

DEA를 통한 공공데이터의 효율성 분석

An Efficiency Analysis of the Public Data by DEA

김 동 찬*, 옥 영 석**
 기술경영협동과정*, 기술경영전문대학원**
 부경대학교

Kim dong-chaun*, Ock Young-Seok**
 Interdisciplinary Program of Management of
 Technology*,
 Graduate School of Management of Technology**,
 Pukyong National University

요약

정부의 공공데이터 활성화 전략에 의해 행정기관과 공공기관의 정보공개 및 재이용이 활발해 지고 있다. 이러한 분위기와 함께 민간에서의 공공데이터 활용이 점점 늘어나고 있다. 하지만 당초 기대치에 못 미치는 성과와 공공데이터 개방하는 기관에는 정보화시스템을 운영하는 비용은 증가하였다. 따라서 상대적으로 민간에서 선호하는 문화관광 공공데이터를 통해 공공데이터의 효율성을 분석한 후 시사점을 도출한다.

I. 서론

행정기관과 공공기관을 중심으로 공공데이터 개방이 활발하게 진행되고 있다. 정부의 강력한 개방정책에 의해 공공데이터 개방은 단기간에 양적 확대를 이루었다. 하지만 양적 개방 확대와 투입하는 예산에 비해 민간의 활용도가 아직은 저조한 것으로 보인다. 본 연구는 공공기관의 공공데이터 효율성을 DEA(Data Envelopment Analysis, 자료 포락 분석) 모형으로 상대적 효율성을 서로 비교해 보고, 공공데이터 운영에 관한 투입요소와 산출요소를 활용하여 비효율성 정도를 도출하여 효율성 개선을 위한 정보를 제시하고자 한다.

II. 본론

1. 연구 배경

개방, 공유, 소통, 협력이라는 정부 운영의 전략변화로 행정기관과 공공기관이 소유한 공공데이터를 개방하도록 강력한 드라이브를 걸고 있다. 이를 통해 민간이 공공데이터를 자유롭게 재사용할 수 있도록 기술적, 법적, 행정적 지원도 진행 중이다. 다만 각 공공기관에서 경쟁적으로 비용을 들여 공공데이터를 개방하는 것에 비해 민간에서 활용 낮은 수준이다. 따라서 공공데이터 개방에 대한 효율성을 점검해볼 필요성이 있다.

2. 분석 방법

공공데이터포털(www.data.go.kr)에서 분류한 14개 유형의 민간 활용사례를 분석한 결과 문화관광(25.2%)에 집중된 모습을 보였다. 이는 문화관광 분야의 공공데이터가 활용성과 접근성이 상대적으로 쉽기 때문으로 보인다. 따라서 DEA 분석을 문화체육관광부 소속 산하기관 43개에 기관을 중심으로 DMU(Decision Making Unit)로 선정하였다.

표 1. 공공데이터 활용 사례

순위	분야	활용건수	비율
1	문화관광	228	25.2%
2	교통물류	130	14.4%
3	환경기상	105	11.6%
4	산업고용	83	9.2%
5	공공행정	69	7.6%
6	교육	56	6.2%
7	국토관리	45	5.0%
8	과학기술	43	4.8%
9	재정금융	38	4.2%
10	농축수산	32	3.5%
11	사회복지	31	3.4%
12	재난안전	30	3.3%
13	건강식품	12	1.3%
14	통일외교안보	2	0.2%
합계		904	100%

매년 문화체육관광부에서 수집하는 산하기관 문화데이터 개방 조사와 정보통신산업진흥원이 수집하는 공공부문 SW사업 수요예보 조사 그리고 공공데이터포털의 데이터를 활용하여 투입/산출변수를 설정하였다. 투입변수는 유지보수비용, 기관 보유 서비스 건수이며, 산출변수는 문화데이터 수집 건수, 데이터셋 개방 건수, 메타표준 연계항목수, 공공데이터 이용건수이다. DEA에서 많이 활용하는 인력 변수는 공공데이터를 관리하는 기관에서 1~2명이 다른 업무를 겸임하며 공공데이터를 담당하는 방식이 주를 이루기 때문에 제외했다.

표 2. DEA 모형에서의 투입변수와 산출변수

투입변수		산출변수	
구분	단위	구분	단위
유지보수비용	천원	문화데이터 수집 건수	개

기관 보유 서비스 건수	개	데이터셋 개방 건수	개
		메타표준 연계항목수	개
		공공데이터 이용건수	개

SPSS를 활용하여 투입과 산출 변수의 기술통계량을 분석한 결과 최소값, 최대값, 표준편차, 분산이 모두 큰 차이가 있었다. 이는 문화관광 데이터 특성상 다양한 데이터 유형이 발생할 수밖에 없고 따라서 기관이 보유하는 데이터의 정보량에서 편차가 많이 발생하기 때문이다.

표 3. 기술통계량

유지보수비용

N	최소값	최대값	평균	표준편차	분산
43	3960	3541000	566859	816085	665994565217

기관보유서비스건수

N	최소값	최대값	평균	표준편차	분산
43	200	107552880	4258668	17340089	300678681344354

문화데이터수집건수

N	최소값	최대값	평균	표준편차	분산
43	279	29489432	1484744	6145047	37761605115657

데이터셋개방

N	최소값	최대값	평균	표준편차	분산
43	0	22	5	6	34

메타표준연계항목수

N	최소값	최대값	평균	표준편차	분산
43	2	159	41	37	1357

공공데이터이용건수

N	최소값	최대값	평균	표준편차	분산
43	0	8867	483	1390	1931428

유지보수비용처럼 고정된 투입요소를 바탕으로 최대 산출을 얻기 위한 효율성을 분석하는 CCR(Charnes, Cooper, & Rhoe의 이니셜)모형으로 수행한다. DEA 분석 프로그램은 EnPAS를 활용하였다.

3. 분석결과

분석결과 전체 DMU 평균점수는 0.49 수준이며, 표준편차는 0.35 수준이다. 기술 효율성 값이 1인 DMU는 9개(20.9%)이다. DEA 모형에서는 효율성의 값이 1에 가까울수록 효율적이라고 판단한다.

DMU13, DMU17, DMU27, DMU37, DMU42 등 0.1 이하의 낮은 효율성을 지는 DMU의 경우는 대중의 관심도가 비교적 적은 문화컨텐츠인데다 기관에서 보유한 데이터에 비해 정보 공개가 미흡하여 효율성이 낮게 나온 것으로 판단된다. DMU28의 경우는 민간이용률이 가장 많았지만 유지보수비용이 3번째로 커서 효율성 0.69정도로 나왔다. 반면 효율성이 1인 DMU의 경우에는 공공데이터 이용률이 낮더라도 상대적으로 유지보수비용이 낮았다.

표 4. CCR 모형 분석결과

구분	CRS	준거집단	참조횟수
DMU1	0.2837	4,16,33	0
DMU2	0.2718	5,29,33	0

DMU3	0.2029	7,33,38	0
DMU4	1	4	11
DMU5	1	5	7
DMU6	0.186	7,16,33	0
DMU7	1	7	3
DMU8	0.1317	33	0
DMU9	0.268	4,16,33	0
DMU10	0.7128	16	0
DMU11	0.2161	5,29,33	0
DMU12	0.1051	16,33	0
DMU13	0.0726	4,16,33	0
DMU14	1	14	0
DMU15	0.5585	5,29,33	0
DMU16	1	16	25
DMU17	0.0651	4,16,33	0
DMU18	0.2056	4,7,16	0
DMU19	0.1091	16,25,29	0
DMU20	0.2237	4,16,33	0
DMU21	0.6011	16,29,33	0
DMU22	0.5375	4,16,33	0
DMU23	0.2272	16,33	0
DMU24	0.5084	16,25,29	0
DMU25	1	25	2
DMU26	0.4971	4,16,33	0
DMU27	0.0952	16,33	0
DMU28	0.6959	16,29,33	0
DMU29	1	29	11
DMU30	0.2274	5,33	0
DMU31	0.5242	5,16	0
DMU32	0.1894	16,29,33	0
DMU33	1	33	28
DMU34	0.9302	5,33	0
DMU35	0.7264	33,38	0
DMU36	0.9089	4,16,33	0
DMU37	0.0157	33	0
DMU38	1	38	2
DMU39	0.8896	4,16	0
DMU40	0.5449	5,16,29,33	0
DMU41	0.289	4,16,33	0
DMU42	0.0141	16,29,33	0
DMU43	0.1073	16,29,33	0

III. 결론

공공데이터를 개방하는 업무가 기관 입장에서는 무시할 수 없는 비용과 노력이 들어가지만 민간의 활용도는 낮은 수준이다. 이용율이 높더라도 공공데이터를 생산 및 운영하기 위한 유지보수가 많이 소요되기 때문에 역설적으로 이용률이 적더라도 비용이 적게 소요되는 DMU가 더 효율적이었다. 이는 공공데이터의 민간 활용률의 절대치가 낮기 때문에 생기는 문제로 보인다. 이러한 문제점을 해결하기 위해서는 공공데이터의 이용을 활성화하고 이용접근성을 편리하게 개편해야 될 것으로 판단된다.

■ 참고 문헌 ■

- [1] 박만희 “효율성과 생산성 분석” 한국학술정보(주)
- [2] 이정동, 오동현 “효율성분석이론” (주)지필미디어
- [3] 공공데이터 이용활성화를 위한 비용부담방안 연구, 한국정보화진흥원