

Activities to facilitate the authentic interpretation of archived databases

Aleksandra Mrdavšič
Archives of Republic of Slovenia
Zvezdarska 1
1127 Ljubljana
+386 (0)1 24 14 256
aleksandra.mrdavsic@gov.si

Jože Škofljanec
Archives of Republic of Slovenia
Zvezdarska 1
1127 Ljubljana
+386 (0)1 24 14 248
joze.skofljanec@gov.si

ABSTRACT

Experience in archiving has shown that context information needs to be provided also by records in a digital form. The paper presents the activities and tools used by public archival services during the appraisal and advising records creators on how to create SIPs in a way that information needed for authentic interpretation of records is included.

1. INTRODUCTION

A considerable part of digital records that are of permanent and thus archival value are kept as databases. According to the Reference Model for an Open Archival Information System - OAIS the key purpose is to preserve not only the content of the records but their usability as well. In the paper the authors aim to present the method that has been tested in practice and by means of which the Archives of the Republic of Slovenia is attempting to ensure the integrity and authenticity of the ingested database records as well as include the information needed to ensure their authentic interpretation.

To monitor the method of facilitating authentic interpretation, a special tool was designed in the form of a check list. The check list is used by authorized archival employees when discussing the transfer of records with records creators on behalf of the archive. It can also be used as a template for the preparation of submission agreement.

2. PRE-INGEST ACTIVITIES

Records creators and archives to which their records will be transferred both face the challenge of how to ensure the records' accessibility, usability, integrity, authenticity and durability. The former are in charge of this task for the time when digital records are in their custody and the latter take over once such records are transferred to them. Based on the experience acquired by public archive services when dealing with such challenges, a number of procedures have been developed to provide integrity, authenticity and accessibility of records. However, methods for providing usability of records have not been developed to the same degree mostly due to the fact that the use of such records is not very

frequent. In compliance with the Slovenian archival act records need to be entered in the public register of records and made available to the public two weeks after their transfer to archives. Keeping in mind this legal provision we completed our check list by adding some requirements to ensure the records' usability.

Requirements observed in the preparation of our check list were as follows:

- Providing control over the preparation of SIP and its transfer to archives;
- Providing sufficient documentation of the process to preserve authenticity;
- Providing comparability of implemented procedures of creating individual SIP and its transfer to archives;
- Providing repetitiveness of individual stages of the procedure
- Preserving the needed level of records' usability and ensure it in the shortest possible time.

Experience acquired during the ingest of records to the Archives of the RS and best practices of some of the European¹ public archives have revealed that ingesting only records from databases is not enough, and that we also need to ingest additional documentation. The purpose of such documentation is to ensure authentic interpretation of data. An example from Slovenia demonstrated that just one missing table with certain codes, which in the original environment was not included in the database, can have a profound effect on our understanding of all records within the submitted database.

As far as practical application is concerned, in most cases it turns out that records creators do not have all the additional documentation that the archive would like to ingest, so compromise is often necessary. An example given is the documentation on Graphical User Interface – GUI, which was used by the records creator for entering and changing data in the database as well as for the printing out of the data needed. Typically, records born digitally as databases (such as public registers) have separate GUI designed for managing each such register. Usually, instructions for GUI are precise enough in describing activities performed during the capture and export of data from and into database. Among other things, they include screenshots which provide clear insight into the work with the

¹ In recent years a closer look at e-archiving at the national archives in Austria, Denmark, Estonia, Germany, Hungary, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom has been taken.

interface. Assisted by user manual for GUI, a user will be able to understand how the database was managed by records creator. When user manuals are not preserved or were not even prepared, they can be replaced by screenshots of some of the key stages of capture, accompanied by short explanations. These can be prepared by records creator if he has a working version of the application used during the capture of data into database. When working GUI is not available, it is recommended that a user makes a short description of how GUI was used, mainly for the capture and changing of data, and also what the most frequent forms of query and reports on screen and/or paper were. When employees who worked with GUI had different competencies or roles, it is sensible to include also the explanations about such competencies. If it is not possible to obtain any information about the working of GUI, the sensible thing to do is to make a note that no such information was available to the archive. The same applies to the description of external sources in automatic capture of certain data.

The check list prepared by the Archives of the Republic of Slovenia includes the following check points for facilitating authentic interpretation of the records' content:

- User documentation – these are user manuals and instructions for end users of applications for entering, changing and output of data;
- Use or its description and screenshots of important functionalities (when there are no suitable user documentation available) – it should include the description of the actual management and use of data in the database;
- Codelists – it should include all codelists and values for the each of them. When individual value was in a codelist only for a certain period of time, this time period should also be indicated;
- A list of data sources – it should include information on the sources and scope of data obtained from individual source. Also included should be the information on when individual source was used and how often data was captured from them; such information is particularly important for interconnection of databases, i.e. public registers;
- Forms, reports and queries (when no other suitable user documentation is available) – these are important for understanding how a user accessed data in the original environment. Also included here should be the description of queries in the original application, either as screenshots or examples of reports;
- Typical queries – apart from the actual description of a query (as mentioned above) it is also recommended for the records creator to add technical description of typical queries. Those are the queries that administrators, personnel entering the data, or users regularly used in the course of their work and for which standard output was formed (in the application) to display the results. Whenever possible, segments of code with SQL commands from the original application should be enclosed;
- Security scheme in the original environment – it describes restricted access to data in the original environment. Also stated are groups of users (for

example according to their roles) with their access rights or permissions to change data.

In the check list each of the check points consists of three elements: instruction, description and references. The latter are particularly important since a reference to one proper part of documentation can replace the entire description. When data sources of official registers or its metadata structure are legally regulated, records creator only has to refer to a single provision in a legal act. Records creators can also omit from their SIP any publicly available documentation, such as public regulations.

3. PROACTIVE ACTIVITIES

Drawing from our experience with ingestion of digital records, the Archives of the Republic of Slovenia also had an effect on field-related regulations regarding capture and storage of digital records by records creators. Keeping up with our proactive activities, several requirements for database records and public registers have been included in our regulation titled Uniform Technological Requirements, which is a regulation that needs to be observed by all Slovenian public sector institutions in the course of their digital records management. These requirements comply with check points in our check list. The list helps archivists to guide their records creators when preparing their records for transfer and, even more importantly, it guides records creators themselves to be more attentive to providing certain level of documentation already during production.

Uniform Technological Requirements demand that for each of the databases, especially those that are the basis for the keeping of public registers, records creators design the so-called public register dossier. Included in such dossiers is key documentation on managing individual register, i.e. managing data in individual database. The files preserve:

- Legal regulations (public and internal) that had an effect on management of registers;
- List of roles and competencies for individual roles;
- List of software and periods of their use;
- Description of data structure, which also includes data model, data sources (i.e. eventual other public registers, application forms ...) and codelists;
- Description of technical terms used (semantics) which enables long-term use of data from databases or public registers outside their original environment;
- Specification or description of changes for data structure, including data model and the software used;
- Description of any deletion or cancellation of the chosen data transfer, for example when technical environment or data model is changed.

4. CONCLUSIONS

The key tasks of any archive institution is to make the records it keeps available to users and do so in a way that enables users authentic interpretation of data. For this purpose, it is necessary to have tools that encourage standardization of procedures. Such tools need to provide enough room for flexibility in terms of implementation of procedures necessary due to diversity of business operation of individual creators. At the same time, the tools must enable discussion on all important topics and acquisition of available information to ensure enough data about

the context in which records were created and to facilitate authentic interpretation of such records.

When advising and directing business activities of our records creators and when this is allowed by state legislation, proactive action is sensible since it ensures preservation of documentation that will accompany records when they are transferred to archives and enable their authentic interpretation. Check list is independent of the technology used at the time of transfer.