

A Data Governance Framework for Industry 4.0

J. R. Yebeles and M. Zorrilla

Abstract— The fourth industrial revolution, or Industry 4.0, represents a new stage of evolution in the organization, management and control of the value chain throughout the product or service life cycle. This digitization of the industrial environment is characterized by the connection of Information Technologies (IT) and Operations Technologies (OT) through cyber-physical systems and the Industrial IoT (IIoT). One of the main consequences of this integration is the increasing amount and variety of data generated in real time from different sources. In this environment of intensive generation of actionable information, data becomes a critical asset for Industry 4.0, at all stages of the value chain. However, in order to data become a competitive advantage for the company, it must be managed and governed like any other strategic asset, and therefore it is necessary to rely on a Data Governance system. Industry 4.0 requires a reformulation of governance since the data is a key element and the backbone of the processes of the organization. This paper proposes a Reference Framework for the implementation of Data Governance Systems for Industry 4.0. Previously, it contextualizes data governance for Industry 4.0 environments and identifies the requirements that this framework must address, which are conditioned by the specific features of Industry 4.0, among others, the intensive use of big data, cloud and edge computing, artificial intelligence and current regulations.

Index Terms— Data governance, Data-Centric architecture, Industry 4.0, Big Data, Real Time.

I. INTRODUCCIÓN

La "Industria 4.0" (I4.0) supone la transformación de los procesos de producción aprovechando el uso de la abundante información disponible en cada fase de la cadena logística, de producción y de consumo [1].

I4.0 se caracteriza principalmente por la integración de las Tecnologías de la Información (IT) y las Tecnologías de Operaciones (OT) [2], esto es, la unión del mundo físico y digital mediante sistemas ciberfísicos y el IoT Industrial lo que supone un cambio en el modelo productivo, que pasa a estar basado en la ubicuidad y conectividad de datos, personas, procesos, servicios y sistemas ciberfísicos, como si fuera una red social en la que todos los actores (nodos de la red) intercambian y explotan la información generada en cada nivel y cuya principal consecuencia es el incremento en la cantidad y variedad de datos generados en tiempo real desde diferentes fuentes. En este entorno, el dato pasa a ser un activo crítico para la I4.0, en todas las etapas de la cadena de valor.

Para lograr la funcionalidad que requiere la I4.0, es necesario desarrollar y adoptar métodos, tecnologías y herramientas destinadas a gestionar las características específicas de los procesos industriales, tales como disponer de alta capacidad de cómputo escalable dinámicamente; la posibilidad gestionar

grandes volúmenes de datos en tiempo real, procedentes de un gran número de dispositivos heterogéneos, muchos de ellos heredados; facilitar la interacción dinámica de las aplicaciones con entornos inteligentes y el uso de inteligencia artificial y aprendizaje automático (IA y AA) [39]. Actualmente se considera que las plataformas digitales de tercera generación (3P) [3], son la solución más adecuada al reto tecnológico que supone la I4.0.

En este entorno tan complejo, el dato se convierte en un activo clave para el negocio, por lo que adoptar un modelo centrado en los datos (*Data-Centric*) resulta de vital importancia. En este modelo, los datos se separan de las aplicaciones y las plataformas tecnológicas, lo que evita la aparición de silos y permite que los datos sean compartidos y aprovechados por la totalidad de la empresa [4].

Sin embargo, para que el dato pueda convertirse en una ventaja competitiva, debe ser gestionado y gobernado como cualquier otro activo estratégico, y para ello se hace necesario crear un sistema de Gobernanza de Datos (GD) [5].

Aunque la mayoría de las organizaciones llevan a cabo cierto grado de GD [6], los cambios que conlleva la I4.0, como son la adopción de tecnologías que soportan las 3P [7], la colaboración interempresarial, los aspectos regulatorios y los acuerdos de niveles de servicio contratados con terceros, agregan una complejidad adicional al sistema de GD. Esto, unido a las características específicas del entorno industrial, nos ha llevado a elaborar un marco para el desarrollo de un sistema de GD adecuado a la I4.0 después de comprobar, mediante una revisión sistemática [8], su inexistencia. A partir de la literatura y de nuestra interacción con empresas del sector, se identificaron y especificaron previamente los requisitos del marco de GD tanto generales como específicos.

Conviene señalar que, de los marcos de GD analizados ninguno está específicamente dirigido a la I4.0. De aquellos que hacen un enfoque más general y que se pueden considerar más completos [5,25,41,42,43], ayudan a elaborar un programa de GD. [42] incluso presenta una hoja de ruta de siete fases para el desarrollo de un programa de GD y [43] describe una serie de pasos para su implementación, pero ninguno presenta una metodología para pasar de la elaboración de un programa de GD a su implementación en la empresa utilizando una arquitectura empresarial de GD. Una vez definido un programa de GD, nuestro marco proporciona una arquitectura de referencia que facilita la implantación del sistema de GD considerando los requisitos de la I4.0. Esto permite elaborar un plan de GD más detallado, comunicarlo mejor, determinar las capacidades que la empresa necesita y verificar que está alineado con los objetivos y estrategias empresariales. No

obstante, nuestro marco es complementario a los anteriores y se puede considerar como una herramienta de ayuda a su implementación.

El marco que se presenta forma parte de la arquitectura de referencia RAI4.0 [9], donde los datos y su gobierno se sitúan en el centro y se formula en base a tres principios: el Dato como Servicio (*DaaS*), la Plataforma como Servicio (*PaaS*) y la Monitorización como Servicio (*MaaS*).

El método de investigación seguido en este trabajo es Investigación-Acción [40]. Se trata de un método iterativo que involucra a investigadores y profesionales que actúan juntos en un ciclo particular de actividades, incluyendo el diagnóstico de problemas, la intervención de acción y el aprendizaje reflexivo.

Después de esta introducción, en la sección II se especifican los requisitos que el marco de GD debe abordar y que están condicionados por las características específicas de la I4.0. La sección III expone los aspectos, tanto de carácter conceptual como normativo, en que se basa el desarrollo del marco propuesto. La sección IV presenta el marco para la construcción de sistemas de GD en I4.0 y desarrolla algunos aspectos del mismo. Por último, la sección V recoge las conclusiones y aportaciones de este trabajo, al tiempo que propone futuras líneas de investigación.

II. REQUISITOS DEL SISTEMA DE GOBERNANZA DE DATOS

Un sistema de GD tiene como finalidad habilitar y establecer en la organización las capacidades necesarias para ejercer la toma de decisiones consensuada y comunicada, la autoridad y el control sobre la gestión de los activos de datos y definir quién tiene los derechos de toma de decisiones y las responsabilidades en los procesos relacionados con los datos [5].

El sistema de GD está condicionado por la misión, estrategia, normas y cultura de la organización [10] de forma que ésta pueda gestionar sus datos como un activo estratégico. Para ello, el sistema de GD debe orquestrar personas, procesos y tecnologías [11] y debe ser transversal a los principales departamentos de una organización [12].

La GD también guía las actividades relacionadas con la gestión de los datos. En definitiva, la GD se asegura de que los datos se gestionan de la forma adecuada, mientras que la Gestión de Datos (DM) se responsabiliza de gestionar los datos para alcanzar los objetivos de la organización [5], siguiendo las directrices establecidas por la GD.

El sistema de GD debe establecer [13]:

- El ámbito de la toma de decisiones en relación con el gobierno de los datos. Se refiere a identificar qué actividades relacionadas con los datos y qué aspectos específicos de los mismos deben ser objeto de gobierno.
- Los roles involucrados en el proceso de toma de decisiones. Algunos roles que se mencionan frecuentemente en la literatura son *data stewards*, *data owners*, y *data committees*, entre otros.
- Cómo se relacionan los roles involucrados con la toma de decisiones. Se refiere a los derechos de toma de decisiones, autoridad y responsabilidades asignadas a los roles.

A. Requisitos Generales del Sistema de GD

En [14], se establecen las condiciones que debe cumplir un sistema de gobernanza, que asumimos y aplicamos a nuestro sistema de GD, a saber: satisface las necesidades de los grupos de interés y genera valor; es dinámico; distingue entre actividades y estructuras de Gobernanza de Datos y las de Gestión de Datos; y, se ajusta a las necesidades de la empresa y tiene en cuenta toda la cadena de valor.

B. Requisitos Específicos del Sistema de GD

El sistema de GD debe satisfacer una serie de requisitos propios de I4.0, lo que supondrá evolucionar las capacidades actuales de la organización e incorporar otras nuevas; algunas de tipo organizacional; otras enfocadas a los procesos; mientras que otras corresponden al uso de tecnologías de automatización, AA e IA. Todo ello teniendo en cuenta las complejidades reales a las que la GD tiene que enfrentarse en un entorno I4.0, como son: sistemas masivamente distribuidos, uso de sistemas heredados e integración de servicios de terceros, principalmente.

A continuación, se especifican los requisitos que debe satisfacer el sistema de GD organizados según apliquen a principios, gobierno, gestión o monitorización del sistema de gobernanza.

1) Principios

a. Requisitos asociados a Principios de GD

Se deben fijar los principios [15] que guían la conducta, el comportamiento y la filosofía de la empresa respecto al uso, gestión y gobierno de los datos. Estos deben orientarse hacia una arquitectura centrada en el dato y estar alineados y soportar las metas y objetivos de la GD y de la empresa.

2) Gobernanza

a. Requisitos de Alineamiento estratégico

La GD debe definir metas, objetivos y estrategias alineados con los de la organización [5] y contar con métodos de monitorización para verificar y asegurar que esto se cumple. También debe recoger las necesidades de información de la organización y facilitar su uso para la toma de decisiones [16].

b. Requisitos Organizacionales

Debe definir el modelo de organización con el que se va a operar y los órganos de gobierno, órganos gobernados y roles. Estos últimos definidos por las actividades relacionadas con los datos, que sean objeto de gobierno [17].

Los roles deben asignarse a perfiles con capacidad de evolución dinámica, fácilmente adaptables a los cambios normativos y tecnológicos [18] y deben estar alineados con la arquitectura industrial que se esté aplicando para implantar el modelo I4.0 (IIRA, RAMI, etc.) [19]. Igualmente, se debe extender el sistema de GD (incluidos roles) [20] a las organizaciones externas que formen parte de la Integración Vertical y Horizontal de la I4.0 (proveedores, distribuidores,

etc.). Esto conllevará la firma de contratos con compromisos de implantación y cumplimiento con el sistema de GD estipulado por la empresa [21].

c. *Requisitos de Gobierno y Administración de los datos*

El sistema de GD debe establecer un modelo de gobierno basado en las funciones de Evaluar, Dirigir y Supervisar [22] y concretar qué actividades relacionadas con los datos deben ser objeto de gobierno (Ciclo de Vida del Dato, CVD), y qué aspectos específicos de los datos se van a gobernar (calidad, seguridad, metadatos) [25] y definir políticas a tal efecto. Estas políticas deben contemplar el cumplimiento de las legislaciones, normativas y regulaciones que afecten a la GD y deben considerar la distribución de los datos a través de los diferentes niveles de integración Vertical y Horizontal de la I4.0 y a lo largo de todas las etapas del CVD.

Por otra parte, debe asignar a los roles definidos, derechos de decisión, autoridad y responsabilidades sobre los activos de datos [5] y establecer una forma de gobierno en función de modelos *Agile*, *DataOps* y *DataGovOps* [24].

3) *Gestión*

a. *Requisitos de Clasificación y Metadatos*

El sistema de GD debe establecer un Glosario de Negocio, un Diccionario de Datos y un Catálogo de Datos, así como las políticas y procesos para su gestión, de forma que facilite la comprensión del contexto, importancia y asociaciones de los datos, al tiempo que establece un lenguaje común y un modelo de referencia para abordar la gran variedad de datos que definen los conceptos del negocio y las relaciones entre ellos. Igualmente, debe recabar los metadatos necesarios para registrar el *data lineage* incorporando procesos para el análisis del impacto que provoquen los cambios que se realicen en los artefactos de datos. Esto implica registrar y organizar todos los metadatos relacionados con los datos, incluido el código que actúa sobre los datos [24].

Las actualizaciones del Glosario, Diccionario y Catálogo se deben automatizar incluyéndolas en el proceso de gestión de cambios (como si se tratara de actualizaciones de código) [26].

Debe contemplar también la automatización de procesos de búsqueda, captura, interpretación y enriquecimiento de metadatos e incluir técnicas y enfoques avanzados en la catalogación y clasificación de los metadatos mediante el uso de tecnologías de AA e IA. Por último, se deben explorar modelos federados de gestión de metadatos, debido a la gran cantidad de sistemas implicados.

b. *Requisitos de Calidad de los Datos*

Debe identificar los parámetros por los que se medirá la calidad de los datos (v.g. precisión, fiabilidad, integridad, oportunidad, etc.) y considerar el uso de tecnologías escalables de automatización para la evaluación y validación automática, continua y en tiempo real, de la calidad de los datos en cada fase del CVD, emitiendo alertas en tiempo real o incluso corrigiendo el problema en su caso.

Debido a la gran variedad de fuentes de datos, su volumen y diferente naturaleza, será necesario especificar reglas de calidad de los datos para cada caso de uso y tipología [27].

c. *Requisitos de Seguridad, Privacidad y Riesgo*

Debe identificar los datos sensibles y establecer clasificaciones de seguridad de los mismos, teniendo en cuenta los requisitos de negocio, regulaciones, normativas y legislación que le son aplicables y, en función de esta clasificación, establecer políticas sobre qué se puede hacer con los datos, quién puede hacerlo, bajo qué términos y en qué condiciones (autenticación, autorización y no-repudio).

Debe garantizar la confidencialidad, integridad y alto nivel de disponibilidad de los datos y establecer mecanismos para validar que tanto la fuente de datos de entrada como el consumidor de los mismos están autenticados.

Será necesaria la aplicación de modelos de seguridad adaptados a I4.0 [29] que protejan también los datos operativos usados, almacenados o movidos en los diferentes niveles de integración de la I4.0; datos de configuración de los sistemas; de operación de las redes e interconectividad y datos relativos a la monitorización de los sistemas. También debe utilizar diferentes mecanismos y enfoques de protección de datos dependiendo de si los datos están en uso, en movimiento o en reposo [30].

Debe establecer políticas específicas y diferentes a las aplicadas en los sistemas centralizados, para protección de la información, copias de respaldo y para hacer frente a posibles deterioros de los datos y recuperación ante desastres [28], así como establecer una capacidad de respuesta extremadamente rápida a los problemas y amenazas de seguridad.

d. *Requisitos del Ciclo de Vida de los Datos*

En relación a la fase de **Planificación y diseño**, se debe establecer los requisitos de la arquitectura de datos de la organización y gestionar los diseños, herramientas y ciclo de vida relativos a la misma. Esta arquitectura debe estar alineada y tener en cuenta la arquitectura industrial que se esté aplicando para implantar el modelo I4.0 y el intercambio y la compartición de datos a lo largo de todos los niveles de Integración Vertical y Horizontal de la I4.0, tanto dentro como fuera de la empresa.

Para soportar distintos casos de uso de la I4.0, la arquitectura debe establecer la necesidad de un repositorio de metadatos común y distribuido y tener en cuenta las distintas tipologías de datos (Estructurados, Semiestructurados, No-estructurados).

En la etapa de **Captura y distribución**, se deben establecer políticas y estándares sobre la Integración e Interoperabilidad de los datos que se incorporan y cómo éstos se deben extraer y tratar para la creación de un dato útil (limpio, corregido, completo y riguroso) y que se puede almacenar. Para ello debe tener en cuenta las diferentes estrategias de captura y transferencia de datos, latencia y otros requisitos no funcionales.

Se deben establecer *Data Sharing Agreements* que estipulen las responsabilidades y el uso aceptable de los datos que se

capturan y distribuyen. En este sentido, se considera conveniente la aplicación de estándares (v.g. International Data Spaces) que permitan facilitar el intercambio y compartición de datos a lo largo de la cadena de valor, dentro y fuera de la empresa, al tiempo que permite a los distintos actores, definir contratos legibles por software adjuntos a los datos (*Self-Aware Contracts*).

Se deben establecer políticas respecto de la gestión de fuentes de datos (registro y aceptación, modificación, etc.) e incorporar sistemas de validación automática y en tiempo real de los datos que se capturan, emitiendo alertas en tiempo real cuando sea necesario, además de establecer políticas que permitan detener la entrada de datos de una fuente con errores [31].

Se debe evaluar la criticidad de los datos que llegan de sistemas ciberfísicos y establecer políticas, reglas y procesos al respecto.

En la etapa de **Almacenamiento**, se deben establecer políticas, estándares y procesos para el almacenamiento, mantenimiento y mejora de los datos (el dato no cambia intrínsecamente) o para su archivo, así como para la auditoría de los datos almacenados, teniendo en cuenta los requisitos legales al respecto y los distintos niveles de Integración Vertical y Horizontal de la I4.0; principalmente en el *Edge*, debido a su complejidad técnica.

En la fase de **Preparación y uso**, se debe establecer políticas, estándares y procesos para la preparación, transformación y uso de los datos en los distintos niveles de Integración de la I4.0, (v.g. selección y filtrado de datos en el *Edge*).

En cuanto a la **Toma de decisiones**, se debe definir un proceso de delegación que asegure que las decisiones tomadas, manuales o automáticas, consecuencia del análisis o procesado de un conjunto de datos, se ajustan al nivel de responsabilidad del rol que las toma y establecer los controles apropiados, incluida la intervención manual, para hacer frente a cualquier sesgo o discriminación en el proceso de toma de decisiones.

Debe establecer procesos para valorar la utilidad de los datos usados para la toma de decisiones y utilizar este nuevo metadato para enriquecer estos datos y así mejorar la toma de decisiones futuras.

En la última etapa, **Destrucción**, se deben establecer políticas, estándares, procesos, etc. para la eliminación de los datos de forma permanente

4) Monitorización

a. Requisitos de Supervisión y Evaluación

Se deben establecer políticas, procesos y un sistema de medidas e indicadores que permitan monitorizar el rendimiento del uso de los datos en la organización y asegurar que las estrategias relacionadas con los datos se han implementado correctamente, así como que el uso y gestión de los datos se lleva a cabo conforme a las políticas internas y requisitos externos establecidos.

Se deben automatizar los procesos de monitorización mediante el uso de tecnologías de AA e IA a lo largo del CVD, haciendo posible la monitorización y respuesta en tiempo real.

III. LA GOBERNANZA DE DATOS COMO SISTEMA

Según la definición de “sistema” propuesta en [32] y tal y como se recoge en [33], la gobernanza en general y la GD en particular se pueden concebir como un sistema. Por otra parte, se debe tener en cuenta que la GD es una función empresarial [34] y como tal, debe estar alineada y ser coherente con las metas, objetivos y estrategias de la empresa [5]. Además, como se indica en [35], para definir y representar un sistema se utiliza una arquitectura, la cual se expresa mediante una “descripción de la arquitectura” que a su vez, identifica el sistema en cuestión. Esto es válido para cualquier sistema, incluido un sistema de GD. En este sentido, *The Open Group*, tomando como base la ISO/IEC/IEEE 42010:2011, desarrolló TOGAF® Standard [15], un *framework* para construir arquitecturas empresariales cuyo alcance puede ser la totalidad de la empresa o áreas o partes concretas de la misma, como puede ser el caso de la función empresarial de GD. Por ello, y según lo expresado anteriormente, podemos plantear el desarrollo de un sistema de GD y representarlo mediante una arquitectura empresarial.

Este es el objeto de este trabajo de investigación que propone el desarrollo de un Marco para la Construcción de Sistemas de GD en I4.0 y su representación mediante una arquitectura de GD, tomando como base los conceptos establecidos en la norma ISO/IEC/IEEE 42010:2011 y el *framework* TOGAF® Standard V9.2, para la construcción de arquitecturas empresariales. Además, este marco pretende ser complementario y compatible con el uso de otros marcos de gobernanza IT, como COBIT® 2019, o de gestión de datos como DAMA-DMBOK® y el modelo MAMD para mejora de los Datos [36], entre otros.

IV. MARCO PARA LA CONSTRUCCIÓN DE SISTEMAS DE GD EN I4.0

Una vez especificados los requisitos y establecidas las bases de nuestro trabajo, se procede a definir formalmente un Marco para la Construcción de Sistemas de GD en I4.0 (Fig. 1). Este marco está formado por una Arquitectura de Referencia para la representación del sistema de GD, un Método que indique los pasos a seguir para el desarrollo de la arquitectura, una Lista de estándares recomendados y un Modelo de madurez [15].

Todo ello, nos permitirá instanciar sistemas de GD alineados con la estrategia empresarial de una organización y describir la arquitectura del sistema de GD en una Industria 4.0.

Por motivos de espacio, a continuación se hace un compendio de la arquitectura de referencia y de la descripción de un Bloque Constitutivo de Arquitectura (ABB) del modelo de referencia, quedando fuera del alcance de este documento el resto de elementos del marco.

La arquitectura de referencia está formada por un Metamodelo de Contenidos y un Modelo de Referencia que contiene un conjunto de Bloques Constitutivos de Arquitectura

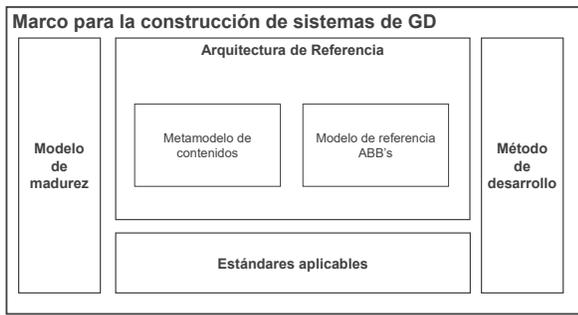


Fig. 1. Marco de referencia.

(ABB). Esta arquitectura de referencia permite describir de forma consistente la arquitectura de un sistema de GD.

El metamodelo de contenidos elegido es el definido en el *framework* TOGAF® Standard V9.2 (Fig. 2), al que hemos añadido una nueva entidad “*Policy*” derivándola a partir de la entidad “*Principle*” del metamodelo inicial. Este metamodelo establece una estructura formal de entidades con sus atributos; un esquema de relaciones entre ellas y las reglas que rigen dichas relaciones, lo que permite definir, estructurar y presentar contenidos de la arquitectura de forma consistente. Por lo tanto, este metamodelo nos proporciona una herramienta para definir los ABB que conforman la arquitectura.

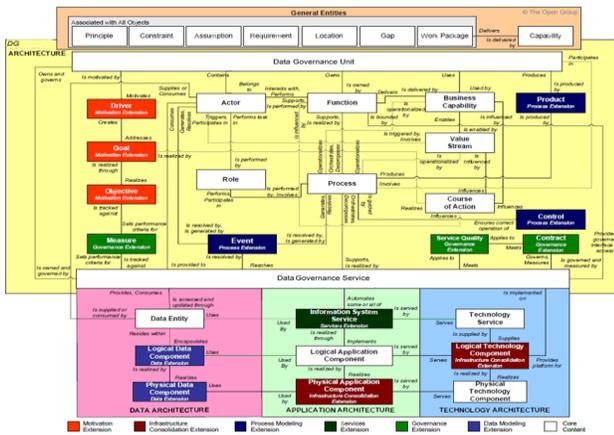


Fig. 2. Metamodelo de contenidos TOGAF® Standard V9.2.

Estos ABB y las relaciones entre ellos se describen y representan utilizando el modelo de *Architecture description* descrito en la norma ISO/IEC/IEEE 42010:2011.

Nuestro modelo de referencia para la arquitectura de GD (Fig. 3) define un conjunto de ABB, cada uno de los cuales es una parte de la arquitectura y especifica una funcionalidad y unas capacidades que la arquitectura debe implementar. Estos ABB modelan los requisitos expresados en la Sección II.

Para especificar, describir y gestionar de manera formal ABB reusables, hemos creado *ABB-profile* (Fig. 4), una extensión del “*Default Profile*” del estándar *Reusable Asset Specification (RAS), Version 2.2* [37].

Para su adaptación a nuestro propósito, hemos extendido el *default profile* de *RAS* con las clases siguientes:

- **Requirement.** Describe cada requisito que el ABB contribuye a cumplir.
- **Architecture-description.** Esta clase ayuda a describir y comunicar distintas partes de la arquitectura definidas por el ABB, siguiendo la norma ISO/IEC/IEEE 42010:2011. Está

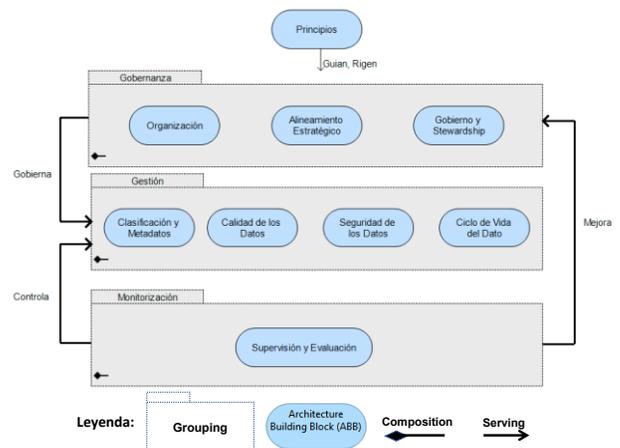


Fig. 3. Modelo de referencia.

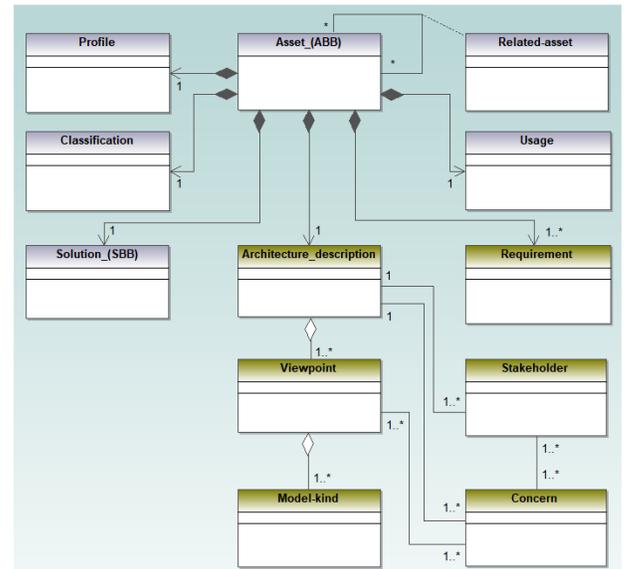


Fig. 4. *ABB_profile*. (Modelo UML).

compuesta por las clases siguientes:

- **Stakeholder.** Contiene información sobre un individuo, equipo, organización o categorías de los mismos, que tenga interés en el sistema de GD y para el cual se construyen los *Viewpoint*.
- **Concern.** Describe un aspecto del sistema que es objeto de preocupación o de interés para algún Stakeholder.
- **Viewpoint.** Establece una especificación de las pautas para un tipo particular de vista de la arquitectura del ABB. Tiene, al menos, los atributos siguientes: *Name*, *Type*, *Description*.
- **Model-kind.** Esta clase describe las normas de modelado utilizadas por el *Viewpoint*, tomando como referencia el Metamodelo de Contenidos.

En este artículo, describimos el *Asset_(ABB)* Políticas y estándares, que se enmarca dentro del *Asset_(ABB)* Gobierno y *Stewardship*, (Fig. 5) perteneciente al grupo de Gobernanza.

1) *Asset_(ABB)* Políticas y estándares

La Tabla I describe el *Asset_(ABB)* Políticas y estándares, que tiene la estructura reflejada en la Fig. 6.

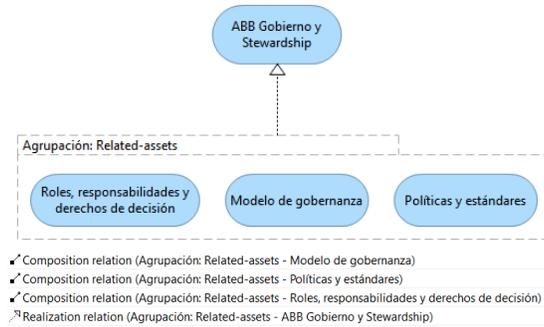


Fig. 5. Asset (ABB) Gobierno y stewardship.

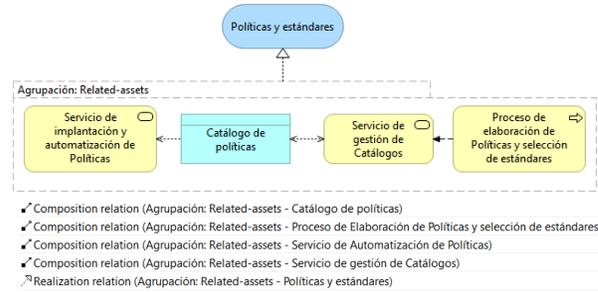


Fig. 6. Asset (ABB) Políticas y estándares.

La clase *Architecture-description* en este ABB está compuesta por los *Viewpoint* Catálogo de políticas (Tabla II) y Diagrama de Políticas y Principios a los que contribuye (Tabla IV).

TABLA I.
ASSET (ABB) POLÍTICAS Y ESTÁNDARES

Name	Políticas y estándares
<i>Description</i>	Permite pasar de los principios a la especificación de las políticas y estándares que se deben aplicar para las actividades del CVD, en los aspectos específicos de los datos que serán objeto de gobierno (calidad, seguridad, metadatos, etc.) y a los niveles de Integración Vertical y Horizontal de la I4.0. Facilita la identificación de los aspectos a tener en cuenta para la monitorización del desempeño y cumplimiento de las políticas fijadas. Promueve la automatización de los procesos de GD con la implantación de conceptos como “ <i>Continuous Governance</i> ”, metodologías como “ <i>DataGovOps</i> ” y tecnologías de “ <i>Governance as code</i> ”.
<i>Requirement</i>	Este bloque de funcionalidad implementa los requisitos de definición de políticas y estándares; de cumplimiento de las normativas y regulaciones relacionadas con los datos, tanto internas de la empresa, como externas, que afecten a la organización; de identificación de los aspectos a tener en cuenta para la monitorización del desempeño y cumplimiento de las políticas fijadas y de automatización de las políticas de GD.
<i>Classification</i>	Realización de la estrategia; Políticas de GD; <i>DataGovOps</i> ; Automatización de procesos; <i>Continuous Governance</i> .
<i>Usage</i>	Sirve para elaborar, gestionar e implantar las políticas de GD. Se usa como una referencia para la toma de decisiones, para traducir los Principios en Políticas y Reglas y para demostrar la coherencia entre las Políticas y los Principios de las que se derivan. También para desarrollar la automatización de las políticas.
<i>Related-asset</i>	Name: <i>Servicio de gestión de Catálogos</i> . Permite la realización de altas, bajas y modificaciones en el catálogo. Permite la gestión del acceso a la información del catálogo. Relationship-type: Composición.
<i>Related-asset</i>	Name: <i>Proceso de elaboración de Políticas y selección de estándares</i> . Define la secuencia de actividades a llevar a cabo para la elaboración de las políticas y selección de estándares relacionados con las mismas. Relationship-type: Composición.
<i>Related-asset</i>	Name: <i>Servicio de implementación y automatización de Políticas</i> . Permite la implementación de las políticas y estándares mediante la aplicación de reglas que se ejecutan bien de forma manual, bien de forma automática leyendo las políticas del catálogo, en lenguaje natural y transformarlas en reglas en lenguaje técnico, reconocidas a nivel máquina y aplicables de forma automática a los datos. Relationship-type: Composición.
<i>Related-asset</i>	Name: <i>Catálogo de Políticas</i> . Define la entidad de datos que contiene las políticas especificadas en el <i>viewpoint</i> Catálogo de Políticas. Relationship-type: Composición.

TABLA II.
VIEWPOINT CATÁLOGO DE POLÍTICAS

<i>Viewpoint name</i>	Catálogo de Políticas.
<i>Description</i>	Este catálogo recoge las políticas que se han definido para cumplir con los principios y aplicar las estrategias de GD por cada actividad del CVD y cada aspecto específico de los datos que sea objeto de gobierno (calidad, seguridad, metadatos, etc.). El catálogo contiene las políticas y los estándares relacionados con las mismas, según una estructura fijada (ver <i>Model-kind</i>).
<i>Type</i>	Catálogo.
<i>Stakeholders</i>	<i>Data stewards</i> , arquitectos empresariales y TIC.
<i>Concerns</i>	Misión y estrategia, aplicación de los principios, motivación. Está compuesto por entidades del tipo <i>Policy</i> , derivada de la entidad <i>Principle</i> del metamodelo y contiene los atributos siguientes: (ver ejemplo en Tabla III)
<i>Model-kind</i>	<ul style="list-style-type: none"> • Id. Identificador exclusivo y único de la política. • Entity name. En este caso <i>Policy</i>. • Description. Las Políticas son directrices que gobiernan, rigen y orientan la actuación de la empresa en un asunto o campo determinado, en este caso respecto de los datos y su gobierno [5]. Las políticas permiten traducir los principios a reglas que gobiernan la gestión de los datos. Las políticas describen el “Qué” de la GD y los estándares y procedimientos el “Cómo”. • Category. Viene determinadas por las actividades del CVD, y los aspectos específicos de los datos que serán objeto de gobierno (calidad, seguridad, metadatos, etc.). Por lo tanto, una categoría podría ser: “<i>Políticas de seguridad en el almacenamiento de datos</i>”. • Owner. Responsable de la definición y gestión de la política. • Name. El nombre que le asignamos a la política. • Purpose/Objective. Finalidad que tiene esta política y objetivos que se pretende cumplir con esta Política. • Statement. Declaración que expone la política de forma inequívoca, concisa y clara. • Scope/affected audience. Indica a quién afecta esta política. • Procedures. Relaciona de forma sucinta los procedimientos que se derivan de esta Política. • Standards. Se referencian aquí los estándares que contribuyen y ayudan a implementar la Política y que estarán incluidos en un Repositorio de Estándares de la arquitectura. • Date. Entrada en vigor y, en su caso, fecha de revisión y motivos. • Metric. Identifica los mecanismos que se utilizarán para evaluar el cumplimiento de la Política.

TABLA III
EJEMPLO DE ENTRADA EN EL CATÁLOGO DE POLÍTICAS

Id.	IM6_11
Entity name	Política
Description	Las políticas son directrices que gobiernan, rigen y orientan la actuación de nuestra empresa respecto de los datos y su gobierno. traducen los principios a reglas que gobiernan la gestión de los datos.
Category	Política de seguridad en la distribución de información.
Owner	Director de Protección de Datos.
Name	Política de intercambio de información (general).
Purpose /objective	Proporcionar orientación al personal de ACME sobre cuándo se debe realizar una Evaluación de Impacto de Protección de Datos y bajo qué circunstancias puede ser necesario un acuerdo para el intercambio de información.
Objetivos:	<ul style="list-style-type: none"> • Proporcionar un marco para aclarar los procedimientos locales relacionados con el intercambio de información del usuario del servicio. • Asegurar que solo se comparta la información mínima necesaria para el propósito. • Proporcionar un mecanismo para que los signatarios de esta política acuerden que cumplirán con términos y directrices contenidos en esta política.
Statement	Esta política describe los estándares de ACME para el intercambio de información. Es obligatorio informar fehacientemente a la persona cuya información se quiere compartir sobre la posibilidad de compartirla y las opciones que tiene para limitar ese intercambio. Si la persona dice NO a compartir la información confidencial, esta no se compartirá.
Scope	Todo el personal de ACME
Procedures	P01_IM_11 Procedimiento de evaluación e intercambio de información
Standards	REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016, relativo a la protección de al tratamiento de datos personales y a la libre circulación de estos datos.
Date	Fecha de vigor: 20/09/2019
Metric	Se verificará la autorización sobre el intercambio de información. Se registrarán las No conformidades por incumplimiento de la política

TABLA IV
VIEWPOINT DIAGRAMA DE POLÍTICAS

Viewpoint name	Diagrama de Políticas y Principios a los que contribuye.
Description	Permite justificar de forma visual que una política determinada contribuye a la aplicación de un principio.
Type	Diagrama.
Stakeholders	Data stewards, arquitectos empresariales y TIC.
Concerns	Aplicación de los principios, metas y objetivos, motivación. Se utilizan las entidades <i>Policy</i> , <i>Principle</i> , <i>Goal</i> y <i>Objective</i> del metamodelo, asociándolas mediante las relaciones <i>Influence</i> y <i>Realization</i> que indican que una entidad <i>Policy</i> puede tener una influencia en la aplicación de una entidad <i>Principle</i> y por ende en la realización de una entidad <i>Goal</i> .
Model-kind (ver ejemplo en Fig. 7)	Principle. Existe una relación <i>Influence</i> entre una entidad <i>Policy</i> y una entidad <i>Principle</i> . Es decir, las políticas contribuyen a la aplicación de los principios. Objective y Goal. La aplicación de los principios mediante las políticas es fundamental para que se cumplan los objetivos y las metas.

V. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Este trabajo presenta el desarrollo de un Marco de Referencia para la construcción de sistemas de GD para la I4.0 soportada por tecnologías 3P (IoT, Tecnología Social, Dispositivos

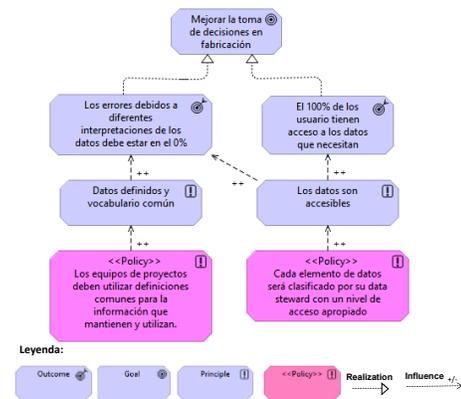


Fig. 7. Ejemplo de Diagrama de políticas y principios (usando Archimate® 3.1).

Móviles, Big Data y Computación en la Nube y en el Edge). Así mismo, se incluye la especificación de los requisitos que dan soporte a este marco, así como una definición formal del mismo.

En lo que se refiere a los requisitos, se ha puesto de manifiesto la importancia que para la GD en entornos I4.0 tiene el contar con perfiles y roles adaptados a las nuevas tecnologías disruptivas y a su evolución futura. También la necesidad de automatización de las políticas y procesos bajo el enfoque “Continuous Governance”, “DataGovOps” y “Governance as code”.

Por otra parte, la GD se ve afectada y debe tener en cuenta la integración Vertical y Horizontal en la I4.0 lo que supone la extensión del sistema de GD a todos los niveles, incluyendo a empresas proveedoras y distribuidoras y que debe reflejarse en contratos y establecimiento de acuerdos de niveles de servicio (SLA) entre las partes.

La gran cantidad y variedad de datos que se generan y se procesan en el entorno I4.0 requiere del uso de tecnologías y herramientas Big Data y Cloud Computing, lo cual representa también un reto en cuanto a asignación de responsabilidades sobre los datos, las políticas de seguridad y de protección de los datos (en uso, en movimiento o en reposo). A esto hay que añadir la necesaria integración de estas tecnologías con sistemas heredados masivamente distribuidos y heterogéneos.

La especificación de requisitos ha dado paso a definir un Marco de Referencia con una arquitectura que incluye los Bloques Constitutivos de Arquitectura (ABB) [38] a partir de los cuales puedan instanciarse los servicios, procesos y artefactos software necesarios para materializar el sistema de GD utilizando distintos sistemas de información y tecnologías que se seleccionen para su implantación. Así mismo, se ha desarrollado un modelo basado en [37] para especificar los ABB, y se ha descrito uno de ellos en detalle mostrando un ejemplo de instanciación. Como continuación en la línea de investigación se definirá un modelo de madurez que permita conocer a cada organización en qué nivel se encuentran los distintos procesos relacionados con la GD. Finalmente, se instanciarán el marco de desarrollo y el modelo de madurez en casos de uso reales en el ámbito industrial.

REFERENCIAS

- [1] Deloitte Consulting LLP. Industry 4.0. Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies, 1-32, (2015).
- [2] Velásquez. Cloud Computing, Big Data and the Industry 4.0 Reference Architectures. *Journal of Computer Science and Technology*, 18.03 (2018), 258-266. <<https://doi.org/10.24215/16666038.18.e29>>.
- [3] Kevin Kwang. “Third Platform” Shift Triggers Enterprise Software Evolution. *ZDNet*, 1-4 <<https://www.zdnet.com/article/third-platform-shift-triggers-enterprise-software-evolution/>> [accessed 26 October 2018].
- [4] Stratio. Data-Centric Architecture A Model for Embracing the Machine Age, 1-7, (2018).
- [5] DAMA International. DAMA-DMBOK2: Data Management Body of Knowledge, ed. by Deborah Henderson and Susan Earley, 2nd edn, (2017)
- [6] Alhassan. Data Governance Activities: An Analysis of the Literature. *Journal of Decision Systems*, 64-75, (2016) <<https://doi.org/10.1080/12460125.2016.1187397>>.
- [7] Mabkhot. Requirements of the Smart Factory System: A Survey and Perspective. *Machines*, 6.2 (2018), 1-22. <<https://doi.org/10.3390/MACHINES6020023>>.
- [8] Yebenes & Zorrilla. Towards a Data Governance Framework for Third Generation Platforms. *Procedia Computer Science*, 151 (2019), 614-21 <<https://doi.org/10.1016/j.procs.2019.04.082>>.
- [9] López. RAI4 Deployment Tool and Metamodel. <<https://github.com/istr-uc/RAI4DeploymentTool>> [accessed 28 September 2020].
- [10] Weber. One Size Does Not Fit All - A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, 1.1 (2009), 4:1-4:27 <<https://doi.org/10.1145/1515693.1515696>>.
- [11] MDM Institute. Data Governance Definition. <<http://www.tcdii.com/whatIsDataGovernance.html>>. [accessed 23 October 2018].
- [12] Gray & IBM Corporation. An Overview of Data Governance Elevator Pitch (Final), 1-4, (2011).
- [13] Otto, 2011. Data Governance. *Business and Information Systems Engineering*, 3.4 (2011), 241-44 <<https://doi.org/10.1007/s12599-011-0162-8>>.
- [14] ISACA, 2018. COBIT 2019 Framework. Introduction and Methodology. 1-64 <https://doi.org/10.1057/9780230299283_1>.
- [15] The Open Group, 2018. TOGAF® Standard Version 9.2. 197.
- [16] MIT Technology Review Insights, 2020. Data on Demand: Dynamic Architecture for a High-Speed Age, 4.
- [17] Felici, 2013. Accountability for Data Governance in Cloud Ecosystems., *Proceedings of the 2013 IEEE International Conference on Cloud Computing Technology and Science-Volume 2*, 2013, 327-332. <<https://doi.org/10.1109/CloudCom.2013.157>>.
- [18] Deloitte, 2018. Data Governance for Next-Generation Platforms, 1-11.
- [19] IIC. 2018. Industrial Internet Consortium. The Industrial Internet of Things Volume G2: Key System Concerns, 2018, pp. 1-43.
- [20] Sohail, 2018. 4 Pillars to Guide Data Governance for New Platforms. <<https://deloitte.wsj.com/cio/2018/10/10/4-data-governance-pillars-for-modern-data-platforms/>>, [accessed 26 July 2019].
- [21] Wallis, 2020. Agreements Between Enterprises Digitized by Smart Contracts in the Domain of Industry 4.0. *Computer Science & Information Technology (CS & IT)*, 23-32 <<https://doi.org/10.5121/csit.2020.101003>>.
- [22] ISO, 2017b. International Standard Organization - ISO/IEC 38505-1. Application of ISO/IEC 38500 to the Governance of Data, pp. 1-28.
- [23] ISO, 2018. International Standard Organization - ISO/IEC TR 38505-2. Implications of ISO/IEC 38505-1 for Data Management.
- [24] Strod, 2020. Continuous Governance with DataGovOps. *DataKitchen Blog*, 2020. <<https://blog.datakitchen.io/blog/continuous-governance-with-datagovops>> [accessed 15 September 2020]
- [25] Ballard, 2014. IBM Information Governance Solutions, ed. by IBM Corporation, 1-264.
- [26] Madsen & Bergh, 2020. Redefining Data Governance with DataGovOps (Webinar) <<https://info.datakitchen.io/watch-on-demand-webinar-redefining-data-governance-with-datagovops>>.
- [27] Williams & Tang, 2020. Data Quality Management for Industry 4.0: A Survey. *Data Quality*, 22.2 (2020), 26-35.
- [28] Collibra, 2018. The Need for Big Data Governance, 1-8.
- [29] IIC. 2016. Industrial Internet Consortium. The Industrial Internet of Things Volume G4: Security Framework, Industrial Internet Consortium, 2016, pp. 1-173.
- [30] IIC. 2109. Industrial Internet Consortium. Data Protection Best Practices, 1-38.
- [31] DataKitchen, 2020. What Is-Dataops. Ten Most Common Questions. *Blog of DataKitchen* <<https://blog.datakitchen.io/blog/what-is-dataops-ten-most-common-questions>> [accessed 2 March 2020].
- [32] ISO, 2017a. International Standard Organization - ISO/IEC/IEEE 24765 Systems and Software Engineering - Vocabulary, 2017, pp. 1-530.
- [33] ISO, 2008. International Standard Organization - ISO/IEC 38500:2008(E) Corporate Governance of Information Technology).
- [34] Brous, 2016. Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. *Lecture Notes in Computer Science*, 9820 LNCS, 115-125. <https://doi.org/10.1007/978-3-319-44421-5_9>.
- [35] ISO, 2011. International Standard Organization - ISO/IEC/IEEE 42010:2011 Systems and Software Engineering — Architecture Description, 2011, pp. 1-46 <<https://doi.org/10.1080/21670811.2017.1279978>>.
- [36] Carretero, 2017. MAMD 2.0: Environment for Data Quality Processes Implantation Based on ISO 8000-6X and ISO/IEC 33000. *Computer Standards and Interfaces*, 54. September 2016 (2017), 139-51 <<https://doi.org/10.1016/j.csi.2016.11.008>>.
- [37] Object Management Group, 2005. Reusable Asset Specification, Version 2.2, pp. 1-121.
- [38] The Open Group, 2014. TOGAF® 9.1 Translation Glossary: English – Castilian Spanish, 1-45.
- [39] C González García, ER Núñez-Valdez, V García-Díaz, Juan Manuel Cueva Lovelle A Review of Artificial Intelligence in the Internet of Things, *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(4), pp. 9-20, 2019.
- [40] Avison, D.; F. Myers L.M.; Nielsen A. Action Research. (1999). *Communications of the ACM*, 42(1): 94-97.
- [41] Katherine Downing, AHIMA Information Governance & The Information Governance Adoption Model (IGATM), 2017 <[http://www.himconnect.ca/Documents/The Information Governance Adoption - AHIMA.pdf](http://www.himconnect.ca/Documents/The%20Information%20Governance%20Adoption%20-%20AHIMA.pdf)>.
- [42] Gwen Thomas, ‘The DGI Data Governance Framework’, The Data Governance Institute, Orlando, FL (USA), 2006, 20.
- [43] Majid Al-Ruithe, ‘Development and Evaluation of a Holistic Framework and Maturity Assessment Tools for Data Governance in Cloud Computing Environments’ (Staffordshire University, 2018).



Juan R. Yében es Ingeniero Informático por la Universidad Politécnica de Madrid y Máster en Dirección Comercial y Marketing por el IE Business School. Ha ocupado cargos ejecutivos en diferentes empresas nacionales y multinacionales relacionadas con las TIC y en la actualidad está realizando su tesis doctoral sobre la Gobernanza de Datos en la Industria 4.0 en la Universidad de Cantabria.



Marta Zorrilla es Profesora Titular de Universidad en el Grupo de Ingeniería de Software y Tiempo Real de la Universidad de Cantabria (España). Ha participado en más de 30 proyectos de investigación nacionales y europeos. Sus intereses de investigación son las tecnologías de gestión de datos, ciencia de datos y big data, aplicados a la Industria 4.0. Es autora de un libro de bases de datos y más de 70 trabajos publicados.