

# 정부 여론조사자료 아카이브 구축방안에 관한 연구\*

## A Study on Archiving of Government Survey Data

남 영 준(Young-Joon Nam)\*\*

서 만 덕(Man-Deok Seo)\*\*\*

### 목 차

- |                       |                        |
|-----------------------|------------------------|
| 1. 서 론                | 4. 여론조사자료 아카이브 사례분석    |
| 2. 여론조사자료의 개념과 기록적 가치 | 4.1 국외 여론조사자료 아카이브     |
| 2.1 여론조사자료의 개념        | 4.2 국내 여론조사자료 아카이브     |
| 2.2 정부 여론조사자료의 기록적 가치 | 5. 정부 여론조사자료 아카이브 구축방안 |
| 3. 정부 여론조사자료 관리 현황    | 5.1 수집정책               |
| 3.1 정부 여론조사자료 관리실태    | 5.2 통합 메타데이터(안)        |
| 3.2 정부 여론조사자료 관리시스템   | 5.3 자료보존               |
|                       | 6. 결 론                 |

### <초 록>

정부 및 공공기관은 현재 사회, 문화, 복지, 교육 등 국가 전 영역에 걸쳐 과학적인 정책결정을 도출하는 수단으로 여론조사를 실시하고 있다. 이와 같은 중요성에도 불구하고 현재 생산되는 정부 여론조사자료는 정부부처별로 산재하여 관리되고 있으나 수집 및 관리체계 부재로 인하여 일부자료가 누락되거나 소실되고 있다. 왜냐하면, 정부 여론조사자료가 대부분 인쇄자료 형태로 분산관리되기 때문에 통계적 활용이 불가능하며 자료접근에 제한이 있다. 따라서 이 연구에서는 여론조사자료의 장기적인 보존·활용을 위해 여론조사자료의 관리·보존현황에 대한 국내외 사례 분석, 문헌연구, 관련 표준분석을 실시하여 수집정책, 평가기준, 통합메타데이터, 자료변환절차, 보존포맷으로 구분하여 아카이브 방안을 도출하였다.

주제어: 여론조사, 정부여론조사자료, 아카이브, 메타데이터, 보존포맷, 원자료

### <ABSTRACT>

The government and public institutions are conducting public-opinion surveys as means of drawing the scientific policy decisions throughout the country's whole sectors : society, culture, welfare, education, etc. The Government Survey Data have been under management by each governmental department, but the improper ways of data collection and management have led to omissions and losses of some data. Also there are limitations on the statistic usage and access to the data due to the fact that the data are managed dispersedly in the form of printed materials. Accordingly, to make the survey data available for the long-term preservation and usage, this study suggests collection policy, standard of evaluation, metadata, the procedure and method of converting materials, preservation format.

Keywords: survey, government survey data, archive, metadata, preservation format, raw data

\* 이 논문은 2009년도 중앙대학교 우수연구자연구비 지원에 의한 것임.

\*\* 중앙대학교 대학원 기록관리학과 교수(namyj@cau.ac.kr)(제1저자)

\*\*\* 중앙대학교 대학원 기록관리학과 석사, 한국사회과학데이터센터 선임연구원(seomd@ksdc.re.kr)(공동저자)

■ 접수일자 2009년 5월 16일 ■ 수정일자 2009년 6월 10일 ■ 게재확정일자 2009년 6월 18일

## 1. 서론

사회가 복잡하게 변하고, 사회현상도 다양해짐에 이를 해석하고, 사회현상을 체계적이고 효율적으로 분석하기 위한 객관적인 데이터가 필요하다. 또한 현상에 대한 정확한 인식과 더불어 과거 현황과 비교분석을 통해 어떻게 변화하고 있는지를 진단할 필요성도 있다. 이에 따라 정부 및 공공기관은 현재 사회, 문화, 복지, 교육 등 국가 전 영역에 걸쳐 과학적인 정책 결정을 도출하는 수단으로 여론조사를 실시하고 있다. 특히 청와대는 2003년에 여론조사 예산으로 8억 9천만원을 집행하였으며, 2007년에 19억 2천만원으로 이에 대한 비용이 지속적으로 증가하고 있고 정부부처별로 수행되는 여론조사 관련 집행비용도 방대한 실정이다. 이러한 예산집행의 결과, 사회의 다양한 여론 지표가 여러 가지 관점에서 도출되고 있으나 여론조사자료의 결과를 도출한 설문지 및 보고서는 국가정책포털을 통해 부분적으로만 관리·서비스되고 있다. 특히 이차활용이 가능한 원자료(raw data)의 경우에는 별도 관리되고 있지 않다. 결과적으로 정부부처에서 실시한 여론조사자료의 양은 방대하나, 이를 효율적으로 생산, 관리, 보존하기 위한 체계도 제대로 구축되어 있지 않고 기관별로 임의적인 기준을 적용하여 관리하고 있는 실정이다. 궁극적으로 이는 원자료의 소실 및 사장으로 이어지고 있기 때문에 정책결정자 및 학술연구자의 이차적 활용을 위한 자료로 거의 활용하지 못하고 있다. 또한 이 원자료는 해당 정책에 대한 적정한 분석 혹은 판단을 위한 검증자료로서 가치를 갖고 있기 때문에 그 보존의 필요성과 당위성을

갖고 있지만, 실제적으로 이 원자료를 체계적으로 관리하지 않고 있다. 왜냐하면, 해당 연구를 수행한 정부기관에서 결과를 도출하기 위해 사용한 원자료는 수집대상에 포함되지 않고, 단순하게 이의 해석만이 보고서에 수록됨으로써 그 활용목적을 달성하였다고 판단하기 때문이다. 이는 결과에 대한 합리적인 해석을 검증하거나 혹은 다른 관점에서 해당 원자료를 분석하여 기존 결과에 대한 재해석 등을 하는 것을 불가능하게 하였다. 원자료를 다른 목적으로 재활용할 수 없다는 경제적 문제보다 분석 결과에 대한 투명성을 검증할 수 없다는 신뢰도에 관련한 문제를 의미하였다.

따라서 정부는 여론조사자료의 보존과 이차적 활용, 분석결과에 대한 신뢰도 확보를 위해, 여론조사자료에 대한 자료수집원칙, 메타데이터, 장기보존포맷 등의 아카이빙에 대한 체계적인 방안을 마련해야 한다. 이 연구에서는 이러한 사회적 요구를 고려하여, 여론조사자료의 관리·보존현황에 대한 국내외 사례 분석, 문헌연구, 관련 표준분석을 통해 정부 여론조사 자료를 효과적으로 보존·활용할 수 있는 아카이브 구축방안을 제시한다.

## 2. 여론조사자료의 개념과 기록적 가치

### 2.1 여론조사자료의 개념

일반적으로 여론조사는 넓은 의미로 사회조사의 한 형태이다. 사회조사는 사회현상을 객관적이고 정확하게 파악하며 그러한 현상 속에

내재하는 법칙을 규명하는 것이다.

브리태니커 백과사전에서는 ‘여론조사는 일반적으로 사회문제나 특정화제에 대해 사회구성원이 어떠한 의견이나 태도를 갖고 있는가를 알아내는 통계적 조사’라고 정의하고 있다.

이해용은 여론조사에 대하여 ‘생활주변에서 발생하고 있는 생활전반이나 관심사에 대하여 그 시대를 살고 있는 사람들의 의식을 조사하고자 하는 목적으로 실시하는 조사활동 전반을 총칭하는 것’이라고 설명하였다(이해용 1998, 131-132).

김광웅은 ‘여론을 파악하기 위하여 대상 집단을 전수 조사하거나, 또는 소규모 하위 집단인 표본으로부터 모집단에 대한 정보를 체계적이면서 과학적으로 수집하는 방법’이라고 정의하였다(김광웅 1996, 43).

이를 종합하면 ‘여론조사는 사회문제 및 특정이슈에 관하여 사회구성원이 가지고 있는 하나의 의식과 판단을 통계적으로 조사하는 방법’이라고 정의할 수 있다. 여론조사자료는 여론조사의 산출물로서 다양한 정책 및 사안에 대해 사회구성원, 즉 개인의 대한 의견을 자료화한 것을 수록한 것이다. 여론조사자료의 범위는

여론조사 수행과정에서 생산되는 자료의 유형을 살펴보면 알 수 있다. 여론조사를 수행하는 절차는 조사설계, 자료수집, 자료처리, 자료분석, 결과해석 순으로 전개된다. 이 과정에서 <표 1>과 같이 설문지, 코드북, 원자료, 결과보고서 등 다양한 형태의 자료가 생산된다.

- ① 설문지(questionnaire): 설문지는 여론조사 수행 시 측정할 수 없는 추상적인 개념을 계량화하여, 정량적인 분석이 가능하도록 작성된 일련의 질문들을 체계적으로 수록한 자료이다.
- ② 원자료(raw data): 원자료는 조사기관이 의도하는 데이터를 수집하기 위해 질문조사표나 면담에 의해 수집된 원래의 가공되지 않은 정보이다. 원자료의 표현은 문자 또는 숫자 기호로 나타내며, 인위적인 처리활동을 배제한 정보의 손상이 없는 형태의 정보이다. 여론조사자료에서 원자료는 데이터셋을 의미한다.
- ③ 코드북(codebook): 생산된 여론조사자료 분석을 위해 원자료 파일에서 변수의 위치를 파악하고 각 코드의 의미를 해석하는 지침이다.

<표 1> 여론조사 단계별 생산자료

| 조사단계         | 생산자료  | 생산포맷   |
|--------------|-------|--|
| 조사설계         | 조사계획서 | 한글 파일, MS-Word 파일, PDF 등                                       |
|              | 설문지   | 한글 파일, MS-Word 파일, PDF 등                                       |
| 자료수집<br>자료처리 | 원자료   | SPSS 파일, SAS 파일, MS-Access 파일<br>SQL 파일, MS-Excel 파일, 텍스트 파일 등 |
|              | 코드북   | 한글 파일, MS-Word 파일, 텍스트 파일<br>통계S/W 명령문 파일, PDF 등               |
| 자료분석         | 결과집계표 | 한글 파일, MS-Word 파일, 텍스트파일<br>MS-Excel 파일 등                      |
| 결과해석         | 결과보고서 | 한글 파일, MS-Word 파일, PDF 등                                       |

- ④ 결과보고서: 여론조사가 완료된 후 조사 데이터를 통계처리 및 분석하여, 최종으로 해석하여 작성된 조사의 결과물이다.

## 2.2 정부 여론조사자료의 기록적 가치

정부여론조사자료는 사회적 관점에 따라 다양한 해석이 가능하기 때문에 해석방법에 따라 다양하게 정책적 판단을 도출하는 중요한 자원이다. 따라서 이러한 여론조사자료는 이를 검증하거나 혹은 해석을 달리하여 처리할 수 있는 새로운 관점과 연구가 차후에 제기될 때, 증거적 가치와 함께 보존적 가치를 함께 갖고 있다. 또한 정부여론조사자료는 자료의 공공성과 더불어 다음과 같은 기록적 가치를 지닌다.

- ① 정책수혜자, 즉 국민들의 생활여건 및 욕구를 정확하게 파악하여 정책결정에 반영할 수 있다.
- ② 이미 집행된 정책에 대한 국민들의 반응을 살펴봄으로써 그 효과를 측정할 수 있으며, 이를 바탕으로 새로운 정책 입안을 위한 참고자료로 활용할 수 있다.
- ③ 정책에 대한 국민들의 의견이 상반되는 사안에 대해 여론조사 수행은 국민들의 의견 수렴 뿐 아니라 국정홍보효과의 수단으로 활용할 수 있다.
- ④ 학계 및 관련연구기관의 연구활동에 필요한 참고자료로 활용한다.

즉, 정부 여론조사자료는 당시 국가정책사안에 대한 내용과 정책입안을 위한 정부활동내용을 포함하고 있기 때문에 업무활동의 증거로서의 가치가 있다. 또한 이는 사회 및 국민들의 삶과 가치관에 대한 정량적인 정보를 담고 있어 당시 시대상을 유추할 수 있는 증거자료이다. 특히 여론조사자료는 가공되지 않은 통계분석용 원자료를 포함하고 있기 때문에, 정책 및 의사결정자의 분석목적에 따라 자료에 대한 재 활용이 가능하며 결과적으로 문제에 대한 재해석이 가능하다. 요약하면 여론조사자료는 기록물의 일반적 가치인 증거성을 가지고 있으며 나아가 향후 정부 뿐 아니라 국민, 학계, 연구소 등 다양한 기관에서 이용될 수 있는 활용적 가치가 있다.

## 3. 정부 여론조사자료 관리 현황

### 3.1 정부 여론조사자료 관리실태

정부 여론조사는 정부부처에서 실시하는 것이 아니라, 여론조사기관에 직접 의뢰하여 조사를 수행하거나 학술연구용역<sup>1)</sup> 및 정책연구용역<sup>2)</sup>을 통해 선정된 업체와 기관이 여론조사를 수행한다. 전자의 경우에는 대체로 수행한 여론조사의 결과물인 최종보고서만을 수집 대상으로 하고 있으며, 조사의 신뢰성과 객관성을 담보하는 원자료는 수집되고 있지 않은 실정이다. 특히 청와대의 경우 2003년 8억 9천만

1) 학문분야의 기초과학과 응용과학이 광범위하고 심층 있게 적용되어야 할 비정형화된 연구용역 및 이에 준하는 용역으로서 정부정책이나 정부 시책의 자문에 제공되는 용역(조달청일반용역적격심사세부기준 제2조).  
 2) 중앙행정기관이 정책의 개발 또는 주요 정책현안에 대한 조사·연구 등을 목적으로 정책연구과제를 선정하고, 연구자와 연구수행에 대한 대가를 지급하는 내용의 계약을 체결하는 방식으로 정책연구를 추진하는 용역사업(정책연구용역관리규정 제2조).

원, 2004년 10억, 2005년 6억 9천만원이 여론조사 비용으로 집행되었으나, 극히 제한적으로 대통령기록관을 통해 여론조사 관련 자료를 서비스 하고 있다. 또한 대부분이 보도자료 중심의 웹기록물이며, 보고서 및 원자료에 대한 자료는 거의 보존하지 않고 있다.

학술연구용역 및 정책연구용역을 통해 수행된 여론조사의 경우에도 여론조사 포함여부를 확인할 수 있는 정보가 없어 생산된 여론조사 자료의 정확한 수를 판단하기는 어렵고 연구용역의 세부내용 확인을 통해 여론조사 실시횟수를 종합하여 유추할 수 있을 뿐이다. 정부부처에서 연구요구가 발생하고 입찰을 통해 사업자를 선정하여 용역이 진행될 경우, 계약체결 이후에는 용역수행자에 의해 모든 산출물이 생산된다. 정부연구용역은 정책연구용역관리규정에서 정책연구 전반에 대한 관리 및 활용에 대한 조항을 마련하고 있으며 정책연구용역종합관리시스템을 통해 관리하도록 규정하고 있다.<sup>3)</sup> 이러한 규정이 존재하나, 여론조사 관련 연구용역의 경우 생산되는 자료의 유형 및 내용이 특수하므로 이를 표준화하여 관리 및 보존할 수 있는 법·제도적 장치는 아직까지 구체적으로 마련되어 있지 않다. 대부분 여론조사의 결과물인 최종보고서만이 구축·관리되고 있으며, 관련 데이터셋인 원자료는 관리대상에서 제외하고 있다.

여론조사의 활용도가 높은 통일부의 관리실태를 살펴보면 통일부에서 현재 생산·관리되고 있는 여론조사보유량은 2006년 당시 1990년대 이후 생산된 통일관련 75 건이며, 관련 자료

는 결과보고서와 설문지의 인쇄본으로만 보관하고 있다. 특히 대부분의 여론조사자료는 2000년대의 최신자료만 보유하고 있어 90년대 전후의 자료는 파악하기 힘든 것으로 조사되었다(통일부 2006, 6-7). 통일부에서 보관되고 있는 여론조사자료의 관리 문제점을 살펴보면, 첫째, 통일부 내의 자료는 내부적으로 관리되고 있어 인쇄본에 대한 물리적 보관공간 확보의 어려움이 있다. 둘째, 담당자별로 분산되어 관리되고 있기 때문에 정부조직의 특성상 담당자가 교체될 때나 자료보관 장소 이동 등으로 인하여 자료의 분실 가능성이 있다. 셋째, 기존 여론조사 자료를 찾기 위해서는 모든 자료를 열람해야 하기 때문에 자료의 재활용성이 낮다. 넷째, 여론조사자료를 구성하는 다양한 형태의 자료 중 조사보고서와 설문지에 대한 인쇄본으로만 보관하고 있기 때문에, 원자료 부재에 따른 자료분석 및 자료 재가공이 불가능하다.

한편 정부부처별로 산재하여 관리되고 있는 여론조사자료는 업무담당자 변경과 자료관리 책임부재로 인한 자료의 소실우려가 있으며, 보관되고 있는 자료는 조사보고서, 설문지의 인쇄본에 국한하고 있다. 또한 활용적 가치가 높은 원자료 및 코드북에 대한 보존이 이루어지고 있지 않기 때문에 자료의 이차적 활용이 제한된다. 오히려 민간기관인 한국사회과학데이터센터 및 한국사회과학자료원에서 선별적으로 정부·공공기관과의 자료기탁협의를 통해 원자료를 포함한 완전한 여론조사자료를 보존·서비스하고 있는 실정이다. 정부에서 생산한 여론조사자료를 완전하게 보존·관리하기

3) 정책연구용역관리규정 [국무총리훈령 일부개정 2007.3.30] 제3장.

위해서는 정부가 주체가 되어 생산한 여론조사 자료를 수집하고 자료의 평가기준을 마련해야 한다. 이와 함께 자료보존을 위한 표준메타데이터와 자료보존포맷을 기반으로 하는 통합 아카이브를 구축해야 한다.

### 3.2 정부 여론조사자료 관리시스템

정부 여론조사가 정책연구용역 및 학술연구용역으로 수행된 경우, 여론조사를 포함한 연구보고서는 대부분 행정안전부의 정책연구용역종합관리시스템(PRISM: Policy Research Information Service & Management)에서 관리하거나 정부부처별 연구용역 관련 시스템으로 관리하고 있다.

정책연구용역종합관리시스템은 우리나라 중앙부처에서 수행하는 정책연구용역 과정을 관리하고, 연구용역 결과물에 대한 대국민 서비스를 제공하는 시스템이다. 정책연구용역종합관리시스템의 연구용역에 관한 메타정보는 연구용역정보, 연구결과정보, 용역평가정보, 활용정보, 관리정보의 영역으로 구분된다. 연구용역 메타데이터는 연구용역의 효율적 관리를 위한 용역과정에 대한 전반적인 내용정보를 기술할 수 있다. 이에 반해 연구용역 일반적 특성 이외의 여론조사의 특성을 기술할 수 있는 기술요소는 없다. 또한 관련 산출물이 연구보고서에 국한되기 때문에, 연구수행 과정에서 생산된 여론조사의 원자료, 코드북 등 특수형태의 자료는 관리대상에서 제외된다. 즉, 연구결과에 대한 최종적인 결과물을 등재하고 보관하는 시스템이며, 해당 결과가 도출되기 위해 사용되었던 원자료에 대한 보존은 이루어지지 않고

있다. 이와 같이 원자료에 대한 관리체계가 마련되어 있지 않기 때문에 여론조사의 검증 혹은 재활용을 위한 수요가 발생할 때 이에 대한 대처방안이 시스템적으로 마련되어 있지 않다.

또한 정부에서 직접 여론조사기관에 의뢰하여 여론조사를 실시한 경우, 부분적으로 대한민국정책포털을 통해 관리·서비스하고 있다. 대한민국정책포털은 정부부처에서 분산 관리되는 다양한 정부정책정보를 제공하고, 더불어 정부부처와 공공기관에서 제공하고 있는 각종 민원·경제·직업·여가·생활정보에 대한 국민들의 이용편의성을 제고하기 위해 구축된 시스템이다. 대한민국정책포털은 2008년 8월 개편작업을 통해 정부부처에서 실시한 정책관련 여론조사자료 약 500여건을 구축하여 서비스하고 있다. 구축된 여론조사자료는 주제, 유형, 출처, 등록일, 조회수, 내용 등의 메타정보만을 기술하고 있으며 조사의 객관성과 신뢰성을 판단하는 조사방법 및 조사과정에 관한 정보는 누락되어 있다. 제공하는 자료의 유형은 설문지, 원자료, 조사결과표, 보고서로 구성되어 있다. 현재 정부에서는 해당 조사자료를 구성하는 파일에 대해 유형별로 표준화하여 관리하지 않고 첨부파일형태로 통합관리하고 있어 구체적인 파일 관리 기준과 지침을 제안하지 않고 있다.

## 4. 여론조사자료 아카이브 사례분석

### 4.1 국외 여론조사자료 아카이브

#### 4.1.1 ICPSR

미국은 여론조사자료를 체계적으로 관리하는

전담기관으로서 사회과학자료보관소인 ICPSR (The Inter-University Consortium for Political and Social Research)을 통해 여론조사자료를 아카이브로 구축·관리하고 있으며, 이를 위해 자료수집기준, 메타데이터, 보존절차를 아래와 같이 제시하고 있다.

### 1) 자료수집 및 평가기준

ICPSR의 자료수집기준은 첫째, 연구 및 교육을 위한 실질적 가치와 영구적인 기록적 가치를 평가한다. 둘째, 수집된 자료는 개방적·범용적 서비스가 가능해야하며 미래 활용가치가 높아야 한다. 셋째, 최신의 연구 및 통계기법이 활용된 자료를 수집한다. 넷째, 사회과학 연구기법인 질적연구와 양적연구가 동시에 가능한 자료를 수집한다. ICPSR은 사회과학 영역에서 사회과학 연구기법이 활용 가능하고 학술적 중요도가 높은 자료를 수집하고 있다. 내부 연구진을 통해 자료의 현재가치와 영구적 가치를 판단하고 있으며, 특히 여론조사자료의 경우 종단조사자료,<sup>4)</sup> 다양한 국가 및 인종에서 조사된 자료, 국가의 현황분석 및 국가간 비교 연구가 가능한 자료, 순차(Sequential)<sup>5)</sup> 방식, 컨버전(Conversion)<sup>6)</sup> 방식 등 하나 이상의 연구방법을 통해 분석된 자료는 가치 있는 양적 자료로 평가받고 있다.

ICPSR은 미 정부의 자료기탁을 통해 학술적, 연구적 가치가 있는 자료에 대해 별도의 아카이브를 구축하여 수집, 보존, 제공의 역할을 수행하고 있다. 분야별로 살펴보면 교육(International

Archive of Education Data), 노인(National Archive of Computerized Data on Aging), 범죄와 재판(National Archive of Criminal Justice Data), 약물남용과 정신건강(Substance Abuse & Medical Health Data Archive), 인구(Census 2000 at ICPSR), 범죄(CrimeStat) 등의 아카이브 사례가 있다. 그러나 ICPSR는 미 정부에서 생산한 모든 여론조사자료를 보존·관리하지 않고 있다.

수집된 여론조사자료의 평가는 ① 자료의 역사적 가치, ② 자료의 완전성, ③ 자료의 신뢰성 확인을 위한 조사방법 관련 기술문서 포함여부, ④ 자료포맷의 범용성과 호환성, ⑤ 원 자료 형태의 데이터파일 보유여부의 기준으로 판단한다.

### 2) 메타데이터

ICPSR은 초기에 자료수집 시 데이터제출양식을 이용하여 자료의 내용과 구조를 기술하였다. 그러나 국제 사회과학 연구자들과 정보관리자들은 사회과학 분야의 자료를 상세하게 기술할 수 있는 구조화된 메타데이터 표준의 필요성을 인식하였다. ICPSR의 관리자인 Richard Rockwell은 1995년 5월 SGML을 이용하여, 국제적인 코드북 표준을 위한 DTD(Data Type Documentation)를 개발하기 위한 위원회를 결성하였다. 이후 ICPSR에서는 1996년 SGML을 이용한 DDI(Data Documentation Initiative) 버전 1.0을 작성하였고 수많은 개정작업을 거쳐, 2005년 Sanda Ionescu, Mark Diggory에

4) 동일한 조사를 일정한 시간간격을 두고 반복해서 조사하는 것으로 연속형 조사라고 한다.  
5) 질적 접근 다음에 양적 접근 또는 그 반대로 실시하는 방법.  
6) 하나의 연구방법에서 다른 연구방법으로 전환되는 것.

의해 XML기반의 버전 2.1을 구축하였으며 2008년 이후 버전 3.0이 개정되었다. 본 연구에서는 활용성이 높은 2.1을 기준으로 해당 메타데이터를 분석하였다. DDI 메타데이터는 <표 2>와 같이 ① 문서기술(Document Description), ② 연구기술(Study Description), ③ 데이터파일 기술(Data Files Description), ④ 변수기술(Variable Description), ⑤ 관련 연구자료(Other Study-Related Materials)의 다섯 가지 영역으로 구분하고 있다.

문서기술 영역은 DDI 메타데이터 표준에 의해 마크업된 여론조사자료 객체에 대한 서지정보를 기술하는 영역으로 표제, 부표제를 포함한 타이틀 요소와 자료생산 및 수집과 관련된 기탁정보와 생산정보를 포함한다. 연구기술 영역은 양적연구에 관한 내용, 절차, 방법론을 기술하는 요소와 자료접근요소, 기타 관련자료 요소로 구성되어 있다. 또한 자료의 수집시기, 지역, 분석단위 등 여론조사자료의 조사범위를 기입하는 하위요소와 표본설계, 표본오차, 조사방법, 응답률 등 자료의 수집과정에 관한 일련의 절차와 표집방법을 기술하는 하위요소 등을

포함한다. 데이터파일기술 영역은 통계 및 여론조사자료와 같은 양적자료 파일에 대한 구조와 형태를 기술하는 요소로서, 원자료를 구성하는 레코드 길이, 변수수, 사례수, 결측값 등과 같은 요소를 포함한다. 또한 원자료파일 자체에 대한 파일명, 파일유형, 포맷, 사용소프트웨어, 버전 등 파일형태에 관한 정보를 포함한다. 변수기술 영역은 변수그룹과 그룹 내의 개별변수에 대한 기술요소이다. 여론조사자료에서 변수는 설문항과 대응되는 개념이며 통계자료에서 변수는 통계항목에 대응되는 개념이다. 변수기술 영역은 변수에 대한 내용, 형태, 권한에 대한 하위요소를 모두 포함한다.

3) 자료보존

ICPSR은 효율적인 자료보존을 위해 다음의 자료보존절차를 수행한다.

최초 자료제출자는 온라인 자료제출양식을 통해 SPSS 및 SAS 등의 데이터파일과 코드북, 설문지 등의 기술문서를 포함한 제출정보 패키지(Submission Information Package)를 구성한다. 자료가 제출되면 수집된 자료에 대

<표 2> DDI 메타데이터 표준

| 영역  | 설명                             | 요소                                 |
|---|--------------------------------|------------------------------------|
| 문서기술<br>(Document Description)            | 마크업된 코드북에 대한 기본정보를 기술          | 표제, 생산정보, 기탁정보 등                   |
| 연구기술<br>(Study Description)               | 연구의 내용과 맥락을 기술                 | 연구방법, 조사방법, 조사절차, 자료처리, 자료접근 등     |
| 데이터파일기술<br>(Data Files Description)       | 데이터파일(원자료)의 구성과 구조를 기술         | 파일유형, 파일포맷, 파일사이즈, 변수수, 사례수, 결측값 등 |
| 변수기술<br>(Variable Description)            | 데이터파일을 설명하는 변수정보를 기술           | 변수그룹, 변수명, 분석단위, 응답단위, 통계 요약 등     |
| 관련연구자료<br>(Other Study-Related Materials) | 여론조사자료 관련 서지자료 및 연구자료 서지 사항 기술 | 자료유형, 표제, 생산자, 생산일자 등              |



한 평가작업을 통해 구축대상자료를 선별하고 자료의 활용성과 호환성을 높이기 위해 자료변환작업을 실시한다. 이 때 변환작업은 파일포맷의 컨버전, 표현정보의 컨버전 등의 구조적 변환작업을 실시하여 장기보존을 위한 자료포맷으로 전환한다. 원자료의 경우 대부분 SPSS, SAS와 같은 상용통계프로그램 파일형태로 수집되므로 자료처리과정에서 우선적으로 소프트웨어에 비의존적인 영구적 포맷으로 전환하고 이 때 원자료는 텍스트파일로 저장하며 통계적 해석을 위해 필요한 통계처리정보는 통계 소프트웨어의 명령문 파일(Setup File)로 저장한다. 이와 같이 저장 할 때에, 원자료와 명령문 파일이 동시에 저장된 패키지 파일이 사용자의 이용편의성을 높이기 때문에, 원자료와 통계처리정보를 장기보존이 가능한 통합포맷(SPSS, SAS, Stata 등)으로도 전환한다. 또한 ICPSR은 파일변환과 별도로 자료의 접근·보존·분석을 위해 양적자료와 관련한 내용정보와 기술정보를 ICPSR 메타데이터 기술방식인 DDI Variable-Level XML 파일로 저장하여 변수에 대한 상세정보를 기술하고 있다. 일반적인 문서자료인 코드북 및 설문지는 ISO 장기보존포맷인 PDF/A로 일괄 변환한다. 처리와 변환작업이 완료되면 입수자료는 최종적으로 보존정보 패키지(Archival Information Package)로 재생산하며, 최초 기탁된 제출정보패키지의 경우에도 자료의 진본을 유지하기 위해 보존한다.

#### 4.1.2 UKDA

UKDA(UK Data Archive)는 영국 에섹스

대학에 소재한 연구기관으로 인문·사회과학 분야의 데이터를 발굴, 수집, 보관하는 영국의 최대의 디지털 자료를 보관하고 있는 아카이브이다. UKDA는 유럽 최대 사회과학데이터 아카이브인 CESSDA(Council of European Social Science Data Archives)<sup>7)</sup>의 회원이자, 미국 최대 사회과학 아카이브인 ICPSR의 회원이다. 다양한 학문영역의 연구 및 교육에 필요한 약 5,000여개의 여론조사자료, 통계자료 등의 양적 자료와 심층면접자료, 관찰자료, 멀티미디어자료(이미지, 사진, 오디오)와 같은 질적 자료를 제공하고 있다.

##### 1) 자료수집 및 평가기준

UKDA는 사회과학분야에서 장기간 활용될 수 있는 양적자료 및 질적자료를 수집한다. 수집된 자료는 주제의 적절성 뿐 만아니라 시계열 분석이 가능한지에 대한 중단적 범위와 비교연구가 가능한지에 대한 횡단적 범위가 자료 가치 판단에 중요한 요소이다. 또한 양적자료는 통계분석을 통한 결과재해석이 가능하므로 자료의 재활용성과 접근성에 대한 부분도 중요한 가치판단기준으로 설정하고 있다. 그러나 표본부족 및 데이터소실로 인한 통계분석 접근이 어려운 경우 아카이브 대상에서 제외한다.

##### 2) 메타데이터

UKDA의 경우에도 수집된 자료를 DDI 메타데이터를 통해 구조화한다. 자료수집 이후 자료처리와 보존활동을 지속적으로 기록·관리하기 위해 자원관리 메타데이터를 사용한다. UKDA

7) 1976년에 설립된 유럽연합의 아카이브로서 유럽의 사회과학 전반에 걸친 사회조사자료를 수집 및 제공하며 각국의 대표적인 아카이브 링크 제공.

자원관리메타데이터는 식별정보(Identification), 자원내용정보(Resource content), 처리이력정보(Processing history), 자료전달정보(Data delivery)의 4가지 영역으로 구성된다.

자원관리메타데이터의 식별정보는 자료자체를 식별하기 위한 식별자와 연구명을 기술한다. 둘째, 자원내용정보는 자원의 형태와 구조적 특징을 나타내는 것으로서 파일명, 파일버전, 파일포맷에 대한 요소를 기술한다. 또한 자원의 수집, 관리과정상에서 생산된 파일에 관한 형태정보를 기술한다. 셋째, 업무이력정보는 자료수집, 변환, 복제 등의 자료처리과정상 나타나는 일련의 업무활동내용을 기술하는 요소이다. 자료 컨버전 업무를 수행할 경우, 변환된 파일명, 사용된 소프트웨어, 발생한 오류, 변환이유 등의 업무처리과정 및 결과를 기술한다. 넷째, 자료전달정보는 이용자에게 자료를 서비스할 때 유용한 부가정보와 이용자에게 자료서비스 중 발생한 비일상적인 업무를 기록하는 요소이다.

### 3) 자료보존

UKDA 자료보존정책의 일련의 과정은 양적 자료의 수집, 메타데이터 등록, 보존과 이용에 필요한 자료변환업무, 장기보존포맷 전환으로 수행된다.

우선 양적자료가 수집되면 자료관련 메타데이터는 DDI 표준에 따라 구축하고, 장기적인 자료보존을 위해 파일을 표준포맷으로 변환하며, 진본의 보존을 위해 저장매체로 이중 백업한다. 자료의 포맷은 자료의 장기적 이용을 위해 환경 변화에 따라 손실위험이 적은 보존포맷으로 변환하고 영구기록매체에 저장한다. 특히 보존대

상 자료의 저장공간과 파일포맷은 영구불변의 속성을 갖지 않기 때문에, 모든 데이터유형을 일관성 있게 관리할 수 있도록 최소한의 보존포맷으로 한정하여 관리해야 한다. 양적자료의 보존포맷은 소프트웨어에 비의존적이고 범용적으로 사용되는 형식이어야 하며, 적절한 보존포맷으로 전환이 쉽게 가능해야 한다.

UKDA 자료보존포맷은 변수명이나 컬럼값이 없는 일반 표형자료일 경우는 탭 또는 콤마로 분리된 텍스트파일형태로 저장한다. 텍스트 파일은 불필요한 서식정보를 기술할 필요가 없어 파일이 가볍고, 소프트웨어에 의존적이지 않은 공개문서포맷이기 때문에 범용적으로 활용 가능한 포맷이다. 변수 또는 컬럼이 있는 경우에는 원자료의 수치데이터는 구분자로 분리된 텍스트로 저장하고 컬럼 및 변수에 관한 기술정보는 상용통계프로그램의 명령문 파일로 저장하거나 변수정보를 DDI XML File로 저장하여 보존한다.

기타 보고서, 설문지, 질적자료 등의 문서자료의 경우에는 PDF, RTF, XML 마크업파일 등 다양한 문서포맷으로 보존·관리한다.

## 4.2 국내 여론조사자료 아카이브

### 4.2.1 한국사회과학자료원

한국사회과학자료원(KOSSDA)은 2006년 에스콰이아학술문화재단이 각각 독립적으로 운영하던 한국사회과학도서관과 한국사회조사데이터아카이브를 통합하여 설립한 비영리 사회과학 자료전문기관이다. 국내외에서 생산하는 한국과 한국관련 사회과학 원자료와 문헌자료를 전문적으로 수집하여 제공하고 있으며, 원

자료와 문헌자료를 통합하여 검색하고 이용할 수 있는 온라인 서비스를 제공하고 있다.

### 1) 자료수집

KOSSDA의 자료수집은 연구자의 기탁 또는 기관과의 기탁협약을 통해 이루어지며, 기관의 경우 협약체결 이후 공개 가능한 여론조사자료에 대하여 자료를 기탁 받아 수집하고 있다. 수집대상이 되는 여론조사자료의 종류는 원자료, 설문지, 코드북, 코딩가이드, 조사원지침, 표본틀 설명에 관한 자료이다. 자료가 수집되면 내부 기탁신청등록폼을 통해 자료를 등록하고 표준화한다. KOSSDA의 수집대상기관은 대학연구소, 법인체 연구기관, 정부출연 연구기관이며 정부출연기관의 경우 약 20여개 기관과의 자료기탁협정을 통해 원자료, 설문지, 보고서 자료를 수집하고 있다. 그러나 지속적인 기관협정에도 불구하고, 자료기탁에 의존하고 있기 때문에 정부 여론조사자료를 포괄적으로 구축하기는 어렵다.

### 2) 메타데이터

여론조사자료 메타데이터는 ICPSR과 UKDA에서 준수하고 있는 DDI 메타데이터 표준에 기초하고 있다. 메타데이터는 연구 기술(Study Description), 데이터파일 기술(Data Files Description), 변수 기술(Variable Description)의 세 가지 영역으로 구분하여 등록한다. 단 한국사회과학자료원에서 수집되는 여론조사자료는 자료기탁을 통한 수집이기 때문에 여론조사의 생산과정과 연구의 진행과정을 포괄하여 기술하기는 어렵다. 또한 보존측면에 있어 자료의 컨버전 업무, 자료수정 업무와 같은 업무활동

이력에 관한 내용을 기술하는 요소는 없다.

### 3) 자료보존

보존대상이 되는 자료의 유형은 원자료, 설문지, 코드북, 코딩가이드, 조사원지침, 표본틀이며, 문서형태로 수집되는 자료의 경우에는 문서장기보존포맷인 PDF로 변환하여 보존한다. 원자료의 경우에는 명령문파일과 데이터파일을 별도 분리하지 않고, 통계소프트웨어에서 활용가능한 SPSS, SAS, Stata 완성형파일로 보존·관리하고 있다.

특히 수집된 양적자료는 설문지의 변수정보를 변수 기술 메타데이터로 변수명, 변수그룹, 변수값, 보기정보 등의 상세내용을 기술하여 데이터베이스로 구축한다. 또한 원자료를 통계분석용 데이터로 변환하여 온라인 통계분석툴인 NESSTAR(Networked Social Science Tools and Resources)를 통해 이용자들이 웹상에서 여론조사자료를 분석할 수 있는 서비스를 제공한다.

### 4.2.2 한국사회과학데이터센터

한국사회과학데이터센터(KSDC)는 1997년에 설립된 교육과학기술부 소관 비영리 연구소로, 국내외에서 수집된 여론조사자료 및 통계자료를 내부표준화 작업을 수행한 뒤 아카이브로 구축하고 있다. 정치, 선거, 행정, 경제, 사회, 문화 등의 16개 분야의 1,800여건이 여론조사자료가 구축되어 있으며, 관련 설문지, 코드북, 원자료, 보고서로 구성되어 있다. 또한 정부 및 공공기관에서 생산된 국내외 통계자료 33,000여건의 통계항목을 제공하고 있다.

1) 자료수집

KSDC의 수집대상이 되는 자료는 정부, 공공기관, 연구기관, 언론사에서 생산된 여론조사 자료이다. KOSSDA와 마찬가지로 자료공동활용에 관한 협정이 체결된 기관을 대상으로 여론조사자료를 수집하고 있기 때문에, 정부에서 생산된 모든 여론조사자료를 포괄적으로 수집하기 어렵다. 또한 협정이 체결된 기관인 경우에도 대외적인 공개를 희망하지 않는 자료의 경우에는 그 수집대상에서 제외된다.

자료협정을 통해 수집된 여론조사자료는 학술적 가치, 활용가능성, 파일완전성으로 자료를 평가한다. 여론조사의 표본설계, 표집오차, 응답률, 조사방식에 대한 분석을 통해 조사의 신뢰성을 판단하고, 내용적으로 학술연구 및 의사결정을 위한 가치를 지니고 있는지 판단한다. 또한 수집된 자료가 원자료, 설문지, 보고서를 포함하고 있는지에 대한 자료의 완전성 검증과 수치데이터의 누락 및 오류에 대한 논리적인 검증을 실시하여 구축대상 자료를 최종적으로 선정한다.

2) 메타데이터

자료가 평가되면 여론조사의 내용 및 절차를 포함하는 정보를 분석하여, KSDC 여론조사 메타데이터 요소에 적합한 데이터를 추출하여 저

장한다.

KSDC 여론조사자료 메타데이터 요소는 <표 3>과 같이 조사개요정보, 파일정보, 질문정보로 구분된다. 조사개요정보는 여론조사의 방법 및 내용에 대한 간략정보를 기술하는 영역이며, 파일정보는 관련파일을 일괄기술하지 않고 코드북, 설문지, 원자료 파일로 구분하여 기술한다. 질문정보는 설문지에 대한 질문내용과 보기내용을 구분하여 기술한다. KSDC 여론조사자료 메타데이터 요소는 여론조사 자체에 대한 기술요소로서 자료의 관리·보존·활용에 대한 내용과 업무처리이력을 기술하는 정보는 부재하다. 또한 자료의 생산과정과 관리형태에 대한 구체적인 기술요소는 없다.

3) 자료보존

자료수집 이후 여론조사 메타데이터가 구축되면, 여론조사자료의 관리포맷을 표준화하기 위해 자료변환작업을 수행한다. 설문지는 최초 PDF, 한글, MS-Word, 텍스트 포맷으로 입수한다. 수집된 설문지에 대한 내용분석을 통해 질문정보와 답변정보로 구분하고 질문정보는 질문유형에 따라 재그룹하고 답변정보는 보기번호와 보기명으로 구분하여 모든 정보는 구분된 텍스트파일 형식으로 변환한다. 원자료는 SPSS, SAS, Stata, MS-Excel, 텍스트 등의 포맷으

<표 3> KSDC 여론조사자료 메타데이터

| 영역     | 요소  |
|--------|---|
| 조사개요정보 | 식별자, 조사제목, 주제분야, 연구주관기관, 수행기관, 조사목적 및 내용, 조사시기, 조사지역, 모집단, 표본크기, 표집방법, 표집오차, 조사방법, 등록일, 공개등급 등 15개 요소 |
| 파일정보   | 코드북, 설문지, 원자료 등 3개 요소   |
| 질문정보   | 질문코드, 질문명, 질문유형, 답변코드, 답변내용 등 5개 요소   |

로 입수하며, 원자료를 구성하는 설문코드 및 응답코드 모두 구분된 텍스트 파일 형식으로 변환한다. 변환된 원자료 및 설문지는 SQL 쿼리를 통해 조사자료 관리데이터베이스에 축적된다. 변환작업 수행과 동시에 수집된 여론조사자료의 원본을 보존하기 위해 파일원본을 저장매체에 저장하고 파일의 논리적 위치를 데이터베이스에 기록한다. 조사계획서 및 조사보고서의 경우는 원본을 그대로 보존·관리하고 있다. 변수와 원자료가 자료변환과정 이후 조사자료 관리데이터베이스에 축적되면 KSDC의 온라인 분석 도구를 이용하여 이용자는 다양한 통계분석기능을 별도의 통계프로그램 없이 웹상에서 실시간으로 활용할 수 있다.

국내의 민간기관의 여론조사자료 아카이브는 여론조사자료의 대한 수집과 보존의 역할을 수행하기 위해 여론조사 메타데이터와 보존포맷을 구축하여 관리하고 있다. 이와 더불어 여론조사자료의 이차적 활용을 높이기 위해 변수와 원자료에 대한 구성정보를 가공 및 분석하여 이를 표준화하여 데이터베이스에 구축함으로써 온라인상에서 다양한 통계분석기능을 지원하고 있다. 정부 여론조사자료의 아카이브 구축은 자료의 장기보존과 더불어 정부와 민간에서 이차적 활용이 이루어 질수 있도록 웹상에서 통계분석을 지원할 수 있어야 한다. 더불어 국내외 아카이브의 경우, 자료기탁 및 협의하에 정부 여론조사자료를 수집하고 있으므로 포괄적인 자료관리를 위해서는 아카이브 구축 및 운영주체는 정부가 되어야 한다.

## 5. 정부 여론조사자료 아카이브 구축방안

여론조사자료 아카이빙 구축을 위해서는 수집정책, 메타데이터, 자료변환, 자료보존의 네 가지 측면이 고려되어야 한다. 우선 분산 관리되고 있는 여론조사자료의 수집방안과 평가기준을 마련해야 한다. 둘째, 연구용역, 여론조사, 원자료파일, 설문내용, 관리 및 활용을 모두 고려하여 여론조사의 생산·수집·관리·보존·활용의 전 과정에 대한 상세내용을 기술할 수 있는 통합 메타데이터 요소가 필요하다. 셋째, 자료보존과 활용을 위해 필요한 자료변환업무에 대한 방법과 절차가 마련되어야 하고, 자료의 장기적 보존을 위한 보존포맷이 도출되어야 한다.

### 5.1 수집정책

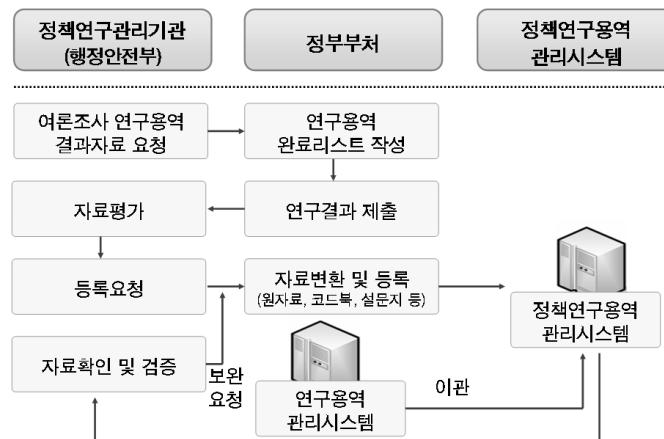
정부부처에서 산재하여 생산된 여론조사자료를 수집하기 위한 최선의 방법은 정부부처의 기록을 총체적으로 관리해야 할 책임이 있는 국가기록원이 정부부처의 여론조사자료를 아카이빙 하는 것이다. 현재 공공기록물 관리에 관한 법률은 여론조사관련 정책연구용역의 최종산출물에 해당하는 연구보고서만 기록적 가치판단에 따라 부분적으로 보존대상으로 정의하고 있으며, 여론조사를 구성하는 원자료, 설문지, 코드북 등의 자료는 수집대상이 되지 않는다. 이를 위해 정부부처의 중요한 여론조사자료를 스스로 수집하고 관리하며 보존할 수 있도록 관리지침을 각 기관에 제공하는 것이 필요하며, 국가기록원은 장기보존 가치를 가진

여론조사자료를 이관 받기 위한 이관과 처분에 대한 지침을 마련할 필요가 있다.

두 번째 방안은 정책연구용역종합관리시스템(PRISM)의 기능을 확장하는 것이다. 정부 여론조사는 학술연구용역 및 정책연구용역으로 수행된다. 정책연구용역은 정책연구용역 관리규정에 따라 정책연구용역 추진과정을 투명하게 관리하고 정책연구용역 결과 등을 공동으로 이용할 수 있도록 정책연구용역종합관리시스템을 구축·운영하여야 한다. 또한 종합관리시스템과 각 중앙행정기관 및 유관 연구기관에서 구축·운영하는 자체 정책연구용역 관련 시스템간의 연계방안을 강구하여야 한다. 또한, PRISM에서 관리하는 연구결과에 대하여 용역 과제 유형에 따라 구분할 수 있는 분류체계가 마련되어야 한다. 분류체계는 여론조사 관련된 연구용역을 구분하기 위한 필수적인 정보이기 때문이다. 다음으로, 현재 운영중인 연구평가체계를 보완하여 여론조사자료의 완전성과 중요도를 판단할 수 있는 평가기준이 마련되어야 한다. 마지막으로, 시스템의 총괄관리 역할을 수

행하는 행정안전부는 중앙행정기관에서 자체적으로 수집되는 연구결과물에 대해 각 부처시스템과의 연계를 통하여 모든 여론조사 관련 연구용역결과를 수집하여야 한다. 이를 위해 각 정부부처에 대해 여론조사 관련 연구용역의 목록을 요청하고 온·오프라인의 자료이관작업을 실시한 후 제출목록과의 확인작업이 필요하다. 이때 연구용역을 수행하는 정부부처는 실제 연구용역을 수행하는 당사자에게 여론조사자료와 관련된 모든 자료유형(용역보고서를 포함한 여론조사 설문지, 코드북, 원자료)을 제출받아야 하고 이를 모두 정책연구관리기관에 제출하여야 한다(그림 1 참조).

수집된 자료의 평가기준을 마련하기 위해 정부여론조사자료의 일반적 특성과 ICPSR 자료평가기준, UKDA의 자료평가기준, PRISM의 용역평가기준을 종합하였다. 종합된 평가요소는 조사내용 및 결과의 질적 가치를 판단하기 위해 내용 평가기준으로 범주화하고, 자료의 신뢰성과 완전성을 판단하기 위해 여론조사 절차 및 자료 평가기준으로 범주화하였다.



〈그림 1〉 정부여론조사자료 수집절차(안)

1) 내용 평가기준

- ① 내용의 적절성: 정책결정 및 의사결정에 필요한 사안에 대해 국민여론이 적절하게 수렴되었고 연구결과에 반영되었는지 평가한다. 자료의 내용이 전문적이고 범용적으로 활용될 수 있는 가를 평가하고 자료의 역사적 가치 또한 평가대상이 된다.
- ② 내용의 공공성: 여론조사의 목적 및 내용이 정부정책수립 및 정책평가 조사자료, 사회현안 조사자료, 공공행정 및 서비스 조사자료 등 공공성에 부합하는 자료인지를 평가한다.
- ③ 업무 활용성: 정책결정 및 의사결정에 필요한 기초자료로서 활용가능한 지를 평가하고 업무수행을 위한 장기적 활용가치에 대해 판단한다.
- ④ 민간 활용성: 향후 민간에서의 학술·연구적 활용가치에 대해서 판단한다.

2) 여론조사 절차 및 자료 평가기준

- ① 자료의 정확성과 신뢰성: 자료수집방법, 응답률, 자료의 상태, 샘플링 방법, 표본의 수, 측정도구, 자료분석방법의 분석을 통해 자료의 정확성을 판단한다. 연구설계, 표본 설계, 자료수집, 자료분석, 결과 해석에서 발생될 수 있는 오차를 평가한다. 또한 여론조사 수행절차상의 오류 및 문제점에 대하여 평가하여 자료의 신뢰성 정도를 판단한다.
- ② 자료의 범위: 자료의 횡단적 범위와 종단적

인 범위가 넓은 지를 판단하고, 그렇지 않을 경우 조사의 범위는 적절한지 판단한다.

- ③ 자료의 완전성: 여론조사자료를 구성하는 자료유형(설문지, 코드북, 원자료, 보고서)은 완전하게 갖추어졌는지를 평가한다.
- ④ 자료의 호환성: 자료가 다양한 기술 환경에서 사용가능한지, 관리 및 보존을 위해 쉽게 전환이 가능한 자료인지를 판단하여 평가한다.
- ⑤ 자료의 가용성: 자료의 접근이 용이하고, 장기적인 이용가능성이 있는지 평가한다.

5.2 통합 메타데이터(안)

수집된 여론조사자료는 부처별로 여론조사의 메타정보가 상이하게 기술되고 분산관리되고 있기 때문에, 여론조사자료의 내용과 구조 정보를 일관성 있게 기술할 수 있는 메타데이터 요소가 필요하다.

정부 여론조사자료는 학술연구용역 및 정책연구용역 수행과정 상에서 생산되므로, 여론조사 메타데이터 뿐 아니라 생산·관리·보존과 같은 업무활동에 관한 메타데이터 기술요소도 필요하다. 따라서 여론조사자료의 통합 메타데이터(안)을 제안하기 위하여 다음과 같은 분석을 실시하였다. 우선 ICPSR 및 UKDA에서 준수하고 있는 DDI 표준이 양적자료에 대한 상세 기술요소를 포괄하고 있으나, 기술요소가 광범위하고 세부적이기 때문에 <표 4>에서 제시한 CESSDA 메타데이터 권고안<sup>8)</sup>에 따라 DDI 표

8) 2006년 11월 5일~6일 양일간 영국 엑식스에서 열린 CDG(CESSDA DDI working Group) 회의에서 DDI 표준에서 기술하고 있는 요소 중 자료관리 및 활용을 위한 공통적이고 최소한의 요건을 추출하였다. DDI 표준이 방법론적으로 너무 많은 수의 요소로 표현되어 있기 때문에, DDI의 모든 상세요소에 대한 분석과 논의를 통해 자료표현을 위해 적절한 필수요소, 권고요소를 구분하여 메타데이터 권고안을 마련하였다.

〈표 4〉 CESSDA의 DDI 메타데이터 권고안

| 범 주      | 필수요소  | 권고요소  |
|----------|---|---|
| 도큐먼트 기술  | Title(표제)<br>Identification Number(식별자)<br>Date of Production(발행일)  | Parallel Title(대등표제)<br>Authoring Entity(저작자)<br>Producer(발행자)<br>Copyright(저작권)<br>Distributor(배포기관)   |
| 연구 기술    | Series Name(시리즈명)<br>abstract(조사개요)<br>Time Period Covered(수집기간)<br>Date of Collection(수집날짜)<br>Country(국가)<br>Weighting(가중치) | Keywords(주제어)<br>Topic Classification(핵심분류)<br>Geographic Coverage(지리적 범위)<br>Unit of Analysis(분석단위)<br>Universe(모집단)<br>Time Method(연구방법의 시간적 특성)<br>Data Collector(자료수집자)<br>Sampling Procedure(표본추출과정)<br>Mode of Data Collection(자료수집방법)<br>Location(자료저장위치)<br>Availability Status(이용상태)<br>Extent of Collection(자료의 범위)<br>Restrictions(접근제약사항) |
| 데이터파일 기술 | File Name(파일명)  | Overall Case Count(전체사례수)<br>Overall Variable Count(전체변수수)  |
| 변수 기술    | Variable(변수)<br>Variable Label(변수명)   | Variable Group(변수그룹)<br>Variable Group Label(변수그룹명)<br>Question(질문명)<br>Summary Statistics(통계 요약)<br>Version Statement(변수버전정보)  |

준을 4개 범주로 구분하여 필수요소와 권고요소를 다음과 같이 추출하였다.

다음으로 〈표 4〉에서 추출한 DDI의 12개 필수요소와 25개 권고요소를 기준으로, KSDC 메타데이터 중 통계분석을 위한 요소와 PRISM의 연구용역 기술요소, UKDA 자원관리메타데이터의 자료보존 및 이력에 관한 요소를 종합하였다.

KSDC 메타데이터에서 CESSDA의 권고안과 중복되는 요소를 제거하고, DDI 메타데이터와의 비교를 통해 조사신뢰성을 나타내는 표본 크기, 표본오차, 응답률에 관한 요소를 추가하

였고 웹상에서 통계분석을 지원하기 위해 변수를 기술하는 요소로서 질문유형, 답변코드, 답변내용에 대한 요소를 포함하였다. 특히 DDI 표준에서 제시한 원자료파일기술 영역 중 파일 사이즈, 파일유형, 파일버전, 포맷, 소프트웨어 등 원자료의 장기적 활용을 위한 요소를 포함하였다. 마지막으로 PRISM의 연구용역정보 및 평가정보와 UKDA의 자료보존 및 이력에 관한 정보를 종합하여 이를 연구용역정보, 여론조사정보, 설문정보, 원자료파일정보, 관리 및 활용정보로 범주화하여 〈표 5〉와 같이 통합 메타데이터(안)을 제시하였다.



〈표 5〉 정부여론조사 연구용역 영역 메타데이터(안)

| 영역          | 하위영역       | 요소   |
|-------------|------------|--|
| 연구용역 영역     | 용역개요       | 연구용역ID, 연구용역과제명, 초록, 용역수행유형, 연구기간, 연구비, 발주기관, 계약일, 계약방법            |
|             | 용역담당       | 담당부서, 담당자명, 담당자 연락처, 담당자 이메일, 담당자 직급                               |
|             | 용역수행       | 용역수행기관, 참여연구원명, 참여연구원소속, 참여형태                                      |
| 여론조사 영역     | 연구범위       | 여론조사ID, 여론조사명, 초록, 키워드, 주제, 조사수행기관, 수집기간, 수집날짜, 수집지역, 수집방법, 모집단    |
|             | 방법론 및 자료처리 | 표본크기, 표본추출방법, 표본오차, 응답률, 가중치, 분석단위                                 |
|             | 자료접근       | 접근권한, 제약사항, 자료구성, 자료등록일  |
| 원자료 파일기술 영역 | 원자료 파일기술   | 파일ID, 파일명, 전체 변수수, 전체사례수, 파일사이즈, 파일유형, 파일버전, 포맷, 결측값, 소프트웨어        |
| 설문정보 영역     | 변수그룹       | 변수그룹ID, 변수그룹명, 변수그룹내용  |
|             | 변수         | 변수ID, 변수명, 질문내용, 전체응답수, 변수유형, 척도유형, 보기코드, 보기명, 접근권한                |
|             | 통계요약       | 빈도값, 최대값, 최소값, 평균, 표준편차, 유효값 범위, 비유효값 범위                           |
| 관리 및 활용 영역  | 자료처리 이력    | 처리유형, 처리방법, 처리단계, 처리근거, 처리날짜, 처리자, 변경파일명, 변경파일포맷, 처리검증, 처리에러, 해결방법 |
|             | 활용이력       | 활용구분, 생산자료, 활용평가, 활용날짜, 활용자  |

연구용역 영역은 여론조사자료를 생산하기 위해 수행되는 정책연구용역에 대한 정보를 관리하기 위한 영역이다. 용역수행의 절차, 방법, 예산에 관한 행정적 업무처리과정을 기술하는 요소로 구성되며, 이는 업무활동의 증거적 가치를 보존한다. 본 영역은 용역개요, 용역담당, 용역수행의 3개의 하위영역으로 구성된다.

여론조사 영역은 연구과정 중 실시되는 여론조사에 대한 내용, 시기, 절차, 방법론에 대한 내용을 기술하는 영역이다. 본 영역은 연구범위, 방법론 및 자료처리, 자료접근, 관련자료의 하위영역으로 구분된다.

원자료 파일기술 영역은 실제 응답수치자료를 저장하는 원자료 파일에 대한 논리적 속성과 포맷, 통계적 특징을 기술하는 영역이다. 원자료의 재활용을 위해 파일의 포맷, 유형, 버전 정보를 기술하고 이를 실행할 수 있는 소프트

웨어 정보도 포함한다. 통계적 특징은 향후 통계분석을 위해 필요한 변수수와 사례수를 기술한다.

설문정보 영역은 여론조사자료의 질적 가치를 판단하기 위해 설문의 내용, 구조, 척도 등의 확인을 위한 요소를 기술한다. 여론조사는 알고자 하는 바에 대해 구조화된 설문지를 사용하여 자료를 수집하게 된다. 여론조사자료를 재활용하기 위한 필수요건은 연구목적에 적절한 설문정보를 포함하고 있는지 여부이다. 따라서 활용적인 측면에서 설문지는 설문항, 보기정보 등의 모든 요소에 대해 접근 가능해야 한다. 더불어 웹을 통한 통계분석을 지원하기 위해 설문정보를 구조화하여 DB로 구축해야 한다.

설문정보 영역은 변수그룹, 변수, 통계요약의 하위영역으로 구분한다. 변수그룹영역은 측정하고자 하는 변수간의 내용적 측면이 유사한

경우와 인구통계학적 변수와 같이 배경변수로 활용되는 경우 변수그룹으로 설정한다. 변수는 조사대상의 경험적 속성을 나타내며 그 속성에 계량적인 수치를 부여할 수 있는 개념으로 설명될 수 있으며, 변수의 내용, 구조, 보기 등의 내용으로 구성되어 있다. 통계요약은 개별변수에 대한 간단한 통계처리 결과를 저장하는 필드로서, 기본적인 빈도값과 기술통계를 확인하기 위한 영역이다.

관리 및 활용 영역은 업무활동에 관한 이력 정보를 영구적으로 기록하기 위해 여론조사자료의 관리 및 활용정보 메타데이터이다.

### 5.3 자료보존

정부에서 수행하는 여론조사자료는 모두 전자적 형태로 생산된다. 특히 원자료는 통계분석을 위한 기초파일이므로, 컴퓨터 기술의 업그레이드 및 노후화 등에 직접적으로 종속되며 기술적 변화에 대비하지 못할 경우 원자료의 장기적 활용은 불가능하게 된다. 따라서 원자료의 장기간 유지·보존을 위해 지속적으로 복사, 리포맷, 변환, 마이그레이션 작업을 수행해야 한다. 현재 생산되는 원자료의 유형은 다양한 소프트웨어 및 포맷으로 생산되고 되고 있기 때문에 이를 효과적으로 관리하기 위한 자

료보존포맷 및 자료변환체계를 다음과 같이 제안한다.

#### 5.3.1 자료보존포맷

자료보존포맷은 자료의 보존과 활용적 가치를 지속할 수 있어야 한다. 여론조사자료 중 조사보고서, 설문지, 코드북은 문서자료 형태로 입수된다. 원자료의 경우 통계분석을 위해 활용한 소프트웨어에 의존한 통계분석용 파일형태로 입수된다. 국제 문서보존포맷과 UKDA 양적자료보존포맷을 분석한 결과, 자료유형별 입수포맷과 보존포맷을 정의하면 아래 <표 6>과 같다.

조사보고서, 설문지, 코드북 등 문서자료의 경우 포맷의 범용성, 상호운용성, 표현력, 검색기능, 안정성의 요인을 고려하여 PDF/A를 보존포맷으로 정의하였다. 원자료의 경우, 국제적으로 양적자료를 보존하기 위한 표준이 수립되어 있지 않다. 원자료는 구성된 응답값 자체에 보존가치가 있기 때문에 표현양식 및 서식에 독립적이다. 또한 특정소프트웨어에 의존적이지 않아야 한다. 원자료는 UKDA 자료보존포맷에 근거하여 구분된 텍스트파일로 관리되어야 한다. 텍스트파일은 문자 및 숫자 데이터만으로 구성된 파일으로서 어떤 H/W 및 S/W에서도 이용가능한 호환성이 높은 파일형식이다.

<표 6> 정부 여론조사자료 보존포맷

| 자료유형  | 입수포맷   | 보존포맷                               |
|-------|--|------------------------------------|
| 조사보고서 | 한글 파일, MS-Word 파일, PDF 등                             | PDF/A                              |
| 원자료   | SPSS 파일, SAS 파일, MS-Access 파일, MS-Excel 파일, 텍스트 파일 등 | Delimited Text File 통계소프트웨어 명령문 파일 |
| 설문지   | 한글 파일, MS-Word 파일, PDF 등                             | PDF/A                              |
| 코드북   | 한글 파일, MS-Word 파일, PDF 등                             | PDF/A                              |

개행, 탭, 콤마 등 데이터작성에 필요한 것을 빼고는 제어코드를 포함하지 않아서 파일이 가볍다. 또한 원자료를 구성하는 문자 및 숫자데이터를 표현에 제약사항이 없다. 원자료의 해석을 위해 필요한 변수구분은 탭 또는 콤마로 구분된다. 또한 원자료의 온라인 통계분석을 지원하기 위해 DB로 이전할 수 있는 SQL 쿼리를 포함하여 관리될 필요가 있다. 또한 이용자의 다양한 환경에서의 통계분석 지원을 통해 SPSS, SAS, STATA에서 사용가능한 명령문 파일을 첨부하여 보존한다.

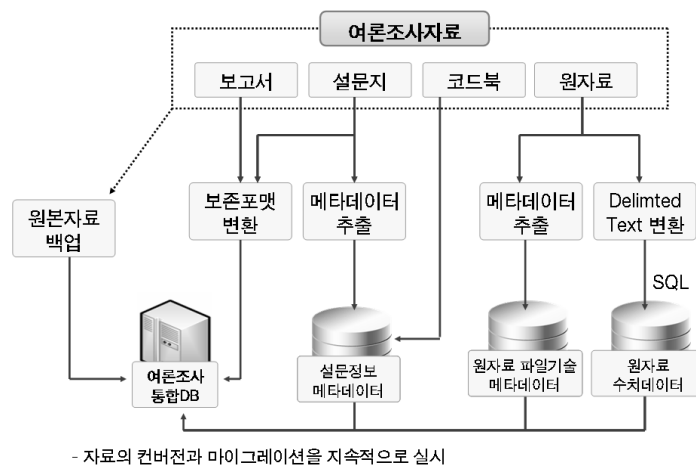
### 5.3.2 자료변환절차

자료를 변환하는 목적은 정보기술의 외부환경 변화에 대해 영향을 최소화할 수 있는 표준포맷으로 전환하여 그 보존가치를 유지함에 일차적 목적이 있다. 또한 여론조사자료 중 원자료, 설문지의 경우 다양한 소프트웨어에서 생산된 파일에 대하여 동일한 환경 내에서 언제든지 통계적 재활용이 가능하도록 하는 활용적

가치를 유지함에 이차적 목적이 있다.

여론조사자료를 구성하는 자료의 유형을 보존·활용목적에 따라 변환되는 과정을 도식화하면 <그림 2>와 같다.

여론조사자료가 입수되면, 조사보고서, 설문지, 코드북, 원자료의 원본자료는 모두 백업하여 보관한다. 여론조사 통합시스템에 등록이 되면 지속적인 컨버전과 마이그레이션 등의 보존활동을 통해 자료의 활용성을 유지한다. 조사보고서는 텍스트 기반의 문서파일이므로, 별도의 자료변환과정 없이 보존포맷으로 전환한다. 설문지는 설문지의 내용 및 구조분석을 통해 설문정보 메타데이터 요소에 부합하는 데이터를 추출한다. 추출된 데이터의 검증 및 확인 작업을 수행하고 최종적으로 메타데이터 DB에 등록한다. 원자료의 특성 및 구조에 관한 내용은 원자료 파일기술 메타데이터 DB에 등록하고, 실제 분석대상이 되는 변수코드 및 응답값에 대한 데이터는 구분된 텍스트형식으로 그 포맷을 변환하여 SQL 질의어를 통해 원자료



<그림 2> 여론조사자료 변환 과정

수치DB에 구축한다. 원자료 수치DB와 설문정보 메타데이터는 온라인상에서 통계분석을 바로 실시할 수 있는 기반이 된다. 최초 수집된 여론조사자료는 자료의 진본성을 유지하기 위해 여론조사 통합DB에 백업하여 보존한다.

## 6. 결론

이 연구에서 수행한 정부 여론조사자료의 장기적 보존과 활용을 위한 아카이브 요건을 정리하면 아래와 같다.

첫째, 정부 여론조사는 정책연구용역과 학술연구용역을 통해 수행된다. 현재 연구용역관리 시스템인 PRISM의 시스템 확장을 통해 모든 여론조사자료를 의무적으로 수집해야 한다. 이를 위해 유관기관과의 자료이관이 실시되어야 하며, 수집된 자료는 내용의 적절성, 공공성, 업무 활용성, 민간 활용성 등의 내용적 평가와 자료의 정확성, 신뢰성, 완전성, 호환성을 포함한 여론조사의 절차 및 자료에 대한 평가가 병행되어야 한다.

둘째, 본 연구에서는 해외 사회과학 양적자료 메타데이터 표준인 DDI와 국내 KSDC, KOSSDA 메타데이터 기술요소의 분석을 통해 여론조사 자료의 내용과 구조를 포괄하여 기술할 수 있는 통합메타데이터 요소(안)을 제안하였다. 여론조사자료는 DDI와 표준과 같은 양적자료에 대한 국제표준을 준수하면서, 자료생산 및 보존에 대한 국내의 환경을 반영한 통합 메타데이터를 구축해야 한다.

셋째, 여론조사자료의 시스템을 통한 이차적

활용을 제공하기 위한 자료변환이 필요하다. 특히 본 연구에서는 설문지와 원자료의 통계분석 지원을 위한 데이터베이스 구축방안을 제안하였다.

넷째, 자료의 장기보존을 위해서는 유형별로 보존포맷이 제시되어야 한다. 최초 수집자료는 자료의 진본성 유지를 위해 백업하고, 소프트웨어와 하드웨어환경변화에 적합한 장기보존 포맷으로 전환해야 한다. 특히 통계적 특징이 있는 원자료는 구분된 텍스트파일로 보존하고, 기타 문서자료의 경우에는 PDF/A를 보존포맷으로 제안하였다.

이상 본 연구에서는 여론조사자료 아카이브 구축을 위한 수집에서부터 보존까지의 일련의 과정 순으로 방안을 제시하였다. 완전한 여론조사자료 아카이브 구축을 위해서는 자료의 생산과정에서부터 표준화된 절차와 방법으로 관리해야 할 필요가 있다. 생산과정에서 자료의 완전성과 신뢰성이 결여되는 경우 보존과 활용 가치는 저하되기 때문이다.

더불어 여론조사자료는 공공기록물의 범주에 포함하여 관리할 수 있는 방안이 모색되어야 한다. 여론조사자료는 특수자료이므로 공공기록물의 범위에 포함되지 않으며, 이를 효과적으로 구축할 수 있는 메타데이터와 보존포맷도 제시되어 있지 않다. 여론조사자료 역시 기록물의 유형에 포함되므로, 의사결정과정, 업무처리과정 등 업무맥락정보에 대한 정보도 연계되어 관리되어야 한다. 정부 연구용역으로 생산된 여론조사자료는 기록물 관련 법·제도 정비를 통해 체계적으로 관리되어야 하며, 향후 이에 대한 활발한 연구가 필요하다.

## 참 고 문 헌

- 김광웅, 임동욱, 황중우, 우병렬. 1996. 『춤 여론 바로 보기』. 서울: 나남.
- 남영준. 2005. 국가 통계표준 메타데이터 설계에 관한 연구. 『정보관리연구』, 36(1): 33-56.
- 박무익. 1988. 한국여론조사 실태와 조사방법. 『산업과경영』, 8(1): 711-719.
- 이기성, 안승철. 1997. 여론조사의 실태와 방법. 『기초과학연구소 논문집』, 2: 105-115.
- 이남영. 2001. 『사회과학데이터분석시리즈』, 서울: 한국사회과학데이터센터.
- 이소연. 2002. 디지털 아카이빙의 표준화와 OAIS 참조모형. 『정보관리연구』, 33(3): 45-68.
- 이해용. 1998. 일간지에 발표된 여론조사의 변천에 관한 연구. 『기초과학연구』, 16(1): 131-138.
- 임진희. 2006. 전자기록의 장기보존을 위한 보존정보패키지(AIP) 구성과 구조. 『기록학 연구』, 13: 41-90.
- 통일부. 2006. 『통일부 여론조사자료 DB 구축에 관한 연구』, 연구보고서.
- 황중우. 2006. 『여론조사가 정책에 미치는 영향과 조사오류에 관한 연구』. 서울대학교 석사학위논문.
- 홍기선. 1988. 여론과 여론조사. 『산업과경영』, 8(1): 749-767.
- Altman, M., L. Andreev, M. Diggory, M. Krot, G. King, D. Kiskis, E. Kolster, A. Sone and S. Verba. 2001. "A Digital Library for the Dissemination and Re-  
plication of Quantitative Social Science Research." *Social Science Computer Review*, 19(4): 458-471.
- Blank, G. and K. B. Rasmussen. 2004. "The Data Documentation Initiative: The Value and Significance of a Worldwide Standard." *Social Science Computer Review*, 22(3): 307-318.
- Gallup, G. 1972. *The Sophisticated Poll Watcher's Guide*. Gallup Press.
- O'Toole, J. M. 1990. *Understanding Archives and Manuscripts*. 이승익 역. 2004. 『기록의 이해』. 서울: 진리탐구.
- Rasmussen, K. B. and G. Blank. 2007. "The Data Documentation Initiative: A Preservation Standard for Research." *Archival Science*, 7(1): 55-71.
- Ragin, C. 1996. *Guide to Resources and Services*. ICPSR.
- Ryssevik, J. and S. Musgrave. 2001. "The Social Science Dream Machine." *Social Science Computer Review*, 19(2): 163-174.
- Vardigan, M. and C. Whiteman. 2007. "ICPSR Meets OAIS: Applying the OAIS Reference Model to the Social Science Archive Context." *Archival Science*, 7(1): 73-87.

**【참고사이트】**

브리태니커백과사전.

〈<http://premium.britannica.co.kr/>〉

대한민국정책포털. 〈<http://www.korea.kr/>〉

정책연구용역종합관리시스템(PRISM).

〈<http://www.prism.go.kr/>〉

한국사회과학데이터센터(KSDC).

〈<http://www.ksdc.re.kr/databank>〉

한국사회과학자료원(KOSSDA).

〈<http://www.kosdda.or.kr>〉

CESSDA(Council of European Social Science Data).

〈<http://www.nsd.uib.no/cessda/>〉

DDI(Data documentation initiative).

〈<http://www.ddialliance.org>〉

ICPSR. 〈<http://www.icpsr.umich.edu/>〉

UKDA. 〈<http://www.data-archive.ac.uk>〉