

Przechowywanie obiektów elektronicznych w długim czasie

– nowe wyzwanie dla archiwów

26 V 2004

Quod non est in actis non est in mundo

dr Hubert Wajs
hubert.w@wp.pl



Aksjomaty

- „*Slajdware*” – materiały
- „Długi czas” > 100 lat
- Naruszenia:
 - Ustawa o języku polskim
 - Zakaz reklamy
- Problemy, idee, a nie droga wyjścia
 - Brak wewnętrznej spójności

Układ prezentacji

- Pojęcia:
 - *Preservation; Records*
- Archiwa
 - Co tam jest wewnątrz?
- Technologie AD 2004
 - XML
 - Doświadczenia innych (GB)
 - „Trzecia Droga”

Preservation - przechowywanie+zabezpieczenie

- *Preservation encompasses the activities that prolong the usable life of materials. Preservation activities are designated to minimize the physical and chemical deterioration of materials and to prevent the loss of informational content. These activities include providing a stable environment for materials of all media types.....*
 - [Care and Handling of CDs nad DVDs, CLIR 2003]

Preservation

- Dla bibliotekarzy oznacza ono [proste] zabezpieczenie; np. zmikrofilmowanie i udostępnianie w czytelni mikrofilmu już jest „preservation”.
- Dla archiwistów jest to pojęcie występujące z przyimotnikiem „long term” i obejmujące znacznie bardziej złożone działania (mikrofilmowanie wprawdzie też, ale przede wszystkim stworzenie właściwych warunków – temperatura, światło, wilgotność, bezpieczeństwo – dla przechowywania materiałów archiwalnych w długim czasie).

Preservation -

- Zapewnienie, że udostępnienie w czasie będzie dotyczyło, materialnie, tego samego obiektu, niezmienionego przez czas lub użycie. Przyszli badacze muszą mieć dostęp do tych samych materiałów w niezmienionych warunkach. Tylko tak można zagwarantować, że jakichś dwóch czytelników będzie się mogło ze sobą porozumieć (polemizować), gdyż widzieli ten sam obiekt.

ISO 15489

- 1. Scope (TR 1)
- 2. Normative reference
- 3. Terms and Definitions
- 4. Benefits of records management
- 5. Regulatory environment
- 6. Policy and Responsibilities (TR 2)
- 7. Records management requirements
- 8. Design and Implementation of a records systems (TR 3)
- 9. Records Management process and control (TR 4)
- 10. Monitoring and Auditing (TR 5)
- 11. Training (TR 6)



COMMITTEE
DRAFT

COMMITTEE DRAFT ISO/CD 15489	
Date 08/12/99	Reference number ISO/TC 46 /SC 11 N 1748Std
	Supersedes document

WARNING: This document is not an International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

ISO/TC 46/SC 11

Title
Archives / Records Management
Secretariat
SAA

BSI DISC Private Circulation
Document Number: 99/625284
Date: 1999-12-09
Committee Reference: IDT/2/17

COMMENT BY 2000-02-15

Circulated to P- and O-members, and to technical committees and organizations in liaison for:

- discussion at [venue/date of meeting]
 comments by 15/03/20 (international)
 UB BY 2000-02-15

 approval for registration as a DIS in accordance with
 2.5.6 of part 1 of the ISO/IEC Directives, by
 15/03/20 (as above)
 [date]

(P-members vote only: ballot form attached)
P-members of the technical committee or subcommittee concerned have an obligation to vote.

Title (English)

Records Management

Title (French)

Reference language version: English French Russian

Introductory note

745

1. I have great pleasure in presenting you with a copy of the draft of the International Standards Organisation's proposed Records Management Standard, catalogued as ISO15489. I do so in the hope that you will find it of interest but also to solicit your comments and recommendations for any improvements you may wish to see made.

2. The document is the end result of 18 months' deliberation by a sub-committee of archivists and records managers, described in ISO nomenclature as ISO/TC46/SC11, from a dozen or so countries within Europe, North America and Australasia. It follows publication in 1995 of the Australian Standard for Records Management AS4390 on which it was originally based.

3. The draft for the international standard, however, takes a widely differing approach. As you will see, the proposed standard comprises only the high-level principles of records management and contains little or no guidance on compliance procedures. This is quite deliberate.

4. While the principles of records management practice are observed internationally, processes for achieving them vary widely around the globe. These application variances, which neither cause nor need be standardised, are being identified by the SC11/AM3 group.

ISO 15489 - *Records Management*

- Historia:
 - 1997 (BSI o AS 4390 → australijskie rozwinięcie normy ISO 9000)
 - 1999 TR (wersja 1)
 - 3 X 2001 – publikacja
- Układ:
 - ISO 15489:1 – 2002 – norma: Records Management (for non specialists, but useful for records professionals)
 - ISO 15489:2 – Technical Report (for records practitioners) + bibliography

ISO 15489

- 2. Normative references:
 - ISO 5127:1983 – Documentation and Information
 - ISO 5963:1985 – Documentation – Methodes for examining documents, determining their subjects and selecting indexing terms
 - ISO 11799 Information and Documentation – Document storage requirements for archives and library materials

ISO 15489

- 3. Terms:
 - **Electronic record:** *Record on electronic storage media, produced, communicated, maintained and/or accessed by means of electronic equipment.*
 - **Records:** *Documents created, received and maintained as evidence and information by an agency, organization or person in pursuance of legal obligations or in the transaction of business.*
 - **Records management:** *Field of management responsibilities for their efficient and systematic control of the creation, reception, maintenance, use and disposition of records, including processes for capturing and maintaining evidence and information of business activities and transactions in the form of records.*

ISO 15489

- Recommendatory standard („Should“)
- Management of records in their originating organisation (public or private, for internal and external clients.
 - [it does not include the management of archival records within archival institutions.]
- International value of ISO (v. AS 4390)
- „Records management“ not „recordkeeping“
- Records as information not only as evidence

Definicja z ISO 15489

- **Records:** *Documents created, received and maintained as evidence and information by an agency, organization or person in pursuance of legal obligations or in the transaction of business.*
- **Electronic record:** *Record on electronic storage media, produced, communicated, maintained and/or accessed by means of electronic equipment.*

Historia problemu

- W 1972 r. pierwszy projekt w Stanach Zjednoczonych, w 1974 r. w Kanadzie.
- W 1978 r. Francuskie *Archives Nationales* poleciło przeprowadzenie studium tego zagadnienia, w 1983 r. powołano przy Centrum Dokumentacji Współczesnej (Centre des Archives Contemporains) w Fontainebleau grupę Constance.
- W 1986 r. Amerykański Komitet do spraw Przechowywania Dokumentów Historycznych (Committee on Preservation of Historical Records) opublikował *Preservation of Historical Records*. Była to pierwsza próba odniesienia się do problemów wynikających z nietrwałego charakteru informacji utrwalonej w formie elektronicznej.
- W 1990 r. został opublikowany pierwszy raport *Management of Electronic Records: Issues and Guidelines*, przygotowany przez Komitet Doradczy Organizacji Narodów Zjednoczonych do spraw Koordynacji Systemów Informacyjnych (United Nations, The Advisory Committee for the Coordination of Information Systems - ACCIS). Było to pierwsze opracowanie ponadnarodowe, ujmujące zagadnienia społeczeństwa informacyjnego z perspektywy archiwalnej.

Po co przechowywać?, czyli Zadania Archiwów

- Przechowanie i udostępnianie dokumentów administracji publicznej (w tym elektronicznych) w długim czasie.
 - Samo gromadzenie bez udostępniania nie ma sensu !
 - Udostępnienie dokumentów elektronicznych w długim czasie → zmiana „filozofii” archiwalnej

Zmiana „filozofii”

- **ANALOGOWE**
 - Przechowywanie nośnika
 - Archiwa na końcu „cyklu produkcji” dokumentu
- **CYFROWE**
 - Przechowywanie informacji
 - Wymagania archiwalne muszą być respektowane od początku tworzenia systemów: obiegu dokumentów, baz danych, stron www...

25°C (77°F) & 50% RH

Co mają gromadzić archiwa?

DOKUMENT:

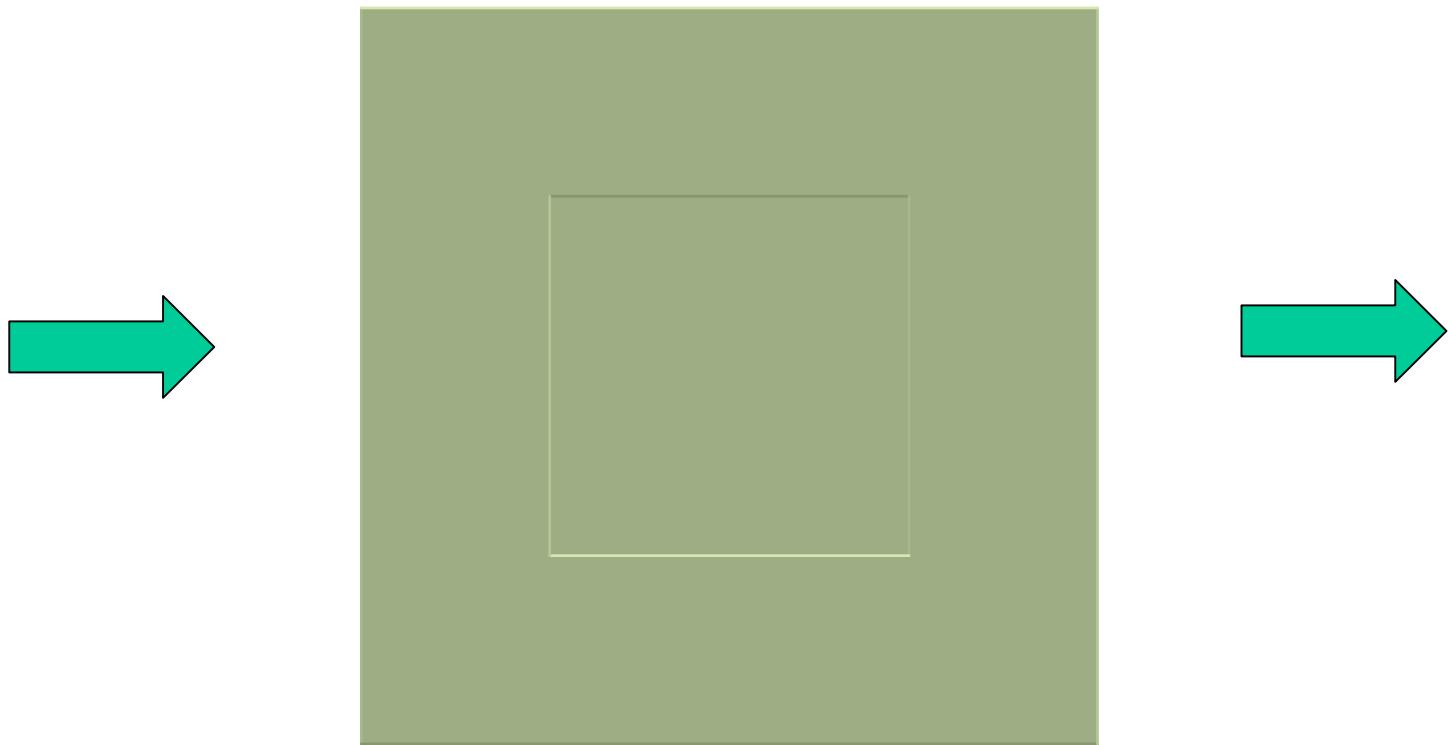
- Autentyczny
- Wiarygodny
- Integralny
 - można go będzie odczytać i zinterpretować,
 - zachowuje logiczną strukturę i fizyczną integralność,
 - zachowuje treść, kontekst i wygląd taki, jak w chwili utworzenia / otrzymania.

[Według pracy Ch. M. Dollar'a, *Authentic Electronic Records: Strategies for Long-Term Access*, Chicago 2000]

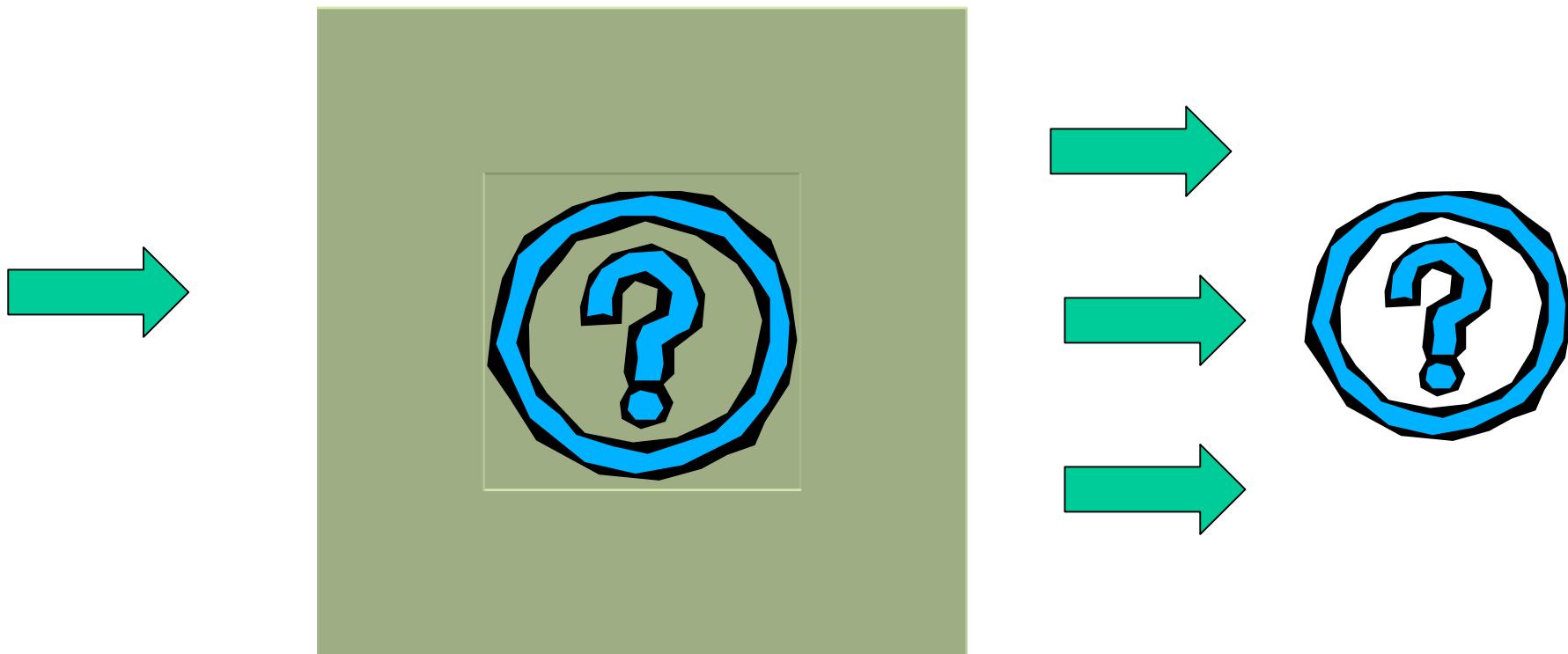
Jak to osiągnąć w świecie cyfrowym?

- ???? „podpis elektroniczny”
 - Środek uwierzytelniania dogodny dla handlu elektronicznego, a nie dla długotrwałego przechowywania dokumentów elektronicznych.
 - Samo określenie "podpis cyfrowy" to kolejna pułapka semantyczna.
- A może inaczej sformułować problem?
 - „*Paper minds, electronic records*”
 - Co trzeba przechować – Rzeczowy Wykaz Akt:
 - Rozporządzenie Ministra Kultury z dnia 16 września 2002 w sprawie postępowania z dokumentacją i zasad jej klasyfikacji i kwalifikowania oraz zasad i trybu przekazywania materiałów archiwalnych do archiwów państwowych. (Dz. U. Nr 167/2002 poz. 1375).

Maszyna trywialna według Heinza v. Förstera



Maszyna nie-trywialna według Heinza v. Förstera



Klasyfikacja

- Wykaz akt (strukturalno-rzeczowy):
 - **0 – Zarządzanie**
 - 00 – Organy kolegialne
 - 01 – Organizacja
 - 02 – Akty normatywne, pomoc prawna
 - 04 – Informatyka
 - 05 – Skargi i Wnioski
 - **1 - Kadry**
 - **2 - Środki rzeczowe**
 - **3 – Ekonomia**

Kategoria Archiwalna Dokumentacji

- A – materiały archiwalne
- B – dokumentacja niearchiwalna
- B0 – z cyfrą arabską – czasowe znaczenie praktyczne
- Bc – krótkotrwałe znaczenie praktyczne
- BE0 – z cyfrą arabską – dokumentacja, która po określonym czasie podlega ekspertyzie

Emulacja lub migracja (konwersja)

- Raport *Digital Culture* z 1999 r. wyróżnia następujące strategie przechowywania:
 - przechowywanie technologii,
 - emulację, czyli przechowywanie programu oraz
 - migrację (wg nomenklatury Ch. M. Dollar'a - konwersję) odwołującą się do powszechnie przyjętych standardów.
 - wg Ch. M. Dollar'a: w długim czasie można to zapewnić jedynie albo przez:
 - odnowienie nośnika, czyli reformatowanie
 - konwersję

CRC / HASH

- komunikat cyklicznej kontroli nadmiarowej (*Cyclical Redundancy Checksum - CRC*).
- jednokierunkowa suma kontrolna mieszana lub jednokierunkowa funkcja skrótu (*hash*).
- Zasada: "*Document what you do, do what you document*".

Jak z tego wyjść ? (Technologia AD 2004)

- Metadane
 - SGML - *Standard Generalized Markup Language*
 - XML - *eXtensible Markup Language*

XML

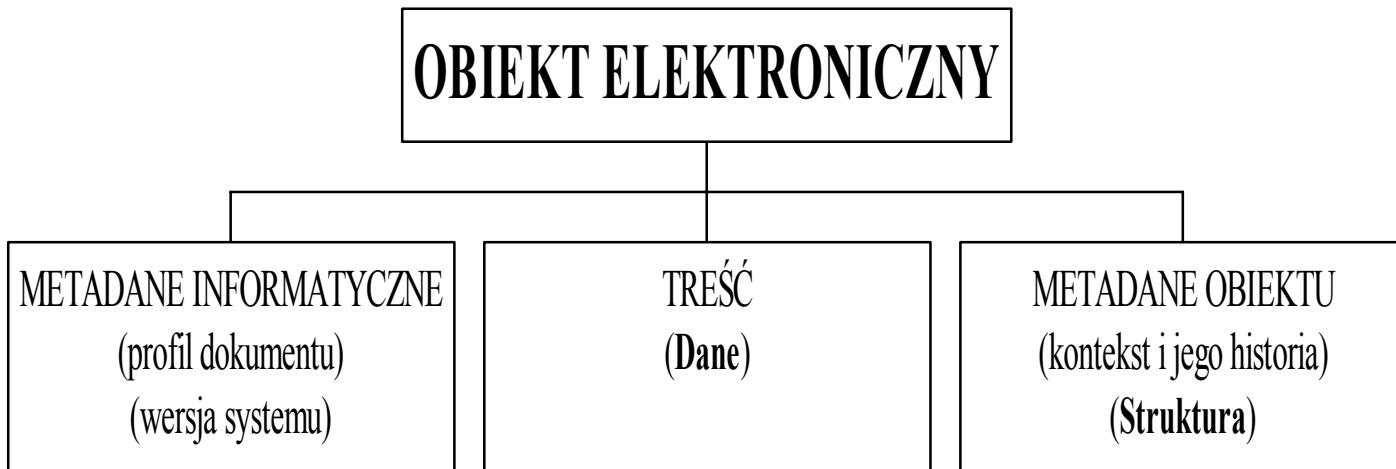
eXtensible Markup Language

- 1998 r. - *World Wide Web Consortium* - W3C
- XSL (*eXtensible Stylesheet Language*)
- Jest to metajęzyk mający wspólną gramatykę i jednolita metodologia opartą na SGML. Za pomocą tego metajęzyka można stworzyć uniwersalny format dla dokumentów mających strukturę i zawierających dane - można opisać nie tylko dane, ale i ich strukturę (metadane). W odróżnieniu od języka HTML, który pozwala tylko na formatowanie tekstu, XML pozwala opisać to, co faktycznie zawiera informacja tekstowa.

Elementy dokumentu (obiektu)

1. Treść (zawartość dokumentu, informacja);
2. Logiczna struktura dokumentu (układ); logiczna struktura może być odmienna od fizycznej struktury dokumentu.
3. Metadane czyli kontekst,
 - a) powiązania z innymi dokumentami,
 - b) dane techniczne opisujące *hardware*, *software* i środowisko.

‘Obel’ - zasadnicze elementy



XML pliki:

- *.xml (plik z danymi)
- pliki do prezentacji w sieci www
 - *.html lub
 - *.css lub
 - *.xsl



- C:\.....\ra1.sxv (=.zip)
- Taka jest zawartość:

Nazwa	Typ	Rozmia...	Ma h...	Roz...	St...	Data
META-INF	Folder plików	0 KB		0 KB	0%	
content	Dokument XML	6 KB	Nie	48 KB	88%	2004-04-29 06:15
layout-cache	Plik	1 KB	Nie	1 KB	0%	2004-04-29 06:15
meta	Dokument XML	2 KB	Nie	2 KB	0%	2004-04-29 06:15
mimetype	Plik	1 KB	Nie	1 KB	0%	2004-04-29 06:15
settings	Dokument XML	2 KB	Nie	8 KB	81%	2004-04-29 06:15
styles	Dokument XML	2 KB	Nie	10 KB	82%	2004-04-29 06:15

Obiekty elektroniczne przechowywanie

- ERPANET
- *Digital Preservation Coalition*
- *Digital Curation Centre*
- Inna droga: DAW



erpanet

<http://www.erpanet.org/>

ER PANET

ER PANET is concerned with addressing the lack of awareness, fragmentation of knowledge and skills amongst the stakeholder communities about how to handle existing digital preservation problems, and how to plan effectively for the future. It aims to address the lack of identification and focus on core research/problem areas and bring coherence and consistency to activities in this area.

- Digital records and resources will form the building blocks for learning and leisure, for scientific and historical research and for new business in the "ambient intelligence landscape". The effective and affordable preservation of digital resources (whether digital objects, eContent, or electronic records) of European culture and science, as well as ubiquitous records of social and economic change, is fundamental to the overall aim of providing "new tools and business models for service design and provision and for content creation and delivery".
- Increasing amounts of Europe's cultural and scientific heritage is being created or represented in digital form. The preservation and reuse of these digital assets forms both the cornerstone of future economic growth and development, and the foundation for the future of memory. This material represents Europe's heritage and is its future intellectual capital. The fast pace of change in the technological landscape makes ensuring long-term access to this material a challenge. As a result of these technological advances preserving digital assets cannot happen as an after-thought, it needs to be planned. Policies, technical methods and strategies are required because media degrade (e.g. magnetic particles lose their properties and dye layers on optical media break down), technological developments make systems obsolete, or information is rendered inaccessible by changes in encoding formats. For instance, even where it is possible to recover data from the obsolete media it will be in a range of file formats including word processing, sound, text, image, and database file formats. Little will be generic enough to be accessed without the original application. Some of these will be proprietary and others will require particular versions of software that was long-since superseded.
- We widely recognise the benefits of using digital information and a result of its prevalence is the emerging vision of Europe as an information rich society whose record is just waiting to be harvested and processed by the technology-enabled researcher of the future or by emerging eContent industries. Ensuring this vision depends upon the survival of digital data in accessible and usable form

ERPANET - aims

- **ERPANET project aims to establish an expandable and self-sustaining European Initiative, which will serve as a virtual clearinghouse and knowledge-base in the area of preservation of cultural heritage and scientific digital objects.**
- **The dominant feature of ERPANET will be the exchanging of knowledge on state-of-the-art developments in digital preservation and the transfer of expertise among individuals and institutions. More specifically ERPANET will deliver a range of services (e.g. content creation, advisory service, training and thematic workshops and fora), both to information creation and user community. It will make accessible tools, knowledge, and experience. ERPANET will not directly carry out new research to develop such tools, but it will create a coherent platform for proactive co-operation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects. It will bring together research institutions, memory organisations, ICT industry, entertainment and creative (e.g. broadcasting) industries and provide effective, multidisciplinary, knowledge and resource-sharing infrastructure.**



<http://www.dpconline.org/>



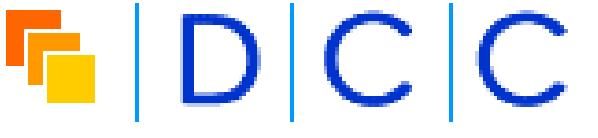
- The Digital Preservation Coalition was established in 2001 to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base. These pages provide background information on our mission, programmes, administration and membership.

Mission and Goals

- The aim of the Digital Preservation Coalition is to secure the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base.
- In order to achieve this aim, the Coalition has the following long-term goals:
- producing, providing, and disseminating information on current research and practice and building expertise amongst its members to accelerate their learning and generally widen the pool of professionals skilled in digital preservation.
- Instituting a concerted and co-ordinated effort to get digital preservation on the agenda of key stakeholders in terms that they will understand and find persuasive.
- Acting in concert to make arguments for appropriate and adequate funding to secure the nation's investment in digital resources and ensure an enduring global digital memory.
- Providing a common forum for the development and co-ordination of digital preservation strategies in the UK and placing them within an international context.
- Promoting and developing services, technology, and standards for digital preservation.
- Forging strategic alliances with relevant agencies nationally and internationally, and working collaboratively together and with industry and research organisations, to address shared challenges in digital preservation.
- Attracting funding to the Coalition to support achievement of its goals and programmes.

Principles of the Coalition

- The activities and members of the Coalition will operate by the following principles:
- *Openness*: The Coalition and its members commit to promoting and disseminating information and sharing outcomes so that we can all learn and benefit as quickly as possible from transferable lessons and experience (both positive and negative).
- *Collaboration*: Digital preservation has become so significant a phenomenon (in scope, complexity, and investment), that no single organisation can address all the challenges alone.
- The Coalition provides a forum for members to identify relevant issues and support to pursue collaboration across organisations and sectors to mutual benefit.
- *Collective benefit*: Core Coalition activities supported by resources from its membership must be of common interest and benefit to them. Projects or the activities of individual members may have narrower collective benefit and can contribute to the wider goals of the Coalition but are separately funded.
- *Vendor neutrality*: The goals of the Coalition are generic and will be vendor neutral. It will support the development of standards and generic approaches to digital preservation, which can be implemented by a range of hardware, software, and service vendors.



Digital Curation Centre (DCC)

<http://www.dcc.ac.uk/>

Digital Curation Centre (DCC)

- The JISC and the eScience Core Programme have entrusted the task of establishing the Digital Curation Centre (DCC) to a consortium comprising four partner institutions: the [University of Edinburgh](#) (lead partner) and the [University of Glasgow](#), which together host the NeSC; [UKOLN](#), at the University of Bath; the Council for the [Central Laboratory of the Research Councils](#) (which operate the Rutherford and Daresbury Laboratories).
- Using the integrating ideas of collaboration, curation and continuing access for data held in institutional and national data repositories, we will be supporting a '[Collaborative Associates Network of Data Organisations](#)' in order that the DCC engages with and benefits from the progress being made by leaders within communities of practice across the wide range of scholarly and scientific disciplines, nationally and internationally.
- We see our overriding purpose to be continuing improvement in the quality of data curation and digital preservation, recognising that data have importance as the evidential base for scholarly conclusions, and for the validation of those conclusions. In interests of securing consensus, we propose the term curation to cover the active management and appraisal of data over the corresponding lifecycle of scholarly and scientific interest: it is thus the key to reproducibility and reuse.
- The DCC is not itself to be a digital repository, nor an attempt to impose policies and practices of one branch of scholarship upon another. Rather, based on insight from a vibrant research programme that addresses wider issues of data curation, it will develop and offer programmes of outreach and practical services to assist those who must curate data. The challenge is to build the DCC and its associate network in ways that promote the positive interplay between research, development, services and outreach.

What is digital curation?

- The funders of the Digital Curation Centre began the work of defining the concepts, goals, tasks and research needs associated with this challenging area in their revised circular of calls for proposals, in June, 2003.
- *The term digital curation is used in this call for the actions needed to maintain digital research data and other digital materials over their entire life-cycle and over time for current and future generations of users. Implicit in this definition are the processes of digital archiving and preservation but it also includes all the processes needed for good data creation and management, and the capacity to add value to data to generate new sources of information and knowledge.*
- *Curation and long-term preservation of digital resources will be of increasing importance for a wide range of activities within research and education. Through sensors, experiments, digitisation and computer simulation, digital resources and data are growing in volume and complexity at a staggering rate. The cost of producing these resources is very high: satellites, particle accelerators, genome sequencing, and largescale digitisation and electronic publishing collectively represent a cumulative investment of billions of pounds in digital research and learning.*
- *Long-term curation and preservation of digital resources is seen as a challenge which is difficult if not impossible for individual institutions to resolve on their own due to the complexity and scale of the challenges involved. The JISC has therefore been active on behalf of the sector in promoting research, partnership, and services in this field.*
- A bibliography for further reading is appended to the document.

Our vision statement: Evidence and Enlightenment

- Data have importance as the evidential base for scholarly conclusions, and for the validation of those conclusions, a basic tenet of which is reproducibility. Curation is the active management and appraisal of data over the life-cycle of scholarly and scientific interest; it is the key to reproducibility and re-use. This adds value through the provision of context and linkage: placing emphasis on 'publishing' data in ways that ease re-use, with implications for metadata and interoperability. Metadata for resource discovery and retrieval are also important, with mark-up on time/place referencing as well as subject description and linkage to discipline-based ontologies. Special emphasis is required on the descriptive information that allows effective re-analysis of datasets of scientific and scholarly significance, and re-use in new and unexpected contexts, e.g. e-Learning or science history. The demands for linkage to two further domains of scholarly communication and e-Learning must also be understood.
- Science and scholarship cut across disciplinary boundaries. So too does digital curation: appreciation of differences between disciplines is essential for understanding and consensus building. An open and creative culture is necessary to foster the flow of ideas between research and practice, provoking research with practical challenges and informing practice from theory and experiment. This lays the grounds for leadership and advocacy; for continuing professional development; for matters requiring mediation; for building consensus and for the adoption of common standards. Use of open standards enables re-use and re-purposing by avoiding dependencies of platform and application software.
- Digital curation, broadly interpreted, is about maintaining and adding value to, a trusted body of digital information for current and future use. The digital archiving and preservation community now looks beyond the preservation, cataloguing and cross referencing of static digital objects such as documents. The scientific community has data characterised by structure, volatility and scale. These require us to extend our notions of curation. We must also investigate the principles that underlie appraisal, and lessons learnt about the economics of preservation.

The UK Census of Population and Housing

Brytyjski Spis Powszechny - stan na 29 kwietnia 2001

- 30 million public
- 20 typów formularzy o rozmiarach od 1 do 20 stron

Uzyskane materiały to :

- 37 miles półek
- 0.4 mln zeskanowanych stron
- 28 terabytes of image and data storage
 - 250 - 300 million frames of 16mm roll microfilm (w 6 miesięcy)
 - for long term archiving and delivery to Public record Office (PRO)

„Trzecia droga” - digital to analog

Kodak Document Archive Writer i9600 Series Writers



Σ

- Czy zawsze konieczne są dokumenty oryginalne?
- Centrum certyfikacji (ang.: *trusted third part* - TTP → elektroniczny notariat).
- Nieprzerwany profesjonalny łańcuch działań ludzi obdarzonych społecznym zaufaniem (archiwistów, IT, TTP).