Cybersecurity und Safety in Smart Manufacturing

Standardization Landscape

ÖVE, Tagung zur Cybersicherheit, 2022-10-06

Erwin Schoitsch, AIT Austrian Institute of Technology

Co-funded by the European Union

What is Smart Manufacturing?

Acknowledgement and Disclaimer:

Most of this work on SM was done in IEC TC65 WG23, WG24 and JWG21 in developing SM standards IEC TR 63283-x, in IEC SC65A MT 61508 (Functional safety) and ISO/IEC JTC1 SC42 WG03 (AI Trustworthiness). Figures and tables are taken from working documents to show the directions in which concepts and ideas are developing. Beware that most of them are intermediate results of work still ongoing, not final standards and recommendations, and the selection represents only my personal view.

AIT contributions were co-funded by the European Commission mainly in ECSEL-projects.

"Smart manufacturing" is defined by ISO/TR 22100-4:2018 as follows:

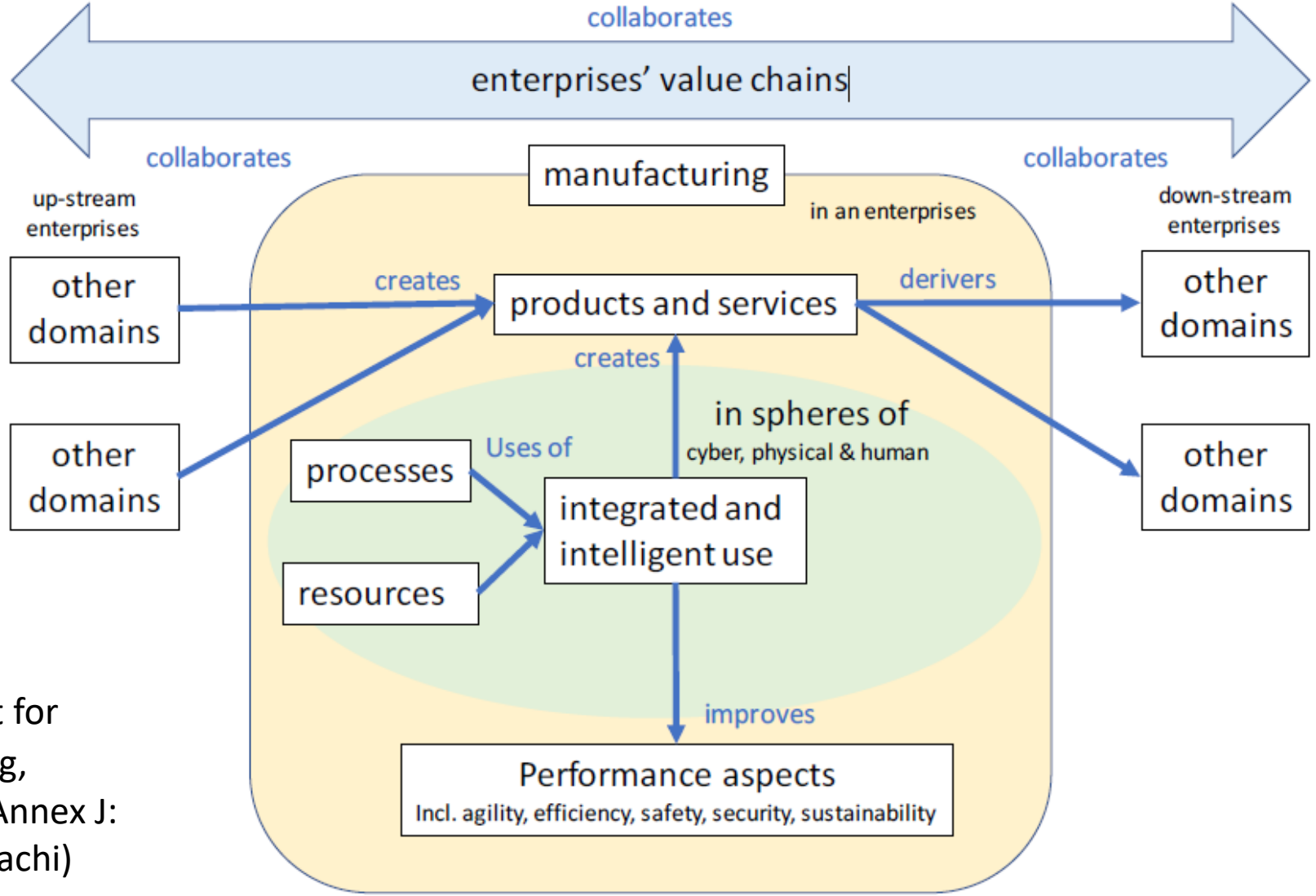
- manufacturing that improves its performance aspects with integrated and intelligent use of processes and resources in cyber, physical and human spheres to create and deliver products and services, which also collaborates with other domains within enterprises' value chains.
  - Note 1 to entry: Performance aspects include agility, efficiency, safety, security, sustainability or any other performance indicators identified by the enterprise.
  - Note 2 to entry: In addition to manufacturing, other enterprise domains can include engineering, logistics, marketing, procurement, sales or any other domains identified by the enterprise.

Smartness includes Digitalization and Intelligence

- **Note:** ISO/TR 22199-4:2018 - **Safety of machinery — Relationship with ISO 12100 — Part 4: Guidance to machinery manufacturers for consideration of related IT-security (cyber security) aspects**
- **Note:** IEC TR 63283-1 ED1; Smart Manufacturing – Part 1: Terms and definitions

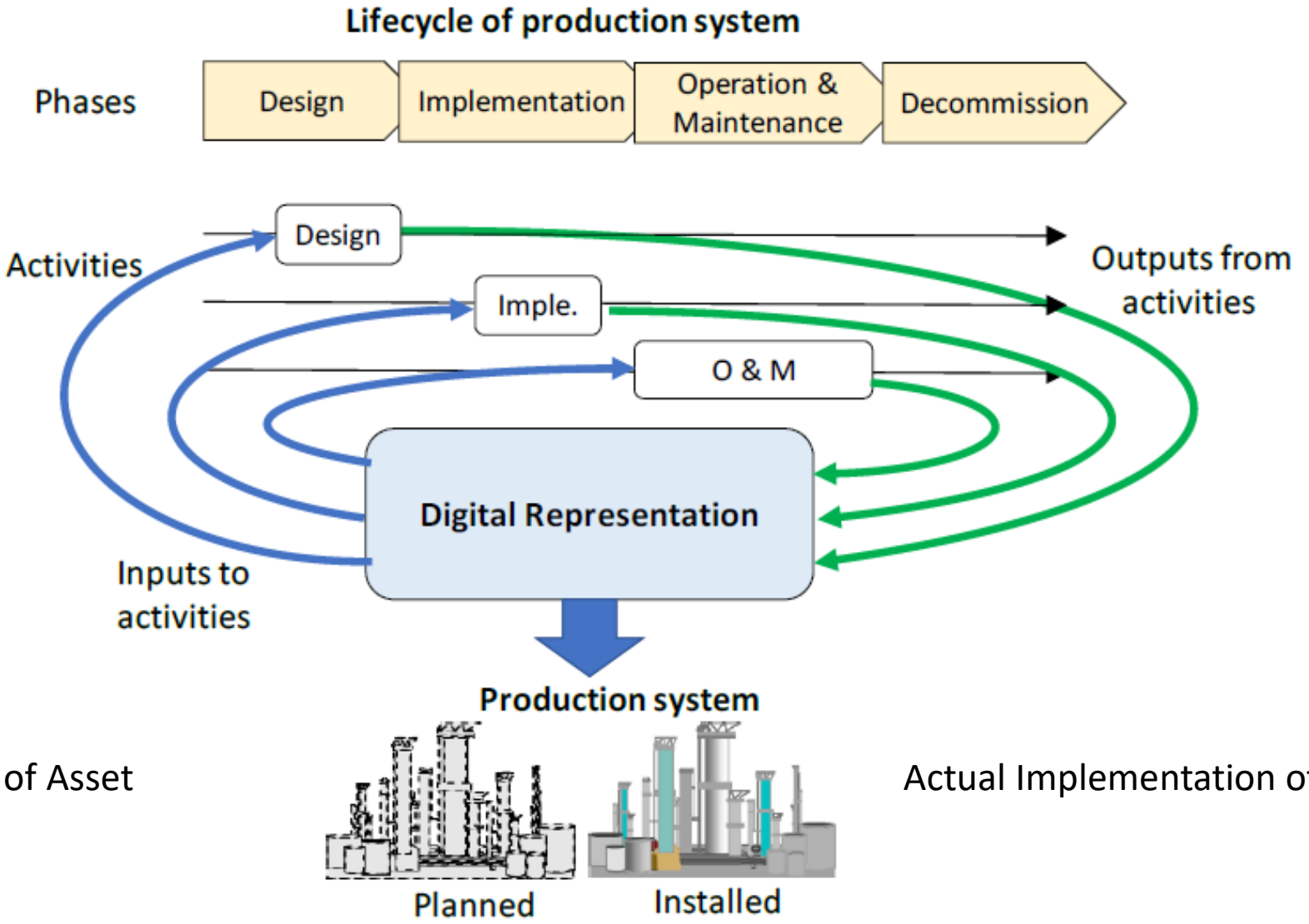
What is Smart Manufacturing?
Value Stream between Enterprises

Graphical representation of the definition of Smart Manufacturing



Source: IEC TC65 WG23, Working draft for IEC TR 63283-1, Smart Manufacturing, Part 1: Terms and definitions, former Annex J: Background and motivation (Koji Demachi)

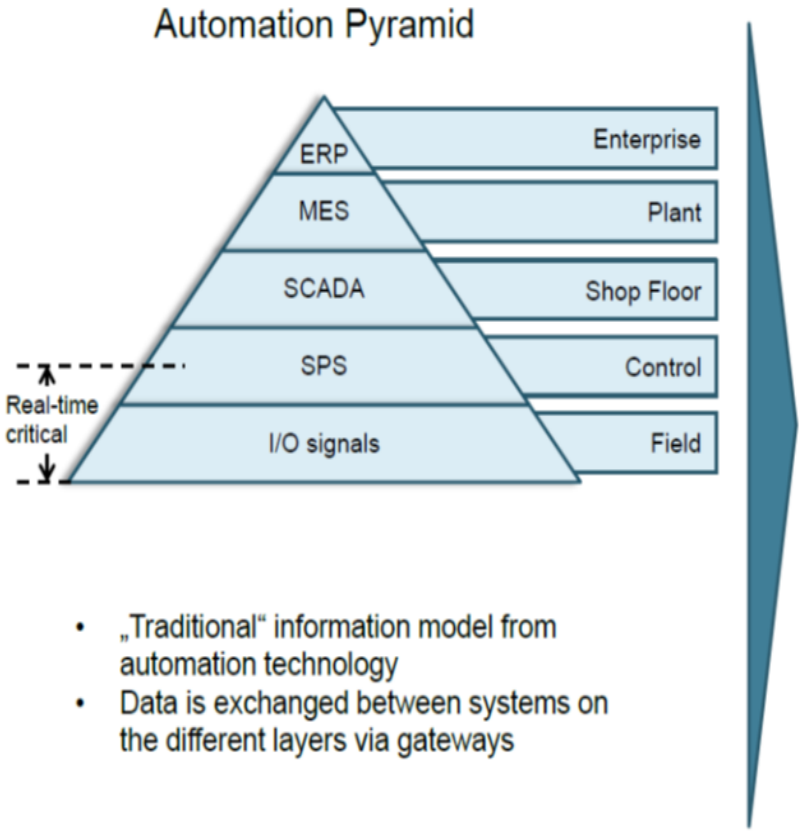
Digitalization: Value Stream over Life Cycle Phases
Digital representation of production system in the lifecycle



Type of Asset

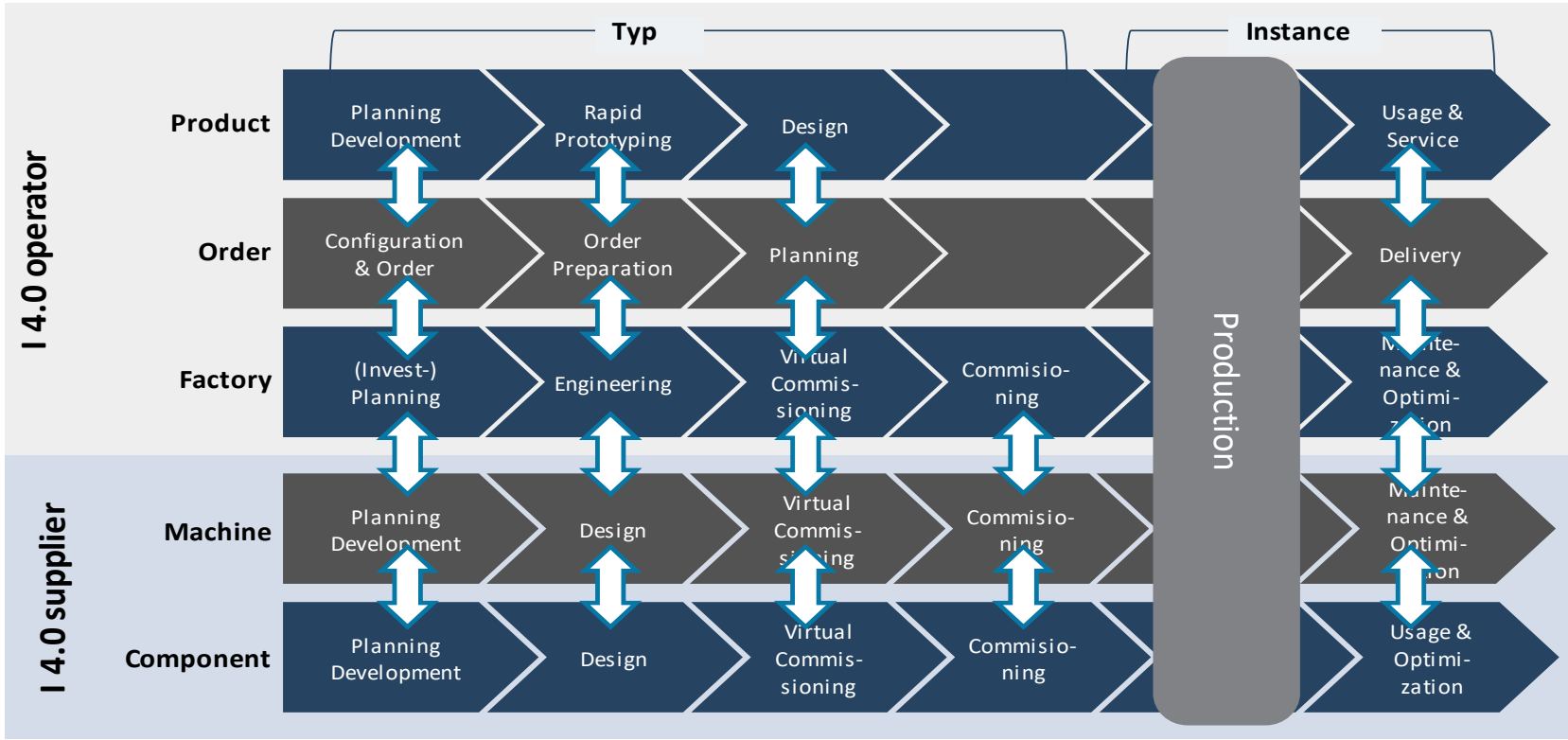
Actual Implementation of Asset of Asset

From Automation Pyramid to Value networks (IEC TR 63283-3)



- „Traditional" information model from automation technology
- Data is exchanged between systems on the different layers via gateways

Smart Manufacturing
From Value Streams to Value Networks

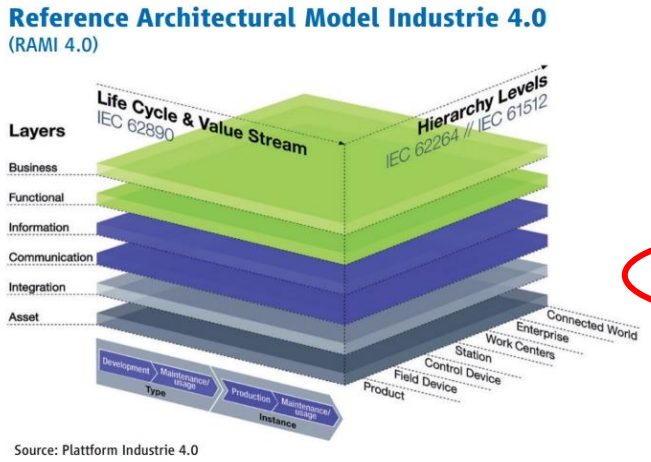


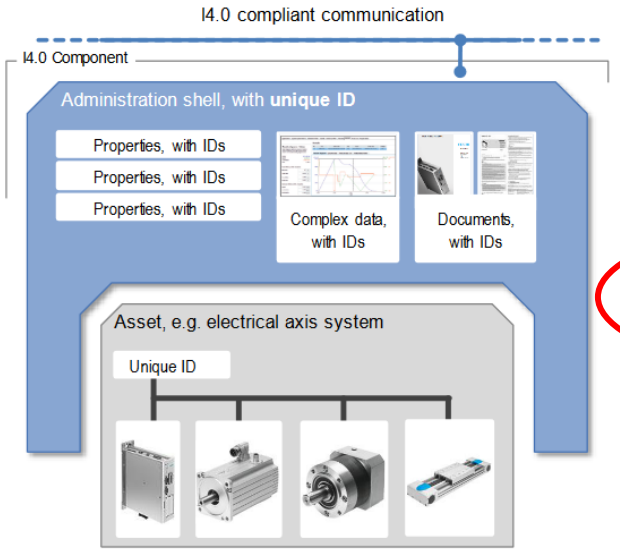
Interaction between Lifecycles is key

Source: based on Plattform Industrie 4.0 AG 1 / based on Prof. Bauernhasl, Fraunhofer IPA

Overview Smart Manufacturing Standards (SM) & Asset Administration Shell

RAMI 4.0/Asset (Component) ID





Asset Administration Shell Standards (RAMI4.0)

- IEC 63278 Asset Administration Shell for Industrial Applications (AAS)
- IEC 63278-1 SM Asset Administration Shell structure
- IEC 63278-2 SM Information meta model
- IEC 63278-3 SM Security provisions for AASs

SM Framework of Standards

standards proposed to be used in SM architecture (existing standards may need to be **enhanced to fully support SM**)

- IEC 62443 Security
- ISO 270xx Security
- IEC 61508 Functional Safety
- IEC 61511 Functional Safety
- IEC 63164 Reliability
- IEC 62890 Lifecycle M.

- IEC 62832 Digital Factory
- IEC 62264 MOM (MES) IEC 62424 P&ID
- IEC 62714 AutomationML
- IEC 63280 MTP
- IEC 62541 OPC UA

- IEC 61360 CDD
- IEC 62720 Units
- IEC 61987 List of Properties
- IEC 63365 Digital Nameplate
- IEC 61406 Identification Link
- ISO 13584-42
- ISO 22400 KPI

- IEC 61131 PLC
- IEC 61499 Function Block
- IEC 61512 Batch
- IEC 62682 Alarm Management

...

SM Base Standards

Standards especial supporting and used as base for SM architectures

- (IEC PAS 63088 RAMI 4.0 as long it exist)
- IEC TR 63319 SM Meta Reference Model
- IEC 63339 Unified Reference Model SM

- IEC 63283-1 SM Terms & Definitions
- ~~IEC 63283-2 SM Use Cases~~
- IEC 63283-3 SM recommendations for cybersecurity
- ~~SM recommendations for safety~~
- SM Framework and Concepts

Overview IEC TC65 WG23 Smart Manufacturing

- IEC TR 63283-1 ED1: Smart Manufacturing – Part 1: Terms and definitions
- IEC TR 63283-2 ED1 (under development): Industrial-process measurement, control and automation – Smart Manufacturing – Part 2: Use cases
- IEC TR 63283-3 ED1 (under development): Industrial-process measurement, control and automation – Smart Manufacturing – Part 3: Challenges for Cybersecurity
- IEC TR 63283(-4) WD (under development): Industrial-process measurement, control and automation – Smart Manufacturing – Part 4: New technologies

  This document is a "Smart manufacturing trend analysis". The identified new technologies relevant for AIOTI are:
  - AI
  - Edge computing;
  - Cloud technology;
  - Digital twin;
  - New communication protocols, 5G, TSN;
  - Big data and data analytics;
  - IoT and IIoT;
  - Privacy technology, etc.

  Each chapter has a subchapter on "Technology description", "Use case analysis" and "Standardization needs".
- **IEC TR 63283-5 ED1** (under development): Industrial-process measurement, control and automation – Smart Manufacturing – Part 5: Market and innovation trends analysis
  - Industrial IoT, (I)IoT devices, Edge, Cloud, 6G and AI are the key topics and sections,
- IEC TR 63283-6 ED1 (CD under development): Evaluation of intelligence properties for SM
- *IEC TR 63283-7 ED1 (draft under development(DE)): SM recommendations for safety*
- *IEC TR 63283-8 ED1 (TF started): SM Gap Analysis*
- **IEC TC65 WG24:** Asset Administration Shell for Industrial Applications
- **IEC TC65/ISO TC 184: JWG21:** Smart Manufacturing Reference Model(s)

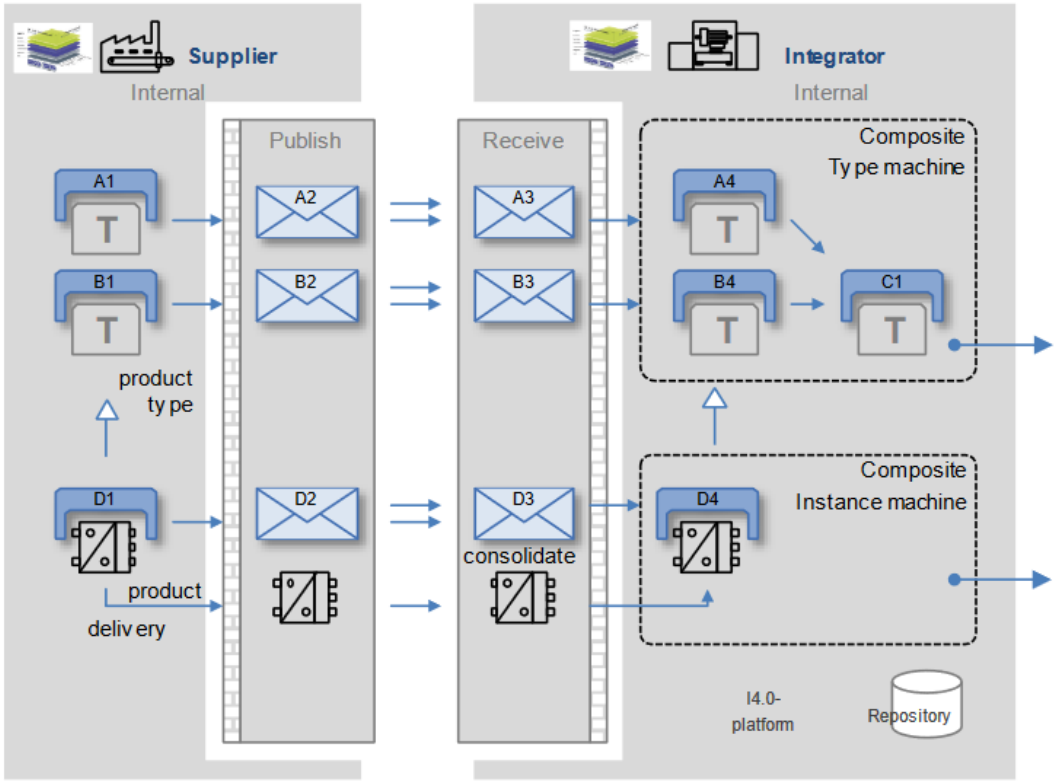
IEC TC65 WG24 – Asset Administration Shell

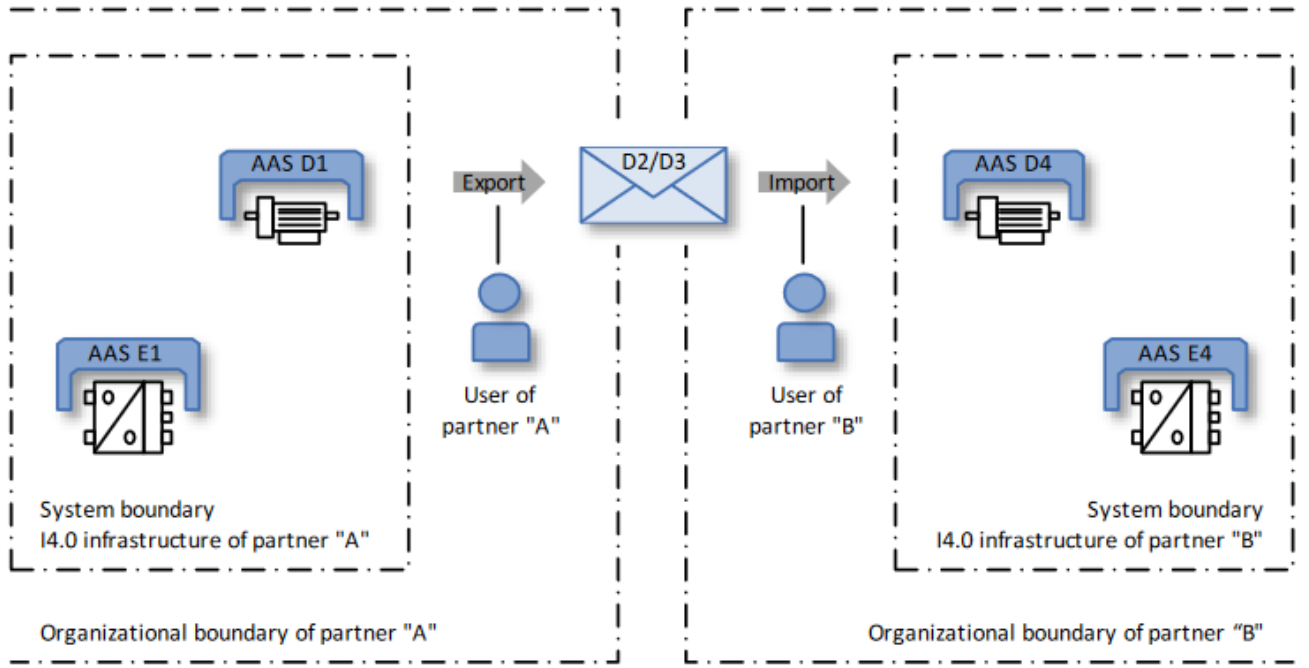
WG24 Asset Administration Shell for Industrial Applications:

- IEC 63278-1 Ed1: Asset administration shell for industrial applications – Part 1: Administration shell structure: CDV was circulated from 2022-05-13 to 2022-08-05
- IEC 63278-2 Ed1: Asset administration shell for industrial applications – Part 2: Information meta model: CD is in preparation
- IEC 63278-3 Ed1: Asset administration shell for industrial applications – Part 3: Security provisions for AAS: NP is approved, CC and CD is in preparation –
  - Work could start because of my nomination as expert, because a fifth participating NC was needed
- **JWG21 Smart Manufacturing Reference Model(s): IEC TC65 & ISO TC184 SC4**
  - modelling framework to support various arrangements of manufacturing elements into conceptual configurations deemed pertinent to domains of manufacturing enterprises
  - conceptualization of semantic models that reside within the modelling framework
- IEC PAS 63088:2017 Ed. 1.0: Smart manufacturing - Reference architecture model industry 4.0 (RAMI4.0): published 2017-03-07
- IEC TR 63319: A meta-modelling analysis approach to smart manufacturing reference models: publication is expected soon (2023-03)
- IEC 63339: Unified reference model for smart manufacturing: CC with observations is circulated, CDV is in preparation

Asset Administration Shell

(Example The exchange of information between partners in the value chain of Industrie 4.0)





Source: Plattform Industrie 4.0

Picture Hoffmeister, Jochem, according Epple, 2016

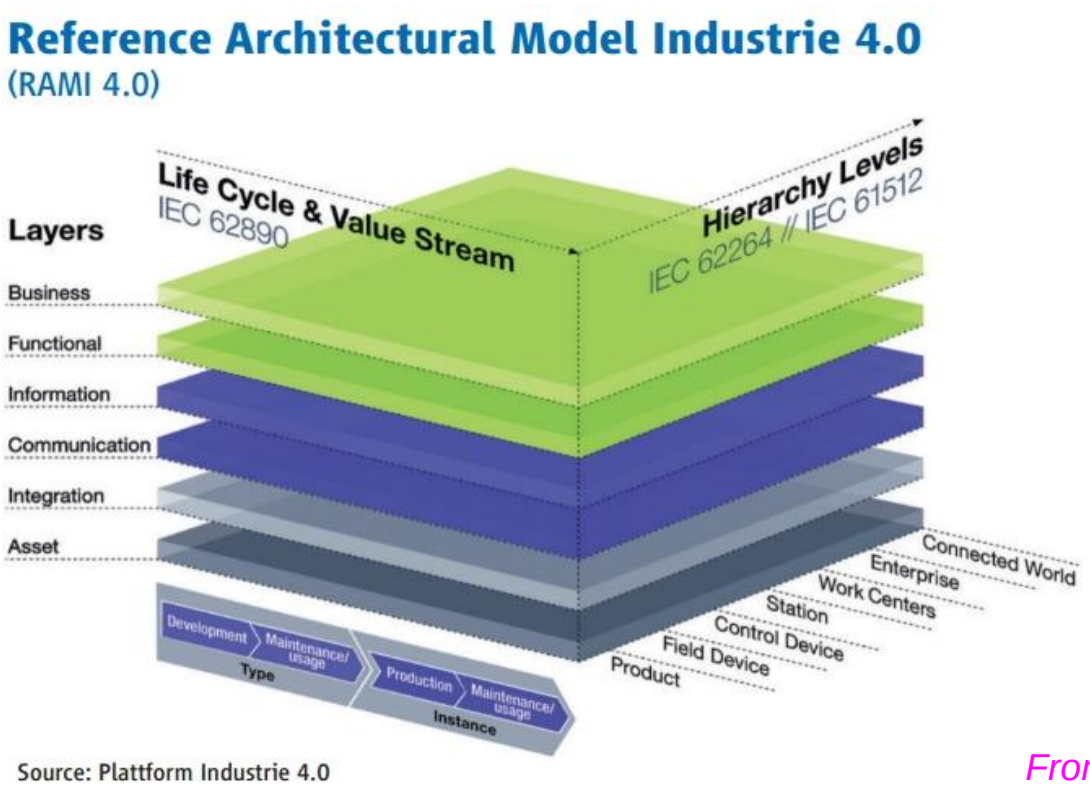
Two types of Assets are exchanged: one for the asset being the type of a product, one for the assets being the actual product instances

File exchange between two value chain partners, container format with predefined structure;
Usability & Security requirements!

Source: Plattform Industrie 4.0, Specification: Details of the Asset Administration Shell, Part 1 (BMWK)

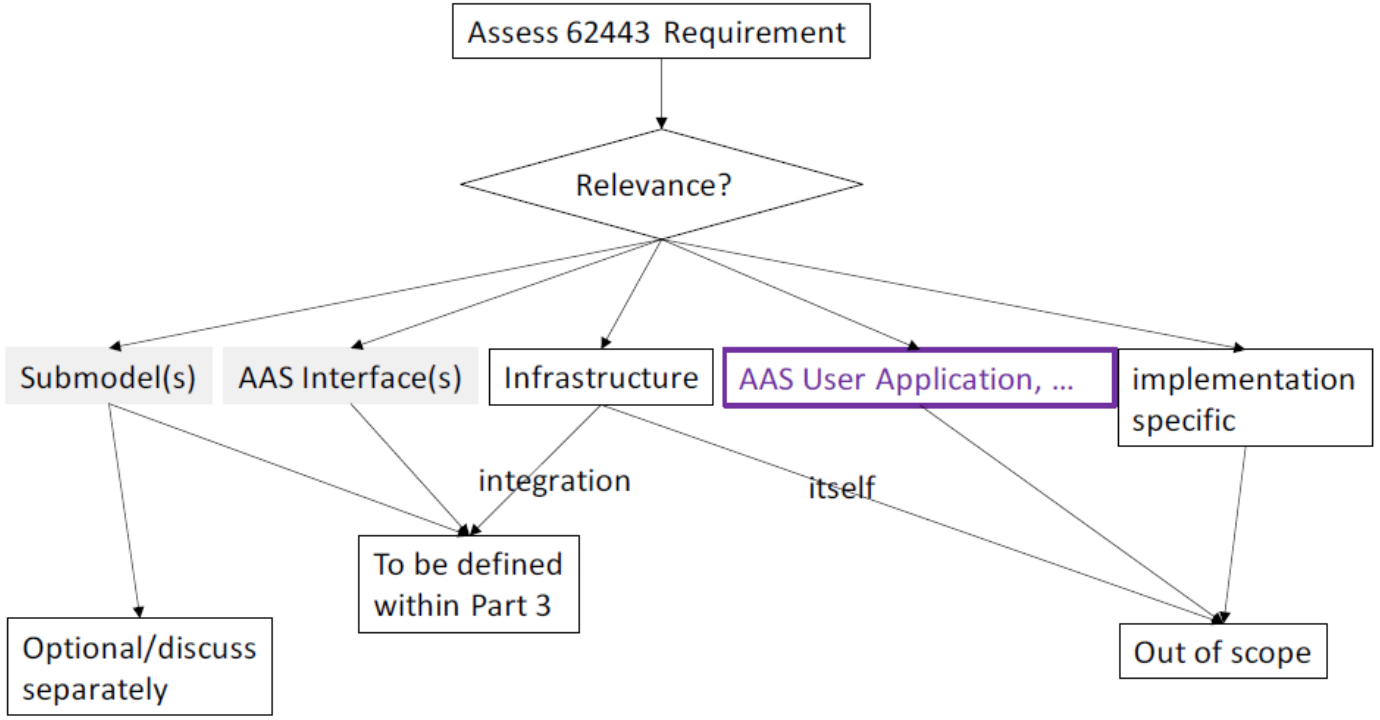
Asset Administration Shell – NEW: Cybersecurity Provisions

Classification scheme for IEC 62443 requirements wrt. AAS



Source: Plattform Industrie 4.0

Life Cycle & Value Stream:
Type: Development –> Maintenance/Usage
Instance: Production -> Maintenance/Usage



From NP – first considerations: Security Requirements with respect to:

- *objects of the meta model, e.g., access to asset information)*
- *services of the interface, e.g., exploration of a subtree, use of a specific REST API to interact with AAS);*

Interfaces are described as infrastructure objects – so they are assets themselves.

Security: Foundational Requirements – Tables (example from NP)

Table 7 – Classification of IEC 62443 FR7 Resource Availability w.r.t. AAS

FR	Title [62443-4-2 CR]	Submodel(s)	AAS Interface(s)	Infrastructure	Other	Implementation specific
FR 7 – Resource Availability						
7.1	Denial of service protection		•	•		•
7.2	Resource management		•			•
7.3	Control system backup	•		•		•
7.4	Control system recovery and reconstitution	•		•		•
7.5	Emergency power					•
7.6	Network and security configuration settings					•
7.7	Least functionality	•				
7.8	Control system component inventory					•

Industrial IoT – IIoT (Industrial Internet Consortium)

Intersection & Collaboration with ISO/IEC JTC1 SC41 (IoT), IEC TC 65 JWG 21

- **IEC PAS 63441 - Functional architecture of industrial internet system for industrial automation applications**

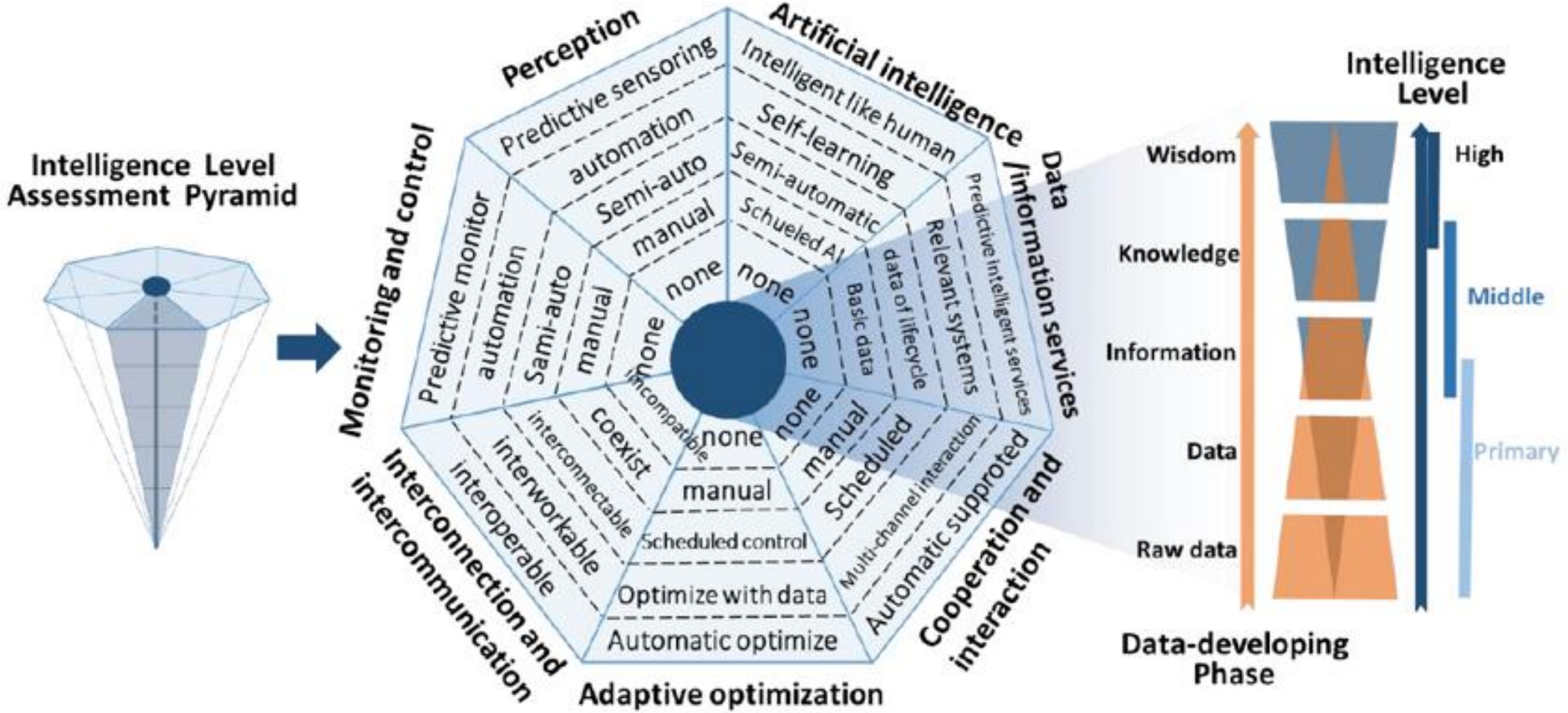
  "IEC PAS 63441:2022 (EN) defines
  - *the functional architecture and functional model of the Industrial Internet System for industrial applications.*
  - *presents the models, structures, activities, and interaction contents between layers of the end, edge, and cloud: infrastructure as a service (IaaS), platform as a service (PaaS), and software as service (SaaS), respectively."*
- **ISO/IEC JTC1 SC41 (IoT and Digital Twin) with ISO/IEC TR 30166: 2020**
- *Describes:*
  - ***general Industrial IoT (IIoT) systems and landscapes*** *which outline characteristics, technical aspects and functional as well as non-functional elements of the IIoT structure and*
  - *a listing of standardizing organisations, consortia and open-source communities with work on all aspects on IIoT;*
  - *considerations for the* ***future standardization perspective of IIoT*** *including risk analysis, new technologies and identified collaboration*

Others:

- **ISO/IEC JTC1 SC41 – IoT and Digital Twin (Host: IEC)**
- **ISO TC 184 SC4 Automation systems and integration – Industrial data – Digital Twin**

  Link to IoT:
  - *The Digital Twin reference architecture is based on the Internet of Things (IoT) reference architecture defined in ISO/IEC 30141, which provides guidance for the architect developing an IoT system and aims to give a better understanding of IoT systems to the stakeholders of such systems, including device manufacturers, application developers, customers and users.*
  - *The User entity can access/manipulate IoT devices (e.g., sensors, actuators) using protocols such as OPC-UA, OCF, LwM2M, and oneM2M. Edge computing is addressed in context of security, cloud, and IoT data.*

Smart (=Intelligent) Manufacturing – Intelligent Properties & Assessment



Smart Manufacturing – Intelligent Properties & Assessment

Example for Requirements: Perception (sensoring and processing) → Challenges for Safety & Security

- **Data acquisition requirements**
  - Automation device and systems should have sensor elements to support data acquisition. Data acquisition support digital conversion of analog quantity, frequency and digital quantity acquisition. Data acquisition requirements include timeliness, integrity, accuracy, resolution, sensitivity and stability requirements.
- **Information processing requirements**
  - The requirements of intelligent equipment information processing at the sensor level include timeliness, consistency, reliability and durability.

Example for Requirements: Artificial Intelligence → Challenges for Safety & Security

- **Machine Learning**
  - The capability to acquire knowledge from users or input data.
- **Model based inference**
  - The capability to categorize, make matches and deductions based on deep learning models that have been trained.
- **Decision-making**
  - Actively response to various requests or commands based on reasoning results.
- **Memory (accumulation of knowledge base)**
  - The capability to search and execute relevant knowledge and model parameters in time according to the task requirements from the knowledge base and model base learned in the training process.

Smart Manufacturing – Intelligent Levels (Examples)

Intelligent property	Level	Requirements
Perception	**0**	None. Could not collect data of the device of system.
	1	Human perception. Collect the data and preliminary information processing by manuals.
	2	Semi-automation. Collect the data and preliminary information processing with semi-automation.
	3	Automation perception. Collect the data and preliminary information processing with automation.
	4	Predictive collect the data and preliminary information processing.
Artificial Intelligence	**0**	Lack of AI application.
	1	AI are applied in several scheduled scenarios and given functions, but it could not make automatic reasoning.
	2	Provide semi-automatic reasoning with programmatic and scheduled cases based on AI algorithms and models, such as auto-compensation in given operation conditions.
	3	Provide self-reasoning, self-studying, self-decision methodology in open scenarios, which based on past and present test data to automatic calculate the parameters of AI algorithms and model, such as predictive adjustments, predictive maintenance, etc.
	4	Be able to have the intelligence capability like human intelligence with innovation and originality, which would not be discussed in this standards.

Cybersecurity and Profiles (conventional manufacturing as basis for SM)

“Cybersecurity” as defined in IEC TS 62443-1-1 Edition 1

- actions required to preclude unauthorized use of, denial of service to, modifications to, disclosure of, loss of revenue from, or destruction of critical systems or informational assets
  NOTE: The objective is to reduce the risk of causing personal injury or endangering public health, losing public or consumer confidence, disclosing sensitive assets, failing to protect business assets or failing to comply with regulations. These concepts are applied to any system in the production process and include both stand-alone and networked components. Communications between systems may be either through internal messaging or by any human or machine interfaces that authenticate, operate, control, or exchange data with any of these control systems. Cybersecurity includes the concepts of identification, authentication, accountability, authorization, availability, and privacy.

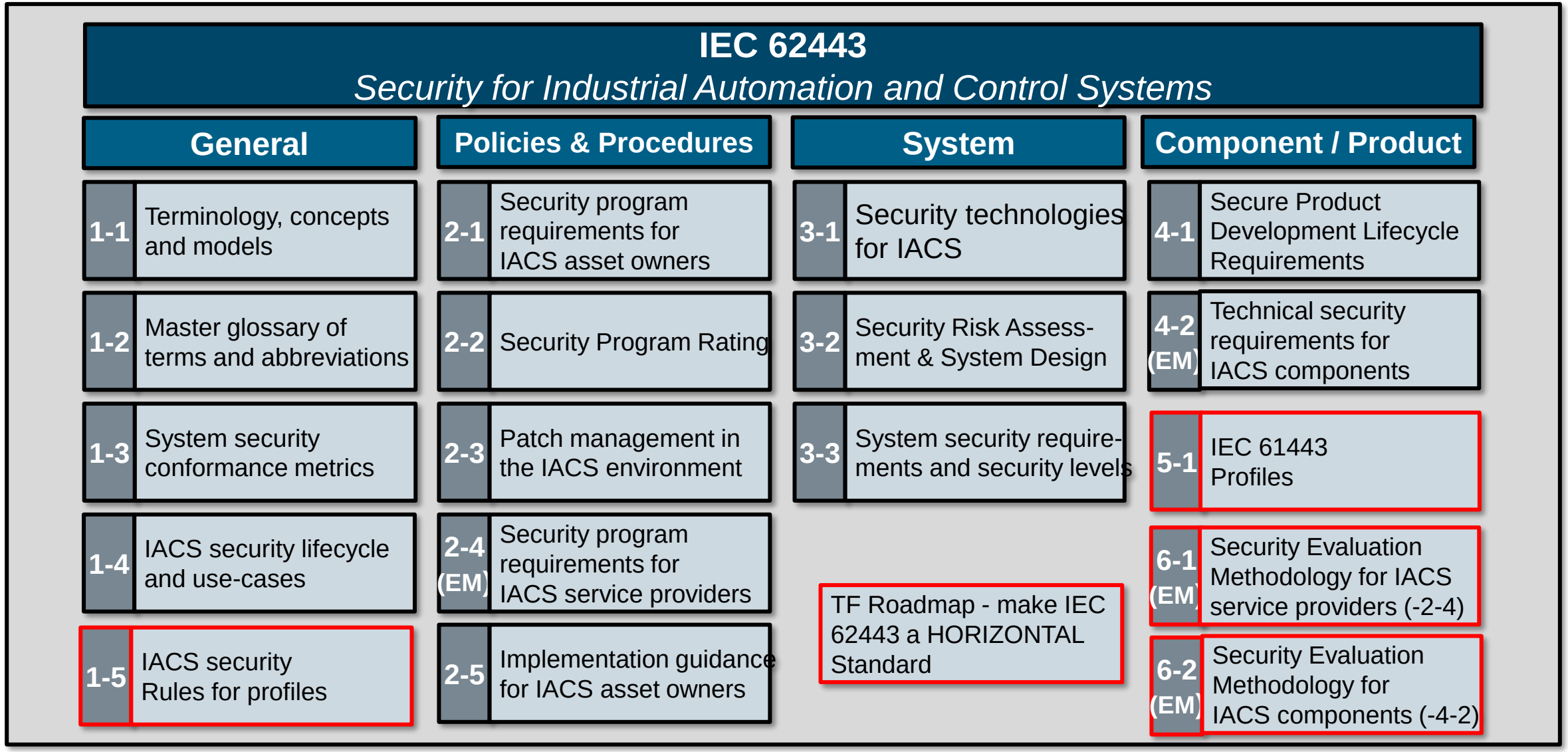
Profiles:

Primary objectives of standardization: to minimize variation and encourage single common standards for worldwide use. But sometimes it is necessary to choose (sub)sets of characteristics from a common defined framework for specific applications.

- EXAMPLE 1: Application-specific variants of a standard or set of standards.
- EXAMPLE 2: User profiles, which are a defined subset that is valid for a specific type of user.
- EXAMPLE 3: A subset of characteristics designed for one specific function.

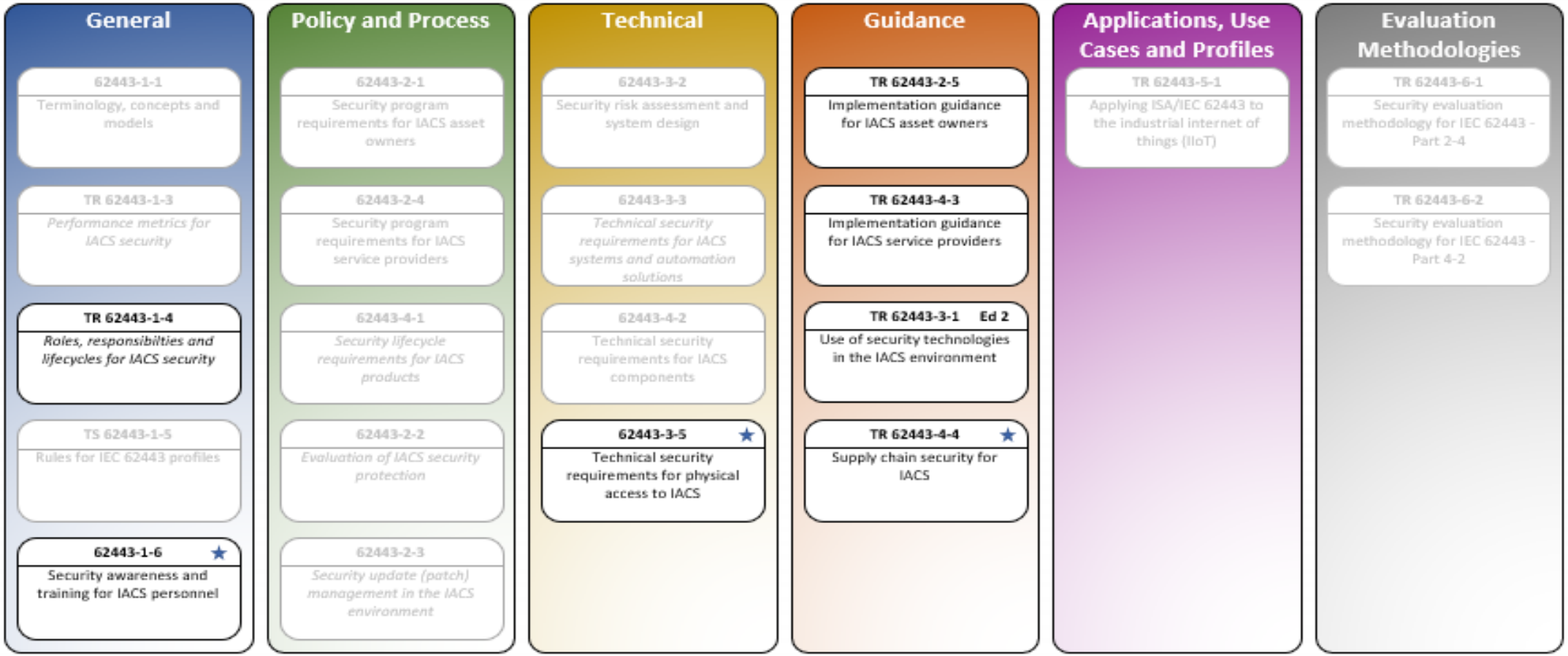
IEC 62443 series – Cybersecurity for classical IACS

- IACS (Industrial Automation and Control Systems) Security - ISA 99 (US) and IEC TC 65 WG10



IEC 62443 series Restructuring – Cybersecurity for classical IACS

Future Roadmap



★ indicates a new document

Existing documents (9/2020) retain their document number

Italics indicates a changed title

Smart Manufacuring: particular Cybersecurity challenges

- Classic manufacturing systems follow IACS Security along IEC 62443 series (industrial) and ISO 27000 series (organisational)
- **SM characteristics with impact on Cybersecurity:**
  - **Multiple Stakeholders** (asset owners (production, equipment, process, data analytics, services) – balance and protect potentially opposing interests
  - **Continuous Change and Reconfiguration** (enhancement, equipment, product to be produced, process optimization, overlapping life cycles, security during transition and adjustment)
  - **Digitalization:** intensive use of digital data (product, process, sensors, engineering data, simulation models, maintenance data, machine-internal and descriptive data,plannin, designe, supplies, …) – direct benefits for attackers (data acquisition, competitive advantages) or negative impact on production, blackmailing (ransomware):
  - **Architecture:** Automation Pyramid transformed to a structured automation network based on service-oriented paradigms (new architectural concepts, e.g., IEC PAS 63088:2017 (RAMI 4.0), BS ISO/IEC 30131:2018 IoT Reference Architecture, ISO/IEC 30166 ED1 IoT – Industrial IoT) – need to adjust security architecture
  - **Application of new communication technologies:** Wireless networking, 5G, protocols for communication across the different system layers, from field device sensors to enterprise level.
  - **Smart means „intelligent“, partial autonomy and decision making** – impact of AI and ML paradigms on safety & security (ISO/IEC JTC1 SC42 WG3 & IEC 61508-3 on TR 5469 „Functional safety of AI systems“; ETSI ISG SAI Securing AI – report) https://www.etsi.org/deliver/etsi_gr/SAI/001_099/004/01.01.01_60/gr_SAI004v010101p.pdf)
  - **Systems engineering** (complex integration of facilities/assets → Systems-of-Systems (ISO/IEC/IEEE 15288:2015 processes)

Recommendation for Improvement of IEC 62443 for SM (IEC CD TR 63283-3)

- Align on use of specific terms and definitions, esp. as far as components, products, and systems are concerned to avoid misunderstandings (cf. IEC 63283-1)
- Extension of virtual zones and conduits concept of IEC 62443 to adopt to logical/virtual topologies which are equivalent to that of the physical manufacturing system (e.g., ref. Digital Twin)
- Adjustment of Risk Assessment and Security Levels of IEC 62443, e.g., a manufacturing system may be able to switch between different (predefined) security levels depending on the actual production context (e.g., product being produced, configuration).
- Aligning Security Lifecycle of IEC 62443 with lifecycle of Smart Manufacturing systems, esp. giving guidance to what extent flexible reconfiguration is to be considered in the initial system design and can be handled within the maintenance process.
- Considering the product and its data being an integral part of the production system (e.g., carrying production data, providing feedback data), hence the product being manufactured may take part in the manufacturing process.
- Extension of Auditing and Logging of IEC 62443.

Smart Manufacturing Security Threats

IACS Protection goals: „CIA“

- Availability – the automation system should be able to execute its intended function. The production must not be disturbed by (intentional) attacks
- Integrity – the automation system should behave as intended and all data used in the production system should not by tampered with or modified by unauthorized entities
- Confidentiality – certain data, e.g., product, processing, or machine intellectual property, customer private data, KPIs, should not be disclosed to unauthorized entities

Viewpoints:

- Use Case View – Discusses potential threats relevant for the Use Cases of IEC TR 63283-2
- Technology View – Discusses potential threats related to individual newtechnologie that are used in smart manufacturing to efficently archive the goals stipulated by the use cases (IEC TR 63283-4 – under development)
- Lifecycle View – Discusses potential threats caused by additional life-cycle interdependencies of smart manufacturing systems.

List of 16 Use Cases in DTR 63283-3 → selected as example
→ **“Modularization of Production System”**

Summary of SM Cybersecurity Challenges – Apply Dev & Ops Measures

- ✓Identification and Authentication Control (AC)
- ✓Use Control (UC) (self-descriptive information of system and device)
- ✓Data and System Integrity (DI) Challenges) (data processed and exchanged)
- Data Confidentiality Challenges (DC) regarding Privacy and other issues
- Timely Response to Events Challenges (TRE)
- Restricted Dataflow Challenges (RDF) and
- Resource Availability Challenges (RA)

✓ {used in example, detailed description included, details from others omitted)}

Summary of SM Cybersecurity Challenges – Apply Dev & Ops Measures

Identification and Authentication Control (AC)

AC01	Authentication of the customer
AC02	Authentication of devices/sensors
AC03	Authentication of simulation data providers
AC04	Authentication of configuration change providers
AC05	Authentication of network participants
AC06	Authentication of rescheduling demanders
AC07	Authentication of providers for functional enhancements
AC08	Non-repudiation of audit information The increased flexibility in smart manufacturing (e.g., temporarily deployed 3rd party production systems entering and leaving the system) may require an increased level of trustworthiness of traceability information. Non-repudiation for traceability information among different stakeholders has to be ensured.
AC09	Authentication of sensor data

Summary of SM Cybersecurity Challenges

Use Control (UC) (self-descriptive information of system and device)

UC01	Use Control of production plan
UC02	Use Control of information
UC03	Use Control of simulation model
UC04	Use Control of production capabilities
UC05	Use Control of field devices
UC06	Use Control of network
UC07	Use Control of configuration

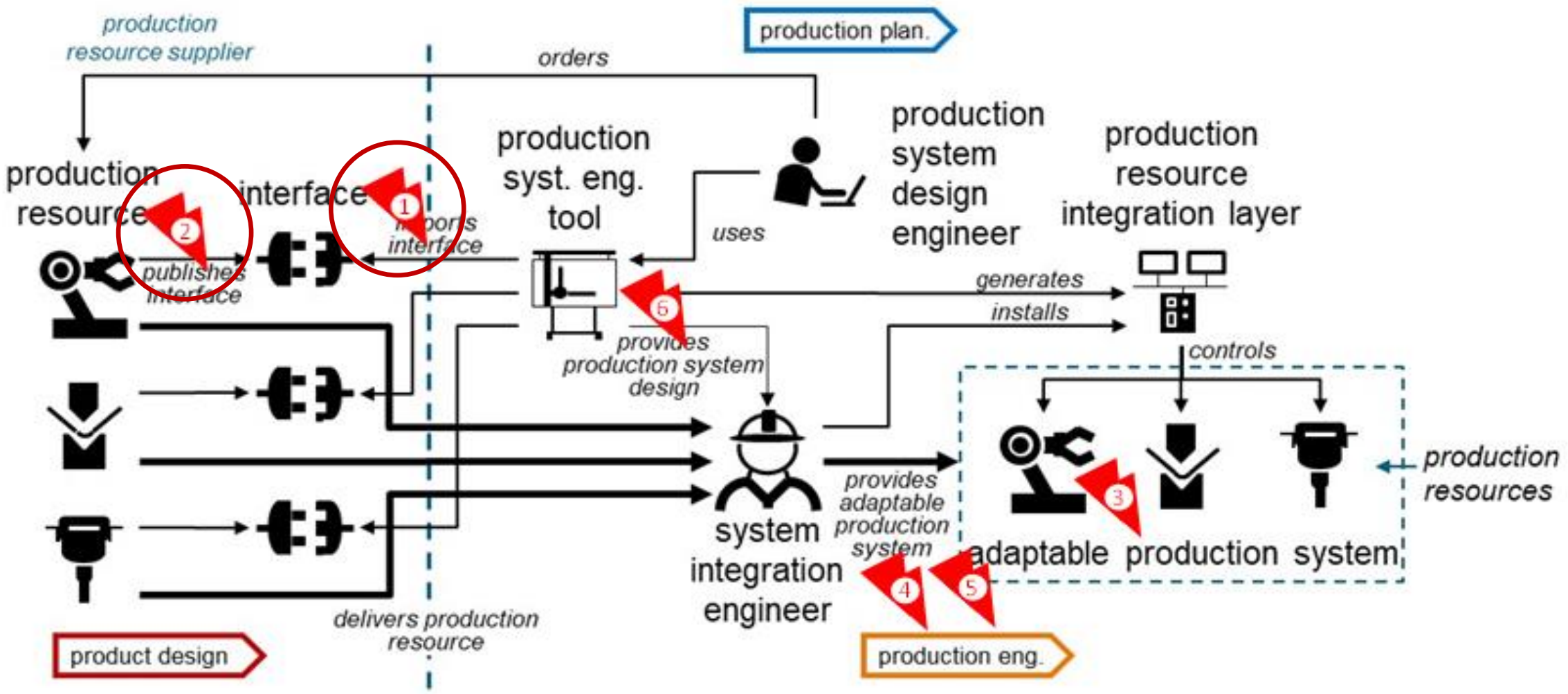
Summary of SM Cybersecurity Challenges

Data and System Integrity (DI) Challenges) (data processed and exchanged)

DI01	Data integrity of exchanged data
DI02	Data integrity of simulation model
DI03	Data integrity of new functions/configurations
DI04	Data integrity of production plans
DI05	Data integrity of engineering data
DI06	System integrity during transitions
DI07	End-point self-contained basic protection
DI08	Integrity of collected and aggregated sensor data
DI09	Ensure genuineness of installed devices/sensors
DI10	Integrity of (collected and aggregated sensor) data used for ML training and testing
DI11	Ensure genuineness of semi-finished and finished products
DI12	Data integrity with respect to known boundary conditions

USE Case Example: „Modularization of Production System“

Interchangeable production resources for flexible adaptation of production (different product or equipment replacement): A new field component adds itself automatically into an existing production system.



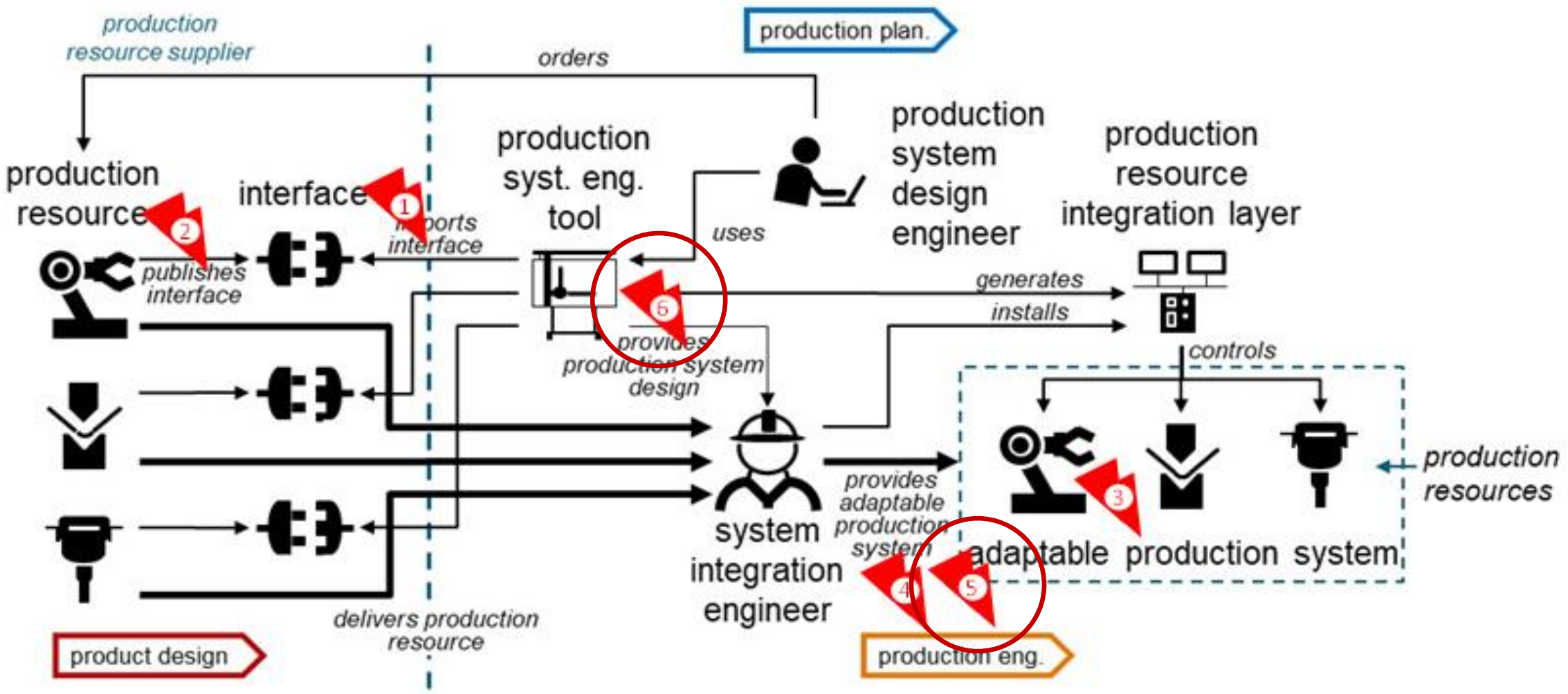
USE Case Example: „Modularization of Production System“

Interchangeable production resources for flexible adaptation of production (different product or equipment replacement): A new field component adds itself automatically into an existing production system.

Ref	Threat		Challenge	SM Specific?
1	Field device adds itself to the production system and accesses confidential information not intended for this device.	Confidentiality	**UC05 Use Control of field devices** – The new field device access permissions need to be set according to its intended task in the production process.	Security bootstrapping process. The device is initially unknown but requires sufficient permissions to integrate/interact with the existing system.
2	New field device impersonates another field device (e.g., offers functionality it cannot actualy provide) in order to get access to confidential information.	Confidentiality	**AC02 Authentication of devices/sensors** – Device needs to provide authenticated information about its identity and properties/functionalities.	No dedicated engineering processes.

USE Case Example: „Modularization of Production System“

Interchangeable production resources for flexible adaptation of production (different product or equipment replacement): A new field component adds itself automatically into an existing production system.



USE Case Example: „Modularization of Production System“

Ref	Threat		Challenge	SM Specific?
5	Attacker takes benefit of protection gap of individual devices during transition, e.g., zone protection may fail while a device is relocated.	Integrity	**DI07 End-point self-contained basic protection -** For ACS systems with static engineering it is often assumed that a device is located in a dedicated zone of trust and protected by the perimeter of that zone. This assumption may no longer hold for smart manufacturing systems. Each device needs to provide some basic self-protection on a restricted functional level. Advanced functionality may only be offered after the device was able to verify that certain (security) conditions are met by the environment.	Individual devices/endpoints can no longer rely on a stable operational environment
6	Attacker defines a new configuration state (or redefines an old one) which can be exploited more easily	Integrity	**DI03 Data integrity of new functions/configurations** – New configuration should be validated and authorized before being implemented **AC04 Authentication of configuration change providers** – Only authorized sources should be able to initiate configuration changes.	Configuration changes are normal and therefore not always suspicious

Trustworthiness (1) - Functional Safety and AI (ISO/IEC DTR 5469)

Joint approach in IEC 61508-3 (Basic Functional safety of E/E/PE Systems Standard, SW part) with ISO/IEC JTC1 SC42 WG03 (AI, Trustworthiness): Matrix on applicability of AI technology in safety-related system context – ***under development***

AI Technology Classes

Class I developed and reviewed using existing functional safety methods and standards,

Class II cannot be fully developed and reviewed using existing functional safety methods and standards, but it is still possible to identify a set of available methods and techniques satisfying the properties.

Class III cannot be developed and reviewed using existing functional safety methods and standards and it is also not possible to identify a set of available methods and techniques satisfying the functional safety properties.

AI Application and Usage Classes

A1: Used in safety relevant E/E/PE system and automated decision making possible.

A2: Used in safety relevant E/E/PE system and no automated decision making (e.g., for uncritical diagnostics).

B1: Used to develop safety relevant E/E/PE systems (offline support tool). Automated decision making of developed function is possible.

B2: Used to develop safety relevant E/E/PE systems (offline support tool). No automated decision making of the developed function is possible.

C: AI technology is not part of a safety function in the E/E/PE system. Has potential impact on safety (e.g., increase demand placed on a safety system)

D: AI technology is not part of a safety function in the E/E/PE system. No impact on safety due to sufficient segregation.

Current DTR-version:

Even more rigid as in my last year‘s presentation!

AI Technology Class => AI application and usage level	AI technology Class I	AI technology Class II	AI technology Class III
Usage Level A1 (1)	Application of risk reduction concepts of existing functional safety International Standards possible	Appropriate set of requirements (5)	Not recommended
Usage Level A2 (1)		Appropriate set of requirements (5)	
Usage Level B1 (1)		Appropriate set of requirements (5)	
Usage Level B2 (1)		Appropriate set of requirements (5)	
Usage Level C (1,3)		Appropriate set of requirements (5)	
Usage Level D (2)	No specific functional safety requirements for AI technology, but application of risk reduction concepts of existing functional safety International Standards (4)		

1 Static (offline) (during development) teaching or learning only

2 Dynamic (online) teaching or learning possible

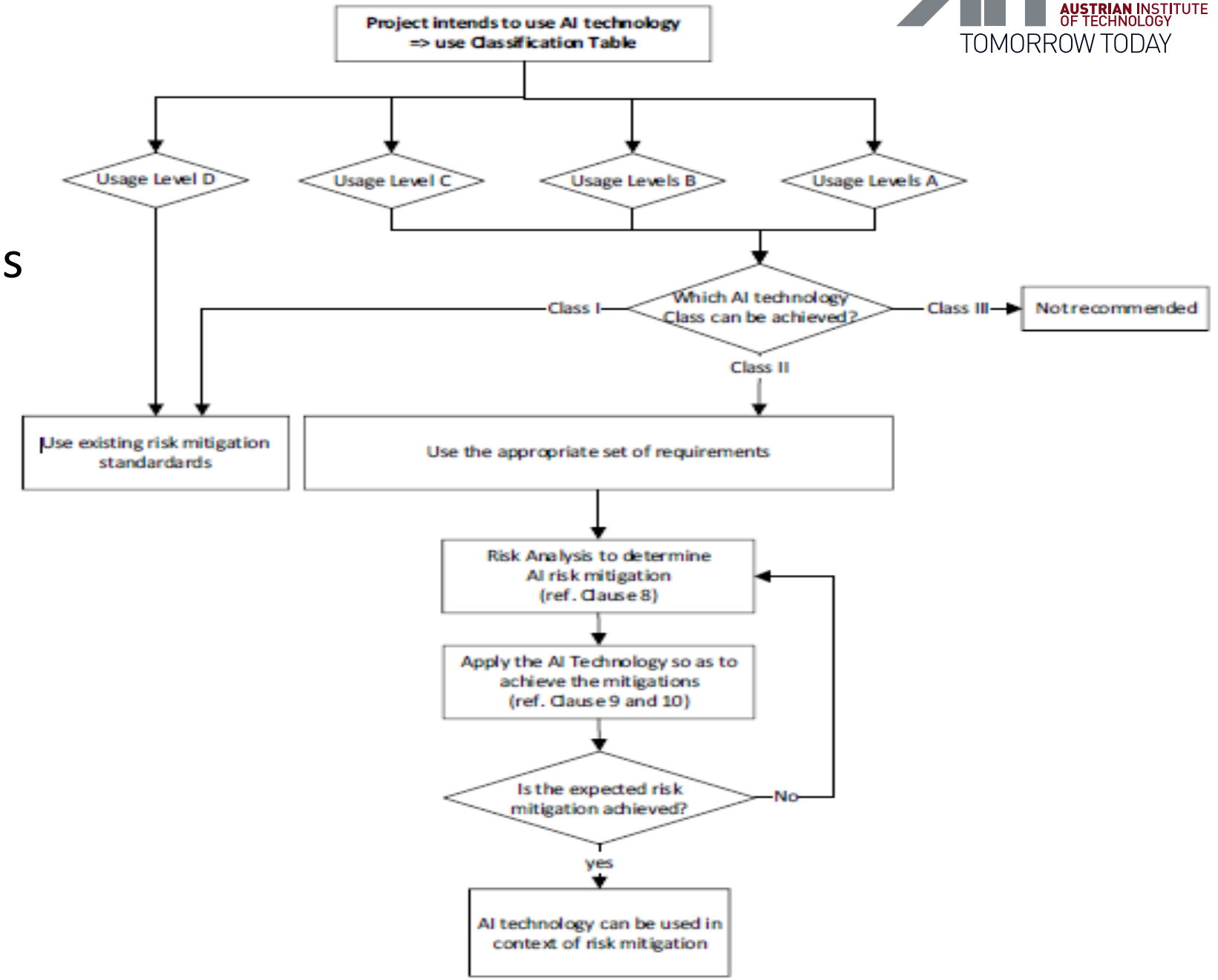
3 AI techniques clearly providing additional risk reduction and whose failure is not critical to the level of acceptable risk.

4 Additionally, other safety aspects (not being addressed with functional safety methods) can possibly be adversely affected by AI usage.

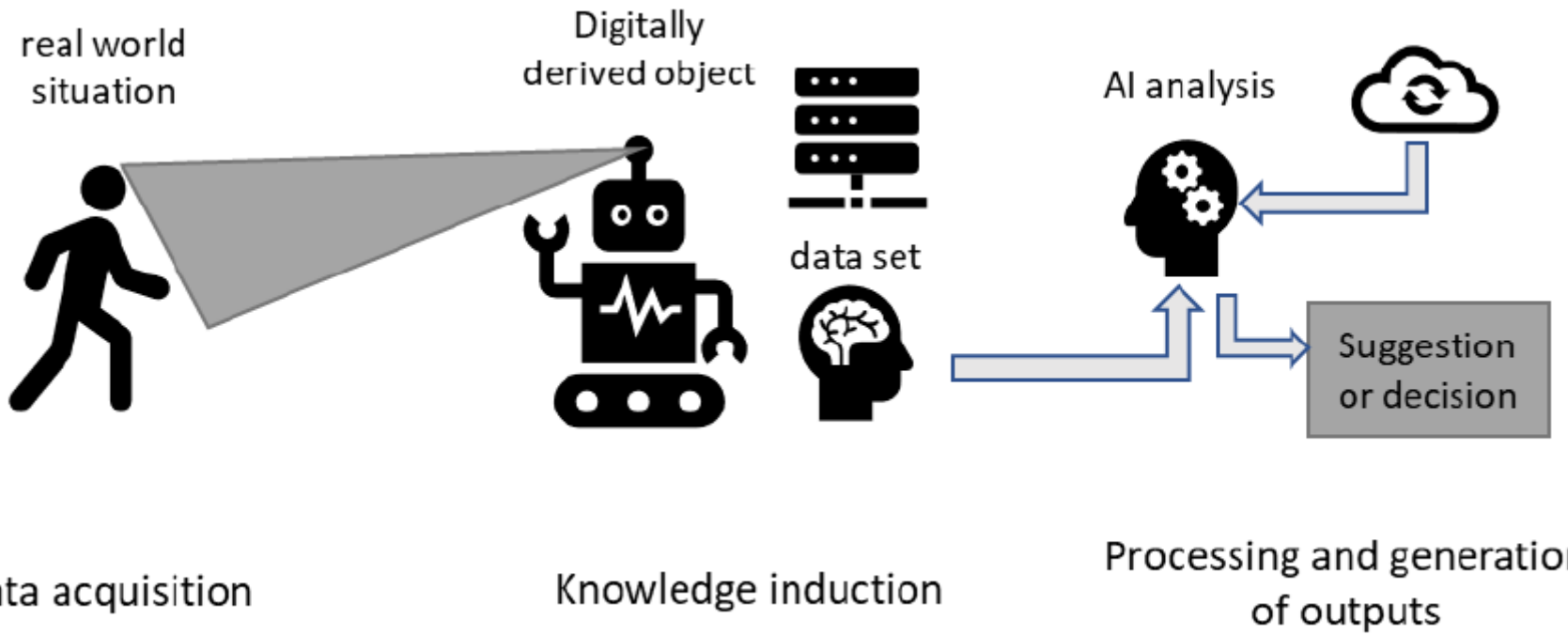
5 The appropriate set of requirements for each usage level can be established in consideration of Clauses 8, 9, 10 and 11. Examples are provided in Annex B.

Classification Scheme:

(Example how to assess Applicability of AI in Safety-related E/E/PE Systems)



Properties and 3 stages of realization



Three-stage realization principle:

- data acquisition;
- knowledge induction from data and human knowledge;
- processing and generation of outputs.

Deriving acceptance criteria:

- Desirable properties are defined for each of the three stages (depending on use cases)
- The properties are related to topics and eventually to detailed methods and techniques addressing those topics.
- Acceptance criteria are identified from the set of the detailed methods and techniques.

Properties and related risk factors (Examples)

Considerations on:

- Algorithms and models
- Level of automation and control („autonomy“) (ISO 22989)
- Degree of transparency and explainability
- Complexity of environment and vague specifications (data drift, concept drift, reward hacking, safe exploration, Resilience (inputs!), model attacks, HW issues, etc.)

		Level of automation	Comments
Automated system	Autonomous	Autonomy	The system is capable of modifying its operating domain or its goals without external intervention, control or oversight.
	Heteronomous	Full automation	The system is capable of performing its entire mission without external intervention.
		High automation	The system performs parts of its mission without external intervention.
		Conditional automation	Sustained and specific performance by a system, with an external agent being ready to take over when necessary.
		Partial automation	Some sub-functions of the system are fully automated while the system remains under the control of an external agent.
		Assistance	The system assists an operator.
		No automation	The operator fully controls the system.

Trustworthiness (2) – Cybersecurity and AI - ETSI Report GR SAI 004

V1.1.1 (2020-12) – Securing AI

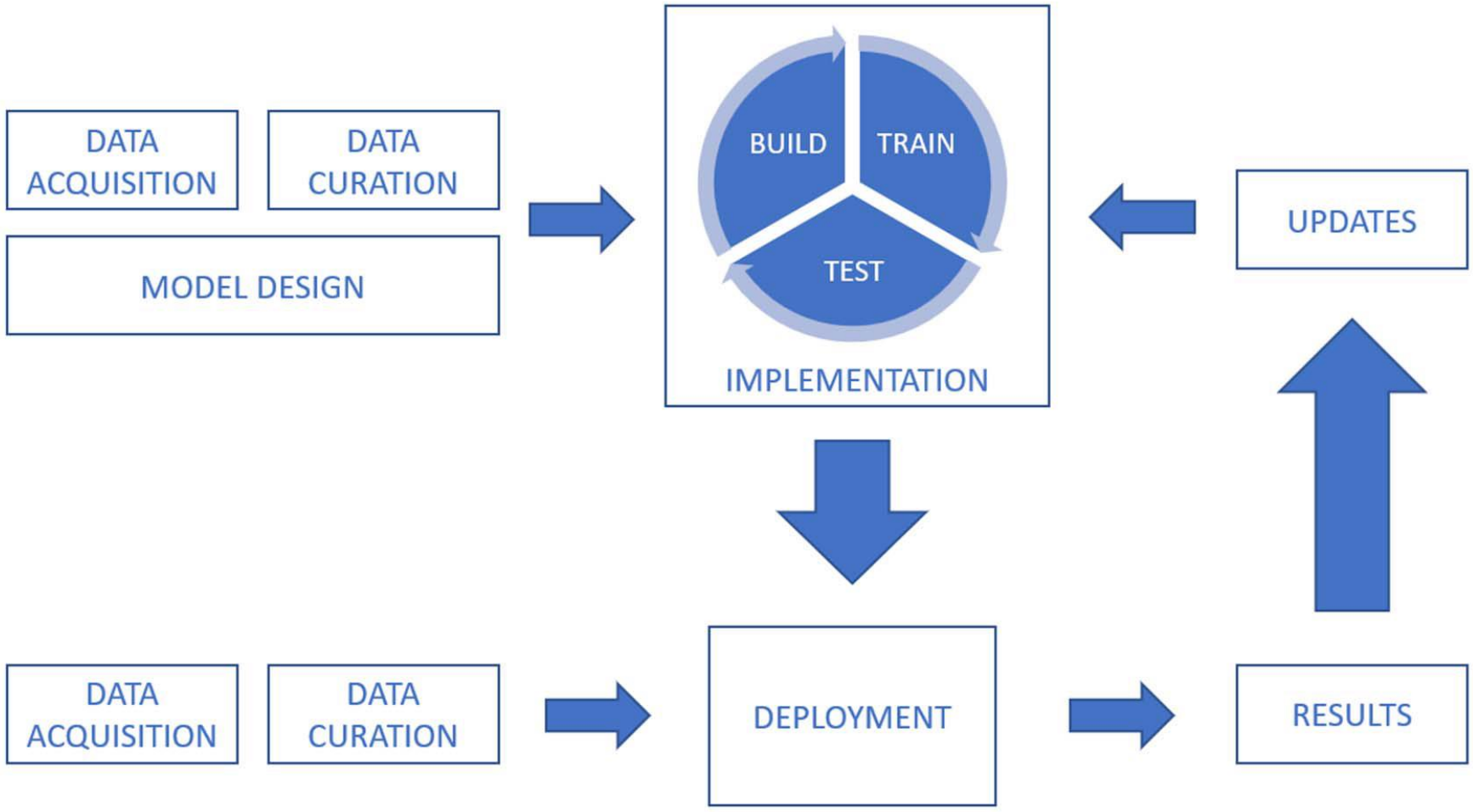
- https://www.etsi.org/deliver/etsi_gr/SAI/001_099/004/01.01.01_60/gr_SAI004v010101p.pdf

Typical Machine Learning Life Cycle

Specific Challenges?

<Model check>

<Reality check>



Trustworthiness (2) – Cybersecurity and AI - ETSI Report GR SAI 004
Specific Challenges for ML – Attacker Impact in certain LC Phases

Clause	Lifecycle Phase	Issues
4.3.2	Data Acquisition	Integrity
4.3.3	Data Curation	Integrity
4.3.4	Model Design	Generic issues only
4.3.5	Software Build	Generic issues only
4.3.6	Train	Confidentiality, Integrity, Availability
4.3.7	Test	Availability
4.3.8	Deployment	Confidentiality, Integrity, Availability
4.3.9	Upgrades	Integrity, Availability

Design challenges and unintentional factors:

- Bias (balanced data set) (confirmation, selection (subjective), outliers (extrem values), under- or overfitting (model too simplistic or complex) bias)
- Explainability (trust & transparency, understandability → assurance)
- Ethics (data access/privacy vs. Performance; decision making – Ethics Guidelines)
- Attack awareness/misuse (e.g. handwriting, voice, image deep fakes; data/algorithm/Model poisioning; Ad-blocker attacks, Malware obfuscation; DDOS; ….

Co-funded by
the European Union

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

Acknowledgements

Part of the work received funding from the EC via Horizon 2020, the ECSEL Joint Undertaking and the partners' national funding authorities (in Austria FFG (Austrian Research Promotion Agency) on behalf of BMK, The Federal Ministry of Climate Action, Environment, Mobility, Innovation and Technology): Productive4.0 (Grant agreement n° 737459-2), AutoDrive (737469-2), SECREDAS (783119), iDev40 (783163), AfarCloud (783221) and AI4CSM (101007326-2). ADEX (Automated Driving EXaminer) was funded by the Austrian Research Promotion Agency.

Thank You for your kind attendance !!