

디지털 콘텐츠 식별 시스템(DOI)의 구축

Adoption of Digital Object Identifier system in Korea

정상원, 오상훈. 한국데이터베이스진흥센터

Jeong, Sang-Won; Oh, Sang-Hoon. Korea Database Promotion Center

bstone@dpc.or.kr; ash@dpc.or.kr

인터넷을 통해 유통되는 디지털 콘텐츠의 식별과 접근을 위한 고유 식별체계로서 DOI(Digital Object Identifier)의 개발, DOI의 구문구조, DOI의 관리와 운영, DOI 메타데이터와 장르, DOI변환과 핸들시스템(Handle System)을 소개하고, DOI의 국내 도입을 위한 DOI 등록 시스템, DOI 변환 시스템, DOI 검색시스템, INDECS 메타데이터를 이용한 디지털콘텐츠 유통관리시스템, DOI 참조링크 시스템의 개념적 설계와 구축내용을 기술한다.

1. 서언

모든 정보의 유통은 식별을 전제로 한다. 현재, 인터넷 환경에서 디지털 정보자원의 위치를 찾아내고 이에 접근하기 위한 식별체제로 URL이 널리 활용되고 있다. 그런데 URL(Uniform Resource Locations)은 논리적인 내용물을 지정하는 것이 아니라 정보자원의 주소를 나타내기 때문에 정보자원의 주소가 바뀌거나 그 정보자원이 들어있는 도메인의 디렉토리가 변경되거나 또는 서버의 위치가 변경될 경우 식별과 접근이 불가능하게 된다. 이러한 문제점을 해결하기 위해 1994년 IETF(Internet Engineering Task Force)는 디지털 정보자원의 위치와는 상관없이 정보자원의 이름을 부여하는 URN(Uniform Resource Names) 체계를 제안하였다. URN은 정보자원의 수명이나 위치, 접근방법과는 무관하게 해당 정보자원에 이름을 부여하는 영구식별 체계이다. 현재, URN 체계를 가장 활발하게 활용하고 있는 사례는 DOI(Digital Object Identifier)이다. 이 글은 DOI의 개념과 구조 그리고 이의 국내 도입을 위해 추진중인 '디지털콘텐츠 식별 시스템'의 구축에 대해 다루고자 한다.

2. DOI; 디지털콘텐츠의 고유 식별체계

2.1 DOI 체계의 개발

DOI(Digital Object Identifier)는 URN체계의 한 사례로서 디지털 콘텐츠에 대해 고유한 식별기호를 부

- 1) 디지털 객체는 디지털 콘텐츠의 구성요소로서, 하나의 디지털 콘텐츠는 다양한 종류의 디지털 객체로 구성되어있다. 예컨대, 하나의 전자저널 기사에는 글자, 이미지, 동화상, 소리가 포함될 수도 있다. 이때 이 전자저널을 구성하는 각각을 디지털 객체(Digital Object)라 한다. DOI는 이러한 의미에서 디지털 콘텐츠뿐만 아니라 이를 구성하는 각 객체에 대해서도 부여할 수 있다.

여하고 이를 URL로 변환하여 인터넷상의 해당 정보의 위치에 접근할 수 있는 체계다. 1990년대 이후 인터넷의 보급이 확대되면서 다양한 콘텐츠가 인터넷을 통해 유통되기 시작하였다. 한편, 과학기술관련 학술잡지 출판사들은 전자저널형태로 유통시키기 시작하였으나 인터넷을 통해 유통되는 콘텐츠의 식별체제로서 ISBN이나 ISSN 등 기존의 식별체계가 한계가 있으며, URL또한 콘텐츠 자체가 아니라 콘텐츠의 위치를 식별하는데 이용되기 때문에 콘텐츠의 고유한 식별시스템으로는 적절치 않다는 것을 알게 되었다. 이에 1997년 미국출판협회는 디지털 형태로 유통되는 콘텐츠의 고유한 식별체계의 개발을 모색하던 중 URN 체계를 따르는 고유식별시스템으로서 DOI를 개발하게 되었다.

DOI는 하나의 통합 시스템으로서 다음의 4가지 요소로 구성되어있다. 열거(enumeration): 식별자의 할당, 기술(description): DOI로 식별된 콘텐츠의 메타데이터(INDECS 체계), 변환(resolution): DOI의 해당 콘텐츠로의 변환(Handle System), 정책(policy): 시스템의 운용을 관할하는 규칙

2.2 DOI의 구문구조

DOI 구문은 <Prefix>와 <Suffix>로 구성된다. Prefix는 <등록관리기관 번호>와 <등록기관번호>로 구성되는데 등록관리기관번호와 등록기관번호는 “.”로 구분한다. Suffix는 Prefix가 끝난 다음 사선(/)뒤에 오게되는데 등록기관이 자체적으로 부여한다. Suffix에는 통상 각 콘텐츠별 기존 식별코드를 선언하고 그 식별코드의 실제 값이 오게된다. DOI 구문구조를 실례로 들면 다음과 같다.

doi:10.1082/TEST-KOSIM_PROC/1999-133-136

이때 “10.1082”는 Prefix가 되고 “TEST-KOSIM_PROC/1999-133-136”은 Suffix가 된다. Prefix는 앞으로 전체적인 DOI 등록관리기관의 지정과 맞물려

'10.1082.10/1234'와 같이 두 번째 부분이 세분될 수도 있다.

2.3 DOI의 관리와 운영

DOI의 관리와 운영은 IDF(International DOI Foundation)가 담당하고 있다. IDF는 1998년 결성되어 DOI 체계의 개발과 정책을 수립하고, DOI 등록관리를 담당하고 있다. 회원은 창립회원과 일반회원으로 구분되며, 현재 콘텐츠 보유자, 출판사, 기술회사, 정보증개기관 등 36개 기관이 회원으로 가입되어 있다. 운영은 회원기관의 투표에 의해 선출되는 이사회와 IDF 전체를 총괄하는 총재가 있다.

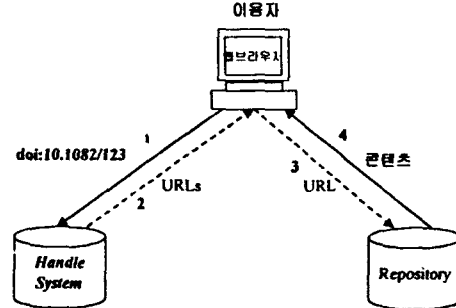
DOI 시스템은 현재 하나의 등록관리기관(Registration Agency)과 100여 개의 등록기관(Registrant)체제로 운영되고 있다. 등록관리기관은 DOI prefix의 할당, DOI의 등록, 등록기관이 메타데이터와 상태데이터를 유지하는데 필요한 기반을 제공하는 역할을 한다. DOI 등록기관은 콘텐츠를 보유하고 있는 기관으로서 DOI 등록관리기관(Prefix의 첫 번째)으로부터 DOI 등록기관 번호(Prefix의 두 번째 부분)를 할당받는다.

향후 DOI의 운영은 하나의 유지관리기관(IDF, Maintenance Agency)과 다수의 등록관리기관 그리고 각 등록관리기관에 DOI를 등록하는 다수의 등록기관의 모습을 띠게 된다. 등록관리기관은 DOI 장르와 지역별로 지정하게 되는데 등록관리기관으로 지정받기 위해서는 기본적으로 특정분야 DOI 장르의 개발과 함께 시스템 관리능력, 전문인력, 응용개발 아이디어, 유지관리를 위한 예산이 확보되어야 한다. 등록관리기관이 다수 지정되면 DOI의 운영과 관리는 장기적으로 등록관리기관의 연합체의 모습을 띠게 된다.

2.4 DOI의 변환과 핸들시스템

DOI는 기본적으로 역동적인 식별체계이다. 즉, DOI는 변환(Resolution)을 통해 사용자가 실제 콘텐츠에 접근할 수 있도록 해주는 시스템이다. DOI 구문의 변환은 CNRI(Corporation for National Research Initiatives)의 핸들시스템을 이용한다. 핸들시스템(Handle System)은 디지털 콘텐츠의 이름, 즉 핸들(Handle)을 저장하는 분산 컴퓨터 시스템으로서 DOI 코드에 해당되는 디지털 콘텐츠의 위치를 확인하여 접근하는데 필요한 정보로 재빨리 라우팅해주는 역할을 한다. 사용자가 DOI를 클릭하면 메시지가 핸들시스템에 전송되어 DOI와 관련된 URL을 이용자의 인터넷 브라우저에 전송하고 이용자는 콘텐츠 자체를 볼 수 있게 되거나 콘텐츠에 대한 정보나 접근방법에 대한 정보를 획득할 수 있게 된다. 디지털 콘텐츠가 다른 곳으로 이동하거나 저작권 소유권자에 변경사항이 생겼을 경우 변경내역이 핸들시스템의 등록관리시스템에

기록되므로 변경 후 접근하는 이용자는 새로운 사이트로 자동으로 안내되게 된다. 이때 브라우저가 DOI구문을 인식하기 위해서는 DOI plug-in이 설치되어야 한다. DOI의 변환과정을 도식화하면 아래와 같다.



현재 DOI의 변환은 1:1 변환체제이지만 앞으로 하나의 DOI에 대해 복수의 콘텐츠에 접근할 수 있는 1:n의 변환이 가능한 시스템으로 발전하고 있다.

핸들시스템은 하나의 Global Handle Registry(GHR)와 분산된 다수의 Local Handle Services(LHS)로 구성되어 있다. GHR은 Naming Authority를 등록하고 LHS는 각 핸들을 등록한다. 예를 들면 10.1082/123라는 DOI구문에서 10.1082는 Naming Authority가 되고 123은 LHS가 관할하는 핸들(handle)이 된다. DOI의 변환은 핸들시스템 내의 클라이언트가 10.1082이라는 핸들의 변환을 요청하면 어떤 핸들서비스에 요청하지를 알아보기 위해 GHR에 Naming Authority 10을 질의하게 되고 GHR은 Naming Authority 10에 대한 서비스 정보를 제공하게 된다. 클라이언트는 10.1082가 있는 서비스 사이트를 알아내서 해당 로컬핸들서비스 사이트로 핸들의 변환을 요청하게 된다.

2.5 DOI 메타데이터와 장르

DOI 문자열내의 요소는 식별된 객체에 대한 정보를 포함하고 있지 않다. 따라서 기술정보(description)가 없으면 이용자는 DOI가 무엇을 식별하는지를 명확히 알 수 없게 된다. 따라서 DOI가 보다 완전한 기능을 하기 위해서는 식별되는 콘텐츠에 대해 기술하는 메타데이터가 필요하다. 이를 위해 DOI 코드와 함께 콘텐츠의 명확한 식별을 위해 최소한의 메타데이터를 규정하고 있는데 이를 핵심 메타데이터(Kernel Metadata)라 하며, 그 요소는 다음과 같다.

- DOI: DOI
- DOI 장르: 공동의 속성들을 공유하는 콘텐츠의 유형
- 식별자(Identifier): 기존의 식별체계가 있는 경우 그 식별체계
- 표제(Title): 디지털 콘텐츠의 이름
- 유형(Type): 식별되는 디지털 콘텐츠의 유형(추상적

- 인 “저작물”, 디지털 혹은 물리적 형태의 “구현”, 실행)
- 원시성(Origination) : 저작물이 원본인지 아니면 다른 저작물의 파생물인지에 대한 정보
- 일차 역할자(Primary agent) : “일차적 역할자”의 신원, 통상 디지털 콘텐츠의 창작자
- 역할자의 기능(Agent role) : 일차적 역할자가 그 디지털콘텐츠의 창작에 행한 역할

그러나 서로 다른 유형의 콘텐츠를 기술하는데 필요한 메타데이터는 그 구조가 서로 다르기 때문에 IDF에서는 이들 서로 다른 유형의 콘텐츠를 장르로 규정하고 있다. DOI 장르(Genre)란 공동의 속성들을 공유하는 콘텐츠의 유형으로서 메타데이터 스키마로 정의된다. DOI로 식별되는 모든 콘텐츠는 적어도 하나 이상의 DOI 장르가 부여되며, 하나의 콘텐츠가 여러 개의 장르를 가질 수도 있다. DOI 장르는 공백장르와 기본장르 그리고 앞으로 개발될 미지의 장르가 있다. 공백장르(Zero Genre)는 메타데이터의 선언 없이 등록된 장르로서 DOI가 처음 개발되었던 때부터 지금까지 부여된 콘텐츠를 위한 장르다. 그러나 공백장르는 DOI와 관련된 메타데이터가 없기 때문에 매우 제한된 기능만을 갖게된다. 기본장르(Base Genre)는 핵심 메타데이터를 선언해야 하며 모든 등록관리기관이 기본 장르를 제공하게 된다. 그러나 기본장르의 운영에 필요한 규칙은 아직 정해지지 않았다. 한편, IDF에서는 새로운 DOI 장르를 등록관리기관의 지정과 연계시키고 있으며, 새로운 DOI 장르의 개발을 위해 관심집단이 장르 정의그룹(Genre Definition Group)을 결성하도록 권고하고 있다. 또한 새로운 장르 스키마의 정의와 값들은 DOI가 식별하는 콘텐츠의 확실한 구분과 콘텐츠 유통의 상호운용성을 보장하기 위해 <Index> 메타데이터 체계에서 규정하는 데이터 사건을 반드시 차용해야 한다.

3. 디지털콘텐츠 식별시스템의 설계

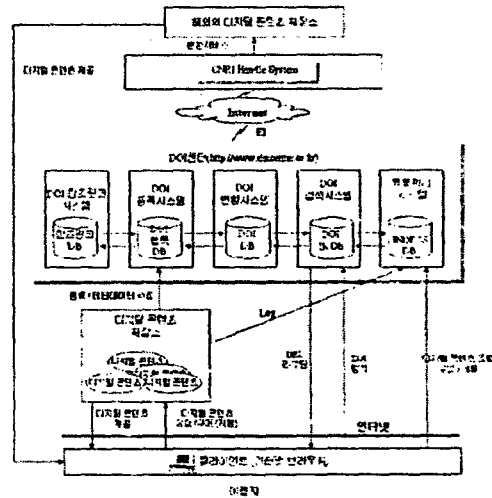
DOI는 인터넷에서 유통되는 콘텐츠의 정확한 식별과 접근을 위한 인프라 기능을 할 수 있다. 이러한 관점에서 국내에 DOI를 도입하는데 의의가 있다. 한국데이터베이스진흥센터에서는 국내에 DOI 도입을 위해 「디지털 콘텐츠 식별시스템 구축사업」을 수행하고 있다. 이 글에서는 현재 이 사업에서 추진하고 있는 DOI 등록, 변환, 검색, 유통관리, 참조링크 시스템의 설계내용과 기능 그리고 각각의 역할에 대해 간략히 소개하고자 한다.

3.1 DOI 등록 시스템

DOI 등록시스템에서는 디지털콘텐츠를 보유한 등록자가 DOI 코드를 부여받기 위한 시스템으로 등록기관

에 DOI Prefix를 할당하고 등록기관이 할당된 Prefix로 Suffix를 부여하고 관리하는 시스템이다. DOI를 등록할 때는 앞서 기술한 핵심장르의 메타데이터와 앞으로 개발될 DOI 장르의 메타데이터를 함께 등록하게 된다. 또한 Dublin Core, MARC 등으로 가공된 콘텐츠의 경우 이들 메타데이터가 핵심장르나 DOI 장르로 변환될 수 있는 환경을 제공하게 된다. 이 시스템은 크게 DOI 등록신청 시스템, DOI 등록처리 시스템, DOI 등록관리 시스템, 메타데이터 생성시스템으로 구성된다.

한편, DOI의 원활한 등록을 위해서는 Prefix의 할당 규칙과 DOI Suffix 부여규칙이 필요하기 때문에 현재 이에 대한 표준과 지침을 개발하고 있다.



3.2 DOI 변환 시스템

변환시스템은 디지털콘텐츠에 부여된 DOI를 해당하는 URL로 변환해준다. 변환시스템은 URN을 받아서 URN에 의해 식별된 정보의 목록이나 하나 이상의 URL을 되돌려준다. 여기서 변환이란 네트워크서비스에 식별자를 넘겨주고 그 식별자와 관계된 현재의 정보를 하나이상 되돌려 받는 것을 의미한다. 예컨대, DNS의 경우 변환은 도메인명에서 단일 IP 주소로 변환한다. Handle System의 경우 변환은 하나의 핸들에서 하나 이상의 데이터로 변환해준다.

3.3 DOI 검색 시스템

DOI 검색시스템은 DOI로 디지털 콘텐츠에 접근할 수 있는 검색환경을 제공해주는 시스템이다. 디지털 콘텐츠 주제, 형태, 유형별로 탐색할 수 있는 일종의 탐색엔진의 기능을 수행한다. 이용자는 해당 디지털 콘텐츠의 메타데이터 DB에서 추출된 색인DB를 검색엔진을 통해 탐색하여 DOI로 인코딩된 디지털 콘텐츠를 검색하고 이후 DOI 변환시스템을 통해 실제 디지털

콘텐츠에 접근하게 된다. 이 시스템은 DOI가 부여된 디지털 콘텐츠의 메타데이터를 색인하여 검색시스템을 구축함으로써 디지털콘텐츠 정보서비스의 신뢰성과 효율성을 극대화시킬 수 있을 것으로 기대된다.

3.4 디지털콘텐츠 유통관리 시스템

DOI를 활용한 유통관리 시스템은 디지털콘텐츠를 판매하는 일종의 쇼핑몰에서 전자거래용 메타데이터인 INDECS를 이용하여 콘텐츠의 거래내역 등을 관리하는 역할을 한다.

INDECS(Interoperability of Data in E-Commerce System)는 디지털콘텐츠의 전자거래 시스템에서 이용되는 다양한 메타데이터의 상호운용성을 위해 개발된 구조다. 물리적인 재화와는 달리 하나의 디지털 저작물은 수백 수천 건의 독립된 콘텐츠로 구성될 수도 있으며, 또 다른 콘텐츠의 제작, 이용, 거래와 매우 밀접하게 관계를 맺게된다. 그러나 CIDOC, Dublin Core, MPEG21 등 디지털 콘텐츠를 기술하는 메타데이터 스키마는 그 출발점이 다르기 때문에 상호호환이 되지 못한다. 특히 디지털 콘텐츠의 유통에서 발생된 관계는 매우 복잡하기 때문에 콘텐츠의 제작, 이를 만들고 이용하는 집단, 이러한 이용을 지원하는 거래내역과 관련된 모든 종류의 메타데이터가 서로 호환되지 않고서는 다른 시스템과 자동적으로 서로 호환되지 못한다. INDECS 체계는 이러한 메타데이터 표준간의 충돌을 막고 이들 표준들이 의미적으로 상호운용이 가능한 기반구조를 제공하는데 그 목적이 있다. 물론 메타데이터간의 상호호환을 위해 XML과 RDF가 활용되고 있으나 의미적 본질과 연관된 근본적인 문제를 다루지는 못한다.

유통관리시스템은 DOI등록시스템을 통해 등록된 디지털 콘텐츠를 DOI 메타데이터와 함께 INDECS 메타데이터로 관리된다. INDECS는 객체를 저작(work), 실행(performance), 실현(manifestation) 형태로 세분하여 처리가능하고 특정거래를 사건 중심으로 기술할 수 있기 때문에 콘텐츠의 저작권 처리, 계약, 판매 등 유통시 발생할 수 있는 모든 형태의 처리과정을 기술할 수 있다. INDECS 메타데이터를 조직적으로 관리하기 위해 INDECS 서버를 구축한다. 디지털 콘텐츠의 저작자 및 저작권 소유자, 콘텐츠 제공기관간의 관계변화, 이용자의 콘텐츠 구매내역 등에 대한 이벤트를 유지한다. 모니터링 시스템은 디지털 콘텐츠의 이용자와 디지털 콘텐츠 제공기관간에 발생한 이벤트를 추적하여 해당 디지털 콘텐츠의 유통에 참여한 기관(자)에 통보한다. DOI를 활용한 유통관리시스템을 쇼핑몰에 적용함으로써 디지털콘텐츠 거래의 투명성과 신뢰성을 높일 수 있을 것이다.

3.5 DOI 참조링크 시스템

DOI를 이용한 참조링크 시스템은 학술 커뮤니케이션에서 발생하는 인용사항을 DOI로 식별하여 접근할 수 있게 하는 시스템이다. 예컨대, 특정분야의 이용자가 온라인화된 학술저널 기사, 석·박사학위논문, 학술회의 자료집, 전자저널의 참고문헌을 클릭했을 때 그 해당 문헌의 초록, 전문, 또는 서지 레코드에 곧바로 링크될 수 있도록 하는 시스템이다. 참조링크 서비스의 형태는 초록만 제공하는 형태, 전문(fulltext)을 보여주는 형태, 서지레코드를 보여주는 형태가 있다. 이 시스템은 학술문헌을 브라우징하거나 읽기 쉽게 하고 논리적으로 연관된 콘텐츠에 접근할 수 있도록 해주는 기능을 한다. 참조링크 시스템이 구축될 경우 SCI 서비스와 유사한 인용색인 DB 구축을 포함한 다양한 응용서비스가 가능해질 것이다. 이러한 참조링크 서비스가 잘 정착되기 위해서는 국가전자도서관 프로젝트 참가 기관 및 일반 색인초록작성기관의 적극적인 참여가 필요하다. 한편, DOI를 이용하여 참조 링크시스템을 구축한 대표적인 사례는 Crossref(www.crossref.org)과 Link Openly(<http://www.openly.com>)에서 찾아볼 수 있다.

4. 결론

DOI는 디지털 콘텐츠의 고유한 식별체계로서 역할을 하며 DOI를 활용한 다양한 서비스가 구축된다면 디지털 콘텐츠에 대한 접근성 향상과 이용의 편의성을 도모할 수 있을 것이다. 또한 DOI와 연계된 INDECS 메타데이터는 디지털콘텐츠의 거래시 발생하는 메타데이터간의 상호운용성을 확보해줌으로써 디지털 콘텐츠의 제작자와 이용자를 보호해줄 수 있음은 물론 디지털 콘텐츠의 효과적이고 투명한 상거래 환경을 조성할 수 있을 것이다.

참고문헌

- [1] IDF. 2000. *The DOI Handbook*. version 0.3 July 2000. <doi:10.1000/182>
- [2] IDF. 2000. *About the DOI*. <<http://dx.10.1000/7>>
- [2] Sun, Sam X; Reilly, Sean; Lannom, Larry. 1999. *Handle System Namespace and service definition*. <<http://www.handle.net/draft-sun-handle-system-def-01.html>>
- [3] R. Moats. 1997. *URN syntax RFC 2141*, May 1997. <<http://www.ietf.org/rfc/rfc2141.txt>>
- [4] ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier
- [5] 한국데이터베이스진흥센터. 2000. 디지털콘텐츠 식별시스템구축사업. <<http://www.doicenter.or.kr>>