

빅데이터 유통과 거래 모델

최재영 빅데이터 프로젝트 그룹(PG1004) 위원, 성균관대학교 교수

1. 머리말

빅데이터는 특성상 어느 한 분야의 데이터만 분석하는 것보다는 이종 데이터를 결합 분석할 때 새로운 부가가치를 발견할 가능성이 크다. 이를 위하여 전 세계적으로 공공/민간 및 동종/이종 간의 데이터 연계 활용에 막대한 노력을 투자하고 있으며 이를 위한 데이터 개방·유통 거래에 대한 관심과 중요성도 함께 커지고 있다. 더욱이 빅데이터 유통 단계는 생산에서 소비에 이르는 일련의 과정을 이어주는 중개인의 역할을 하면서 이용자가 필요로 하는 데이터를 더 쉽게 얻을 수 있게 하고, 데이터 간의 연계·융합 편의성을 높여준다는 점에서 데이터의 이동과 함께 데이터의 가치를 높일 수 있는 매우 핵심적인 단계이다.

본고에서는 다양한 데이터 상품이 유통, 거래되는 데이터 유통 생태계 모델의 일례를 제시하고, 이러한 생태계에서 이동하는 데이터 관점에서 데이터 거래 모델, 방식, 포맷, 고려사항 등을 서술하고자 한다.

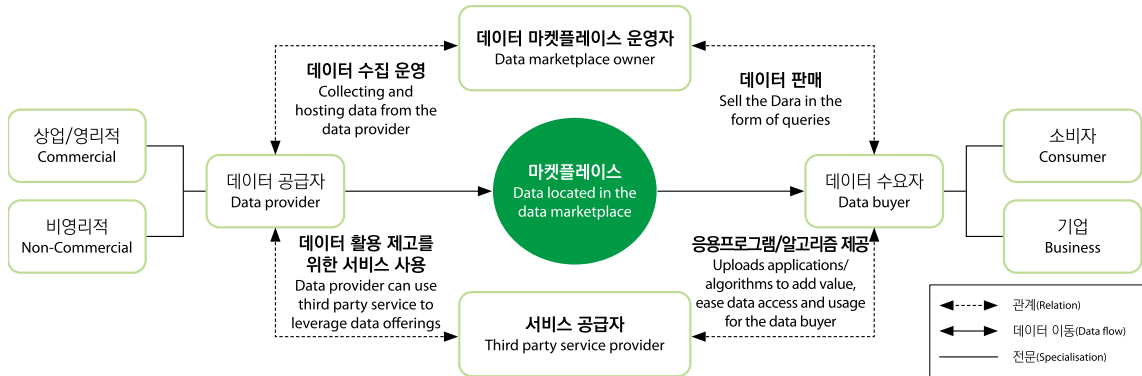
2. 데이터 유통 생태계와 거래 모델

2.1 데이터 유통 생태계 모델

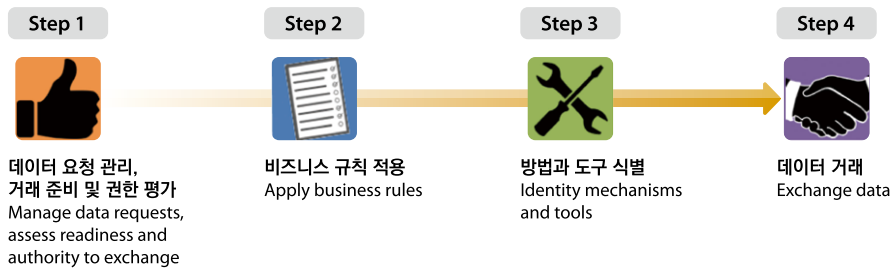
데이터 유통의 개념은 매우 포괄적으로, 데이터 흐름(거래포함)에 관계되는 모든 참여자의 연결 및 상호 작용을 통하여 가치사슬을 만드는 것을 뜻하며, 이러한 참여자들이 협력하거나, 때로는 경쟁하면서 상생하는 곳을 유통 생태계라 볼 수 있다.

[그림 1]은 데이터 마켓플레이스(Data Marketplace)이자 유통·거래 플랫폼 모델의 일례를 보여준다[1]. 유통·거래 플랫폼은 상품(데이터)을 거래하는 공급자와 수요자가 만나기 위하여 개인 혹은 그룹(기업)을 온/오프라인으로 연결함으로써, 표준화된 인터페이스와 서비스를 통해 좀 더 효율적인 상호작용이 가능하게 한다.

이러한 플랫폼은 중립적인 중개자 역할을 해야 하며, 플랫폼을 통해 누구나(또는 최소한 많은 잠재적으로 등록된 수요자) 데이터 제품을 업로드하고 판매할 수 있어야 한다. 또한, 정적 데이터 또는 (동적)데이터 스트림을 모두 제공할 수 있으며 개별 파일 다운로드, API 또는 맞



[그림 1] 데이터 유통 생태계 모델(예)



[그림 2] 데이터 거래 모델

충형 웹 인터페이스와 같은 다양한 액세스 유형을 통해 액세스를 허용함으로써 모든 시장 거래를 처리할 수 있는 데이터 중심의 시스템으로 구성돼야 한다. 본고에서는 일례로 제시한 [그림 1]의 생태계 모델상에서 데이터 관계(relation), 데이터 이동(data flow) 관점에서 발생하는 데이터 거래 모델, 방식, 포맷, 고려사항 등을 기술하고자 한다.

2.2 데이터 거래 모델

데이터 거래 모델의 대표적인 사례 중 하나는 호주의 빅토리아주다. 빅토리아 주정부는 데이터 공유 평가, 관리, 권한 부여 및 수행에 대한 지침을 마련하여 데이터 소유자 및 관리자(비즈니스 내에서), 데이터 요청자, 데이터 및 정보

관리 실무자와 정보 기술 전문가 등이 지원 도구 및 템플릿을 사용하는 한편, 데이터 거래 표준(Data Exchange Standard)을 이해하고 실행하는 데 도움을 주고 있다. [그림 2]는 해당 모델의 개요를 보여준다. 이 지침은 데이터 교환을 위해 고려해야 할 최소한의 고려사항, 적용해야 하는 비즈니스 규칙 및 데이터 교환의 평가와 관리에 도움이 되는 도구 사용에 대한 부가적인 조언과 설명을 제공한다[2].

우리나라에서는 주로 비즈니스 모델에 기반하여 데이터 거래를 고려함에 따라 가격, 계약 중심으로 서술하는 경우가 많으나, 본고에서는 공공데이터를 포함한 데이터 관점에서의 거래를 주로 다루고자 한다. 상기 거래 모델의 1단계에서 성공적인 데이터 거래 또는 전송에 중요한 것

은 데이터 요청(수신된 요청과 전송된 요청 모두)을 적절하게 관리하여 데이터를 공유할 준비가 되었는지 이를 공유할 권한이 있는지를 평가하는 것이다. 2단계에서는 비즈니스 규칙을 식별하여 안정적이고 일관되며 지속 가능한 데이터 거래 및 의사 결정을 보장하는 데 있는데, 여기에 계약 등의 내용도 포함된다. 3단계에서는 데이터 거래가 점점 더 많은 서비스와 기능을 지원하기 위하여 사용 가능한 항목, 대화 대상, 사용할 도구 및 템플릿을 파악함으로써, 간소화되고 안전하며 승인된 데이터 거래가 이루어질 수 있다. 4단계에서는 각 데이터 거래가 특성에 따라 다르지만 설계, 테스트 및 BAU(Business-As-Usual)에서는 유사한데, 이를 위한 지침(예를 들어 데이터 거래를 설계하고 수행하는 과정, 거래 종단간 테스트 수행 방법, 데이터 거래 계약을 BAU에 넘기기 위한 단계 등)을 제시하여 표준화된 데이터 거래방식을 통한 유통·거래 활성화를 제고한다. 단계별 간략한 주요 고려항목은 <표 1>과 같다.

2.3 데이터 거래 방식

데이터 거래 방식은 이해당사자 사이에 데이터를 제공할 때 적용하는 데이터 전달 방식으로서, <표 2>에서 보는 바와 같이 API(Application Programming Interface), ETL(Extract, Transform, and Load), 파일 전송(File Transfer/Download), 원격 프로시저 호출(Remote Procedure Call), 이벤트 기반/중개 메시징(Event Based/ Brokered Messaging), 데이터 스트리밍(Data Streaming) 방식 등이 있다[3].

그러나 일반적으로 데이터 마켓플레이스에서는 데이터 품질 및 메타데이터 수준이 향상된 덕분에 다양한 비즈니스에서 사용할 수 있도록 준비된 데이터를 제공한다. 이에 따라 해당 데이터를 분류하고, 그 속성을 설명하고, 또한 추가로 ETL 프로세스에 시간을 할애하지 않고 인사이트를 얻을 수 있는 상태의 데이터를 취급하고 거래함으로써 주로 API(open API) 방식과 파일 전송/다운로드 방식이 쓰인다[5].

<표 1> 데이터 거래 단계별 주요 고려 항목[4]

단계	주요 고려 항목
데이터 요청 관리, 거래 준비 및 권한 평가 (Step 1)	거래 준비 상태 평가 (Assess readiness to exchange (guideline)) 거래 권한 평가 (Assess the right to exchange (guideline)) 데이터 요청 관리 (Managing data requests (guideline))
비즈니스 규칙 적용 (Step 2)	데이터 거래 표준 (Data exchange standard) 데이터 거래 계약 표준 수립 (Establishing a data exchange arrangement standard)
방법과 도구 식별 (Step 3)	데이터 서비스 및 인프라 (Data services and infrastructure) 데이터 소스 및 스키마 (Data sources and schemas) 데이터 전문가 (Data experts) 규정 준수 점검 (Compliance) 데이터 거래 위험 평가 모델 (Data exchange risk assessment model) 데이터 요청 (Data request (template)) 데이터 거래 점검표 (Data exchange checklist (template)) 데이터 거래 계약 (Data exchange agreement (template)) 데이터 거래 의도 (Data exchange statement of intent (template)) 데이터 거래 기술 사양 (Data exchange technical specification (template))
데이터 거래 (Step 4)	데이터 거래 설계 (Designing a data exchange (guideline)) 데이터 거래 테스트 (Test a data exchange (guideline)) 일상적 영업 진행 (Transition to BAU(Business-As-Usual) (guideline))

데이터 거래 방식은 사전에 결정된 특정 방식을 따르기보다, 각 방식의 장단점과 수요자의 요구사항 간 균형을 맞춰 조정되는 경우가 많다. 이러한 경우 전송할 데이터 개체에 여러 관련 요소가 포함되어 있거나 특정 구성 요소 등이 있는지, 데이터 세트가 자주 업데이트되고 업데이트 수가 어느 정도인지, 데이터 세트 크기를 고려한 파일 전송의 성능은 어떠한지 등 데이터 세트의 특성을 고려해야 한다. 또한 다양한 계획과 잠재적 데이터 흐름에 대한 분석을 통한 데이터 거래 방식의 설계, 시간·비용·품질과 맞물리는 사용 가능한 기술, 비즈니스 전략 및 조직 문화와 같은 요소도 거래 방식을 결정하는 제약조건 중의 하나다.

따라서 단순 기술적인 성능만으로 특정한 거래 방식을 선택하기보다 다양한 고려사항을 점검하고 테스트하여 최적의 교환 방법을 선택하

는 것이 중요하다. <표 3>에서는 데이터 유통·거래에서 이러한 데이터 거래방식 선택 시 고려사항과 이에 대한 구체적 내용을 간략히 기술했으며, <표 4>에서는 사용 사례에 따라 일반적인 데이터 거래 방식의 선호 순위를 제시했다.

2.4 데이터 거래 포맷

데이터의 거래 포맷은 다른 데이터와 연결 가능성, 검색 용이성 및 특정 소프트웨어에 대한 의존성 등을 기준으로 등급을 나누어 관리한다. 본고에서는 팀 버너스리(Tim Berners-Lee)에 의한 ‘5수준 별점 모델(5-star deployment scheme for Open Data)’의 데이터 제공 규격을 차용하여 마켓플레이스에서 데이터를 제공할 때의 포맷을 기술했다[6]. 데이터 거래 포맷의 각 단계별로 데이터 구매자들이 얻는 기대 효과 및 특징은 <표 5>와 같은데, 수준 5 거래

<표 2> 유통과정에서의 데이터 거래(Data exchange) 방식

방식	설명
API (Application Programming Interface)	SOAP, RESTful 등의 웹서비스 API를 통하여 데이터를 송·수신
ETL (Extract, Transform, and Load)	데이터베이스에 직접 연결하여 데이터를 읽고 쓸 수 있도록 하여 전송
파일 전송 (File Transfer)	JSON, XML, CSV 또는 기타 여러 텍스트 기반 또는 이진 파일 형식을 사용하여 대상 위치로 전송
원격 프로시저 호출 (Remote Procedure Call)	프로시저가 다른 주소 공간 (일반적으로 공유 네트워크의 다른 컴퓨터에서)에서 실행되도록 하여 데이터 교환
이벤트 기반/중개 메시징 (Event Based/Brokered Messaging)	데이터가 포함된 메시지를 생성하여 전달할 서비스에 제공
데이터 스트리밍 (Data Streaming)	데이터를 수신 프로세스로 지속적으로 전송

<표 3> 유통상에서 데이터 거래방식 선택시 고려사항


고려사항	내용
데이터 세트 특성	데이터 복잡성(Data complexity), 데이터 업데이트 빈도(Frequency of data update), 데이터 세트 크기(Data set size)
데이터 환경 특성	데이터 흐름과 광범위한 솔루션(Data flows and breadth of solution), 데이터 사용 빈도(Frequency of data usage), 데이터 버전(Data versions), 데이터 보안(Data security), 데이터 변환 복잡성(Data transformation complexity), 연결 지속성(Connection persistence)
범위 제약	시간(time), 비용(cost) 및 품질(quality)의 삼각범위 균형(trade-off) 고려
조직 고려 사항	조직 요구사항의 균형(프로젝트 및 응용 프로그램별 요구 사항과 함께 더 광범위한 조직 요구 사항의 균형을 맞추는 시도)
소비자 특성	소비자 시스템 혹은 선호 방식 고려(비즈니스 프로세스와 관련된 시스템)

포맷은 데이터 연결을 통해 다른 데이터와 쉽게 링크가 가능한 수준(link your data to other data to provide context)을 뜻한다[7]. 수준 4 거래 포맷은 URI를 통해 사람들이 쉽게 검색할 수 있는 수준(use URIs to denote things, so that people can point at your stuff)이며, 수준 3은 특정 소프트웨어를 벗어나 어디서든 열어보고 편집 가능한 수준(make it available in a non-proprietary open format, e.g., CSV instead of Excel)을 말한다. 최근의 마켓플레이스에서는 팀 버너스리의 권고 수준에 따라 데이터의 유통 거래를 활성화하기 위해 최소 수준 3 이상의 포맷으로 유통·거래한다.

3. 맺음말

데이터 유통에서 거래는 점점 더 중요한 요소로 자리잡고 있으며, 가격책정, 계약 이슈, 개인정보 여부 등과 맞물려 유통 과정 중 가장 고려사항이 많은 과정이라 할 수 있다. 본고에서는 데이

터의 생성, 처리, 저장, 배포 등의 전체 유통 생태계 속에서 데이터 공급자, 수요자, 서비스 공급자, 마켓플레이스 운영자 간 데이터 이동을 위한 데이터 거래 관점에서 모델, 방식, 포맷 등의 참조요소들을 기술하였다.

국내에서는 2019년 10대 분야 빅데이터 플랫폼을 시작으로 2020년 6개 플랫폼을 추가로 신규 구축하여 데이터의 수집·유통 및 활용 기반을 광범위하게 마련함과 동시에, 활발한 데이터 유통을 위하여 분야별/분야간 데이터 표준화를 추진하고 있다. 이를 통해 올 상반기에 약 3만 6천 건의 데이터가 이용됐으며 연말까지 약 3천 종의 데이터가 구축된다. 현재 이러한 플랫폼은 데이터 거래 및 데이터 관련 서비스에 필요한 인프라를 제공함으로써 데이터 경제의 중심으로 이동하고 있으며, 본고에서 기술한 기본적인 요소들의 검토 이외에도 데이터 공유·거래의 주요 장애물인 보안, 표준, 정책 등의 당면한 현안들에 대해 기반한 해법을 통하여 다양하고 활발한 데이터 유통·거래가 펼쳐지기를 기대해본다. 

<표 4> 사용사례에 따른 데이터 거래방식

사용사례 (Use cases)	방법별 순위			
	API	메시징	ETL	파일전송
여러 형식이 요구되는 데이터 (The data is required in multiple formats)	높음	중간	중간	낮음
클라이언트 프론트엔드에서 사용되는 데이터 (The data is used in a client front-end)	높음	중간	중간	낮음
특징을 지원하는 데이터 (The data supports a feature)	높음	높음	중간	낮음
빈번히 요청되는 데이터 (The data is frequently requested)	높음	높음	중간	낮음
빈번히 변경되는 데이터 (The data changes frequently)	높음	높음	중간	낮음
백엔드 시스템에서 사용되는 데이터 (The data is used in a back-end system)	낮음	높음	높음	중간
다중 흐름이 있는 데이터 (The data has multiple flows and receivers)	중간	높음	낮음	낮음
다중 버전/스키마로 요청되는 데이터 (The data is requested in multiple versions/schemas)	낮음	낮음	높음	낮음
다중 속성과 로직을 포함한 데이터 결합 (Assembling the data involves multiple entities and/or variable logic)	낮음	낮음	높음	낮음
데이터 세트가 큰 경우 (The data set is very large)	낮음	중간	높음	높음
대형 플랫폼의 기초를 형성하는 데이터 (The data forms the basis of a larger platform)	낮음	낮음	높음	높음
사람이 읽을 수 있어야 하는 데이터 (The data must be human readable)	낮음	낮음	낮음	높음

<표 5> 데이터 거래 포맷의 개방수준별 특징

개방수준	포맷	기대효과 및 특징(costs & benefits)
5	LOD	<ul style="list-style-type: none"> • 데이터를 쉽게 이용하며, 관련된 데이터 검색이 용이 • 데이터 스키마에 대한 직접적 이해 가능 • 데이터가 검색(discoverable)되어 데이터의 가치 제고 가능
4	RDF, URI	<ul style="list-style-type: none"> • 어떤 곳(로컬이나 웹 상)에서라도 링크, 북마크가 가능 • 데이터 일부를 재사용할 수 있음 • 데이터 항목에 대한 정교한 제어가 가능하고, 검색을 최적화(로드 밸런, 캐싱 등)할 수 있음
3	SHP, GML, JSON, JSON-STAT, GEOJSON, CSV, TSV, XML, KML	<ul style="list-style-type: none"> • 어떤 독점 소프트웨어 패키지도 소유할 필요없이 자신이 원하는 방법으로 데이터를 다룰 수 있음 • 독점 포맷 데이터를 내보내기 위해서 변환기나 플러그인 필요 • 비교적 데이터 개방을 통한 제공이 쉬움
2	XLS, XLSX, ODS, HWP, JPG	<ul style="list-style-type: none"> • 데이터 수집, 계산, 시각화 등 독점 소프트웨어를 사용해 데이터를 직접 처리할 수 있음 • 다른(구조화된) 포맷으로 데이터를 내보낼 수 있음 • 판매자 입장에서 보유 데이터를 그대로 개방하는 강점 보유
1	PDF, ODF, ZIP, HTML, XHTML	<ul style="list-style-type: none"> • 직접 보거나, 인쇄가 가능 • 저장장치에 저장이 가능하고, 다른 전산시스템에 입력이 가능 • 원하는 대로 데이터를 수정하거나 공유가능

- LOD : Linked Open Data
- RDF : Resource Description ramework
- URI : Uniform Resource Identifier
- SHP : Shapes File
- GML : Geography Markup Language
- JSON : JavaScript Object Notation
- JSON-STAT : JavaScript Object Notation - STAT
- GEOJSON : GeoJSON File
- CSV : Comma Separated Values File
- TSV : Tab Separated Values File
- XML : Extensible Markup Language
- KML : Keyhole Markup Language
- XLS : Excel Spread sheet
- XLSX : Microsoft Excel Open XML Spreadsheet
- ODS : Open Document Spreadsheet
- HWP : Hanword Document
- JPG : Joint Photographic Experts Group : Images files
- PDF : Portable Document Format
- ODF : Open Document File
- ZIP : Zipped File
- HTML : HyperText Markup Language
- XHTML : Extensible HyperText Markup Language

참고문헌

- [1] Data Marketplaces: Trends and Monetisation of Data Goods, Markus Spiekermann, Intereconomics 2019, pp.208~216
- [2] Data Exchange Framework IT Strategy Action Plan 2017-18, State Government of Victoria, 2018
- [3] Data Exchange Methods and Considerations, Greg Charest and Mitch Rogers, Enterprise Architecture at Harvard University, 2020
- [4] <https://www.vic.gov.au/data-policies-and-standards#data-quality-guidelines>
- [5] <https://research.aimultiple.com/data-marketplace/>
- [6] <https://5stardata.info/en/>
- [7] <https://www.w3.org/DesignIssues/LinkedData.html>