

오픈 사이언스 실현을 위한 참조표준의 공유와 협력

채균식 (한국표준과학연구원 국가참조표준센터)

목 차

1. 서 론
2. 국가참조표준 체계를 통한 과학기술데이터의 신뢰성 확보
3. 오픈사이언스 실현을 위한 참조표준의 역할
4. 결 론

1. 서 론

다양한 형태로 생산한 과학기술 연구결과물을 개별 연구자 또는 기관에서 제한적으로 관리하여 오던 것을 인류 모두에게 공개하여 누구나 자유롭게 이용할 수 있게 한다는 의미의 오픈사이언스 운동이 활발히 진행되고 있다. 오픈소스, 오픈액세스 등 공개와 공유 개념을 가진 오픈이라는 용어 보다 넓은 의미로 오픈사이언스가 등장한 것 같다. 과학자들끼리 정보를 공개하여 연구를 진행하는 것 뿐만 아니라 일반대중에게도 정보 접근이 가능하도록 개방하고 비단 정보 또는 데이터 공개에 국한하지 않고 과학 실험장비까지 공개하여 실험을 돕는 운동이 전개되고 있다.

오늘날 우리는 다양한 개인 정보통신기기의 발달과 편리한 소프트웨어의 활용으로 수없이 많은 정보를 접하고 있고, 어떤 정보를 어디에 사용할 것인가에 대한 판단이 어려울 정도로 정보의 홍수 속에 살고 있다. 덕분에 많은 사람들이 신속하게

전세계에서 일어나고 있는 일들을 금방 알 수 있게 되었다. 이러한 일상에서 생산되는 다양한 정보들은 쉽게 접근이 가능하지만 과학기술정보를 비롯한 고급 정보의 접근은 쉽지가 않다. 어렵게 자금을 마련하여 오랜 실험결과물인 데이터를 다른 사람에게 아무 헤택 없이 공개한다는 것에 누구나 선뜻 나서기가 어려울지 모른다. 연구결과물인 데이터 및 관련한 정보를 이용하는데도 어려움이 많은데 연구에 필요한 다양한 리소스 즉 실험장비, 관련 데이터 분석 등 소프트웨어와 유사한 과학기술 인프라를 이용하는 것은 더욱 어려울 수 있다. 과학자가 가진 다양한 정보와 과학기술 인프라의 공유와 개방으로 다른 과학자와 교류를 활발히 할 수 있는 촉매제가 될 수 있다. 공개한 연구성과는 후속검증 및 추가연구로 이어질 가능성이 높아 해당분야의 연구발전을 획기적으로 성장시킬 수 있을 것이다. 또한 중복연구로 인한 시간과 예산을 획기적으로 줄일 수 있을 것이다. 우리나라의 경우 2006년부터 2011년까지 유사 중복 연

구로 지적받은 국가 R&D 사업은 419건에 이른다 (국가과학기술위원회, 2012a:2).

참조표준(Standard Reference Data)은 과학기술 연구개발과정에서 생산된 측정데이터 및 정보를 믿고 사용할 수 있도록 해당 분야 전문가 평가를 통해 신뢰성을 확보하여 국가가 공인한 표준데이터이다(국가표준기본법 제3조 7항)[1]. 용어에서 알 수 있듯이 “표준”은 합리적인 기준을 정하기 위해 절차와 방법에 따라 마련된 것으로 일반적이고 보편적으로 사용할 수 있는 기준인 것이다. 연구자들은 자신의 실험으로 인해 생성한 데이터가 얼마나 정확한지 그리고 믿을 수 있는지 비교 대상의 데이터를 찾는 경향이 있다. 참조표준은 해당 분야 표준데이터를 제시한 것으로 같은 값을 얻기 위해 중복연구를 최소화하여 시간과 예산의 낭비를 줄일 수 있을 뿐 아니라 자신의 연구결과 신뢰도를 객관적으로 비교해 볼 수 있는 표준데이터이다.

한가지 예로 같은 재료를 가지고 다른 실험실에서 연구자가 물성 값을 얻고자 할 때 그 값이 약간씩 다르게 나올 수 있다. 이러한 결과 데이터를 또 다른 연구자가 사용하고자 할 때 누구의 실험 결과를 사용해야 할지 혼란이 생길 수 있다. 따라서 실험과 측정된 결과 값의 다양한 변수들을 고려하고 과학적인 방법으로 결과 값을 평가하여 제시해 줌으로써 그 실험결과 값은 정확도와 신뢰도를 확보되어 믿고 사용할 수 있게 된다. 과학자가 가지고 있는 연구결과데이터를 공개하는 것이 오픈사이언스의 첫출발이라고 할 수 있지만, 엄청나게 생산되는 과학기술데이터에 신뢰성을 부여함으로써 후속연구에서 검증되지 않은 데이터의 사용으로 발생할 수 있는 전혀 엉뚱한 결과를 방지할 수 있다.

2. 국가참조표준 체계를 통한 과학기술 데이터의 신뢰성 확보

2.1 참조표준 정의

참조표준은 “측정데이터 및 정보의 정확도와 신뢰도를 과학적으로 분석·평가하여 공인한 것으로서 국가사회의 모든 분야에서 널리 지속적으로 사용되거나 반복 사용할 수 있도록 마련된 물리화학적 상수, 물성값, 과학기술적 통계 등을 말한다”(국가표준기본법 제3조). 국가표준기본법에서 국가표준을 측정표준(Measurement Standards), 성문표준(Document Standards), 참조표준(Standard Reference Data)을 3대 국가표준으로 정하여 국가사회의 모든 분야에서 정확성, 합리성 및 국제성을 높이기 위하여 국가적으로 공인된 과학·기술적 공공기준으로 기술하고 있다.

참조표준은 유사한 시험 연구에 참조할 수 있어 중복실험을 최소화할 수 있어 국가적으로 시간과 비용을 절감할 수 있는 핵심 요소이다, 산업적인 측면에서는 제품의 설계 및 신제품 개발에 기준데이터로 참고함으로써 생산비용을 절감하고 신속한 상용화를 가능하게 할 것이다. 과학기술 분야에서는 기초과학연구에 활용됨으로써 중복실험을 최소화하여 연구시간과 비용을 획기적으로 줄일 수 있다. 국민의 삶의 질 향상을 위한 의료분야에서는 한국인 고유의 생체 참조표준데이터 등을 개발함으로써 의사의 경험이 아닌 과학적인 근거에 의해 진단이 가능함으로써 진단의 정확도를 높일 수 있다. 또한 한국인 고유생체 참조표준데이터의 활용은 의료 장비의 국산화로 사회적인 비용 경감과 의료산업 발전에 핵심역할을 할 수 있다. 우리의 일상생활에서는 일조량, 기온, 먹는 물의 안전성, 물 관리 등에 대한 참조표준데이터를 제정하여 활용함으로써 안전하고 편리한 생활환경을 유지할 수 있다.

미국을 비롯한 선진국에서는 과학기술데이터를 중심으로 국가적인 지식자산을 보호하고 효율적인 과학기술데이터를 활용하기 위해 1960 대부터 참조표준 사업을 시작하였다. 미국은 1968년 참조표준법(Standard Reference Data Act, Public Law 90-396)을 마련하여 미국 국립표준연구소(NIST)로 하여금 참조표준 사업을 전개해 오고 있다. 연방정부기관 및 대학, 민간 기관에 참여하여 방대한 참조표준 데이터를 제정해 오고 있다. 대부분의 데이터는 공익을 위해 공개되어 있고 산업적인 활용 가치가 높은 참조표준은 고가의 판매 전략으로 상용화한 후 수요자에게 판매되고 있다. 분야별로 약 170여 데이터베이스 구축과 200만개 이상의 참조표준데이터를 확보하고 있으며 질량분석 참조표준 같은 데이터는 미국 국립보건원(NIH) 등과 협력으로 패키지화하여 데이터 산업화에도 많은 노력을 기울이고 있다.

독일은 1975년부터 참조표준 진흥사업을 연방정부 차원에서 추진하고 있으며 측정표준대표기관인 독일이공학연구소(PTB)가 중심이 되어 약 580만개의 참조표준을 개발해오고 있다. 대표적인 데이터베이스로서 DDB(Dortmund Data Bank)ST GmbH 개발하여 글로벌 회사에서 판매를 대행해 오고 있다.

일본의 경우 참조데이터 수준에서 산업기술종합연구소 소속 기관에서 생산한 다양한 과학기술분야의 데이터를 수집 생산하여 산업 및 연구개발 현장에서 사용할 수 있도록 지속적인 개발을 해오고 있다. 대표적으로 다음과 같은 데이터베이스를 제공하고 있다.

- Spectral Database for Organic Compounds (SDBS),
- Network Database System for Thermophysical Property Data
- Physical Hazard Database of chemicals

- Relational Information System for Chemical Accidents Database. (RISCAD)
- Geological Survey of Japan (GSJ) Database Collection
- Solid-State NMR Spectral Database (SSNMR_SD)[3]

2.2 참조표준의 현황

한국표준과학연구원에 국가참조표준센터가 2006년 8월에 설립되어 본격적인 국가참조표준 개발 및 보급 사업이 진행되고 있다. 한국표준과학연구원이 참조표준 사업을 위해 중심적인 역할을 하는 이유는 데이터 신뢰성 확보 때문이다. 참조표준의 신뢰성 확보는 측정소급성 (Measurement Traceability), 측정 불확도 (Measurement Uncertainty)를 핵심으로 하는 데이터 평가 기술이 확보되어야 한다. 따라서 측정데이터의 평가 기술은 국가측정표준대표기관 (National Measurement Institute)인 한국표준과학 연구원에서 확보한 기술을 바탕으로 이루어진다.

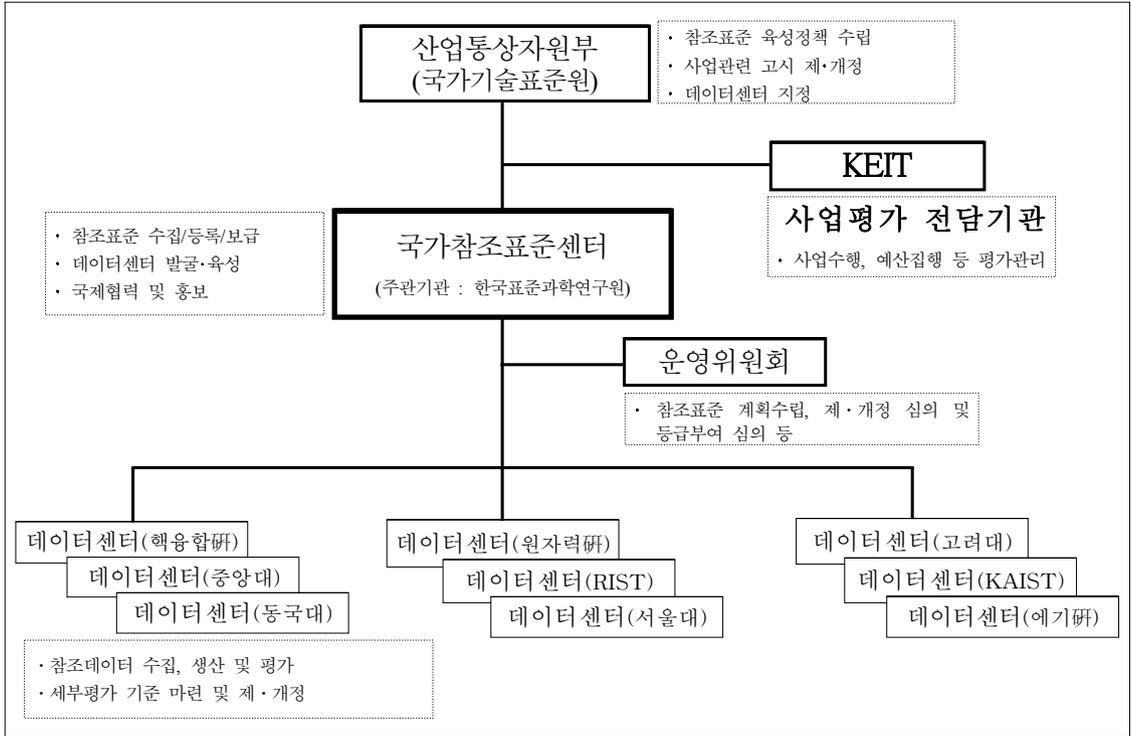
2.2.1 참조표준의 개발 현황 및 과정

참조표준으로 개발한 데이터는 물리화학, 금속 재료, 보건의료, 에너지, 환경 등 다양한 분야의 데이터가 개발되어 보급되고 있다[4]. 일반적으로 매년 10여종의 참조표준이 개발되어 현재까지 약 52종의 참조표준 DB가 구축되어 있으며, 전체적으로는 현재까지 약 32,000여건의 참조표준데이터가 제공되고 있다.

□ 참조표준 개발 과정

참조표준을 개발하는 것이 마치 수없이 많이 흠어져 있는 구슬을 실로 꿰어야 보배가 된다는 말이 생각난다. 같은 재료를 가지고 다양한 실험실에서 여러 연구자가 실험을 하더라도 결과값이 조금씩

〈표 1〉 국가참조표준 사업 추진체계



다르게 나타난다. 이를 활용하고자 하는 연구자는 어느 실험실에서 누가 실험한 데이터를 사용해야 할지 혼란스럽게 된다. 또한 자신이 실험한 데이터가 어느정도 믿을 수 있는지도 비교해 보아야 하는데 비교대상을 찾기가 쉽지 않다. 참조표준데이터는 이러한 관점에서 시작되었다고 해도 과언이 아니다.

정부는 다양한 분야에서 요구되는 믿을 수 있는 측정데이터를 수집 생산하여 전문가 평가를 거쳐 중복 사용이 가능하도록 공인한 참조표준을 국가에 제공함으로써 사회적 경제적 비용과 시간 절감의 필요성 인식하였다. 참조표준이 일반적인 과학기술데이터와 다른점은 평가를 거쳐 신뢰성을 확보한 공인한 데이터라는 것이다. 그러면 다양한 분야의 데이터의 평가는 어떤 방법으로 할 것인가에 대한 기준으로 다음과 같이 제시하고 있다.

참조표준은 <표 2> 기술평가 총괄기준에 따라 1에서 7항목까지 만족하는 데이터부터 참조표준으로 등록할 수 있다. 평가의 핵심 요소는 첫째 측정량(Measurand)으로서 무엇을 측정하고자 하는가가 분명해야 하며, 측정하고자 하는 내용이 잘 표현되도록 구체화 할 필요가 있다. 둘째는 측정소급성(Measurement Traceability)로서 교정한 장비를 사용하여 측정결과 값을 얻었는지에 관한 것이다. 셋째는 측정불확도(Measurement Uncertainty)로서 측정 결과에 대한 신뢰도를 나타내는 정량적인 과학적인 표현 방법이다. 측정결과에 대한 참값을 인간은 알 수 없기 때문에 참값에 가까운 측정 결과값이 제시한 값 범위내 어디에는 포함되어 있다는 추정 값을 과학적으로 표현(측정값 = 측정평균값 ± 측정불확도)한 방법이다. 일반적으로 측정값은 측정량에 대한 근사값 또는 추정값일 뿐이기 때문에 그

〈표 2〉 참조표준 제정 보급에 관한 운영요령 제12조 기술평가 총괄 기준

평가기준	등급기준
1. 측정하고자 하는 양이 명확하게 명시되어 있는지 여부 2. 측정방법과 절차 및 이론계산에 대한 설명내용 명시 여부 3. 측정방법의 적절성 및 그 근거의 명기 여부 4. 측정결과에 영향을 주는 요인의 제어 여부	참조데이터 (1~4까지 만족한 데이터)
5. 측정방법의 불확도 평가 및 측정소급성 확보 여부 6. 측정결과와 불확도 추정에 대한 적정성 여부 7. 측정의 상세절차와 측정의 재현조건 명시 여부	유효참조표준 (1~7까지 만족한 데이터)
8. 다른 경로에 의해 얻어진 결과와 측정결과와의 일관성 여부 9. 연관식과 모델링을 통한 데이터의 예측가능성 검증	검증참조표준 (1~9까지 만족한 데이터)
10. 2인 이상 관련분야 제3자의 종합검토를 받았는지 여부	인증참조표준 (1~10까지 만족한 데이터)

값에 대한 불확도(불확실성)가 함께 명시되어야 완전해 질 수 있다. 이밖에도 참조표준으로 제정하기 위해서는 측정방법, 재현성, 일관성, 또 다른 전문가에 의해 평가를 통해 유효참조표준(Qualified SRD)), 검증참조표준(Validated SRD), 인증참조표준(Certified SRD)로 구분한다.

참조표준의 제정은 각 분야별 데이터센터, 국가참조표준센터, 기술위원회, 운영위원회의 역할에 따라 이루어진다. 데이터센터는 해당 분야 참조표준을 개발하기 위해 우선 활용도를 전제로 한 활용로드맵을 수립하여 단계적으로 참조표준을 개발하게 된다. 참조표준데이터는 평가를 위해 상기 총괄평가기준서에 따라 해당 분야 데이터의 특성을 감

안한 세부적인 평가요소를 도출하고 적용한 세부평가기준서를 마련하게 된다. 이를 바탕으로 수집생산한 데이터는 세부평가기준서의 절차에 따라 평가한 후 각 데이터별 등급을 부여하여 참조표준(안)을 구성하게 된다. 참조표준(안)이 마련되면 국가참조표준센터에서 운영하는 기술위원회의 기술적인 심의를 거쳐 부적합 사항 또는 미비 사항을 개선 보완 조치 한 후 최종적인 참조표준(안)이 마련된다. 최종 참조표준(안)은 최종 심의 의결 기구인 운영위원회의 심의를 통해 참조표준으로 등록하게 된다. 이처럼 복잡한 등록과정을 거치게 되는 이유는 ‘표준데이터’이기 때문에 누구나 해당 데이터를 활용할 때 기준으로 삼을 수 있을 정도로 정확도와

〈표 3〉 참조표준 제정 과정에서 조직별 역할

수행조직	주요내용	세부내용
① 데이터 센터	품질매뉴얼, 생산절차서 수립	생산절차서(안), 세부평가기준서(안) 유효성 검증
	데이터세부평가기준서 수립	데이터 평가 기준서
	데이터 수집 생산	공개문헌, 직접생산
	데이터 편집, 가공	최상의 값도출, 유효데이터 확보
	데이터 자체 평가	유효, 검증, 인증 참조표준 등급부여(안) 마련
	최종 참조표준(안)	국가참조표준센터에 등급부여안 심의요청
② 국가참조표준센터	기술위원회 운영	데이터센터에서 요청한 참조표준(안) 기술 심의 생산절차서, 세부평가기준서 유효성 검증
	운영위원회 개최	참조표준(안) 기술평가 등 최종 심의
	고시 요청	국가기술표준원

신뢰성을 갖추고 있어야 한다.

2.2.2 참조표준 데이터센터 지정 운영

참조표준을 제정하기 위해서는 해당 분야 데이터를 수집, 생산, 평가할 수 있는 품질시스템 - ISO 9001, 17025 - 을 갖춘 기관에서 참조표준데이터 센터를 산업통상자원부 장관을 대신해 국가기술표준원장이 평가하여 지정한다. 데이터센터의 기능은 해당분야 참조표준 개발 계획 수립, 참조데이터의 지속적인 수집, 생산 및 평가, 참조데이터 평가를 위한 세부평가기준서 제정 및 개정, 국내 관련기관과 협력을 주요 기능으로 하고 있다. 데이터센터로 지정받기 위해서는 신청서 제출에서 현장평가, 부적합 사항 보완 등 일련의 과정을 거쳐 국가로부터 지정을 받게 된다. 데이터센터 지정조건은 해당분야 참조표준을 생산하기 위해 요구되는 「공인기관 인증제도 운영요령」에 따른 공인시험기관으로 인정을 받거나 동등한 품질시스템을 운영하고 있는 조직을 가장 우선적으로 평가하여 지정하게 된다. 또한 참조데이터 수집 생산 능력, 참조표준 데이터의 지속적인 유지, 관리 능력, 불확도 평가능력, 측정데이터 생산 장비의 소급성 확보 등을 갖춘 기관을 지정한다.

참조표준데이터센터로 지정된 기관은 지속적으로 품질시스템을 유지하여야 해당분야 참조표준을 생산할 수 있다. 따라서 지정 후 2년 이내 사후관리를 실시하게 된다. 품질시스템 운영여부, 참조데이터 수집가공 절차 수립 여부 등 경영기준과 지정분야 전문지식과 측정기술, 불확도 평가능력 등 기술기준에 적합한지를 현장평가를 통해 이루어진다. 데이터센터로 지정된 기관은 5년간 유효기간동안 해당분야 참조표준을 개발할 수 있는 자격을 갖추고 있고, 유효기간 6개월 이전에 갱신평가를 신청하여 데이터센터 신규 신청에 준하는 평가를 거쳐

자격을 획득할 수 있다.

참조표준데이터센터는 국가적인 참조표준 개발 수요에 따라 27개소를 지정 운영하고 있으며, 물리화학, 금속재료, 보건의료, 생명과학, 에너지 등 다양한 분야에서 참조데이터 수집 생산, 평가 능력을 갖춘 기관을 지정하여 운영하고 있다.

2.2.3 참조표준 개발 사례

국가참조표준은 국가 첨단산업, 기간산업, 국민의 삶의질 향상 분야로 나누어 개발을 해오고 있으며 수요조사에 근거하여 개발 분야를 선정하고, 개발분야 중에 국가적으로 개발능력이 있는 분야를 개발대상으로 한다. 선진국에서 이미 개발한 참조데이터 또는 참조표준은 개발 대상에서 제외되며, 한국적이고 한국인 고유의 참조표준데이터를 개발 우선 대상이다.

참조표준으로 개발한 “한국인 허혈뇌지도 참조표준 도판”은 한국인의 의료 MR영상 빅데이터를 수집하여 참조표준으로 제정 후 의료진단 현장에서 Reference로 사용할 수 있는 표준도판을 제작하여 활용하는 사례이다. 통계청의 2014년 한국인 사망원인 통계에 따르면 뇌혈관 질환 관련 질병은 세 번째로 높은 사망률을 나타내고 있다[5]. 뇌혈관 질환은 뇌경색, 일과성 뇌허혈증, 뇌출혈을 총칭하는 것이며, 환자에게는 심각한 장애를 초래할 뿐만 아니라 가족의 고통, 사회적으로는 노동력 상실, 환자 부양비용 증가 등 국가 사회 경제적으로 많은 부담이 되는 질병이다. 특히 뇌경색 진료에 있어 뇌 MR 영상이 광범위하게 사용되고 있으나 환자에게 정량적으로 정확히 소견을 전달하기가 어려운 경우가 많다. 따라서 뇌허혈 병변이 발견된 경우 환자의 성별, 위험인자를 고려하여 “건강인 기준으로 평균 몇 살 정도의 노화된 뇌에서 관찰되는 소견이다”라고 진단을 할때 뇌 MR 영상 참조표준이 있으면 진

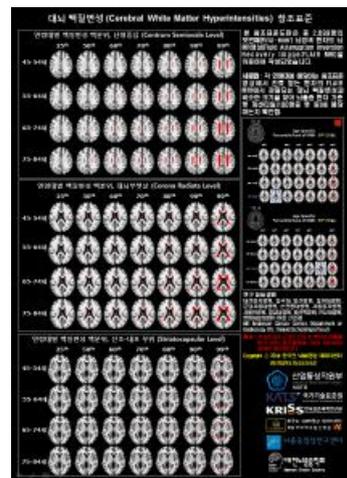
〈표 4〉 예시) 전체뇌경색환자 Diffusion MRI상 급성 뇌경색 병변 (Volume%) 참조표준

Age	35~44	45~54	55~64	65~74	75~	Total
No	186	568	917	1551	1666	4935
Median	0,1615	0,1107	0,1205	0,1153	0,2174	0,1431
Mean	0,8103	0,9827	1,0213	1,2257	1,9008	1,3697
S.E.M	0,1119	0,1043	0,0776	0,0747	0,1091	0,0481
S.D	1,5259	2,4846	2,3491	2,9402	4,4517	3,3794
Ua(a-type) ¹⁾	0,1119	0,1043	0,0776	0,0747	0,1091	0,0481
U = k*Uc ²⁾	0,2238	0,2085	0,1551	0,1493	0,2181	0,0962
1	0,0	0,0	0,0	0,0	0,0	0,0
10	0,004	0,005	0,007	0,007	0,014	0,009
20	0,022	0,020	0,020	0,021	0,034	0,025
30	0,045	0,038	0,036	0,040	0,061	0,045
40	0,080	0,063	0,063	0,068	0,107	0,076
50	0,161	0,110	0,121	0,115	0,217	0,143
60	0,337	0,236	0,246	0,238	0,435	0,303
65	0,460	0,343	0,377	0,389	0,606	0,450
70	0,607	0,504	0,534	0,571	0,890	0,656
75	0,828	0,716	0,787	0,836	1,278	0,923
80	1,235	0,969	1,140	1,244	2,038	1,347
85	1,548	1,519	1,806	1,901	3,349	2,188
90	2,198	2,278	2,960	3,559	5,393	3,850
95	3,941	4,804	5,323	6,841	10,689	7,398
97	4,520	8,634	7,420	9,995	15,303	10,692
99	7,803	11,995	12,657	14,918	23,257	17,344

단을 하는 의사가 환자에게 쉽게 설명이 됨으로 적극적이고 효율적인 치료가 가능할 것으로 판단된다.

한국인 뇌 MR 영상 참조표준을 제정하기 위해 전국적인 11개 대학병원에서 뇌질환 환자 뇌 MR 영상을 수집(급/만성 뇌경색, 만성허혈, 출혈성병변 등) 하여 병변이 있는 부분을 컬러로 표기하고 병변의 정도를 픽셀수로 계산한다. 그런 후 이미 공개된 표준 Template과 정합하여 병변의 정도를 비율로 나타낸다. 데이터의 신뢰성을 확보하기 위해 신경과 전문의, 의료정보학자, 불확도 및 소급성에 관한 표준 전문가로 구성된 위원회에서 데이터 평가기준서를 제정하여 3차례 이상의 검증과정을 거쳐 참조표준을 제정하게 된다. 참조표준으로 제정한 데이터를 바탕으로 의료 현장에서 쉽게 활용될 수 있도록 (그림 1)과 같이 “한국인 뇌허혈 참조표준 도판”을 제작하였다. 참조표준 도판은 지금까지는 MRI 검사 후 허혈성 뇌손상이 발견되면 그 심한

정도를 ‘없다, 조금 있다, 많다, 아주 많다’ 등 의사의 주관적인 진단으로 이루어졌다면, 표준도판은 1~100 등까지 각각의 표준화 등수에 해당하는 허혈성 뇌손상의 크기와 위치를 동시에 보여주는 참



(그림 1) 한국인 뇌 허혈 참조표준 도판

조 영상 자료를 제시해 줌으로써 연령대 별로 제공되는 뇌허혈 뇌지도를 참조하면, 허혈성 뇌손상 관련 ‘뇌 건강나이’도 추정할 수 있게 된다. <표 4>의 참조표준데이터를 (그림 1)과 같이 의료 현장에서 쉽게 사용할 수 있도록 가공 편집 후 보급한 사례이다. 표준도판은 전국 110여개 병원에서 의료진단 참고데이터로 사용되고 있다.

3. 오픈사이언스 실현을 위한 참조표준 역할

오늘날의 연구는 데이터 집약형(Data-Intensive) 연구실로서 기존에 공개된 데이터를 수집하거나 새롭게 생성한 데이터를 분석함으로써 새로운 과학지식을 창출하는 것 뿐 아니라 학문 분야간 융복합을 통해 문제 해결을 시도하고 있다. 인터넷, SNS 등에서 생산되는 무수히 많은 정형, 비정형데이터를 실제로 사용하기 위한 선별과 활용을 위한 가치 부여에 많은 노력이 필요하다. 국가 연구개발 예산이 2016년 19조 이상이 투자되어 다양한 연구결과물이 쏟아지고 있다. 2004년에서 2014년까지 주요 SCI 학술지에 게재된 논문중 상위 1%에 해당하는 우수논문은 세계 15 위(3,302 건)이며, 논문의 수는 12 위(421,829)로 발표되었다[6]. 그러나 아직까지 논문을 발표하기 위해 다양한 실험을 한 결과데이터의 관리는 거의 되지 않고 있다. 한국과학기술정보연구원(KISTI)의 과학데이터 관리에 대한 조사에 따르면 데이터 생산은 하고 있으나 약 77%가 관리가 되지 않고 있으며, 연구자들이 활발한 데이터 공유를 위해서는 데이터 권리보호 및 인센티브 제공, 데이터 신뢰성 확보를 통한 품질보장, 분야별 특성을 고려한 단계적 추진, 자발적 데이터 개방, 공유문화 정착이 되도록 제도적인 뒷받침이 되어야 공유가 가능할 것이라고 조사되었다 [7].

조사에서 처럼 오픈사이언스 실현을 위해서는

각 연구자, 기관이 보유하고 있는 데이터를 공개하도록 제도와 사회적인 문화가 성숙되어야 할 것이다. 연구자 입장에서는 자신이 어렵게 생성한 데이터를 아무 조건없이 공개하는 것에 대한 것 뿐 아니라 공개된 데이터에 대한 품질에 대한 이용자들의 좋지 않은 평가로 인해 발생할 수 있는 부담을 느낄 수 밖에 없는 것이다.

참조표준은 국가 예산으로 투입되어 개발한 공공의 표준데이터이다. 따라서 누구나 쉽고 편리하게 이용할 수 있도록 공개를 원칙으로 하고 있다. 참조표준은 어떤 측면에서 보면 신뢰도와 정확도가 평가된 Raw data 수준이다. 이용자들은 이러한 참조표준 데이터를 가져다가 2차 가공 및 편집 작업을 통해 목적에 맞게 활용할 수 있도록 공개하고 있다. 과학데이터 현황 조사에서 데이터 공개와 활용의 어려움중에 ‘신뢰성 확보를 통한 품질보장’을 해결하기 위해 국가참조표준체계를 활용하면 가능할 것으로 생각된다. 앞서 밝힌 것처럼 참조표준으로 등록되기 위해서는 엄격한 기준을 통과하여야 한다. 그러나 이미 생산된 많은 과학기술데이터는 이러한 기준에 만족하기는 어렵다. 그렇지만 데이터 평가를 통해 이용자가 신뢰할 수 있을 정도의 품질을 보장해 줄 수 있다.

품질이 부족한 데이터를 활용하게 되면 전혀 다른 결과를 초래할 수 있어 더 큰 문제가 발생할 수 있다. 데이터를 기반으로 한 산업 분야에서도 품질에 대한 중요성을 강조하고 있다. 한가지 예로 반도체 장비 설계용 플라즈마 물성 참조표준의 활용에서 반도체 장비 설계 공정용 시뮬레이터를 국내 중소기업에서 개발하여 실제로 공정에 적용한 결과 정확한 결과 예측 뿐만 아니라 공정기간을 획기적으로 단축할 수 있게 되었다. 핵심은 신뢰할 수 있는 물성데이터이며, 이러한 데이터를 국가에서 지속적으로 제공해 줌으로써 고부가가치의 데이터 기반 산업 활성화가 앞당겨질 것으로 생각된다.

4. 결 론

오픈사이언스 실현의 첫 단계는 국가 R&D 결과에 의해 생산된 많은 데이터의 수집과 활용에서부터 시작되어야 한다. 데이터는 품질관리에 의해 믿고 사용할 수 있어야 한다. 참조표준은 데이터의 객관적인 프로세스에 의해 신뢰성을 부여함으로써 중복실험으로 인한 낭비요소를 줄이고, 데이터를 이용하여 새로운 연구를 할 경우 정확한 결과 값을 얻을 수 있다. 과학기술데이터를 생산 단계부터 신뢰성을 확보하는 것이 필요한데 예를 들면 측정하고자 하는 량 정의, 데이터의 생산이력 정보로서 측정 및 실험장비의 교정(Calibration)을 통해 측정의 소급성 확보, 불확도 표현 등 핵심적인 요소가 포함되어야 사용이 가능하다. 오픈사이언스 실현에서 참조표준의 역할은 데이터의 신뢰도를 객관적으로 보장해 주는 역할 일 것이다.

또한 과학기술데이터는 통합된 플랫폼에서 저장, 등록, 관리보존, 검색, 다운로드 후 분석가공 등 데이터 생애주기 전과정을 국가적으로 관리되도록 지원 하여야 한다. 그러나 이러한 통합운영체계를 갖추기 위해서는 ‘데이터 거버넌스’를 위해 관련 법률정비, 제도 및 정책, 품질, 표준화 등을 크게 추진할 내용을 구분하고 각각의 활동에 대한 중장기 계획이 수립되어야 한다.

오픈사이언스를 비롯한 오픈엑세스, 오픈 소스 등의 용어는 우리에게 이미 익숙한 단어들이다. 뭔가를 공개하고 공유하기 위해서는 먼저 자신이 가진 것을 내놓아야 하고, 그렇게 하기 위한 자발적인 문화 성숙도 필요하지만, 과학자의 참여를 유도할 수 있는 국가적인 제도와 지원이 갖추어졌을 때 가능하다. 이제는 우리도 전통적인 실험실 연구 뿐 아니라 데이터 기반의 스마트한 연구에 많은 관심을 가져야 할 시점이다. 그러기 위해서는 품질이 갖추

어진 풍부한 데이터를 모으고 활용할 수 있는 체계가 갖추어져야 한다.

참 고 문 헌

- [1] 측정불확도 표현 지침서(Guide to the Expression of Uncertainty in Measurement), 한국표준과학연구원, 2010
- [2] 국가표준기본법, 법률 제12925호, 법제처, 2014
- [3] http://www.aist.go.jp/aist_e/list/database/riodb/
- [4] 국가참조표준센터 <http://www.srd.re.kr>
- [5] 2014년 사망원인 통계, 통계청 보도자료, 2015.
- [6] 각국의 SCI 게재 우수논문 분석결과, 한국연구재단, 2015
- [7] 오픈사이언스의 시작, 국가 과학데이터 공유, 참조표준 연구교류회 발표자료, 이상환, 2016

저 자 약 력



채 군 식

이메일 : cks@kriss.re.kr

- 1989년 계명대학교 문헌정보학과 학사
- 1996년 충남대학교 문헌정보학과 석사
- 2006년 충남대학교 문헌정보학과 박사
- 2013년~현재 한국표준과학연구원 국가참조표준센터장
- 관심분야: 과학기술데이터 신뢰성 평가, 참조표준 개발 보급