

# Central data archive (CDA)

University Library in Bratislava and TEMPEST have built a nationwide digital archive. It is used for preservation and storage of the cultural heritage. The archive will store digitised historical papers and documents, digital copies of collections of art and artworks, as well as historical audiovisual recordings.

25<sup>PB</sup> / 75<sup>PB</sup>

25 PB is the total available capacity of the archive, 75 PB is the total gross storage capacity (about 75,000 computers with disk capacity of 1,000 GB)



The system will check quality of data and metadata, possible presence of viruses or malicious code in real time

381  
million

381 million is the expected number of stored objects

150<sup>TB</sup>

50 TB is the total weekly gross data throughput

3x

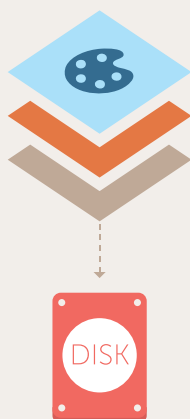
data redundancy

3x data redundancy in order to preserve the data in the event of a disaster or outage

1<sup>s</sup>

It takes 1 second to process, check and store a standard document

## BENEFITS



- consistent, standardized methodology (OAIS) and technological environment for the processing and storing of digital records
- triple redundancy of data in geographically separate locations
- validation, quality control and multiple data protection at all levels of operations
- flexibility in data storage formats adapted to the requirements of individual depositors
- flexibility of data entry (online or through logistics solution)
- The solution is prepared for ISO 16363:2012 certification and also the operating Organization is preparing to get ISO 27000 certification

The main benefit of the project is the long-term preservation and protection of cultural heritage, in some cases of irreplaceable scope and significance, for future generations. The institutionalization of long-term digital archive meeting the international norms and standards in the field is also a benefit to no small extent. The definition of a legislative, procedural and methodological framework for the long-term preservation of digital objects, their availability and processing, is another benefit. A community of experts in the field of digitization and preservation of digital objects whose know-how will contribute significantly to the development of the use of the area for the general public, is being formed during the project implementation. By implementing this project, Slovakia reaches the top in the long-term preservation of digital data within the EU.

---

## Project

---

**Current users of the service are important institutions in the field of digitisation of cultural heritage:**

- The Monuments Board of the Slovak Republic
- Museum of the Slovak National Uprising
- Slovak National Gallery
- Slovak National Library
- State Scientific Library in Prešov
- Slovak Film Institute

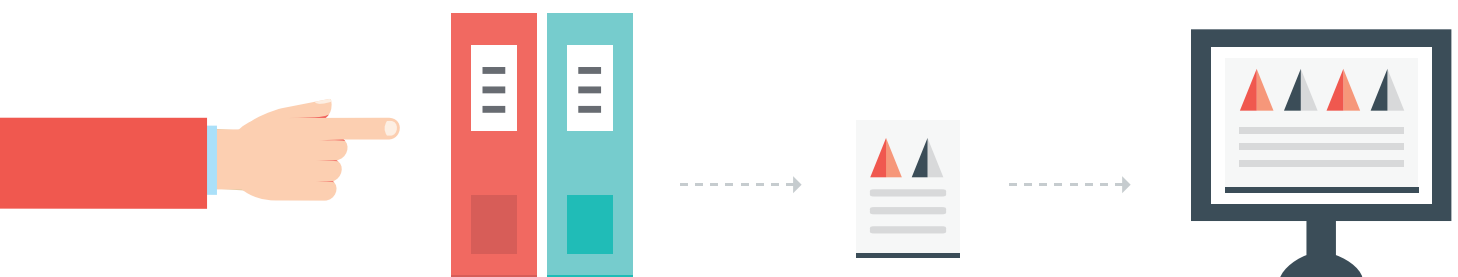
The target group is the community of institutions that carry out the digitization of cultural heritage in Slovakia. The archive is ready to open for other institutions or for commercial or non-commercial entities outside the field of culture.

---

## University Library in Bratislava (ULB)

---

University Library in Bratislava (ULB) is the oldest and largest research library in Slovakia. It was established in 1919 as a library of Comenius University. Until 1954, in addition to its academic mission, it served as a national library. Since 1954 it has been an independent scientific library serving the general public. At the end of 2013, ULB archive contained a total of 2,650,000 library units.



---

## Status

---

Until 2012, there was no reliable, adequate and sustainable infrastructure for long-term archiving. The Operational Programme Information Society (OPIS) and its main priority axis 2 cover approximately 3.5 mil. cultural heritage objects, for which the previous solution was not ready. The Ministry of Culture of the SR considered the existing conditions as unsatisfactory and insufficient – in terms of both technical facility parameters and low level of interoperability and standardisation, security and integrated methodical management of relevant work procedures.

Fragmented data storages that are part of informatization projects of individual memory and fund institutions (MFIs) from previous years (KIS3G, CEDVU, CEMUZ, etc) were outdated, underpowered in performance and insufficient in capacity. Thus the starting situation for the CDA was the absence of a reliable data storage for launching and emerging digitization projects, and the lack of a functional solution for the comprehensive and reliable backup of program systems and data.

---

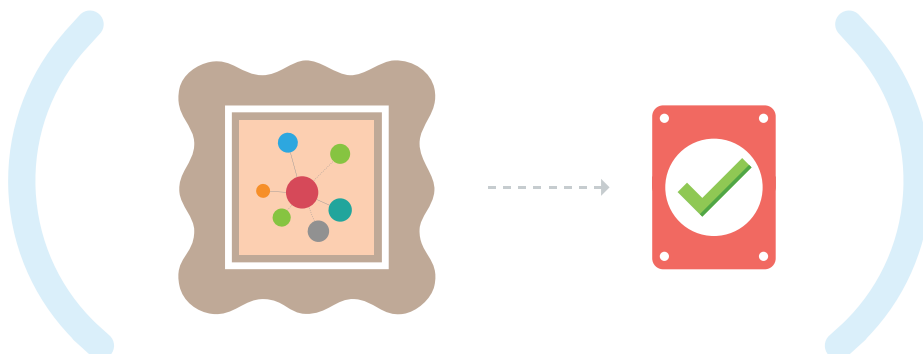
## Project objectives

---



- **developing data storages at determined locations**
- **recording, collecting, long-term archiving and protection of digital content**
- **implementation of web harvesting and web archiving**
- **compliance with the defined parameters, indicators, criteria and standards**

The project is aimed at developing a comprehensive integrated system of the long-term preservation and protection of digital content, its acquisition, processing, protection and utilization. A central data archive was built within the project, which under the prescribed standards ensures the storing of copies of digital objects in at least two geographically separate locations at least 50 km distant. The locations are equipped with the appropriate HW and SW equipment, interconnected through data telecommunication networks. The data archive also includes a passive media storage, which is located in the ULB object. These aims form the basis for systematic support regarding the digitization of cultural, scientific and intellectual heritage, which will address the relevant memory and fund institutions (MFIs) within digitization projects and campaigns.



# Basic technical parameters of the project and scheme

## DOCUMENT LIFECYCLE, STANDARDS AND PROTECTION

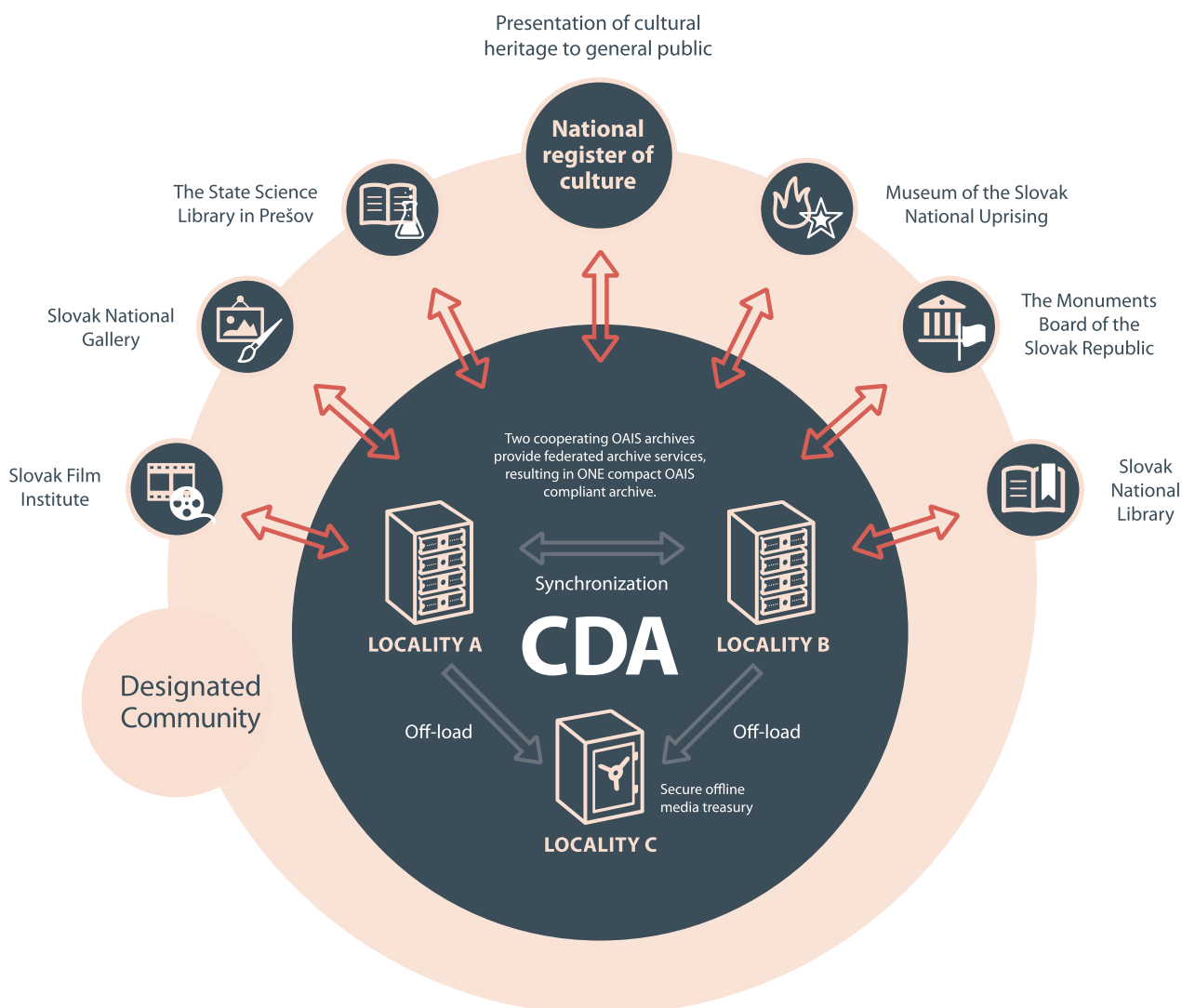
The solution provides long-term archiving services for digitized or digital content in accordance with the international OAIS standard (Open Archival Information System). The idea of a long-term archive is not only to store data, but mainly to maintain their legibility and clarity. The CDA therefore focuses on tracking the life cycle of stored formats, their conversion or emulation, as well as bit protection. The CDA solution also includes the support of depositors within the selection and identification of suitable formats for archiving, securing legislative support for data storage, and of course, the variability in the data selection. Part of the CDA is also the protection of intellectual property in accordance with applicable legislation and safety of stored objects.

## REDUNDANCY

Each data is stored in three separate media, which are located in three geographically separate locations. Two of these sites are living systems that can meet the community demands placed on services. The third site is a secure repository designed for the long-term protection of the media.

## CAPACITY

The gross storage capacity is 75 PB \* Usable storage capacity: 25 PB \* Expected objects: 381,346,900 \* System scalability: unlimited



## **COMPATIBILITY WITH OAIS (OPEN ARCHIVAL INFORMATION SYSTEM)**

The CDA is built as a solution that fully meets the OAIS requirements. It consists of two separate archives, each of which complies with the OAIS. These archives act as cooperating and associated archives offering full compatibility and a closed solution. The data are regularly synchronized, thus ensuring stable redundancy and availability. The solution is ready for certification according to ISO 16363: 2012 and the operating organization is ready to be certified ISO 27000. All internal processes are designed to respect and comply with the recommendations and requirements of the framework ITIL (Information Technology Infrastructure Library).

---

## **HIGH FLEXIBILITY**

Every member of the community signed a formal agreement with the CDA on data input. This agreement contains data formats, validations, conversions, and details about the data processing tailored to the specific needs of each member. These details are then converted to a specific archiving profile.

---

## **LOGISTICS**

Data can be entered via the Internet (online) or through specialized logistics system, provided as a part of CDA solution. The consignment obtains guarantees of suitable environmental conditions for transported media and the proper handling and protection of media. The logistics system further provides a convenient media packaging and the correct labeling of packaging. The logistics system is focused on the handling of magnetic tape, but the solution is open enough to introduce the support of new data carriers as needed in the future.

---

## **SECURITY**

All online communication of external entities is protected by encryption and all sensitive transactions are protected by two-factor authentication. The CDA operates its own certification body offering PKI services (Public Key Infrastructure) to all members of the specified community. The CDA solution further implements stable mechanisms for digital content signing, thus ensuring the high consistency and authenticity of data.

---

## **DATA PRIVACY**

In addition to a multiplicity of data redundancy at the physical level, the CDA applies modern and technologically advanced principles of logical data protection, thanks to which it provides the long-term preservation of data in terms of availability, quality and reliability. Adjustable processing procedures based on a plug-in architecture offer scalability and future improvement of usability. The same principles apply to periodic LTP process (long-term potentiation), format management, preservation planning and risk assessment. Thanks to modern tools for managing storage plans, the solutions offer their own system to manage data formats. This tool collects information about formats processed in the archives from various sources and external databases, and allows the automatic evaluation and processing of formats that are particularly new and not well known. The system also allows management of the entire life cycle from the planning stage to the end of life of the given format and its conversion.

---

## **ADVANCED TECHNOLOGY**

The CDA solution is built on advanced technologies with the use of all modern principles and architectures used to process voluminous data. The main aim of this approach is to ensure the further improvement of the solution that will be able to meet future needs for growth, expansion and modernization. The main objectives when building architecture were to guarantee modularity, scalability, robustness and expandability.

---

## **OPEN SYSTEM**

The CDA solution is built as an open system and is a combination of open source and tools at the enterprise level.

Partners

---



univerzitná knižnica  
v bratislave



VTGroup

 **Tempest**

I T m a k e s s e n s e

TEMPEST a.s.  
GBC IV  
Galvaniho 17/B  
821 04 Bratislava 2  
Slovak republic

Tel.: +421 2 502 67 111  
Fax: +421 2 502 67 100

[info@tempest.sk](mailto:info@tempest.sk)  
[www.tempest.technology](http://www.tempest.technology)