ce/papers Proceedings in civil engineering

Ernst & Sohn A Wiley Brand

ORIGINAL ARTICLE

EUROSTRUCT

Importance of Digitalization and Standardization for Bridge and Tunnel Monitoring and Predictive Maintenance

Matthias Weise[1] | Michel Böhms[2] | Diego Allaix[2] | Ana Sánchez-Rodríguez[3] | Mattia Rigotti[4]

Correspondence

Dr.-Ing. Matthias Weise
AEC3 Deutschland GmbH
Archivstr. 21
01097 Dresden
Email: mw@aec3.de

[1] AEC3, Dresden, Germany
[2] TNO, Delft, The Netherlands
[3] ICITECH, Valencia, Spain
[4] IBM Research, Zurich, Switzerland

Abstract

The Horizon 2020 IM-SAFE project aims to provide input for a mandate for further CEN standardization towards the use of inspection, monitoring and testing for the safety assessment of bridges and tunnels and the implementation of predictive maintenance strategies. Although the focus of this project is on the engineering domain, the use of digitalization is a necessity for the effective and efficient implementation of the proposed work- and dataflows. This paper shows what topics related to digitalization are of particular interest, explains the state-of-the-art in those areas and discusses the need and readiness for further ICT standardization. In particular, it focuses on the importance of data interoperability and emphasizes the challenge to consolidate, reuse and harmonize existing standards from BIM, GIS and IoT, rather than developing new standards.

Keywords

BIM, GIS, IoT, interoperability, semantic web, monitoring and predictive maintenance.

1 Introduction

The transport infrastructure in Europe is currently affected by ageing, increasing growth of traffic loads, natural hazards and deficiencies of the maintenance practices. European countries are therefore facing a huge challenge to keep the transport infrastructure networks in good conditions and to avoid accidents like the tragic collapse of the Morandi bridge in Genoa in 2018. This requires huge investments in the coming years, if not decades. Thus, it is important to develop new strategies to improve asset management of existing infrastructure, meaning to increase their quality while also looking into potential cost savings.

The IM-SAFE project provides the technical input for the preparation of a mandate for CEN standardization aiming to improve the assessment of structural performance and the proactive maintenance practices for the European transport infrastructure by making optimal use of information from inspection, monitoring and testing. More details about those proposals are reported in Allaix et al. 2023 [1] and other partner publications at the EUROSTRUCT conference. This paper is dealing with the importance of digitalization, being seen as a necessary complementary effort for further standardization as proposed within the IM-SAFE project.

1.1 Monitoring and predictive maintenance of bridges and tunnels

The IM-SAFE project provides the technical input for the following future European standards:

- <u>new standard on structural monitoring</u>, addressing the principles of setting the objectives of structural monitoring, the principles and requirements for the design of monitoring systems and methodologies used for translating data into useful and meaningful information relevant for diagnostics of structures, safety assessment and maintenance approaches,
- <u>further amendment to the existing EU standards on safety assessment taking into account inspections, monitoring and testing</u>, focussing on approaches to integrate diagnostics of structures based on data from inspection, monitoring and testing with the evaluation of the structural condition and the assessment of the structural performance,
- <u>new standard on proactive maintenance of the transport infrastructure</u>, promoting the transition from corrective towards preventive and condition-based maintenance strategies and the implementation in the long term of risk-based maintenance management of infrastructure assets.

The proposals that have been developed in this project heavily rely on high quality, up-to-date digital data about existing assets. Such databases, which potentially include massive amounts of measurements, need to be combined with other data sources, evaluated regarding structural health, and shared with partners to agree on (predictive)

https://doi.org/10.1002/cepa.2144

maintenance activities. Accordingly, digitalization plays an important role in the implementation of the proposed processes, which, to meet the goal of efficiency and effectiveness, must be automated to a significant degree.

1.2 Need for complementary standardization on digitalization

The overall process targeted by the IM-SAFE project starts from onsite measurements and goes via all kinds of quality control and data analysis to final decisions about required maintenance. It therefore involves collaboration of different stakeholders and requires seamless data exchange between different tools. However, a high degree of automation requires highly integrated systems. Today, setting up such solutions requires a lot of experience and effort, and tends to lead to an inflexible ad-hoc solution.

In order to improve the current situation, aiming to a more flexible, configurable solution, standardization of data interfaces and services are needed. However, not everything needs to be standardized and a combination between proprietary solutions with the right level of open interfaces yet needs to be agreed.

2 Digitalization in monitoring and predictive maintenance

A work package of the IM-SAFE project was dedicated to an analysis on the state-of-the-art of digitalization, in particular to recommend further steps in complementary standardization. Four essential questions have been addressed in that survey:

1. How to get correct and up-to-date asset measurements?
2. How to efficiently and effectively use the amount of digital data to support decision making?
3. What common data language are needed to support collaboration?
4. How to provide secure and reliable access to distributed data?

Digitalization permeates nearly all areas of modern business cases, in particular where a huge amount of data needs to be handled. Beside looking into the availability and use of Information and Communication Technologies (ICT), the analysis also aimed to evaluate the timeliness for further standardization.

2.1 Data acquisition, processing, and quality assurance

One of the main issues in Architecture, Engineering and Construction (AEC) is the collection and exploitation of meaningful data throughout the lifecycle of infrastructure assets. Before the analysis and decision-making process can take place (see chapter 2.2), the data needs to be collected, filtered, formatted, and stored in secure ways.

The analysis defines in detail each of the 19 surveying technologies studied within the IM-SAFE project. These deal with both bridges and tunnels, located in roads and railways. Since the performance of these huge structures results from complex relationships between different systems and elements, specific monitoring systems are used to explore each of their aspects in their current state. These systems differ greatly from one another, with some capturing the overall surface geometry of the element (e.g. LiDAR), while others inspect the interior of the asset in search for defects (e.g. GPR). They are tailored for their specific task, meaning that their software must operate within the parameters of the hardware required for their measurement. Furthermore, they often lack standardization of their own, proving unification of data collection under a same standard extremely challenging.

Following the data collection, the techniques applied to data are entirely dependent on the data itself, as they come from different systems with different purposes. The nature of the data obtained with one technology in a survey, as well as their format, might be completely different from those acquired with another technology, or even the same one, in a different survey. The analysis details the pre-processing and correction methodologies that are available for each surveying technology. These methodologies, however, are applied depending on the results of the data collection because of their quality or quantity.

Like data collection, these processes do not have standards on their own or as a whole. However, in this case, these shortcomings are further compounded by the lack of standardization in the data collection stage upon which it is building. Nevertheless, the analysis provides general guidelines for safe and efficient data monitoring, enumerating the characteristics that the tool must possess, such as working automatically and unattended.

Conclusions and recommendations

As shown in the analysis, there are different surveying technologies that are applicable to different scenarios or purposes. These technologies are completely unique to their role and are not comparable to one another since they are specially designed for their task, both hardware and software-wise.

Furthermore, the processing of the data captured by those systems is equally diverse depending on the interest behind the following analysis. Even the case where the same equipment is used in different surveys might require different processing techniques that are adequate for the location, conditions, and characteristics of the survey that retrieved the data. In the same manner, the data can be presented in a wide variety of formats.

Current technologies are dominated by proprietary solutions, whose variety and heterogeneity make standardization difficult at this point in time. However, it is recommended to prepare further standardization activities by proposing a common ground data format and structure for monitoring data. A starting point for better comparison of available technologies can be agreements about:

– Key characteristics about used surveying technologies as used in the IM-SAFE reports to describe the various technologies,
– Meta-data about the measurements, monitoring system and data processing to be used for documentation purposes.

It is suggested that such agreements can be documented in a separate knowledge base to be encoded as an ontology and/or dictionary. Further information can be found in Sánchez Rodríguez et al. 2022 [3].

2.2 Data analytics with Artificial Intelligence

Artificial Intelligence (AI) and Machine Learning (ML) have received a lot of attention lately. It is expected that they also provide value and capabilities in the civil engineering and construction domains. Such solutions are required to support users in dealing with the amount of data produced by deployed monitoring technologies creating a continuous stream of measurement data, possibly including images or other types of data requiring a rather high processing power. Further examples are the use of more intelligent algorithms for completing planning and construction, smart asset tracking, predictive maintenance (identify damages together with possible repair strategies), as well as for aiding in mitigating risks and ensuring safety.

Deployment of AI, with focus on ML, is a challenging task that is based on a set of recurring tasks, that also require continuous calibration. Main tasks are: (1) data collection, (2) data preparation (both already addressed in chapter 2.1), (3) model building and (4) model deployment. There are a couple of general findings:

- Iterative refinement: The data analytics and ML infrastructure must encompass a continuous deployment approach that considers the iterative process going from data collection to model building and deployment.
- Quality of data is crucial: The data analytics and ML infrastructure must focus on data curation and linking data, as a single untreated source of data are typically not enough to provide the necessary knowledge to train useful AI models.
- Multicriteria for selecting the right approach: Choice of AI models and their development must be driven by multiple dimensions beyond accuracy, such as explainability, training time and cost, robustness. Key metrics need to be quantified along all these dimensions.
- Model monitoring is essential post-deployment to ensure its continuous deployment.
- Fairness and explainability of models ensure trustworthiness among domain experts (like civil engineers) when applied in real-world use cases.
- Proper documentation and understanding of use cases and data is essential to guide data preparation, as well as model building and deployment.

Conclusions and recommendations

It is expected that AI will play an important role in data evaluation. However, they are still young technologies, and people are currently learning how to integrate them in our monitoring use cases in an efficient, reliable, and transparent manner. Accordingly, documentation of experiences for further evaluation and reuse in other projects should be a key aspect when applying this technology.

2.3 Semantic integration of IoT, BIM and GIS

Being able to share digitally available data within a project team is a crucial prerequisite for efficient collaboration. Accordingly, standards about shared data structures and semantic integration are a necessary basis for optimizing monitoring and maintenance processes. So far, main areas for development of open standards have been:

- Data about measurements via IoT,
- Data about the building/infrastructure asset/structure itself (BIM),
- Data about the integration of the building/infra into an infrastructure network (GIS).

Historically, developments in those areas have been independent from each other and along with those standards' further agreements, partially a standard of its own, have been developed to organize data sharing processes and to support quality control.

The scope of our analysis has been on existing and ongoing standardization efforts tailored into above mentioned main areas including latest developments in linking/combining existing standards using Semantic Web technology. For such analysis it is crucial to properly understand the business cases as they define what data needs to be shared in collaborative workflows.

Proper documentation of business cases thus become a crucial necessity to proceed with standards for semantic data integration. Such workflow documentation, yet to be agreed and standardized, should follow existing standards like ISO 29481 (IDM) and can be published in a central place, like the Use Case Management system from buildingSMART.

A clear message and result of our analysis is that existing standards must be reused and combined to take advantage of current achievements. Emerging and rapidly developing Digital Twin concepts show how integration of different data sources and up-to-date knowledge about the status of the building could be implemented and used [13]. It is therefore recommended to focus on solutions being able to reuse and integrate existing standards, rather than developing highly customized new standards. Such harmonisation remains a challenge of the whole industry and should be implemented using same technical solutions, ideally based on Semantic Web technology and model linking solutions.

Conclusions and recommendations

Data integration is a complex topic that needs a roadmap for step-by-step implementation. Main driver will be the business cases and the technical set-ups used in pilot projects that are going to be accepted by the industry as state-of-the art.

When it comes to semantic integration, important standards have already been developed and provide a sound basis for support and further development of monitoring and predictive maintenance use cases. However, not all use cases are sufficiently supported yet. Further standardization work is therefore suggested in following areas ordered by priority:

- Documentation of the condition and degradation of the

building and its parts by using an agreed damage ontology.
- Documentation of monitoring systems and measurement set-ups, including for instance the network of installed sensors, flight path for UAVs, knowledge about the meaning of measurements, etc.
- Product specification sheets for sensors and other equipment for instance with information about necessary maintenance, range of valid measurements, lifespan of sensors etc.
- Documentation of maintenance activities that can help to make further decisions. This should include changes during lifecycle of the asset, hazard events, etc. thus creating a Digital Building Logbook (DBL) or passport.

A lot of data can already be stored in IFC (ISO 16739) and other data structures, even with the possibility of using additional bilateral agreements about proxy elements and user defined properties. Accordingly, existing standards could be used already today in well-defined scenarios, but cutting-edge state-of-the art technologies will likely need further integration efforts to be properly implemented within the available family of data sharing standards. It is therefore suggested, to start with most urgent use cases to be tested and documented in pilot projects in order to consolidate requirements and to derive further requests for standardization.

2.4 IT platforms for monitoring data of transport infrastructure

IT data platforms provide an essential functionality when using ICT in shared environments. While it is not the aim to standardize a software system, it is important to understand the main features of available solutions and what becomes relevant for monitoring and maintenance use cases.

Secure, reliable and extendable data storage and management is a topic being addressed by a couple of domain independent initiatives for digital platforms such as DigiPLACE RAF, GAIA-X/IDS or FIWARE, also introducing a proposal for a conceptual data architecture for asset management dealing with specific requirements such as "Network level" interconnection between individual assets, entire lifecycle support or Level of Detail (LoD) definitions.

Important general issues like data security, dealing with confidentiality, integrity, and availability of data need to be addressed specifically in ways that meet our domain requirements, meaning for instance to agree on access control settings (who can see/modify what data), logging of data access, versioning, rollback and backup strategies etc. A more universal agreement for monitoring and predictive maintenance use cases could be the identification of roles and basic level of access rights settings.

Conclusions and recommendations

Key factors for the design of a software platform for data management in construction sector are to focus on data, function and usability for our domain. The main recommendations are:

- Employ user-centric design practices in designing and developing tools and platforms for standardizing the monitoring and managing of data, to make sure to address the needs expressed by the IM-SAFE Community of Practice and accommodate consolidated workflows and already established standards.
- Rely on established standards and best practices in data security and data governance, preferably by leveraging open standards.
- Leverage abstraction layers to automate resource allocation.
- Design modular and extensible platforms to favour customization and adaptability to varying and various user needs (see e.g. Bartezzaghi et al. [14]).

3 Situation in ICT standardization

3.1 Introduction

In general, the scope of 'standardization' covers two areas:

- *Real things:* assets, products, materials, sensors/actuators).
- *Abstract things:* processes, methodologies, methods, data about real things or coming from real things, software, both Commercial Off-The-Shelf and Open-Source Software (OSS) or combinations in the form of digital twins (data + software + sensors & actuators).

The focus in this paper is on infra-assets & sensors and the data about/coming from them. This data can range from being totally static ('time-invariant') or highly dynamic (like changing a thousand times per second). It can be data on the asset itself or on its environment in which it operates (like weather and traffic loads).

One of the main problems with standardization in this area is the fact that the use of proprietary data formats, which have led to incompatible isolated solutions and should be replaced by new standards, have now themselves led to data standardization chaos with competing or overlapping standards.

In the remainder of this paper, key relevant standards are identified, and recommendations are given on how to deal with them. First, a data architecture is presented, showing what data aspects needs standardization in the first place.

3.2 Data architecture

ISO 8000 Part 110 ('Master data — Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification') [8] defines a data architecture for the 'exchange of data without loss of content and meaning'. This data architecture identifies the following data aspects:

(Master) data set is a kind of reference/project-independent data bound to a specific stakeholder. The architecture can however be applied to 'data sets in general', so even, data sets relevant for multiple stakeholders in the building sector, covering both internal master data and internal/external project data. Data sets can refer to data resources in any form like databases and data files.

A *data specification* that defines all the possibilities and/or impossibilities for the content in the (master) data sets.

This way it defines the meaning or 'semantics' of the data. Such specification typically takes the form of a graphical diagram or a lexical schema or ontology. This specification typically defines concepts/classes, attributes, datatypes, relations and constraints/rules that should hold for the data.

Said otherwise: a data specification is used to correctly interpret the data. The backbone of any ontology is often formed by two hierarchies of concepts ('classes'): a specialization hierarchy (also known as a 'taxonomy') and a typical decomposition hierarchy (also known as a 'meronomy'). The meronomy should not be confused with an actual decomposition of the instances of concepts/classes, often called a System Breakdown Structure (SBS) or "object tree".

A *formal syntax* or 'format' or in 'linked data'-speak, a 'serialisation'. The format can be abstracted by defining a functional direct access method like a (low-level, schema/ontology-agnostic) Application Programming Interface (API) or more flexibly, a Query Language (QL). Often predefined queries in some QL can be grouped into a more static 'snapshot' API that can be regarded as a schema/ontology-aware high-level API as alternative for the schema/ontology itself. The same content of a data set in a different format is often referred to as a 'distribution'.

A *data dictionary* that defines all the terms used as names or labels for the concepts, attributes and relations used in the data specification. Often a data specification and a data dictionary are integrated into one specification. Dictionaries can be multi-lingual and also indicate synonyms and homonyms for terms. A data dictionary and an ontology can be interrelated in a standard way.

An *identification scheme* is needed for all elements in the master data, data specifications and data dictionaries. The actual IDs following this scheme can be distributed over the various stakeholders/geographic levels. IDs can take the form of meaningful names (for humans) or meaningless codes like the generated Universally Unique IDs (UUIDs) [9]. UUIDs are unique by themselves but the other way round, the buildings, their parts or aggregates can still have multiple UUIDs. One of them that is unique regarding the asset/part/aggregate is often referred to as a Unique Object Identifier (UOI).

A *data specification/dictionary language* is needed to provide the language constructs to define a data (or meta-data) specification and/or a data (or meta-data) dictionary. Typically, such a language contains elements like 'term' and 'definition' (for the data dictionary) and 'concept', 'attribute', 'datatype', 'relationship' and 'constraint' (for the data specification).

In a simplified way one could state that there are underlying "technologies" (formats, access methods and languages) and "specifications" (data specifications, dictionaries and identification schemes), making use of those underlying technologies.

Note that the focus is here on the syntax and semantic aspects of data. In general, a layer on top and below have to be added as shown in the 'bigger' picture below (Figure 1).

- **WHY?** Pragmatics, Purpose, Use Case
- **WHAT?** Specifications ('semantics')
- **HOW?** Technologies (syntax, language)
- **WHERE?** The medium: 'federated data spaces'

Figure 1 Adding 'why?' and 'where?'' questions.

All these data aspects are a prime candidate for standardization. For many of these aspects there are already many alternatives and typically incompatible standards available. In the next chapter the most relevant ones will be introduced for our scope from the various standardization fora/sources like ISO, CEN, W3C, bSI, OGC, OMG, IETF.

3.3 Overview standardization fora and related existing standards

Besides the overview presented in Table 1 there are many other organizations and initiatives involved in more technical standards or standards more related to software aspects:

- Internet Engineering Task Force (IETF) with basic standards like Uniform Resource Identifiers (URIs) and Universally Unique Identifiers (UUIDs)
- GAIA-X/International Data Spaces Association (IDSA)
- Linux Foundation (including OAI: Open API Initiate and the GraphQL Foundation (producing GraphQL and GraphQL Schema)
- The international industry association for ICT (ECMA)

There are also a lot of technologies that are not 'standard" but just used very often, like for instance Microsoft Excel (.xlsx files), Autodesk Revit (.rvt files), Adobe Portable Document Format (.pdf files) etc.

All these identified fora have technical committees, task groups, working groups, community groups or even joint technical committees producing many formal/industry standards/guidelines for the various data aspects: both standard technologies & standard specifications, sometimes complementary, often overlapping or competing.

Typically, "worldwide" coverage is preferred because there is a great need for focus and critical mass, rather than fragmentation, on both the user side (international users/usage) and the vendor side (international software vendors) in order to invest in implementation and use by the industry.

A good example of the implementation of an international standard is CEN SML [10]. This standard, started as a National (NL) standard (NEN2660), is now in the process of becoming a European CEN standard where it might evolve further into a world-wide ISO standard in future. Luckily, many relevant data standards are already world-wide (those developed by bSI, OGC and W3C).

Below, a selection of data standards is shown that are most important including standards from Open BIM (ISO/CEN with liaison to OGC) and Open GIS (ISO/CEN with liaison to bSI).

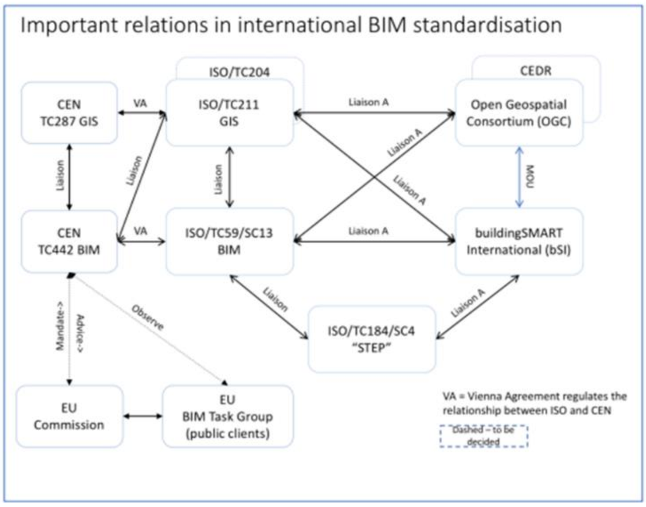

Figure 2 Key fora for data standardization and their links [15]

Within this picture, there is a key role of CEN TC442 and buildingSMART International (bSI) to come to a common approach on how to model data (syntax and semantics). The generic W3C Linked Data approach as defined by CEN SML could play an integrating role here, hopefully complementing the perceived inadequacies of the existing construction-specific ISO 12006-3 (IFD).

Table 1 Most relevant European and international standardization fora and standards in context of data digitalization.

Forum	Description	Standards
ISO/IEC	International standardization Organization also covering BIM and related standards. IEC being Electrotechnical area.	ISO 19650 ISO 12006-3 (IFD) ISO 23386 & 23387 ISO 29481 (IDM) ISO 10303 STEP incl. SPFF, EXPRESS(-G) JavaScript Object Notation (JSON) ISO 21597 Information Container for linked Data delivery (ICDD)
CEN/CENELEC	European Committee for Standardization. CENELEC being the electrotechnical area.	EN 17412 (LOIN) EN 17632-1 (SML)
EC	European Commission.	INSPIRE specification in UML/XSD
bSI	BuildingSmart International. A non-profit organization defining standards for BIM-based working, bSI standards are also published as ISO, CEN and national standards.	IFC (ISO 16739), bSDD, BCF, mvdXML, ifcXML, ifcOWL, ifcJSON, IDS Use Case Management (UCM)
OGC	Open Geospatial Consortium, International voluntary consensus standards organisation for geospatial content and services.	(City)GML, CityJSON, WFS/WMS GeoSPARQL in RDFS (together with W3C)
W3C	World Wide Web Consortium. Standardization body for all internet-related standards including semantic-web and model-linking approaches.	XML/XSD RDF/RDFS/OWL/RDF SHACL, SHACL-AF SKOS RDF-XML, Turtle, JSON-LD SPARQL Time, wgs84_pos (GPS), LOCN, SSN/SOSA, DCAT
OMG	Object Management Group	BPMN, UML, MOF, XMI, SysML
gbXML Schema	green building XML. Industry driven standard for energy-efficiency	gbXML

3.4 Recommended data standardization work

There is an enormous potential in the W3C linked data/semantic web technology. A drawback is its complexity and slow uptake by software developers. Recently however the semantic and software worlds have come closer by the development of the common syntax form JSON-LD. It is in fact 100% JSON (that is liked so much by web developers) but fully web-based and usable for the serialisation of semantic data and their controlling ontologies and dictionaries.

Independent of such a format is the query language GraphQL that could be used as an alternative or frontend to the sometimes hard to use SPARQL. Note that the graphs are exchanged for tree making things easier but not necessarily better.

On the language side SHACL as closed world option is a powerful way to specify your data including, in the near future, standard derivation and validation rules (involving SHACL-AF (AF from Advanced Features) that uses the procedural power of SPARQL construct queries. Such a choice could also settle the earlier mentioned bSI/CEN TC442 one (generic) language issue.

On the modelling side itself there is work to be done to better harmonize semantics:

- From BIM and GIS (the latest versions of IFC/bSDD and GML/CityGML).
- Reuse of W3C semantic resources (SML Top Level Model, FOAF, QUDT, DCAT), but especially SSN/SOSA [11] for monitoring and control data.
- Reuse of existing EC directive's modelling efforts like INSPIRE [12].

This way compatible semantics is obtained forming the future fundament for a broad semantic data eco-system.

4 Conclusions and next actions

The paper briefly shows the plethora of topics that need to be considered when dealing with digitalization supporting monitoring and predictive maintenance use cases. While there is clearly a need for further standardization, it is still however yet too early to start with a specific ICT standardization activity. There are several reasons for such kind of overall conclusion:

(1) Existing standards provide a sound basis to get started, in particular with semantic integration and basic data sharing scenarios. Instead of starting yet another standardization effort, industry first must implement those standards and users need to get familiar with its use in daily practice. There is a critical mass of tools, users and interoperable digital data about our assets before starting to extend available standards based on well documented pilot projects.

(2) Existing standards like IFC provide some level of flexibility to adopt its actual usage to specific needs. Thus, major work that needs to be done before starting further standardization is to settle on the most important use cases in form of process specifications including an agreement on data exchange requirements. For this, standards

being already available for generic process specification should be used, also to work on harmonisation with other activities of the AEC industry.

(3) Available technology related to data acquisition and data processing are still rapidly evolving, in particular when it comes to the use of Artificial Intelligence. Those developments need to settle down and must reach some level of consolidation, common agreement and ideally consistent adoption by the industry before going ahead with further standardization. In order to properly describe and classify such technologies a commonly agreed vocabulary linked with the targeted business cases and related standardization not only for required information about the asset but also for types of data acquisition and processing is suggested.

From all covered topics semantic data integration and data sharing should get highest priority to prepare further standardization, also following a road map derived from business cases. All developments and proposals should be guided by the FAIR principle (where data is Findable, Accessible, Interoperable and Reusable, [7]) and should try to reuse and combine existing standards. This is seen as a major challenge not only for this development but for the whole AEC industry and, thinking about universal data protocols, shared data spaces and linking technologies, even beyond our AEC domain.

Today, the situation in standardization related to semantic data integration in the AEC industry can be characterized by following conclusions:

- There are (already) too many data standards (existing or in development) drafted by the various fora.
- This causes both technical and conceptual fragmentation and incompatible solutions that can only be partially solved by data transformation or linking hampering critical mass for end-users and software vendors implementing the standards. No critical mass means no willingness to invest. Resulting in not solving the data chaos keeping incompatible data sets in isolated silos.
- Therefore, there is an urgency to decide on the governance of the alignment of key data standardization going on in ISO TC59/CEN TC442, bSI, OGC and W3C related to the built environment.
- This means, decide on a common technological basis (currently being an incompatible mix of 'Relational', ISO STEP, UML, XML, JSON, Linked Data/Semantic Web technology and own meta-models like ISO 12006-3). Decide one format, one direct access method and one data modelling language (that are compatible to each other).
- Next, or in parallel, decide on common top-level semantics integrating existing specs like IFC, CityGML, TC442 semantics for data templates and the SML Top Level model) and W3C generic resources. This semantics should also or especially cover geometric representations in 3D and meta-data.
- Where no choices can be made, provide standard linking between semantic resources (data set and/or data specification/dictionary level).
- Finally, define a common data environment as data space supporting the choices made above and underlying data space decisions (distributed/federated, encrypted, supporting secure single sign-on identification, authentication and authorisation).

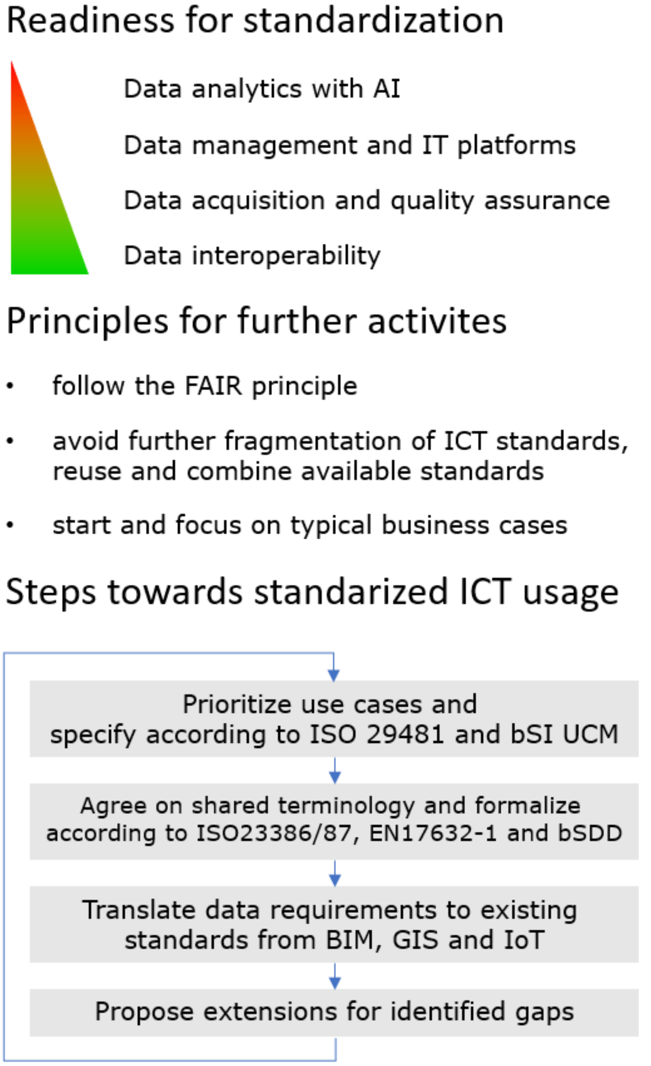

Figure 3 Conclusions and recommendations towards complementary standardization activities.

While acknowledging the importance of digitalization and the long-term need for further standardization of monitoring and predictive maintenance specifics, it is however recommended to start with definition of business requirements and, for validation in pilot projects, start with using the flexibility of existing standards. Such domain and business knowledge about IM-SAFE processes and workflows has been addressed in our deliverables and can be used as a starting point. The list of requirements contains for instance:

- Classification of information relevant for maintenance management, including data about the inventory, legal aspects, the condition of the asset, structural assessment and review, and inspection.
- Information requirements on organisational and asset level.
- Information quality assurance and control within the acquisition phase, processing and sharing phase and supporting decisions phase.

Those requirements need to be mapped to existing standards and further tested in pilot projects in order to settle down ICT requirements and specifications for standardization. Further details can be found in chapter 4.6 in Allaix et al. 2023 [1].

Acknowledgements

The authors acknowledge that this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958171. The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the Innovation and Networks Executive Agency (INEA) nor the European Commission are responsible for any use that may be made of the in-formation contained therein.

References

[1] Allaix, D.L., Daró, P., & Strauss, A. (2023). *Formulation of the EC mandate to CEN and plan of approach for the execution of the mandate*. H2020 CSA IM-SAFE report D5.1, Available at: https://im-safe-project.eu/wp-content/uploads/2021/01/958171_IM-SAFE_D5.1_V1.0.pdf.

[2] Weise, M., Böhms, H.M. & Allaix, D.L. (2023). Action plan and recommendations for complementary standardisation in digitalisation. H2020 CSA IM-SAFE report D5.4.

[3] Sánchez Rodríguez, A., Varela, M., Garrido, S., Lago, D., Sánchez Rodríguez, A., Sebastian, R., Böhms, H.M. & Giurgiu, I. (2022). *Guidelines for data acquisition, processing, and quality assurance.* H2020 CSA IM-SAFE report D4.1, Available at: https://im-safe-project.eu/wp-content/uploads/2021/01/958171_IM-SAFE_D4.1_V1.0.pdf.

[4] Weise, M., Sebastian, R., Mitsch, N. & Sánchez Rodríguez, A. (2022) *Framework for semantic integration of IoT, BIM and GIS of transport infrastructure.* H2020 CSA IM-SAFE report D4.2, Available at: https://im-safe-project.eu/wp-content/uploads/2021/01/958171_IM-SAFE_D4.2_V1.0-1.pdf.

[5] Rigotti, M., Elseicy, A.M.I., Sánchez Rodríguez, A., Justo Domínguez, A., Malossi, C., Giurgiu, I., Alovisi, I., Sanecka, P. & Sebastian, R. (2022). *Design of IT platforms for monitoring data of transport infrastructure*. H2020 CSA IM-SAFE report D4.3, Available at: https://im-safe-project.eu/wp-content/uploads/2021/01/958171_IM-SAFE_D4.3_V1.0.pdf.

[6] Giurgiu, I., Rigotti, M. & Sanecka (2022) *Data analytics and Artificial Intelligence*, H2020 CSA IM-SAFE report D4.4, Available at: https://im-safe-project.eu/wp-content/uploads/2021/01/958171_IM-SAFE_D4.4_V1.0.pdf.

[7] The FAIR principle: https://www.go-fair.org/.

[8] ISO 8000-110:2021, *Data quality — Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification*, https://www.iso.org/standard/78501.html.

[9] A Universally Unique IDentifier (UUID) URN Namespace, RDF 4122, IETF, https://www.ietf.org/rfc/rfc4122.txt.

[10] EN 17632-1:2022, *Semantic Modelling and Linking (SML) standard, Part 1*, CEN TC442, WG4, TG3, https://standards.cen-cenelec.eu/dyn/www/f?p=CEN:110:0::::FSP_PROJECT,FSP_ORG_ID:67839,1991542&cs=1BCE2DE4154949B53FB406C2D4EE5011D.

[11] Semantic Sensor Network Ontology, OGC/W3C Recommendation, 19 October 2017, https://www.w3.org/TR/vocab-ssn/.

[12] Infrastructure for Spatial information in Europe (INSPIRE), https://inspire.ec.europa.eu/.

[13] Sebastian R., Böhms H.M., Luiten G.T. (2021), *Developing a standardized approach of Asset Lifecycle Information Modelling for Semantic Digital Twin in the built Environment.* In: Proceedings of the Joint Conference of CIB W78 and Linked Data in Architecture and Construction, Luxembourg, 11-15 October 2021.

[14] A. Bartezzaghi, I. Giurgiu, C. Marchiori, M. Rigotti, R. Sebastian, C. Malossi (2022) *Design of a Cloud-Based Data Platform for Standardized Machine Learning Workflows with Applications to Transport Infrastructure.* IEEE 21st Mediterranean Electrotechnical Conference, MELECON 2022.

[15] Business Plan, CEN/TC 442 Building Information Modelling, Version 2020-12-15.