

# DOI 구문 식별 코드 개발

## The Development of Prefix/Suffix Code of DOI Syntax

°김세정, 안계성, 한국데이터베이스진흥센터  
Kim, Sei-Jung, Ahn, Gye-Sung, Korea Database Promotion Center

본 연구는 인터넷상에서 제공되는 지적콘텐츠에 대한 영속적이고 고유한 식별 체계인 DOI 상용 서비스의 일환으로 DOI 구문을 구성하고 있는 Prefix와 Suffix 식별코드를 개발하였다. 이를 위해서 현재 인터넷상에서 지적콘텐츠를 유통시키고 있는 기관들을 조사 분류하여 콘텐츠 보유기관 식별코드를 개발하였다. 또한 지적 콘텐츠의 속성 및 유형 분석을 토대로 Suffix 코드의 구조 및 식별코드를 개발하였으며 관련 콘텐츠간의 연계를 위한 저작권 식별기호를 고려하였다.

### 1. 연구의 개요

DOI(Digital Object Identifier)란 인터넷상에서 제공되는 지적콘텐츠에 대한 고유 식별기호를 의미하며, DOI 서비스는 핸들시스템을 사용하여 특정 콘텐츠의 DOI를 URL로 변환해주고 그 콘텐츠와 관련된 메타데이터를 함께 제공함으로써 콘텐츠의 검색 및 지적재산권 보호의 기능을 제공한다.

본 연구는 DOI 서비스 상용화를 위한 선행 조건으로써 DOI 구문을 구성하고 있는 Prefix와 Suffix 코드 표준 개발을 목적으로 한다.

DOI Prefix 코드 표준을 통해 디지털 콘텐츠 보유기관의 DOI 등록 및 관리에 있어서 일관성 및 체계성을 확보할 수 있으며, 콘텐츠 산업에 관한 통계조사 및 기타 현황파악을 위한 자료로 사용될 수 있다.

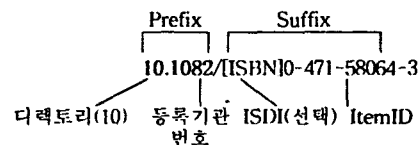
현재 DOI Suffix 코드를 부여하는데 있어 특정한 기준이 적용되고 있지 않은 상태이며 기존 국제표준을 이용하거나 등록기관내에서 자체적으로 개발한 의미없는 식별자(dumb number)를 사용하고 있는 실정이다. Suffix 코드 부여는 등록기관이 임의적으로 부여할 수 있지만 상이한 형식의 Suffix 코드의 사용은

상당한 콘텐츠 등록 및 관리에 있어서 상당한 혼란을 야기할 수 있으므로, 관리의 효율성을 위해서 다양한 유형의 콘텐츠에 적용할 수 있는 Suffix 코드의 개발이 요구된다.

본 연구에서 제시하는 Suffix 코드의 표준은 등록기관 및 이용자를 위한 최소한의 상호운용성을 제공할 수 있으며, 인터넷상에서 유통되는 콘텐츠 유통 및 거래의 활성화를 촉진할 수 있다.

### 2. DOI 구문구조

DOI 시스템은 DOI 식별기호, 디렉토리, 콘텐츠와 관련된 메타데이터를 저장한 데이터베이스의 세 가지 요소로 구성되어 있으며, DOI의 식별기호(Identifier)는 DOI 레코드에서 재사용되지 않는 고유값을 가지는 Prefix와 Suffix 두 가지 요소로 구성되어 있다.



Prefix와 Suffix로 구성된 DOI 식별기호 구분형식 중에서, 앞 부분의 Prefix는 DOI 레코드를 관리하고 Prefix의 할당을 담당하고 있는 디렉토리 관리기관(Directory)과 콘텐츠 보유기관에 대한 등록기관 코드(RegistrantPrefix)로 구성된다. 콘텐츠를 보유하고 있는 기관은 그들의 지적콘텐츠에 대한 한 개 이상의 Prefix 코드를 할당받을 수 있다.

DOI Suffix는 Prefix 다음부분의 '/' 뒤에 위치하며 특정 콘텐츠를 소유하고 있는 DOI 등록기관이 자체적으로 부여한다. 대개의 경우 ISBN, ISSN, ISAN 등과 같은 국제 표준 중에서 선택하여 사용하고 있으며 미국국립표준협회(NISO)에서는 ISDI (International Standard Digital Identifier)를 사용할 것을 권고하고 있다.

### 3. DOI Prefix 코드 표준

현재 Prefix는 디렉토리 관리자를 나타내는 디렉토리기관 코드에 해당하는 값으로 오직 '10'하나만이 존재하고 있다. IDF가 유일하게 전세계의 모든 DOI를 관리하는 디렉토리 관리기관(인)으로 등록되어 DOI를 등록하기 원하는 세계 각국의 콘텐츠 보유기관에 대해 일괄적으로 등록순에 따라 순차적인 번호를 부여하고 있다. 하지만 가까운 시일내에 각국 또는 장르별 대표 디렉토리 관리기관이 지정될 예정이며, 등록기관 코드 부여의 권한 및 콘텐츠의 관리는 해당 디렉토리 관리기관으로 이양될 것이다.

따라서 본 연구에서는 한국데이터베이스진흥센터를 디렉토리 관리기관으로 가정하고 '20'을 디렉토리 번호로 사용하여 콘텐츠 보유기관에 Prefix 코드(등록번호)를 개발하였다.

DOI Prefix코드 표준을 개발하기 위해서는 현재 인터넷상에서 제공되고 있는 콘텐츠 산업에 대한 특성 및 동향 파악이 선행되어야 할 것이다. 따라서 본 연구는 이러한 조사를 토대

로 이미 제시되어 있는 국내외의 기관 분류표준을 참고하여 콘텐츠 제작 및 유통이 가능한 각 기관별 항목을 추출하였다.

#### (1) 등록기관 분류 원칙

첫째, 인터넷상에서 제공되는 모든 유형의 콘텐츠를 제공하는 영리 혹은 비영리 기관들을 대상으로 한다.

둘째, 정보통신산업 및 전자기술, 생명공학, 신소재 등 첨단기술 분야를 포함할 수 있도록 새로운 산업분야에 대한 분류항목을 제공한다.

셋째, 기본적으로 이미 국가규격으로 제시된 기관코드 및 교육기관코드, 한국표준산업분류(2000. 1. 7 고시)의 항목을 따르되 소분류 항목에 대해서는 콘텐츠 산업의 특성을 반영할 수 있도록 기관을 세분화 시켰다.

넷째, 전문적인 콘텐츠 제작 업체의 활성화를 고려하여 데이터베이스 및 온라인 정보 제작 및 제공업과 소프트웨어 자문, 개발, 공급업을 대분류 항목으로 추가하였다.

#### (2) 등록기관 분류기준 및 코드

디지털 콘텐츠 보유기관을 대분류, 중분류, 소분류로 구분하고, 코드길이는 대분류는 알파벳 1문자, 이하 분류 단위에서는 아라비아 숫자 2자리씩 부여하였다.

먼저 '국가정부기관', '교육기관', '연구기관', '공공기관', '기업 및 상용기관', '개인'으로 대분류하였으며, 다음과 같이 중분류, 소분류 하고 그에 대한 코드를 개발하였다.

A	국가정부기관	
	A10	행정기관
		A1001 대통령직속기관
		A1002 국무총리직속기관
		A1003 중앙 정부기관(부,청)
		A1004 광역자치기관
		A1005 기초자치기관
		A1099 기타행정기관
	A20	입법기관

### 4. DOI Suffix 코드 표준 연구

다양한 유형의 콘텐츠를 식별하기 위한 Suffix 코드를 개발하기 위해서 실제로 인터넷 상에서 제공되고 있는 지적콘텐츠의 속성을 분석하고 이를 각 유형별로 분류하였다.

또한 DOI Suffix는 단행본, 학술잡지 기사, 초록, 텍스트 파일, 멀티미디어 콘텐츠, 비디오, 이미지, 소프트웨어 프로그램 등 다양한 유형의 콘텐츠와 그것의 구성요소에 개별적으로 부여할 수 있는 Suffix 코드구조를 개발하였다.

DOI 서비스 대상콘텐츠는 단지 디지털형태의 콘텐츠뿐만 아니라 전자상거래를 목적으로 하는 물리적인 형태의 지적 콘텐츠를 포함하므로 Suffix 코드는 이 모든 형태의 지적콘텐츠에 적용 가능해야 한다. 따라서 DOI 식별체계란 디지털 콘텐츠 식별기호가 아니라 디지털 형식의 지적콘텐츠 식별기호가 되어야 한다.

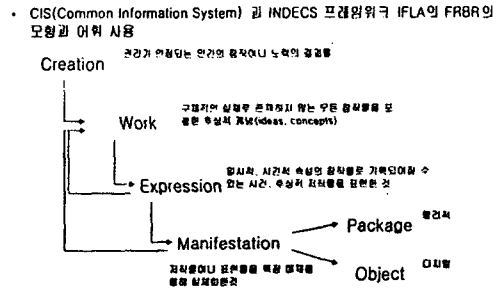
(1) 지적콘텐츠의 형태 및 속성 정의

인터넷상에서 제공되는 모든 지적콘텐츠의 형태를 크게 다음과 같이 4가지 용어로 정의할 수 있다. 현재 DOI에서는 모든 형태의 지적 콘텐츠를 규정하기 위한 공통의 용어로 CIS(Common Information System)에서 개발된 지적콘텐츠에 대한 용어정의를 사용하고 있으며, DOI에서 채택하고 있는 INDECS 메타데이터에서도 이와 유사한 IFLA의 FRBR (Functional Requirement for Bibliographic Record) 모형을 사용하여 지적콘텐츠를 정의하고 있다.

지적콘텐츠란 이 모형에서 정의한 “창작물(Creations)”을 의미한다. “창작물(Creation)”이란 “권리가 인정되는 인간의 창작이나 노력의 결과물”로 정의되며, 구체적인 실체로 존재하지 않는 모든 창작물을 포괄한 추상적 개념(ideas, concepts)을 “저작물(Work)”이라고 한다. 저작물(Work)을 특정매체에 상관없이 기록되어지는 창작물 자체를 표현물(Expression)이라 하며 이것을 디지털 이나 물리적인 매체로 실체

화 한 것을 실체물(Manifestation)이라 한다.

지적콘텐츠 형태 및 속성



(2) 식별기호를 통한 지적콘텐츠의 연계

위에서 정의된 바와 같이 하나의 지적콘텐츠는 여러 가지 형태의 표현물이나 실체물로 존재하고 있다. 각 형태의 지적콘텐츠 대한 개별적인 식별기호가 존재하며, 또는 모든 형태의 지적콘텐츠에 적용될 수 있는 식별기호가 있다. DOI는 서적, 음반물, 전자파일 등과 같은 실체물(Manifestation) 형태의 지적콘텐츠 뿐만 아니라 저작물(Work)이나 표현물(Expression)에 모두 적용가능하다.

이와 같이 동일 지적콘텐츠에 관한 저작물과 표현물 또는 여러 가지 유형의 실체물들이 존재함으로써 이것들간의 연계 필요성이 커지고 있다. DOI에서는 이러한 관계 정보를 식별기호를 통해 연계할 수 있는데 각각의 물리적 형태의 콘텐츠(학술잡지)와 디지털 형태의 저작물(온라인 학술잡지)에 DOI식별기호를 부여하고 각 콘텐츠의 메타데이터에서 관련된 패키지와의 관계를 기술하고 이를 변환해준다.

하지만 이것은 패키지 형태의 콘텐츠와 객체 형태의 콘텐츠간의 1:1의 동등관계밖에 표현하지 못한다. 중요한 것은 객체 형태의 표현물과 패키지 형태의 표현물간의 관계를 나타내 주는 것이 아니라 이 표현물들과 이것의 저작물(Work)에 대한 관계를 표현하는 것이다.

DOI는 동일 저작물에 대한 여러 가지 표현

물들간의 관계를 표현하기 위한 방안으로 DOI Suffix 코드내에서 “저작물 식별기호”(예: 10.1082/(work)abcde..)를 포함하는 것이다.

지적콘텐츠간의 연계를 위한 DOI Suffix 코드를 개발함에 있어서 가장 중요한 관건은 인용문헌간의 연계를 위한 저작물 식별기호의 필요성과 전자상거래를 위한 객체 식별기호의 필요성간의 충돌문제이다. 이러한 문제점을 해결하기 위해서 모든 유형의 창작물(Creation)을 수용하고 이들간의 관계를 식별할 수 있는 기법과 어휘들의 구축이 필수적으로 요구된다.

(3) DOI Suffix 코드 표준

본 연구에서는 다음과 같은 지적콘텐츠 및 구성요소에 대한 식별체계 표준초안을 제시하며, 이를 DOI Suffix코드로 사용할 것을 권고한다. 간단한 기본적인 정보를 제공하고 있는 의미없는 식별자 체계의 DOI Suffix 코드체계를 개발 원칙으로 하였다. Suffix코드는 다음과 같이 콘텐츠 유형코드, 콘텐츠 식별번호, 콘텐츠 구성요소 식별코드, 형식코드의 4가지 데이터 항목으로 구성된다.

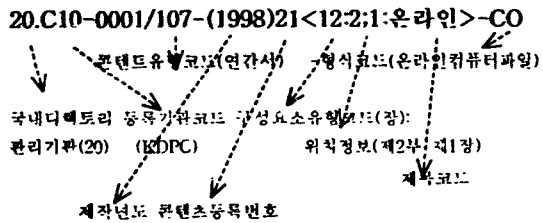
① 콘텐츠 유형코드	② 콘텐츠 식별번호	③ 구성요소 식별코드	④ 형식 코드
콘텐츠 유형코드	(제작년도)콘텐츠 등록순 일련번호	<구성요소 유형코드: 위치정보:제목코드>	형식 코드

DOI 서비스에서 제공되는 지적콘텐츠의 유형을 크게 ‘서적 및 일반 간행물’, ‘음악 및 음향물’, ‘영상물’, ‘소프트웨어 프로그램물’로 대분류하고 각각의 소분류 항목에 대한 콘텐츠의 유형코드를 개발하였다. 콘텐츠 식별번호는 콘텐츠의 제작년도와 콘텐츠 등록순 일련번호로 구성되며, 제작년도가 제시되지 않을 경우라도 공괄호로 표기한다. 구성요소 식별기호는 단계성(Granularity)과 관련된 사항으로 이것 역시 콘텐츠 보유기관이 임의적으로 결정할 수 있다. 즉 특정콘텐츠를 구성하고 있는 세부적인 콘텐츠, 예를 들어 도표, 초록, 목차, 이미지 등의 콘텐츠에 개별적인 식별기호를 부여하여

유통하고자 할 경우 ‘각괄호(< >)’안에 구성요소에 대한 데이터를 제공한다. 만약, 콘텐츠 구성요소에 대해 DOI 식별기호를 부여하지 않을 경우 공괄호로 둔다. 형식코드는 DOI 서비스에서 제공되는 지적콘텐츠의 형식을 식별하기 위한 것으로 크게 패키지(Package) 형태의 콘텐츠와 객체(Objects) 형태의 콘텐츠로 구분하고 ANSI/NISO Z39.56(1996)에서 제공되고 있는 서브셋 코드를 사용하였다.

(4) DOI Prefix/Suffix 할당 예

한국데이터베이스진흥센터, 1998. “데이터베이스백서 : 제 2부 데이터베이스산업의 현황 ; 제1장 온라인데이터베이스 부문별 현황”, <http://www.dpc.or.kr/whitepaper/wp98/21.html> => DOI 부여 예



5. 결론

DOI 서비스의 상용화를 위한 DOI식별코드의 개발은 궁극적으로 우리나라의 콘텐츠 산업의 경쟁력을 키울 수 있는 기반구축을 의미한다.

콘텐츠 Suffix 코드를 개발함에 있어서 저작물 식별기호의 도입은 하나의 저작물에 관련된 여러 가지 유형의 표현물들을 연계시키고 인용문헌 연계를 위해서 반드시 요구되며, DOI 성공의 중요한 요소로 작용할 것이다.

추후 실질적인 DOI 상용서비스의 운영을 통해 Prefix/Suffix코드에 대한 확장 및 수정작업이 요구된다.

참고문헌

ANSI/NISO Z39.84-XXXX, "Syntax for the Digital Object Identifier", 1999.