

과학기술데이터 신뢰성 평가를 통한 참조표준 확립에 관한 연구

A study on establishment of standard reference data through reliability evaluation of science and technology data

참조표준을 생산함으로써 국가고유의 지적자산을 확보할 뿐만 아니라 국가 과학기술의 인프라로 활용될 것이며, 이러한 중요성을 생각해 볼 때 아직까지 과학기술 정보관리와 이용에 대한 낮은 인식을 탈피하여 본격적인 참조표준 구축체계를 갖추기 위해서는 관련 기관을 비롯한 정부의 적극적인 예산과 인력지원이 필요한 시점이라고 할 수 있다.

*채 균 식_한국표준과학연구원 지식정보팀

초 록

과학기술데이터는 엄격한 평가기준에 의해 신뢰성과 정확성이 보장된 참조표준(standard reference data) 값을 지닐 수 있어야 한다. 참조표준은 측정표준을 바탕으로 구하여진 결과 값을 표준화시키는 작업으로서 데이터의 생산조건, 실험환경, 측정방법, 데이터 처리 등이 검토되어 참조데이터(reference data)와 구분된다. 참조표준은 “공인된 수치데이터”로 표현되고 이는 측정결과의 신뢰도를 정량적으로 나타내는 불확도(uncertainty)로 표기된다. 본고에서는 참조표준에 대한 의미를 이해하고, 평가되지 않은 과학기술 데이터가 어떤 평가 과정을 거쳐 참조표준으로 분류되는지를 알아보기 위해 소재물성분야를 예로 살펴보았다.

주제어 : 참조표준, 과학기술데이터, 데이터 평가, 데이터 평가절차

I. 서론

1. 과학기술데이터 평가의 필요성

과학기술 분야의 연구는 점차 다양하고 규모가 방대해지고 있으며, 연구결과로 생산된 정보역시 폭발적으로 증가하고 있다. 이러한 정보들은 일부 체계적인 조직에 의해 관리되기도 하지만 아직도 많은 부분에 있어서 관리의 미흡으로 이용에 어려움이 많다. 이러한 막대한 예산을 투입하여 생산된 국내 과학기술데이터는 연구에 활용하기 위해 자세히 검토해 보면 대부분 정확한 평가를 거치지 않은 신뢰도와 정확도가 결여된 저급데이터이다. 특히 측정에 의해 생산된 과학기술 수치데이터는 활용가능성이 매우 높음에도 불구하고 환경적, 정책적인 요인으로 인해 관리와 활용이 매우 어려운 상태이다.

*한국표준과학연구원 지식정보팀, 책임기술원 (cks@kriss.re.kr)



과학기술데이터는 엄격한 평가 기준에 의해 신뢰성과 정확도가 보장된 데이터로 공인함으로써 반복 사용이 가능한 참조표준(standard reference data) 값을 지닐 수 있어야 한다. 즉, 참조표준이라 함은 측정데이터 및 정보의 정확도와 신뢰도를 과학적으로 분석·평가하여 공인함으로써 국가사회의 모든 분야에서 널리 지속적으로 사용되

거나 반복사용이 가능하도록 마련된 자료로서 물리화학적 상수, 공인된 물성값, 공인된 과학기술적 통계를 말한다.¹⁾

참조표준은 모든 산업분야에 연관되어 있으며 기술수준이 고도화되면서 자체 설계능력을 갖추게 될 때 필요하게 되는 분야이다. 우리나라가 선진국으로 진입하기 위해서는 부가가치가 높은 첨단산업 위주로 개편해야 하며 이러한 첨단산업에서 기술의 성패를 좌우할 수 있는 요소중의 하나가 물질의 특성치를 얼마나 잘 활용할 수 있는가에 달려 있다. 물질의 특성에 관한 자료는 참조표준이라는 용어가 정의되기 이전에 reference book, table, data card, handbook 등 여러 가지 형태로 수 천만건의 정보가 산업체에 이용되어 왔으나 과학기술이 급격히 발전하고 실험자료가 방대해짐에 따라 다음과 같은 문제점이 나타나기 시작하였다. 첫째, 실험의 양이 급격히 증가하였으나 이미 알려진 자료에 대한 정보의 부족으로 같은 수치자료를 얻기 위해 중복실험이 많아지게 되었다. 둘째, 종래의 수치자료들이 여러 곳에 분산되어 수집에 많은 시간과 노력이 소요되었다. 셋째, 비록 필요로 하는 자료를 찾더라도 그 자료에 대한 신뢰도를 알 수 없는 경우가 많았다.

참조표준은 측정표준으로 구하여진 결과 값을 표준화시키는 작업으로서 데이터의 생산조건, 환경, 측정방법, 데이터처리 등이 검토되어야 한다. 따라서 데이터의 불확도(uncertainty) 평가가 제대로 이루어졌는지의 여부를 우선 확인하고, 관련 자료와 비교 검토하여 신뢰성 평가가 이루어진 데이터들이 참조표준으로 활용할 수 있게 된다.

국가 산업규모가 커지고 고부가가치 산업위주로 정책지원이 이루어지고 있는 상황에서 과학기술데이터의 이용이 더 이상 해외데이터에 의존하는 것에서 탈피할 시점이다. 장기적인 투자와 정책으로 한국적인 고유한 과학기술데이터를 갖출 필요가 있다. 고품질의 과학기술데이터로 대변되는 참조표준은 미국을 중심으로 한 선진국에서는 막대한 투자와 체계적인 정책지원으로 국가차원에서 관리되고 있다. 구축된 참조표준데이터는 이를 필요로 하는 국가에 온·오프라인을 통해 고가로 제공하고 있으며 일부 주요 데이터는 공개를 하지 않고 있다. 따라서 우리나라에서도 독자적인 참조표준데이터를 확보하여 국가차원의 지식정보를 구축할 필요가 있다.

1) 국가표준기본법 제3조 7항(정의), 제16조(참조표준의 제정 및 보급), 등 시행령 14조(참조표준의 제정 및 보급), 1999.

1999년 국가표준기본법이 제정되면서 “참조표준의 제정과 보급”에 대해 법제화되어 있는 상태이다. 그러나 법령에는 아직까지 국가참조표준체계 구축에 필요한 세부 시행령이 마련되어 있지 않아 국가참조표준센터 설립, 데이터센터 지정 등 주요사업을 펼칠 수 있는 제도적 장치가 미약한 상태이다. 특히 신뢰성이 평가되지 않은 과학기술수치데이터의 품질관리를 위한 기초연구로 각 분야별 참조표준평가시스템 개발이 선행되어 있어야 한다. 참조표준 구축사업이 국가적인 중요한 과학기술 인프라이기 때문에 장기적으로 정부가 투자해야 하는 사업 분야 중의 하나이다. 최근 들어 국내에서는 6T(IT, BT, NT, CT, ET, ST)기반 고부가 산업으로 움직이고 있기 때문에 과학기술 수치데이터베이스의 필요성에 대한 인식이 점차 고조되고 있다.

신뢰도와 정확도가 보장된 참조표준데이터의 사용은 고정밀 산업설계 등의 과정에서 오류를 최소화하여 공산품의 품질향상과 의료, 보건, 환경 분야의 기초정보로 사용함으로써 삶의 질을 향상시킬 수 있다. 또한 과학기술 연구분야에는 직접적으로 참조표준을 보급함으로써 중복실험을 방지할 수 있고, 부정확한 실험결과에 따른 손실을 줄일 수 있다. 따라서 본 고에서는 범국가적으로 참조표준을 구축, 보급할 경우 고려해야 할 여러 가지 사항을 살펴보고, 참조표준을 구축하기 위해 가장 중요한 과학기술데이터의 신뢰도와 정확도를 평가하기 위해 소재물성 분야를 대상으로 평가시스템을 예를 들어 분석하였다. 그리고 과학기술자뿐만 아니라 관련자들의 참조표준에 대한 부족한 이해도를 높이고, 선진국들의 정보유출 제한에 대응하기 위해 국가적 차원의 지적 자산을 확보하기 위한 방안을 살펴보았다.

II. 참조표준의 분류와 국·내외 동향

1. 참조표준의 분류

과학기술데이터는 각 분야별 전문가들에 의해 구축된 엄격한 평가 기준에 의해서 신뢰도와 정확도를 보장받을 수 있다. 평가과정에서 고려될 여러 가지 요인들을 검증함으로써 데이터의 등급이 주어지게 된다. 참조표준의 영역은 기초 물리상수뿐만 아니라 과학적 수치 데이터 모두를 포함하고 있다. 참조표준과 관련된 국제기구인 “과학기술 데이터 위원회”(CODATA, Committee on Data for Scientific and Technology)에서는 <표 1>과 같이 참조표준을 3종류로 구분하고 있다.

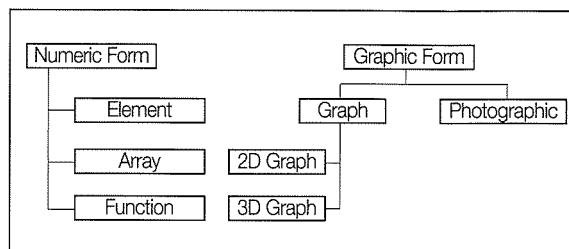
데이터 종류	특성	예
Class A	잘 정의된 시스템에 의해 얻어지는, 재현성 있는 데이터(repeatable measurement on well defined system)	물리, 화학, 재료 등
Class B	관측에 의해 얻어지는 데이터 (observational data)	생물, 천문, 자원, 지구 과학 등
Class C	조사에 의해 얻어지는 통계적 데이터 (statistical data)	유전자, 인체 골격, 심전도 등

<표 1> 참조표준의 분류

이중에서 과학기술정보의 가장 상위부류인 기초상수와 같은 Class A 정보는 과학기술의 근간이 되는 정보가 될 것이지만 수요도 측면에서는 Class B, C도 중요한 정보로 취급되어야 한다. 마찬가지로 이들 데이터에 대한 정확한 평가로 신뢰도를 확보하여 산업에 널리 사용될 수 있도록 해야 한다.

참조표준의 핵심인 신뢰도의 정량적인 평가는 불확도로 표기된다. 불확도는 측정결과의 신뢰도를 나타내는 정량적인 지표로서 데이터를 어느 정도까지 믿을 수 있는가를 보여준다. “공인된 수치데이터”라고 참조표준을 표현한다면 수치정보가 평가되고 공인받음으로써 생산된다는 것을 고려할 때 각 데이터의 불확도가 포함되지 않으면 데이터의 의미는 크게 상실될 수밖에 없다.²⁾ 측정결과와 신뢰성은 정확성과 관련된 소급성(traceability)에 의하여 확보된다. 소급성이란 “명시된 불확도를 갖고 끊이지 않는 비교사슬을 통하여 통상국가 또는 국제적 표준과 연관될 수 있게 하는 측정결과 또는 표준값의 특성³⁾”으로 규정하고 있다. 즉, 측정결과와 정확성을 불확도 기준으로 확인하게 되는데 이는 측정표준 대표기관과 연계하여 정확성 여부를 확인한다는 의미이다. ISO/IEC 17025⁴⁾의 기술요건 항목을 보면 측정의 소급성 유지에 대한 기본지침이 상세히 기술되어 있으며 이는 궁극적으로 불확도를 나타낸 것이다.

참조표준은 크게 특성적, 형태적, 기능적 관점으로 분류할 수 있다. 첫째, 특성적(property)분류는 특성에 따라 크게 4가지로 분류되는데 물질의 특성 및 성능데이터인 기본상수와 열역학적 특성, 열화학적 특성 등 열적성질, 전기적 특성, 자기적 특성, 광학적 특성 등 전자기적 성질, 기계적 특성, 마찰 및 윤활 특성 등 기계적 성질, 중성자 단면, 핵구조 등 핵공학 및 초전도 관련 등의 데이터가 이에 속한다. 그리고 구조(spectral & structural)데이터로는 물질의 구조와 존재를 확인하기 위한 데이터로 결정대칭(crystal symmetry), 분자구조, 질량분석(mass spectra), 엑스레이분석(x-ray spectra) 등이 있다. 상호작용데이터로는 물질의 상호작용에 관한 것으로 가스화학 반응율(gaseous chemical reactions rates), 동력학(kinetics) 등이 있다. 다음은 생물·의료 관련데이터로 의료용 인증표준물질(certified reference materials)의 물성데이터, 유전자정보데이터 등이 있다.



〈그림 1〉 형태별 참조표준의 종류

둘째, 형태적 관점에서의 분류는 수치형태데이터로 대부분의 데이터는 이와 같은 형태이며 수치, 단위 및 오차로 표현하면 다음과 같은 종류가 있다. 요소 형태의 자료로 오차한계를 지니는 값(예 : 빛의 속도), 배열형태의 자료로 2개의 요소가 어울린 형태(예 : JANAF table에서 온도), 기능형태의 함수식형태로 보관되는 자료(예: gibb's free energy of formation)이다. 그래픽 형태데이터로는 판독성을 증가시키기 위해 수치데이터를 사용하여 이해가 쉽도록 만든 것이며, 그래프에는 2차원 및 3차원 그래프가 있다.

셋째, 기능적 관점으로는 크게 3가지로 분류되는데 사실정보 형태로 엄격히 평가된 자료 자체를 직접 이용할 수 있는 숫자형태의 자료, 서지자료로 상기의 자료를 생산하는데 사용한 문헌의 원본을 수록하고 있는 자료, 데이터베이스로 상기의 자료들을 재가공하여 검색 시간 단축 및 이용성을 높인 자료이다.

2. 국내동향

국내의 경우 참조표준 제정 및 보급에 관한 규정이 국가표준기본법⁵⁾에 상세히 기술되어 있어 구축에 필요한 제도적 장치는 모두 마련되어 있다고 볼 수 있으나 정부 차원의 구체적인 지원 정책이 시행되지 않아 참조표준 연구활동 및 데이터 신뢰성 평가에 대한 연구는 미약한 실정이다.

참조표준은 물리상수와 같은 순수 기초과학에 속하는 것도 있으나 산업체에 직접적으로 활용될 수 있는 재료나 물질에 대한

2) Qian Dong, et al., "Data quality assurance for thermophysical property databases- applications to the TRC SOURCE Data system", Journal of Chemical Information Computer Science, Vol.42 No.3(2002), pp. 473-480.

3) 정낙삼 등. 국제측정학용어집(VIM), KRISS/SP-2003-069, 2003.

4) ISO/IEC17025, "General requirements for the competence of testing and calibration laboratories.", 1999.

5) 국가표준기본법시행령 (일부개정 2000. 6. 23 대통령령 제16851호)제14조에 참조표준의 제정 및 보급에 대해 다음과 같이 상세히 기술하고 있다. 1. 측정데이터의 수집·분석·평가 체계의 확립 2. 참조표준의 제정절차의 수립 및 사후관리 3. 참조표준 및 측정데이터의 축적 및 보급체계 확립 4. 측정표준과의 소급성 체계 유지 5. 또한 표준과학연구원으로 하여금 참조표준에 관한 측정데이터의 수집, 축적 및 평가에 관한 사업을 할 수 있도록 규정하고 있다.

것도 많다. 참조데이터와 참조표준은 산업의 모든 분야에 걸쳐 활용될 수 있으며 대표적으로 재료나 물질에 관련이 깊은 철강산업, 비철금속산업, 석유화학산업, 파인세라믹산업, 부품소재산업 등을 들 수 있다. 또한 최근에는 국민의 건강과 환경에 대한 참조표준의 수요가 증가하고 있는 실정이다. 이러한 다양한 국내 수요에도 불구하고 아직까지 수요를 충족시킬 수 있는 참조표준이 없을 뿐만 아니라 참조표준평가시스템 구축조차 확립되어 있지 않은 상태이다.

최근 성장 동력 산업분야에 연구비 지원이 집중되고 있으며 2004년도 미래첨단 기술분야의 연구개발 예산은 17,839억원으로 전년도 대비 6.1% 증가한 투자규모이다. 이처럼 산업 전 분야에 걸쳐 참조데이터 및 참조표준은 기초정보로 활용되어야 함에도 불구하고 관련 연구결과로 생산된 과학기술데이터의 구축 뿐만 아니라 이를 효율적으로 이용하기 위한 체계가 마련되어 있지 않은 상태이다.

2.1 참조표준 국내 수요

국내에서는 아직까지 과학기술 실험데이터의 체계적인 수집, 가공, 축적, 보급에 대한 체계가 수립되어 있지 않아 산업이나 과학기술 발전 및 국가경쟁력 강화에 심각한 장애요인으로 대두되고 있다. 그리고 연구개발비용이 지속적으로 증가하고 있지만, 부정확한 실험데이터의 생산과 투자의 중복으로 낭비가 크다. 이와 같이 부정확한 실험데이터가 생산현장에서 직접 사용하게 될 때는 제품의 품질이 저하되고 경쟁력을 잃게 됨으로써 막대한 손해가 예측된다. 따라서 참조표준에 대한 국가적인 관리 체계 구축이 연구 기술 개발 못지않게 중요한 요소이다.

2003년도에 실시한 참조표준 수요도⁶⁾ 조사에 의하면 응답자 3,377명 중 연구자의 2,364명(70%)정도가 참조데이터 또는 참조표준이 과학기술 인프라로 구축할 가치가 있으며, 구축되었을 경우 연구개발에 참조표준을 지속적으로 이용할 계획인 것으로 나타났다. 또한 참조표준을 구축하는데 중요한 요소로는 내용의 정확성 1,080명(32%), 이용의 편리성 912명(27%), 이용자의 활용도 607명(18%) 순으로 응답하였다. 이는 참조표준이 다른 정보에 비해 내용의 정확성과 신뢰성을 생명으로 하고 있다는 것이다. 연구자들이 가장 우선적으로 선호하는 분야는 화학, 재료분야의 참조표준데이터이다. 이는 첨단산업이 화학, 재료 등의 기반산업을 바탕으로 하여 발전한다는 사실과 과학기술분야의 기본 인프라 자원이라는 사실에 기인한다.

그리고 참조표준 관련 정보중에서 가장 우선적으로 구축해야 할 분야는 화학물질정보 296명(8.8%), 열물성정보 249명(7.4%), 반응 및 합성정보 236명(7.0%), 무기결정 구조정보 233명 (6.9%) 물리상수 정보 220명 (5.6%) 순으로 나타났다. 본 조사에서는 <표 2>와 같이 20개 분야를 설문자에게 제공하여 시급하게 필요한 분야를 복수로 선택하게 하는 방식으로 조사되었기 때문에 분야별로 응답자 수가 다르게 나타났다. 예를 들면 열물성 분야 응답자는 전체 249명(7.4%)이고 그중에서 가장 우선적으로 구축해야 한다고 응답한 것이 65명(2%), 그 다음 순으로 구축해야 한다고 응답한 응답자 수는 34명(1%)으로 나타났다.

6) 한국과학기술정보연구원(KISTI)에서는 참조표준 수요도 조사를 2003년에 실시하여 3377명으로부터 응답을 받았으며 구성은 기업체가 60%, 연구소 및 대학 35%, 기타 5% 정도가 응답하였다. 전공분야는 화학, 재료, 전기전자, 금속, 물리 등 참조표준이 생산되는 분야이며 참조표준을 실제로 사용하는 이용자들이므로 의견이 충분히 반영된 것으로 볼 수 있다.

내용	순위																				합계
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
화학물질 정보	81	82	45	33	19	10	8	5	5	1	1	1	1	1	2	0	1	0	0	0	296(8.8%)
열물성 정보	65	34	28	27	21	19	11	5	8	10	5	4	0	3	2	2	1	3	1	0	249(7.4%)
반응/합성정보	29	44	51	29	26	17	10	9	4	7	0	3	2	2	1	2	0	0	0	0	236(7.0%)
결정구조정보	42	37	33	23	19	16	17	9	11	7	5	2	5	3	2	0	1	0	0	1	233(6.9%)
물리상수정보	38	29	30	24	19	14	12	7	10	6	3	4	1	7	4	4	3	1	1	3	220(6.5%)
스펙트럼 정보	24	27	21	22	19	24	7	13	12	5	6	6	2	1	3	3	1	3	2	2	203(6.0%)
독성/안전성	13	26	28	22	21	12	12	10	1	9	4	11	5	6	3	3	3	1	2	0	192(5.7%)
촉매물질정보	8	16	27	21	21	20	16	16	11	8	5	6	5	4	2	1	1	0	0	0	188(5.6%)
폴리머정보	35	23	11	16	11	14	11	5	3	9	4	5	7	6	3	2	4	1	3	2	175(5.2%)
시약/카탈로그	4	16	15	22	19	10	12	15	11	8	14	4	6	3	0	0	3	2	3	3	170(5.0%)
3D분자구조정보	4	12	12	10	12	10	13	13	15	11	4	9	5	3	7	2	1	4	1	5	153(4.5%)
의약 정보	20	13	6	15	8	6	9	10	4	5	4	8	4	8	17	3	7	5	1	0	153(4.5%)
생약정보	13	14	6	7	5	5	2	2	4	8	5	2	1	4	10	9	19	9	4	1	130(3.8%)
식품첨가물 정보	7	4	9	9	10	3	5	6	6	1	8	7	5	18	4	5	8	4	4	2	125(3.7%)
액정정보	5	8	11	6	8	5	9	3	10	4	10	6	7	5	5	3	2	8	3	3	121(3.6%)
내분비 교란물질	3	8	2	7	7	4	7	5	4	4	8	2	10	3	4	18	8	8	5	1	118(3.5%)
방사선 동위원소	4	7	7	2	2	0	0	4	3	5	4	2	4	3	2	5	9	8	34	2	107(3.2%)
다이옥신	1	7	1	3	4	3	6	4	2	4	4	8	1	1	7	10	9	21	4	5	105(3.1%)
농약/살충제 정보	1	6	5	0	3	4	1	3	4	2	4	6	20	5	9	8	4	3	6	9	103(3.1%)
핵반응정보	3	7	1	4	1	1	3	1	4	0	4	3	1	3	2	7	2	5	10	38	100(3.0%)

〈표 2〉 참조표준 수요 조사

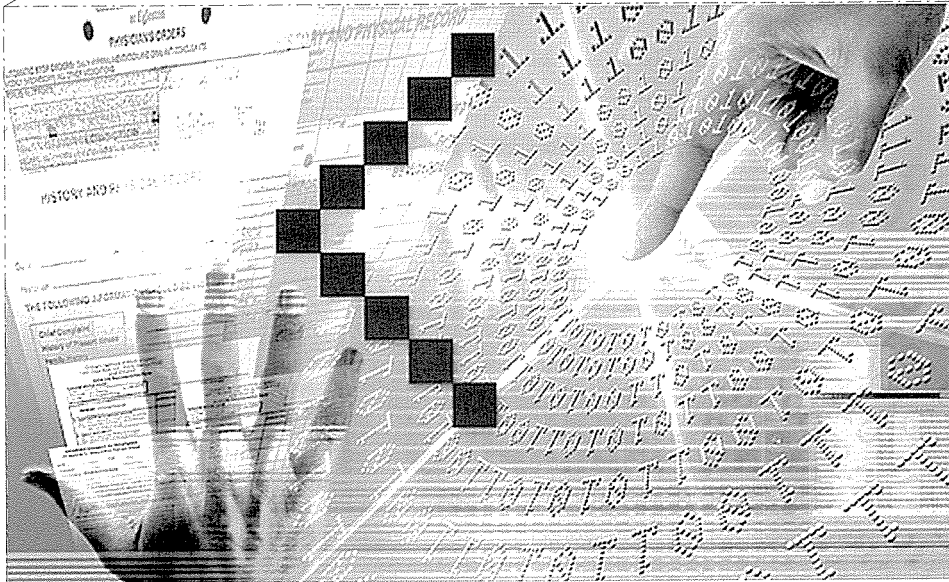
3. 해외동향

3.1 미국

참조표준 활동을 활발히 전개하고 있는 나라는 그 특성상 근대 과학기술의 역사가 긴 선진국들에 국한된다. 이 중에서 가장 오랫동안 참조표준을 축적해온 나라는 독일이라고 볼 수 있으며 재료물성 핸드북의 경우 100여 년간 걸쳐 수백 권으로 데이터를 축적해 오고 있다. 미국의 경우 과학기술데이터 축적에 대한 필요성을 인식하여 이러한 활동을 전개하였는데 유럽지역의 참조표준 활동과 달리 단순히 과학기술데이터를 축적만 하는 것이 아니라 데이터의 신뢰성을 평가하였다는 것이다. 따라서 진정한 의미의 참조표준은 미국에서 시작하였다고 볼 수 있다.

과학기술에 사용되는 다양한 물질의 물리화학적 특성에 대한 수치데이터는 연구개발, 공학적 설계, 상업적 응용을 위한 편집과 평가가 필요하다. 이러한 참조표준의 필요성 인식이 1963년 미국 연방정부의 연방과학기술위원회(Federal Council for Science and Technology)에서 “국가참조표준데이터시스템에 관한 연방정책”(National Policy on National Standard Reference Data System)을 수립하면서 시작되었다. 당시 과학기술 수치데이터를 엄격하게 평가하고 신뢰도를 보증하기 위한 업무를 담당할 적절한 기관으로 국립표준기술원(NIST-National Institute of Standard and Technology)을 주관기관으로 선정하였다. 참조표준의 본격적인 활동은 1968년 미 의회에 의해 제정된 Public Law 90-396(Standard Reference Data Act)으로 합법화하여 본격적인 업무가 시작되었고, NIST는 “표준참고자료 프로그램(Standard Reference Data Program)⁷⁾”에서 전체적인 업무를 관장하고 있다.

7) NIST SRD program의 특징은 공개된 문헌들에서 참조데이터를 수집하여 잘 갖추어진 평가절차와 기준에 따라 전문가에 의해 평가한다. 평가결과에 의해 참조표준데이터도 있지만 참조데이터도 제공함으로써 수요자의 요구에 부응하고 있다. 연간 예산은 약 \$3,000만 정도가 지원되고 있다.



NIST SRD 프로그램은 에너지, 환경, 보건, 산업디자인, 재료의 내구성, 자원의 재활용 등에 응용하기 위한 과학기술 수치데이터를 수집하여 그 신뢰도를 평가한다. 이 프로그램의 목적은 “신뢰성 있는 평가된 수치데이터의 보급”이다. 구체적으로 NIST SRD 프로그램의 목적은 다음과 같이 크게 3가지로 요약할 수 있다. 첫째, 편리하고 접근이 용이한 형태로 과학기술분야 종사자들에게 엄격하게 평가된 수치데이터를 제공하는 것이다. 둘째, 측정의 신뢰도를 향상시키는데 도움이 되도록 데이터를 평가하는 과정에서 얻어진 경험을 널리 알리는 것이다. 실험자들은 자신이 수행한 실험결과와 NIST에서 평가한 결과를 비교함으로써 실험의 절차나 방법을 개선할 수 있을 것이다. 셋째, 과학기술 수치데이터의 정확한 평가를 거친 참조표준을 제공함으로써 미국의 산업경쟁력을 강화하는 것이다.⁸⁾⁹⁾

데이터의 수집, 생산, 평가, 보급이 초창기에는 SRD 사무국에서 일괄 처리하였으나 최근에는 데이터센터별로 데이터의 수집, 가공, 생산, DB 설계를 담당하고 표준보급센터(Office of Measurement Service)에서는 검색소프트웨어 개발, 웹 기반의 정보시스템 개발, DB유동, 홍보 등의 업무를 수행한다. 특히 연구부서에 있는 데이터센터를 지정할 경우 다음과 같은 사항이 지정의 고려대상이 된다. 구축하려는 데이터의 수요는 있는가? NIST만이 그 데이터를 구축할 수 있는가? 시설 인력 등으로 구축 능력은 있는가? 구축이 완료되었을 경우 산업계를 비롯한 관련 분야에 영향력이 얼마나 되는가?(Lide, Jr., David R, 1984) 등이 주요 검토사항이다.

NIST가 미국내 참조표준에 대한 주요기관이기는 하지만 기타 연방정부(DoD, DoH, NASA, NSF 등)와 밀접한 협력을 통해 범국가적 차원에서 수행하고 있으며 개인이나 학술단체(ASM International, ACS, American Ceramic Society, ACTIC, Gas Research Institute 등)와 협력하고 있다.¹⁰⁾

8) L. Franz, "Information Handling in the National Standard Reference Data System," (NBS Technical Note 290, 1966)

9) Edward L. Brady, Wallenstein A. "National Standard Reference Data System Plan of Operation," (NSRDS-NBS, 1964)

10) NIST이외의 대표적인 과학기술데이터센터는 Center for Information Numeric Data Analysis and Synthesis(CINDAS), Coblenz Society, National Nuclear Data Center, Institute for Materials Science and Engineering, Mossbauer Effect Data Center, Thermodynamics Research Data Center 이다.

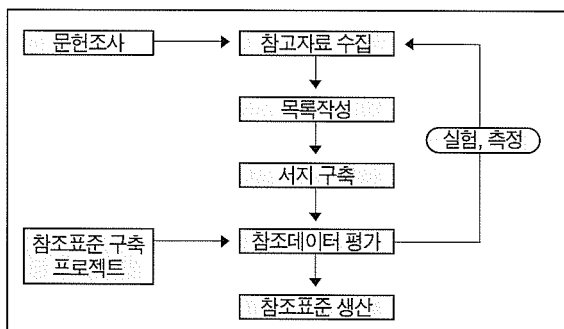
A standard reference method of standard reference data through preliminary evaluation of science and technology data

현재 구축된 참조표준 데이터베이스는 약 200여종에 이르며 특별히 주제별로 구분되어 있지는 않다. 구축된 데이터베이스는 미국화학회와 미국물리화학회가 공동으로 발간하는 “Journal of physical and chemical reference data”를 통해 수록되거나, “NIST standard reference data report”, “NIST special report”를 통해 발행된다. 최근에는 온라인을 이용하여 구축된 데이터를 판매하고 있으며 웹 인터페이스는 NIST data gateway를 통해 제공된다. 구축한 참조표준의 적극적인 활용을 위해 측정기기 산업체와 연계하여 보급하게 되는데 예를 들면 질량분석기와 같은 첨단 측정 장비에 참조표준 데이터를 탑재하여 실 수요자에게 제공하기도 한다.

NIST의 참조표준 데이터 평가방법은 학술잡지에 공개된 과학기술 수치데이터를 수집하여 평가한다. 따라서 참조표준 데이터를 직접 생산하기 보다는 공개된 데이터를 수집하여 평가함으로써 소요예산이 적게 들며 단기간에 많은 량의 데이터를 확보할 수 있는 장점이 있다. 데이터 평가 전문가들은 관련분야 데이터를 지속적으로 수집하여 측정방법, 재료의 특성, 사용규격 등을 비교·분석하여 결과를 통계적으로 처리함으로써 데이터의 신뢰도를 평가할 수 있는 기반을 수립하는데 대략 10년 정도 걸린다고 한다.

객관적인 신뢰성을 판단하기 위한 주요 평가기준으로 첫째, 데이터의 생산과정이 실험절차나 표준규격과 얼마나 잘 부합하는가 인데 여기에는 모든 관련변수는 확인되었는가? 데이터는 통계처리 되었는가? 증명은 어떻게 하였는가? 둘째, 생산과정이 잘 명기 되었는가? 셋째, 물리적 법칙과 일치하는가? 넷째, 다른 측정방법이나 계산결과와 비교하여 어떤가? 등이다. 이러한 과정을 살펴보면 데이터의 생산과정 뿐만 아니라 이론적인 예측 값과도 비교함으로써 데이터의 신뢰도를 이중으로 확인하고 있음을 알 수 있다.

NIST SRD Program에서는 <그림 2>와 같이 공개된 문헌에서 데이터를 수집하여 일반적인 서지정보를 구축하게 된다. 구축된 서지정보는 평가를 위한 기초정보로 이용하게 된다. 이 중에서 가장 중요한 검토 요소는 데이터 생산 환경과 조건이다.



<그림 2> NIST SRD 데이터 구축 과정

3.2 일본

일본에서는 통상산업성 지원으로 1972년 RIPS(Research Information Processing System) program을 통하여 참조표준과 관련한 타당성 조사를 실시하였으며, 1973년 3월 에 “Advanced Information Processing of Large Scale Data Over a Board Area”라는 시스템을 개발한 것이 국가적 차원의 참조표준 체계 확립 노력의 시초라고 할 수 있다.

현재 통상산업성은 참조표준과 관련하여 AIST(National Institute of Advanced Industrial Science and Technology)기관 산하 연구소인 NMIJ(National Metrology Institute of Japan), 및 NIMS(National Institute of Material Science)등 3개 기관을 집중적으로 지원하고 있다. NMIJ는 최근 대형 국가과제로 “고체연료의 분산형 열물성 DB”를 구축하고 있다. 최근 경향은 고체뿐만 아니라 유체 분야의 DB 개발에도 눈에 띄게 증가하고 있다. AIST에서 개발한 RIO-DB(Research Information Data Base)는 다양한 연구분야의 멀티미디어 데이터베이스이며 AIST 소속 연구기관들의 연구 산출물을 데이터베이스한 최대 규모 일반 과학기술데이터베이스이다. RIO-DB의 목적은 학술분야 및 산업체에 과학기술의 발전을 도모하기 위해 인터넷을 통해 각종 정보를 보급하는 것이다. 이 DB는 TACC(Tsukuba Advanced Computing Center)에서 관리하고 있으며 지구과학, 화학, 에너지, 침적물, 재료, 정보기술, 생물 및 안전에 관한 정보를 수록하고 있다. NIMS는 재료관련 분야의 DB를 전문적으로 구축하고 있으며 폴리머, 고자장공학(high magnetic field engineering), 핵재료(nuclear materials), 결정구조(crystal structure), 전자구조(electronic structure) 등 11종의 DB를 개발하고 있다.

3.3 유럽

유럽국가들은 EU를 중심으로 CRDIS(Community Research and Development Information Service)를 설립하여 연구 활동에 대한 통합정보서비스를 실시하고 있으며 CODATA DB의 지적재산권 부여에 관한 것을 ICSU에 상정하는 등 CODATA 활동에 적극적으로 참여하고 있다. 독일의 경우 과학기술데이터의 관리가 가장 활발히 이루어지고 있으며 특히 재료분야 경우 100년 이상 많은 투자와 노력으로 데이터 복을 생산해 오고 있다. 유럽지역에서의 참조표준 활동은 미국과 달리 데이터에 대한 직접적인 평가를 하지 않고 수집에만 주력하는 것이 차이점이라 할 수 있다.

과학기술데이터에 대한 대표적 유통기관인 Fiz-Karlsruhe(Fachinformationszentrum Karlsruhe)는 학계와 산업계에 과학기술데이터를 제공하고 있다. 미국과 일본의 협력으로 운영되는 STN망을 통해 과학기술 정보를 제공하고 있고, 미국과 일본에 있는 STN 분소는 유럽의 STN센터로서의 역할을 하고 있다. 약 200개 이상의 데이터를 제공하고 있으며 그 중에서 참조표준과 관련한 데이터로는 ENTEC(The Germany Energy database), ENERGY, INIS(Nuclear Research Information Technology), ICSD(Crystallography) 등이 있다.

III. 참조표준데이터 품질관리 방안

1. 데이터 품질평가 기준 및 절차

참조표준을 생산하기 위한 데이터 평가기준은 분야별 특성에 따라 조금씩 다르다. NIST에서는 일반적으로 모든 분야에 공통적으로 적용할 수 있는 기준으로 데이터 생산과정에 대한 절차, 변인검토, 생산과정 표기, 적절한 통제, 다른 측정법과 비교 등을 검토하게 된다. 각 분야별 데이터센터에서 생산한 참조표준 데이터는 상위 조직인 전문위원회에 평가결과를 의뢰하여 심의를 거쳐 참조표준센터에서 최종 등록을 하는 절차이다. 최종 참조표준으로 등록하기에 앞서 이해 당사자간의 의견을 조정하기 위해 공개적으로 참조표준 대상 정보를 인터넷에 공개할 필요가 있다. 수집한 평가 결과는 아래 <표 3>와 같이 데이터 평가 등급이 부여된다.¹¹⁾

11) Munro, Ronald G. Data evaluation theory and practice for materials properties., (NIST Special Publication 960-11 2003), p.22.

등급	내용
인증(certified)데이터	참조표준값
검증(validated)데이터	상호연관성 및 모델링에 의해 확인된 데이터
유효(qualified)데이터	기본 수용기준을 만족하는 데이터
상용(commercial)데이터	제조업자의 데이터
일반(typical)데이터	유사한 재료들의 일반적인 경향을 보여 주는 데이터
연구(research)데이터	개발중인 소재 데이터
미평가(unevaluated)데이터	기타

〈표 3〉 데이터 평가 등급

참조표준으로 확립하기 위해 일반적으로 고려해야 할 평가기준으로는 첫째, 제3자에 의해서 실험방법의 재현이 가능하도록 시험방법과 절차가 자세히 명기되었는지를 검토한다. 최종 데이터가 생산되기까지 사용한 규격은 무엇이고, 실험실의 온도와 주변 환경은 적절한가 등을 검토하게 된다. 둘째, 시험방법의 불확도 평가 및 소급성의 유지 여부를 검토한다. 실험데이터의 신뢰성을 갖기 위해서는 측정결과와 관련하여 측정량을 합리적으로 추정된 값의 분산 특성을 나타내는 파라미터로 불확도를 평가하고, 정확성을 보장하기 위해서는 명시된 불확도를 갖고 끊이지 않는 비교사슬을 통하여 소급성을 유지한다.¹²⁾ 셋째, 시험방법의 적절성 및 한계를 명기한다. 넷째, 시험방법에서 결과에 영향을 줄 수 있는 변수의 적절한 제어 여부를 검토한다. 다섯째, 다른 방법에 의해 얻어진 결과와 일관성 여부를 검토한다. 여섯째, 이미 알려진 법칙과 위배되는지를 검토한다. 일곱째, 잘 알려진 연관식과 모델링에 대한 예측가능성을 검증한다. 여덟째, 실험결과와의 상충과 쟁점 현안에 대한 전문가 평가를 검토하게 된다. 이러한 평가기준에 의해 생산된 데이터를 과학기술데이터베이스로 구축할 때 이용자 측면에서 본다면 다음과 같은 기준으로 구축되어야 한다. 이용자 질의에 어떤 데이터를 보여줄 것인가에 대한 검색의 정확성(relevance), 이용자들이 찾는 데이터를 얼마나 많이 보유하고 있는지에 대한 포괄성(comprehensiveness), 접근할 수 있는 데이터가 얼마나 편리하게 구성되어 있고 재현성이 있는지에 대한 기준으로 신뢰성(reliability)이 검토된다.¹³⁾

데이터의 평가기준 및 절차에 있어서 데이터 평가과정에서 상호보완적으로 적용되는 두가지 원칙이 있는데 하나는 데이터의 신뢰도를 확인하는 것과 다른 하나는 수집된 데이터를 충분한 이유없이 배제해서는 안된다는 것이다. 이러한 것을 기본원칙으로 하여 데이터를 평가하는데 고려해야 할 사항은 다음과 같다. 첫 번째는 데이터의 재현성으로 여러 독립적인 연구과정에서 동일한 결과가 얻어져야 한다. 그 다음으로는 상호연관성이 있는 데이터들 간에 관계와 이론적인 모델을 활용하여 평가하는 과정을 거쳐 검증하는 것이다.¹⁴⁾

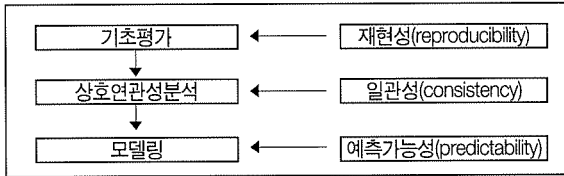
데이터의 평가과정은 3단계로 나눌 수 있다. 최초 평가단계는 기초평가로서 평가의 주안점은 재현성에 있다. 기초평가단계에서 추가되는 기준은 정체성(identification), 기술성(description), 양립성(comparability)으로 비교순회시험을 주로 이용하여 측정방법의 상대적 신뢰도를 평가하고 측정된 값의 정량적 신뢰도를 평가한다. 그 다음 단계는 상호연관분석단계로 일관성에 주안점을 두고 평가한다. 마지

12) 정낙삼 등. 측정불확도표현지침, KRISS-99-070-SP, 1999.

13) Peter, Kuhn, Rene Deplanque, Ekkehard, Fluck "Criteria of quality assessment for scientific database", Journal of Chemical Information Computer Science, Vol.34 No.3(1994), pp. 517-519.

14) Ibid., p. 518

막 단계로는 예측가능성인데 이론적 모델을 근거로 하여 예측한 결과와 데이터가 일치하는가를 기준으로 평가한다.



〈그림 3〉 참조데이터의 평가과정

데이터의 등급부여는 데이터의 성격이나 상황에 따라 달라질 수 있으나 어떤 기준이 적용되든지 간에 “수용가능(acceptable)” 과 “수용불가(unacceptable)” 로 나누어진다. 수용불가라는 데이터는 틀렸다는 의미가 아니라 다른 데이터들과 호환성이 없다는 의미이다. 예를 들면 A라는 재료의 경도가 31이었고 B라는 재료의 경도가 45이었다면 B가 A보다 단단하다는 것은 알 수 있어도 어떤 기준으로 경도가 사용되었는지 알 수가 없으므로 데이터의 호환성이 없으며 따라서 수용이 불가능한 데이터가 된다.¹⁵⁾

〈그림 4〉는 참조데이터가 일반적인 평가기준과 절차에 의해 참조표준으로 생산되기까지의 과정을 나타낸 것이다. 각 단계별로 데이터의 상태에 따라 〈표 3〉의 데이터 등급이 주어지게 된다. 공개된 학술정보원인 논문, 학회발표 자료 등에서 수록한 결과 데이터를 수집하여 〈그림 4〉와 같은 절차로 평가를 해 보면 많은 논문들이 결과데이터가 얻어지기까지 과정을 수록하지 않는 경우가 많아 평가 자체가 어려운 경우가 많다. 그러나 최근에는 논문심사가 보다 엄격해짐으로 인해 이러한 문제는 다소 해소되었다고 할 수 있다.

과학기술자들은 어떤 특정재료에 대한 정확한 특성값을 찾을 경우 공개된 자료를 주로 이용하게 되는데 자료마다 조금씩 결과값이 차이가 나기 때문에 보다 정확한 값을 얻기 위해서는 다양한 자료를 수집해서 판별해야 한다. 따라서 이러한 값들에 대해 정확히 표준값을 인증해 주는 것이 참조표준의 역할이라고 할 수 있다.

〈그림 4〉는 수집한 참조데이터의 일반적인 평가절차를 나타낸 것이다. 평가절차에 따라 수집한 참조데이터는 재료규격의 명시 여부에 따라 평가대상 데이터가 되지만 그렇지 않은 경우는 평가를 할 수 없는 기각데이터가 된다. 연구데이터는 측정방법이 명시되고 연구가 종료된 데이터이며 측정방법이 명시되지 않은 데이터는 상용데이터나 평가 불능의 데이터가 된다. 재현가능성과 일관성이 있는 데이터는 유효데이터로 분류되며 데이터를 보고 결과를 예측 가능한 데이터는 검증된 데이터로 본다. 이러한 과정을 거쳐 최고 우위에 있는 데이터는 참조표준으로 등록된다.

2. 데이터 품질향상 사례 - 소재물성 분야

소재 물성 데이터는 설계기술, 소재의 신뢰성, 설비의 건전성 관리에 필요한 인프라이다. 소재물성 데이터 평가는 많은 시간과 비용이 소모되는 반면 사업성이 낮기 때문에 일반 산업체에서는 추진하기 어렵고 국가적 차원에서 집중적이고 체계적으로 추진해야 한다. 국산제품이 점차 고정밀도가 요구되면서 부가

¹⁵⁾ A. J. Barrett, "Data evaluation, validation and quality" ASTM manual on the building of material databases, Crystal H, Newton, Ed., (ASTM manual series MNL1993), pp. 53-67.

가치가 높은 제품생산에 주력하고 있다. 그러나 국내의 경우 국산소재에 대한 물성DB가 없으므로 외국의 유사 재질DB를 사용해야 하기 때문에 설계 최적화와 내구수명 예측이 어려워 신뢰성 하락으로 대외 경쟁력이 취약한 상태이다. 따라서 부품설계 단계에서는 정밀설계기술, 부품 생산단계에서는 신뢰성 평가기술, 완성품사용단계에서는 건전성 평가기술이 필요하다. 이러한 기술단계마다 소재물성 데이터가 필수적인 요소이다.

〈표 4〉는 소재물성을 평가하는 기초평가 항목과 세부평가 항목을 구분하고 있다. 세부 평가 분야의 경우 열처리온도의 영향, 부하속도의 영향, 온도의 영향으로만 검토해 보아도 소재간의 특성 차이가 있다. 국내의 경우 동일강종, 동일제품이지만 시험기관에 따라 물성의 차이가 있다. 국외의 경우에서도 같은 강종과 같은 시험기관 소재의 제작사에 따라 물성차이가 있다.

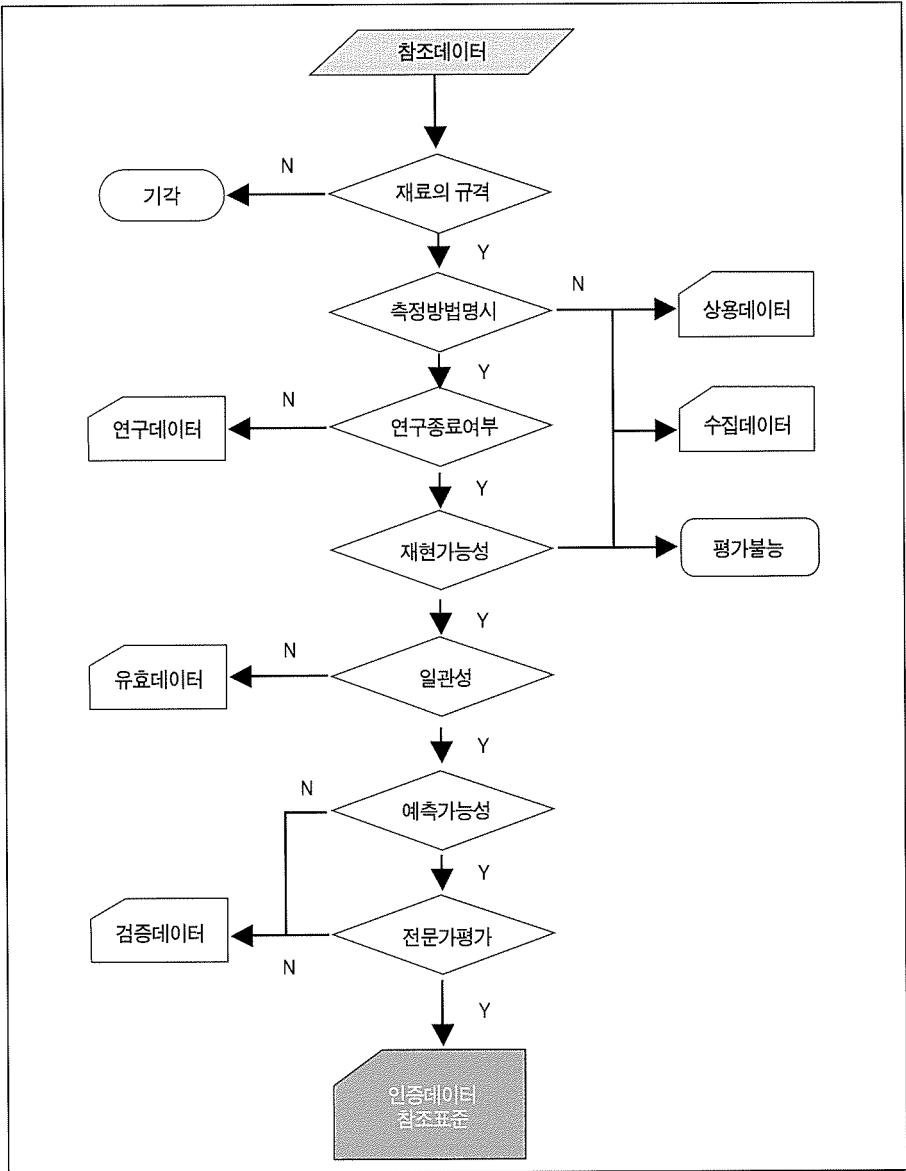
평가기준 항목	세부평가분야
소재관련	적용규격, 화학조성, 용해방법, Ingot크기, 탈산공정, 제품형태, 제품치수, 가공 및 열처리, 결정입도, 비금속 개재물, 편석, 미세조직
시험기 관련	시험기정보(제작사, 모델, 용량, 사용 온도 범위), 장비의 교정, 실험장비측정렬 정도, 열전대 유형 및 교정, 열전대 사용법, 자동화 사용시 사용SW 유효성
시험법 관련	적용규격, 시험환경(온도 및 습도), 측정렬, 시험편 치수 측정, 그립, 연신계 게이지길이, 표점 표시방법, 연신계 부착방향, 샘플링 속도, 부하속도 관련, 사용하중, 항복강도, 연신률, 단명수축률 결정방법
시험자 관련	적용규격, 결과의 통계적 처리여부, 시험온도, 유효 숫자처리, 피단연신을 보고시의 표점길이 기술여부, 응력변형률곡선, 불확도 평가

〈표 4〉 소재물성 기초평가 항목

일반적으로 해당분야의 참조데이터를 평가하기 위해서는 학술논문, 연구보고서, 공개 가능하며 평가되지 않은 데이터베이스 등이 수집대상 정보원이 된다. 참조데이터에 평가는 수집된 참조데이터의 참조표준 평가기준의 적합성을 검토하고 관련 기술위원회에 참여한 전문가들에 의해 평가기준이 제정된다. 제정된 평가기준은 참조데이터를 직접 적용해 봄으로써 유효성 검증실험을 하게 된다. 이때 참조표준 평가를 위한 요소별 영향도를 분석하여 평가기준을 보완해 나간다. 유효성 검증을 통해 평가기준이 완료되면 참조데이터의 등급이 부여된다. 소재분야 기술위원회에서 평가한 참조데이터는 인증절차와 평가요인들을 상세히 기술하여 상위 평가위원회인 전문위원회에 확인을 거쳐 최종등록하게 되는데 평가한 참조표준을 등록 전에 공개함으로써 이해 당사자간의 의견을 조율할 필요가 있다. 모든 등록절차가 종료되면 본격적인 보급을 위해 기술위원회에서 개발한 데이터베이스 알고리즘을 개발, 설계하고 참조표준DB를 구축한다.¹⁶⁾

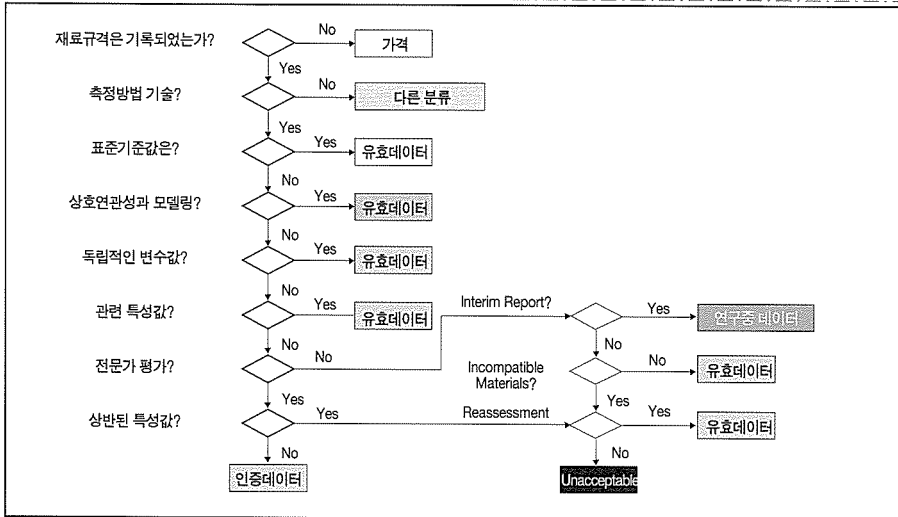
16) Jorge Luis, Romeu, "Data quality and pedigree", AMPTIAC Material Ease-9, 1998, http://amptiac.alionscience.com/pdf/1999_MaterialEase9.pdf. [Cited 2005. 4.20]

A study on establishment of standard reference data through...
 and the application of reference and technical... data



〈그림 4〉 참조데이터 평가절차

〈그림 4〉는 각 분야별 평가절차에 사용되어야 할 기초적인 요소와 절차를 표기한 내용이다. 그러나 아래 〈그림 5〉는 〈그림 4〉의 일반적인 평가기준과 절차를 소재 철강분야의 참조데이터의 평가절차에 적용하여 예시로 제시한 것이다. 따라서 소재 철강 분야평가절차를 제정하는데 있어 이를 기초로 작성되었다. 재료의 규격, 측정방법, 표준기준값, 상호연관성과 모델링, 각각의 특성값 등을 주요 평가요소로 사용되며 각 단계별로 만족하지 못할 경우 기각, 유효, 인증데이터로 등급이 부여된다.



〈그림 5〉 소재물성 분야 데이터 평가 절차

V. 결론

과학기술 연구개발 사업이 대형화되고 삶의 질에 대한 관심이 점차 높아짐에 따라 고급의 과학기술정보에 대한 수요가 크게 증가하고 있다. 참조표준은 이러한 시대적인 흐름에 가장 적합한 고급 과학기술정보이다. 한 국가가 참조표준은 생산하고 관리할 수 있는 능력을 가진다는 것은 선진과학기술 능력을 보유하고 있다는 증거이다. 따라서 우리나라도 국가고유의 신뢰성 높은 참조표준을 확보하여 연구개발에 중복 이용함으로써 예산, 시간 등의 낭비를 최소화하고 나아가 제품의 경쟁력을 갖출 수 있는 기초정보로 활용되어야 한다. 본 연구에서는 국가참조표준체계를 구축하기 위한 방안으로 국내외 참조표준 개발 현황과 참조표준에 대한 실질적인 수요도가 있는지는 살펴보았다. 참조표준을 구축하기 위해서는 가장 중요한 “신뢰성 있는 평가되지 않은 과학기술 수치데이터”를 체계적으로 평가하기 위한 기준으로 재현성, 일관성, 예측가능성을 가장 큰 틀로 구성해 보았고, 평가되지 않은 데이터의 등급을 부여하기 평가절차를 정리해 보았다. 이러한 일반적인 평가기준과 절차가 잘 구성되었는지를 확인하기 위해 소재물성 분야를 시범 적용해 보았다. 시범 적용해 본 결과 본격적인 데이터 생산은 이루어지지 않아 평가절차와 기준이 완벽하게 잘 수립되었는지는 파악할 수 없으나, 기본적인 체계는 갖추었다고 할 수 있다.

각 분야별로 평가절차와 기준은 다를 수 있겠지만 장차 참조표준 구축사업이 범국가적 차원에서 본격적으로 수행될 것을 대비하여 기반환경 조성이 필요하다. 즉 참조표준이 생산되기까지 요구되는 다양한 기준과 절차를 미리 확보해 둘 필요가 있다. 따라서 본 연구에서는 이러한 기반 환경을 조성하기 위해 새롭게 구성해야 할 국가참조표준센터, 데이터센터, 각종위원회의 역할과 임무를 정의하였고, 참조표준 등록과 보급을 위한 방안, 법적 제도적 장치에 대한 사항을 살펴보았다. 참조표준의 활용은 일반적인 제품의 품질향상 뿐만 아니라 국민의 삶의질 향상을 위해서도 점차 그 수요는 증가하고 있다. 참조표준을 생산함으로써 국가고유의 지적자산을 확보할 뿐만 아니라 국가 과학기술의 인프라로 활용될 것이며, 이러한 중요성을 생각해 볼 때 아직까지 과학기술 정보관리와 이용에 대한 낮은 인식을 탈피하여 본격적인 참조표준 구축체계를 갖추기 위해서는 관련 기관을 비롯한 정부의 적극적인 예산과 인력지원이 필요한 시점이라고 할 수 있다.