

Ontology Management in an Industrial Environment: The BASF Governance Operational Model for Ontologies (GOMO)

Ana Iglesias-Molina¹, José Antonio Bernabé-Díaz², Prashant Deshmukh², Paola Espinoza-Arias², Aarón Ayllón-Benítez², Alba Fernández-Izquierdo², José María Ponce-Bernabé², Sara Pérez², Edna Ruckhaus¹, Oscar Corcho¹ and José Luis Sánchez-Fernández²

¹Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain

² BASF Group, Ludwigshafen am Rhein/ Madrid, Germany/ Spain

ABSTRACT

Governance on ontology development and maintenance practices within an organization has many advantages over ungoverned, siloed approaches that many organizations exhibit nowadays. This paper presents the BASF Governance Operational Model for Ontologies (GOMO), which addresses all stages of the ontology lifecycle and provides a framework for the development and maintenance of ontologies within BASF. GOMO is comprised of Principles, Standards and Quality Assurance criteria, Best Practices, Training and Outreach and is the result of a collaborative effort between industry and academia in the semantic web field. GOMO Principles, Standards and Best Practices are being applied to all running ontology-based projects in BASF. Through outreach presentations with sections of the BASF community, GOMO has reached a wider audience to foster understanding on the utility and implementation of ontologies. Finally, GOMO stands as a framework that is fit for adoption by other organizations facing similar challenges in ontology governance.

1 INTRODUCTION

Ontologies have been widely used for different purposes and applied to a wide variety of domains. They have been developed and implemented in both academic and industrial contexts (Poveda-Villalón, Fernández-Izquierdo, Fernández-López, & García-Castro, 2022). Some of the reasons for their adoption are (i) their potential for data semantification, (ii) their provision of standardized meta-, reference- and master data, (iii) their ability to make curated, granular domain knowledge reusable and standardized, (iv) their ability to infer new knowledge, and (v) their use in data integration: enabling interoperability through alignment of schema and controlled vocabularies across multiple data/information source systems.

In large organizations such as BASF, which are adopting ontologies as a way to describe data across the organization, it is not uncommon to see the application of ungoverned ontology development methods, relying on ad-hoc practices, hence producing ontologies that are not interoperable, discoverable or understandable. Indeed, this generates obstacles to data integration and results in duplication of efforts and poor coordination across stakeholders such as data and knowledge stewards, data scientists, application owners and users.

Given this starting point, which is common in many organizations, the need for a holistic approach for ontology governance has led BASF to create GOMO, a Governance Operational Model for BASF ontologies that enables (i) interoperability, (ii) development and use of common tools, (iii) the application of best practices and (iv) addressing the skills gap. The model, its components and how it is being used at BASF is described in the remaining of this paper.

2 RELATED WORK

In recent years the increase in the volume of data handled by organizations has had an impact on how data is governed, since it has become a critical asset in industrial, scientific and public administration settings. For that reason, data governance has gained importance and several frameworks have been developed and adopted by organizations (Alhassan, Sammon, & Daly, 2016). One of the most relevant frameworks is DAMA (Brackett & Earley, 2009). This framework defines Data Governance as “the exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets”.

In relation to ontologies, several communities have developed general guidelines that are applied in the creation, development and maintenance of ontologies. This is the case of the OBO Foundry (Smith, et al., 2007) for biological sciences, or the Industrial Ontologies Foundry (IOF) (Karray, et al., 2021) for industrial ontologies, among others. These communities aim at creating an interoperable

portfolio of ontologies that have the same fundamental characteristics (e.g., IRI structure, IDs, formats, availability). The International Organization for Standardization (ISO) has provided standards for the development of top-level ontologies (International Organization for Standardization, 2021). In addition, the FAIR Data Principles (Wilkinson, et al., 2016) provide measurable principles to enhance the reusability of published data; they are being extensively adopted by the community and can closely apply to ontologies (Poveda-Villalón, Espinoza-Arias, Garijo, & Corcho, 2020). For instance, the Consultative Group on International Agricultural Research (CGIAR) is bringing together stakeholders to ensure the generation of FAIR Agrifood data and knowledge using ontologies (Arnaud, et al., 2020).

GOMO integrates concepts from data governance, FAIR principles, and the OBO and IOF Foundries. However, to the best of our knowledge, there is no *holistic* ontology governance approach that enables the management of individual and collective ontology-related goals, encompasses all stages of the ontology lifecycle (see section 3.2), and connects business requirements, common ontology development methodologies, technical standards, tool and infrastructure requirements, and people and skills.

3 GOMO - ONTOLOGY GOVERNANCE MODEL

The BASF Governance Operational Model for Ontologies (GOMO) provides a framework for the development and maintenance of ontologies within an organization. It adds value by enabling the build-up of the organization’s capacity to govern and manage ontologies. In this way the organization can coordinate and control the development, management, and curation of ontologies in their various scenarios of use. This model provides guidelines on how to create, develop and maintain ontologies, delivered as a set of interlinked documents and training sessions that tackle the tasks of the ontology lifecycle from different perspectives to ensure the governance of ontologies.

3.1 Components of GOMO

GOMO is comprised of four interlinked components that support ontology development throughout the stages of the ontology lifecycle. These components are listed as follows:

- (1) **Governance Principles** define a set of high-level fundamental rules on how the organization should develop, publish, maintain, and consume ontologies. They refer to the availability, accessibility, persistence, documentation, FAIRness, and the community of people involved with the ontology.
- (2) **Standards** are a set of agreed specifications to be followed during the ontology lifecycle. Each Standard is associated with a **Quality Assurance** (QA) method to allow human or software-based evaluation of its correct implementation. Standards enable the development of tools and ensure ontology reusability and interoperability. They refer to mandatory metadata for ontologies and its entities, naming conventions, ontology languages, versioning, diagram notation, and IRI structure.
- (3) **Best Practices** comprise a set of recommendations and guidelines that explain and illustrate how to follow the principles and implement the standards, while providing the user with the background knowledge required to perform specific activities. These guidelines apply to the ontology lifecycle and are developed to address each activity performed in the process, e.g., curation, deployment.
- (4) **Trainings and Outreach** comprise a series of interactive, expert-led workshops to explain the concepts and guidelines from Best Practices

and Standards with tailored examples for the target audience. These Trainings include sessions corresponding to the Best Practices, a 101 series to present ontologies, its infrastructure and consumption possibilities, sessions on how to organize hack-type events oriented to develop ontologies (such as VoCamps¹), and tutorials on how to use tools, such as WebProtégé. Outreach refers to applying GOMO within the organization through Trainings or ongoing project work.

Fig.1 depicts the relations among these components. As an example, Principles establish that all ontologies must have global, unique, persistent, and resolvable IRIs; Standards define the IRI structure; QAs validate that IRIs are correct; and Best Practices and Trainings show how to create ontology IRIs.

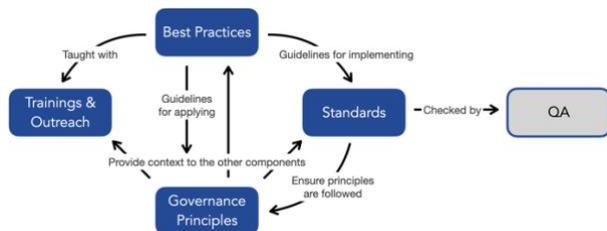


Fig. 1. Interaction among GOMO components.

3.2 GOMO and the ontology lifecycle

GOMO is integrated in an ontology development lifecycle. This lifecycle is based on the workflow proposed by the LOT Methodology (Poveda-Villalón, Fernández-Izquierdo, Fernández-López, & García-Castro, 2022). It is illustrated in Fig. 2. Each stage is composed of several tasks and activities, which are governed with GOMO through its components. The different components may apply to the entire lifecycle, or to specific steps. For instance, the Principle on Documentation only applies to the Publication stage whereas the Principle on FAIRness applies to all stages. Additionally, GOMO defines the roles that apply to the ontology's community and proposes a RACI matrix to define the degree of responsibility of each role in the development tasks that are part of the lifecycle.

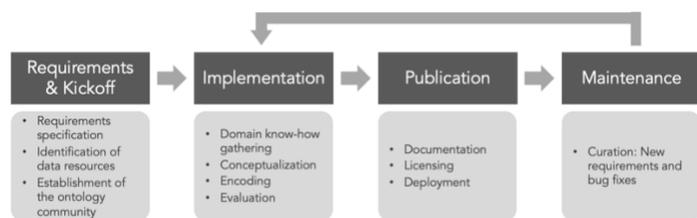


Fig. 2. Ontology lifecycle developed in GOMO. Adapted from LOT (Poveda-Villalón, Fernández-Izquierdo, Fernández-López, & García-Castro, 2022).

3.3 Methodology

GOMO is the result of an industry-academy collaboration between BASF, Universidad Politécnica de Madrid (UPM), and Universidad de Murcia (UM). The team is composed of ontology engineers, data engineers, and knowledge architects who use ontologies in knowledge graphs as well as develop both of them. An agile iterative methodology has been followed, with constant, direct communication between stakeholders and the team to ensure the model suits the company's needs.

The model was developed following these steps: (1) agreement and documentation of ontology lifecycle and Govern the requirements and correct media for trainings, and development of training materials. The generation of the model was iterative and, once released, will continue to be updated regularly. All the materials are being made available on an internal wiki. Current work involves the application of the governance framework to all ontology use-case related projects, either run by the central team of knowledge architects or by decentralized IT departments.

4 CONCLUSIONS AND FUTURE WORK

The GOMO framework provides a solid foundation for enabling organizations to develop, consume and maintain ontologies. GOMO has been developed from both

a domain- and technology-agnostic perspective. Thus, we believe this framework is not only applicable for BASF but may be adopted by other organizations seeking a governance framework to support ontology development and maintenance.

The undertaking of GOMO was met by organizational challenges and lessons learned. For example, considering the needs of all stakeholders was complex, particularly when committing to agreements on ontology standards. Usual academic considerations in approaching ontology development had to make way for more pragmatic approaches in an industrial context – for example regarding confidentiality and differential treatment of particular ontologies. Finally, the adoption of GOMO principles and standards in ongoing projects led to the identification of blind spots and continue to help improve the GOMO framework.

GOMO is a living project, and as such, activities are underway to refine the content and enable adoption, not only within BASF but also externally, within organizations developing ontologies for BASF. Within BASF, tools are currently under development – based on GOMO Standards – to support the ontology lifecycle at BASF, e.g., an automated quality assurance pipeline. To foster appreciation and adoption of ontologies internally, a series of outreach sessions is scheduled, where the content of GOMO will be presented to the community. It is also planned to form an internal technical consortium within BASF to ensure continuous alignment, foster further refinement of GOMO and identify and bring new stakeholders into dialog. Finally, new updates of the content will be released externally² for the benefit of the wider community.

ACKNOWLEDGEMENTS

We would like to thank Dr. Alexander García Castro for the GOMO conceptualization and Prof. Jesualdo Tomás Fernández-Breis (UM) for his input. We thank Dr. Jürgen Müller as the BASF sponsor of this project.

REFERENCES

- Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems*, 25, 64–75.
- Arnaud, E., Laporte, M.-A., Kim, S., Aubert, C., Leonelli, S., Miro, B., . . . King, B. (2020). The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. *Patterns*, 1, 100105. doi:https://doi.org/10.1016/j.patter.2020.100105
- Brackett, M., & Earley, P. S. (2009). The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide). International Organization for Standardization. (2021). *Information technology — Top-level ontologies (TLO)*. Retrieved from (ISO/IEC 21838.1): https://www.iso.org/standard/71954.html
- Karray, M., Otte, N., Rai, R., Ameri, F., Kulvatunyou, B., Smith, B., . . . Arista, R. (2021). The Industrial Ontologies Foundry (IOF) perspectives. Industrial Ontology Foundry (IOF) - achieving data interoperability Workshop, International Conference on Interoperability for Enterprise Systems and Applications, Tarbes, . Retrieved from https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=925879
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53, 148–152.
- Poveda-Villalón, M., Espinoza-Arias, P., Garijo, D., & Corcho, O. (2020). Coming to terms with FAIR ontologies. *International Conference on Knowledge Engineering and Knowledge Management*, (pp. 255–270).
- Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An industrial oriented ontology engineering framework. *Engineering Applications of Artificial Intelligence*, 111, 104755. doi:https://doi.org/10.1016/j.engappai.2022.104755
- Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., . . . others. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25, 1251–1255.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., . . . others. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 1–9.

¹ http://vocamp.org/wiki/Main_Page

² <https://github.com/protegeproject/basf-gomo>