

국가 지식정보의 디지털화 전략

Digitisation Strategy for National Collections

정상원, 한국데이터베이스진흥센터
Jeong Sang-Won, Korea Database Promotion Center
bstone@dpc.or.kr, http://dpc.or.kr/~bstone

초 록

국가 지식정보자원의 디지털화를 위한 기초로서 국내외의 디지털화 현황을 고찰하고, 디지털화 대상 및 우선순위, 디지털화 방법, 디지털정보에 대한 접근, 디지털정보의 보존 분야의 핵심 문제점과 해결방안을 모색하였으며, 효과적인 디지털화를 추진하기 위한 디지털화 추진단의 구성 및 운영방안을 제안하였다.

1. 서언

디지털 자원은 지식정보사회를 이루는 핵심의 하나로 그 가치가 날로 높아지고 있다. 디지털화는 단순히 아날로그나 인쇄형태의 정보자원을 디지털형태로 변환하는데 의의가 있기보다는 총체적 문화창달을 위한 조건을 조성한다는 데 의의가 있다. 디지털화는 사회·문화적 가치를 확대시킴과 동시에 지식기반사회의 성공적인 전환을 보장할 수 있다. 디지털화는 또한 국가적 지식과 경험에 대한 인식증진과 접근향상에 기여할 뿐만 아니라 콘텐츠 및 멀티미디어 산업분야의 발전과 고용창출의 기회를 확대시킬 수 있다.

지식정보자원 디지털화의 효과를 최대화하기 위해서는 디지털화 대상과 우선순위, 디지털화 형태, 절차·방법·추진체계, 그리고 디지털정보의 보존·접근·활용을 위한 종합적인 마스터플랜이 필요하다. 디지털화에 대한 종합적인 전략이 있어야만 지식정보자원의 디지털화 중·장기 추진계획의 수립이 가능하고 디지털화의 일관성, 연속성, 효율성을 제고시킬 수 있으며, 예산·시간·인적 낭비를 미연에 예방할 수 있게 된다.

본 연구는 국가 지식정보자원의 효과적인 디지털화를 위한 기초로서 디지털화 대상 및 우선순위, 디지털화 방법, 디지털 정보의 접근 그리고 디지털 정보의 보존으로 디지털화 이슈를 나누고 각각에 대한 추진전략을 모색하고자 한다.

2. 국내외의 디지털화 현황

디지털화를 성공적으로 추진하고 있는 나라의 공통점은 디지털 정보를 총체적 문화유산의 계승과 산업발전의 차원에서 접근하고 있으며, 또 디지털화 마스터플랜을 수립하고 있다는 것이다. 미국은 1993년 NII's Agenda for Action을 수립하고 digital library program을 비롯하여, 국회도서관의 American Memory Project, NSF의 Digital Library Initiative 등을 추진하고 있다. 민간에서도 RLG의 Making of America 등 여러 가지 프로젝트를 통해 국가 내에 유통되는 정보, 특히 역사 및 문화와 관련된 자료를 디지털화하고 있다.

캐나다는 1996년 'Building the Information Society: Moving Canada into the 21st Century'라는 국가 정보화 계획을 수립하고 4개의 실행계획 중 콘텐츠의 확충 프로그램을 수행하였다. 이 중 콘텐츠 확충의 일환으로서 Task Force on Digitization을 구성하여 디지털화 자료의 선정, 디지털화 자금지원, 수익창출, 탐색도구, 표준, 연계, 저작권, 보존, 협력 문제 등에 대한 전략을 제시하고 있으며, 그 하부 프로그램으로 Canadian Heritage Information Network(CHIN), Access Amicus Project(<http://www.nlc-bnc.ca>), Schoolnet Digital Collections Program(<http://www.schoolnet.ca/collections/>), CINCH(Canadian Initiatives on Network Clearing House, http://strategis.ic.gc.ca/sc_indps/sectors/engdoc/cinch_hpg.html) 등을 추진하고 있다.

유럽연합에서는 Fifth Framework Programme for Research and Technological Development(<http://www.cordis.lu/fp5/>) 중 정보사회 기술 프로그램을 수립하고 '멀티미디어 콘텐츠 및 도구'라는 하부 활동지침으로 '과학 문화 유산에 대한 접근 및 문화유산의 디지털보존 사업'을 추진하고 있다.

프랑스는 1998년 'Action Plan for Information Society'의 6개분야 정보화계획의 하나로써 '프랑스 문화유산의 디지털화 프로젝트'를 추진하고 있다.

노르웨이는 정보기술 국가위원회의 'The Norwegian Way to the Information Society. Bit by Bit'라는 보고서를 토대로 디지털 공간에서의 노르웨이 언어와 문화의 전파를 위해 1991년부터 도큐멘테이션 프로젝트를 추진하고 있다.

호주의 경우 'Digitisation of National Collections'(<http://dms.sas.gov.au/digitn/index.html>)를 통해 정보자원의 디지털화 및 전자접근과 관련된 주요 해외활동에 대한 조사, 주요 디지털화 프로젝트의 감사, 디지털화 및 전자접근에 대한 의사결정 정책 구조의 수립, 효과적이고 효율적인 디지털화를 위해 기관 및 정부가 해결해야 할 주요 전략적·정책적 이슈 등을 해결하고 있다. 또한 통신정보예술부(DCITA)에서는 Digitisation

Forum Online(<http://www.digitisation.net.au/index.html>)을 구축하여 디지털화의 의미, 디지털화 절차 및 디지털 객체의 관리, 변환과정 관리, 저작권관리 등에 관한 정보를 제공하고 있다.

이상과 같이 정보선진국에서는 정보화의 핵심으로서 디지털화에 관한 종합적인 계획을 수립하여 디지털화 프로젝트를 보다 효율적으로 추진하고 있다.

우리나라의 「정보화촉진계획」(1996)은 공공기관의 정보화와 정보통신산업, 초고속정보통신망을 축으로 하고 있기 때문에 지식정보자원의 디지털화를 위한 종합적인 정책이나 전략의 수립에 소홀할 수밖에 없었다.

디지털화의 범주에 들어갈 수 있는 사업중 「공공DB 개발·보급사업」(1994~1997)은 DB개발자의 제안에 따라 과제를 선정함으로써 전체적인 정보수급관계를 고려한 총괄적 정보개발계획이 없으며, 비체계적인 선정으로 개발비용이 중복되거나 개발분야가 편협하며, 개발 과제수가 지나치게 많아 공공DB의 의미가 희석되었다.

「정보개발지도제작사업」(1995)은 개발되어 서비스중이거나 개발을 추진중인 정보, 미개발 개발 공공정보자원을 조사하고 분류하여 년차별 개발분야 및 우선순위 설정을 위해 정보지도도를 작성하였다. 그러나 상업화 및 경제논리에만 치중하여 정보를 분류하고 우선순위를 설정하였기 때문에 정보자원의 디지털화의 방향과 목표, 범위에 대한 전략적 접근이라 볼 수 없다.

현재는 실업자문제의 해결과 정보화의 촉진을 위해 「정보화근로사업」(1998년도 11개, 99년도 23개 과제)이 추진되고 있으나 디지털화에 대한 종합적인 마스터플랜에 따라 추진되고 있지 않기 때문에 정보의 중복 및 누락현상이 발생할 수 있으며, 구축된 디지털정보간의 상호운용성과 표준의 결여로 정보에 대한 원활한 접근이 어려울 것으로 예상된다. 더욱이 강기적인 관점에서 디지털정보에 대한 접근과 이용에 필요한 보존에 대해서는 전략이 마련되어 있지 않을 뿐더러 디지털화와 관련된 아이디어와 전략을 종합적으로 공유하고 있지 못하고 있다.

요컨대, 우리나라의 디지털화에 대해서는 보유정보의 효율적인 접근과 이용을 위한 종합적인 청사진 부재, 디지털화 대상, 우선순위 등 디지털화에 대한 총체적 방법론의 부재, 정보자원에 대한 상호연계 및 접근체제의 전략부재, 정보자원의 보존체계 미비 등을 문제점으로 지적할 수 있다.

3. 디지털화 대상 및 우선순위

디지털화의 핵심문제 중에 하나는 어떠한 지식정보를 디지털화 대상으로 선정하고 어떤 것을 우선적으로 디지털화 해야 할 것인가 하는 문제이다. 기본적으로 디지털화할 대상은 총체적인 문화역량을 재고할 수 있는 정보, 인프라적 성격이 강한 정보, 국가발전을 위한 공동기반의 확충에 이바지할 수 있는 정보, 교육과 학

습에 도움이 되는 정보, 국가의식과 국민의 지적 향상, 경제성장과 고용창출에 이바지 할 수 있는 정보이어야 한다. 이러한 측면에서 정부 및 공공부문은 지식정보자원을 가장 많이 보유하고 있는 주체이다. 따라서 우선적인 디지털화의 대상은 공공부문의 정보가 되어야 한다. 그러나 현행 예산 현실에 비추어볼 때 지금까지 공공부문에서 수집하거나 생산한 모든 지식정보자원을 소급하여 디지털화하기란 불가능하기 때문에 디지털화할 자료를 선정기준을 개발하고 우선순위를 부여해야 함은 당연한 귀결이다.

디지털화자료를 선정하기 위한 기준을 개발하기 위해서는 현재 진행되고 있는 디지털화 활동에 대해 파악이 선행되어야 한다. 정부 및 공공기관의 정보자원 보유실태, 현재 진행되고 있거나 계획하고 있는 디지털화 활동의 특성과 범위, 디지털화의 목적(보존, 수익창출, 전시)에 대한 정확한 파악 없이는 디지털화 대상선정을 위한 기준과 우선순위를 제대로 설정할 수 없기 때문이다. 같은 이유로 디지털 정보의 목표이용자를 확인하고, 이용자의 요구와 기대를 파악하기 위해 이용자가 원하는 정보의 형태, 정보추구행위, 선호하는 접근 방식에 대한 조사도 병행해야 할 것이다.

한편, 디지털화된 정보를 장기적으로 유지하기 위해서는 상당한 비용과 인적자원이 필요하기 때문에 디지털화 대상자료의 선정에 있어 비용대 효과를 측정할 수 있는 모델이 개발되어야 한다. 디지털화된 정보에 대한 접근과 효과적인 이용과 관련된 기술수준도 디지털화 대상선정에서 고려해야 한다. 예컨대, 이용자가 미술작품이나 지도와 같이 용량이 큰 그래픽파일을 볼 수 있는 기술을 활용할 수 없다면 디지털화된 정보의 배포는 불가능하기 때문이다. 반대로 디지털정보에 대한 이용자의 기대수준에 못 미칠 수 있는 가능성도 있기 때문에 디지털화 대상자료의 우선순위와 이용자의 기대수준을 충족시키는 것도 중요하다.

이상과 같은 사항을 염두에 두고 디지털화 대상자료의 현황을 파악한 다음 이를 데이터베이스로 구축하고 DB가 국가 지식정보자원의 디지털화에 관한 종합적인 인벤토리와 우선순위 설정에 소용될 수 있도록 해야 할 것이다.

4. 디지털화 방법

디지털화할 대상자료가 선정되고 우선순위가 부여되고 나면 디지털화 절차, 목표, 범위 등에 대한 구체적 계획을 수립하여 추진해야 할 것이다. 디지털화 과정에서 시간적·인적낭비를 없애고, 디지털 정보의 접근과 이용에 있어 상호운용성과 호환성을 확보하기 위해서는 최소한의 합의와 표준화가 필요하다. 따라서 표준화는 디지털화에 있어 가장 큰 비중을 차지하게 된다. 그러나 새로운 표준이 계속 출현하고 있고, 표준의 개발과 이용에 있어 이해당사자가 증가하고 있으며, 디지털화 기술이 지속적으로 변화하고 있기 때문에 어떤 것

이 디지털화에 있어 최상의 표준인지를 판단하기 어렵게 되었다. 그럼에도 불구하고 적절한 표준을 적용함으로써 디지털화 과정과 그 결과에 있어 불필요한 낭비를 최소화하고 효과를 극대화할 수 있으며, 표준화된 형식의 정보검색이 가능하므로 디지털정보 이용의 편의성과 정보요구에 대한 만족도를 높일 수 있다. <표 1>은 현재 디지털화에 적용 가능한 표준 또는 사실표준이다.

<표 1> 디지털화 관련 표준

표준범주	표준
OSI 및 인터넷 데이터 전송	TCP/IP, Telnet, FTP, FTAM-OSI, Gopher, HTTP, MIME, SMTP, X.400 Mail(OSD), MARC
문자집합	ASCII, KS X 1001, ISO/IEC 10646 UNICODE
문서교환	ASCII, RTF, Word processing Format(DWP, DOC.), Desktop Publishing Format(PageMaker, QuarkExpress), Structured Format(SGML, HTML, XML), Page Descriptor, Format(LaTeX/TeX, PostScript, PDF)
소리교환	WAV, RA, ALI, AIFF, MD
이미지 교환	CITF Group 4 Fax, CGM, GIF, JPEG, PhotoCD, PNG, TIFF
동영상 및 멀티미디어	AVI, GIF89a, MPEG, MOV, RealVideo, Shockwave
3차원 데이터흐름	VRML, QuickTime Virtual Reality
탐색 및 검색	Common Indexing Protocol, Z39.50(Z39.50Command Language), ISO10160/10161(ILL), SQL, SQLMM
Directory	ISO 9594, LDAP, WHOIS, WHOIS++
메타데이터	LC/Warwick Framework, EAD(SGML DTD), GILS, MCF, PKCS, SCIF, TEI header(SGML DTD), URC
Locator	ISBN, ISSN, URL, PURL, SICI(Z39.50), URN

어떠한 목적으로 디지털화를 추진하든지 간에 디지털화 프로젝트의 추진에 있어 디지털화 기획 및 관리, 디지털화 과정 및 절차, 이용자 요구사항분석, 정보의 특성 및 유통방식에 대한 분석, 디지털화 도구·인력 등 자원의 준비, 적절한 디지털화 기술의 선정, 디지털화 대상자료의 준비, 적용가능 표준 및 최신기술, 정보의 형태 및 매체별 디지털화 방법, 디지털화 모델, 디지털화 실무를 담당할 기관 또는 업체의 자격심사, 디지털화 비용산정 및 회수, 디지털화 재원확보, 디지털 정보의 기술 및 식별, 디지털 정보에 대한 접근방법, 디지털화 도구 평가, 디지털화 완료후 디지털 정보의 평가 등 디지털화와 관련된 종합적인 지침을 개발할 필요가 있다.

현재는 디지털화에 대한 종합적인 청사진이 제시되어 있지 않을뿐더러 디지털화를 실지로 담당할 기관에서조차도 자체적인 디지털화 지침을 개발한 사례는 극히 드문 실정이다. 이에 반해 미국의 국회도서관, Research Library Group, 영국의 국립도서관 등 많은

기관에서 '디지털화 지침'을 개발하여 이용하고 있다. 또, 호주의 경우에는 Digitisation of National Collections 나 Digitisation Forum Online은 디지털화 지침을 보급하고 있으며, NINCH Working Group on Best Practices(<http://www.ninch.org>)에서도 전문가로 구성된 실무조사·분석반을 운영하여 디지털화 방법·과제 등을 제안하고 디지털화 지침을 제정하고 있다.

5. 디지털 정보에 대한 접근

디지털정보에 대한 접근의 문제는 기술적 측면과 정책적 측면으로 구분하여 생각할 수 있다. 기술적 측면에서 디지털정보는 분산된 디지털저장소를 통해 접근할 수밖에 없다. 따라서 디지털 정보에 대한 단절없는 접근을 위해서는 이들 정보저장소간 상호운용성의 확보와 상호연계가 필요하다. 상호운용성이 확보됨으로써 이용자는 시공간적 제약 없는 원활한 접근이 가능하게 된다. 디지털저장소 또는 데이터베이스간 상호운용성 확보를 위해서는 디지털 콜렉션의 연계, 통합의 수단으로서 메타데이터, 시스템간 탐색, 디지털 콜렉션간의 접근 등과 관련한 일정한 합의와 최소한의 지침이 필요하다.

한편, 디지털 정보의 접근 문제는 디지털 정보 접근 기술의 사용 편의성과 밀접하게 관련되기 때문에 디지털 콘텐츠의 탐색도구나 접근을 위한 인터페이스 설계 지침에 관한 연구도 있어야 할 것이다.

결과적으로 디지털 정보의 기술적 접근의 문제 해결은 상호운용성의 확보 및 DB간 통합, 그리고 접근의 편의성을 확보하기 위한 합의도달과 지침개발이 관건이다.

정보정책의 측면에서 디지털정보에 대한 접근의 원활화는 국가적 일체성과 사회적 통합의 촉진, 문화향유와 경제발전에 크게 이바지할 수 있다. 디지털 콘텐츠에 쉽게 접근할 수 있게 됨으로써 정부기관간의 협동이 활성화되고 보다 적극적인 서비스가 가능하며, 국민의 참여기회를 확대시킬 수 있게된다. 이와 관련하여 국가나 공공기관에서 보유하고 있는 정보는 크게 필수정보, 핵심정보, 주문정보로 구분된다. '필수정보'는 정부기관이 법률에 의해 세금의 지원을 받아 국민에게 제공해야 하는 정보로서 건강, 사회안전, 국민의 권리와 의무에 관련된 정보이다. '핵심정보'는 정보제공에 대한 법률상의 명문규정은 없으나 경제적 기회, 사회통합 등 국민생활에 중요한 정보로서 무료 또는 실비로 제공되어야 한다. '주문정보'는 개인 또는 법인에 이득이 되며, 개인적 목적으로도 이용될 수 있는 정보로서 기본적으로 각각의 이용자가 고객의 입장에 서게 된다. 이러한 정보의 접근과 활용에 있어 필수정보와 핵심정보의 경우는 보편적이고 균등한 서비스를 위해 무료 또는 실비로 접근하게 하고, 주문정보는 민간의 콘텐츠 산업을 발전시키기 위해 상업화해야 할 것이다. 참고로, 이상과 같은 정보의 구분은 디지털화 대상자료를

선정하는데 있어서도 고려사항이 될 수 있을 것이다.

6. 디지털정보의 보존

디지털화는 일차적으로 디지털정보의 접근과 배포에 초점을 맞추고 있다. 그러나 디지털정보는 근본적으로 역동적인 정보이기 때문에 전통적인 체계에 비해 훼손되기 쉽다. 또 단계적으로 조작, 배포 복제, 연결시킬 수는 있지만 정보기술의 지속적인 변화로 인해 장기적인 보존이 어렵다는 특성이 있다. 따라서 이러한 어려움을 극복하고 장기적인 보존과 접근을 위한 수단과 전략을 개발할 필요가 있다.

디지털정보는 그 양의 방대성과 인적·물적 자원의 제한 때문에 모두를 보존하기는 사실상 거의 불가능하다. 따라서 어떤 정보를 보존해야 할 것인가에 대한 선택의 문제가 대두된다. 장기 보존할 디지털정보를 선택할 때에는 정보의 가치에 대한 질적 평가가 요구된다. 그러나 선택과정에서 주관성이 개입됨으로써 잘못된 판단으로 이어질 위험성이 있으며, 동시에 선택된 정보가 미래에도 유용할 것이라 미리 예측할 수 없다. 이를 해결하기 위해 국가차원의 디지털자료 선택기구가 일정기간을 정해놓고 선택기준에 따라 보존되어야 할 자료를 선정하는 방법이 하나의 대안이 될 수 있을 것이다. 또한 디지털정보의 특성측면에서 어떤 데이터는 그 가치를 계속유지하기 위해 끊임없는 갱신이 필요가 있는 반면 어떤 자료는 그렇지 않기 때문에 정보의 성격에 따른 보존기준이 달리 설정되어야 한다. 보존되는 정보가 향후에 얼마나 이용되고 이용될 수 있기 위해서는 상황정보가 포함되어야 한다. 이 때문에 디지털정보와 관련된 상황정보를 어떻게 포함시킬 지에 대한 기준의 설정 또한 필요하다.

일반적으로 디지털정보를 보존하는 방법에는 크게 복사(copying), 이송(migration), 에뮬레이션(emulation), 방치(archiving) 등이 있으며, 디지털저장소의 구축에 있어서도 집중형, 분산형, 협력형, 혼합형 모델이 있으며 나름대로 장·단점이 있다. 따라서 이처럼 다양한 보존방법과 모델을 비교·분석하여 적합한 보존 모델을 도출해야 할 것이다.

이상과 같이 디지털정보의 보존을 둘러싼 다양한 문제가 존재함에도 불구하고 현재는 디지털정보의 보존과 관련된 표준화된 가이드라인이 없는 실정이다. 따라서 디지털 정보의 보존을 책임지고 있는 기관이 공동으로 보존대상의 선택 및 기준, 선정절차, 보존전략, 보존방법, 메타데이터, 저장포맷, 매체의 형태, 플랫폼의 형태, 보존비용 등에 대한 지침과 표준을 개발해야 할 것이다. 현재, 미국의 Internet Archive, 영국의 CEDARS, 호주의 PANDORA, EU의 NEDLIB, 덴마크의 INDOREG, 핀란드의 EVA, 스웨덴 Kulturarw3 프로젝트 등 많은 국가에서 디지털 정보에 대한 보존 프로젝트를 추진하고 있으며, 이를 위해 나름대로 디지털 정보보존에 관한 지침을 개발하여 이용하고 있다. 특히

RLG의 RLGPRESERV(<http://www.rlg.org/preserve/>), Society of American Archivists, Task Force on Archiving of Digital Information 등에서 제시한 지침은 우리의 디지털 보존을 위한 지침개발에 있어 유용한 참고자료가 될 것이다.

한편, 보존과 관련하여 많은 디지털 정보 특히, 원래부터 디지털로 생산된 정보는 납본대상이 되지 않고 있기 때문에 디지털정보의 망라적이고 효율적인 보존을 위해서는 디지털 납본제도의 도입이 시급하다.

7. 결론 및 제언

국가 지식정보자원의 디지털화는 지식정보사회로의 성공적인 이행을 위한 필요조건이다. 디지털화를 성공적으로 추진하고 그 효과를 극대화하기 위해서는 앞서 제시한 4가지 이슈에 대한 충분한 논의와 협의가 필요하다. 이와 관련하여 캐나다나 호주의 예처럼 범정부간 기구로서 가칭 [디지털화 추진단]의 구성을 제안하고자 한다.

추진단에서는 정보자원의 디지털화계획 및 프로젝트 조사, 디지털화의 타당성 연구, 이용자 연구, 디지털화 추진기관의 협력체계 수립, 디지털화 대상 자료선정을 위한 기준설정 및 디지털화 우선순위 설정, 디지털화 표준 및 지침의 개발과 보급, 디지털화 장비와 인력의 개발 및 훈련, 정보자원 디지털화 및 전자접근에 대한 정책구조 개발, 디지털정보의 보존을 위한 모델의 개발, 지적소유권과 관련된 정책의 수립 등 디지털화와 관련된 모든 사항을 망라적으로 취급해야 할 것이며, 또 이를 위해 각 분야별 전문가로 구성된 연구위원회를 조직하여 운영해야 할 것이다.

참고문헌

- [1] Federal Task Force on Digitization. 1997. Towards a learning nation: the digital contribution. ; Recommendation Proposed by the Federal Task Force on Digitization. Final Report. <<http://www.nrc.ca/dtf-gtr/english/report/report.doc>>
- [2] Digitisation of national collections-Australia. [1999]. <<http://dino.slsa.sa.gov.au/digitnc/inrlex.html>>
- [3] European Commission. 1999, Key Action III Multimedia content and tools III. 2 Interactive publishing, digital content and cultural heritage: Action Line III.2. 4 Digital preservation of cultural heritage, <<ftp://ftp.echo.lu/pub/digitalc/ev/fp5/al324.pdf>>
- [4] Industry Canada. 1997. Building the Information Society: Moving Canada into the 21st Century <http://strategis.ic.gc.ca/pics/lh/21st_e.pdf>