ManagementSolutions
Making things happen

Report on Big Data and Advanced Analytics

European Banking Authority (EBA)

Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AML/CFT	Anti-Money Laundering/Countering the financing of terrorism
API	Application Programming Interface
BD&AA	Big Data and Advanced Analytics
EBA	European Banking Authority
ECB	European Central Bank
ESAs	European Supervisory Authorities
EU	European Union
FinTech	Financial Technology
GDPR	General Data Protection Regulation
ICT	Information and Communication Technology
ML	Machine Learning
NIST	US National Institute of Standards and Technology
NLP	Natural Language Processing
RegTech	Regulatory Technology
SupTech	Supervisory Technology

Index

Context

Key pillars

Elements of trust in BD&AA

Key observations, risks and opportunities

Annex

1 Context

Background and objectives

The EBA's interest in financial innovation and FinTech is motivated by the growing significance in the use of Big Data Analytics by entities across EU

In March 2018, the European Supervisory Authorities (ESAs) issued a Joint Committee **Final Report on Big Data** and the EBA published its **Roadmap on FinTech**. In both documents the ESAs tackled the issues about data-driven approach emerging across the banking sector, which are affecting banks' business strategies, risks, technology and operations. Furthermore, through the use of Big Data and Advanced Analytics (BD&AA) techniques, institutions are exploring more efficient ways to save costs and ensure regulatory compliance, as well as for the calculation of regulatory capital requirements.

In this context, the EBA has published the **Report on Big Data and Advanced Analytics**, with the aim of sharing knowledge among stakeholders on the current use of BD&AA by providing useful background on this area, along with key observations, and presenting the key pillars and elements of trust that could accompany their use. This report focuses on BD&AA techniques and tools, such as machine learning (ML), that go beyond traditional business intelligence to gain deeper insights, make predictions or generate recommendations using various types of data from different sources.

Objectives

- This report provides **background information on BD&AA**, along with an educational perspective, and describes the current landscape as regards their use in the banking sector.
- It aims to share knowledge about and enhance understanding of the **practical use of BD&AA**, noting the **risks and challenges** currently arising from the implementation of such solutions, such as the integration and coordination of institutions' legacy infrastructures with new big data technologies.
- This report covers **four key pillars** (i.e. data management, technological infrastructure, organization and governance & analytics methodology), **elements of trust** (e.g. ethics and security) and **key observations, risks and opportunities**.

This **Technical Note** summarizes the main aspects included in the Report on Big Data and Advanced Analytics (EBA/REP/2020/01).

1 Context

Current landscape

The EBA has made a number of observations in the area of BD&AA that are relevant to the financial sector

Level of BD&AA adoption

Institutions are currently developing or implementing **digital transformation programmes**, which include the growing use of advanced analytics across business functions.

Use of cloud services

Greater use of BD&AA is perceived to be facilitated by the **use of cloud services**, which promises high levels of availability, scalability and security.

Leveraging data sources

To support BD&AA applications, some institutions are exploring the **use of algorithms and ML models** available from open-source libraries.

Human involvement and explainability

The involvement of humans is required to make decisions based on ad**vanced analytics-related techniques**. Furthermore, new skills in data science are required and a gap has appeared between business and IT experts.

Data protection and data sharing

There is an increasing convergence on the adoption of **data protection rules and principles closely mirroring the GDPR model**. The GDPR can be regarded as an opportunity to strengthen customers' trust in how institutions process their data.

Bias detection

Various statistical techniques are explored to help in **detecting bias**, while an iterative approach may help to gradually strengthen models against bias.

Software tools

Institutions frequently use **open-source frameworks** to implement BD&AA solutions. This covers programming languages, code versioning, and big data storage and management.

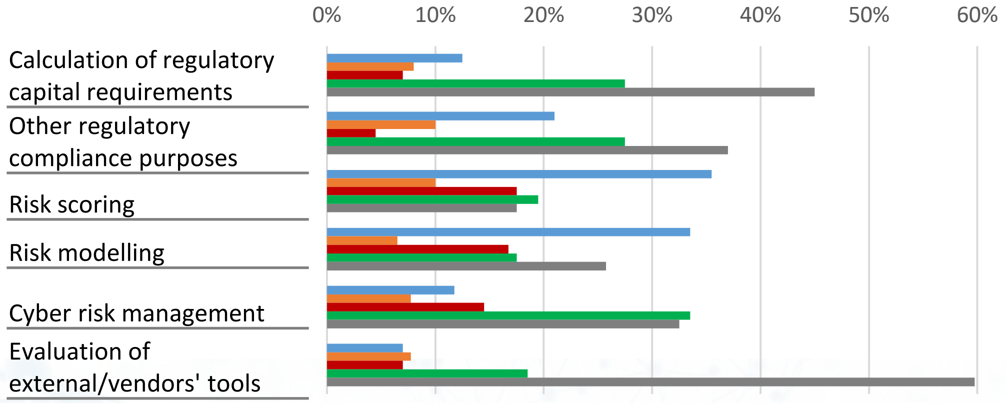
1 Context

Current landscape

BD&AA in financial services may have multiple applications, such as risk management and customer engagement, reflecting data pervasiveness and advanced analytics adaptability

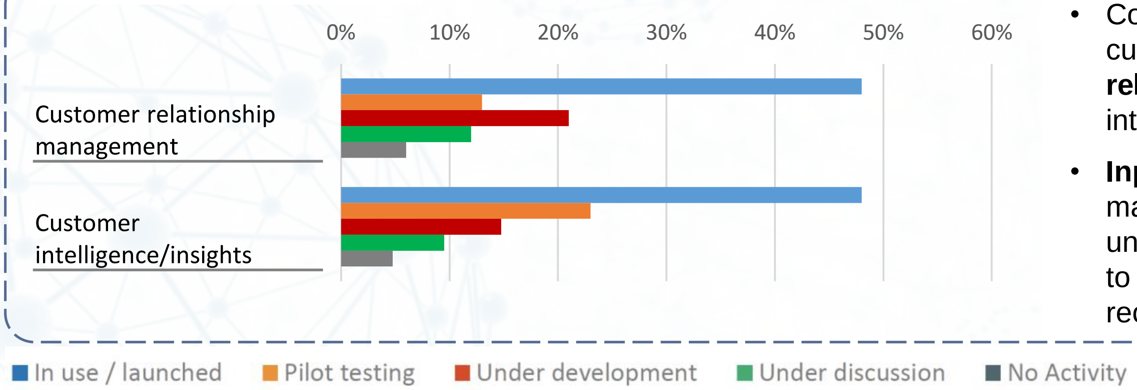
Current use of BD&AA (1/2)

Risk management



- The use of BD&AA tools in the area of risk mitigation appears to be increasing. In particular, growing use was observed for **risk-scoring** and **risk-modelling purposes**.
- Institutions acknowledge the growing focus on **operational risks**, such as **cyber-risk**, and **fraud** and anti-money laundering/countering the financing of terrorism (**AML/CFT**) issues. The use of BD&AA techniques also lead to save costs and ensure compliance.

Customer engagement



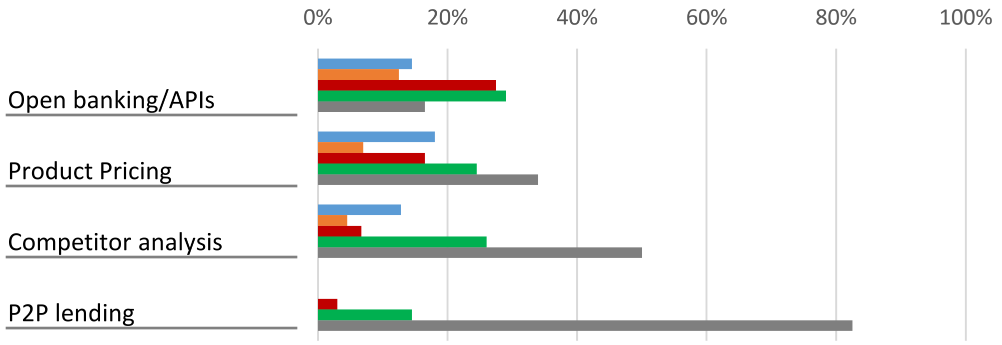
- Considerable use of BD&AA is observed in the area of customer engagement, with a focus on **customer relationship management** as well as improving customer intelligence and gaining better customer insights.
- **Inputs from users** can be processed (e.g. via NLP) and is a major source of data, embracing both structured and unstructured formats, from digital forms to paper documents to be processed for subscription and biometric data for facial recognition provided during digital onboarding.

Source: EBA risk assessment questionnaire (spring 2019)

BD&AA in financial services may have multiple applications, such as product transformation and process optimisation, reflecting data pervasiveness and advanced analytics adaptability

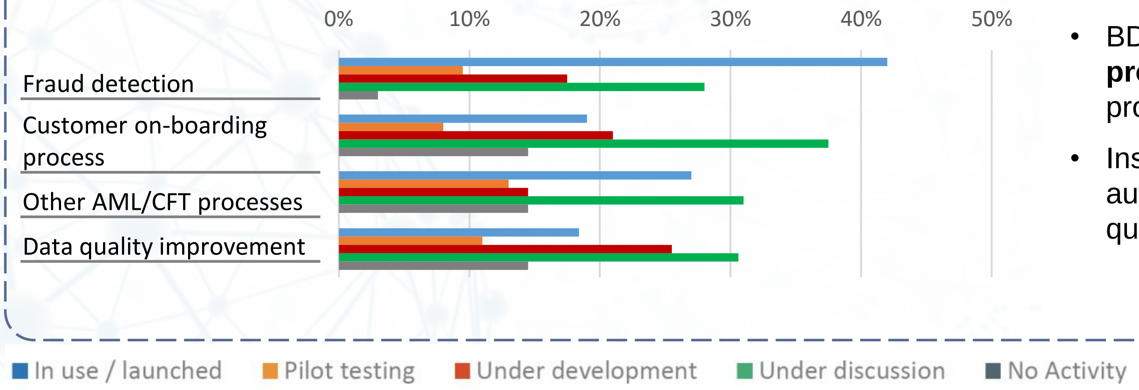
Current use of BD&AA (2/2)

Product transformation



- **Customer insight** is used to develop new business or maintain existing business. It aims to improve **customer understanding** through better customer **segmentation analysis** backed with ad hoc models that allow the use of AA.
- **Customer churn or customer behaviour** is reflected through dedicated analytics fed by customer interaction data. It should be combined with sales analysis, product analysis and network marketing analysis.

Process optimisation



- BD&AA solutions are mostly used in **optimising the process of fraud detection**, as well as other AML/CFT processes.
- Institutions are also exploring the use of BD&AA to automate customer onboarding processes and improve data quality.

Source: EBA risk assessment questionnaire (spring 2019)

Index

Context

Key pillars

Elements of trust in BD&AA

Key observations, risks and opportunities

Annex

2 Key pillars

Introduction

The EBA identifies four key pillars for the development, implementation and adoption of BD&AA: data management, technological infrastructure, organization and governance & analytics methodology

DATA MANAGEMENT

- Data types and data sources: structured, unstructured and semi-estructured data.
- Data security and data protection: protect the confidentiality, integrity and availability of data.
- Data quality: ensure data quality throughout the whole BD&AA lifecycle.

TECHNOLOGICAL INFRASTRUCTURE

- Infrastructure: networking resources to transmit Big Data
- Data platform: manages all the data used by an advanced analytics system
- Processing:provides the necessary software to support the implementation of advanced analytics applications

ANALYTICS METHODOLOGY

- Data preparation: feature engineering and feature importance analysis
- Analytics: model training, tuning, validation and selection, testing and deployment.
- Operations: model monitoring and update.

ORGANISATION AND GOVERNANCE

- Internal governance structures and measures: governance structure, strategy and risk management, transparency, external development and outsourcing.
- Skills and knowledge: level of understanding of management body and senior management, second and third line, developers of advanced analytics-enabled systems.

ManagementSolutions
Making things happen

Index

Context

Key pillars

Elements of trust in BD&AA

Key observations, risks and opportunities

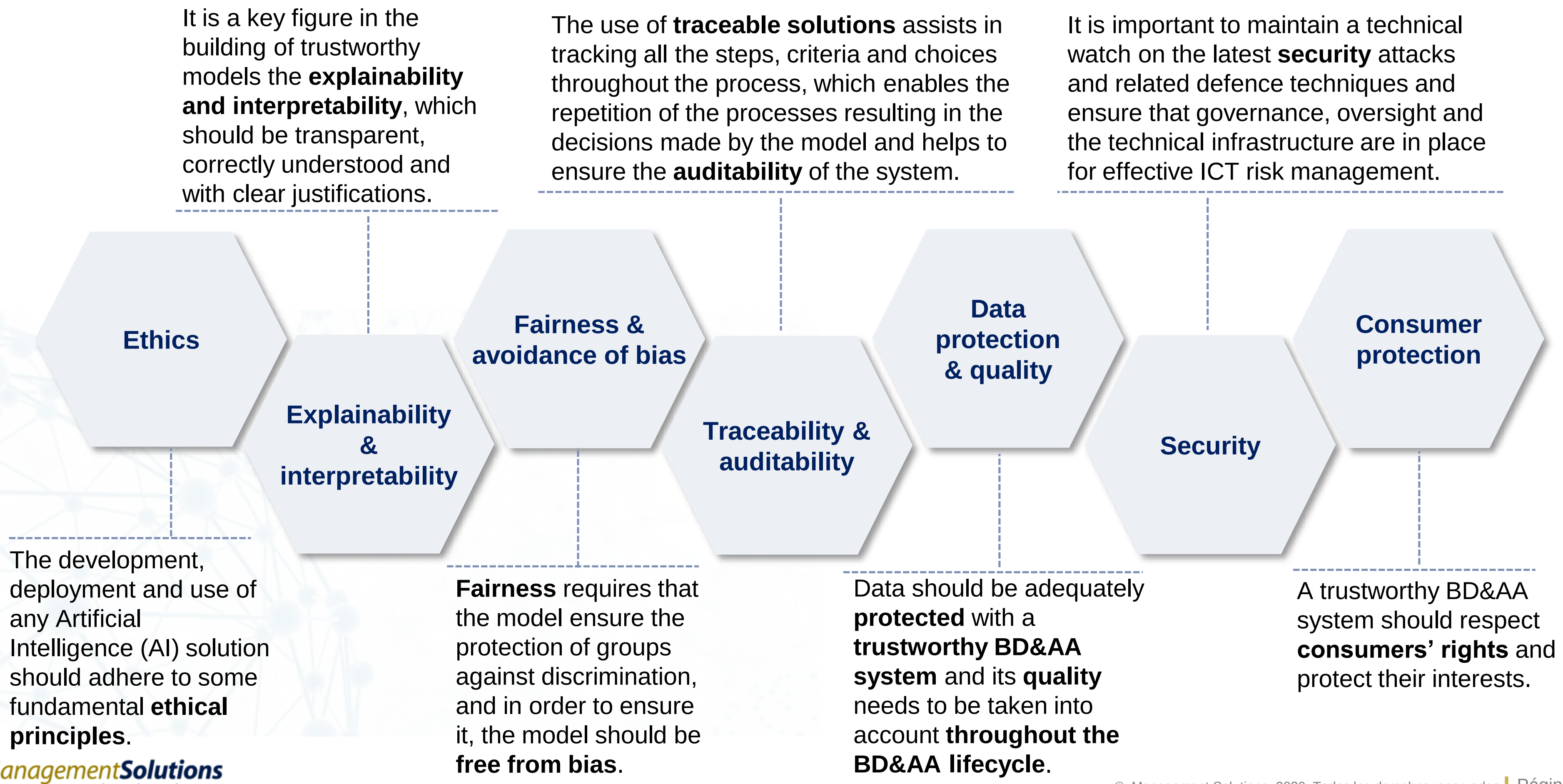
Annex

3 Elements of trust in BD&AA

Introduction

The roll-out of BD&AA affects issues around trustworthiness and shows a number of fundamental trust elements that need to be properly and sufficiently addressed and which cut across the four key pillars

Introduction



Ethics

The development, deployment and use of any Artificial Intelligence (AI) solution should adhere to some fundamental **ethical principles**.

Explainability & interpretability

It is a key figure in the building of trustworthy models the **explainability and interpretability**, which should be transparent, correctly understood and with clear justifications.

Fairness & avoidance of bias

Fairness requires that the model ensure the protection of groups against discrimination, and in order to ensure it, the model should be **free from bias**.

Traceability & auditability

The use of **traceable solutions** assists in tracking all the steps, criteria and choices throughout the process, which enables the repetition of the processes resulting in the decisions made by the model and helps to ensure the **auditability** of the system.

Data protection & quality

Data should be adequately **protected** with a **trustworthy BD&AA system** and its **quality** needs to be taken into account **throughout the BD&AA lifecycle**.

Security

It is important to maintain a technical watch on the latest **security** attacks and related defence techniques and ensure that governance, oversight and the technical infrastructure are in place for effective ICT risk management.

Consumer protection

A trustworthy BD&AA system should respect **consumers' rights** and protect their interests.

The elements of trust to be respected throughout the development, implementation and adoption of BD&AA cover ethics, explainability and interpretability, fairness and avoidance of bias…

(1/2)

Ethics

- **“Ethical by design” approach** to cover some fundamental **ethical principles** such as respect for human autonomy, prevention of harm, fairness and explainability.
- **Ethical policy** in place enforcing the principles and setting the boundaries for acceptable and unacceptable use cases.
- It is highly recommended to set up an “**ethics committee**” or integrate it into an existing similar one to, for example, validate new AI use cases or periodically review fairness metrics from live models.

Explainability and interpretability

- A model is **explainable** when it is possible to generate explanations that allow humans to understand (i) how a result is reached or (ii) on what grounds the result is based.
  i. In the first case the model is interpretable, since the internal behavior (representing how the result is reached) can be **directly understood by human**.
  ii. In the second case, techniques exist to provide **explanations** for the main factors that led to the output.

Fairness and avoidance of bias

- **Fairness** requires that the model ensures the protection of groups against (direct or indirect) **discrimination**. To ensure fairness (non-discrimination), the model should be **free from bias**.
- **Bias** is ‘an **inclination of prejudice** towards or against a person, object, or position. It can be introduced in the input dataset, via the online learning process or algorithmic bias, among others. There are techniques to prevent or detect bias, by setting controls, applying statistical analysis and monitoring to ensure that it has not deviated into discriminatory behaviour.

…traceability and auditability, data protection and data quality, security and consumer protection

(2/2)

Traceability and auditability

- **Traceability**. Steps and choices made throughout the entire data analytics process need to be clear, transparent and traceable to enable its oversight; including, inter alia, model changes, data traceability and decisions made by the model.
- **Auditability**. A traceable solution, for which there are detailed audit logs for all phases of the process that can be used to identify 'who did what, when and why', facilitates oversight of the system, as it makes it possible to follow the whole process and gain better insights.

Data protection and quality

- **Data protection**. Entities should comply with the current regulation on data protection when managing personal data a trustworthy BD&AA system. According to the GDPR, institutions should have a lawful basis for processing personal data.
- **Data quality**. The purchase of data from third parties should be collected complying with the GDPR. The data controller should be accountable for, among other things, the lawful, fair and transparent processing of personal data.

Security

Main types of attack affecting ML:

- **Model stealing/extraction attacks**. Used to 'steal' models by replicating their internal functioning.
- **Poisoning attacks**. Attackers deliberately influence the training data to manipulate the results of a predictive model.
- **Adversarial attack**. Providing a sample of input data that has been slightly perturbed to cause the model to misclassify it.

Consumer protection

- A trustworthy BD&AA system should **respect customers' rights and protect their interests**.
- **Alternative sources of data**. Entities should be aware of how this data is used, because it can lead to **financial inclusion** when customers gain access to financial services that they could not access before, or can lead to **financial exclusion** for some customers unfairly excluded if they do not share the data required or do not have that data at all.

Index

Context

Key pillars

Elements of trust in BD&AA

Key observations, risks and opportunities

Annex

4

Key observations, opportunities and risks

Key observations and opportunities

The EBA outlined the key observations (e.g. different stages of BD&AA development), opportunities (e.g. improvement of customer satisfaction)…

Key observations

- Institutions are at different stages of BD&AA development. Typical use cases found in **fraud detection, CRM and process automation**.
- More reliance on **internal data**, rather than external data or social media. Incorporation of open source solutions. **Limited use of complex** algorithms.
- **Different level of integration and governance** of advanced analytics in the institution.
- Increasing reliance on technology companies for the provision of both **infrastructure and cloud services**.

Key opportunities

- Financial services customers from leisure and retail sectors expecting a **more personalized service**. Trust on financial sector with regards to GDPR compliance.
- Improvement of **customer satisfaction** and use of insights to improve the offering, reduce churn, optimize processes, and assist **risk mitigation and fraud detection**.
- **Many possible uses and opportunities** arising from the use of **interpretable models**.

…and key risks and proposed guidance such as model accuracy and performance regular monitoring

Key risks and proposed guidance

- The output of models can be complex, non deterministic, and correct according to a probability measure, which could harm the institution or customers. It has to be ensured that the outputs of these systems do not violate **institutions' ethical standards**. In addition, a human in the loop should be involved in the decisions, and therefore there is a need for a proper **employee training**.
- The implementation of a **governance and methodological framework on BD&AA** could promote its responsible use, which should include appropriate documentation, sufficient justification, and other explainability and traceability techniques, including the use of traceable solutions. The explainability should be based on a risk-based approach.
- There is a need for **model accuracy and performance regular monitoring**.
- The use of ML solutions could raise **ICT risks**: data security, model security, data quality, change management, and business continuity and resilience.
- As a consequence of the reliance on **open source frameworks**, **or on tools and systems developed by third parties**, both their potential risks (lack of third-party knowledge and control, vendor lock-in, concentration risk, model maintenance, etc.) and the liability, which will always remain with the entity, must be assessed.
- Finally, the importance of **data quality, protection and security** is underlined, both for regulatory purposes (including compliance with GDPR) and to ensure the adequacy of the model.

Index

Context

Key pillars

Elements of trust in BD&AA

Key observations, risks and opportunities

Annex

Data management enables an institution to control and secure data used for enterprise purposes. To be able to manage data, it is important to know its location, origin, type, content and who has access to them

Data types and data sources

- Data types: many different forms of data:
  - **Structured data**: data that exists in a format that has been sorted or organised into standard fields and categories.
  - **Unstructured data**: data not sorted or organised in a predetermined way and that consists of a wide variety of data.
  - **Semi-estructured data**: data that contains semantic tags but does not conform to the typical relational databases structure.
- Data sources: the origins of the data used for BD&AA can be either internal data derived from the institution itself or external data collected or acquired from external entities. Institutions predominantly collect and use internal data for their BD&AA models.

Data security and data protection

- Information security: the protection of information from **unauthorised access, use, disclosure, disruption, modification or destruction** in order to provide **confidentiality, integrity and availability**.
- Data security focuses on protecting the confidentiality, integrity and availability of data. To ensure data security:
  - The **protection needs** have to be identified and classified.
  - Appropriate **safeguards** for data security need to be defined and implemented, including appropriate **technical and organisational measures** to ensure a level of security appropriate to the risk.
- Organisational and management levels of institutions need to comply with data protection on BD&AA, and especially comply with the GDPR throughout the entire lifecycle of a BD&AA application when using **personal data** for training.

Data quality

- Data quality need to be considered throughout the whole BD&AA lifecycle, especially during data collection and preparation. Data that are doubtful or derived from unknown sources and ingested into analytics may result in **erroneous outputs** and lead to wrong decisions.
- Categorising aspects of data quality: **accuracy and integrity, timeliness, consistency and completeness**.

The technology of BD&AA is based on three components: infrastructure, data platform and processing

Technological infrastructure refers to the technology foundation in place for developing, implementing and adopting BD&AA. According to the US National Institute of Standards and Technology (NIST) Big Data reference architecture, the technology of BD&AA is based on three components, which are **infrastructure, data platform and processing**.

Infrastructure

The **infrastructure component** includes networking resources to transmit Big Data either into or out of the data centre, computing resources (e.g. physical processors and memory) for executing the software stack required for BD&AA and storage resources (e.g. storage area network and/or network-attached storage) to ensure the persistence of the data.

Data platform

The **data platform component** manages all the data used by an advanced analytics system and provides the necessary API to enable data to be accessed.

Processing

The **processing component** provides the necessary software to support the implementation of advanced analytics applications. The processing component enables the processing of the data according to its volume and/or velocity, in support of advanced analytics applications. Different types of processing can be applied to the **three distinct processing phases** of Big Data, namely data collection, analytics and access.

5 Annex

Organisation and governance

The establishment of appropriate internal governance structures and measures, as well as the development of sufficient skills and knowledge is key for the development, implementation and adoption of BD&AA

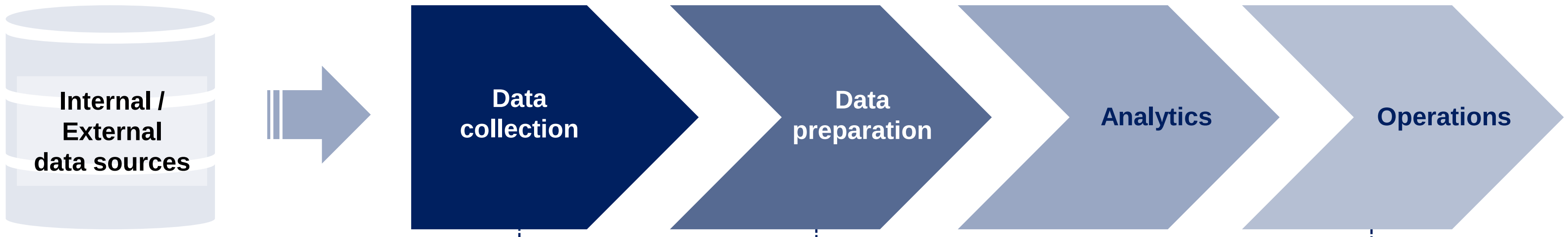
Internal governance structures and measures

- Governance structure, strategy and risk management: **clear roles and responsibilities** within the governance structure and an adequate understanding by the board of directors of the adoption and use of BD&AA, taking accountability for the related risks.
- Transparency: adherence to the fundamentals of **explainability and interpretability** to enable adequate risk management and internal audit, as well as effective supervision, supported by systematic documentation, sufficient justification and communication of important elements of BD&AA applications.
- External development and outsourcing: adhere to the EBA's Guidelines on outsourcing arrangements applies to the use of **externally developed and/or sourced BD&AA** applications; institutions cannot outsource responsibility to external providers and thus they remain accountable for any decisions made.

Skills and knowledge

- Level of understanding of management body and senior management: **relevant, up-to-date competence,** and **specific training** for the managers responsible for the use of BD&AA can help to understand the risks associated with BD&AA.
- Level of understanding of the second and third line: important to ensure that they sufficiently understand the **risks to core business processes** and address the right challenges.
- Level of understanding of developers of advanced analytics-enabled systems: data scientists could be trained to understand the **impact of the input parameters** used in BD&AA applications; and senior management need to be able to understand the **explanations** provided.
- Level of understanding of staff working day to day with advanced analytics-enabled systems: appropriate training could be provided to staff working day to day with BD&AA applications to improve awareness regarding the **responsible use of BD&AA applications**.

The fourth pillar refers to the analytics methodology that is in place for the development, implementation and adoption of BD&AA



During the **data collection phase**, data are collected from internal data sources and/or external sources.

In the **data preparation phase** the raw data are transformed to make them ready for analysis (data validation, cleansing, scaling and aggregation. During this phase, having good data quality and good data governance in place are clear success factors.

The **analytics phase** uses techniques, such as ML, to develop models that extract knowledge from data.

The **operations phase** enables the end user or another system to request and access the results and insights gained by the model and includes monitoring and maintenance of the advanced analytics solution to ensure that results remain accurate over time.