



디지털 보존시스템 현황 및 상호운용성 개념 이해

2019. 8. 28.
안대진 (아카이브랩)

국가기록원 2019년 제4차 전자기록관리 연구포럼

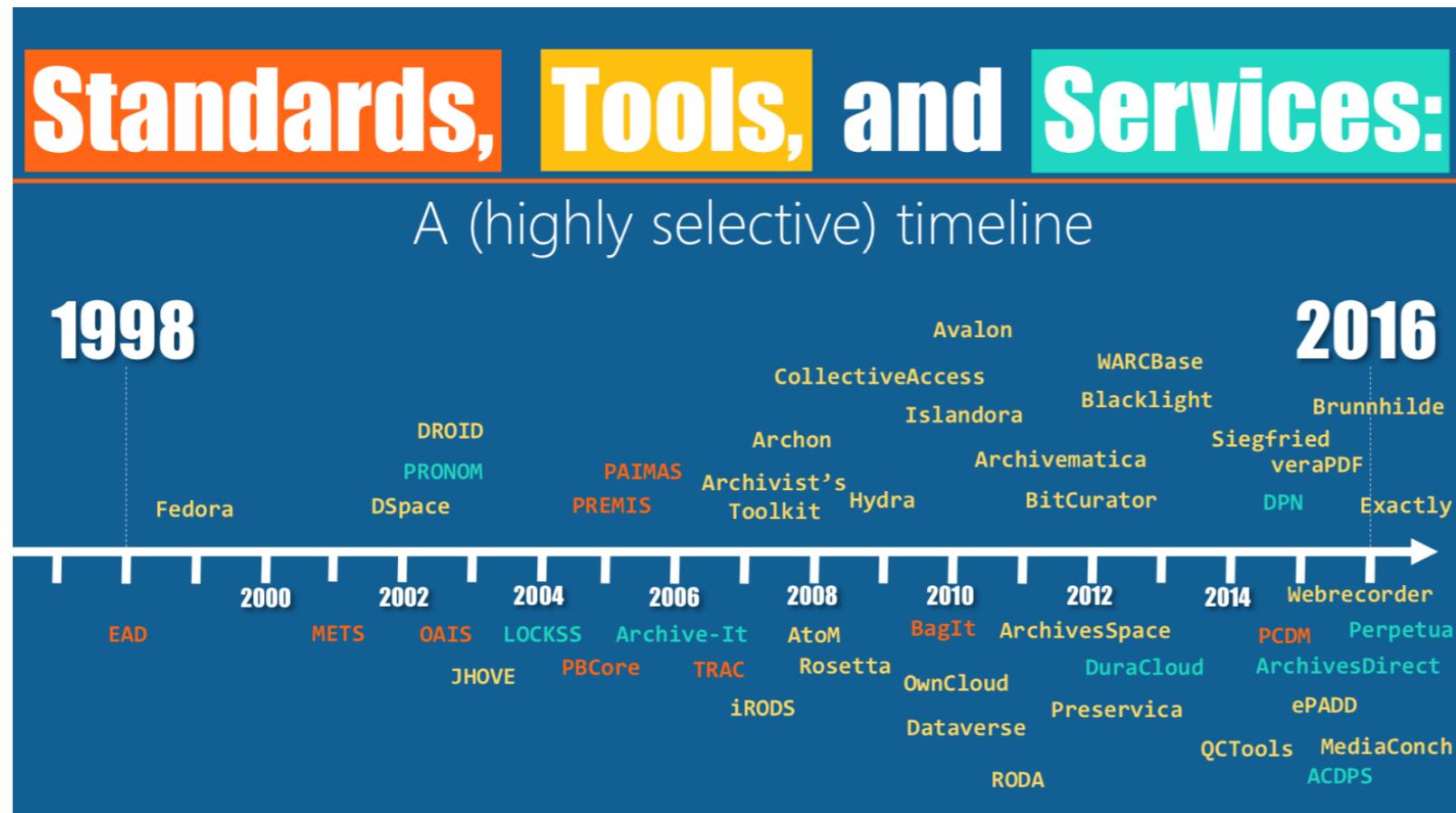
목차

1. 표준, 툴, 서비스
2. 시스템
 - Archivematica
 - Preservica
 - ERA 2.0
 - SDAS
3. 상호운용성
4. 결론

1. 표준, 툴, 서비스

디지털 보존 분야의 표준, 툴, 서비스

90년대 말부터 디지털 보존 커뮤니티는 표준을 공동 개발하고 이를 툴과 서비스로 발전시킴



주요 표준, 툴, 서비스

표준

- EAD, METS, PREMIS(메타데이터)
- OAIS, TRAC, PAIMAS (장기보존)
- BagIt(패키지)
- PCDM(데이터모델)

툴

- Fedora, Dspace 등 (리포지터리)
- Archon, AtoM 등 (카탈로그)
- DROID, JHOVE 등 (식별도구)
- veraPDF 등 (검증도구)
- Preservica 등 (아카이브시스템)

서비스

- PRONOM 등 (포맷레지스트리)
- Archive-It 등 (웹 아카이빙 서비스)
- ArchivesDIRECT 등 (보존 서비스)

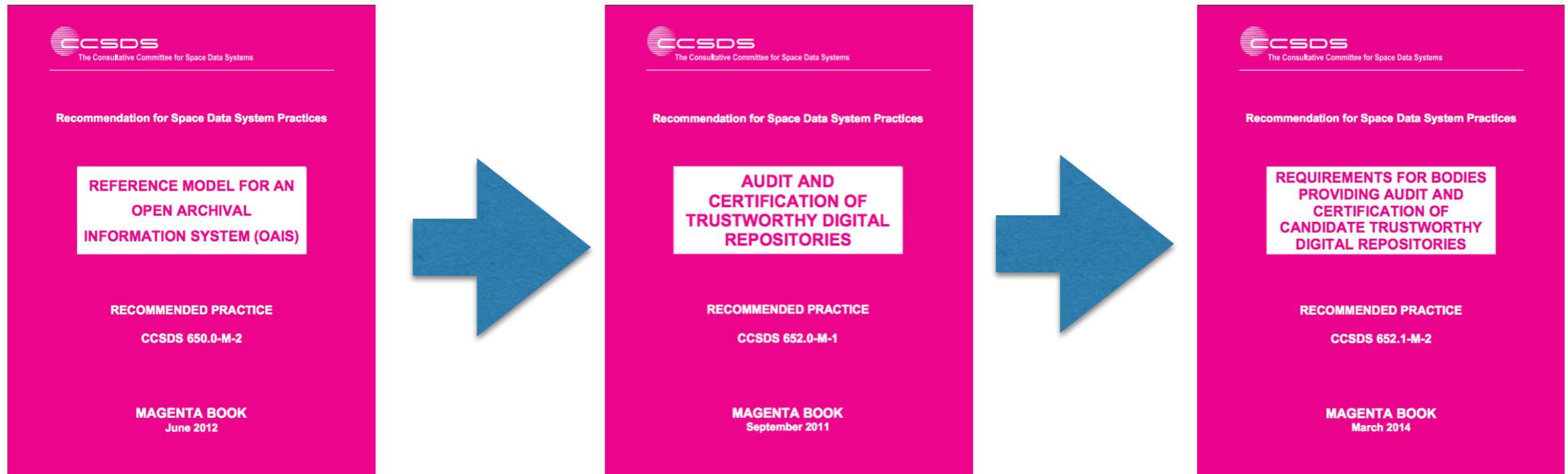
1. 표준, 툴, 서비스

1) 표준

OAIS와 후속표준

디지털 보존 분야의 대표적 표준은 ISO 14721(OAIS 참조모델)과 ISO 16363(디지털저장소 신뢰성 감사 및 인증)임. 대부분의 디지털 아카이브시스템은 두 표준을 근간으로 설계됨.

- ISO 14721은 전자기록의 장기보존과 지속적 접근을 위한 가장 포괄적인 표준(InterPARES 역시 OAIS를 기반으로 함)
- ISO 16363은 OAIS의 후속표준이며, TRAC(RLG & NARA)과 DRAMBORA 등의 평가도구를 근간으로 만들어짐.
- ISO 16919는 ISO 16363을 사용하여 디지털 저장소의 신뢰성을 평가/인증하는 조직의 요구사항을 정의함.



ISO 14721:2002
(OAIS)

ISO 16363:2012

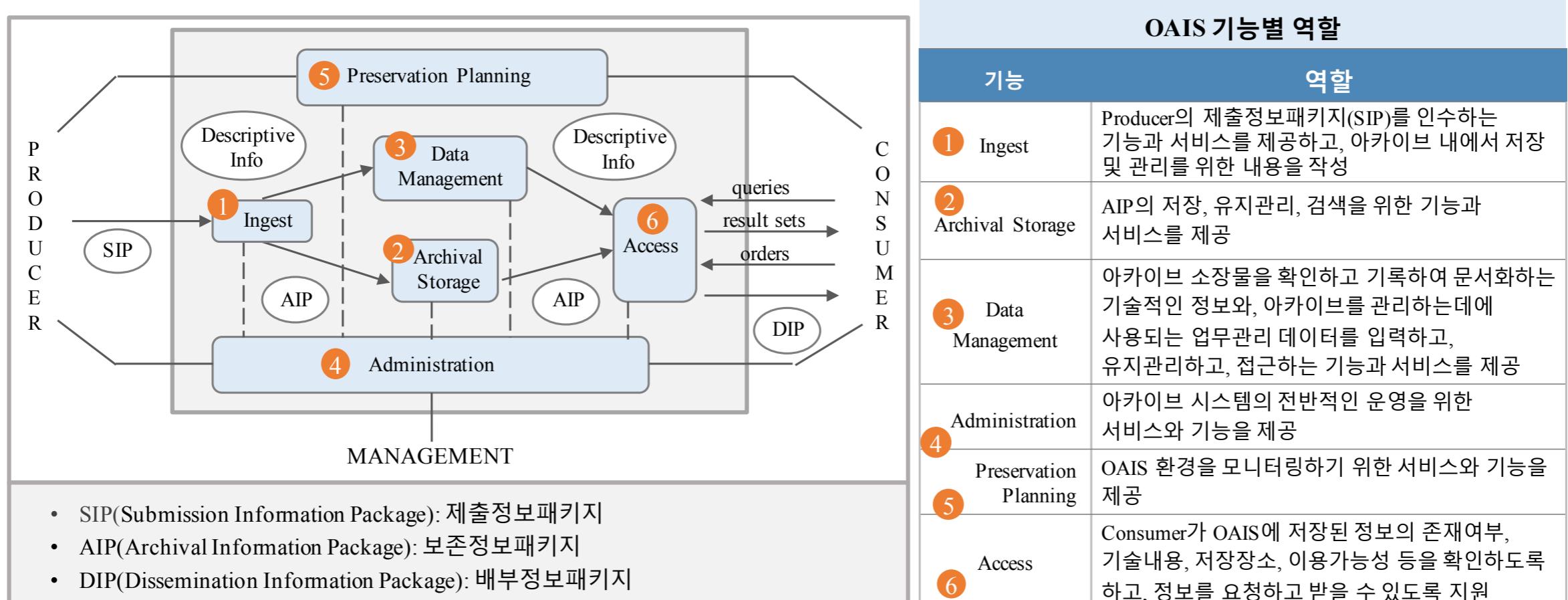
ISO 16919:2014

1. 표준, 툴, 서비스

1) 표준

OAIS 개요

ISO 14721(OAIS)은 전자기록을 장기보존하고 지속적 접근을 제공하기 위한 참조모델 제시



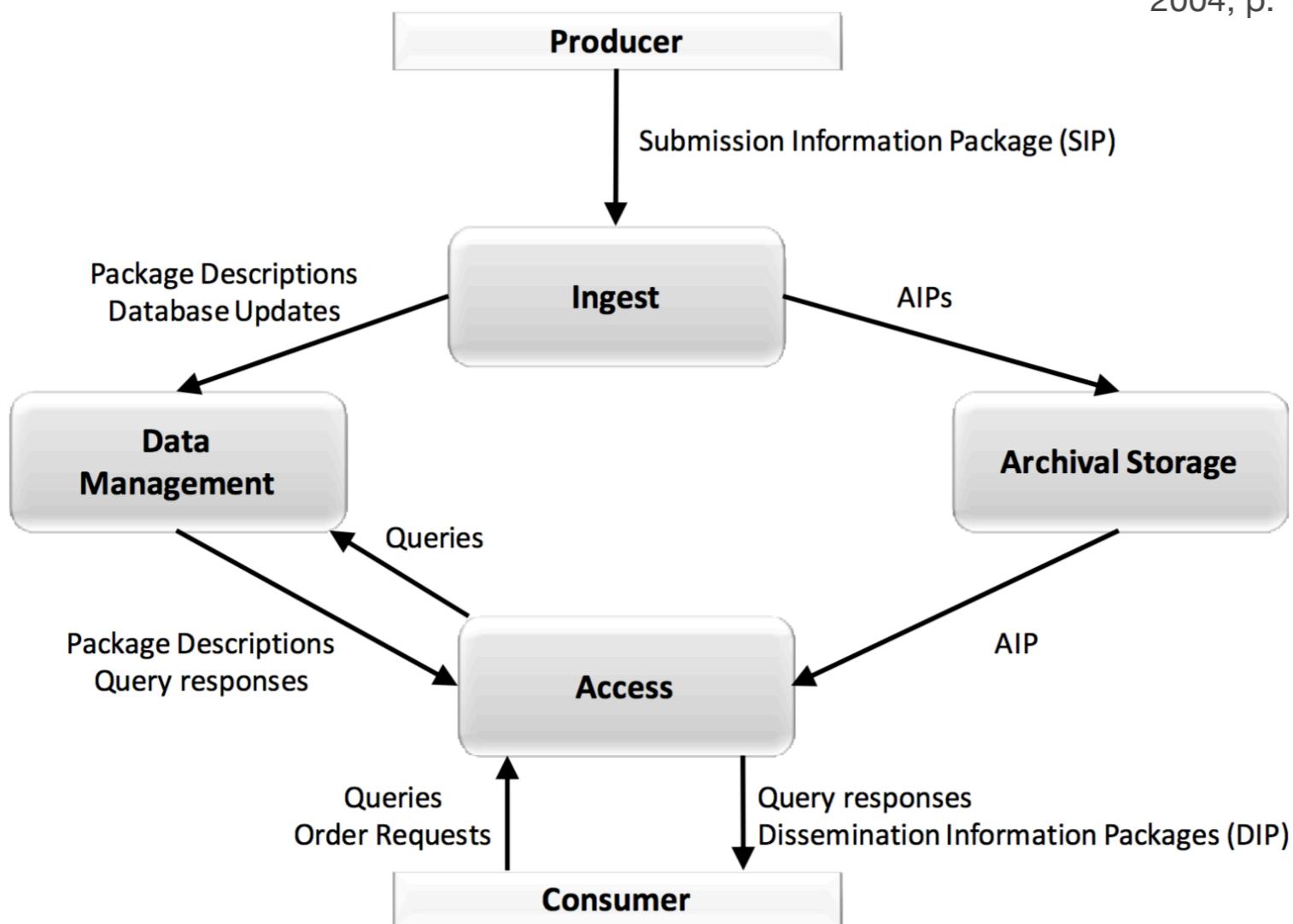
1. 표준, 툴, 서비스

1) 표준

OAIS의 문제

OAIS의 요구사항은 매우 추상적이어서 해석, 번역, 구체화 필요

“ Because of the OAIS model’s abstract character, the notion of OAIS compliance is, as has been pointed out repeatedly, “necessarily vague” ” (Lavoie, 2004, p. 17).



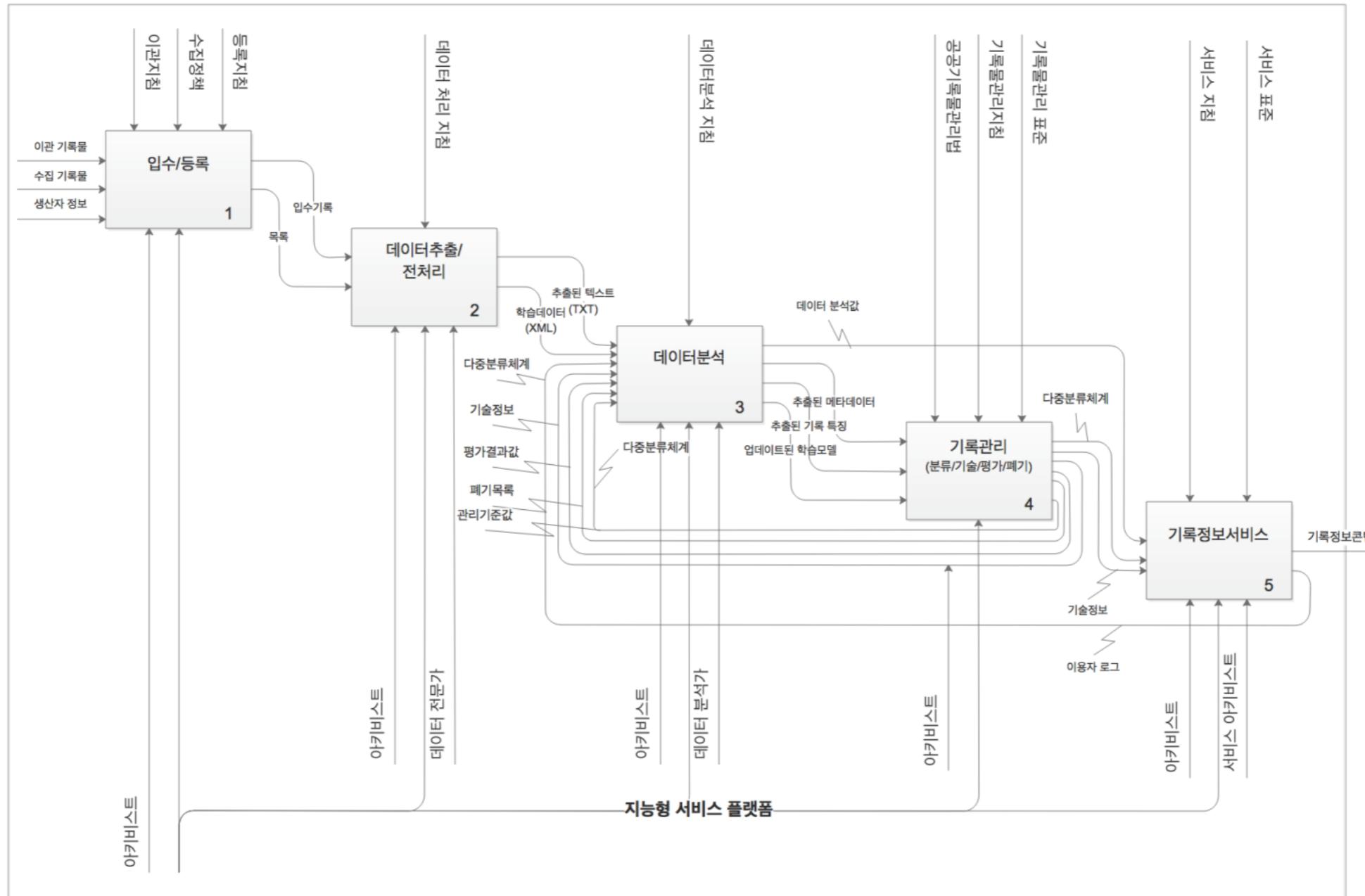
[그림] High Level 데이터 흐름도

1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : RM + AM

국가기록원 차세대 연구에서는 OAIS 참조모형에 따라 차세대 기록관리시스템의 서비스 모델 수립
기록물 입수, 데이터 분석, 기록물 분류, 평가폐기, 기록물 보존까지 RM과 AM의 기능을 수행하도록 해석 확장



*출처: 국가기록원. 2017. 차세대기록관리모델 재설계.

1. 표준, 툴, 서비스

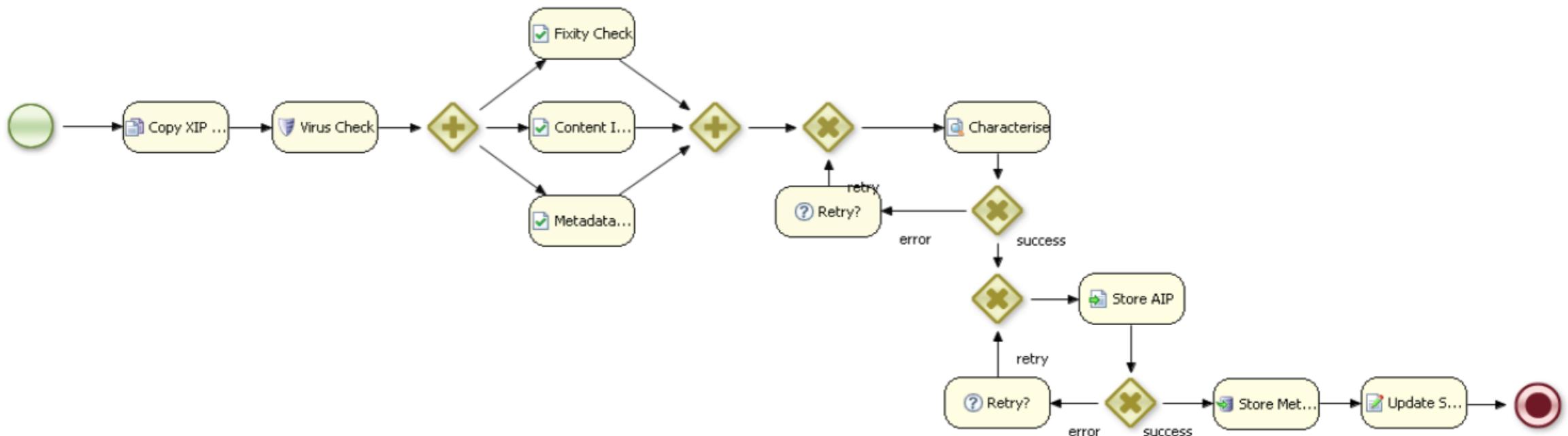
1) 표준

OAIS 해석 및 구현 : 워크플로우

워크플로우는 특정한 비즈니스 요구사항을 충족하기 위해 시스템 상의 여러 가지 처리 작업(job)을 구체적으로 정렬한 것

워크플로우 맥락(contexts)은 워크플로우를 정의하는 것과 이것을 실행하는데 필요한 구성 정보의 조합을 설명시 사용

워크플로우 인스턴스(instance)는 워크플로우 맥락의 특정한 단일 행위에 따라 실행



[그림] 일반적인 워크플로우 다이어그램

1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : 정보 패키지

OAIS 참조모델에서는 정보패키지의 유형을 SIP, AIP, DIP로 구분하고 있으나 CAMS 등에는 이의 구분이 모호함

| 구분 | SIP | AIP | DIP | 시사점 |
|-------|--|--|--|--|
| | 제출정보패키지(Submission) | 보존정보패키지(Preservation) | 배포정보패키지(Dissemination) | |
| 정의 | <ul style="list-style-type: none">생산자가 OAIS로 보내는 정보패키지 (형태와 내용은 OAIS와 생산자 간의 협약에따름) | <ul style="list-style-type: none">SIP를 기반으로 OAIS에 보존하기 위해 변형된 정보패키지 | <ul style="list-style-type: none">이용자의 주문에 대해 AIP의 전체 또는 일부를 제공하는 정보패키지 | 메타데이터 구분 어려움 <ul style="list-style-type: none">OAIS에서는 메타데이터 모형을 제공하지 않아 SIP, AIP, DIP별로 메타데이터를 나누기 어려움 |
| 정보 요소 | <ul style="list-style-type: none">내용정보와 보존기술정보로 구성표현정보와 연결 | <ul style="list-style-type: none">내용정보와 보존기술정보의 묶음, 혹은 AIP들의 집합PREMIS/METS 메타데이터 적용 | <ul style="list-style-type: none">AIU(AIP들의 집합)이거나 PDI(보존기술정보) 없는 AIP 등AtoM 등 서비스 시스템의 계정정보 및 보안정보(DRM 등)METS 메타데이터 적용 | 규격 확정 필요 <ul style="list-style-type: none">시스템 구축과 병행하여 세 규격 확정 필요(특히 SIP는 생산자와 AMS 간의 협약된 규격에 따르며, 유연하게 적용 가능)시청각기록, 데이터세트 등이 포함될 수 있는 패키지 규격 필요(현재는 문서보존포맷 중심으로 설계됨) |
| 비고 | <ul style="list-style-type: none">다수 SIP가 동일한 표현정보를 이용하면 표현정보는 한 번만 제공함다수 SIP에 동일한 보존기술 정보가 적용되면 내용정보 없는 SIP 형태로 제공함기술정보는 SIP 제출시점에 관계없이 제공됨 | <ul style="list-style-type: none">패키징 정보는 내부의 표준을 따름AIP 기술정보를 이용해 이용자가 원하는 내용정보를 찾고주문 가능하도록 함 | <ul style="list-style-type: none">배포매체와 이용자 요구에 따라 패키징정보가 다양함DIP 기술정보는 DIP 전송 이전, 이후, 전송시 제공가능 | 적용사례(Implementation) 벤치마킹 필요 <ul style="list-style-type: none">OAIS는 추상적 수준 요건이므로 실제 구현 사례 검토 필요각국 아카이브즈 및 연구 (E-Ark), rm, AMS 소프트웨어 (Archivematica 등)의 SIP, AIP, DIP 규격 벤치마킹 필요 |

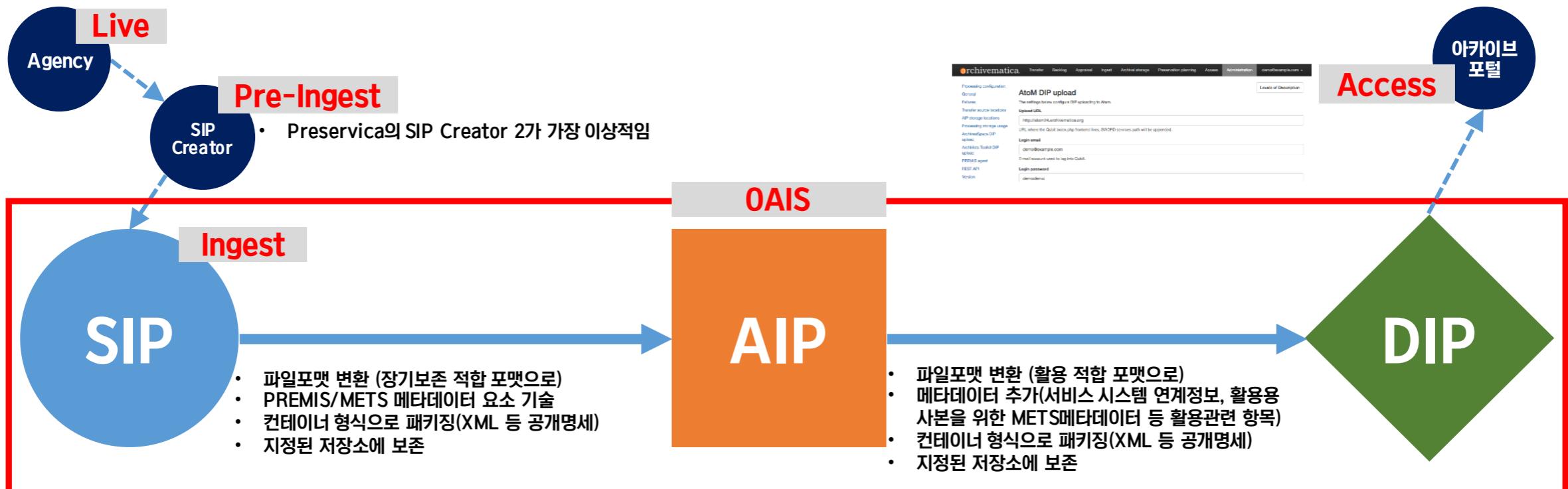
*출처: 대통령기록관. 2019. 대통령기록 정보시스템 고도화 ISP

1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : 정보 패키지

E-Ark 프로젝트에서는 Preservica, Archivematica 등 현행 7종의 AIP와 3종의 컨테이너 포맷(.ace, .rar, .tar)를 분석한 후 범 유럽 공통의 SIP, AIP, DIP 명세서를 개발 및 공개(장기보존, 벤더독립, 사본호환, 재난복구 등 4대 목표)



시사점

E-Ark의 SIP, AIP, DIP 명세 최우선 참고 요망

- 공개된 명세 최신버전 확인 필요 (<https://www.eark-project.com/resources/project-deliverables>)

E-Ark의 AIP 비교연구 참고 요망

- Project 620998: European Archival Records and Knowledge Preservation – E-ARK (<https://www.eark-project.com/resources/project-deliverables/93-d34-1/file>)

*출처: 대통령기록관. 2019. 대통령기록 정보시스템 고도화 ISP

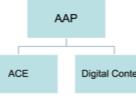
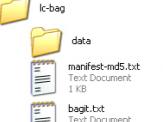
1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : 정보 패키지

각국의 기록소장기관들은 OAIS 참조모형의 정보패키지를 기반으로 SIP, AIP, DIP 설계
XML로 묶는 형태와 폴더 형태로 구분

- 보존패키지 유형

| METS | PREMIS | XIP | AAP | NEO / VEO | EAD / EAC | SIARD | XML | PDF/A | Windows folder | BagIt |
|---|---|---|---|--|--|--|--|--|---|--|
|  <small>METS fileSec -> file inventory dmdSec -> descriptive metadata amdSec -> administrative metadata behaviorSec -> behavior metadata structMap -> structural metadata</small> |  <small>PREMIS data model Agent Object Event</small> |  <small>Tessella Technology IT Consulting AAP ACE Digital Content</small> |  |  <small>NEO / VEO Record Metadata</small> |  <small>EAD / EAC Header Finding aid Administrative Publication Content Structural Relationships Citation Creation Revision</small> |  <small>SIARD</small> |  <small>XML</small> |  <small>PDF/A</small> |  <small>Windows folder</small> |  <small>BagIt manifest-md5.txt bagit.txt data</small> |

- 장기보존포맷 요건 및 특징 (OAIS 참조모형 기준)

| | |
|--------------|--|
| 개요 | <ul style="list-style-type: none">대부분 아카이브는 OAIS를 따르며, 정보 패키지는 SIP, AIP, DIP로 구성국가별 정보 패키지 규격을 설계하여 XML로 표현 |
| AIP 구성요소 | <ul style="list-style-type: none">콘텐츠정보(Content Information), 보존기술정보(PDI) |
| PDI 구성요소* | <ul style="list-style-type: none">참조정보, 출처정보, 맥락정보, Fixity 정보 (기관별 재량) |
| Fixity check | <ul style="list-style-type: none">MD5, SHA-1, SHA-256, SHA-512 등 (기관별 재량) |
| 인캡슐레이션 | <ul style="list-style-type: none">ZIP, TAR, Base64, 또는 무압축(Unzipped Bag 등) (기관별 재량) |

* PDI 내용구성은 개별 아카이브의 재량이며, 전자서명(PKI) 인증정보를 패키지 내에 포함하기도 함(NEO, VEO 등). 전자서명 대신 해시로 위변조 여부 파악하기도 함

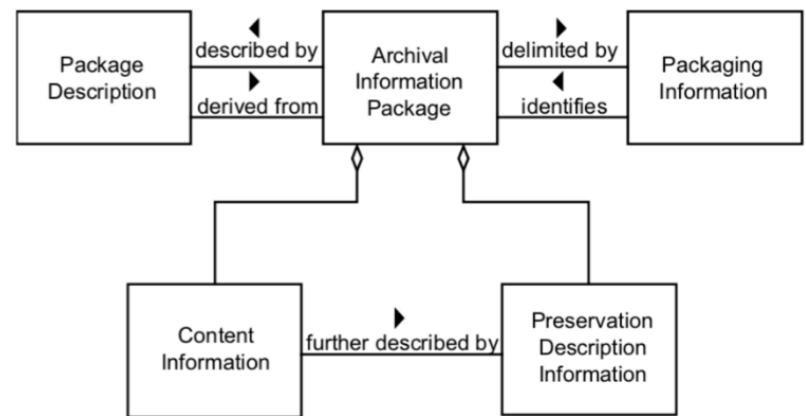


Figure 4-15: Archival Information Package (AIP)

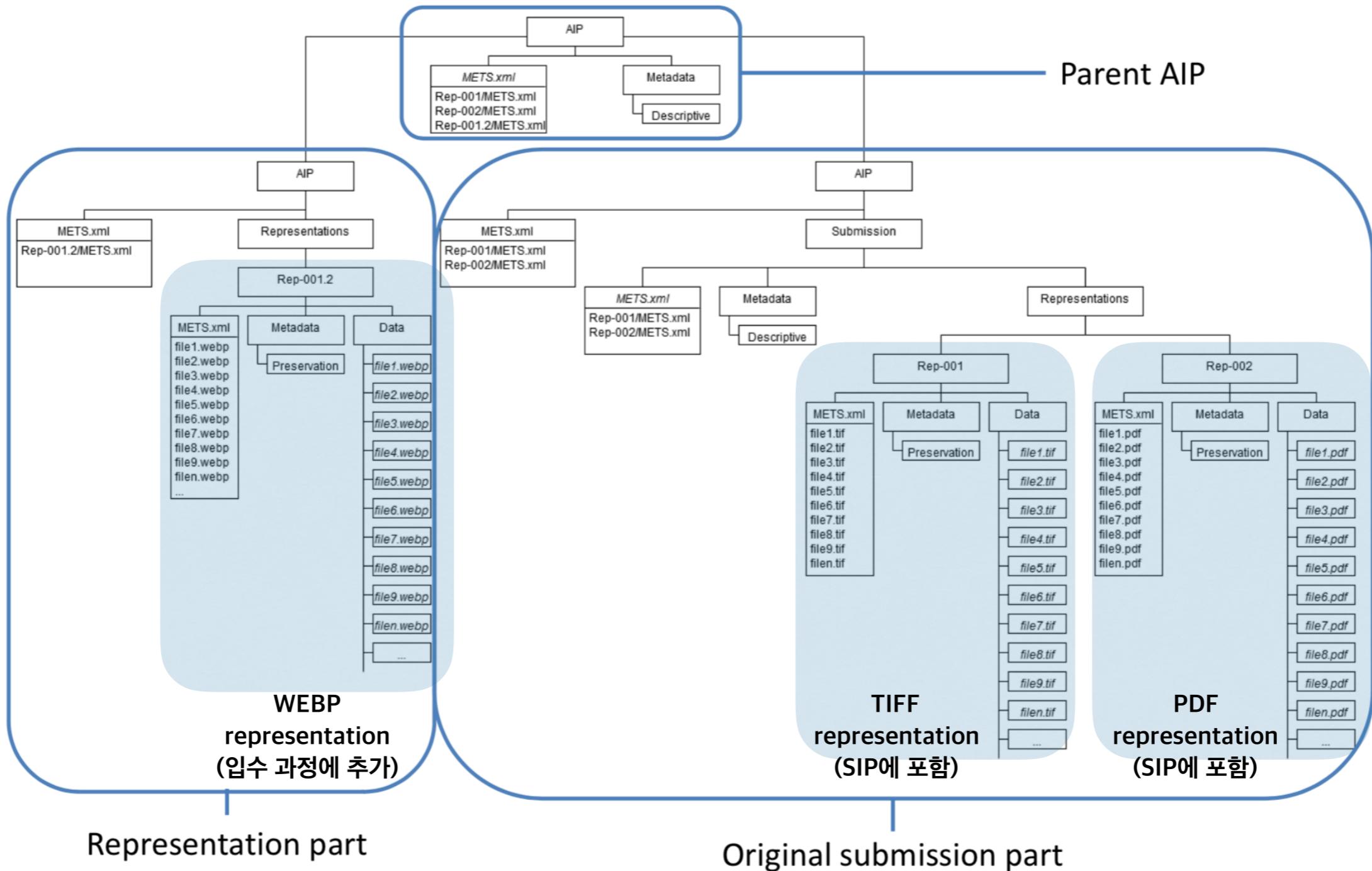
<그림> OAIS의 보존정보패키지 개념도

1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : 정보 패키지 (E-Ark)

E-Ark는 SIP, AIP, DIP 명세서와 변환 툴 제공(변환 워크플로우 커스터마이징 가능)



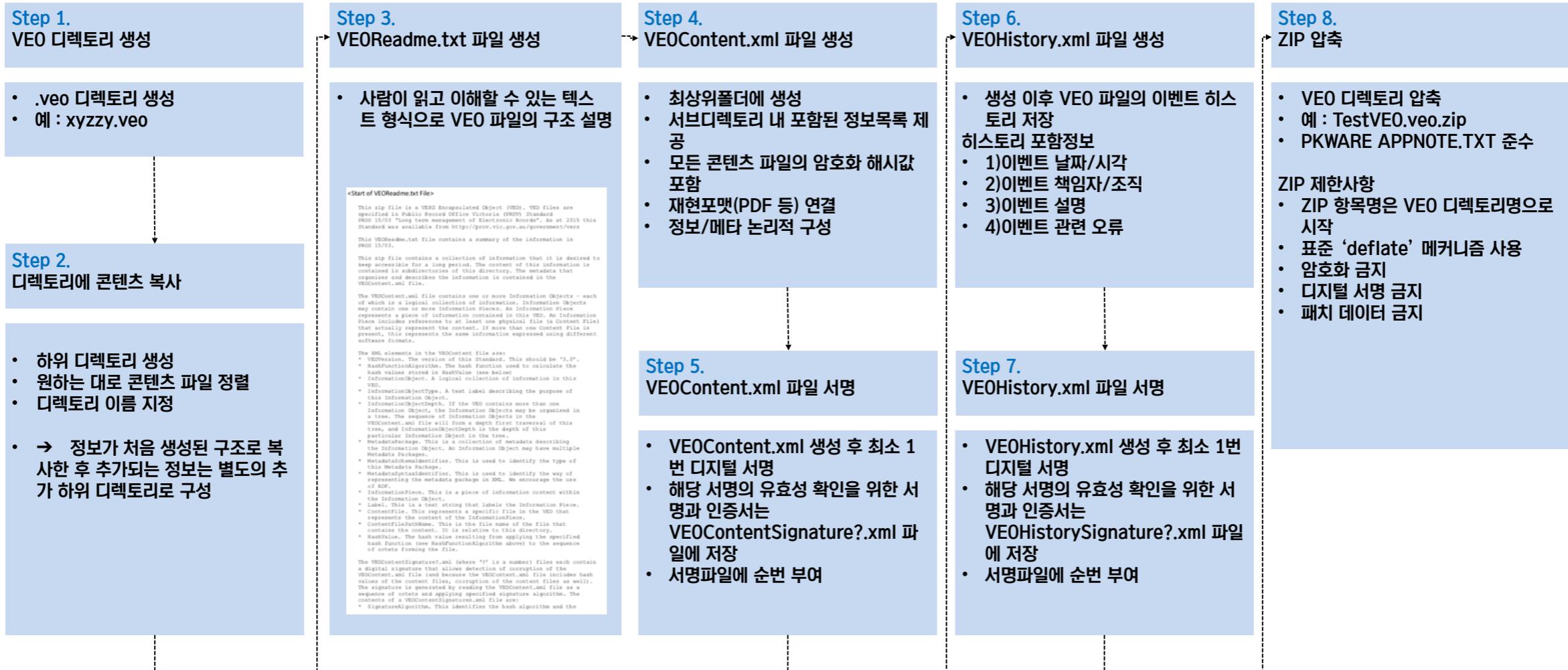
1. 표준, 툴, 서비스

1) 표준

OAIS 해석 및 구현 : 정보 패키지 (VEO3)

VEO3 생성과정은 8단계를 거침

VEO3는 콘텐츠와 메타데이터 디렉토리 전체를 ZIP으로 인캡슐레이션



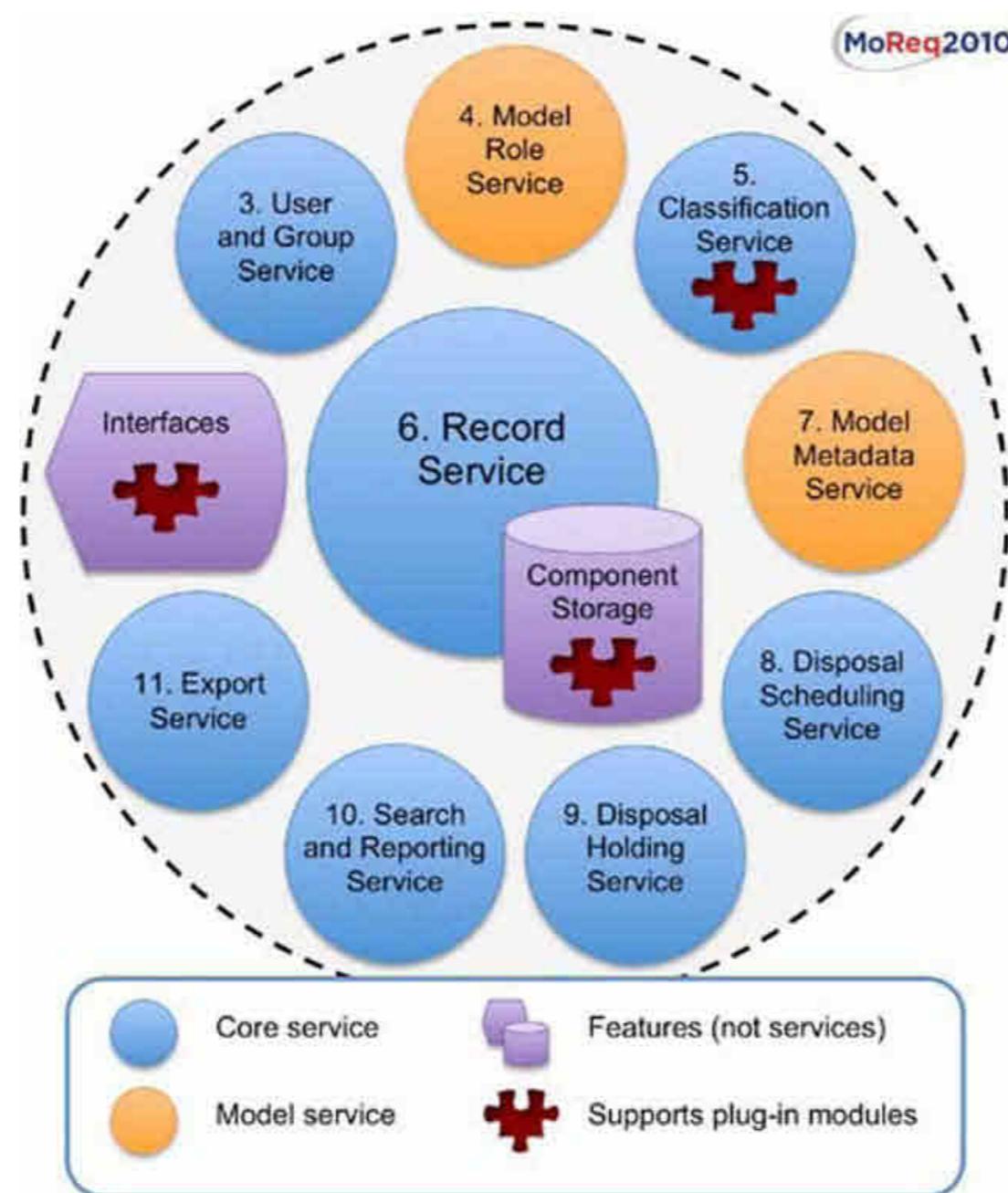
1. 표준, 툴, 서비스

1) 표준

MoReq2010

DLM Forum 등이 주도적으로 개발한 기록시스템 기능요건(장기보존과 구분)

MoReq2010은 모듈화, 집합체(Aggregation), 메타데이터와 콘텐츠 구분, 콘텐츠 단위 식별과 통제, 진본성 확인체계 등 기록시스템의 이상적인 개념 제시

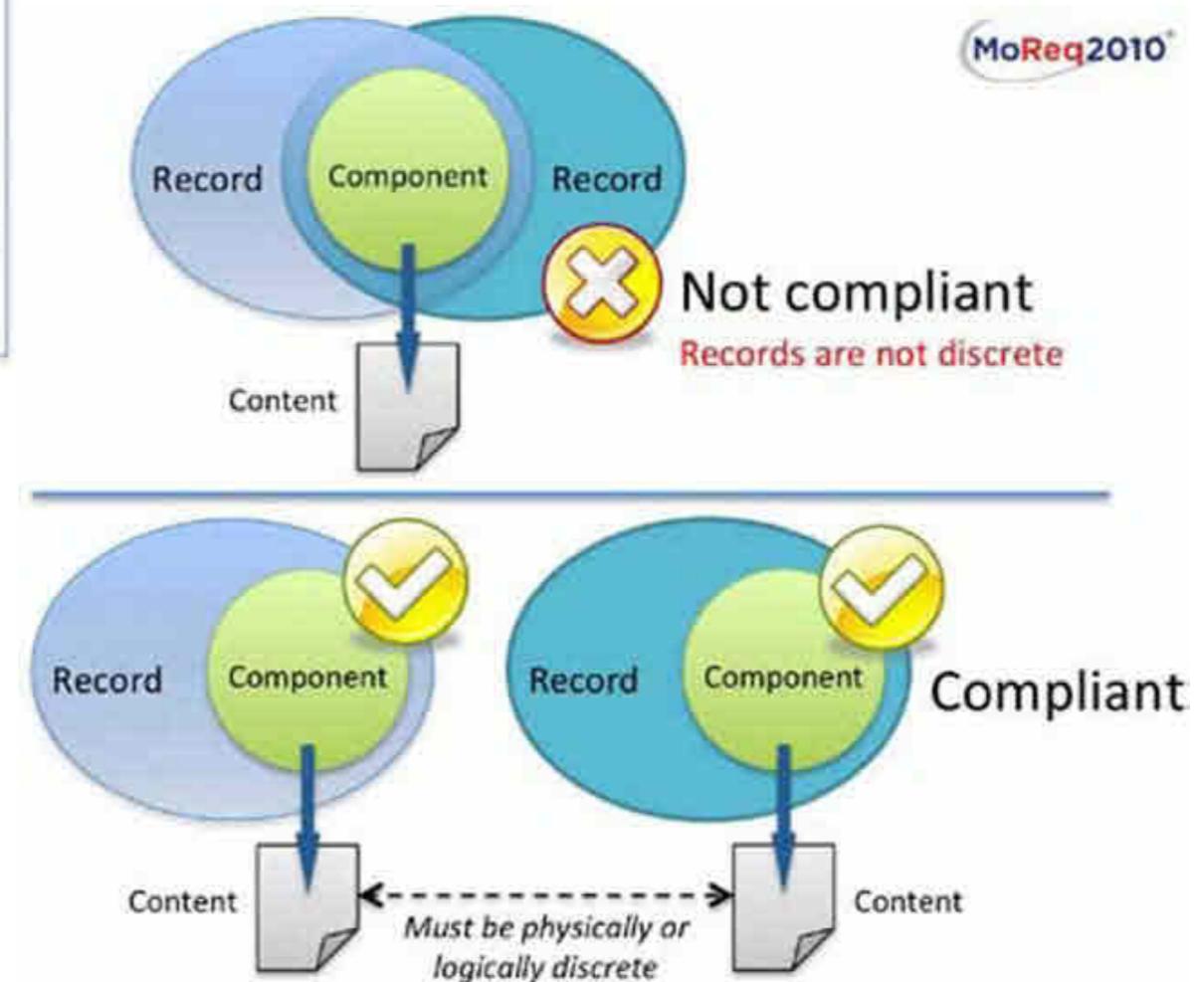
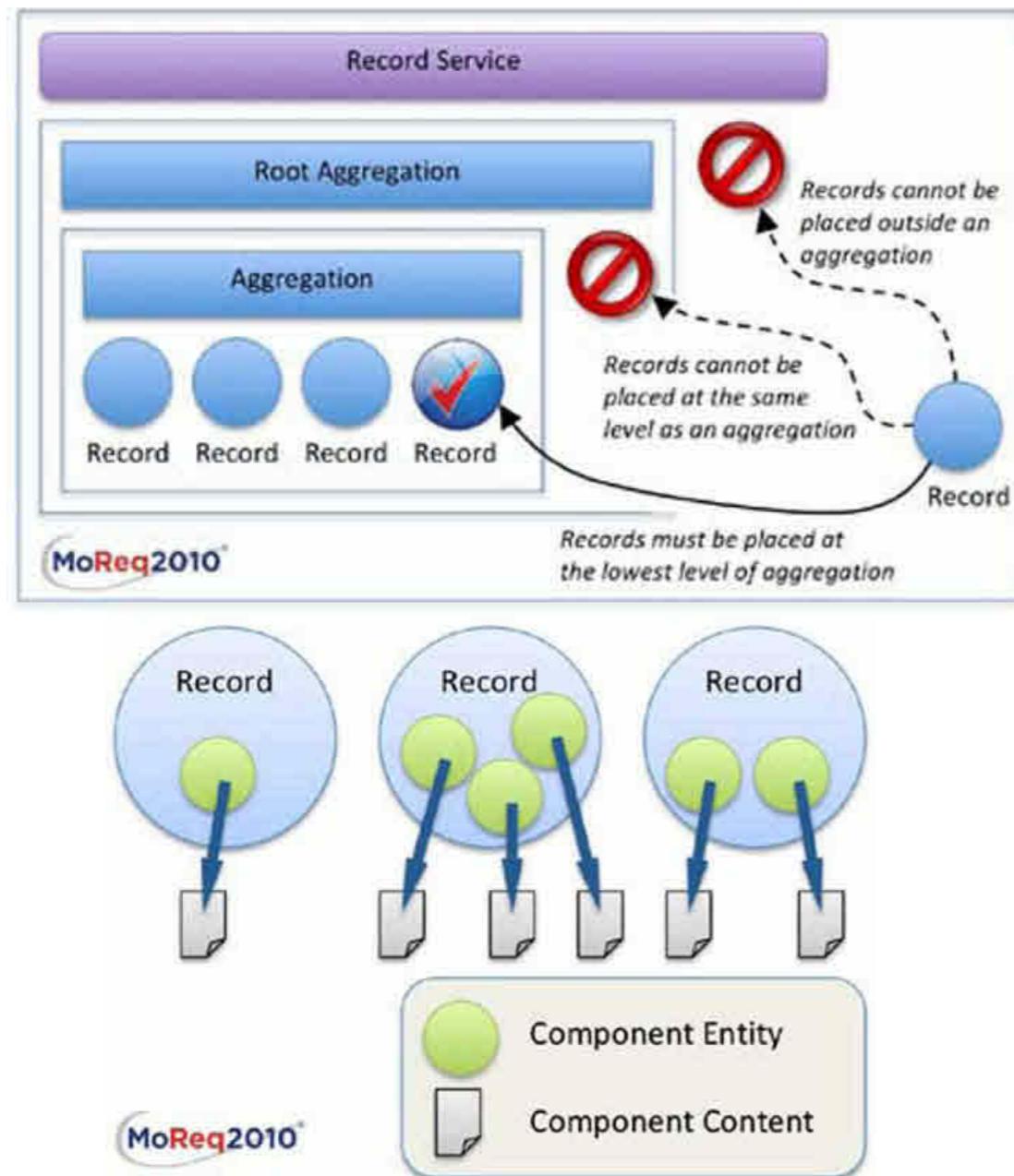


1. 표준, 툴, 서비스

1) 표준

MoReq2010

집합체(Aggregation)와 레코드, 컴포넌트, 콘텐츠 개념은 RM과 AM 단계에 공통 적용 가능



1. 표준, 툴, 서비스

2) 툴

OAIS와 ISO 16363 컴플라이언스 측정 : Langley, 2018

최근 ISO 14721과 ISO 16363 표준을 기반으로 디지털 보존 수준을 평가/인증하는 다양한 지표들이 개발됨

2018년에는 이 지표들의 항목을 비교한 연구가 수행됨(Langley, 2018)

- Langley는 신뢰저장소(TDR), 성숙도, 위험관리, 워크플로우, DPOC 등 5개 분야 12개 평가지표의 123개 요건을 크로스워크 분석
- 보존계획 수립 및 시스템 설계 시 디지털 보존 정책, 전략, 계획 등을 총체적으로 고려하여 시스템 요구사항 개발 필요

| Categories | Trusted Digital Repository (TDR) Certification | | | Digital Preservation Maturity Models | | | | | Risk Management Model | | Digital Preservation Workflow Models | DPOC Requirements |
|--|---|---|--|---|---|---|--|---|---|--|---|---|
| Module | CoreTrustSeal (CTS) formerly Data Seal of Approval (DSA) harmonised with World Data System (WDS) | Trusted Repositories Audit and Certification (TRAC) | Nester Seal (based on the DIN J1644 standard) | Digital Preservation Capability Maturity Model (DPCMM) (Dollar & Ashley) | National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation | Three-Legged Stool for Digital Preservation (Survey of Institutional Readiness) | Assessing Organisational Readiness (AOR) formerly Assessing Institutional Digital Assets (AIDA) | Five Organisational Stages of Digital Preservation (Kenney & McGovern) | Digital Repository Audit Method Based on Risk Assessment (CLOCKSS) Threats Model and Mitigation Strategy (DRAMBORA) | Controlled Lots Of Copies Keeps Stuff Safe (CLOCKSS) Threats Model and Mitigation Strategy (Langley) | Digital Stewardship End-to-End Workflow Model (Langley) | Digital Preservation at Oxford and Cambridge (DPOC) |
| Context | Context – RD | | | | Background Information | | | | | | | |
| Acknowledge | | | | | | | | Organisational Stages | | | | |
| Act | | | | | | | | Acknowledge: Understanding that digital preservation is a local concern | | | | |
| Consolidate | | | | | | | | Act: Initiating digital preservation projects | | | | |
| Institutionalise | | | | | | | | Consolidate: Scaling from projects to programs | | | | |
| Externalise | | | | | | | | Institutionalise: Incorporating the larger environment | | | | |
| Identify Organisational Context | | | | | | | | Externalise: Embracing inter-institutional collaboration and dependency | | | | |
| Document Policy and Regulatory Framework | | | | | | | | Risks Stages | | | | |
| Identify Activities, Assets and Owners | | | | | | | | Stage 1 - Identify Organisational Context | | | | |
| Identify Risks | | | | | | | | Stage 2 - Document Policy and Regulatory Framework | | | | |
| Assess Risks | | | | | | | | Stage 3 - Identify activities, assets and their owners | | | | |
| Manage Risks | | | | | | | | Stage 4 - Identify Risks | | | | |
| Mission | Organisational Infrastructure | Organizational Infrastructure | | | Organisational Infrastructure | | | Stage 5 - Assess the Identified Risks | | | | |
| Mandate | Mission/Scope – R1 | | | | Mission | ORG 01: Mission Statement | | Stage 6 - Manage Risks | | | | |
| Goals | | | | | | | | T1: What is the mandate of your repository or the organisation in which it is embedded? | | | | |
| Policy | Procedural accountability & policy framework | | | Policy | Policies and Procedures | ORG 02: Policies and Procedures | | T2: List goals and objectives of your repository | | | | |
| Strategy | | | | Strategy | | | | T3: List your repository's strategic planning documents | | | | |
| Governance | Governance & organizational viability | | | Governance | | | | | | | | |
| Collaboration | | | | Collaboration | | | | | | | | |
| Authority | | | | Authority | ORG 03: Sharing of Digital Content | | | | | | | |
| Implementation | | | | Implementation | | | | | | | | |
| Licences | Licences – R2 | Contracts, Licenses and Liabilities | CB Legal and contractual basis | | ORG 10: Intellectual Property Rights and Rights Management | | | | | | | Access Rights |

디지털 보존 평가지표 크로스워크

출처: Langley, S. (2018). Digital Preservation Maturity Model, Risks, Trusted Digital Repository Certification and Workflow Crosswalk [Dataset]. <https://doi.org/10.17863/CAM.34266>

1. 신뢰저장소

- CTS
- TRAC
- Nester Seal

2. 성숙도

- DPCMM
- NDSA레벨
- Three-Legged Stool for DP
- AOR
- Five Organisational Stages of DP

3. 위험관리

- DRAMBORA
- CLOCKSS

4. 워크플로우

- Digital Stewardship End-to-End Workflow Model

5. DPOC 요건

- DPOC

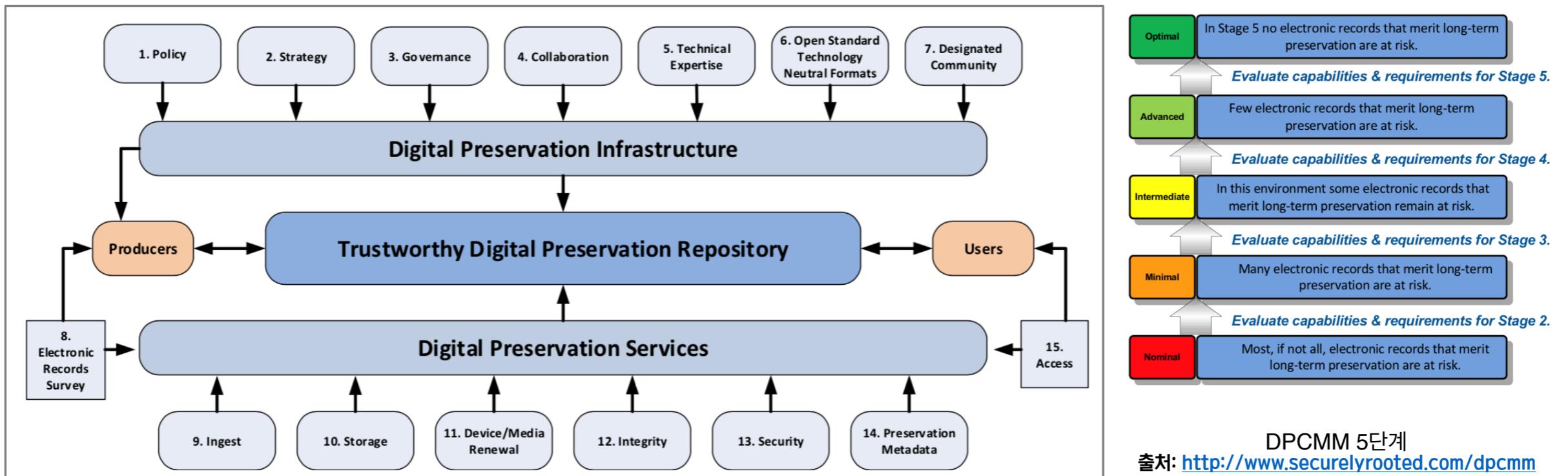
1. 표준, 툴, 서비스

2) 툴

OAIS와 ISO 16363 컴플라이언스 측정 : DPCMM

DPCMM(Digital Preservation Capability Maturity Model)은 OAIS와 ISO 16363 두 표준의 추상성을 15개 구성요소로 상세화하여 실무담당자 스스로 평가 및 지속 모니터링이 가능하도록 함

- OAIS(ISO 24721)와 신뢰저장소인증기준(ISO 16363)의 보존 요구사항을 측정하는 메트릭(xlsx 서식) 제공



DPCMM 개념도
출처: <http://www.securelyrooted.com/dpcmm>

1. 표준, 툴, 서비스

2) 툴

OAIS와 ISO 16363 컴플라이언스 측정 : Drupal TRAC Review Tool

최근엔 TRAC 등 평가모델을 실행할 수 있는 웹 기반 소프트웨어나 호스팅 서비스가 배포됨

- MIT 도서관에서 개발 및 배포, Artefactual에서 호스팅 서비스 제공 (Drupal 7.5 버전 기반)
- ISO 16363의 체크리스트를 온라인 구현하여 각 요구사항별 조직의 역할을 통해 TRAC 준수 여부를 5단계 점수(0~4점)로 평가

The diagram illustrates the integration of the CCSDS Audit and Certification of Trustworthy Digital Repositories document (Magenta Book) with the DRAC Review Tool. Two blue arrows point from the document to the tool's interface. The left arrow points to the 'Status Summary' section, which displays compliance ratings for various organizational infrastructure and management sections. The right arrow points to the 'Requirement Status' section, which provides detailed status for specific requirements under the 'Governance and Organizational Viability' category.

CCSDS Audit and Certification of Trustworthy Digital Repositories (Magenta Book, September 2011)

DRAC Review Tool Interface

- Status Summary:** Shows Total Requirements and Average Compliance Rating for sections like Organizational Infrastructure, Digital Object Management, and Infrastructure and Security Risk Management.
- Requirement Status:** Provides detailed status for requirements under categories such as Governance and Organizational Viability, Organizational Structure and Staffing, and Procedural Accountability and Preservation Policy Framework.
- Responsible:** Lists roles including Senior Management, Coordination Group, Operations Group, Information Technology, Administration: Finance or HR, Acquisitions, Preservation, Dissemination, Rights Management, and External Advisory Group.
- Accountable:** Lists roles including Senior Management, Coordination Group, Operations Group, Information Technology, Administration: Finance or HR, Acquisitions, Preservation, Dissemination, Rights Management, and External Advisory Group.

TRAC 리뷰 툴 –개요 페이지
출처: Dan Gillean(2017). Digital Preservation with Archivematica: An Introduction.

TRAC 리뷰 툴 –실행 페이지
출처: Dan Gillean(2017). Digital Preservation with Archivematica: An Introduction.

1. 표준, 툴, 서비스

2) 툴

소프트웨어 레지스트리 : COPTR, POWRR

COPTR : 52개 기능별 600여 개 툴/서비스 목록 제공

POWRR : 24단계 보존 프로세스별 66개 툴 매핑

| | <i>Access, Use and Reuse</i> | <i>Create or Receive (Acquire)</i> | <i>Cross-Lifecycle Functions</i> | <i>Dispose</i> | <i>Ingest</i> | <i>Preservation Action</i> | <i>Store</i> |
|------------------------------------|------------------------------|------------------------------------|----------------------------------|----------------|---------------|----------------------------|--------------|
| Audio | 2 | 4 | 3 | | 11 | 10 | |
| Binary Data | | | 4 | | | | |
| Container | | | | | | 5 | |
| Database | | 1 | 3 | | 3 | 13 | |
| Disk Image | | 6 | 3 | | 2 | 1 | |
| Document | 3 | 1 | 4 | | 33 | 14 | |
| EBook | | | | | 5 | 2 | |
| Email | | | 5 | | 2 | 4 | 1 |
| Geospatial | | | | | 1 | | |
| Image | 2 | 2 | 3 | | 23 | 23 | |
| Project Management Data | 1 | | | | | 1 | |
| Research Data | 2 | 8 | 13 | | 4 | | 16 17 |
| Software | | | | | 1 | 2 | |
| Spreadsheet | | | | | 6 | 3 | |
| Video | 1 | 3 | 1 | | 8 | 8 | |
| Web | 3 | 21 | 3 | | 7 | 3 | 1 1 |
| ~Not Content Type Specific~ | 18 | 37 | 82 | 9 | 65 | 58 | 29 49 |

| Digital POWRR Tool Evaluation Grid | Copy | Ingest | | Processing | | Access | | Storage | | Maintenance | | Other | | |
|---|------|--------------|-------------|----------------|-----------------------|------------------------|------------------|-------------------|-------------------|-------------------|------------|--------------------------------------|--|-------------------|
| | | Fixity Check | File Dedupe | Auto Unique ID | Auto Metadata Harvest | Auto Metadata Creation | Public Interface | Auto SIP Creation | Auto AIP Creation | Auto DIP Creation | Redundancy | Reliable, Long-Term Bit Preservation | Geographically Dispersed Data Storage Node | Cost |
| ACE (Audit Control Environment) | | x | | | | | | | | | | | | Free |
| AFF Open Source Computer Forensics Software | | | | | | | x | | | | | | | Free |
| Amazon S3 | x | x | | x | x | | | | | x | x | x | x | Varies |
| Archive-It | x | x | x | x | x | x | x | x | x | x | x | x | x | Varies |
| **Archivematica | x | x | x | x | x | x | x | x | x | x | x | x | x | Free |
| BagIt Library | x | | x | x | | | x | | x | | x | | | |
| BagIt Transfer Utilities | | | | | | | | | | | | | | Free |
| BitCurator | x | x | | x | x | x | x | x | x | | | | | free |
| BWF MetaEdit | x | | | | | x | x | | | | | | | free |
| Carbonite | x | x | | | | | | | | x | x | x | x | \$149 – \$599 |
| Chronopolis | x | x | x | x | x | x | x | x | x | x | x | x | x | \$1,500 – \$2,200 |
| Cinch | x | x | x | x | x | x | | | | | | | x | Free |
| ContextMiner | | | | | | | | | | | | | x | Free |
| **Curator's Workbench | x | x | | x | x | x | x | x | x | | | | x | Free |
| DAITSS | x | x | x | x | x | | x | x | x | x | x | x | x | |
| DCape (ingest only) | x | x | x | x | x | x | x | x | x | x | x | x | x | |
| DataVerse | x | | x | | x | x | x | x | x | | | | x | Free |

COPTR, <http://www.digipres.org/tools/#dcc-ref>

POWRR, <http://digitalpowrr.niu.edu/tool-grid/>

1. 표준, 툴, 서비스

2) 툴

포맷 검증 : VeraPDF

유럽연합의 PREFORMA 프로젝트에서는 PDF 인코더별 변환 품질의 문제를 해결하고자 PDF 검증도구인 veraPDF를 개발하여 오픈소스로 배포

- 리포트 유형 : Validation, Features, Validation and Features, Policy 중 선택
- PDF/A 유형 선택 : ISO 19005 (PDF/A)의 세부항목 선택
- 실행 결과 메시지 (2번에서 선택한 요건의 적합여부)
- 리포트 저장 및 열람 (XML 또는 HTML)
- ISO 19005 요건별 적합여부 및 Failed의 원인 설명

The image shows the veraPDF Conformance Checker application interface on the left and its generated validation report on the right, connected by a large blue arrow.

Application Interface (Left):

- File, Configs, Help** menu bar.
- veraPDF logo**.
- Message:** "Please choose a PDF and a Validation Profile, then press "Execute"
- Input:** "/Users/daejinan/Downloads/대통령기록관의 역할과 활동방향.pdf".
- Buttons:** "Choose PDF", "Choose Profile", "Choose Policy".
- Report type dropdown:** "Validation and Features" (highlighted with red box).
- Message:** "Policy file not chosen".
- Validation Result:** "PDF file is not compliant with Validation Profile requirements" (highlighted with red box).
- Output Options:** "**XML**" and "**HTML**" (highlighted with red boxes).
- Policy dropdown:** "Auto-detect" (highlighted with red box), followed by a list of PDF/A flavours: Auto-detect, PDF/A-1A, PDF/A-1B, PDF/A-2A, PDF/A-2B, PDF/A-2U, PDF/A-3A, PDF/A-3B.

Validation Report (Right):

- Header:** veraPDF
- Title:** Validation Report
- File:** /Users/daejinan/Downloads/대통령기록관의 역할과 활동방향.pdf
- Validation Profile:** PDF/A-1B validation profile
- Status:** Failed
- Statistics:**

| | |
|-------------------------|---------------------------|
| Version: | 1.12.1 |
| Parser: | GreenField |
| Build Date: | 2018-05-08T22:06:00+09:00 |
| Processing time: | 00:00:02.868 |
| Total rules in Profile: | 103 |
| Passed Checks: | 201227 |
| Failed Checks: | 626 |
- Validation information:**

| Rule | Status |
|---|----------------|
| Specification: ISO 19005-1:2005, Clause: 6.2.3, Test number: 3 DeviceCMYK may be used only if the file has a PDF/A-1 OutputIntent that uses a CMYK colour space 402 occurrences | Failed Show |
| Specification: ISO 19005-1:2005, Clause: 6.1.3, Test number: 1 The file trailer dictionary shall contain the ID keyword. The file trailer referred to is either the last trailer dictionary in a PDF file, as described in PDF Reference 3.4.4 and 3.4.5, or the first page trailer in a linearized PDF file, as described in PDF Reference F.2 1 occurrences | Failed Show |

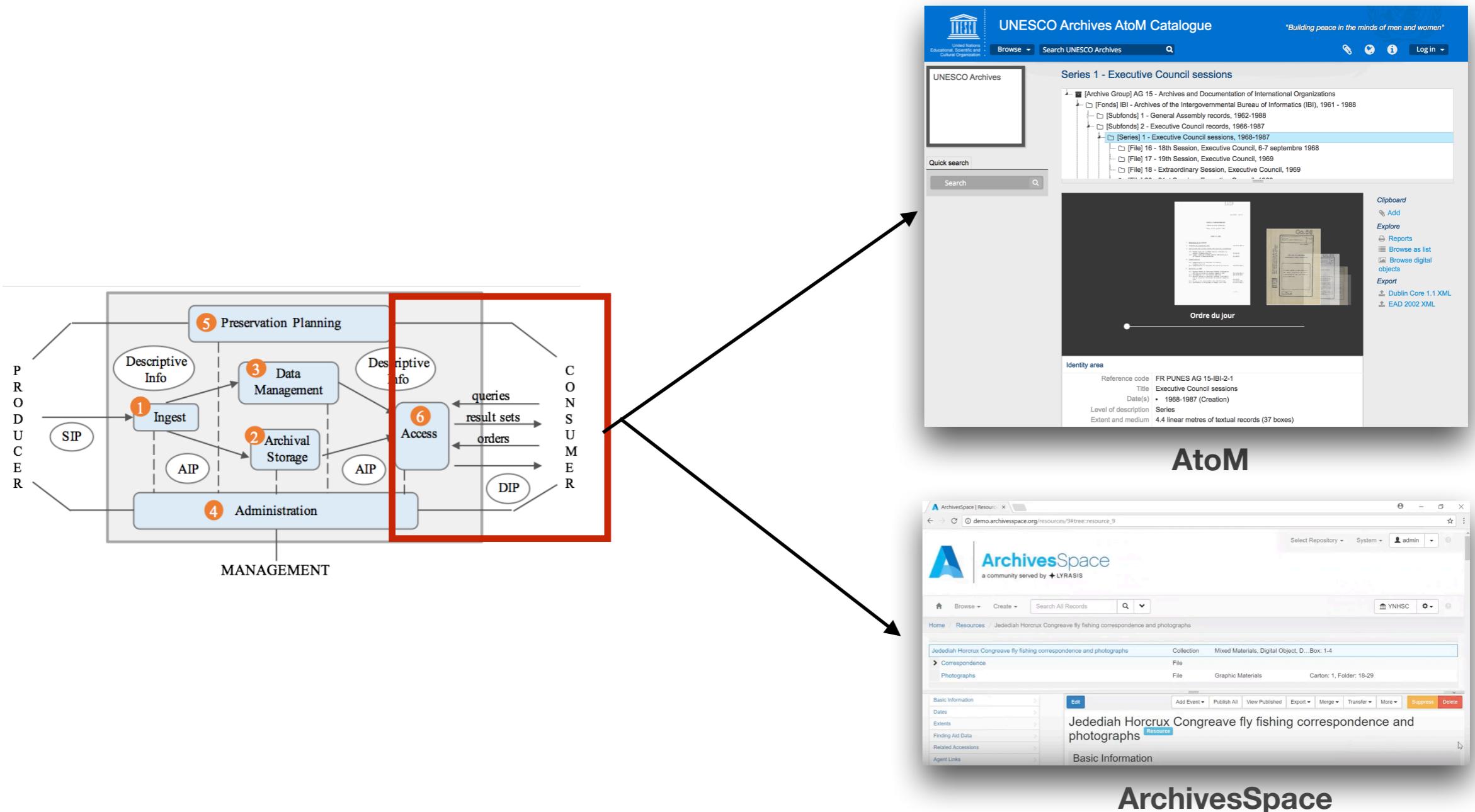
1. 표준, 툴, 서비스

2) 툴

카탈로깅 : AtoM, Rosetta, ArchivesSpace, CollectiveAccess 등

디지털 보존시스템과 카탈로깅시스템 연계 활용(ex. Archviematica to AtoM, Preservica to ArchivesSpace)

카탈로깅 시스템 : 국제/국가기록물기술표준, 더블린 코어 등을 지원하는 온라인 카탈로깅 소프트웨어



1. 표준, 툴, 서비스

2) 툴

가상화 : Docker, Kubernetes 등

도커 등 컨테이너 기술은 OAIS를 지원하는 툴을 가상화하여 서비스 형태로 제공

- Docker : 컨테이너 기술
- Kubernetes : 컨테이너 관리(오케스트레이션) 툴

Docker Client

- 도커 조작 인터페이스(CLI)

Dockerfile

- 도커 이미지 조합을 위한 명령어
(텍스트 파일)

Images

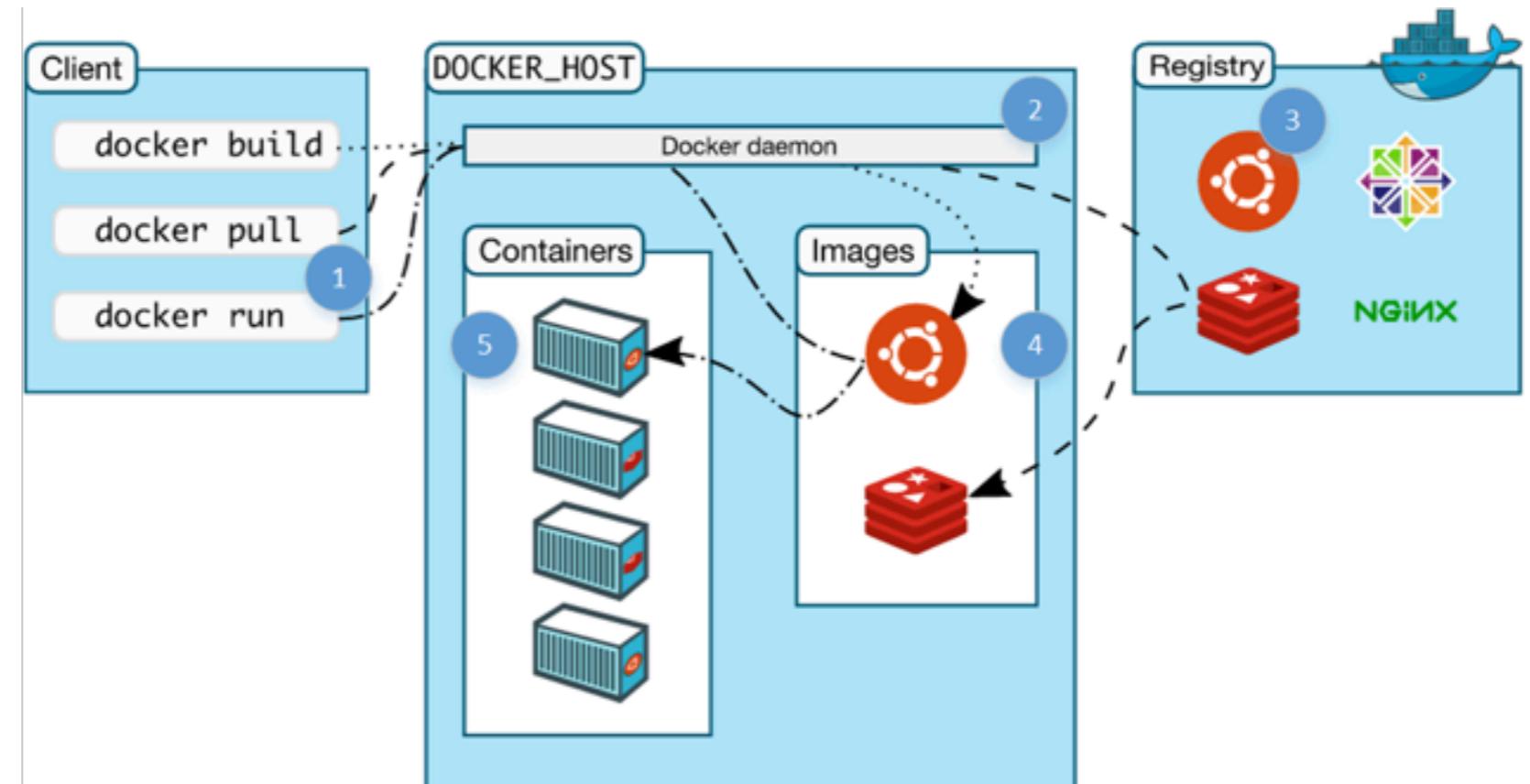
- 도커파일로부터 정의된 도커 이미지(OS, App, DBMS 등)

Containers

- 도커 이미지 실행 인스턴스

Registry

- 도커 이미지 저장소

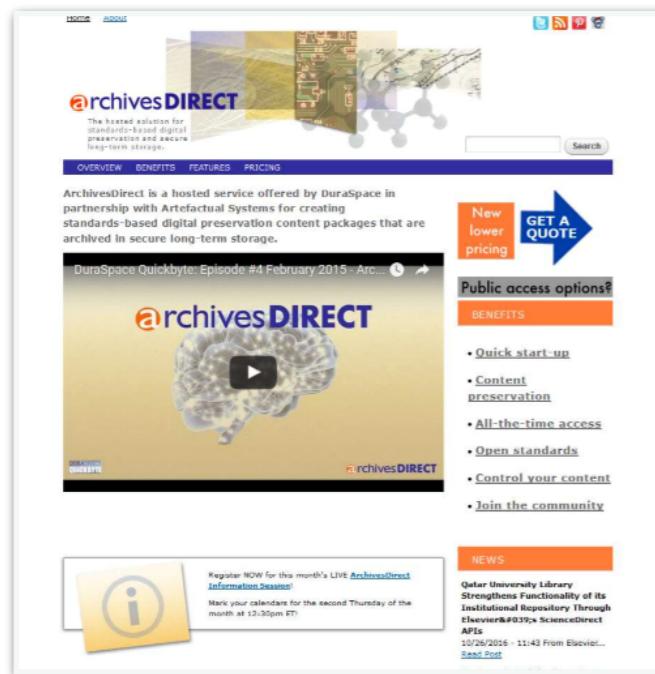


1. 표준, 툴, 서비스

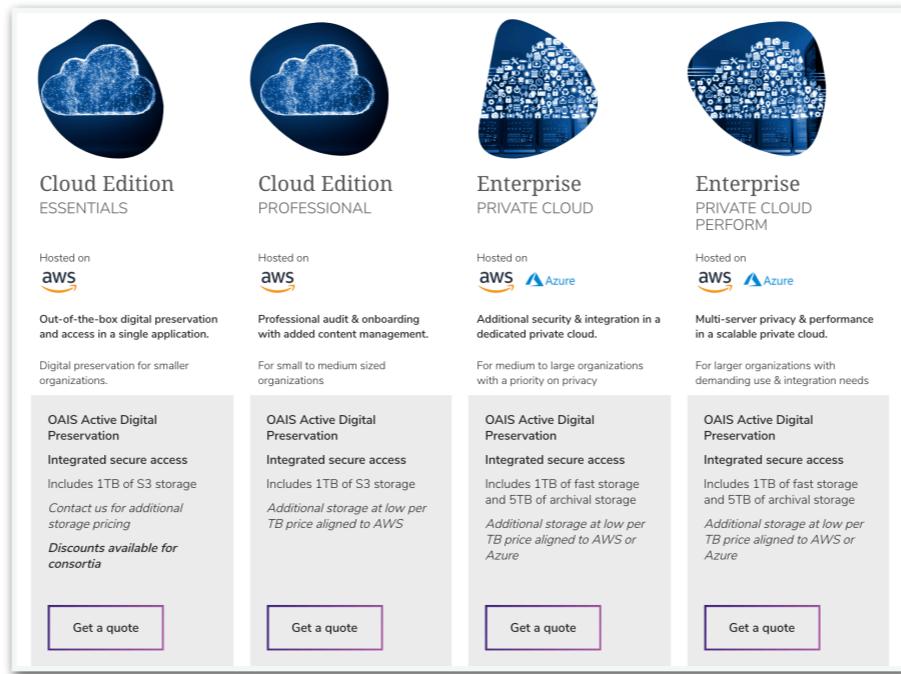
3) 서비스

디지털 보존시스템

유럽, 북미의 기록소장기관을 중심으로 OAIS 요건을 충족하는 디지털 보존시스템, 저장소 등을 클라우드 형태의 구독 서비스로 활용



ArchivesDIRECT
(Archivematica 디지털 보존 서비스)



Preservica
(디지털 보존 서비스))



DuraCloud
(클라우드 스토리지 서비스)

*출처: 국가기록원. 2017. 차세대기록관리모델 재설계.

1. 표준, 툴, 서비스

3) 서비스

웹 아카이빙

2010년 이후 웹 아카이빙 기관들은 웹 크롤링, 보존, 재현 등 웹 아카이빙 업무를 전문 서비스 구독 형태로 전환

- Archive-It : 미국 샌프란시스코(웹/소셜미디어)
- ArchiveSocial : 미국 노스캐롤라이나(소셜미디어)
- Erado : 미국 댈러스(웹/소셜미디어)
- MirrorWeb : 영국 맨체스터(웹/소셜미디어)

The screenshot shows the Archive-It dashboard for a monthly crawl. It displays a summary of 24.1 GB total data, 473,832 total docs, and 453,771 new docs. A circular chart shows the distribution of data by host, with 24,640 hosts contributing to 24.1 GB. Below this is a bar chart of top hosts by document count.

The screenshot shows the ArchiveSocial interface for an entry titled "Visit North Carolina". It includes the original timestamp (2012-06-07T16:25:03Z), file type (POST), and status (signed; Indexed). It also shows the URL of the original news article and a thumbnail image.

The screenshot shows the MirrorWeb dashboard. It displays statistics for a crawl: 1,254 URLs, 109 size (MB), and a publish date of 02 Mar 2018. The 'CRAWL DETAILS' section shows a completed crawl from March 2nd, 2018, at 11:09PM. The 'LOGS' section lists various log files like alerts.log, crawl.log, and frontier_recover.gz. A preview of the archived website is shown at the bottom.

- 인터넷 아카이브(IA, Internet Archive)에서 제공하는 클라우드 웹 아카이빙 서비스
- 이용 기관들은 사용자 친화적인 웹 기반 인터페이스와 자동화된 품질평가도구 등을 이용하여 쉽게 웹 자원을 수집, 보존, 관리할 수 있음
- 수집 과정에서의 인덱싱을 통해 보존된 자원의 풀 텍스트 검색 제공
- Archive-It을 통해 수집된 웹 자원은 인터넷 아카이브에 축적되며 Wayback-machine 툴을 통해 일반인들에게 공개됨
- 2016년 5월 현재 전 세계 16개국의 도서관, 박물관, 아카이브, 정부기관, 자치단체, 비영리단체 등 459 기관이 파트너로 참여하여 3,487개 컬렉션을 수집함
- Internet Archive와의 협약을 통해 서비스 이용이 가능하며 클라우드 저장 용량에 따라 1년 단위 과금

- 미국 노스캐롤라이나 주 더럼(Durham) 소재의 웹 아카이빙 서비스 제공자로 정부기관, 법률기관, 금융기관, 교육기관 대상 웹 아카이빙 서비스 제공
- Facebook, Twitter, Instagram, Youtube 등 소셜 미디어 아카이빙 기능에 특화되어 있으며 오바마 대통령의 소셜 미디어 아카이브(Obama White-House Social Media Archive) 구축 사례로 널리 알려짐
- 월별 웹 기록 수집 건 수에 따라 절약형, 표준형, 고급형의 요금제를 제공하며 경제형의 경우 10개의 사이트만을 아카이빙할 수 있고 모든 요금제에 무제한 저장 용량 제공

- 영국 맨체스터 소재의 웹 아카이빙 서비스 제공자로 정부기관, 금융기관, 기업 웹페이지 및 소셜 미디어 아카이빙 서비스 제공
- eDiscovery 지원 – MirrorWeb을 통해 아카이빙된 웹 기록을 eDiscovery 전문가, 변호사, 기타 조사원들이 검색/활용하는 기능 제공
- 영국국립기록관리청(TNA), 영국은행(Bank of England) 등 유럽의 다수 정부기관 활용
- Facebook, Twitter, Youtube, LinkedIn, Instagram, Vimeo 등 소셜미디어 6종 지원

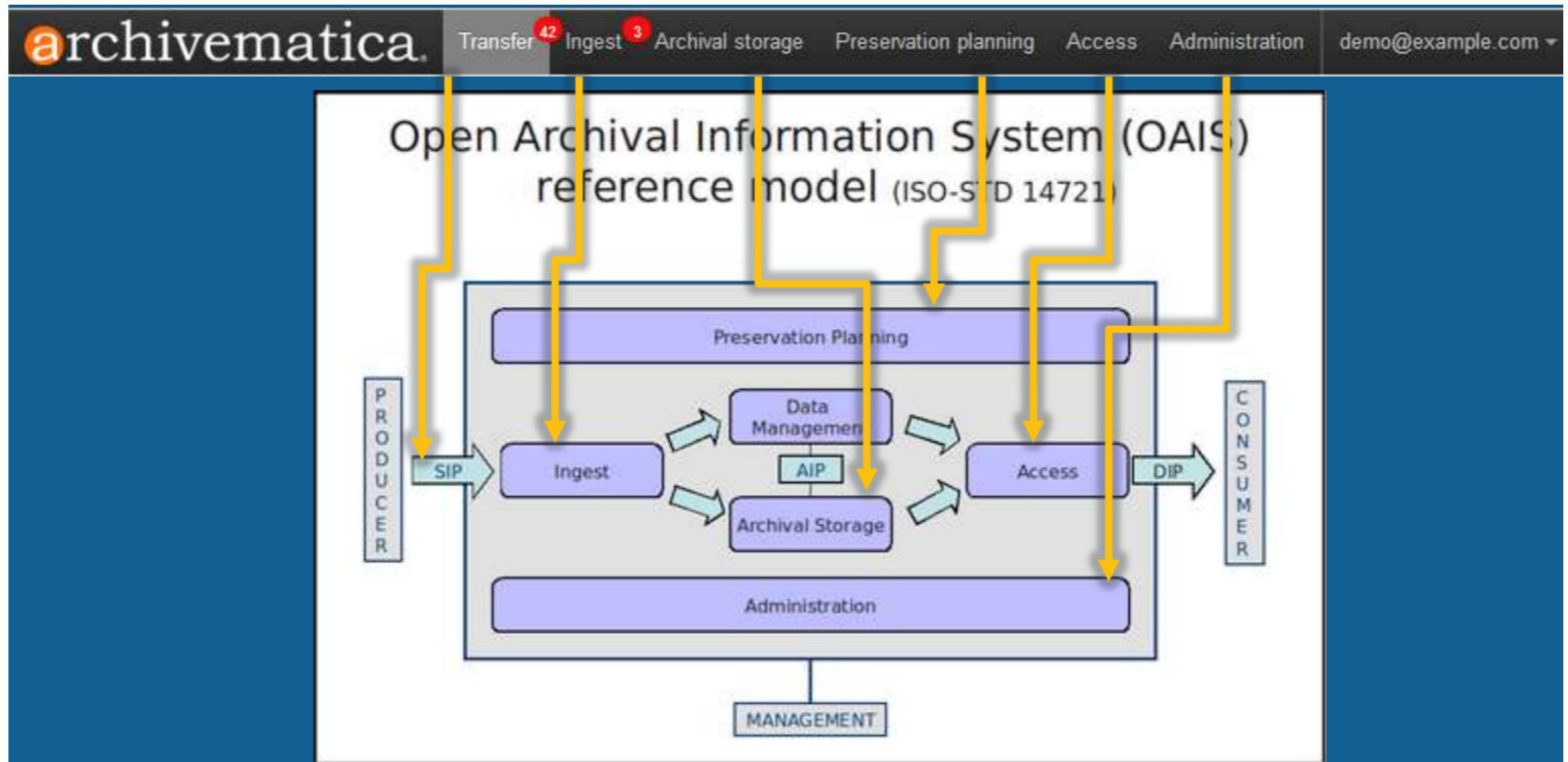
*출처: 대통령기록관. 2019. 대통령기록 정보시스템 고도화 ISP

2. 시스템

1) Archivematica

개요

UNESCO 등의 펀딩으로 2012년 캐나다의 Artefactual 사에서 개발
OAIS의 주요 기능 엔티티를 시스템의 메뉴로 구성
(Ingest, Archival Storage, Preservation Planning, Access, Administration)



2. 시스템

1) Archivematica

사용자

주로 북미지역 대학도서관, 국제기구, 미술관/박물관/아카이브에서 사용



2. 시스템

1) Archivematica

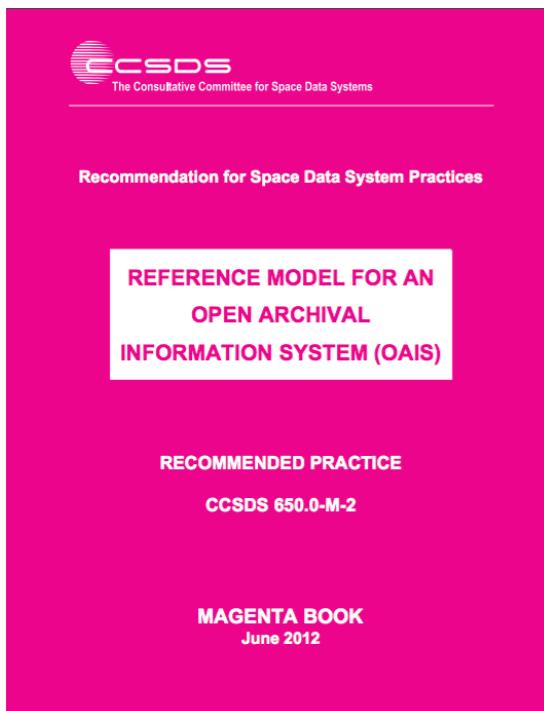
특징

표준 : OAIS, PREMIS, METS, BagIT 등 국제기록관리 커뮤니티 표준 적용

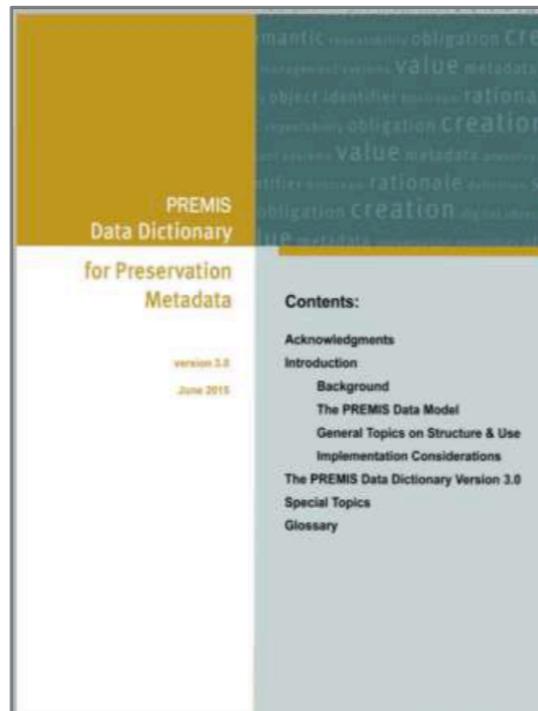
오픈소스 : 기 개발된 30여 종의 오픈소스 활용

FPR 서비스 : 수백 종의 파일포맷 식별 및 처리

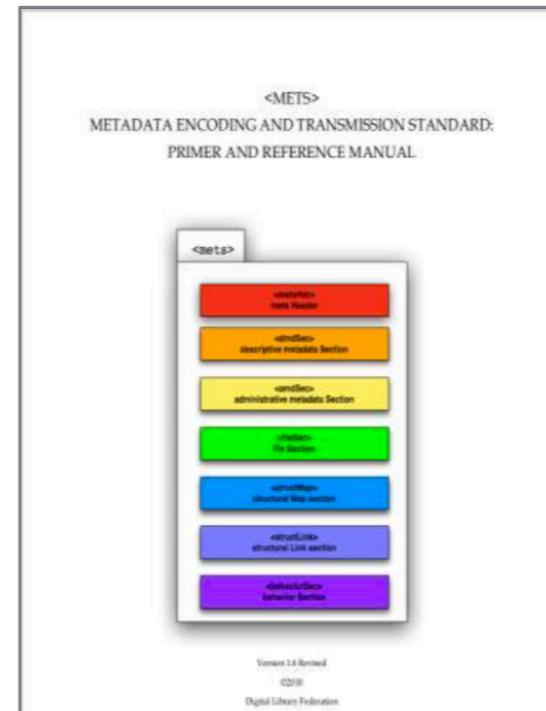
マイ크로서비스 : 각 기능을 세분하여 독립적 처리



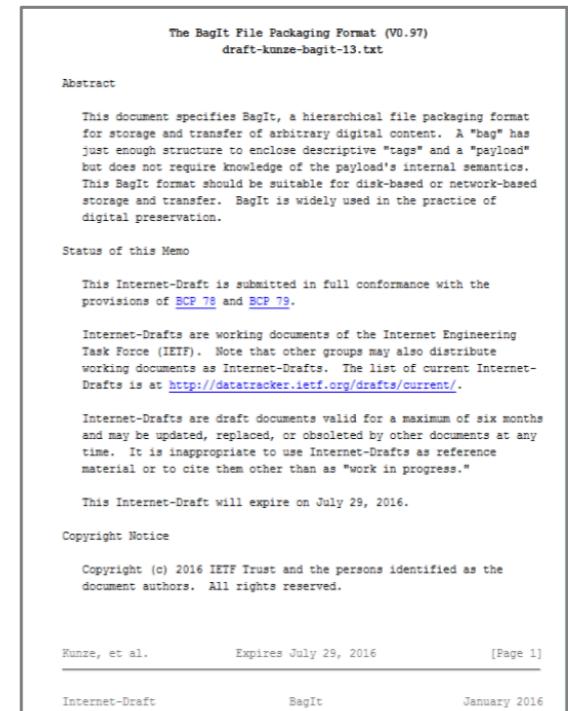
ISO 14721 (OAIS)



PREMIS



METS



BagIT

2. 시스템

1) Archivematica

마이크로서비스 아키텍처

전자기록 입수부터 제공까지의 각 기능을 마이크로서비스 수준까지 잘게 나누어 처리

마이크로서비스는 각 기능의 독립적 실행, 배포, 운영, 개선이 가능하여 MoReq2010의 핵심개념을 실현

| Submission Information Package | UUID | Ingest start time | |
|---|--------------------------------------|-------------------|---|
| <ul style="list-style-type: none"><input checked="" type="checkbox"/> Test_transfer<ul style="list-style-type: none">▶ Micro-service: Store AIP▶ Micro-service: Upload DIP▶ Micro-service: Prepare AIP▶ Micro-service: Prepare DIP▶ Micro-service: Generate AIP METS▶ Micro-service: Verify checksums▶ Micro-service: Process metadata directory▶ Micro-service: Process submission documentation▶ Micro-service: Transcribe SIP contents▶ Micro-service: Add final metadata▶ Micro-service: Process manually normalized files▶ Micro-service: Normalize▶ Micro-service: Clean up names▶ Micro-service: Rename SIP directory with SIP UUID▶ Micro-service: Verify SIP compliance▶ Micro-service: Remove cache files▶ Micro-service: Include default SIP processingMCP.xml▶ Micro-service: Verify transfer compliance | b7d5f60d-3f40-4ac5-b49b-b8079c52be55 | 2016-10-06 14:50 |   |

2. 시스템

1) Archivematica

마이크로서비스 아키텍처

전자기록 입수부터 제공까지의 각 기능을 마이크로서비스 수준까지 잘게 나누어 처리

마이크로서비스는 각 기능의 독립적 실행, 배포, 운영, 개선이 가능하여 MoReq2010의 핵심개념을 실현

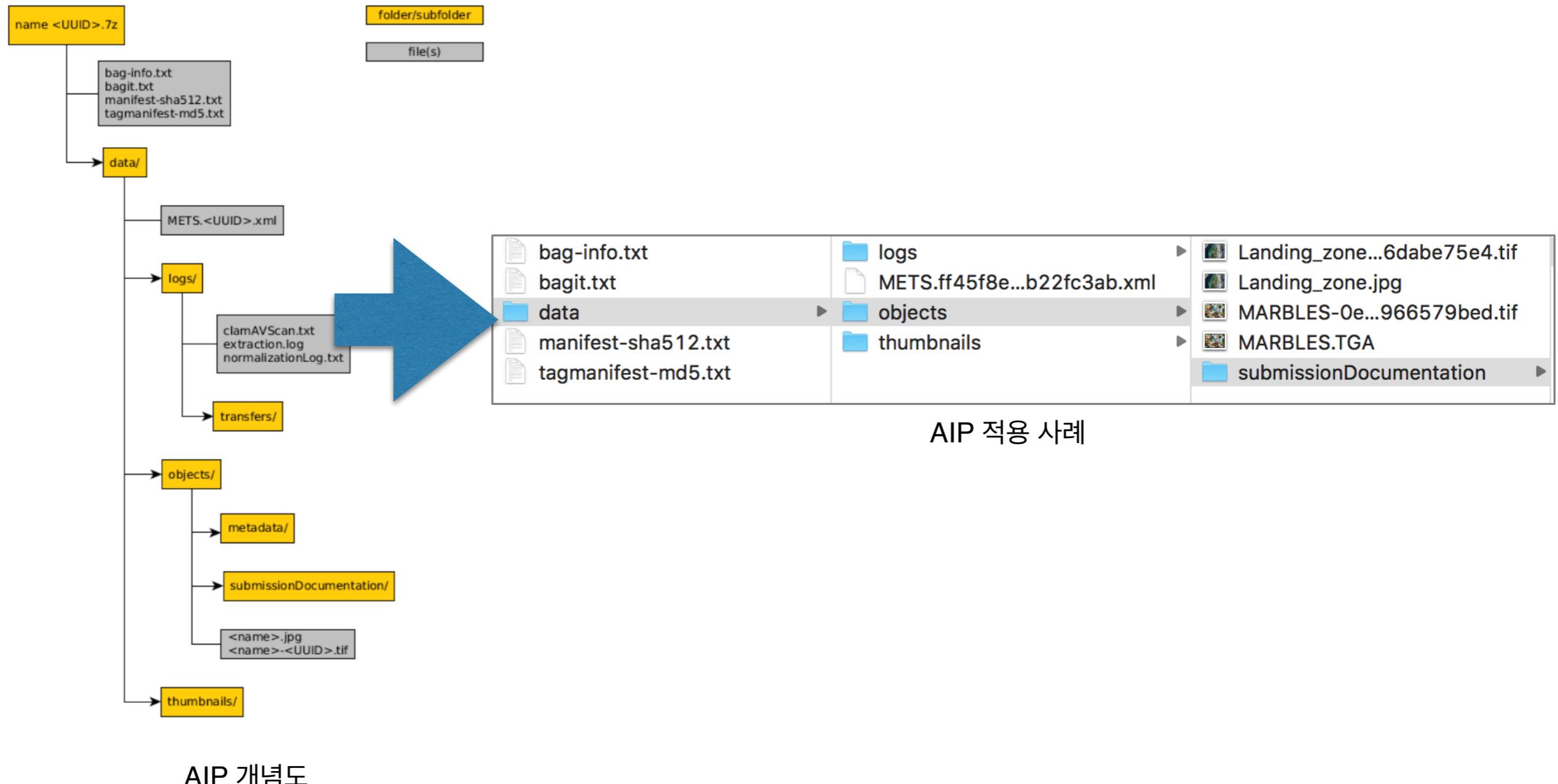
| ▼ Micro-service: Normalize | | |
|--|------------------------|--|
| Set file permissions | Completed successfully | |
| Move to processing directory | Completed successfully | |
| Approve normalization [?] | Completed successfully | |
| Move to approve normalization directory | Completed successfully | |
| Remove files without linking information (failed normalization artifacts etc.) | Completed successfully | |
| Set file permissions | Completed successfully | |
| Normalize for preservation | Completed successfully | |
| Normalize for access | Completed successfully | |
| Normalize for thumbnails | Completed successfully | |
| Create thumbnails directory | Completed successfully | |
| Create DIP directory | Completed successfully | |
| Move to processing directory | Completed successfully | |
| Normalize [?] | Completed successfully | |
| Resume after normalization file identification tool selected. | Completed successfully | |
| Identify file format | Completed successfully | |
| Select pre-normalize file format identification command | Completed successfully | |
| Move to select file ID tool | Completed successfully | |
| Grant normalization options for no pre-existing DIP | Completed successfully | |
| Set remove preservation and access normalized files to renormalize link. | Completed successfully | |
| Check for Access directory | Completed successfully | |
| Check for Service directory | Completed successfully | |
| Identify manually normalized files | Completed successfully | |

2. 시스템

1) Archivematica

SIP, AIP, DIP (정보 패키지)

Archivematica는 BagIt, METS, PREMIS 기반으로 SIP, AIP, DIP를 구현



2. 시스템

1) Archivematica

FPR(Format Policy Registry, 포맷정책 레지스트리)

FPR서버를 통해 파일 포맷의 식별, 특성 추출, 불변성체크(fixity), 포맷변환 수행

FPR은 디지털 보존 계획(Preservation Planning)의 핵심도구로 쓰임

The screenshot shows the Archivematica web interface for managing format policy rules. The top navigation bar includes links for Transfer, Backlog, Appraisal, Ingest, Archival storage, Preservation planning, Access, Administration, and support. The main title is "Format policy rules / List". On the left, there are filters for Formats (with a "Create new rule" link), Groups (showing 10 entries), and Identification (Tools, Rules, Commands). The main table lists rules categorized by tool (e.g., Tools, Rules, Commands) and format (e.g., Generic JPEG, 1987a, 1989a, Adobe Photoshop, Bitmap 3.0 NT, Canon CIFF raw image data, Canon CR2 Raw, Digital Negative Format (DNG)). Each row includes columns for Purpose, Format, Command, Success, Enabled, and Actions (View | Replace | Disable). The table shows 344 entries, with page navigation at the bottom.

| Formats | Create new rule | | | | | | | | | | | |
|------------------------|--------------------------------|-------------------------------|---------------------------------|------------|---------|--|---|---|---|-----|----|------|
| Groups | Show | 10 | ▼ entries | Search: | | | | | | | | |
| Identification | Purpose | Format | Command | Success | Enabled | Actions | | | | | | |
| Tools | Access | Generic JPEG | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Rules | Access | 1987a | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Commands | Access | 1989a | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Format policy registry | Access | Adobe Photoshop | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Tools | Access | Bitmap 3.0 NT | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Characterization | Access | Canon CIFF raw image data | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Rules | Access | Canon CR2 Raw | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Commands | Access | Digital Negative Format (DNG) | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Event Detail | Access | Digital Negative Format (DNG) | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Rules | Access | Digital Negative Format (DNG) | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Commands | Access | Digital Negative Format (DNG) | Transcoding to jpg with convert | 0 out of 0 | Yes | View Replace Disable | | | | | | |
| Extraction | Showing 1 to 10 of 344 entries | | | Previous | 1 | 2 | 3 | 4 | 5 | ... | 35 | Next |

Archivematica 의 포맷정책규칙(FPR)

출처: https://www.archivematica.org/en/docs/archivematica-1.9/_images/normalization-rules.png

2. 시스템

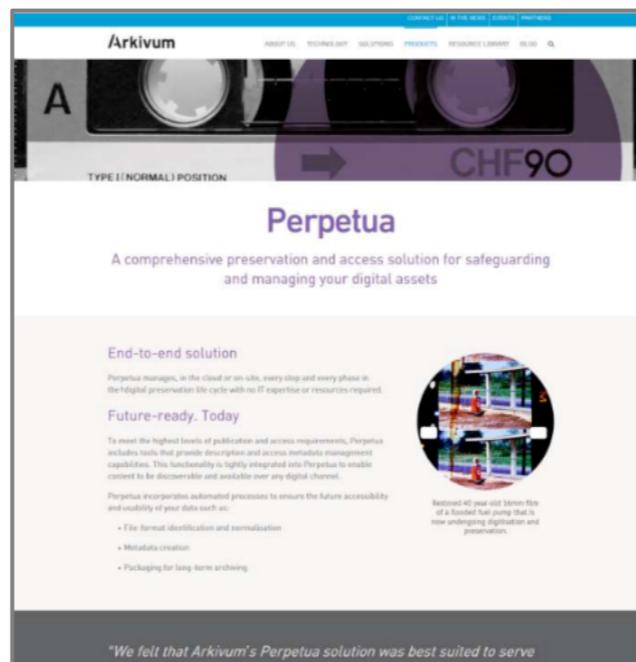
1) Archivematica

클라우드 서비스

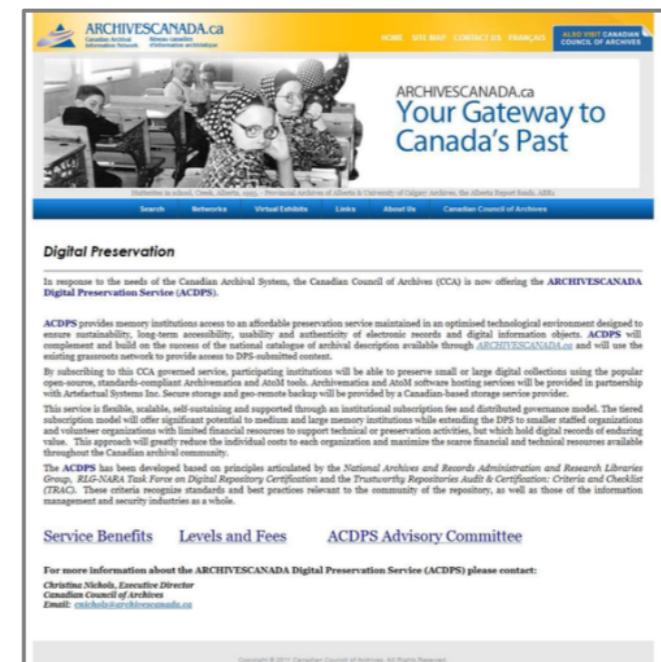
Artefactual사는 2014년 미국(ArchivesDIRECT), 2016년 영국(Perpetua)과 캐나다(ArchivesCANADA)의
로컬 클라우드 서비스 공급자와 제휴하여 클라우드 기반 디지털 보존 호스팅 서비스 제공



ArchivesDIRECT
(미국)



Perpetua
(U.K.)



ArchivesCANADA
(캐나다)

구독모델 (ArchivesDIRECT)

- **디지털보존 평가 플랜** : 3개월 컨설팅, - 500GB 제공 (\$4,500)
- **표준 플랜** : Archivematica 전기능 활용, - 1TB 제공 (\$9,999)
- **프로페셔널** : 대규모 또는 복잡한 구현

2. 시스템

2) Preservica

개요

TNA는 영국의 아카이브시스템 개발업체인 Tessela와 함께 Preservica를 개발
이후 Preservica는 이를 엔터프라이즈 버전으로 확장하여 유럽 다수의 내셔널 아카이브에 클라우드 서비스 제공

National & Pan-National



Libraries, Museums & Education



State & Government

Business & Corporate

2. 시스템

2) Preservica

OAIS 확장

Preservica는 OAIS의 모든 요건을 만족시킬 뿐만 아니라, 대부분의 업무를 워크플로우로 자동화하고, Access 등 아카이브 소장물 검색 및 이용자 편의기능을 대폭 향상시킴

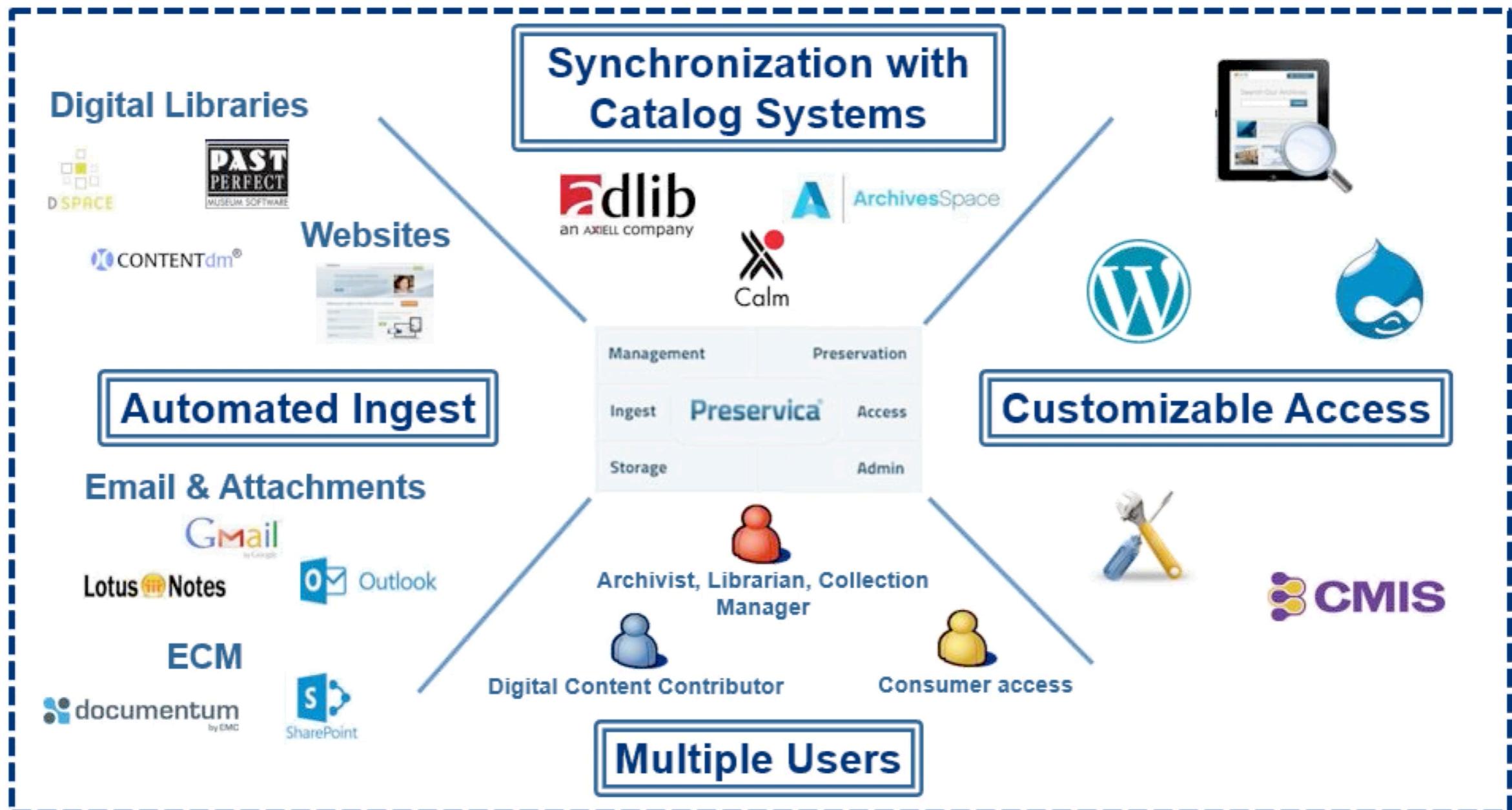


2. 시스템

2) Preservica

CMIS

주요 기록 생산시스템, 서비스시스템 등과 연결 인터페이스 제공

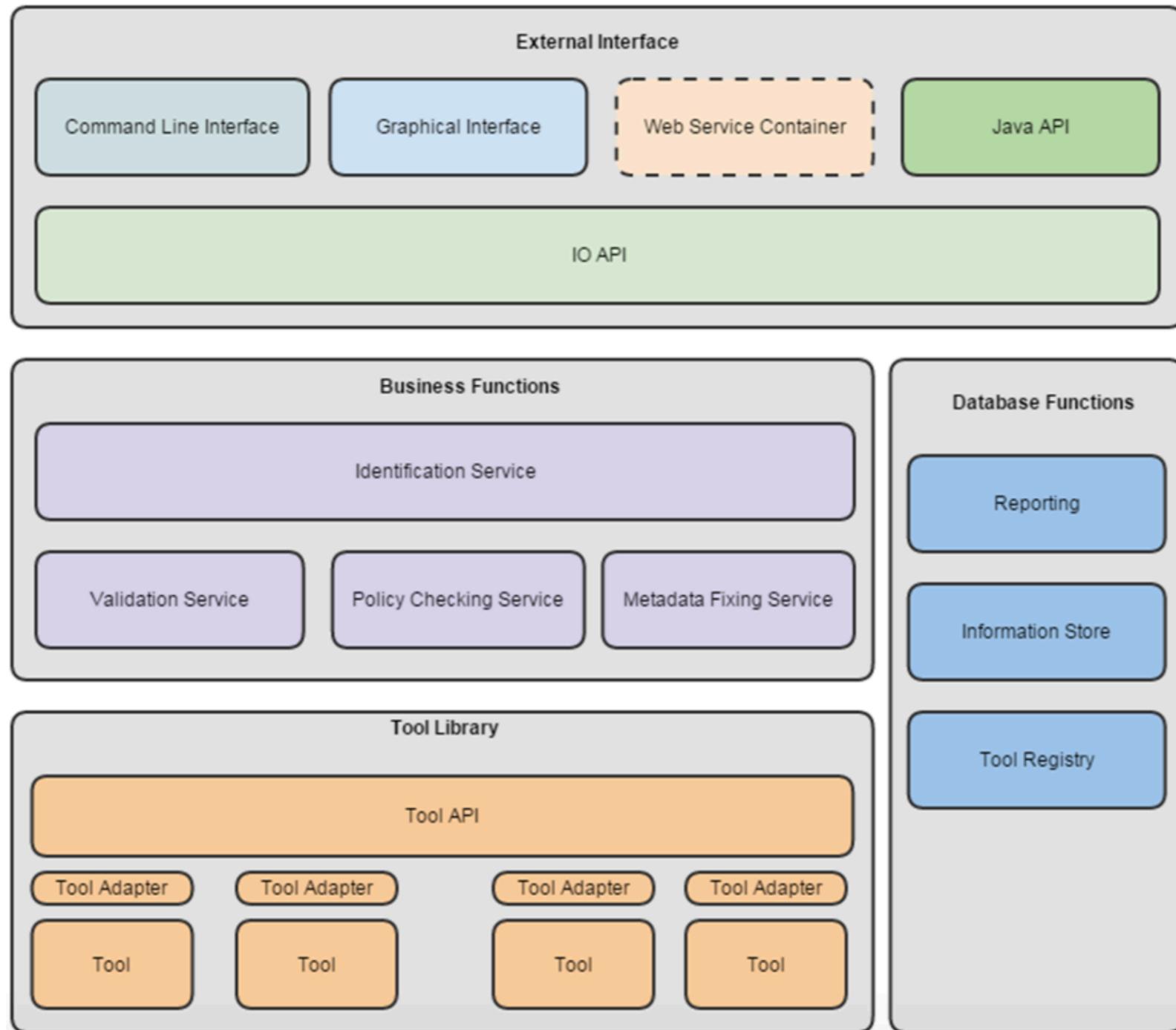


2. 시스템

2) Preservica

아키텍처

Tool Library, Business Function, Database Function, External Interface로 구성



2. 시스템

2) Preservica

워크플로우

아카이비스트는 개별 업무를 워크플로우로 정의하고, Preservica 시스템이 실행

The diagram illustrates the workflow process. On the left, a screenshot of the 'Workflow Definitions' interface shows a list of existing definitions, with a large blue arrow pointing to the right towards the execution interface. On the right, a screenshot of the 'Workflow Details' interface displays the context and progress of a specific workflow named 'Package Upload'. The 'Step Progress' table shows the status of various steps, all of which have been completed successfully.

Workflow Definitions (Left):

- Schema Management
- Transfer Agreements
- Workflow Definitions
- Security
- Reports
- About

Add New Workflow Definition No file selected.

Existing Definitions

| Name | Version | Uploaded By | Registered On | Active | Action |
|---|---------|-------------------|-------------------|-------------------------------------|---|
| Batch Ingest Workflow with Summary Email | 3 | System Admin User | 06.04.14 14:59:31 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| FTP Full Ingest Workflow | 6 | System Admin User | 06.04.14 14:59:44 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Full Ingest Workflow | 5 | System Admin User | 06.04.14 15:00:19 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Full Ingest Workflow (Delete Source XIP) | 3 | System Admin User | 06.04.14 15:00:12 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Full Ingest Workflow (Manual Selection) | 6 | System Admin User | 06.04.14 15:00:40 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow (Manual Upload) | 2 | System Admin User | 06.04.14 14:59:58 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow (No Virus Scan) | 3 | System Admin User | 06.04.14 11:05:33 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow with Preservation | 6 | System Admin User | 06.04.14 15:00:33 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow with Preservation and Forced Ingest | 8 | System Admin User | 06.04.14 15:00:26 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow with Preservation via Selected Pathways | 2 | System Admin User | 06.04.14 15:01:00 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Ingest Workflow with Preservation via Selected Pathways and Forced Ingest | 2 | System Admin User | 06.04.14 15:00:53 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Manual Batch Ingest Workflow with Summary Email | 1 | System Admin User | 06.04.14 14:59:37 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Post-Migration Ingest | 5 | System Admin User | 06.04.14 15:00:05 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Sharepoint Ingest Workflow (Manual Selection) | 2 | System Admin User | 06.04.14 14:59:51 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| Website Ingest Workflow (Full) | 6 | System Admin User | 06.04.14 15:00:46 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| ZIP Upload Ingest (Server-Side SIP Creation) | 4 | System Admin User | 06.04.14 15:01:07 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |
| ZIP Upload Ingest with Preservation | 2 | System Admin User | 06.04.14 15:01:14 | <input checked="" type="checkbox"/> | <input type="button" value="Download"/> |

Workflow Details (Right):

Workflow Context: Package Upload
Workflow Definition: Standard Ingest Workflow (Package Upload)
Workflow ID: 30299
Workflow State: Active
Date Started: 25.08.17 14:46:37
Date Finished: 8
Number of Files: 30 MB
Collection Code: 1234
Submission name: if
Top Level Record: if

Step Progress:

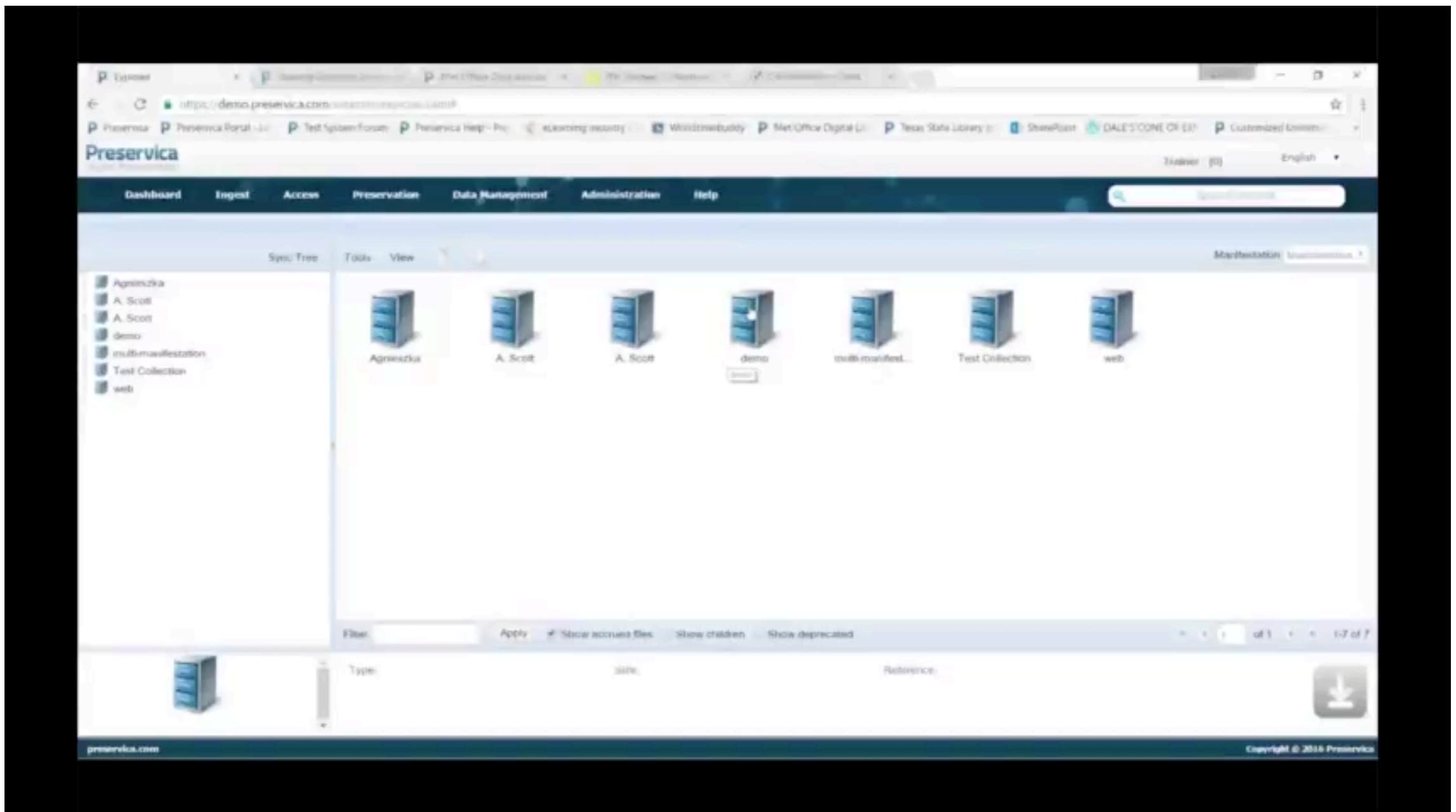
| State | Name | Progress | Started | Finished | Messages |
|-------|---|---|-------------------|-------------------|----------------------|
| | Upload Package File | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:46:37 | 25.08.17 14:48:48 | |
| | Process Package | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:48:49 | 25.08.17 14:48:54 | |
| | Virus Check | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:48:55 | 25.08.17 14:49:45 | |
| | Metadata Integrity | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:49:46 | 25.08.17 14:49:48 | |
| | Content Integrity | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:49:49 | 25.08.17 14:49:51 | |
| | Fixity Check | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:49:52 | 25.08.17 14:49:54 | |
| | SIP Validation | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:49:55 | 25.08.17 14:49:57 | |
| | SIP Validation with Database Crosscheck | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:49:58 | 25.08.17 14:50:00 | |
| | Characterise | <div style="width: 100%; background-color: green;"></div> | 25.08.17 14:50:01 | 25.08.17 14:50:06 | View |
| | Store Files | <div style="width: 100%; background-color: grey;"></div> | 25.08.17 14:50:07 | | |

워크플로우 정의 **워크플로우 실행**

2. 시스템

2) Preservica

시연 (입수 워크플로우)



2. 시스템

3) ERA 2.0

개요

NARA는 2017년부터 클라우드 기반의 ERA 2.0으로 고도화 추진 중
(소장기록 : 백악관, 의회 및 연방기관 전자기록(160억건/8억철, 400TB) 및 비전자기록)

ERA 1.0

개발시기 : 2005 ~ 2011 (예산 4천 억원, 록히드 마틴)

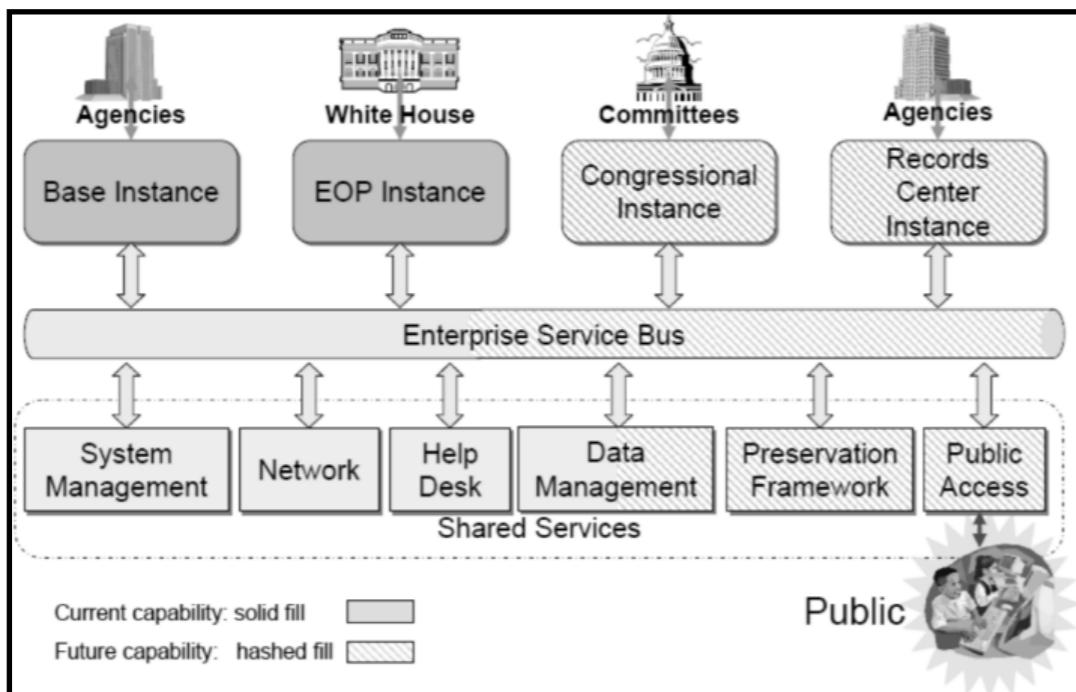
주요 이용자 : 백악관, 의회, 연방기관

개발배경 : 90년대 말 전자정부 도래로 2000년 클린턴 행정부 종료 시
방대한 전자기록 이관을 대비하여 ERA 프로그램 준비

기능 범위 : RM(전자/비전자) + AM(자동화된 영구기록관리)

아키텍처 : ESB 기반의 서비스 지향 아키텍처(SOA)

ERA와 각 기관 RMS의 관계 : ERA는 GRS와 패키징/이관 수행



<그림> ERA 1.0 개요

출처: Thibodeau, Kenneth(2009) "NARA's Electronic Records Archives Status and Plans"

ERA 2.0

개발시기 : 2017 ~ 2011

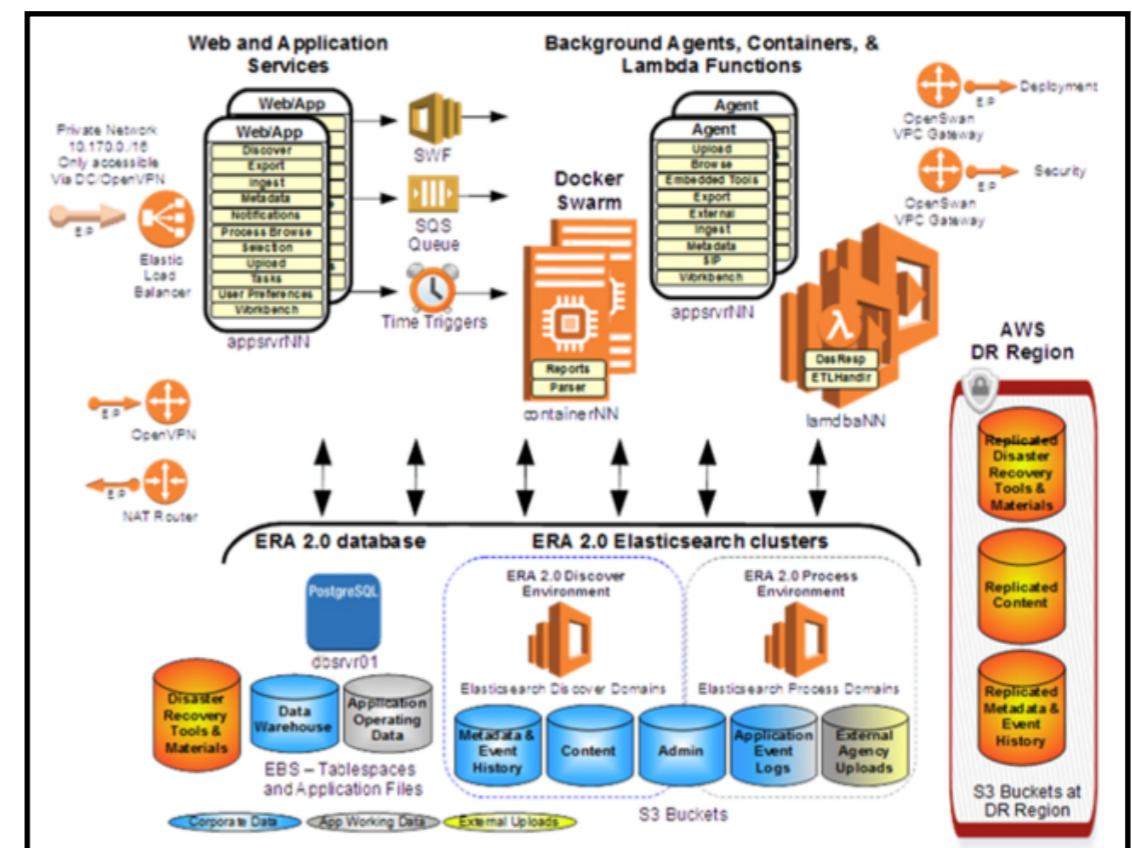
주요 이용자 : 백악관, 의회, 연방기관

개발배경 : 클라우드 환경의 기민한 개발 프로세스로 고도화

기능 범위 : RM(전자/비전자) + AM(자동화된 영구기록관리)

아키텍처 : 애자일 프로세스 기반의 마이크로서비스(MSA)

기능 아키텍처 : DPE(처리), DOR(보존), BOM(업무지원) 구분



<그림> ERA 2.0 개요

출처: Leslie Johnston (2018) "ERA 2.0: A Cloud-Based Preservation Repository at the NARA"

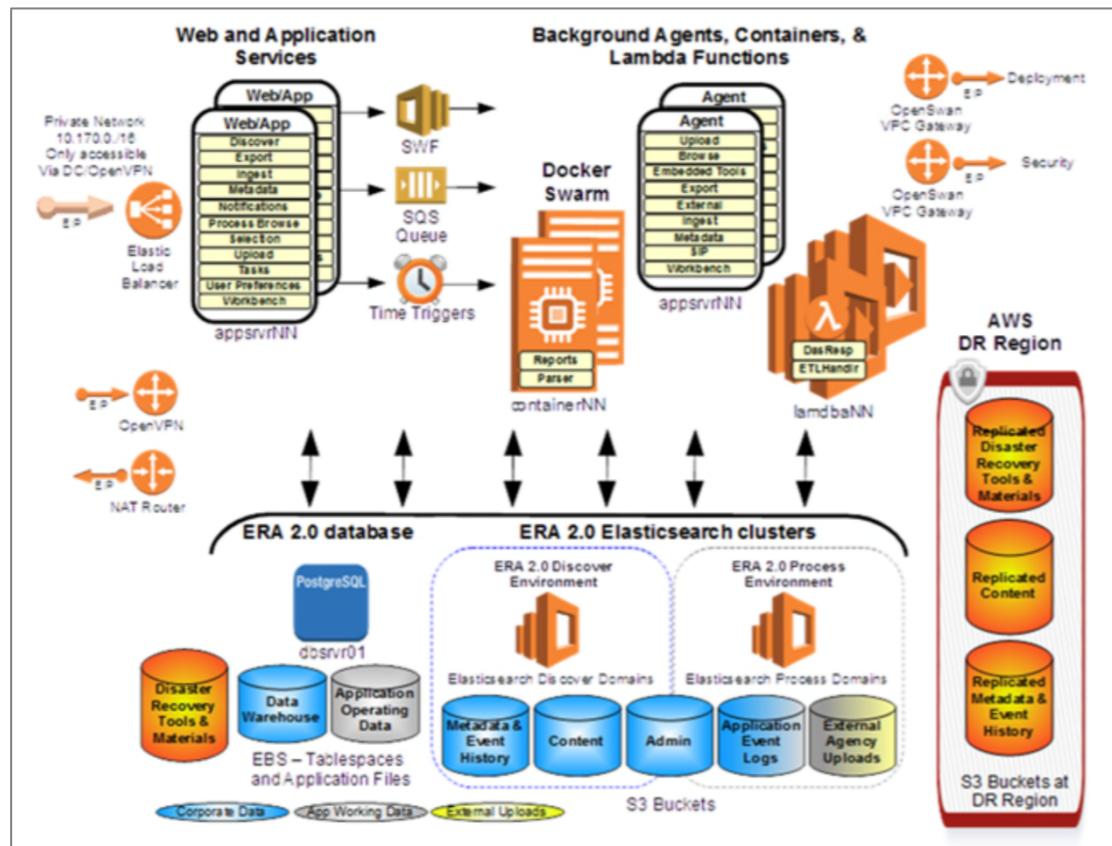
2. 시스템

3) ERA 2.0

특징

1. 모듈 구성 : 입수 단계를 별도 모듈로 분리

2. 클라우드 : 클라우드 컴퓨팅 기반의 인프라, MicroService 소프트웨어 아키텍처, 배포(Docker 컨테이너 기반) 등 차세대 기술로 전면 전환



<그림> ERA 2.0 아키텍처

출처: Leslie Johnston (2018) "ERA 2.0: A Cloud-Based Preservation Repository at the NARA"

DPE (입수)

- DPE(Digital Processing Environment)는 입수 단계에 필요한 소프트웨어 툴과 메타데이터 처리 등의 기능을 수행

DOR (보존)

- DOR(Digital Object Repository)은 불변성 검사, 평가, 리포팅 등 소장기록의 보존과 관리를 위한 모든 기능을 수행

BOM (업무지원)

- BOM(Business Object Management)는 아카이브의 업무를 지원하는 기능을 수행(현재 단일(Federal) 워크플로에서 멀티(Federal, Presidential, Legislative, Judicial)로 전환)

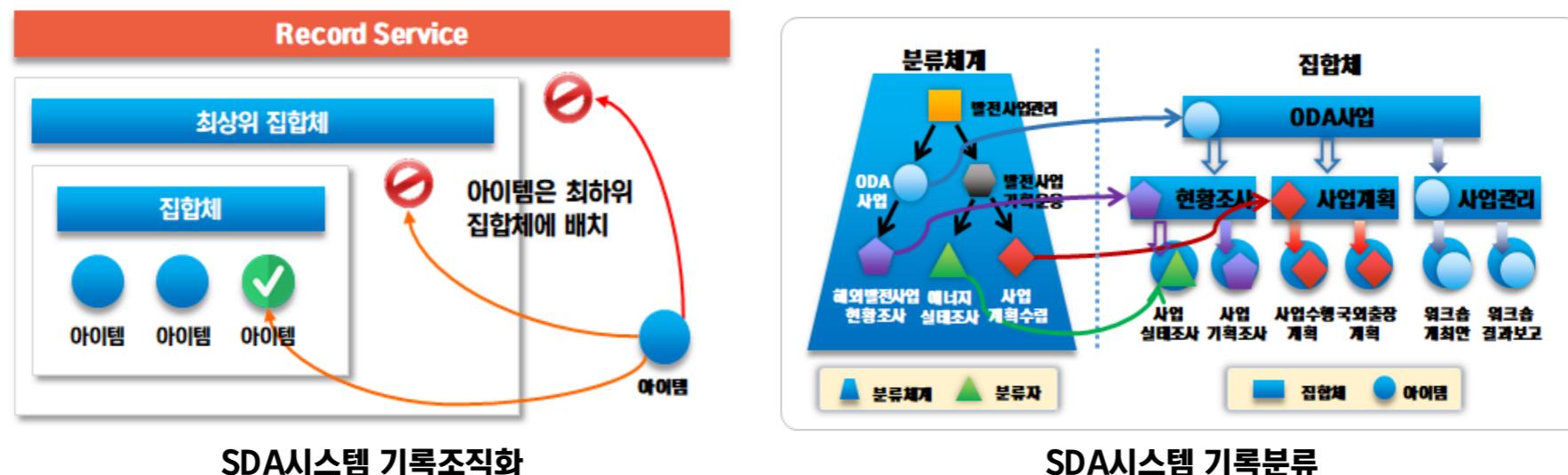
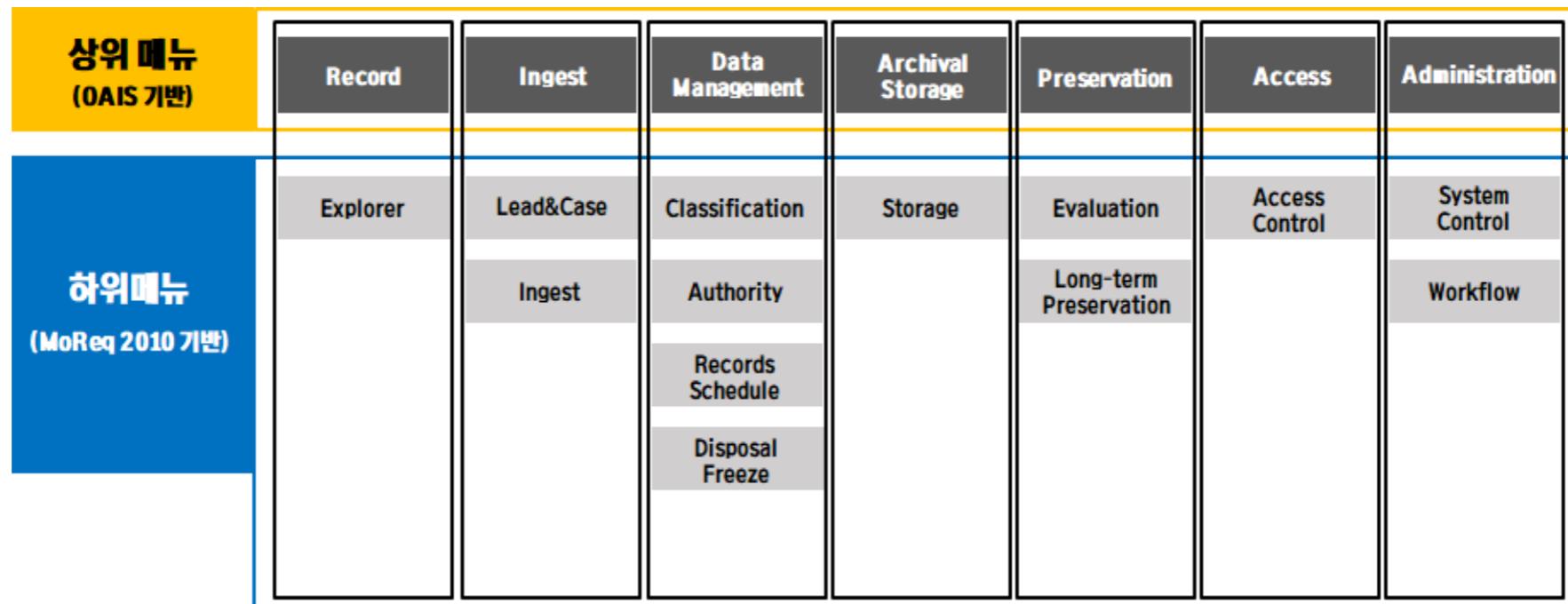
<그림> ERA 2.0 모듈 구성

2. 시스템

4) SDAS(서울기록원)

개요 : 2019년 아카이브시스템 개발 3년차 수행 중

- OAIS 참조모형을 기반으로 시스템 상위 메뉴 구성
- Moreq2010에서 제시하는 모듈화된 서비스 기반하여 하위 메뉴 구성
- Preservica의 Explorer, 워크플로우 기능 등을 응용하여 기록관리 기능 구현
- 기록물 집합체(Aggregation) 개념 도입하여 기록조직화 및 기록분류

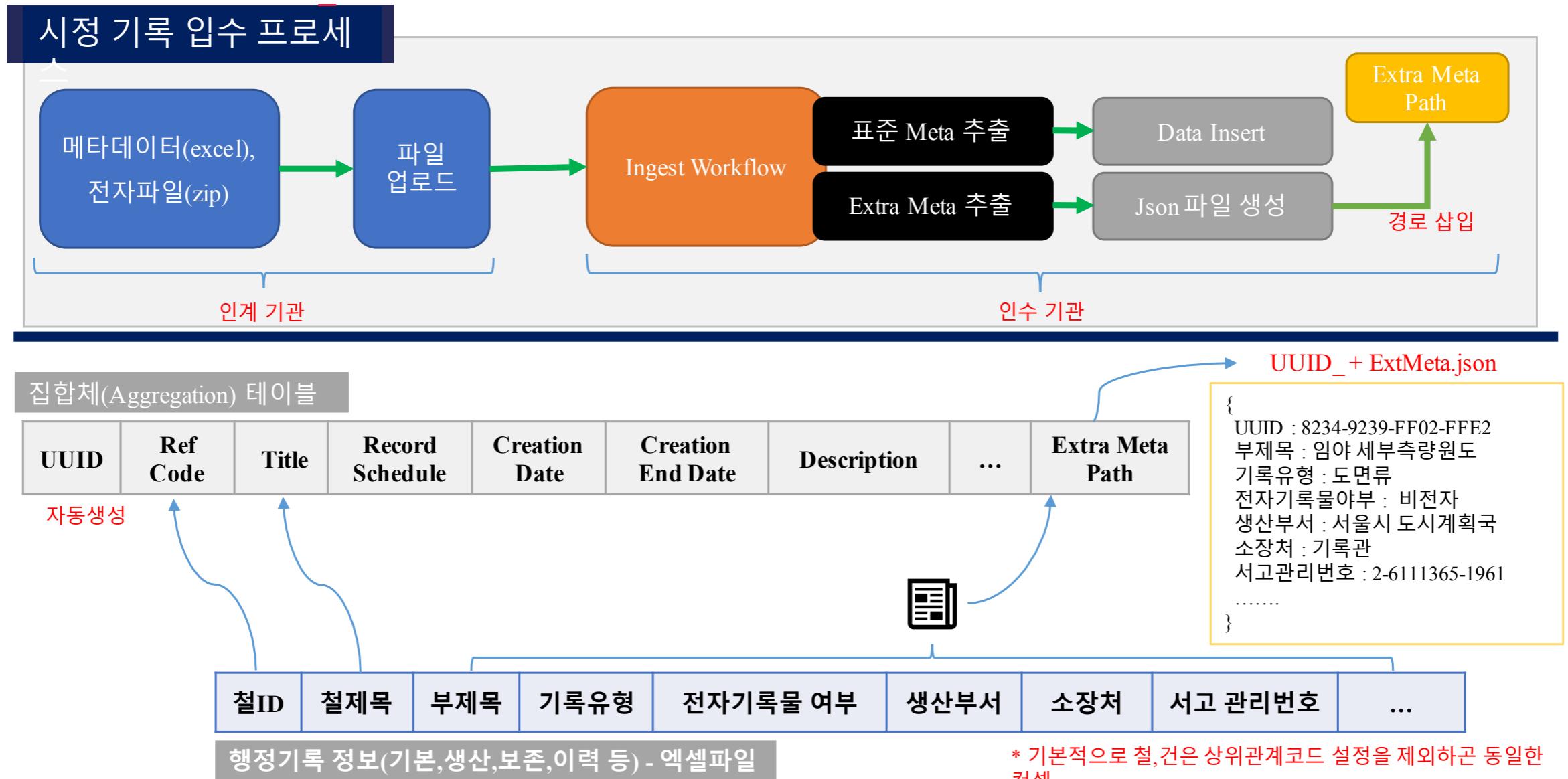


2. 시스템

4) SDAS(서울기록원)

워크플로우

시정기록 입수 워크플로우 사례



2. 시스템

4) SDAS(서울기록원)

디지털 카탈로그

2018년부터 디지털 카탈로그 모형을 수립하고 시범구축 진행
기록활용의 핵심도구로 내외부 이용자에게 서비스 예정

- 외부이용자 : 검색도구, 기록자원 네비게이션, Primary 기록콘텐츠 제공
- 내부이용자 : 기록조직·통제도구, 기록맥락 확보, 기록활용의 기본도구로서 활용
- 전거(기관, 지역, Topic, 인물 업무기능) Authority와 기록기술을 통해 기록의 통합적인 활용 가능

The screenshot displays the SDAS digital catalog interface, illustrating the transition from a classification description to an archival description.

Classification Description: On the left, there is a detailed classification description for '주택.도시계획 관련 기록, 1946-' (SC-010). It includes sections for '분야' (Category), '관리번호' (Number), '분야' (Category), '표제' (Title), '날짜' (Date), '설명' (Description), and '구성' (Composition). A red dotted line highlights the '구성' section, which lists '25개 시리즈' (25 series).

Archival Description: In the center, there is a detailed archival description for '주택재개발 및 재건축 관련 기록, 1986-1992' (DS-0021). It includes sections for '기본정보' (Basic Information), '기록유형' (Record Type), '기록상태' (Record Status), '내용정보' (Content Information), and '기록내역' (Record History). A red dotted line highlights the '구성' section, which lists '11개 파일로 구성' (11 files) and '구성: 11개 파일로 구성' (Composition: 11 files).

File List: On the right, a file list is shown under the heading '파일' (File). It lists various records with their file numbers, titles, and descriptions. Some entries are preceded by a red icon indicating they are part of the highlighted series.

| 파일 | 제작자 | 제작일 | 제작장소 | 제작내용 |
|----|---|-----------|-------------------|----------------|
| 파일 | 공동주택관리 기록, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 건설관리국 주택개량과 | 기록번호 : DF-0007 |
| 파일 | 도시정비기획 기록, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 도시계획국 도시계획과 | 기록번호 : DF-0008 |
| 파일 | 재건축 도면 모음, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 건설관리국 재개발과 | 기록번호 : DF-0009 |
| 파일 | 임대주택관리 기록, 1986-1989 From 시리즈 DS-0021 | 1986-1989 | 서울특별시 건설관리국 주택개량과 | 기록번호 : DF-0010 |
| 파일 | 저소득층 주거 지원 관련 기록, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 건설관리국 주택기획과 | 기록번호 : DF-0011 |
| 파일 | 한옥단지 조성 관련 기록, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 건설관리국 주택개량과 | 기록번호 : DF-0012 |
| 파일 | 재건축 관련도로정비기록, 1986-1992 From 시리즈 DS-0021 | 1986-1992 | 서울특별시 건설관리국 도로과 | 기록번호 : DF-0013 |

Classification Description

Archival Description

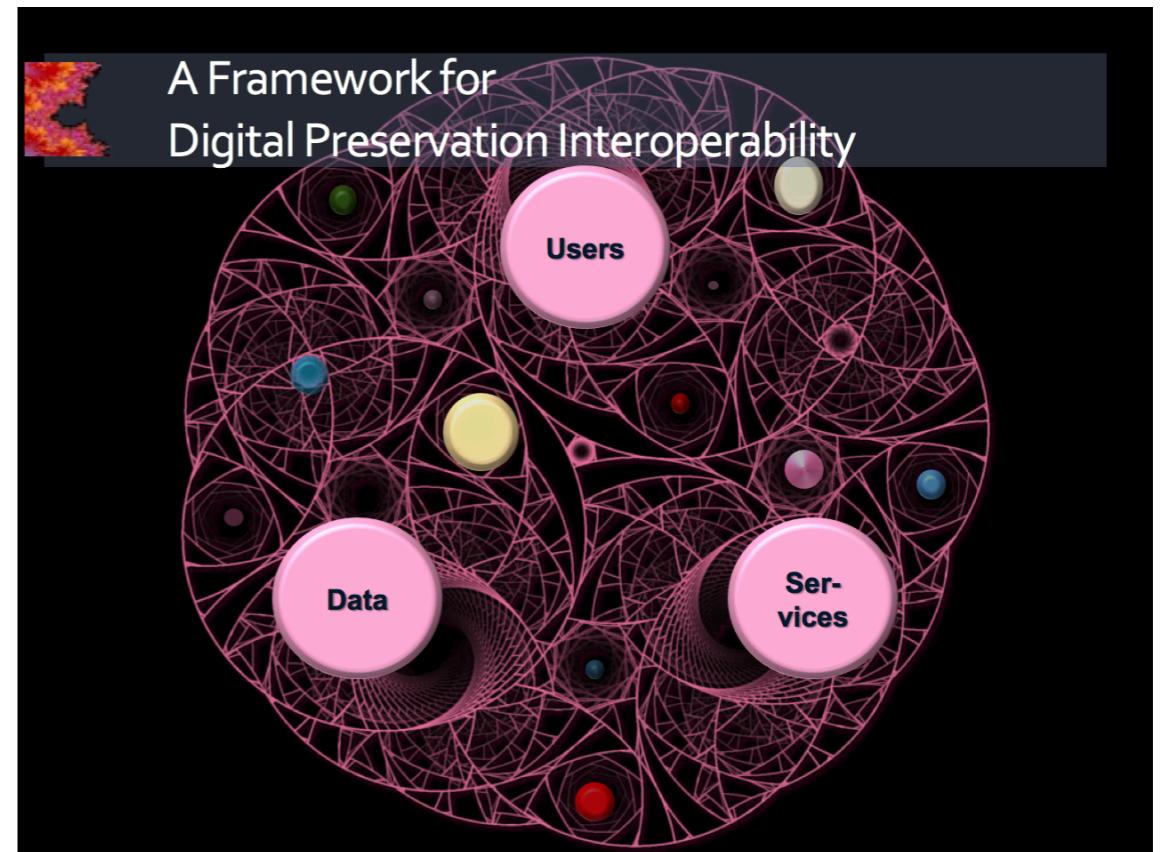
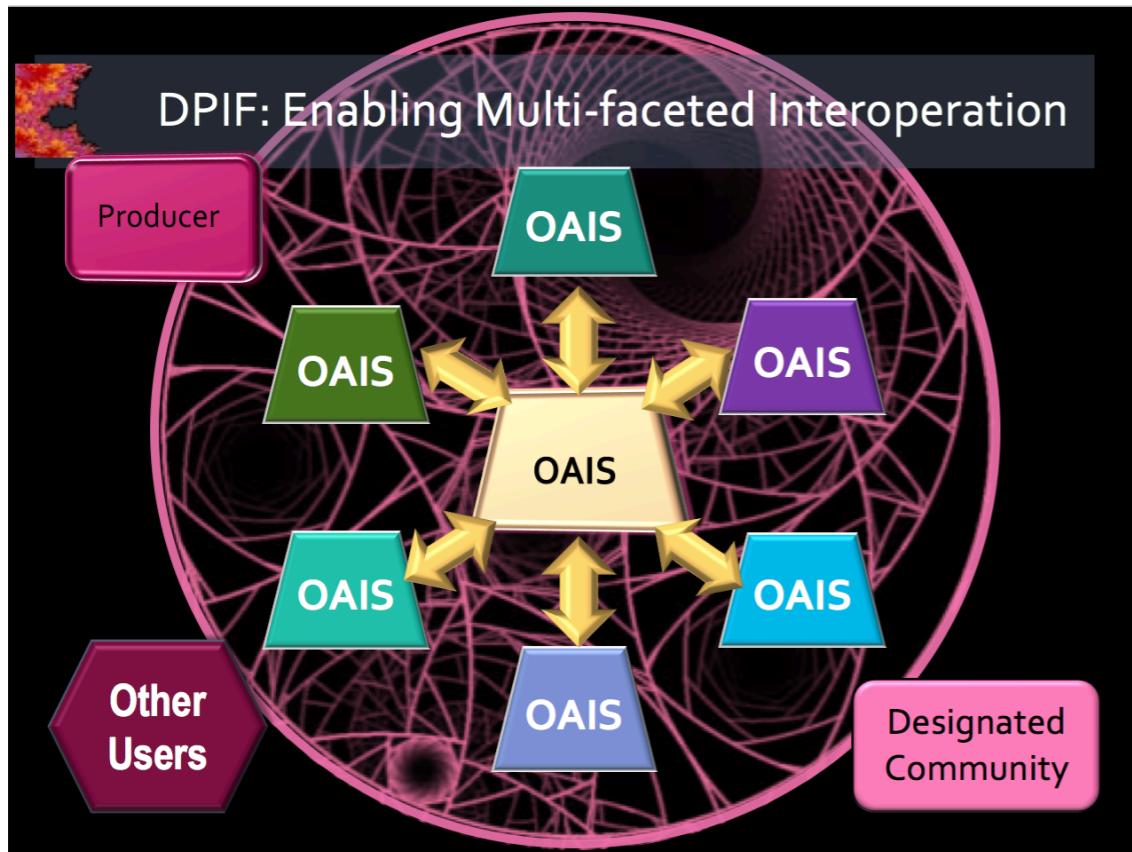
3. 상호운용성

DPIF

DPIF

ISO/IEC Digital Preservation Interoperability Framework(DPIF)

OAIS를 참조하는 복수 시스템 간 이용자의 요구에 따른 데이터 보존처리 방법과 도구, 서비스를 상호 운용 가능하도록 하기 위한 프레임워크 (2009~2010)

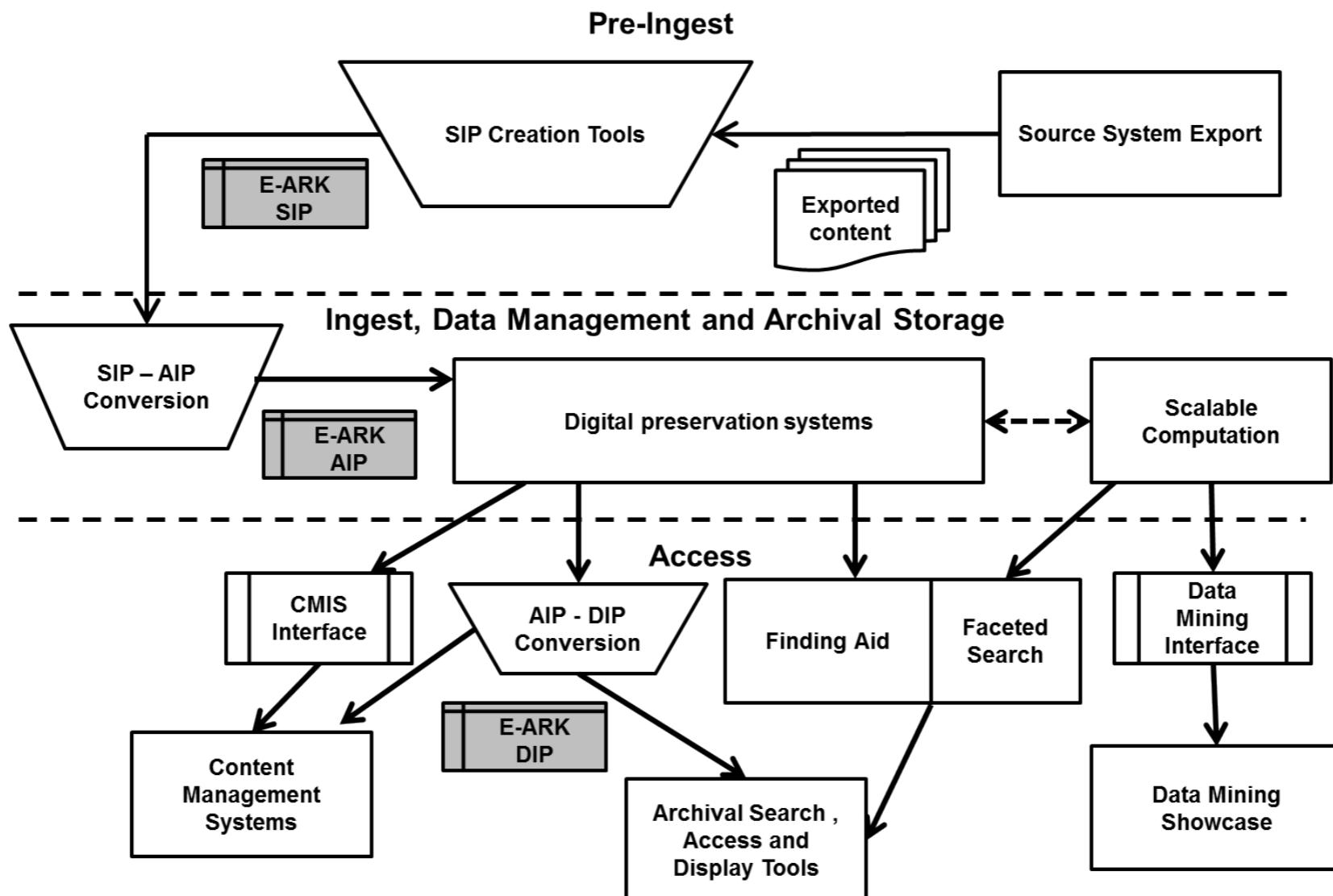


3. 상호운용성

E-Ark

E-Ark : OAIS의 실행을 위한 툴

E-Ark 프로젝트는 OAIS를 확장하고 공유하기 위한 EU의 공동연구로, 디지털 보존 실무를 보다 상세하게 정의하고 실무에 쓰일 수 있는 도구들을 식별 또는 직접 개발하여 오픈소스로 배포함



[그림] E-Ark 아키텍처

3. 상호운용성

E-Ark

E-Ark : 오픈소스 툴

OAIS의 주요 엔티티별 34종의 오픈소스 툴과 포맷 명세서 제공

| E-ARK Tools and Integrated Prototype Elements | | |   |
|---|--|--|---|
| | E-ARK Tools | Integrated Prototype | Third party tools used by E-ARK |
| Pre-Ingest | <ul style="list-style-type: none">● Database Preservation Toolkit (DBPTK)● ERMS Export Module● Universal Archiving Module (UAM)● RODA-In● ESSArch Tools for Producer | <ul style="list-style-type: none">● SIP Creator (E-ARK web) | |
| Ingest | <ul style="list-style-type: none">● ESSArch Tools for Archive | <ul style="list-style-type: none">● SIP to AIP Converter (E-ARK web) | |
| Storage | <ul style="list-style-type: none">● ESSArch Preservation Platform● RODA Repository | <ul style="list-style-type: none">● HDFS Storage (E-ARK web)● SOLR index (E-ARK web) | |
| Access | <ul style="list-style-type: none">● Search and Display GUI● Order Management Tool | <ul style="list-style-type: none">● Lily-Ingest● E-ARK Web Search● AIP2DIP (E-ARK Web) | <ul style="list-style-type: none">● QGIS● Geoserver |
| Viewer | <ul style="list-style-type: none">● IP Viewer (also ERMS and single file viewer)● Database Visualisation Toolkit (DBVTK)● CMIS portal/viewer | | <ul style="list-style-type: none">● Oracle OLAP Viewer● Peripleo |

3. 상호운용성

E-Ark

E-Ark : 통합 프로토타입

OAIS 전반을 수행하는 통합된 프로토타입 제시

The diagram illustrates the integration of several OAIS components:

- Hadoop Map/Reduce Administration:** Shows a cluster summary with 342 MB/899 MB heap size.
- Scheduling Information:** Displays a queue named "default" with one running job.
- Completed Jobs:** Lists two completed jobs, both named "batchIndexBuild" with 100% completion.
- Solr Search Interface:** Shows a search results page for "DNA_AVID_SA_18001.01_141104/DNA_AVID_SA_18001.01_141104/Metadata".
- ESSArch Preservation Platform (EPP):** Shows a dashboard with various preservation metrics and a list of information packages.
- E-ARK Integrated Prototype:** A central component connecting the others, labeled "repository", "item search and access", and "archival storage". It includes sections for "preservation platform" and "archive storage".
- NGDATA lily:** Shows a user interface for managing clusters, indices, schemas, and roles.

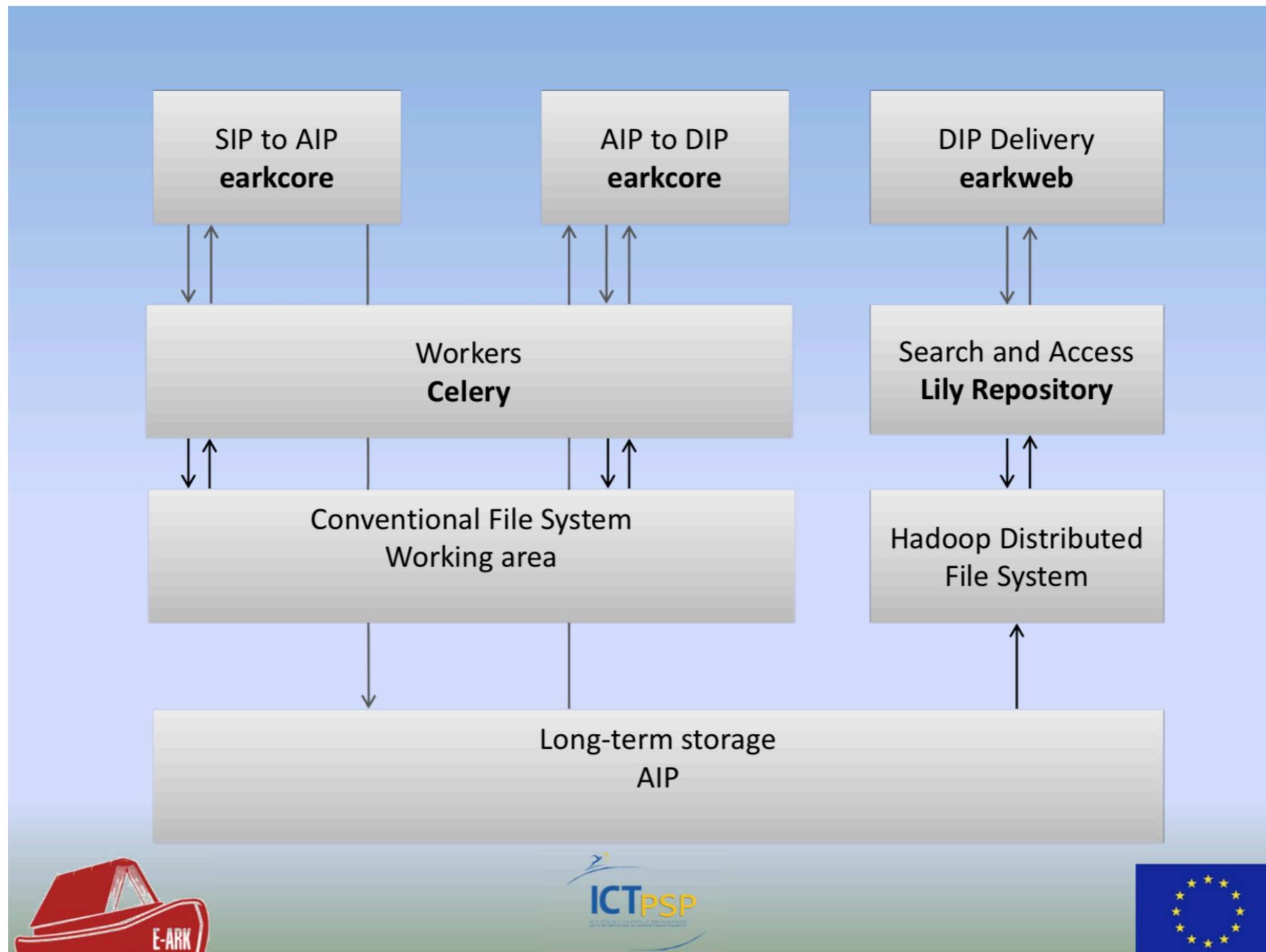
출처 : <https://www.eark-project.com/resources/conference-presentations/dlm-oct15/43-earkips-1/file>

3. 상호운용성

E-Ark

E-Ark : 통합 프로토타입

OAIS 전반을 수행하는 통합된 프로토타입 제시



3. 상호운용성

E-Ark

E-Ark : 워크플로우

E-Ark 프로젝트에서는 입수 단계의 베스트 프랙티스를 아래 6단계로 정의하고, 각 단계에서 쓰일 수 있는 아카이브 커뮤니티 내의 식별된 도구들을 배치함

| Workflow step | Name of the tool |
|----------------------------|---|
| Transfer to SIP | Elev SIP Creator; Preservica, MetsCreation; UAM; DRI; rsync*. |
| Ingest | DRI; Archaeology Data Service (ADS) Data Seal of Approval (after ADS DSOA); kleio; SFTP*, maior memorix. |
| Identification | DROID; maior memorix. |
| Normalisation | Preservica; METS, DRI, ADS DSOA; maior memorix; AdobePhotoshop; AdobePremiereCS5.5.2; Matrox MAX H264 Capture. |
| Characterisation | JHOVE (via kleio); Preservica; MODS; ADS DSOA; PRONOM; DROID. |
| Additional metadata | Preservica, DRI, UAM, MARC to MODS; maior.memorix; PIT+AIS; Adobe Bridge; EZID, gencat (Catalonian metadata); FTK Imager. |

* file transfer protocol not specific to archiving

E-Ark 프로젝트에서 정의한 이관 프로세스 및 도구

출처: <https://www.eark-project.com/>

3. 상호운용성

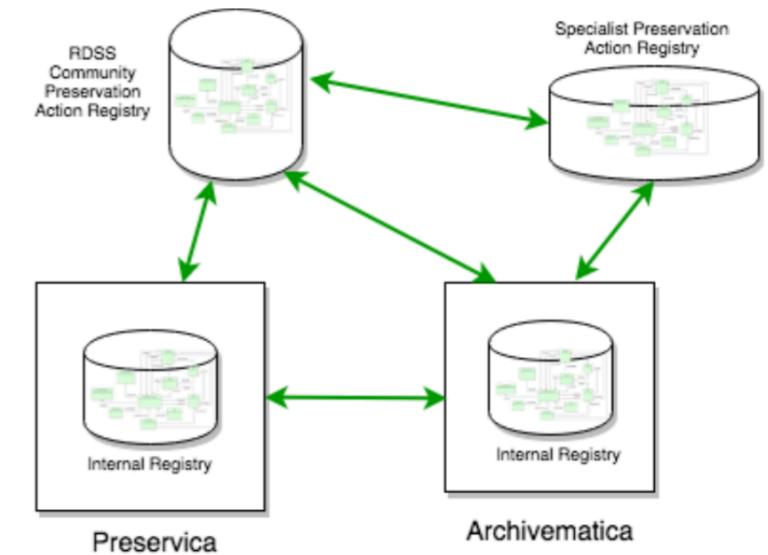
RDSS-PAR

개요

- ▶ 영국 JISC의 최신 프로젝트 (<http://parcore.org/>)
- ▶ 디지털 보존의 실무/기술을 주도하는 5개 기관/업체가 디지털 객체 보존기술과 모범사례를 포착하고 공유하기 위한 보존조치 레지스트리(Preservation Action Registry, PAR)를 연구

개발 배경

- ▶ 배경 : Archivematica의 FPR (Format Policy Register)과 Preservica의 Linked Data Registry (LDR)는 디지털 보존을 수행할 때 사용할 도구와 규칙을 정의하는데, 이 두 가지 방법은 유사한 문제를 해결하기 위한 것임에도 상호운용 안됨
- ▶ Artefactual, Arkivum, Preservica, Jisc 및 Open Preservation Foundation은 디지털 객체를 보존하기 위한 기술적 모범 사례를 포착하고 공유할 수 있는 PAR(Preservation Action Registries)을 연구하기로 함



데이터 모델

- ▶ PAR 레지스트리의 데이터 모델은 다양한 디지털 객체의 1)처리도구, 2)속성, 3)보존기능을 정의
- ▶ 이 정보에는 버전이 지정되며 다양한 보존 조치를 실행하는 데 필요한 도구 매개 변수 및 소프트웨어 환경이 포함되어 있어 보존 시스템에서 직접 실행 가능

3. 상호운용성

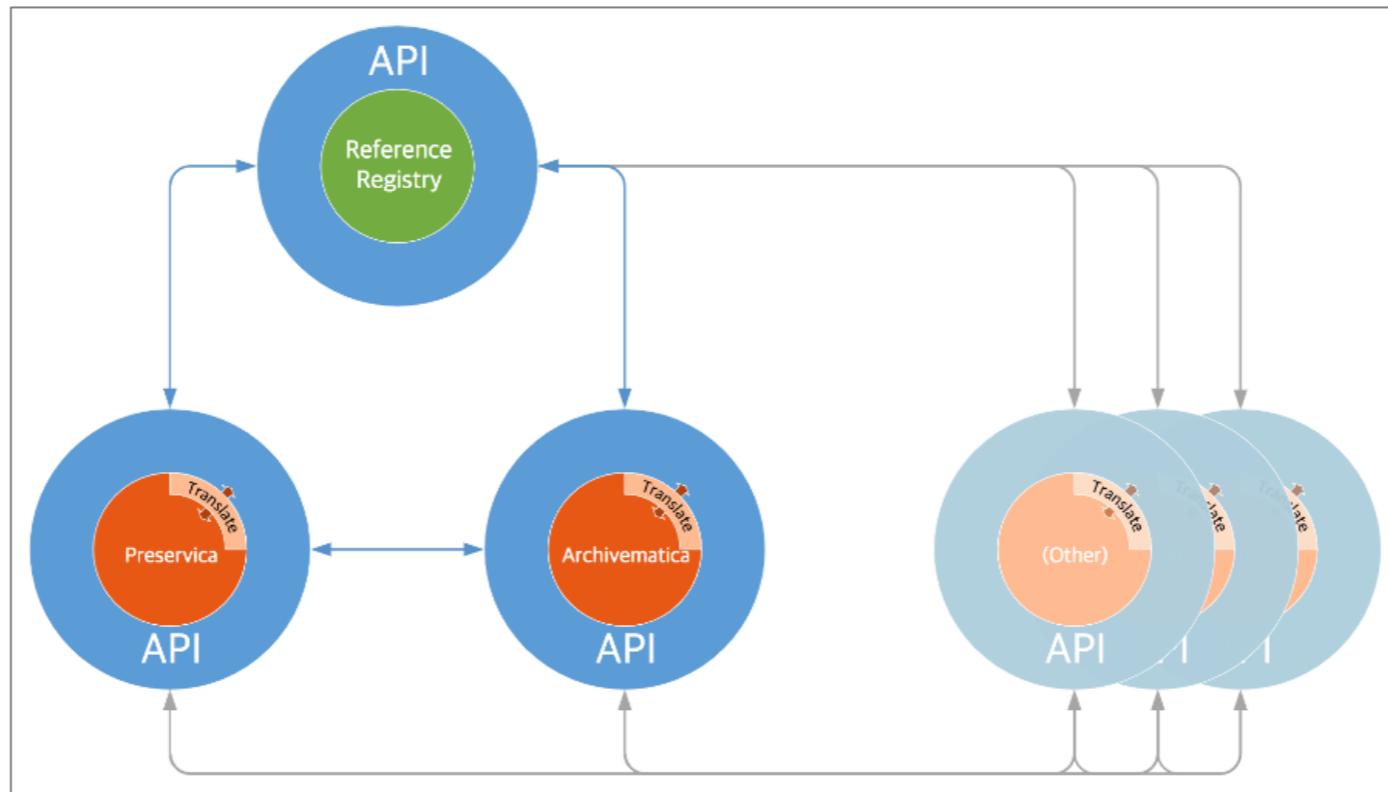
RDSS-PAR

산출물

| 산출물 | 내용 |
|--------------|--------------------------|
| 개념모델 | 프레임워크, 논의이슈 (다국어 제공) |
| JSON 스키마 | API 페이로드에 적용될 PAR 모델 정의 |
| API | 보존시스템 간 교환 가능한 범용인터페이스 |
| 수행 가능한 DP 액션 | 플랫폼 간 배포/실행 가능한 툴 (벤더독립) |
| PoC | 개념증명 결과물 |

개념도

- ▶ Archivematica, Preservica 등 보존시스템들이 보존정책 및 조치방법을 상호 참조
- ▶ 향후 다양한 보존시스템 및 기관들이 선진실무를 상호 공유할 수 있도록 API 연계 확대 예정

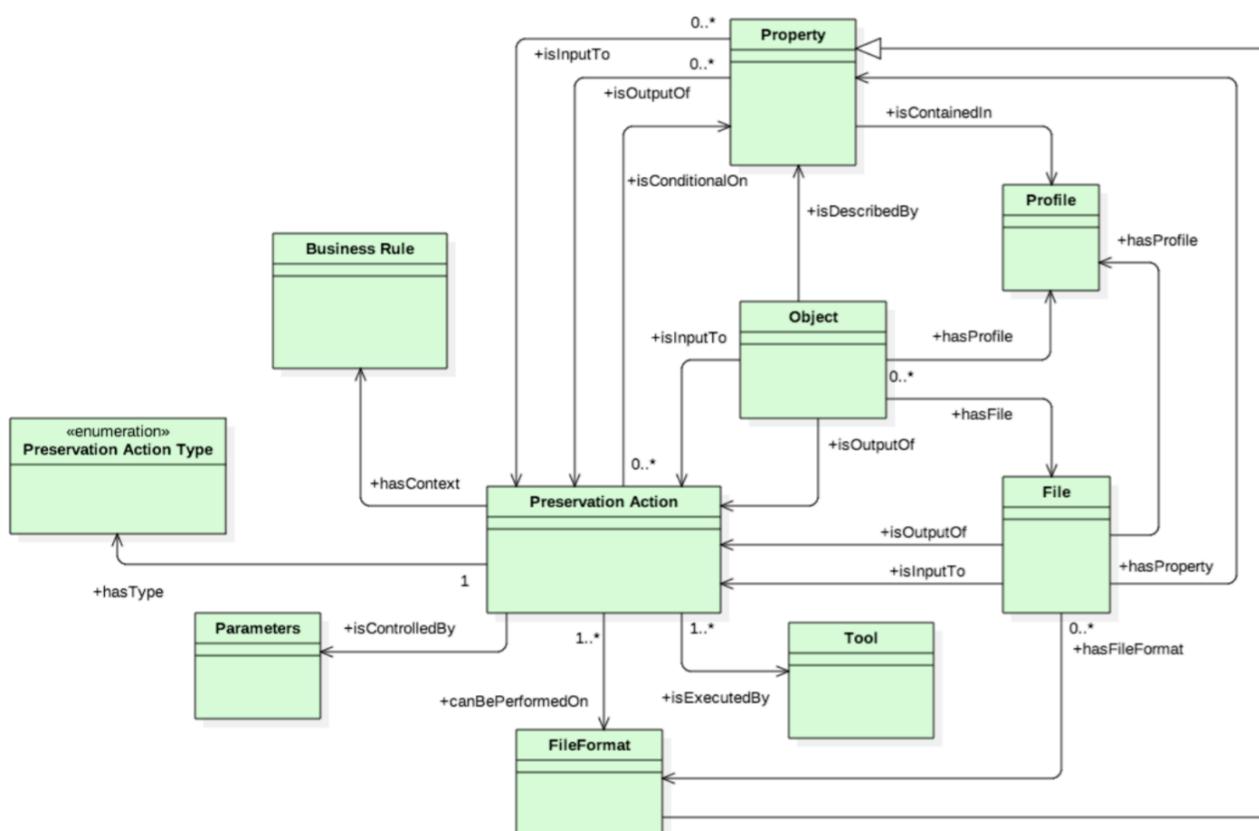


3. 상호운용성

RDSS-PAR

개념모델

- 보존조치는 md5sum, DROID, JHOVE, ffmpeg 등 웹서비스, 또는 컨테이너 형태의 툴을 호출하여 Object 또는 File에 디지털 보존조치 실행(명령줄 실행 또는 그 이상의 조치)



| Preservation Action | par:GenerateMD5checksum | Preservation Action Type | premis:fixity |
|---------------------|-------------------------------|---|---------------|
| | | Input Type | pcdm:File |
| | | Output Type | par:Property |
| Tool | par:toolRegistryName | COPTR | |
| | par:toolRegistryKey | Md5sum Unix Command | |
| | par:toolVersion | md5sum (GNU coreutils) 8.25 | |
| | par:toolDocumentation | https://man.cx/md5sum | |
| Property (Output) | par:propertyType | premis:messageDigest | |
| | premis:messageDigestAlgorithm | MD5 | |
| | xsd:anyURI | http://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions/md5.html | |

Simplified example of a Preservation Action (generating a MD5 checksum on a File)

<그림> PAR 개념모델

<그림> 파일의 MD5 체크섬을 생성하는 보존조치 사례

4. 결론

1. OAIS 해석 및 기능요건 구체화
2. 표준/툴/서비스 (METS, PREMIS, DROID, JHOVE, Archive-It 등)
3. 신기술 (클라우드, 마이크로서비스 아키텍처, 도커/쿠버네티스, 오픈소스)
4. 모니터링과 협력



감사합니다

안대진

daejin@archivelab.co.kr