

El papel de la gestión de datos en la tecnología de gemelos digitales

Ignacio Marco-Pérez 

igmarco@unirioja.es

Dpto. de Matemáticas y Computación, Universidad de La Rioja

Palabras clave: Data Management, Digital Twin, Literature Review, Taxonomy

Iniciativas de digitalización en el Sector Agroalimentario

En los últimos años hemos visto cómo el sector agroalimentario en la Unión Europea se encuentra en una situación especialmente compleja. Por un lado, existen pérdidas e ineficiencias que afectan a la productividad global del sistema, mientras que la sociedad demanda de un modo creciente medios para la protección del medio ambiente y el suministro de alimentos seguros, nutritivos y saludables, exigiendo productos de alto valor añadido, de producción sostenible y verificables [1].

En este contexto, tecnologías propias de la Industria 4.0 relacionadas con la sensorización y digitalización de procesos, como el internet de las cosas (IoT por sus siglas en inglés) [2], el aprendizaje automático [3] y herramientas de almacenamiento y verificación como el blockchain [4], juegan un importante papel en el desarrollo del sector [5].

Gemelos digitales y gestión de datos

El concepto del gemelo digital fue presentado por primera vez por Grieves en 2003 como un “constructo de información digital de un sistema físico, que incluye (en su forma óptima) toda la información relevante requerida para

El trabajo se enmarca dentro de la Tesis Doctoral "Gobernanza de datos en el desarrollo de Modelos computacionales de simulación de Procesos Alimentarios", desarrollada bajo un contrato con la entidad *Centro Nacional de Tecnología y Seguridad Alimentaria - Laboratorio del Ebro* (CNTA), beneficiario de la convocatoria de *Ayudas predoctorales para la investigación en Centros Tecnológicos, de Investigación y Empresas de la Comunidad Autónoma de La Rioja* de acuerdo con la *Resolución 1361/2023*.

completar una tarea determinada, y que está vinculada con dicho sistema físico” [6]. Posteriormente, Glaessgen y Stargel definieron formalmente este concepto aplicado a la industria aeroespacial [7]. Generalizando dicha definición, podemos considerar un gemelo digital como una simulación integrada, multifísica, multiescala y probabilística de un sistema que utiliza los mejores modelos físicos disponibles, sensores, datos históricos, etc., para reflejar la vida de su gemelo físico.

El propio Grieves presentó un esquema general basado en tres componentes generales comunes a cualquier gemelo digital: un elemento o sistema físico, un modelo virtual y una conexión de datos entre ambas partes [6]. Según este modelo, el flujo de datos del componente físico al modelo virtual requiere de un procesamiento para convertirlo en información valiosa. Por otro lado, el flujo de datos del componente virtual al físico se corresponde con información procesada que puede utilizarse para controlar y gestionar el sistema físico [8].

Una siguiente aproximación por Enders y Hofbach agrega a estas tres partes dos más: el catálogo de servicios ofrecidos al usuario y los propios datos interconectados entre los otros componentes [9]. Las distintas aproximaciones a la definición de gemelo digital se acercan progresivamente en el tiempo a las ideas de integración, almacenamiento, calidad, seguridad, etc. de los datos involucrados, es decir, a la gestión de datos en esta tecnología.

En este contexto, la tesis doctoral "Gobernanza de datos en el desarrollo de Modelos computacionales de simulación de Procesos Alimentarios" tiene como objetivo el estudio de los distintos aspectos de gestión de datos, en particular la gobernanza de datos, en el contexto del diseño y desarrollo de gemelos digitales. De hecho, entre sus objetivos principales se encuentra la aplicación de esta tecnología a una planta de extrusión de proteínas [10].

Una primera revisión bibliográfica

Como primer apartado en el desarrollo de la tesis, se ha considerado llevar a cabo una revisión sistemática de la literatura (Systematic Literature Review en inglés) siguiendo las directrices propuestas por la guía de Kitchenham y Charters [11]. El objetivo de esta revisión sistemática es plantear las bases de la aplicación de la tecnología de gemelos digitales con independencia del dominio de aplicación bajo la perspectiva de la gestión de datos, con la intención de elaborar posteriormente un framework aplicable a sus componentes más relevantes. A pesar de que existen trabajos que ya abordan esta cuestión (e.g. [12, 13]), nuestra perspectiva trata de hacer énfasis en las distintas partes que componen la gestión de datos definidas por el estándar propuesto por DAMA en la guía DMBOK [14].

En esta línea, un primer esfuerzo ha sido llevar a cabo un reconocimiento

bibliográfico basado en el concepto de estudio de mapeo sistemático (Systematic Mapping Study en inglés) [11] de las diferentes clasificaciones y taxonomías (y por tanto dimensiones y características) de gemelos digitales con las siguientes preguntas de investigación:

1. RQ1: ¿Existe una taxonomía unificada en materia de gemelos digitales?
2. RQ2: ¿Se consideran los aspectos de la gestión de datos en las taxonomías de gemelos digitales?
3. RQ3: ¿Se considera la gestión de datos como una dimensión concreta en las taxonomías de gemelos digitales?
4. RQ4: ¿Cuáles de los aspectos de la gestión de datos propuestos por el DMBOK [14] son los considerados en mayor medida en las taxonomías de gemelos digitales?

Para llevar a cabo este estudio se consideraron las bases de datos científicas Scopus, ScienceDirect, SpringerLink, IEEE Xplore y ACM Digital Library con la siguiente expresión de búsqueda:

**((TITLE-ABS (taxonom* OR classific*) AND
TITLE (“Digital Twin”* OR “Digital Shadow”
OR “Cyber-Physical”)) AND PUBYEAR >2019**

Siguiendo las líneas propuesta por Kitchenham y Charters [11], los artículos seleccionados para el estudio fueron 861; 692 tras eliminar duplicados; 134, 110 y 59 tras aplicar tres criterios de exclusión respectivamente. El estudio de los artículos relevantes demostró lo siguiente:

1. RQ1: No existe todavía una taxonomía o clasificación de gemelos digitales unificada, aunque su definición se considera uno de los principales retos actuales en este contexto.
2. RQ2: Distintos aspectos de la gestión de datos se tienen en cuenta como dimensiones o características por lo general no fundamentales, aunque con términos y enfoques no estandarizados.
3. RQ3: Existen pocos estudios que consideran la gestión de datos como una dimensión concreta en su taxonomía o clasificación propuesta.
4. RQ4: Los aspectos de la gestión de datos considerados en mayor medida en los estudios incluidos son, a grandes rasgos: el almacenamiento, la integración, la interoperabilidad, la seguridad, la calidad y la minería de datos.

Estas conclusiones nos permiten identificar una necesidad por parte de la comunidad científica de homogeneizar y estandarizar los conceptos relativos a la gestión de datos y de proponer un marco de trabajo integral y agnóstico respecto al dominio que cubra estas cuestiones, trabajo que se plantea abordar en las posteriores fases del proyecto de tesis.

Bibliografía

- [1] Adithya Sridhar, Muthamilselvi Ponnuchamy, P. Senthil Kumar, Ashish Kapoor, Dai Viet Nguyen Vo, and Gayathri Rangasamy. Digitalization of the agro-food sector for achieving sustainable development goals: a review. *Sustainable Food Technology*, 1:783 – 802, 10 2023.
- [2] Revathi Nukala, Krishna Panduru, Andrew Shields, Daniel Riordan, Pat Doody, and Joseph Walsh. Internet of things: A review from 'farm to fork'. 2016. Cited by: 84.
- [3] N Mohanasuganthi, J Kanimozhi, V Preethi, M S Rohini, and V Swathilakshmi. A state-of-the-art survey of machine learning in the agricultural sector. 2022.
- [4] F Dal Mas, M Massaro, V Ndou, and E Raguseo. Blockchain technologies for sustainability in the agrifood sector: A literature review of academic research and business perspectives. *Technological Forecasting and Social Change*, 187, 2023.
- [5] Mario Lezoche, Herve Panetto, Janusz Kacprzyk, Jorge E. Hernandez, and M^a Mar Eva Alemany Díaz. Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. *Computers in Industry*, 117, 2020. Cited by: 367; All Open Access, Bronze Open Access, Green Open Access.
- [6] Michael Grieves and John Vickers. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, pages 85–113, 1 2016.
- [7] E. H. Glaessgen and D. S. Stargel. The digital twin paradigm for future nasa and u.s. air force vehicles. *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2012.
- [8] Ashwin Agrawal, Martin Fischer, and Vishal Singh. Digital twin: From concept to practice. *Journal of Management in Engineering*, 38:6022001, 3 2022.

- [9] Martin Enders and Nadja Hoßbach. *Dimensions of Digital Twin Applications - A Literature Review*. 3 2019.
- [10] Tianyu Zhang, Shengjuan Yu, Yihao Pan, He Li, Xinqi Liu, and Jinnuo Cao. Properties of texturized protein and performance of different protein sources in the extrusion process: A review. *Food Research International*, 174:113588, 2023.
- [11] Barbara Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. 2, 3 2007.
- [12] Jaqueline B Correia, Mara Abel, and Karin Becker. Data management in digital twins: a systematic literature review. *Knowledge and Information Systems*, 65:3165–3196, 3 2023.
- [13] Galiveeti Poornima, Vinay Janardhanachari, and Deepak S. Sakkari. Data management for iot and digital twin. pages 28–45, 10 2022.
- [14] Dama International. *DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition)*. Technics Publications, LLC, 2017.