

3장. 디지털 객체 관리

부산대학교 문헌정보학과

이수상 교수

sslee@pusan.ac.kr



1. 디지털 객체의 정의

■ 객체 (object)

- 디지털 컴퓨팅 환경에서 많이 사용하고 있는 개념. 속성과 행위로 설명된다.
- 디지털정보 시스템에서의 디지털 객체
 - ✓ 메타데이터라는 속성으로 정의되고,
 - ✓ 다양한 행위로 구성되는 기능에 의하여 처리된다.
- 디지털 객체의 속성
 - ✓ 서명, 저자, 출판년 등의 색인 또는 목록 정보, 파일의 크기나 매체 유형 등과 같은 기술적인 정보
 - ✓ 보존에 필요한 정보, 저작권관련 정보, 또는 시스템에서의 각종 관리적 목적의 정보
- 디지털 객체의 행위: 수집, 조직, 저장, 배포 등과 같은 처리동작으로서 시스템의 주요 기능

1. 디지털 객체의 정의

■ 디지털 객체 (digital object)

- 디지털 문헌에 대한 관점 중의 하나인 객체로서의 문헌
- 디지털도서관 시스템에서 처리되는 정보의 한 단위
- 시스템과 관련된 설명에서 사용하는 용어로 비트열(bit stream) 또는 바이트열(byte stream)로 표현
- 2가지 유형
 - ✓ 단일의 파일로 구성되는 단순객체(simple object)로 존재하거나
 - ✓ 복수의 파일, 또는 복수의 멀티미디어 파일로 구성되는 복합객체 (complex object)
- 주소표현이 가능한 정보자원(addressable information resource) 또는 디지털 문헌 → URI(Uniform Resource Identifier)
- 관련 용어: 디지털 매체, 전자기록물, 이프린트(e-print), 이텍스트(e-text), 디지털 자원, 전자정보, 디지털 지식정보 등

1. 디지털 객체의 정의

■ 디지털 문헌

- 2가지 형식의 파일
 - ✓ 비트열 또는 바이트열 형식의 물리적 파일
 - ✓ 뷰어(viewer)를 거쳐 문장, 단락, 장과 절 등 인간이 이해할 수 있는 형식으로 재현되는 논리적 파일
- 콘텐츠의 인코딩 형식에 따라 텍스트, 이미지, 멀티미디어로 구분
- 생산방식에서 본 디지털 문헌의 2가지 생산 유형
 - ✓ 디지털생산(born digital) 문헌
 - ✓ 디지털변환(being digital) 문헌

1. 디지털 객체의 정의

- 디지털 문헌의 핫이슈: 장기적인 보존과 접근 보장
 - 기계종속적이다.
 - 매체가 손상되기 쉽다.
 - 변경이 쉽다.
 - 인쇄문헌보다 관심을 더 기울여야 함.
 - 보존활동은 지속적인 투자와 기술적인 해결책, 그리고 비용에 제약을 받음.
 - 디지털보존은 새로운 워크플로우, 기술, 전문가들의 협력체제 등을 필요로 함.
 - 디지털 문헌의 생명주기 상에는 여러 유형의 이해당사자들이 있음.
 - 기록 및 보존 기능을 개선하기 위한 새로운 시스템과 S/W 산학협력이 필요함.
 - 앞으로 디지털 정보의 생산과 활용이 비약적으로 증가할 것임.

1. 디지털 객체의 정의

■ NISO의 보고서에서 정의하는 디지털 객체의 형태

- 첫째, 객체는 하나의 파일로 구성된다(예: PDF파일로 발간된 보고서).
- 둘째, 객체는 다중 링크된 파일로 구성된다(예: HTML 페이지와 페이지 내의 이미지).
- 셋째, 다중 파일과 그것을 연결하는 구조적 메타데이터로 구성된다(예: 페이지 단위로 이미지화된 도서).

■ 디지털 객체의 형태 (2)

- 보존용 객체(master copy 또는 preservation copy): 최초의 디지털 객체 원본을 의미하며, 대체적으로 가장 품질이 높은 버전이다.
- 서비스용 객체(access 또는 use copy): 특정한 목적이나 계획에 따라 이용자의 서비스를 위해 파생된 것이다.

1. 디지털 객체의 정의

■ NISO의 디지털 객체를 위한 6가지 기본원칙

- <객체 원칙 1> 우수한 디지털 객체는 상호운용성과 재사용성을 위한 품질 수준을 유지하면서, 장서의 우선순위를 보장하는 방식으로 생산되어야 한다.
- <객체 원칙 2> 우수한 객체는 영구적으로 보존되어야 한다. 즉, 개인 또는 기관의 의지에 의해 기술적인 변화에도 불구하고 우수한 객체의 접근성이 보장되어야 한다.
- <객체 원칙 3> 우수한 객체는 현재와 미래의 이용을 지원하거나, 또는 이러한 이용을 지원하는 접근용 복제물(서비스용 객체)을 변환하는 것을 지원하는 형식(포맷)이 디지털화가 된다.
- <객체 원칙 4> 우수한 객체는 영구적이고 고유한 식별자를 이용한 명칭이 부여될 것이다.
- <객체 원칙 5> 우수한 객체는 최소한 다음의 3가지 관점에서 진본성을 인증 받는다. → 첫째, 이용자가 객체의 원본, 구조, 개발이력(버전 등)을 결정할 수 있다. 둘째, 이용자는 자신의 취지나 의도에 알맞은 객체를 결정할 수 있다. 셋째, 이용자는 객체가 비인가된 방식으로 수정되거나 변경되지 않았음을 결정할 수 있다.
- <객체 원칙 6> 우수한 객체는 관련된 메타데이터를 가진다.

■ 물리적 객체(아날로그 정보자원)의 유형

- 1차 자원/2차 자원/3차 자원, 학술정보/교양정보 등 유형구분
- MARC의 유형구분(Leader 06필드-Type of record) 사례
 - ✓ a - 문자로 된 인쇄자료(Language material)
 - ✓ c - 인쇄악보(Notated music)
 - ✓ d - 필사악보(Manuscript notated music)
 - ✓ e - 인쇄지도(Cartographic material)
 - ✓ f - 필사지도(Manuscript cartographic material)
 - ✓ g - 평면영사자료(Projected medium)
 - ✓ i - 비음악자료(Nonmusical sound recording)
 - ✓ j - 음악자료(Musical sound recording)
 - ✓ k - 평면비영사자료(Two-dimensional nonprojectable graphic)
 - ✓ m - 컴퓨터 파일(Computer file) 등

- 디지털 객체의 유형구분 사례: DCMI 유형어휘 (Type Vocabulary)
 - 장서(Collection) : 아이템의 집합체
 - 데이터세트(Dataset)
 - 사건(Event)
 - 이미지(Image)
 - 대화식 자원(Interactive Resource)
 - 물리적 객체(Physical Object)
 - 서비스(Service)
 - 소프트웨어(Software)
 - 사운드(Sound)
 - 정지화상(Still Image)
 - 텍스트(Text)

■ MODS “typeOfResource” 요소의 유형값

- 텍스트(text)
- 지도(cartographic)
- 악보(notated music)
- 녹음자료-음악(sound recording-musical)
- 녹음자료-비음악(sound recording-nonmusical)
- 녹음자료(sound recording) *음악/비음악 이외의 자료가 혼재하는 경우
- 정지화상(still image)
- 동영상(moving image)
- 입체자료 및 실물(three dimensional object)
- 소프트웨어(software), 멀티미디어(multimedia)
- 혼합 자료(mixed material)

■ 포맷의 정의

- 자원을 물리적 또는 디지털 형식으로 구성한 것으로서, 매체의 유형과 특징을 포함한다.
- 자원을 재현하고, 동작하는데 필요한 S/W, H/W, 다른 장치를 인식하는데 사용하는 정보(<http://dublincore.org>)

■ 디지털 포맷의 유형

- 디지털 매체의 유형(media type)에 따라 구분(MIME(Multi-purpose Internet Media Type)에서 정의한 것을 많이 사용)
- MIME에서 포맷의 유형은 2가지로 나누어 구분하고 있다.
 - ✓ 첫 번째 구분은 문헌의 콘텐츠 유형(type)에 따라 8가지로 구분한다. 'application, audio, image, message, model, multipart, text, video'이며,
 - ✓ 각각은 다시 하위유형(sub-type)들이 존재한다.[관련 표준 참조: RFC 2045, RFC 2046] 예를 들면 video는 그 하위유형으로 avi, mpeg, mov 등이 있다.

3. 디지털 객체의 포맷

■ 디지털 포맷의 표현: 텍스트, 이미지, 멀티미디어

■ 텍스트

- ✓ 단순한 아스키 파일(TXT 확장자 파일), RTF 파일(TXT보다 고급의 텍스트 파일), 워드프로세서 파일(HWP, DOC 등), 식자형 포맷의 파일(Tex, LaTeX 등), 인쇄용 파일(PDF, Postscript 등) 등으로 세분 가능
- ✓ 텍스트는 SGML, XML, HTML 등과 같은 마크업(markup) 형식으로 구성 가능

■ 이미지

- ✓ 저장공간을 많이 차지하는 단순한 비트맵 포맷(BMP, PCX), 압축저장 포맷(GIF, JPG), 문서교환 포맷(TIFF) 등으로 구분된다.

■ 멀티미디어

- ✓ 오디오 포맷(ASF, WAV, MIDI, MP3, AU), 애니메이션 및 동영상 포맷(MPEG, AVI, FLI, QuickTime 등), 컴퓨터그래픽 파일(CGM), 가상현실 모델 언어(VRML), 마크업 표준(HyTime) 등으로 표현된다.

3. 디지털 객체의 포맷

■ 디지털 포맷과 보존 문제

- 디지털 아카이빙에서 파일포맷에 관한 문제는 제출포맷, 보존포맷, 그리고 배포포맷의 3가지 관점에서 구분하여 접근하여야 한다.

■ 제출포맷

- 생산자는 자신이 선호하는 포맷을 사용(제출포맷의 유형)
 - ✓ 공개 또는 표준포맷을 사용하는 경우
 - ✓ 워드프로세서, 모델링도구 등과 같은 전용포맷을 사용하는 경우

■ 보존포맷

- 제출포맷을 보존포맷으로 변환
 - ✓ 자료의 접수단계에서 포맷전환(migration on Ingest)하거나
 - ✓ 해당 포맷의 버전이 변경되거나 기술적인 퇴화가 나타날 경우, 보존 상태에서 수시로 포맷전환을 해 주어야 한다.

■ 배포포맷

- 디지털 문헌의 콘텐츠를 식별할 수 있는 포맷
 - ✓ 보존포맷 그대로 배포할 수 있으며,
 - ✓ 배포시 포맷변환을 하는 경우도 있다. 이것은 배포단계 포맷전환 (migration on demand) 방식이라 한다.
- 배포포맷은 별도의 노력없이 디지털 문헌의 콘텐츠를 식별할 수 있는 포맷이어야 한다.

4. 디지털 객체 관리 주요 기능

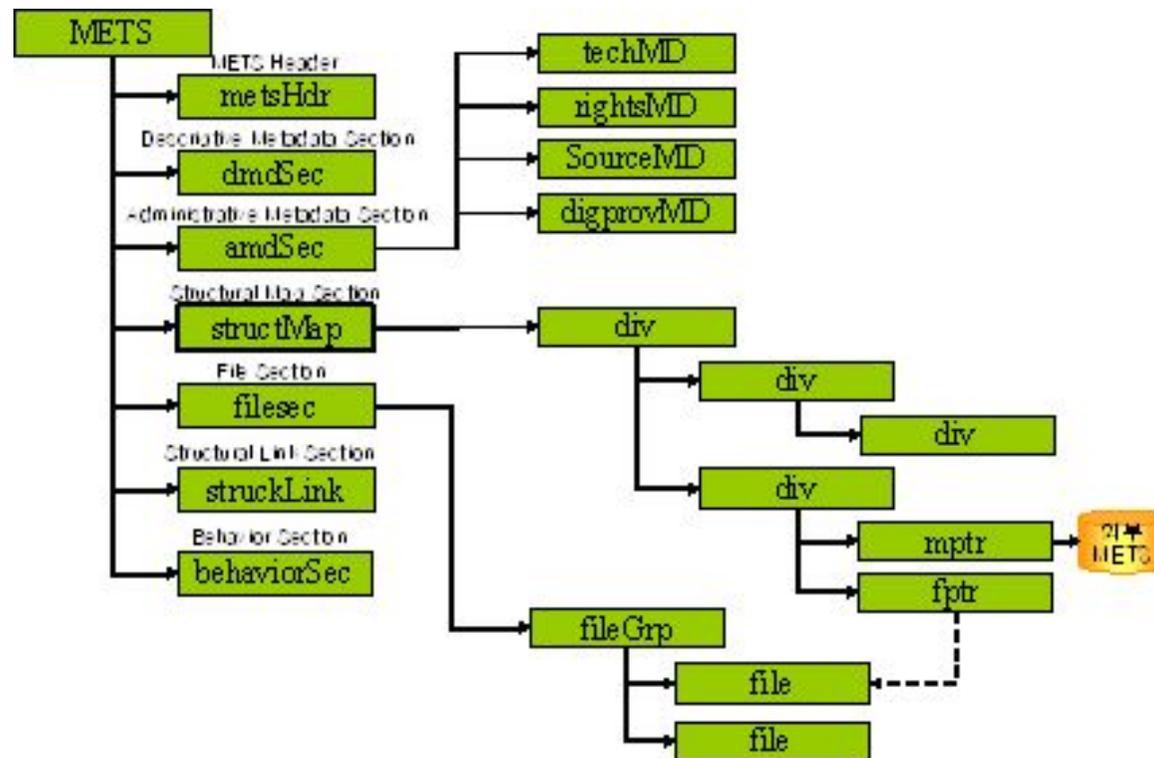
■ 객체 모델링

- OAIS 정보객체모델
- OAIS 참조모델(Reference Model)을 구성하는 중요한 개념으로서, OAIS 정보객체모델(Information Object Model)은 정보시스템에서의 정보객체가 수행하는 각종 행위(또는 동작)를 개념화한 것
- 정보객체가 정보패키지를 구성한다.
- 정보패키지의 3가지 유형
 - ✓SIP(Submission Information Package)
 - ✓AIP(Archival Information Package)
 - ✓DIP(Dissemination Information Package)

4. 디지털 객체 관리 주요 기능

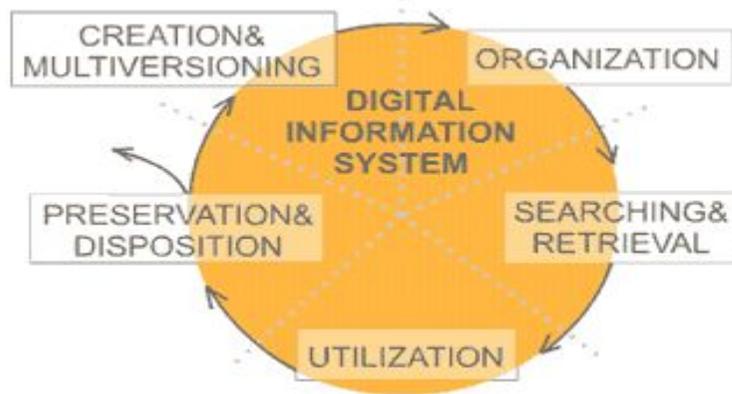
■ 정보객체 패키징: METS

- METS(Metadata Encoding and Transmission Standard)는 시스템간의 '복합디지털 객체'의 전송 및 아카이빙을 위한 디지털 자원에 대한 허브(hub)문서의 인코딩 규칙을 규정하는 XML 스키마 기반의 명세
- METS 기본 구조 다이어그램

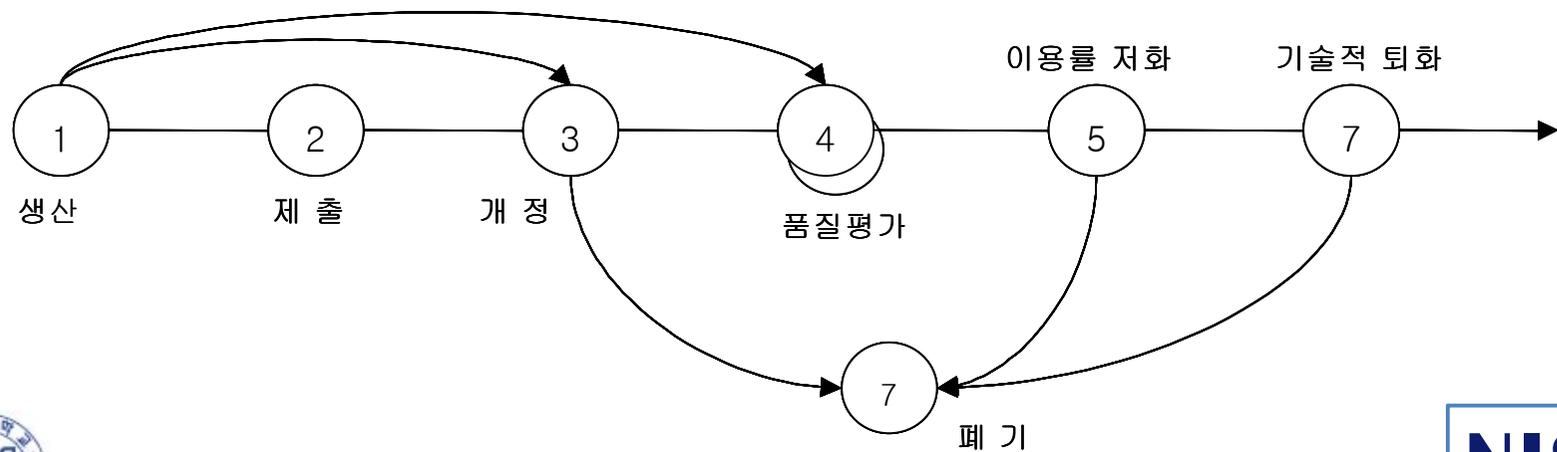


5. 객체관리 워크플로우

■ 스웨트랜드의 디지털 객체 수명주기

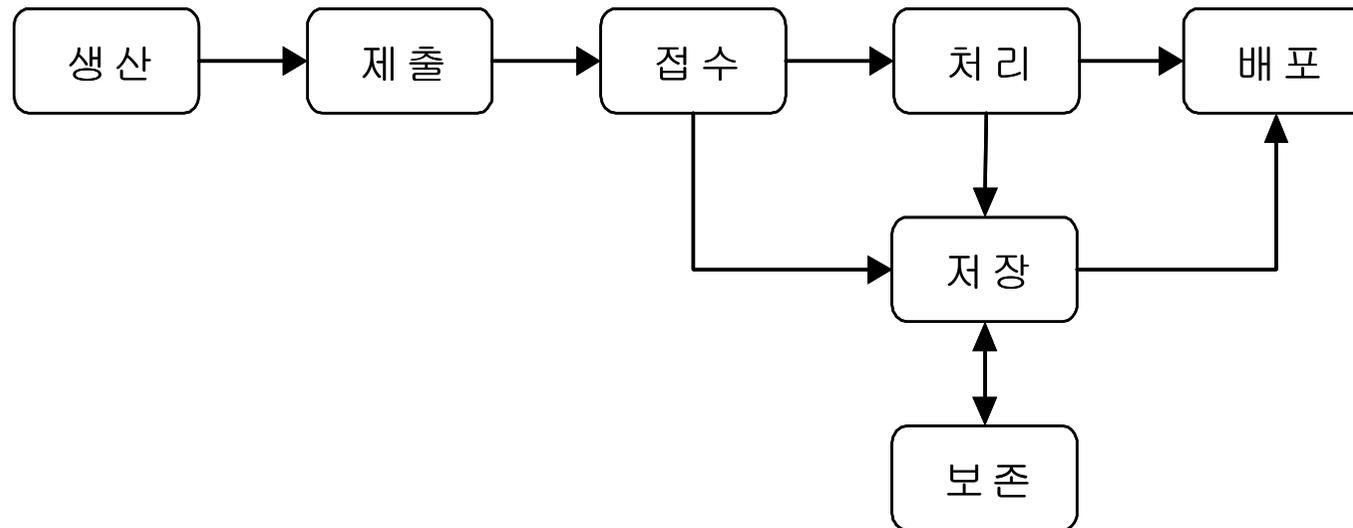


■ ISC 보고서의 디지털 객체 수명주기



5. 객체관리 워크플로우

■ 아카이빙 워크플로우



6. 디지털 콘텐츠의 평가

■ 평가방법

- 사서들은 기존의 장서평가방법의 틀 안에서 디지털 콘텐츠를 평가하는 방법을 연구해 오고 있으며,
- 디지털 콘텐츠의 평가는 이용 중심적 평가방법을 적용하는 것이 바람직하다.
- 이용 중심적 평가방법
 - ✓ 이용자 의견조사, 이용자관찰, 포커스그룹 등
 - ✓ 디지털 콘텐츠의 포괄성, 수록범위의 깊이 및 접근과 이용의 수월성 등에 대한 이용자의 의식수준을 파악할 수 있다.
- 디지털 콘텐츠를 평가하기 위한 방법 중 가장 주목받고 있는 것이 **이용 통계에 의한 평가방법**이다.

■ 이용통계에 의한 평가

- 사서들은 출판사나 벤더에게 이용통계자료의 제출을 요구
- 출판사나 벤더는 이용통계의 제공을 꺼리거나, 비용을 요구
- 관련 프로그램
 - ✓ ICOLC(The International Coalition of Library Consortia)
 - ✓ E-Metrics(Measures for Electronic Resources)
 - ✓ COUNTER(Counting Online Usage of NeTworked Electronic Resources)
- 단위도서관에서는 로그데이터를 이용하여 디지털 콘텐츠의 이용통계를 수집 가능함 - 이용로그에 의한 이용행태 분석

■ 보존의 문제

- 디지털 콘텐츠의 보존전략은 디지털 콘텐츠의 생산방법에 따라 달리 할 수 있다.
 - ✓ 콘텐츠를 제작한 자관의 책임
 - ✓ 국가도서관의 책임
- 보존전략을 수립하기 위한 프로젝트 사례: 영국의 CEDARS 프로젝트

■ 참고 – 전자저널의 보존 문제

- 구독 도서관의 책임: 저작권, 기술 등의 문제로 보존책임은 무리
- 출판사의 책임: 영업 기간 동안은 저작권을 가지고 있으므로 보존의 책임이 있지만, 보존이 지속적으로 유지될 수 없다.
- 국가도서관의 책임: 납본법을 통하여 디지털 자료를 수집, 보존함으로써 보존의 주체가 되는 방법
- 관련 단체가 협의체에서 보존