A Framework to Specify Agent-Based Models Using ODD* Protocol

Cédric Grueau
Department of Systems and Informatics
EST Setúbal, Polytechnic Institute of Setúbal
Setúbal, Portugal
0000-0003-3445-4070

Thyago Romagna Bendo
Department of Systems and Informatics
EST Setúbal, Polytechnic Institute of Setúbal
Setúbal, Portugal
0000-0003-2164-4270

Alan Gavioli
Paraná Federal University Of Technology
Medianeira, Brazil
0000-0002-5325-0754

João Junior Araújo
Department of Computer Sciences
NOVA School of Science and Technology
Caparica, Portugal
0000-0001-5914-1631

***Abstract*—The use of Agent-Based Models (ABM) is a popular approach to develop simulations in fields such as social sciences, geography and natural sciences. However, due to the complexity of the models, the sharing of models' specification becomes an arduous process, which makes it difficult to validate and replicate this type of model. To assist in this process, experts from the domain have developed the ODD (Overview, Design concepts, Details) protocol to help systematize ABM descriptions. However, because it is a protocol, the latitude allowed in the description of the models and the format in which these descriptions are made only address part of the problem. On the one hand, it may remain difficult to understand the description, and on the other hand, the access and processing of the ODD descriptions are still a manual task. To solve the problem, a framework, based on an ontology, was developed to allow traceability of ABM described with ODD. Concepts from other metadata initiatives, such as Dublin Core and Codemeta, were included to add important data to the ODD protocol. An application was developed to facilitate the access to the models, which allows the analysis of the data source of the model and the extraction of the protocol components. An API was also developed that gives access to ABM descriptions and promotes the processing of ABM descriptions. The application is ready for a series of tests to validate our approach, the usability of the application and the utility of the framework.**

***Index Terms*—ODD Protocol; ODD+D; ODD+2D; Linked Data; Agent-Based modelling.**

I. Introduction

For the last two decades, individuals are in the center of research social and geographical systems research [1]. Networks and interactions of individual entities, called agents, contribute for the emergence of the system's behavior. Agent-based models (ABM) have become an efficient technique to simulate complex systems. The popularity of the field began to increase, leading to an increasing number of applications of ABM for Geographical Systems (see [2] or [3] for some examples).

After two decades of development and the creation of an enlarged user community, many challenges still need to be addressed in developing and advancing ABM applicable to geographical systems. Some of these challenges were first presented a decade ago in a publication by [4] and were recently revisited [5]. One of these challenges is Replication and Experiment. Replication of agent-based models is difficult to achieve due to multiple parameters, methods and contexts that pertain to a certain situation [5].

Another challenge is related with the vocabulary used by modellers and programmers to describe the concepts used in their model. The lack of standardization may lead to misunderstanding of the exchange of software metadata. [6] suggested a minimum standard of model description for good modelling practice, namely the provision of source code and an accessible natural language description, and argue for the development of a common standard.

However, as noted by [7], code is often only understandable by other modellers and even this can be problematic if it is badly written and poorly documented. It is only through such activities that we can replicate and experiment with agent-based models. Similar efforts with respect to reproducibility of results are also being called on in the geocomputation.

The ODD protocol [8] (Overview, Design concepts, and Details) was created to define a series of information that should accompany the description of each ABM. However, the first version of the protocol is limited by its ability to describe only one version of the model and does not have a direct relationship with the code for the model. These aspects were improved with the new versions of the protocol, ODD+D (ODD + Decision), with the latest version ODD + 2D (ODD + Decision + Data) having a considerably greater descriptive capacity [9]. In this paper, we will refer to OOD protocol and its different extensions by ODD*. Using ODD*, modellers should fulfill a series of fields to describe models' structure and data, simulations components and agents' rationale. Some questions to answer are provided by the protocol's authors in order to help modellers inform ABM descriptions. Modellers must organize this information into a document that will be used by other experts of the field to understand and replicate the model.

 ISBN: 978-1-61208-898-3

Despite the improvement brought by the odd* protocol, it still remains challenging for modelers to compare two ABM or to process quickly and safely an ABM description in order to reuse it. Ones will still need to manually collect descriptions, interpret them, compare them and adapt the code.

In section 2, we will see how researcher have been working on ODD* to improve models specification. Section 3 will describe our proposal and section 4 will present first results.

II. Related Work

Different approaches have been followed to reduce the gap between modellers and the community of users interested in replicating, understanding and validating ABM. Early adopters of the ODD protocol have advocated that the more information about an ABM is given to the community, the more transparent the model is. This approach has evolved towards the development of repositories such as CoMSES [10], which is a network with the common goal of improving the way experts develop, share, and use computational modeling in the social and ecological sciences. However, suggestions of researchers [6] have headed towards a minimum standard of model description, namely the provision of source code and an accessible description. Yet, ABM specifications can be found in repositories of models and scientific articles, these specification only remain accessible by manual searches.

In [7], authors proposed an extension to the ODD protocol to build, document, and compare AMB for Coupled human and natural systems (CHANS). They included two new sections a "Model Verification and Validation" section and a "CHANS Characteristic Features" section. Authors also propose a technical development that facilitates CHANS- related agent-based modeling, including the online pseudo-code and preliminary library of reusable modules in Netlogo [11], a multi-agent programmable modeling environment. Authors presented a cross-site comparison in relation to CHANS models include reinforcing the need for standardized modules for CHANS ABMs. Yet, authors do not provide any suggestion on how to store and provide automatic comparisons of cross-sites ABM. Models stored in separate repositories are still hard to search for and compare.

The latest addition to ODD* is ODD+2D with a greater description of data used for the simulations. This new version implied a reorganization of the protocol's fields related to data, which highlighted a crucial aspect of ODD* protocol: the information to provide for the model description vary according to the version of ODD* the author is using to document his model. This point usually implies field removal or adaptation, damaging the development of a pattern on MBA documentation.

Despite efforts promoted to normalize ABMs' description Modelers still need better support for ABMs' specification, comparison and discovery. The development of a common standard remains thus limited to natural language. On the other hand, the analysis and comparison between models stored in such repositories remains a task that can only be performed by users. Different software repositories, software languages and scientific domains denote this information in different ways, which makes it difficult or impossible for tools to work across these different sources without losing valuable information along the way.

III. Proposal

As stated before, the ODD* protocol is the most consistent proposition to describe ABM. It organizes information of an specific instance of an ABM in a document, according to a specific ODD version. As a document, the ABM description using ODD[a] should also be described with other metadata. Data about authors, date and keywords are essential for the management of models' specifications. On the other hand, computer code has been identified as a key element for the ABM replication and validation, we thus advocate that metadata about code should also be included in ODD* descriptions. We propose in these work to include these two components, document and code metadata, into models specifications described with ODD*. For the last decade different initiatives have led to the development of metadata sets that can describe both document and code. Dublin core [12] for documents and the CodeMeta Project [13], for computer code, are two initiative that are widely used to describe such data. For this work, we will take advantage of both. A subset of Dublin Core will be used to add data to identify and link ODD* specifications to authors and scientific papers. On the other hand, codeMeta metadata will allow to describe code versions and dependencies to facilitate software replication and code analysis.

To address some of the limitations and to enable better use and greater adoption of the ODD* protocol, we developed a framework to enhance the capacity of describing ABM provenance and provide a more flexible analysis and comparison of models. To prevent patterns on MBA documentation we propose a model for the ODD* based on an ontology. This will enable the analysis and reasoning on models components in order to compare models instances. The framework allows ABM users to better manage the version of the ODD* protocol associated with each model. It also allow to describe models authoring and computer code related to the implementation of simulations. Through the platform, MBA descriptions are stored as linked data that can easily be accessed and processed. The extended description is supported by an ontology called ODD* that can be extended according to new versions of the protocol that may be proposed. Figure 1 presents the architecture of the framework that will be detailed in Section III-B.

A. Domain Model

To develop an ABM documentation model, it is important to provide support for all concepts existing in ODD*. At the present time there are four versions of protocols considered between ODD and ODD+2D. The differences between each versions consist in changes of fields' name, relocation of fields, and expansion of fields.

 ISBN: 978-1-61208-898-3

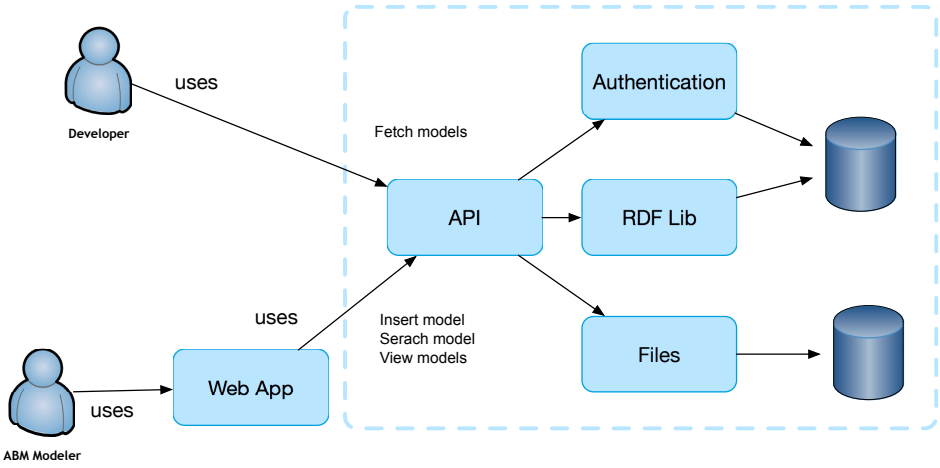

Fig. 1. Architecture of the framework

To provide a support for all the versions of ODD*, it was necessary to map all the concepts and constructs between versions. We used the descriptions of each versions presented in [8], [9], [14], [15]. The mapping process resulted in a feature model partially presented in Figure 2.

The feature model only describes concepts present in the ODD* protocol. Considering that the description of an ABM using the ODD protocol is a document, it is also relevant to provide metadata to describe this resource and provide provenance data. We chose to include a subset of the Dublin Metadata Element Set to achieve that. We added the following attributes from the Dublin Core schema for our domain model: Title; Creator; Contributor; Date; Subject; Language; Source; Rights; Publisher; Bibliographic Resource; and Identifier.

Another important concept included in the ODD* ontology is the the information regarding computer code. We incorporated the following fields from the CodeMeta schema to our model: URL; Maintainer; Version; Reference Publication; Programming Language; File Format; File Size; License; and Read Me.

The ontology which aggregates ODD* concepts, including selected fields from Dublin Core and CodeMeta, is presented in Figure 3.

It defines a structure of components that can be used and combined to build a description of an ABM, according to a certain version of the ODD protocol.

B. ODD* as linked data

Linked Data is structured data linked to other data in order to create a larger set of data that can be used through semantic queries. It uses Web standards such as URIs, HTTP and RDF format to identify a resource object in the internet. The possibility to structure ODD* components as linked data offers the opportunity to define resources that can be accessed, analyzed and manipulated by both human and machines. For these reasons, we decided to represent and store ABMs as Linked Data.

C. Framework

The framework is composed by a Web application that can be used by users to insert, search and visualize ABMs. A REST API manages the access to models repository that are store in a SQL Database after being transformed to RDF Format using the RDFLib Library. In the actual version of the application, some of the models' data are stored as files. This is the case for tabular data and images. Other application can also access models stored in the framework taking advantage of the API included in the framework.

The Web application provides access for users without any special requirement on the software use. This Web client encapsulate a didactic approach to explain to users how the model are organized and combined, according to ODD versions. It uses colors to explain the transition between protocol versions and facilitate models creation and visualization. The emphasis have been directed towards performance reading ease, allowing different screen sizes and navigation using keyboard keys.

To provide machine access, developers access the API, allowing access to models components according to the ontological model.

IV. FIRST RESULTS

After the implementation of all components, we chose to test the usability of the Web application before any other aspect of the platform. We want to ensure that users are able to create and search for models using the Web application. The objective of the test are to verify how easy it is to create a new model, without errors, and store it using the Web application. We also want to understand if users could search, read and analyze a model stored in the platform. To evaluate the usability, a group of students from Computer Science and Software Engineer courses from bachelor and master degrees will be invited to use the application and execute two different tasks. A first group will be assigned the task to insert an ABM description and the second group will be assigned the task to retrieve information from a model already existing in the database. For the test,

 ISBN: 978-1-61208-898-3

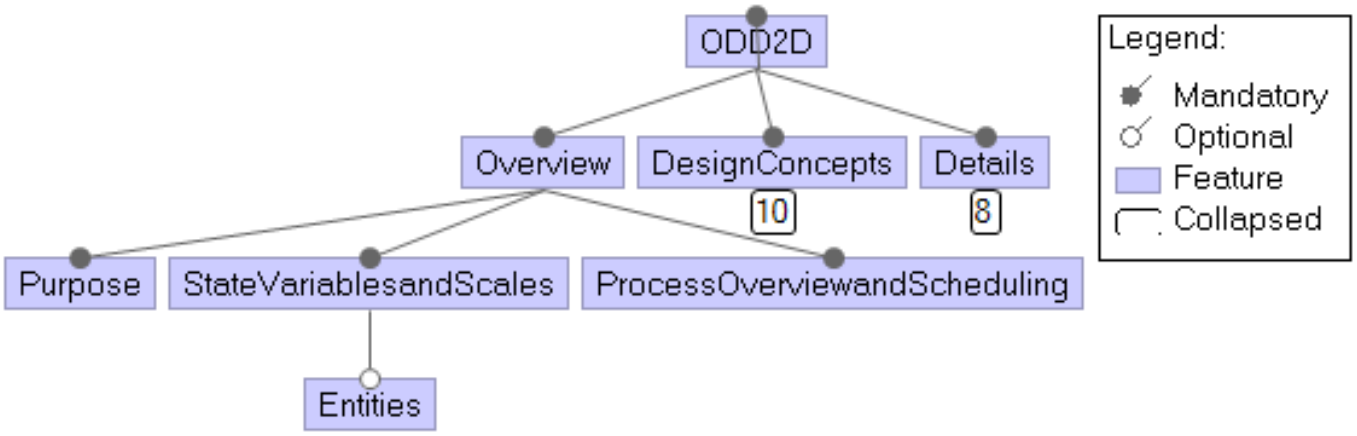

Fig. 2. Partial view of the feature model for ODD* protocol.

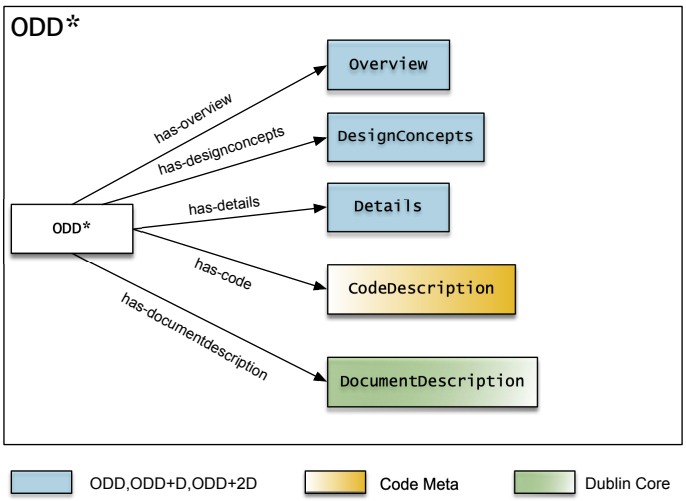

Fig. 3. Main concepts of ODD* Ontology.

we will use the ODD description of an ABM about Urgent Diffusion in Social Media, published by [16]. For both tests, the objectives of the experience will be clearly explained and some basic concepts about ABM will be introduced before start.

V. Conclusions and Future Work

In this paper, we presented a project that aims at improving the documentation of ABM using the ODD* protocol. We introduced a framework, based on an ontology, to create and store ABM specifications available as Linked Data. This framework will contribute to enhance models provenance and manipulation to perform more complex tasks on those specifications. In addition, we presented a Web application that enables the creation, storage and retrieval of ABMs. The platform will be tasted to verify if the Web application can be used efficiently to analyze and understand ABMs based on their ODD descriptions. Future work will focus on testing the Web Application with experts from the domain in order to evaluate the utility of our approach. Tests will also be performed on the API in order to evaluate its usability by programmers.

References

[1] M. Batty, *The New Science of Cities*. The MIT Press, 2013.
[2] A. J. Heppenstall, A. T. Crooks, L. M. See, and M. Batty, *Agent-based models of geographical systems*. 2012.
[3] J. Rocha, *Multi-agent Systems*. InTech, sep 2017.
[4] A. Crooks, C. Castle, and M. Batty, "Key challenges in agent-based modelling for geo-spatial simulation," *Computers, Environment and Urban Systems*, vol. 32, pp. 417–430, nov 2008.
[5] A. J. J. Heppenstall, A. Crooks, N. Malleson, E. Manley, J. Ge, and M. Batty, "Agent-based Models for Geographical Systems: A Review." 2019.
[6] B. Müller *et al.*, "Standardised and transparent model descriptions for agent-based models: Current status and prospects," *Environmental Modelling & Software*, vol. 55, pp. 156–163, may 2014.
[7] L. An, A. Zvoleff, J. Liu, and W. Axinn, "Agent-Based Modeling in Coupled Human and Natural Systems (CHANS): Lessons from a Comparative Analysis," *Annals of the Association of American Geographers*, vol. 104, pp. 723–745, jul 2014.
[8] V. Grimm *et al.*, "A standard protocol for describing individual-based and agent-based models," *Ecological Modelling*, vol. 198, pp. 115–126, sep 2006.
[9] A. Laatabi, N. Marilleau, T. Nguyen-Huu, H. Hbid, and M. A. Babram, "ODD+2D: An ODD based protocol for mapping data to empirical ABMs," *JASSS*, vol. 21, no. 2, 2018.
[10] "Comses network - abm." https://www.comses.net/, 2021. Accessed: 2021-01-29.
[11] "Netlogo." https://ccl.northwestern.edu/netlogo/, 2021. Accessed: 2021-01-29.
[12] D. Hillmann, "Using Dublin Core," 2008.
[13] "codemeta." https://codemeta.github.io/, 2021. Accessed: 2021-01-29.
[14] V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, "The ODD protocol: A review and first update," *Ecological Modelling*, vol. 221, pp. 2760–2768, nov 2010.
[15] B. Müller *et al.*, "Describing human decisions in agent-based models – odd + d, an extension of the odd protocol," *Environmental Modelling & Software*, vol. 48, pp. 37 – 48, 2013.
[16] W. Rand, J. Herrmann, B. Schein, and N. Vodopivec, "An agent-based model of urgent diffusion in social media," *Journal of Artificial Societies and Social Simulation*, vol. 18, no. 2, p. 1, 2015.

 ISBN: 978-1-61208-898-3