swo/SWO_0000002)

★ http://purl.obofoundry.org/obo/IAO_0000136, ≡ http://www.w3.org/2002/07/owl#owl:equivalentClass, and ∗ http://purl.obolibrary.org/obo/BFO_0000051

equivalent using MCRO as the scaffold conceptual schema. Both the textual form and the ontology artifact form are available at the GitHub repository [23].

Abbreviations

FAIR   Findable, Accessible, Interoperable, and Reuse
NIH   National Institute of Health
AI   Artificial intelligence
ML   Machine learning
FDA   Food and Drug Administration
OBO   Open Biological and Biomedical Ontology
BFO   Basic Formal Ontology
OWL2   Web Ontology Language
RDF   Resource Description Framework
MCRO   Model Card Report Ontology
IAO   Information Artifact Ontology
PROV-O   Provenance Ontology
SWO   Software Ontology
SKOS   Simple Knowledge Organization System
EHR   Electronic health record
MSM   Men who have sex with men
NLP   Natural language processing
SAO   System Accountability Ontology
RAInS   Realising Accountable Intelligent Systems
OBI   Ontology of Biomedical Investigation
URI   Uniform resource identifier

Author contributions
MA, CT conceptualized the research. MA, CT, LC, XJ, DZ, KR contributed to development of the model. FL, EY collected and annotated data. All authors read and approved the final manuscript.

Availability of data and materials
Model Card Report Ontology: https://github.com/UTHealth-Ontology/MCRO.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1.   ...Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da SilvaSantos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes

Amith *et al. BMC Bioinformatics* (2022) 23:281

Page 16 of 17

E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1): 160018. https://doi.org/10.1038/sdata.2016.18.

2. National Institutes of Health: Data Generation Projects for the NIH Bridge to Artificial Intelligence (Bridge2AI) Program (OT2) (2021). https://www.commonfund.nih.gov/bridge2ai/programresources. Accessed 9 Sept 2021

3. National Institutes of Health: Bridge to Artificial Intelligence (2021). https://www.commonfund.nih.gov/bridge2ai. Accessed 9 Sept 2021

4. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, Spitzer E, Raji ID, Gebru T. Model cards for model reporting. In: Proceedings of the conference on fairness, accountability, and transparency. FAT* '19, pp. 220–229. Association for Computing Machinery, New York, NY, USA;2019. https://doi.org/10.1145/3287560.3287596. event-place: Atlanta, GA, USA. https://doi.org/10.1145/3287560.3287596

5. Berners-Lee T, Hendler J, Lassila O, et al. The semantic web. Sci Am. 2001;284(5):28–37.

6. Insight Centre for Data Analytics: The Linked Open Data Cloud (2021). https://lod-cloud.net/. Accessed 10 Mar 2022

7. Hitzler P. A review of the semantic web field. Commun ACM. 2021;64(2):76–83. https://doi.org/10.1145/3397512.

8. McCrae JP, Abele A, Buitelaar P, Cyganiak R, Jentzsch A, Andryushechkin V, Debattista J, Nasir J. The linked open data cloud (2019). http://lod-cloud.net/. Accessed 9 Sept 2021

9. The Open Biological and Biomedical Ontology (OBO) Foundry: about the OBO Foundry (2021). http://www.obofoundry.org/about-OBO-Foundry.html. Accessed 9 Sept 2021

10. Arp R, Smith B, Spear AD. Building ontologies with basic formal ontology. Cambridge: MIT Press; 2015.

11. Grau BC, Horrocks I, Motik B, Parsia B, Patel-Schneider P, Sattler U. OWL 2: the next step for OWL. J Web Semant. 2008;6(4):309–22.

12. Klyne G, Carroll JJ, McBride B. Resource description framework (RDF) 1.1 concepts and abstract syntax;2014. https://www.w3.org/TR/rdf11-concepts/. Accessed 1 Jan 2021

13. Beckett D, Berners-Lee T, Prud'hommeaux E. Turtle-terse RDF triple language;2014. https://www.w3.org/TR/turtle/. Accessed 9 Sept 2021

14. Harris S, Seaborne A. SPARQL;2008. http://www.w3.org/TR/sparql11-overview/. Accessed 15 Aug 2014

15. O'Connor M, Das A. SQWRL: a query language for OWL. In: Proceedings of the 6th international conference on owl: experiences and directions;2009; vol. 529, pp. 208–215. CEUR-WS.org, Chantilly VA

16. Zheng J. Information artifact ontology (2020). http://www.obofoundry.org/ontology/iao.html. Accessed 10 Mar 2022

17. Ceusters W. An information artifact ontology perspective on data collections and associated representational artifacts. Stud Health Technol Inform. 2012;180:68–72. https://doi.org/10.3233/978-1-61499-101-4-68.

18. Lebo T, Sahoo S, McGuinness D. PROV-O: the PROV Ontology;2013. https://www.w3.org/TR/prov-o/. Accessed 10 Mar 2022

19. Ciccarese P, Soiland-Reyes S, Belhajjame K, Gray AJ, Goble C, Clark T. PAV ontology: provenance, authoring and versioning. J Biomed Semant. 2013;4(1):37. https://doi.org/10.1186/2041-1480-4-37.

20. Lister A. Software ontology;2019. https://obofoundry.org/ontology/swo.html. Accessed 10 Mar 2022

21. Malone J, Brown A, Lister AL, Ison J, Hull D, Parkinson H, Stevens R. The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation. J Biomed Semant. 2014;5(1):25. https://doi.org/10.1186/2041-1480-5-25.

22. Tsarkov D, Horrocks I. FaCT++ description logic reasoner: system description. In: Automated reasoning: third international joint conference, vol 4130. IJCAR 2006, Seattle, WA, USA, August 17–20, 2006, Proceedings. Seattle, WA: Springer; 2006. p. 292–7.

23. Amith M, Tao C. Model card report ontology (2022). http://github.com/UTHealth-Ontology/MCRO. Accessed 10 Mar 2022

24. Dentler K, Cornet R, ten Teije A, de Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semant Web. 2011;2(2):71–87. https://doi.org/10.3233/SW-2011-0034.

25. Mungall C. mereotopologically related to 2022. http://purl.obolibrary.org/obo/RO_0002323. Accessed 10 Mar 2022

26. Obrst L, Ceusters W, Janssen T. Ontologies, semantic technologies, and intelligence: looking toward the future. In: Proceedings of the 2010 conference on ontologies and semantic technologies for intelligence;2010. pp. 213–224. IOS Press, NLD

27. Peters B, Consortium O et al. Ontology for biomedical investigations. Nat Preced. 2009; 1

28. Horridge M, Bechhofer S. The owl api: a Java API for owl ontologies. Semant Web. 2011;2(1):11–21.

29. McBride B. Jena: a semantic Web toolkit. IEEE Internet Comput. 2002;6(6):55–9. https://doi.org/10.1109/MIC.2002.1067737.

30. Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z. HermiT: an OWL 2 reasoner. J Autom Reason. 2014;53(3):245–69.

31. Lanthaler M, Gütl C. On using JSON-LD to create evolvable restful services. In: Proceedings of the third international workshop on RESTful design;2012. pp. 25–32

32. Kuhn T, Chichester C, Krauthammer M, Queralt-Rosinach N, Verborgh R, Giannakopoulos G, Ngonga Ngomo A-C, Viglianti R, Dumontier M. Decentralized provenance-aware publishing with nanopublications. PeerJ Comput Sci. 2016;2:78. https://doi.org/10.7717/peerj-cs.78.

33. Naja I, Markovic M, Edwards P, Cottrill C. A semantic framework to support AI system accountability and audit. In: European semantic web conference;2021. pp. 160–176. Springer

34. Nguyen A, Weller T, Färber M, Sure-Vetter Y. Making neural networks fair. In: Iberoamerican knowledge graphs and semantic web conference;2020. pp. 29–44. Springer

35. Nguyen A, Weller T, Faber M, Sure-Vetter Y. The FAIRnets ontology. Institute of Applied Informatics and Formal Description Methods;2019. https://w3id.org/nno/ontology. Accessed 1 Mar 2022

36. Markovic M, Naja I, Edwards P. The system accountability ontology;2020. https://rains-uoa.github.io/SAO-Ontology/index-en.html. Accessed 1 Mar 2022

37. Naja I, Markovic M, Edwards P. The RAInS ontology;2021. https://rains-uoa.github.io/RAInS-Ontology/v2.0/index-en.html. Accessed 1 Mar 2020
38. Noy NF, McGuinness DL. Ontology development 101: a guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA;2001
39. Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: a tool for automating ontology workflows. BMC Bioinform. 2019;20(1):1–10.
40. Miles A, Bechhofer S. SKOS simple knowledge organization system reference;2009. http://www.w3.org/TR/skos-reference/. Accessed 10 Mar 2022
41. Google, Inc.: Model Card Toolkit (2021). https://github.com/tensorflow/model-card-toolkit
42. Musen MA. The ProtéGé project: a look back and a look forward. AI Matters. 2015;1(4):4–12. https://doi.org/10.1145/2757001.2757003.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.