

## “CKAN DCAT, 전 세계와의 데이터 공유 가속화할 수 있는 기술”

김형근 모비젠 연구소장

### Q. CKAN DCAT 방식이 선호되는 이유는?

CKAN 과 DCAT 은 이미 해외에서 검증된 사례가 많고, 정부에서 추진하고 있는 데이터 공유·유통 플랫폼 구축에 최적화된 ‘표준’과 ‘개방’을 지키는 플랫폼이다.

CKAN 에서는 서로 다른 데이터센터들 간에 데이터를 공유·연계할 수 있고 이를 ‘하베스팅(harvesting)’이라는 용어로 표현한다. 하베스팅을 하면 CKAN 으로 구축한 데이터 센터 간에는 서로 손쉽게 데이터를 ‘수확’하고 활용할 수 있다.

이는 비단 국내에만 한정된 게 아니다. 전 세계에서 많은 기관들이 CKAN DCAT 방식을 지켜서 데이터 플랫폼을 구성하고 있으며, 다양한 데이터 표준 중 가장 국제적인 협력이 잘 일어나고 있다. 따라서 국내에서 CKAN DCAT 을 준수해 구성한 데이터 플랫폼에서는 CKAN 으로 구현된 해외 플랫폼의 데이터도 손쉽게 활용할 수 있다. 한국정보화진흥원에서는 ‘데이터스토어’의 ‘글로벌 데이터’ 페이지에서 CKAN DCAT 으로 구성된 다양한 해외 데이터 플랫폼들을 안내하고 있으며, 해당 플랫폼에 올라가있는 데이터 상품들을 확인할 수 있다.



▲ 김형근 모비젠 연구소장

### Q. 모비젠이 보유한 오픈 데이터 플랫폼 기술은?

모비젠은 지난 2017 년부터 국책과제인 ‘국제표준 기반 오픈 데이터 유통 플랫폼 확장 기술 개발’에 참가해 ‘아이리스(IRIS) 오픈 데이터 플랫폼’을 개발했다. ‘아이리스 플랫폼’은 국제 표준에 기반한 유통 플랫폼으로서 CKAN 을 고도화하고 국내 환경에 맞게 다양한 관리 및 유통 기능을 추가한 솔루션이다.

가령 유료 결제 기능 등이 새롭게 추가됐다. 본래 CKAN 은 데이터의 무료 공유·개방을 위해 만들어진 플랫폼이다. 그래서 CKAN 자체에 별도의 결제 기능은 포함돼 있지 않다. 그런데 국내에서는 유료 데이터를 가지고 있는 기관들을 영입하고 데이터의 개방과 거래를 모두 실행하기 위해서는 유료 결제 모듈을 더할 필요가 있었다. 그래서 결제 기능을 CKAN 과 연계해 유통 플랫폼으로 재조정하는 기능을 추가했다.

한편 해당 국책 과제의 목표에는 CKAN 을 우리나라 환경에 맞게 조정하는 것 뿐만 아니라 오픈소스 CKAN 에 대해 코드 기여를 하는 것도 포함돼 있었다. 현재 모비젠은 CKAN 에서 사용하고 있는 성능 관련 이슈와 실시간 스트림 하베스팅 기능 등에 대해 코드 기여를 하고 있다.

### Q. 공공데이터는 사용하기 어렵고 품질이 낮다는 불만에 대해 어떻게 생각하는가?

전 세계 어디를 가나 처음부터 깔끔하게 만들어져 있는 데이터는 단 하나도 없다. 무료로 공유되는 데이터의 품질이 낮으니 우리는 못하겠다고 내던지는 게 아니라, 질 낮은 데이터를 가져다가 가다듬고 가공해서 사용할 수 있도록 만들어서 써야 한다. 그렇게 하고 싶지 않다면 적절한 비용을 지불하고 데이터 가공 업체를 이용해도 될 것이다. 어느 쪽이든 비용이 드는 작업이다.

데이터를 가공하는 데에는 비용이 들고, 따라서 외부에서 생산된 데이터를 필요로 하는 기업이라면 데이터에 대한 구매 비용을 예산으로 할당해야 한다. 우리가 1년 예산을 짤 때 서버 구매비용이나 인프라 유지 비용 등을 고려하는 것처럼, 데이터 구매에 어느 정도 예산을 투자할 것인지 고민할 필요가 있다. 데이터에 대한 소비가 늘어나면 데이터를 생산하는 생태계도 확대될 것이고, 이는 전체적인 데이터의 품질 향상으로 이어질 것이다.

## “데이터 유통, 모두의 업무로 이해해야”

김형근 모비젠 연구소장은 “데이터 개방과 공유 정책을 각 기관의 중요한 업무로, 각 개인의 중요한 역할로 인식해야 한다”고 강조했다. 공공기관과 기업들이 자사의 이익만 신경쓰며 데이터를 공공 숨기는 것은 바람직하지 못하며, 이는 해당 데이터의 가치를 떨어트린다는 점에서 국가적인 손실일 뿐만 아니라 해당 기관의 손실이기도 하다.

데이터를 공유하기 위해서는 해당 데이터를 생산하고 공유하려는 담당자가 최소한의 가공을 해야 한다. 이는 본래라면 하지 않아도 될 업무이기에 당연히 일은 늘어나고, 처음 해보는

일이라면 쉽지 않을 수도 있다. 하지만 데이터를 공개하기 위해 정리하고 손보는 과정에서 그 데이터가 가지고 있는 가치를 재조명하게 되고, 그렇게 가공돼서 가치를 부여받은 데이터가 유통되면서 서로 다른 가치를 가진 데이터를 만나 시너지를 일으키게 될 것이다.

데이터의 공유와 개방, 민간 데이터까지 포함한 유통 생태계를 만들고자 하는 정부의 방침에 대해서는 공공과 민간 모두 긍정적인 반응을 보이고 있다. 멀지 않은 미래가 데이터 중심 사회로 예측되는 가운데, 정부의 데이터 중심 정책이 어떠한 결과로 이어질지 귀추가 주목된다.