

## डिजिटल डेटा के संरक्षण की चुनौती और भारत में इस दिशा में हो रहे प्रयास

### Challenges in Preservation of Digital Data and Indian Initiatives in this Direction

प्रतापानन्द झा

निदेशक (कल्चरल इन्फॉर्मेटिक्स), इन्दिरा गाँधी राष्ट्रीय कला केन्द्र, नई दिल्ली

Pratapanand Jha, Director (Cultural Informatics),

Indira Gandhi National Centre for the Arts, New Delhi

(pjha@ignca.nic.in)

सारांश :

डिजिटल डेटा आज हमारी सभी गतिविधियों का एक महत्वपूर्ण हिस्सा है। हम अपने सभी दस्तावेजों को डिजिटल रूप से बनाते हैं। इसलिए हर दिन हम बड़ी मात्रा में डिजिटल डेटा बना रहे हैं। इसके अलावा, डिजिटल प्रारूप में उपलब्ध एनालॉग डेटा के रूपांतरण के लिए एक विशाल कार्यबल दुनिया भर में कार्यरत है। डिजिटल डेटा के संरक्षण के प्रति हमारा व्यवहार आज भी उतना दायित्वपूर्ण नहीं है। इसीलिए, डिजिटल प्रारूप में डेटा के स्थायी क्षति का प्रतिशत अभी भी बहुत अधिक है। आवश्यकता है कि इसे व्यक्तिगत, संस्थागत तथा राष्ट्रीय स्तर पर एक औपचारिक प्रणाली से जोड़ा जाय जो डेटा बनने के साथ ही, किसी न किसी प्रणाली के तहत सुरक्षित रह सके। डिजिटल डेटा के दीर्घकालिक संरक्षण के लिए, ट्रस्टेड डिजिटल रिपोजिटरी (टीडीआर) के मानकों को आईएसओ (ISO) द्वारा 2012 में अनुमोदित किया गया था। लेकिन अब तक दुनिया भर में केवल दो रिपोजिटरी को ही आईएसओ 16363 के अन्तर्गत टीडीआर के लिए प्रमाणित किया गया है। एनसीएए (National Cultural Audiovisual Archives, NCAA) दुनिया का पहला आईएसओ प्रमाणित टीडीआर है, जिसका विवरण यहाँ संक्षेप में प्रस्तुत किया गया है। इसकी प्रक्रिया किसी भी आपदा की स्थिति में डेटा के रिकवरी का ध्यान रखती है। हालांकि, एक ट्रस्टेड डिजिटल रिपोजिटरी की संगठनात्मक और/या वित्तीय विफलता की संभावना हमेशा बनी रहती है इसलिए, हरेक टीडीआर के साथ उसका एक स्टैंडबाय टीडीआर, जो आपदा के समय उसके डेटा को एक विश्वसनीय उत्तराधिकारी के रूप में सुरक्षित रख सके, एक समाधान के रूप में प्रस्तावित किया गया है। राष्ट्रीय स्तर पर ट्रस्टेड डिजिटल रिपोजिटरी की एक ग्रिड तैयार करने की जरूरत है जो सभी डेटा के संरक्षण और उसकी सुरक्षा को सुनिश्चित कर सके।

**Abstract:**

*Today, digital data is part of all our activities. We create all documents digitally. Hence everyday we generate large volume of digital data. In addition, a huge work force is employed worldwide for conversion of available analog data in digital format. Our behavior towards the protection of digital data is not very responsible. Therefore, the percentage of permanent loss of data in digital format is still very high. It is required that with the creation of data it must be linked to a formal system at the individual, institutional and national levels, to ensure its long term availability. For preservation of digital data, standards for Trustworthy Digital*

*Repository (TDR) have been framed under ISO in 2012 but till date only two repositories (worldwide) have been certified for ISO 16363 to become a TDR. The National Cultural Audiovisual Archives (NCAA) has been certified as world's first ISO Certified TDR, has been presented here in brief, as a case study. The process takes care of data recovery in case of any disaster. However, there is always a finite possibility of organizational and/or financial failure of a Trustworthy Digital Repository. Therefore, creating a standby Digital Repository which, in case of the failure of TDR, would function as a trusted inheritor of its data has been proposed as a solution. At national level, there is a need to create a grid of trusted digital repositories that can ensure the protection and security of all the data.*

**मुख्य शब्द:** आपदा प्रबंधन, डिजिटल संरक्षण, डिजिटल आर्काईव्स, डिजिटल लेखागार, ट्रस्टवर्दी डिजिटल रिपोजिटरी, ट्रस्टेड डिजिटल रिपोजिटरी, टीडीआर, नेशनल कल्चरल ऑडियोविजुअल आर्काईव्स

**Keywords:** Disaster Management, Digital Preservation, Digital Archive, Trustworthy Digital Repository, Trusted Digital Repository, TDR, National Cultural Audiovisual Archives, NCAA,

**पृष्ठ भूमि :**

डिजिटल डेटा आज हमारी सभी गतिविधियों का अंतरंग हिस्सा है। अपने सभी दस्तावेजों को हम डिजिटल रूप से बनाते हैं। लेखन (टेक्स्ट), फोटोग्राफ, ऑडियो, वीडियो, एनीमेशन आदि बनाने के लिए बाजार में उपलब्ध सभी गैजेट्स डिजिटल प्लेटफॉर्म पर काम करते हैं और डिजिटल आउटपुट देते हैं। मोबाइल फोन या कंप्यूटर पर आधारित हमारे सभी संचार तंत्र पूरी तरह से इसी माध्यम पर आधारित हैं। अधिकांश लेनदेन चाहे वह आधिकारिक, निजी, बैंकिंग, परिवहन, चिकित्सा, मनोरंजन या खरीदारी का हो, हम इसे ऑनलाइन कर रहे हैं। डिजिटल तकनीक ने आधुनिक जीवन के लगभग हर पहलू को बदल दिया है। यात्रा, कार्यक्षेत्र, खरीदारी, मनोरंजन और संचार ऐसे क्षेत्र हैं, जिनमें हाल के दशकों में अभूतपूर्व क्रांति हुई है। अब कोई भी उपकरण या मशीनरी, जो किसी भी प्रकार के डिजिटल तकनीक को शामिल नहीं करता है, मिलना दुर्लभ हो गया है। डिजिटल तकनीक का मतलब है कि उपकरण या डिवाइस का अधिक कॉम्पैक्ट, कुशल, हल्का (आसानी से उठाने योग्य), सुविधाजनक और उपयोगी होना। इसकी सहायता से बड़ी मात्रा में सूचना को स्थानीय

या दूरस्थ रूप से संग्रहीत किया जा सकता है और लगभग तुरंत ही प्रकाशित किया अथवा संचारित किया जा सकता है। यहां तक कि "सूचना" शब्द अब केवल शब्दों और संख्याओं को संदर्भित नहीं करता अपितु उसका विस्तार अन्य मीडिया, जैसे फोटो, ऑडियो और वीडियो तक को शामिल करने के लिए भी होने लगा है। परिणामतः, हर दिन हम बड़ी मात्रा में डिजिटल डेटा का निर्माण कर रहे हैं।

इसके अलावा, उपलब्ध एनालॉग डेटा के डिजिटल प्रारूप में रूपांतरण के लिए एक विशाल कार्यबल दुनिया भर के कई संस्थानों में कार्यरत है। तथापि, अनेक संस्थान इस ज्ञान को भविष्य की पीढ़ियों तक संरक्षित करने के लिए पूरी तरह से तैयार नहीं दिखते हैं। डिजिटल प्रारूप में डेटा की स्थायी क्षति होने का प्रतिशत अभी भी बहुत अधिक है। वर्ल्ड इकोनॉमिक फोरम (World Economic Forum) के एक रिपोर्ट के अनुसार, सन् 2020 तक डिजिटल संसार का विस्तार 44 जेट्टाबाइट तक पहुंचने की उम्मीद है। साथ ही यह अनुमान है कि सन् 2025 तक प्रतिदिन 463 एक्साबाइट डेटा

का निर्माण होगा।<sup>1</sup> डिजिटल डेटा का भंडारण (storage) बाईट्स (मेगा, गीगा, टेरा, पीटा, एक्सा ...) में होता है। हरेक बाईट आठ बिट से बनी होती है और हर बिट की वैल्यू 0 (जीरो) या 1 (वन) होता है। इसे बाईनरी कोड (Binary code) प्रणाली कहते हैं। कम्प्यूटर इसी भाषा में हर कमांड को समझता है चाहे वह text (लेखन) हो या विजुअल हो या निर्देश हो, प्रत्येक उसके अनुरूप ही कार्य करते हैं। डिजिटल डेटा को बिना तकनीकी सहायता से पढा नहीं जा सकता है। तकनीकी अप्रचलन (outdating) के कारण डिजिटल दस्तावेज हमेशा विलुप्ति की कगार पर रहते हैं। डिजिटल सामग्री के आपदा प्रबंधन के लिए आज हमारे पास अनुकरणीय उदाहरण नहीं है। डिजिटल में सूचना बनाए जाने के साथ ही उसके संरक्षण के लिए योजना बनाना तथा तदनुसार कार्य शुरू करना आवश्यक है।

#### डिजिटल डेटा के संरक्षण की चुनौतियाँ :

डिजिटल डेटा के संरक्षण की प्रमुख चुनौतियाँ हैं:—

1. डिजिटल मीडिया का जल्दी अप्रचलन
2. डेटा का मालिकाना स्वरूप (proprietary format)
3. फाइल को खोलने के लिए आवश्यक सॉफ्टवेयर का अप्रचलन
4. मानकों में निरन्तर बदलाव
5. सॉफ्टवेयर, हार्डवेयर और ऑपरेटिंग सिस्टम के अप्रचलन के पूर्वाभास के साथ उसके उन्नयन (Upgrade) की अग्रिम योजना बनाना, एवं
6. कॉपी राइट्स और बौद्धिक संपदा अधिकार (IPR): डिजिटल रूप में लोगों तक सूचना उपलब्ध कराने में सबसे बड़ी बाधा प्रौद्योगिकी से संबंधित नहीं है, बल्कि कॉपीराइट के मुद्दों से है। एक डिजिटल लाइब्रेरी अपने संग्रह को संरक्षित करने के साथ-साथ उसे इंटरनेट पर

उपलब्ध कराती है। यह डेटा कई जगहों से संग्रहीत होता है, और इसके स्रोत एवं स्वरूप भी बहुविध होते हैं। ऐसा बहुत सा डेटा है जिसके कॉपीराइट की स्थिति स्पष्ट नहीं है किन्तु सूचना के विचार से वह अत्यन्त महत्वपूर्ण है। बिना अपेक्षित अनुमति के उसे इंटरनेट पर उपलब्ध कराना अनुचित है क्योंकि डिजिटल डेटा की नई प्रतियाँ डिजिटल मूल के समान होती हैं।

#### वैश्विक स्तर पर किये जा रहे प्रयास :

Memory of the World Programme के अन्तर्गत, यूनेस्को (UNESCO) से डिजिटल डेटा सहित दुनिया की दस्तावेजी विरासत की सुरक्षा के लिए एक मानक—सेटिंग साधन विकसित करने का आग्रह किया गया है। प्रारंभिक अध्ययन के आधार पर यूनेस्को जनरल कॉन्फ्रेंस ने अपने 37वें सत्र में महानिदेशक को इस विषय पर एक मसौदा तैयार करने के लिए संस्तुति प्रदान की है। 1 और 2 जुलाई 2015 को आयोजित 38वें सत्र में, “Recommendation Concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form” की सिफारिशों को मंजूरी दी गई, जो यूनेस्को के कार्यों में दस्तावेजी विरासत के संरक्षण के लिए मील का एक पत्थर है। सिफारिश का उद्देश्य सदस्यों को राष्ट्रीय स्तर पर, दस्तावेजी विरासत और नीतिगत उपायों की पहचान कराना, और उसके उपयोग और संरक्षण के लिए द्विपक्षीय या बहुपक्षीय अनुसंधान परियोजनाओं और प्रकाशन दिशानिर्देशों, नीतियों और सर्वोत्तम अभ्यास मॉडल के माध्यम से अंतर्राष्ट्रीय पैमानों पर सहायता प्रदान करना है। यह नया मानक—सेटिंग साधन, समय की बरबादी, प्राकृतिक आपदाओं, मानव व्यवहार, प्रौद्योगिकी अप्रचलन, संसाधनों को जुटाने जैसे खतरों के लिए उपयुक्त समाधानों की पहचान करने के लिए साझेदारी बनाने में मदद करेगा, ताकि मूल्यवान संग्रह और रिकॉर्ड कभी खो न सकें। इसलिए, नया साधन स्पष्ट रूप से सदस्यों को अपने विशिष्ट संदर्भों के

1. <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/>

साथ इसे संरेखित करने के लिए उचित कदम उठाने की सिफारिश करता है, इसे अपने राष्ट्रीय क्षेत्रों में व्यापक रूप से प्रसारित करता है, और साथ ही यह समर्थन नीतियों, रणनीतियों और कानून के निर्माण और अपनाने के माध्यम से इसके कार्यान्वयन की सुविधा प्रदान करता है। यह भी निर्णय लिया गया कि हर चार साल में सदस्य इस सिफारिश को प्रभावी करने के लिए की गई कार्रवाई पर रिपोर्ट देंगे।<sup>2</sup>

इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय, भारत सरकार ने 2007 में डिजिटल परिरक्षण की आवश्यकता को देखते हुए, C-DAC पुणे को National Digital Preservation Programme के कार्यान्वयन के लिए अनुमोदन प्रदान किया। इसके अन्तर्गत अभी तक अनेक सफलताएं मिली हैं जिनमें डिजिटल प्रिजर्वेशन में इंटरनेशनल ट्रेंड्स पर कार्यशाला का आयोजन, C-DAC में डिजिटल प्रिजर्वेशन की उत्कृष्टता के लिए एक केंद्र स्थापित करना, भारत के डिजिटल परिरक्षण आवश्यकताओं पर राष्ट्रीय अध्ययन रिपोर्ट का प्रकाशन आदि प्रमुख हैं।<sup>3</sup>

डिजिटल डेटा के दीर्घकालिक संरक्षण के लिए, ट्रस्टेड डिजिटल रिपोजिटरी (टीडीआर) के मानकों को 2012 में अनुमोदित किया गया। आईएसओ (ISO) 16363 ऑडिट एक भंडार को भरोसेमंद और जिम्मेदार डेटा प्रबंधन के लिए सक्षम होने के लिए अच्छी तरह से परीक्षण की गई प्रक्रिया का हिस्सा है। आईएसओ 16363 प्रमाणन प्राप्त करने से एक भंडार को वहाँ संरक्षित डेटा की गुणवत्ता और स्थिरता, डेटा की अखंडता और उनकी देखभाल के लिए सौंपी गई जानकारी के लंबे समय तक संरक्षण और उपयोग के लिए अपनी प्रतिबद्धता दिखाने की अनुमति मिलती है। लेकिन, अब तक दुनिया भर में केवल दो रिपोजिटरी को ही आईएसओ (ISO)

16363 के अन्तर्गत टीडीआर के लिए प्रमाणित किया गया है। दुनिया के पहले आईएसओ प्रमाणित टीडीआर, National Cultural Audiovisual Archives (NCAA)<sup>4</sup>, जिसे नवंबर 2017 में प्रमाणित किया गया, तथा The United States Government Publishing Office (USGPO) को दिसंबर 2018 में टीडीआर के लिए Primary Trustworthy Digital Repository Authorisation Body Ltd (PTAB)<sup>5</sup>, United Kingdom द्वारा प्रमाणित किया गया। PTAB दुनिया का पहला संगठन है जिसे दुनिया भर में (ISO) 16363 ऑडिट और प्रमाणन करने के लिए National Accreditation Board for Certification Bodies of India (NABCB) द्वारा मान्यता प्राप्त है।

भारतीय परिदृश्य:

भारत की सांस्कृतिक सम्पदा का एक बड़ा हिस्सा दृश्य-श्रव्य संग्रह के रूप में है, जो पिछले छह दशकों में बनाई गई, विभिन्न सरकारी और गैर-सरकारी संस्थानों और निजी संग्रहों में संचित है। इनमें भारत के कुछ महान कलाकारों की प्रतिभा भी संरक्षित है। यह एक अमूल्य राष्ट्रीय धरोहर है जिसे संरक्षित रखने और देश के नागरिकों को सुलभ रूप में उपलब्ध कराने की आवश्यकता है। तकनीकी में लगातार बदलाव, जागरूकता की कमी, उचित रखरखाव का अभाव तथा चुम्बकीय माध्यम की नाजुकता (जिनमें ये डेटा संग्रहित है), के कारण इन सामग्रियों पर हमेशा के लिए खो जाने का खतरा बरकरार है। इस मुद्दे की गंभीरता को समझते हुए, 12वीं पंचवर्षीय योजना में संस्कृति मंत्रालय ने सांस्कृतिक श्रव्य-दृश्य सामग्रियों के संरक्षण के लिए National Cultural Audiovisual Archives (NCAA) परियोजना का प्रस्ताव रखा, जिसके कार्यान्वयन का दायित्व इन्दिरा गाँधी राष्ट्रीय कला केन्द्र को दिया गया।

2. <https://en.unesco.org/programme/mow/recommendation-documentary-heritage>

3. [https://www.cdac.in/index.aspx?id=mc\\_hc\\_national\\_digi\\_pre](https://www.cdac.in/index.aspx?id=mc_hc_national_digi_pre)

4. <http://www.iso16363.org/iso-certification/certified-clients/>

5. <http://www.iso16363.org>

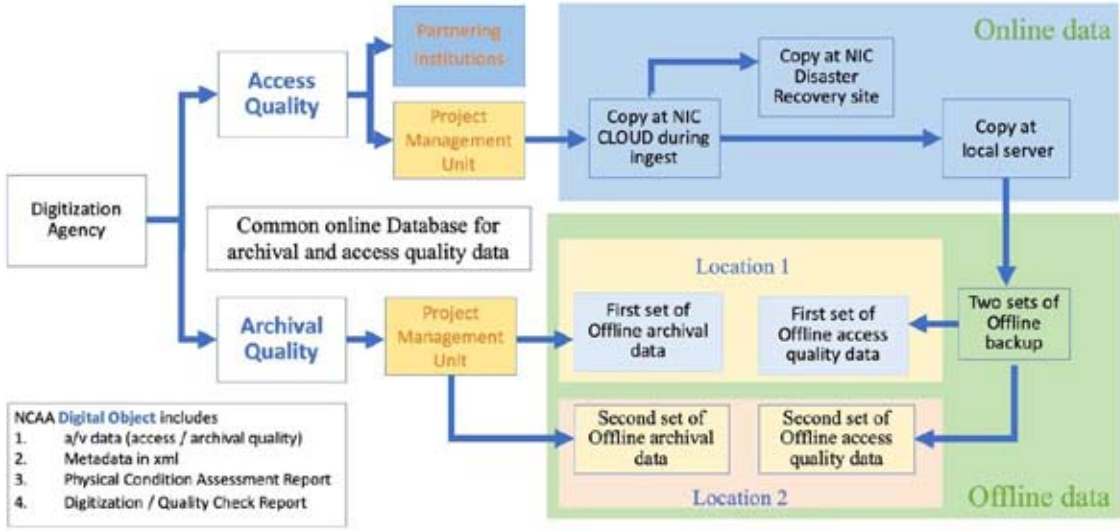
परियोजना के प्रारंभिक (पायलट) चरण में लगभग 32,000 घंटे के ऑडियो और वीडियो को डिजिटलाइज कर वेबसाइट <http://ncaa.gov.in> पर उपलब्ध कराया गया है। देश भर के कई सरकारी और गैर-सरकारी संस्थानों में उपलब्ध अप्रकाशित, गैर-व्यावसायिक ऑडियो-वीडियो रिकॉर्डिंग को प्रलेखित और डिजिटलाइज कर आनलाईन प्रस्तुत किया गया है। इन रिकॉर्डिंग्स में निहित भारत की सांस्कृतिक विरासत को जनमानस तक सुलभ रूप में पहुँचाने का यह एक अनूठा प्रयास है। मुख्य रूप से कलाकारों, शोधकर्ताओं, विद्वानों, छात्रों, वृत्तचित्र फिल्म निर्माताओं, फिल्म और टेलीविजन पेशेवरों, लेखकों, और GLAM (गैलरी, पुस्तकालय, अभिलेखागार और संग्रहालय) पेशेवरों, संस्कृति के प्रति उत्साही और पारखी लोगों के लिए National Cultural Audiovisual Archives (NCAA) एक प्रमुख संसाधन है। NCAA का उद्देश्य भारत में मौखिक परंपराओं, प्रदर्शन कलाओं, अनुष्ठान और सांस्कृतिक प्रथाओं और भारत के पारंपरिक ज्ञान प्रणालियों तक दीर्घकालिक पहुंच सुनिश्चित कराना है।

NCAA डिजिटल रिपोजिटरी का विकास ओपन-आर्काइव इन्फॉर्मेशन सिस्टम (OAIS) रेफरेंस मॉडल के अनुपालन में C-DAC, पुणे द्वारा डिजिटल संरक्षण केंद्र के सहयोग से किया गया है। एनआईसी ने संग्रह को ऑनलाइन सुलभ बनाने के लिए आवश्यक बुनियादी ढाँचे को बढ़ाया है। डिजिटलीकरण का मानक इंटरनेशनल एसोसिएशन ऑफ साउंड एंड ऑडियोविजुअल आर्काइव्स (आईएएसए IASA) और मेटाडेटा डबलिन कोर द्वारा निर्धारित मानकों पर आधारित है। जैसा कि ऊपर बताया गया है, इसे विश्व के प्रथम आईएसओ ISO प्रमाणित टीडीआर के रूप में प्रमाणित किया गया है। ऑडिट की प्रक्रिया के माध्यम से टीडीआर प्रमाणीकरण, डेटा संग्रहण नीति और आपदा रिकवरी योजना के महत्वपूर्ण हिस्सों को मंजूरी दी गई है।

NCAA के अन्तर्गत डेटा प्रबंधन योजना बनाई गई जिसे नीचे चित्र में दर्शाया गया है। डिजिटलाइज्ड डेटा को अभिलेखीय (Archival) और एक्सेस क्वालिटी (कम्प्रेस्ड, छोटे आकार में अभिलेखीय से संप्रेषित) गुणवत्ता में संग्रहीत किया जाता है। "अभिलेखीय" शब्द का अर्थ है उच्च गुणवत्ता वाली वस्तु, जो दीर्घकालिक संरक्षण के लिए उपयुक्त है। NCAA में, अभिलेखीय डेटा एनालॉग रिकॉर्डिंग से प्राप्त किया गया उच्चतम गुणवत्ता वाले डिजिटल मास्टर है, जो इस परियोजना के लिए निर्धारित मानकों के अनुसार हो। ऑडियो के लिए यह एलपीसीएम (LPCM) में इनकोडेड बीडब्लूएफ (BWF) फाइल फॉर्मेट है। वीडियो के लिए एमएक्सएफ (MXF) आवरण में जेपीईजी2000 (JPEG2000) फाइल फॉर्मेट है। आम तौर पर, ये फाइलें आकार में बहुत भारी होती हैं और इसलिए ऑनलाइन प्रकाशन के लिए उपयुक्त नहीं होती हैं। उसकी एक कम्प्रेस्ड फॉर्मेट, जिसे एक्सेस क्वालिटी डेटा कहा जाता है, का उपयोग ऑनलाइन प्रकाशन के लिए किया गया है। ऑडियो और वीडियो के लिए एक्सेस क्वालिटी के रूप में क्रमशः MP3 और MP4 प्रारूप का उपयोग किया गया है। इस परियोजना के लिए निर्धारित डिजिटलीकरण और मेटाडेटा मानकों<sup>6</sup> को एनसीए की वेबसाइट पर देखा जा सकता है।

अभिलेखीय गुणवत्ता मुख्य रूप से आगे के अनुसंधान और संपादन के लिए है, जबकि एक्सेस क्वालिटी डेटा ऑनलाइन प्रकाशन के लिए है। प्रत्येक डिजिटल ऑब्जेक्ट में (i) ऑडियो/वीडियो डेटा (अभिलेखीय/कम्प्रेस्ड), (ii) xml में मेटाडेटा, (iii) स्थिति मूल्यांकन रिपोर्ट और (iv) डिजिटलीकरण /गुणवत्ता जांच रिपोर्ट शामिल हैं। अभिलेखीय और कम्प्रेस्ड डेटा के लिए एक ही डेटाबेस है। एक्सेस क्वालिटी (कम्प्रेस्ड) डेटा हार्ड ड्राइव के दो समान सेटों में प्राप्त होता है। एक सेट को संबंधित संस्था (मूल डेटा धारक) और दूसरे सेट को NCAA में

6. <http://ncaa.gov.in/repository/common/digitizationmetadastandards>



एनआईसी क्लाउड पर डेटाबेस के साथ कॉपी करने के लिए जमा किया जाता है। प्राधिकृत निदेशक द्वारा डेटा और मेटाडेटा का सत्यापन एवं आनलाईन प्रकाशन के बाद उसे सार्वजनिक रूप से उपलब्ध कराया जाता है। एनआईसी क्लाउड की डुप्लीकेट कॉपी एनआईसी के डिजास्टर रिकवरी साइट पर रखी गई है। आईजीएनसीए के स्थानीय सर्वर पर भी एक प्रति रखी जाती है, जिसे एनआईसी क्लाउड से समय-समय पर अद्यतन किया जाता है। ऑनलाइन एक्सेस क्वालिटी डेटा के उपरोक्त तीन सेटों के अलावा, और डिजिटलीकरण एजेंसी से प्राप्त हार्ड ड्राइव के डेटा का एलटीओ टेप में ऑफलाइन माध्यम में स्टोर करने के लिए दो सेट बनाये जाते हैं। अभिलेखीय डेटा LTO के दो समान सेटों में प्राप्त होता है। अभिलेखीय और एक्सेस क्वालिटी डेटा के इन दोनो सेटों को दो अलग-अलग स्थानों पर, अलग-अलग भूकंपीय क्षेत्रों में तथा अलग-अलग पहुंच और पर्यावरण नियंत्रित क्षेत्रों में रखा जाता है। NCAA के लिए ऑफलाइन डेटा सेंटर, IGNSA के नई दिल्ली एवं बंगलुरु ऑफिस में स्थित है।

NCAA के कार्यान्वयन में जो प्रमुख चुनौतियाँ थी उनमें, मेटाडेटा बनाना एक था। खासकर कई व्यक्तिगत संग्रहों में वस्तु (टेप, कैसेट..आदि) के अलावा

किसी भी तरह का दस्तावेज (documentation) नहीं होने के कारण, उसे बनाना लगभग असंभव था। प्लेयर के अभाव में रेकॉर्डेड कंटेंट देखा/सुना नहीं जा सकता था। अतः हमने मेटाडेटा को तीन पार्ट (administrative – जो संग्रहकर्ता के द्वारा प्रारंभ में, descriptive – जो शोधार्थी के द्वारा डिजिटिजेशन के बाद, तथा technical – डिजिटिजेशन एजेंसी के द्वारा डिजिटिजेशन के बाद) में बनवाया। हरेक स्तर पर संग्रहकर्ता और विषय विशेषज्ञों का सहयोग लिया गया जिससे हम संदर्भ को सत्यता के साथ प्रस्तुत कर सकें। शोधार्थियों को कई स्तर पर प्रशिक्षण दिया गया, तथापि विषय-वस्तु का क्षेत्र इतना बड़ा है कि संशोधन की विकल्प हमेशा बना रहेगा। दूसरी प्रमुख चुनौती थी डिजिटिजेशन की। आडियो और विडियो के संग्रह जो विभिन्न वातावरण में दशकों से रखे हैं, अलग-अलग फॉर्मेट में, कई कम्पनी द्वारा बनी हुई मीडिया में, अलग-अलग स्पीड में रेकॉर्डेड हों, को एक मानक के तहत एक संग्रह में लाना। कोई भी अपने संग्रह को बाहर भेजना नहीं चाहता है। साथ ही डिजिटिजेशन के अन्तर्गत अधिकतम सफलता के लिए इन माध्यमों को महिनो वातानुकूलित कक्षों में रखना था, जिससे डिजिटिजेशन चक्र तीन से चार महिने का हो गया था। ऐसे में संग्रहकर्ताओं

को विश्वास दिलाना कि आपका संग्रह सकुशल वापस आयेगा, शुरु में आसान नहीं था। बाद में जब हमने कुछ संस्थानों का कार्य कुशलतापूर्वक किया, तदुपरांत कई संस्थायें आगे आयीं। तीसरी और सबसे महत्वपूर्ण चुनौती थी, संग्रहों में उपलब्ध डेटा का कॉपी राइट्स और बौद्धिक संपदा अधिकार (IPR) पता करना। संस्थाओं के पास ऐसा बहुत सा डेटा है जिसकी कॉपीराइट स्थिति स्पष्ट नहीं है किन्तु सूचना के विचार से वह अत्यन्त महत्वपूर्ण है। अतः NCAA के अन्तर्गत प्रोफेसर दामोदरन (MHRD IPR Chair Professor, IIM Bangalore) की अध्यक्षता में एक परामर्शी दस्तावेज बनाया गया, जो प्रकाशनाधीन है और जल्द ही NCAA website पर उपलब्ध होगा। इसके अन्तर्गत संसाधनों के सार्वजनिक रूप से डिजिटल प्रसार की संभावना के बारे में स्पष्टता हासिल करने के लिए, एक विस्तृत दिशा-निर्देश प्रस्तुत किया गया है। इस दिशा में यह एक अमूल्य योगदान है जिसका उपयोग देश भर के सांस्कृतिक संस्थान कर सकेंगे।

अभी केवल कुछ वर्ष बीते हैं और डेटा संरक्षण से जुड़ी कुछ गतिविधियाँ जैसे पुराने संस्करण से नए एलटीओ में माइग्रेशन के कार्य को शुरु नहीं किया गया है। टीडीआर के लिए यह एक निरन्तर प्रक्रिया है। साथ ही मानकों को प्रत्येक दो-तीन वर्ष पर उन्नयन/सत्यापन करना आवश्यक है, जिससे फॉर्मेट और मीडिया माइग्रेशन संबंधित फ़ैसला लिया जा सके। साथ ही, टीडीआर का प्रमाणन भी सावधिक है, जिसका पुनर्माणन अनिवार्य है। अपनाई गई प्रक्रिया में डेटा रिकवरी के लिए पर्याप्त प्रावधान है। इसलिए डिजिटल डेटा के आपदा प्रबंधन के लिए आज टीडीआर विकसित करना सबसे अच्छा उपाय है। NCAA की तरह डिजिटल पाण्डुलिपियों के लिए एक टीडीआर बनाने का कार्य शुरु हो रहा है, जिनमें पाण्डुलिपियों के साथ-साथ अभिलेख और ताम्रपत्र भी होंगे। संभव हुआ तो दोनों टीडीआर एक दूसरे के लिए एक स्टैंडबाय डिजिटल रिपोजिटरी की तरह भी काम करेंगे। जिसमें एक का डेटा दूसरे टीडीआर में

नियमित रूप से (periodically) अपडेट होगा और सुरक्षित रहेगा। आपदा की स्थिति में वहाँ पर सुरक्षित डेटा का उपयोग किया जा सकेगा। विजुअल्स के लिए भी एक टीडीआर सोचा जा सकता है जिसमें सभी तरह के 2D और 3D चाहे वह चित्रकला, शिल्पकला, वास्तुकला, कलाकृति या किसी रूप में उसका छायाचित्र सम्मिलित हों। कार्य की विशालता को देखते हुए इसके लिए कई संस्थानों को एक साथ मिलकर कार्य करना होगा।

किसी भी डिजिटल रिपोजिटरी की संगठनात्मक और/या वित्तीय विफलता की संभावना हमेशा बनी रहती है। इसलिए एक स्टैंडबाय डिजिटल रिपोजिटरी का निर्माण जो विफलता के समय डेटा को एक विश्वसनीय उत्तराधिकारी के रूप में सुरक्षित रखे, जरूरी है। प्रतिदिन बढ़ते डिजिटल डेटा के साथ-साथ अनेक टीडीआर बनाने की आवश्यकता है। डिजिटल डेटा के संरक्षण के लिए एक राष्ट्रीय नीति की आवश्यकता है तथा नीति के प्रभावी कार्यान्वयन के लिए तदनु रूप एक अधिनियम की आवश्यकता है, जो निम्न कार्यों का निष्पादन कर सके—

- A. पूरे भारत में कई ट्रस्टेड डिजिटल रिपोजिटरी की स्थापना।
- B. सभी डिजिटल डेटा को किसी न किसी ट्रस्टेड डिजिटल रिपोजिटरी में संरक्षित करना।
- C. प्रत्येक ट्रस्टेड डिजिटल रिपोजिटरी के लिए एक या अधिक विश्वसनीय उत्तराधिकारी होना।
- D. हमारी सामूहिक स्मृति को डिजिटल रूप में संरक्षित करने के लिए सभी डेटा की सुरक्षा के लिए ट्रस्टेड डिजिटल रिपोजिटरी की एक ग्रिड बनाना।
- E. युवाओं, जो कि तकनीकी के सबसे बड़े उपयोगकर्ता हैं, को जोड़ने के लिए डिजिटल संरक्षण को शिक्षा की मुख्यधारा में लाने की आवश्यकता है।

**निष्कर्ष :**

ट्रस्टेड डिजिटल रिपोजिटरी किसी भी आपदा में डिजिटल डेटा के संरक्षण का ख्याल रखती है। साथ ही यह सुनिश्चित करती है कि डिजिटल डेटा का संरक्षण लंबे समय तक हो एवं किसी भी तरह से डेटा का दुरुपयोग न हो सके। ट्रस्टेड डिजिटल रिपोजिटरी सभी डेटा को ओपेन फॉर्मेट में संरक्षित करती है एवं प्रयुक्त साफ्टवेयर भी ओपेन प्लेटफॉर्म में बनी होती है। इसीलिए, डेटा तक पहुँच हमेशा बनी रहती है। ट्रस्टेड डिजिटल रिपोजिटरी संरक्षित डेटा पर आने वाले संकटों के प्रति हमेशा सचेत रहती है और तकनीकी रुझान और आने वाले संकटों के पूर्वानुमान से प्रेरित समयोचित कार्यवाही करती है। साथ ही, तकनीकी उन्नयन के साथ-साथ ट्रस्टेड डिजिटल रिपोजिटरी में संरक्षित डेटा का मीडिया, फॉर्मेट एवं साफ्टवेयर अपग्रेड नियमित रूप से करना अनिवार्य है। इसीलिए, ट्रस्टेड डिजिटल रिपोजिटरी का ISO 16363 प्रमाण-पत्र भी समयावधि के लिए होता है। तदुपरान्त ऑडिट की प्रक्रिया पुनः करानी होती है। NCAA के लिए भी प्रमाण-पत्र नवंबर 2020 तक है। डिजिटल संरक्षण में चुनौती केवल तकनीकी नहीं है, बल्कि संगठनात्मक भी है। साथ ही डेटा के संरक्षण के लिए लगातार प्रचुर धन की आवश्यकता होती है। अलग-अलग संस्थानों के द्वारा संस्थागत रूप से ट्रस्टेड डिजिटल रिपोजिटरी बनाना आसान नहीं है। लेकिन, जैसा कि NCAA के लिए संभव हो सका, कई संस्थान एक साथ मिलकर इस कार्य को कर सकते हैं। इससे डेटा का संरक्षण राष्ट्रीय स्तर पर हो सकेगा। साथ ही, किसी एक संस्था के उपर वित्तीय बोझ भी कम होगा। अतः ट्रस्टेड डिजिटल रिपोजिटरी की वित्तीय कारणों से विफलता की संभावना कम हो जायेगी। जबतक हम डिजिटल डेटा का संरक्षण सुनिश्चित नहीं करेंगे, हमारा आज, कल के इतिहास में सुरक्षित नहीं रह सकेगा।

**Terminologies:**

Disaster Management	आपदा प्रबंधन
Digital Preservation	डिजिटल संरक्षण
Digital Archive	डिजिटल आर्काइव्स / डिजिटल लेखागार
Trustworthy Digital Repository (TDR)	ट्रस्टवर्दी डिजिटल रिपोजिटरी (टीडीआर)
Trusted Digital Repository (TDR)	ट्रस्टेड डिजिटल रिपोजिटरी (टीडीआर)
National Cultural Audiovisual Archives (NCAA)	नेशनल कल्चरल ऑडियोविजुअल आर्काइव्स
Offline data centre	ऑफलाइन डेटा सेंटर
Access Quality	एक्सेस क्वालिटी
Archival Quality	अभिलेखीय / आर्काइवल क्वालिटी
Open format	ओपेन फॉर्मेट
Open Platform	ओपेन प्लेटफॉर्म
LPCM – Linear Pulse Code Modulation	एलपीसीएम
BWF	Broadcast Wave Format – बीडब्ल्यूएफ
MXF	Material Exchange Format –एमएक्सएफ
JPEG2000 – Motion JPEG2000 file format for video	–जेपीईजी2000
Open Archive Information System (OAIS)	ओपेन-आर्काइव इन्फॉर्मेशन सिस्टम
Reference Model	रेफरेंस मॉडल
Standards for Digitization	डिजिटलीकरण का मानक
International Association of Sound and Audiovisual Archives (IASA)	इंटरनेशनल एसोसिएशन ऑफ साउंड एंड ऑडियोविजुअल आर्काइव्स (आईएसए)
Metadata – Information about content	मेटाडेटा



Outdated	अप्रचलन
LTO- Linear Tape Open- A media for digital data storage.	एलटीओ
Upgrade	अपग्रेड / उन्नयन
Copyright	कॉपीराइट (सर्वाधिकार)

### References:

1. Goodman, Paul, 16 Advantages of Digital Technology,
2. <https://turbofuture.com/computers/Advantages-of-Digital-Technology>
3. Sinha, Dr. Mukul, Digital Repository: Audit

& Certification - (Position Paper Template)  
Study Report on Requirements of Digital Preservation in India

4. <https://en.unesco.org/programme/mow/recommendation-documentary-heritage>
5. Lavoie, Briam 2014, The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)
6. Gladney, Henry M. 2007, Preserving Digital Information, Springer, Heidelberg
7. <http://ncaa.gov.in/repository/>
8. <http://www.iso16363.org>
9. [https://www.cdac.in/index.aspx?id=mc\\_hc\\_national\\_digi\\_pre](https://www.cdac.in/index.aspx?id=mc_hc_national_digi_pre)