

संरक्षणीय वाङ्मय के सांख्यिकीकरण  
की मानदण्ड-निर्देशिका

## Guidelines for Digitization of Archival Material



National Mission for Manuscripts

### **Technical Committee**

#### **Dr. Gautam Bose (Chairperson)**

Deputy Director General  
National Informatics Centre

#### **Members**

##### **Mr. Pratik Shrivastava**

Sr. Technical Director  
National Informatics Centre

##### **Mr. P. Jha**

Director, CIL  
IGNCA

##### **Mr. N.S. Mani**

Microphotographer  
National Archives of India

##### **Mr. Biswaranjan Mallick**

Coordinator, Digitization  
National Mission for Manuscripts

### **Editor**

#### **Prof. Dipti S. Tripathi**

Director, National Mission for Manuscripts  
New Delhi

### **Publisher**

Director, National Mission for Manuscripts  
11 Mansingh Road  
New Delhi – 110 001  
Tel: +91 11 23383894  
Fax: +91 11 23073340  
Email: [director.namami@nic.in](mailto:director.namami@nic.in)  
Website: [www.namami.org](http://www.namami.org)

**Design:** Macro Graphics Pvt. Ltd.  
([www.macrographics.com](http://www.macrographics.com))



# संरक्षणीय वाङ्मय के सांख्यिकीकरण की मानदण्ड-निर्देशिका

## १. परिचय

वर्ष २००३ में स्थापना के बाद से राष्ट्रीय पाण्डुलिपि मिशन ने बड़े पैमाने पर भारत की पाण्डुलिपि-सम्पदा के संरक्षण तथा उन पाण्डुलिपियों तक पहुँच प्रदान करने का काम शुरू किया है।

यह निर्देशपत्र पाण्डुलिपियों, मुद्रित पुस्तकों, मानचित्रों, चित्रों, स्लाइड, नेगेटिव, माइक्रोफिल्म आदि सामग्रियों की सांख्यिकीय (डिजिटल) छवियों में अभिलेखीय (आर्काइवल) गुणवत्ता उत्पन्न करने के लिए दिशानिर्देश निर्धारित करता है। ये दिशानिर्देश उन सभी संगठनों के लिए हैं जो इन सामग्रियों के सांख्यिकीकरण की योजना बना रहे हैं।

छवि की गुणवत्ता, स्कैनिंग की पद्धति और उन तक पहुँच की एक रूपता को बनाये रखने के लिए ये दिशानिर्देश निर्धारित किये गये हैं। दीर्घकाल तक इन छवियों का उपयोग करने के लिए और पुनः उस सामग्री को स्कैन करने की आवश्यकता को कम करने के लिए इन दिशानिर्देशों का पालन करने की सलाह दी जाती है।

दृश्य एवं श्रव्य रिकॉर्डिंग के लिए सांख्यिकीकरण मानक, मूल सांख्यिकीय सामग्री, सांख्यिकीकरण के लिए सामग्री तैयार करने के अन्तर्गत संरक्षण की प्रक्रिया और पाण्डुलिपियों को पढ़ने के लिए विद्वानों का आवश्यक योगदान, जैसे फोलियो नं. आदि इन दिशानिर्देशों के दायरे से बाहर हैं।

## २. राष्ट्रीय पाण्डुलिपि मिशन की पृष्ठभूमि

भारत के पास विश्व का सबसे बड़ा पाण्डुलिपि-संग्रह है। वे देशभर में विभिन्न पुस्तकालयों, शैक्षणिक संस्थानों, संग्रहालयों, मन्दिरों, मठों और निजी संग्रहों में फैली हुई हैं और विदेशों में भी अनेकत्र पुस्तकालयों में रखी हैं। भारत की समृद्ध पाण्डुलिपि-सम्पदा आज विलुप्त होने के खतरे का सामना कर रही है। पाण्डुलिपियों के रूप में विद्यमान भारत की इस अमूल्य विरासत को सूचीबद्ध रूप में संरक्षित करने और उन्हें आज तथा आनेवाली पीढ़ियों के लिए सुलभ बनाने की आवश्यकता है जिससे उनमें निहित ज्ञान का प्रसार हो सके।

सांख्यिकीय प्रौद्योगिकी का उपयोग करके भावी पीढ़ी के लिए पाण्डुलिपियों का संरक्षण राष्ट्रीय पाण्डुलिपि मिशन का प्राथमिक लक्ष्य है। राष्ट्रीय पाण्डुलिपि मिशन ने राष्ट्रीय तथा अन्तर्राष्ट्रीय स्तर पर चल रही कई सांख्यिकीय परियोजनाओं के द्वारा अपनायी गयी पद्धतियों का अध्ययन करने के उपरान्त इस क्षेत्र के विशेषज्ञों के साथ लम्बा परामर्श कर सांख्यिकीकरण के लिए ये दिशानिर्देश प्रस्तुत किये हैं।



### ३. सांख्यिकीकरण

सांख्यिकीय प्रौद्योगिकी ने डाटा भंडारण के क्षेत्र में एक नयी दृष्टि प्रदान की है। वर्ल्ड वाइड वेब में लाखों वेबसाइट्स हैं और इंटरनेट अब एकीकृत एवं सार्वभौमिक स्तर पर अनुसंधान, शिक्षण, अभिव्यक्ति, प्रकाशन और सूचना के संचार के साथ जुड़ा हुआ है। जहाँ तक डाटा के सूचीकरण और प्रसंस्करण का सम्बन्ध है, पुस्तकालय तथा अभिलेखागार समाज के, सूचना प्रदान करनेवाले, प्राथमिक केन्द्र हैं। बड़ी संख्या में पुस्तकालयों में आज मूल सांख्यिकीय सामग्री को संरक्षित करने तथा उस तक पहुँच प्रदान करने के अलावा उनके वर्तमान संसाधनों से सांख्यिकीय प्रतिनियुक्त (सरोगेट) बनने लगे हैं। सांख्यिकीकरण का अर्थ है – सांख्यिकीय स्वरूप में किसी जानकारी को प्राप्त करना, परिवर्तित करना, भंडारण करना और माँग होने पर उसे मानक और संगठित रूप में उपलब्ध कराना।

#### ३.१ लक्षित उपयोगकर्ता

इन दिशानिर्देशों का निर्धारण-निर्णायक अधिकारियों, पुरालेखपालों, पुस्तकालय-प्रबन्धकों, क्यूरेटों और तकनीकी कार्यकर्ताओं के उद्देश्य से किया गया है।

### ४. सांख्यिकीकरण क्यों?

सांख्यिकीकरण परियोजना को लागू करने अर्थात् गैर-सांख्यिकीय स्रोत सामग्री को सांख्यिकीय बनाने के अनेक कारण हैं।

#### सुविधापूर्वक उपलब्धि

अनुसन्धान प्रयोजनों के लिए यह सबसे स्पष्ट और प्राथमिक कारण है। जहाँ, उपयोगकर्ताओं को आवश्यकता हो तथा पुस्तकालयों एवं अन्य संस्थाओं को सामग्री की उपलब्धि सुविधाजनक बनाने की इच्छा हो, वहाँ सांख्यिकीकरण पहला विकल्प होता है। इसके द्वारा किसी संस्था के संसाधनों को अधिक से अधिक लोगों तक पहुँचाकर शिक्षा एवं आजीवन शिक्षण के मार्ग को प्रशस्त किया जा सकता है।

#### पहुँच तथा उपयोग के नये मार्ग प्रशस्त करने के लिए

जो मूल सामग्री संग्रहकर्ता तक बिना गये उपलब्ध नहीं हो सकती सांख्यिकीकरण के द्वारा उसे उपलब्ध कराना महत्वपूर्ण लक्ष्य है। साथ ही अत्यधिक जर्जर और क्षतिग्रस्त सामग्री की भी आकृति और विषयवस्तु इस माध्यम से उपलब्ध करायी जा सकती है।

#### संरक्षण हेतु

लम्बे समय तक चलने वाले माध्यम में मूल सामग्री की सटीक प्रतिकृतियाँ बनाकर सांख्यिकीकरण के माध्यम से उन्हें संरक्षित किया जा सकता है। ये प्रतिकृतियाँ वर्तमान तथा भविष्य के उपयोगकर्ताओं के लिए सन्तोषजनक होनी चाहिए। अतः ये उच्च गुणवत्ता सम्पन्न तो होनी ही चाहिए इनका भौतिक स्वरूप भी ऐसा होना चाहिए जो कालक्रम में स्थिर बना रहे।



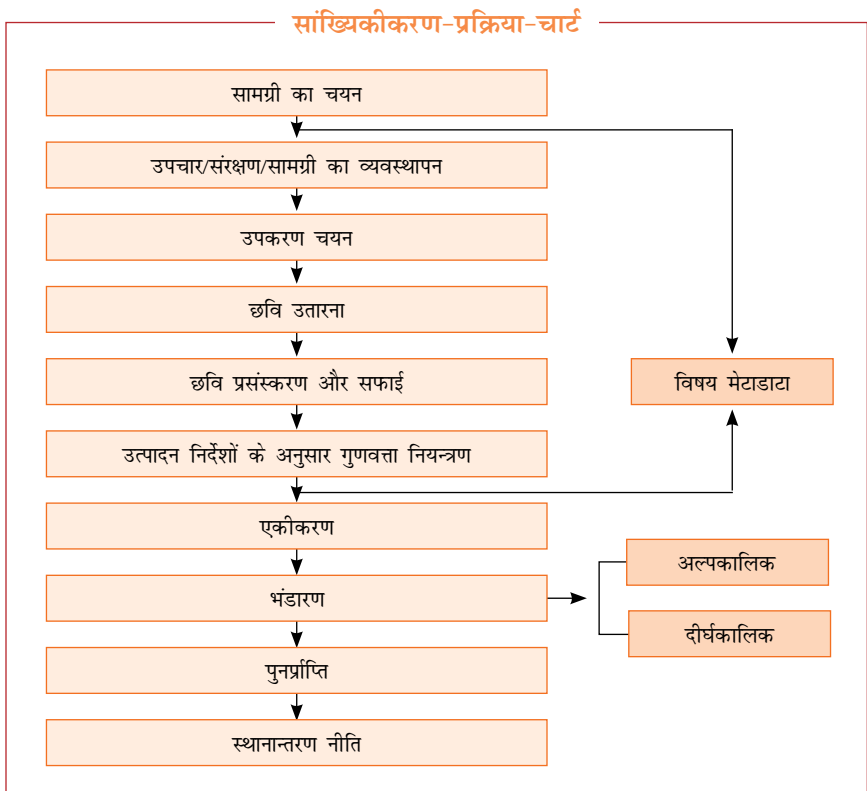
### सामग्री को अनेकशः स्पर्श से बचाने के लिए

अत्यधिक उपयोग में लायी गयी मूल सामग्री अथवा जर्जर सामग्री को बार बार स्पर्श से बचाने के लिए, साथ ही इस प्रकार की विनष्टप्राय सामग्री की सुरक्षा-प्रति बनाने के लिए भी सांख्यिकीकरण का प्रयोग होता है।

## ५. सांख्यिकीकरण की प्रक्रिया

सांख्यिकीकरण प्रक्रिया में स्रोत सामग्री को सांख्यिकीय स्वरूप प्रदान कर उसे सांख्यिकीय माध्यम में सुरक्षित किया जाता है। इस सामग्री तक पहुँचने के लिए पुनरुपलब्धि सॉफ्टवेयर का प्रयोग होता है।

निम्नलिखित चार्ट में सांख्यिकीकरण के विभिन्न चरणों को दिखलाया गया है।





## ५.१ सामग्री का चयन

सांख्यिकीकरण एक ऐसी चयन प्रक्रिया है जिसमें आवश्यकताओं और माँग के बीच में सन्तुलन स्थापित करने की आवश्यकता होती है। इसके लिए स्रोत-सामग्री का चयन करते समय तीन मूल प्रश्न उठते हैं:

१. स्रोत-सामग्री के रूपान्तरण की आवश्यकता है या नहीं
२. स्रोत-सामग्री के रूपान्तरण का औचित्य क्या है, तथा
३. स्रोत-सामग्री के रूपान्तरण की कितनी सम्भाव्यता है

अतः सामग्री का चयन इस प्रकार किया जाना चाहिए जिसमें चयनित सामग्री के महत्त्व, उसकी विषयवस्तु में रुचि पर विचार करने के साथ उसकी माँग, तकनीकी सम्भावना और संस्थागत आवश्यकता पर भी विचार किया गया हो।

### अवस्था

सांख्यिकीकरण के लिए सामग्री का चयन उसकी भौतिक अवस्था और पठनीयता से प्रभावित होता है। जो सामग्रियाँ जर्जर, क्षतिग्रस्त तथा दुरवस्था में हों उनका सांख्यिकीकरण से पूर्व आवश्यक उपचार करने के साथ-साथ उन्हें हाथ से उठाते समय विशेष ध्यान देना चाहिए ताकि सामग्री और अधिक क्षतिग्रस्त न हो। सांख्यिकीकरण के लिए चयनित सामग्री का यदि सविस्तर सूचीकरण नहीं हुआ हो तो पहले उसकी साधारण सूची एवं सविस्तर सूची बनायी जानी चाहिए। इसका खर्च सांख्यिकीकरण परियोजना के बजट में ही अन्तर्भुक्त किया जाना होगा।

### माँग

सांख्यिकीकरण के लिए सामग्री चयन के समय तत्सम्बन्धी माँग का विशेष महत्त्व है। सांख्यिकीकरण के लिए सामग्री के चयन तथा चयनित सामग्री की प्राथमिकता निर्धारित करने के लिए विद्वानों अभिलेखपालों, पुस्तकालयाध्यक्षों, संग्रहालयाध्यक्षों तथा शोधकर्ताओं का सहयोग लिया जाना चाहिए।

### विषयवस्तु

सांख्यिकीकरण परियोजना का उद्देश्य चाहे जो हो सामग्री-चयन विषयवस्तु के आधार पर ही निर्णीत होगा।

## ५.२ उपचार/संरक्षण/सामग्री का व्यवस्थापन

सांख्यिकीकरण के लिए चयनित सामग्री में निम्नलिखित प्रक्रियाओं को पूरा किया जाना चाहिए।



1. आवश्यकतानुसार संरक्षण
2. आवश्यकतानुसार क्रम-निर्धारण
3. मानक संरक्षण तकनीक के द्वारा सामग्री की आवश्यकतानुसार सफाई जिससे वह सुपाठ्य हो सके।

मूल सामग्री विशेषतः पाण्डुलिपियों को क्षति से बचाने के लिए निम्न सावधानियाँ अपेक्षित हैं:

- क. विशेषज्ञों के द्वारा निर्धारित दिशानिर्देशानुसार सामग्री का व्यवस्थापन होना चाहिए। अतः सांख्यिकीकरण प्रक्रिया के दौरान सामग्री को स्कैनिंग या फिलिमिंग पटल पर रखने का काम स्वयं पाण्डुलिपिधारक को, विशेषज्ञ को अथवा विशेषज्ञ की कड़ी देख रेख में किसी अन्य को करना चाहिए।
- ख. सामग्री की बैंथाई तोड़ना अनुमत नहीं है क्योंकि उससे मूल प्रति क्षतिग्रस्त हो सकती है। यदि बैंथाई तोड़ना अनिवार्य हो तो इस बात का ध्यान रखा जाना चाहिए कि निर्धारित प्रक्रिया के अनुसार ही वह खोली जाय और कार्य समाप्त होने पर फिर बहाल कर दी जाय।
- ग. स्कैनर या कैमरा संचालक शल्यक्रिया के दस्ताने का उपयोग करें ताकि सामग्री को कोई क्षति न पहुँचे।
- घ. कोमल तूलिकाओं से आवश्यकतानुसार धूल और गंदगी साफ की जानी चाहिए।
- ङ. लम्बी तथा क्षितिजीय सामग्री विशेष व्यवस्थापन की अपेक्षा रखती है।
- च. खुले पृष्ठों वाली सामग्री (प्रायः पाण्डुलिपियों के पृष्ठ एक दूसरे से अलग होते हैं) के क्रम को सही रखने के लिए तथा अनुपलब्ध पृष्ठों की सूचना निबद्ध करने के लिए स्थानीय विशेषज्ञों या विद्वानों की सहायता ली जा सकती है। कार्य पूरा होने पर ये ही इन्हें पुनः संग्रहित करें। सामग्री को लापरवाही से व्यवस्थापित करने, विशेष रूप से पुस्तकों की बैंथाई पर जोर डालने से बचना चाहिए। सही और बारीक चित्रांकन (जैसे मूल प्रति में छपाई या लेख की बारीकियों को चित्रांकित करने) की अपेक्षा मूल को क्षतिग्रस्त होने से बचाने पर अधिक बल होना चाहिए।

### ५.३ उपकरण का चयन

उपकरण का चयन छवि की गुणवत्ता को प्रभावित करता है। समान तकनीकी क्षमताओं के बावजूद विभिन्न निर्माताओं के उपकरणों के परिणाम अलग अलग हो सकते हैं। छवि लेने के लिए सम्पर्की उपकरण की बजाय असम्पर्की तथा ठंडी रोशनी वाले उपकरण का प्रयोग किया जाना चाहिए। स्पर्श करके छवि लेने वाले उपकरण मूल प्रति को क्षति पहुँचा सकते हैं, अतः उनका उपयोग नहीं होना चाहिए। छवि लेने के सभी उपकरण तथा प्रकाश व्यवस्था ऐसी होनी



चाहिए जिससे मूल सामग्री को किसी भी प्रकार की क्षति न पहुँचे। इसके अलावा रोशनी का उपयोग भी इस प्रकार किया जाना चाहिए जिससे मूल सामग्री क्षतिग्रस्त न हो।

## ५.४ छवि उतारना

### स्रोत सामग्री की विशेषताएँ

प्रायः स्रोत सामग्री का वर्गीकरण उसके आकार, अवस्था, सचित्र अथवा चित्ररहित आदि आधारों पर किया जाता है। स्रोत सामग्री देश के विविध संग्रहालयों में अलग अलग स्थानों पर उपलब्ध है उनका सांख्यिकीकरण उसी स्थान पर किया जाना चाहिए।

स्रोत सामग्री मुख्यतः विविध प्रकार के कागज, तालपत्र, भोजपत्र, बाँस के पत्ते, कपड़े, मिट्टी के फलक, चर्मपत्र, ताम्रपत्र, लकड़ी के फलक, लकड़ी के मनके, हाथीदाँत के फलक, मृगचर्म आदि पर उपलब्ध होती है। गौणरूप से यह माइक्रोफिल्म, फोटोग्राफिक प्लेट एवं नेगेटिव स्लाइड आदि पर भी उपलब्ध हो सकती है। यह समस्त सामग्री प्रायः बहुत पुरानी और जर्जर होती है अतः उनका उपयोग करते समय अत्यन्त सावधान रहने की अपेक्षा है।

कुछ स्रोत सामग्रियों में अंकित चित्र/नक्शे आदि प्राचीन स्याही, सब्जियों के प्राकृतिक रंग एवं सोने तथा चाँदी जैसे धातु से तैयार किये गये होते हैं। तेज रोशनी और ताप के सम्पर्क से इन पदार्थों के ऑक्सीकरण की सम्भावना रहती है।

चित्र सामग्री जैसे कि पाण्डुलिपि में दिये हुए नक्शे आदि अलग से स्कैन किये जाने चाहिए और उसके बाद समुचित स्थल पर मूल में जोड़े जाने चाहिए। धातु से बने हुए चित्रों पर ऑक्सीकरण का प्रभाव यथाशक्य कम हो इसके लिए सचेष्ट रहने की आवश्यकता है।

**उतारी गयी छवि** की गुणवत्ता, स्कैनिंग विभेदन (रिजोल्यूशन), स्कैन की गयी छवि में बिट की गहराई, छवि को बढ़ाने की प्रक्रिया एवं प्रयुक्त सम्पीडन, स्कैनिंग के उपकरण तथा स्कैनिंग करने वाले की कुशलता पर निर्भर करती है।

**विभेदन** का निर्धारण पिक्सेल की संख्या पर आधारित है जो प्रति इंच बिन्दुओं अथवा प्रति इंच पिक्सेल के रूप में गिने जाते हैं। पिक्सेल की संख्या बढ़ाने से छवि ऊँचे विभेदन की होगी और उसमें सूक्ष्म विशेषताएँ उभर कर सामने आयेंगी। लेकिन केवल विभेदन बढ़ाने से ही छवि में वैशिष्ट्य नहीं आयेगा क्योंकि एक सीमा के बाद पिक्सेल बढ़ाने पर केवल फाइल का आकार बढ़ा हो जाता है, छवि के स्वरूप में कोई अन्तर नहीं आता।

पिक्सेल का लक्षण स्पष्ट करने के लिए बिट की गहराई मानक है। जितनी अधिक बिट की गहराई होगी उतने ही भूरे रंग की विविधता एवं अन्य रंगों की रंगत अभिव्यक्त हो सकेगी।

- ❖ **बाइटोनल स्कैनिंग** में सफेद एवं काले रंग को व्यक्त करने के लिए प्रति पिक्सेल १ बिट का उपयोग किया जाता है।





- ❖ **ग्रेस्केल स्कैनिंग** में भूरे रंग के विभिन्न स्तरों को अभिव्यक्त करने के लिए प्रति पिक्सेल एकाधिक बिट का प्रयोग हो सकता है। प्रायः भूरे रंग में ८ बिट प्रति पिक्सेल का उपयोग होता और इस स्तर पर कोई भी आकृति भूरे रंग के २५६ स्तरों में से चयन कर सकती है।
- ❖ **रंगीन स्कैनिंग** में रंग का प्रतिनिधित्व करने के लिए प्रति पिक्सेल एकाधिक बिट का उपयोग किया जाता है। २४ बिट प्रति पिक्सेल, रंग के वास्तविक स्तर को, अभिव्यक्त करता है और इसके द्वारा १६.७ लाख रंगों में से एक रंग का चयन करना सम्भव होता है।

### सांख्यिकीय रूपान्तरण

चयनित सामग्री स्कैनर या सांख्यिकीय कैमरे के द्वारा बिट-मैप छवियों के रूप में बदली जाती है। छवि उतारने के दौरान दिये गये मानदण्ड (टेबल ए) के अनुसार समुचित विभेदन एवं बिट गहराई में दस्तावेज़ को स्कैन किया जाना चाहिए, जिससे उत्पाद (रॉ मास्टर इमेज) उपलब्ध हो सके। इससे उपलब्ध प्रत्येक पिक्सेल के बिट वाली सांख्यिकीय फाइल का स्वरूपण एवं अंकन किया जाता है, जिससे संगणक में उसको रखने तथा पुनः उपलब्धि में सुविधा हो। इससे उपलब्ध सांख्यिकीय आकृति टिफ फॉरमैट (टेबल ए) में रॉ मास्टर इमेज शीर्षक से सुरक्षित होती है। इन फाइलों से संगणक ऐनालॉग प्रतिकृति उत्पन्न कर सकता है जो स्क्रीन पर दिखायी अथवा मुद्रित की जा सके। उच्च विभेदन वाली फाइलें अत्यधिक विशाल होती हैं इसलिए उनका आकार छोटा करने के लिए सम्पीडन की आवश्यकता हो सकती है ताकि संगणक एवं उपयोगकर्ता उसे सुविधापूर्वक काम में ला सकें।

जब स्रोत सामग्री की स्कैनिंग हो जाय तो समस्त डाटा, विशिष्ट फाइल आकार में, परिवर्तित कर सुरक्षित रखा जाता है। (टेबल ए)।

### ५.५ सांख्यिकीय छवि का प्रसंस्करण और सफाई

किसी एक स्रोत प्रलेख में अथवा प्रलेखों के समूह में उपलब्ध सूचना को सांख्यिकीय आकृति में उपलब्ध कराना, किसी भी सांख्यिकीकरण कार्यक्रम का लक्ष्य होना चाहिए। महत्वपूर्ण अंशों का अंकन करते समय सांख्यिकीय आकृतियों की गुणवत्ता का निर्धारण सांख्यिकीय आकृति और मूल आकृति की तुलना के आधार पर किया जाना चाहिए। यह ऐच्छिक मानदण्ड पर आधारित नहीं होना चाहिए।

मूल रॉ आकृति मास्टर आकृति के मानदण्डों (टेबल ए) के अनुसार सुरक्षित की जानी चाहिए। रॉ आकृति से धूल, कीटचिह्न, जलचिह्न, शोर, छाया, खरोंच के चिह्न, टेपेन आदि को



हटाने के लिए प्रसंस्करण आवश्यक है। प्रसंस्करण का दूसरा पक्ष है – चमक और विषमता का संयोजन, गामा प्रसंस्कार, तीक्ष्णता और धुंधलेपन का शोधन, आकृतियों को हटाना और रंगों का समायोजन। साफ की हुई छवि संस्कारित मूल छवि (टेबल ए) के रूप में सुरक्षित की जानी चाहिए।

### सम्पीडन

प्रसंस्करण के लिए, फाइल का आकार छोटा करने के लिए, भंडारण के लिए और सांख्यिकीय आकृति के प्रसारण के लिए प्रायः सम्पीडन का प्रयोग किया जाता है। इसके लिए बार-बार आनेवाली सूचना को अथवा उन सूचनाओं को, जिन्हें देखना मानवीय नेत्रों के लिए कठिन है, संक्षिप्त किया जाता है। सम्पीडन की तकनीक एवं सम्पीडन के स्तर से आकृति की गुणवत्ता प्रभावित हो सकती है। सम्पीडन की तकनीक दो प्रकार की होती है १. असम्पीडित आकृति सम्पीडन की प्रक्रिया के बाद भी अपनी मूल आकृति के समान होती है क्योंकि फाइल के आकार को कम करते समय उसमें कोई भी सूचना हटायी नहीं जाती। २. जहाँ अनपेक्षित सूचनाओं को सम्पीडन के समय निकाल दिया जाता है वहाँ असम्पीडित? प्रायः प्रथम प्रकार का सम्पीडन मास्टर फाइल के लिए और दूसरे प्रकार का सम्पीडन पहुँच (एक्सेस) फाइल के लिए किया जाता है।

पहुँच आकृति स्वच्छ की गयी आकृति से प्राप्त की जानी चाहिए।

### ५.६ उत्पादन विनिर्देश

१. असंस्कृत मास्टर छवि (मूल बिना साफ की हुई तथा असम्पीडित)
२. स्वच्छ मास्टर छवि (साफ तथा हानिरहित सम्पीडित छवि)
३. पहुँच छवि/पीडीएफ-ए (व्युत्पन्न हानियुक्त छवि)





इन छवियों का सविस्तर विनिर्देश निम्नलिखित है:

### तालिका - ए

सामग्री	असंस्कृत मास्टर छवि	बिट गहराई	स्वच्छ मास्टर छवि	पहुँच छवि
पाण्डुलिपियाँ	टिफ (असम्पीडित) ३०० डीपीआइ (उत्पादन)	२४ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी स्थानिक विभेदन: ८" X १०" ३०० डीपीआइ	जेपेग/ पीडीएफ-ए स्थानिक विभेदन: १०२४ X ७६८ पिक्सेल
मुद्रित पुस्तकें	टिफ (असम्पीडित) ३०० डीपीआइ (उत्पादन)	२४ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी	जेपेग/ पीडीएफ-ए
श्वेत/श्याम		८ बिट		
तस्वीरें	टिफ (असम्पीडित) ३००/६०० डीपीआइ (उत्पादन) विशेष स्थिति में १२००	२४ बिट ८ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी	जेपेग/ पीडीएफ-ए
श्वेत/श्याम	डीपीआइ और उससे अधिक			
स्लाइड्स/ नेगेटिव	४००० डीपीआइ टिफ (असम्पीडित)	२४ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी	जेपेग/ पीडीएफ-ए
श्वेत/श्याम		८ बिट		
माइक्रोफिल्म/ माइक्रोफिश	३०० डीपीआइ टिफ	८ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी	जेपेग/ पीडीएफ-ए
मानचित्र/ रेखाचित्र	टिफ (असम्पीडित) ३००/६०० डीपीआइ (उत्पादन) विशेष संदर्भ में १२०० डीपीआइ और उससे अधिक	२४ बिट	टिफ (सम्पीडित) ३०० डीपीआइ (उत्पादन) समूह ६ सीसीआईटीटी	जेपेग/ पीडीएफ-ए
श्वेत/श्याम		८ बिट		



## ५.७ मेटाडाटा

### पाण्डुलिपियों का विषय-मेटाडाटा

- |                       |                                       |
|-----------------------|---------------------------------------|
| १. सामग्री संख्या     | १३. गुच्छ संख्या                      |
| २. ग्रन्थ नाम         | १४. पत्र संख्या                       |
| ३. ग्रन्थ के अन्य नाम | १५. पृष्ठ                             |
| ४. लेखक               | १६. सामग्री                           |
| ५. संस्था             | १७. लुप्त अंश                         |
| ६. टीका               | १८. चित्र                             |
| ७. टीकाकार            | १९. अवस्था                            |
| ८. लिपिकार            | २०. सूचीपत्र का आधार                  |
| ९. भाषा               | २१. टिप्पणी                           |
| १०. लिपि              | २२. पाण्डुलिपि की तिथि                |
| ११. पूर्ण/अपूर्ण      | २३. पाण्डुलिपि की लम्बाई (इन्चों में) |
| १२. विषय              | २४. पाण्डुलिपि की चौड़ाई (इन्चों में) |

ऊपर निर्दिष्ट विषय-मेटाडाटा एनआइसी द्वारा विकसित मानुसग्रन्थावली सॉफ्टवेयर पर आधारित है। राष्ट्रीय पाण्डुलिपि मिशन पाण्डुलिपियों का रिकार्ड बनाने के लिए इसका उपयोग करता है। अन्य संस्थाओं में यह सूची इससे भिन्न भी हो सकती है।

### तकनीकी मेटाडाटा

तकनीकी मेटाडाटा सांख्यिकीय फाइल की विशेषताओं का विवरण देता है। यह, आकृति की फाइल बनाते समय, स्वयं उत्पन्न होने का सामर्थ्य रखता है और उसी समय फाइल के साथ संलग्न हो जाता है।

फाइल का नाम	:	स्कैनिंग के समय निर्धारित
बनने की तिथि	:	
संशोधन की तिथि	:	
व्यवहृत उपकरण	:	
छवि का स्वरूपण	:	
चौड़ाई	:	पिक्सेल्स
ऊँचाई	:	पिक्सेल्स



रंग की विधा	:	आरजीबी, ग्रेस्केल
विभेदन	:	इंच प्रति पिक्सेल
फाइल का आकार	:	
रंग का प्रोफाइल	:	आइसीसी रंग प्रोफाइल
निर्माता	:	स्केनर का सॉफ्टवेयर
X विभेदन	:	X के अक्ष में स्कैनिंग विभेदन
Y विभेदन	:	Y के अक्ष में स्कैनिंग विभेदन
विभेदन की इकाइयाँ	:	
सॉफ्टवेयर	:	इमेजिंग सॉफ्टवेयर

छवि-सम्पादन के उपकरणों से तकनीकी मेटाडाटा को देखा जा सकता है।

## ५.८ गुणवत्ता-नियन्त्रण

किसी भी सांख्यिकीय आकृति-निर्माण-परियोजना का महत्वपूर्ण घटक गुणवत्ता का नियन्त्रण है। इसके बिना सांख्यिकीय छवि की गुणवत्ता को सुनिश्चित नहीं जा सकता।

### पद्धति

आज सांख्यिकीय आकृति के मूल्यांकन के लिए उपलब्ध स्वचालित उपकरण बहुत मँहगे हैं और सम्भवतः असानी से उपलब्ध भी नहीं होते, इसलिए यह सलाह दी जाती है कि निम्नलिखित बिंदुओं पर दृश्य गुणवत्ता का मूल्यांकन किया जाय। सांख्यिकीय आकृतिदर्शन स्क्रीन पर अथवा नमूने के तौर पर मुद्रित रूप में भी किया जा सकता है।

### मूल्यांकन के लिए प्रस्तावित विधि

#### स्क्रीन पर

- ❖ स्कैन्ड छवियों को १ : १ (शतप्रतिशत वर्धित) आकार में देखें।
- ❖ भूरे का स्केल और रंग के प्रजनन का मूल्यांकन कैलिबरेटेड चार्ट के प्रयोग से करें।
- ❖ स्थान सम्बन्धी विभेदन तथा टोनल प्रजनन के मूल्यांकन के लिए विभेदन लक्ष्य (रिजोल्यूशन टारगेट) और हिस्टोग्राम का प्रयोग करें।
- ❖ छवि के गुणवत्ता नियन्त्रण की परीक्षा-सूची (अनुबन्ध ए)

#### मुद्रित प्रति का मूल्यांकन (नमूने के आधार पर)

- ❖ मूल प्रति के साथ मुद्रित प्रति की तुलना करें।



## छवियों का नामकरण

छवियों के नामकरण के लिए दो दृष्टिकोण हैं १. पहले से विद्यमान सूचीकरण पद्धति में दिये गये अंक अथवा २. सार्थक नाम। दोनों ही पद्धतियाँ ग्राह्य हैं। पद्धति का चयन औचित्य और सुविधा के आधार पर किया जाना चाहिए।

राष्ट्रीय पाण्डुलिपि मिशन सूचीकरण के लिए उपयोग में लाये जाने वाले सॉफ्टवेयर मानुस-ग्रन्थावली में स्वतः उत्पन्न होने वाली आइडी के अनुरूप पाण्डुलिपि की छवि का नाम प्रदान करता है। प्रत्येक स्कैन्ड पाण्डुलिपि का सूचना-सम्बन्धी मेटाडाटा डाटाबेस में सुरक्षित किया जाता है और उसकी पहचान उसके मानुस आइडी से होती है। इस प्रकार मानुस आइडी तथा जहाँ सांख्यिकीकरण हो रहा है, उस संग्रहकर्ता संस्था के द्वारा दी गयी पहुँच (एक्सेशन) संख्या, सांख्यिकीकृत आकृतियों को पहचानने का आधार बनती हैं।

**फ़ोल्डर नामकरण:** जिस फ़ोल्डर में स्कैन की गयी पाण्डुलिपियाँ संगृहीत हों उसका नाम, वह मानुस आइडी होगा, जो राष्ट्रीय पाण्डुलिपि मिशन के द्वारा पहले से दिया जायेगा। तदनन्तर शीर्षक, पहुँच संख्या और छवि के प्रकार (क्लीन टिफ, रॉ टिफ, पीडीएफ) सूचित किये जाएँगे।

**उदाहरण के लिए:** IGO 001333948 Krushnalila Ext\_96 Clean Tiff.

यदि पहुँच संख्या उपलब्ध नहीं है तो उसके बिना भी फोल्डर का नाम हो सकता है।

**उदाहरण के लिए:** IGO 001333948 Krushnalila Clean Tiff.

यदि पाण्डुलिपि का नाम डायक्रिटिक या देवनागरी लिपि में लिखा गया हो तो मेटाडाटा में भी उसी स्वरूप में लिखा जायेगा।

**उदाहरण के लिए:** BOO0000740792 Rāmāyaṇa S17-58-571 PDF.

इसी प्रकार उन सभी संस्थाओं को, जो बड़े पैमाने पर सांख्यिकीकरण कर रही हों, छवियों के नामकरण की प्रणाली सुनिश्चित करनी होगी। यहाँ पर ध्यान देने योग्य बात यह है कि नामकरण प्रणाली ऐसी हो जिसे समझना प्रयोगकर्ता के लिए सरल हो और डाटा तक पहुँचना भी आसान हो।

## छवियों का विवरण

सांख्यिकीय छवियों के वर्णन के लिए मेटाडाटा की आवश्यकता होती है इससे हमारा तात्पर्य उस व्यवस्थित डाटा से है जो अन्य डाटा को समाहित करता है। मेटाडाटा का दूसरा लक्षण उस डाटा के रूप में भी दिया जा सकता है जो दूसरे डाटा के प्रबन्धन को सुविधाजनक बनाता है। राष्ट्रीय पाण्डुलिपि मिशन पाण्डुलिपियों की सूचना मानुस-ग्रन्थावली सॉफ्टवेयर के अन्तर्गत सुरक्षित करता है।



इसी प्रकार अन्य सांख्यिकीय सामग्रियों का मेटाडाटा भी संग्रहकर्ता संस्थाओं के द्वारा मानकीकृत होना चाहिए। यहाँ हम डबलीन कोर मानक की संस्तुति करते हैं।

सर्वोत्तम गुणवत्ता वाली फाइल को सांख्यिकीय मास्टर कहते हैं। ये फाइलें आकृति को सीधे उतार कर तैयार की जाती हैं। मास्टर फाइल मूल आकृतियों को यथाशक्य सही और निकट रूप में प्रतिफलित करती है। अन्य छवियाँ मास्टर छवि से बनायी जाती हैं जिसमें छवि सम्पादन के सॉफ्टवेयर का प्रयोग होता है।

### ५.९ सांख्यिकीय छवियों का एकीकरण, भंडारण एवं प्रबन्धन

सांख्यिकीय छवियों का लक्षित उपयोग केवल रूपान्तर-मानक और गुणवत्ता नियन्त्रण पर ही निर्भर नहीं करता अपितु इस पर भी निर्भर करता है कि उन छवियों का भंडारण और प्रबन्धन किस प्रकार हुआ है। यदि अल्पकालिक उपयोग के साथ दीर्घकालिक उपयोग भी लक्ष्य है तो वर्तमान की आवश्यकताओं के साथ भविष्य की आशाओं को ध्यान में रखकर समुचित कदम उठाने होंगे।

#### इसके लिए निम्न प्रकार से योजना बनानी होगी

- ❖ लक्षित उपयोग के लिए स्कैनड छवियों को उपलब्ध कराना।
- ❖ समय-समय पर उपयोगकर्ता इंटरफेस की उपयोगिता को संवर्धित करना।
- ❖ सांख्यिकीय सूचना के प्रसंस्करण और उपयोग के लिए छवियों को नये माध्यम में स्थानान्तरित करना।
- ❖ स्कैन की हुई आकृतियों को उपयोग के योग्य बनाने के लिए उनके भंडारण पर ध्यान दिया जाना जरूरी है। सभी फाइलें जो किसी भी सांख्यिकीय छवि परियोजना के अन्तर्गत बनायी जाती हैं सुव्यवस्थित और इस प्रकार वर्णित होनी चाहिए जो उस योजना के लक्ष्य के अनुकूल हो।

#### छवियों का सुनियोजन

किसी फाइल का नाम तथा विवरण निर्धारित करने से पहले उसके भंडारण की प्रक्रिया एवं प्रारूप को निर्धारित करने की आवश्यकता होती है। पाण्डुलिपियों के सन्दर्भ में मूल प्रलेख स्कैन करने के बाद उन्हें पुस्तकालय प्रबन्धन के सिद्धान्तों के अनुसार सुनियोजित किया जाता है। सामान्य रूप से संगृहीत सामग्री को संग्रहकर्ता पुस्तकालय अथवा संग्रहालय द्वारा संख्या दी जाती है। इन्हें इस प्रकार व्यवस्थित किया जाता है कि कोई व्यक्ति सिर्फ नाम देखकर किसी सांख्यिकीकृत पाण्डुलिपि के विषय में सविस्तर बता सकता है।

#### अल्पकालिक भंडारण

सांख्यिकीय छवियाँ, डीवीडी, नीली किरण, टेप एवं उत्तम तथा प्रख्यात गुणवत्ता वाली हार्डडिस्क पर भंडारण की जा सकती हैं।



## दीर्घकालिक भंडारण ( अभिलेखीय )

दीर्घकालिक भंडारण के लिए सामान्य अनुशांसा

- ❖ चमकीले सोने की परत वाला डिस्क अभिलेखीय रिकॉर्ड के लिए उपयुक्त है। ऐसा देखा गया है कि वर्तमान में तकनीकी दृष्टि से प्रोन्नत डिस्क में सोने की चमकीली परत लगाना धीरे-धीरे अप्रचलित होता जा रहा है।
- ❖ सामान्य भंडारण के लिए तापमान  $4^{\circ}$  डिग्री सेलसियस, ( $39^{\circ}$  फारेनहाइट) और  $20^{\circ}$  डिग्री सेलसियस, ( $68^{\circ}$  डिग्री फारेनहाइट) के बीच होना चाहिए। आपेक्षिक नमी  $2^{\circ}$  प्रतिशत से  $5^{\circ}$  प्रतिशत होनी चाहिए।
- ❖ लम्बी अवधि के भंडारण के लिए  $18^{\circ}$  डिग्री सेलसियस ( $64.4^{\circ}$  फारेनहाइट) तापमान और 40 प्रतिशत आपेक्षिक नमी होनी चाहिए।
- ❖ लम्बी अवधि के अभिलेखीय भंडारण के लिए उपर्युक्त से भी कम तापमान और आर्द्रता होनी चाहिए।
- ❖ प्रायः कम से कम प्रकाश वाले क्षेत्र में भंडारण अधिक अच्छा होता है।
- ❖ सांख्यिकीय डाटा को लम्बी अवधि तक सुरक्षित रखने का एक सही विकल्प है, उसका माइक्रोफिल्म तैयार करना।

## ५.१० पुनः प्राप्ति

सांख्यिकीय डाटा को पुनः प्राप्त करने के लिए एक खोजीयन्त्र का विकास कर उसके माध्यम से चयनित सामग्री को प्राप्त किया जा सकता है। खोजीयन्त्र इतना विकसित होना चाहिए कि वह मेटाडाटा में उपलब्ध सम्पूर्ण सामग्री को खोज सके। लोकल सर्वर एवं वेब साइट के माध्यम से उस खोजीयन्त्र द्वारा सामग्री की उपलब्धि हो सकती है।

## ५.११ स्थानान्तरण ( सांख्यिकीय एवं भौतिक )

सांख्यिकीय डाटा को लम्बी अवधि तक पहुँच के योग्य बनाये रखने के लिए उसे नयी फाइल की आकृति अथवा भौतिक आधार में स्थानान्तरित किया जा सकता है। प्रत्येक संस्था को पाँच वर्षों में एक बार इस प्रकार के स्थानान्तरण के लिए सिद्धान्त और विधि सुनिश्चित करनी चाहिए। आवश्यकता होने पर स्थानान्तरण पाँच वर्ष से कम अवधि में भी किया जा सकता है।

## राष्ट्रीय पाण्डुलिपि मिशन के मामले में

- ❖ पाण्डुलिपियों की संख्या 100 या उससे अधिक होने पर सांख्यिकीकरण के लिए उन्हें उनके मूल स्थान से हटाया नहीं जायेगा।





- ❖ अगर पाण्डुलिपियों की संख्या १०० से कम है तो उपयुक्त अधिकारी की लिखित अनुमति से उसे मूल संग्रहालय से उस स्थान तक स्थानान्तरित किया जायेगा जहाँ सांख्यिकीकरण की प्रक्रिया हो रही है।
- ❖ पाण्डुलिपि को सुरक्षित ले जाने का खर्च राष्ट्रीय पाण्डुलिपि मिशन का होगा और मूल संग्रहालय या संग्रहकर्ता के संरक्षण में यह कार्य किया जायेगा।
- ❖ पाण्डुलिपि को ले जाने या ले आने की जिम्मेदारी संग्रहकर्ता की होगी।
- ❖ लाने ले जाने की प्रक्रिया में यदि पाण्डुलिपि को कोई क्षति पहुँचती है तो उसकी जिम्मेदारी मिशन की नहीं होगी। वे दूसरी समस्त संस्थाएँ जो दस्तावेज़ के सांख्यिकीकरण की इच्छुक हैं उन्हें दस्तावेज़ की सुरक्षा के लिए नियम बनाने चाहिए तथा लाने ले जाने सम्बन्धी नियम भी स्पष्टतः परिभाषित करने चाहिए।

उपर्युक्त नियम मिशन को उदाहरण के रूप में प्रस्तुत करते हुए दिये गये हैं।

## ६. परियोजना की लागत

अंकन के उपकरणों का चयन करते समय वित्त की महत्वपूर्ण भूमिका होती है। छवि निर्माण में आने वाला खर्च केवल अंकन के उपकरण पर ही आधारित नहीं होता उसके साथ साथ परिधीय उपकरण, प्रकाश व्यवस्था, श्रम की लागत, प्रसंस्करण उपकरण, भंडारण की लागत आदि उसमें जुड़े होते हैं।

खर्च कम करने के लिए छवि उतारने, प्रसंस्करण और भंडारण की प्रक्रिया जहाँ तक संभव हो स्वचालित होनी चाहिए। ऐसा स्कैनर जिसमें मूल वस्तु ऊपर से अंकित की जाती है और जिसमें प्रसंस्करण में समय कम लगता है, एक अच्छा विकल्प हो सकता है। सांख्यिकीय फोटोग्राफी में प्रकाश व्यवस्था वस्तु को बेहतर प्रकाशित करती है। इस तरह स्कैनर और कैमेरा में किसी एक का चुनाव, इस बात पर निर्भर करता है कि स्रोत सामग्री, जिसका सांख्यिकीकरण किया जाना है, कितनी है? यह, अलग अलग संस्थाओं के द्वारा लिया जाने वाला, स्वतन्त्र निर्णय का, विषय है।

किसी भी सांख्यिकीकरण परियोजना पर विचार करने के समय सांख्यिकीय सामग्री को नये माध्यम और स्वरूप में स्थानान्तरित करने के खर्च को भी अनिवार्यतः ध्यान में रखा जाना चाहिए।



## अनुबंध १

### छवि गुणवत्ता परीक्षण सूची

न्यूनतम छवि गुणवत्ता सुनिश्चित करने के लिए यह सांकेतिक गुणवत्ता परीक्षण सूची है। कुछ आकलनों को मूल के साथ प्रत्यक्ष तुलना की आवश्यकता होगी।

१. छवि का सही आकार लम्बे आयाम में विभेदन
    - क. सांख्यिकीय मास्टर ३०००-५००० = पिक्सेल (असम्पीडित)
    - ख. स्वच्छ मास्टर ३०००-५००० = पिक्सेल (सम्पीडित)
    - ग. पहुँच फाइल १०२४ X ७६८ = पिक्सेल (सम्पीडित)
  २. फाइल का नाम सही है (इस्तेमाल किया गया नामकरण प्रतिमान देखें)
    - क. सांख्यिकीय मास्टर
    - ख. स्वच्छ मास्टर
    - ग. पहुँच फाइल
  ३. फाइल स्वरूप सही है
    - क. डिजिटल मास्टर - टिफ (असम्पीडित)
    - ख. स्वच्छ मास्टर - टिफ (सम्पीडित)
    - ग. पहुँच फाइल - जेपेग/पीडीएफ-ए (संपीडित)
- निम्न के लिए सांख्यिकीय प्रतियों की मूल के साथ तुलना करें
४. छवि सही रंग से बनी है (८ बिट ग्रे स्केल आदि)
  ५. सही ढंग से कटी है
  ६. घूमी हुई फाइल नहीं है
  ७. टेढ़ी नहीं है
  ८. तीक्ष्णता का अभाव/अत्यधिक तीक्ष्णता है
  ९. लहरिया आकृति नहीं है
  १०. पिक्सलीकरण नहीं है
  ११. रंग नहीं पड़ा है
  १२. हिस्टोग्राम (नहीं काटा गया हो - २५६ प्रकार के रंगों की छाया का प्रतिनिधित्व करने वाली स्याही के भेद में विवरण अंश की कोई क्षति नहीं हो)।
  १३. रंग अत्यधिक गहरा/अत्यधिक हल्का
  १४. असमान रंगत वाला/चमक वाला
  १५. अत्यधिक शोर
  १६. टिप्पणियाँ



## तकनीकी शब्दों की सूची तथा हिन्दी अनुवाद

Access	:	पहुँच
Archival	:	अभिलेखीय
Computer	:	संगणक
Compression	:	सम्पीडन
Contact	:	सम्पर्क/सम्पर्की
Contrast	:	विषमता
Digit	:	संख्या
Digital	:	सांख्यिकीय
Digitize	:	सांख्यिकीकरण
Document	:	प्रलेख/प्रलेखन/दस्तावेज़
Format(v)	:	स्वरूपण
Handling	:	व्यवस्थापन
Imaging/Tagging	:	अंकन
Migration	:	स्थानान्तरण
Moire	:	लहरिया
Non contact	:	असम्पर्क/असम्पर्की
Oxydation	:	ऑक्सीकरण
Processing	:	प्रसंस्करण
Reproduction	:	प्रजनन
Resolution	:	विभेदन
Search Engine	:	खोजीयन्त्र
Spatial Resolution	:	स्थानिक विभेदन
Specification	:	विनिर्देश
Surrogate	:	प्रतिनियुक्त
Technology	:	प्रौद्योगिकी



## आभार:

### तकनीकी विशेषज्ञता

राष्ट्रीय सूचना विज्ञान केन्द्र  
सांस्कृतिक सूचना विज्ञान प्रभाग  
नई दिल्ली

इन्दिरा गांधी राष्ट्रीय कला केन्द्र  
नई दिल्ली

राष्ट्रीय अभिलेखागार  
भारत सरकार  
नई दिल्ली

राष्ट्रीय संग्रहालय  
नई दिल्ली

राष्ट्रीय पाण्डुलिपि मिशन  
नई दिल्ली

### संदर्भ

- |                                    |   |                              |
|------------------------------------|---|------------------------------|
| लाइब्रेरी ऑफ कांग्रेस              | : | उत्पादन विनिर्देश            |
| न्यूजीलैण्ड का राष्ट्रीय पुस्तकालय | : | छवि गुणवत्ता की परीक्षण सूची |
| आस्ट्रेलिया का राष्ट्रीय पुस्तकालय | : | उत्पादन विनिर्देश            |
| यूनेस्को: सांख्यिकीकरण             | : | चयन मापदंड आदि।              |



# Guidelines for Digitization of Archival Material

## 1. Introduction

The National Mission for Manuscripts since its inception in the year 2003 has undertaken the work of preserving and providing access to the manuscripts wealth of India in a big way.

This document sets the guidelines for creating archival quality digital still images of materials which include manuscripts, printed books, maps, photographs, slides, negatives, microfilms etc. These guidelines are meant for any organization planning to digitize these materials. The guidelines specify factor affecting image quality, file formats, storage and access standards for images.

The guidelines are prescribed to maintain consistency of image quality, scanning methods and accessibility. It is advisable to use these guidelines to create images of long-term usage and reduce the need to rescan material.

Digitization standards for audio and video recording, born digital material, conservation processes for preparation of material for digitization and scholarly inputs required in deciphering manuscripts like folio numbers etc. are out of the scope of these guidelines.

## 2. Background of the NMM

India has the largest collection of manuscripts in the world. They are spread all over the country and also abroad in different libraries, academic institutions, museums, temples and monasteries and in private collections. The rich manuscript wealth of India today faces a threat of extinction. The invaluable heritage of India in the form of manuscripts



needs to be documented, preserved and made accessible to the present and to successive generations for dissemination of knowledge contained therein.

National Mission of Manuscripts (NMM) has the primary objective of using digital technology to preserve the manuscripts for posterity. NMM has studied the best practices being adopted in several digitization projects at national and international level and after long consultation with experts in the field has come up with this document that presents guidelines of digitization.

### 3. Digitization

Digital technology has opened up a new perspective in the field of data storage. The World Wide Web holds millions of websites and the Internet is now integrally and universally linked with research, teaching, expression, publication and communication of information. Libraries and Archives are society's primary information providers so far as cataloguing and processing of data is concerned. Besides preserving and providing access to 'born digital material' a great number of libraries nowadays have also turned to creating digital surrogates from their existing resources. Digitization means acquiring, converting, storing and providing information in digital format that is standardized, organized and available on demand.

#### 3.1 Target Audience

These guidelines are aimed at decision makers, archivist, library managers, and curatorial and technical staff members.

### 4. Why Digitize?

The reasons for implementing a digitization project, or more precisely for digital conversion of non-digital source material, are varied.



### Enhanced access

This is the most obvious and primary reason, mainly for research purposes. Where there is high demand from users and the library or source has the desire to improve access to a specific collection, digitization is the first choice. Services to an expanding user's group can be improved by providing enhanced access to the institution's resources with respect to education and life long learning.

### To facilitate new forms of access and use

The main purpose in this case is to enable the use of such originals that cannot be consulted without a visit to the specific repository, as also those originals that are fragile or damaged and where technology is needed to reveal the content and shape.

### For preservation

The purpose is, in the first place, to create accurate reproductions of the original documents on a long lasting medium. These reproductions need to satisfy both users of today and future potential users, and must, therefore, both be of high quality and possess a physical stability that can be maintained over time.

### Reduction in handling

Reduce the handling and use of fragile or heavily used original material and create a "back up" copy for endangered material.

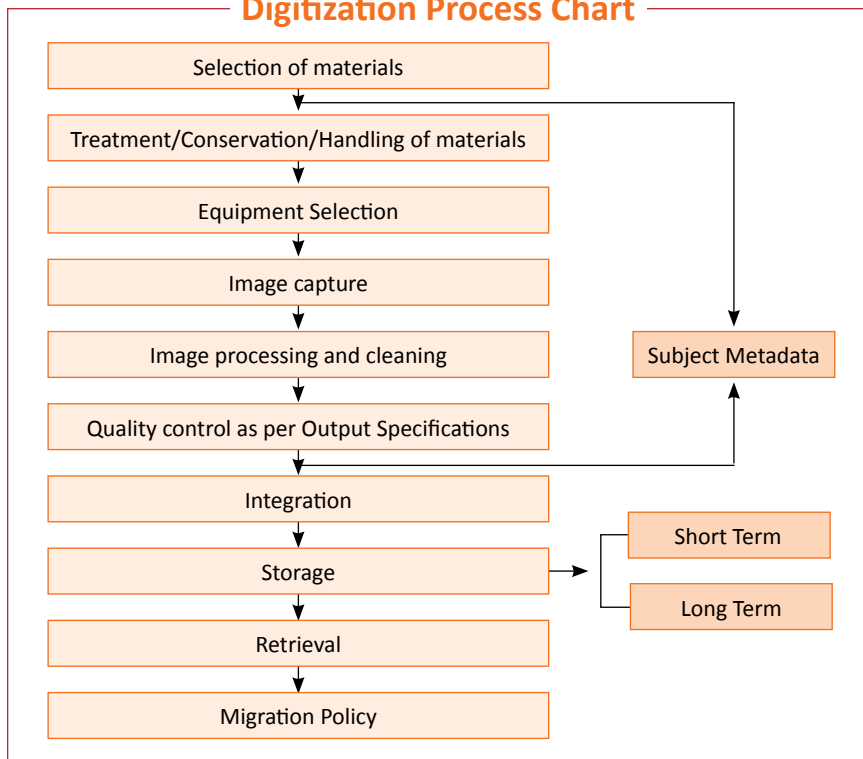
## 5. Process of Digitization

It is the process of converting the target material into digital format, storing the same in a digital medium and providing access using retrieval software.

The following chart highlights various steps involved in the digitization process.



## Digitization Process Chart



### 5.1 Selection of Material

It is important to see digitization as a series of choices where competing requirements and demands have to be balanced. When selecting source material for digitization it comes down to three basic questions: whether the source material

1. Needs to be converted?
2. Should be converted?
3. Can be converted?

The selection therefore has to be made in such a way that it will assure that not only issues like the value of the selected material and interest





in its content are considered but also demands concerning technical feasibility and institutional requirements are taken into account.

### Condition

Selection of material for digitization will be affected both by its physical condition and by readability. Material, which is fragile, damaged and in poor condition may present too many risks of further damage being caused by handling to allow it to be scanned without special care or without some basic conservation techniques. Similarly, if the material being considered as a candidate for digitization lacks detailed cataloguing or descriptive data, it is essential to create such data for future access to such material. In such cases the necessary cost of preparing the descriptive data would need to be included in the budget of digitization project.

### Demand

The level of demand is of course of great interest when selecting source material for digitization. Scholars, archivists, librarians, curators and researchers etc. should be involved in the process of selection and prioritization of the materials for digitization.

### Content

Regardless of the purpose for implementing a digitization project, the selection of source material will always be more or less content driven.

## 5.2 Treatment/Conservation/Handling of Materials

Materials taken up for digitization should have undergone

- ❖ Conservation, wherever required.
- ❖ Sequencing of materials, wherever required.
- ❖ Cleaning through standard conservation techniques to increase legibility of materials as and where necessary.

The following precautions need to be followed to spare damage to the originals especially in case of manuscripts:



- ❖ Materials should be handled in the best way as per guidelines of the experts. Therefore, during digitization process the material shall be placed on the scanning, filming platform by the custodian/experts or strictly under their guidance.
- ❖ In general, binding is not allowed to be taken off as it may damage the original, however in some cases where it is absolutely necessary, due care shall be taken to remove the binding and rebind them using recommended methods.
- ❖ Scanner/Camera operators should wear surgical gloves so as not to damage any of the materials.
- ❖ Soft bristled paint brushes should be used to wipe away the years of accumulated dust and dirt as necessary.
- ❖ Long, horizontal format requires special handling considerations.
- ❖ To maintain the sequence of loose leaf materials, local scholars should be engaged to remove the thread, enumerate the pages and record the missing folios and rethread the materials after digitization.

Rough handling or placing the original under stress, especially the binding of a book, should be totally avoided. Damage avoidance from handling or equipment shall have priority over the requirements such as capturing of details of printing or writing on originals.

### 5.3 Selection of equipment

It has an important impact on the quality of image. Equipment from different manufacturers can perform differently, even if the specified technical capability appears to be the same. Non-contact device should be used to capture images with cool light. Use of touch devices should not be allowed as it might harm the originals. The equipment (including lights) used for all image capture should be such as not to damage original material. Besides, light should be handled in a manner which is not damaging to the source material.



## 5.4 Image capture

### Input Specification

The source materials are generally categorized by material, size of materials, condition of materials, illustrated and non-illustrated materials etc. Materials are available at various repositories located in different parts of the country and need to be digitized at site.

Materials are primarily available on paper (various types), palm leaf, bhoj patra, bamboo leaf, cloth, clay tablet, parchment, tamrapatra, wood, wooden bead, ivory cover/sheet, deer skin, and also secondarily available on microfilm, photographic plate and negative slide etc. They are generally very old and brittle and need special and sophisticated handling techniques.

Some materials have illustrations/charts etc which are created using ancient inks, vegetable dyes, metals such as silver, gold etc. They are likely to get oxidized if exposed to bright light and heat.

Illustrated materials like manuscript charts should be scanned separately and merged with the text at the appropriate location. While capturing images of illustrations especially when they are created using metal such as silver, special care should be taken to avoid oxidation.

**Image quality** at capture can be defined as the cumulative result of the scanning resolution, the bit depth of the scanned image, the enhancement processes and the compression applied, the scanning device or technique used, and the skill of the scanning operator.

**Resolution** is determined by the number of pixels used to present the image, expressed in dots per inch (dpi) or pixels per inch (ppi). Increasing the number of pixels used to capture the image will result in a higher resolution and a greater ability to delineate fine details, but just continuing to increase resolution, after certain extent, will not result in better quality, only in a larger file size.



**Bit depth** is a measurement of the number of bits used to define each pixel. The greater the bit depth used, the greater the number of grey and colour tones that can be expressed.

- ❖ *Bitonal scanning* uses one bit per pixel to represent black and white.
- ❖ *Greyscale scanning* uses multiple bits per pixel to represent shades of grey, the preferred level of grey scale is 8 bits per pixel, and at this level the image displayed can select from 256 different levels of grey.
- ❖ *Colour scanning* uses multiple bits per pixel to represent colour. 24 bits per pixel is called true colour level, and it makes possible a selection from 16.7 million colours.

## Digital Conversion

Selected materials are transformed to bit-mapped images by scanner or digital camera. During image capture these documents are “read” or scanned at appropriate resolution and bit-depth to produce the output (Raw Master Images) as per the defined specification (please refer table A). The resulting digital files, containing the binary digits (bits) for each pixel, are then formatted and tagged in a way that makes it easy for a computer to store and retrieve them. The resulting digital image is saved in Tiff format as per specification given in ‘Table A’ under the heading ‘Raw Master Image’. From these files the computer can produce analog representations for on-screen display or printing. Because files with high- resolution images are very large it may be necessary to reduce the file size (compression) to make them more manageable both for the computer and the user.

When a source material has been scanned, all data is converted to a particular file format for storage as defined in Table A.

## 5.5 Digital Image processing and cleaning

The goal of any digitization programme should be to capture and present in digital formats the significant informational content contained in a



single source document or in a collection of such documents. To capture the significant parts, the quality assessment of the digital images have to be based on a comparison between those digital images and original source documents that are to be converted, not on some vaguely defined concept of what is good enough to serve the immediate needs.

Original raw image should be saved as per Master Image Specification (please refer table A). The raw image should be processed to remove dirt, worm marks, water marks, noise, shadow, scratch marks, skew etc. Adjustment of brightness and contrast, gamma correction, sharpening and blurring, removing patterns and adjusting colors should also be part of image processing. Cleaned image should be saved as per Clean Master Specification (please refer table A).

Compression is normally used to reduce file size for processing, storage and transmission of digital images. Methods used are, for example, to abbreviate repeated information or eliminate information that the human eye has difficulty in seeing. The quality of an image can, therefore, be affected by the compression techniques that are used and the level of compression applied. Compression techniques can be either “loss less”, which means that a decompressed image will be identical to its earlier state because no information is thrown away when the file size is reduced, or “lossy” when the least significant information is averaged and decompressed image in this case will not be identical to its earlier state as part of information is discarded in this process. In general “loss less” compression is used for master files and “lossy” compression techniques for access files.

Derivative images namely Access Image will be derived from the cleaned image.

## 5.6 Output Specification

1. Raw Master Image (Original Uncleaned and Uncompressed)
2. Clean Master Image (Cleaned loss less compressed image)
3. Access Image/PDF-A (Derivative lossy image)



The detail specifications of these images are as follows:

**Table A**

Material	Raw Master Image	Bit depth	Clean Master Image	Access Image
Manuscripts	Tiff (Uncompressed) 300 dpi (output)	24 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT Spatial Resolution: 8" X 10" at 300 dpi	JPEG/PDF-A Spatial Resolution: 1024 x 768 pixels
Printed Books	Tiff (Uncompressed) 300 dpi (output)	24 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT	JPEG/PDF-A
B/W		8 bit		
Photographs	Tiff (Uncompressed) 300/600 dpi (output) 1200 dpi or higher for special cases	24 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT	JPEG/PDF-A
B/W		8 bit		
Slides/ Negatives	4000 dpi Tiff (Uncompressed)	24 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT	JPEG/PDF-A
B/W		8 bit		
Microfilm/ microfiche	300 dpi Tiff	8 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT	JPEG/PDF-A
Map/ Drawing	Tiff (Uncompressed) 300/600 dpi (output) 1200 dpi or higher for special cases	24 bit	Tiff (compressed) 300 dpi (output) Group 6 CCITT	JPEG/PDF-A
B/W		8 bit		



## 5.7 Metadata

### Subject Metadata of Manuscripts

- |                         |                                   |
|-------------------------|-----------------------------------|
| 1. Material number      | 13. Bundle number                 |
| 2. Title                | 14. Folio number                  |
| 3. Other title          | 15. Pages                         |
| 4. Author               | 16. Material                      |
| 5. Organization         | 17. Missing portion               |
| 6. Commentary           | 18. Illustrations                 |
| 7. Commentator          | 19. Condition                     |
| 8. Scribe               | 20. Catalogue source              |
| 9. Language             | 21. Remarks                       |
| 10. Script              | 22. Manuscript date               |
| 11. Complete/Incomplete | 23. Manuscript length (in inches) |
| 12. Subject             | 24. Manuscript width (in inches)  |

Subject metadata referred to as above has been created according to the specific Manus Data Record using Manus Granthawali software of NIC already in use within the National Mission for Manuscripts. It may vary in case of other institutions.

### Technical metadata

Technical Metadata is that which describes the features of the digital file. Technical Metadata is automatically generated and assigned to the image file at the time of creation.

File Name	:	Assigned at the time of scanning
Date Created	:	
Date Modified	:	
Equipment Used	:	
Image Format	:	
Width	:	Pixels
Height	:	Pixels
Color Mode	:	RGB, grayscale



Resolution	:	Pixel per inch
File Size	:	
Color Profile	:	ICC color profile
Make	:	Scanner Software
X Resolution	:	Scanning Resolution in the X axis
Y Resolution	:	Scanning Resolution in the Y axis
Resolution Units	:	
Software	:	Imaging Software

Using image editing tool these technical metadata can be viewed.

## 5.8 Quality control

It is an important component of a digital imaging project. Without this activity it may not be possible to ensure the quality of the digitized images.

### Methods

The automated digital image evaluation tools that are available today are quite expensive and may not be easily accessible. Therefore, it is suggested to at least undertake visual quality evaluation through digital image viewing on screen or print outs (on sample basis).

### Recommended methods for:

#### On-screen evaluation

- ❖ View scanned images at 1:1 (100% enlargement).
- ❖ Use of calibrated chart to evaluate greyscale and colour reproduction (colour calibration).
- ❖ Use resolution targets and histograms to evaluate spatial resolution and tonal reproduction.
- ❖ Image Quality Control Checklist at Annexure I

#### Print-out evaluation (on sample basis)

- ❖ Compare the printouts with the source documents.





## Naming of images

There are two approaches for this: (1) to use a numbering scheme that reflects numbers already in use in an existing cataloguing system, or (2) to use meaningful names. Both approaches are valid, and what fits a certain collection or source document should be chosen.

In the case of National Mission for Manuscripts images will be named according to Manus Id generated by the cataloguing software of the mission called Manus Granthavali. Meta data information for each manuscript scanned will be stored in the database and will be identified by its Manus Id. So the Manus Id and the Accession Number given by the institute where the digitization is taking place forms the basis of naming the digitized images of each manuscript page.

**Folder Naming:** Name of the folder containing the manuscript images must be the Manus ID (which will be provided by NMM) following title, Acc.no, type of image (Clean Tiff/Raw tiff/PDF).

**For Example:** IGO 001333948 Krushnalila EXT\_96 Clean Tiff.

If the accession number is not available then it may be excluded from the name of the folder.

**For Example:** IGO 001333948 Krushnalila Clean Tiff.

If the manuscript name is provided in diacritic or devanagari in the metadata; then the folder name should contain the name in the concerned format only.

**For Example:** BOO0000740792 Rāmāyaṇa S17-58-571 PDF.

Similarly other institutions that go in for large scale digitization would need to create a system of naming images. It is worth noting here that the system of naming should be such as would facilitate access to data as well as would be easy to understand for the lay users.

## Description of images

To describe digital images there is a need for metadata that is structured data about data. Metadata can also be defined as data that facilitates the



management and use of other data. In case of manuscripts, metadata information is stored through the Manus Granthavali software for each manuscript by National Mission for Manuscripts.

Similarly, the metadata structure for other digital material should be standardized by the custodian institution. It is recommended to use Dublin core standards.

The highest quality file produced is referred to as Digital Master. These files are created through direct result of image capture. Master file represents the original images as closely and correctly as possible. Derivative images are generated from the master image using image editing software.

## 5.9 Integration, Storage and Management of digitized images

The possibility of being able to use a collection of digital images in the way it was intended, depends not only on conversion standards and quality controls, but also on how the digital images are stored and managed. If the purpose is to meet not only short term needs but also to provide accessibility over time, steps have to be taken to satisfy both current users and the expectations of future users.

Plan must be made to:

- ❖ Make scanned images available for the intended use.
- ❖ Periodically upgrade the user interface functionality.
- ❖ Transfer images to new media to meet increasing capacity for processing and handling of digital information.

To make scanned images usable, great care should be taken relating to their storage. All image files that are produced by a digital image project must be organized, named and described in a way that fits the purpose of the project.

### Organization of images

The process and format of storage needs to be decided before a name and a description of an image file is considered. In case of manuscripts,



the source documents being scanned are physically organized according to principles of library management. Collections of materials generally have numbers given by the library or repository where they are stored. Therefore, it is organized in such a way that just by looking at the name one can tell about the details of digitized manuscript.

### Short term storage

The digitized images may be stored on DVD's, blue ray, tapes and hard disks of superior and reputed quality.

### Long term storage (archival)

General recommendation for long-term storage conditions

- ❖ For archiving recordable (R) discs, discs that have a gold metal reflective layer are recommended. It has been observed in the technically advanced current disks that the use of gold as a reflecting layer is slowly becoming less important.
- ❖ For general storage, a temperature between 4°C (39°F) and 20°C (68°F) is recommended with a relative humidity of 20% to 50%.
- ❖ For long term storage, 18°C (64.4°F) and 40% RH are recommended.
- ❖ For extended term archival storage even lower temperature and humidity are recommended.
- ❖ Storage in the dark, while not absolutely, is required.
- ❖ Microfilming of digital data can be an option to preserve the digital data for long term.

## 5.10 Retrieval

Digital data can be retrieved by developing a search engine which can be used to select the material. The search engine may be made intelligent to search the full text on metadata. The data can be accessed in house through a local server or through web access using the search engine.



## 5.11 Migration Policy (Digital and Physical)

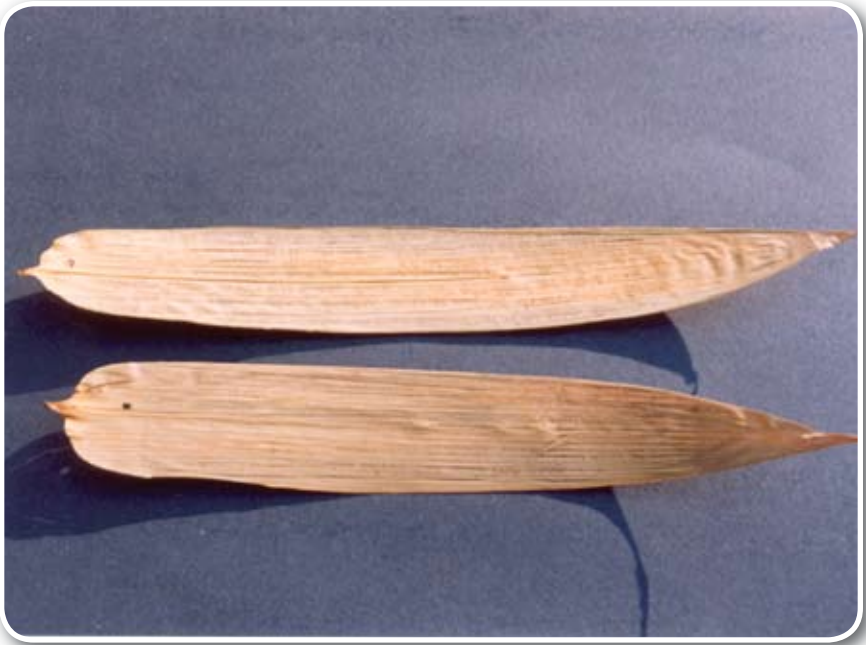
Digital data may be migrated to new file formats or physical media to ensure long-term accessibility. The institution should have a migration policy in place for migration to new file format or media (established) every five years or earlier, if required.

In the case of National Mission for Manuscripts:

- ❖ No manuscript will be relocated or transferred temporarily from the repository to any other location if the number of manuscripts to be digitized is more than or equal to 100.
- ❖ If the number of manuscripts is less than 100 then it may be transferred from the original repository to the digitization station with prior and written permission of the competent authority of the manuscript repository concerned and if necessary, with the consent of the state/district administration.
- ❖ Transporting cost and package cost will be borne by NMM and supervised by original repository/manuscript holder.
- ❖ The responsibility of transporting will be that of the manuscript holder.
- ❖ NMM will not be responsible for any damage to manuscripts in the process of transportation. Other agencies wanting to get their data digitized need to formulate rules governing safety and transportability of original material. The above is only a specific case study of NMM.

## 6. Project Costing

While selecting the imaging equipments cost plays a crucial role. The total cost of imaging not only involves the cost of the capture device but also associated peripheral devices, lighting equipments, labour cost, processing equipments, storage cost both long term and short term etc.



The workflow of image capturing, processing and storing should be automated, as far as possible, to reduce the cost. Face up scanners with low processing time can be a better choice, whereas in digital photography flexible lighting arrangements can make the object better lit. The choice between a scanner and a camera for digitizing would largely depend on the quantum of data to be digitized and it is upto the individual institution to take a decision in this matter.

While conceptualizing any digitization project of archival importance the cost involved for migrating the digital content to new media/format should be considered.



## Annexure I

### Image Quality Check List

This is an indicative quality checklist for ensuring minimum image quality. Some of the assessments will need to be made with direct comparison with the original.

1. Image is correct size/resolution in the long dimension
  - a. Digital Master = 3000-5000 pixels (uncompressed)
  - b. Clean Master = 3000-5000 pixels (compressed)
  - c. Access File = 1024x768 pixels (compressed)
2. File Name is correct (refer naming convention used)
  - a. Digital Master
  - b. Clean Master
  - c. Access File
3. File Format is correct
  - a. Digital Master - Tiff (uncompressed)
  - b. Clean Master - Tiff (compressed)
  - c. Access File – Jpeg/PDF-A (Compressed)

Compare Digital copies with the original for:

4. Image is correct color made (8 bit gray scale etc.)
5. Cropped Correctly
6. Not rotated/flipped
7. Not skewed
8. Lack of sharpness/excessive sharpness
9. No moiré patterns
10. Not Pixilated
11. Not color cast
12. Histogram (Not clipped – No loss of detail in high- lights shadows – 256 shades of gray represented – tonal variation)
13. Over all Too Dark/Too Light
14. Uneven tonal values/flare
15. Excessive noise
16. Remarks

## Credits

### Technical Expertise

National Informatics Center  
Cultural Informatics Division  
New Delhi

Indira Gandhi National Centre for the Arts  
Cultural Informatics Lab  
New Delhi

National Archives of India  
New Delhi

National Museum  
New Delhi

National Mission for Manuscripts  
New Delhi

### References

- Library of Congress : Output specifications
- National Library of New Zealand : Image quality checklist
- National Library of Australia : Output Specifications
- UNESCO: Digitization : Selection criteria etc.



National Mission for Manuscripts  
11 Mansingh Road, New Delhi – 110 001

Tel: +91 11 23383894; Fax: +91 11 23073340

Email: [director.namami@nic.in](mailto:director.namami@nic.in); Website: [www.namami.org](http://www.namami.org)