

DOI 메타데이터의 역할 및 개발 현황

The Role and the Current Development Status of DOI Metadata

°조소연, 조현주
한국데이터베이스진흥센터

Cho So-Youn, Cho Hyun-Joo
Korea Database Promotion Center

국내외적으로 DOI에 대한 논의와 이의 적용 노력이 점차 확산되고 있다. DOI 메타데이터는 DOI 이용의 효용을 높여주고 DOI와 관련한 각종 응용 서비스 개발을 가속화시킬 수 있는 중요한 수단이다. DOI 메타데이터의 필요성 및 역할을 고찰하고 이와 관련한 개발 원칙 및 개발 현황을 살펴봄으로써 DOI 메타데이터와 관련한 향후 과제 및 역할을 모색한다.

1. 서론

DOI(Digital Object Identifier)는 디지털 콘텐츠에 대한 영구 식별자이다. 1994년 미국출판협회의 이사회는 인터넷 상에서의 저작권 보호 문제를 검토하기 위해 특별위원회를 설치했고, 특별위원회는 저작물의 불법 이용 방지에 필요한 기술을 검토하기 시작하였다. 그 결과 각종 저작권 보호 기술을 유효하게 작동시키기 위해서는 콘텐츠를 식별하는 표준 방안이 필요하다는 결론에 도달했고, 이에 DOI의 개발이 시작되었다(內田尚子 1999).

DOI는 URN의 한 응용모델로서, 모든 식별 가능한 디지털 콘텐츠에 부여되어 영구적이고 고유 식별기능을 수행한다. 한편 DOI는 URN 변환기와 같은 역할을 하는 DOI 핸들 시스템에 의해 URL로의 변환이 가능한데, 이로써 이용자는 DOI를 가지고 콘텐츠의 현 해당 위치

(URL)에 접근할 수 있게 된다(Norman Paskin 1999b).

이처럼 DOI는 기본적으로 콘텐츠에 대한 영구적인 식별자를 부여하고, 이를 통해 콘텐츠의 현 해당 위치를 찾아주는 역할을 수행한다. 그러나 향후 DOI의 이용을 전세계적으로 확산시키기 위해서는 이러한 기본적인 기능 외에 각종 응용 서비스를 개발시킬 필요가 있다. DOI의 기본적인 기능이 유용한 것이기는 하지만 DOI를 활성화시키기 위해서는 사업자 및 이용자들을 매료시킬 만한 확실한 유인책이 필요하기 때문이다.

DOI 서비스의 다양화를 가능하게 하는 요소 중 가장 중요한 수단이 DOI 메타데이터이다. DOI 메타데이터는 DOI의 활용 가치를 높여줄 뿐 아니라, 각종 응용서비스를 가능하게 하는 필수 요소로서, IDF(The International DOI Foundation)에서도 DOI 메타데이터의 개발에

많은 관심을 기울이고 있다. 그러나 DOI 메타데이터의 개발이 그다지 용이한 것만은 아니다. DOI를 부여할 수 있는 콘텐츠의 종류 및 범위가 광범위할 뿐만 아니라, 시스템간의 호환성을 확보하기 위해서는 다양한 메타데이터 간의 상호운용성을 고려하지 않을 수 없기 때문이다. 현재 IDF는 INDECS를 기반으로 DOI 메타데이터 개발에 힘쓰고 있다. 이미 모든 DOI에 적용할 수 있는 기본이 되는 최소한의 메타데이터인 Kernel 메타데이터를 개발하여 발표한 상태이며, 특정한 콘텐츠(DOI 장르)에 구체적으로 적용할 수 있는 확대(extension)된 메타데이터 개발을 서두르고 있다.

국내에서도 DOI를 도입하기 위하여 기술적, 정책적 노력을 기울이고 있는 이 시점에서, DOI 메타데이터에 대한 정확한 이해와 이를 바탕으로 한 적절한 적용은 시행착오를 줄이고 비용 효율화를 이룰 수 있도록 하는 기본 방안이 된다.

2. DOI 메타데이터의 필요성 및 역할

DOI는 콘텐츠의 유일무이성을 확인해주는 영구적이고 고유한 식별자이기는 하지만 그 콘텐츠에 대한 어떠한 정보도 제공하지 않는다. 즉 DOI 그 자체는 '의미없는 식별자(dumb number)'로서, DOI가 부여된 해당 콘텐츠의 제목이 무엇인지, 콘텐츠의 타입은 무엇인지(책, 전자책, MP3, 공연 등), 저자는 누구인지 등에 대한 관련 정보를 지니고 있지 않다.

그러나 DOI의 활용 가치를 높이고 유용한 서비스를 제공하기 위해서는 DOI가 부여된 콘텐츠에 대한 관련 정보를 제공해 줄 수 있는 일정한 수단이 반드시 필요하다. IDF에서는 메타데이터가 이러한 역할을 수행할 수 있다고 결론내렸고, 메타데이터를 DOI 서비스의 다양

화를 촉진시킬 필수적인 요소로 보았다.

DOI 메타데이터는 DOI에 대한 효율적인 검색을 가능하게 한다. 또한 메타데이터는 DOI를 다중 변환(multiple resolution : 하나의 DOI를 URL 등 다양한 위치 및 서비스에 매핑시키는 것)하는 과정에서 의미있는 선택 작업(원하는 위치 및 서비스로 매핑)이 이루어질 수 있도록 매개변수를 제공한다. 뿐만 아니라 현재 DOI의 응용 서비스 중 가장 주목받고 있는 참조링크 서비스를 가능하게 하기 위해서도 메타데이터의 활용은 중요하다(Norman Paskin 1999b).

이처럼 메타데이터는 i) DOI의 폭넓은 이용을 가능하게 할 뿐만 아니라 ii) DOI와 관련한 각종 응용 서비스 개발을 가속화시킨다(IDF 1999).

한편 DOI의 등록요건에 메타데이터를 추가함으로써 메타데이터의 수집을 효율적으로 수행할 수 있다. 메타데이터의 무결성을 확보하고 효과적인 활용을 위하여 등록기관은 메타데이터의 형식 및 내용에 대한 품질 관리에 힘써야 한다.

3. INDECS 데이터 모델 채택

DOI 메타데이터가 유용하게 활용되기 위해서는 DOI 메타데이터가 기존의 또는 개발되고 있는 다양한 메타데이터 스키마 간의 상호운용성을 보장하고 있어야 한다. IDF는 이러한 상호운용성을 보장해 주는 DOI 메타데이터를 "잘 구성된(well-formed)" 메타데이터라고 한다. 메타데이터가 "잘 구성"되기 위해서는 데이터 사전에 의해 지원되는 통제된 용어를 사용할 필요가 있다. 즉 용어가 통제된다는 것은 상이한 메타데이터 스키마 간에 값을 서로 매핑시킬 수 있다는 것을 의미한다.

DOI는 "잘 구성"된 메타데이터를 개발하기

위하여 INDECS(Interoperability of Data in E-Commerce Systems) 메타데이터 모델을 채택하고 있다. INDECS 프로젝트는 디지털 콘텐츠의 전자상거래를 위한 기반 마련과 지적재산권 관리를 위한 메타데이터가 필요하다는 인식하에 1998년 유럽연합의 지적재산권 관련 기관들 주도로 시작된 프로젝트이다. INDECS는 다양한 메타데이터 스키마를 조화시킬 수 있는 공통된 프레임워크로서, 서로 다른 메타데이터 스키마간의 상호운용성 확보를 목표로 한다. INDECS는 DOI와 마찬가지로 콘텐츠의 모든 타입을 수용할 수 있도록 고안되었고, 개방형 모델로서, 다양한 콘텐츠의 구체적인 요구 사항을 충족시킬 수 있도록 확장성을 부여하고 있다(IDF 2000).

INDECS 메타데이터는 i) 유일한 식별성(unique identification : 모든 엔터티는 유일하게 식별될 필요가 있음)과 ii) 기능적인 세분성(functional granularity : 구별할 필요가 있는 모든 엔터티는 식별될 필요가 있음)을 기본 원칙으로 하고 있다(Norman Paskin 1999b). INDECS의 기본 원칙 중 '유일한 식별성' 원칙은 DOI 할당의 기본적인 원칙과 일치하고, '기능적인 세분성' 원칙은 등록자가 선택하는 어떠한 엔터티(예를들면, 논문 중에 수록되어 있는 표나 그림 등)에든지 DOI를 할당할 수 있다는 DOI의 가이드라인과 일치한다.

한편 INDECS 프로젝트는 콘텐츠의 기술적인 내용을 표현하는 '고정된(static) 데이터'보다는 권리관리를 위한 '동적인(dynamic) 데이터'의 표현에 중점을 두고 있다. 반면 DOI 메타데이터에 대한 IDF의 관심은 콘텐츠의 식별성을 제공해 줄 수 있는 최소한의 기술 메타데이터를 작성하는 것이다(IDF 2000).

DOI 메타데이터는 필수요소 선정을 위해 INDECS 데이터모델을 활용하고 있고, 데이터

표현을 위한 데이터모델로는 응용범위가 넓고 구조도 합리적이라는 이유로 RDF를 사용하고 있다. 교환 형식은 XML을 사용한다(Norman Paskin, 1999a).

4. DOI 커널(Kernel) 메타데이터

DOI 메타데이터는 다양한 미디어 환경을 지원할 수 있도록 상호운용성을 보장하고 확장성을 가져야 한다. 이러한 문제를 해결하기 위하여 최소한의 공통된 메타데이터 요소를 개발하고 통일된 용어를 고려하는 것이 무엇보다도 중요하다. 물론 최소한의 요소라고는 하지만 적어도 해당 콘텐츠를 식별하기에 충분한 정도, 즉 해당 콘텐츠를 정확히 검색할 수 있는 정도는 만족하여야 한다.

이에 IDF는 INDECS 데이터모델을 기반으로 하여 DOI 메타데이터 개발을 진행시켰고, 1999년 상반기에 상호운용성을 보장하고 어떠한 DOI에도 응용할 수 있는 최소한의 데이터 요소로 구성된 메타데이터를 개발하여 발표하였다. IDF는 이러한 최소한의 메타데이터를 DOI 커널(Kernel) 메타데이터로 명명하였다.

DOI 커널 메타데이터는 <표1>과 같이 8개의 최소한의 요소로 구성되어 있다(IDF 2000).

DOI 커널 메타데이터는 '권리'를 표현하기 위한 데이터 요소는 가지고 있지 않으며, 데이터 값이 변하지 않는 고정된(static) 값으로만 구성되어 있다. DOI는 항상 하나 이상의 DOI 장르를 반드시 가져야 한다.

DOI 커널 메타데이터는 모든 DOI에 적용되는 필수 사항임과 동시에 확장성을 제공한다. 즉 개별 DOI 장르에 대한 메타데이터는 DOI 커널 메타데이터의 요소는 반드시 포함하고 있어야 하지만, 각 장르의 특성을 고려하여 DOI 커널 메타데이터에 좀 더 구체적인 요소를 '추

구성요소 (상위 레벨)	정의	허용되는 값	지위	수
DOI	DOI	DOI	확정적	1
DOI 장르	IDF에 의해 정의된 공통되는 특성을 가진 엔터티의 범주	DOI 장르 테이블로부터	확정적	n
식별자(Identifier)	DOI 이외의 식별자	어떠한 문자와 숫자	장르별 구체화	n
제목(Title)	엔터티가 알려진 이름	어떠한 문자와 숫자	장르별 구체화	n
타입(Type)	엔터티의 주요한 구조 타입	작품(work) 물리적 기술 (physical manifestation) 디지털 기술 (digital manifestation) 실연(performance)	확정적	1
창작 방식 (Origination)	엔터티가 만들어지는 과정	창작(original) 개작(derivation, version) 인용(extract) 편집(compilation) 복사(replica)	확정적	n
주요 행위자 (primary agent)	주요 행위자의 이름이나 식별자	허용된 장르 명칭공간으로부터의 식별자 또는 이름	확정적	n
행위자 역할 (agent role)	주요 행위자에 의해 행해진 역할	허용된 장르 명칭공간으로부터의 역할 코드	확정적	n

<표1> DOI Kernel 메타데이터 요소

가(addition)'하거나 '한정(qualification)'함으로써 '확장(extention)'된 메타데이터를 채택할 수 있다(IDF 2000).

DOI 커널 메타데이터는 공개적으로 접근가능하여야 한다. 즉 DOI 이용자는 자유롭게 커널 메타데이터를 이용하여 DOI 검색을 할 수 있어야 한다.

5. DOI 장르와 메타데이터 확장(extention)

DOI 장르란 공통되는 속성을 가지는 콘텐츠(창작물)의 한 부류를 의미한다. DOI 장르는 공통의 특성을 가진 콘텐츠에 대하여 공통의

관심을 지닌 단체들의 요구를 수용하는 과정에서 생겨난다. DOI 장르의 예로는 학술저널논문, MP3 음악파일, e-book, 의학사진 등이 있다.

DOI 장르가 필요한 이유는 각각의 DOI 장르가 필요로 하는 메타데이터가 다르기 때문이다. 비록 모든 DOI 장르가 최소한의 메타데이터인 DOI 커널 메타데이터를 가지지만, 대부분의 장르는 해당 콘텐츠를 구체적으로, 유용하게 관리하기 위해서 확장(extention)된 메타데이터가 필요하다. 음악 파일을 표현하기 위한 메타데이터와 논문을 표현하기 위한 메타데이터가 상이할 수 밖에 없다는 점을 생각한다면 이해하기 쉽다.

확장된 메타데이터는 각각의 장르에 대한 구체적인 응용을 지원할 수 있는 데이터 요소로서 구성되어야 하며, 이는 DOI 커널 메타데이터를 '추가(addition)'하거나 '한정(qualification)'함으로써 이루어진다. 추가란 8가지 요소 외에 주제(subject)라든지, 형태(form), 모드(mode), 청중(audience) 등의 요소를 덧붙이는 것을 의미한다. 한정이란 메타데이터 요소를 좀 더 구체화하는 것을 의미하는데, 예를 들어 문헌 간에 참조 링크를 가능하게 하는 장르에서는 '식별자'라는 최상위 커널 요소를 '저작물(work) 식별자', 그리고 이 저작물의 표현물에 대한 '논문 식별자', '저널 식별자', '출판일', '페이지 수' 등의 구체적인 값을 갖는 요소로 한정할 수 있다(IDF 2000).

6. 메타데이터 관련 IDF 정책

DOI 등록자는 자신의 판단에 의해 DOI 장르를 선택할 수 있다. DOI가 등록된 콘텐츠는 최소한 하나 이상의 DOI 장르를 가져야 한다.

새로운 DOI 장르의 생성을 필요로 하는 단체는 누구나 이의 필요성을 요청할 수 있으며, IDF는 이러한 요청에 대하여 새로운 DOI 장르의 필요성 여부를 검토하게 된다. 새로운 DOI 장르에 대한 기술적, 절차적, 상업적인 규정의 개발은 그 장르를 등록하고자 하는 단체의 몫이지만, 이러한 개발 작업은 반드시 IDF와의 협력과 대화를 통하여 진행되어야 한다.

DOI 장르 메타데이터 스키마 역시 INDECS 메타데이터 스키마의 원칙 및 구조와의 일치성을 유지하면서 개발되어야 한다. 정의(definition)와 값(value)에 대한 리스트는 가능한 한 인덱스 프레임워크를 기반으로 개발된 데이터 사전(ex. EPICS)에서 추출할 것을 권장하고 있다. IDF는 DOI 장르에 대한 메타데이

터 스키마가 DOI 커널 메타데이터와 인덱스 데이터모델과의 일치성을 확보하고 있는지를 검토하며, 어떠한 DOI 장르 메타데이터 스키마도 IDF의 승인이 없이는 선언될 수 없도록 하고 있다. 또한 DOI 장르와 관련한 모든 절차적, 상업적인 규정 역시 IDF 정책과의 일치성을 확보하기 위하여 IDF의 승인을 받도록 하고 있다.

적어도 하나 이상의 등록기관이 DOI 장르에 관한 업무를 수행하여야 하며, 그렇지 않을 경우 IDF는 해당 DOI 장르를 공식적인 DOI 장르로 인정하지 않는다.

이처럼 IDF는 DOI 장르의 생성 및 이의 응용 개발은 철저히 이해 단체의 주도하에 진행되도록 개방형 정책을 채택하고 있으나 상호운용성 확보와 일관된 정책 추진을 위하여 IDF의 검증 및 승인 절차를 거칠 것을 원칙으로 하고 있다.

7. 결론

DOI는 디지털 저작물을 비롯한 각종 콘텐츠의 영구적인 식별자로서, 이를 이용한 다양한 응용서비스의 개발은 더욱 풍요로운 정보통신 서비스 환경을 제공하여 줄 것으로 기대된다. 따라서 DOI의 응용서비스 개발을 촉진시키기 위한 강력한 수단 중 하나인 DOI 메타데이터의 개발 의의는 매우 크다.

이미 IDF가 최소한의 요소로 구성된 DOI 커널 메타데이터를 발표하였으나, 실제 서비스에 활용되기 위해서는 각각의 장르가 필요로 하는 추가적인 요소를 포함한 확대된 메타데이터가 개발되어야 한다.

장르별 메타데이터에 대한 상호운용성 확보와 정책적 일관성 유지를 위하여 IDF 정책 및 개발 동향을 지속적으로 견지할 필요가 있다.

그러나 우리나라의 콘텐츠 유통 환경을 고려한 메타데이터를 마련하고 이의 적용을 확산시키기 위해서는 콘텐츠 관련 기관들이 수동적인 입장에서 벗어나 적극적으로 메타데이터 연구 개발에 참여하여야 한다.

참고문헌

- Godfrey Rust, Mark Bide. 1999. The <Indecs> metadata model. May 1999. <<http://www.indecs.org>>
- IDF. 1999. "DOI Metadata Principles". March 1999. <<http://www.doi.org>>
- IDF. 2000. DOI Handbook. Version 0.3 July 2000. <<http://www.doi.org>>
- John S.Erickson. 1999. "The Role of Metadata Supply Chains in Doi-Based, Value-added Services" April, 1999. <<http://www.ybp.com/yps/papers/doiservicesapr99.htm>>
- Norman Paskin, Godfrey Rust. 1999a. "Digital Object Identifier initiative : metadata implication". DOI discussion paper Number 2 (Version 3, 10 FEBRUARY 1999). <<http://www.doi.org>>
- Norman Paskin. 1999b. "DOI: Current Status and Outlook". D-Lib Magazine. May 1999.
- 内田尚子. 1999. "디지털·オブジェクト 識別子(DOI)". 情報管理. Vol.42 No.1.