기록관리업무 표준

NAK/TS 2:2008(v1.0)

National Archives Standard

전자기록물 문서보존포맷 기술규격

Digital Document File Format for Long-term Preservation

Version 1.0

국가기록원
National Archives of Korea

2008년 11월 4일 제정

○ 제 정 자 : 행정안전부 국가기록원장
○ 제 정 일 : 2008년 11월 4일(행정안전부 고시 제 2008-43호)
○ 심의부회 : 국가기록관리위원회, 표준전문위원회
○ 원안작성 :
• 국가기록원 기록정보화과 김동명(공업연구사)
○ 검토·관리 :
• 국가기록원 표준협력과 김형국(학예연구관), 이지영(공업연구사)
○ 자 문 :
• 충북대학교 오항녕(연구교수)

(1) 이 표준에 대한 의견 또는 질문은 아래 전화로 연락하거나 홈페이지를 이용하여 주십시오.

• 표준열람 : 국가기록원(http://www.archives.go.kr)→기록관리자 서비스→기록관리표준→표준화현황
• 행정안전부 국가기록원 기록정책부 표준협력과(042-481-6248, 6250)
기록정보서비스부 기록정보화과(042-481-8970)

(2) 이 표준에 대한 저작권은 국가기록원에 있으며, 이 문서의 전체 또는 일부에 대하여 상업적 이익을 목적으로 하는 무단 복제 및 배포를 금지합니다.

Copyrightⓒ National Archives of Korea(2008). All Rights Reserved.

목 차

머리말 ····· ii

1 적용범위 ····· 1

2 인용표준 ····· 1

3 용어정의 ····· 2

4 표기법 ····· 4

5 문서보존포맷 개요 ····· 4
5.1 문서보존포맷 정의 ····· 4
5.2 문서보존포맷 필요성 ····· 4
5.3 문서보존포맷 고려요소 ····· 5
5.4 문서보존포맷 규격 ····· 7

6 문서보존포맷 세부규격 ····· 9
6.1 파일 구조(File structure) ····· 9
6.2 그래픽(Graphics) ····· 13
6.3 폰트(Fonts) ····· 18
6.4 투명도(Transparency) ····· 23
6.5 주석(Annotations) ····· 24
6.6 행위(Actions) ····· 25
6.7 메타데이터(Metadata) ····· 27
6.8 논리적 구조(Logical structure) ····· 37
6.9 대화식 폼(Interactive Forms) ····· 41

부속서 A (참고) 문서보존포맷 해외사례 ····· 43

참고문헌 ····· 46

머리말

이 표준은 문서가 생산된 시점에 가지고 있던 내용과 구조를 그대로 재현하여 전자문서의 내용보기가 가능하도록 문서보존포맷의 기술규격을 ISO 19005-1:2005의 PDF/A-1을 b수준으로 제정하는 것을 목적으로, 표준전문위원회의 전문심의 및 국가기록관리위원회의 심의를 거쳐 제정한 공공표준이다.

이 표준의 보다 구체적인 법률적 근거는 다음과 같다.

- 공공기록물 관리에 관한 법률 제20조(전자기록물의 관리)
- 공공기록물 관리에 관한 법률 시행령 제36조(기록관 및 특수기록관의 전자 기록물 보존)
- 공공기록물 관리에 관한 법률 시행령 제46조(영구기록물관리기관의 전자 기록물 보존)

이 표준은 국가기록원에 의해 유지 및 관리되며, 관련 법령의 개정, 기술의 발전, 관계기관의 요청 등으로 인해 개정이 필요할 경우에는 필요성 및 타당성 검토를 거쳐 개정안을 마련하고 전문가 검토 및 의견수렴 절차를 거쳐 개정을 추진한다.

이 표준은 저작권법에서 보호대상이 되는 저작물이다.

전자기록물 문서보존포맷 기술규격

1 적용범위

이 표준은 기술규격의 적용범위 및 대상으로 기록물관리기관에서 소장하고 있는 전자기록물 중 보존기간 10년 이상의 기록물을 대상으로 한다.

2 인용표준

이 표준은 다음의 표준을 참조하여 관련 조항을 구성하였다.

- ISO/IEC 9541-1, *Information technology - Font information interchange - Part 1: Architecture*
- ISO/IEC 10646-1, *Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane*
- ISO 15930-4, *Graphic technology - Prepress digital data exchange using PDF - Part 4: Complete exchange of CMYK and spot colour printing data using PDF 1.4 (PDF/X-1a)*
- *Date and Time Formats, W3C Note, 15 September 1997. Available from internet <http://www.w3.org/TR/NOTE-datetime>*
- ISO 19005-1:2005(PDF/A-1) - *Document management Electronic document file format for long-term preservation, Part 1: Use of PDF 1.4 (PDF/A-1)*
- ICC.1:1998-09, *File Format for Color Profiles, International Color Consortium. Available from Internet <http://www.color.org/ICC-1_1998-09.PDF>*
- ICC.1A:1999-04, *Addendum 2 to Spec. ICC.1:1998-09, International Color Consortium. Available from Internet <http://www.color.org/ICC-1A_1999-04.PDF>*
- *PDF Reference: Adobe Portable Document Format, Version 1.4, Adobe Syste*

ms Incorporated - 3rd edition. (ISBN 0-201-75839-3) Available from Internet <http://partners.adobe.com/asn/acrobat/docs/File_Format_Specifications/PDFReference.pdf>

- *XMP Specification, January 2004, Adobe Systems Incorprated Available from Internet <http://partners.adobe.com/asn/tech/xmp/pdf/xmpspecification.pdf>*
- *Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation, 4 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-xml-20040204>*
- *RDF/XML Syntax Specification (Revised), W3C Recommendation, 10 February 2004. Available from Internet <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>*

3 용어정의

이 표준의 목적을 위하여 다음의 용어와 정의를 적용한다.

3.1 그리프(Glyph)

특정 디자인으로부터 독립적으로 인식되는 이론적 그래픽 심볼
[ISO/IEC 9541-1]

3.2 사전(Dictionary)

키-값 쌍을 포함한 연합 표. 일반적으로 복잡한 객체의 속성들을 모으고 묶는데 사용되는 객체의 속성 이름과 값을 명시
[ISO 15930-4]

3.3 상호참조표(Cross reference table)

파일의 간접 객체의 시작 바이트(byte) 위치를 포함하고 있는 PDF 데이터 구조
[ISO 19005-1:2005]

3.4 적합성 수준(Conformance Level)

파일과 판독기가 반드시 준수해야 하는 제약사항과 요구사항들의 식별된 집합

[ISO 15930-4]

3.5 파일 끝 표시(End-of-file marker)

PDF 파일 끝을 표시하는 다섯 개의 문자열인 '%%EOF'

[ISO 19005-1:2005]

3.6 판독기(reader)

파일을 적절하게 읽을 수 있고 처리할 수 있는 소프트웨어 응용프로그램

[ISO 15930-4]

3.7 폰트(Font)

그리프 또는 다른 그래픽 요소일 수 있는 그래픽의 수집

[ISO 15930-4]

3.8 행 끝 표시(End-of-line marker)

EOL(End Of Line) 표시 문자 행의 끝을 표시하는 하나 또는 두개의 연속 문자. 캐리지 리턴(Carriage Return, 0Dh), 라인피드(Line Feed, 0Ah), 또는 캐리지 리턴 바로 다음의 라인피드

[ISO 19005-1:2005]

3.9 ICC 프로파일(ICC profile)

ICC 설계서와 부록에 순응하는 컬러 프로파일

[ICC.1: 1998-09]와 [ICC.1A:1999-04]

3.10 PDF(Portable Document Format)

일반적인 모든 사용자의 운영시스템 환경에 구속되지 않고 문서를 생산하고 읽고 출력이 가능하도록 지원하는 포스트스크립트(PostScript) 언어기반의 문서의 포맷 형식

[ISO 15930-4]

4 표기법

이 표준에서는 PDF 연산자, PDF 키워드, PDF 사전의 키 값과 이미 정의된 이름은 진한 휴먼고딕체 폰트로 표기하였다. PDF 연산자의 피연산자(operand) 혹은 사전 키의 값들은 이탤릭(기울임) 휴먼고딕체 폰트로 표기하였다.

보기 **TR2** 키에 대한 *Default* 값

PDF Reference 3.1에 정의된 대로 PDF 파일의 구조를 기술하고 객체를 구분하는데 사용되는 토큰 문자는 진한 휴먼고딕체 폰트의 대문자로 표기하며 괄호에 삽입된 두 자리의 16진수 숫자와 접미사 "h"를 붙여 표기한다.

보기 **CARRIAGE RETURN** (0Dh)

5 문서보존포맷 개요

5.1 문서보존포맷 정의

문서보존포맷은 문서가 생산된 당시의 애플리케이션이 없어도 해당문서의 내용과 외형을 그대로 재현하여 내용보기를 가능하게 하는 포맷이다. 이 포맷은 문서의 보존을 위해 필수적인 요소들을 규정하여 문서의 내용과 구조를 보존한다. 또한 이 포맷은 문서의 생산, 저장 또는 표현에 사용된 도구와 시스템에 관계없이 시각적 모양을 장기간 지속적으로 유지하도록 전자문서를 표현하며 필요시 정확하게 재현하여 이용자에게 제공한다.

5.2 문서보존포맷 필요성

전자문서는 문서를 생산했던 애플리케이션이 더 이상 존재하지 않거나, 운영

체제나 플랫폼 환경의 차이에 의해 원문 그대로의 모습을 볼 수 없거나, 심할 경우 파일 자체를 열어볼 수 없는 경우가 발생할 수 있다. 이런 문제점을 해결하기 위해 해당 애플리케이션의 버전 및 소멸에 영향 받지 않고 전자문서의 내용보기를 가능하게 하는 문서보존포맷이 필요하다[1].

5.3 문서보존포맷 고려요소[2]

5.3.1 공개용 표준

특정업체가 독점적으로 소유권을 가지고 있지 않아야 하며, 누구나 참조하고 이용할 수 있게 공개되어야 한다. 그리고 사용자에게 경제적으로 영향을 주지 않아야 한다.

5.3.2 편재성

오랜 기간 동안 사용될 가능성이 있어야 하며 많은 곳에서 사용되는 포맷이어야 한다. 그래서 포맷의 이용에 대해 지역이나 기간의 제한을 받지 않아야 한다.

5.3.3 안정성

원래 생산된 문서를 문서보존포맷으로 변환할 때 원래 문서의 내용, 구조, 맥락정보 등을 훼손시키지 않고 보존해야 한다. 즉 문서의 내용이 시간이 지남에 상관없이 그대로 유지될 수 있고, 버전의 지속적 변화에 상관없이 호환성이 유지되어, 구 버전의 문서 또한 그대로 볼 수 있어야 한다.

5.3.4 메타데이터 지원

문서를 장기보존할 때 필요한 메타데이터를 지원하여야 한다. 메타데이터 지원은 문서보존포맷의 필수적인 사항이다. 메타데이터 항목으로는 생산자, 출처, 내용, 주제, 키워드, 보존과 기술적(記述的) 특성에 관한 사항 등 전자문

서와 관련된 각종 정보를 지원한다.

5.3.5 상호운용성

운영체제나 플랫폼에 독립적이어서 서로 다른 시스템 간에 문서의 마이그레이션을 쉽게 하여 한 기관에서 생성된 문서보존포맷은 다른 기관 혹은 외부 이용자도 사용할 수 있어야 한다.

5.3.6 진본성

문서보존포맷은 문서의 내용, 외형 등이 시간의 경과에 상관없이 원래의 모습과 일치하며, 생산자가 생산하려고 했던 원래의 취지에 맞는 그 전자문서라는 증명이 가능하도록 문서가 훼손, 위조, 변조가 되지 않도록 하는 포맷이어야 한다.

5.3.7 표현력

문서보존포맷은 원래 생산된 전자문서의 내용뿐만 아니라 문서의 외형과 구성 그 자체가 그대로 표현되어야 하며 복원될 수 있어야 한다. 즉 이는 문서의 진본성과도 연결되는 조건으로서 원문적 특성이 원래의 문서 그대로 보존될 수 있어야 한다는 것이다.

5.3.8 검색기능

문서보존포맷은 문서 내부에서 이용자가 원하는 문서내용에 대한 검색 기능을 제공하여야한다. 그래서 전자문서를 열어보지 않고도 검색에 의해 문서가 검색과 관련된 내용을 포함하고 있는지에 대한 여부를 찾을 수 있게 하여 이를 통해 이용자는 자신에게 적합한 문서를 찾아낼 수 있다.

5.4 문서보존포맷 규격

5.4.1 개요

이 표준은 문서보존포맷이 갖춰야할 요건을 충족하는 PDF/A-1(ISO 19005-1:2005 Part 1: Use of PDF 1.4)을 문서보존포맷으로 규정하고 있다. 다만, 문서의 장기보존에 위배될 가능성이 있는 일부 요소(암호화, 내장파일, LZW압축, 투명성, 멀티미디어, 자바스크립트)를 금지하여 사용한다. 그리고 장기보존을 위해 필요한 부분(PDF/A를 위한 확장 스키마)을 추가하였다.

PDF/A-1의 "A"는 기록물 즉 Archive를 나타내며, "1"은 인쇄가 가능한 모든 매체를 대상으로 하고 있다는 의미이다. 그래서 동영상, 비디오, 오디오 매체에는 적용이 불가능하다.[1)]
다음 사항은 기술규격 범위에서 제외한다.

- 종이 또는 전자문서를 변환하는 특정 방법
- 기술적 디자인, 사용자인터페이스, 렌더링에 대한 구체적인 방법
- 저장 방법 및 저장 매체
- 필요한 컴퓨터 하드웨어 또는 운영체계

5.4.2 적합성 수준

5.4.2.1 적합성 A 수준

적합성 A 수준은 ISO 19005-1의 모든 기술규격을 수용한다. 이 수준에 따르는 파일을 "PDF/A-1a"라 한다.

5.4.2.2 적합성 B 수준

적합성 B 수준은 ISO 19005-1에서 유니코드 문자 맵(ISO 19005-1 6.3.8 참조)

1) 동영상, 오디오, 3D 그래픽, JPEG-2000압축 등에 대한 보존포맷인 PDF/A-2는 PDF 1.6 버전을 기반으로 한다. 현재 PDF/A-2에 대한 국제적인 표준화 작업이 진행 중이며 자바스크립트와 실행파일 같은 동적인 콘텐츠를 포함하는 PDF/A-3도 논의 중이다.

과 논리적 구조(ISO 19005-1 6.8 참조)을 제외한 모든 조건을 만족한다. 이 수준을 따르는 파일을 “PDF/A-1b"라 한다.

5.4.2.3 적합한 판독기

PDF/A-1 문서를 표현하는 적합한 판독기 규격을 정의하는데 기술적인 디자인, 사용자 인터페이스, 렌더링에 대한 구체적인 세부사항은 규정하지 않는다.

- PDF/A-1a용 판독기: PDF/A-1a, PDF/A-1b 수준의 문서를 처리한다.
- PDF/A-1b용 판독기: PDF/A-1b 수준의 문서를 처리한다.

5.4.3 이 표준에 적용된 적합성 수준

이 문서보존포맷의 규격은 태그 붙은 PDF(6.8.2 참조)가 가지는 특성으로 인해 “논리적 구조(6.8 참조)”의 일부를 충족할 수 없으므로 적합성 B 수준으로 선정한다.

태그 붙은 PDF가 가지는 특성을 살펴보면 다음과 같다.

- 기술규격에서 배재된 태그 붙은 PDF는 이용자가 원문을 편집할 때 태그 정보 및 스타일시트 정보를 분석, 적용하여 만든 문서 태그 정보가 들어있는 PDF로 태그 붙은 PDF가 적용된 PDF 문서는 사용자가 지정한 태그 및 스타일시트 정보에 따라 PDF 문서의 본문 텍스트 내용을 추출할 수 있다. 그래서 검색 텍스트 추출, TTS(Text To Speech) 텍스트 추출 등에서 정확성을 보장 받을 수 있다. 태그 붙은 PDF가 아닌 경우, 행끊김[2] 글자들에 대해 정확한 해석이 불가능하지만 태그 붙은 PDF는 행이 끊겨 다음 줄로 넘어간 글자들에 대한 정확한 해석이 가능하다는 것이다. 하지만 태그 정보를 PDF 문서에 유지하기 위해 많은 정보가 필요하며, 이는 PDF 문서의 용량 증가 및 판독기의 프로세싱 증가를 초래한다.
- 태그 붙은 PDF 적용 시 문제점으로는 원문의 태그 정보 및 스타일시트 정보를 반영해야 하기 때문에 PDF 문서를 변환할 때 원문 편집 프로그램의 태그 및 스타일시트 정보를 얻을 수 있어야 하는데 흔글 편집기와 기타 문서 편집기의 경우에는 태그 정보 및 스타일시트 정보를 얻을 수 있는 어떤 방법도 제공하지 않으므로 태그 정보 및 스타일시트 정보를

2) 단어가 다음 줄로 넘어감으로 분리되는 경우

얻어서 PDF 문서를 만드는 것이 원천적으로 불가능하다. 그리고 MS 오피스 계열(워드, 엑셀, 파워포인트)의 원문 편집 프로그램은 태그 정보 및 스타일시트 정보를 제공 받을 수 있는 프로그램 라이브러리 모듈을 제공하고 있으나 태그정보 및 스타일시트 정보를 반영하여 PDF 문서를 변환하는데 상당한 시간을 필요로 한다.[3)]

6 문서보존포맷 세부규격[4)]

6.1 파일 구조(File structure)

6.1.1 일반 사항

파일 포맷 이슈들의 전반적인 사항과 규격에 적합한 파일들의 일반적 구조를 형성하는 기본 요소들에 대해 언급한다.

6.1.2 파일 헤더(File header)

파일 헤더의 % 문자는 파일의 바이트 오프셋 0 에 있어야 한다.
파일 헤더 다음 행은 % 문자 다음에 최소 4 문자(각 문자의 인코딩된 바이트 값은 십진수 127보다 커야 한다)를 포함한 설명이 있어야 한다.

상세설명[5)] 파일 시작에 가까운 십진수 127보다 큰 인코딩된 바이트 값의 존재는 다양한 소프트웨어 툴들과 프로토콜에 의해 파일이 처리과정에서 보존되어야 할 8비트 이진(binary) 데이터를 포함하고 있음을 표시하기 위해 사용된다.

3) 태그는 의미 있는 키워드로써 서로간의 약속이 필요하다. 예를 들면 XML 스키마 또는 DTD는 태그의 의미를 정의한 것으로 이에 해당되는데 서로간의 약속된 태그가 아닌, 사용자들이 임의로 적용한 태그의 경우는 의미해석이 불가능하므로 이런 경우 PDF 문서의 태그 정보는 의미 없는 데이터일 뿐이다.

4) 이 문서보존포맷의 세부규격은 ISO 19005-1:2005의 기술규격을 포함한다. 국가기록원이 적용한 기술규격에 대한 설명은 [비고]에 기술된다.

5) 상세설명은 ISO 19005-1:2005의 [note]부분으로 기술규격 조항에 대한 상세설명이다.

6.1.3 파일 트레일러(File trailer)

파일 트레일러(trailer) 사전은 **ID** 키워드를 포함한다. **Encrypt** 키워드는 트레일러 사전에서 사용되어서는 안 된다. 하나의 임의의 EOL 표시 외에는 어떤 데이터도 마지막 EOL 표시 이후에 있을 수 없다.

파일 트레일러는 *PDF Reference*의 3.4.4, 3.4.5에서 묘사된 것 같이, PDF 파일의 마지막 트레일러 사전을 참조하거나, *PDF Reference* F.2에서 묘사된 것 같이 선형화된(linearized) PDF 파일의 경우 첫 페이지 트레일러 사전을 참조해야 한다.

선형화된 PDF 파일에서 **ID** 키워드는 첫 페이지 트레일러와 마지막 트레일러 사전 모두에 존재해야 하며 두 경우 모두의 키워드 값은 동일해야 한다.

상세설명 **Encrypt** 키워드의 명백한 금지는 암호화와 비밀번호-보호된 접근 허가를 금하는 암시적인 효과가 있다.

6.1.4 상호 참조표(Cross reference table)

상호 참조(Cross reference) 서브섹션 헤더(Subsection Header)에서 시작 객체 번호와 범위는 하나의 **SPACE**(20h) 문자로 구분된다.

xref 키워드와 교차참조 서브섹션 헤더는 하나의 EOL 표시로 구분된다.

상호 참조표에서 참조되지 않는 오프셋을 가진 객체는 이 표준의 모든 조건들로부터 제외된다.

6.1.5 문서 정보 사전(Document information dictionary)

문서 정보 사전은 규격에 적합한 파일에 정의될 수 있다. 만약 정의되었다면, 요소들은 6.7.3에 명시된 것 같이 유사한 XMP 메타데이터 속성들과 일치해야 한다.

6.1.6 문자열 객체(String objects)

16진법 문자열은 짝수의 비-공백 문자들로 구성한다. 각 문자는 **0** - **9**, **A** - **F** 또는 **a** - **f** 범위안의 문자이다.

6.1.7 스트림 객체(Stream objects)

stream 키워드 바로 다음에는 **CARRIAGE RETURN**(0Dh)과 **LINE FEED**(0Ah)의 세트 또는 하나의 **LINE FEED** 문자가 나타나야 한다. **endstream** 키워드 앞에는 EOL 문자가 있어야 한다.

스트림 사전에 명시된 **Length** 키의 값은 **stream** 키워드 다음의 **LINE FEED** 문자 이후부터 **endstream** 키워드 이전에 있는 EOL 문자 이전까지의 파일 바이트 숫자이다.

상세설명 1 이 조건들은 스트림 내용의 끝과 관련해 가능한 모호함을 제거한다.

스트림 객체 사전은 **F**, **FFilter**, 또는 **FDecodeParams** 키들을 포함하지 않는다.

상세설명 2 이 키들은 파일의 외부에 해당하는 문서 내용을 가리키는데 사용된다. 이 키들의 명백한 금지는 외부 의존과 보존 노력을 복잡하게 하는 외부 내용을 금하는 암시적인 효과가 있다.

6.1.8 간접 객체

객체 번호와 생성 번호는 하나의 공백 문자로 분리된다. 생성 번호와 **obj** 키워드는 하나의 공백 문자로 분리된다.
객체 번호와 **endobj** 키워드는 EOL 문자가 앞에 있어야 한다.
obj 그리고 **endobj** 키워드들 다음에는 EOL 문자가 있어야 한다.

6.1.9 선형화된 PDF

선형화(Linearization)는 허용되지만 규격에 적합한 판독기는 파일 안에 제공되는 모든 선형화 정보를 무시해야 한다.

6.1.10 필터

LZWDecode 필터는 허용되지 않는다.

상세설명 LZW 압축 알고리즘의 사용은 지적 재산권의 보호 대상이다.

비고 ASCIIHexDecode, ASCII85Decode, FlateDecode , RunLengthDecode, CCITTFaxDecode, JBIG2Decode, DCTDecode 중에서 FlateDecode, CCITTFaxDecode, DCTDecode 필터만을 사용한다.

6.1.11 내장 파일(Embedded files)

파일 명세 사전(File specification dictionary)에는, *PDF Reference* 3.10.2에 정의된 것과 같이, **EF** 키를 사용하지 않는다. 파일의 이름 사전(Name dictionary)에는 *PDF Reference* 3.6.3에 정의된 것과 같이 **EmbeddedFiles** 키를 사용하지 않는다.

비고 이 키들은 임의의 내용을 가진 파일들을 PDF 파일을 캡슐화하는데 사용된다. 이 키들의 명백한 금지는 외부 의존과 보존 노력을 복잡하게 하는 파일들의 내장을 금하는 암시적인 효과가 있다.

6.1.12 구현 한계

규격에 따르는 파일들은 *PDF Reference*의 표 C.1에 명시한 어떤 구성적 제한도 위배해서는 안 된다.

상세설명 규격에 따르는 파일들은 이 제한들을 준수함으로 가능한 많은 범위의 판독기들과 호환될 수 있다.

6.1.13 선택적 컨텐츠(Optional content)

문서 카탈로그 사전은 **OCProperties** 키를 포함 하지 않는다.

상세설명 *PDF* 1.5[4]에서는 허용된 **OCProperties** 키의 명백한 금지는 문서의 또 다른 렌더링(rendering)을 생성하는 옵션 내용을 금하는 암시적인 효과가 있다.

6.2 그래픽(Graphics)

6.2.1 일반 사항

규격에 적합한 파일과 판독기 모두에게 주어진 제한을 설명한다. 이 제한들은 글꼴들과 상호작용적 요소들을 수반하지 않는 그래픽적인 표현 이슈들을 다루기 위한 것이다.

6.2.2 출력 의도(Output Intent)

규격에 따르는 파일은 PDF/A-1 OutputIntent를 사용해서 표현하고자 하는 장치의 색 특성들을 명시할 수 있다. PDF/A-1 OutputIntent는 *PDF Reference* 9.10.4에 정의된 것 같이 파일의 **OutputIntents**열에 포함된 OutputIntent 사전이며, **S**키의 값으로는 *GTS_PDFA1*, **DestOutputProfile** 키의 값으로는 유효한 ICC 프로파일 스트림(Profile stream)을 가져야 한다.

만약 파일의 **OutputIntents** 열이 하나 이상의 항목을 포함하고 있다면, **DestOutputProfile**키를 포함한 모든 항목들은 키의 값으로 동일한 간접객체(Indirect object, 유효한 ICC 프로파일 스트림)를 가지고 있어야 한다.

비고 문서 전체에 하나의 OutputIntent를 사용하고, 그 값은 sRGBIccProfile

이다.

6.2.3 칼라스페이스(Colour spaces)

6.2.3.1 일반 사항

모든 색들은 장치와 독립적인 칼라 스페이스(colour space)를 사용해서 직접적인 방법으로 명시하거나, **OutputIntent**를 사용해서 간접적인 방법으로 명시하여야 한다. 규격에 따르는 파일은 ICCBased 칼라 스페이스, Uncalibrated 칼라 스페이스에서 제한한 내용을 제외하고 *PDF Reference*에 명시된 모든 칼라 스페이스를 사용할 수 있다.

상세설명 6.2.3 에서 설명된 것 같이 장치-독립적인 방식으로 색을 명시하는 것은 규격에 적합한 파일이 가정 및 외부의 정보에 의존하지 않고, 색채계(colorimetric) 정의에 기반한 예측할 수 있는 색 표현을 가능하게 한다. 또한 색채계 정의가 장치-의존적 색 데이터와 결합하는 방식도 제공한다.

비고 문서 전체에 걸쳐 DeviceRGBColorSpace를 사용하며, OutputIntent로 sRGBIccProfile을 사용한다.

6.2.3.2 ICCBased 칼라 스페이스(ICCBased colour spaces)

모든 **ICCBased** 칼라 스페이스는 *PDF Reference* 4.5에서 설명한 것 같이 ICC 프로파일 스트림으로 파일에 내장된다.

규격에 따르는 판독기는 **ICCBased** 칼라 스페이스를 ICC 프로파일 명세에서 명시한 것 같이 표현한다. 그리고 ICC 프로파일 스트림 사전에 명시된 **Alternate** 칼라 스페이스를 사용하지 않는다.

6.2.3.3 Uncalibrated 칼라 스페이스(Uncalibrated colour spaces)

규격에 따르는 파일은 **DeviceRGB** 또는 **DeviceCMYK** 칼라 스페이스 중 하

나를 사용할 수 있지만 둘 다의 사용은 안 된다. 만약 파일에 Uncalibrated 칼라 스페이스가 사용되었다면 그 파일은 6.2.2에 정의된 것 같이 PDF/A-1 OutputIntnet를 포함한다. 만약 파일이 **RGB** 칼라 스페이스를 사용하는 PDF/A-1 출력의도를 가지고 있다면 **DeviceRGB**가 사용될 수 있다. 만약 파일이 **CMYK** 칼라 스페이스를 사용하는 PDF/A-1 출력의도를 가지고 있다면 **DeviceCMYK**가 사용될 수 있다.

OutputIntent RGB 프로파일인 파일의 **DeviceGray** 칼라 명세 표현할 때에는, 규격에 적합한 판독기는 *PDF Reference* 6.2.1에서 설명된 방법으로 **DeviceGray** colour 명세를 **RGB**로 변환한다.

OutputIntent가 **CMYK** 프로파일인 파일의 **DeviceGray** colour 설계서를 표현할 때에는, 규격에 적합한 판독기는 *PDF Reference* 6.2.2에서 설명된 방법으로 **DeviceGray** 칼라 명세를 **DeviceCMYK**로 변환한다.

장치-의존적 칼라 스페이스에서 명시된 색들을 표현할 때에는, 규격에 적합한 판독기는 파일의 PDF/A-1 출력의도 사전을 *PDF Reference* 6.2.2에서 정의한 것 같이 원본 칼라 스페이스로 사용한다.

6.2.3.4 Separation과 DeviceN 칼라 스페이스(Separation and DeviceN colour spaces)

규격에 따르는 판독기는 **DeviceN** 또는 **Separation** 칼라 스페이스에 기반해서 칼라 스페이스를 표현할 때, 아래의 규칙을 따라야 한다.

- 만약 칼라 스페이스의 지명된 색소들이 모두 **Cyan**, **Magenta**, **Yellow**, **Black**리스트의 것들이라면, 파일은 **OutputIntent**를 가지고 있으며, 그 **OutputIntent**는 **CMYK**프로파일이다. 그러면 이 색소들은, 6.2.2에서 정의한 것 같이, PDF/A-1 출력의도 사전에서 명시된 대로 칼라 스페이스의 구성처럼 취급한다. 그리고 선택적 칼라 스페이스는 사용되지 말아야 한다.

- 또한 출력장치가 **Separation** 칼라 스페이스 또는 **DeviceN** 색소들을 지원하지 않으면, **Alternate** 칼라 스페이스가 사용된다.

Separation 또는 **DeviceN** 칼라 스페이스의 **Alternate** 칼라 스페이스는 6.2.3.2와 6.2.3.3에 명시된 칼라 스페이스에 대한 모든 제한들을 따라야 한다.

6.2.4 이미지(Images)

이미지 사전은 **Alternates** 키 또는 **OPI** 키를 포함하지 않는다.
만약 이미지 사전이 **Interpolate** 키를 포함한다면, 그 값은 *false*여야 한다.
Intent 키의 사용은 6.2.9의 규칙에 규격에 따라야 한다.

비고 이미지 데이터는 이미지 스트림이나 인라인 이미지 형태로 파일에 내장되어 있다.

6.2.5 폼 X객체(Form XObjects)

Form XObject 사전은 다음을 포함 하지 않는다.

- **OPI** 키
- 값이 *PS*인 **Subtype2** 키
- **PS** 키

상세설명 PDF 초기 버전에서 값이 *PS*인 **Subtype2** 키 그리고 **PS**키는 임의로 실행 가능한 포스트스크립트 코드 스트림을 정의하는데 사용되었는데, 이것은 신뢰하고 예측할 수 있는 표현을 방해하는 가능성이 있었다.

6.2.6 참조 X객체(Reference XObjects)

규격에 적합한 파일은 어떤 참조 X객체도 포함하지 않는다.

상세설명 참조 X객체는 외부 PDF 파일들의 임의적 문서 내용이며, 외부 의존도를 생성하여 보존 노력을 복잡하게 한다.

6.2.7 포스트스크립트 X객체(PostScript XObjects)

규격에 따르는 파일은 어떤 포스트스크립트 X객체도 포함하지 않는다.

상세설명 포스트스크립트 X객체는 임의적으로 실행할 수 있는 포스트스크립트 코드 스트림(PostScript code streams)을 포함하고 있어, 믿을 수 있으며 예측할 수 있는 표현을 방해할 가능성이 있다.

6.2.8 확장된 그래픽영역(Extended graphics state)

ExtGState 사전은 **TR** 키를 포함하지 않는다. ExtGState 사전은 값이 *Default* 이외의 **TR2**키를 포함하지 않는다. 규격에 적합한 판독기는 ExtGState 사전의 어떤 경우의 **HT**키도 무시할 수 있다.

RI 키의 사용은 6.2.9의 규칙에 따라야 한다.

6.2.9 렌더링 의도(Rendering intents)

렌더링 의도(Rendering intent)가 명시되면, 값은 아래와 같은 *PDF Reference* *RelativeColormetric, AbsoluteColorimetric, Perceptual* 또는 *Saturation* 에 정의된 4개의 값 중 하나여야 한다.

비고 기본 랜더링 의도의 값은 *RelativeColorimetric*이다.

6.2.10 컨텐츠 스트림(Content streams)

컨텐츠 스트림은 *PDF Reference*에 정의되지 않은 연산자는 포함하지 않는다. 그 연산자가 호환 연산자인 **BX/EX**에 의해 괄호로 묶여 구분되어 있어도 마찬가지이다.

ri 연산자의 사용은 6.2.9의 규칙에 따라야 한다.

상세설명 1 Contents 스트림은 페이지 설명을 위해 사용된다. 예) 페이지

객체의 컨텐츠 스트림, 폼 X객체의 스트림, 폼(Form) 필드 또는 위젯주석(Widget annotation)을 포함한 주석의 모양 스트림.

상세설명 2 PDF 포맷의 초기 버전에서 포스트스크립트 연산자 **PS**가 정의되었다. 이 연산자가 *PDF Reference*에서는 정의되지 않아서, 연산자의 사용은 6.2.10에 의해 암시적으로 금지되었다.

6.3 폰트(Fonts)

6.3.1 일반 사항

조건들이 의도하는 바는 규격에 따르는 파일의 문자 모양이 원래 생성된 대로 일치하는 것을 보장하고 문자 내용의 의미적 속성들의 복구를 허용하는 것이다.

6.3.2 폰트 유형(Font types)

규격에 적합한 파일에 사용된 모든 폰트들은 *PDF Reference* 5.5에 정의된 폰트 명세에 따라야 한다.

복합 마스터 폰트는 Type 1 폰트의 특별한 경우로 간주한다. Type 1에 관해 명백하게 명시된 모든 조건들은 복합 마스터 폰트에게 암시적으로 요구된다.

상세설명 모든 폰트의 규격 적합성 정도를 확보하는 것은 파일 생성 프로그램의 책임이다. 이 표준은 폰트의 규격 적합성 정도를 확인하는 방법을 규정하지 않는다.

비고 흔글 전용 HFT 폰트는 CIDFontType0, 윈도우 TrueType 폰트는 CIDFontType2로 폰트 포맷으로 변환해서 사용한다.

6.3.3 복합 폰트(Composite fonts)

6.3.3.1 일반

규격에 따르는 파일 안에서 참조되는 모든 복합 (Type 0) 폰트에서 CIDFont와 CMap 사전의 **CIDSystemInfo** 항목들은 *PDF Reference* 5.6.2에서 설명된 것 같이 호환되어야 한다. 다시 말해서, 그 폰트의 CIDSystemInfo사전의 **Registry**와 **Ordering** 스트링은 CMap사전의 **UserCMap**키의 값이 *Identity-H* 또는 *Idnetity-V*가 아닌 경우 동일해야 한다.

비고 1 CMap은 Identity-H 만을 사용하고, CIDSystemInfo 사전에는 항상 "/Registry (Adobe) /Ordering (Identity) /Suppliment 0"를 사용한다.

6.3.3.2 CIDFonts

모든 Type 2 **CIDFont**들은 CIDFont 사전에 **CIDToGIDMap** 항목을 포함해야 한다. 이것은 *PDF Reference*의 표 5.13에서 설명된 것 같이 CID들에서 그리프 인덱스나 이름 식별자로의 스트림 매핑을 위해 필요하다.

비고 2 CIDFontType2의 경우 "/CIDToGID /Identity"만을 사용한다.

6.3.3.3 CMaps

규격에 적합한 파일에서 사용된 모든 CMap들은, *Identity-H*와 *Identity-V*를 제외하고, *PDF Reference* 5.6.4에 설명된 것 같이 그 파일에 포함된다. 포함된 CMap들은, CMap 사전의 **WMode** 항목의 정수 값이 포함된 CMap 스트림의 **WMode** 값과 동일해야 한다.

비고 3 CMap은 Identity-H 만을 사용한다.

6.3.4 내장된 폰트 프로그램(Embedded font programs)

규격에 적합한 파일에 사용된 모든 폰트의 폰트프로그램들은, 폰트들이 오직 문자 표현 모드(mode) 3으로만 사용됐을 때를 제외하고는, *PDF Reference* 5.8에 정의된 것 같이, 그 파일에 포함되어야 한다. 폰트는 아래와 같은 정황

에서 폰트의 어느 그리프라도 참조되면 사용되었다고 간주된다.

- 페이지 객체의 **Contents** 스트림
- 폼 X객체의 스트림
- 폼 필드를 포함한 주석의 모양 스트림
- Type 3 폰트 글리프의 컨텐츠 스트림
- 타일(tiling) 패턴의 스트림

파일에 합법적으로 제한 없이, 전체 표현이 가능하게 포함될 수 있는 폰트만을 사용한다.

규격에 적합한 모든 판독기들은 로컬에 상주, 대치 또는 흉내낸 폰트들 대신에 파일에 포함된 폰트들을 표현에 사용한다.

비고 페이지 객체의 컨텐츠 스트림에서만 폰트를 사용하며, CIDFontType0에 대해 "/FontFile3"을, CIDFontType2에 대해 “FontFile2"를 사용하며, 모든 폰트는 파일에 내장되어 있다.

상세설명 1 *PDF Reference* 5.2.5에서 논의된 것 같이, 문자 표현 모드 3은 그리프가 선을 긋거나, 채우지 않고 그리프 범위로 사용되지 않는 것을 명시한다. 이 모드로만 참조되는 폰트는 표현되지 않으므로 포함되는 조건에서 제외된다.

상세설명 2 14가지 표준 Type 1 폰트들은 6.3.4의 조건들로부터 제외되지 않는다. Type 3 폰트들은 *PDF Reference* 6.3.4의 조건들로부터 제외된다. 이유는 Type 3 폰트들의 정의된 방법이 PDF 파일에, 포함되는 방식이 *PDF Reference* 5.8의 것과 다르지만, 폰트들이 항상 포함되게 보증하기 때문이다.

상세설명 3 폰트 프로그램 메타데이터의 조건들은 6.7.10에 설명된다.

상세설명 4 6.3.5에서 명시된 것 같이, 폰트 부분집합(Font subset)은 포함된 폰트 프로그램이 파일에서 참조되는 모든 문자들의 그리프 정의를 제공하는 한 수용이 가능하다. 폰트 프로그램을 포함시키는

것은 규격에 적합한 모든 판독기가 모든 그리프를 일시적일 수 있는 외부 자원을 참조하지 않고 출판된 원본 그대로의 방법으로 재생하는 것을 허용한다.

상세설명 5 이 규격의 이 장은 폰트 저작권자의 특별한 동의에 합법성을 의존하는 그런 폰트들의 포함 배제를 설명하고 있다.

6.3.5 폰트 부분집합(Font subsets)

6.3.4에서 언급한 것 같이, 포함된 폰트 프로그램들은 규격에 따르는 파일이 표현하기 위해 참조하는 모든 그리프들을 정의한다. Type 0 **CIDFont**와 Type 1 그리고 TrueType 폰트 부분집합은, *PDF Reference* 5.5.3 에서 설명한 것 같이 포함된 폰트 프로그램이 모든 적절한 그리프들을 정의한다면 사용될 수 있다.

규격에 적합한 파일에서 참조되는 모든 Type 1 폰트 부분집합은, *PDF Reference* 표 5.18에서 설명한 것 같이, 폰트 기술 사전(Font descriptor dictionary)에 폰트 부분집합에서 정의된 문자명을 목록화하는 **CharSet** 문자열을 포함한다.

규격에 따르는 파일에서 참조되는 모든 **CIDFont** 부분집합은 *PDF Reference* 표 5.20에서 설명한 것 같이, 폰트 기술어 사전(font descriptor dictionary)에 포함된 **CIDFont** 파일에 어느 CID들이 존재하는지 확인하는 **CIDSet** 스트림을 포함한다.

비고 사용된 모든 그리프에 대한 폰트 부분집합을 정의하여 파일에 내장하며, 윈도우 GDI 방식으로 파일을 변환함으로 포스트스크립트 폰트인 Type 1 폰트를 사용하지 않는다. 또한 CIDSet 스트림을 생성해서 Font Descriptor 사전에 포함한다.

상세설명 폰트 부분집합의 사용은 규격에 따르는 파일 크기를 상당히 줄이는 가능성을 허용한다.

6.3.6 폰트 메트릭(Font metrics)

규격에 적합한 파일에 포함된 모든 폰트에서, 폰트 사전의 **Widths** 항목에 저장된 그리프 너비 정보와 폰트 프로그램에 포함된 그리프 너비 정보는 일치해야 한다.

비고 국가기록원 문서보존포맷은 폰트 사전에서 사용된 모든 그리프의 너비 정보를 W키를 통해 정의한다.

상세설명 이 조건은 주어진 판독기가 **Widths**의 메트릭 또는 폰트 프로그램의 것 중 어느 것을 사용하는 것과 관계없이 예측할 수 있는 표현을 보장하기 위해 필요하다.

6.3.7 문자 인코딩(Character encodings)

모든 non-symbolic 트루타입 폰트는 폰트 사전의 **Encoding** 항목의 값으로 *MacRomanEncoding* 또는 *WinAnsiEncoding*을 명시해야 한다. 모든 symbolic Truetype 폰트는 폰트 사전에 **Encoding** 항목을 명시하지 않는다. 그리고 폰트 프로그램의 "cmap" 표들은 정확히 하나의 인코딩을 포함한다.

상세설명 이 조건은 *PDF Reference* 5.5.5에서 설명한 제안된 지침을 규범으로 만든다.

6.3.8 유니코드 문자 맵(Unicode character maps)

6.3.8은 A 수준 규격 적합성을 준수하는 파일들에게만 적용된다. 6.3.8의 조건들은 B 수준 규격 적합성에서는 무시될 수 있다.

폰트 사전은 아래의 세 가지 상황 중에 속하지 않는 한, *PDF Reference* 5.9에서 설명된 것 같이 문자 코드들을 유니코드 값[5]들에 할당하는 CMap 스트림 객체가 값인 **ToUnicode** 항목을 포함한다.

- 미리 정의된 인코딩인 **MacRomanEncoding**, **MacExpertEncoding** 또는

WinAnsiEncoding을 사용, 또는 미리 정의된 **Identity-H** 또는 **Idnetity-V** CMap들을 사용하는 폰트들

- *PDF Reference* 부록 D에서 정의된 것 같이, 문자 이름들이 Adobe standard Latin 문자셋 또는 symbol 폰트의 지정된 문자들의 집합(set)을 따르는 Type 1 폰트들

- 파생된 **CIDFont**가 **Adobe-GB1**, **Adobe-CNS1**, **Adobe-Japan1** 또는 **Adobe-Korea1** 문자 셋을 사용하는 Type 0 폰트들

참고 사용된 모든 그리프에 대한 유니코드 매핑정보를 폰트 사전의 ToUnicode 키 값으로 정의하여 파일에 포함한다.

상세설명 유니코드 매핑은 파일에서 참조된 모든 문자의 의미적 특성들의 검색을 허용한다.

6.4 투명도(Transparency)

만약 **ExtGState** 또는 **XObject** 사전 안에 **SMask**키가 있으면, 값은 *None* 이다.

*투명도*의 값인 **S** 키가 있는 **Group** 객체는 **XObject** 폼에 포함되지 않는다.

만약 아래의 키들이 **ExtGState** 객체에 존재한다면, 값들은 다음과 같다:

- **BM** *Normal* 또는 *Compatible*
- **CA** *1.0*
- **ca** *1.0*

비고 국가기록원의 문서보존포맷은 ExtGState 객체의 값으로 "/BM /Normal /CA 1.0 /ca 1.0" 만을 사용한다.

상세설명 이 조항들은 규격에 따르는 파일에서 투명도의 사용을 금지한다.

부분적으로 투명한 그래픽의 시각적 효과는 전에 렌더링된 데이터 또는 플래튼(flatten) 벡터 객체들을 포함한, *PDF Reference* 투명도 키들의 사용 외에, 다른 기술을 사용해서 얻을 수 있다. 이런 기술의 사용은 파일이 규격에 따르는 것은 방해하지 않는다.

6.5 주석(Annotations)

6.5.1 일반 사항

이 표준의 이 항에 의해 수정된 *PDF Reference*에 정의된 표현 행동에 더하여, 규격에 적합한 상호작용적인 판독기는 주석 사전의 **Contents** 키의 값들을 표시하기 위한 방법을 제공한다.

상세설명 이 항은 이 기능의 조건을 구현하기 위해 상호작용적인 판독기가 사용할 수도 있는 특별한 행동 또는 기술적 상세 구현 내용을 규정하지 않는다.

6.5.2 주석 유형(Annotation types)

*PDF Reference*에 정의되지 않은 주석 유형은 허용되지 않는다. 추가로, **FileAttachment**, **Sound** 그리고 **Movie** 유형도 허용되지 않는다.

상세설명 멀티미디어 내용에 대한 지원은 이 규격에서 다루지 않는다.

6.5.3 주석 사전(Annotation dictionaries)

주석 사전은 *1.0* 이외의 값을 가진 **CA** 키를 포함하지 않는다.

주석 사전은 **F** 키를 포함한다. **F** 키의 **Print**를 표시하는 비트는 *1*로 지정되며 **Hidden**, **Invisible** 그리고 **NoView**를 표시하는 비트들은 *0*으로 지정한다.

문자 주석은 **NoZoom**과 **NoRotate**을 표시하는 **F** 키의 비트를 *1*로 지정한다.

상세설명 1 주석 플래그(flag)들에 대한 제한은 숨겨진 또는 보이지만 출력할 수 없는 주석들의 사용을 방지한다. **NoZoom**과 **NoRotateflag**들은 허용되어서, 일반적으로 사용되는 문자 주석 종류와 같은 행동을 가지고 있는 주석 종류들의 사용을 허용한다. 정의에 의하면, 문자 주석은, *PDF Reference* 8.4.5에 설명된 것 같이, 플래그가 지정되지 않아도 **NoZoom**과 **NoRotate**의 행동을 보인다. 이 플래그들의 값을 명백하게 지정하는 것은 주석 사전 값과 판독기 행동과의 사이에 있을 수 있는 가능한 모호함을 없앤다.

주석 사전은 PDF/A-1 출력의도 사전의 **DestOutputProfile**의 칼라 스페이스, 6.2.2에서 정의 된 것과 같이, RGB가 아닌 한, **C** 배열 또는 **IC** 배열을 포함하지 않는다.

상세설명 2 이 조항들은 주석에 모양 스트림(Appearance stream)외의 방법으로 사용된 장치 칼라 스페이스들이 PDF/A-1 출력의도에 의해 간접적으로 정의되는 것을 보장하려는 의도이다.

만약 주석 사전이 **AP** 키를 포함하다면, 값으로 정의하는 주석의 모양 사전은 **N** 키만을 포함하며 **N** 키의 값은 주석의 모양을 정의하는 스트림이다.

상세설명 3 6.5.3의 모든 조항들은 폼 필드에 사용된 **Widget** 유형을 포함한 모든 주석 유형에 적용된다.

6.6 행위(Actions)

6.6.1 일반 사항

Launch, **Sound**, **Movie**, **ResetForm**, **ImportData** 그리고 자바스크립트 행위들은 허용되지 않는다. 추가로, 지원 중지된 **set-state** 그리고 **no-op** 행위들도 허용되지 않는다. **NextPage**, **PrevPage**, **FirstPage** 그리고 **LastPage**외의 지정된 action들은 허용되지 않는다. 허용된 4개의 지정된 행위들에 대한

반응으로, 규격에 적합한 상호작용적 판독기는 *PDF Reference* 표 8.45에 설명된 적절한 행위를 실행한다.

상호작용적 폼 필드(Interactive form fields)는 어떤 종류의 행위도 실행하지 않는다.

상세설명 1 멀티미디어 내용에 대한 지원은 이 규격의 범위가 아니다. **ResetForm** 행위는 표현되는 폼의 모양을 변화한다. **ImportData** 행위는 외부 파일로부터 폼 데이터를 가져온다. **JavaScript** 행위들은 신뢰하고 예측할 수 있는 표현을 방해할 가능성이 있는 임의적 실행 코드를 허용한다.

상세설명 2 상호작용적 폼 필드의 추가적 조건들은 6.9에 명시된다.

6.6.2 트리거 이벤트(Trigger events)

위젯 주석 사전 또는 필드 사전은 추가행위 사전으로 **AA**항목을 포함하지 않는다. 문서 카탈로그 사전은 추가행위 사전으로 **AA**항목을 포함하지 않는다.

상세설명 이 추가행위 사전들은 임의적 자바스크립트 행위들을 정의한다. **AA**항목의 명백한 금지는 외부 의존을 생성하고 그리고 보존 노력을 복잡하게 하는 자바스크립트 행위들을 허용하지 않는 암시적인 효과가 있다.

6.6.3 하이퍼텍스트 링크(Hypertext links)

규격에 적합한 상호작용적 판독기는 하이퍼링크들을 실행하지 않게 만드는 것을 선택할 수 있다. 그러나 *PDF Reference*에서 정의된 표현 행동에 더하여 이 항에서 수정된 것과 같이, 판독기들은 **GoToR** 행위 사전의 **F**와 **D** 키들, **URI** 행위 사전의 **URI** 키, 그리고 **SubmitForm** 행위 사전의 **F**키를 표시하는 방법을 제공해야 한다.

상세설명 하이퍼링크들이 실행의 쓰레드를 상호작용적 판독기의 관리 밖으로 건네기 때문에, 이 하위조항은 상호작용적 판독기가 하이퍼링크들을 실행하지 않도록 선택하는 것을 허용한다. 규격에 따르는 파일들의 전체 정보 내용의 기록 공개의 목적을 위해, 상호작용적 판독기가 모든 하이퍼링크의 목적지를 공개하는 방법을 제공하는 것이 중요하다.

6.7 메타데이터(Metadata)

6.7.1 일반 사항

규격에 따르는 파일들의 메타데이터를 위한 조건들을 명시한다. 메타데이터는 파일의 생애주기 동안에 파일의 효과적인 관리를 위해 필수적이다. 파일은 식별자 그리고 설명뿐 아니라 적절한 기술적 그리고 관리적 요소들을 설명하기 위해 메타데이터에 의존한다. 이 결과로 규격에 적합한 파일의 생성 프로그램들은 이 규격의 외부에서 정의한 다양한 도메인, 특정한 메타데이터 조건들에 따라야할 수도 있다. 이 장의 규격은 다양한 메타데이터 조건들을 지원하는 구조화되고 일관된 틀을 약술한다.

6.7.2 속성(Properties)

규격에 적합한 파일의 문서 카탈로그 사전은 **Metadata** 키를 포함한다. 메타데이터 스트림을 형성하는 그 키의 값은 *XMP Specification*에 따른다. 파일에 포함된 모든 메타데이터 속성들은, 6.7.3에서 정의된 것 같이, XMP 유사한 항목이 없는 문서 정보 사전 항목들을 제외하고는 XMP 형식이어야 한다. XMP 형식으로 명시된 속성들은 *XMP Specification* 4에 정의된 미리 정의된 스키마 또는 *XMP Specification* 4와 6.7.8에 따르는 확장 스키마를 사용한다. 메타데이터 객체 스트림 사전은 **Filter** 키를 포함하지 않는다.

비고 문서 카탈로그 사전에 Metadata 키를 포함하며, 메타데이터 스트림은 *XMP Specification*를 준수한다.

상세설명 1 **Filter** 키의 명백한 금지는 XMP 메타데이터 스트림의 내용을

PDF를 인식하지 못하는 툴들에게 보여줄 수 있는 일반 텍스트로 보존하는 암시적인 효과가 있다.

상세설명 2 확장 스키마는 *XMP Specification*에 정의되지 않은 모든 XMP 스키마이다.

6.7.3 문서 정보 사전(Document information dictionary)

문서 정보 사전은 규격에 따르는 파일에 있을 수 있다. 만약 없다면, 미리 정의된 XMP 스키마에 유사한 속성들을 가지고 있는 모든 항목들은, 표 1에 정의된 것 같이, 동등한 값들과 함께 XMP 형식으로 파일에 포함된다. 표 1에 없는 모든 문서 정보 사전의 항목은 미리 정의된 XMP 스키마 속성을 사용해서 포함되지 않는다.

상세설명 1 규격에 적합한 파일에 문서 정보 사전이 허용되기 때문에, 하나의 파일이 PDF/A-1 (ISO 19005-1) 그리고 PDF/X (ISO 15930-4, ISO 15930-6[6]) 등 모든 규격에 만족하는 것이 가능하다.

표1 - 문서 정보 사전과 XMP 속성 비교

문서 정보 사전		XMP	
항목	PDF 종류	속성	XMP 종류
Title	*text string*	**dc:title**	*Text*
Author	*text string*	**dc:creator**	*seq Text*
Subject	*text string*	**dc:subject**	*Text*
Keywords	*text string*	**pdf:keywords**	*Text*
Creator	*text string*	**xmp:CreatorTool**	*Text*
Producer	*text string*	**pdf:Producer**	*Text*
CreationDate	*date*	**xmp:CreateDate**	*Date*
ModDate	*date*	**xmp:ModifyDate**	*Date*

비고

접두사 **dc**의 XML 네임스페이스 URI는 <http://purl.org/dc/elements/1.1/>;
접두사 **pdf**의 네임스페이스는 <http://ns.adobe.com/pdf/1.3/>; 그리고
접두사 **xmp**의 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/>

문서 정보 사전 항목들의 값과 유사한 XMP 속성들의 값은 동일하다. PDF *Text string* 유형에서 XMP *Text* 유형으로 매핑되는 속성들에 값의 동일성은 문자들의 numeric ISO/IEC 10646-1 코드 포인트를 비교할 때 인코딩에 의존하지 않는 문자-대-문자에 기반을 둔다.

비고 문서 정보 사전에 Creator, Producer, CreationDate, ModDate를 정의하여 포함하며, 그 값은 XMP 메타데이터 항목과 일치한다.

상세설명 2 문서 정보 사전 항목들의 값과 유사한 XMP 속성들의 값 사이의 동일성의 명백한 조건은 속성 값의 명백한 해석을 제공하는 암시적인 효과가 있다.

만약 **dc:creator** 속성이 XMP 메타데이타에 있다면, 이 속성은 길이가 하나인 정돈된 *Text* 배열로 표현된다. 배열 하나의 항목은 하나 이상의 이름들을 포함한다. **Author**와 **dc:creator** 사이의 동일성은 문자들의 ISO/IEC 10646-1 코드 포인트(code point)를 비교할 때 인코딩에 의존하지 않는 문자-대-문자에 기반을 둔다.

보기 1 문서 정보 사전 항목:

/Author (Peter, Paul and Mary)

는 XMP 속성과 동일하다:

```
<dc:creator>
 <rdf:Seq>
  <rdf.:li>Peter, Paul, and Mary</rdf:li>
 </rdf:Seq>
</dc:creator>
```

Date 속성들은 년, 달, 일, 시, 분, 초 로 나뉘는 시간적 요소들의 변하는-길이의 연속(sequence)의 형식으로 구성된다. *PDF Reference* 3.8.2에서 정의한 *PDF date* 유형 그리고 *Date and Time Formats*에서 정의한 XMP Date 종류 사이를 매핑하는 속성들은, 값의 동일성이, 협정 세계시(Coordinated Universal Time, UTC)에 비교하여, 컴포넌트(component)-대-컴포넌트에 기반한다.

보기 2 문서 정보 사전 항목들:
/CreationDate (D:20040402)
/ModDate (D:20040408091132-05‘00’)
는 XMP 속성들과 동일하다:

<xmp:CreateDate>2004-04-02</xmp:CreateDate>
<xmp:ModifyDate>2004-04-08T14:11:32Z</xmp:ModifyDate>

6.7.4 정규화(Normalization)

모든 XMP 스키마는 속성들에 적용될 수 있는 정규화 규칙을 정의한다. 정규화 규칙들을 제공하는 스키마들에 정의된 모든 메타데이터 속성들은, 규격에 적합한 판독기들이 메타데이터의 교환을 쉽게 하고 일관된 해석을 위해 스키마들에 의해 정의된 정규화된 방법으로 속성 값들을 입력, 저장 그리고 유지한다.

6.7.5 XMP 헤더(XMP header)

bytes와 **encoding** 속성들은 XMP 패킷(Packet)의 헤더에 사용되지 않는다.

상세설명 bytes와 **encoding** 속성 모두 *XMP Specification*에서 제외되었다.

6.7.6 파일 식별자(File identifiers)

규격에 따르는 파일은 파일의 특성, 분류 및 식별을 위해 하나 이상의 메타데이터 속성들을 가진다. 이 규격의 이 장은 어떤 특정한 식별 방법도 요구하지 않는다. 식별자들은 국제 표준 도서 번호(International Standard Book Number, ISBN)[7] 또는 디지털 객체 식별자(Digital Object Identifier, DOI)와 같이 외부에 기반을 둔다. 또는 전 세계적 단일 식별자/범용 단일 식별자(Globally Unique Identifier/Universally Unique Idnetifier, GUID/UUID) 또는 작업 공정 중에 할당한 또 다른 명칭과 같이 내부에도 기반을 둔다. 식별자들은 **xmp:Identifier** 속성의 사용, **xmpMM:DocumentID**,

xmpMM:VesionID 그리고 **xmpMM:RenditionClass** 속성의 사용, 또는 확장 스키마의 속성들의 사용을 통해 포함된다. XMP 조건들과 이 규격의 이 장에 따르는 한, 어떤 식별 시스템도 사용될 수 있다.

만약 규격에 따르는 파일이 어떤 방식으로라도 변화됐다면, 6.7.7에서 설명한 것처럼 **xmpMM:History** 항목이 추가된 것 뿐 이더라도 변화하는 식별자 부분을 반영하기 위해 *PDF Reference* 9.3에서 설명된 것 같이 파일 트레일러 사전의 **ID** 키가 수정되어야 한다.

상세설명 xmp 접두사의 XML 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/> 이다, **xmpMM** 접두사의 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/mm/> 이다.

6.7.7 파일 출처 정보(File provenance information)

규격에 적합한 파일을 생성, 변형, 또는 설명하기 위한 고급 사용자 행위를 설명하기 위해서, 각각의 행위들은 **xmpMM:History** 속성에 기록되어야 한다. 기록되는 각각의 행위들은 아래와 같다.

- **action**, **parameters**, **when** 필드는 명시되어야 한다.
- **softwareAgent** 필드는 명시하여야 한다.
- **instanceID** 필드는 명시하지 않는다.

상세설명 1 접두사 **xmpMM**의 XML 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/mm/> 이다.

상세설명 2 특정한 감사 조건들이 있는 응용프로그램들은 미리 정의된 XMP 스키마들에 정의된 것들 이외의 행위의 추가적 종류 또는 행위에 대한 추가적 상세정보를 기록할 필요가 있을 수 있다. 행위의 추가적 종류의 예제는 손실압축(downsampling) 또는 폰트 대치와 같이 문서의 모양을 변화시키는 것을 포함한다. 추가적 상세정보의 예제는 행위를 실행하는 사람에 관한 식별자 또는 행위가 일어난 환경을 포함한다.

종이, 마이크로폼 또는 전자 파일등의 원천 소스들이 규격에 따르는 파일들

로 변형됐을 경우, **xmpMM:History**는 모든 high-level 처리 (예. PDF 1.4에서 PDF/A-1로의 변형); 파일 내용 또는 기능에 대한 변경 (예. 포함된 자바스크립트 그리고 오디오 객체들은 유지되지 않았다); 이미 존재하는 메타데이터의 취급 (예. XMP로 변환된 모든 문서정보 사전 값들); 그리고 이외의 모든 변형 과정의 중요한 면들을 설명한다.

모든 규격에 적합한 파일들은, 원래의 파일로 생성되었거나 종이, 마이크로폼(Microform) 또는 다른 전자 포맷 소스의 변환으로 생성되었던 간에, **xmpMM:History**는 수반하는 모든 high-level 워크플로우 처리들 [예. 활동(activities) 그리고 핸드오프(handoffs)의 설명]; 파일 처리를 관리하는 정책들의 열거 (예. 파일들이 모아지고, 처리되고, 사용되는 공식 명령의 제목들); 소프트웨어 툴들의 이름과 버전; 파일의 생성과 사용의 배경을 알려주기에 필요한 다른 요소들을 설명한다.

XMP 메타데이터 속성들이 파일의 생애주기를 따라 이동하는 동안 수정 또는 삭제된 경우, **xmpMM:History**는 속성들의 이름과 속성들의 이전 값들을 명시하는 **parameters** 필드가 있는 항목들을 포함하는 것으로 그 변화들을 설명한다. 이 권고는 **xmpMM:History** 자신을 제외한 모든 메타데이터 속성들에 적용된다. 만약 메타데이터 속성이 삭제되었다면, **xmpMM:History**의 항목의 **action** 필드의 값은 *pdfa:deleted* 이다.

6.7.8 확장 스키마(Extension schemas)

규격에 적합한 파일에 사용된 모든 확장 스키마들은 6.7.2에서 정의된 메타데이터 스트림(Metadata stream)을 가지고 있는 파일 내에 포함한 설명을 가지고 있어야 한다. 이 설명들은 이 조항에 정의된 PDF/A 확장 스키마 설명 스키마를 사용하여 명시한다.

상세설명 1 확장 스키마는 *XMP Specification*에 정의되지 않은 모든 XMP 스키마이다.

<표2>에 정의된 확장 스키마 설명 스키마는 네임스페이스 URI <http://www.aiim.org/pdfa/ns/schema>를 사용한다. 필요한 스키마 네임스페이스 접두사는 pdfaSchema 이다.

상세설명 2 W3C XML 네임스페이스 권고[8] 에 의하면, 네임스페이스URI들은 식별 목적만을 위한 것이며 실행가능한(actionable) 링크여야 할 필요는 없다. 이 규격의 이 장에서 XMP 확장 스키마의 네임스페이스 URI들은 실행가능한(actionable) 링크가 없다. 이 링크들의 값을 알아내거나 또는 따라 가려는 시도는 유효한 웹 페이지의 결과를 주지 않는다.

표2 - 확장 스키마 설명스키마

속성	값 종류	카테고리	설명
pdfaSchema:schema	*Text*	외부	스키마의 옵션 설명
pdfaSchema:namespaceURI	*URI*	외부	스키마 네임스페이스 URI
pdfaSchema:prefix	*Text*	외부	우선의(Preferred) 스키마 네이미스페이스 pre-
pdfaSchema:property	*seq Property*	내부	스키마 속성들의 설명
pdfaSchema:valueType	*seq ValueType*	내부	schema-specific 값 종류의 설명

표 3에 정의된 **Property** 타입은 스키마 속성의 설명을 포함한 XMP 구조이다. 필드 네임스페이스 URI는<http://www.aiim.org/pdfa/ns/property>이다. 필요한 필드 네임스페이스 접두사는 **pdfaProperty**이다.

표3 - PDF/A 속성 종류 스키마

필드 이름	값 종류	설명
pdfaProperty:name	문자	속성 이름
pdfaProperty:valueType	자유 선택 문자	XMP Specification 4로 부터의 값 종류, 또는 포함된 PDF/A 값 종류 확장 스키마
pdfaProperty:category	Closed Choice of Text	속성 카테고리: 내부 또는 외부
pdfaProperty:description	문자	속성 설명

pdfaProperty:valueType의 선호하는 값은 *XMP Specification* 2004, 4에서 정의한 non-deprecated 속성 값 종류들이다. 배열 종류들은 컨테이너 종류(***alt, bag*** 또는 ***seq***)가 앞에 있으며, 기본 종류와 하나의 공백 문자로 분리된다.

표 4에 정의된 **ValueType** 유형은 포함된 확장 스키마에 의해 사용되지만 *XMP Specification* 4에 정의되지 않은 모든 속성 값 유형의 설명을 포함한다. 필드 네임스페이스 URI는 <http://www.aiim.org/pdfa/ns/type> 이다. 필요한 네임스페이스 접두사는 **pdfaType**이다.

표4 - PDF/A 값 유형 스키마

필드 이름	값 종류	설명
pdfaType:type	문자	Property value type name
pdfaType:=namespaceURI	URI	Property valute type field namespace URI
pdfaType:prefix	문자	Preferred value type field namespace prefix
pdfaType:description	문자	property value type의 설명
pdfaType:field	seq Field	structured field의 옵션 설명

표 5에 정의된 **Field** 유형은 속성 값 유형 필드의 설명을 포함한 XMP 구조이다. 필드 네임스페이스 URI는 <http://www.aiim.org/pdfa/ns/field> 이다. 필요한 네임스페이스 접두사는 **pdfaField**이다.

표5 - PDF/A 필드 스키마

필드 이름	값 종류	설명
pdfaField:name	문자	필드 이름
pdfaField:valueType	자유 선택 문자	XMP Specification 2004, 4로 부터의 필드 값 종류, 또는 포함된 PDF/A 값 종류 확장 스키마
pdfaField:description	문자	필드 설명

6.7.9 검증(Validation)

모든 XMP 패킷의 모든 내용은 *Extensible Markup Language(XML) 1.0(Third Edition)*, 2.1, 그리고 *RDF/XML Syntax Specification(Revised)*, 7에서 정의한 것 같이 체계화 된다. 가능하다면, 파일 생성 프로그램이 규격에 따르는 파일을 생성 또는 다시 저장할 때, 그 파일의 XMP 패킷들의 모든 내용은 확인되어야 한다.

6.7.10 폰트 메타데이터(Font metadata)

포함된 모든 Type 0, Type 1, 또는 TrueType 폰트 프로그램들은, 포함된 폰트 파일 스트림 사전에 XMP 메타데이터 스트림이 값인 **Metadata** 항목을 포함한다. 다음의 XMP 메타데이터 항목들을 제공해야 한다: **xmp:TItle**, 폰트의 font descriptor 사전으로 부터 **FontName**키의 값을 제공한다; **xmpRights:Copyright**, 저작권 문구를 제공한다; **xmpRights:Marked**, Boolean 값 *true*; **xmpRights:Owner**, 폰트의 합법적 소유자를 표현한다; **xmpRights:UsageTerms**, 폰트가 사용되는 라이센스 조건 문구를 제공한다. 파일 생성 프로그램의 판단에 따라 추가적인 XMP 메타데이터가 포함될 수 있다.

상세설명 1 폰트 저작권 정보는 폰트 저작권 소유자의 지적 재산권의 정체성과 범위를 보존하는데 도움이 된다. 많은 폰트들이 저작권 그리고 라이센스 조건을 폰트 안에 포함하지만, 이것은 통합된 방식이 아니다. 그러므로 규격에 적합한 파일에 저작권 문구의 명백한 표현이 요구된다. 이것이 불필요한 정보일 수 도 있지만, 미래의 시스템이 폰트 프로그램의 특정한 내부 구조를 분석하는 기능을 가져야 하는 필요성을 미리 방지한다.

상세설명 2 **xmp** 접두사의 XML 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/>이다; **xmpRights** 접두사의 네임스페이스 URI는 <http://ns.adobe.com/xap/1.0/rights/>이다.

6.7.11 버전과 적합성 수준 식별자

파일의 PDF/A 버전과 적합성 수준은 이 하위조항에 정의된 PDF/A 식별자 확장 스키마를 사용해서 명시된다.

표6에 정의된 식별자 스키마는 네임스페이스 URI <http://www.aiim.org/pdfa/ns/id>를 사용한다. 요구되는 스키마 네임스페이스 접두사는 **pdfaid**이다.

표6 - PDF/A 식별자 스키마

속성	값 종류	카타고리	설명
pdfaid:part	자유선택 정수	내부	PDF/A 버전 식별자
pdfaid:amd	자유선택 문자	내부	선택의 PDF/A 수정 식별자
pdfaid:conformance	Closed Choice of Text	내부	PDF/A 적합성수준: A 또는 B

pdfaid:part의 값은 파일이 따르는 규격의 부(part) 번호이다. 만약 파일이 수정된 부에 정의된 규격의 버전에 따른다면, **pdfaid:amd**의 값은 콜론(:)으로 분리된 수정 번호와 연도이다.

A 수준 적합성 파일은 **pdfaid:conformance**의 값을 *A*로 명시한다. B 수준 적합성 파일은 **pdfaid:conformance**의 값을 *B*로 명시한다.

pdfaid:part, **pdfaid:amd**, 그리고 **pdfaid:confomance** 속성들의 값들은 자체로는 이 규격에 적합한 지를 결정하지 않는다. 적합성의 실제 결정은 조항 5에서 명시된 것 같이 실행된다.

비고 국가기록원 문서보존포맷은 아래와 같은 XMP 정보를 포함하고 있다.

```
<rdf:Description rdf:about=''
xmlns:pdfaid='http://www.aiim.org/pdfa/ns/id/'>
  <pdfaid:part>1</pdfaid:part>
  <pdfaid:conformance>B</pdfaid:conformance>
</rdf:Description>
```

6.8 논리적 구조(Logical structure)

6.8.1 일반사항

A 수준 적합성에 준수하는 파일들에게만 적용된다. B 수준 적합성은 6.8의 조건들을 무시할 수 있다.

아래의 조건들의 의도는 규격에 따르는 파일의 문자 내용을 저장된 언어 자연 읽기 순서에 정의된 단어들의 순서로의 복귀를 보증하는 것이다. 유사하게, 각 단어의 글자들도 본래의 읽기 순서로 복귀되는 것을 보증한다. 더욱이 이 조건들은 문서의 논리적 구조와 관련된 고수준의 의미 정보 복귀를 허용한다.

PDF/A-1 생성 프로그램은 규격의 적합성을 위한 목적만으로 소스 자료에 명백히 또는 암시적으로 존재하지 않는 구조 또는 의미 정보를 추가하지 않는다. 이런 정보의 예로는 구조 계층, 자연 언어 설계서, 대안의 설명, 비-문자 주석, 대체 문자 그리고 약자와 두문자어의 확장이다.

상세설명 파일의 생성 프로그램은 구조 또는 의미 정보를 적절한 확인없이 자동 처리로 생성하는 것은 권하지 않는다.

6.8.2 태그 붙은 PDF(Tagged PDF)

6.8.2.1 일반

A 수준 적합성 파일은 *PDF Reference* 9.7에서 태그 붙은 PDF(tagged PDF)을 위해 지정한 모든 조건들을 준수해야 한다.

상세설명 태그 붙은 PDF는 문서 내용의 논리적 구조적 면을 명백히 선언하고 설명하는 규정을 정의한다.

6.8.2.2 표시 정보 사전(Mark information dictionary)

문서 카탈로그 사전은 **MarkInfo** 사전을 포함한다. 이 사전의 하나뿐인 항목, **Marked**는 *true* 값을 가진다.

상세설명 이 지정은 파일이 태그 붙은 PDF 규정에 적합하다는 것을 나타낸다.

6.8.3 아티팩트(Artifacts)

6.8.3.1 일반

난외표제(running head) 또는 페이지 번호 등의 페이지 매김 기능들, 각주 규칙 또는 배경 화면 등의 편집 기능들, 그리고 절단선 표시 또는 색 막대 등의 생산 지원 기능들은 *PDF Reference* 9.7.2에서 설명된 것과 같이, 각각 페이지 매김, 레이아웃(layout), 그리고 페이지 아티팩트(Artifacts)라 명시한다.

6.8.3.2 단어 분할(Word breaks)

보통 간격 문자를 사용해서 단어 분할을 나타내는 언어들 그리고 스크립트 시스템들은, 다음의 추가적 제한들이 적용된다.

문자열 보이기 안에서, 단어 분할은 문자열 보이기의 모든 단어들 사이에 하나 이상의 간격 문자의 존재로 명백히 표시된다. 만약 단어가 문자열 보이기 경계의 끝에 있으면, 하나 이상의 간격 문자가 문자열 보이기 끝에 추가된다. 하나의 단어가 하나 이상의 문자열 보이기에 걸칠 수 있는 것에 주의한다. 단어 분할(Word breaks)은 하나 이상의 간격 문자들의 명백한 존재만으로 표시되며, 문자열 보이기의 경계들에 의해 표시되지 않는다. 단어 분할을 표시하기 위한 목적에서는, 둘 이상의 연속 간격 문자들은 의미적으로 하나의 간격 문자와 동일하다.

6.8.3.3 구조 계층(Structure hierarchy)

규격에 따르는 파일의 논리적 구조는 *PDF Reference* 9.6에 설명된 것 같이,

문서 카탈로그 사전의 **StructTreeRoot** 항목 안에 위치한 구조 계층에 의해 설명된다.

구조 계층의 각각의 구조 요소 사전은 **StructElem** 값을 가지고 있는 **Type** 항목을 가지고 있다.

규격에 적합한 파일의 생성 프로그램은 문서의 논리적 구조 계층을 가능한 최소 단위로 포착하려고 시도한다. *PDF Reference* 9.7.4에 정의된 것 같이, 그룹핑 요소들, 블록 수준의 구조 요소들, 단락과 유사한 요소들, 리스트 요소들, 표 요소들, 인라인 수준의 구조 요소들, 링크 요소들 그리고 일러스트레이션 요소들의 표준 구조 유형들을 가능한 최대한 사용한다.

상세설명 문서의 논리적 구조의 명백한 설명은 표현 또는 다른 데이터 포맷으로의 변환 등의 목적을 위한 문서의 완전한 의미 값의 복원 노력이 미래에는 중요하다는 것을 알게 될 것이다.

6.8.3.4 구조 유형(Structure types)

블록 수준(block-level) 구조 요소들의 정의는, *PDF Reference* 9.7.4에 설명된 것 같이 엄격한 구조 패러다임을 따라야 한다.

모든 비표준 구조 유형은, *PDF Reference* 9.7.4에 정의된 것 같이, 구조 트리 루트의 롤맵(role map) 사전 안의 기능상 가장 동일한 표준 종류에 매핑된다. 이 매핑은 간접적일 수 있다; 롤맵안에서 비표준 유형은 또 다른 비표준 유형에 직접 매핑될 수 있다. 하지만, 최종적으로 매핑은 표준 유형에서 끝나야 한다.

6.8.4 자연어 명세(Natural language specification)

파일안의 모든 문자의 기본 자연 언어는 문서 카탈로그 사전의 **Lang** 항목에 의해 명시된다.

기본 언어와 다른 파일안의 모든 문자 내용은, *PDF Reference* 9.8.1에 설명된

것 같이 표기된 컨텐츠 순서(marked-content sequence)에 연결된 **Lang** 속성의 사용, 또는 구조 요소 사전의 **Lang** 항목에 의해 표시된다.

만약 **Lang** 항목이 문서 카탈로그 사전 안에 또는 구조 요소 사전 또는 속성 리스트 안에 존재한다면, *PDF Reference* 9.8.1에서 설명되고 RFC 1766와 *Tags for the Identification of Languages*에 정의된 것 같이 값은 언어 식별자이어야 한다.

파일의 언어가 기본 자연 언어를 사용하지 않거나 innermost enclosing 구조 요소 또는 표시된 내용순서에 의해 정의된 자연 언어가 아닌 유니코드로 인코딩된 모든 문자열은 그것의 언어를 *PDF Reference* 3.8.1에 설명된 것 같이 내부 escape 순서를 사용해서 표시한다.

상세설명 언어에게 이질적인 단어와 일반적 사용에 의해 언어에 합병된 이질적 단어의 구별은 문제이다. 이 조건들은 문자 내용의 미래의 명백한 의미 해석을 허용하려는 의도이다.

6.8.5 부가 기술(Alternate descriptions)

실물과 동일한 미리 정의된 문자 유사어가 없는 내용을 담고 있는 이미지나 공식과 같은 모든 구조 요소는 *PDF Reference 9.8.2*에 기술된 것 같이 구조 요소 사전에 **Alt** 입력어를 사용하여 부가적인 문자 기술(alternative textual description)을 제공해야 한다.

상세설명 아티팩트 기술은 분명하지 않은 비-문자 내용의 적절한 해석을 도와주는, 문자 설명을 제공한다.

6.8.6 비-문자 주석(Non-textual annotations)

문자를 표시하지 않는 모든 주석 종류들에게, 주석 사전의 **Contents** 키가 명시되어야 하며 사람이 판독할 수 있는 형태로 주석 내용의 부가적인 기술과 함께 명시되어야 한다.

6.8.7 대체 문자(Replacement text)

비-표준 방식으로 표현되는 모든 문자 구조 요소들은, 예) 맞춤 글자 또는 인라인 그래픽들, *PDF Reference* 9.8.3에서 설명한 것 같이 구조 요소 사전의 **ActualText** 항목을 사용하여 대체 문자를 제공한다.

상세설명 대체 문자는 분명하지 않고, 문자 구성 요소들이 일반적이지 않은 표현들의 적절한 해석을 도와주는 동등한 문자를 제공한다.

6.8.8 약어와 두문자어의 확장(Expansions of abbreviations and acronyms)

문자 내용의 약어와 두문자어(acronym)의 모든 경우들은, *PDF Reference* 9.8.4에서 설명한 것 같이, **E** 속성이 약자 또는 두문자어의 문자 확장을 제공하는 **Span** 태그가 있는 표시된 내용순서 안에 위치한다.

상세설명 약자와 두문자어 확장은 분명하지 않은 명칭의 적절한 해석을 도와주는 문자와 동등한 것을 제공한다.

6.9 대화식 폼(Interactive Forms)

이 하위 조항의 조건들의 의도는 폼 필드의 표현에 애매함이 없게 보증하는 것이다.

규격에 적합한 판독기는 파일의 페이지 또는 내용의 변화를 수반하는 어떤 폼 필드를 사용해서도 안 된다. 위젯 주석 사전 또는 필드 사전은 **A** 또는 **AA**키들을 포함하지 않는다.

대화식 폼 사전의 **NeedAppearances** flag는 존재하지 않거나 값이 *false*이다.

모든 폼 필드는 필드 데이터와 연계된 모양 사전을 가지고 있다. 규격에 적합한 판독기는 폼 데이터와 상관없이 모양사전에 따라 필드를 표현한다.

상세설명 모양 사전의 필요는 폼의 신뢰할 수 있는 표현을 보증한다.

부속서 A
(참고)
문서보존포맷 해외사례

A.1 호주 NAA

NAA(National Archives of Australia)의 문서보존포맷은 XML이다. XML은 요소와 스키마를 이용한 표준화된 텍스트 구조로 문서를 설명하는 것이 가능하며, 비독점적 포맷으로 전 세계적으로 다양한 소프트웨어에 의해 지원되기 때문에 선정하였다.

A.2 호주 VERS

호주의 VERS(Victorian Electronic Records Strategy)에서는 TEXT, PDF/A, PDF, TIFF, JPEG, JPEG-2000, MPEG-4를 문서보존포맷으로 선정하였다.

A.2.1 텍스트(Text)

사람이 직접 읽을 수 있는 일련의 문자와 ASCII나 Unicode 같이 컴퓨터가 인식하여 처리할 수 있는 형식으로 바꿀 수 있는 텍스트를 문서보존포맷으로 선택하였다.

A.2.2 PDF/A

문서에 사용된 폰트, 레이아웃 등과 같은 정보를 내장하여, 전용판독기만 있으면 문서가 처음 만들어졌을 때의 모습과 느낌 그대로 볼 수 있으며, 문서보안과 검색 기능을 제공한다.

A.2.3 PDF

문서에 사용된 폰트, 레이아웃 등과 같은 정보를 내장하여, 전용판독기만 있으면 문서가 처음 만들어졌을 때의 모습과 느낌 그대로 볼 수 있으며, 문서 보안과 검색 기능을 제공한다.

A.2.4 JPEG, JPEG-2000

사진과 같은 톤 이미지 보존포맷이다. JPEG는 풀 컬러(full-color)와 그레이스케일(gray-scale)의 압축을 위하여 고안되었다. JPEG 압축기술을 이용하면 이를 적절히 조절하여 이미지에 손상에 가지 않도록 이미지를 압축할 수 있다. JPEG는 1,600만 색상을 표시할 수 있어 고해상도 표시장치에 적합하다.

A.2.5 MPEG-4

멀티미디어에 대한 보존포맷으로 오디오 프레임, 비디오 프레임, 스프라이트(sprite), 3D 객체 등을 기술할 수 있다.

A.3 영국 EROS(Electronic Records from Office Systems)

영국은 EROS 프로젝트에서 문서보존포맷을 정의하나, 특정한 문서보존포맷을 지정하지 않고 있다. 그러나 향후 PDF로 지정할 계획이 있음을 문서상에 기술하고 있으며, 이관 및 보존포맷으로 포스트스크립트, PDF, TIFF, CSV, SGML 등을 권고하고 있다. 이를 권고한 이유는 다음과 같다.

A.3.1 포스트스크립트(PostScript)

1) 컴퓨터에서 생산하는 텍스트 및 그래픽 기반의 문서들에 유용하며 컴퓨터상에서 구동되는 모든 애플리케이션은 포스트스크립트 출력을 지원한다.
그리고 포스트스크립트의 사용은 기존의 IT 애플리케이션 발전을 거의 요구

하지 않으므로, 전자기록물의 이관에 중요한 포맷이다.
2) 포스트스크립트는 이미 20년이 넘게 존재하였고, 초기버전으로 만들어진 파일들은 아직도 가장 최신버전 소프트웨어를 사용해서 읽을 수 있다.

A.3.2 PDF(Portable Document Format)

PDF는 시스템독립적인 문서기술을 위한 표준포맷으로 담당부서가 PDF 형식으로 전자문서를 저장한다면, TNA는 이를 이관 및 저장에 적절하도록 고려한다.

A.3.3 TIFF(Tagged Image File Format)

TIFF는 공개된 이미지 표준으로, 이미지 기반의 기록물을 이관하는데 사용될 수 있다.

A.3.4 CSV(Comma Separated Value)

CSV는 스프레드시트나 데이터베이스로 유지되는 기록물들과 같은 구조화된 기록물들의 이관에 적절하며, 데이터입력을 애플리케이션으로 보존 후에 스프레드시트 또는 데이터베이스를 재생산할 수 있도록 한다.
공급업자들은 스프레드시트나 데이터베이스의 한 테이블로부터 다른 테이블로 구조화된 데이터을 이관하기 위한 포맷을 현재까지 지원하므로 보존 시 고려된다.

A.3.5 SGML(Standard Generalised Mark-up Lasnguage)

SGML은 문서기술을 위한 ISO 표준이며, 기존의 문서를 SGML로 변경하는데 드는 비용은 너무 높은 단점이 있다. 만일 문서가 SGML포맷으로 존재한다면 TNA는 이를 문서 이관 및 저장용으로 적절하도록 고려한다.

참고문헌

[1] Looney, Michael., "The Need for Digital Archiving Standards", Campus Technology, 2003, Retrieved from <http://www.campustechnology.com/article.aspx?aid=39100>

[2] Adrian Brown, "『Digital Preservation Guidance Note 1: Selecting File Formats for Long-Term Preservation』, The National Archives 2003, DPGN-01, Issue1, 19 June 2003."

[3] Susan Sullivan, "PDF/A Worldwide Collaboration to Preserve Electronic Documents," US technical advisory group for the PDF/A ISO standard.

AIIM, Frequently Asked Questions (FAQs) ISO 19005-1:2005 PDF/A-1

[4] PDF Reference: Adobe Portable Document Format, Version 1.5, Adobe Systems Incorporated -4th edition. Available from Internet <http://partners.adobe.com/asn/acrobat/sdk/public/docs/PDFReference15_v6.pdf>

[5] The Unicode Standard, Unicode Consortium. Available from Internet <http://www.unicode.org/versions/>

[6] ISO 15930-6, Graphic technology - Prepress digital data exchange using PDF - Part 6: Complete exchange of printing data suitable for colour-managed workflows using PDF 1.4 (PDF/X-3)

[7] ISO 2108, Information and documentation - International standard book number (ISBN)

[8] Namespaces in XML 1.1, W3C Recommendation, February 4, 2004. Available from Internet <http://www.w3.org/TR/2004/REC-xml-names11-20040204>