

기록시소러스 구축지침 개정에 관한 연구*

A Study on the Revision Archival Thesaurus Construction

박 지 영 (Zi-young Park)**

윤 소 영 (SoYoung Yoon)***

이 혜 원 (Hyewon Lee)****

목 차

- | | |
|------------------|---------------------------|
| 1. 서 론 | 4. 기록 시소러스 지침 개선을 위한 고려사항 |
| 2. 이론적 배경 | 5. 결 론 |
| 3. 시소러스 활용시스템 사례 | |

〈초 록〉

시소러스는 분류체계의 분류명을 제어해 주고, 분류표의 색인 역할을 수행할 수 있다. 기록관리에 있어서 시소러스는 기능어를 제어하고, 검색의 접근점을 확장시켜 기능분류체계를 보완해 준다. 그런데 최근에는 정보환경의 변화와 시소러스 구축 및 활용 시스템의 발전으로 시소러스 구축에 관한 국제표준인 ISO 25964가 개정되었다. 또한 기록관리 전반에 대한 국제표준인 ISO 15489의 제1부도 2016년에 개정되었다. 또한 영국의 IPSV나 유럽의 EuroVoc, 뉴질랜드의 FONZ 등은 최근의 경향을 반영하여 효과적으로 시소러스를 구축하고 연계하고 있었다. 이에 본 연구에서는 체계적인 기록관리를 위한 시소러스 구축지침을 관련 표준과 사례 측면에서 살펴보고, 현재의 우리나라 기록관리 분야 시소러스 구축지침의 개선 방안을 제안하였다.

주제어: 시소러스, 기능 시소러스, 주제 시소러스, 시소러스 상호운용성, ISO 15489-1:2016, ISO 25964

〈ABSTRACT〉

The thesaurus can control the headings in a classification scheme and can serve as an index of the classification scheme itself. In records management, the thesaurus controls functional terms and expands the access point of search to complement the functional classification scheme. In recent years, ISO 25964, the international standard for thesaurus construction, has been revised because of changes in the information environment and the development of thesaurus construction and utilization systems. Part 1 of ISO 15489, the international standard for overall records management, was also amended in 2016. In addition, the Integrated Public Sector Vocabulary (IPSV) in the UK, EuroVoc in Europe, and Functions of New Zealand (FONZ) in New Zealand have been effectively building and linking thesauri to reflect recent trends. In this study, we propose a thesaurus construction guideline for systematic record management in terms of related standards and cases, and suggest an improvement plan for the thesaurus construction guideline in Korea.

Keywords: thesaurus, functional thesaurus, subject thesaurus, thesaurus interoperability, ISO 15489-1:2016, ISO 25964

* 본 연구는 한성대학교 교내학술연구비 지원 과제임.

** 한성대학교 인문대학 응용 인문학부 문헌정보전공(zgpark@hansung.ac.kr) (제1저자)

*** 한국고전번역원 특수고전번역실(syoon@itkc.or.kr) (공동저자)

**** 서울여자대학교 사회과학대학 문헌정보학과(hwlee@swu.ac.kr) (교신저자)

■ 접수일: 2017년 1월 22일 ■ 초심사일: 2017년 2월 7일 ■ 게재확정일: 2017년 2월 24일

■ 한국기록관리학회지 17(1), 117-141, 2017. <<http://dx.doi.org/10.14404/JKSARM.2017.17.1.117>>

1. 서론

과거에 시소러스는 종이에 인쇄하여 발행했고, 색인과 탐색에 관한 교육을 받은 전문가들이 활용하는 도구였다. 그런데 최근에는 대부분의 시소러스가 전자적으로 발행되고 있어 시소러스를 탑재하고 활용할 수 있도록 지원하는 소프트웨어가 반드시 필요하다. 또한 시소러스의 이용자도 전문가에서 일반 이용자로 확대되었다. 이제는 시소러스를 구축할 때, 시소러스가 탑재된 시스템에서 제공하는 탐색기능이 먼저 정의되어 있어야 한다. 또한 일반 이용자가 정보탐색 과정에서 직관적으로 시소러스를 이해하고 활용할 수 있어야 한다.

그리고 이와 같이 변화하는 시소러스의 이용 방식은 시소러스 구축을 위한 표준에도 반영되어야 할 것이다. 현재 우리나라의 기록 시소러스 지침으로는 'NAK/A 8:2009(v1.1) 국가기록원 시소러스 지침'이 있다(국가기록원, 2009). 국가기록원의 시소러스 지침은 시소러스 구축을 위한 국제표준인 "ISO 2788:1986: Guidelines for the establishment and development of monolingual thesauri"와 기록관리 국제표준인 "ISO 15489: 2001 - Information and documentation - Records management"를 번역하여 국가표준으로 제정한 KS X 15489:2007을 인용표준으로 밝히고 있다. 그런데 ISO 2788은 지난 2011년에 ISO 25964로 개정되었으며, ISO 15489도 2016년에 제1부가 개정되었다. 국제표준도 최근의 변화된 정보환경을 반영한 내용을 추가하였기 때문이다. ISO 25964는 상이한 어휘 제어 시스템 간의 상호운용성을 확보하기 위한 내용이 중점적으로 추가되었다(ISO, 2011). ISO 15489

도 기술 환경의 변화와 ISO 30300(기록경영시스템)과의 연계와 관련된 내용 등이 추가되었다. 이와 같이 새로운 표준이 제정되면 기존 표준은 폐기될 것이다.

이에 본 연구에서는 개정 전의 표준을 바탕으로 작성된 국가기록원의 시소러스 구축지침도 새롭게 검토하여, 추가나 수정이 필요한 내용을 도출하고자 한다. 이를 위한 연구방법으로 문헌 연구와 사례 연구를 병행하였다. 문헌연구로는 우선 최근에 개정된 기록관리 국제표준인 ISO 15489-2:2016을 시소러스 및 색인과 관련된 부분을 중심으로 분석하였다. 그리고 역시 개정된 시소러스 구축 표준인 ISO 15964-1:2011과 ISO 25964:2013을 분석하였는데, 개정 버전에서 새롭게 추가된 부분과 기록관리 분야의 시소러스에 대한 지침을 중점적으로 살펴보았다. 그리고 사례 연구로는 최근의 정보환경과 개정된 국제표준의 경향에 부합하는 공공분야의 시소러스를 중심으로 조사하였다. 조사 대상으로는 영국의 IPSV를 비롯한 관련 공공영역의 제어어휘 목록과 유럽의 EuroVoc, 뉴질랜드의 공공부분 시소러스인 FONZ를 선정하였다.

2. 이론적 배경

2.1 기록관리 국제표준

2.1.1 기록관리 국제표준의 개정

기록관리 일반 원칙과 지침을 제시한 국제표준이며 우리나라의 국가표준으로도 제정된 ISO 15489의 1부가 지난 2016년에 개정되었다(ISO, 2016). 개정된 1부는 제목이 '기록관리-일반'에

서 '기록관리-개념과 원칙'으로 변경되었는데, 개정작업이 진행 중인 2부도 '기록관리-지침'에서 '기록관리를 위한 평가 표준'(Standard on Appraisal for Managing Records)과 '기록을 위한 시스템 설계 표준'(Standard on System Design for Records)으로 이어질 전망이다 (Bustelo, 2016). 기존의 표준인 2001년 판 ISO 15489는 호주의 기록관리 표준이었던 AS 4390을 바탕으로 제정되었는데, ISO 15489의 개정 과정에서는 도출된 기존 표준의 한계는 다음과 같다(ISO, 2001; Reed, 2013).

- 기록관리(recordkeeping)에 대한 통일된 관점의 결여
- 암묵적으로 문헌(document)과 기록(records)을 구분함
- 기록관리에 대한 상이한 개념적 뷰를 피상적으로 표현함
- 평가를 지속적인 절차로 간주하지 않음

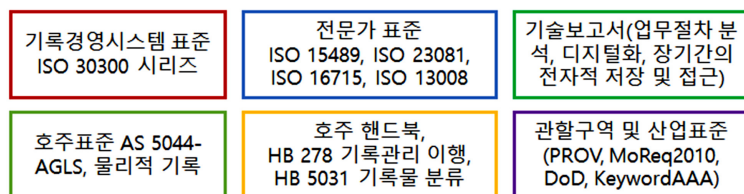
ISO 15489의 개정과정에서는 우선적으로 기존의 표준을 검토하는 작업이 진행되었다. 검토 대상인 표준은 <그림 1>과 같은데, 이러한 표준은 앞으로 ISO 15489와 연계되어 기록관리 시스템 전반을 지원할 것이다.

기록관리 환경과 ISO 15489 개정의 고려사항을 연계하면, 기록의 유형 변화와 기록이 생산되

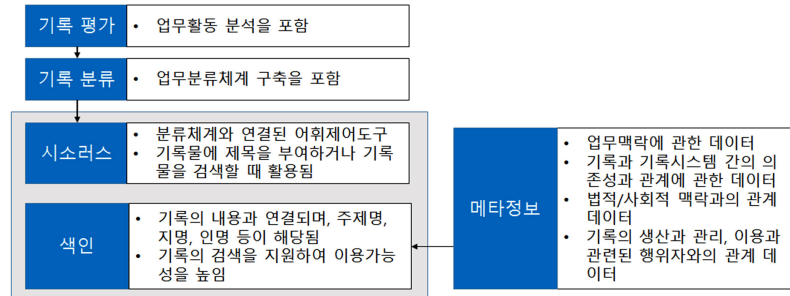
는 시스템의 다양화를 들 수 있다. 디지털로 생산되는 기록이 대량으로 증가하고, 기록이 생산되는 시스템 자체가 전자문서관리시스템(Electronic document and records management system, EDRMS)에만 한정되지 않기 때문이다. 특히 EDRMS 이외의 기록생산시스템에는 웹을 플랫폼으로 활용하는 유형의 증가가 컸다(Reed, 2013).

2.1.2 ISO 15489 개정판의 특징

개정된 ISO 15489 제1부를 보면, '기록'은 업무활동(business activity)의 증거임과 동시에 정보자산(information assets)이 된다(ISO, 2016, p. v). '기록'을 '기록'이 아닌 유형의 정보와 명확히 구분하지 않는 것이다. 또한 개정된 ISO 15489에서는 '평가'(appraisal)에 관한 지침이 강화되었다. 용어의 의미도 기존의 '평가'에서 업무 맥락과 업무 활동의 분석까지 포함하도록 확장되었다. ISO 15489의 개정 과정에서 도출한 기존 ISO 15489의 한계 중에서 '평가'를 지속적인 과정으로 간주하지 않았다는 지적이 있었다. 평가 개념의 확장은 이에 대한 개선 결과이며, 앞으로는 생산하고 획득해야 할 기록을 확정하고 장기간에 걸친 적절한 기록관리 방안을 결정하는 것도 평가의 영역이 될 것이다. 따라서 기록물 평가를 위해서는 기록물을 생산하는 기관과 해당 기관의 업무활동에 대해 이해해야 할 것이다.



<그림 1> ISO 15489 개정 과정에서 참고한 관련 표준(Reed, 2013)



〈그림 2〉 ISO 15489에 규정된 제어어휘의 유형과 활용

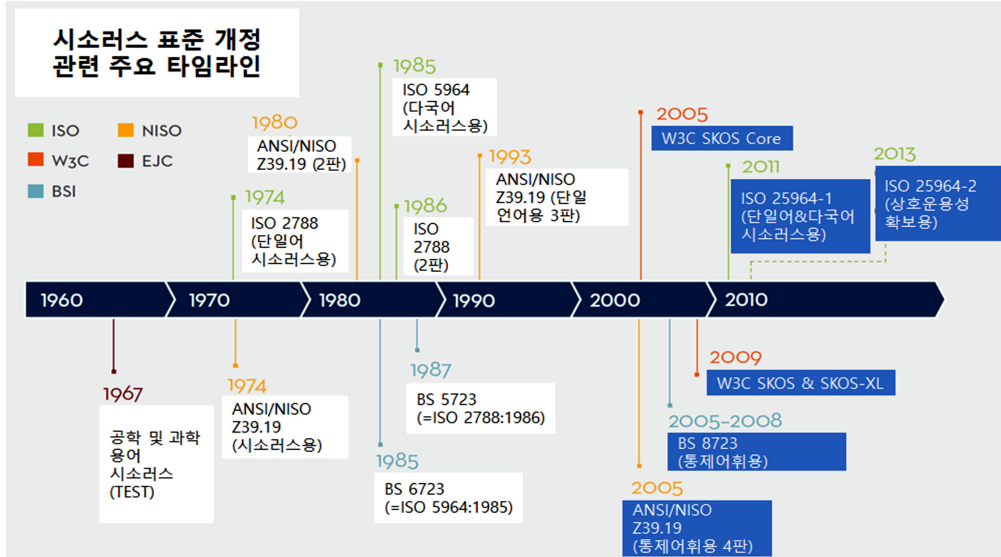
업무활동에 대한 이해는 업무분류체계(Business classification schemes) 구축의 바탕이 되며, 기능, 활동, 업무절차를 분석하여 구축한 업무분류체계는 기록물에 적용된다(ISO, 2016, p. 11, 14). 그리고 업무분류체계는 다시 시소러스와 같은 어휘제어도구와 연계된다. 이 점에서 기록 시소러스는 업무분류체계를 지원하여, 기록물의 제목을 정하거나 기록물을 검색하는 데 활용된다(ISO, 2016, p. 14).

ISO 15489에서는 업무분류체계를 바탕으로 구축된 시소러스 외에도 색인(indexing)에 대한 지침을 제시하고 있는데, 색인은 메타데이터의 요소로서 기록의 검색을 지원함으로써 기록의 이용가능성을 높일 수 있다. 시소러스가 업무분류체계와 연계되는 것과 대조적으로 색인 메타데이터의 유형으로는 주제명, 지명, 인명 등이 제시되어 있으며, 기록의 획득 시점에서부터 부여될 수 있고 필요에 따라 그 후에도 추가할 수 있다(ISO, 2016, p. 17). ISO 15489에 제시된 시소러스와 색인의 특징을 정리하면 〈그림 2〉와 같다. 즉, 시소러스는 확장된 기록평가의 결과 중 하나인 기록분류체계와 연계되는 도구이며, 색인은 기록의 메타 정보 중에서 기록물 내용과 연계되는 도구로 제시되어 있다.

2.2 시소러스 관련 국제표준

2.2.1 ISO 25964의 개요

시소러스의 기본 개념과 구축 원칙은 단일어 시소러스 지침인 ISO 2788과 다국어 시소러스 지침인 ISO 5964로 구분되어 있었다. 그리고 국내외의 많은 시소러스가 이 표준을 기초로 구축되었으며, 2009년에 발표된 우리나라의 국가기록원 시소러스 구축지침도 ISO 2788을 인용표준으로 삼고 있다. 그런데 국제표준기구(International Organization for Standardization, ISO)에서는 지난 2011년과 2013년에 시소러스 구축 표준을 개정하였다. Clarke와 Zeng(2013)에 따르면, ISO 2788을 기준으로 삼은 시소러스 표준 체계와 ISO 25964로 변경된 시소러스 체계의 구분은 2005년부터 나타났다. 〈그림 3〉과 같이 2005년 이후에는 ANSI/NISO Z39.19 제4판과 BS 8723이 발표되었으며, W3C의 SKOS가 발표되었다. ANSI/NISO Z39.19 제4판에서는 기본 개념과 원칙을 유지하면서도 인쇄물 중심의 가이드라인이 제공되었던 제3판에 비해 새로운 정보의 저장 매체와 검색 방법을 반영할 수 있도록 확장된 내용이 추가되었다(ANSI/NISO, 2010, pp. viii-ix). BS 8723은 데이터 모형이나



〈그림 3〉 ISO 25964 관련 주요 시소러스 관련 표준(Clarke & Zeng, 2013)

SKOS 적용 등 ISO 25964와의 연관성이 매우 깊다(Clarke, Will, & Cochard, 2008; NISO, 2017). ISO 25964의 개발과 관련된 표준의 특징으로는 디지털 정보환경을 반영하고, 상이한 시소러스 간의 상호운용성을 확보하기 위한 데이터 모형 설계나 SKOS 기반 연계 체계를 구축했다는 점이 있다.

2.2.2 ISO 25964의 주요 특징

개정된 시소러스 표준인 ISO 25964는 총 2부로 구성되어 있다. 제1부는 기존의 시소러스 구축 표준이었던 ISO 2788:1986과 ISO 5964:1985의 내용을 대체하면서 새로운 내용을 일부 추가하였으며, 제2부는 전체 내용이 새롭게 추가된 부분에 해당된다. 즉, ISO 25964의 1부에서는 시소러스의 형식과 데이터 교환 프로토콜을 포함한 개발과 관리를 포함하며, 단일언어와 다국어 시소러스에 관한 지침을 모두 다루고, 2부에서는 분류체계나 이름전거과일, 온톨로지와 같

은 유사한 다른 제어어휘 표준과 시소러스의 상호운용성에 관한 지침을 제공한다(ISO, 2011). 제1부를 기존 표준과 비교하면 1장부터 13장까지는 기존의 시소러스 구축 표준을 종합한 것이고, 14장부터 18장까지는 시소러스 구축을 위한 S/W와 데이터 모형에 관한 지침을 추가한 새로운 부분에 해당된다. 제2부도 두 가지로 구분되어 있는데, 1장부터 6장까지는 상이한 시소러스 간의 맵핑 유형과 방법에 대해 다루고 있고, 17장부터 24장까지는 제어어휘 목록을 유형별로 정리하였다. 기존의 표준과 개정된 표준을 비교하여 정리하면 〈그림 4〉와 같다(ISO, 2011, 2013).

1) 시소러스 구축 S/W 및 데이터 모형 측면의 특징

2011년에 개정된 ISO 25964의 제1부에서는 시소러스의 구축이나 활용을 시스템을 통해 수행하고, 복수의 시소러스 간에 상호운용성을

새로 추가된 내용		
ISO 2788: 1986	ISO 25964-1: 2011	ISO 25964-2: 2013
ISO 5964: 1985	1-13 장 원칙과 구축절차	1-16 장 맵핑 유형 및 방식
기존 표준에 해당되는 내용	14-18 장 구축S/W 및 데이터 모형	17-24 장 유형별의 제어 어휘목록

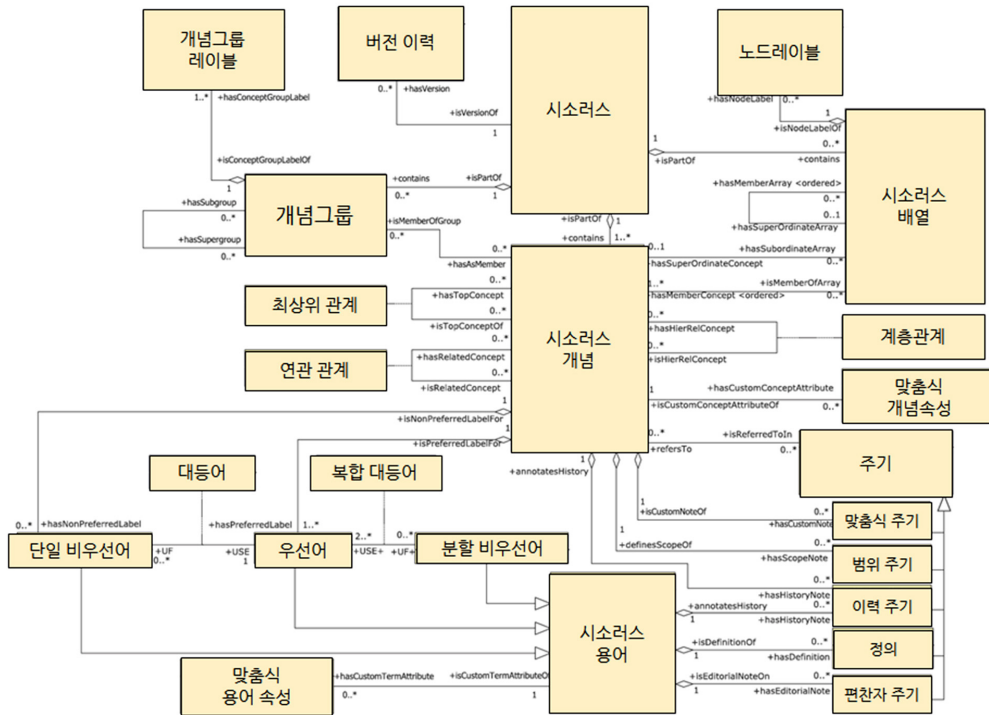
〈그림 4〉 ISO 2788/5964와 ISO 25964의 구조적 특징 비교

확보하기 위한 부분이 새로 추가되었다. 시소러스 통합을 위한 데이터 모형에 대한 연구는 1990년대로 거슬러 올라간다. Soergel(1996)은 복수의 시소러스를 하나의 데이터베이스로 통합하여 저장하기 위한 두 가지 모형을 제안하였다. 첫 번째는 용어기반 통합모형으로서 모든 관계를 용어간 관계로 저장하는 방식이다. 이 방식은 유연성이 확보되나 중복 데이터가 저장될 수 있다. 동일 개념을 2개 이상의 용어로 표현한다면, 용어 DB 내에 중복성이 생기기 때문이다. 두 번째는 개념기반 통합모형으로서 개념을 개념번호로 식별하고, 개념간 관계를 저장한다. 이 방식은 동일 개념에 대한 모든 용어를 그룹화할 수 있다는 장점이 있으나, 유연성이 부족할 수 있다. 두 모형의 차이는 관계가 용어 간에 존재하는지와 개념 간에 존재하는지에 대한 구분이다.

ISO 25964에서도 데이터 모형의 중요성과 상세성이 이전 버전보다 강조되어 있다. 기존의 표준에 제시된 데이터 모형은 비교적 단순하다. 용어와 용어관계(USE/UF, BT/NT, RT)가 하나의 엔트리로 묶여 있고, 개별 엔트리가 자모순과 같은 특정 순차에 따라 배열되는 구조였다. ISO 25964에서는 기본 구조를 계승하면서도 부가적인 기능을 담을 수 있는 데이터 모형

을 제안하고 있다. 이 표준에서는 시소러스를 일종의 지식조직시스템(Knowledge Organization System, KOS)로 간주하고, 특정 KOS를 기계처리가 가능한 형태로 발행하며 시맨틱웹 환경에 적용 가능하도록 설계하기 위해 도입하고자 하였다. 그리고 이를 위해 적용한 구조가 SKOS(Simple Knowledge Organization System)이다(Will, 2012). ISO 25964의 데이터 모형은 통합 모델링 언어(unified modelling language, UML)로 표현할 수 있는데, 〈그림 5〉와 같이 시소러스의 관계는 시소러스 개념과 연계되어 있으며, 시소러스 개념은 다시 시소러스 용어에 연결되어 있다.

2) 시소러스 간 맵핑 유형 및 방식 상의 특징
 시소러스의 구조화 방식은 용어를 구성하는 문자열이 아니라 용어로 표현된 개념 간을 연결하는 것이다. 관습적으로 상위어(broader term)나 하위어(narrower term)와 같이 용어관계로 시소러스 내의 관계를 표시하지만, 실제로는 상위개념(broader concept)과 하위개념(narrower concept)을 나타내는 것이다(Bill, 2012, p. 49). 이와 같이 시소러스 구축 과정에서는 개념의 식별과 연계가 핵심이 되는데, ISO 25964에서는 개념 간의 관계가 이전 표준보다 정교해졌



〈그림 5〉 ISO 25964의 데이터 모형

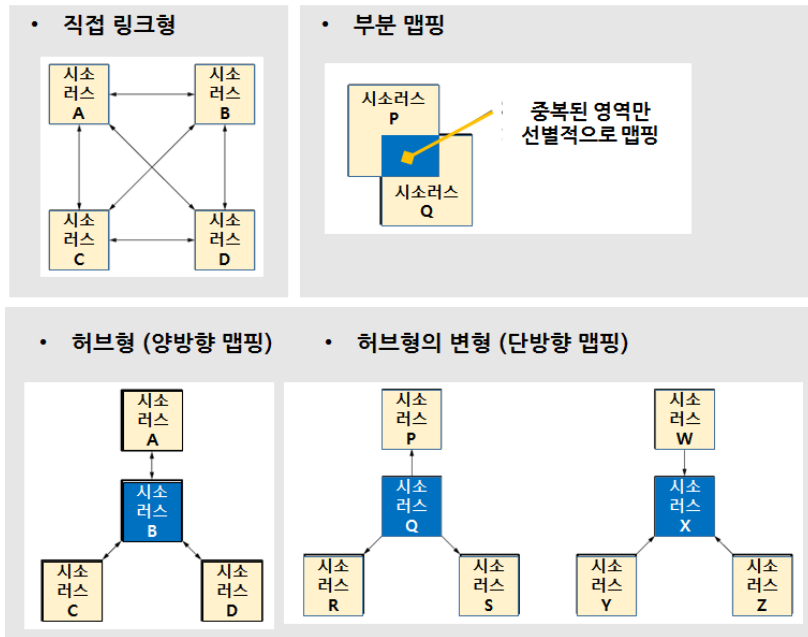
다. ISO 25964에서는 우선어/비우선어 관계에도 '역할'(role)이라는 속성을 추가할 있도록 규정하고 있는데, 즉, 약어/완전명, 공식/비공식 표현, 퇴화한/현행의, 학명/속명 등과 같은 역할어를 넣어 대응관계를 상세히 나타낼 수 있다. 역할어가 일종의 패시 기능을 수행하는 것이다. 더불어 ISO 25964에서는 '복합 대응관계'(compound equivalence)라는 장치를 도입하여 복합어의 대응관계도 표현할 수 있게 하였다. 예를 들면, 단일 개념어를 바탕으로 구축되는 시소러스에서는 일반적으로 '석탄 채굴'(coal mining)과 같은 복합어가 디스크립터로 등재되지 않는다. 따라서 '석탄 채굴'이라는 용어는 〈표 1〉과 같이 '석탄'이나 '채굴'로 연계되어야 하며, 그 반대로는 '석탄'이라는 디스크립터의

비대표어로 '석탄 채굴'이 존재할 수 있다. 이를 일종의 복합 대응관계라고 하는 것이다.

또한 ISO 25964에서는 상이한 시소러스 및 관련 제어어휘 도구에 대한 상세한 맵핑 지침을 제공하고 있다. ISO 25964에서 제시한 맵핑 방식은 SKOS의 관계 어휘를 이용한 것으로서, 복수의 시소러스에 존재하는 유사한 용어를 1:1로 연계하는 기존의 방식보다 명확하다. ISO 25964의 2부에서 맵핑 모형을 유형별로 구분하여 제시하고 있는데, 〈그림 6〉과 같다. 맵핑 유형 중에서 직접 링크형은 둘 이상의 제어어휘 도구가 매우 상이한 구조로 이루어진 경우에 사용하는데, 예를 들면 분류표와 주제명표를 연계하는 경우가 해당될 수 있다. 반면에 허브형 맵핑은 하나의 제어어휘도구가 다른 도구를 모두 포함할 정도

<표 1> 시소러스 디스크립터에서의 복합어의 처리방식 비교(Will, 2012, p. 49를 재구성)

'석탄 채굴'(coal mining) USE+ 석탄(coal) USE+ 채굴(mining) → '석탄 채굴'이라는 복합어가 시소러스 디스크립터로 등재되지 않으므로, USE는 '석탄'이나 '채굴'과 같은 단일개념어에만 해당되고, '석탄 채굴'은 비디스크립터로만 존재함	석탄(coal) UF+ 석탄 채굴(coal mining)
① '화석 연료'(fossil fuels) USE+ 석탄(coal) USE+ 천연가스(natural gas) USE+ 석유(petroleum) → ①은 '석탄 채굴'과 마찬가지로, '화석 연료'가 디스크립터로 존재하지 않는 경우임. 실제 시소러스를 구축하여 활용할 때는 ②와 같이 복합어인 '화석 연료'를 디스크립터로 지정하고, 하위 개념관계를 활용하는 것을 권장함	② '화석 연료'(fossil fuels) NT 석탄(coal) NT 천연가스(natural gas) NT 석유(petroleum)



<그림 6> ISO 25964의 제어어휘 간 맵핑 유형 비교

로 광범위할 때 적용할 수 있다. 가운데는 '허브'가 되고, 허브의 주변은 '위성'의 역할을 수행한다. 허브형 맵핑 유형은 단방향으로도 변형될 수 있는데, 이 경우에는 위성이 되는 제어어휘 도구간의 상호운용성에 제한이 있다. 이 외에도 맵핑이 가능한 부분만을 선별적으로 연계하는 부분 맵핑 방식이 있다(ISO, 2013, pp. 17-19).

2.3 기록 시소러스 구축지침

2.3.1 기록관리를 위한 분류체계와 시소러스

ISO 25964의 제2부에서는 상이한 제어어휘 도구 간의 맵핑 방식과 함께 분야별 시소러스에 대한 특징을 제시하였다. 이 중에서 기록관리 영역은 제18장에 해당되는데, 이 장의 제목인

‘기록관리용 분류체계’(Classification schemes used for records management)에서는 기능분류체계를 기본적인 기록분류의 기준으로 삼고, 기록물의 제목을 명명하거나 검색에 필요한 어휘제어를 위해 시소러스를 연계할 것을 제안하였기 때문이다(ISO, 2016).

ISO 25964에 따르면 각 기관은 자관의 공문서 관리를 위해 별도로 개발한 분류체계를 사용하는데, 이는 파일 플랜(file plan)이나 업무분류체계(business classification scheme)에 해당된다. 그리고 이와 같은 체계는 문헌의 도서관에서 사용하는 주제 기반 분류체계와 구분되어야 한다. 기록분류에서 강조하는 것은 ‘업무 맥락’(business context)이며, 이 점에서 업무 기능, 활동, 처리행위의 결과물인 기록의 분류 체계가 도출될 수 있다. 물론 ‘주제’라는 개념은 다양한 맥락에서 사용되므로 도서관의 분류체계는 ‘주제’(subject matter)보다는 주로 ‘학문 분야’(discipline)를 기준으로 설계되었다고 간주하는 것이 더 정확하다.

1) 기록물 탐색을 위한 시소러스의 필요성

ISO 25964에서는 기록관리 분야의 분류체계는 일반적으로 브라우징 기능을 제공하며, 기록물의 탐색 역할은 제한적으로만 수행된다고 지적했다(ISO, 2013, pp. 56-57). 브라우징 도구로서 계층형 분류체계는 각 계층집단 간의 수직적인 관계를 제시한다. 예를 들어, 어떤 시리즈 단계를 찾은 이용자는 그 하위의 폴더를 모두 브라우징할 수 있다. 그런데 전자기록관리시스템에서는 이용자가 파일명이나 폴더명을 알아야 해당 명칭을 탐색어로 사용할 수 있을 것이다. 그러나 기록으로서의 파일이나 폴더명은 대부

분 탐색어로 사용되기에 적합하지 않으므로, 특정 주제별로 기록을 탐색할 때는 전문(fulltext) 검색이나 시소러스를 통한 색인을 비롯한 별도의 탐색도구가 필요하다.

2) 전자기록환경에서 파일 및 폴더명의 제어 필요성

대부분 인쇄물 형식으로 기록이 생산될 경우에는 기록이 물리적인 단위의 파일로 묶이고, 계층형 분류체계에 따라 레이블이 부착되어 파일 저장소에 함께 보관되었다. 그런데 컴퓨터 기반 업무환경이 도입되면서 대부분의 기록이 전자적인 형태로 생산되고 있다. 전자기록관리 환경에서는 기존의 종이 기반 환경에서 개발된 계층형 분류체계가 적합하지 않을 수 있다. 또한 전자기록관리시스템에서는 파일이나 폴더명을 분류체계의 항목명과 결합해서 관리할 수 있다. 분류체계의 클래스명에 추가 요소를 부가하거나 식별기호를 부여하여, 분류명을 유일하게 식별할 수 있다.

3) 분류체계와 시소러스의 구성요소 비교

분류체계와 시소러스는 지식의 구조화 과정에서 정보자원의 주제정보(또는 맥락정보)를 개념화하고 이를 색인언어로 표현하는 공통 단계를 지닌다. 그런데 이 두 도구가 어휘를 제어하여 개념구조를 구축하는 방식에는 차이점이 있다. 분류표는 상위수준의 지식을 세분화하는 하향식 방법을 주로 사용하는데, 대부분의 계층형 분류체계가 이 경우에 속한다. 그러나 시소러스는 개별 개념에서부터 상위개념으로 이동하는 상향식 방법을 주로 사용하므로 용어의 특정성이 높다. 따라서 분류체계와 시소러스는 상

〈표 2〉 시소러스와 연계되는 기록 분류체계 구성요소(ISO, 2013, pp. 57-58)

요소명	설명
클래스명 (분류표목)	<ul style="list-style-type: none"> • 분류명이나 레이블이 부여됨. 일부 분류체계에서는 분류명을 '표목'(headings)이나 디스크립터, 키워드, 제목, 레이블이라고도 함. 단일 분류체계에서는 분류명이 중복되지 않아야지만, 중복이 허용되는 분류체계도 존재함. 예를 들어, '예약'이라는 항목은 '회의'나 '전시' 항목의 하위에 중복해서 위치할 수 있음.
식별기호 (분류기호)	<ul style="list-style-type: none"> • 전자기록관리시스템에서는 일반적으로 각 분류명이나 파일(기록물 단위)에 유일한 식별기호를 부여함. 이 식별기호는 분류기호에 해당되는데, 기록관리시스템에서는 식별기호가 분류명뿐 아니라 기록에까지 부여된다는 특징이 있음
기타 요소	<ul style="list-style-type: none"> • 기술정보(description, 시소러스의 범위주기에 해당됨) • 색인어(시소러스 용어와 맵핑됨) • 날짜(파일의 시작일자과 종료일자, 삭제일자) • 접근권한 <ul style="list-style-type: none"> • 처분지침 • 이력주기 <ul style="list-style-type: none"> • 상호참조(도보라 참조)

호보완적인 관계를 지니게 된다(Sager, 1990, pp. 37-39; 박지영, 2010, p. 79). ISO 25964에 제시된 분류체계와 시소러스 구성요소를 살펴 보면 〈표 2〉와 같다.

또한 ISO 25964에서는 다음과 같이 분류체계와 시소러스를 상호보완적으로 활용하기 위한 4가지 방안을 제시하고 있다(ISO, 2013, pp. 58-59).

- ① 시소러스를 분류체계의 색인으로 활용
- ② 시소러스를 기록물의 색인어 부여에 활용
- ③ 시소러스를 기록관리시스템 내의 기록물 검색 기능에 활용
- ④ 시소러스 용어와 분류체계의 분류체계를 맵핑하여, 기록관리시스템에 입력되는 질의어의 변환에 활용

2.3.2 국가기록원 시소러스 구축지침

우리나라에서 제정한 기록관리 분야의 시소러스 지침에는 2009년에 발표된 '국가기록원 시소러스 지침(1.1)'이 있다(국가기록원, 2009). 이 지침의 인용표준에는 'ISO 2788:1986 - 단일언어 시소러스 제정 및 개발 지침'과 'KS X 15489 문헌 정보 - 기록관리 제1부: 일반 및 제2부: 지침'이

있다. 그런데 ISO 2788은 개정되어 지난 2011년과 2013년에 각각 제1부와 제2부가 발표되었다. 그리고 ISO 15489도 2016년에 제1부가 개정되었으며, 추가적인 개정작업이 진행되고 있다.

시소러스 구축에 관한 상세 지침은 크게 일반 주제시소러스에 관한 지침과 기능시소러스에 관한 지침으로 구분된다(국가기록원, 2009). 시소러스 엔트리 구축을 위한 용어사용이나 표기법, 용어 간의 관계지시기호는 일반주제시소러스와 기능시소러스가 유사한 지침을 공유하며, 대표어의 선정이나 용어관계의 설정 기준에서는 일반주제시소러스와 기능시소러스의 지침에 차이점이 있다. 이를 도식화하면 〈그림 7〉과 같다.

공통의 지침이 적용되는 부분은 주로 시소러스 구축에 관한 기본적인 원칙과 방법에 속하므로 ISO 2788과 연계된다. 그런데 2011년과 2013년에 걸쳐 ISO 2788이 ISO 25964로 개정되었으므로, 개정된 시소러스 관련 국제표준의 내용 중에서 우리나라의 시소러스 개선에 필요한 사항을 검토하여 적용하는 방안을 고려해 볼 수 있다. ISO 25964의 주요 개정 내용은 시소러스 관리 시스템의 상호운용성과 상이한 분야의

공통 구분	일반주제시소러스	기능시소러스
용어사용	<ul style="list-style-type: none"> 대표어: 인명, 기관/단체명, 지명, 사건명, 법률명 등 	<ul style="list-style-type: none"> 대표어: 기능어(단위 과제, 단위업무)
표기법	<ul style="list-style-type: none"> 등가관계: 유의어 및 대응 외국어 	<ul style="list-style-type: none"> 등가관계: 정제 전 기능어(UF)
용어관계	<ul style="list-style-type: none"> 계층관계: 의미상 포괄정도에 따른 상하관계 연관관계: 등가 및 계층관계 이외의 연관어 	<ul style="list-style-type: none"> 계층관계: 업무분류 체계의 상위계층어 연관관계: 관련 기능어

〈그림 7〉 국가기록원 시소러스 구축 지침 도식화

시소러스의 연계 방법에 관한 것이었다(ISO, 2011, 2013). 기록관리 분야도 유관 분야와의 협력력이 증진되고 있으며, 기록관리 분야 내에서도 복수의 시소러스가 관리되고 있다. 이와 같은 추세를 반영할 수 있는 방안을 개정된 국제표준에서 도출할 수 있는지 검토해야 할 것이다.

또한 일반주제시소러스에 대한 지침과 기능시소러스에 관한 지침이 달라지는 부분은 ISO 15489에 제시된 기능분류체계를 바탕으로 하고 있다. 기록관리 분야의 특성을 나타내는 중점적인 부분에 해당될 것이다. 그런데 ISO 15489도 2016년에 개정되었고, 주요한 개정 부분인 '평가가 기관의 기능분석까지 확대된 넓은 개념이 되었다는 점에 주목할 필요가 있다. 앞으로는 기록관리 기능을 정리하고 도구를 개발할 때, 기록평가를 위한 조직의 기능분석, 기능분류체계, 시소러스의 연계를 고려해야 할 것이다.

3. 시소러스 활용시스템 사례

정보를 다루는 모든 분야는 많은 정보를 효율적으로 모으고, 관리하는 것을 가장 중요하게 다루고 있다. 이러한 고민은 제어어휘 활용으로 연

결되며, 구체적으로는 시소러스나 전거시스템의 구축을 요구하게 된다. 본 장에서는 온라인 환경에서 시소러스의 활용의 사례를 분석하였다. 특히 ISO 25964를 적용할 수 있는 시소러스 시스템을 중점적으로 조사하였다. 시소러스 활용시스템으로 영국의 지방정부기록의 제어어휘시스템(Integrated Public Sector Vocabulary, IPSV), EU의 다국어 시스템(EuroVoc), 뉴질랜드의 기능어 시스템(Functions of New Zealand Thesaurus, FONZ) 등을 선정하였다.

3.1 Integrated Public Sector Vocabulary(IPSV)

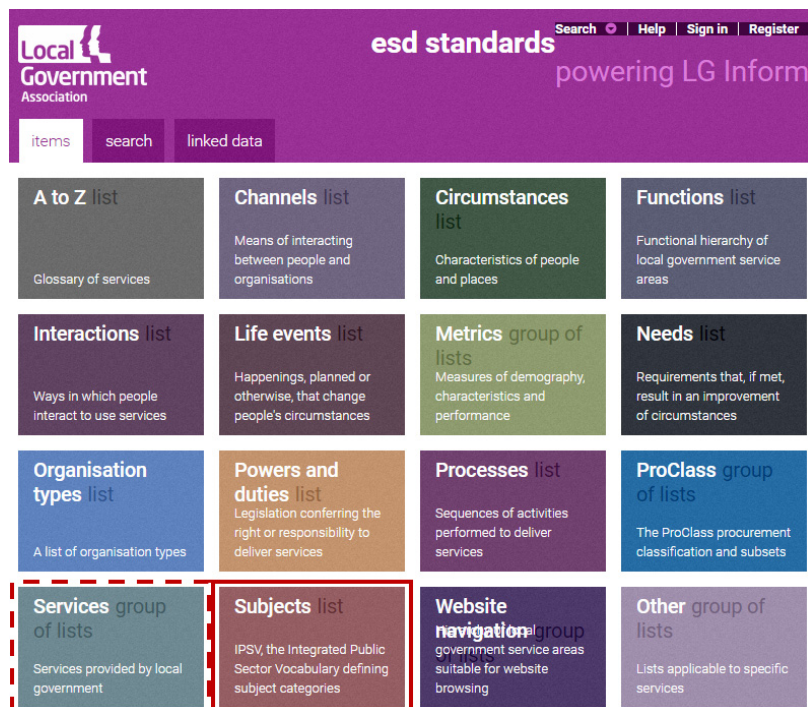
IPSV는 영국 공공기관에서 생산되는 웹페이지나 문서들의 레이블(label) 작성을 위한 것이며, e-Government Metadata Standard (e-GMS)의 주제 요소를 부여하기 위한 입력체계이다. IPSV는 단일어 시소러스의 국제표준인 ISO 2988과 BS 8723을 수용하여 구조화되었는데, 이러한 두 표준은 현재의 시소러스 구축 표준인 ISO 25964의 근간이 되었다. Government Category List(GCL), Local Government Category List(LGCL), seamlessUK taxonomy를 통합한 버전 1.0이 2005년 4월에

공표되었으며, 2006년 4월에 버전 2.0으로 업데이트되었다. 버전 1.0은 2,732개의 디스크립터와 4,230개의 비디스크립터로 구성되었으며, 버전 2.0은 3,080개의 디스크립터와 4,843개의 비디스크립터로 구성되었다. IPSV의 용어 수준은 공공성을 지향해야 하며, 뿐만 아니라 지방정부 내부에서 다루어지는(internal-facing) 토픽들도 포함시켜야 하고, 이를 위해 지속적인 업데이트 계획이 수립되어 있다.

IPSV의 일부 콘텐츠는 esd-standards 사이트(<http://standards.esd.org.uk>)의 하위 셋으로 접근이 가능하다. esd-standards 사이트는 영국 지방정부(UK Local Government)와 관련된 정보의 유형을 16가지로 구분하고, 그 아래에 제어어휘 목록을 제시하였다(〈그림 8〉 참고).

IPSV는 〈그림 8〉의 실선 사각형으로 표시된 바와 같이 16가지 중 주제(subjects)에 해당되며, 지방정부 관련 토픽을 표현한 756개의 디스크립터와 1,333의 비디스크립터, IPSV 전체를 추약하여 표현한 549개의 디스크립터와 1,472개의 비디스크립터 등의 데이터를 포함하고 있다.

IPSV에서 제공하는 데이터의 내용과 구조를 바탕으로 도출한 IPSV의 특징은 다음과 같다. 첫 번째는 시소러스의 관계 구조 안에서 대등 및 계층관계를 표현한다. 대등관계는 디스크립터와 비디스크립터와의 구별이며, IPSV에서는 'also known as'로 표현하였다. 계층관계는 탭 메뉴를 통한 개념간의 링크로 표현된다. 두 번째는 외부 시스템과의 상호운용성을 고려한 메타데이터 체계를 수용하고, 디스크립터에 부여될

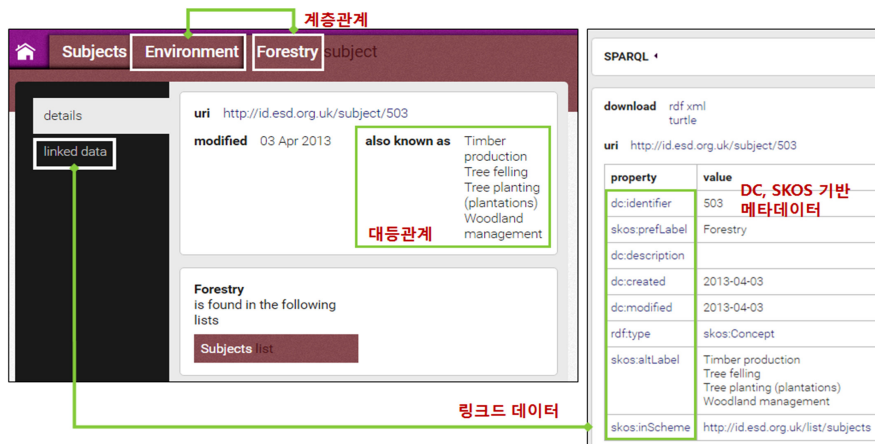


〈그림 8〉 UK Local Government esd-standards 웹사이트

요소 및 값을 제공하였다. 특히 더블린코어 메타데이터와 SKOS 운영체계를 준수하여 IPSV 지식체계를 외부 시스템과 공유할 수 있는 기반을 제공하였다. 세 번째로는 링크드 데이터를 발행하였다. IPSV는 개별 식별기호(Uniform Resource Identifier, URI)를 통해 개별 용어를 식별하고, 각 URI에 링크 정보를 부여하여 외부 시스템과의 연결을 지향하고 있다(〈그림 9〉

참고).

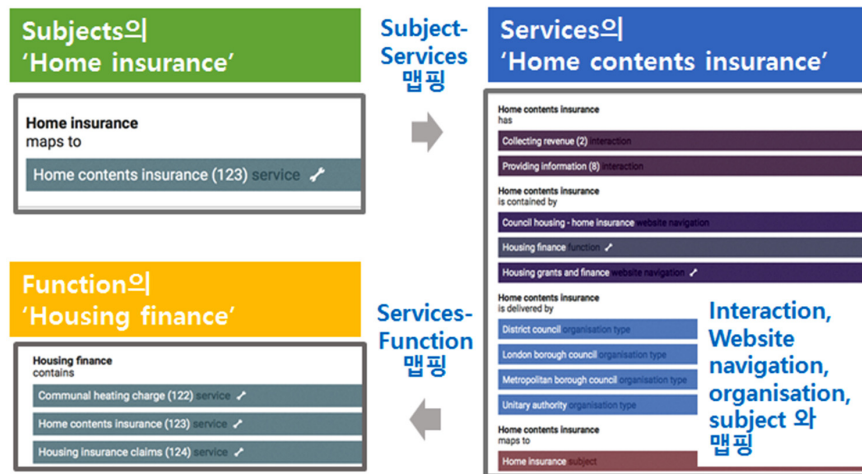
네 번째는 UK Local Government Association에서 제공하는 다른 시스템간의 연결을 보장한다. 〈그림 9〉에서 제시된 16가지의 시스템들 중 IPSV는 서비스(services)와 연결되고 해당 서비스에 의해 다양한 용어 목록인 기능(functions), 조직(organizations), 환경(circumstances) 등으로 확장된다(〈그림 10〉 참고). 또한 영국 지



〈그림 9〉 IPSV 데이터의 용어관계 표현 및 링크드데이터 발행 예시
 (http://standards.esd.org.uk/?uri=subject%2F503&tab=linked)



〈그림 10〉 서비스 어휘목록을 중심으로 한 지방정부 관련 어휘목록의 연계



〈그림 11〉 UK Local Government; IPSV(subject)-Services-Functions 맵핑 화면

방정부의 표준 어휘목록 간의 연결은 맵핑 정보로 이루어지는데, 〈그림 11〉과 같이 IPSV(주체: subjects)는 서비스(services)를 통해 기능(functions)과 연결될 수 있다.

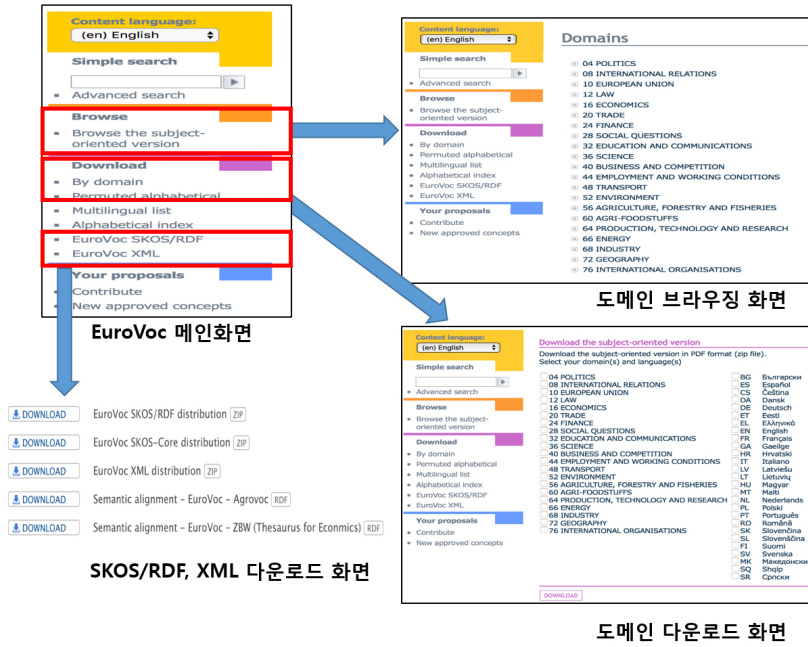
3.2 EuroVoc

Eurovoc은 유럽연합 출판국(Publications Office of the European Union)에서 운영하는 다국어 시소러스로, 불가리아어, 크로아티아어, 체코어, 덴마크어, 네덜란드어, 영어, 에스토니아어, 핀란드어, 프랑스어, 독일어, 그리스어, 헝가리어, 이탈리아어, 라트비아어, 리투아니아어, 몰타어, 폴란드어, 포르투갈어, 루마니아어, 슬로바키아어, 슬로베니아어, 스페인어, 스웨덴어 등의 23개 언어가 제공된다.

유럽연합 출판국은 EuroVoc 체계가 온톨로지 기반 시소러스로 발전해야 하며, 새로운 시소러스 기준과 W3C에서 제안하는 웹 발전의 흐름에 맞는 시맨틱웹 기술을 수용해야 한다고

강조하였다. 또한 유럽연합이나 유럽 국가와 의 회뿐만 아니라 전 세계의 정부와 개인 이용자가 지도 포함시키는 EuroVoc의 이용자 범위를 제안하였다(〈그림 12〉 참고).

〈그림 13〉은 EuroVoc의 주요 화면들을 제시한 것으로, EuroVoc에서 제시한 새로운 시소러스 지침과 시맨틱웹이 지향하는 방향에 맞춰 시소러스를 개발하고 있다는 것을 보여주고 있다. 주요 화면 중 도메인 브라우징 화면은 정치(04)부터 국제기구(76)까지 21개의 상위주체를 제공하고, 도메인 다운로드 화면에서는 상위주체와 언어를 선택하여 PDF 파일로 데이터를 다운로드 받을 수 있다. SKOS/RDF, XML 다운로드 화면을 통해서도 EuroVoc의 데이터를 SKOS/RDF, XML 형식으로 일괄 다운로드 받을 수도 있으며, 다른 시소러스와의 연계 데이터를 제공받을 수도 있다. 〈그림 14〉는 SKOS/RDF 파일의 일부분으로 다양한 지식체계와 연동하고 확장할 수 있는 가능성을 네임스페이스로 확인할 수 있다.



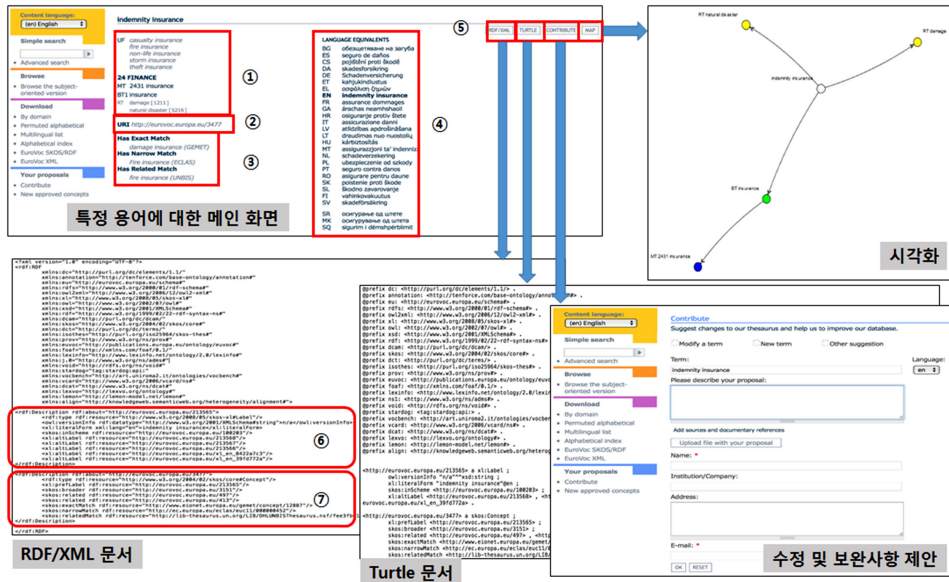
<그림 12> EuroVoc 인터페이스

```
<rdf:RDF
  xmlns:isothes="http://purl.org/iso25964/skos-thes#"
  xmlns:lemon="http://lemon-model.net/lemon#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:j.0="http://www.w3.org/ns/adms#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:prov="http://www.w3.org/ns/prov#"
  xmlns:euvo="http://publications.europa.eu/ontology/euvo#"
  xmlns:void="http://rdfs.org/ns/void#"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  xmlns:lexvo="http://lexvo.org/ontology#"
  xmlns:lexinfo="http://www.lexinfo.net/ontology/2.0/lexinfo#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:vocbench="http://art.uniroma2.it/ontologies/vocbench#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:skosxl="http://www.w3.org/2008/05/skos-xl#"
  xmlns="http://eurovoc.europa.eu/schema#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:stardog="tag:stardog:api:"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
```

<그림 13> EuroVoc; SKOS/RDF 파일의 네임스페이스 선언부

<그림 14>는 특정 용어에 대해 제공되는 정보들과 관련된 서비스를 구조화 한 것이다. 특정 용어 'indemnity insurance'에 대한 정보와 확장된 서비스를 제시한 것이 '특정용어에 대한 메인

화면'이며, 이 화면은 크게 다섯 가지의 기능을 제공한다. 첫 번째 서비스는 기본적인 시소러스 구조로 대응, 계층, 연관관계를 제공하여, 용어에 대한 개념을 구조화 하였다. 두 번째 서비스는



〈그림 14〉 EuroVoc; 특정 용어에 대한 인터페이스

해당 용어에 대한 정보를 웹 주소인 URI로 발행하여 내·외부 시스템과 연계될 수 있는 기반을 제공하였다. 세 번째 서비스는 EuroVoc과 연결되어 타 시소러스와의 의미 상호운용성을 제시하였다. 해당 용어와 연결된 ECLAS (European Commission Library Thesaurus), GEMET (GENERAL Multilingual Environmental Thesaurus), Unbis (United Nations Thesaurus) 뿐만 아니라, AGROVOC, ZBW (Thesaurus for Economics), LCSH (Library of Congress Subject Headings), INSPIRE feature concept dictionary, Rameau (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) 등의 의미 매핑을 제공하고 있다. 네 번째는 EuroVoc이 제공하는 고유 서비스 중 하나로 다국어 지원을 제공하는 것이다. 다섯 번째 서비스는 최신의 웹 기술을 수용하고, 정보 공개 및 공유를 강조하는 것으로, XML/RDF 문서, Turtle 문

서, 이용자가 직접 용어에 대한 정보를 수정하고 추가할 수 있는 인터페이스, 용어의 계층 및 연관관계를 지도로 표현한 시각화 등을 제공하였다. EuroVoc에서는 특정 용어 자체에 대한 식별자 URI와 개념에 대한 URI를 구분하여 주고 있다. 〈그림 14〉 안의 XML/RDF 문서를 살펴보면, EuroVoc은 ⑥번과 같이 'indemnity insurance'라는 영어 자체에 대한 식별자를 'http://eurovoc.europa.eu/213565'으로, ⑦번과 같이 'indemnity insurance'의 개념에 대한 식별자를 'http://eurovoc.europa.eu/3477'로 나누어 제공하였으며, 속성을 표현한 구문도 용어 자체는 '〈rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/〉'로, 개념은 '〈rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/〉'로 표현하였다.

3.3 Functions of New Zealand Thesaurus (FONZ)

Functions of New Zealand Thesaurus (FONZ)는 뉴질랜드 정부가 구축하여 Australian and New Zealand Society of Indexers(ANZSI)에서 제공하는 시소러스이다. FONZ는 2004년에 구축된 시소러스로, 1,543개의 디스크립터를 제공하고 있다(〈그림 15〉 참고).

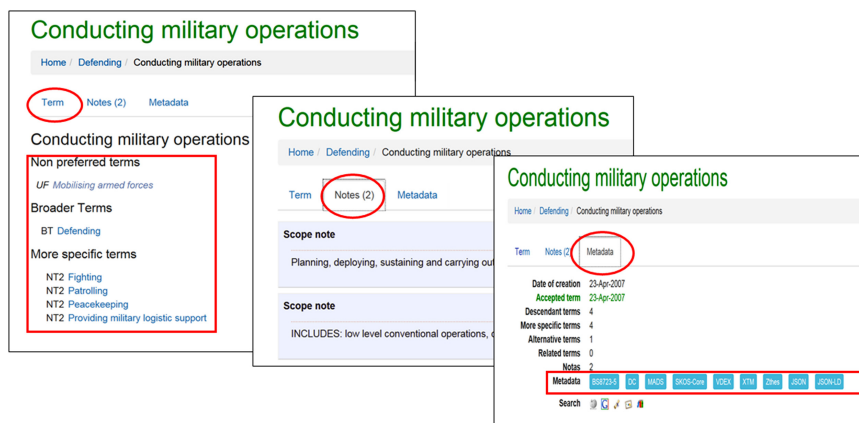
FONZ에서는 〈그림 16〉과 같이 특정 용어에 대한 정보를 크게 세 가지로 나누어서 제공한다. 먼저, 'Term'에서는 해당 용어의 대등, 계층 관

계를 제시하고, 'Notes'에서는 범위주기를 통해 용어 개념 범위 및 정의를 제공하고 있다. 마지막의 'Metadata'는 BS8723-5, DC, MADS, SKOS-Core, VDEX, XTM, Zthes, JSON, JSON-LD 등 타 메타데이터 요소들을 연결하는 탭으로, 특정 용어가 타 메타데이터 요소들로 표현될 때 사용되는 속성 및 값을 제시한 구문을 링크로 제공하고 있다.

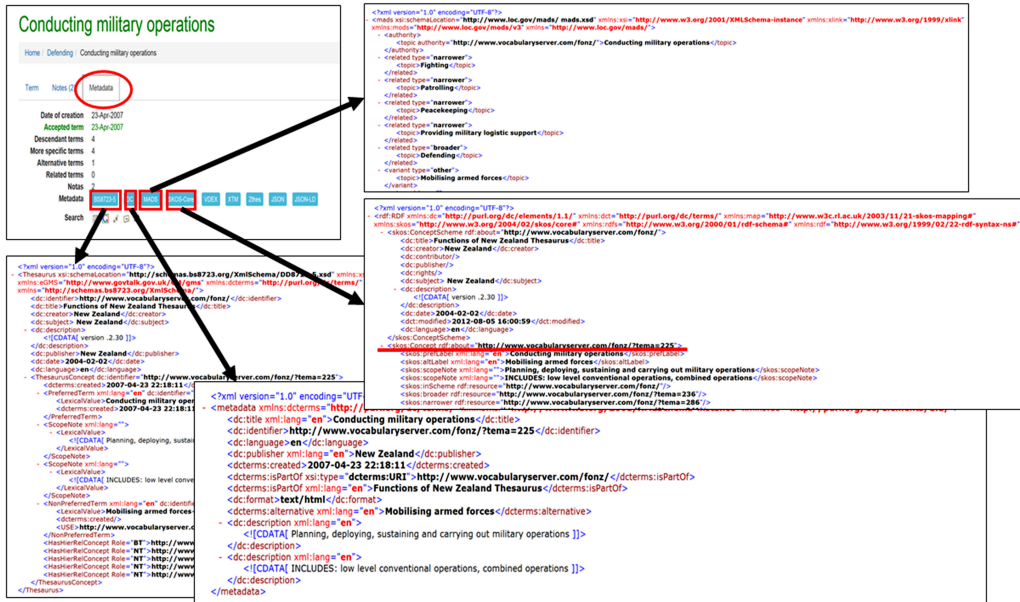
〈그림 17〉은 'Metadata' 탭에서 확장되는 의미구조의 실제 예시이다. 'Metadata' 탭에서는 위에서 제시한 바와 같이 BS8723-5를 비롯한 아홉 개의 메타데이터 셋으로 확장될 수 있는데,



〈그림 15〉 Functions of New Zealand Thesaurus(FONZ) 설명



〈그림 16〉 FONZ 인터페이스



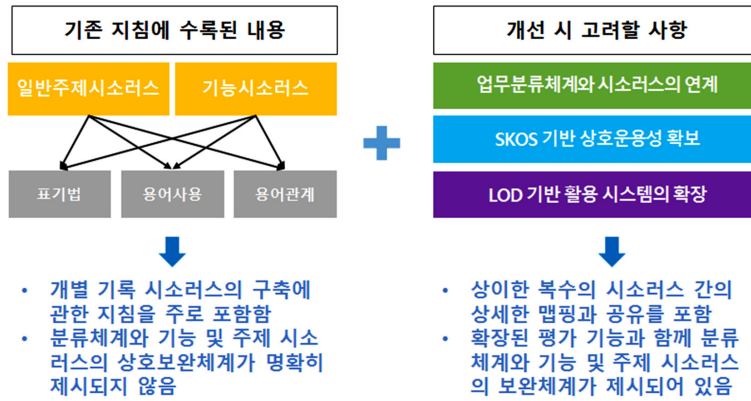
〈그림 17〉 FONZ; Metadata 탭으로 연결되는 다양한 문서들

〈그림 17〉은 그 중 BS8723-5, DC, MADS, SKOS-Core 등 네 개의 셋만을 제시하였다. 먼저, 'BS8723-5'는 BS8723 (Structured vocabularies for information retrieval - Guide)의 Part 5 (Exchange formats and protocols for interoperability)에 해당하는 것으로 이종의 시스템 간의 상호운용성을 강조하는 것이다. FONZ는 'BS8723-5'를 준수한 데이터 구문을 표현함으로써, 영국에서 생산 및 관리되는 정보의 속성을 활용할 수 있으며, 동시에 FONZ가 보유하고 있는 정보를 영국에 제공하는 정보 환경의 기틀을 제공하고 있다. 두 번째로는 DC 메타데이터 요소와의 상호운용성을 유지하고 있다. 해당 페이지에서는 DC 스키마와 속성을 활용하여 특정 용어에 대한 정보를 설명하였다. 세 번째로는 MADS 요소와 연결되는 링크를 제공하고 있다. MADS는 미국 국회도서관에서 선언한

메타데이터 셋으로, MADS와의 연결은 도서관의 서지데이터와의 공유 및 교환의 가능성을 보장하는 것이다. 마지막으로 SKOS-Core와의 연결은 FONZ와 온라인으로 접근 가능한 다양한 지식체계와의 호환성을 보장하는 것이다. FONZ는 〈SKOS:Concept〉을 통해 해당 용어가 표현하고자 하는 '개념' 중심의 정보를 제공하고 있다.

4. 기록 시소러스 지침 개선을 위한 고려사항

최근에는 기록관리 분야도 유관 분야와의 협력 필요성이 증대되고 있으며, 기록관리 분야 내에서도 복수의 시소러스가 관리되고 있다. 이러한 환경 변화에 맞춰 개정된 시소러스 관련 국



〈그림 18〉 기록 시소러스 지침 개선을 위한 고려사항

제표준인 ISO 25964 및 기록관리 표준인 ISO 15489를 반영하여 우리나라의 기록시소러스 개선에 필요한 사항을 검토하여 반영하였다(〈그림 18〉 참고).

4.1 업무분류체계와 시소러스의 연계

ISO 15489-1:2016과 ISO 25964-2:3013에 따르면 기록관리를 위한 업무분류체계는 시소러스를 통해 보완될 수 있다. 이를 위해서는 업무분류체계의 최하위 요소를 중심으로 기능 시소러스와의 연계가 필요하다. 업무분석을 바탕으로 구축된 업무분류체계는 크게 기능, 활동, 업무절차의 계층구조를 이루게 되는데, 업무절차로 생산된 기록물의 제목을 포함하여 기능어 전반을 대상으로 기능 시소러스를 구축할 수 있다(ISO, 2016, p. 14). 기록물의 맥락정보에 해당되는 기능은 기록관리를 위한 분류와 시소러스 구축을 위한 핵심적인 기준 역할을 수행하기 때문이다. 또한 ISO 15489에서는 기능시소러스를 통한 업무분류체계의 지원 외에 주제어로 구성된 색인(indexing)에 대한 지침도 제시하고

있다. 주제색인 정보는 기능분류체계와 직접으로 연계되는 것이 아니라, 기록물에 대한 별도의 메타정보로 부여된다. 그리고 기록물을 중심으로 하여, 기능분류체계-기능시소러스 및 색인어의 연결 구조가 작성될 수 있다. 이는 업무 맥락 중심의 기능분류와 이용자의 검색 접근점인 주제색인의 연계를 통한 기록관리 및 기록물 검색을 지원하기 위한 것이다. 한편 영국의 지방정부기록에 대한 제어어휘시스템인 IPSV는 서비스(services)를 중심으로 주제(subject), 기능(functions), 조직(organizations), 환경(circumstances) 등으로 연결하여 다른 시스템 간의 연동 및 확장을 제공하고 있다. IPSV를 참조하여 기능 시소러스와 주제 시소러스간의 직접적인 연동이 아니라 업무 서비스를 통한 중계 방식의 연계 방법도 고려해 볼 수 있다.

현재 국가기록원의 시소러스 지침(2009)에서는 기능 시소러스를 구축할 때, 기능분류체계의 최하위 단위업무(또는 단위과제)를 참고하도록 규정하고는 있다. 그러나 시소러스를 통해 기능분류체계를 보완하는 방식에 대한 구체적인 언급이 없다. 또한 일반주제시소러스에 대한

지침과 기능시소러스에 관한 지침이 별개로 제시되어 있으므로 이 두 시소러스의 상호 보완 기능에 대해서도 구체적으로 제시하고 있지 않다. 앞으로는 시소러스 구축 지침에 분류체계와 시소러스, 주제색인과 기능시소러스 간의 효과적인 연계체계의 필요성과 방법에 대한 내용이 추가되는 것이 바람직할 것이다. 더불어 개정된 ISO 15489에서는 '평가'의 범위가 기관의 기능분석까지 확대되었다는 점도 분류체계나 시소러스 구축 지침에서 언급되어야 할 것이다.

4.2 SKOS 기반 시소러스 간의 상호운용성 확보

ISO 25964의 개발과 관련된 표준의 특징으로는 디지털 정보환경을 반영하고, 상이한 시소러스 간의 상호운용성을 확보하기 위한 데이터 모형 설계나 SKOS 기반 연계 체계를 구축했다는 점이 있다. 이 표준에서는 시소러스를 일종의 지식조직시스템(Knowledge Organization System, KOS)으로 간주하고, 특정 KOS를 기계처리가 가능한 형태로 발행하여 시맨틱웹 환경에 적용 가능하도록 설계하고자 하였다. 이를 위해 적용한 구조가 SKOS(Simple Knowledge Organization System)로 이는 온라인으로 접근 가능한 다양한 지식체계와의 호환성을 보장하는 방법이라 할 수 있다. 이와 더불어 상이한 시소러스 및 관련 제어어휘 도구를 맵핑하는데 SKOS의 관계 어휘를 이용하여 상호운용성을 확보할 수 있다. 복수의 시소러스에 존재하는 유사한 용어를 1:1 혹은 1:n으로 연계하는 기존 방식이 아니라 n*n으로 연계하여 맵핑 정보를 명확하게 부여할 수 있다.

또한 더블린코어(DC)를 활용하여 특정 용어

에 대한 정보를 설명하여 외부 시스템과의 상호운용성을 유지할 수도 있다. MARC21이나 미국 국회도서관에서 선언한 메타데이터 셋인 MADS 요소와 연결되는 링크를 제공하여 서지 데이터와의 공유 및 교환 가능성을 보장할 수도 있을 것이다. 또한 FONZ에서 영국의 이종 시스템과 상호운용성을 위해 'BS8723-5'를 지원하는 것처럼 관련된 해당 표준 어휘 지원을 통해 상호운용성을 확보할 수 있다. 그리고 웹 주소 URI 연계를 통해 해당 용어에 대한 정보를 다른 시소러스와의 연결하여 간략한 수준의 의미적 상호운용성을 지원하는 방법도 가능하다. IPSV에서는 URI를 통해 개별 용어를 식별하고 각 URI에 링크 정보를 부여하여 외부 시스템과 연계를 제공하고 있다. EuroVoc에서는 특정 용어 자체에 대한 식별자 URI와 개념 URI를 구분하여 주고 있다.

시소러스 간 상호운용성 확보를 위해서는 다국어 처리에 대한 부분도 고려해 봐야 할 사항에 해당한다. 현재 국가기록원 기록시소러스 지침(2009)에서는 시소러스 용어관계 중 하나로 대응 외국어 관계지시기호를 이용하여 대표어와 동일한 의미를 가진 대응 외국어를 사용하도록 하고 있다. 이는 다양한 언어로 기술된 기록물의 검색을 제한할 뿐만 아니라 다른 제어어휘들과의 상호운용성에도 상당한 제약을 초래하는 원인이 될 것이다. ISO 25964에 따르면 다국어 시소러스의 모든 언어 각각은 동일한 구조를 공유해야 하며 용어관계, 계층관계 등도 대칭적으로 다루도록 하고 있으나 예외적으로 언어 습관이나 문화의 특성에 따라 비대칭적인 구조가 발생할 수도 있다. ISO 25964에서는 이러한 다국어 시소러스를 구축하는 방법으로 다음과 같이

세 가지 방법을 제안하고 있다. 각각의 방법은 가능한 인력, 시간, 비용에 따라 선택가능하며 기존에 단일언어 시소러스가 존재하는지, 색인에 이용되는 시소러스가 있는지와 같은 상황에 따라서도 달라질 수 있다.

첫 번째 방법은 단일언어 시소러스 번역으로, 다국어 시소러스를 구축하는 일반적이고 상대적으로 경제적인 방법이다. 검색하고자 하는 기록이 기존의 단일언어 시소러스로 색인되어 있을 경우에 특히 비용 면에서 효과적이다. 그러나 이 방법론을 채택하게 되면 기존 단일언어 시소러스가 우선 언어가 되는 상황에 놓이게 된다. 번역과정에서 언어 및 사회가 가지는 문화적, 개념적 차이로 인해 언어별로 용어 및 그에 해당하는 관계 구조가 항상 일치하지 않을 수도 있다.

두 번째 방법은 몇 개의 단일언어 시소러스를 병합하는 것으로, 복잡한 과정을 거쳐야 하지만 다른 개념적, 언어적 체계를 심사숙고하여 모든 언어가 본연의 역할을 할 수 있도록 허용한다. 그러나 각 언어들에 가지는 전형적인 특수성 및 사전 조합 수준의 차이 등으로 인해 관리차원에서 가장 어려운 접근 방법이라 할 수 있다.

세 번째 방법은 다양한 언어 버전의 다국어 시소러스를 동시에 구축하는 방법으로, 모든 언어는 동등한 지위를 가지게 된다. 각 언어는 차례로 원본 및 목표 언어가 되며 계층 및 용어관계 구축에 있어서도 영향을 미칠 수 있다. 교차언어간 동등성을 유지하기 위해 용어적, 구조적으로 채택한 방법이 동일언어에서는 항상 성립되지 않는 방식으로 관리되어야 한다. 성공적인 결과물을 도출하기 위해 처음부터 세심한 조정과 팀워크가 필요하다.

4.3 LOD 기반 시소러스 활용 시스템의 확장

시소러스를 탑재한 시스템은 탐색기능과 함께 일반 이용자가 정보탐색 과정에서 직관적으로 시소러스를 이해하고 활용할 수 있도록 지원하는 소프트웨어가 필요하다. 이러한 시소러스의 이용 방식은 시소러스 구축을 위한 표준에도 반영되어야 할 것이다. ISO 25964에서도 이러한 환경 변화에 맞게 시소러스 구축을 위한 S/W와 데이터 모형에 관한 지침을 추가하고 상이한 시소러스 간의 맵핑 유형과 방법에 대해서 제시하고 있다. 시소러스 활용 시스템은 시소러스 데이터의 관리 및 입출력이 가능해야 하며 필요에 따라 동시에 복수의 시소러스를 관리할 수 있어야 한다. ISO 25964 표준에서 제시한 시소러스 활용 및 관리 시스템이 갖춰야 할 기본적인 기능 요건을 제시하면 다음과 같다.

첫째, 각 용어 및 개념은 유일한 식별기호를 가져야 한다.

둘째, 용어나 개념, 개념그룹 혹은 개념배열은 그와 연관된 최소 1개 이상의 코드, 숫자 혹은 다른 지시기호와 결합이 가능하며 복수개의 주제 분류와 연결될 수 있어야 한다.

셋째, 용어개수, 글자 수, 계층의 레벨수의 제한이 없어야 한다.

넷째, 개념 및 용어간의 관계, 즉 BT, NT, RT, USE, UF, SN 등을 모두 지원해야 한다.

다섯째, 시소러스 용어 및 개념에 대한 범위 주기의 길이에 제한이 없어야 하고, 주기 내용에 마커나 하이퍼링크 지원이 가능해야 하며 해당 링크의 유효성도 점검해야 한다.

여섯째, 시소러스 데이터의 구조를 원형대로 유지하면서도 데이터의 손실이 없이 대량의 입출

력이 가능해야 하고 일괄 수정 및 삭제 기능 뿐 아니라 부분 수정, 삭제 기능도 지원해야 한다.

마지막으로, 시소러스 관리자를 위한 내비게이션 기능, 브라우징 기능, 편집 기능, 삭제 확인과 같은 실수방지 기능 등이 지원되어야 하며, 이 외에 보안기능, 동시제어 기능, 작업취소 기능 등 적절한 관리 기능이 지원되어야 한다.

그런데 위와 같은 요건은 향후 시소러스를 링크드데이터로 발행하기 위한 기반 구축과도 연계될 수 있다. 시맨틱웹 기술을 적용하여 시소러스의 온톨로지 기반 링크드 데이터 발행을 통해 정보 공개 및 공유를 확대하는 것이다. 또한 위키 기반 인터페이스를 제공하여 이용자가 직접 용어 정보를 수정, 추가할 수 있도록 하고 용어의 계층 및 연관관계를 지도로 표현한 시각화 등을 제공할 수 있다.

유럽연합 출판국은 EuroVoc 체계가 온톨로지 기반 시소러스로 발전해야 하며 새로운 시소러스 기준과 W3C에서 제안하는 웹 발전의 흐름에 맞는 시맨틱웹 기술을 수용해야 한다고 강조하였다. ISO 25964는 SKOS를 통해 시소러스를 기계처리가 가능한 형태로 발행하여 시맨틱웹 환경에 적용 가능하도록 하였다. IPSV도 더블링크어 메타데이터와 SKOS 운영체계를 준수하여 IPSV 지식체계를 외부 시스템과 공유할 수 있는 기반을 마련하고 링크드 데이터를 발행하였다. Getty의 AAT 시소러스도 처음에 문화유산 정보의 분류, 기술, 색인 작업을 위한 도구로 만들어졌으나 현재는 시맨틱웹을 적용하여 링크드 데이터로 발행하여 누구나 자유롭게 30년이 넘게 축적되어 온 연구 및 학술성과를 공유할 수 있도록 제공하고 있다.

5. 결론

본 연구에서는 시소러스의 구축과 활용, 상호 운용성 확보에 관한 최근의 국제 동향을 파악하고, 이를 우리나라의 기록 시소러스 구축 지침의 개선에 활용하고자 하였다. 시소러스 구축과 활용에 대한 국제 동향으로는 ISO 25964의 개정을 들 수 있다. 이 표준에 따르면 단일 시소러스의 구축을 위해서는 종이 기반 시소러스 구축 및 관리에서 시스템 기반 구축 및 관리체제로 발전해야 한다. 그리고 이를 위해서는 시소러스 디스크립터를 비롯한 개별 구성요소에 식별기호(URI)를 부여하고, DC나 SKOS와 같이 국제적으로 호환되는 메타데이터 체계를 도입해야 할 것이다. 또한 앞으로는 정보환경의 변화와 다양한 분야의 시소러스 개발로 인해 단일 시소러스의 구축뿐 아니라 복수의 시소러스 간 연계 방식도 이전보다 정교해질 필요가 있다. 이에 대한 대응은 상이한 시소러스 어휘를 1:1로만 맵핑하는 것이 아니라, SKOS의 다양한 연결 어휘를 활용하여 명확히 연계하는 방안을 비롯하여 시소러스 자체를 온톨로지로 변환하여 링크드 데이터로 발행하는 것도 고려될 수 있다.

기록관리 분야의 측면에서는 ISO 15489의 개정이 기록 시소러스의 구축과 관리에 영향을 줄 수 있다. 2016년에 개정된 ISO 15489에서는 기록물의 평가를 거시적인 관점에서 기관의 기능분석까지 포함하는 넓은 개념으로 확장시켰다. 그리고 그 결과로 업무분류체계를 수립하고, 업무분류체계를 기능시소러스와 연계하여 활용하도록 제안하였다. ISO 15489에서는 기능어의 제어 외에도 주제색인을 통한 기록의 검색 성능

개선도 제시하고 있어서 향후 기록 시소러스는 기능 시소러스와 주제 시소러스의 연계를 통해 발전해야 할 것이다.

우리나라의 시소러스 관련 지침의 개선을 위해서 국제표준과 더불어 최근의 정보환경의 변화를 적용하여 시소러스를 효과적으로 관리하고 있는 사례도 함께 살펴보았다. 이 중 영국의 IPSV에서는 공공영역을 다양한 분야로 구분하고 각각의 분야별로 제어어휘를 개발하여 통합 서비스하고 있었다. 여기에는 공공영역의 기능어나 주제어도 포함이 되어 있었는데, 다양한 공공분야의 제어어휘 목록을 공공분야의 서비스 어휘를 통해 연계했다는 특징이 있었다. 또한 IPSV를 비롯한 영국의 지방정부 제어어휘는 링크드 데이터로도 발행되고 있었다. 이 외에 EuroVoc 유럽의 환경에 맞게 다국어 시소러스를 지원하면서 다양한 언어로 된 시소러스

를 연계하는 것에 중점을 두었다. 또한 시소러스를 시맨틱웹 기반의 온톨로지로 관리하며, 링크드데이터로도 제공하고 있었다. 뉴질랜드의 FONZ는 용어의 계층관계를 체계적으로 표현할 뿐 아니라, 복수의 시소러스 용어를 SKOS를 비롯한 표준 메타언어로 연계하고 있었다.

이와 같이 국제적인 표준과 주요 국가의 공공분야 시소러스가 변화하고 있다. 시소러스 자체를 시스템을 통해 구축하고, 시맨틱웹을 지원하는 링크드데이터로 발행할 수 있어야 한다. 또한 기록 시소러스는 문화유산 관리기관이나 공공분야라는 상위 분야의 하위 어휘목록일 수 있으며, 이 경우 해당 분야에 개발되어 있는 복수의 시소러스 간의 상호운용성 확보가 중요하다. 2009년에 구축된 우리나라의 기록 시소러스 구축 지침도 새로운 정보환경과 유관 분야와의 협력을 위한 개정 작업을 준비할 필요가 있다.

참 고 문 헌

- 국가기록원 (2009). NAK/A 8:2009(v1.1) 국가기록원 시소러스 구축지침.
- 박지영 (2010). 시소러스를 연계한 문학류 폐쇄 분류체계 개발. 한국비블리아학회지, 21(3), 77-89.
- ANSI/NISO (2010). ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Retrieved January 5, 2017, from http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf
- Bustelo, Carlota (2016). ISO 15489. Records management standard updated.
- Clarke, Stella G Dextre., Will, Leonard D., & Cochard, Nicolas (2008). The BS8723 thesaurus data model and exchange format, and its relationship to SKOS. 21. July. 2008. ISKO UK Conference: Sharing Vocabularies on the Web via Simple Knowledge Organization System (SKOS). Retrieved January 15, 2017, from http://www.iskouk.org/sites/default/files/dextre_clarke_SKOSmtg21Jul2008.pdf

- Clarke, Stella G. Dextre and Zeng, Marcia Lei (2013). From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling. *Information Standards Quarterly*, 24(1), 20. DOI: 10.3789/isqv24n1.2012.04.
- EuroVoc. Retrieved January 3, 2017, from <http://eurovoc.europa.eu/drupal/>
- Functions of New Zealand Thesaurus (FONZ). Retrieved January 5, 2017, from <http://www.vocabularyserver.com/fonz/index.php>
- Getty Vocabularies: LOD. Retrieved January 5, 2017, from http://www.getty.edu/research/tools/vocabularies/lod/aat_semantic_representation.pdf
- IPSV. Retrieved January 5, 2017, from <http://standards.esd.org.uk/?uri=list%2Fsubjects&tab=details>
- ISO (2001). ISO 15489-1:2001. Information and documentation - Records Management - Part 1: General.
- ISO (2011). ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval.
- ISO (2013). ISO 25964-2: 2013 Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies.
- ISO (2016). ISO 15489-1:2016 Information and documentation - Records management - Part 1: Concepts and principles.
- NISO (2017). ISO 25964 Official Website (ISO 25964 - the international standard for thesauri and interoperability with other vocabularies). Retrieved January 6, 2017, from <http://www.niso.org/schemas/iso25964/>
- Reed, Barbara (2013). Revision of ISO 15489-Records Management. Retrieved January 3, 2017, from http://www.records.com.au/media/upload/HP_Information_Governance_15489_review_2013.pdf
- Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- Soergel, Dagobert (1996). Data Models for an Integrated Thesaurus Database. Paper presented at the Research Seminar on Compatibility and Integration of Order Systems, Warsaw, Poland, September 13-15, 1995. Issued by International Society for Knowledge Organization: Polish Library Association: Society for Professional Information. Warsaw: Wydaw. SBP.: 1996. pp. 47-52. Retrieved January 7, 2017, from https://www.researchgate.net/publication/2365344_Data_Models_for_an_Integrated_Thesaurus_Database

Will, Leonard (2012). The ISO 25964 data model for the structure of an information retrieval thesaurus. *Bulletin of Association for Information Science and Technology*, 38(4), 48-51. Retrieved January 3, 2017, from https://www.asis.org/Bulletin/Apr-12/AprMay12_Will.pdf

• 국문 참고자료의 영어 표기

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

National Archives of Korea (2009). NAK/A 12:2009 (v1.0) Guide for Authority Record of National Archives of Korea.

Park, Ziyong (2010). Developing a Faceted Classification Scheme Integrated with a Thesaurus for Literature. *Journal Of The Korean Biblia Society For Library And Information Science*, 21(3), 77-89.