

디지털 연구데이터 장기보존의 편익에 대한 연구

A Study on Benefits of Digital Preservation of Research Data

현 문 수(Moonsoo Hyun)*

목 차

- | | |
|---------------------------|--------------------------------|
| 1. 서론 | 3.2 산출물 득점표와 비용 템플릿을 이용한 편익 산출 |
| 1.1 연구의 배경과 목적 | 3.3 직·간접적 비용 절감과 잠재적 혜택 |
| 1.2 선행연구 | 3.4 세 가지 측면에서 바라본 편익 |
| 2. 디지털 연구데이터 보존의 필요성 | 4. 디지털 연구데이터 편익 규명을 위한 제안 |
| 2.1 연구 결과 검증의 전통적 기능 | 4.1 편익 확인 및 산출을 위한 객관적인 도구 개발 |
| 2.2 다양한 측면에서의 보존 요구 증가 | 4.2 편익 확인 및 산출을 위한 다양한 연구 |
| 3. 디지털 연구데이터의 보존 편익 개발 사례 | 5. 결론 및 제언 |
| 3.1 편익 개발의 필요성 | |

<초록>

본 연구는 디지털 연구데이터의 보존을 통해 얻을 수 있는 편익을 확인하기 위한 것이다. 경제학적 측면에서 디지털 연구데이터의 보존을 바라보고 있는 연구 사례를 살펴봄으로써 양적인 측면과 질적인 측면에서 디지털 연구데이터 보존을 통해 얻을 수 있는 광범위한 혜택을 분석한다. 이를 통해 국내의 디지털 연구데이터를 체계적으로 보존·관리하기 위한 시작점으로, 보존활동에의 투자를 이끌어낼 수 있는 편익을 확인하고 산출하기 위한 방안을 제시하고자 한다.

주제어: 디지털 연구데이터, 디지털 보존, 보존 편익

<ABSTRACT>

The study identifies benefits arising from the digital preservation of research data. Through analysing studies on digital preservation and digital research data in a economic perspective, broad benefits from digital preservation of research data have been identified and analysed quantitatively as well as qualitatively. It provides the starting point of systematic management and preservation of digital research data. Furthermore, it attempts modelling for investigating and producing benefits both in quantity and quality that is fundamental to promote investments.

Keywords: digital research data, digital preservation, benefit

* 한국외국어대학교 정보·기록관리학과 박사과정(수료)(puremoon@hanmail.net)

■ 접수일: 2011년 5월 21일 ■ 최종심사일: 2011년 6월 1일 ■ 게재확정일: 2011년 6월 22일

1. 서론

1.1 연구의 배경과 목적

연구자 10명 중 3명이 연구데이터를 분실하거나 소실한다(BRIC 2009). 2009년 BRIC의 조사에 따르면 조사대상 연구자 중 1/3 가량이 연구실에서 생산한 실험데이터를 소실 또는 분실한 경험이 있으며, 그 중 대부분은 분실 또는 소실로 인하여 문제가 발생하였다고 답한 것이다. 또한, 조사 대상의 반 수 가량은 그 주요 원인을 데이터 관리에 대한 인식 부족으로 보고 있었다(BRIC 2009).

연구데이터(research data)는 실험의 재료나 과정 및 결과, 관찰이나 현장조사 및 설문조사의 결과 등의 원(raw) 자료로(서울대학교 2010), 연구의 기초가 되는 다양한 데이터를 일컫는다. 과학 데이터(scientific data)라고도 불리는데, 천체망원경 등의 관찰 장비에서 생산되는 데이터나 인구조사 등의 통계데이터, 실험이나 시뮬레이션 등을 통해 시스템이 생산한 데이터 등이 이에 속한다. 따라서 이러한 연구데이터가 디지털 형식으로 생산되어 관리되는 것, 즉 연구의 과정에서 디지털 형태로 생산되는 원 자료를 통칭하여 디지털 연구데이터(digital research data)라 말할 수 있겠다.

정보기술 및 네트워크 기술의 발달과 함께 디지털 연구데이터의 효과적이고 효율적인 관리와 이용이 학계의 중요한 관심 대상이 되었다. 연구데이터가 연구 수행과 결과물의 배포 및 이용에 중요한 부분을 차지할 것이라는 인식도 크게 증가하고 있다. 이에 미국과 유럽에서는 연구의 최종결과물 뿐 아니라, 연구과정

에서 생산된 디지털 연구데이터에까지 관심을 넓혀, 그 보존과 활용에 대한 방안을 모색하고 있다(Fry et al. 2008; National Academy of Science 2009; Koski et al. 2009, 3; Key Perspectives Ltd 2010). 또한 디지털 연구데이터의 보존을 바라보는 시각도 정보기술의 측면에서 조직의 정책이나 활동의 지속가능성의 측면에까지 확대되고 있다.

디지털 보존을 경제적 활동으로 보고, 이를 지속적으로 수행하기 위한 연구를 수행하고 있는 Lavoie(2004)는 디지털 보존을 담당하는 조직이 확대·분화됨에 따라 보존에 대한 보상이 불명확하게 되면서 디지털 보존을 둘러싼 문제는 더욱 심각해지고 있다 지적한다. 지속적인 디지털 보존 활동을 보장하기 위해서는 디지털 보존을 통해 창출되는 가치를 제공하는 철학에 기초해 세심하게 계획을 수립하는 기반이 마련되어야만 가능하다는 것이다. 나아가 이러한 가치 평가는 자료 이용의 구체적 맥락과 이용자에게 전달할 수 있는 잠재적 이익, 자료 보존에 책임이 있는 기관의 미션 및 주요 목적과 밀접하게 관련된다(Currall & McKinney 2006).

물론 디지털 연구데이터의 보존은 중요하다. 그러나 “중요하기 때문에 디지털 연구데이터를 보존해야 한다”라는 주장만으로는 서로 다른 설립·운영 목표를 지닌 기관이 디지털 연구데이터를 보존하도록 하는 동기를 불러넣기에 미약하다. “중요하다”에서 더 나아가 “왜 중요한가”, 그리고 “왜 해야 하는가”라는 질문에 답을 할 수 있어야 한다.

이에 본 연구에서는 여러 각도에서 디지털 연구데이터 보존의 필요성을 살펴보고, 디지털 연구데이터 보존을 통해 실현시킬 수 있는 편

익을 확인하여 이를 양적·질적으로 표현함으로써 디지털 연구데이터를 지속가능하게 보존하고 미래에까지 이용할 수 있도록 하는 기반을 제공하고자 한다.

1.2 선행연구

디지털 연구데이터 보존은 고사하고 디지털 보존에 대한 비용이나 편익에 대한 연구는 2000년대 이후에서야 조금씩 이루어지고 있을 정도로 그 역사가 짧다. 대부분 연구의 초점도 디지털 보존 활동을 기반으로 비용 요소를 도출하는 데에 두고 있는 실정이다. 경제적인 측면에서의 연구는 아니지만 국내의 과학데이터(scientific data)의 보존과 관련된 연구로는 대표적으로 김선태, 한선화, 이태영, 김용(2010)의 연구가 있다

한편, Lavoie(2004)를 시작으로 Lavoie와 Dempsey(2004), Currall과 McKinney(2006: 2007), Fry 외(2008), Beagrie, Chruszcz와 Lavoie(2008), Beagrie, Lavoie와 Woollard(2010) 등 소수의 연구자들이 경제학적 관점에서 디지털 보존 및 디지털 연구데이터의 보존을 연구하고 있다.

기존의 연구가 주로 기술적(technological) 관점에서 디지털 보존을 바라보았다면 Lavoie(2004) 및 Lavoie와 Dempsey(2004) 등은 경제학·문화사회학적 관점이라는 새로운 시각에서 디지털 보존을 연구한다.

Lavoie와 Dempsey는 디지털 보존은 따로 떼어 해결할 수 있는 기술적 문제가 아니라, 디지털 정보환경을 구성하는 서비스와 정책, 이해관계자가 상호 연결된 광범위한 집합체의 한

부분이라 정의하였다. 따라서 디지털 보존을 위해서는 정보의 생명주기에 분산된, 주의 깊은 디지털 자산 관리 실무 준수를 포함해야 한다는 인식이 필요하다 강조하며, 경제·사회·문화적 입장에서 13가지의 서로 다른 디지털 보존의 정의를 열거하였다. 한편 Lavoie는 경제학적 관점에서 디지털 보존의 3가지 영역을 책임(responsibilities), 보상과 장려(incentives), 보존활동의 체계(organisation)로 재정의 하였다. 전통적인 자료나 기록관리기관 밖에서까지 보존 행위가 이루어져야 하는 환경 변화로 인해 그 책임이 분산됨에 따라 책임을 확인하고, 책임 기관에게 돌아가야 할 보상을 만들어 내야 함은 물론, 이들을 아우를 수 있는 체계가 필요하다는 것이다. 즉, 보존활동은 투자인데, 조직의 1차적인 목표가 디지털 자원의 보존이 아닌 기관의 입장에서는 의사결정자가 투자에 대한 보상을 예상할 수 있어야 한다는 것이 그 핵심이다.

구체적으로 실제 사례에서 디지털 보존의 편익을 확인하고, 이를 일반화하려는 노력도 Currall과 McKinney(2006: 2007), Lyon(2007), Fry 외(2008), Beagrie, Chruszcz와 Lavoie(2008), Beagrie, Lavoie와 Woollard(2010) 등의 연구에서 이뤄지고 있다. Currall과 McKinney는 소요되는 비용 뿐 아니라 비용을 지출하게 된 근거와 이유의 중요성을 강조하였다. 이에 연구 기금을 지원할 프로젝트를 객관적으로 평가함으로써 기금 지원으로 수행되는 연구에서 산출되는 혜택을 공정하게 예측하기 위한 도구를 개발하고자 하였다. 특히 Lyon이나 Fry 외, Beagrie, Chruszcz와 Lavoie, 그리고 Beagrie, Lavoie와 Woollard는 디지털 연구데이터의 보

존에 중점을 두어 편익을 바라보고 있다. Lyon은 디지털 연구데이터 큐레이션과 보존 인프라 구축을 위해 연구 재정지원 기관과 협력하여 비용-편익 분석을 수행해야 함을 제안하였다. Fry 외는 연구데이터의 큐레이션과 공유를 통해 얻을 수 있는 혜택을 확인하고, 이를 계산할 수 있는 방법론을 도출하고자 하였으며, Beagrie, Chruszcz와 Lavoie 및 Beagrie, Lavoie와 Woollard는 디지털 연구데이터의 보존비용 연구와 더불어 그 편익을 다양한 측면에서 분석하고자 하였다.

이러한 연구들을 배경으로 하여 본 연구는 경제학적 측면에서 디지털 연구데이터 보존을 바라보고 이를 통해 얻을 수 있는 편익을 고찰해보고자 한다.

2. 디지털 연구데이터 보존의 필요성

2.1 연구 결과 검증의 전통적 기능

전통적으로 연구데이터는 연구결과를 검증하는 중요한 기능을 수행해왔다. 학술저널 기사나 보고서 등을 통해 발표되는 수많은 연구 결과의 검증에 연구데이터는 필수적이다. 실험 데이터를 재현할 수 없다면 실험결과에 대한 오류 판단이 어려우며, 관측 데이터가 부재하다면 이를 바탕으로 한 결과를 신뢰하기 힘들어진다. 사회과학분야의 연구를 통해 생산된 데이터를 제시할 수 없다면 연구결과에서 도출한 일반화된 명제를 객관적으로 납득시키기 어렵다. 이렇듯 데이터를 이용하는 연구나 연구

과정에서 데이터를 생산하는 연구의 경우에는 연구결과 뿐만 아니라 연구데이터도 중요한 산출물로 여겨져, 실험실이나 연구자 개인 차원에서 관리와 보존이 이루어져왔다.

나아가 연구데이터의 부재나 비공개를 용인하는 사회 분위기가 팽배해질 경우에는 연구결과의 검증을 벗어나 연구윤리의 문제까지 대두될 수 있다. 일례로 지난 2005년~2006년에 걸친 황우석 사건은 우리나라의 연구 프로세스를 재검토하고 연구과정의 부산물인 연구 노트와 연구데이터에 대한 중요성을 각인시켜 주는 동시에, 학계나 사회에서 연구윤리라는 문제를 돌아보게 되는 중요한 계기를 마련해주었다.

2.2 다양한 측면에서의 보존 요구 증가

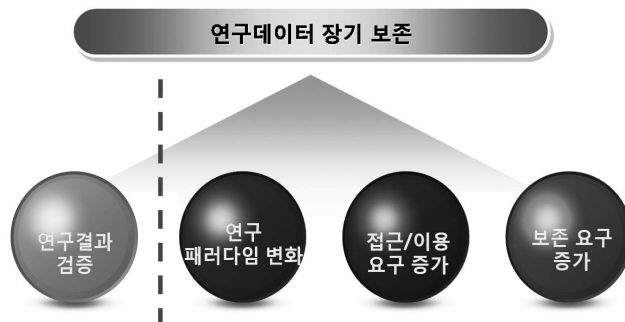
그런데 점차 전통적인 검증 기능 외에 다른 측면에서 연구데이터의 보존을 요구하는 목소리가 높고 있다. 그 첫 번째는 연구패러다임의 변화로 인한 연구데이터 보존 요구의 증가다. 점차 거의 모든 학문분야에서 데이터를 이용하여 연구문제를 도출하며, 데이터를 이용하는 연구방법을 활용하는 등 데이터를 중심으로 연구를 진행하고 있다(Koski et al. 2009, 3). 다시 말하면 일부 학문분야에서 기존의 가설을 검증하기 위해 연구데이터를 활용하였던 과거와는 달리, 거의 모든 학문분야에서 연구데이터를 이용해 새로운 가설을 제안하거나 신규 연구를 창출하는 연구 환경으로 변화한 것이다. 주로 컴퓨터 기기와 성능의 급속한 발전으로 인한 변화라 하겠는데, 예전보다 더욱 빠르게 신규 연구의 자료가 되는 대용량 데이터베이스

를 분석할 수 있게 되고, 나아가 더 많은 연구에 기존의 연구데이터를 활용할 수 있게 된 것이다. 나아가 영국을 비롯한 유럽과 호주 등의 사례에서처럼 연구 재정지원 기관들이 기금 지원 시 연구데이터까지도 결과물에 포함해 제출하도록 하거나, 연구수행기관에서 일정 기간 연구데이터를 보유하도록 강제하는 경우가 늘고 있어, 체계적인 연구데이터 보존의 중요성은 더욱 증가하고 있다.

두 번째 변화는 기존의 연구과정에서 생산된 연구데이터에 대한 이용이 급속하게 증가하고 있다는 점이다. 점점 국제적 공조로 이루어지는 연구가 늘어나고 있고, 이에 따라 시공간을 초월하여 연구데이터를 활용하고자 하는 요구도 늘어나고 있다. 그러나 현재 상당수의 연구데이터는 고립된 레파지토리에서 보관하고 있거나 연구자 개인 또는 실험실의 개인 저장소에 위치하는 경우가 많아 외부에서 자유롭게 이용하기 힘든 실정이다. 이렇게 늘어가는 이용 요구를 만족시킬 수 있을 만큼 충분히 연구데이터가 공유되지 않음으로 인해, 연구데이터에 대한 접근가능성의 문제는 물론 자원의 상호연결가능성과 상호운용가능성으로까지 문제

가 확산되고 있다. 결국은 수많은 시간과 노력, 인력, 예산이 투입되어 수행되는 연구의 산출물을 효과적으로 활용하여 연구의 효율성을 향상시키고자 하는 요구가 나날이 증대되고 있는 실정이다.

마지막으로는 점차 제도적으로 연구데이터 보존에 대한 요구를 반영하고 있다는 점이다. 호주의 경우 *Australian Code for the Responsible Conduct of Research*(Australian Government, 2007, 2.1.1)에 따라 기본적으로 5년간 연구데이터를 보유하도록 강제하고 있으며, 우리의 경우는 「연구개발사업 연구노트 관리지침」(교육과학기술부 2009) 제11조 2항에 따라 국가연구개발사업에서 생산한 연구 노트를 연구기관에서 30년간 보유하도록 하고 있다. 연구과정에서 생산된 연구데이터세트는 학문적 지식기반의 일부분을 구성하며 미래 세대에 현대의 지식을 전달한다. 그러므로 법규에서 정한 연구데이터의 보유기간을 준수함은 물론, 생명주기 이상의 장기보존할 수 있다면 동일한 데이터를 다시 생산하는데 드는 비용 절감 효과까지 누릴 수 있게 된다(〈그림 1〉 참조).



〈그림 1〉 연구데이터 장기 보존 필요성

3. 디지털 연구데이터의 보존 편익 개발 사례

3.1 편익 개발의 필요성

디지털 연구데이터의 장기 보존에 대한 요구는 늘어가고 있는 반면, 아직 현실적으로 충분한 여건을 마련하지 못하고 있는 것이 사실이다. 여전히 소재파악도 불가능한 디지털 연구데이터가 허다하며 보존의 책임도 구체적으로 확립되지 않아 체계적인 관리나 보존을 기대하기는 어렵다. 가장 근본적인 이유는 디지털 연구데이터의 보존활동은 필수불가결하게 투자를 동반하기 때문이다. 어떤 방식으로든 디지털 보존 활동을 수행하기 위해서는 비용과 자원이 투입되어야 한다. 그런데 현실에서 이러한 투자를 결정하는 주체는 보존활동의 담당자가 아니라 조직의 의사결정자이거나 연구에 재원을 지원하는 연구기금 지원기관이다. 따라서 투입 자원, 즉 투자를 통해 객체 보존에 영향을 줄 수 있도록 하기 위해서는 의사결정자가 투자에 대한 보상을 예상할 수 있어야만 한다(Lavoie 2004). 물론 그 보상은 조직의 금전적 이윤과 같은 가시적인 것일 수도 있고, 명예 상승이나 위험 감소와 같이 무형의 혜택일 수도 있다.

이렇듯 투자에 대한 보상, 즉 혜택이나 편익을 예상하기 위해서는 이에 대한 체계적인 분석과 연구가 필요한데, 현재까지는 비교적 편익 분석에 대한 조명이 없는 상태이다. 여전히 “미래 세대를 위해 사회의 디지털 기록을 보존해야 한다”는 상식선에서의 가정이나 의사결정 과정에서 이용하기에는 너무 일반적인 편익만

이 흔히 제시되고 있다(Beagrie, Lavoie, & Woolard 2010). 또 편익을 확인할 수 있다고 하더라도, 그 혜택이 오랜 시간 이후에 실현될 다거나 디지털 연구데이터를 보존하는 조직을 훨씬 넘어서는 범위에 있는 등 의사결정자가 조직 내부에서 단기간에 실현될 것으로 기대되는 보상을 상상하기란 어렵다.

이에 편익 분석 연구에 선행하여 “과연 얼마나 비용을 투입해야 하는가”라는 의문에서 여러 가지 각도에서 디지털 연구데이터의 보존비용을 예측하고자 하는 시도가 있었다. 활동의 단계별로 비용 요소를 연구하거나 마이그레이션·에물레이션 등의 서로 다른 보존 전략을 적용했을 때의 비용에 대한 연구 등이 그것이다.

그런데 보존 활동이 지속적으로 수행되어야 하는 일상의 실무라는 점을 고려해볼 때, 보존활동에 소요되는 비용만이 아니라 “왜” 그만큼의 비용을 투자해야 하는지라는 질문에 답할 수 있는 연구를 함께 수행할 필요가 있다. 점차 디지털 객체의 관리·보존 책임이 분산되고 있는 현대의 환경(Lavoie 2004)을 고려한다면, 연구데이터 관리·보존이 그 설립 목적이 아닌 기관인 경우 디지털 연구데이터의 유형·무형의 가치를 충분히 드러내고 전달하고자 하는 노력은 곧바로 보존 활동의 정도와 직결된다. 따라서 연구데이터 보존의 비용 편익 뿐 아니라 무형의 편익을 다양한 시각에서 바라보고자 한 3개 연구를 대표적으로 살펴봄으로써 디지털 연구데이터 장기보존의 혜택을 객관적으로 드러낼 수 있는 방법을 모색해보고자 하였다.

3.2 산출물 득점표와 비용 템플릿을 이용한 편익 산출

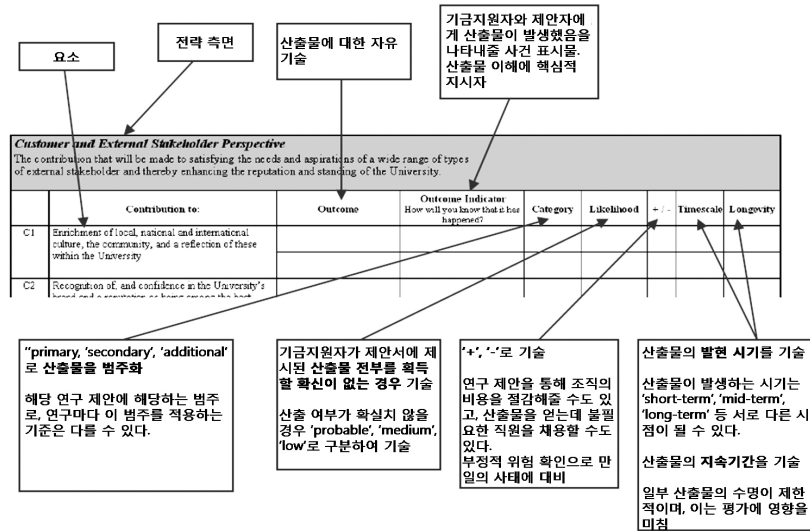
JISC 지원으로 진행된 글래스고 대학의 espida 프로젝트는 정보 객체의 수명을 보장할 수 있는 충분한 비즈니스 케이스를 만들기 위해 무형의 자산 가치를 어떻게 평가할 수 있는가를 연구했다(Currall & Mckinney 2007). 특히, 연구에의 지원이나 투자를 통해 회수될 것으로 예상되는 혜택을 계량화된 투자 수익으로 측정할 수 없는 경우에도 충분한 정보를 기반으로 의사결정을 할 수 있도록 지원하는 도구를 개발하고자 하였다. 비록 espida 프로젝트가 연구 데이터로 그 영역을 한정하지는 않았으나, 연구데이터를 산출하는 연구를 효과적으로 지원하기 위한 결정을 돕는, 큰 틀에서의 편익을 확인할 수 있는 객관적인 도구를 개발했다는 점에서 주목할 필요가 있다.

이 프로젝트의 관심은 의사결정자와 연구 제안자 사이의 관계를 재조정하는 것이다. 즉, 객관적인 도구를 사용함으로써 의사결정자의 측면에서는 첫째, 무형의 수익을 가져올 제안서에 대해 충분한 정보에 기초하여 기금 지원을 결정할 수 있으며, 둘째, 전문가들이 수행할 프로젝트의 제안서를 이해할 수 있고, 셋째, 긍정적·부정적인 결과물의 가능성을 예측할 수 있게 되어 결국 투자 효용을 측정할 수 있게 된다. 연구 제안자의 경우 조직의 전략 아젠다에 부합하도록 제안서를 평가해 조정할 수 있고, 의사결정자가 이해할 수 있는 용어를 사용하여 제안서의 가치를 전달할 수 있어 무형의 혜택을 효과적으로 전달할 수 있게 된다. 결국은 왜 보존활동을 수행해야 하며, 기관의 주요 업무가 아닌 보존활동에 왜

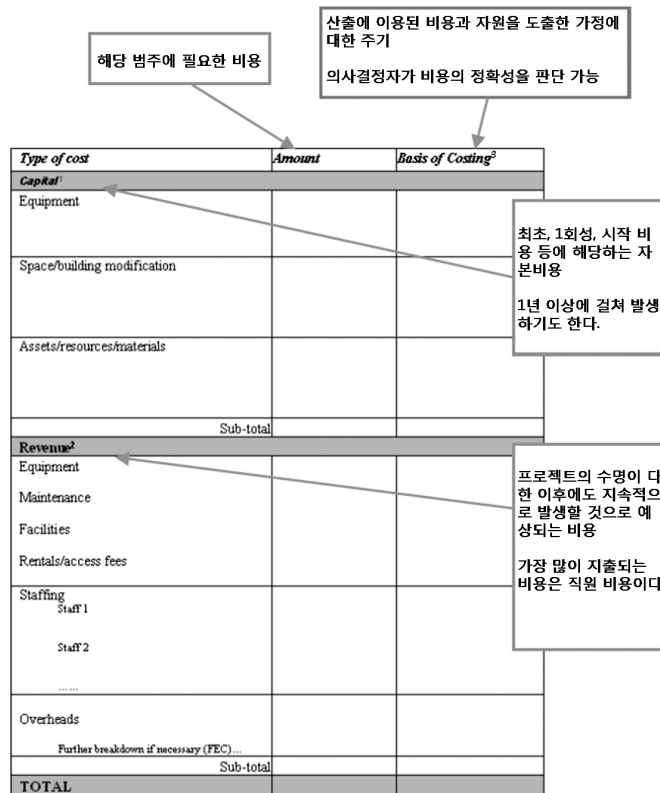
비용을 지출해야 하는지 등 보존활동이 주요 사명이 아닌 조직의 상위 관리자가 의문을 품고 있을 질문에 효과적으로 대응할 수 있다.

Currall과 Mckinney(2007)는 연구의 목표가 얼마나 조직의 전략 목표와 동일 선상에 있는지를 보여주는가에 따라 대부분 제안서의 통과 여부가 결정된다고 판단했다. 이에 고객과 외부 이해관계자의 경험 향상, 신규 연구 수행의 길 마련, 비용 절감 및 수입 발생, 효율성·유연성·능력 증가 등을 조직의 핵심 전략으로 보았으며, 이를 기반으로 이용자 및 외부 이해관계자, 내부 업무 프로세스, 재무, 혁신 등 4개 측면에서 프로젝트를 평가하는 산출물 득점표(outcome scorecard)를 개발하였다. 더불어 투입 비용을 개괄적으로 표현하기 위해 비용 템플릿(cost template)을 활용해 연구 지원을 결정하는 도구를 개발하였다. 산출물 득점표는 그림과 같이 각 측면별 평가 요소에 대해 어떤 산출물이 생산될 것이며 이에 대한 다양한 평가 점수를 기입할 수 있도록 구성되었다(〈그림 2〉 참조).

투입 비용과 관련하여서는, 보존 활동에 필요한 비용 범주를 디지털 객체의 수명을 보장할 수 있는 시스템이 디지털 객체를 입수하는 비용과 입수된 디지털 객체의 관리·보존·이용 비용, 즉 필요한 기간만큼의 기술(technology) 갱신·서비스 운영·신규 실무 적용을 위한 직원 훈련·프로세스 평가 비용으로 나누어 설명하였다. 이에 구체적인 비용으로 제시할 수 있는 편익 산출의 바탕이 될 비용 템플릿을 함께 개발하고 있다. 비용 템플릿은 그림과 같이 비용의 유형과 그 양, 비용 산정의 근거를 자본 비용(capital cost)과 재정비용(revenue cost)으로 나누어 기술할 수 있도록 구성되었다(〈그림 3〉 참조).



<그림 2> 산출물 득점표 설명(Currall & McKinney 2007; 재구성)



<그림 3> 비용 템플릿 설명(Currall & McKinney 2007; 재구성)

3.3 직·간접적 비용 절감과 잠재적 혜택

JISC의 디지털 레퍼지토리 프로그램에 따라 2005년 이래 연구데이터에 대한 다양한 연구가 수행되고 있다. 그 중 Fry 외(2008)는 양적·질적 방식으로 연구데이터 큐레이션 및 공유에서 오는 혜택을 확인하고 객관적으로 이를 측정하기 위한 연구를 진행하였다. 이들은 문헌연구를 바탕으로 European Bioinformatics Institute (EBI) 및 Economic & Social Data Service의 Qualidata 대상의 사례연구를 수행하였으며, 비용 절감이라는 직접적인 혜택 이외에도 효율성 개선, 새로운 연구를 통한 가치를 창출할 수 있는 신규 기회 제공 등 다양한 방식의 혜택이 존재함을 밝혔다.

드러난 주요 혜택으로는 1) 데이터 수집에 소요되는 투자 극대화, 2) 개별 연구자나 기관의 부담 비용 없이 광범위한 접근 확대, 3) 기

존의 데이터로부터 새로운 발견을 이끌어낼 수 있는 가능성 확대, 4) 데이터 컬렉션 중복 생산 비용 감소 및 과학기록의 투명성 증가, 5) 연구의 영향력 증가 및 이러한 영향 파급에 소요되는 시차 감소, 6) 신규 협력 및 신규 지식기반 산업 확대 등을 들 수 있다. 더욱 광범위한 사회의 측면에서 바라본다면 1) 연구 재정지원 상의 투명성 향상, 2) 학생들의 데이터에 대한 인지력 향상을 위해 교육계에서 데이터세트 이용 가능, 3) 광범위한 데이터에 접근하게 되어 연구자의 전문성 향상, 4) 데이터의 질을 향상시킬 잠재력을 지닌 도구와 표준 개발 가능, 5) 기관과 연구자의 가시성·명망 증가 등의 혜택 까지도 누릴 수 있다고 밝혔다(〈표 1〉 참조).

오랜 시간에 걸쳐 누적되는 혜택의 성격과, 대부분 미래에 실현된다는 점을 고려한다면 표로 제시한 편익을 모두 계량화하기에는 무리가 있으며 심지어 불가능하다 하겠다. 그럼에도

〈표 1〉 비용 대 편익(Fry et al. 2008)

	비 용	편 익
직·간접 (재정적)	<ul style="list-style-type: none"> • 큐레이션과 보존에 필요한 기반 제공 <ul style="list-style-type: none"> - 직원 - 데이터 저장·검색용 h/w와 s/w - 데이터 조직 및 접근을 위한 도구와 표준 개발·적용 - 데이터학자 훈련 - 학계의 주의를 끌기 위한 외부활동과 훈련 	<ul style="list-style-type: none"> • 데이터 수집에의 투자 극대화 <ul style="list-style-type: none"> - 개별 연구자·기관에게는 비용 없는 접근 확대 - 기존 데이터를 이용한 신규 연구문제 도출 - 데이터 중복 생산비용 감소 • 연구의 영향력 증가 <ul style="list-style-type: none"> - 신규 협력 - 영향력 인지에 소요되는 시간 감소 및 연구 속도 증가 • 신규 지식기반 산업
확 대 (사회문화적)	<ul style="list-style-type: none"> • 차이 <ul style="list-style-type: none"> - 데이터 소유권 - 문화 - 기탁에 드는 시간과 노력 • 윤리 및 기밀 • 연구자 및 데이터학자를 위한 인지와 보상 • 재사용에 필요한 전문기술 	<ul style="list-style-type: none"> • 연구 재정지원의 투명성 • 교육 도구로 데이터세트 이용 • 풍부한 데이터 이용으로 연구자의 전문성 향상 • 데이터 질을 향상시킬 도구와 표준 • 기관과 연구자의 가시성/명망 증가

이 연구에서는 제시된 편익 중 계량화 가능한 편익을 중심으로 가상의 사례를 들어 계산식을 제안하고 있다(〈표 2〉 참조).

10년간 6회에 걸쳐 이용된 어떤 연구데이터의 사례에 의하면 데이터의 수집과 관리에 소요되는 10년 치의 비용은 데이터 수집/생산비용(C1)과 레퍼지토리의 이용 지원 준비 비용(C2), 이용자들이 6번에 걸쳐 이용시 소요 비용(C3 * N), 10년간 데이터 저장 비용(C4 * L)을 더한 값(DC)이다. 한 번의 데이터 수집/생산비용을 지출함으로써 인해 6번의 이용이 이루어진 것이므로, 이를 달리 말하면 6번을 더 수집/생산했을 수도 있는 비용(C1 * 6 = DB)을 절감한 셈이다. 따라서 직접적 편익율은 편익을 직접 비용으로 나눈 3.7(DBCR)이다. 즉, 이 연구데이터에 대해서는 소요 비용의 3.7배에

해당하는 이득을 취하게 되었다.

한편, 이 연구데이터를 수집하여 관리하지 않았더라면 더 지출했어야 할 비용을 다른 연구 지원에 지출(DB - DC = ARD)할 수 있으며, R&D 지원의 기본 회수율을 20퍼센트¹⁾로 보았을 때 추가 연구지원(ARD)에서의 투자회수 비용(ARD * 0.2 = AR)도 간접 이득으로 발생한다. 그러므로 이렇게 발생한 간접 편익(AR)을 직접 편익(DB)에 더해 총 편익 가치(DB + AR = TB)를 산출할 수 있다. 결국 총 편익의 면에서 이 연구데이터는 소요 비용에 비해 무려 4.2배의 이득을 취했다는 결과를 도출할 수 있다.

물론 직·간접 비용 절감의 효과 이외에도 연구데이터 보존과 공유를 통해 얻을 수 있는 다수의 잠재적 혜택이 존재한다. 연구자·학

〈표 2〉 연구 비용 절감 사례(Fry et al. 2008; 재구성)

데이터 이용/재사용에서 오는 직접 비용 절감		가치*
데이터 조건 :		
데이터 수집/생산 비용(기탁자)	C1	£200,000
공유를 위한 추가 준비 비용(기탁자/레퍼지토리)	C2	£10,000
데이터 검색/접근 비용(이용자)	C3	£2,500
대상 데이터의 1년간 저장 비용(레퍼지토리)	C4	£10,000
데이터의 존속 년 수	L	10
데이터 생명주기동안 이용/재사용된 회수	N	6
직접 비용 절감 :		
집적 비용 = C1 + C2 + (C3 * N) + (C4 * L)	DC	£325,000
직접 편익 = C1 * N	DB	£1,200,000
직접 비용-편익 비율 = (C1 * N) / (C1 + C2 + (C3 * N) + (C4 * L))	DBCR	3.7
간접 비용 절감 :		
타 R&D에 효과적인 추가지원 = DB - DC	ARD	£875,000
지원으로부터 R&D로 추가 회수(기본 비율 20%) = ARD * 0.20	AR	£175,000
총 편익 = DB + AR	TB	£1,375,000
총 비용-편익 비율 = TB / DC	TBCR	4.2

* 해당 연구의 추정치임.

1) 30여 년간의 연구를 분석한 Martin·Tang(2007)에 따르면 공공 연구 개발 지원의 투자 회수율은 20%에서 50%인 것으로 드러났다. Fry 외(2008)는 이 중 가장 보수적인 수치를 적용하여 투자 회수율을 20%로 가정하였다.

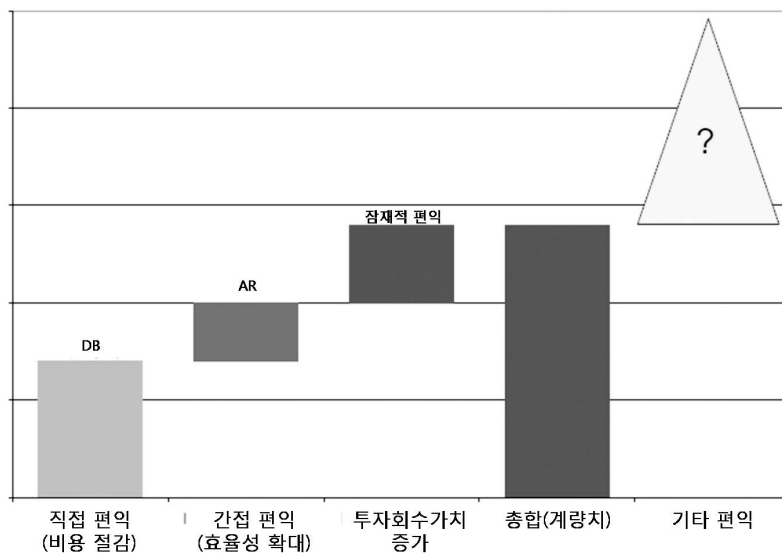
계·국가 간 협력으로 향상된 결과물을 산출할 수 있으며, 더 나은 교육·연구 훈련이 이루어질 수 있음은 물론, 신규 연구기회와 신규 이용자를 창출할 수 있고, 더욱 투명한 과학 기록을 생산하고 관리할 수 있게 되며, 연구·연구자의 가시성과 명망 및 연구에 따른 보상을 향상시킬 수 있다. 그러나 이러한 잠재적 혜택은 장기간에 걸쳐 연구의 질과 효율에 영향을 미치는 효과로, 계량화하기 힘든 단점이 있다(Fry et al. 2008).

이상의 직·간접적 비용 편익과 잠재적 편익의 추정치를 상대적으로 표현하면 다음과 같다. 단, 계량화하기 힘든 잠재적 혜택, 즉 기타 혜택은 그래프에서는 삼각형으로 표시하여 계량 가능한 여타의 편익과 구분하였다(그림 4 참조).

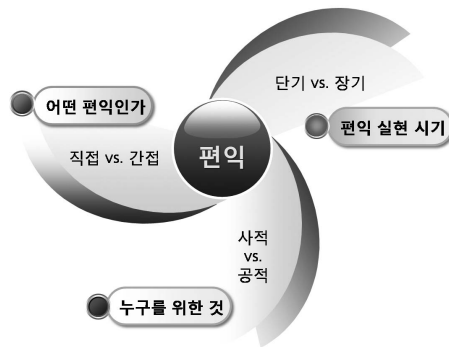
3.4 세 가지 측면에서 바라본 편익

유럽의 연구데이터 관리·보존 활동 기관을

대상으로 지속적으로 디지털 연구데이터의 보존비용과 편익에 대해 연구하고 있는 Beagrie, Lavoie와 Woollard는 2010년 JISC의 지원으로 디지털 연구데이터에 대한 보존비용 모델 수립 연구의 일부분으로 편익 분석을 수행하였다. 이 연구는 영국의 Data Archive(UKDA)와 National Crystallography Service(NSC), Oxford 대학 대상의 사례연구를 바탕으로 디지털 연구데이터 보존으로 인한 혜택을 확인하였다. 그 결과 디지털 연구데이터에 대한 편익을 3가지 관점에서 바라보며 구분하였는데, 어떠한 편익인가에 따라 직접적 편익과 간접적 편익으로, 편익이 실현되는 시기가 언제인가에 따라 단기적 편익과 장기적 편익으로, 실현되는 편익은 누구를 위한 것인가에 따라 사적 편익과 공적 편익으로 나누어 분석하였다(그림 5 참조).



〈그림 4〉 가상의 사례를 바탕으로 한 편익(Fry et al. 2008; 재구성)



〈그림 5〉 디지털 연구데이터 보존의 편익(Beagrie, Lavoie, & Woolard 2010; 재구성)

3.4.1 직접적 편익과 간접적 편익

디지털 연구데이터 보존이 가져오는 직접적인 혜택은 '디지털 자료에 대한 지속적인 접근을 유지함으로 인해 창출되는 가치에 대한 긍정적인 평가'와 같이 대부분 학자들이 '디지털 보존'을 떠올릴 때 생각하는 가치다. 예를 들면 현재의 연구를 통해 생성된 지식을 미래 세대의 학자들에게 전달한다거나, 연구의 생산성을 향상시키는 것, 생산된 주요 연구 데이터세트를 중심으로 학제간 연구가 가능하다는 것, 나아가 집적된 과학 지식의 상용화 가능성이 증가할 수 있다는 것 등이 이에 해당한다. 다시 말하면, 디지털 연구 데이터의 보존을 통해 창출되는 직접적인 결과(outcome)에 초점이 맞춰진 편익이 직접적 편익이다.

이에 반해 간접적 편익은 디지털 연구데이터 보존으로 인해 잠재적으로 발생할 수 있는 또 다른 형태의 혜택이다. 특히 미래에 발생할 것으로 예상되는 비용의 감소나 연구 투자비용의 보호 측면에서 주로 부각된다. 특정 연구데이터를 보존하는데 소요되는 현재의 투자비용보다 미래 데이터가 사라진 후 이를 재생산하는 데 지출될 것으로 예상되는 비용이 엄청나게 클

때 이러한 비용 억제 효과가 나타날 것이다. 비록 연구데이터를 완전히 분실하거나 소실할 가능성이 극히 적다고 하더라도 디지털 생명주기 초반에 보존활동을 수행하는 것이 미래 일회적으로 보존활동을 수행하는 것보다 비용 지출이 적다는 사실을 부인할 수 없다(〈표 3〉 참조).

3.4.2 단기적 편익과 장기적 편익

언제 혜택이 발생하는가의 기준에서 바라보았을 때에는 단기적 편익과 장기적 편익으로 구분할 수 있다. 근본적으로 '디지털 보존'을 통해 발생하는 편익은 장기적 편익을 가정한 것으로, 현재 이루어지는 투자로 인해 미래 세대가 투자이익을 취할 것이라는 것을 전제로 한다. 그러나 UKDA를 대상으로 한 사례 연구에 따르면 단기적 이용을 제공하는데 소요되는 비용이 실상은 장기 보존이나 이용 준비비용과 상당부분 중복되며, 따라서 지속적인 이용의 측면에서 본다면 장기 보존에 할당하는 비용도 사실 단기적 이용 제공에까지 혜택이 확대된다는 사실이 드러났다. 결국 단기적 편익과 장기적 편익을 완전하게 구분하는 것은 불가능하며, 이 둘은 본질적으로는 연결된 혜택이다.

〈표 3〉 직접적 편익 대 간접적 편익(Beagrie, Lavoie, & Woollard 2010)

직접 편익	간접 편익
<ul style="list-style-type: none"> • 새로운 연구 기회 • 학자간 커뮤니케이션/데이터 접근 • 데이터 재활용 • 연구 생산성 증가 • 신규 네트워크/협력 조장 • 산업계로의 지식 전달 • 기술 기반 • 상품 생산성 증가/시장 확대 • 연구 및 연구 무결성 검증 • 규제 준수 	<ul style="list-style-type: none"> • (동일)데이터 재생산 방지 • 미래 연구 기회 상실 방지 • 낮은 미래 보존비용 • 신규 독자(이용자)의 목적에 맞게 데이터 재사용 • 방법론 재활용 • 신규 독자의 이용 • 초기 투자에 대한 보상 보호

물론 장기적 편익과 뚜렷하게 구분되는 단기적 편익도 분명히 존재한다. 잘 보존된 연구데이터는 현 세대의 연구자나 학생에게 가치 있는 자료로 이용되기 때문이다. Oxford 대학을 대상으로 한 사례연구에 따르면, 현재의 연구 환경이나 기관 인프라 내에서는 연구데이터를 효율적으로 보존할 수 없는 것으로 드러났다. 박사 후 연구자들의 잦은 이직으로 인해 연구 과정에서 생산되는 데이터 손실의 문제가 발생하지만 이에 대응할 수 있는 기관 내 구조나 제도는 미비한 상태이다. 최초 데이터 생산에 대한 이해를 도울 충분한 맥락정보를 보유하지 못해 몇 년 만 흘러도 데이터를 이해하기 힘들게 된다. 전자망원경 등의 전자장비에서는 다량의 데이터를 생산하고 있지만, 이를 안전하게 저장할 수 있는 저장소는 충분하지 않다. 일

부 임상연구센터에서 진행하는 연구의 경우는 수 십 년에 걸쳐 데이터를 생산하며, 따라서 노화방지를 위해 수개월에 걸쳐 데이터 포맷 마이그레이션을 실시해야 하는 경우도 허다하다. 대다수 연구자는 연구결과를 발표하는 학술기사의 첨부 데이터도 온라인에서 이용되길 바라지만 기관의 인프라는 이를 뒷받침할 수 있는 수준에 미치지 못하고 있다. 바꾸어 말하면 디지털 연구데이터 보존에 대한 투자가 이루어져 원활한 보존 활동이 수행된다면 이러한 문제를 해결할 수 있는 시작점이 된다(〈표 4〉 참조).

3.4.3 사적 편익과 공적 편익

디지털 연구데이터 보존으로 인해 발생하는 혜택이 누구에게 실현되는가의 관점에서 보면 보존으로 인해 발생하는 편익을 사적 편익과

〈표 4〉 단기적 편익 대 장기적 편익(Beagrie, Lavoie, & Woollard 2010)

단기 편익	장기 편익
<ul style="list-style-type: none"> • 현 연구자/학생의 가치 • 박사 후(Post Doc) 연구자들의 이직으로부터 데이터 유실 방지 • 잘 큐레이트된 데이터의 단기 재사용 • 데이터 집적 연구를 위한 안전한 저장소 • 미출간 저널기사 데이터 이용가능성 	<ul style="list-style-type: none"> • 미래 연구자/학생에게 안전한 가치 전달 • 장기간 컬렉션 및 양 증가에 따른 부가가치

공적 편익의 범주로 나눌 수 있다. 디지털 연구 데이터 보존에 투자하거나 보존을 지원하는 각각의 객체에게 직접적으로 누적되는 편익이 사적 편익이라면, 광범위하게 확산될 수 있는 혜택이 공적 편익이다.

UKDA가 정의한 직접적 이해관계자를 살펴보면, 우선 데이터 아카이브가 대행하여 보존하는 데이터의 직접적 이용자, 둘째, 연구 산출물의 미래 이용가능성이나 검증가능성을 기대하는 데이터 생산자나 주로 정부부처로서 비용을 지출 없이 공유 가능한 공적 재이용 가치를 지닌 데이터의 생산자, 마지막으로 기금을 지원한 연구에서 산출된 데이터가 학계에서 활발하게 재사용되길 원하는 투자자 등으로 나뉜다 (Beagrie, Lavoie, & Woollard 2010). 즉, 일정한 연구를 통해 생산된 연구데이터가 어떠한 방식으로든 재이용되길 바라는 연구 수행자, 연구데이터 생산자, 그리고 연구 기금 지원 기관 등에게 디지털 보존을 통해 실현되는 혜택이 돌아가는 경우, 이를 사적 편익으로 간주할 수 있다. 예를 들면, *Australian Code for the Responsible Conduct of Research* (Australian Government 2007)의 '최종 연구결과물 출판 후 5년간 연구데이터를 보관해야 한다'는 조항을 준수할 수 있고, 안전한 저장소에 데이터를 기탁해야 한다는 조항이 기재된 연구보조금 지원 계약을 충족시킬 수 있다거나, 연구과정에서 산출된 연구데이터를 외부 연구자들이 이용함으로써 해당 연구의 가시성과 영향력이 증대되어 해당 연구를 수행한 연구자와 모기관의 명망과 명예가 향상되는 경우 등이 이에 해당한다.

반면, 타 기관의 연구자에게 미래 연구 동기를 부여한다거나 연구데이터가 미래 연구의 투

입 자료로서 역할을 수행하는 경우도 발생한다. 타 기관의 연구자뿐만 아니라 디지털 연구데이터 보존 계획시 "의도하지 않았던 결과(unintended consequences)"에 따라 학계를 넘어선 동반 이익이 발생하거나 일반 대중까지도 연구데이터셋을 통해 혜택을 받을 수 있는데, 예를 들면 학습용 교육교재로 재가공해 사용하는 경우 등이 그것이다. 나아가 과거 연구를 검증한다거나 신규 연구를 조장하는 등 현재의 연구로 축적된 지식을 시공간을 초월하여 전달할 수 있다는 등의 혜택을 공적 편익으로 볼 수 있다.

이렇게 구분해 보면 사적 편익에 비해 공적 편익은 공공의 이익만을 반영하는 것으로 보인다. 그럼에도 공적 편익이 대변하는 광범위한 혜택 안에는 투자 회수 가치도 여전히 존재한다. 연구수행을 통해 산출된 과학적 발견을 상업화할 수 있는 가능성이 향상된다거나, 나아가 신규 회사 창업을 촉진할 수 있고, 전문적 기술을 보유한 직원을 고용할 기회가 확대된다는 점 등이 그 예이다(〈표 5〉 참조).

이상의 연구를 살펴본 결과 디지털 연구데이터를 보존하는 주체가 다양해지면서, 과거 권위 있는 집중 레퍼지토리나 아카이브 외에도 연구데이터 생산기관 등 다양한 조직에까지 디지털 연구데이터의 보존 책임이 확산되고 있다는 사실을 확인할 수 있었다. 디지털 연구데이터 보존이 조직의 1차적 사명이 아닌 경우에는 보존 활동에 대한 더 명확한 투자 근거가 필요하다. 기존의 연구 성과를 살펴보면, 체계적인 디지털 연구데이터의 보존을 통해 얻을 수 있는 다양한 편익을 구체적으로 확인하고 비용 절감과 효율성 확대의 측면에서 이를 객관적으로 계량화하기 위한 노력이 유럽을 중심으로

〈표 5〉 사적 편익 대 공적 편익(Beagrie, Lavoie, & Woollard 2010)

사적 편익	공적 편익
<ul style="list-style-type: none"> • 연구 및 아카이브 스폰서/재정지원가에게 이득 • 연구자에게 이득 • 기금 의무사항 충족 • 연구의 상업성 	<ul style="list-style-type: none"> • 미래 연구를 위한 투입 • 신규 연구동기부여 • 신규 기업 및 고기술 보유 직원 활성화

〈표 6〉 양적 편익과 질적 편익

계량 가능	질적 평가 가능
<ul style="list-style-type: none"> • 연구데이터 중복 생산 방지 • 연구데이터에 대한 접근/이용 확대 • 연구데이터로 인해 발생하는 각종 수입 증가 (소유권, 수수료 등) • 연구데이터의 상업화 가능성 증가 • 미래 보존비용 감소 • 기존 데이터를 이용한 신규 연구 • 학계에서의 신규 협력 증가 • 영향력 인지 시간 감소 • 연구 속도 증가 • 신규 지식 기반 산업 창출 및 시장 확대 • 상품 생산성 증가 	<ul style="list-style-type: none"> • 연구 및 연구의 무결성 검증 • 연구 재정지원의 투명성 증가 • 안전한 저장으로 데이터 집적 연구의 효율성 향상 • 미래 연구자/학생에게 가치 전달 • 미래 신규 연구 동기 부여 • 미래 부가가치 증가 • 학습/교육에의 활용 • 풍부한 연구데이터 이용으로 연구자 전문성 향상 • 데이터 질을 향상시킬 도구와 표준 개발 • 기관·연구자 가시성/명망 증가 • 직원의 전문성 향상 • 규제 준수

활발하게 진행되고 있음을 알 수 있었다. 또한 비용의 가치로 표현할 수 없는 무형의 편익도 객관적으로 측정하고 표현하고자 하는 시도가 일부 이루어지고 있었다. 앞서 정리한 다양한 연구데이터 장기보존의 편익을 비용의 측면에서 계량 가능한 것과 질적 평가만이 가능한 것으로 구분하여 다시 정리해보면 〈표 6〉과 같다.

종합해 보면, 디지털 연구데이터 보존에 소요되는 비용을 산출하여 데이터 이용이나 연구 확산 등 비용 면에서 산출할 수 있는 편익을 확인하고자 하는 노력이 행해지고 있는 한편, 정확한 비용으로 산출할 수 없는 질적 측면의 편익까지 확인하고 이를 객관적으로 제시하고자 하는 연구도 진행되고 있는 것으로 드러난다. 따라서 비용의 측면에서 계량화하여 드러낼 수

있는 편익의 경우와 질적 평가만이 가능하여 별도의 평가 지표를 개발해야 하는 혜택으로 나누어 디지털 연구데이터 장기보존의 편익을 측정해야 할 필요가 있다.

4. 디지털 연구데이터 편익 규명을 위한 제안

다양한 연구에서 디지털 연구데이터 보존의 편익을 다각도로 확인한 결과 디지털 연구데이터를 장기보존함으로써 얻을 수 있는 혜택은 크게 비용의 측면에서 수치화하여 드러낼 수 있는 편익과 그렇지 않은 질적 편익으로 크게 구분할 수 있었다. 따라서 이러한 편익을 객관

적으로 드러내기 위한 도구 역시 비용 산정을 동반하는 편익 산출 방법과 질적 편익 평가 방법을 함께 사용하여 개발해야 할 것이다.

4.1 편익 확인 및 산출을 위한 객관적인 도구 개발

디지털 보존의 궁극적인 목표가 미래 세대의 공공 이용임을 고려할 때, 디지털 연구데이터의 장기보존을 통해 발생하는 유형의(tangible) 혜택을 드러내는 일은 힘든 작업이다. 그런데 동일 연구데이터의 중복 생산을 방지한다거나 중복 생산으로 인한 관리·보존비용의 절감, 타 연구 지원 기회 증가 등 계량화가 가능한 혜택도 분명히 존재한다. 만약 풍부한 사례 연구를 통해 특정 연구데이터 세트의 생산·수집·관리·보존에 투입된 비용을 산출할 수 있다

면, 앞선 연구에서와 같이 해당 연구데이터 세트의 총체적인 보존에서 오는 편익을 계량화할 수도 있다. 특정 연구데이터에 대해 생산 비용, 아카이브나 관리기관으로의 입수 비용, 관리·보존비용, 이용 지원 비용을 전체 보존비용으로 본다면 그 편익을 산출하는 모형(안)을 다음과 같이 제시할 수 있다(〈그림 6〉 참조).

한편, 디지털 연구데이터의 보존을 통해 얻을 수 있는 혜택 중에는 수치화하여 표현할 수 없는 무형의(intangible) 혜택이 더 많다. 그러므로 질적 방식으로 그 혜택을 표현하여 의사 결정 지원 정보로 활용함은 물론 보존 활동에 투자하는 주체가 디지털 연구데이터 보존에서 오는 혜택을 충분히 이해할 수 있도록 해야 한다. 디지털 연구데이터 보존에서 오는 혜택을 평가할 수 있는 템플릿을 작성하고 그 정도를 제시할 수 있다면 최대한 객관적으로 질적 편익을

<p>특정 연구데이터 세트의 직·간접비용 절감치(총 편익) $TB_t = DCS_t + ICS_t$ 직접비용 절감치 $DCS_t = (C + I_n)_t \times n_t$ 간접 효용치 $ICS_t = AE \times r$ 추가 지출 가능치 $AE = DSC_t - TC_t$</p> <hr style="border-top: 1px dashed black;"/> <p>총 보존비용 $TC_t = C + I_n + P_t + A_n$</p> <p>TB : 특정 연구데이터 세트의 직·간접비용 절감치(총 편익) Total Benefit DCS : 직접비용 절감치 Direct Cost Savings ICS : 간접 효용치 Indirect Cost Savings AE : 추가 지출 가능치 Additional Expenditure r : 투자회수율 Return rate n : 이용 회수 t : 시간 TC : 총 보존비용 Total Cost C : 생산 Creation I : 입수 Ingest P : 관리·보존 Preservation A : 이용지원 Access</p>

〈그림 6〉 비용 - 편익 산출 모형(안)

드러낼 수 있을 것으로 예상된다. 이에 앞서 살펴본 espida의 산출물 득점표를 기반으로 아래와 같이 연구데이터 보존의 질적 편익 평가 템플릿 개발을 제안하였다(〈표 7〉 참조). 기존의 영역에 대해서는 연구데이터 편익의 특성을 반영하여 ‘이용자 및 외부 이해관계자’ 영역과 ‘내부 업무 프로세스’ 영역의 명칭을 각각 ‘학계 및 사회·문화적 영향’과 ‘연구 프로세스’로 변경하였다. 각각의 기여 항목 역시 디지털 연구데이터 장기보존을 통해 연구 프로세스 개선은 물론 연구 산출물, 나아가 공공에 기여할 수 있는 혜택의 측면을 반영하여 다음과 같이 재구성하였다.

이 평가표는 디지털 연구데이터의 장기보존을 통해 일반적으로 얻을 수 있는 혜택에 대한 평가 항목으로 구성된 것이다. 따라서 무형의 편익을 확인하고자 하는 조직에서는 해당 조직에서 디지털 연구데이터를 보존함으로써 얻을 수 있는 각각의 전략적 측면과 집중 평가 요소를 확인하고, 평가 템플릿을 수정하여 적용해야 할 것이다.

더불어 이 절에서 제안한 비용·편익 산출모형(안)과 질적 편익 평가 템플릿(안)은 실제 적용에 앞서 검증과정을 거쳐야 한다. 이를 위해서는 다음 절에서 제안하는 바와 같이 연구데이터를 생산하고 관리하는 연구기관이나 데이터 레파지토리 등을 대상으로 디지털 연구데이터 생산에서부터 관리를 아우르는 총체적인 보존비용에 대한 연구는 물론, 디지털 연구데이터의 생애 주기에 연루되는 다양한 이해관계자를 대상으로 하는 질적 연구가 이행되어야 할 것이다.

4.2 편익 확인 및 산출을 위한 다양한 연구

디지털 연구데이터 보존에서 오는 유·무형

의 편익을 확인하고 객관적으로 이를 산출하기 위한 도구 개발을 위해서는 다양한 연구가 필요하다. 우선, 디지털 연구데이터의 보존비용에 대한 연구가 시급하다. 앞서 강조한 바와 같이, 디지털 연구데이터의 보존에 필수적인 투자를 위해서는 최대한 미래에 소요될 비용과 자원에 대한 예상치가 필요한데, 이를 위해 선행되어야 하는 것이 디지털 연구데이터의 생명주기 동안 소요되는 비용에 대한 연구이다. 디지털 연구데이터의 생산, 심지어 연구데이터를 산출한 연구 활동 단계에서부터 시작되는 관리·보존의 국면을 확인하고, 이를 요소화하여 비용을 산출할 수 있다면 더 객관적이고 정확한 비용 대비 편익 분석을 수행할 수 있는 바탕이 된다.

둘째는 풍부한 사례 연구이다. 관측 데이터, 실험 데이터, 사회조사 데이터 등 다양한 유형의 디지털 연구데이터를 생산·활용하는 연구 분야가 증가하고 있으며, 다양한 위상의 보존 책임을 지는 조직에서 디지털 연구데이터를 관리하고 보존하고 있다. 조직의 1차적 설립 목적이 디지털 연구데이터의 보존이나 디지털 보존인 경우는 비용이나 편익과 무관하게 조직의 사명을 수행하기 위해 자원을 투입할 것이다. 그러나 디지털 연구데이터의 보존이 부차적인 목표이거나 보존과는 전혀 무관한 조직일 경우라면, 디지털 연구데이터 보존에 자원을 투입할 적절한 근거가 필요하게 된다. 따라서 연구데이터 중앙관리기관의 위상을 지닌 조직 외에도 다양한 조직을 대상으로 한 풍부한 비용 연구와 비용-편익 연구 및 질적 편익 연구를 수행한다면 조직의 상위 의사결정자가 충분히 이해할 수 있는 정교하고 객관적인 디지털 연구데이터의 편익을 도출해낼 수 있을 것이다.

〈표 7〉 질적 편의 평가 템플릿(안)

학계 및 사회·문화적 영향								
	기여 항목	결과	결과 지표	범주	실현 가능성	+/-	실현 시기	결과 지속 기간
E1	기관 및 국가·국제적 차원에서의 연구데이터 공개·이용	데이터 공개 및 이용 확대	이용회 수증가	1차	상	+	10년 이하	장기
E2	연구자 및 기관 인지도 및 신뢰도 향상							
E3	안정적인 데이터 이용·활용 체계 구축							
E4	학문적 발전 촉진 - 미래 연구자·학생에게 가치 전달 - 학계 신규 협력 증가 - 신규 연구 및 연구투자 증가							
E5	학문적 성과의 상업화 촉진 - 상업화 가능성 증가 - 신규 시장 확대 및 지식 기반 산업 창출							
	평가자 요약							
연구 프로세스								
P1	연구데이터 접근가능성							
P2	연구데이터의 안전한 저장·집적을 통한 연구 효율, 속도 및 생산성							
P3	규제 준수 및 재정 지원 투명성							
	평가자 요약							
혁신 / 발전								
I1	연구데이터 및 데이터 생산·관리의 질 향상을 위한 도구 표준 개발							
I2	미래 연구 동기부여, 연구자 및 직원 전문성 향상							
I3	연구·산출물의 질, 잠재력 및 부가가치 증가							
I4	데이터세트의 학습 및 교육 활용							
I5	변화 민감성(변화관리능력/유연성)							
	평가자 요약							
재 무								
	수익 발생							
F1	연구데이터 판매							
F2	연구데이터 라이선스/권한							
F3	교육/연구							
F4	계약, 보조금, 수수료, 기부금							
	비용 절감							
F5	연구시간							
F6	공간 및 저장소							
F7	직접 지출							
	평가자 요약							
	총합계							

주 : E1의 기술내용은 템플릿 작성의 이해를 돕기 위한 본 연구자의 가정치 임.

셋째, 다양한 비용 연구와 비용 편익 및 질적 편익에 대한 연구를 통해 서로 다른 조직에서 디지털 연구데이터를 보존함으로써 얻을 수 있는 광범위하고 포괄적인 편익 풀을 개발할 수 있다. 이는 범용 비용 - 편익 산정 모델 및 질적 편익 평가 템플릿 개발의 기반이 된다. 다양한 사례 연구를 통해 개발된 범용 편익 풀에서 조직의 사명과 설립·운영 목표에 따라 편익을 선택한다면 디지털 연구데이터 보존에 대한 비즈니스 모델을 수립할 수 있으며, 디지털 연구데이터의 보존 목표까지 확립시킬 수 있는 훌륭한 도구가 될 것이다. 나아가 범용 비용 - 편익 산정 모델 및 질적 편익 평가 템플릿을 좀 더 용이하게 조직의 목표와 정책에 맞도록 수정할 수 있게 되어 현실적인 편익 도출을 지원할 것이다.

5. 결론 및 제언

정보기술과 네트워크 기술의 급속한 발달은 연구과정에서 생산되는 디지털 연구데이터 양의 폭발적 증가로 이어지고, 이는 다시 생산된 연구데이터의 이용가능성이나 상호운용가능성에 대한 기대의 증가로 이어지고 있다. 그러나 이러한 기대와는 달리 연구의 최종 결과물에 비해 연구과정에서 산출된 디지털 연구데이터의 체계적인 관리나 보존에 대한 고려는 상대적으로 미약했던 것이 사실이다. 이에 미국이나 유럽 등의 선진국은 연구데이터의 관리와 활용 보존의 중요성을 인식하고 점차 이를 위한 연구와 투자를 증가하고 있다. 그 연구 내용도 마이그레이션이나 에뮬레이션 등의 보존 방

식 개발에 집중해 기술적인 해결책을 찾는 쪽에서 점차 보존 대상의 가치를 탐구하고 지속 가능한 보존활동의 당위성을 찾는 부분까지 연구 분야를 확대하고 있다.

실험실에서, 사회조사 기관에서, 여러 전자 장비에서 매일같이 생산되는 디지털 연구데이터는 그 양이 방대하고, 분실했을 경우 재생산 비용이 상당하거나 재생산이 불가능한 경우가 대부분이다. 보통은 다른 자료로 대체할 수 없다. 따라서 이러한 디지털 연구데이터를 대상으로 적절하고 지속적인 투자가 이루어져야만 미래 이용을 보장할 수 있다. 그러나 정작 투자를 결정하는 주체 기관이나 기관의 상위 의사결정자의 경우는 이러한 디지털 연구데이터 보존에 대한 인식이나 보존 편익에 대한 인지가 부족한 것이 현실이다.

이에 본 연구에서는 그간 이루어진 디지털 연구데이터 보존의 편익 분석에 대한 연구를 살펴봄으로써 다양한 각도에서 디지털 보존의 편익을 이해하고 미약하나마 이를 객관적으로 확인하고 산출할 수 있는 방법을 제안해보고자 하였다. 살펴본 결과, 여전히 편익을 양적으로 계량화하는 작업은 힘들고, 심지어 불가능한 경우도 존재하는 것으로 드러났다. 다만, 미미하나마 질적인 방식으로 디지털 연구데이터 보존의 편익을 표현할 수 있다면 연구데이터 보존에 대한 연구 기금 지원 기관이나 의사결정자의 인지를 향상시키는데 도움을 줄 수 있을 것이다. 나아가 디지털 연구데이터 보존을 통해 획득할 수 있는 다양한 편익의 특성을 더 깊게 이해시킬 수 있을 것이다.

이제 디지털 연구데이터 보존 활동을 수행하는 국내 기관을 대상으로 다양한 사례연구를

수행함으로써 현실에 맞는 편익 개발이 필요한 시점이다. 풍부한 사례연구를 통해 객관적으로 디지털 연구데이터 보존의 혜택을 확인하고,

정교한 도구를 개발하여 올바르고 효율적인 국내의 디지털 연구데이터 보존 활동의 기초를 다져야 할 것이다.

참 고 문 헌

- 과학기술부. 2009. 『국가연구개발사업 연구노트 관리 지침』. [교육과학기술부훈령 제128호].
- 김선태, 한선화, 이태영, 김용. 2010. 과학데이터 보존 및 활용모델에 관한 연구. 『한국비블리아학회지』, 21(4): 81-93.
- 생물학연구정보센터(BRIC). 2009. 『연구실 실험데이터 관리 실태조사』. 포항: SciON.
- 서울대학교. 2010. 『서울대학교 연구윤리 지침』. 서울: 서울대학교.
- Australian Government. 2007. Australian Code for the Responsible Conduct of Research. [cited 2011. 5. 2].
<http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/r39.pdf>.
- Beagrie, Neil, Julia Chruszcz, & Brian Lavoie. 2008. Keeping research data safe: a cost model and guidance for UK universities. [cited 2011. 5. 3].
<<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>>.
- Beagrie, Neil, Brian Lavoie and Mathew Woollard. 2010. Keeping research data safe 2. [cited 2011. 4. 8].
<[ments/publications/reports/2010/keepingresearchdatasafe2.pdf](http://www.jisc.ac.uk/media/docu)>.
- Currall, James and Peter McKinney. 2006. Investing in value: a perspective in digital preservation. D-Lib Magazine, 12(4). [cited 2011. 5. 9].
<<http://www.dlib.org/dlib/april06/mckinney/04mckinney.html>>.
- _____. 2007. Expressing project costs and benefits in a systematic way for investment in information and IT: espida Handbook. [cited 2011. 3. 21].
<https://dspace.gla.ac.uk/bitstream/1905/691/1/espida_handbook_web.pdf>.
- Key Perspectives Ltd. 2010. Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study. Edinburgh: Digital Curation Centre. [cited 2011. 4. 11].
<<http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP-Synthesis.pdf>>.
- Fry, Jenny, Suzanne Lockyer, Charles Oppen-

- heim, John Houghton and Bruce Rasmussen. 2008. Identifying benefits arising from the curation and open sharing of research data produced within UK Higher Education and research institutes: exploring costs and benefits. JISC. [cited 2011. 5. 6]. [〈http://ie-repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf〉](http://ie-repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf).
- Koski, Kimmo, Claudio Gheller, Stefan Heinzl, Alison Kennedy, Achim Streit and Peter Wittenburg. 2009. Strategy for a European Data Infrastructure. Espoo: PARADE. [cited 2011. 5. 4]. [〈http://www.csc.fi/english/pages/parade/whitepaper〉](http://www.csc.fi/english/pages/parade/whitepaper).
- Lavoie, Brian F. 2004. Of mice and memory: economically sustainable preservation for the twenty-first century. In *Access in the Future Tense*. Washington, D.C.: Council on Library & Information Resources. [cited 2011. 4. 21]. [〈http://www.clir.org/pubs/reports/pub126/lavoie.html〉](http://www.clir.org/pubs/reports/pub126/lavoie.html).
- Lavoie, Brian and Lorcan Dempsey. 2004. Thirteen ways of looking at...digital preservation. *D-Lib Magazine*, 10(7/8). [cited 2011. 4. 21]. [〈http://www.dlib.org/dlib/july04/lavoie/07lavoie.html〉](http://www.dlib.org/dlib/july04/lavoie/07lavoie.html).
- Lyon, Liz. 2007. *Dealing with Data: Roles, Rights, Responsibilities and Relationships*. Consultancy Report. UKOLN. [cited 2011. 5. 2]. [〈http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf〉](http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf).
- National Academy of Science. 2009. Executive Summary, In *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington D.C.: The National Academies Press. [cited 2011. 4. 9]. [〈http://www.nap.edu/html/12615/12615_EXS.pdf〉](http://www.nap.edu/html/12615/12615_EXS.pdf).