

도서관 희귀 자료의 보존과 접근을 위한 디지털화의 표준화

Standardizing Digitization for Preservation and Access to Rare Library Materials

아돌프 크놀(Adolf Knoll)*

목 차

- | | |
|---------------------------|----------------------------------|
| 1. 서 론 | 5. 메타데이터의 미래 |
| 2. 표준화 | 6. 데이터솔루션: 어떻게 인코딩 할 것인가 |
| 3. 메타데이터 솔루션: 무엇을 기술할 것인가 | 6.1 아카이벌 파일 |
| 3.1 식별메타데이터 | 6.2 이용자 파일 |
| 3.2 내용기술 메타데이터 | 6.3 해상도와 데이터 흐름 |
| 3.3 구조 메타데이터 | 6.4 손실 압축 |
| 3.4 기술적 메타데이터 | 6.5 Mixed Raster Content(MRC) 기술 |
| 3.5 커뮤니케이션 메타데이터 | 6.6 다중해상 이미지 |
| 3.6 추가된 메타데이터 속성과 성질 | 6.7 이미지 서버 |
| 4. 메타데이터 솔루션: 어떻게 기술할 것인가 | 7. 연구, 개발과 생산 |

<초 록>

체코공화국 국립도서관은 1992년 유네스코의 제안을 받아들여, 소장품에서 가장 귀중한 자료들을 보여주기 위하여, 파일럿 CD ROM 발행을 준비하기 시작했다. 1993년 4월에 발행된 이 CD는 세계 기록 문화유산 프로그램에 대한 대중적 관심을 불러일으키려는 목적을 갖는 일련의 유사한 유네스코 프로젝트들 가운데 하나였다. 1993년부터 1995년 사이에, 두 가지의 중세 문서를 전부 보여주는 두 개의 간행물들이 출간되었다. 이들은 모든 이미지 문서화와 원본 문서의 다양한 특성에 대한 과학적 기술을 제공하였다. 체코 국립도서관의 디지털화센터는 Czech AIP Beroun Company와 협력하여 고문서와 고인쇄본을 전문적으로 처리하고 있다. 현재, 체코 국립도서관은 십여 개의 체코 및 여러 외국 기관들과 협력하고 있다. 두 개의 간행물은 디지털 도서관을 통해 이용할 수 있다.

주제어: 도서관 희귀자료, 보존, 디지털화

<ABSTRACT>

The beginning of digitization activities in the National Library of the Czech Republic are interconnected with the first periods of existence of the UNESCO Memory of the World programme. It was in 1992 when the library accepted the UNESCO proposal to prepare a pilot CD ROM publication for the programme showing the most precious items from its collections. In 1993 - 1995, the publication was followed by two another titles, which presented two medieval manuscripts in their entirety, providing both all the image documentation and scientific description of various features of the original documents.

The digitization centre specialized in processing of old manuscripts and old printed books in co operation with the Czech AIP Beroun Company. Nowadays, these two national programmes in which dozens of Czech and also several foreign institutions take part - are represented on their access side by two digital libraries: Manuscriptorium and Kramerius.

Key words: rare library material, preservation, digitization

* 체코 국립도서관(adolf.knoll@nkp.cz)

1. 서론

체코공화국 국립도서관에서 디지털화 활동이 시작된 것은 유네스코의 세계기록문화유산 프로그램이 생겨난 초기에 상호 연결되어 있다. 도서관이 유네스코의 제안을 받아들여, 소장품에서 가장 귀중한 자료들을 보여주기 위하여, 이 프로그램을 위한 파일로트 CD ROM 발행을 준비한 것은 1992년이였다.

CD¹⁾는 도서관 자체에 선정된 자료의 기술(記述) 및 희귀본 및 사라질 위기에 처한 도서관 희귀문서의 접근 개선과 간접적 보존을 위한 디지털 기술의 실행가능성을 입증하기 위한 표본 이미지를 보여주었다. 1993년 4월에 발행된 CD는 세계 기록 문화유산 프로그램에 대한 대중적 관심을 불러일으키려는 목적을 갖는 일련의 유사한 유네스코 프로젝트들 가운데 하나였다.

1993년부터 1995년 사이에, 두 가지의 중세 문서를 전부 보여주는 두 개의 간행물들²⁾³⁾이 이어졌고, 이들은 모든 이미지 문서화와 원본 문서의 다양한 특성에 대한 과학적 기술을 제공하였다. 최초의 디지털화 제품이 아직은 보다 높은 품질의 이미지를 얻기 위해 스캔된 컬러 슬라이드를 매개로 하여 제작되었을 때, 동시에 원본을 직접 디지털화하기 위한 다양한 장치들을 시험하기 위해 많은 노력이 기울여졌

다. 당시는 최초의 디지털카메라가 실험실용으로만 등장하고 있는 때였다. 그것들은 매우 비쌌지만, 그 같은 것만으로도 그것들을 적용시킴으로써 디지털화 비용이 상당히 줄어들 수 있게 하였다. 그것의 해상도는 오늘날 일반적으로 사용되는 카메라의 해상도와 비견할 만하였고, 그것들 가운데 가장 뛰어난 것은 256 회색조로 디지털 이미지를 제공하는 것으로 시작하였다. 이것은 우리가 1996년 우리의 최초의 디지털화 센터에서 회색조로만 디지털화하기 시작했다는 것을 의미하지는 않는다. 객체의 부분적 RGB 이미지를 진짜 컬러 이미지로 컴퓨터화하여 통합하는 가능성 덕분에, 우리가 일상적 디지털화를 시작한 것은 또한 상당히 뛰어난 컬러 이미지를 제공하였다. 따라서, 우리가 사용한 최초의 디지털 카메라인 KODAK DCS 460은 텍스트 고문서의 일부뿐만 아니라, 그 크기와 세부사항의 풍성함이 우리 카메라가 제공할 수 있는 것보다 더 나은 디지털화 매개변수를 요구하지는 않았던 일부의 채색 고문서들까지도 디지털화할 수 있게 하였다.

디지털화센터는 Czech AIP Beroun Company⁴⁾와 협력하여 고문서와 고인쇄본을 처리하는 것을 전문으로 하였다. 그 이후에 최초의 카메라보다 훨씬 뛰어난 매개변수를 가진 새로운 장비(컬러 스캐너인 Betterlight와 Cruse)가 설비되었다. 또한 디지털화하는 동안 고문서를 보

1) Paměť světa Memory of the World Mémoire du Monde. Praha, Národní knihovna Albertina icome Praha, 1993(s UNESCO Memory of the World CD ROM publication).

2) Antiphonarium Sedlecense, MS XIII A 6. Praha, Národní knihovna Albertina icome Praha, 1995. Memoriae Mundi Series Bohemica, 1(1st entirely published digitized manuscript in the UNESCO Memory of the World programme).

3) Chronicon Concilii Constantiniensis, MS VII A 18. Praha, Národní knihovna Albertina icome Praha, 1995. Memoriae Mundi Series Bohemica, 2(publication of a digitized manuscript).

4) <<http://www.aipberoun.cz>>

호하기 위하여 물리적 취급과 빛⁵⁾에 대한 노출을 다에 대하여 하드웨어 측면에서 많은 작업이 수행되었다. 국립도서관의 프로젝트는 문화부가 지원하기 시작한 국가적 프로그램으로 단계별로 발전하고 있었다. 이러한 발전은 다른 기록 기관들의 주의를 끌어서 공동 디지털도서관을 구축하기 위해 자기들의 소장품을 기여하도록 하였다. 2000년에, 대등한 보존과 디지털화 프로젝트가 산성지로 된 도서관자료를 특별히 강조하면서 시작되었다. 이것은 표준화에 관한 최초의 프로젝트의 권장사항들을 대부분 준수하였다.

오늘날 이 두 개의 국가적 프로그램은 십여 개의 체코 및 여러 외국 기관들이 또한 참여하고 고, 두 개의 디지털 도서관에 의하여 접근된다: 약 700,000여 디지털화된 페이지를 온라인으로 제공하는 Manuscriptorium⁶⁾(더 오래된 Memoriae Mundi Series Bohemica programme을 말하는 것으로, 현재는 짧게 Memoria라고 불리고 있음)과 주로 구간 정기간행물을 대략 120만 페이지 정도 온라인으로 제공하는 Kramerius⁷⁾이 그것이다.

2. 표준화

최초의 시험 프로젝트들이 완료된 후에, 기 대하던 일상적 디지털화는 접근 어플리케이션

과 실재하는 소프트웨어 및 하드웨어 플랫폼에 대해서 디지털 생산물의 상대적 독립성을 보장할 수 있는 확고한 개념에 기반해야 한다는 것이 명백하였다. 따라서 모든 생산된 디지털 문서들, 즉, 디지털 파일에 관한 기술적 정보를 포함하여, 내용에 관한 필요한 기술(記述)을 갖는 모든 이미지 파일들을 어떻게 구조화할 것인가 하는 방법을 찾기 위한 방법이 모색되었다. 솔루션은 최종적으로 일명 복합문서(compound document)의 개념에 기반하였다. 복합문서는 디지털화함으로써 생산된 데이터 파일(주로 이미지이지만, 또한 오디오나 비디오 파일과 같은 다른 파일도 있음)을 나중에 추가된 모든 정보(메타데이터)와 결합하여 이용자들이 하여금 데이터 파일을 갖고 작업할 수 있게 하는 것이다. 동시에 가장 중요한 사항은, 예를 들어 이미지 파일에 기록된 IPTC나 EXIF 표준 혹은 MP3나 심지어 WAV 오디오 파일에도 저장될 수 있는 ID3 정보 태그와 같이, 비록 많은 데이터 포맷이 상이한 유형의 메타데이터가 직접 그 자체에 기록되고, 그것과 함께 저장되는 것을 용납하지만, 메타데이터 정보는 데이터 파일 외부에 있어야 한다는 것이었다.

복합문서의 개념은 새로운 것은 아니다. 왜냐하면 예를 들어 우리가 가장 보편적인 오디오CD판을 볼 때, 디지털오디오파일을 콤팩트 디스크에 저장하고, 때때로 많은 기술(메타)데이터를 매체나 종이 표지와 삽입지에, 혹은 삽

5) 이들 중 일부는 FAO와 UNESCO가 2005년에 간행한 새로운 전자학습 코스인 Information Management Resource Ki에 예시되어 있다(<http://www.fao.org/imark/>: National Library of the Czech Republic가 작성한 디지털화 동안 회귀 문서의 처리에 관한 챗터를 보시오).

6) <<http://www.manuscriptorium.com>> or <<http://www.memoria.cz>>

7) <<http://kramerius.nkp.cz>>. 저작권 없는 이전 문서는 <<http://kramerius.nkp.cz/kramerius/PShowVolume.do?id=3676&it=1>>를 참조하십시오.

입된 소책자에 인쇄한 것을 발견한다. 우리가 이 모든 것을 단지 전자적으로 저장하려고 한다면, 모든 제품들이 서로 비교되고 호환될 수 있게 함으로써 더 잘 선전되고, 판매되고 접근될 수 있게 하는 방식으로 발행되도록 어떤 규칙이 도입되어야 한다는 것은 명백하다.

기술(記述, descriptive) 메타데이터는 거의 언제나 객체(확인, 목록 혹은 서지 레코드)를 식별하지만, 기술적(technical) 메타데이터를 포함하여, 상이한 전문화된 내용을 갖는 많은 다른 종류의 정보를 가져올 수 있다. 더욱이, 그 같은 기술정보의 어떤 선택된 부분으로부터 데이터 객체들로 참조할 수 있어야 하고, 따라서 이용자는 그 객체가 무엇인지에 관하여 알게 될 뿐 아니라, 그것의 가상 표현을 보고 그것의 속성과 내용에 관하여 판단을 내릴 수 있게 해야 한다.

1990년대 중반에 권장된 도구는 SGML이었지만, 동시에 그것이 크게 확장되어 거의 어떤 이용자를 복합 SGML 문서를 쉽게 조작할 수 있게 될 것인지에 대하여 상당한 불확실성이 있었다. SGML의 강점은 어떤 실재하는 플랫폼과 별로 상관없이, 디지털 이미지나 다른 디지털 데이터파일에 대한 주요한 접근 게이트웨이 역할을 할 수 있는 상당히 안정적인 가독형으로 문서를 구조화하는 데 있었다. 한편 SGML의 약점은 그것이 매우 복잡하다는 것이다. 이것은 문서를 읽고 프린트하기 위해 포맷하는 방법이 상당히 어려웠던 것에서 나타났고, 또한 메타데이터 입력과 데이터 파일 참조가 현재

이용자의 지식 및 그가 일반적으로 이용할 수 있는 도구들로 판단할 때 너무 복잡하였다. 동시에 인터넷은 미래의 데이터 전송과 접근을 위해 가능성이 있는 결정적인 매체로서 중요한 역할을 하기 시작하고 있었다.

우리가 어떻게 진행할지 결정해야 하는 상황은 구조화하고, 링크하고 포맷화 하는 언어로서 HTML이 매력적인 단순성을 갖고 있는 것과, 내용 기술에 있어서 완전 SGML이 우세하다는 것 때문에 발생했다. 우리는 이 두 가지 특성을 결합하도록 노력하기로 결정했고, 그 결과는 우리가 DOBM⁸⁾이라고 부르는 일종의 재정의 된 HTML로 나타났으며, 이 두 원칙들을 결합시킬 수 있게 했던 것은 실제로 특수한 DTD 기반 단순 SGML 어플리케이션이었다. 이것은 DOBM으로 필요한 기술규칙, 표준, 혹은 적절한 실행방식을 매핑할 기회를 갖게 하는 한편, 주된 이용자 접근도구로서 인터넷 브라우저에 의존할 가능성을 제공하였다. 이 방식으로 우리는 디지털화된 고문서와 정기간행물을 구조화하는 복잡한 문서 포맷 혹은 명세들을 작성하였다. 이 두 가지 명세는 2002 2003년까지 두 개의 국가적 프로그램에 성공적으로 공헌하였다. 이 접근방법을 CD ROM으로 발행한 것은 또한 유네스코에 의하여 세계기록문화유산 프로그램에서 따를 만한 적절한 실행방식이라고 승인되었다(1999).

실제로 완전한 복합 디지털 문서는 두 부분을 갖는다. 메타데이터와 데이터 자체. 이들의

8) Digitization of Rare Library Materials: Storage and Access to Data / Project Management by Adolf Knoll and Stanislav Psohlavec. Authors: Adolf Knoll, Stanislav Psohlavec, Jan Mottl, Jan Vomlel, Tomáš Mayer, ... (Preservation and Access International Newsletter, June no. 6, 199에서 중요한 소스로 강조). Prague, National Library Albertina icome Praha, 1999. Memoriae Mundi Series Bohemica. <<http://digit.nkp.cz/rare99/enter.htm>>도 보시오.

각각은 다량(수백)의 엔티티들(서브파트)로 구성될 수도 있고, 혹은 단 하나의 엔티티에 해당할 수도 있다. 만약 이 부분들 중 하나가 비었으면(null), 데이터 부분이 없을 때에는 우리는 보통 적어도 하나의 식별 레코드 혹은 완전한 복합기술구조를 갖는다. 그렇지 않고 기술 부분이 없을 때에는, 영성한 데이터 파일(예를 들어, 이미지)을 갖는다.

디지털화 프로그램의 적절한 발전을 보장하기 위해서는 메타데이터와 데이터 부분 둘 다에 특별한 주의를 기울일 필요가 있다.

3. 메타데이터 솔루션: 무엇을 기술할 것인가

메타데이터는 위에서 언급한 것처럼 식별을 위해서, 구조화를 위해서(이들을 구조 메타데이터라고 부를 수 있음), 디지털 문서의 다양한 계층적 수준에서 콘텐츠를 기술하기 위해서, 데이터 파일에 관한 기술적 정보의 저장을 위해서, 그리고 또한 다른 어플리케이션과의 커뮤니케이션을 위해서 필요하다. 메타데이터 기술의 다양한 부분들이 특정한 역할을 갖기 때문에, 가장 중요한 결정 가운데 하나는 필요한 메타데이터의 최소 집합을 정의하는 것이다. 이 범주는 필수 메타데이터 요소들(이들은 반드시 있어야 한다) 및 매우 권장되거나 이용 가능하면 필수적인(mandatory if available) 요소들을 포함할 것이다.

필수적 및 이용 가능하면 필수적인 요소들은 다음과 같은 목표를 갖고 선택되어야 한다:

- 우리가 커뮤니케이션 하고 싶은 환경과 기본적인 호환성을 보장하도록;
- 상이한 장치들에서 데이터 및 그것들의 복제물에 적절하게 접근할 수 있게 하도록.

이 두 가지 측면은 디지털화 프로젝트나 프로그램의 목표를 반영한다. 예를 들어, 목표가 또한 문서의 원래 모양이나 음향을 진짜로 복제하는 것일 때, 기술적 메타데이터에 많은 주의가 기울여져야 한다. 그러나 단순한 텍스트의 가독성만이 요구될 때는(예를 들어, 스캐너의 ICC 프로파일이나 정지 디지털 이미지에 관한 다른 메타데이터를 포함하는) 앞에서 말한 메타데이터 세트가 반드시 요구되지 않는다.

3.1 식별 메타데이터

식별 메타데이터의 영역은 비교적 잘 정의되어 있다. 왜냐하면 이들이 도서관, 박물관, 혹은 아카이브들이 그들의 컬렉션 객체들을 목록화하고 배열하거나 분류할 필요성을 반영하기 때문이다: 따라서, 다른 유형의 기관들은 수작업으로 기록되었든, 보다 잘 정리된 기계가독형태(아카이브용의 EAD; 혹은 AMICO, CIDOC, REACH 요소 세트 등과 같은 다양한 박물관 표준)로 되었든, 다른 내용기술 규칙들에 의존할 수 있는 한편, 일부 실재하는 목록 규칙 혹은 다양한 MARC 포맷이나 더블린 코어로 표현한 것이 도서관에게는 매우 익숙하다는 것은 의심할 바가 없다. 우리의 선택은 여기서 우리가 협력하고 싶은 기관들과 함께 달성할 수 있도록 동의한 바를 반영해야 한다.

우리의 디지털화 프로그램에서, 우리는 역사

적 도서관 자료가 아닌 자료의 식별을 위해서는 현재 사용되는 목록규칙을 따르고, 고문서 용으로는 특수한 MASTER포맷(European 프로젝트⁹⁾의 결과물)을 적용한다. 우리가 단지 약간만 다루고 있는 박물관 객체에 관해서는 이 영역에서의 여러 접근방법들(주로 미국의 RLG가 마련한 REACH 요소 세트¹⁰⁾ 및 French Systèmes descriptifs¹¹⁾) 및 일부 확증된 도서관 목록방법들에 의하여 영감을 받아왔다.

식별 메타데이터의 역할은 디지털 객체를 원래의 객체(예를 들어, 책)와 일치시키는 것이다. 우리의 디지털 도서관이 또한 목록과 대등하겠다는 사명을 갖지 않는 경우에, 그러면 식별 메타데이터 요소들의 수는 단지 매우 기본적인 수 있다. 한편 이용자가 완전한 서지 레코드를 보고 싶어 할 때에는 OPAC이나 종합목록으로 진행될 수 있다. 그 역으로, OPAC의 이용자는 이용 가능한 문서의 완전한 가상표현이 있을 경우에 디지털 도서관 응용 프로그램으로 진행될 수 있다. 이러한 생각은 Kramerius programme에서 이행되는 중이고, 한편 Memoria에서는 Manuscriptorium Digital Library가 참여기관들의 역사적 문서들의 종합목록 역할도 또한 수행하고 있다: 따라서 여기에서는 많은 서지 레코드들이 상당히 포괄적이다.

3.2 내용기술 메타데이터

이 영역은 준수될 수 있는 일반적으로 채택된 규칙이 없지만, 대개는 우리의 적절한 실행

방식이 디지털화된 문서의 비식별 기술수준을 분류하거나 그것을 따르기 위하여 무엇이 중요할 것인지 지시할 수 있다. 예를 들어서, 오래된 문서들의 다양한 버전의 상이한 시작이나 끝의 텍스트 덩어리들을 비교하는데 관심이 있을 수도 있고, 고문서 채색장식으로부터 객체들을 비교하고 분류하는 데 관심이 있을 수도 있다. 전문(畵文)에서 발견되고 정의될 수 있는 객체에 관해서는, 보다 깊은 텍스트 분석과 구조가 권장될 수 있고, 그 결과는 전문과 다른 종류의 데이터(특히 이미지)가 상호 연결된, 혼합된 고도로 복잡한 디지털 문서일 수도 있다.

3.3 구조 메타데이터

이 메타데이터들은 식별 부분에 부분적으로 포함될 수 있지만, 그것들 가운데 어떤 것이 요소들로 정의될 것인가는 대체로 우리가 선택한다. 이 메타데이터들은 단순히 원래 문서(연간물 세트, 페이지, 권, 부록, 기사, 호 등)의 물리적 구조를 표현할 수 있게 하고, 거기에서 우리가 이용자에게 그것들을 표현하는 데이터(예를 들어, 이미지나 전문)를 보낼 수 있다. 많은 경우에, 우리가 포맷을 작성해야 하는 입장이 되는 경우에 우리는 우리의 디지털 문서가 어떻게 구조화될 것인가를 결정하기 위해서 모형을 해야만 한다. 물론 동시에 우리는 다른 필수적 혹은 비필수적 속성들을 갖는 하위 구조수준에서 식별 메타데이터를 필요로 할 수도 있다.

9) <<http://www.cta.dmu.ac.uk/projects/master/>>

10) <<http://www.rlg.org/reach.elements.html>> or <http://www.rlg.org/en/page.php?Page_ID=432>

11) <<http://www.culture.gouv.fr/culture/inventai/extranet/sysdesc.htm>>

3.4 기술적 메타데이터

우리는 우리가 디지털화한 것을 정확하게 표현하기 원할 때, 그 같은 메타데이터를 필요로 할 것이다. 정확한 표현은 다음과 같은 것에 영향을 받을 수 있다:

- 디지털화 장치의 속성들과 비교하거나 속성들 사이에 비교된, 디스플레이, 재생 혹은 프린트 장치의 구체적 속성들;
- 그것들의 정확한 디코딩에 영향을 미칠 수도 있는 데이터 포맷의 노화.

그 같은 경우에, 특히 이미지화 영역에서 진짜 컬러 표현이 우리에게 결정적 역할을 할 때, 우리는 스캐닝 과정 자체에 관한 많은 정보를 필요로 할 것이다. Memoria 프로그램에서 그 같은 정보는 나란히 세 가지 방법에 의하여 지원된다. ICC 프로파일의 저장, 이미지에 관한 선택된 기술적 메타데이터의 저장, 및 캘리브레이션(calibration, 분도 혹은 각도)에 관한 정보의 저장(캘리브레이션 바 속성 및 처리된 고문서로부터 폴리오와 함께 스캔되고 저장된 그것의 이미지). 만약 그 같이 높은 요구조건들이 없다면, 우리는 기술적 요소들의 수를 제한할 수 있다. 음향과 비디오 영역에서 적용된 부호기(코덱)에 관한 정보는 나중에 재생하는 것과 그것의 품질을 위해 매우 적절한 중요성을 가질 수도 있는 한편, 이미지화하는 데 있어서, 우리는 매우 적은 수의 매개변수들만을 필요로 할 수도 있다.

3.5 커뮤니케이션 메타데이터

이 메타데이터들은 보통 Z39.50¹²⁾ 혹은 OAI PMH¹³⁾와 같은 커뮤니케이션 프로토콜들의 요구조건과 하베스터와 검색도구들에 의하여 제안된 프로파일의 요구조건에 맞추기 위하여 이용 가능한 식별 메타데이터로부터 주로 선택된다. 또한 여기에서 일부 즉시 이용 가능한 커뮤니케이션 프로파일들이 있지만, 그것들은 우리가 일차적으로 관심을 갖는, 협력하는 이용자/전문가 커뮤니티가 아닌 다른 커뮤니티를 반영할 지도 모른다. 예를 들어, 고문서 기술에 대한 우리의 최고의 요구조건은 MASTER 포맷을 적용시키는 것인데, 동시에 우리는 MARC 기반 프로파일링이 더 퍼져있는 보다 일반적 도서관 포털에서 표현되는 것에 관심이 있을 수도 있다. 무엇을 선택하고, 우리의 솔루션을 어떻게 정당화할 것인지 등에는 다시 위험이 있다. 오늘날 OAI PMH가 더욱 더 중요성을 얻기 시작하는 한편, 동시에 대부분의 도서관들은 아직 Z39.50으로 커뮤니케이션 할 것이다.

OAI PMH용 디폴트 프로파일은 Dublin Core Unqualified이지만, 이밖에 추가적 프로파일을 추가할 수도 있다. Manuscriptorium Digital Library는 예를 들어 MARC21 및 특수한 공공용 DTD(OpenM.dtd)¹⁴⁾에 기반한 그 자체의 단순한 식별 레코드를 또한 만들었다.

커뮤니케이션을 위해서, 그것이 포털과 직접 커뮤니케이션 할 디지털 도서관 어플리케이션 일지, 아니면 그 같은 커뮤니케이션이 완전한

12) <<http://www.niso.org/z39.50/z3950.html>>

13) <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>

14) <<http://digit.nkp.cz/MSSXchange/DTD/1.00/OpenM.dtd>>

서지 레코드를 갖는 전자목록을 통해서만(혹은 또한 그것을 통해서도) 가능할 것인지 결정하는 것이 중요하다. 어쨌든, 물론 커뮤니케이션 프로파일들이 상이한 정도나 완전성을 가질 수도 있는 한편, 현실은 두 가지 솔루션을 다 요구할 것으로 보인다.

3.6 추가된 메타데이터 속성과 성질

복합 디지털 문서를 위한 포맷은 도서관 업무 혹은 다른 기록기관들의 업무로부터 발견했거나 아는 것을 어느 정도 확장하는 것을 또한 가능하게 해야 한다. 예를 들어, 사서는 작가를 사진작가, 삽화가, 편집자, 작곡가 등으로 제한하기 때문에, 우리는 그 같은 생각을 따라서 실질적으로 모든 가능한 기술요소들을 제한하고, 그래서 우리의 접근방법을 보다 체계화할 수도 있다. 예를 들어서 디지털화된 정기간행물 포맷에서 우리는 개별 페이지를 한정하는데, 이용자들을 위해서 그것들을 분류하는 것이 체계적으로 되기만 하면, 이용자에게는 매우 이로운 것일 수 있기 때문이다.

```
<!ATTLIST PeriodicalPage
  Type(Advertisement | Blank | Index |
  ListOfIllustrations | ListOfMaps |
  ListOfTables | NormalPage | Table |
  TableOfContents | TitlePage)
  "NormalPage">
  우리 포맷에서, 유사하게 확장한 것이 매우
```

빈번하며, 그것들은 우리의 경험뿐만 아니라, 다른 기관들, 특히 당시에 디지털 정기간행물에 관한 유럽의 프로젝트 DIEPER¹⁵⁾에서 모인 기관들의 경험에도 기반한다.

4. 메타데이터 솔루션: 어떻게 기술할 것인가

위에서, 우리는 복합 디지털 문서에서 무엇을 기술해야 하는지, 혹은 기술할 수 있는지를 논하였고, 이것이 어떤 플랫폼에서 수행될 수 있는지 이미 언급하였다. 당시에 우리는 SGML을 선택하였고, 나중에, 즉 XML이 등장한 후에, 우리의 접근방법을 XML 플랫폼으로 다시 매핑할 것인지, 그리고 언제 해야 할지에 대하여 많은 논의가 있었다.

XML은 사실상 우리가 1995년에 갖고 싶다고 생각했고, 우리가 만든 SGML 하부언어가 표현할 수 있었던 것을 해결했다. 그러나 메타데이터 플랫폼의 변경은 XML의 등장보다는, 예를 들어서 고문서의 전자적 기술을 위한 MASTER 포맷, 디지털화된 정기간행물을 위한 DIEPER 프로젝트의 결과, 혹은 정지 디지털 이미지의 기술을 위한 새로운 표준초안¹⁶⁾과 같은 새로운 내용기술 접근방법들이 등장함으로써 야기되었다.

우리가 XML 플랫폼으로 복합문서 포맷을 재정의한 것은 그 같은 내용기술 때문이었다. 오늘날, 우리는 문서를 다음과 같이 정의한다.

15) <<http://gdz.sub.uni-goettingen.de/dieper/>>

16) Data Dictionary Technical Metadata for Digital Still Images - NISO Z39.87. <http://www.niso.org/committees/committee_au.html>.

- 고문서, 고판본, 및 회귀 고인쇄도서¹⁷⁾
- 디지털화된 근대 단행본(1800년 이후 간행)¹⁸⁾
- 디지털화된 정기간행물¹⁹⁾
- 디지털화된 오디오 문서(초안)²⁰⁾
- 박물관 객체(이행중)²¹⁾
- Manuscriptorium 커뮤니케이션 프로파일
- 역사적 전문(TEI 플랫폼)²²⁾

이 모든 표준들은 서버 <http://digit.nkp.cz>에 공개되어 있고, 'Technical Standards'를 참조하면 된다.

5. 메타데이터의 미래

연구 혹은 개발 활동의 모든 일상적 적용 영역에서 그렇듯이, 최선의 기술이 특정한 목적을 위해 적용되는 때를 결정해야 한다. 한편, 동시에, 그것은 나중에 시작했거나 실험실 테스트 혹은 소량의 데이터로 작업한 것에 기반한 결과를 겨우 발표하고 있는 사람들의 활동과 비교할 때, 실재하는 기술이 노화되고, 뒤처지게 되는 시발점이기도 하다.

발전된 결과를 일상적으로 활용하는 동안 어떤 중요한 기간들이 있다. 어느 정도의 성공적 생산 기간 후에, 우리가 이런 저런 표준, 접근방법, 혹은 방법들을 왜 적용하지 않았는가? 질문

을 받을 수도 있고, 우리가 선택한 기술이 이행되는 순간에 그 새로운 접근방법이 아직 존재하지 않았기 때문에 아무런 과실이 없음에도 불구하고, 비판에 직면해야 하는 때가 보통 발생한다. 그 같은 발전 방식과 대화가 상당히 자연스러운 것이긴 하지만, 또한 심리적으로 우리가 곤란함을 겪을 수도 있는 것이 또한 사실이다.

복합 디지털 문서를 위한 복잡한 포맷의 이행은 생산을 위한 도구들(메타데이터 편집기와 구조화 어플리케이션)의 발전 및 모든 요구된 기능을 충족시키기 위하여 우리 포맷을 디지털 도서관 어플리케이션에 매핑 하는 것에 기반한다. 전체적으로, 이것은 상당히 돈이 많이 드는 시도이다. 차후의 보다 작은 변화들은 또한 비용이 많이 들고, 주요한 변화는 전체 환경의 새로운 발전이나 수정을 위해서 대단한 비용을 요구한다. 그렇기 때문에, 우리가 계속 진행하고 싶으면, 우리의 솔루션은 제 값을 하도록 상당한 기간 동안 가동되어야 한다.

더욱이, 생산된 데이터의 증대된 양과 그것들의 유지 관리는 변화나 수정이 제안될 때, 우리로 하여금 보다 심각하게 고려하게 한다. 왜냐하면 이것은 우리의 경우에는 DOBM 포맷에서 XML 기반 포맷으로 전환했던 이년 전에 그랬던 것처럼, 보통 우리의 모든 생산물이 대대적으로 이동되어야 한다는 것을 의미하기 때문이다.

확실히 새로 생겨난 접근방법들은 계속해서 나타날 것이고, 그것들은 반드시 고려되어야

17) <<http://digit.nkp.cz/MMSB/1.0/msnkaip.xsd>>

18) <<http://digit.nkp.cz/Monographs/DTD/1.0/Monograph.xsd>>

19) <<http://digit.nkp.cz/DigitizedPeriodicals/DTD/1.00/Periodical.xsd>>

20) <<http://digit.nkp.cz/audiodoc/Audio/DTD/1.00/Dokumentace/Audio.xsd>>

21) <<http://digit.nkp.cz/MaterialObjects/1.0/DTD/MuseumObject.xsd>>

22) <http://digit.nkp.cz/MSSFullText/DTD/1.00/mss_fulltext.dtd>

한다. 그러나 그것들이 우리가 해왔던 것을 어느 정도로 구체적으로 개선시킬 것인가에 대해서는 언제나 매우 실질적으로 평가되어야 한다. 때때로 솔루션들은 동일한 방식으로 작동된다는 조건을 갖는 것들보다 더 높은 활용 수준을 가질 책임이 있다.

6. 데이터 솔루션: 어떻게 인코딩할 것인가

데이터 파일들은 디지털화 과정의 직접적 결과이고, 그것들은 실제로 원래 객체의 대용물이나 가상 표현이다. 그것들은 그 주요한 목표 내에서 디지털화 프로젝트에서 정의된 원본 정보의 범주를 용납할 수 있게 해야 한다. 데이터 파일의 품질은 예측된 커뮤니케이션의 주요한 목적을 충족시켜야 한다. 예를 들어, 원래 문서의 인쇄된 복사본을 제공하는 것을 목표로 하는 프로젝트와, 디지털화된 정기간행물을 위한 우리의 Kramerius 프로그램에서 그런 것처럼, 그 목표가 이용자에게 읽을 수 있는 텍스트를 갖는 스캔된 이미지를 제공하는 것인 프로젝트는 상이할 것이라는 점은 명백하다. 여기에서 후자의 경우에, 진짜 컬러의 표현이 요구되지는 않는다. 따라서 이미지는 회색조로만 제작될 수 있고, 그 자체로 혹은 심지어는 원본이 품질이 나쁘지 않을 경우에는 흑백으로 접근될 수도 있다. 이미지의 품질은 따라서 우리의 결정에 달려있다. 이것은 아카이벌 이미지 및 이용자 이미지에 대하여 권장된 기본적 매개변수들로 변형될 것이고, 특히 색상의 수(색도), 해상도, 및 말하자면 손실이 많은 압축 체계가 적

용되었을 경우에는 압축률에 관하여 그렇게 할 것이다.

6.1 아카이벌 파일

아카이벌 파일은 생산된 디지털 문서, 즉, 우리의 작업의 보존을 가능하게 해야 한다. 그리고 그것들은 필요하면, 일정기간이 지난 후에 그것들에게 돌아가서, 변경되거나 개선된 매개변수를 갖는 이용자 품질의 이미지를 다시 만들 수 있게 하는 정도의 품질이어야 한다. 컴퓨터 환경은 매우 빠르게 발전한다. 따라서 인터넷 연결속도가 학술기관에서만 증대되고 있는 것은 아니므로, 얼마 후에는 이용자들이 더 나은 품질을 요구할 수도 있다. 그러므로 아카이벌 이미지는 프로젝트에서 예측되는 나중의 이용을 위해 충분한 여분 정보를 제시할 수 있는 품질로 저장되어야 한다.

이 경우에 적용된 데이터 포맷은 자주 사용되는 것이어야 하고, 그럴 수 있으면, 이미지를 위한 TIFF나 JPEG 같은 ISO 표준인 것이어야 한다. 오디오와 비디오 영역에서, 디지털화는 인정된 보존 방법이고, 파일은 되도록이면 (오디오와 최근에는 비디오조차) 압축되지 않거나, 혹은(비디오는) 무손실 체계로만 압축되어 저장된다. 그럼에도 불구하고, 이것들이 최상의 아카이벌 권장 사항들이다. 대다수의 프로젝트를 위해서 오디오 파일은 WAV 포맷으로 압축하지 않고 저장하는 것이 용납되는 한편, 비디오용으로는 MPEG 그룹에서 선택한 것을 사용하도록 권장될 수 있고, 최근에는 대부분이 아마 더 높은 데이터 흐름(8 mbps 정도)을 갖는 MPEG2가 해당될 것이다. 물론 예

를 들어 VHS 카세트나 Video 8 테이프와 같은 일부 원본의 경우에는 그 같은 요구조건들이 너무 높을 수도 있다. 왜냐하면 원본의 품질 덕분에 3.5mbps와 8mbps 사이에 거의 별 차이를 알 수 없을 것이기 때문이다.

이미지화 하는데 있어서, 진짜 컬러 이미지(16.7백만 색상)와 회색조로 된 이미지를 위해서는 압축하지 않은 TIFF나 약간 압축한 JPEG 파일이 또한 우리의 최상의 요구조건들을 충족시킬 것이다. 흑백 이미지용으로는 TIFF이 CCITT Fax Gr. 4 압축체계를 갖고도 매우 뛰어난 솔루션이 된다.

우리의 디지털화 프로그램에서 이미지는 거의 모든 경우에 아주 약간의 손실(Adobe Photoshop에서 5까지)로 압축되는 JPEG 포맷으로 아카이빙 된다.

따라서, 아카이빙 포맷의 그룹은 매우 축소되고, 우리가 다루어야 하는 이 소수의 아카이빙 포맷들은 변화되어야 할 때, 나중에 과도하게 조작하는 것을 줄여주기 때문에 미래를 위해서는 유리하다.

6.2 사용자 파일

이용자들에게는 보통 축소된 품질의 접근 파일²³⁾이 제공된다. 그 이유는 가능한 저작권 위반과 인터넷 상에서 서비스 속도 때문이다. 오

늘날 일반적 경향은 특정한 목적을 위한 허용 한계보다 약간 높은 접근품질을 제공하는 것이다. 예를 들어, 특정한 목적이 기본적 가독성을 제공하는 것이라면, 우리는 화소의 수와 해상도를 줄이고, 혹은 심지어 더 높은 압축률을 적용시킬 수 있다.

최근, 이용자 접근 파일의 크기를 줄이기 위한 기본적 도구들은 다음과 같다.

- 해상도의 축소(이미지화 및 비디오) 혹은 데이터 흐름의 축소(오디오 및 비디오)
- 손실 압축
- 더 높은 전통적 압축률
- 새로운 압축체계
- 혼합 점방식 이미지 기술(Mixed Raster Image Technology)
- 다중해상(Multiresolutional) 이미지
- 이미지 서버

6.3 해상도와 데이터 흐름

이용자들은 아카이빙용 해상도의 요구조건들을 상당부분 필요로 하지 않는다. 그들은 보통 자기들이 요청한 데이터가 실제로 필요한 것인가를 결정하기 위해 빠른 정보를 필요로 한다; 따라서 그들은 보통(보통은 GIF나 JPEG 포맷으로 된) 썸네일 및/혹은 소형 미리보기

23) 예를 들어: Knoll, Adolf: Quality Levels of Reproduction, In: Atti della Conferenza Internazionale, Scelte e strategie per la conservazione della memoria = Proceedings of the International Conference Choices and Strategies for the Preservation of the Collective Memory = Internationale Tagung, Erhaltung der kollektiven Erinnerung: Strategien und Lösungen. Bolzano, Centro di fotocoproduzione legatoria e restaruro degli Archivi di Stato e dall'Archivio di Stato di Bolzano/Staatsarchiv Bozen, 2005. 512 pp. Pp.317-323.(ISBN 88 7574 048 8), Electronic version in: Passati al futur, Scelte e strategie per la conservazione della memoria, Conferenza Internazionale, Dobbiaco (prov. Bolzano), 25-29 giugno 2002.

이미지, 그리고 약간의 낮은 품질 이미지를 갖는 것을 이해한다. 문서들이 디지털 도서관에 업로드 되기 전에 우리가 모든 사전에 정의된 품질 수준을 생성할지, 아니면 이미지 서버를 통하여 바로 것처럼 낮은 품질 이미지의 일부를 생성할지는 우리에게 달려 있다.

어떤 경우에 우리는 또한 이용자가 더 낮은 품질의 미리 보기 선택권을 갖지 않도록 결정할 수도 있다. 왜냐 하면 어쩌면 원본 객체의 속성이 그렇기 때문에 그렇게 하는 것이 말이 안 될 수도 있기 때문이다(너무 많은 세부사항을 갖거나 예를 들어, 오래된 신문처럼 단조로운 객체일 경우).

오디오 용으로는, MP3나(낮은 데이터 흐름을 갖는) WMA 포맷이 적절한 솔루션일 수 있다: 우리는 재생될 때 문제를 야기할지도 모르는 포맷(MP3pro 혹은 OGG)을 피해야 한다. 인터넷에서 비디오 영역은 WMA나 MPEG4(아마 대개는 DivX)가 또한 용납할 만한 솔루션이다.

우리의 디지털화 프로그램에서, Manuscriptorium은 디지털화된 고문서나 고인쇄본 페이지를 위해 썸네일(GIF), 소형 미리보기(JPEG), 낮은 품질(JPEG), 정상 품질(JPEG), 및 흑백(GIF) 이미지들을 제공한다.²⁴⁾ 그러나 Kramerius는 웨이브렛 기반 이미지로 각 페이지 당 하나의 품질 수준만을 제공한다(아래를 보시오).

6.4 손실 압축

- 전통적 접근방법: 주로 JPEG으로, 우리

는 더 높은 압축률을 적용시키도록 노력해야 한다. DCT JPEG 압축 알고리즘이 이미지를 처리하고, 이용자를 혼란하게 할지도 모르는 전형적인 가공물을 생산하는 방법을 염두에 두면서, 여러 유형의 원본들을 위한 전형적 압축률을 확정하는 것이 권장될 수 있다. 이미지에서 객체들의 매끄러운 표면은, 예를 들어 매우 다양한 상이한 객체들이 있는 정원의 이미지보다 덜 적극적인 압축을 요구한다.

- 새로운 압축 체계: 새로운 압축체계는 흑백과 진짜 컬러 영역에서 따로따로 발전해왔다. 그것들은 JBIG 및 JPEG ISO 이니셔티브에서, 그리고 많은 독점적 비 ISO 솔루션들에서 찾아볼 수 있다. 우리가 새로운 압축체계용으로도 적절한 포맷을 필요로 하므로, 미래를 위해서 유일하게 추천할 만한 제품은, 그것이 인터넷 포맷으로 용납되는 경우에는, JPEG 2000일 수도 있다. 모든 새로운 진짜 컬러용 손실 압축 솔루션은 대체로 일명 웨이브렛 기술에 기반하고 있는데, 이는 전통적인 손실 체계로 압축되는 대형 파일보다는, 유사한 주관적 품질 효과를 제공하는 소형 파일들의 제작을 가능하게 한다. 그러나 더 높은 압축률로, 또한 여기서 전형적 가공물들이 등장하고 있다: 이들은 고전적 JPEG(스퀘어)으로 제작된 것들과는 다른 특성(웨이브)을 갖는다. 어떤 것이 우리를 더 혼란하게 하는가, 그리고 이용자 파일의 크기에서 차이가 얼마나

24) 다음 사이트에서 보기에 자유롭게 접근할 수 있다. <http://www.memoria.cz/special_en/free_docs2.asp>

중요한가는 우리에게 달려있다. 흑백 이미지용²⁵⁾으로는 아래의 Mixed Raster Technology(MRC) 부분에서 언급한 것 말고는, 손실 부문에서 권장할 만한 솔루션은 없다. 당분간 우리의 디지털화 프로그램은 아래 언급한 기술(MRC와 다중 해상 이미지)의 일부로서만 새로운 압축 체계를 사용한다.

6.5 Mixed Raster Content(MRC) 기술

이 접근방법²⁶⁾은 이미지를 전형적으로 전경과 배경 부분으로 나누어진 층들로 수직 분할한 것으로 구성된다. 이 기술은 이미지와 텍스트를 둘 다 포함하는 스캔된 문서의 경우에 가장 적합하지만, 텍스트 문서의 경우에도 뛰어난 결과를 가져온다.

전경은 보통 두 개의 층을 갖는다: 흑백 층(보통 텍스트 층이라 불림)과 그 층에 대한 컬러 배경. 텍스트 층은 보통 더 높은 해상도 영역(전형적으로 300 dpi)에서 처리되고, 그 배경은 그 목적이 단지 채색 착시를 추가하는 것일 뿐이기 때문에, 어떤 솔루션들에서는 심지어 25 dpi 이상 정도만을 가질 수도 있다. 그 배경이 일부 효과적 웨이브렛 압축기로 처리되는 한편, 텍스트 층은 매우 효과적인 최신 무손실 혹은 손실 압축 체계(CCITT Fax Gr. 4 혹은 주로

JBIG 계통)로 압축될 수 있을 것이다.

배경은 단 하나의 층(적절한 이미지 층)만을 가질 수도 있고, 혹은 또한 더 많은 층들로 분할될 수도 있다. 모든 배경 층들은 항상 효과적인 웨이브렛 압축기로 압축되고, 주로 낮은 해상도 수준(전형적으로 100 dpi)을 갖는다.

상기한 사실들로부터 보면, 컬러의 수는 가능하면 축소되고, 해상도도 그런 한편, 이미지의 각 부분은 이용할 수 있는 가장 효과적인 도구들로 처리된다는 것이 명백하다. 그 같은 접근방법은 따라서 우수한 가독성을 제공하는 소형 파일에 훌륭하게 작용한다. 그 같은 파일의 크기는 보통 비 MRC 접근방법의 경우에서 파일의 크기보다 몇 배 작다. 크기는 물론 이미지에 들어있는 객체들의 특성에도 영향을 받는다.

이용 가능한 솔루션들을 비교하는 데 있어서 중요한 점은 압축 전에 발생하는 자동화된 사전처리 일정일 수도 있다. 이렇게 사전 처리하는 동안, 이미지는 층들로 분할되고, 또한 모두 전체 이미지 파일의 최종 크기에 영향을 미치는 다른 작업이 발생할 수도 있다. 다른 중요한 점은 이미지를 처리하는 사람이 MRC 기술이 어떻게 작동하는지에 관하여 갖는 지식일 수도 있다. 그 이유는 부적절한 매개변수가 매우 불만족스러운 결과를 가져올 수 있기 때문이다. 보통 MRC 솔루션은 우리가 소스 이미지의 특성이 그것을 요구할 수 있다고 판단할 때마다,

25) Knoll, Adolf: Compression of bi level images: compressor performance report, CD RO으로도 간행: Informace na dlani 2000, Praha, Albertina icome Praha, 2000 프라하에서 2005년 5월 23일 25일에 열린 INFORUM 2000 Conferenc에서 발표된 영어 논문들 가운데. <http://digit.nkp.cz/knihcin/digit/vav/bi-level/Compression_Bi_level_Images.htm>에서도 볼 수 있음.

26) 다음 자료를 더 보시오: Knoll, Adolf: New Image Formats and Approaches for Document Delivery and Their Comparison with Traditional Methods, CD RO으로도 간행: Informace na dlani 2001, Praha, Albertina icome Praha, 2001 프라하에서 2005년 5월 23일 25일에 열린 INFORUM 2000 Conferenc에서 발표된 영어 논문. <<http://www.inforum.cz/inforum2001/English/prispevky/knoll.ht>>에서도 볼 수 있음.

전경과 배경 분할을 또한 전환시킬 수 있게 한다. 예를 들어, 텍스트 이미지는(언제나 그렇지 않지만) 배경을 끄고, 따라서 소스를 흑백 이미지만으로 진동하게 할 수 있는 한편, 실사 이미지²⁷⁾는 전경을 끄고도 적절하게 처리될 수 있다. 따라서 아래에서 지적한 특수한 MRC 포맷은 흑백과 실사 이미지로도 성공적으로 사용될 수 있다.

오늘날 시장에서 살아남은 이용 가능한 솔루션은 DjVu²⁸⁾와 LDF²⁹⁾ 포맷이다. DjVu는 모두 일곱 개의 층(전경에 2, 배경에 4)을 갖고, LDF는 세 개의 층(전경에 2, 배경에 1)을 갖는다. 흑백층을 위한 압축 솔루션은 DjVu 에서 독점적 JBIG 유사 압축기와 LDF에서 CCITT Fax Gr. 4 혹은 새로운 JBIG2이다. 컬러 층용으로는 LDF가 LWF 이미지나 새로운JPEG 2000을 사용하는 한편, DjVu는 독점적 웨이브렛 IW44 솔루션을 사용한다.

MRC 솔루션은 ISO 기반이 아니지만, 우리가 볼 수 있는 것처럼 LDF는 두 구성요소들이 ISO 표준인 포맷을 제공하기 시작했다.

체코공화국 국립도서관은 Kramerius 프로그램에서 디지털화된 콘텐츠에 대한 모든 접근을 위해서는 DjVu 이미지를 사용한다.

6.6 다중해상 이미지

어떤 원본은 매우 크고, 많은 미세한 세부사항을 포함한다. 이것은 특별히 지도자료의 경우

에 그렇다. 만약 이들을 갖고 작업을 하고 그 같은 세부사항을 볼 수 있기 원하면, 고해상도 이미지를 생성할 필요가 있다. 그러나 원본의 물리적 크기가 크다는 사실을 감안하면, 비압축 디지털 이미지는 수백 메가바이트를 가질 것이다. 그 같이 큰 이미지는 - JPEG으로 압축된다 하더라도 압축을 푼 후에 컴퓨터를 마비시키는 엄청나게 큰 파일을 갖고 작업을 해야 하기 때문에, 다루기가 어렵다. 이 문제에 대한 해결책은 - 작은 것부터 최대의 것까지 - 여러 이미지들이 함께 저장되는 이미지 파일이 될 수 있다. 그 같은 이미지 파일은 다중해상 이미지 파일이라고 불리고, 적절한 접근도구들과 결합될 때, 대형 데이터 파일을 조작하기 쉽게 만들 수 있다. 접근될 때 작은 이미지만이 매우 빨리 디스플레이 되기 때문에 우리는 디지털화된 자료의 특성에 관한 어느 정도의 정보를 즉시 가질 수 있다. 예를 들어, 지도에서 구체적 지역처럼, 미세한 세부사항까지 보도록 이미지에 깊이 들어가고 싶으면, 파일로부터 더 높은 해상도의 이미지가 사용되고, 전체 이미지 표면으로부터 지적된 덩어리만이 전달된다.

이 분야에서 가장 성공적인 기술은 MrSID³⁰⁾ 포맷(Multiresolutional Seamless Image Databank)이다. 이것은 특수한 웨이브렛 압축 체계를 사용하고, 접근을 위해서 이용자는 이 기술을 효과적으로 이용하기 위해 특수한 브라우저 플러그인을 사용해야 한다.

MrSID 포맷(*.sid)은 스캔된 역사지도 접근

27) 다음 사이트에서 보기를 볼 수 있음. <<http://digit.nkp.cz/knihcin/digit/vav/wavelet/Samples.html>>

28) <<http://www.lizardtech.com>> 혹은 <<http://www.djvuzone.org>>를 보시오.

29) <<http://www.luratech.com>>을 보시오.

30) <<http://www.lizardtech.co>>을 보시오.

용으로 Manuscriptorium Digital Library에서 사용된다. 일반적으로 우리의 MrSID 파일은 동일한 이미지의 10개 해상도 수준까지 포함하고 있다.

6.7 이미지 서버

이미지 서버는 이미지 품질 수준을 요구에 따라 전달할 수 있게 한다. 이들은 보통 바로 요청된 접근 데이터를 생성한다. 이들은 요구에 따라 압축하고, 색도를 변경하거나, 보다 효과적인 전송 및/혹은 디스플레이를 위해 포맷할 수 있다. Manuscriptorium Digital Library에서는 디지털화된 역사지도에 대해 이용자가 편하게 접근할 수 있도록 Lizardtech Image Express Server를 사용한다. 이 서버는 요구 시에 요구된 데이터 덩어리를 전송하는 MrSID 파일을 갖고 작업한다. 이 방식으로 대형 이미지 파일 - 우리 경우에 한 이미지가 0.5 GB의 비압축 데이터까지 - 이 느린 인터넷 연결에서조차도 상당히 빠르고 쉽다. 서버는 특수한 MrSID 플러그인과 함께 작동하고, 플러그인이 설치되지 않은 경우에는, 비록 그 작업이 특수한 플러그인과 하는 것처럼 편안하지는 않지만, 기본적인 확대와 절단도 가능하게 한다.

7. 연구, 개발과 생산

위에서 언급한 모든 활동은 다양한 재원이 지원되는 연구개발 프로젝트에 적극적으로 참여하지 않는다면 이루어질 수 없었다. 국립도서관은 보존(아날로그와 디지털)과 접근 영역

(현대 및 역사자료)에서 인정된 연구개발 주체이다. 재정지원 기구는 주로 문화와 교육의 국가부처이고, 또한 Framework Programmes, eContent, Eureka! 등과 같은 다양한 프로그램 하에 일명 European 프로젝트에 우리가 참여하는 경우에는 유럽 연합위원회(European Commission) 이기도 하다.

역사 및 희귀 소장품에서 디지털 데이터를 생산하고 유지관리하는 것은 우리가 모든 것을 다하려고 애쓰기보다는, 훨씬 융통성 있는 체코의 민간기업들과 협력하여 부분적으로 운영되고 있다. 많은 실제적 지식은 디지털 문서의 생산과 전달의 모든 단계에서 만들어졌다. 직원과 이용자로 하여금 문서를 생성하고 그것에 접근하도록 돕는 다양한 도구의 준비뿐만 아니라, 스캐닝 과정, 그 동안의 희귀문서의 처리에 다 관련된다.

접근은 저작권 문제 때문에 때때로 문제가 되기도 한다: 예를 들어, 우리는 저자가 사망한 지 칠십 년 이상 된 문서만 디지털로 무료로 이용하도록 할 수 있다. 이것은 수많은 저자가 나타나고, 모든 저작권 문제들을 해소하기가 어려운 디지털화된 정기간행물의 영역에서 특히 큰 장애를 야기한다. 따라서 정기간행물에 대해서는 원본으로 발행된 것이 1880년 이전인 권호이면 무료로 접근할 수 있게 한다. Manuscriptorium은 보다 나은 품질 이미지에 대해서는 제한된 접근을 허용한다. 이들은 오직 등록된 라이선스를 갖는 이용자들만이 이용할 수 있다.

다른 중요한 문제는 특별히 채색 고문서에 관련된, 디지털 이미지에 대해 가능한 저작권 위반이나 남용에 관한 것이다: 따라서, Manu-

scriptorium Digital Library에서 이른바 정상 품질 이미지들은 MMSB(Memoriae Mundi Series Bohemica)라는 문자로 가시적으로 워터마크가 되어 있다.

디지털화된 원본의 숫자가 증대하는 것과 Manuscriptorium이 역사적 컬렉션들의 공유 종합목록으로서 기능하는 것은 외국의 기관들로 하여금 역사연구의 가상환경을 만드는 것에 참여하도록 관심을 불러일으켰다. 우리는 리투아니아, 폴란드, 슬로바키아, 크로아티아, 헝가리, 오스트리아 등과 같은 여러 나라들로부터 관심의 표현, 협정 및 실제로 기여하는 기관들을 이미 갖고 있다.

상이한 기관들이 다양한 표준을 사용하므로, Manuscriptorium - 호환적 데이터를 얻거나, 그들의 데이터의 변환과 업로드를 위해(대부분의 경우에) 특수한 변형과정을 만드는 것이 언제나 필요하다. 이 작업은 또한 비용이 들기 때문에 주로 그 같은 확대는 공통의 국제 프로젝트 내에서 가능하다. 또한 호환되는 표준을 공

공연히 적용하는 것조차도 데이터가 아무런 향후의 작업 없이 호환된다는 것을 의미하지는 않는다. 대개는 동일한 표준도 동일한 결과를 의미하지는 않는다.

국립도서관을 둘러싸고, 여러 가지 유형의 문서를 디지털화하기 위한 표준화된 접근방법이 발전되었고, 이들 가운데 상당수가 공유될 수 있거나, 예를 들어 우리의 문서 포맷이나 무료로 다운로드할 수 있는³¹⁾ Manuscriptorium 호환 문서용 편집기와 인증기와 같이 공개적으로 이용할 수 있다. 시간이 흐르면서 모든 것이 발전하고 또한 변화하고 있다. 그러나 어떤 타당한 표준을 준수하고 그것을 일관성 있게 하는 것이 중요하다. 왜냐하면, 일관성이 적어도 상위의 정보 수준에서, 특히 국가적³²⁾ 혹은 국제적 수준의 공동 포털을 통하여, 호환성을 보장하기 때문이다.

▷ 번역: 윤정옥(청주대학교 문헌정보학과 교수)

31) <http://www.memoria.cz/site_en/download.asp>

32) 예를 들어, Czech Information Gateway의 <<http://www.jib.cz>> 이나 The European Library(TEL)의 <<http://www.theeuropeanlibrary.org>>.