Archiving by Design requirements

In Archiving by Design, the overall quality requirements (sustainably accessible information described as findable, available, readable, interpretable, reliable and future-proof) are translated into practically feasible improvement measures in the process of procuring (e.g. through procurement) and/or implementing a new system or making adaptations to an existing system. This is done through an evaluation of ten requirements that are also known as AbD requirements. They are used in relation to user needs in various tools that are built around the AbD concept and methodology, like the AbD scan procedure.

- Information model
- Information value
- Life cycle planning
- Search function and representation
- Choice of format
- Metadata
- Deletion
- Export
- Right of access
- Security

The requirements are understood in relation to the Archiving by Design life-cycle perspective on information and its management. Archiving needs to be understood as both the activity that adds information to the archive – archiving – while archiving is also the process of managing information over time while maintaining sustainable accessibility throughout the information's life cycle.

In order to examine the requirements and how well they are implemented in an information system, checkpoints are used that together form a checklist. The checklist can be explored in workshops with users and information management experts, but documentation about the system, operations, and context is also a resource in the survey. The survey is carried out under the leadership of an AbD team that continuously analyzes the respective requirements in relation to intended system solutions and the users' needs.

Information model

Description	The organisation maintains an information model that describes all the information objects that the information systems contain.
Objective	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.
Implementation	An explicit listing and description of all information objects is the cornerstone for many archival (incl. Archiving by Design) actions. An information model allows stakeholders to understand the context of the information object, assess their value and need for restrictions, assess if information is complete and correct, and how it should be interpreted. For now the AbD requirements do not pose a specific template or notation for an information model. However, any reasonable information model should: * Describe the semantics of the information object, ideally be based on a solid business vocabulary; * Explain the structure of and relations between the information objects; * Include basic information about the use and management of the information objects (i.e. who are responsible and who are the main users of the object(s)); * Describe both the data and metadata of an information object; * Be developed using a standardised notation (e.g. UML); * Be available in an open format which is easily accessible for all relevant stakeholders (data managers, archivists, users, business owners); * Be regularly checked and updated in order to uphold its usability and accuracy; * Be automated if possible and reasonable (e.g. that updates in a database model are automatically reflected in the information model). Please note, that for larger information systems an information model might consist of multiple sub-models!
Questions	* Are all information objects covered by the information model? * Is all metadata covered by the information model? * Are the descriptions of the information objects sufficiently understandable? * Does the information model provide an overview of the owners and users of the information objects? * Is the information model available in a sufficiently standardised and accessible format? * Is the information model up to date?

Information value

Description	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.
Objective	An organisation must have a clear understanding of "why" it gathers and maintains the information objects and "how" these information objects must be available. Such insights allow the organisation to determine

	which information object must be sustained for the long-term and which must not.
Implementation	* Information value can be recorded as part of the information model described above; * Organisation should make an effort to document the justifications for the assigned value (e.g. is there a legal act requiring the collection of the information, which business function(s) make use of the information, which stakeholder groups reuse the information); * The value of information can be determined by using risk assessment techniques (i.e. what are the risks if the data is not available any more); * Accessibility levels should be recorded using a consistent taxonomy (e.g. public, available for registered users, available after 5 years, available if anonymised); * Information value is often seen as subjective. Make sure to involve ALL stakeholders into the assessment!; * In practice the assessment of information value can be done in parallel with assigning retention periods, and to some extent it is possible to also apply local, national or international appraisal techniques for this task and, depending on legal requirements involvement of archive, it is necessary or recommended to set the enduring value of information.
Questions	* Is the information value recorded for all information objects? * Are accessibility levels recorded for all information objects? * Are the explanations and reasoning for information value appropriate and understandable?

Lifecycle planning

Description	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.
Objective	The definition of retention periods allows organisations to implement appropriate information lifecycle events into information systems (destruction and export for archiving or migration to other systems). The timely destruction of information objects is also a crucial component of GDPR compliance.
Implementation	* Many countries have established rules for determining retention periods (i.e. appraisal). Make sure to check for your national rules with the appropriate records management or archival agency; * Note that in some cases it is necessary to have a different retention period for data and metadata (i.e. some metadata is left available after the retention period for data is over); * Retention periods should also be recorded within the information system, as the basis for executing relevant destruction and export events; * Retention periods (including the retention periods as recorded within information systems) should be possible to be reviewed and changed in order to meet with changing user needs and legislation; * Note that retention periods should be implemented across all relevant information systems (e.g. if the information object is maintained as versions in various systems).

Questions	* Do all information objects have a retention period? * Is the retention period logical and appropriate (in regard to the information value determined in information value requirement)? * Have relevant national appraisal rules and legislation been respected? * Have retention periods been implemented into the information system(s)?

Search and representation

Description	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.
Objective	Long-term accessibility of information means that at each point in time information objects can be searched for, and can be represented in a meaningful manner.
Implementation	* A search function should include both full-text and metadata based advanced search; * A search function should be possible to be limited on specific information objects; * One information object can have multiple representations, however there should be at least one full representation including all relevant data and metadata of this information object; * The search and representation must be available for all users who are entitled to do so; * The search and representation should be based on user needs analysis, ideally developed including UX design competencies; * If reasonable, search and representation capabilities could be structured across core user groups (e.g. employees and customers can have different possibilities to search for and view the information object); * Ideally one information system implements only one search engine for all its information objects (i.e. users do not have to switch between different search engines when looking for information); * The representation of an information object should be connected to a permanent URI (URL, DOI etc.) which allows it to be referred or linked.
Questions	* Does the full-text search cover all information objects? * Does the advanced search cover all information objects? * Are search and representation capabilities appropriate for defined user groups? * Are search results consistent and reliable? * Are representations of information objects using a permanent URI (URL, DOI etc.)?

Preferred/open formats

Description	Information objects (i.e. their representations) are stored or can be easily exported in open and standardised formats.

Objective	The use of preffered/open formats in information storing and/or export extends the practical accessibility and usability period of the information.
Implementation	Binary file formats: * Establish a list of preferred formats (open standards or formats that best meet the objective); * Prefer well standardised and widely used formats within this list; * Lists of "archival file formats" have been prepared by many national archives across Europe; * Implement file format identification procedures in appropriate information creation, information receipt and/or export workflows; * When receiving files in other formats it is recommended to keep the original file, and implement appropriate migration procedures for export purposes only; * Make sure to test any file format migration software thoroughly before implementing it; * Regularly review the list of preferred formats, appropriate information creation and file migration tools. Data formats: * Use open and easily reusable formats for data exports (TXT, XML, RDF, JSON, CSV, SIARD with attached templates); * Make sure that appropriate metadata (e.g. column names) are possible to be exported along with the data; * Use widely accepted XML schemas and other data structures whenever possible; * Use standard (SQL, XML, JSON or JSON-LD) data types as much as possible.
Questions	* Is an preffered/open format representation or export available for all information objects? * Is a list of preferred/open formats available and regularly reviewed? * If file migration tools are implemented, have these been thoroughly tested? * Are data (export) structures and formats defined and maintained? * Are data (export) structures and/or file formats appropriate for their intended users?

Metadata

Description	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 Search and representation.
Objective	Appropriate metadata allows the information objects to be searched and managed. Appropriate metadata provides context and provenance information relevant for the long-term usability of the information.
Implementation	* Check your national records management, archival and data governance bodies for possible national metadata requirements;

	* The quality and availability of metadata is especially crucial for the usability of information objects with long retention periods - make sure to prioritise these objects when developing metadata rules; * Automate metadata creation as much as possible; * Where metadata needs to be added manually, there are clear instructions on how and by whom to do so; * Create metadata during the creation of data. The later metadata is added, the greater is the chance of errors and/or high costs; * Make sure that appropriate administrative and provenance metadata is created and stored (in addition to descriptive metadata); * Note that the boundaries between data and metadata are not always clear in the case of structured data. In this case it is recommended to make sure that the information object itself includes sufficient data for object discovery, context and provenance detection purposes.
Questions	* Which metadata standards are implemented in the organisation and within the specific system that is being examined? * Are they compliant with legislative or business requirements if requested?

Destruction

Description	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.
Objective	If the information value and retention periods have been assigned appropriately, the destruction of unnecessary information objects helps the organisation to "clean" the information system and save resources in long-term data management actions, hardware resources, etc. Timely destruction of PII (personally identifiable information) is also key in achieving GDPR compliance.
Implementation	* The retention period should be defined in the metadata of an information object; * The expiry of that period should be automatically flagged, after which destruction follows; * Make sure to allow for a review of the destruction decision, especially if unsure about the appropriateness of retention periods; * Maintain clear destruction agreements for information objects for which automatic signalling is not possible with the administrators; * A declaration of destruction should be available; * Make sure that all appropriate versions of the information object are destroyed; * Make sure that the destruction of information objects does not destroy data needed by other information objects (with longer retention periods).
Questions	* Is destruction capability implemented within the system? * Has the destruction capability been appropriately tested? * Are flexible mechanisms available for the review of destruction decisions? * Is a destruction certificate being created?

Export

Description	Information objects can be exported in a standardised and controlled manner for transfer to an archive or any other information system.
Objective	A reasonable, trusted, tested, and easy to use export capability allows for information objects to be migrated into other systems whenever necessary - for example for storage in dedicated long-term repositories, or when changing the technical platform of the information system (potentially due to technology obsolescence). This, in turn, allows for continuing providing long-term access to the information across generations of soft- and hardware. In certain cases (e.g., large scientific datasets), it may be acceptable to export source data along with parameters, enabling reproduction of results (e.g. AI model or algorithms, cross-validation, fairness, bias analysis, etc.).
Implementation	* Make sure that all information objects are covered with export functionality; * Test the export functionality and make sure that the integrity of exported information is not endangered; * Create bulk export capability (i.e. to allow for the export of multiple (thousands) of information objects at once); * If the export to an external system is planned in the information object's lifecycle (e.g. transfer to an external digital archive after 5 years), develop appropriate automatic reminders; * Note that these export reminders are very similar to retention period reminders (though the outcome is export, not destruction); * Make sure to not delete exported information objects from within the information system before you have verified the quality of the export.
Questions	* Are all information objects covered with export functionality? * Are pseudonymised data transferred with key to their interpretation? * Has the export functionality been appropriately tested? * Is a bulk export capability available?

Right of access

Description	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.
Objective	For information to be long-term accessible it must be made available to the largest possible extent - users should have access to information as much as possible, subject to any restrictions on public access. This principle also includes information objects that contain parts that are not public but can be published after anonymisation, pseudonymisation, etc.
Implementation	* Provide a list of permissible grounds for restriction of publicity, backed up by legislation and policy, so that it is clear to everyone who has access rights; * Enable organisation-wide identity management and access control. Connect that to the applications in which information objects are

	stored, allowing an easy check if employees (and partners in the information chain) have access to certain information objects; * When 'blacking out' data it is not just the information object which might be anonymised but also metadata, for example the title of a record or an anonymised view / query; * Make sure that the information model defines which representations are temporary and which should be preserved.
Questions	* Have access rights been evaluated? * Is a list of user groups, their rights and permissions available and implemented within the system? * Is a list of possible restrictions available and implemented within the system? * Have means for data anonymisation, pseudonymisation, etc. been implemented within the system? * Are restrictions recorded within object metadata in a consistent and trusted manner?

Security

Description	The information system complies with applicable institutional, national and international information security standards.
Objective	The application of sufficient organisational, procedural and technical security methods is crucial for the long-term authenticity of information. A user must be confident that an information object is as previously recorded by authorised employees and that it has been protected from unauthorised changes throughout its lifecycle (e.g. ensuring proper use and preventing abuse in cloud environments).
Implementation	* In most cases specific security standards and guidelines are available on a national level. Make sure to be aware of the standards that apply to your institution.
Questions	* Is the organisation aware of the security standards and guidelines that apply? * Does the organisation possess a valid security certificate? * Has the organisation undergone a recent security audit (both internal and external)?