3. Framework Overview

As presented in Figure 1, The BLIND proposes a wide range of processes for curating metadata in Data Ecosystems, where each process describes a coherent set of engineering and management activities related to metadata curation. The BLIND structure is flexible and may be easily adapted to the needs of the actors interested in curating Data Ecosystem metadata. In this sense, processes are organized in functional dimensions, enabling modularization of the framework. BLIND also provides a set of best practices aligned with principles of agile and open collaboration for managing curation work through the collaborative effort of self-organizing actors. Finally, BLIND also defines a set of roles to guide the curation work.
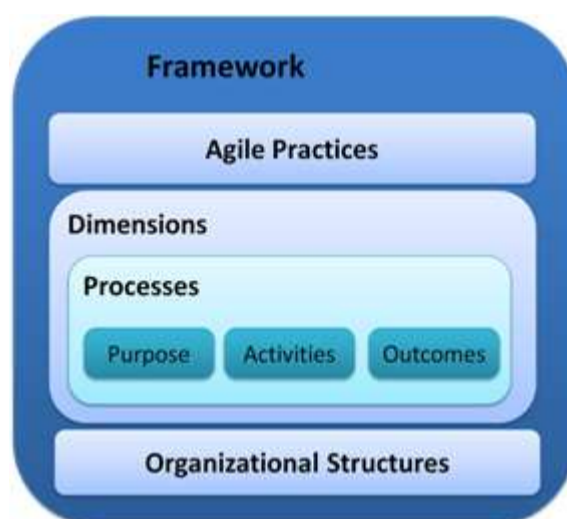

Figure 1: BLIND Framework

These elements can be seen as enablers for metadata curation, through which the curation is guided and goals can be achieved. These elements are detailed as:

- **Actors and Roles:** represent the people part of metadata curation work. The BLIND recommends how actors may be organized to curate metadata as well as defines a set of roles involved in performing and supervising the curation work.
- **Agile Practices:** adopt and tailor strategies from mainstream agile methods to support curation of metadata.
- **Dimension:** groups a set of correlated processes related to one or more stages of metadata lifecycle.
- **Process:** groups a set of correlated activities that are executed in order to generate expected outcomes. It indicates key curation areas where actors should focus on enabling curation of metadata. The order that the activities are presented in a process does not imply any prescriptive order in their use. The order of activities is influenced by multiple factors, including organizational and technical considerations, each of which can vary according to the Data Ecosystem context.
- **Activity:** represents a set of actions defined to achieve a specific result.
- **Purpose:** represents a high-level goal related to a process.
- **Outcome:** describes an observable result of the achievement of the process purpose, including the production of an artifact, a significant change in Data Ecosystem, project or environment, meeting of specified requirements or goals.

The Agile Practices provide a set of best practices that enable, encourage and guide contributors in the iterative, incremental and self-organizing curation of metadata. Best practice is a concept which refers to a set of techniques, procedures, and activities that has been shown by research and experience to produce good results and that is established or proposed as a standard suitable for widespread adoption. The Agile Practices are: **User Story, Persona, Backlog, Continuous Integration, Continuous Refactoring, Automated Tests, Collective Ownership and Burndown Chart**

Besides the agile practices, the BLIND Framework should ensure that all of the necessary stages in the metadata lifecycle are covered. In consequence, the overall curation of metadata is conceived regarding an initial set of six dimensions; each one contributes to one or more stages of metadata lifecycle. The scope of each dimension are:

- **Metadata Curation Planning:** Provides the basis for the exercise of planning, monitoring the metadata curation initiative.
- **Metadata Curation Monitoring and Controlling:** Provides the basis for enabling and promoting collaborative efforts from contributors.

- **Metadata Curation Platform Administration:** Provides the basis for planning, development, and management of a metadata curation platform.

- **Metadata Acquisition:** Provides the basis for planning, control and support the acquisition and selection of metadata.

- **Metadata Quality Management:** Provides the basis for planning, implementation and control the quality assurance and improvement of metadata.

- **Metadata Preservation and Dissemination:** Provides the basis for preserving metadata and keeping them discoverable and accessible to Data Ecosystem actors.

Each one of the proposed Agile Practices, Dimensions and Processes will be presented in the next steps of this survey. In particular, each process also contains a purpose and a set of outcomes, work products and activities.

3.1 Actors and Roles

Description: The people part of metadata curation involves actors who curate metadata, which they will make available to contributors and non-contributors alike (from here, all the actors who contribute to the curation of metadata will be called contributors).

BLIND also supports team-building and teamwork. Contributors have varying areas of expertise, with their diverse set of skills, the contributors should be able to curate metadata as a group. Hence, teams and the characteristics of teams are central to metadata curation into Data Ecosystems. Contributors should work in collaboration with each other to curating metadata into Data Ecosystems. Several agile approaches emphasize that collaboration is the key to success for agile project delivery.

Roles clearly defined likely makes coordination easier (though non-trivial), while communication and coordination patterns often align with the structure of the software itself. Inspired by Scrum and Disciplined Agile Delivery frameworks, BLIND deemphasizes roles based strictly on skill sets in favor of primary roles that can include a variety of skills. Accordingly, the primary roles are:
- **CurationMaster:** Represents an authority to exist beyond the potential crowd of curators who organizes the whole metadata curation work providing a way for any actor to contribute to the common effort. The CurationMaster identifies and explains the metadata curation requirements, prioritizes requirements, provides feedback on the curatorial strategy.
- **Stakeholder:** Represents someone who is materially impacted by the outcome of the metadata curation work. A stakeholder could be

any Data Ecosystem actor who directly or indirectly consumes curated metadata. Stakeholder represents the needs and desires of the Data Ecosystem actors regarding metadata and curation needs.
- **MetadataCurator:** Represents contributors who focus on performing the actual curation of metadata. MetadataCurators will carry out all tasks required to curate the metadata, including acquisition, appraisal and selection, quality assurance and preservation.
- **TechnicalExpert:** Since MetadataCurators Team should perform tasks in one or more disciplines, this doesn't imply that every MetadataCurator needs to be an expert at everything. Sometimes the team needs the help of a TechnicalExpert to overcome a difficult problem and to transfer their skills to one or more MetadataCurator.
- **PlatformOwner:** Represents the authority that provides the services and functions for the overall creation and operation of the metadata curation platform. The platform used to support metadata curation tasks is a key source for successful curation and someone needs to be responsible for mitigates this risk.
- **TeamLeader:** Often, a team needs an informal leader, called a TeamLeader, who emerges from the team because of a passion for the subject. The TeamLeader is a kind of metadata curation coach, helping to keep the team focused on curation activities work items and fulfilling their iteration goals and commitments made to the Stakeholders.

3.2 Agile Practices

Purpose: The Agile Practices provide a set of best practices that enable, encourage and guide contributors in the iterative, incremental and self-organizing curation of metadata. Best practice is a concept which refers to a set of techniques, procedures, and activities that has been shown by research and experience to produce good results and that is established or proposed as a standard suitable for widespread adoption.

Practices:
- **User Story:** It is a technique widely used by several agile methods to capture system requirements [74]. In metadata curation context, the simplicity of user stories lends itself to the articulation of metadata curation requirements. The informal language may contribute to create a shared understanding with those contributors not familiar with metadata curation and metadata management concepts.
- **Persona:** It is a detailed, synthetic biography of fictitious user. In metadata curation context, personas are useful when both MetadataCurators and CurationMaster don't have easy access to Stakeholders, helping to guide your decisions about metadata needs and curation requirements.
- **Backlog:** Backlog is a list of all the work necessary for the product. It lists all requirements, features, improvements, and fixes that constitute the changes to be made to the product in future releases. In metadata curation context, the backlog can be used to coordinate

the contributors to perform tasks. In this sense, a curation backlog lists technical work related to metadata curation. The backlog items reflect tasks related to BLIND's dimensions and processes.

- **Continuous Integration:** In Software Engineering, continuous integration is the practice in which developers integrate their work frequently, and testing the modifications, as early and often as possible. In metadata curation context, it is recommended that all metadata acquired should be as soon as possible transferred/ingested to a metadata repository. Once ingested in the repository, the metadata must be evaluated against compliance with quality policies, standards and other regulations. The idea is to find issues quickly; giving each MetadataCurator feedback on their work and automated mechanisms evaluate that work quickly.
- **Continuous Refactoring:** Refactoring is a controlled technique for restructuring the internal structure of the existing program's source code without changing its external behavior [52]. In BLIND, refactoring practice is viewed as the refactoring metaphor proposed by [13]. In that sense, refactoring should be seen as a restructuring and possible optimization of metadata curation work. For example, planned metadata curation actions that are not being properly performed or resulting in a poor performance should be reviewed as soon as the problem is identified.
- **Automated Tests:** In software development, an automated test is a technique to automate some repetitive but necessary tasks in a formalized testing process, or perform additional testing that would be difficult to do manually. In metadata curation context, an automated test suite enables curation teams to verify the quality of metadata safely. Moreover, such a practice can continuous assure quality.
- **Collective Ownership:** In software development projects, collective ownership is the explicit convention that any team member is allowed to make changes to any code file as necessary. In metadata curation context, the metadata can be owned by all the contributors. Everyone can have access and authorization to access, edit, and enhance any metadata. Ownership is collective and everyone is equally responsible to all parties.
- **Burndown Chart:** Burndown chart is a visual measurement tool to measure the development team progress. It shows the completed work per day against the projected rate of completion for the current project release. Burndown charts can be applied to metadata curation to measure progress over time. Thus, the curation work can be represented regarding either time or curation backlog items.

3.3. Metadata Curation Planning Dimension

Description: The Metadata Curation Planning Dimension consists of those processes performed to establish the total scope of the effort, define and refine the goals and requirements, and develop the course of action for curating metadata into Data Ecosystems. The Planning processes develop the metadata curation plan and the curation artifacts (e.g., policies, standards, and procedures) that will be used to carry out the curation initiative. The distributed nature of Data Ecosystems may require the use of repeated feedback loops for additional analysis. As more the experience and information are gathered and understood from metadata curation initiative, additional planning will likely be required. Significant changes occurring throughout the metadata curation initiative may trigger a need to revisit the action plan.

Processes:
- **Curation Requirements Engineering**
- **Curation Planning**

3.3.1 Curation Planning Dimension

Process: Curation Requirements Engineering

Purpose: Recommends activities for elicitation (collecting, creating), analysis (aligning, prioritizing), and validation (monitoring, enforcing) of requirements involving metadata (e.g., business rules, metadata ownership, metadata classification, metadata quality, metadata usage, metadata access, authentication, entitlements, etc.).

Outcomes:
- The needs and expectations of stakeholders concerning metadata curation are collected.
- The needs and expectations are refined into requirements.
- Requirements are analyzed to determine their feasibility and priority.
- Requirements are organized and available to stakeholders and other contributors.

Activities:
- **Envisioning requirements:** Envision high-level goals for coming to a common understanding about the scope of metadata curation work.
- **Elaborating requirements:** The information obtained during requirements envisioning is expanded and modified during elaboration.
- **Validating requirements:** Certify that the requirements are an acceptable description of the stakeholders' expectations and needs related to metadata curation.
- **Classifying and prioritizing requirements:** This activity grouping requirements into different priority groups with each group representing something stakeholders can relate to a concept or category.
- **Managing requirements:** This activity organizes and manages requirements in order to communicate them between stakeholders and other contributors.

3.3.2 Curation Planning Dimension

Process: Curation Planning

Purpose: Recommends activities for establishing the basis for creating and maintaining a metadata curation action plan that aligns with the requirements with a strategy and a set of policies, standards and procedures.

Outcomes:
- Scope and target are defined for metadata curation in accordance with metadata curation requirements.
- A metadata curation strategy is created, describing the vision, long-term goals, and an implementation road-map.
- Policies, procedures, standards and licenses are defined according to metadata curation strategy.
- Metadata curation platform architecture is established to support the activities towards the metadata curation activities.

Activities:
- **Defining and prioritizing metadata needs:** Each Data Ecosystem must identify what metadata would benefit most from curation, and determine if potential returns would support the required efforts.
- **Defining and reviewing metadata curation strategy:** A strategy, on the other hand, is a blueprint, layout, design, or idea used to accomplish a specific goal. A strategy is very flexible and open for adaptation and change when needed. Thus, a strategy is an outline of the steps to curate metadata. A strategy is a solution that helps curators plan the metadata curation.
- **Defining and reviewing metadata policies and procedures:** Policies are formal, brief, and high-level statements on how the contributors must run the acquisition, quality assessment, preservation, security, and use of metadata [67]. Procedures define how to achieve requirements and are the mechanisms to enforce policies. The whole set of procedures also represent the implementation of metadata curation strategy.
- **Defining and reviewing metadata standards:** Metadata standards are documented agreed on properties and rules on representation, format, definition, structuring, dissemination, manipulation, use, and management of metadata.
- **Approving and reviewing metadata curation platform architecture: Metadata** curation platform architecture is an integrated set of high-level structures that govern and define how metadata is used, stored, managed and integrated within a Data Ecosystem.
- **Defining and reviewing and licenses for use and reuse of metadata:** Licenses are legal documents giving official permission to use and reuse of metadata. That is, licenses to clearly explain the conditions under which their metadata may be used.

3.4 Curation Monitoring and Controlling Dimension

Description: The Curation Monitoring and Controlling Dimension consists of those processes required to monitor and control the progress and performance of the metadata curation initiative. These monitoring processes provide contributors insight into the health of the metadata curation initiative and identifies any areas requiring additional attention. This dimension provides the basis for not only monitoring and controlling the work being done but also monitors the adherence to the curation planning.

Processes:
- **Metadata Curation Monitoring**
- **Recruiting and Engagement Management**
- **Communication and Feedback Management**
- **Curation Coordination**

3.4.1 Curation Monitoring and Controlling Dimension

Process: Metadata Curation Monitoring

Purpose: Recommends activities for monitoring metadata curation work. Monitoring allows results, processes, and experiences to be documented and used as a basis to steer decision-making and learning processes.

Outcomes:
- Monitoring indicators and their related procedures are defined for monitoring the ongoing curation activities against the curation plan.
- Indicators measurements are collected and documented.
- Nonconformities related to metadata curation strategy, policies, procedures, standards, and/or architecture are identified.
- The CurationMaster, key stakeholders and other contributors are notified about non-conformities and progress against curation plans.

Activities:
- **Defining monitoring indicators:** Indicators are metrics that are important for any project, particularly for monitoring and evaluation purposes. Through the indicators, contributors can pre-determine how effectiveness will be evaluated in a precise and clear manner.
- **Defining monitoring indicators measurement methods:** Once the indicators have been defined, a practical issue that needs to be addressed is how indicators will be measured and monitored. To evaluate the metadata curation progress and monitoring the plan, clearly defined procedures must be used consistently.
- **Measuring and monitoring indicators:** Measure and monitor indicators to evaluate the compliance with defined metadata curation plan. It is also important to establish appropriate resources to measure indicators.
- **Monitoring conformance with metadata curation strategy, policies, procedures, standards, and architecture:** Part of the curation work is to monitor and ensure conformance with available rules and regulations.
- **Reporting status: Status reporting is one element of the curation controlling process.** Its purpose is to keep the CurationMaster, key stakeholders and other contributors formally communicated and informed about curation work status.

3.4.2 Curation Monitoring and Controlling Dimension

Process: Recruiting and Engagement Management

Purpose: Recommends activities for recruiting contributors and promoting engagement. A collaborative metadata curation initiative relies on a constant stream of curation activity, with multiple contributors contributing in various ways, at various stages. Therefore, it is necessary to motivate, engage and retain new contributors to promote a sustainable curation initiative.

Outcomes:
- Actions to communicate, educate and promote the importance and value of metadata curation are planned and executed.
- A clear mission statement that spells out the metadata curation initiative purpose.
- A group of motivated contributors is recruited to collaborate with metadata curation initiative.
- Contributors are recognized as a valuable resource for the metadata curation initiative.
- Relationships and partnerships among contributors are reinforced and valued, following a more humane and collaborative approach.

Activities:
- **Developing and promoting metadata curation awareness:** Promoting metadata curation awareness means more than ensuring that actors in the Data Ecosystem are aware of the existence of metadata curation issues.
- **Creating a clear and compelling cause:** Contributors must be provided with a convincing reason to be a part of the metadata curation initiative. The more compelling the mission, the easier it is to incentive actors to contribute to the metadata curation initiative.
- **Assigning stewardship:** Curation tasks need to be carried out by someone. Thus, it is crucial identifying who will curate the metadata and assigning to them their proper roles.
- **Developing partnerships:** It is important to know how to develop, maintain, sustain and manage partnerships, which are a key part of any strategic approach to engagement. Strengthening the partnerships can be one of the ways to motivate actors to contribute.
- **Understanding and reducing barriers to contribution and engagement:** It is important to ensure low practical barriers to participation in the work of metadata curation. The emphasis is on facilitating the engagement of new contributors.

3.4.3 Curation Monitoring and Controlling Dimension

Process: Communication and Feedback Management

Purpose: Recommends activities for planning and maintain an effective communication flow between contributors as well as ensuring the ultimate disposition of metadata curation information.

Outcomes:
- Communication strategy and channels are established.
- Communication to be managed is identified.
- Feedback and other metadata curation information are collected and documented.

Activities:
- **Promoting open communication:** Open communication is the ability of anyone, on equal conditions are encouraged to share their thoughts and concerns, both positive and negative, without the worry of retaliation when the feedback is bad. In Data Ecosystems, open communication may contribute to promote trust among contributors, and another natural result is the engagement of contributors.
- **Defining and establishing communication strategy:** A communication strategy defines how to organize and spread information, and ensure control of metadata curation work, while allowing contributors to communicate their progress when appropriate.
- **Defining and establishing communication channels:** Define communication channels to enable communication strategy. Examples of communications channels include websites, social media platforms, blogs, and newsletter.
- **Developing and establishing feedback gathering mechanisms:** Develop a range of methods to gather feedback. The feedback gathering mechanisms including, but not limited to, a contact form, forums, ratings and reviews surveys, or a comment box .
- **Monitoring communication activities:** Monitor communication activities in order to make sure that communications strategy is yielding the expected results, and to determine to what extent communication efforts are effective.
- **Seeking contributors' discussions:** Moreover, one-on-one discussions and focus groups are important to provide the opportunity to dive deeper into understanding the needs of contributors and can help them improve metadata curation initiative and its assets.

3.4.4 Curation Monitoring and Controlling Dimension

Process: Metadata Curation Coordination

Purpose: Recommends activities for organizing, orchestrating, and leading the metadata curation work.

Outcomes:
- Metadata needs are established and documented.
- Metadata needs are prioritized and organized into Metadata Backlog.
- Curation teams are formed.
- Curation Backlog, which contains and organizes a set of metadata curation work, is created.
- Metadata curation work progress data is gathered.
- Metadata curation work progress is tracked and monitored.

Activities:
- **Creating the Metadata Backlog:** The Metadata Backlog serves to connect the contributors and orchestrate curation work. All metadata that should be curated will represent a work item into a Metadata Backlog.
- **Forming curation teams:** Contributors should work in collaboration with each other to curating metadata into Data Ecosystems. Contributors have varying areas of expertise, with their diverse set of skills, the contributors should be able to curate metadata as a group.
- **Grooming of Curation Backlog:** The grooming of Curation Backlog includes the definition of all the tasks related to the curation of selected metadata. These tasks must be aligned with the curation planning.
- **Coordinating metadata curation tasks:** Coordination of activities is one of the ongoing goals throughout the curation effort.
- **Track curation progress:** Monitoring the status of metadata curation work to keep track curation progress and provides the means to recognize deviation from the plan and take corrective and preventive actions and thus minimize risk.

3.5 Curation Platform Administration Dimension

Description: The Platform Administration Dimension consists of those processes for the design, implementation, and maintenance of the platform responsible for support the metadata curation work. Such a platform should enable standardization and integration of metadata curation tasks. It should support the metadata needs of the Data Ecosystem.

Processes:
- **Metadata Curation Platform Design**
- **Metadata Curation Platform Implementation**
- **Metadata Curation Platform Maintenance**

3.5.1 Curation Platform Administration Dimension

Process: Metadata Curation Platform Design

Purpose: Recommends activities for the strategic and technical design and planning of metadata curation platform architecture. Proper implementation of a platform for supporting metadata curation environment is not just about the tools to archive metadata. Rather, it is about creating a strategy to plan, design, and construct a platform capable of addressing curation requirements and allowing supporting the metadata curation strategy.

Outcomes:
- Identified metadata curation concerns are addressed by the Metadata Curation Platform architecture.
- Metadata Curation Platform architecture candidate models are developed.
- A set of technologies alternatives for Metadata Curation Platform are identified and related to architecture elements.

Activities:
- **Initialize the definition of the platform architecture:** Build an understanding of the environment/context of use for which the platform is needed in order to establish platform functions. Therefore, establish a platform architecture roadmap and strategy that should include methods, modeling techniques, tools, need for any enabling curation platform.
- **Developing candidate architectures models:** Using relevant modeling techniques, and in conjunction with the curation needs and Requirement Engineering process, determine architectural entities, which address the different types of curation requirements.
- **Identifying and analyzing available technologies:** Selecting appropriate metadata related technology is an important metadata curation responsibility. It is necessary to identify and analyze such technology to meet curation needs, including total cost, reliability, and other aspects.
- **Relating platform architecture to available technologies:** Define the technologies (e.g., systems, tools, repositories) that reflect the platform architecture.
- **Managing the designed architecture:** Establish and maintain the rationale for all selections among alternatives and decision for the architecture. This includes concordance, completeness, and changes due to the environment or context changes, technological, implementation, and operational experiences.

3.5.2 Curation Platform Administration Dimension

Process: Metadata Curation Platform Implementation

Purpose: Provides the metadata curation platform (i.e., infrastructure and services) to support metadata curation work. This process defines, provides and maintains the facilities, tools, and communications and information technology assets needed for metadata curation work with respect to the architecture designed in the Metadata Curation Platform Design process.

Outcomes:
- Metadata curation technologies are installed and supported.
- Metadata Curation Platform is implemented to support metadata curation work.

Activities:
- **Establish the Metadata Curation Platform:** Identify, obtain and provide infrastructure resources and services that are needed to implement Metadata Curation Platform.
- **Maintain the Metadata Curation Platform.** Evaluate the degree to which Metadata Curation Platform satisfies curation needs. Moreover, identify and provide improvements or changes to the platform as the curation requirements change.

3.5.2 Curation Platform Administration Dimension

Process: Metadata Curation Platform Maintenance

Purpose: Recommends activities for the day-to-day technical supervision of the metadata curation platform. The metadata curation platform must to meet necessary conditions throughout the entire metadata lifecycle. Such management does not have one objective. There are many, including performance, efficiency, security, and privacy.

Outcomes:
- Risks are evaluated and analyzed.
- Risk treatment options are identified, prioritized, and selected.
- Mechanisms are defined and implemented for ensuring metadata security and platform reliability.
- Metadata curation platform is monitored and optimized.

Activities:
- **Performing a risk assessment:** Understand, manage, control and mitigate risk to the metadata curation assets. The risk assessment includes, but not limited to, identify and prioritize assets, identify threats and vulnerabilities, analyze and establish controls to minimize or eliminate the probability of a threat.
- **Developing and establishing metadata security mechanisms:** Establish and maintain policies, controls, and procedures for metadata security. It includes creating plans to ensure that metadata curation platform can recover and continue as failures, defects or serious incidents occur.
- **Monitoring and optimizing platform:** To ensure ongoing access to metadata, it is important to monitor the condition of the platform continually. In particular, it is recommended routinely perform test retrievals or restorations of stored metadata. This activity also aims to optimize metadata curation platform performance both proactively and reactively, by monitoring performance and by responding to problems quickly and competently.

3.6 Metadata Acquisition Dimension

Description: The Metadata Acquisition Dimension consists of those processes required for creating, harvesting and selecting metadata. In particular, 'create' refers to original metadata generated and recorded by participants, and 'harvest' refers to pre-existing metadata collected from other sources. This dimension also includes activities to allow the selection and rejection of metadata that does not meet specified requirements and policies.

Processes:
- **Metadata Creation Management**
- **Metadata Harvesting Management**
- **Metadata Model Management**
- **Metadata Appraisal and Selection Management**

3.6.1 Metadata Acquisition Dimension

Process: Metadata Creation Management

Purpose: Recommends activities for creating appropriate metadata. The metadata produced by human beings is probably what most people assume when they think of a metadata source. In this processes, the metadata will be created manually by the MetadataCurators.

Outcomes:
- A list of standard terminologies to apply to metadata is defined.
- A list of guidelines and documented workflows are created for guiding and support the creation of metadata.
- A set of metadata is created.
- Metadata has its basic quality policies ensured.

Activities:
- **Selecting and developing a metadata model:** Analyze and identify which metadata model should be applied to structure the metadata to be produced.
- **Identifying and selecting standard terminology:** The terms and phrases that are used for creating metadata content should reflect appropriate and accepted vocabularies in Data Ecosystem. Whenever possible use a controlled vocabulary, which provides a consistent way to describe metadata contents.
- **Developing and documenting workflows for metadata creation or capture:** Develop and document metadata creation workflows, i.e., the processes used to prepare and create metadata. These workflows act as metadata templates and provide some guidelines to ensure consistency.
- **Ensuring basic quality control:** Detect and correct general potential problems with metadata that could affect its use. The rejected metadata is ideally reported back to the metadata creator for further analysis to identify and to rectify the incorrect records.

3.6.2 Metadata Acquisition Dimension

Process: Metadata Harvesting Management

Purpose: Recommends activities for harvesting appropriate metadata. Not all metadata should be created from scratch. Indeed, such a notion will be unworkable. Harvesting is the process of collecting metadata from a remote or external source. It should be a scheduled process to keep local copy aligned with remote metadata.

Outcomes:
- A list of metadata to harvest is defined.
- A list of metadata sources are identified
- Mechanisms are defined and implemented for harvesting metadata from metadata sources.
- A set of metadata is periodically harvested.

Activities:
- **Selecting and prioritizing metadata to harvest:** This activity involves the prioritization of metadata to be harvested. Hence, it is important to analyze the feasibility of metadata identified in terms of importance and other aspects, such as cost and timeliness, providing the basis on which determine harvest priority.
- **Identifying, analyzing, and selecting metadata sources:** Identifying, analyzing and selecting potential sources of metadata that best fit the metadata needs of the Data Ecosystem.
- **Designing and developing metadata harvesting mechanisms for selected sources:** Design and develop mechanisms to harvest metadata from selected metadata sources. A properly designed harvesting mechanism extracts metadata from the sources, enforces quality and consistency standards, normalizes and integrates metadata so that separate sources can be used together, and finally delivers metadata in the standardized format so that it can be loaded into the final target repository.
- **Defining a harvesting schedule plan: Define a schedule plan to harvest metadata.** Usually, metadata harvesting is done on a scheduled basis to reflect changes made to the source. Normally, such a schedule plan relies on static intervals, with harvesting processes occurring at daily, weekly, monthly, or other periodic intervals.

3.6.3 Metadata Acquisition Dimension

Process: Metadata Model Management

Purpose: Recommends activities for designing, developing and refining of a metadata model, which is the overall structure for the metadata. Such a model can be viewed as an integrated, subject-oriented set of specifications defining the essential metadata produced and consumed across the Data Ecosystem.

Outcomes:
- A set of subject areas are identified to categorize metadata.
- Subject area metadata models are defined to structure and specify metadata items.

Activities:
- **Defining modeling language:** Several different data modeling languages are available, each using different diagramming conventions or styles. Hence, it is essential to evaluate and select the most suitable language to design a metadata model that support curation needs.
- **Determining metadata to be modeled:** Determine metadata to be modeled. It is essential to collect and restructure base information that will establish metadata model design.
- **Organizing needed metadata into subject areas:** Organize metadata needs into high-level subject areas. All metadata produced and consumed across the Data Ecosystem will be represented within a subject area. Each subject area is a high-level classification of metadata, which represents a group of concepts about a major topic of interest.
- **Defining granularity level of each subject area:** Decide which aspects of metadata are crucial for stakeholders, and how granular each type of metadata need to be.
- **Identifying and selecting relevant metadata standards:** It is necessary to identify those standards that include the metadata items needed to describe a subject area. These standards can be used to design subject area models.
- **Designing subject area models:** Create a subject metadata models, using a high-level conceptual model. Designing subject area model includes the definitions of all the concepts (entities, attributes) of the subject area.
- **Evaluating and refining metadata model:** Metadata model refinement is an iterative process. First evaluate the resulting model using as a reference the requirement documentation, competency questions or usage scenarios. The evaluation results will guide the refinement of metadata model on how effectively and efficiently formalize and structure metadata required.

3.6.4 Metadata Acquisition Dimension

Process: Metadata Appraisal and Selection Management

Purpose: Recommends activities for evaluating metadata and selecting for long-term curation and preservation. An appraisal is the process of evaluating records to determine which are to be retained as archives, which are to be kept for specified periods and which are to be destroyed. Selection is a more general term, usually applied when deciding what will be added to a repository.

Outcomes:
- Policies are identified to guide the appraisal and selection of metadata
- Procedures are defined for the appraisal and selection of metadata according to the related policies.
- Procedures are automated for the appraisal and selection of metadata according to the related policies.
- Metadata are appraised and selected according to the defined policies and procedures.
- Metadata are disposed according to the defined policies and procedures.

Activities:
- **Defining appraisal and selection policies:** Appraisal should be done according to well-defined selection policy. A policy allows informed, consistent and accountable decisions about appraisal and selection of metadata to be made in situations where judgments are subjective and speculative.
- **Defining appraisal and selection procedures:** Appraisal and selection procedures indicate the means and considerations that are taken into account for selecting metadata.
- **Appraisal and re-appraisal of metadata:** Evaluate metadata regarding appraisal and selection criteria and select for long-term curation and preservation. Moreover, re-assessment of the appraisal and selection decisions may be required in order to accommodate changing metadata and curation requirements.
- **Disposing of metadata:** Dispose of metadata, which has not been selected for long-term curation and preservation in accordance with appraisal and selection policies and criteria. Typically metadata is destroyed. In other cases, metadata may be transferred to another archive, repository, or other custodians.

3.7 Metadata Quality Management Dimension

Description: The Metadata Quality Management Dimension consists of those processes required for detecting and preventing inconsistencies and defects in metadata as well as for improving the quality of metadata by cleaning such defects. The quality assurance aims to ensure that curated metadata remains authentic, reliable, usable, accessible to, and understandable over the long term.

Processes:
- **Metadata Quality Control**
- **Metadata Quality Improvement**

3.7.1 Metadata Quality Management Dimension

Process: Metadata Quality Control

Purpose: Recommends activities for measuring, assessing and ensuring the quality of metadata. Metadata Quality control is a set of procedures intended to ensure that the metadata adheres to a defined set of quality criteria, thus meeting the curation requirements.

Outcomes:
- A scope is defined for metadata quality in accordance with the metadata curation requirements.
- Policies are defined for supporting metadata quality management.
- Metrics are defined for metadata quality evaluation according to the related policies.
- Measurement methods are defined by which to determine values for metadata quality metrics.
- Metadata quality measurements are collected and monitored.
- The CurationMaster, MetadataCurators and other contributors are notified about metadata quality issues.

Activities:
- **Analyzing metadata quality requirements:** Metadata quality requirements are also generally considered part of requirements analysis.
- **Developing metadata quality policies:** Develop metadata quality policies that are appropriate for metadata curation strategy, comply with metadata requirements and establish the foundation for the continual improvement of metadata quality.
- **Developing metadata quality metrics and measurement methods:** Develop or select the metrics and correspond measurement methods used to measure/profile the metadata quality policies.
- **Measuring and monitoring metadata quality:** Measure and monitor metadata quality to evaluate the compliance with defined metadata quality policies. It is also important to establish appropriate resources to measure metadata quality.
- **Documenting and reporting metadata quality issues:** To provide awareness of curation activities, there should be periodic reports about most frequent quality issues and common resolutions.

3.7.2 Metadata Quality Management Dimension

Process: Metadata Quality Improvement

Purpose: Recommends activities for planning, implementation and control activities that apply quality management methods to improve and refine the fitness for the use of metadata.

Outcomes:
- Metadata quality issues are registered and manage.
- Metadata quality improvements are executed involving metadata cleansing and metadata enrichment.
- Metadata improvements are reviewed and homologated.
- Records are kept for all improvements made to the metadata.

Activities:
- **Managing metadata quality issues:** Register, track, evaluate and prioritize metadata quality issues and activities for resolving those incidents. A quality issue catalog may be used to log the quality evaluation, initial diagnosis, and subsequent actions associated with metadata quality issues.
- **Cleaning and correct metadata:** Detect and correct (or remove) errors and inconsistencies of metadata as well as identify incomplete, incorrect, inaccurate or irrelevant parts of the metadata and then replacing, modifying, or deleting the dirty metadata.
- **Enriching metadata: Enhance or improve raw metadata.** Although there are different ways to enrich metadata, a common metadata enrichment process could, for example, assign meaning to metadata by semantic enrichment. Another alternative is metadata linkage which is a technique for connecting pieces of metadata that are related.
- **Reviewing and homologating metadata improvements:** Inspect a set of modifications (cleansing and enrichment) over metadata to identify conflicts and the introduction of new defects.

3.8 Metadata Preservation and Dissemination Dimension

Description: The Preservation and Dissemination Dimension consists of those processes required for preserving metadata and ensuring that metadata is discoverable and accessible to both contributors and non-contributors. It includes processes to guide how to integrate metadata, transfer metadata to an appropriate repository and securely store them adhering to relevant standards. In addition, it also recommends processes to guide how to make metadata accessible by displaying publicly or by exposing them to other systems.

Processes:
- **Metadata Ingest Management**
- **Metadata Versioning Management**
- **Metadata Integration Management**
- **Metadata Access Management**

3.8.1 Metadata Preservation and Dissemination Dimension

Process: Metadata Ingest Management

Purpose: Recommends activities for transferring metadata to an appropriate repository for permanent (long-term) storage while maintaining and verifying the integrity of metadata.

Outcomes:
- Contributors are identified and authenticated.
- Metadata deposit agreement is collected.
- Provenance information about metadata is collected and properly managed.
- Metadata are transferred to metadata repository and as a consequence are securely stored.

Activities:
- **Performing authentication:** Confirm the identity of the contributor, who is contributing metadata to the environment. Common methods are password authentication or authorization via a digital signature.
- **Collecting deposit agreement:** Collect a certification by the metadata creator (i.e., the contributor responsible for creation or harvesting of metadata) that the metadata conforms to all policies and conditions (e.g., do not violate any legal restrictions placed on the metadata) and are fit for deposit into the repository.
- **Tracking and maintaining metadata provenance:** Track provenance is essential to the many domains where it can be used to evaluate the quality of metadata source, track the creation of intellectual property, and provide an audit trail for regulatory purposes. In particular, provenance is one kind of information which tracks the steps by which the metadata was created and
- **Storing the metadata securely:** Add the ingested metadata to a well-configured (in terms of hardware and software) metadata repository. Perform routine checks and provide disaster recovery capabilities as needed.

3.8.2 Metadata Preservation and Dissemination Dimension

Process: Metadata Versioning Management

Purpose: Recommends activities for managing metadata version. Metadata are not immutable objects; rather they are subject to changes. Moreover, several contributors may be involved in the creation of the same metadata content. Hence, it is important to identify the different version of metadata. Version management is the ability to manage the change metadata.

Outcomes:
- Metadata versioning strategy is defined for supporting metadata versioning management.
- Procedures are defined for metadata version identification according to the related policies.
- Revision control methods and tools are developed by which allow tracking and controlling over changes to metadata.

Activities:
- **Developing and maintaining the metadata versioning strategy:** Develop and maintain a metadata versioning strategy to keep track of different versions of a metadata. A metadata versioning strategy should envision a logical way to organize and control versions of metadata. It should include how to organize multiple versions within a metadata repository, and how best to describe them so they can be properly discovered and accessed.
- **Developing and maintaining metadata version identification procedure:** Develop and maintain a metadata identification procedure to identify metadata versions in a consistent, clear and transparent manner, so users always know which version of the metadata they have acquired.
- **Developing and maintaining metadata revision control:** Develop and maintain a revision control, which is a kind of practice that tracks and provides control over changes to metadata. It allows reverting a metadata to a previous revision, which is critical for tracking each other's modifications and correct inconsistencies.

3.8.3 Metadata Preservation and Dissemination Dimension

Process: Metadata Integration Management

Purpose: Recommends activities for combining metadata and presenting them in a unified way. Metadata integration consists of a set of techniques used for connecting silos of metadata. Thus, the integration of metadata enables to gain a more comprehensive understanding from metadata.

Outcomes:
- Metadata is analyzed and classified.
- Metadata is linked to another metadata items.
- Metadata is deduplicated.
- Metadata conflicts are resolved.

Activities:
- **Classifying metadata:** Metadata classification is broadly defined as the process of organizing metadata by relevant categories so that it may be used and discovered more efficiently.
- **Linking and combining metadata:** Link and combine the ingested metadata into a single population. In particular, such activity connects pieces of metadata that are thought to relate to the same resource, person, object or abstract entity/concept. Thus it brings separate metadata together to create a richer metadata repository.
- **Deduplicating metadata:** Eliminates redundant or repetitive information from metadata, which can reduce metadata storage costs, as well as speed up the access and discovery of metadata.
- **Identifying and resolving conflicts:** Since metadata may be acquired from different sources, it can provide conflicting information. Conflicts can arise because of incomplete, erroneous, or out-of-date metadata. It is thus critical for metadata curation to resolve conflicts from redundant metadata and identify true values from false ones.

3.8.4 Metadata Preservation and Dissemination Dimension

Process: Metadata Access Management

Purpose: Recommends activities for providing access to metadata. In order to be used, metadata must be accessible for both humans and machines. This process includes activities for creating the means for enabling machine-based search and retrieval functionality that help a user identifies what metadata exist, where the metadata are located, and how can they be accessed (e.g., download).

Outcomes:
- Metadata is indexed for supporting easy and fast access.
- Conditions and controls for controlling access to metadata are defined and established.
- Interfaces for enabling access to metadata are defined and established.
- Metadata is disseminated to external catalogs and platforms.

Activities:
- **Indexing metadata:** Develop or create indexing structures in order to improve the speed of metadata retrieval operations. Indexes are used to quickly locate metadata without having to search every metadata item in a repository every time a metadata is requested.
- **Determining appropriate access controls:** Determine what level of metadata access is required and how stakeholders and metadata consumers may be affected by this condition. In addition, depending on the conditions for access and reuse, place access restrictions on some or all metadata.
- **Designing and developing metadata access interface:** Design and develop access interfaces that support the search-and-retrieval functionality required for metadata assets. The search and retrieval services should help participants to identify what metadata exist, where the metadata are located, and how can they be accessed.
- **Disseminating metadata:** Active dissemination of metadata to search and discovery services (e.g., external metadata catalogs, web-based indexes) for federated search and discovery. Often, metadata is exchanged with external organizations through structured files such as XML and JSON formats.