piql

Gregor ROESLER-SCHMIDT, Beatrice MATUSCHKA, Alexander RYCH*

PIQL PRESERVATION SERVICES
A Holistic Approach to Digital Long-Term Preservation

Summary:

Piql Preservation Services ("Piql") is a turnkey solution designed for secure, migration-free long- term preservation of digital data. Piql sets an open standard for long- term preservation for the future. It consists of equipment and processes needed for writing and retrieving digital data.

Exponentially growing amounts of data demand for logistically effective and cost effective processes. Digital storage media (hard disks, magnetic tape) exhibit limited lifetime. Repetitive data migration to overcome rapid obsolescence of hardware and software bears accelerated risk of data loss, data corruption or even manipulation and adds significant repetitive costs for hardware and software investments.

Piql stores any kind of data in its digital as well as analog form securely for 500 years. The medium that provides this is a film reel. Using photosensitive film polyester base, a very stable material that is known for its immutability over hundreds of years, secure and cost-effective long- term preservation can be provided. The film reel itself is stored in a packaging capable of protecting the optical storage medium. These components have undergone extensive testing to ensure longevity of up to 500 years. In addition to its durability, film is a true WORM (write once- read many) medium. It therefore is resistant to editing or manipulation.

Being able to store any form of data onto the film makes Piql a superior solution for long-term preservation. Paper documents, images, video or audio sequences – all of those file formats and documents can be preserved in its native file structure. In order to restore the encoded digital data, only a film scanner, a digital camera or any appropriate optical reading device will be needed in the future. Every film reel includes an index section describing the data saved on the film. It also contains a content section carrying meta-data, enabling users in the future to rebuild software in order to read and decode the digital information.

Piql has the following key features:

- *Migration-Free. By storing digital data in archival file formats on the unique piqlFilm, repetitive migrations are no longer needed. Users avoid the risk of migration-related corruption and data loss and thus save long-term costs.*

- *Searchable. Operated within a standard IT- environment, the user's data is indexed and fully searchable- today and in the future.*

- *Flexible. Depending on the user's needs, data can either be stored as a human- readable text or image or in a digital format, making data readable and understandable for 500 years.*

- *Unalterable. Data is preserved on a secure true WORM medium, making it impossible to manipulate or delete valuable data.*

- *Permanent. Photosensitive film is known to be robust with proven long-term qualities. PiqlFilm is tested to last for 500 years. Although your time perspective may be shorter, you can be assured that your data will remain intact.*

Key words in English:

long-term preservation, digital data, migration-free, photosensitive film

---

* *In-Vision: Digital Imaging Optics GmbH, Guntramsdorf, Austria, contact: roesler@in-vision.at, matuschka@in-vision.at.*

STORITVE HRAMBE PIQL. HOLISTIČEN PRISTOP K DOLGOROČNI DIGITALNI HRAMBI

Povzetek:

Storitev hrambe Piql (kratko Piql) je rešitev na ključ, oblikovana za varno, dolgoročno hrambo digitalnih podatkov brez migracije. Piql zastavlja odprt standard za dolgoročno hrambo za prihodnost. Sestavljajo ga oprema in procesi, potrebni za zapisovanje in obnavljanje digitalnih podatkov.

Vse večje količine podatkov zahtevajo logistično učinkovite in stroškovno ugodne procese. Mediji za digitalno hrambo (trdi diski, magnetni trakovi) imajo omejeno življenjsko dobo. Večkratne migracije podatkov zaradi zastarelosti strojne in programske opreme predstavljajo vse večje tveganje za izgubo podatkov, za napake ali celo za manipulacijo ter zahtevajo dodatne stroške za investicije v novo strojno in programsko opremo.

Piql hrani varno katerekoli podatke, tako v njihovi digitalni kot tudi analogni obliki, 500 let. Medij, ki to omogoča, je filmska rola. Uporaba fotoobčutljive poliesterske osnove, zelo stabilnega materiala, ki je znan po svoji imunosti že več stoletij, omogoča varno in stroškovno sprejemljivo dolgoročno hrambo. Zvitek se hrani v ovoju, ki zagotavlja zaščito optičnih medijev za hrambo. Omenjene komponente so bile podvržene ekstenzivnim testiranjem s ciljem zagotavljanja življenjske dobe podatkov vse do 500 let. Ne le da je film trpežen, je tudi resnični WORM-medij (write once-read many – zapiši enkrat, beri večkrat). Tako je odporen proti popravljanju ali manipulacijam.

Možnost shranjevanja katerekoli oblike podatkov na film je odlika, zaradi katere je Piql vrhunska rešitev za dolgoročno hrambo. Papirni dokumenti, slike, video ali avdio sekvence – vsi ti formati in dokumenti se lahko hranijo v svoji prvotni strukturi. Za obnovitev kodiranih digitalnih podatkov v prihodnosti bo potreben le filmski skener, digitalni fotoaparat ali katerakoli primerna optična bralna naprava. Vsak filmski zvitek vsebuje tudi indeks z opisom podatkov, ki so na njem shranjeni. Vsebuje tudi vsebinsko sekcijo z metapodatki, ki bodočim uporabnikom omogoča ponovno vzpostavitev programske opreme za branje in dekodiranje digitalnih informacij.

Ključne značilnosti Piqla:

- ***Brez migriranja podatkov***. *S hrambo digitalnih podatkov v arhivskih oblikah datotek na unikatnem filmu piqlFilm večkratne migracije niso več potrebne. Uporabniki se izognejo napakam, pokvarjenim datotekam ali izgubi podatkov in tako prihranijo pri dolgoročnih stroških.*

- ***Možnost iskanja***. *Z uporabo v standardnem IT-okolju so uporabnikovi podatki indeksirani in najdljivi – tako danes kot v prihodnosti.*

- ***Prilagodljivost***. *Odvisno od uporabnikovih potreb so lahko podatki shranjeni kot človeško berljiv tekst oz. slika ali v digitalni obliki, kar podatke naredi berljive in razumljive tudi čez 500 let.*

- ***Nespremenljivost***. *Podatki so hranjeni na varnem WORM-mediju, ki ne omogoča manipulacije ali izbrisa pomembnih podatkov.*

- ***Trajnost***. *Fotoobčutljiv film je znan po svoji robustnosti ter dokazanih dolgoročnih kvalitetah. PiqlFilm zagotavlja glede na rezultate testiranj hrambo 500 let. Čeprav je morda vaša časovna perspektiva krajša, je nedotakljivost vaših podatkov zagotovljena.*

Ključne besede:
dolgoročna hramba, digitalni podatki, brez migracij, fotoobčutljiv film

1 OBJECTIVES OF THE PIQL PRESERVATION SERVICES AND INNOVATIONS

Obstacles of archiving digital data are the regular migration of data as well as the technical obsolescence of storage medium and reading devices. The consequences of migrating data for example are increasing handling-costs as well as the possible loss of the data itself.

To prevent those sorts of problems, main objectives of Piql Preservation Services were to:

- Develop a migration-free preservation medium
- Establish reliable, cost-effective manufacturing methods for the medium and packaging
- Ensure longevity of the medium
- Prove the concept by integrating the Piql system with preservation systems at existing archives according to the accepted OAIS (Open Archival Information System) framework
- Develop a process to secure fast ingest and retrieval of digital objects
- Ensure future access

To obtain a secure solution for the archiving of digital data, different measures had been taken to realize the objectives:

1. To obtain a migration-free archiving system, the stored data is written on the piqlfilm reel only once. The need to migrate the archived information to another system is not necessary anymore. This prevents regular handling-costs and the eventual loss of the data.
2. Piql Preservation Services use a nano-sized photosensitive material, which is known for its stability. To ensure longevity the film reel and its packaging have undergone a series of tests that verified an expected lifetime of 500 years.
3. Piql Preservation Services have been evaluated to satisfy the basic requirements in the OAIS compliant digital repository, which makes preserved information available for a designed group of people.
4. The fast ingest and retrieval of the data is made possible by a logistical process in an automated data-warehouse with the piqlWriter- and Reader.
5. In order to read and decode the digital data stored on the film, only a film scanner, a digital camera or any appropriate optical reading device will be needed.

The solutions for a secure and reliable long-term archiving system were specifically designed with the requirements of digital preservation in mind: the technology of the Piql Preservation Services is flexible, open and has the potential of offering a fully integrated and future-proof solution. The testing of all components in normal storage conditions (temperature and humidity requirements) ensures longevity of up to 500 years of the storage medium.
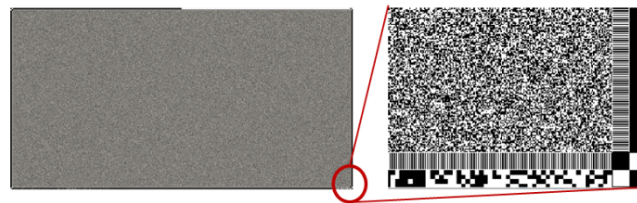
2   THE WORKFLOW OF PIQL PRESERVATION SERVICES

The workflow of the Piql Preservation Services is composed of several processes. The workflow consists in data boxing and integrity check, data recording, data scanning and storage and the data unboxing.

2.1   Data Boxing and Integrity Check

The data boxing is the process of converting digital data so that it can be recorded onto the film. In the OAIS context, only quality controlled SIPs (Submission Information Packages) are accepted as ingests. The SIPs include text files, images, video- and audio files as well as required metadata. At this point, metadata is not only added in order to easily identify and track the film record, but also to ensure data preservation.

Every data will undergo a quality-check of its integrity to prevent unintentional changes. This ensures that the data is recorded and retrieved exactly as intended. After the file is received, checked for virus and normalized to a preservation file format, a check sum is generated for the file, sent back to the data owner for verification and saved in the Piql database for future authentication. In order to fit into the frame on the film reel, the files are decoded and divided into small elements of information (binary form).

Picture 1: Digital data frame. Magnified section to the right.

2.2   Data Recording

The data recording describes the process of exposing the previously boxed data onto high-resolution micrographic film. With the piqlWriter the encoded files are written onto the piqlFilm as square dots together with the file format and the metadata scheme descriptions. The piqlFilm was especially designed for longevity and high-density digital writing. Information like the source code and technical instructions for decoding the content are put on the film as well. This enables future generations to interpret the archived content in a human-readable form.

Picture 2: piqlBox and piqlFilm

2.3   Data Scanning and Storage

After the first two steps of boxing and recording the received data on piqlFilm, it has to be reviewed to guarantee integrity and accessibility in the future. With the piqlReader, an industrial grade data scanner, all frames in the reel are read and decoded. The original files are then restored. To assure the quality and authenticity of the files, each file check sum is verified against the one in the database. After that, the film is packed into the piqlBox, which is then send to final storage.

The Piql system can be integrated with any warehouse management system (WMS). The WMS keeps track of each reel's position in the warehouse. The piqlBox is labeled in order to identify each reel via a standard barcode scanner. As soon as the reel is safely stored, the original uploaded file(s) are deleted from the system, and the data owner is notified.

Picture 3: piqlWriter and piqlReader

2.4   Data Unboxing and downloading

The data unboxing describes the process that fully restores the scanned data. It allows data owners to search for and access files in only a few minutes. This step involves decoding and re-formatting the data to make it fully readable. The data owners can request a specific file through their own integrated document management system or directly through the Piql client interface. The request will be communicated to the WMS where the reel is identified and retrieved. In order to restore the needed

file the reel is placed in the piqlReader where the film is forwarded to the exact position. Then the file is restored and quality-checked as explained above during the data verification stage. Finally, the file is made available for download and the reel is returned to the warehouse.

The download of the file has to be within a limited time before it is made offline again. This measure has to be taken for security reasons, as files should not be kept online longer than needed. As soon as the data owner downloads the file, the recovered file is deleted.

3 FUTURE ACCESSIBILITY AND HUMAN UNDERSTANDING

One of the main objects of the Piql Preservation Services is that all information needed to decode the data is available to those that might need it sometime in the future. Therefore the source code for the decoder is stored at the beginning of the film reel in both human-readable and digital form. For additional security and better understanding in the future, it is also possible to add file format specifications. To restore the information, future users will only need a scanner and a computer as well as follow distinct tasks:

- The square dots have to be read by a film scanner, a digital camera or any appropriate optical reading device
- The source code of the software to encode/interpret the digital data, - and to convert it back to the original file format, - is included on the film in human-readable form and needs to be compiled into an executable software (program) on a computer in the future.
- This program will decode the digital data on the film.

The metadata section on the reel contains the following information:

- Zero reference frame: this frame is used by the reader to automatically identify the start of the reel.
- Control frame: this frame is printed in both digital and human- readable form. It contains all the parameters the decoding software needs to decode the reel.
- Table of content of files in both digital and human- readable form
- Description of the technology used
a. to write the data on the reel
b. to build a system to read back the frames on the reel in human- readable form

Besides the metadata section the reel also contains a content section. It includes a filesystem with digital, human- readable or a combination of both versions of files. The configuration of the reel is defined in a xml and is reflected in the table of content.

4 PIQL PRESERVATION SERVICES IN THE MARKET

In the recent years the archival storage market has emerged as a fast growing segment. Many decision-makers in business, institutions and governmental organizations are confronted with the obstacles within the archiving of data. For smaller and mid-sized companies in particular, concepts of long-term archiving based on systems using magnetic hard disk drives and tapes are much too costly due to their complexity and the amount of staff required to manage the data. The costs associated

with backing up data have risen due to copying and migrating the data to a new storage medium. In many institutions the amount of digital primary data, including elements such as sound, graphics, video, animation etc, increases. It is for these data that Piql as a true migration-free digital preservation solution plays a role. Institutions, organizations or enterprises, that see security and accessibility as main elements for future success, represent market sectors for the Piql Preservation Services.

The following examples indicate "data owners" that have a need for long-term archiving:

- National archives
  National archives typically take over the long-term preservation and archival responsibilities for governmental information and data.

- National Libraries
  The general responsibility of national libraries is to collect, preserve and publish all media content (e.g. books, films, TV programs, radio programs etc). Besides the national libraries also public, academic, institutional and digital libraries are in charge of preserving data.

- Government
  Parliament, government ministries and agencies have a great amount of data that need to be preserved. Economic data, security and pension records, land registers, courts and judicial data and a lot more, have to be preserved not only on a local level but also on a state (region) and federal (national) government level.

- Cultural institutes for the preservation of national and international heritage
  In every country there are a huge number of cultural institutes that are responsible for preserving national and international heritage and therefore have their own archives.

- Insurance and financial services
  Reference to historical financial and economic data, actuarial data for health, life, casualty insurance.

- Medical technologies
  In many countries patient records and medical images are required. The regulations call for a preservation of the data that is kept through the lifetime of the patient and even longer.

- Scientific research programs
  Meteorology, astronomy, particle physics and many more compartments in the scientific research need to have archiving systems for the big amount of collected data.

- Media and entertainment
  Motion picture films, stock video, photography, animation etc. are information that needs to be archived for the future.

The need for long-term archiving exists in other fields like the public and private industry, military bodies and archives, universities and schools, hospitals and much more. All those archives, institutions and organizations accumulate permanent data at rate approaching thousands of Terabytes per year. Piql is able to preserve this data safely and migration-free.

5   KEY INNOVATIONS

Piql Preservation Services manage to combine some of the most important components to ensure a secure and cost-effective long-term archiving system.

- There is no limit of what can be stored onto the piqlFilm. Digital data such as HD audio-visual data, documents or databases can be archived with Piql.

- It is a service that uses a whole new approach to long-term preservation by removing the need for constant data migration, reducing hardware and software maintenance.

- The innovation is to use a permanent media, to adopt standards for formats and interfaces and be independent of proprietary software and specific hardware.

Piql Preservation Services convert digital files into a physical copy that can be stored as the ultimate digital insurance. Data owners can feel confident about the long-term accessibility of their valuable data- today and in 500 years.