Principles for long-term accessibility of information

eArchiving Reference Architecture

v0.3 Preliminary Review Version

(shared within a limited group)

CEF eArchiving Building Block, E-ARK3

CEF-TC-2019-3 eArchiving

# Cover Sheet

**Document Status:**

|  |
| --- |
| **Status** |
| Preliminary Review (shared in a limited group). Second round. |

**Document Approver(s)**

|  |  |
| --- | --- |
| Name | Role |
|  |  |
|  |  |

**Document Reviewer(s)**

|  |  |
| --- | --- |
| Name | Role |
|  |  |
|  |  |

**Summary of Changes**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Created by | Short Description of Changes |
| V0.1 | 17/01/2020 | István Alföldi | First draft version |
| V0.2 | 24/01/2020 | István Alföldi | Preliminary review version |
| V0.3 |  | István Alföldi | Preliminary review version, second round |
|  |  |  |  |

**Contributors**

|  |  |
| --- | --- |
| Name | Organisation |
| Aas, Kuldar | National Archives of Estonia |
| Alföldi, István | Poliphon |
| Holmelund, Frederik | The Danish National Archives |
| Jacobsen, Kristin | The National Archives of Norway |
| Kruithof, Gijsbert | National Archives of the Netherlands |
| Langeland, Øivind | The National Archives of Norway |
| Lund, Audun | Swiss Federal Archives |
| Merenmies, Markus | National Archives of Finland |
| Nielsen, Anders Bo | The Danish National Archives |
| Proença, Diogo | University of Lisbon |
| Schlarb, Sven | Austrian Institute of Technology |
| Skofljanec, Jože | Archives of the Republic of Slovenia |
| Skog, Björn | ESS Solutions |
| Silva, Hélder | KEEP SOLUTIONS |
| Tømmerholt, Phillip | The Danish National Archives |
| Trani, Silvia | Central State Archives |

# 

# Introduction

This document defines and describes the most important digital archiving principles that can serve as the basis of the reference architecture. Along with the eArchiving Reference Architecture – Outline document (setting up the foundation of our work) this document is part of the eArchiving Reference Architecture documentation. It represents the first outcome of the work.

**eArchiving Reference Architecture**

The implementation of an effective digital archiving solution requires an understanding of not only the technological components to be implemented but also the legal, organisational, and business context of an institution. The Reference Architecture workgroup is creating an eArchiving Reference Architecture to simplify the uptake of the eArchiving Building Block. This is:

* aligned with the best practices established by the European Interoperability Reference Architecture (EIRA);
* inspired by Enterprise Architecture conceptual models in the domains of digital preservation and information governance (for example, the outputs of the EC-funded SHAMAN project).

The task will start by describing a neutral core eArchiving Reference Architecture. The architecture will concentrate on mapping crucial strategy and business components of long-term accessibility and present a reference architecture connected to eArchiving specifications and Sample Software components.

Further, this task aims to contact the DG.DIGIT team responsible for EIRA and work together towards describing the CEF eArchiving Building Block as a component within EIRA. Similar actions are planned on the national level, where it is possible to work together with interested European states (e.g. Portugal, Denmark, Estonia, Norway, Finland) on establishing eArchiving as a component in national architecture.

**Background and purpose**

Principles are general rules that have a normative characteristic within a particular domain. We are aware that the principles do not necessarily reflect the current practice; they instead convey a conviction regarding how the desired situation can be reached. Having everyone following the same set of principles facilitates cooperation and sharing of ideas, practices and building blocks. The principles should be general and flexible enough to be stable over time and give room for variations, but at the same time specific enough to give guidance, have practical value and define some absolute boundaries. From a European perspective, it is particularly important to get this balance right to accommodate the differences between member states. The principles defined here apply to the domain of long-term accessibility of information. The principles in their current state may overlap and are not hierarchical.

**Documents of the eArchiving Reference Architecture**

At the time of publishing this v.03 release, the documentation of the reference architecture consists of the following documents:

* Motivation Aspect document – Detailed description of the motivation aspect elements (stakeholders, drivers, goals, and requirements) and their relationships.
* Principles – Separate document defining the eArchiving Principles (this document).

**Our approach**

We are following the guidelines of the Open Group Architecture Framework (TOGAF®). The principles below are classified accordingly into business, data, and application categories. Although the principles are not hierarchical in general, there is an implicit hierarchy in the sense that some of the data and application principles are consequences of the business principles and can be considered as derived principles formulated from a data or application aspect.

We are also following the description structure recommended by TOGAF®. The TOGAF® recommendations related to principles can be found in chapter 20 of the [TOGAF® document](https://www.opengroup.org/togaf).

|  |  |
| --- | --- |
| **Name** | Should both represent the essence of the rule as well as be easy to remember. Specific technology platforms should not be mentioned in the name or statement of a principle. Avoid ambiguous words in the Name and in the Statement such as: "support", "open", "consider", and for lack of good measure the word "avoid", itself, be careful with "manage(ment)", and look for unnecessary adjectives and adverbs (fluff). |
| **Statement** | Should succinctly and unambiguously communicate the fundamental rule. For the most part, the principles statements for managing information are similar from one organization to the next. It is vital that the principles statement is unambiguous |
| **Rational** | Should highlight the business benefits of adhering to the principle, using business terminology. Point to the similarity of information and technology principles to the principles governing business operations. Also describe the relationship to other principles, and the intentions regarding a balanced interpretation. Describe situations where one principle would be given precedence or carry more weight than another for making a decision. |
| **Implications** | Should highlight the requirements, both for the business and IT, for carrying out the principle — in terms of resources, costs, and activities/tasks. It will often be apparent that current systems, standards, or practices would be incongruent with the principle upon adoption. The impact to the business and consequences of adopting a principle should be clearly stated. The reader should readily discern the answer to: "How does this affect me?". It is important not to oversimplify, trivialize, or judge the merit of the impact. Some of the implications will be identified as potential impacts only and may be speculative rather than fully analysed. |

**Table 20-1**: Recommended Format for Defining Principles (Source TOGAF®)

TOGAF® also defines an example set of architecture principles. We consider principles like *Principle 1: Primacy of Principles* or *Principle 7: Compliance with Law* as basic principles fundamental to any approach describing an information system, but we do not copy them into our set unless they are closely related to digital archiving. Even in this case, we do not use the original wording but reformulate according to digital archiving concepts and language.

Our expected audience is not limited to traditional archives. One of the declared goals of the eArchiving Building Block is to broaden the scope and include a wider range of organisations dealing with mid- or long-term preservation issues such as commercial archives. Accordingly, we intend to use clear language that everybody can understand both within and outside the archival community.

As part of this intention, we will use wording like “Archiving is …”, “Archival systems are …”. Please note that the term “archiving” refers to digital archiving throughout the document and covers collecting, preserving, and providing access to information.

Archival terms and concepts are defined in many related standards (e.g. ISO 16175, ISO 17068 ISO 15489, ISO 14721, ISO 16363), and (to say the least) they are not consistent with each other. We do not intend to create a new set of definitions; on the other hand, we would like to use the archival terms consistently throughout the documents of the reference architecture. So, in the Glossary, we provide explanations about the intended meaning of the most important or most ambiguous terms.

As the title of this document shows, we would like to move the focus from long-term preservation towards the long-term access and reuse aspects of digital archiving. However, that does not mean that the significance of the traditional archival approach or the importance of the long-term preservation aspect is decreasing. (And – to prevent misunderstanding – long-term access definitely would not mean the archive constantly or automatically provides access to the content). It is rather a change in the perspective and is in-line with our intention to cover a broader spectrum of institutions providing or using archival services.

**Status of the document**

The principles were the first results of the discussions within the workgroup; however, this set of principles cannot be considered final. We have gathered a consistent initial set of principles sufficient to continue with defining the motivation aspect and the strategic and business layers of the Reference Architecture.

Although the principles belong to the motivation aspect, we created a separate document with the title *Principles for long-term accessibility of information* because of the significant interest in the work within the community. The principles were already published in a limited circle of experts, and the first round of comments was processed. The second revision of the principles document published along with the motivation aspect elements incorporates our answers to the comments.

We would like to share the principles, along with other sections of the reference architecture, with a restricted group of professionals to hear their opinions and suggestions. So, this document represents an intermediate stage of our work. It can be shared with anybody understanding the above limitations.

# 

# Principles

Business Principles

1. Archiving is focussed on use
2. Archiving is preserving information over technologies
3. Archiving ensures the trustworthiness of information
4. Archiving is organising and designing information and processes
5. Archiving is pragmatic, automated, and flexible
6. Archiving is an integral part of information management
7. Archiving takes into account rights, permissions and restrictions

Data Principles

1. Archival data are contextualised
2. Archived data are long-term accessible
3. Archived data are authentic
4. Archived data are reliable
5. Archived data and metadata make use of common vocabularies and definitions
6. Archival data and metadata are machine actionable

Application Principles

1. Archival systems and their components can be replaced with new ones as needed
2. Archival systems are implementation independent
3. Archival systems and their components are interoperable

Business Principles

1. Archiving is focussed on use

|  |  |
| --- | --- |
| Statement | The choices that are made with regard to collecting and preserving information are derived from the purpose for which that information is used, now or in the future. |
| Rational | This is the most important principle from which all other principles are derived. Archiving is essentially making choices. These choices determine which information is recorded by the data providers and must be kept accessible over time. The underlying question is what kind of information requires long-term accessibility. These choices are necessary to be able to target the finite resources needed to achieve sustainable accessibility. This explicitly includes answering the question of how long in time, the information must at least remain accessible. |
| Implications | This principle implies that the archival value of the information is appraised. The appraisal should be risk based, taking into account the needs of current, and if possible, future user groups (designated communities). Within the appraisal process, one should make explicit (as far as possible) what information is to be kept, the level of importance and for how long and for whom. When possible, the identified user groups should be involved in the appraisal process. |
| Notes (if any) |  |

2. Archiving is preserving information over technologies

|  |  |
| --- | --- |
| Statement | The main goal of archiving is to ensure access to information and its context, as long as necessary, without being reliant on originating systems, technologies and media. |
| Rational | Unlike paper-based records, digital records can hardly be kept accessible by default for a long time because systems are replaced, formats become obsolete and physical media deteriorate. To keep the digital records accessible, they must be migrated periodically to both new systems, formats, and media. Because the new system or format often provides different formatting, preservation, or presentation options than the original ones, each of those migration points represents a moment of risk. The goal of digital archiving is to tackle these moments of risk and to prevent losing context and information value over time. This principle should drive the archival decisions behind extracting and reformatting the information when migrating, so the context and information value is preserved in the next period. |
| Implications | This principle implies that when migrating, it is clear how the migration will affect context and information value and whether or not this poses an acceptable risk. It can sometimes be hard to identify the information to be preserved, and how the context and information value is represented in the data source (files, database, system, etc.). Even migrating simple text file formats requires archival decisions (e.g. formatting options and metadata fields on the file level). Also, the context of archived material should be taken care of with migrations (e.g. in the case of databases). |
| Notes (if any) |  |

3. Archiving ensures the trustworthiness of information

|  |  |
| --- | --- |
| Statement | Users must be able to assess the trustworthiness of the preserved information. |
| Rational | Preserved information is going to change in the long-term because of the migrations of software and hardware. These changes should not affect the trustworthiness of the information. Archiving should ensure the trustworthiness of the preserved information over time. |
| Implications | * Both preservation and data management activities should support the ability to trace changes, including migrations made and how those have changed the structure or content of information.   – When designing processes and metadata structures, metadata should be provided to ensure trustworthiness. |
| Notes (if any) | See data principle 10 about authenticity and principle 11 about reliability regarding how to underpin trustworthiness. |

4. Archiving is organising and designing information and processes

|  |  |
| --- | --- |
| Statement | Choices about how the information is archived, are to be made during the design of work processes and the information system. |
| Rational | When designing work processes and designing information systems that support work processes, it must be determined which transactions should be archived in which way. Adding an archiving function after the work processes and information systems are fully operational is too late because it requires changing the process and the systems. It will require extra time, effort, and money. |
| Implications | Records managers and archivists should be involved through all phases of designing work processes and information systems, to ensure that the concern for archiving is taken into account from the very beginning and considered in all relevant design decisions. |
| Notes (if any) | Archiving is primarily a designing activity focused on keeping information components of transactions over time considered as relevant. Archiving is focused on “replaying of sequences of actions in order to prove or disprove the state of something or the conclusions reached.” (Upward, Read, Oliver, Evans: Recordkeeping Informatics for a Networked Age, Monash University Publishing, 2018) |

5. Archiving is pragmatic, automated and flexible

|  |  |
| --- | --- |
| Statement | Archiving implements solutions which are sufficient, as automated as possible, and easily changeable in the future. |
| Rational | Long-term access to information is to be achieved on a sufficient level and should not require too many resources (like storage, electricity, staff, etc.). This means we cannot impose an absolute set of standards and guidelines that will fit any institutional need. An archive must always take into account that the measures and processes implemented are pragmatic, sufficient; and politically, technically, organisationally and financially possible.  One of the key risks in digital archiving is the availability of staff skilled in the relevant aspects of archiving, digital preservation, and IT. In most organisations, it is, therefore, reasonable to automate archival processes to the largest possible extent. |
| Implications | – Information should be categorised in risk categories. So, the efforts in archiving should be appropriate to the value of the information.  – Archiving can either be implemented as a component or set of functionalities within larger information systems or as a separate “digital archives”. Such architectural decisions must be made by weighing the cost and efficiency of either solution.  – Archiving should, as much as possible, make use of international standards, specifications, and tools to lower the total cost of ownership and enhance interoperability.  – Institutions implementing long-term availability measures must analyse internal skills and competencies, and set the minimum requirements for personnel to be available internally (versus external contracts, such as the use of archival service providers).  – Records management functionality and the transfer of content between different information systems (either different generations of the same system or between business information systems and archival components) can be a significant cost component. This relates to practically any data production and management systems with long-term preservation consequences both at the data producer and at the archival institution sites. Institutions can save a lot when taking these needs into account while designing business processes and business information systems (see also Principle 6). |
| Notes (if any) |  |

6. Archiving is an integral part of information management

|  |  |
| --- | --- |
| Statement | Archiving is to be seen as a part of information (lifecycle) management. |
| Rational | A number of tasks have to be carried out to ensure long-term accessibility (for example; the information has to be enriched with sufficient metadata, physical storage has to be reliable and accessible, file format risks and/or system dependencies have to be analysed and solved).  Many of these tasks are not unique to archiving but also relevant to information management in general. As such, “good archiving” does not start once information is delivered to a separated archival repository, but prior to this, with the application of best practices during the creation and initial management of the information. |
| Implications | – Archiving has to make sure that the standards and tasks it uses are in line with the standards and tasks used in information creation and early management.  – Archiving must be capable of collaborating with information management, to discuss and propose best-practices which have to be implemented for the early stages of the lifecycle.  – There is no single “correct” way of integrating archiving and information management. It is rather about choices in organisational and system design if archiving is separated both functionally and architecturally, or indeed natively integrated into information management organisations, functions, and systems (either way, the two implications above remain valid). |
| Notes (if any) |  |

7. Archiving takes into account rights, permissions and restrictions

|  |  |
| --- | --- |
| Statement | Archiving respects legal restrictions to the reuse of data, and makes sure that information is accessible only when appropriate rights and permissions are in place. |
| Rational | Archived information can have a number of different reasons to not be publicly and freely accessible. In many cases, the restrictions fall into the categories of Intellectual Property (IP) and personal information (GDPR). Elsewhere, trade secrets and confidentiality can be relevant.  Regardless of which restrictions apply, archiving has to make sure that the information containing restrictions is correctly identified and appropriate access mechanisms are in place. |
| Implications | – Archiving has to make sure that its setup is sufficient and appropriate to the level of restrictions applicable to the information.  – In most cases, a security policy, governance processes and actors will be required; this must be capable of both avoiding compromises and reducing liabilities.  – Legislation which poses reasons for restricting information or prescribes methods for providing access can change over time. Archiving has to make sure that its processes in regard to managing access rights and restrictions, and appropriate metadata can be updated whenever necessary. |
| Notes (if any) |  |

Data Principles

8. Archival data are contextualised [focusing on the semantics]

|  |  |
| --- | --- |
| Statement | Archival data are sufficiently contextualised to ensure that it can be correctly interpreted and used through organisational, legal, and other changes. |
| Rational | One of the risks in long-term accessibility is the constant change of our legal and organisational environment, both globally, nationally, and locally. This means that even data created 10–20 years ago can easily be misinterpreted.  An archive is different from simple “storage” because it attempts to document the context of data (e.g. how it was created, by whom, why, under what circumstances, and how it has been used and maintained).  Such contextualisation allows future users to understand the processes, institutions and persons involved with the data, helping to establish and maintain the long-term provenance and correct interpretation of data. |
| Implications | Contextualisation implies that sufficient context information (metadata) is provided with the data. How this context information is provided will vary and can range from unstructured traditional archival descriptions to machine-readable strictly structured metadata fields. |
| Notes (if any) | Reworded from the International Council on Archives (https://www.ica.org/en/what-archive). |

9. Archived data are long term retrievable [focusing on the technical aspects]

|  |  |
| --- | --- |
| Statement | Archiving should ensure that information is identifiable, retrievable and renderable. |
| Rational | Digital information is vulnerable because of changes in data preserving technologies. Archiving should ensure that the information is obtainable despite all changes in hardware and software. There is a risk that digital data and its conditions of use have changed to such an extent that storage technologies cannot restore their value and accessibility. |
| Implications | The preservation processes, standards and tools are used to facilitate interpretation of data as information for designated communities.  Proactive preservation planning and sufficient technical metadata are needed to support migration activities. |
| Notes (if any) |  |

10. Data in archives are authentic

|  |  |
| --- | --- |
| Statement | Users must be able to establish the authenticity of archived information. |
| Rational | Authenticity proves that the information object is what it claims to be, that it was created or submitted by the person or organisation that claims to have created or submitted it, and that it was created and submitted at the time indicated by the information object.  Authenticity is dependent on the integrity of the information, meaning that the information has not been subject to any unauthorised changes. |
| Implications | Appropriate organisational and technical measures must be taken to make sure that data are authentic throughout its lifecycle, and that integrity is maintained. Sufficient metadata must be applied to the information object as early as possible in the lifecycle to make it possible for the user to assess the level of authenticity. The information objects must be protected from any unintended changes. |
| Notes (if any) | Taken from ISO 15489. |

11. Data in archives are reliable

|  |  |
| --- | --- |
| Statement | Users must be able to establish the reliability of archived information. |
| Rational | A reliable record is one whose contents can be trusted as a full and accurate representation of the transactions, activities, or facts to which they attest. Reliability is dependent on the integrity of the information, meaning that the information has not been subject to any unauthorised changes. |
| Implications | Appropriate organisational and technical measures must be taken to make sure that data are reliable throughout its lifecycle, and that integrity is maintained. Assessing reliability cannot be assured by metadata alone. It also depends on the culture of the records creator, in their willingness to implement the necessary measures, both organisational and technical, so that the whole personnel of the records creator is willing and able to secure reliable records.  Finally, to demonstrate data reliability, institutions must ensure that they have the means to provide all the objective evidence whenever it is requested for auditing purposes. |
| Notes (if any) | Taken from ISO 15489. |

12. Archived data and metadata make use of common vocabularies and definitions

|  |  |
| --- | --- |
| Statement | Archived data are defined consistently, and the definitions are understandable and available to all users. |
| Rational | Data should have an explicit common definition throughout the archive to enable sharing of data. The use of common vocabularies will facilitate communication and enable dialogue to be effective. |
| Implications | – Resources must be committed to this task. It is key to the success of efforts to improve the information environment.  – The archive administration must establish the initial common vocabulary, and these definitions will be used uniformly.  – Whenever a new data definition is required, the definition effort will be coordinated and reconciled with the archive “glossary” of data descriptions. The archive administrator can provide this coordination.  – Ambiguities resulting from multiple parochial definitions of data must give way to accepted definitions and understanding.  – Multiple data standardisation initiatives need to be coordinated.  – Functional data administration responsibilities must be assigned.  – While the use of external sources (ontologies, vocabularies) is recommended, archives have to be aware that these external sources have to be managed as well, in the sense of long-term accessibility, versioning, etc. Should the external ontologies and vocabularies change or disappear, in the worst case, the archived data might become unusable or lose authenticity. |
| Notes (if any) |  |

13. Archival data and metadata are machine-actionable

|  |  |
| --- | --- |
| Statement | Computational systems must be able to find, access, interoperate, and reuse data with none or minimal human intervention. |
| Rational | To cope with the increase in volume, complexity, and creation speed of data, humans rely on computational support to process data in a meaningful and scalable manner.  This is especially relevant when defining and managing metadata. |
| Implications | Data in archives should be Findable, Accessible, Interoperable and Reusable for both humans and machines. Specifically, data in archives:   * [Are assigned a globally unique and persistent identifier](https://www.go-fair.org/fair-principles/fair-data-principles-explained/f1-meta-data-assigned-globally-unique-persistent-identifiers/). * [Are retrievable by their identifier using a standardised communications protocol](https://www.go-fair.org/fair-principles/542-2/). * [Use a formal, accessible, shared, and broadly applicable language for knowledge representation](https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/). * [Are released with a clear and accessible data usage license](https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/). |
| Notes (if any) | Taken from FAIR <https://www.go-fair.org/fair-principles/> |

Application Principles

14. Archival systems and their components can be replaced with new ones as needed

|  |  |
| --- | --- |
| Statement | Archiving must take into account that the information content of data it gathers, preserves and provides access to, must be maintained across many generations of software in an efficient way. |
| Rational | Digital archiving is technically implemented as a set of hard- and software; it is essentially an information system like any other, prone to IT obsolescence. As such it is known from the moment one starts to develop a digital archive, that the system has to be updated or even replaced regularly in a sustainable and (resource-)effective way, all while making sure that the accessibility and authenticity of data are not endangered. |
| Implications | It is crucial to make sure that the technical design and component architecture of an archive are well thought through and futureproof. For example, one might take into account the following requirements while developing a digital archive: a clear exit strategy must exist; an archival infrastructure is reasonably modular; standardised communication/APIs are used to interact between different modules of an archive; the data are stored in a system-agnostic way; the technologies used within an archive are widely supported (to ensure that sufficient technical expertise exists). |
| Notes (if any) | Inspired by TNA Digital Strategy: “Each new technology brings a fresh challenge for the digital archive, and this change is continual. We were not worrying about big data analytics, distributed ledger technology or the internet of things ten years ago. Digital archives must ride the rise and fall of successive waves of technological change and make decisions about where best to apply their efforts.” |

15. Archiving functionality is system independent

|  |  |
| --- | --- |
| Statement | Archiving of information can take place within any information system. Provided that the criteria of long-term accessibility and the requirements regarding publicity and use determined by the legislator are met. |
| Rational | The technical (architectural) implementation of the archival repository is not relevant. It is not necessary to move digital information to an information system specially designed for archiving. For paper information, there is often a physical reason to move information (cupboards become full, and users want easy access to the pieces). This does not have to be the case for digital information. If the source system has reasonably implemented the core digital preservation/archiving capabilities and functionalities, it is also fulfilling the business principles/requirements of a digital archive. |
| Implications | Systems must be documented in a way that makes it possible to verify that concerns and requirements regarding archiving are met. |
| Notes (if any) |  |

16. Archival systems and their components are interoperable

|  |  |
| --- | --- |
| Statement | Archival systems and their components must conform to standards and specifications to promote interoperability for data, applications, and technology. |
| Rational | Standards help ensure consistency, thus improving the ability to manage systems and improve user satisfaction, and protect existing investments, thus maximising return on investment, and reducing costs. Standards for interoperability additionally help ensure support from multiple vendors for their products. |
| Implications | * Interoperability standards and industry standards will be followed unless there is a compelling business reason to implement a nonstandard solution. * A process for setting standards, reviewing and revising them periodically, and granting exceptions must be established. * The existing solutions must be identified and documented. |
| Notes (if any) |  |

**Glossary**

Archival terms and concepts are defined in many related standards (e.g. ISO 16175, ISO 17068 ISO 15489, ISO 14721, ISO 16363), many of which are not consistent with each other. We do not intend to create a new set of definitions; on the other hand, we would like to use the archival terms consistently throughout the documents of the reference architecture. So, in the Glossary, we provide explanations about the intended meaning of the most important or most ambiguous terms.

**Archive and Archiving**

In professional terminology, *Archive* can mean either a physical repository, a set of (archived) information or an (archival) organisation. The E-ARK Reference Architecture uses the term *Archive* to refer to any group of people and associated infrastructure which implements long-term digital accessibility measures. As such an *Archive* can be anything from a single person to specific large archival organisations.

Accordingly, the term *Archiving* refers to the processes carried out by the *Archive,* which can be implemented either as small functional components or processes within larger information environments, as a specific large-scale digital preservation infrastructure or anything in between.

**Information versus Data**

*Information* is considered to be the main asset of archiving. The word *information* is used whenever the archived content along with its metadata, and the context is referred to. *Data* is only used when we refer to the content bitstreams and files which are managed by archiving.

**Archival Data versus Archived Data**

The term *archival data* refers to any data element anywhere throughout the entire archival process from pre-ingest to access. *Archived data*, on the other hand, only referred to data elements already stored in an archival repository (e.g. in an AIP).

**Available versus Retrievable versus Accessible versus Usable**

These terms are not easy to differentiate. They are synonyms in everyday language but used in very specific meanings in several standards. Unfortunately, these specific meanings are inconsistent in those documents. We have decided to use them in the following meaning:

If *available*, then data in its purest sense are there. If you have methods and tools, you can *retrieve* them. If it is *available* and *retrievable*, then the information is *accessible* if you have the proper rights to access it (or in general, if one could have the rights to access it). Finally, information is only *usable* if you can view and interpret it along with its context.

**Trustworthiness versus Reliability versus Integrity versus Authenticity**

One of the key aspects of any archive is to ensure that the information it keeps can be trusted. More specifically, the E-ARK Reference Architecture understands *Trustworthiness* as a quality of both the Archive and the Information it holds.

For Archives, *Trustworthiness* is about carrying out their processes in a controlled and repeatable fashion.

For Information, *Trustworthiness* does have the sub-characteristics of *reliability*, *authenticity* and *integrity*:

* *Reliable* information is one whose content can be trusted as a full and accurate representation of the transactions, activities, or facts to which it attests and therefore can be depended upon in the course of subsequent transactions or activities.
* *Authentic* information is one that is proven to be what it purports to be and to have been created by the agency with which it is identified.
* *Integrity* of information refers to it being complete and unaltered.