

국제표준이름식별자(ISNI)를 중심으로 한 연구자 식별자의 통합관리체계 연구

A Study on Integrated Management System of Researcher Identifiers based on the ISNI (International Standard Name Identifier)

변 회 균(Hoi-Kyun Byeon)*

오 복 희(Bok-hee Oh)**

초 록

이 연구는 연구자 정보를 관리하기 위해 여러 조직에서 발행하고 관리하는 다양한 식별자를 국제표준이름식별자(ISNI)를 중심으로 융합·관리함으로써 식별자의 중복발행을 방지하고 기관의 연구자 정보의 재정비에 활용할 수 있는 연구자 식별자 통합관리체계 모형을 제시하였다. 여러 식별자 중에 다양한 분야에서도 적용할 수 있는 ISNI의 메타데이터 구조와 관리체계를 살펴보고, ISNI를 중심으로 각 기관에서 발행하는 식별자를 통합하여 관리할 수 있는 체계와 시스템 모형을 설계하였다. 이 모델은 데이터구조 계층, 플랫폼 계층, 어플리케이션 계층, 관리 및 보안계층(4계층)으로 구성되며 주요 기능의 설명 및 통합관리시 고려해야할 이슈를 제시하였다.

ABSTRACT

This study proposed an integrated management system model of researcher identifiers using the ISNI(International Standard Name Identifier) that can prevent duplicate issuance of identifiers and can be used to reorganize the authors' information. We looked at the metadata structure and management system of the ISNI that can be applied in various fields among identifiers. And we have designed the integrated ID system and system model that can manage the identifiers issued by each institute based on the ISNI. This model consists of data structure layer, platform layer, application layer, management and security layer (4 layers) and we have presented some of the key features and issues to consider when managing the integrations.

키워드: 국제표준이름식별자, 연구자 식별자, 저자전거,전거제어, 연구자 통합관리

International Standard Name Identifier, ISNI, Researcher Identifier, Author Authority, Authority Control, Researcher Integration Management

* 숭의여자대학교 문헌정보학 겸임교수(hkbyeon818@daum.net) (제1저자)

** 국회도서관 데이터융합분석과 주무관(bokhee@nanet.go.kr) (공동저자)

논문접수일자 : 2018년 8월 17일 논문심사일자 : 2018년 9월 6일 게재확정일자 : 2018년 9월 18일
한국비블리아학회지, 29(3): 139-155, 2018. [http://dx.doi.org/10.14699/kbiblia.2018.29.3.139]

1. 서론

1.1 연구의 필요성 및 목적

4차 산업혁명의 시대를 맞이하여 도서관은 상호운영성이 확보된 양질의 축적된 정보를 기반으로 다양한 융합·분석 서비스 개발과 고도화에 노력하고 있다. 도서관에서 수집하는 콘텐츠들은 여러 주제가 융합된 형태의 전문화된 내용을 다루고 있으며, 최근의 연구경향은 여러 분야의 융합을 통한 학제적 성격이 강하므로 연구주제가 복잡적이고 연구자간의 지적관계도 과거보다 훨씬 복잡한 수준이다.

이용자에게 특정 저자의 저작물 정보나 저자의 다른 저작물 또는 동일 주제의 연구자 검색을 제공하기 위해서는 연구자의 식별이 필수적이다. 이를 위해서 전거 또는 연구자 정보 그리고 연구자의 저작물 정보를 활용하여 연구자의 중의성을 해소하기 위한 노력이 반드시 필요하다. 도서관의 경우 저자전거는 저작자의 정보에 유일한 식별자를 부여하여 향후 개별 저자에 연관된 저작자 또는 저작물을 정확하게 구별하는 강력한 도구로 활용된다.

이러한 양질의 전거 구축은 많은 비용과 시간이 필요하다. 하지만 도서관의 인력감축을 포함한 운영비용의 축소 및 전거의 활용성의 미비로 인해 전거구축을 포기하거나 운영이 되더라도 기관별 품질 차이가 크다.

이러한 현실적 어려움을 극복하기 위하여 국외에서는 세계 각국의 국립도서관들이 참여하는 VIAF(Virtual International Authority File, 가상국제전거파일) 프로젝트가 2008년부터 시작되었고, 과학 분야의 연구자 식별을 위하여

학술저널 중심의 출판사와 대학들이 참여하는 ORCID(Open Researcher and Contributor ID, 개방형연구자ID) 서비스가 2012년부터 시작되었다. ORCID는 연구자가 자신의 성과물을 스스로 선언하는 셀프 클레임 방식으로 연구자가 자신의 ID를 직접 등록·관리하고 자신의 연구결과물을 찾아 선언하는 방식이다(조제인 2013). 이렇게 VIAF와 ORCID가 국제적으로 통용될 수 있는 식별자로 사용되면서 과거보다 전거의 구축 및 관리가 용이하게 되었다. 하지만 VIAF는 단행본 저자를 중심으로 도서관 분야에서 주로 활용된다는 한계가 있고, ORCID는 학술 분야에서 연구자를 중심으로 사용이 확대되고는 있으나 연구자 스스로 식별자를 관리하기 때문에 식별자가 복수 발행되는 등 관리성 부재의 단점이 있다.

ISNI(International Standard Name Identifier, 국제표준이름식별자)는 학술분야의 연구자뿐만 아니라 작가, 예술가 등 모든 분야의 창작자를 광범위하게 포함하는 표준 식별자의 필요성이 높아짐에 따라 2012년에 ISO 27729 국제표준으로 제정되었다. 2014년 ISNI와 ORCID는 지속가능성과 상호운용을 위한 전략적 파트너십을 체결하였고 2017년 12월부터 레코드를 상호 조회하여 ISNI와 ORCID를 연계할 수 있도록 하였다(Rickenmann 2018). 현재 ISNI는 VIAF, ORCID, BnF(Bibliothque Nationale de France), PCC(Program for Cooperative Cataloging), YouTube 등 47개의 다양한 분야의 주요 기관들이 참여하는 강력한 식별자로 확장해 나가고 있다(MacEwan 2017).

국내의 경우 한국학술진흥재단(現 한국연구재단)에서 대학 및 기관 연구자들의 연구업적정

보를 국가차원에서 공유 및 활용하기 위해 KRI (Korean Researcher Information, 국가연구업적 통합정보) 시스템을 구축하여 2008년부터 KRI ID를 발급하기 시작하였고, 한국과학기술정보연구원(KISTI)에서는 국가연구개발에 참여하는 연구자의 관리를 위해 2011년부터 과학기술인등록번호를 발급하고 있다. 그리고 국회도서관과 국립중앙도서관은 개별적인 저자전거ID를 발행하면서 국내 ISNI 등록기관으로서 연구자 또는 창작자의 ISNI 발급대행을 추진하고 있다.

이렇게 발급된 연구자 식별자는 여러 기관에서 각자의 체계로 발급·관리되어 데이터를 통합하고 연계할 때 식별자의 역할을 충분히 하지 못하고 자관이 발급한 식별자와 연결된 정보만 제공하는데 그치고 있다.

이 연구의 목적은 국회도서관과 같이 방대한 자료의 수집과 서비스를 제공하는 국가도서관에서 연구자 정보의 관리와 협력 기관과의 공유를 위하여 ISNI를 중심으로 한 다양한 식별자를 통합·관리할 수 있는 모형을 제시함으로써 국내 연구자의 식별자의 중복발급을 방지하고 자관의 연구자 정보를 정비하거나 저작물을 식별할 때 활용하여 융합·분석적 링크드 데이터의 구축 기반 마련에도 기여할 것이다.

1.2 관련 연구

지금까지 저자의 중의성을 해결하기 위한 방법 연구나 특정 분야를 벗어나 여러 분야를 포괄하여 저자 정보와 관련된 정보자원을 식별하기 위한 노력들이 진행되었다.

우선 정확한전거데이터의 구축을 위한 연구들은 다음과 같다.

김송이, 정연경(2015)은 한국문학번역원에서 관리하는 한국문학 작가명 데이터베이스(문인 DB)를 대상으로 한국, 미국 및 일본의 한국인 저자전거레코드의 구축현황을 비교분석하여 한국인 저자전거의 현황과 문제점을 살펴보았다. 여기서 국립중앙도서관은 영문과 한자의 이형누락을 발견하였고, 미의회도서관과 일본국립 국회도서관의전거레코드는 잘못된 저자명의 기입, 생몰년의 미표기 등 오기와 항목 누락을 발견하였다. 이러한 문제 해결 방안으로 데이터 관리주체인 한국문학번역원과 국립중앙도서관 등 주요 기관들의 VIAF의 참가를 제시하였다.

이석형(2014)은 저자식별 및 기관 중심으로 연구성과물을 집중시키기 위한 요소로 저자 이메일, 공저자명 그리고 소속기관을 들었다. 하지만 소속기관은 저자의 졸업, 이직 혹은 소속기관의 조직개편 등에 따라 그 변화가 매우 크고, 저자에 따라 그 표현형식이 매우 다양하므로 소속기관명의 통제가 필요하며 이를 위하여 소속정보를 하나의 식별데이터로 이용할 수 있는 저자의 소속기관명 식별데이터 구축 방안을 제시하였다.

한편 다양한 분야에서 활용할 수 있는 글로벌 식별자체계의 도입과 장점을 소개하는 연구는 다음과 같다.

조재인(2013)은전거통제가 단위 시스템에 한정되고 대상이 도서관 소장자원에 국한되어 있어 인터넷 환경에서 상업 출판사, 기관 레포지토리 등의 자유롭게 유통되는 학술 논문의 글로벌한 개념의 식별체계 도입 필요성을 제시하고 도서관계, 출판계를 망라한 글로벌 링크를 지향하는 ORCID를 분석하고 이를 응용하여 핵

심ID 시스템과 파트너 시스템으로 국내 학술연구물의 저자명 식별 시스템 구축방안을 제안하였다.

이미화(2014)는 정보산업 분야 전체를 포괄하는 연계식별자로 ISNI의 중요성을 인식하여 ISNI의 개념 및 전자제어의 활용가능성을 모색하였다. 국내 협력 전자제어를 ISNI의 개념으로 도입하고 도서관 및 정보산업 분야에서 생산되는 모든 전자 데이터를 상호 활용할 수 있도록 추천하고, ISNI를 이용하여 여러 식별자를 연계함으로 링크드 데이터의 구축이 가능하며 도서관의 메타표준인 KORMARC 확장 등을 제안하였다.

김은정과 노경란(2017)은 ICT기술 발전과 디지털 확산으로 학술 커뮤니케이션의 효율화와 고속화, 연구방법의 간소화와 가속화가 진행되고 있으며, ORCID를 사용함으로써 연구자와 연구기관에게 주는 장점과 미국, 영국, 이탈리아, 중국 등 해외 주요 국가들의 ORCID 도입 및 활용 사례를 소개하고 국가 차원에서의 저자 식별을 위한 방안으로 ORCID 활용과 이를 위한 거버넌스, 시스템 요소, 정책적 제도적 측면을 제시하였다.

이성숙 외(2017)는 개별기관의 독립성과 자율성을 전제로 하여 복수의 기관 간의 협업을 지원하는 링크드 데이터가 도서관 및 유관 분야에서 어떻게 구축되고 있는지에 대해 조사하고 국립중앙도서관의 링크드 데이터를 대상으로 인물정보가 표현되고 연계되는 방식을 분석하여 효과적인 인물 정보의 구조화 방안을 제안하였다.

Meadows(2016)는 저자의 모호성 문제를 여러 이유를 들어 설명하고 ORCID가 이러한 문

제를 해결할 것이라 설명하였다.

Krämer, Momeni, Mayr(2017)는 과학 분야에서 저자식별의 분석을 위해 Web of Science (WoS)와 Scopus의 컬렉션 중 1996년-2014년 기간의 6,032,000개의 출판물 데이터 세트에서 Research ID, ORCID 및 Scopus 저자ID를 비교하여 저자들의 분포 및 주제 영역의 차이점 등을 밝히면서 Scopus ID는 전체 식별이 가능하나 중복성 규명의 어려움이 있고, ORCID는 점차 늘고 있는 추세이나 아직 출판물 전체에 차지하는 비중은 낮은 편이라고 밝혔다.

이러한 선행연구를 기반으로 정확한 협력연구데이터 구축과 글로벌 식별자 체계의 장점을 활용하기 위해 이 연구에서는 포괄적 범위의 국제표준 식별자인 ISNI의 특징을 살펴보고, ISNI 외에 여러 기관에서 발급한 연구자의 기존 식별자들을 통합하여 관리할 수 있는 모델을 제시하고 그에 따른 고려사항을 살펴본다.

2. 국제표준이름식별자(ISNI)

2.1 개요

ISNI는 저작물의 창작, 공연, 제작, 관리, 유통에 관련되는 자연인과 법인, 가상의 인물(배역, 작품의 등장인물 등)을 대상으로 고유 식별번호를 부여하여 저작권의 관리, 유통 및 수익 분배 등에 적용할 수 있는 코드이다. ISNI는 중계 식별자(Bridge Identifier)로서 각기 다른 데이터베이스를 연계하여 정보를 제공한다. 그리고 저작권자를 구분할 때 가장 필수적인 최소한의 메타데이터를 사용하여 저작권자를 구

분하고, 나머지 연관 정보는 접근이 제한된 개별 데이터베이스에 남겨두어 각 산업의 파트너들이 기밀정보의 유출 없이 원활한 정보교환을 할 수 있도록 한다(ISO 2012).

ISNI는 <그림 1>과 같이 로컬시스템이나 특정 네트워크 또는 컨소시엄 수준에서 운영되는 식별자를 ISNI로 대체할 필요 없이 기존의 표준 또는 독점적인 식별자를 매핑하면서 상호 운영성이 높은 상위 식별자로 위치하고 있다(Angjeli, Mac Ewan, and Boulet 2014).

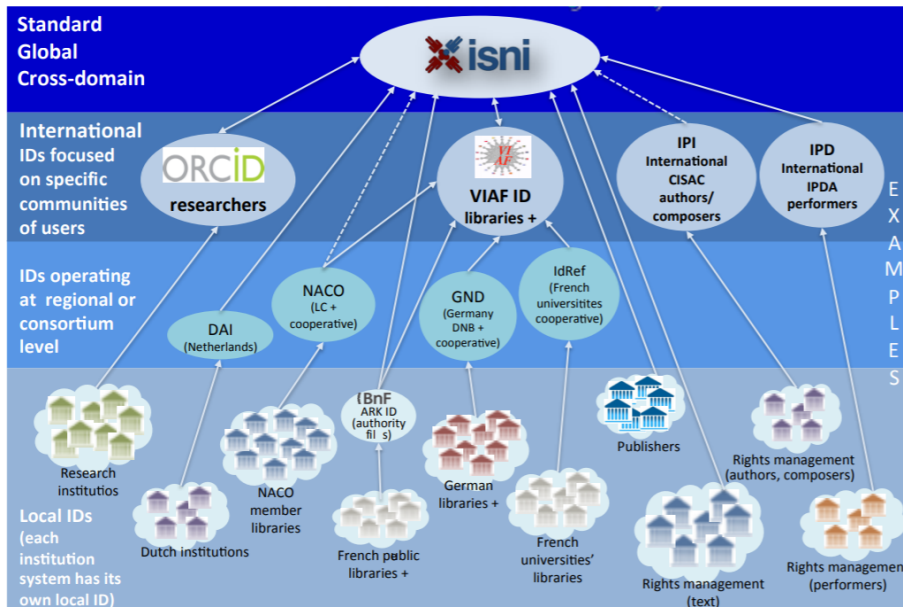
ISNI는 기존의 다양한 데이터 제공자(ORCID, VIAF, CISAC 등)의 로컬식별자를 수집하고 공공객체(Public Identity)로 할당하여 중립적인 최상위 식별자로 연결한 후 중앙관리데이터베이스로 관리되어 식별자 링크의 통합, 유지보수, 업데이트 및 보급을 보장하는 인프라를 제공한다(Angjeli, Mac Ewan, Boulet, 전계서).

<표 1>은 ISNI가 속한 표준화기구 분과와 총괄관리기관 등 ISNI의 간략한 개요이다.

<표 1> ISNI 개요

구분	내용
식별체계명	International Standard Name Identifier, 국제표준이름식별자
표준화기구 분과	ISO TC46/ SC9
표준번호	ISO 27729:2012
총괄관리기관	ISNI International Agency
등록관리기관	ISNI Registration Agency
적용분야	창작자(개인 및 단체)

ISNI의 코드 구조는 0~9까지의 숫자 16자리(15자리 숫자 + 체크숫자)로 구성된다. ISNI의 각 자릿수 혹은 ISNI의 일부분은 특정한 의미를 지니지 않는다. 체크숫자는 앞선 숫자 15 자리에 따라 숫자 0~9 혹은 'X'로 결정된다. <표 2>는 ISNI 예시이다.



<그림 1> 인명·단체명과 관련된 식별자의 계층

〈표 2〉 ISNI 구문구조

구분	설명
구문구조	ISNI [식별번호] - [check digit]
코드예시	ISNI 1422 4586 3573 0476
식별번호	1422 4586 3573 047
Check digit	6

2.2 ISNI 메타데이터 구조

ISNI의 메타데이터는 크게 인명(Person)과 단체명(Organisation)으로 구분된다. 〈표 3〉은 인명과 개인명을 등록하기 위한 레코드 포맷으로 레코드별 탭(Tab)으로 구분된 텍스트 형식의 메타데이터 항목으로 구성되며 메타데이터가 존재하지 않을 경우 'null'로 입력될 수 있다. 텍스트 파일에 대한 인코딩은 UNICODE UTF-8 형식을 준수해야 한다. 인명의 경우 28개의 메타데이터 중에 필수항목으로 구분되는 3개의 필드(Local ID, Forename, Surname)는 반드시 입력해야 하며, 중요항목으로 구분되는 6개의 필드(Other ID, Other ID Type, Birth Date, Death Date, Title, Creation Role)는 가급적 입력해야 한다.

단체명은 기관이나 단체명을 등록하기 위한 레코드 포맷으로 텍스트 형태의 27개 메타데이터 항목으로 구성되며 메타데이터가 존재하지 않을 경우 'null'로 입력될 수 있다. 필수항목으로 구분되는 3개 항목(Local ID, Organisation Type, Main Name of Organisation)은 반드시 입력되어야 하며, 중요항목으로 구분되는 8개

의 필드(Other ID, Other ID Type, Forename, Surname, Birth Date, Death Date, Title, Creation Role, Related Person, Country)는 가급적 입력해야 한다.

2.3 ISNI 발급처리

ISNI 발급을 위해서는 ISNI RAG(Registration Agencies)가 정한 인명과 단체명의 메타데이터 표준을 준수하여 만든 데이터를 ISNI-IA(International Agency)¹⁾로 제출해야 한다. 이때 유의할 부분은 다음과 같다.

첫째, 인명의 경우 성과 이름을 분리하여 신청해야 한다. MARC로 관리할 때 영문명은 '성, 이름'으로 구분하여 관리하지만 한글명은 성과 이름을 붙여서 저장하기 때문에 ISNI 등록 신청할 때 한글명을 성과 이름으로 분리해야 하는 어려움이 있다. 만약 호, 별칭과 같이 성과 이름을 분리할 수 없는 경우 'Surname(성)' 필드에 성과 이름을 모두 묶어 신청할 수밖에 없으나 정확한 관리를 위해 저장구조에는 성과 이름을 분리하여 관리할 수 있도록 해야 한다.

둘째, 단체명인 경우 Country(국가코드)가 중요항목으로 되어 입력을 선택할 수 있으나 가급적 필수로 입력해주는 것이 ISNI 발급할 때 다른 ISNI 데이터와의 중복을 피할 수 있다. 이것은 국가별로 국가기관이나 협회 및 연구소의 경우 비슷한 기능을 가진 유사한 명칭을 사용하기 때문이다.

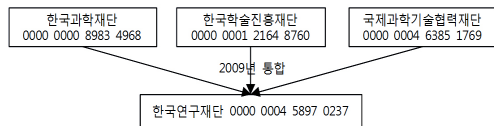
1) ISNI-IA(International Agency, 국제본부)는 ISNI의 정책을 관리하고 ISNI 지원시스템 구조를 결정한다. ISNI의 발급과 메타데이터 수집에 관련한 시스템을 개발, 관리하고 ISNI 등록 메타데이터의 갱신, 업데이트, 확장을 촉진하고 ISNI를 발급받은 기관 또는 개인에게 지침 및 안내를 제공한다. ISNI-IA 시스템은 OCLC 데이터센터에 위탁 운영되고 있다.

〈표 3〉 ISNI 등록 메타데이터 구조(인명, 단체명)

No	필드명		설명		포맷
	인명	단체명	인명	단체명	
0	Local name identifier		데이터 제공기관에서 사용하는 로컬 ID		string
1	Other name ID		로컬 ID 외 다른 식별자		string
2	Other name ID type		로컬 ID 외 다른 식별자 종류		string
3	Name Prefix	-	이름 앞에 붙는 호칭	-	string
4	Forename	-	이름	-	string
5	Middlename	Organisation type	성과 이름 사이 중간 이름	Broadcaster, Publisher, Hospital 등 단체 유형	string
6	Surname	Main name of organisation	성	단체명	string
7	Name Suffix	Subdivision name	이름 뒤에 붙는 호칭	단체의 하위부서와 섹션의 이름들	string
8	Alternative name		다른 표현형(이형)	단체의 약칭 등 이형	string
9	Date of Birth	Date use of the name commenced	출생일	설립일	ISO 8601포맷
10	Date of Death	Date use of the name terminated	사망일	폐쇄일	ISO 8601포맷
11	Title identifier		저작물의 식별자	간행물의 식별자	string
12	Title identifier type		저작물의 식별자 종류	간행물의 식별자 종류	string
13	Title		저작물의 제목	간행물 제목	string
14	Subtitle		저작물의 부제목	간행물 부제목	string
15	Contributed to	-	수록 저널명 또는 참여했던 공동 작품의 이름	-	string
16	Year of publication		출판년도		string
17	Creation class		창작 영역		string
18	Creation role		창작 역할		지정어
19	Publisher		출판사		string
20	Dewey classification number	Organisation relationship type	듀이십진분류코드(DDC)	Affiliated organisation과의 관계	string
21	Affiliated institution	Affiliated organisation	관련 단체명		string
22	Related persons		관련 인물명		string
23	Relationship type	Person relationship type	“Related person”과의 관계	“Related person”과의 관계 유형	지정어
24	-	Country	-	단체가 위치한 국가	-
25	-	Region or state	-	단체가 위치한 주	-
26	-	City	-	단체가 위치한 도시	-
27	URL	URL to Organisation's main web page	정보원의 URL 주소	단체의 웹사이트 주소	url
28	Instrument	-	악기나, 목소리(성악 파트) 정보	-	지정어

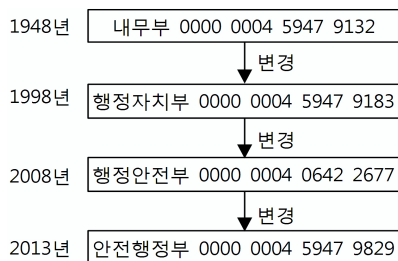
셋째, ISNI의 국가코드는 통합서지용 MARC의 발행국부호표와 달리 ISO 3166의 일부인 ISO 3166-1의 2자로 된 코드를 사용하고 있으므로 같은 코드라도 다른 나라를 의미할 수 있다. 예를 들어 'China'의 경우, 통합서지용 MARC에서의 'cc'는 ISO 3166-1에서는 'Cocos Island'를 의미한다.

넷째, 단체명인 경우 단체의 통·폐합 또는 분리 등으로 인해 복잡하고 다양한 명칭이 생길 수 있으므로 연혁 관리가 필요하다. ISNI는 단체명의 변경된 명칭마다 발급할 수 있고 단체명간의 관계(Organisation to Organisation)를 이전(formerName)과 이후(lastName)로 설정하여 상호 연결할 수 있도록 되어 있다. <그림 2>는 현재 한국연구재단이 통합되기 이전의 단체명과의 관계를 보여준 사례이다.



<그림 2> 통합된 단체명의 ISNI 발급사례

<그림 3>은 현재 행정안전부가 시대에 따라 명칭이 변경된 사례이며 그에 따른 ISNI가 발급된 상황을 보여주는 것이다.



<그림 3> 변경된 단체명의 ISNI 발급사례

다섯째, Creation class는 창작 영역을 가리키는 항목으로 기본 값은 'text'를 사용한다. ISNI에서 내부적으로 관리하고 있는 문자셋은 있으나 지정한 값이 아니므로 등록기관에서 다른 값을 입력해도 무방하다.

여섯째, Creation role은 창작 역할을 나타내며 ISNI가 지정한 코드를 사용하여야 한다. VIAF에서 사용하는 3개의 소문자로 구성된 문자코드를 사용하거나 CISAC(International Confederation of Societies of Authors and Composers, 국제저작권관리단체연맹)에서 사용하는 2개의 대문자로 구성된 문자코드를 선택하여 사용한다. 기본 값은 'aut'(author)이다.

일곱째, ISNI는 저자식별을 위해 중요항목 외에 3건 이상의 저작물을 입력하도록 권장하고 있으나 중요 항목들을 충실히 채울 경우 1건의 저작물을 입력했다더라도 ISNI가 발급될 수 있다.

3. 통합관리체계 모델

3.1 시스템 구성

국회도서관과 같이 방대한 자료의 수집과 서비스를 제공하는 국가도서관인 경우 수많은 분야의 연구자 식별을 위해서는 다양하게 발생되는 연구자 식별자를 통합관리하고 향후 확장성과 신뢰성을 담보로 할 수 있는 연구자 정보의 관리가 필요하다.

국회도서관은 2007년부터 2017년까지 약 61만 건의 인명과 단체명 저자전거를 구축하였다. 그리고 연구자 중심의 ISNI 등록기관으로서 다

양한 식별자를 통합 관리하여 특정 주제와 관련된 학제적 융합이나 연구자들의 지적 연계정보를 협력기관에 제공하는 서비스가 필요하므로 이에 대한 모델을 <그림 4>와 같이 제시한다. 이 모델의 아키텍처는 총 4개의 계층으로 구분하고, 아래 계층부터 '데이터구조 계층', '플랫폼 계층', '어플리케이션 계층'으로 나누고, 3개의 계층과 관련 있는 '관리 및 보안 계층'으로 구성한다.

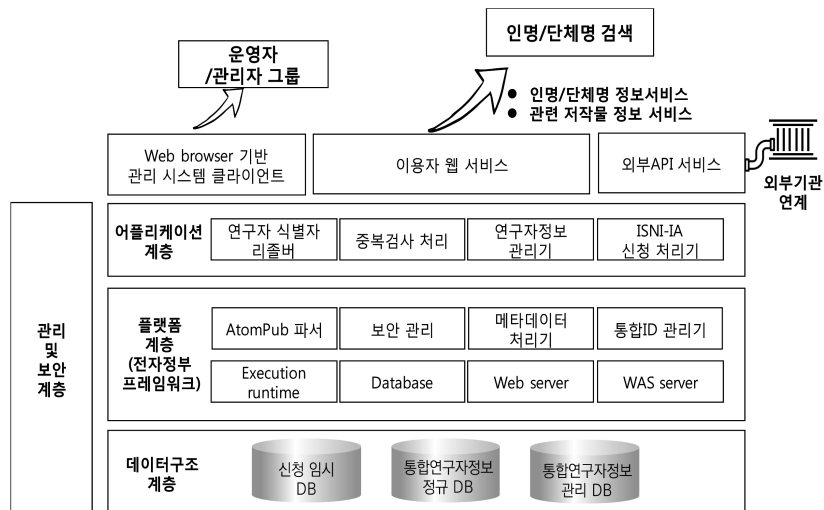
데이터구조 계층은 외부기관으로부터 제공 받은 연구자 정보를 통합하기 전에 임시로 저장하는 '신청임시 DB', 외부기관에서 수집된 정보와 ISNI에 등록되어 있는 해외 연구자 정보를 수집하여 연구자 정보와 저작물 정보를 통합한 '통합연구자정보 정규 DB', 통합관리체계를 관리하기 위한 각종 코드정보와 ISNI로 발급신청 및 변환을 위해 필요한 '통합연구자 정보 관리 DB'로 구성한다.

플랫폼 계층에는 WEB, WAS, Database 등의 인프라 소프트웨어와 ISNI-IA에서 ISNI 발

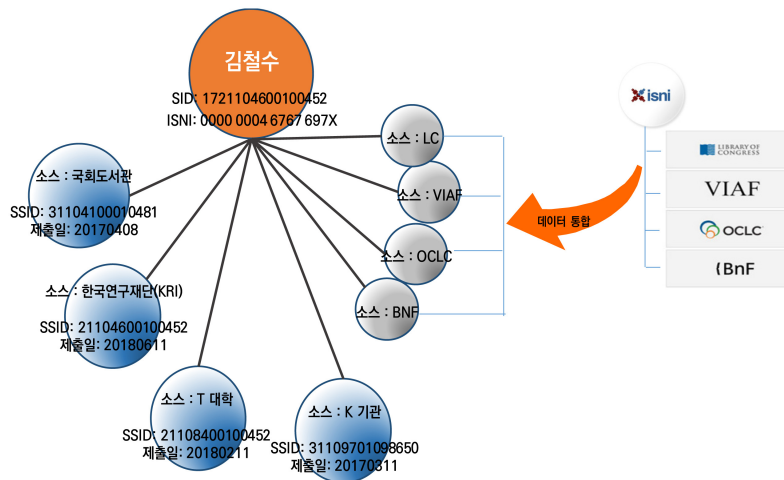
급신청 프로토콜인 'AtomPub API 파서', 연구자 정보를 통합관리하기 위한 '메타데이터 처리기', 비밀번호 등 보안이 필요한 민감한 개인정보를 암호화하는 '보안관리' 그리고 여러 기관에서 관리하고 있는 연구자 식별자를 관리하기 위한 '통합ID 관리기'를 구성한다. 어플리케이션 계층에는 연구자의 로컬ID를 이용하여 타 식별자를 알려줄 수 있는 '연구자 식별자 리졸버', 연구자 정보의 중복검사를 하는 '유사도 검사 모듈', 연구자의 기본 프로필 및 저작물 정보를 관리하는 '연구자 정보 관리기', ISNI 등록 신청 및 발급 결과를 관리하는 'ISNI 신청 처리기'로 구성한다.

3.2 통합ID 체계

연구자 식별자의 통합관리를 위해서 가장 중요한 부분은 통합ID 관리체계를 설계하는 것이다. <그림 5>와 같이 연구자 '김철수'의 정보는



<그림 4> 통합관리체계 모형



〈그림 5〉 ID 통합 모형 예시

국회도서관, 한국연구재단(KRI), T대학, K기관에서 각각 관리되고 있고 이미 VIAF, OCLC, BnF 등과 같은 외국 기관으로부터 ISNI가 발급되어진 상황이라 가정하자. 이런 경우 연구자 정보가 여러 기관에서 자관의 로컬 식별자를 통해 관리되고 사용되고 있으나 본인은 자신에 대한 식별자 정보를 알기 어렵다. 이 모델은 연구자 자신이 여러 기관에서 관리하고 있는 다양한 식별자를 확인할 수 있도록 정보를 제공하고, 타 기관에서 관리되고 있는 정보를 수집하여 소스정보(제공기관 정보)로 관리 및 통합하며 ISNI-IA에서 갱신될 때마다 주기적인 동기화를 통하여 연구자 정보의 최신성을 유지 관리할 수 있는 중앙집중식 데이터베이스이다. 통합ID를 관리할 때 유의사항은 다음과 같다.

첫째, 수집할 때의 기존 데이터와 중복성의 식별 문제이다.

둘째, 소스기관(저자정보제공기관)에서 발생하는 병합 및 삭제 문제이다. 전자의 경우 '김철수' 연구자를 이미 보유하고 있을 때 타 기관

에서 입수된 '김철수'와 동일인인지 중의성을 해결하여 기존 정규 DB에 없는 경우에는 신규 레코드를 생성하고, 만약 동일인이 존재하면 기존 데이터에 제공기관을 추가하고 입수된 데이터를 통합하는 기능이 있어야 한다. 이러한 연구자의 중복성 문제해결은 DB관리에서 매우 중요하고 데이터가 축적될수록 서비스의 품질과 직결된다. 이 모델에서는 어플리케이션의 '중복검사 처리' 모듈에서 담당한다.

후자의 경우는 소스기관에서 데이터를 제공받아 통합처리된 이후 소스기관에서 해당 데이터를 병합 또는 삭제 처리하여 데이터가 변경되었을 때 통합관리체계 모델에서 어떻게 처리할 것인지의 문제이다. 만약 소스기관에서 데이터의 병합 또는 삭제가 발생한 것을 인지하지 못한 상태에서 해당 데이터의 식별자가 이미 타 기관에 제공되어 사용되고 있다면 연계정보가 끊어지는 절단링크(broken link)가 발생할 수 있다.

소스기관의 데이터가 통합처리된 이후에 소스기관은 가능하면 기존 상태를 유지하는 것이

바람직하나 자체데이터 점검 등으로 소스기관에서의 데이터 병합 또는 삭제가 발생한 경우 통합관리체계는 기존 데이터와 식별자를 유지하면서 삭제정보를 관리하고, 병합된 경우에는 어느 데이터와 병합되었는지 병합정보를 저장하여 관련 서비스를 지속할 수 있는 체계를 유지해야 한다.

3.3 구성요소별 주요 기능

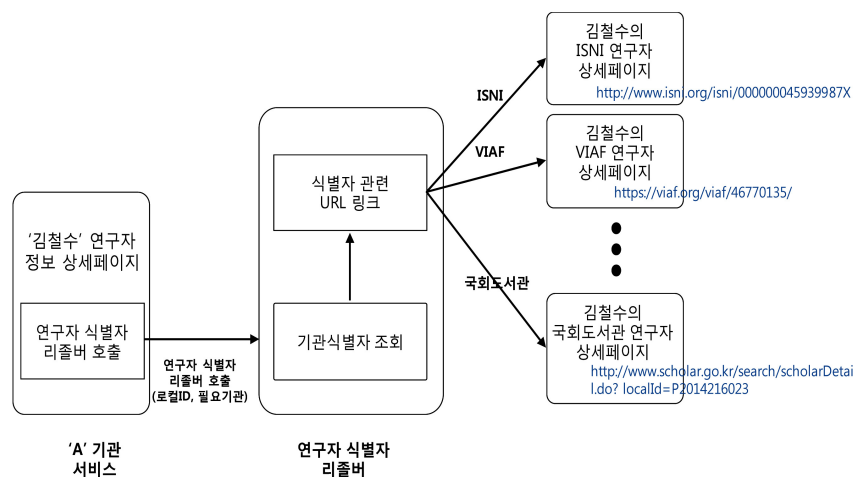
통합관리체계 모형의 데이터 구조 계층, 관리 및 보안 계층, 플랫폼 계층에서의 일반적인 제품이나 솔루션에 관한 모듈설명은 생략하고, 서비스와 관련된 어플리케이션 계층에 '연구자 식별자 리졸버', '중복검사 모듈', '연구자정보 관리기', 'ISNI-IA 신청 처리기'를 중심으로 설명한다.

- 연구자 식별자 리졸버(resolver)
연구자 식별자 리졸버는 자관에서 관리하는

연구자 정보를 이용하여 다른 식별자를 질의하거나 다른 기관에서 관리하고 있는 연구자 상세정보로 링크하기 위한 서비스를 제공한다. <그림 6>에서 'A' 기관에 관리하고 있는 김철수의 식별자인 '0000089892'로 ISNI, VIAF와 같은 식별자 정보를 질의하거나 특정 기관의 식별자를 이용하여 URL정보를 얻기 위해 '연구자 식별자 리졸버'를 호출하면 요청유형을 확인하여 원하는 정보나 URL 링크 정보를 리다이렉트(redirect)하여 해당되는 연구자 상세정보페이지로 연결시켜준다. 이 때 'A' 기관의 정보가 아닌 타 기관의 식별자를 이용하여 호출하는 것도 가능하다.

- 중복검사 모듈

연구자 통합관리체계에서 연구자의 다양한 식별체계를 통합하기 위해 중복검사 모듈은 매우 중요하다. 외부에서 입수되는 연구자 정보를 기존의 연구자와 중복여부를 확인하여 중의성을 해소하고 신규레코드를 생성할 것인지 기



<그림 6> 연구자 식별자 리졸버 개념도

존 연구자의 정보와 통합할 것인지 판단이 필요하다. 이러한 중의성을 해소하기 위해 연구자의 상세정보와 연구자의 저작물 정보를 이용하여 연구자를 식별한다.

연구자의 식별처리는 식별성 높은 주민번호, 이메일, 전화번호 등을 이용하는 것이 가장 효과적이나 민감한 개인정보를 수집·보유하는 것은 「개인정보 보호법」상 현실적으로 매우 어렵다. 그러나 단행본 저자나 연구자의 경우 자신의 저작물에 대한 홍보 및 안내를 위해 출판물과 웹사이트에 이미 공개된 민감하지 않은 개인정보인 한글명, 이형 정보(영문명, 한자명), 소속기관 등은 수집이 가능하다.²⁾ 이러한 연구자의 기본 정보와 저작물 정보를 이용하면 효율적인 식별이 가능하다는 것은 여러 선행논문들에 의해 연구되었다.

Han, Zha and Giles(2003)는 “한 연구자는 제한된 기간 동안 다른 공동연구자들과 함께 유사한 분야의 토픽을 연구하여 그 결과를 연구 분야와 관련된 저널이나 학술대회에 투고한다.”라는 연구결과를 도출하였고, 강인수 외(2008)는 ‘계재지명’, ‘참고문헌’, ‘논문제목’, ‘소속’, ‘전자메일’, ‘공저자’의 자질을 이용하여 저자의 식별성능을 높이는 실험결과를 발표했다.

이러한 선행 연구결과를 바탕으로 연구자의 기본 정보와 저작물 정보에 항목별 가중치를 부여하여 유사도를 분석하면 연구자의 식별이 용이하다.

연구자 기본 정보 중에 대표명, 이형(영문명, 한자명), 생·몰년을 비교하고 저작물 정보 중 서명, 주제, 출판사항을 비교하여 가중치를 반영

한 유사도 점수를 산출하고 ‘완전중복’, ‘유시중복’, ‘신규’로 구분한다. DB의 신뢰성을 높이기 위해서는 신규와 완전중복의 배점차이를 넓게 하여 신규 및 완전중복은 자동으로 처리하고 유시중복은 검증자가 신규 연구자인지 기준에 관리된 연구자와 동일인인지를 규명한다.

연구자 중복검사를 위한 기본적인 고려사항을 검토하면 다음과 같다.

첫째, 연구자의 정보 중에 민감한 정보의 수집을 피하고 연구자정보 정규DB에서도 저장하지 않는다.

둘째, 식별에 활용할 저작물의 속성정보를 자료유형에 맞춰 적용해야 한다. 연구자를 식별하기 위해 저작물 정보를 활용할 때 연구보고서, 논문 등과 같은 자료유형들 간에는 소속기관, 수록잡지명, 발행처, 키워드 및 주제 등이 식별성이 용이하나 비교대상의 저작물이 서로 다른 유형인 경우 식별성이 떨어진다. 즉 기준에 관리하고 있는 연구자의 저작물 정보가 주로 학술 기사, 학위논문이고 비교대상인 입수 저작물이 단행본만 존재하는 경우에는 비교항목을 일치 시키기가 용이하지 않다. 이렇게 자료유형이 상이한 경우에는 저작물 정보 중 공통항목인 제목과 주제를 이용하여 비교하고 연구자정보의 항목들 가중치를 높일 수밖에 없다.

셋째, 연구자의 유형을 개인명과 단체명으로 구분하여 처리한다. 단체명인 경우 단체의 생성, 분리, 합병의 문제로 인한 단체명의 변경이 발생하므로 지속적인 관리가 쉽지 않다. 또한 다양한 약칭 및 영문표기 등의 이형이 존재하여 그에 따른 가중치를 고려해야 한다.

2) 「대법원 2016.8.17. 선고 2014다235080 판결」과 「개인정보보호위원회 결정 제2016-23-72호」내용 참조

• 연구자정보 관리기

연구자 정보를 통합관리하기 위한 DB 관리기이며 처리 상태별 3단계로 구분하여 '정보수집 단계', '검증단계', '승인단계'로 관리한다. '정보수집 단계'는 외부기관 DB 직접반입, 외부기관 신청API 및 개인직접 등록 등의 기능이 필요하다. 입수된 데이터를 '검증 단계'에서는 자동중복검사 절차를 거쳐 '신규'와 '유사중복'과 '완전중복'으로 구분처리하고, 데이터를 쉽게 비교할 수 있는 처리화면을 통해 유사중복 데이터를 신규 또는 중복 자료로 판단한다. 또한 검증처리 결과를 최종 확인하는 '승인 단계'를 거쳐 DB품질의 신뢰도를 높일 수 있도록 한다. 이 외 '동기화'는 ISNI 발급된 국내외 연구자정보의 지속적인 동기화를 통해 글로벌하게 병합 및 갱신되는 최신정보를 수집하여 동기화한다.

• ISNI-IA 신청 처리기

ISNI는 표준 API를 통해 등록기관 시스템 간 연계가 가능하며 ISNI 신청기능과 ISNI 조회기능을 제공한다. ISNI의 발급 신청은 한 건씩 신청이 가능한 AtomPub API 방식과 대용량 신청을 위해 엑셀을 이용한 Bulk 신청방식을 선택하여 처리할 수 있다. 또한 SRU를 통해 ISNI에

등록된 데이터의 검색서비스를 제공한다. 이것을 정리하면 <표 4>와 같다.

ISNI 발급을 신청할 때 AtomPub 스키마는 ISNI 발급을 위한 신규 신청뿐만 아니라 유사중복을 규명하기 위한 재신청, 병합 신청을 할 수 있는 기능이 있다.

Bulk 신청은 연구자의 정제된 정보를 한 번에 대량으로 신청할 수 있는 장점이 있으나 테스트 DB(Aaccept DB)로 신청하여 결과를 확인한 후 서비스DB(Production DB)로 재신청해야 하는 번거로움이 있다.

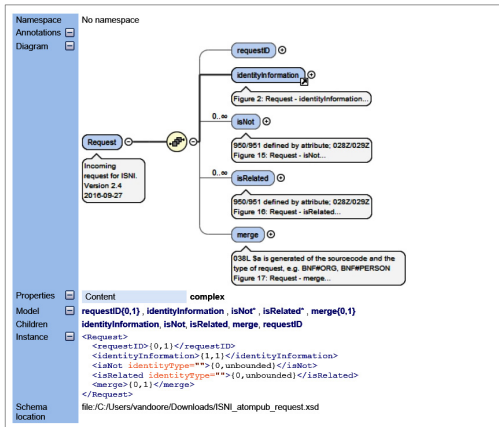
조회할 때 사용하는 SRU는 전통적인 Z39.50의 이론적 개념은 유지하면서 웹 기반 기술을 이용한 프로토콜로서 ZING(Z39.50 International Next Generation) 사업에 포함되어 있으며 현재 가장 많이 사용되는 프로토콜이다.

ISNI-IA는 저장된 ISNI 정보를 SRU 프로토콜을 이용하여 불리언 검색을 포함한 다양한 검색기능으로 제공한다.

<그림 8>은 SRU로 "Artistandmore" 이름을 검색한 결과이며 명령어는 "https://isni-m-acc.org/sru/DB=1.3/username=xxx/password=xxx/?query=pica.na+%3D+Artistandmore*"이다.

<표 4> ISNI 신청 및 조회 방식

구분		설명
신청	AtomPub API 방식	<ul style="list-style-type: none"> • ISNI-IA로 단건 신청시 사용 • AtomPub 프로토콜을 사용하며 포맷은 XML이며 스키마는 <그림 7> 참조
	Bulk 방식	<ul style="list-style-type: none"> • ISNI-IA로 복수건 신청시 사용 • 신청포맷은 엑셀 • <표 3>의 개인명 레코드 구조와 단체명 레코드 구조에 맞춰 엑셀파일을 만들어 FTP로 신청
조회	SRU(Search/Retrieve via URL) 방식	<ul style="list-style-type: none"> • ISNI-IA의 창작자 정보를 조회시 사용 • SRU 프로토콜을 사용하며 포맷은 XML이며 스키마는 <그림 8> 참조



〈그림 7〉 AtomPub API Request 스키마

```
<ZING:searchRetrieveResponse xmlns:ZING="urn:z3950:ZING:Service"
xmlns:dc="http://purl.org/dc/elements/1.1/"
<script type="text/javascript" charset="utf-8" id="zm-extension"/>
<ZING:resultSetId/>
<ZING:numberOfRecords>1</ZING:numberOfRecords>
<ZING:records>
  <ZING:record>
    <ZING:recordSchema>pica</ZING:recordSchema>
    <ZING:recordData>
      <collection xmlns="http://www.loc.gov/MARC21/slim">
        <record>...</record>
      </collection>
    </ZING:recordData>
  </ZING:record>
</ZING:records>
<ZING:diagnostics>
  <ZING:code>OK</ZING:code>
  <ZING:details/>
</ZING:diagnostics>
<ZING:echoedRequest>
  <query>pica.na*330*Artistandmore*</query>
  <maximumRecords>10</maximumRecords>
  <startRecord>1</startRecord>
  <sortKeys/>
</ZING:echoedRequest>
</ZING:searchRetrieveResponse>
```

〈그림 8〉 SRU 검색 결과 스키마

4. 결론

이 연구의 목적은 국회도서관과 같이 방대한 자료의 수집과 서비스를 제공하는 국가도서관에서 연구자 정보의 관리와 협력 기관과의 공유를 위하여 ISNI를 중심으로 한 다양한 식별자를 통합·관리할 수 있는 4가지 계층으로 세분화한 연구자 식별자의 통합관리체계 모형을 제시함으로써 국내 연구자의 식별자의 중복발급을 방지하고 자관의 연구자 정보를 정비하거나 저작물을 식별할 때 효과적으로 활용하여 융합·분석적 링크드 데이터의 구축 기반 마련에도 기여할 것이다.

이를 위해 먼저 도서관 분야뿐만 아니라 다양한 분야의 창작자를 식별할 수 있고 상호운영성을 가진 ISNI의 데이터 구조와 발급에 관한 사항을 살펴보았다.

ISNI는 2012년 ISO 표준(ISO 27729)으로 제정되어 VIAF나 ORCID와 협력을 통해 저작물의 창작, 공연, 제작, 관리, 유통에 관련되는 자연인과 법인(기업, 단체), 가상의 인물을 대상으로

고유 식별번호를 부여하여 저작권의 관리, 유통 및 수익분배 등에 정확하게 식별할 수 있는 기능을 제공한다.

ISNI를 발급받기 위해 ISNI-IA에서 제시한 메타데이터 표준을 준수해야 하며 성과, 이름을 분리하거나, ISO 3166-1의 국가코드 사용, 단체명의 통·폐합 또는 분리 문제 등을 유의해야 한다.

이렇게 발급된 ISNI를 중심으로 연구자 식별자를 통합관리하기 위하여 총 4개의 계층으로 구성된 모델을 제시하였다. 이 구조는 아래부터 데이터구조 계층, 플랫폼 계층, 어플리케이션 계층으로 구성하고, 모든 계층과 관련 있는 관리 및 보안 계층이 있다.

이 모델에서 가장 기본적 토대는 플랫폼 계층의 통합ID 관리기가 처리하는 ID체계이다. 유연하고 체계적인 ID관리를 통하여 여러 기관에서 관리하고 있는 다양한 식별자를 통합하고, 타 기관에서 관리되는 정보를 수집·통합하며 ISNI에서 갱신되는 최신의 정보를 주기적으로 동기화하여 연구자의 기본 정보와 저작물 정보

의 최신성을 유지할 수 있다.

앞에서 언급한 4개의 계층 중에 서비스와 관련된 어플리케이션 계층에는 '연구자 식별자 리졸버', '중복검사 모듈', '연구자정보 관리기', 'ISNI-IA 신청 처리기' 기능을 구성한다.

'연구자 식별자 리졸버'는 자관에서 관리하는 연구자 식별자를 이용하여 타 기관의 식별자를 질의하거나 타 기관에서 관리하고 있는 연구자 상세정보로 링크하기 위한 서비스를 제공한다. '중복검사 모듈'은 외부에서 입수되는 연구자 정보를 기존의 연구자와 일치여부를 확인하여 중의성을 해소하고 신규레코드를 생성할 것인지 또는 기존 연구자의 정보와 통합할 것인지 판단한다. '연구자정보 관리기'는 처리상태별로 구분한 '정보수집 단계'와 '검증 단계' 그리고 '승인 단계'를 통해 연구자 정보 DB의 품질을 높일 수 있도록 관리 기능을 제공한다.

'ISNI-IA 신청 처리기'는 ISNI-IA로 연계할 수 있는 표준 API 연계기능을 가지고 있으며 ISNI 발급신청을 위한 AtomPub API와 ISNI의 연구자 정보를 조회할 수 있는 SRU 기능이

담재된 검색서비스를 제공한다.

이러한 연구자 식별자 통합관리체계 모델은 한 기관에서 저자나 연구자를 관리하는 하나의 정보시스템이 아닌 여러 기관에서 발생된 다양한 식별체계를 통합관리하기 위한 모델로 국가 핵심자원인 국내·외에 활동하는 연구자의 정보 관리에 소비되는 국가적 비용을 절감하고 이용자들에게 정확한 정보검색과 연구자 간의 지적 관계를 밝혀낼 수 있는 기반이 되며 새로운 연구 활동을 지원해 줄 수 있는 기초가 될 수 있을 것이다.

이 연구에서 제시한 연구자 식별자의 통합관리체계 모형은 연구자 정보의 수집·이용·제공 등 서비스에 있어서 「개인정보 보호법」의 준수가 전제되어야 하며 2018년 5월 25일부터 시행되어 모든 EU(European Union)의 회원국에 적용되는 GDPR(General Data Protection Regulation)³⁾과 관련하여 ISNI-IA의 데이터 관리와 서비스 정책의 변화가 예측되므로 그에 따른 법률적 후속연구가 필요할 것이다.

참 고 문 헌

강인수, 이승우, 정한민, 김평, 구회관, 이미경, 성원경, 박동인. 2008. 저자 식별을 위한 자질 비교. 『한국콘텐츠학회논문지』, 8(2): 41-47.
 김송이, 정연경. 2015. 한국인 저자 전거에 관한 국가간 비교 연구. 『한국문헌정보학회지』, 49(1): 379-403.
 김은정, 노경란. 2017. 국가 차원의 ORCID 기반 저자 식별자 활용에 관한 연구. 『한국비블리아학회지』, 28(3): 151-174.

3) GDPR은 유럽연합에서 유럽연합 시민의 개인정보를 보호하기 위하여 2016년에 제정된 기본법으로서 별도의 내국법 편입과정 없이 그 자체로 모든 유럽연합에 회원국에 적용된다.

- 이미화. 2014. 전거제어를 위한 국제표준이름식별자(ISNI)의 활용가능성에 관한 연구. 『정보관리학회지』, 31(3): 133-151.
- 이석형. 2014. 학술논문 저자 소속기관 식별데이터 구축에 관한 연구. 『사회과학연구』, 25(4): 391-410.
- 이성숙, 박지영, 이혜원. 2017. 링크드 데이터에서 인물 정보의 식별 및 연계 범위 확장에 관한 연구. 『정보관리학회지』, 34(3): 7-21.
- 조재인. 2013. ORCID 기반의 학술 연구 결과물 저자명 식별 시스템 구축 방안에 관한 연구. 『한국비블리아학회지』, 24(1): 45-62.
- Angjeli, A., A. Mac Ewan, and V. Boulet. 2014. "ISNI and VIAF - Transforming ways of trustfully consolidating identities."
- Haak, Laure. 2014. [online]. [cited 2018.7.4] <<https://orcid.org/document/orcid-isni-mou>>.
- Hui Han, Hongyuan Zha, C. and Lee Giles. 2003. "A model-based k-means algorithm for name disambiguation." In Proceedings of Semantic Web Technologies for Searching and Restrieving Scientific Data, Oct. 20, Florida:USA.
- ISO. 2012. *ISO 27729:2012: Information and documentation - International standard name identifier (ISNI)*, Switzerland: ISO.
- Krämer, T., F. Momeni, and P. Mayr. 2017. "Coverage of Author Identifiers in Web of Science and Scopus." *CoRR*, abs/1703.01319.
- MacEwan, Andrew. 2017. [online]. [cited 2018.8.8]. <<http://www.isni.org/content/isni-viaf-summit-2017>>.
- Meadows, Alice. 2016. "Everything You Ever Wanted to Know About ORCID." *College and Research Libraries News*. January 2016. 77(1):23-26 and 30.
- Rickenmann, Barbara. 2018. "ORCID-ISNI / DEMO PCC Pilot Group". [online]. [cited 2018.8.2]. <<https://wiki.duraspace.org/pages/viewpage.action?pageId=90977102>>.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Cho, Jane. 2013. "A Study on the Construction Methods for Author Identification System of Research Outcome Based on ORCID." *Journal of the Korean Biblia Society for Library and Information Science*, 24(1): 45-62.
- Kang, In-Su, Seungwoo Lee, Hanmin Jung, Pyung Kim, Heekwan Koo, Mi-Kyung Lee, Won-Kyung Sung, and Dong-In Park. 2008. "Features for Author Disambiguation." *Journal of the Korean Contents Association*, 8(2): 41-47.

- Kim, Eun-Jeong and Kyung-Ran Noh. 2017. "A Study on Utilization of ORCID based Author Identifier at National Level." *Journal of the Korean Biblia Society for Library and Information Science*, 28(3): 151-174.
- Kim, Song-le and Keon Kyoung Chung. 2015. "A Comparative Study on Authority Records for Korean Writers Among Countries." *Journal of the Korean Society For Library and Information Science*, 49(1): 379-403.
- Lee, Mihwa. 2014. "A Study on the Applicability of ISNI for Authority Control." *Journal of the Korea Society for Information Management*, 31(3): 133-151.
- Lee, Seok Hyoung. 2014. "A Study on the Construction of Identified Data of Author's Affiliation in Academic Papers." *Journal of Social Science*, 25(4): 391-410.
- Lee, Sungsook, Ziyong Park, and Hyewon Lee. 2017. "Expanding the Scope of Identifying and Linking of Personal Information in Linked Data: Focusing on the Linked Data of National Library of Korea." *Journal of the Korea Society for Information Management*, 34(3): 7-21.

