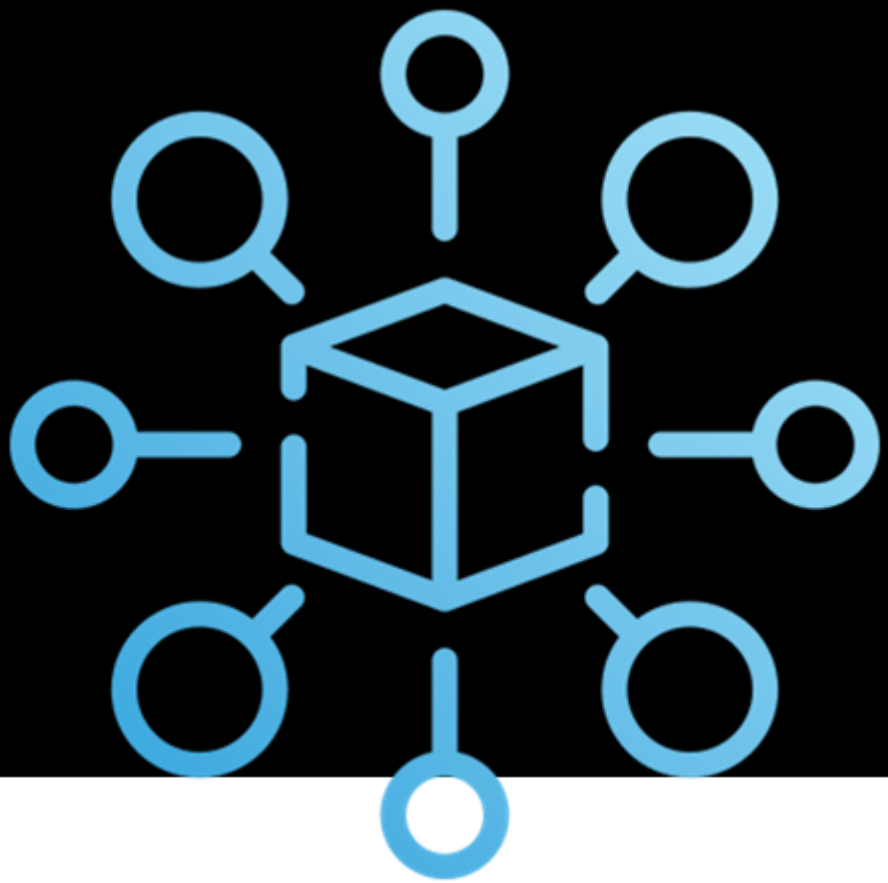
A European Common Digital Manufacturing Infrastructure and Data Space Pathway for Connected Factories 4.0 Data Value Chain Governance

Title	D5.1 - Smart Data Model's Repository
Document Owners	Alberto Abella (FF), Francisco Melendez (FF), Clara Pezuela (FF)
Contributors	
Dissemination	PU
Date	9/9/2024
Version	2.0

Digital Europe EU Grant Agreement: 101083939



Document History

14/11/2022	V0.1 Structure of the deliverable
12/12/2022	V0.2 First draft of the document
16/12/2022	V0.3 INNO Revision
16/12/2022	V0.4 SQS Revision
29/12/2022	Final version
09/09/2024	Reviewed version

Document Fiche

Authors	Alberto Abella (FF), Francisco Melendez (FF), Clara Pezuela (FF)
Internal Reviewers	INNO, SQS
Work Package	WP5: Data prioritisation and roadmapping
Task	T5.1 H2020 Embryonic Data Space Benchmarking
Nature	Other
Dissemination	Public

Project Partners

Participant organisation name	Acronym
ASOCIACIÓN DE EMPRESAS TECNOLÓGICAS INNOVALIA	INNO
FONDAZIONE POLITECNICO DI MILANO	FPM
COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES	CEA
VDI TECHNOLOGIEZENTRUM GMBH	VDI TZI
BRAINPORT INDUSTRIES COOPERATIE UA	BPI
INDUSTRIE 4.0 OSTERREICH – DIE PLATTFORM FUR INTELLIGENTE PRODUKTION	PIA
CHALMERS TEKNISKA HOGSKOLA AB	CHALMERS
INTERNATIONAL DATA SPACES EV	IDSA
ENGINEERING - INGEGNERIA INFORMATICA SPA	ENG
UNPARALLEL INNOVATION LDA	UNPARALLEL
SOFTWARE QUALITY SYSTEMS SA	SQS
FIWARE FOUNDATION EV	FIWARE
IDC ITALIA SRL	IDC ITALIA
SIEMENS AKTIENGESELLSCHAFT	SIE

Executive Summary

The Smart Data Models (SDM) is the result of a joint collaboration between the FIWARE Foundation, TM Forum, IUDX and OASC to support the adoption of common data models for interoperable and replicable smart solutions in multiple sectors. The Smart Data models are published on the site https://smartdatamodels.org and in the public repositories located at GitHub at https://gitub.com/smart-data-models.

The main focus in SDM is to curate open licensed data models, including their structure, definitions, data types and semantic references based on real use cases. This initiative helps the participants publish a data source in a data space and enables convenient means to point to the definition of data models as well as to automatically check the conformance of specific data samples with a predefined data model. The participation is open as long as the best practices of agile standardization[1] are met. Finally, SDM provides tools to check if specific attributes belong to a data model, enabling it to easily implement part of the requirements of the block metadata and discovery services.

The Smart Data Models are organised into domains and data models are grouped into subjects within such domains. The smart manufacturing domain[2] is already active in SDM as well as some subjects which are populated with actual Industry 4.0 data models. Many others are coming such as the Agile Production and Asset Administration Shell[3] related data models. These emerging models can also be found publicly available under the incubated SDM repository.

The document identifies a core problem for data model standardisation in the manufacturing domain. Most of the manufacturers still do not engage in data standardisation. However, the market is evolving to new models in which companies that work in isolation or poorly connected their supply chains are turning less and less competitive. Therefore, some form of standardisation will eventually be established in one way or another to respond to market needs.

Relevant contributions in this direction include the standards listed in the main RAMI 4.0 axes (i.e., IEC 62264, IEC 61512, and IEC 62890), the standardisation activities carried out within the OPC UA, AutomationML, and ECLASS initiatives, or the more recent initiative, the AAS submodel templates. However, none of these specifications except for the submodel templates are 100% open or royalty free. The Smart Data Models initiative will face the necessary challenges to offer a fully open and free means for standardisation in the manufacturing domain which works constructively and promotes as much as possible the synergies not only between the aforementioned initiatives but also with others that may come.

Keywords: data spaces 4.0, smart manufacturing, smart data models, SDM, data interoperability,

[1] https://github.com/smart-data-models/data-models/blob/master/MANIFESTO.md
[2] https://github.com/smart-data-models/SmartManufacturing
[3] https://github.com/smart-data-models/dataModel.AAS/tree/master

Disclaimer

Acknowledgement

This document is a deliverable of the EU DATA SP4CE project. This project has received funding from the European Union's Digital Europe programme under grant agreement N° 101083939.

Table of Contents

Abbreviations and Acronyms

Acronym	Meaning
CA	Consortium Agreement
CPPS	Cyber-Physical Production System
DoA	Description of Action
EC	European Commission
GA	General Assembly
IPR	Intellectual Property Regulations
KPI	Key Performance Indicator
REI	Responsible Exploitation & Innovation Board
RRI	Responsible Research & Innovation
SDM	Smart Data Models
TCC	Technical Coordination Committee
WP	Work Package

List of Figures

1 Introduction to Smart Data Models

The FIWARE Foundation, TM Forum, IUDX and OASC are leading a joint collaboration initiative to support the adoption of a reference architecture and compatible common data models that underpin a digital market of interoperable and replicable smart solutions in multiple sectors, starting with Smart Cities. The Smart Data Models (SDM) is the result of this shared effort, published on the site https://smartdatamodels.org and in the repositories located at GitHub at https://gitub.com/smart-data-models.

The data models are grouped into subjects. These subjects are independent repositories of GitHub. Each subject is a submodule of a domain repository like presented below in Figure 1.

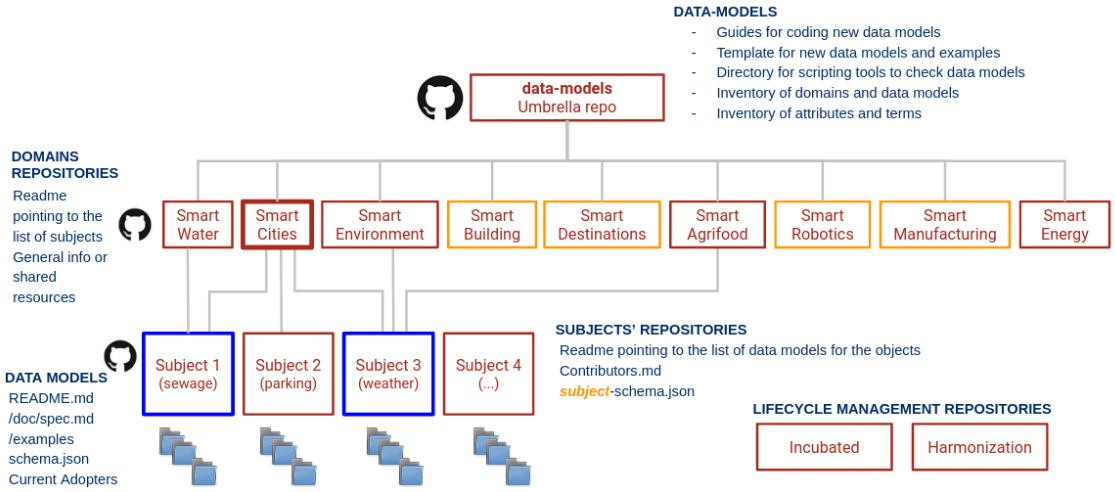


Figure 1. Organisation of data models in the Smart Data Models (SDM) program

What is in a Smart Data Model?

A smart data model includes three elements: First, the schema, or technical representation of the model defining the technical data types, their unique URI and structure. Second, the specification of a written document for human readers translated in 8 languages. And third, the examples of the payloads in several formats.

All data models are public and of royalty-free nature. They are free for all forever and their licensing mode grants 3 rights to the users:

- Free use

- Free modification (customization to local needs)
- Free sharing of the modifications

2 Relevance of SDM for the Data Spaces

The next figure presents the main elements of a data space. The left column of the schema details 3 blocks (light green). The first one addresses the data models & formats of the data to be interchanged in the data space.

Although there are competing solutions about how to define the metadata of a data source available at a data space, it is relevant to note that DCAT-AP is being accepted as a main candidate for the description of these data sources. However, this standard lacks a suitable way to describe the structure of a data source. In fact it relies on the attribute conforms. To that its value is a url pointing to the referenced model. This, in general, does not allow an automatic processing of the conformance with the standard. On the contrary, by using this attribute linking to a SDM schema, it is possible to point to the definition of the data model, a json schema, which is capable of automatically checking the conformance with the defined data model.

On the other hand, in order to enable a quick adoption of the data space, publication of data sources has to be a simple process imposing the minimal barriers to belong to it. Due to the fact that SDM provides a simple way to completely define a data model it is a strong candidate to provide this type of services.

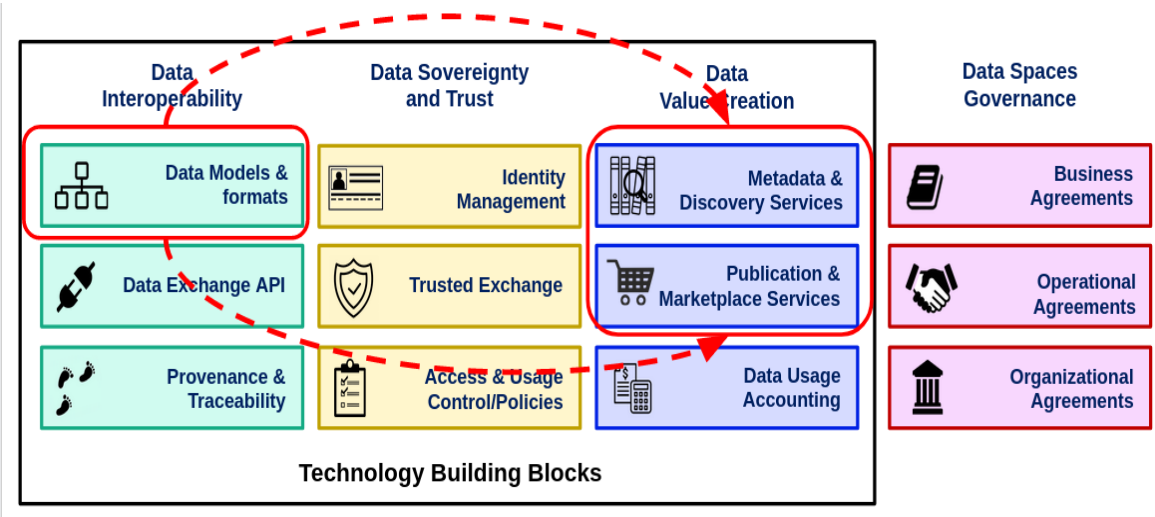


Figure 2. Main data space elements (technology and governance building blocks)

Finally, at SDM there are tools to check if specific attributes belong to a data model, enabling to easily implement part of the requirements of the block metadata and discovery services.

In fact there is a open sourced python package, pysmartdatamodels[4] with all these functions available and some exports in different formats.

[4] https://pypi.org/project/pysmartdatamodels/

Functions to align with data spaces architecture

Data models & formats

The SDM provides functions for:

- Checking the compliance of a payload from a data asset with a specific data model (external)[5]
- Creating a drafted data model from an example of the data asset in csv and json format[6]
- Mapping to alternative semantic references[7]

Metadata and discovery services

The SDM provides functions for:

- Populating the metadata describing the data asset[8] (direct link) and function 23 in pysmartdatamodels[9]

Publication and marketplace services

- Create samples of a data asset without providing the real data[10] and functions 17, 18 and 19 in pysmartdatamodels[11]
- Create a specific @context when the data asset come from different domains[12]

[5] https://www.jsonschemavalidator.net/
[6] https://smartdatamodels.org/index.php/draft-model-based-on-examples/
[7] https://smartdatamodels.org/index.php/generate-acontext-based-on-external-ontologies-iris/
[8] https://smartdatamodels.org/extra/datamodels_metadata.json
[9] https://pypi.org/project/pysmartdatamodels/#description
[10] https://smartdatamodels.org/index.php/generator-of-examples/
[11] https://pypi.org/project/pysmartdatamodels/#description
[12] https://smartdatamodels.org/index.php/generate-a-local-context-based-on-smart-data-models-iris/

3 Existing SDM in Industry 4.0

These are the actual subjects including data models related to industry 4.0. The next list are the official data models already published in December 2022.

- OPC-UA: Data models related to the integration with OPC UA standards. Located at https://github.com/smart-data-models/dataModel.OPCUA. Include these data models:
    - MotionDeviceSystem
    - WoodworkingMachine
- Manufacturing Machine: Data models related to machinery for manufacturing. located https://github.com/smart-data-models/dataModel.ManufacturingMachine. It includes these data models
    - ManufacturingMachine
    - ManufacturingMachineModel
    - ManufacturingMachineOperation
- Data models related to Autonomous Mobile Robots. Located at https://github.com/smart-data-models/dataModel.AutonomousMobileRobot. It includes these data models:
    - CommandMessage
    - CommandReturnMessage
    - StateMessage
    - StopCommandMessage
    - StopCommandReturnMessage
- Data models related to the use of robots for industrial activities. Located at https://github.com/smart-data-models/dataModel.RoboticIndustrialActivities. Include these data models:

- Pallet
- Piece
- RobotArm
- RoboticCell
- VacuumPump

- Data models related to the standard Asset Administration Shell. Located at https://github.com/smart-data-models/dataModel.AAS/tree/master Include these data models:
  - I4AAS
  - I4Asset
  - I4Submodel
  - I4SubmodelElementOperation
  - I4SubmodelElementProperty

Additionally, there are some other models in the incubated repository. This repository is for the development of data models and review before their final approval. Find here a list of the potential ones, but it could happen that some of them would not be finally approved as official data models. all of them are located in different folders of the incubated repository (https://github.com/smart-data-models/incubated)

- 3d_reconstruction
- ActiveProduct
- Asset
- CellStatus
- dummySensor
- GeneralConfig
- HamInspectionBuffer
- I4AAS
- I4Asset
- I4Submodel
- I4SubmodelElementOperation
- I4SubmodelElementProperty
- KUKA robot

- LotInspectionBuffer
- Material
- OEEMetric
- OPCUAfromCobot
- Order
- Process
- Product
- ProductionMeasurements
- ProductionStatus
- Products
- Realtime
- Recipe
- Robot
- robot_control
- Sensor
- SensorStatus
- SortingResultBuffer
- Station
- tracking
- Warehouse
- WorkOrder
- WorkOrderItem
- WorkStation

4 How to extend/define new SDMs

There is a defined workflow for the definition of new data models or for the extension of new ones. The left part of the organigram corresponds to the creation of new data models and the right part to the extension of an existent data model. The first requirement to be assessed is whether the candidate is coming from an actual use case. (A system that is being used in real systems and not for theoretical purposes). If this question is passed the next question is to identify if this data model is already available at any open and adopted standard. This was the case then the open and adopted standard is preferred and the data model is requested to adapt to it. Finally, if this was not the case, the candidate can document the data model according to the contribution manual. The right part of the schema shows the workflow for extending an existing model (already officially published) and it only requires the demonstration of the real use and that the contribution can be valuable for use cases beyond the contributor.

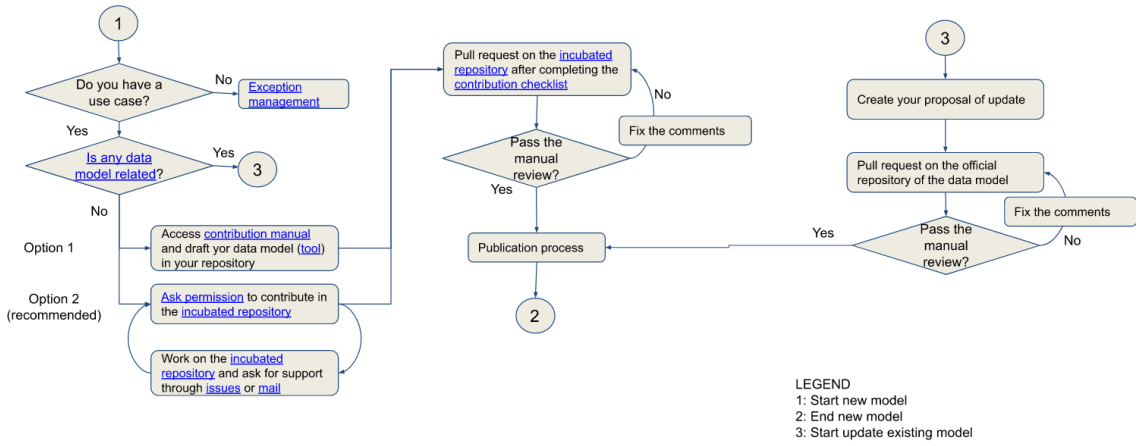


Figure 3. workflow for the definition/extension of data models in the SDM program

The usual way is to work on a public repository of code. SDM offers the incubated repository for this development granting the required permissions to the users. It is also possible to work on his/her own repository. Then it should warn the organisation when the data model is ready, the incubated repository is monitored regularly and this warning is not necessary. Some tools[13] help the contributors to check that the data model is properly documented and that the examples comply with the defined data model.

[13] https://smartdatamodels.org/index.php/data-models-contribution-api/

Finally after a manual review the data model is approved. It implies the publication of the specifications in 8 languages, the extension of the subject's @context, the population of the data models database with the new attributes, data types and descriptions, the creation of additional examples and the update of the README of the data models and the subject.

Likewise when there is a request for the extension of a data model, then the contributions are included in the official repository but the review process is the same and the republication of the specifications is also performed automatically.

5 Challenges for SDM in manufacturing

The scope of data model standardisation in the smart manufacturing domain covers every single dimension of data streams which aim at representing an advance in production means and practices. It is essential to think about standardising data streams at all levels, starting from interconnected machines and systems, through horizontal shop floor and vertical site operations, to external supply chain data exchanges.

The core problem in the manufacturing domain is that a great part of the manufacturers do not engage in data standardisation, and still attempt to manually record and analyse data, using customised solutions to map their data to a unified (but highly tailored) model. The practices associated with this approach have led to complex and hardly replicable data integrations which aim to enable the communication and integration between hundreds/thousands of devices and quite diverse vertical software (CAD/CAM, PLM, MES, etc.) to monitor, automate, and secure production chains in very specific contexts.

Despite this reluctance of manufacturers to engage with standardisation, there is a growing need for openness and agility in the manufacturing domain to remain competitive in the current production markets. As the global competition intensifies, the self-sustaining models are falling down and manufacturers need to evolve the data integration approaches that were designed to operate in closed circuits. Thus, the next challenge is how to cope with the wide disparity in the vast amount of data that such manufacturing systems already produce on a daily basis. There are huge investments behind the existing data integrations, most of them are still in competition and/or not not natively designed to communicate with one another. There is not yet an approach to smart manufacturing that is clear and proven effective enough to contribute a general solution that unifies the immense manufacturing ecosystem.

Some sectors, such as the aerospace industry, faced these problems in the past and developed sector-based exchange platforms and data models. In the smart manufacturing sector, standards such as the ones listed in the main RAMI4.0 axes (i.e., IEC 62264, IEC 61512, and IEC 62890) provide a good basis for such a unified model. Initiatives like the OPC UA, AutomationML, or ECLASS are lowering many barriers and contributing to this need for standardisation in the sector. However, none of these specifications are 100% open or royalty free. A more recent initiative, the AAS submodel templates seem to be more in line with the

openness principle. The Smart Data Models initiative will face the necessary challenges to offer a fully open and free means for standardisation in the manufacturing domain which works constructively and promotes as much as possible the synergies not only between the aforementioned initiatives but also with others that may come.

6 Approach for different stakeholders

Associations of the manufacturing domain

The Smart Data Models initiative offers significant advantages for standardisation bodies in the manufacturing sector, as well as for relevant associations. First, it provides open-licensed, free data models, it facilitates interoperability across systems and organisations, allowing seamless data exchange based on actual use cases and market-adopted standards. These models are customizable, enabling adaptation to local or sector-specific requirements, making them ideal for digital twins and data spaces. Additionally, the initiative supports agile standardisation[14], significantly reducing the time required for creating and updating standards, which aligns well with the fast-paced digital landscape. And due to the open licence approach they can be the sources for further standardisation initiatives of the standardisation bodies, with the advantage of the real cases scenarios validating the results.

Furthermore, Smart Data Models is a collaborative effort involving various organisations and contributors where organisations of the sector can be members of the board[15], as long as they are independent of any vendor, and public or non-profit. Additionally existing standards are mapped once they are adopted providing additional dissemination. OPCUA[16] and the Asset Administration Shell (AAS)[17] for manufacturing are examples of this. This alignment with market needs ensures that standardisation bodies can rely on proven examples.

General public of the manufacturing sector

The Smart Data Models initiative offers valuable benefits for company workers and managers in the manufacturing sector by providing ready-to-use, open-licensed data models that enhance data management and system interoperability. On the contrary to many vendors in the domain these models enable seamless integration of information across different domains[18], tools and platforms, allowing for smoother operations and better decision-

[14] https://github.com/smart-data-models/data-models/blob/master/MANIFESTO.md
[15] https://smartdatamodels.org/index.php/board-member-application/
[16] https://github.com/smart-data-models/dataModel.OPCUA
[17] https://github.com/smart-data-models/dataModel.AAS
[18] https://smartdatamodels.org/index.php/list-of-data-models-3/

making. The models cover not only manufacturing but also other domains like energy, environment, robotics, etc. Besides, the open licensing allows internal customization without the need of asking for additional licensing or disclosing internal knowledge of the company. The smart data models are either based on real use cases, or on open and adopted standards. As the initiative follows an agile approach[19], updates and improvements to these models can be quickly adopted, keeping companies at the forefront of innovation while reducing operational delays. documentation in 8 languages (EN, FR, DE, SP, IT, KO, CH, JP) also helps to the adoption and several examples in different formats also helps to adopt the models.

National technology platforms

The Smart Data Models initiative provides a robust set of data models for national technological platforms promoting technology adoption in the industrial and manufacturing sectors. By offering free and open-licensed data models, it facilitates the integration of standardised digital tools across industries, ensuring interoperability and scalability. Its agile standardisation process allows rapid adaptation to evolving market demands, ensuring that manufacturing ecosystems can leverage state-of-the-art tools for automation, data-driven decision-making, and sustainability and, to some extent, avoid the vendor lock-in.

These models are based on real-world use cases and aligned with widely adopted standards, such as OPCUA and the Asset Administration Shell (AAS), enabling organisations to streamline operations, foster innovation, and support the creation of digital twins and data spaces.

For national platforms focused on advancing industry 4.0, Smart Data Models accelerates the adoption of cutting-edge technologies while reducing the time and cost involved in developing custom solutions.

DSBA members and data spaces related entities

The smart data models cover the needs of the building block of data models in a Data space. Besides this due to the lack of a proper standardisation in a large percentage of the data to be

[19] https://github.com/smart-data-models/data-models/blob/master/MANIFESTO.md

shared in data spaces, some solutions of the initiative can help to cope with that. The services for the generation of a draft data model, based on examples, and profiting for more than 156.000 already existing attributes can ease the documentation of the structure of a data asset to be shared across data spaces. This also can be integrated into any system because the functionality is also integrated in the python package pysmartdatamodels[20] that it is released with an open licence. Additionally all datasets have their metadata available[21] in a structured way, all at the same time or individually also available in the python package.

[20] https://pypi.org/project/pysmartdatamodels/
[21] https://smartdatamodels.org/extra/datamodels_metadata.json

7 Recommendations for the Data Space deployment

The high level architecture of a data space was already presented in Section 2 (Figure 2). Building on this view, main recommendations for the standardisation of the data formats to allow a simple interchange of data include:

- Provide automatically detailed structure of the data to be interchanged. Otherwise, a manual interaction would be required and therefore it would limit the capacity and performance of the data interchange. Eventually, this structure description could also include the definitions. This point is addressed at DCAT-AP standard for metadata in the conforms To attribute, but the precise information to be linked in this attribute is not defined. We suggest linking directly to the json schema of a data model so both information, the description and the technical structure would be available automatically.
- Allow an automatic configuration of the metadata of the data sources to be interchanged in the data space. Provide API or other automatic mechanism to allow the metadata to be filled automatically (and updated as well)
- Unique value for the licence. The licence of a data source has to be explicitly and unequivocally determined. The standard STAT-DCAT_AP provides an extensive list with unique identifiers for multiple licences.. Promote its use and its extension with new licences if necessary.
- Document or draft a model. Description of the information to be published in a data space requires the description of the data assets. The SDM provides a feature[22] that, based on one example (json or csv), drafts a data model based on the available attribute's descriptions (> 156.000)

[22] https://smartdatamodels.org/index.php/draft-model-based-on-examples/