Archiving by design – what it is and how it can be used

A discussion paper presented to the EAG-meeting, June 2021

Date: 31.05.2021.

Background

The EAG meeting in Helsinki, Finland in September 2019 decided to establish a subgroup for developing Archiving by Design [Insert link to mandate]. The subgroup was tasked with addressing a common challenge for European countries - to handle digital born data to ensure that data persists with availability and authenticity over time.

Due to the Covid-pandemic and travel restrictions the subgroup has primarily scoped the work around development of principles and methodologies that can enable National archives implementing Archiving by design in their countries. The baseline is conceptual thinking established in the Netherlands. This paper gives the current thinking of the subgroup on establishing a theoretical framework for Archiving By Design.

The subgroup has developed this discussion paper to elaborate and explore the possibilities of "Archiving by design" (in short, “AbD”) and its relation to information systems, record capture, and submission to the archives.

This paper consists of 3 parts; We present an introduction to the concepts of Archiving by design, a theoretical analysis and some thoughts on how to make the concept *of Archiving by design* operational. Attached to the paper is a draft model of requirements. This model is adapted from the “DUTO requirements”, developed in the Netherlands.

Going forward the subgroup intends to perform pilot studies in several national environments in cooperation with national agencies testing the ideas and requirements. Combining pilots with a theoretical framework can lead to practical methods for working with *Archiving by design* on a European level, in compliance to a diversity of regulations on a national level.

Proposed actions for the AbD subgroup going forward

The AbD subgroup has carried out a few pilots on developing step-by-step approaches on how to work with national agencies when an agency is about to develop a new IT system. Developing a best practice methodology should be on the agenda for further work when the COVID pandemic has ended.

The subgroup could develop the concept and methodology further through pilot projects in participating countries through the following actions:

- **Assessing if information systems are archive-ready:** Try to use and enhance the AbD requirements checklist including the perspective of various purposes - What other questions should be asked? What is relevant to include in developing a workbook for performing an “*AbD scan*”?
- **Recommendations on retention and destruction in information systems:** Try to identify “triggers” regarding when and how information should be “fixed”. Is that even feasible in practice?

- **Auto-capture and auto-classification:** Find out what is “the state of the art” regarding automatic capture. Can we give the national archives any concrete advice, and maybe even point to solutions that work in practice?

Introduction to archiving by design

Archiving by Design (AbD) is a generic term for the mindset of designing information systems that consider from the outset the need for sustainable information accessibility as long as necessary.

The subgroup has worked on the concept of Archiving by design adapting the Dutch thirteen requirements into European ones. We have focused on a general understanding of what digital archiving must consider, and explored the range or boundaries of our common theoretical understanding of digital archiving, which we present in this paper.

We have arrived at the position that Archiving by design is to manage information in systems that have been examined, analyzed and found compliant to aspects of information management and governance according to requirements set in the model grouped in five areas of examination, namely (*here the groups that we arrive at).* It is a method to ensure that all aspects of digital information management concerning archiving is taken account for when establishing and developing information systems in all sectors of society but primarily public sector and public information.

Archiving by design – descriptions of the components

Archiving by design consists of three components:

- Two basic principles, each with a set of characteristics
- Set of standard requirements, linked to the characteristics of the principles
- Methodology of examining the requirements (criteria)

Basic principles

Information must be sustainable and accessible, and these two are the basic principles of Archiving by design. The two principles are basic in the original Dutch concept and have not been questioned in the present adaptation for EU use. On the contrary, the two principles meet a fundamental agreement across the union member countries that participate in the AbD subgroup. The principles should be understood in relation to the following characteristics, that are commonly applicable, but in every case dependent upon values in the legislation of each member state. They therefore need to be interpreted from a genuine knowledge of each context.

The principle of accessibility means that the information should be

- Findable
- Available
- Readable
- Interpretable
- Reliable

The principle of sustainability means that the information should be

- Resilient
- Sufficient

All the characteristics are found in the original Dutch concept except *sufficiency*, that have been added by the AbD subgroup because it is a substantial part of every contemporary discourse dealing

with topics on sustainability (economic, social, environmental etc). It must also be considered in discovering what sustainability implies in the information management context where a sufficient amount of information must be created and kept. Too much would be too expensive, and difficult to overview and therefore risk information-loss. Sufficiency can also be related to the EAG ambition that a sufficient amount, and not too little, of public records must be preserved.

Set of standard requirements

Archiving by design contains ten standard requirements. The requirements are linked to the characteristics of the principles as follows. (For more details on the Archiving by Design requirements checklist, see the separate Excel file):

Accessibility - Information should be

- Findable: Search and representation (Req 2.1)
- Available: Right of access (Req 4.1)
- Readable: Search and representation (Req 2.1)
- Interpretable: Metadata (Req 2.3)
- Reliable: Information value (Req 1.2), Security (Req 4.2)

Sustainability - Information should be

- Resilient: Retention plan (Req 1.3), Export (Req 3.2), Open formats (Req 2.2)
- Sufficient: Information model (Req 1.1), Information value (Req 1.2), Destruction (Req 3.1)

Methodology

When using Archiving by design as a tool the requirements need to be examined regarding the information objects contained in a system. In order to facilitate this process of examination the requirements have been added criteria and are grouped according to aspects examined. The criterias appears in the attached excel sheet.

The requirements are presented in four groups in order to facilitate the examination of compliance. These are 1) Information modelling and classification (req 1.1-1.3), 2) Information representation , format and metadata (req 2.1-2.3), 3) Destruction and export (req 3.1-3.2), and 4) Restrictions and security (req 4.1-4.2).

There are two important guiding stars for using the *Archiving by design* methodology:

1. Don´t impose if there is no need to, instead recommend and help the public bodies define their own needs of keeping information authentic and findable.

2. Perform an Archiving-by-design scan, or examination - a workshop methodology connecting the requirements to user needs. The new provenance is trusting the needs of the users.

Archiving by design – theoretical analysis

Archiving by design has potential implications relating to information systems design, record capture, and submission to the archives. Building on design principles it represents a new approach to archiving where user needs are prioritised and key to the approach. This represents a shift where trusting in building systems based on user needs will also meet requirements of records capture and submission to long term preservation.

Archiving by design should be understood as an equivalent to the well-known concept of privacy by design, that is built-in technology in systems that supports the regulations and prevents errors. By

using the method Archiving by design unwanted effects in the information governance should be avoided, records would be captured, archived, and managed long-term.

At the core of our *Archiving by design* about moves towards understanding is thinking of the concept of archiving as an on-going process through the lifecycle of an information object - , all the way from the creation of an information object, defining the object as part of an passing the point of being added to the archive, and through the measures taken during the course of time in order to preserve or dispose the same information object. Archiving, in the concept of *Archiving by design*, should be understood as both short term and long term preservation and management, and can take place as well in a system used in operation as in a system used only as an archive.

Archiving by design has the potential of connecting issues of GDPR and information security with archiving and solves them through the same approach together. Together with open access and inclusiveness these issues consist of the "big 5" in information governance and management. Archiving by design therefore consists of components supporting all 5.

Archiving by design in information systems

According to the above-mentioned definition, Archiving by design generally means that the accessibility of relevant information in the relevant form on the short, medium and long term should be taken into account when designing and implementing a new information system. Archiving in this context is to be understood both as capture as well as the necessary steps to ensure the long-term accessibility of information.

The term "archiving" needs to embrace the fact that different legislations affect the conditions that determine how and when the shift of status from created document/record to archived data takes place. It can not only be understood as long-term accessibility but also short-term.

Information systems can be understood as systems used by the public bodies for their operations, i.e business systems or information systems apart from e-archives. Archiving does not take place when the system needs to be changed or else information migrates, but much earlier and often close to or at the point of creation of records.

Archiving by design has a major strength in this area due to the principle of "archiving in the source", of keeping the archived data within the system used in the work process, and not migrating it in order to consider it archived. The Dutch concept of *Archiving by design* enhances the need for keeping the information in the source system for as long as possible according to your information management model – every migration and export puts the information at risk. However, in order to comply with GDPR and to avoid format obsolescence, it is critical that the possibility of exporting data to an archive is considered at an early stage.

Archiving by design in records capture and work processes

A well-known challenge for information governance and record keeping is the capture of records. Records management systems are usually based on the manual declaration of records, i.e. the requirement for a caseworker to manually decide which documents should be declared records and, to some extent, manually record metadata. Some classification can be done automatically and some metadata can automatically be extracted from the documents, but usually it has to be manually adapted and corrected.

One important issue here is that the applications and systems that are used for the creation of records are also used for handling information that either definitely not or possibly not should be declared records. E-mails with irrelevant information is an obvious example but also applications such as Word and Excel pose a challenge because we write numerous drafts etc. that never become actual records and never should be "archived". If "archiving by design" also means "record capture

by design", it is obvious that a successful development of the concept will require a lot of research to make sure that the archives eventually only will contain information that should be archived.

A true challenge for Archiving by design would be to support the capture of records in systems by providing methods for using technology to determine what types and kinds of formal and informal records the information system contains or will be containing. Written documents, generated reports, signatures, pictures, sound, video, metadata, registers, functionality, code et cetera. Naming and describing the different records in order to connect and ensure certain actions in the information governance processes, are fundamental tasks in archiving, and can it be made partially by design, much would be gained when it comes to meet with the purpose set out by the EAG, namely to ensure more public information to be preserved within the member states.

Archiving by design in submission to archives

Medium-term accessibility can be achieved through system design. However, indefinite preservation and accessibility can normally not be obtained within one system as IT systems will face technological obsolescence. Long-term accessibility will most likely require a migration to preservation formats and a system-independent information package. In order to be meaningful in an archival context, "archiving by design" needs to include functionality to export data and records in system independent formats and structures that can be submitted and maintained by an archival institution.

However many current IT-systems lack good data model descriptions and functionality to export information in system independent formats. It can therefore be costly and difficult for the records creators to submit data and digital records to an archival institution. Thus, it is worth discussing whether the concept *Archiving by design* could also be used for the process of designing a data model for the best possible submission to an archive, even though it might imply a difference from the data model in the actual it-system where data was created. This standardization is already work in progress in eArchiving building blocks / E-ARK.

For all types of records in every system there must be some kind of strategy that guides the information governance towards the end of the lifecycle of information. The process of submission to a later system environment must take into account that information that is to be preserved forever needs to be transferred eventually.

Archiving by design – thoughts on how to make it operational

When using AbD to help agencies develop new IT systems where data is accessible and sustainable over time there is a need for a set of tools. There are many possible approaches. Building on the work from the Netherlands the subgroup proposes a workshop-oriented method of scanning and mapping the information needed for short, medium and long term.

With the help of the AbD approach, archiving tasks can be addressed at the very beginning of the information lifecycle, through the design of specialised applications, in which the information is created and managed.

In order to make this work in practice, it is important to make sure that we ask (and answer!) the right questions when designing a new IT-system.

Identify purpose for keeping information

For a successful implementation of AbD, it is essential to know the purposes for which we capture and keep information. What are the purposes, and will they change depending on the term?

The requirements could, e.g., be grouped under headings such as "operational", "legal" and historical, and those could again be viewed differently based on whether the requirement is relevant

on short term, medium term or the long term. For example, the documentation needed for maintenance or operation in the short term may need to be more detailed than the documentation needed for historical evidence in the long term, the same ways as there might be specific requirements for legal purposes in the short or medium term that become irrelevant in the long term.

Identify requirements for keeping information

The point of identifying the purposes for capturing and keeping information is that it is necessary in order to define the requirements for the information; requirements that might change in the life-cycle of the information.

Archiving by design methodology (AbD scan/examination) connects the user needs (short term, medium term and long term) to the set of requirements that this concept consists of.

For instance, it might be very important for short term operations reasons to keep information in a very specific and complex file format whereas it might be important that that information can be migrated to a format better suited for long-term preservation when no longer needed for operational purposes.

ARCHIVING BY DESIGN: REQUIREMENTS CHECKLIST

Version: 1.0
Date: 31.05.2021

Acknowledgements

This Archiving by Design requirements checklist is based on the DUTO criteria, first developed by the National Archives of the Netherlands (Nationaal Archief). The initial Dutch requirements have been further developed by the EAG Archiving by Design subgroup in 2020 – 2021.

Purpose and use

Archiving by Design (AbD) is a generic term for the mindset of designing information systems that consider from the outset the need for sustainable information accessibility as long as necessary - even if the information must remain accessible for centuries or permanently.

Information systems, which follow the AbD mindset, are capable of keeping information in an authentic, secure and easily reusable way as long as needed and support natively the controlled destruction or transfer of the information.

This document includes the Archiving by Design requirements checklist - a list of ten most crucial high-level requirements that an organisation must implement within their information systems in order to be compliant with the Archiving by Design mindset.

The Archiving by Design requirements checklist is intended to be a practical tool used in at least (but not limited to) the following scenarios:

- An agency can use the checklist to carry out a self-evaluation of the long-term sustainability and accessibility while preparing for the development (or tender) of a new information system;
- An agency can use the checklist as a reference when testing or accepting new information system developments;
- An agency or an archives can use the checklist for carrying out a gap analysis (i.e. what should be improved in an already established information system);

- An (public sector) archives can use the checklist as an additional tool while consulting agencies in regard to their digital recordkeeping, data management and archiving procedures.

Requirements overview

The ten Archiving by Design requirements are split into four areas. Table 1 provides a basic overview of the areas and requirements. Detailed descriptions of the requirements, including implementation advice and checkpoints, are available in a separate spreadsheet.

Table 1: Archiving by Design requirements checklist overview

Requirement name	***Description***
1. Information modelling and classification	
1.1 Information model	The organisation maintains an information model that describes all the information objects that the information systems contain. The information model must present the information objects in a way that allows them to connect to the organisation's grouping or classification of all its information.
1.2 Information value	The organisation has assessed all its information objects for their reuse value and required accessibility levels. The reuse value of information can for example be legal, administrative, financial, historical etc.
1.3 Retention plan	The organisation has assigned a retention period for each information object and/or group based on the value of the information object and/or group.
2. Information representation, format and metadata	
2.1 Search and Representation	A representation of each information object is defined and available within the information system. A search function is available that makes it possible to easily find the information object.
2.2 Open formats	Information objects (i.e. their representations) are stored or can be easily exported in open and standardised formats.
2.3 Metadata	Each information object is accompanied by and can be easily exported with complete and up-to-date metadata. Sufficient metadata must also be available for the search function defined in requirement 2.1.

3. Destruction and Export	
3.1 Destruction	Information objects are destroyed no sooner and no later than indicated in the retention plan. Such controlled destruction of an information object must be documented.
3.2 Export	Information objects can be exported in a standardised and controlled manner for transfer to an archive or any other information system.
4. Restrictions and security	
4.1 Right of access	Information objects are accessible to anyone who is entitled to access on the basis of regulations and policies. If an information object includes some restricted elements, the system allows for the creation and management of an unrestricted representation of the information object.
4.2 Security	The information system complies with applicable institutional, national and international information security standards.