***This is Application Guidance to FIN-DM (Financial service data mining process model) and should be used in conjunctions with FIN-DM components documentation. Application Guidance describes to potential FIN-DM users its components and guides users in applying the model.***

# 1. FIN-DM Background

FIN-DM is an adaptation and extension of CRISP-DM (Cross-Industry Standard Process for Data Mining); it retains key CRISP-DM terminology and elements structure.

CRISP-DM is a hierarchical process model with four levels of abstraction (***general to specific***) consisting of ***phases, generic tasks, specialized tasks,*** and ***process instances*** respectively. At the top level, process is structured into six ***phases***, each phase consisting of several ***second-level generic tasks*** with respective ***outputs*** (reproduced in Figure 1 below). These two abstraction levels constitute CRISP-DM Reference Model. There are also ***third-level specialized tasks***, which are particular to data mining problem or situation or project specific as well as tool specific. They are further complemented by ***fourth-level process instance*** with account of actual activities within concrete data mining projects. CRISP-DM itself focuses on ***generic phases and tasks level (first and second)***, while detailing ***specialized level tasks and instances (third and fourth)*** are left to Reference Model users [1]. Likewise, FIN-DM covers ***phases and generic tasks*** based on CRISP-DM definitions while its application and scoping at specialized level is left to users.

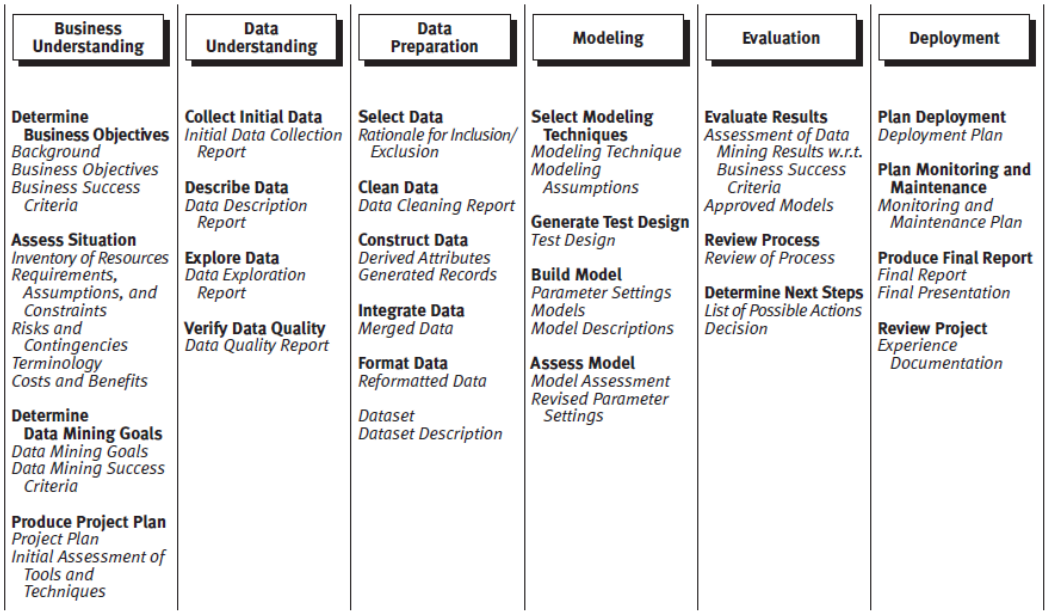



***Figure 1 CRISP-DM Hierarchical View with second-level generic tasks and outputs (as in [1])***

FIN-DM extends CRISP-DM at the three levels with increasing abstraction - starting from adding ***second-level tasks***, and then ***phases***. Lastly, ***frameworks*** or elements thereof originated from other domains are also introduced, they are positioned as specialized **domain extensions** relevant and applicable to whole FIN-DM life cycle, and thus, placed on the same hierarchical level as process itself.

The purpose of these extensions is to cover number of 'gaps' discovered in original CRISP-DM and listed in the *Table 1* below.

| 'Gap' Name | Definition |
|---|---|
| G1 - Requirements management and elicitation | Lack of tasks for validation and modification of existing requirements, elicitation of new ones |
| G2 - Interdependencies | Lack of iterations between different phases of CRISP-DM |
| G3 - Universality | Lack of support for various analytical outcomes, unsupervised and specialized techniques, deployment formats |
| G4 – Regulatory Compliance | Lack of tasks to address regulatory compliance (in particular, GDPR) |
| G5 - Validation | Lack of support for piloting models in real-life settings |
| G6 - Actionability | Lack of support for piloting models in real-life settings |
| G7 – Process | Lack of data mining process controls, quality assurance mechanisms, Critical process enablers (data, code, tools, infrastructure, and organizational factors), required for the effective execution of data mining projects are not taken into consideration |

***Table 1 Consolidated catalogue of CRISP-DM 'gaps' (as in [2])***

# *2. FIN-DM Components*

FIN-DM consists of 5 components Grouped into representation Layers:

1) FIN-DM ***Conceptual View*** – Layer 1
2) FIN-DM ***Hierarchical View*** and ***Key Enablers list*** – Layer 2
3) Accompanying Checklists (4 in total) – Layer 3
4) Glossary – key terms used in the FIN-DM
5) Application Guidance (this document) – description of the FIN-DM and its goals, FIN-DM intended usage and user categories

As noted, FIN-DM is based on original CRISP-DM process model for data mining projects execution and retains its elements. The new FIN-DM elements (additions or modifications) are marked in colours and constructed according to CRISP-DM original taxonomy:

- As additional ***individual generic tasks***,
- As additional ***phases*** (with specified set of ***generic tasks*** and ***outputs***),
- Additional ***frameworks*** and their elements (background note in section *5* below).

FIN-DM ***phases, tasks,*** and ***other frameworks elements*** are to be evaluated in the context of the data mining projects. Only relevant elements are to be picked up while irrelevant can be freely omitted. Also, FIN-DM is accompanied by the four checklists which can be evaluated, and relevant items used (entirely or some parts). FIN-DM allows to iterate between all phases in any sequence. Further, users are not prescribed to start with Business Understanding, but encouraged to evaluate depending on the project which phase to start with, e.g. with Data Understanding or combining Data and Business Understanding.

# *3. FIN-DM User Groups*

Potential users are divided into the three broad user groups depending on the overall role, responsibilities, primary activities on the data mining project and RACI[1] mapping (as presented in the ***Table 2*** below).

| | **User Group 1** | **User Group 2** | **User Group 3** | **User Group 3** |
|---|---|---|---|---|
| **Role** | (Top) Management stakeholders | Project members with business domain knowledge | Project members with project management role | Project members with development role |
| **Profile** | Functional Managers (business and tech domain) | Business users, business domain experts | Project manager(s) | Technical project delivery team - data mining experts, data /business analysts, data scientists, data engineers, software developers, etc. |
| **Primary Activity** | Overall oversight | Busines input (domain knowledge, validation, etc.) | Project management | Technical Development and Deployment end-to-end |
| **RACI designation** | I, C, R | C, R | R, A | R, A |

***Table 2 FIN-DM User Groups (as in [5])***

Three user categories are differentiated as follows:
(1) User Group 1 - management stakeholders,
(2) User Group 2 - domain business users and experts, and
(3) Users Group 3 - technical delivery team (data mining experts, data analysts, data scientists, data engineers, etc.).

FIN-DM representation is mapped to match the respective Users Group as follows:

- Layer 1 representation - intended for all Users Groups, contains model ***Conceptual View***,
- Layer 2 representation - intended for User Group 2 and 3, in additional to ***Conceptual View*** contains model ***Hierarchical Process View*** and detailed ***Enablers*** lists,
- Layer 3 representation - primarily intended for User Group 3, contains all supplementary checklists (4) in addition to ***Conceptual View*** and ***Hierarchical Process View***
- Application Guidance and Glossary are intended for User Group 3 as reference material. They will support the data mining project manager in navigating the framework. Occasionally, it could serve as support for data mining project development team (especially, data scientists) as regards common terminology

FIN-DM Layer 1 is applicable and relevant for all user groups. It provides 'helicopter' view and assists with hands-on understanding on key phases of any data mining project. Also, it provides concise view on key pre-requisites required for such project's execution. Especially, it would assist project managers in explaining data mining project(s) execution to the strategic leaders and functional managers. For the latter, it gives concise overview of the key pre-requisites/enabler required to manage and execute portfolio of data mining projects and initiatives and perform data mining at scale.

FIN-DM ***Hierarchical View*** (Layer 2) is intended for User Group 2 and 3. It equips data mining project participants with the detailed understanding of each data mining project phase, the sequence of activities within each phases and outputs. Moreover, such process view also would assists project managers to explain business and domain experts' key activities where their engagement will be required.

[1] RACI – Responsible, Accountable, Consulted, Informed [5]

Lastly, FIN-DM Layer 3 intended for User Group 3 provides hands-on support in concrete tasks execution based on checklists. It will be of use for project managers and team members providing useful input to execute the project based on structured approach, keep track of its progress, scope and deliverables (manage backlog of the tasks effectively).

*Example: determine the most applicable FIN-DM representation* User might migrate and/or be involved in more than one primary group and assume additional activities beyond primary. For instance, business users (User Group 2) might get involved into testing activities and be closely related to technical development activities at certain stages of data mining project, thus, being involved into User Group 3 too. However, they are likely not to be required to use FIN-DM most specific Layer 3 checklists in these activities. Therefore, in this example, Layer 2 representation will remain most suitable and adequate for User Group 3 given their primary role.

# 4. Note on Additional Frameworks in FIN-DM

FIN-DM combines elements of ITIL[2] and COBIT[3] frameworks and they are present at the highest abstraction level in the ***Conceptual View*** (Layer 1). Further, the ***key enablers*** list is derived based on relevant element of these frameworks and is presented in conjunction with ***Hierarchical View*** (Layer 2). Apart from definitions in Glossary, we provide background and context of their usage below.

## *ITIL and COBIT Background*

ITIL framework covers the whole life cycle of IT services [7], and focuses on defining comprehensive set of best practice processes for IT service management and support [6]. The key components of ITIL 4 framework are the *ITIL Service Value System (SVS)* and the *four-dimensional model* (as presented in reference box to the right). In ITIL paradigm, IT service management functions best when organized as a system. Hence, ITIL SVS describes the inputs to this system, its elements, the outputs, and details how the various components and activities of the organization work together to facilitate value creation through IT-enabled services. To support a holistic approach to service management, ITIL defines *four dimensions* that collectively are critical to the effective and

***ITIL SERVICE VALUE SYSTEM (SVS)***

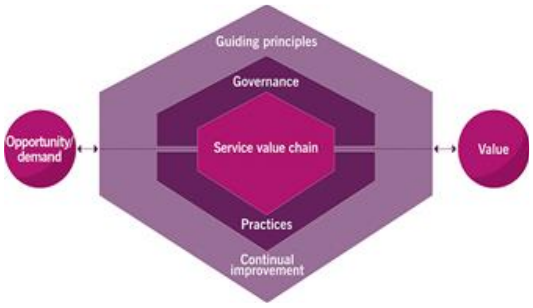
***Guiding principles*** - recommendations that can guide an organization in all circumstances, regardless of changes in its goals, strategies, type of work, or management structure.

***Governance*** - the means by which an organization is directed and controlled.

***Service value chain*** - A set of interconnected activities (operating model) that an organization performs to deliver a valuable product or **service** to its consumers and to facilitate value realization.

***Practices*** - Sets of organizational resources designed for performing work or accomplishing an objective. The ITIL SVS includes 14 general management practices, 17 service management practices, and three technical management practices.

***Continual improvement*** - A recurring organizational activity performed at all levels to ensure that an organization's **performance** continually meets stakeholders' expectations.




***THE FOUR-DIMENSIONAL MODEL***

- Organizations and people
- information and technology
- partners and suppliers
- value streams and processes.

Source: *AXELOS (2020) ITIL ®: Foundation ITIL 4 Edition. TSO, London.*

[2] Information Technology Infrastructure Library, hereinafter, we refer to the latest version - ITIL 4
[3] Control Objectives for Information Technologies

efficient value delivery. They represent perspectives which are relevant to the whole SVS ([8]-[9]).

COBIT is a framework for the governance and management of enterprise information and technology, aimed at the whole enterprise. This is achieved by COBIT [10]:

(1) Defining the components to build and sustain a governance system,
(2) Defining the design factors that should be considered to build a best-fit governance system, and
(3) Grouping relevant components into governance and management objectives.

COBIT is based on a set of key governance principles translated into ***COBIT Core Model***. It consists of 40 high-level ***governance and control objectives*** grouped into five domains. To satisfy governance and management objectives, each enterprise needs to establish, tailor, and sustain a ***governance system*** built from several components proposed by COBIT. It is complemented by ***design factors*** which can influence the governance framework. Finally, ***goals cascade*** supports alignment and prioritization between enterprise goals and management goals. Also, COBIT includes guidance on performance management with focus on improving capabilities and attaining higher maturity levels.

COBIT is a comprehensive control and management framework aimed to ensure holistic IT governance and management throughout organization. COBIT does not include process steps and tasks. Due to such broad scope, COBIT is referred as 'integrator' establishing link between various IT practices and business requirements [6]. COBIT operates from viewpoint of entire enterprise while ITIL focuses entirely on IT and associated service management practices. ITIL can be adapted and used in conjunction with other practices and extensively with COBIT, and both practices are viewed as complementary [7].

***COBIT 2019 CORE MODEL AND KEY COMPONENTS***

***Governance objectives*** – are grouped into EDM domain (Evaluate, Direct and Monitor). In this domain, the governing body evaluates strategic options, directs senior management on the chosen strategic options and monitors the achievement of the strategy.

***Management objectives*** are grouped in four domains:

***Align, Plan and Organize (APO)*** - addresses the overall organization, strategy and supporting activities for I&T.

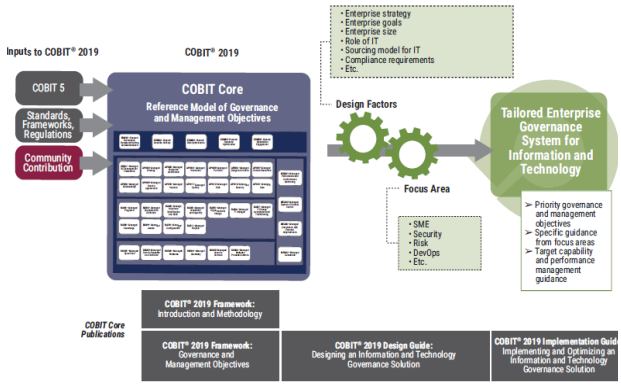
***Build, Acquire and Implement (BAI)*** - treats the definition, acquisition and implementation of I&T solutions and their integration in business processes.

***Deliver, Service and Support (DSS)*** - addresses the operational delivery and support of I&T services, including security.

***Monitor, Evaluate and Assess (MEA)*** - addresses performance monitoring and conformance of I&T with internal performance targets, internal control objectives and external requirements.

***Components of governance system*** – processes, organizational structures, principles, policies, and frameworks, information, culture, ethics and behavior, people, skills and competencies, services, infrastructures, and applications.

***Design factors*** – enterprise strategy and goals, risk profile I&T-related issues, threat landscape, compliance requirements, role of IT, sourcing model for IT, IT Implementation Methods, Technology Adoption Strategy, enterprise size.




***COBIT Overview***

*Source: ISACA, COBIT 2019: Introduction and Methodology. Schaumburg: ISACA, 2018.*

***Using Key Enablers List derived from ITIL and COBIT frameworks***

Number of Foundational capabilities and prerequisites are selected from the ITIL framework as part of ***Key Enablers*** list. ITIL pre-requisites are complemented by COBIT elements related to quality management. The suggested ITIL-derived enablers and COBIT elements are intended to address CRISP-DM process 'gaps' (G7 in *Table 1*).

Based on ITIL typology, the following attributes across four ITIL dimensions are presented:

- Organization and People - stakeholders management and data mining competencies (via knowledge management and codification systems)
- Information and Technology - data management, model development, deployment and self-service technologies encompassing the whole data mining life cycle including sharing the results with users
- Partners and Suppliers - planning of resources and internal coordination
- Value Stream and Processes - delivery models (horizontal vs. virtual cross-functional teams) and the design of service, delivery, and improvement processes

The relevant dimensions and attributes are complemented by sets of selected ITIL Management practices covering general management (portfolio, project, relationships, etc.), service (primarily related to implemented data mining models service management), technical (deployment and software development practices), and delivery approaches (including delivery models, and continuous integration, delivery and deployment practices).

COBIT elements refer to quality management at overall organizational level (data mining process quality principles, policy, and quality management practices of data mining models), and individual data mining projects level, i.e. quality management plans for respective data mining projects based on stakeholders' quality requirements.

The ***Key Enablers*** is especially useful to be considered in the context of establishing industrialized data mining practices at scale and developing portfolio of data mining services and projects. The list is also beneficial to evaluate in the context of individual data mining projects. As noted earlier, only relevant elements are to be picked up while irrelevant can be freely omitted.

> *Example: considering Key Enablers in the context of data mining model to be integrated into customer-facing application* In case of such projects, availability of model results on continuous basis to customers becomes of critical importance. That implies that in case of model failure, fallback process needs to be followed. In this context, by reviewing *Key enablers, incidence management practices* are the most relevant and need to be picked-up. By adopting such practices from the rest of organization towards the given model or establishing such practices anew, organization will ensure required model service level consistently.

For the users' guidance and convenience, ITIL and COBIT elements are also mapped and inherently integrated where applicable into the FIN-DM hierarchical process via respective sub-components and tasks. For COBIT elements, these are data mining project quality management tasks reflected in the ***Risk Management and Quality Assurance sub-component*** of the ***Compliance phase***. For ITIL elements, these are data mining requirements reflected in the ***Requirements phase*****.**

# *5. Note on Project Management Aspects*

FIN-DM primary focus is adapting data mining process towards needs of financial services industry. Therefore, some aspects are left for the users to specify more in the context of existing governance and management practices and frameworks adopted in their specific organizations. One of such aspects is project management where FIN-DM is less centred

beyond Requirements elicitation and management. Typically, project management practices are governed within organization and guided by respective paradigms (e.g. Agile, SAFE, etc.). Therefore, we propose for users to rely on them in conducting data mining projects. Such practices typically specify well the division of responsibilities between various groups of stakeholders in organization when projects are conducted.

At the same time, we include defining and sign-off of Roles and Responsibilities into the Business Understanding phase. Furthermore, we recommend relying on RACI framework (mentioned above) with the tentative categorization of project participants as in BABOK[4] Guide [13] as:

- **Responsible (R)** – person(s) who will be performing the work on the task,
- **Accountable (A)** – person(s) who is ultimately held accountable for successful completion of the task and is the decision maker,
- **Consulted (C)** - the stakeholder or stakeholder group who will be asked to provide an opinion or information about the task (often the subject matter experts (SMEs) are falling into these category),
- **Informed (I)** - a stakeholder or stakeholder group that is kept up to date on the task and notified of its outcome. In case of Informed the communication is one-direction (business analyst to stakeholder) and with Consulted the communication is two-way.

Roles and Responsibilities (as part of Checklist 1) can be detailed for each phase of the data mining projects, emphasizing type of involvement in key process phases of the respective specialists based on their core competences and functions (as in Example).

> *Example:* dedicated SMEs and End Users can be heavily involved in formulating, review and approval of initial data mining requirements. They also are required to be involved and perform validation and business testing of the final data mining model or outcome, as well as validate significant intermediate results of the project from business point of view. Data Scientists and Software/AI Engineers are responsible for technical development and implementation, but with different participation degree: in key FIN-DM phases of Data Understanding and Preparation, Modelling, Testing, Data Scientists lead and perform the work, while in Implementation phase, AI/ML Engineers are responsible to execute key software development and integration associated tasks.

Lastly, we also recommend for the users to consult widely accepted standards and best practice guides in project management field. The closest related, but not exhaustive list of such practice guides as points of initial reference could be PMBOK[5] body of knowledge [11], PRINCE 2[6] methodology [12], BABOK body of knowledge [13], and similar.

# *6. FIN-DM Application Principles and Recommendations*

### *General principles*

**Adaptable and Extendable** FIN-DM is highly adaptable – users are encouraged to contextualize process model to specifics of their concrete data mining project and omit any elements which are not applicable or relevant. As well, any modifications (including extensions) can be introduced by users too.

**Unlimited iterations and accounting for interdependencies** FIN-DM promotes and allows for any number of iterations across all phases and elements as deemed necessary and beneficial by data mining project management and project experts team. Further, users are

[4] BABOK – the globally recognized standard for the practice of business analysis
[5] PMBOK – project management body of knowledge generally recognized as best practice [11], widely accepted and used across the world
[6] PRINCE 2 – widely considered as the leading project management method [12]

free and encouraged when using ***Hierarchical Process Model***, to merge, closely integrate or parallelize phases at their discretion. Based on evaluation with users, we also provided indications of potential parallelization/merger of phases on the diagram. FIN-DM also supports discovering, and ongoing tracking and calibration of interdependencies throughout the whole data mining life cycle [2].

**Compatible, Easy to integrate with other frameworks, practices and methods** Given FIN-DM adaptiveness, flexibility and its concentration on data mining life-cycle, it is fully compatible and can be easily embedded, integrated with other popular product development and delivery frameworks, practices, methods and patterns used across different organizations. For example, it can be fit into Adaptive Software Development/XP programming with focus on software development, or alternatively Agile, Scaled Agile (SAFE), Scrum and its variations focusing on comprehensive flow of work management for complex systems, products and projects, etc.

**Dynamic** FIN-DM follows open architecture principles and is not prescriptive. Therefore, adding, changing, and modifying existing complements and elements in response to external disruptions, technological changes and emerging organizational needs can be implemented by users without impacting framework structure, other components, and content. FIN-DM is referenced and supports periodic realignment with other IT frameworks (ITIL, COBIT) which in turn are upgraded in due course and reflect latest organizational and technological developments in IT/technology domain.

### *Specific application recommendations*

**Practical adoption** To practically adapt FIN-DM, we suggest considering two general application patterns – 'light-weight' mode and 'full' mode. This is especially relevant for two elements – Requirements phase and AI & Ethics subcomponent. Suggested application modes primarily correlate with size and complexity of company operations and its business model. Also, complexity of the data mining project, type of data mining problem and relevance of AI ethics and compliance concerns are key drivers.

For instance, if the company operates in limited geographical scale, is of small size in industry terms and is not delivering full universal banking services, the data mining project scope and data used will be naturally constrained. Therefore, many of Requirements can be simplified and/or omitted which also decreases number of the relevant tasks as well as necessity for many Requirements management and iterations activities.

In the similar way, if due to data mining use case context and data used, there are no triggers for AI ethics and compliance concerns, AI Ethics tasks and Checklist 2 can be approached in 'light-weight' mode, i.e. it is sufficient to use them as checkpoints in the beginning and end of the data mining projects.

**Tools support** Assigning and tracking FIN-DM phases, tasks, and checkpoints progress is best realized with support of respective technologies, such as collaborative issue tracking, project management, and workflow tools.

## *References*

*[1] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. 2000. CRISP-DM 1.0 Step-by-step data mining guide, SPSS Inc.*

*[2] Plotnikova V., Dumas M., Nolte A., Milani F. 'Designing a Data Mining Process for the Financial Services Domain' (submitted).*

*[3] Wirth, R. and J. Hipp (2000). "CRISP-DM: Towards a standard process model for data mining." In: In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, London UK. Springer–Verlag, pp. 29–39.*

*[4] Wiegers, Karl. "More about software requirements: thorny issues and practical advice". Microsoft Press, 2005.*

*[5] Spalek, S. (2018). Data analytics in project management. CRC Press.*

*[6] Guldentops, E., Hardy, G., Heschl, J.G. and Taylor, S., 2010. Aligning Cobit, ITIL and ISO 17799 for Business Benefit: A Management Briefing from ITGI und OGC.*

*[7] Almeida, Rafael, Paloma Andrade Gonçalves, Inês Percheiro, Miguel Mira da Silva, and César Pardo. "Integrating COBIT 5 PAM and TIPA for ITIL Using an Ontology Matching System." International Journal of Human Capital and Information Technology Professionals (IJHCITP) 11, no. 3 (2020): 74-93.*

*[8] AXELOS (2020), ITIL® 4: Create, Deliver and Support. TSO, London.*

*[9] AXELOS (2020) ITIL ®: Foundation ITIL 4 Edition. TSO, London.*

*[10] ISACA, COBIT 2019: Introduction and Methodology. Schaumburg: ISACA, 2018.*

*[11] Project Management Institute (2017). A guide to the project management body of knowledge: PMBOK® GUIDE, Sixth edition, Newtown Square, Pennsylvania, US.*

*[12] AXELOS (2017), Managing Successful Projects with PRINCE2®. TSO, London.*

*[13] International Institute of Business Analysis (2015): BABOK®, version 3, A GUIDE TO THE BUSINESS ANALYSIS BODY OF KNOWLEDGE®, Toronto, Ontario, Canada.*