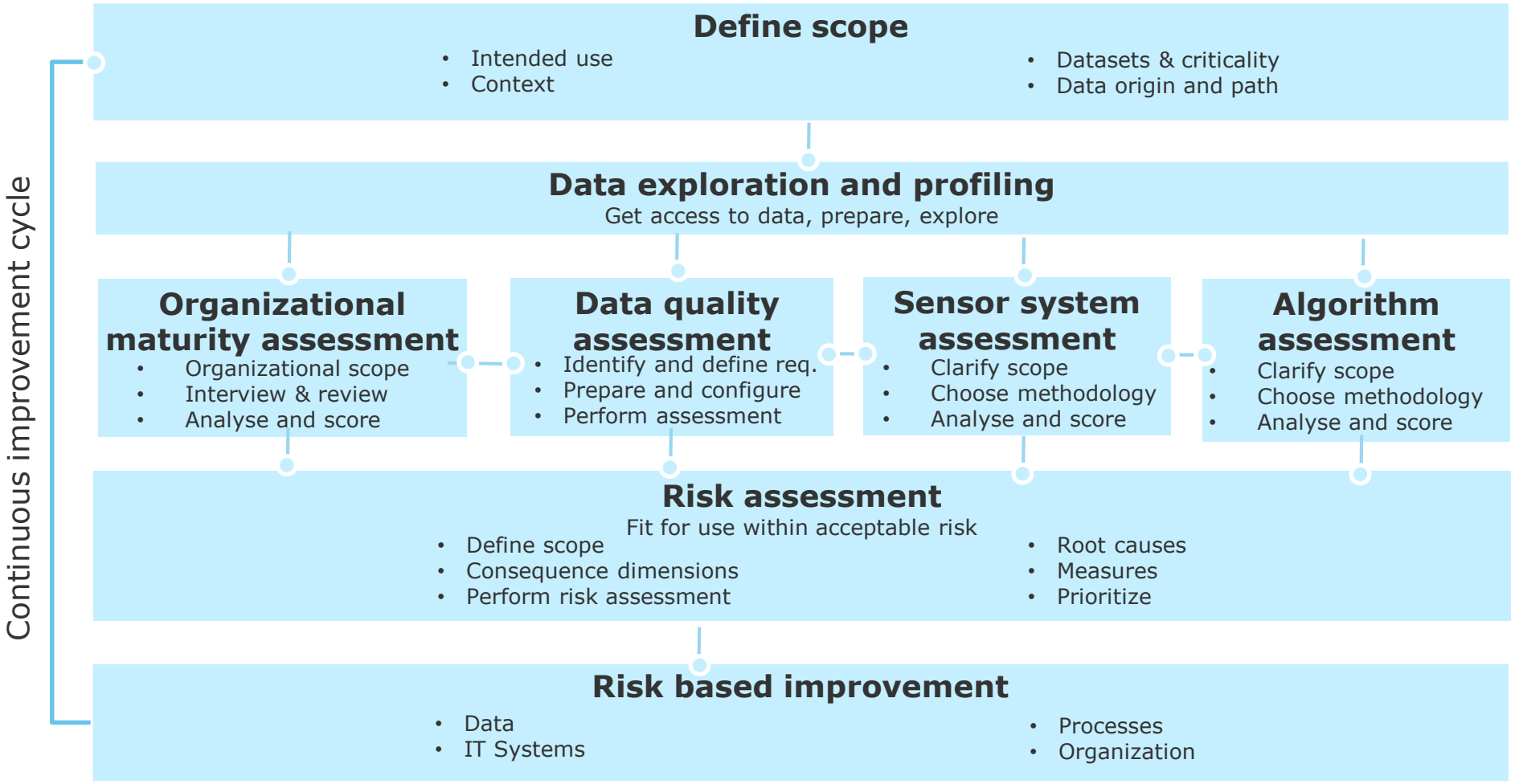
Quality of data and algorithms

Enabling trustworthy digital twins

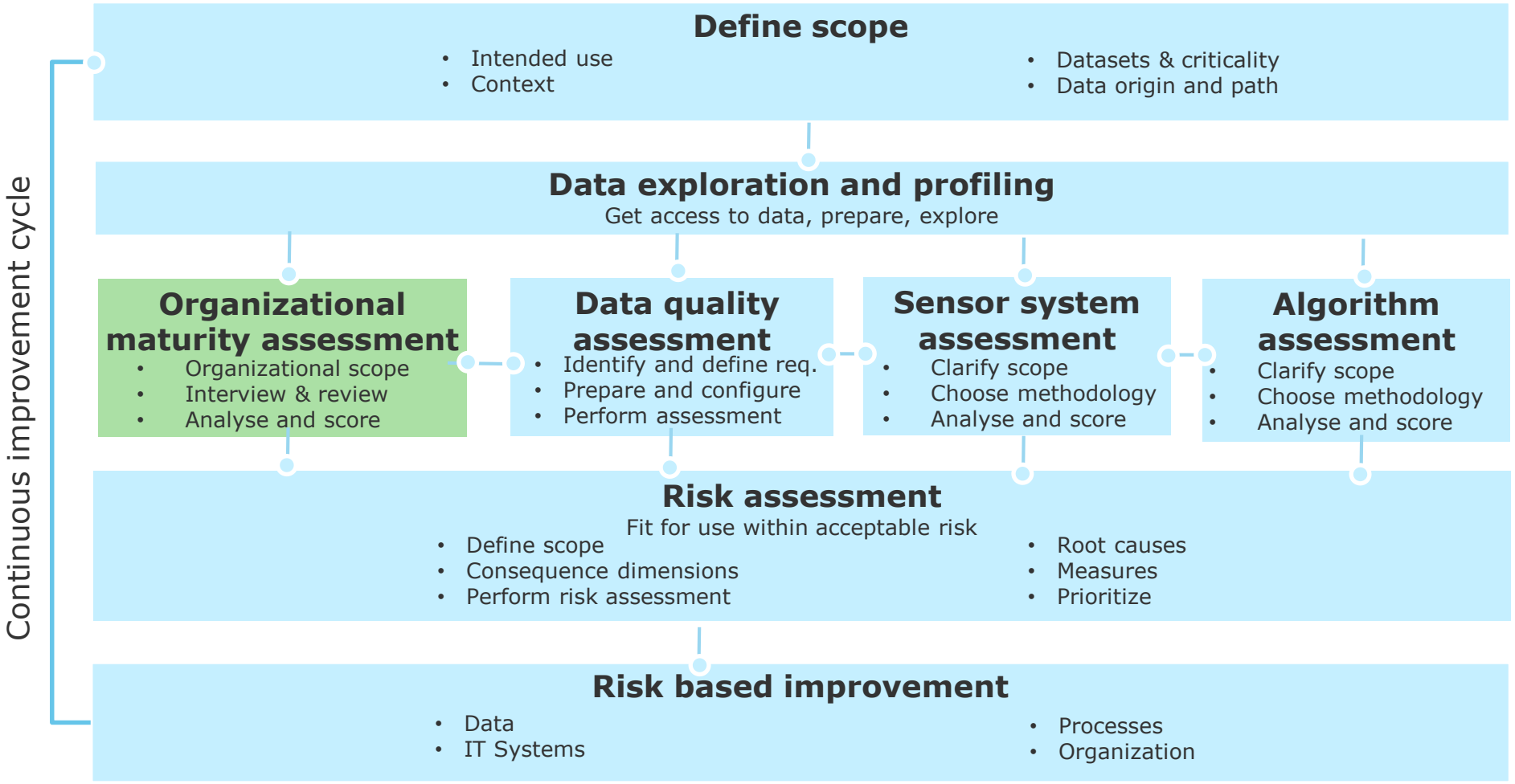
Kjell Eriksson
27 February 2020

SAFER, SMARTER, GREENER

Framework for trustworthy results

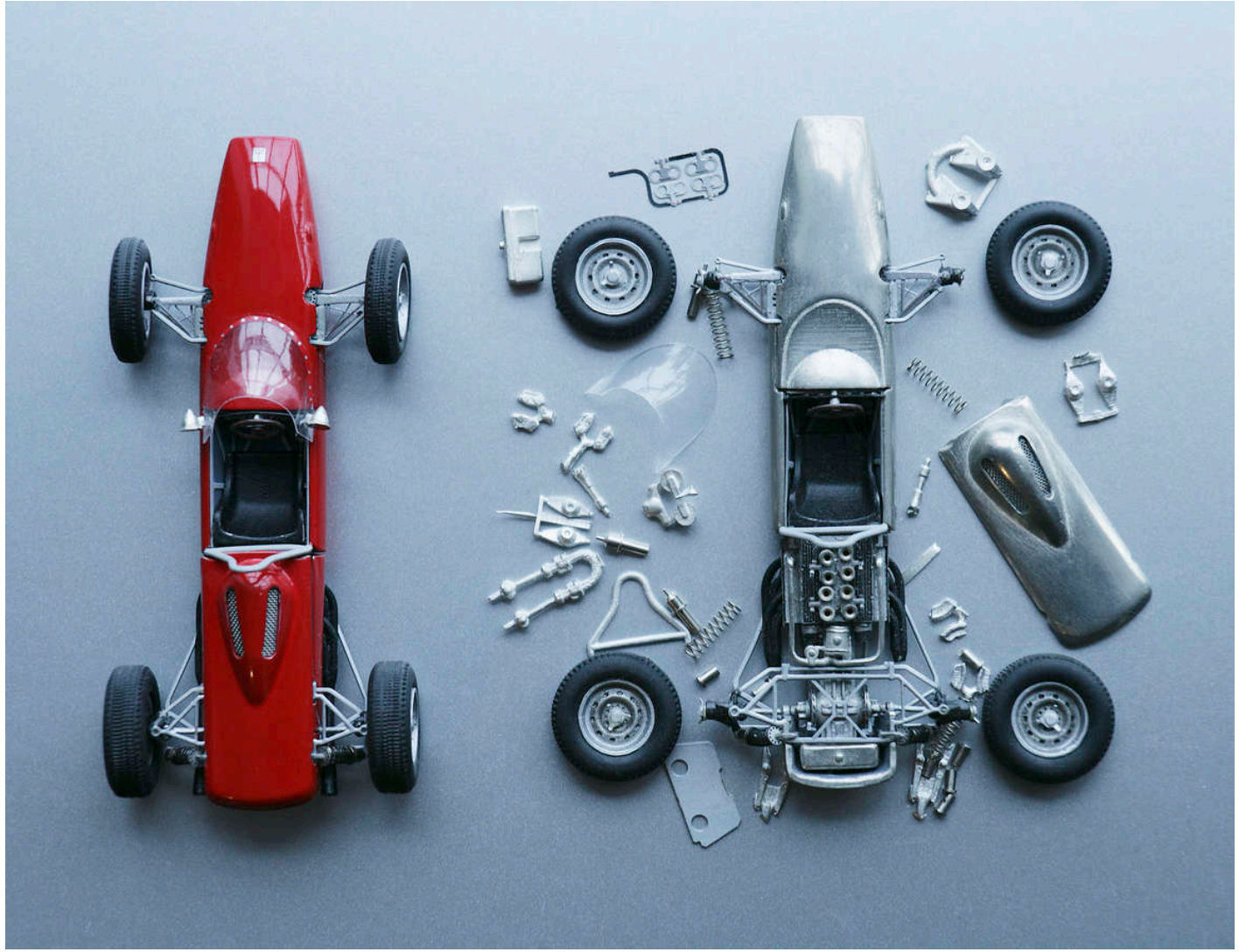


Key building blocks towards trustworthy results

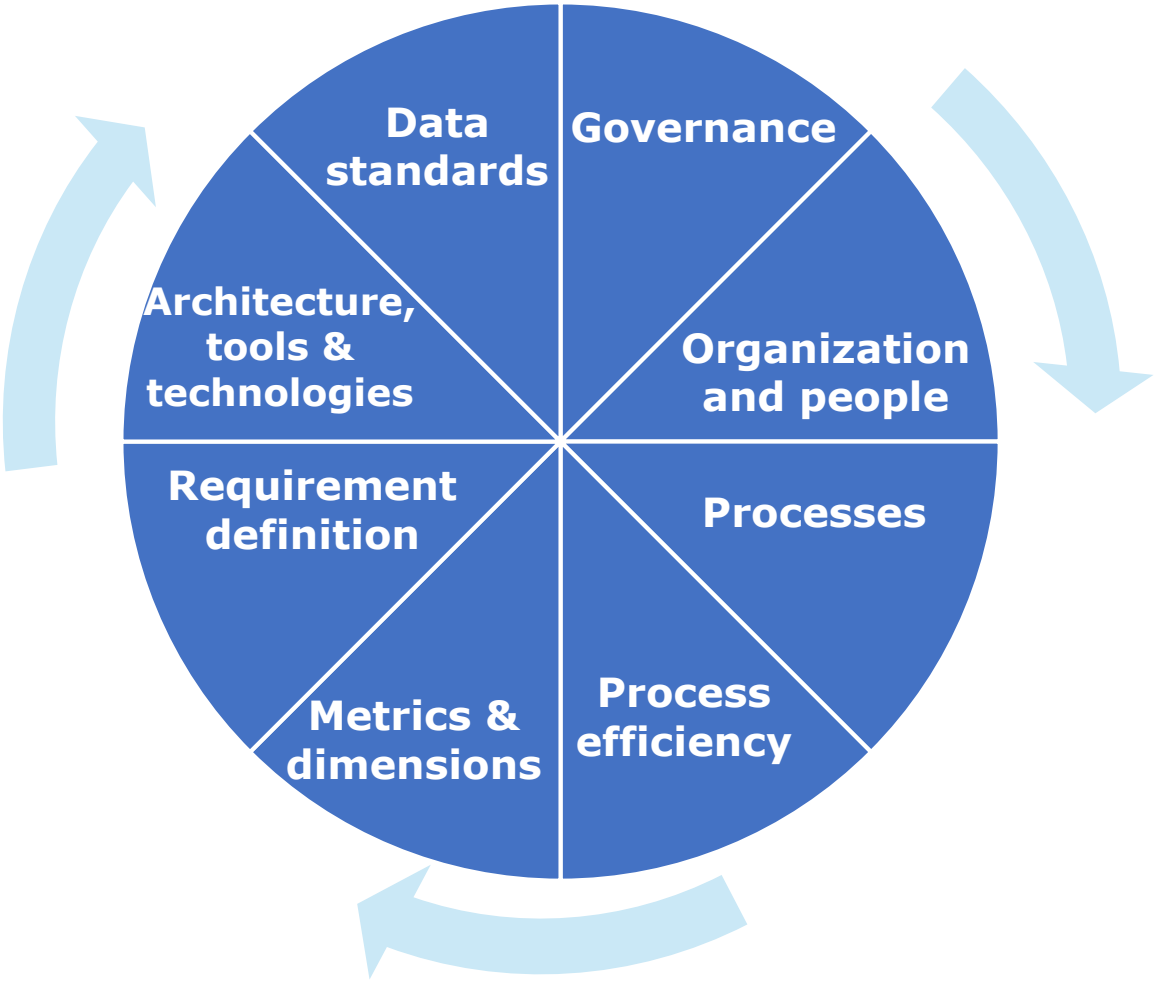


Putting the pieces together – to make quality data

- Business value
- Processes
- Data quality monitoring
- Information Systems
- People
- Knowledge
- Computers
- Networks
- Storage
- Metadata
- Reference Data
- Regulations
- Ownership, Rights
- Data lineage



The “engine” in successful execution is a mature organisation



Level	Maturity
5	Optimized
4	Managed
3	Defined
2	Repeatable
1	Initial

Organizational Maturity Assessment (DNVGL-RP-0497)

Maturity level	Governance	Organization and people	Processes	Process Efficiency	Requirement definition	Metrics and dimensions	Architecture, tools and technologies	Data standards
LEVEL 5 - Optimized	Data management policies governs and drives improvements	Data management board oversees improvement activities	Processes for continuous improvement in place	Processes provides feed-back and feed-forward to support continuous improvement	Baseline established and improvements measured according to requirements	Metrics defines baseline to support continuous improvement	Tools support policy driven continuous improvement cycle	Standard compliance and domain models are subject to continuous improvement
LEVEL 4 – Managed	Policies defined in relation to business objectives	Skillset extended to include risk analysis of quality issues aligned with business objectives	Processes for impact analysis and risk mgmt. in place	Monitoring is performed across enterprise and published as KPI's and trends	Requirements are linked to business impacts	Metrics are linked to business impacts and risk analysis	Tools are driven by business objectives and include support for root cause analysis and risk mgmt.	Standards are used actively to reduce risk for critical business operations
LEVEL 3 – Defined	Policies defined at enterprise level	Roles and required skills defined at enterprise level	Processes are defined and implemented consistently across enterprise	Defined metrics are monitored in advance of business impact	Requirements defined and communicated at enterprise level	Framework for metrics and dimensions defined at enterprise level	Architecture in place at enterprise level supporting full stack data management	Standards, domain models and semantics used at enterprise level
LEVEL 2 – Repeatable	Local initiatives address the requirement for policies	Locally defined roles and some basic skills	Best practices in place but not used consistently	Generic metrics are monitored at point of impact	Local initiatives define requirements	Metrics are reused locally in projects	Tools and technologies used consistently in selected projects	Industry standards and domain models used selectively across projects
LEVEL 1 - Initial	Only ad-hoc or temporal policies in place	No formally defined roles or skillset	Ad-hoc or reactive responses to quality issues	No baseline and no monitoring of quality issues	Re-engineering used to derive requirements	Project specific metrics	Tools are used ad-hoc per project	Ad-hoc and inconsistent use of standards
Objectives,	Policy, Culture, Awareness, Risks, Capabilities to handle DQ issues	Organization, roles, responsibilities, authority, skillsets	Structured and vetted ways of handling and preventing DQ issues	Measure, monitor and use metrics to mitigate DQ issues	DQ Requirements defined, communicated and acted upon	DQ metrics defined, setup, measured and monitored	DQ Tools for processing, analysing and correcting DQ issues with data assets	Use available standards, models, ontologies and taxonomies – a corporate «DQ language»

Approach and benefits

HOW

A gap-analysis is performed to define the improvements required to close the gap between the baseline and the needed maturity level.

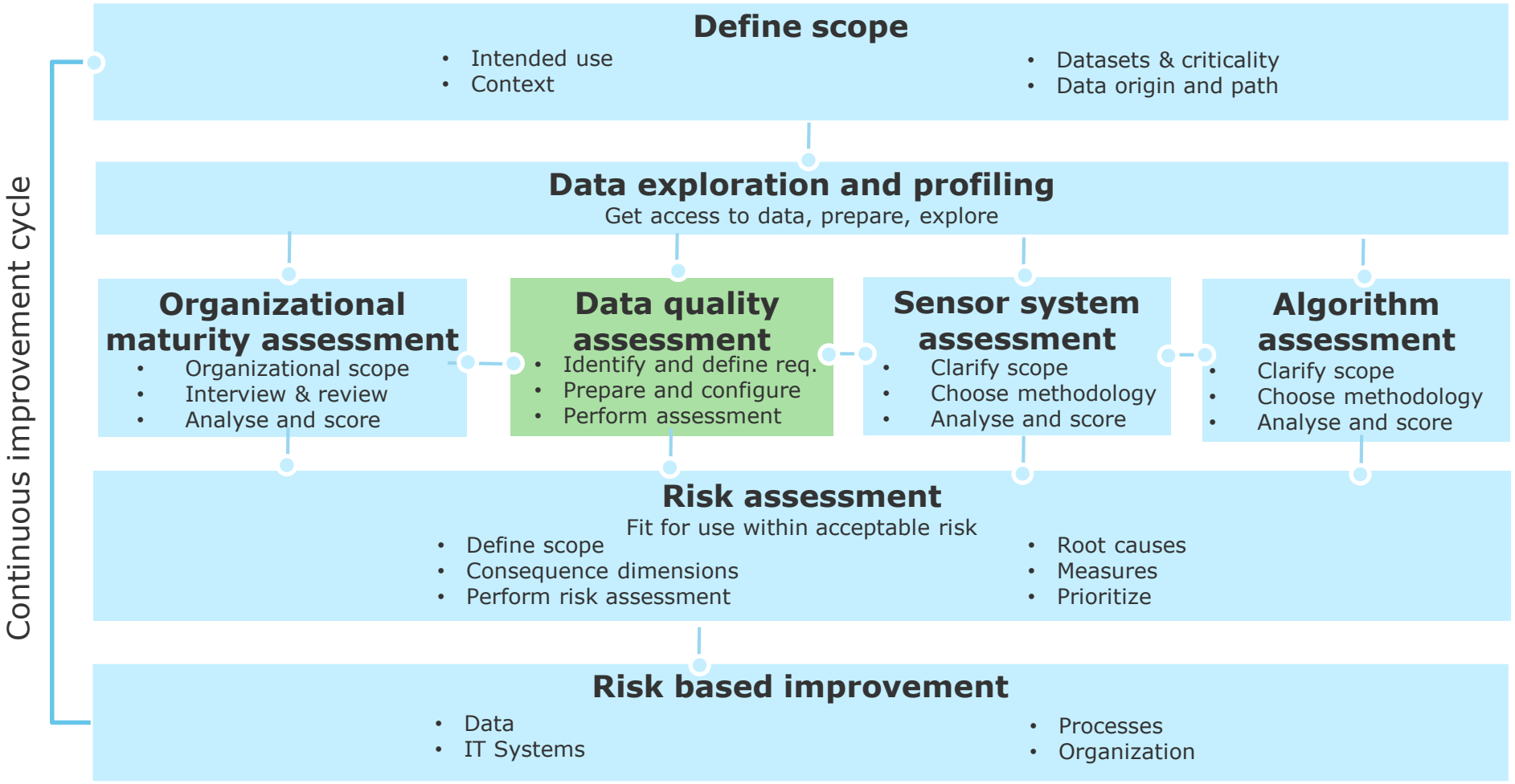
BENEFITS

Understanding and actions on how the organization should:

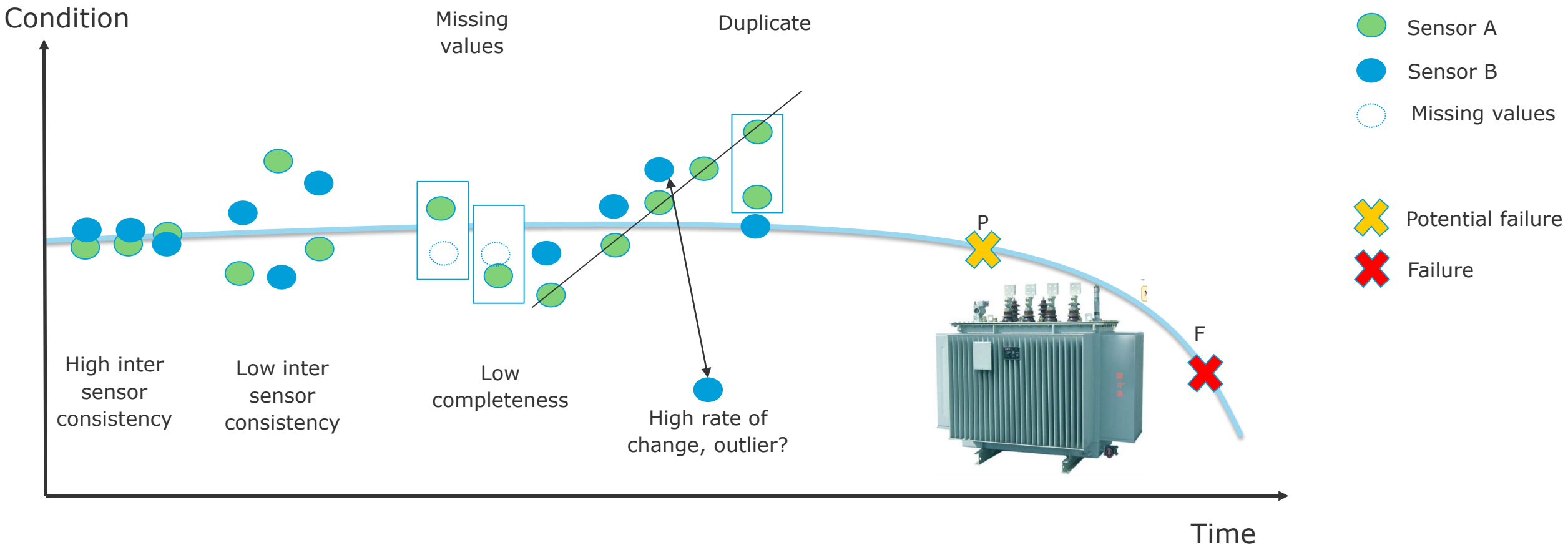
- **capture**
- **manage**
- **use**
- **share**

data without compromising the usefulness of the data or introducing new risks.

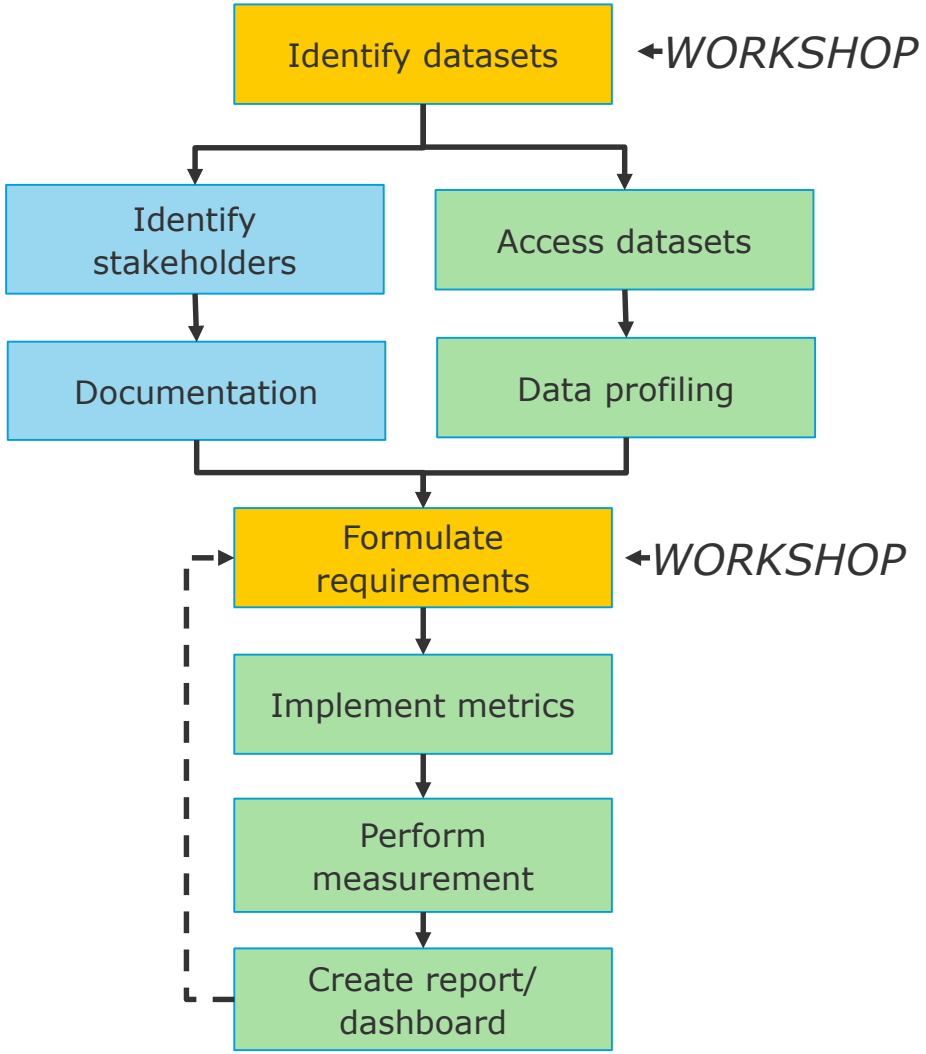
Key building blocks towards trustworthy results



Data quality. Precision, trueness, completeness and duplication



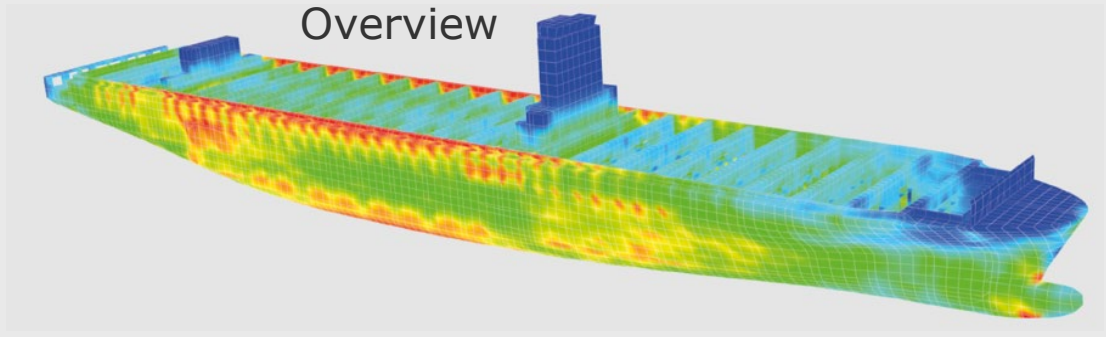
Data quality assessment process and the way it is performed

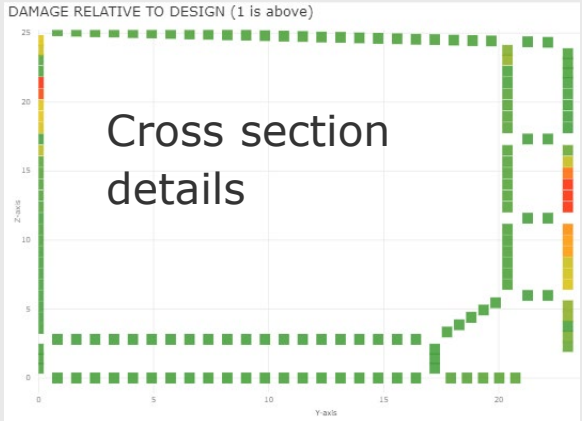


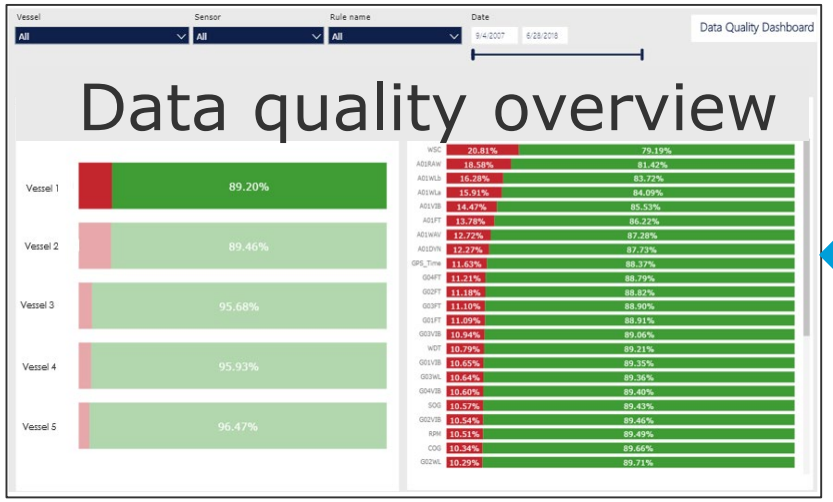
Example

Target inspections of a ship structure based on actual ship movements

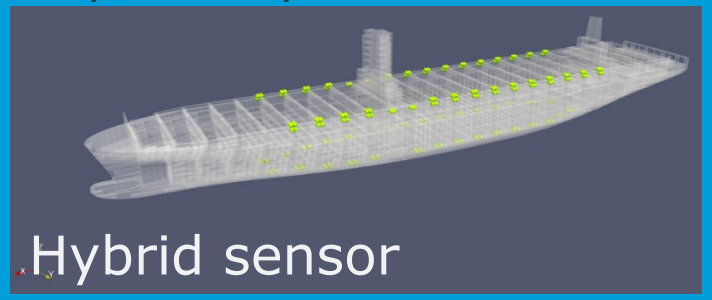
Design models



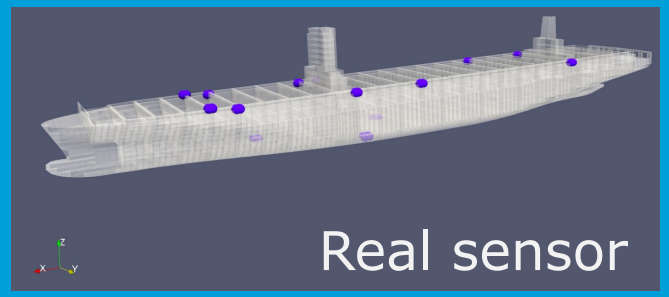


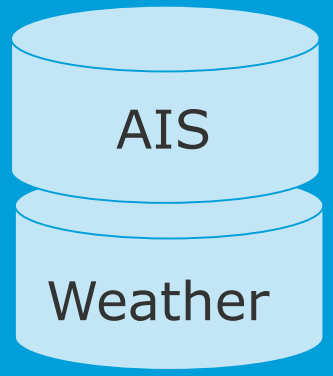


Inspection points

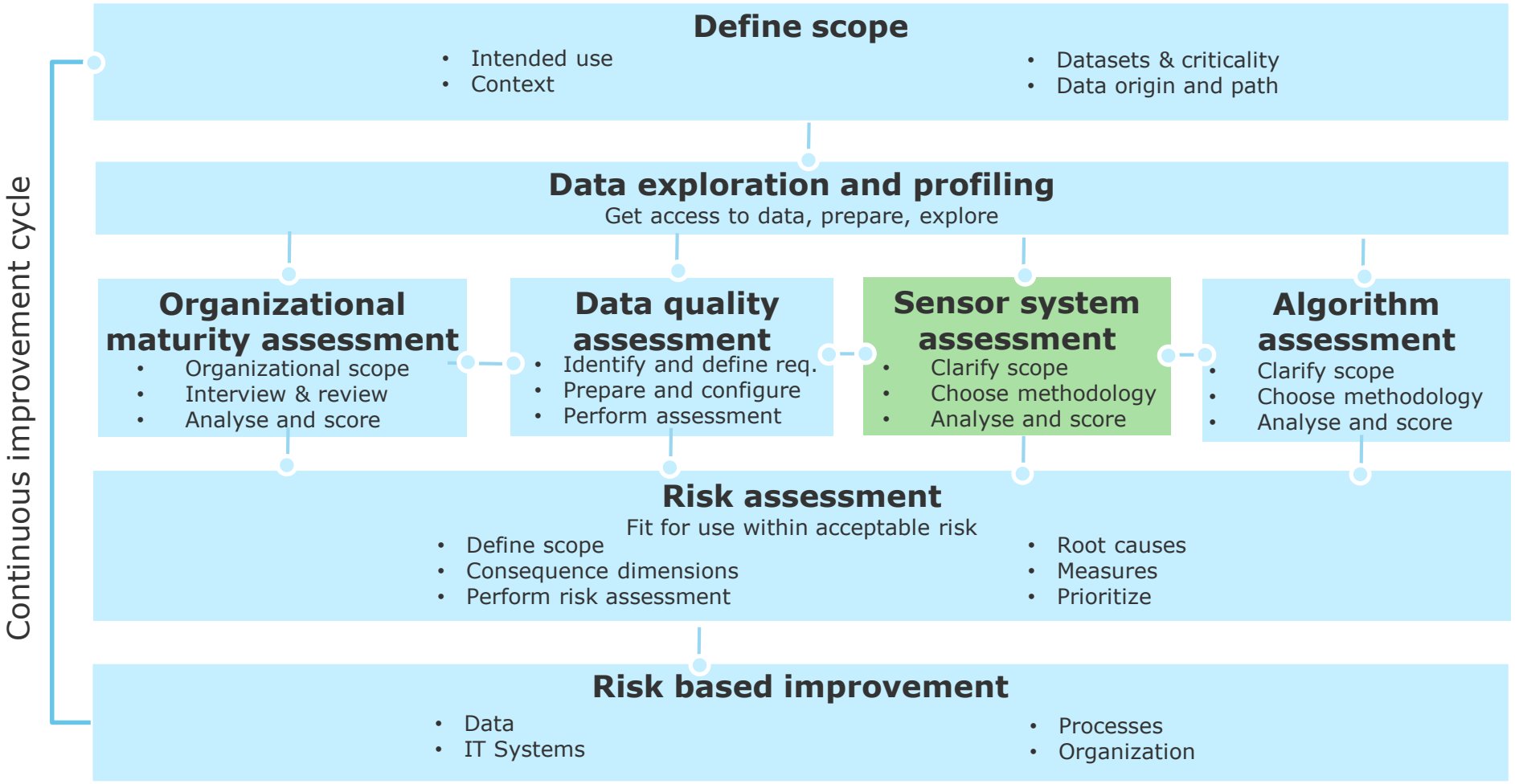


Installation points

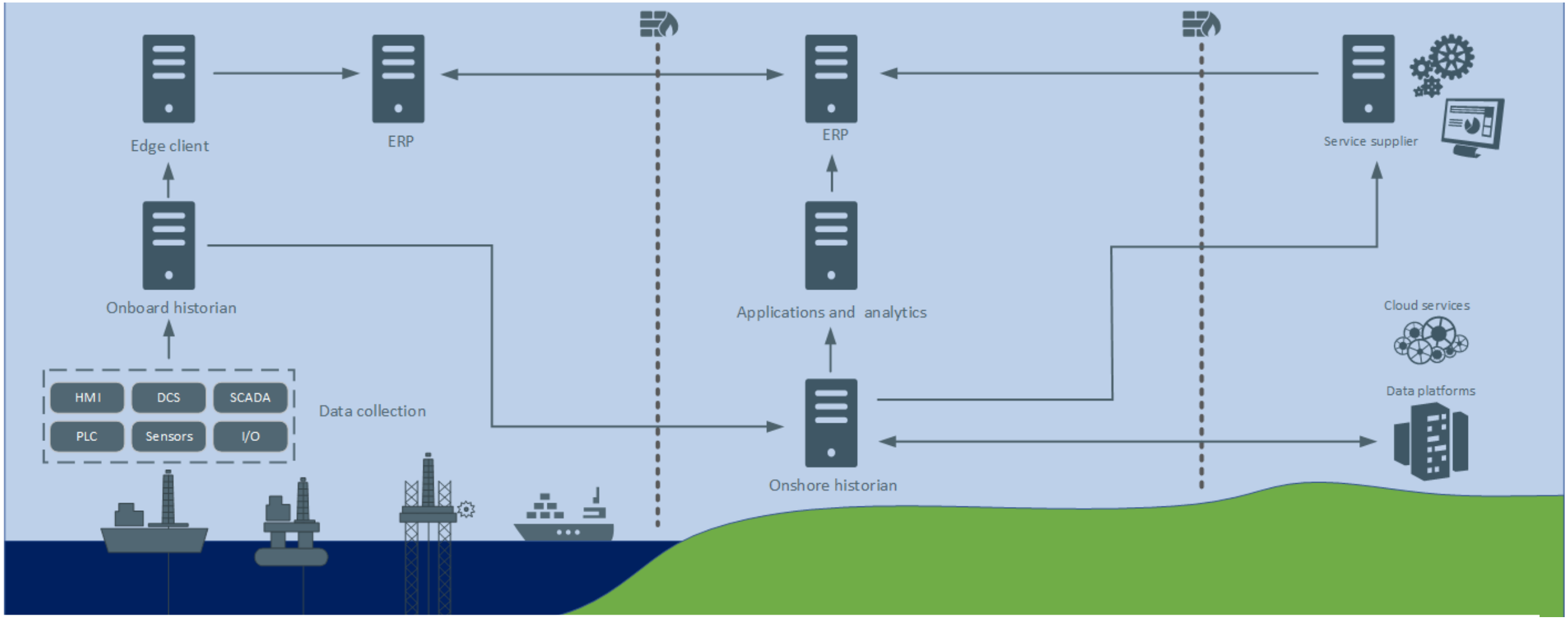




Key building blocks towards trustworthy results



Interoperability, transparency and trust in data value chains



Sensor system data quality

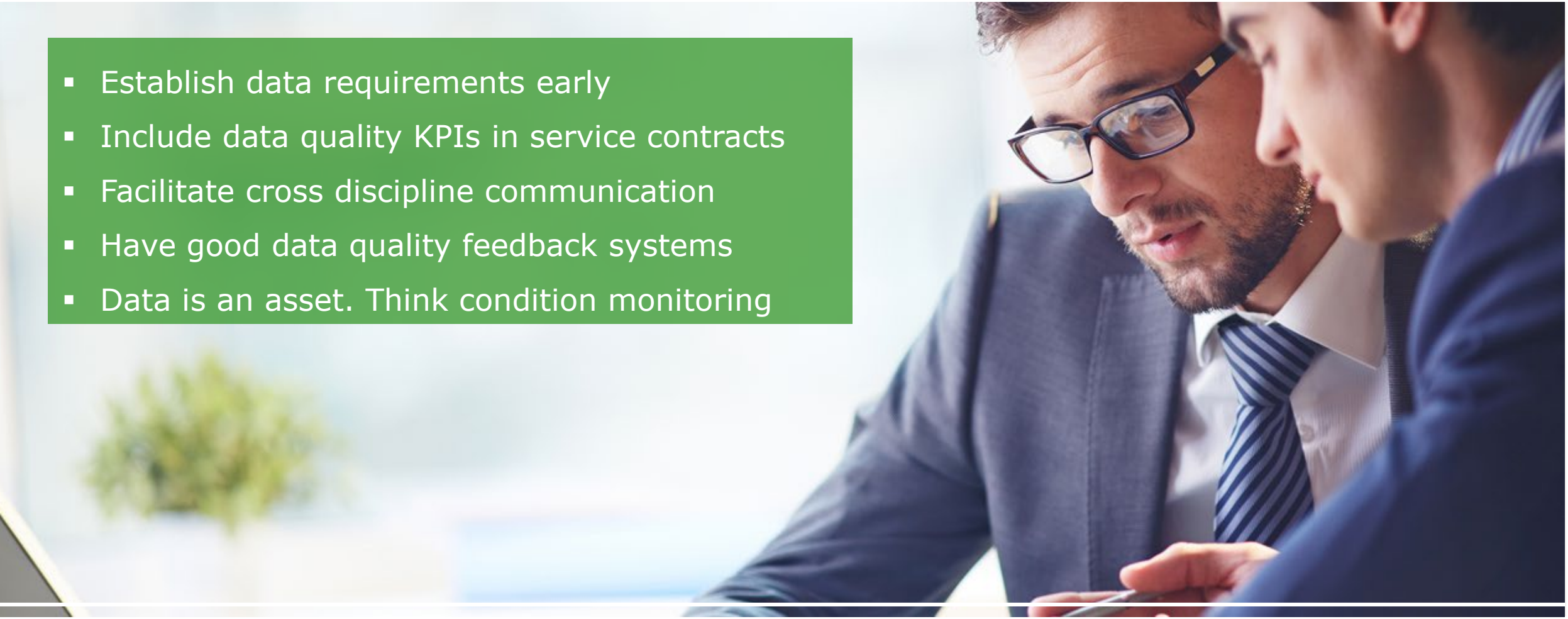
Governance	Organization/people	Processes	Efficiency	Requirement definition	Metrics	Tools and technologies		Data standards
Management focus	End user training	End user work processes	Historization pipeline extent	Legacy system debt	End user quality feedback	Calibration	Sufficient instrumentation	Plant model synchronization and scaling
	System owner training	System owner work processes	Aggregation or sampling freq. loss	Maintenance and service contracts	Heartbeat signals	Structured I/O federation (DCS)	Sufficient granularity of system flags	Meta data synchronization across layers
	Infrastr. and MES training	Infrastr. & MES system work processes	Latency prevention	Upgrade and refitting practices	Asset management systems	I/O Control module load	Sufficient HW/server capacity	Inconsistent data transfer standards and schemas
	ICSS training	ICSS work processes	Design redundancy	Initial planning and development practices	Alarm monitoring and statistics	I/O Control Loop Quality	Bandwidth	Single point of truth
	Instrument training	Instrument work processes	Dynamic and flexible design		Infrastructure service monitoring	Intrinsic instrumentation qualities		
			Graceful buffering and recovery			Time synch-ronization		

Semantic
Syntactic
Pragmatic

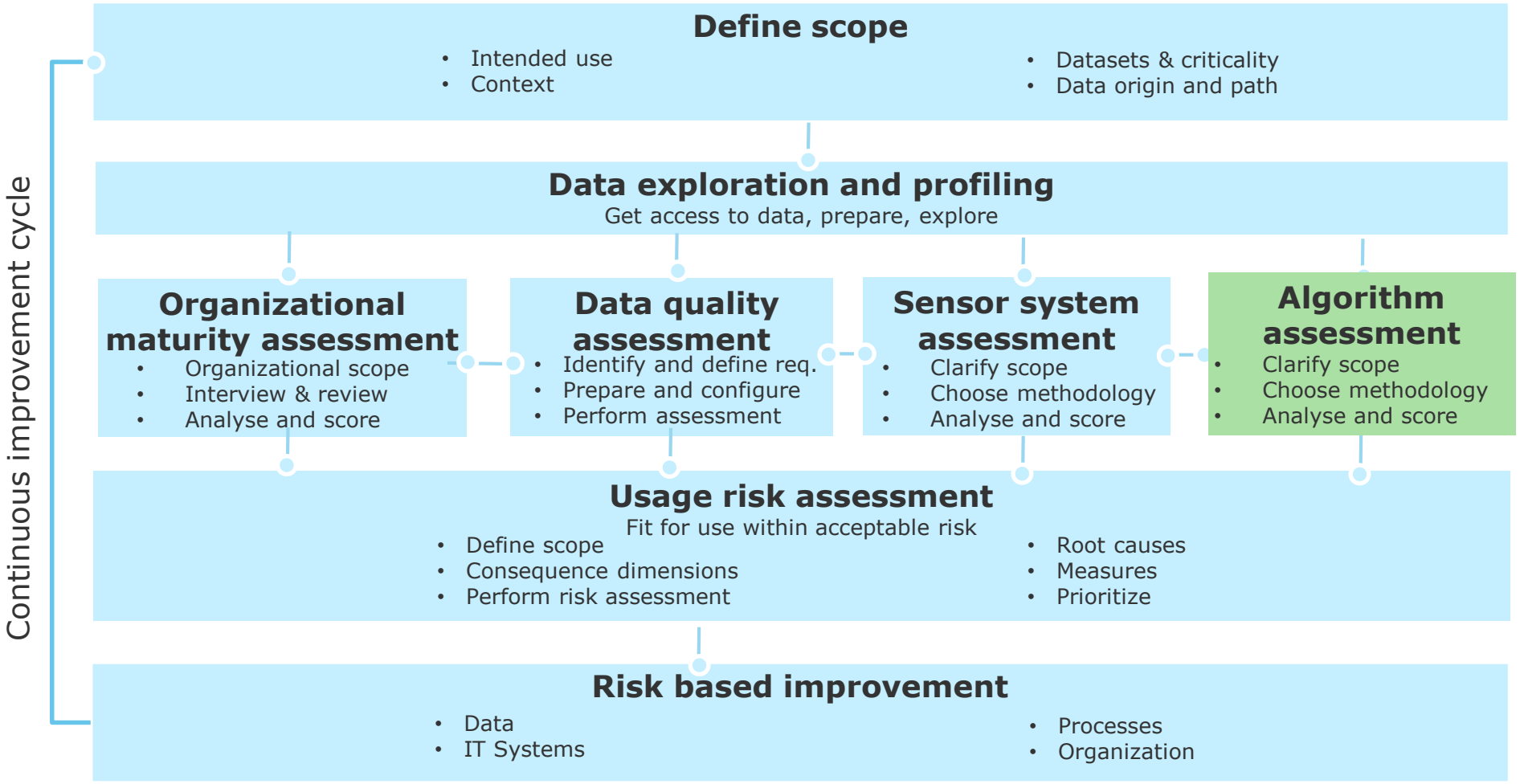
Completeness

Recommendations

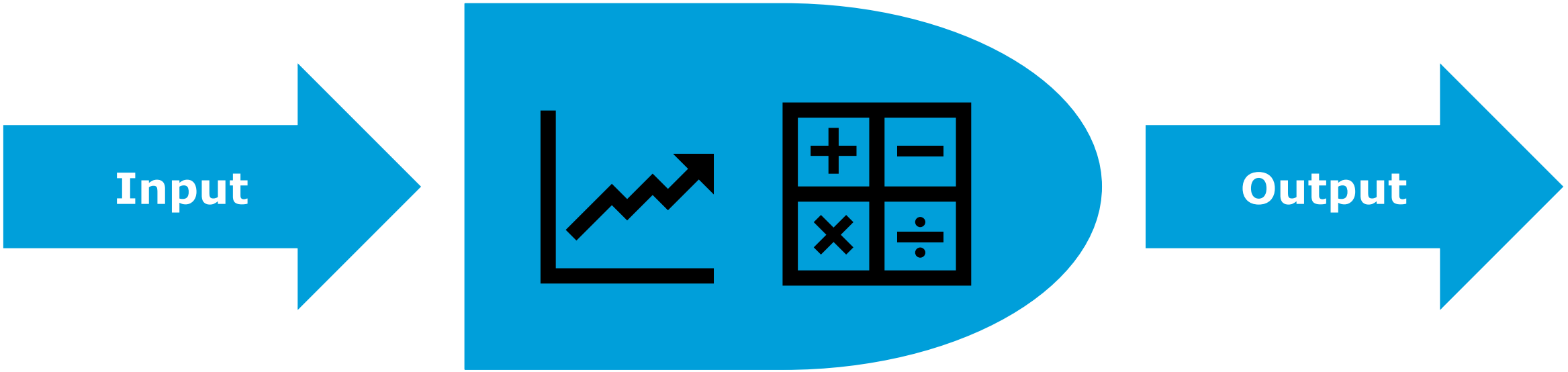
- Establish data requirements early
- Include data quality KPIs in service contracts
- Facilitate cross discipline communication
- Have good data quality feedback systems
- Data is an asset. Think condition monitoring



Key building blocks towards trustworthy results



Building and assuring conventional algorithms



Machine learning is easy to do …

data!

platforms!

tools!

communities!

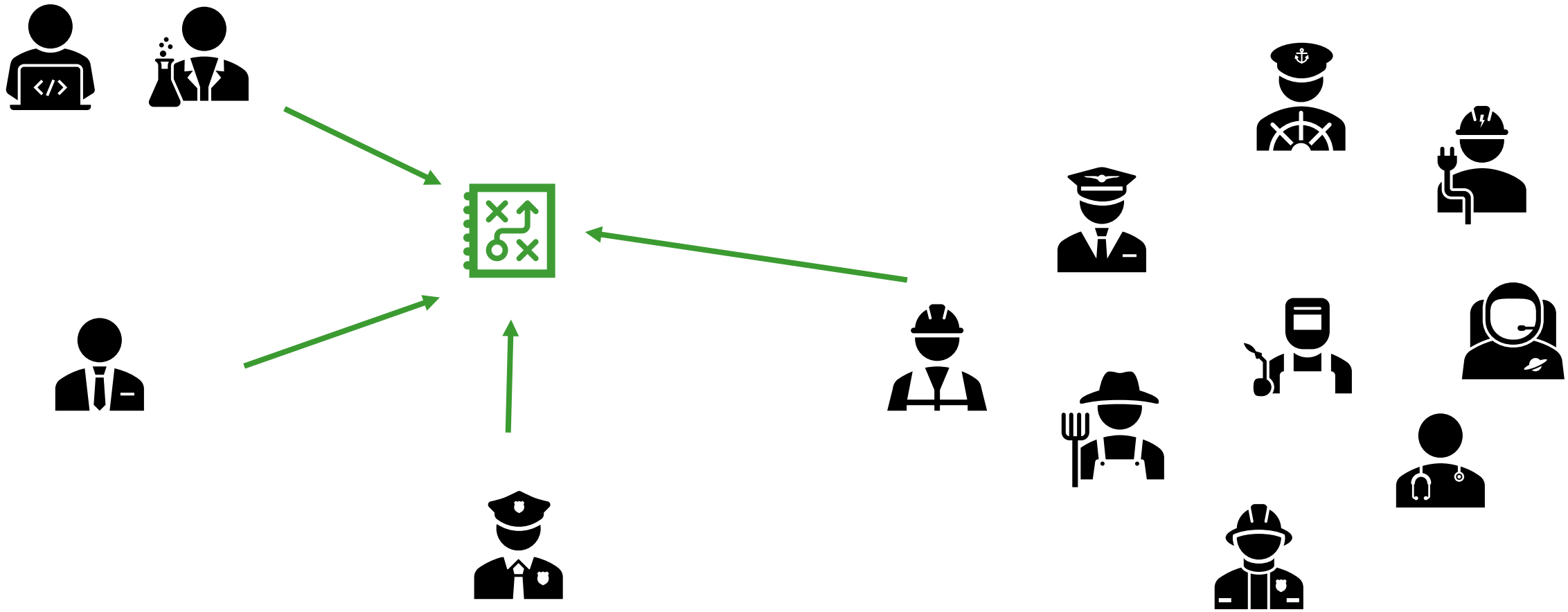
… but Machine learning is hard to do *right*

pitfalls

is it going to work?

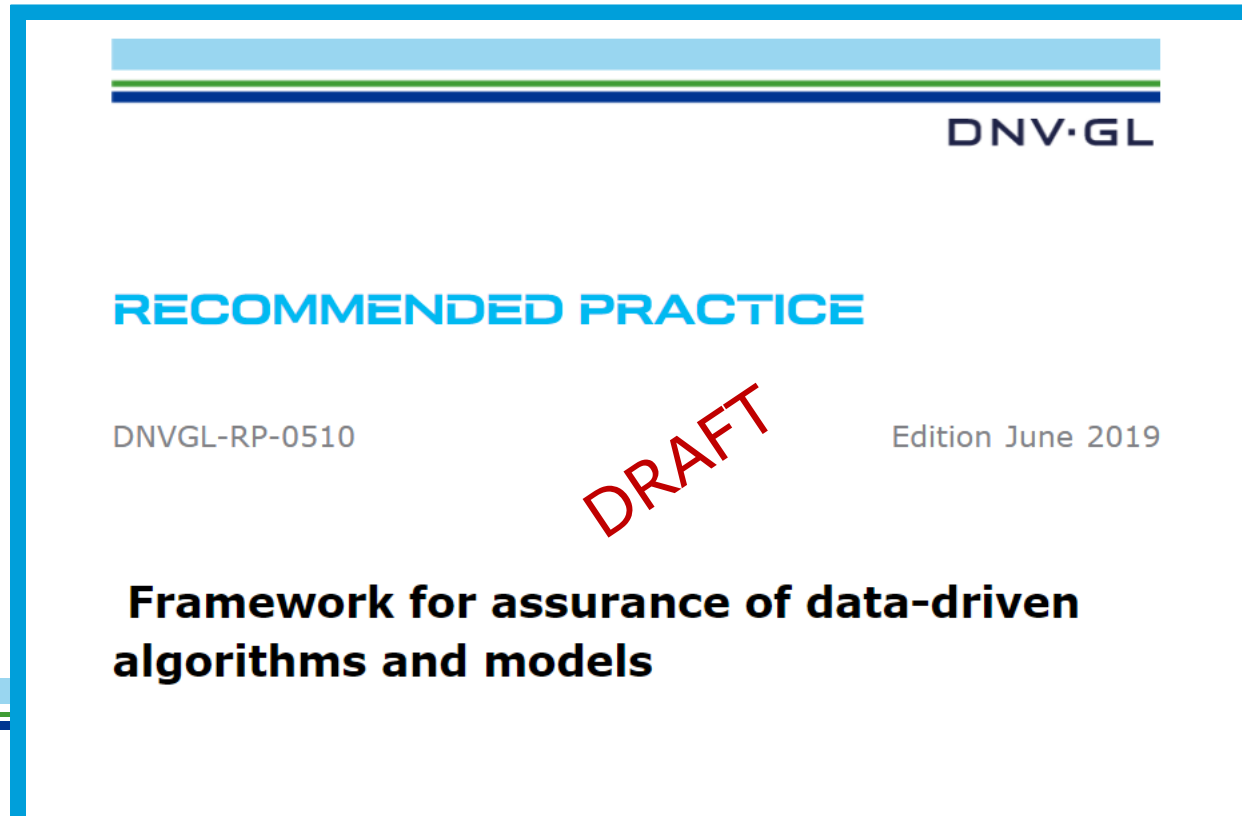
hard to communicate

What is needed is a way to establish *trust* in ML



DNVGL-RP-0510
Framework for assurance of data-driven algorithms and models

- To be published in Q2 2020
  - complements our other RPs on data quality (0497) and guidelines on sensor systems
  - defines the *assurance case*

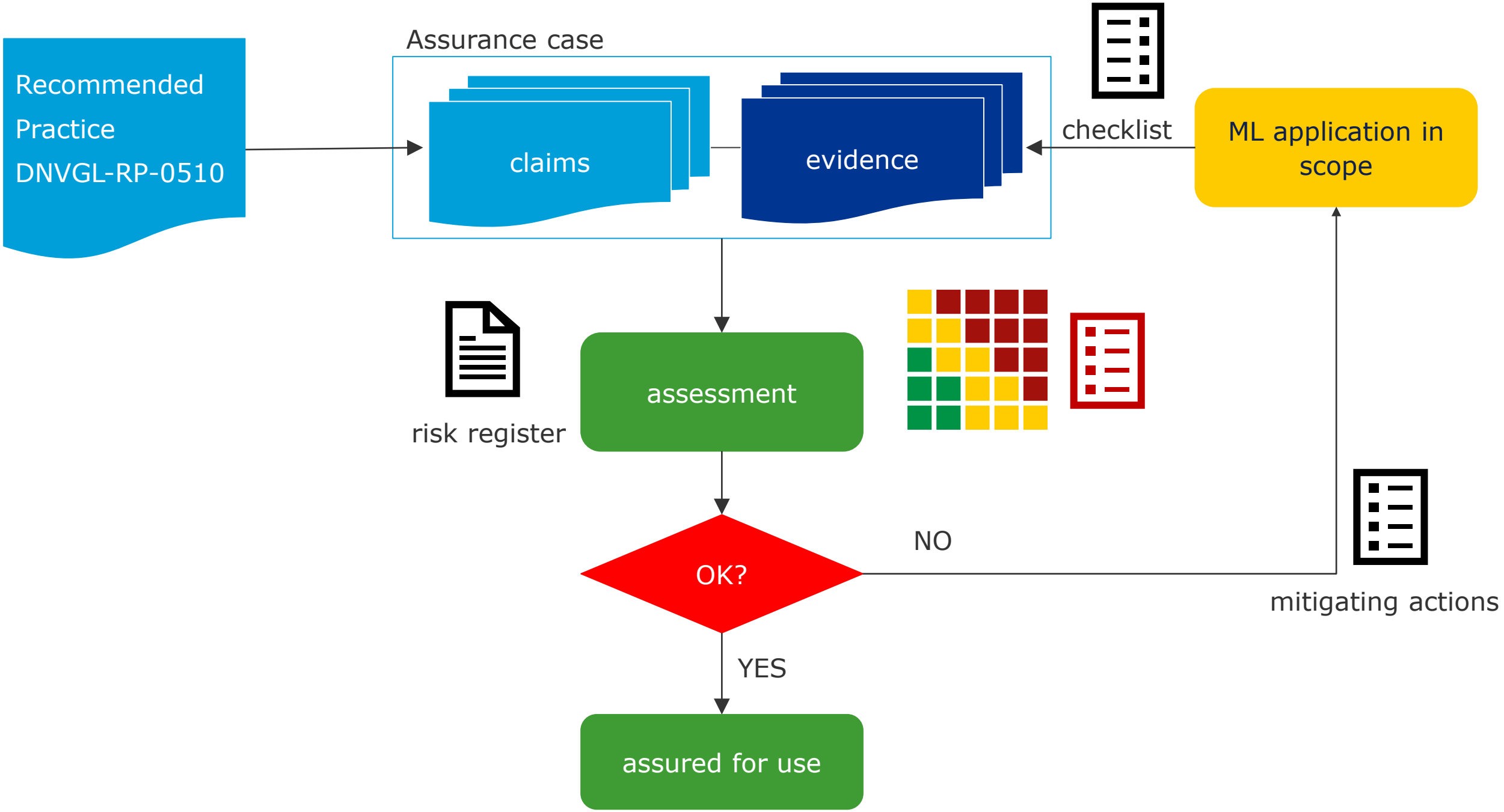

DNV·GL

RECOMMENDED PRACTICE

DNVGL-RP-0510 Edition June 2019

DRAFT

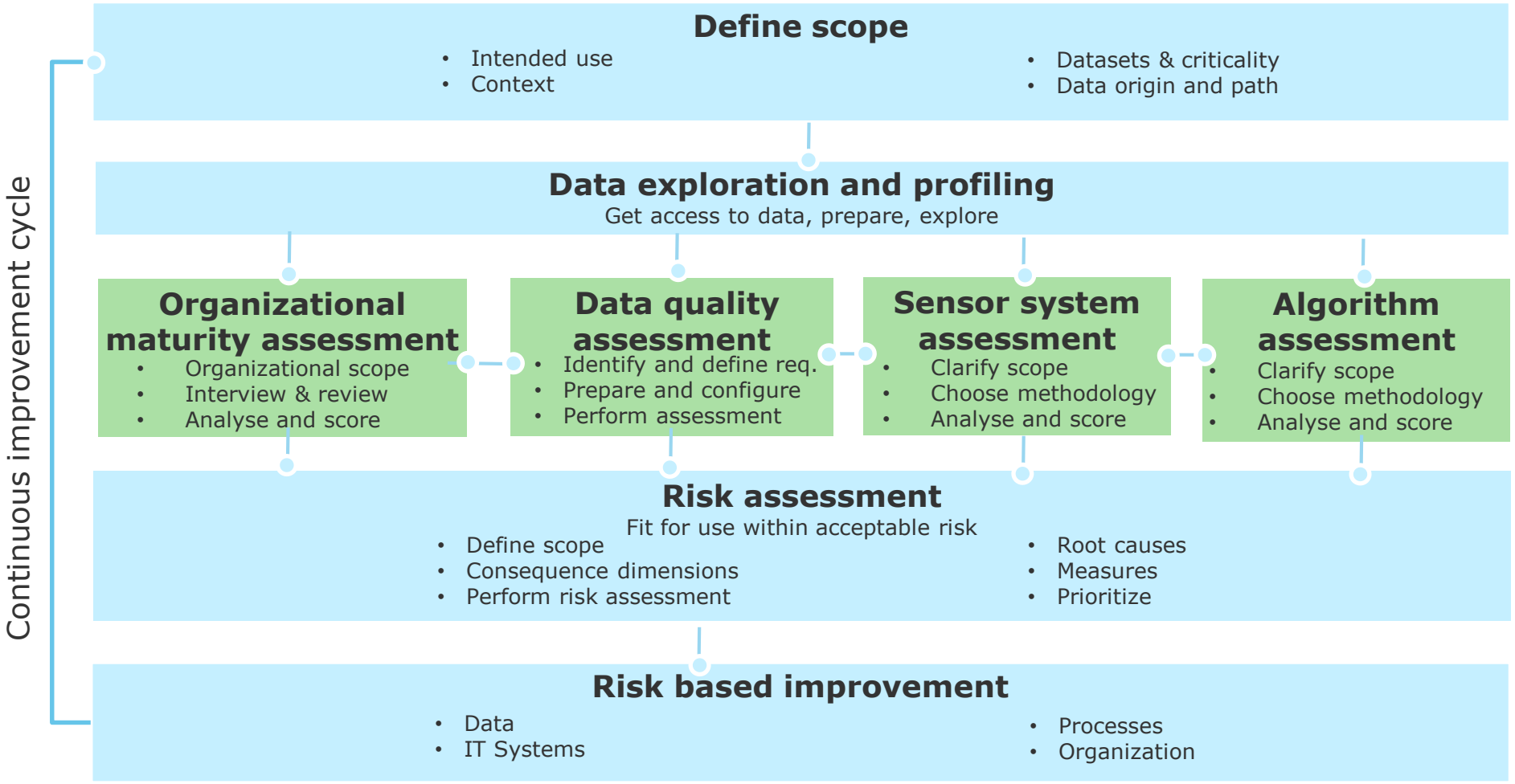
Framework for assurance of data-driven algorithms and models


Recommended Practice DNVGL-RP-0510
Assurance case
claims
evidence
checklist
ML application in scope
risk register
assessment
OK?
NO
YES
mitigating actions
assured for use

Experience from applying the assurance process

- Clearly defined and documented data quality is a prerequisite for enabling trust in machine learning
- Some of it is just good project management
- Enables better communication between data scientists and the other stakeholders
- Works best if introduced at the start of the project, not at the end

Key building blocks towards trustworthy results



Thanks for your attention

Kjell Eriksson
Kjell.Eriksson@dnvgl.com
+47 954 68 338

Per Myrseth
Per.Myrseth@dnvgl.com
+47 418 55 885

Justin Fackrell
Justin Fackrell@dnvgl.com
+47 467 87 123

www.dnvgl.com

SAFER, SMARTER, GREENER

The trademarks DNV GL®, DNV®, the Horizon Graphic and Det Norske Veritas®
are the properties of companies in the Det Norske Veritas group. All rights reserved.