

# 데이터 사이언스와 데이터 리터러시

## Data science and data literacy

이재윤, 명지대학교 문헌정보학과, memexlee@mju.ac.kr

Jae Yun Lee, Dept. of Library and Information Science, Myongji University

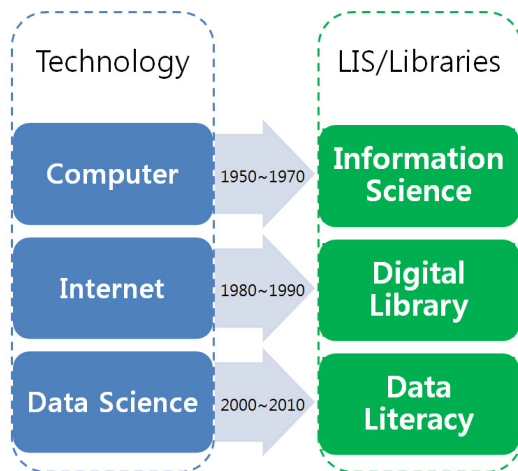
데이터 사이언스가 경영, 학술, 교육 등의 전 영역에 확산되면서 변화를 이끌어내고 있다. 문헌정보학과 도서관의 입장에서 데이터 사이언스라는 대세를 수용하는 전략에 대해서 데이터 리터러시를 중심으로 살펴보았다. 연구 데이터 리터러시, 통계 데이터 리터러시, 실무 데이터 리터러시의 세 가지로 나누어 살펴본 결과, 데이터 리터러시는 데이터와 관련된 기술과 역량을 강조하는 용어로서 문헌정보학과 도서관 현장에서 데이터 사이언스를 주도적인 입장에서 수용하는 브랜드로 육성시킬 필요가 있다고 판단된다.

### 1. 서론

빅데이터가 대중적인 키워드로서 데이터의 중요성을 상징하는 용어로 대표되고 있다면, 학술적으로 데이터의 중요성을 부각시키는 현상은 데이터 사이언스의 대두이다. 엄밀히 따

져서 데이터 사이언스가 하나의 독립된 학문 인가라고 보는 시각에는 아직까지 이론의 여지가 있지만, e-사이언스와 마찬가지로 여러 학문 분야에 걸쳐 영향을 끼치고 있는 학술적 현상임에는 틀림이 없다. 주로 과학기술 분야 위주로 e-사이언스가 강조되었다면, 데이터 사이언스의 부상은 인문사회과학 분야와 여러 실무 현장에까지 영향력이 확산되고 있다는 점에서 실체가 뚜렷하며 더욱 주목되는 흐름이다. 이미 국내에서도 한국데이터사이언스학회가 창립되었고 여러 대학에서 데이터 사이언스 학과를 창설하였으며, 이 중에는 성균관대학교의 경우와 같은 문헌정보학과와의 참여 사례도 나타나고 있다.

전통적으로 ‘정보’를 화두로 삼았던 문헌정보학 분야에게는, 이와 같은 ‘데이터’와 데이터 사이언스의 대두는 새로운 도전이 되고 있다. 이 글에서는 데이터 사이언스라는 시대의 조류에 문헌정보학계가 어떻게 대처할 것인가를 모색해보고자 한다.



<그림 1> 정보기술이 문헌정보학에 미친 영향

## 2. 문헌정보학의 정보기술 수용 방식

데이터 사이언스 이전에 인류 사회와 학계에 대전환을 가져온 주요 정보기술로는 컴퓨터와 인터넷을 들 수 있다. 1950년대 컴퓨터 기술의 급속한 발전과 1980년대 인터넷의 성장은 이후 수 십 년 동안 사회와 학계에 급진적인 변화를 일으켰고, 문헌정보학 분야도 예외가 아니었다(<그림 1> 참조).

컴퓨터의 발전은 다양한 분야에 영향을 끼쳤으며, 구체적으로 컴퓨터 사이언스를 탄생시키기도 하였다. 이에 대해서 문헌정보학(당시는 도서관학)계는 정보학(information science)의 정립으로 대응하였다.

인터넷의 확산은 문헌정보학과 도서관 현장에 심대한 영향을 끼쳤으며, 디지털도서관이라는 새로운 브랜드의 연구 대상과 실무 현장을 낳게 되었다. 물론 디지털도서관이 그 이전의 전자도서관(electronic library)이나 자동화도서관(automated library)과 본질적으로 차별화되는 것은 아니었지만, 인터넷의 대중화라는 흐름을 수용하는 하나의 브랜드로서 가치를 발휘한 것은 사실이다.

이처럼 문헌정보학계는 정보기술의 시대적 흐름에 대해서 한 시대에는 ‘정보학’이라는 브랜드의 정립으로, 또 다른 시대에는 ‘디지털도서관’이라는 브랜드의 정립으로 대응해온 셈이다. 빅데이터와 데이터 사이언스라는 시대적 흐름에 대해서는 문헌정보학계가 어떤 브랜드로 대응해야 할 것인지에 대한 고민이 필요하다. 물론 데이터 사이언스라는 명칭을 그대로 수용하는 것도 한 가지 방법이지만, 다른 학문 분야에서 소유권을 주장하지 않고 문헌정보학의 영역으로 인정받을 수 있는 고유의 브랜드가 별도로 필요하다.

문헌정보학 분야에서 제기되는 브랜드 중에는 데이터 관리(data management), 데이터 큐레이션(data curation), 데이터 서비스(data service),

데이터 사서(data librarian) 등이 있으나, 이들이 데이터와 관련된 문헌정보학의 연구 영역과 실무 활동을 포괄적으로 포함하지는 못하고 있다. 이 글에서는 전통적인 문헌정보학의 영역과 부합되면서 데이터 사이언스를 수용할 수 있는 용어로 ‘데이터 리터러시’를 유력한 후보로 제시하고자 한다.

## 3. 데이터 리터러시

데이터 리터러시는 사서들에게 새로운 용어이지만, 그와 관련된 기술은 정보를 평가하고 큐레이팅하고 분석하고 관리하는 사서들에게 새롭지 않은 것이다. 다만 활용 가능한 데이터의 범위와 깊이가 크게 변화했기 때문에 데이터 리터러시라고 포괄할 수 있는 기술의 중요성이 최근에 크게 부각되었다(Wanner, 2015). 물론 아직까지 데이터 리터러시와 관련하여 합치된 정의나 기준이 제시되었다고 할 수는 없다. 이는 역사가 오래되고 표준화가 잘 되어 있는 정보 리터러시와 비교해보면 더욱 그러하다. Koltay(2015)는 데이터 리터러시가 정보 리터러시와 상당히 유사하지만 여타 리터러시와도 공유하는 요소가 다수 존재한다고 하였다.

데이터 리터러시와 관련된 기술을 가장 단순하게 표현하자면, “가공되지 않은 정보를 어떤 방식으로 처리할 수 있는 능력”(the ability to do something with raw information - to process it in some way)(Johnson, 2012)이라고 할 수 있다. Calzada Prado와 Marzal(2013)은 데이터 리터러시는 개인이 데이터를 획득/이해/평가/관리하고 조작하면서 윤리적으로 활용할 수 있는 능력이라고 정의하였다. 이에 비해 Wanner(2015)가 제시한 내용은, 여러 데이터 리터러시 관련 글에서 언급되는 내용을 느슨하게나마 포괄하고 있다고 생각된다. 이를 나열해보면 다음과 같다.

- 첫째, 데이터 이해 (Understanding data).
- 둘째, 데이터 소재 파악 및 획득  
(Finding and/or obtaining data).
- 셋째, 데이터 읽기, 해석, 평가 (Reading,  
interpreting and evaluating data).
- 넷째, 데이터 관리 (Managing data).
- 다섯째, 데이터 활용 (Using data).

데이터 리터러시의 개념에 대해서는 다양한 이견이 존재하고 있다. 위의 다섯 가지 기술 중에서 어디에 중점을 두는가에 따라서 각자가 주장하는 개념이 달라진다. 또한 데이터 리터러시의 주체가 누구인가에 대해서도 다른 시각이 존재한다.

관련된 용어로는 정보리터러시(information literacy)(The Association of College and Research Libraries, 2000)부터 시작해서 디지털 리터러시(Gilster, 1997), 과학 데이터 리터러시(science data literacy)(Qin & D'Ignazio, 2010), 연구 데이터 리터러시(research data literacy)(Schneider, 2013), 정보 데이터 리터러시(information data literacy)(Carlson et al., 2011), 통계 리터러시(statistical literacy)(Watson, 2006) 등이 있다. 이 글에서는 문헌 정보학의 입장에서 데이터 리터러시를 세 가지 측면으로 나누어 간단히 정리해본다.

### 3.1 연구 데이터 리터러시

문헌정보학 분야에서 지난 10여년 간 데이터 리터러시가 가장 활발하게 언급된 분야는 대학도서관과 연구도서관에서 이루어지는 연구 지원 서비스의 관점에서 다루어지는 데이터 리터러시이다. 이와 관련해서는 과학 데이터의 이해와 활용 및 관리를 위한 능력을 강조하는 과학 데이터 리터러시(science data literacy)라는 용어도 사용된다(Qin, & D'Ignazio, 2010). 과학 분야에서의 데이터 리터러시는 '과학 탐

구를 위해 데이터를 수집, 처리, 관리, 평가, 이용할 수 있는 지식과 기술'을 의미한다(Qin, J., & D'Ignazio, 2010).

데이터 정보 리터러시(Carlson & Johnson, 2015)라는 용어를 대신 사용하는 경우도 있다. 이는 데이터 리터러시 관련 활동에서 사서들이 정보 리터러시에 기반하여 핵심적인 역할을 수행해야 한다는 뜻에서, 정보 리터러시와 데이터 리터러시의 연결고리를 부각시키기 위해 채택한 용어이다(Whitmire & Carlson, 2015).

이와 관련해서는 도서관에서 연구데이터 관리 서비스를 강화해야 한다는 의미에서 연구데이터 관리나 연구데이터 큐레이션이라는 용어를 사용하기도 한다(Corti et al., 2014; Ray, 2014). Schneider(2013)는 "현재의 과학자들 뿐만 아니라 후속 세대에서는 거의 모든 학문 분야에서 연구 데이터를 다루게 될 것이다"라고 말하면서 정보 리터러시의 새로운 하위분야로 '연구 데이터 리터러시(research data literacy)'를 제시하였다.

### 3.2 통계 데이터 리터러시

전통적으로 통계교육과 관련하여 사용해온 개념인 통계 리터러시와 유사한 개념으로 데이터 리터러시를 사용하는 경우도 점차 증가하고 있다. 통계 리터러시는 통계/수치 데이터를 생산하거나 이해할 수 있는 능력을 의미한다. 이와 관련해서는 수량적 사고능력이라는 뜻의 numeracy나 quantitative literacy라는 용어가 사용되기도 한다. 이런 통계 능력을 배양하기 위한 교재들은 최근 데이터 리터러시라는 용어를 사용하고 있다(Bowen & Bartley, 2014; Herzog, 2014).

문헌정보학 분야에서도 통계 리터러시로 데이터 리터러시를 논하는 경우도 상당수이다. Stephenson와 Caravello(2007)는 사회과학 분

야의 학생들이 데이터를 다룰 수 있는 능력을 배양할 수 있도록 정보 리터러시 프로그램에 통계 리터러시라고도 불리는 데이터 리터러시 부분을 강화해야 한다고 주장하였다. 전통적인 정보 리터러시 영역에서도 데이터 리터러시를, “학생들에게 수치 데이터를 찾아서 다루고 해석하는 방법을 교육하는 과정”(Yu & Shrimplin, 2006, p.14)으로 정의하기도 한다. 마이애미 대학의 사서인 이들은 정보 활용법을 강의하는 교양 과목에서 2002년부터 데이터 리터러시에 해당하는 요소를 추가하기 시작하였다. 또한 툴레인 대학 사서인 Beauchamp(2015)도 같은 관점에서 데이터 리터러시를 “통계 정보를 해석하고 평가하고 소통할 수 있는 능력”이라고 규정하고 있다.

학생뿐만 아니라 시민 사회의 일원이 갖춰야 할 소양 중의 하나로 통계 리터러시를 강조하기도 한다. 따라서 이런 개념의 데이터 리터러시는 대학도서관이나 학교도서관 이외에 공공도서관에서도 주목해야 하는 영역이다.

### 3.3 실무 데이터 리터러시

현실적으로 데이터를 다루는 활동은 학문 활동에서 뿐만 아니라 다양한 분야의 실무에서 더욱 광범위하고 활발하게 이루어지고 있으며, 일상적인 업무의 일부분이 되어가고 있다. 빅데이터와 데이터 사이언스가 주목을 받는 이유도 학문보다는 비즈니스와 일상 생활에서 영향력이 급부상했기 때문이다. 이런 상황을 고려하면 ‘연구’ 데이터 리터러시 뿐만 아니라 ‘실무’ 데이터 리터러시라는 용어도 고려해야 할 것이다. 사실상 대부분의 근거중심 실무 영역에서는 실무를 위한 데이터의 확보와 활용을 강조하고 있으므로 실무 데이터 리터러시라는 용어가 낯설 뿐, 이미 실무 현장에서 활동이 이루어지고 있는 개념이다. 물론 이들을 굳이 구분하기 보다는 포괄적인 데이

터 리터러시라는 용어를 사용하는 것이 더 바람직하다.

이는 현장 실무 개선을 위한 연구를 의미하는 실행 연구(action research)에 대해서도 성립된다. 실행 연구 단계는 성찰(reflect), 문제 선정(select a focus), 데이터 수집(collect data), 데이터 분석 및 해석(analyze and interpret data), 실행(take action)로 구분된다(Glanz, 2014). 이와 관련하여 교육 분야에서는 교육 실무를 개선하기 위하여 데이터를 수집하고 활용하는 역량을 일컫는 표현으로 한동안 평가 리터러시(assessment literacy)라는 용어를 사용해 오다가, 최근에는 데이터 리터러시로 이를 대체하는 추세이다(Mandinach & Gummer, 2013).

데이터 리터러시가, 데이터에 기반하여 의사결정할 수 있는 능력(Starobin & Upah, 2014)으로 언급되기도 하므로, 정보관리 전문가의 실무를 위한 기본 소양으로서도 데이터 리터러시가 강조되어야 한다.

## 4. 결론

문헌정보학의 입장에서 데이터 사이언스를 수용하기 위한 브랜드로 데이터 리터러시의 개념과 수용 현황을 간단히 살펴보았다. 연구 현장에서 강조되는 연구 데이터 리터러시는 연구지원 서비스의 일환으로 이해할 수 있다. 학생 및 시민을 위한 통계 데이터 리터러시는 학습 능력과 시민의 기초 소양 향상을 위한 도서관 서비스의 일부로 중요성이 증대되고 있다. 실무 데이터 분석이 강조되는 실무 데이터 리터러시는 정보전문가로서의 실무 개선을 위해 지속적으로 증진시켜야 하는 역량에 해당한다. 이처럼 데이터 리터러시는 데이터와 관련된 기술과 역량을 강조하는 용어로서, 문헌정보학과 도서관 현장에서 데이터 사이언스를 주도적인 입장에서 수용하는 브랜드로 육성시킬 필요가 있다.

## 참고문헌

- Beauchamp, A. (2015). What is Data Literacy? Retrieved from <http://databrarians.org/2015/02/what-is-data-literacy/>
- Bowen, M., & Bartley, A. (2014). *The basics of data literacy: Helping your students (and you!) make sense of data*. NSTA Press.
- Calzada Prado, J., & Marzal, M. A. (2013). Incorporating data literacy into information literacy programs: core competencies and contents. *Libri*, 63(2), 123-134.
- Carlson, J., & Johnston, L. R. (Eds.) (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press.
- Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining data information literacy needs: a study of students and research faculty. *Portal: Libraries and the Academy*, 11(2), 629-657.
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: A guide to good practice*. SAGE Publications Ltd.
- Gilster, P. (1997). *Digital literacy*. New York: Wiley.
- Glanz, J. (2014). *Action research: An educational leader's guide to school improvement*. Third edition. Lanham: Rowman & Littlefield.
- Herzog, D. (2014). *Data literacy: A user's guide*. SAGE Publications.
- Johnson, C. A. (2012). *The information diet: A case for conscious consumption*. O'Reilly Media.
- Koltay, T. (2015) Data literacy: In search of a name and identity. *Journal of Documentation*, 71(2), 401-415.
- Mandinach, E. B., & Gummer, E. S. (2013). Defining data literacy: A report on a convening of experts. *The Journal of Educational Research & Policy Studies*, 13(2), 6-28.
- Qin, J., & D'Ignazio, J. (2010). Lessons learned from a two-year experience in science data literacy education. Proceedings of the 31st Annual IATUL Conference, 20-24 June. Retrieved from <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>
- Ray, J. M. (Ed.) (2014). *Research data management: Practical strategies for information professionals*. Purdue University Press.
- Schneider, R. (2013). Research data literacy. In Kurbanoglu, S. et al. (Ed.), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*, Springer International, Cham, pp. 134-140.
- Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review*, 35(4), 525-540.
- The Association of College and Research Libraries (2000). *Information Literacy Competency Standards for Higher Education*. American Library Association. Retrieved from <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf>
- Wanner, A. (2015). Data literacy instruction in academic libraries: Best practices for librarians. Retrieved from <http://ojs.library.ubc.ca/index.php/seealso/article/view/186335/185517>
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J. : L. Erlbaum Associates.
- Whitmire, A., & Carlson, J. (2015). *Data Literacy*. Retrieved from <http://esciencelibrary.umassmed.edu/data-literacy>
- Yu, J. & Shrimplin, A. K. (2006). Introducing undergraduates to data literacy: How to find, use, and evaluate numeric data. In *Teaching information literacy skills to social sciences students and practitioners: A casebook of applications* (Edited by D. Cook and N. Cooper), Chicago: Association of College and Research Libraries, pp. 14-23.