

# 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

이혜영



## 요약문

### 1. 제목 : 공공 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

현재 국내 공공 데이터 활용을 위한 법제 개선과 정부 주도의 전략적 사업 지원으로 강화되고 있는 반면 빅데이터 분석 플랫폼은 단일화 채널이 아니며 공공 기관들과의 연계성 미비, 데이터 표준체계 부족 등의 이슈를 가지 보여주고 있다.

빅데이터의 링크드 데이터 연계 구축 플랫폼을 통한 끊임 없는 확장된 데이터 중심의 분석플랫폼을 제안한다. 소프트웨어공학 설계기법인 프로덕트라인 공법을 적용한 LOD 적용 플랫폼에서 등록 발행된 링크드 데이터를 재사용하고 데이터 자산을 응용, 활용하여 공공 빅데이터 가치 확산에 기여할 수 있다. 더불어 소프트웨어 아키텍처 모델링 언어인 UML을 적용하여 원활한 분석 플랫폼 개발 소통과 지식 공유를 돕는다.

### 2. 사업의 목적 및 필요성

빅데이터의 새로운 비즈니스 성장엔진의 원동력으로써 IT 정보통신기술의 이중 산업으로의 융복합 진전화와 함께 확산된 빅데이터는 실시간으로 모바일, 센서 디바이스 및 소셜 네트워크, 클라우드 컴퓨팅에 의해 수집되어 가치창출을 위한 분석 활용을 위한 기반이 되고 있다. 빅데이터의 중요성과 잠재적 역량을 이해한 미국, 유럽 등의 세계 여러 국가에서는 공공 개방 빅데이터를 통한 공공 서비스 효율화 및 국가 경쟁력 확보를 위한 공공 서비스 부문에서의 활용에 대한 분석 플랫폼 마스터 플랜 제정, 민간 서비스 연계 확장과 국가 산업 경쟁력 향상을 꾀하고 있다.

미국은 투명하고 열린 정부라는 표제 하에“Big Initiative”아젠다를 기준하여

중앙 부처와 지방정부 개별적인 다양한 정책을 발굴하여 추진하고 있으며 EU는 “ODS: Open Data Strategy”차원에서 모든 공공정보와 데이터를 의무적으로 개방, 2013년 “EU 데이터 포털”을 통해 일반 시민과 기업에서 공공 서비스 빅데이터의 상업적 가치 전환이 가능토록 지원하고 있다.

공공 서비스 데이터를 기반한 재가공 산업 모델은 경제 활성화, 일자리 창출에 따른 생산유발 효과(EU 예측, 시장 고용 규모 증대 약 16조원 이상)를 가져올 뿐만 아니라 공공 데이터의 단계적 개방으로 민간의 가치 창출 및 기업 활동 지원을 강화시킬 수 있다. 대한민국 정부도 빅데이터 산업 국가 전략 수립으로써 공공 데이터 활성화를 위한 법률 제정 및 공공 데이터 개방 정책을 시행 중에 있으나 데이터 개방 방식의 표준화와 분석 연동성에 있어 문서 형태 공개 중심, 부문 간 링크드 데이터의 미동적 활용, 외국 기업의 기술과 오픈 솔루션에 의존, 각 기관들의 정보 공유 협업 미비등의 제약 사항들로 실제 접목 활용한 참조사례들이 부족한 실정에 있다.

공통적인 이슈에 대한 해결방안으로 국내에서 구축된 현재의 플랫폼에서의 빅데이터 활용가능성을 진단, 적합한 데이터 개방에 따른 빅데이터 분석 처리 및 정보 연계 플랫폼 구축을 위한 마스터 플랜 추진을 제안한다. 당 사업은 한국 사회의 공공서비스 개선과 민간으로의 빅데이터 활용 확대의 핵심이자 첫 단계시점으로 공공기관의 정보지식 데이터의 잠재적 가치와 분석 플랫폼 구축으로의 기술적 사안을 중점적으로 조명함으로써 외국과 차별화된 ICT 자산을 기반으로 한 적극적인 마스터 플랜을 제안하고자 한다.

### 3. 사업의 내용 및 범위

빅데이터의 본 사업은 한국의 강점인 ICT 인프라 자산을 활용한 빅데이터 분석 인프라 구축 마스터 플랜 정립에 기여코자 하며 해외 공공 데이터 분석 기술 및 활용 적용 사례를 소개함으로써 공공/민간 서비스 창출에 필요한 적합한 데이터 개방이 필요함에 대한 인식 공유를 목표로 한다.

- (기술 동향) 최신 빅데이터 분석 기술 소개 / 링크드 데이터 조명
- (공공 서비스 데이터 개방 활용) 영국 / 싱가포르 / 미국 공공 빅데이터 분석 인프라 - 활용 효과 사례
- (국내 빅데이터 분석 플랫폼 마스터 플랜 구축 제안) 현황 점검 및 개선 모델링 제안

본 사업 목표를 달성코자 다음과 같은 사업내용과 범위로 구성하였다.

제1장‘서론’은 본 사업의 목표에 따른 내용 및 범위, 활용제안으로 구성한다.

제2장‘해외 공공 빅데이터 활용’은 제1절 해외 정부에서의 공공 빅데이터 가치 제고와 활용 현황을 소개한다.

제3장 사업내용으로‘해외 빅데이터 적용 공공 서비스 동향 및 분석 인프라(가칭 인사이트)’에서는 제1절 EU(영국)의 빅데이터 정책과 데이터 중심 플랫폼 기반 하의 공공 서비스 부문 적용 사례 측면에서 구성하였으며 또한 링크드 데이터 중심의 플랫폼을 통한 공공 데이터 활용 서비스 조명을 통해 국내 공공 서비스에서의 활용 가능한 시사점을 제공한다. 제 2,3절에서는 싱가포르와 미국의 공공분야에서의 빅데이터 분석 인프라 분석 및 도입된 분석 기술과 플랫폼 아키텍처, 공공 빅데이터 활용사례를 설명한다.

4절에서는 기존 구축된 국내 빅데이터 분석 인프라를 진단, 링크드 데이터 플랫폼기반 분석 기술 적용에 대한 필요성을 언급한다. 5절에서는 소프트웨어공학 기법인 UML모델링을 이용한 프로젝트 라인 설계 기법을 통한 데이터 재활용성이 부각된 국내 공공 서비스 빅데이터 활용 추진을 위한 분석 플랫폼 마스터 플랜

모델링을 제시하며 빅데이터 활용 인식 전환에 따른 유기적 협업 체계를 통한 데이터 기반 공공 플랫폼 구축 방향에 대한 제언을 추가한다.

#### 4. 정보조사의 결과

2013년 공공 데이터의 적극적 개방을 통한 신뢰 정부 구현 및 신성장동력을 창출하자는 비전을 담은 정부 3.0 전략을 통해 고가치, 고수요 데이터를 우선적으로 개방하며 민간 주도 개방 로드맵 수립, 빅데이터 데이터 개방체계 마련 및 질적 관리 표준체계 구축으로 진행하고 있으나 해외 선진국 추진 전략과 세부 계획과 비교할 시 기관들과의 데이터 공유와 동기화, 손쉬운 오픈 데이터 플랫폼을 위한 링크드 데이터 인프라구축 고려가 부족한 편이다.

#### 5. 조사의 활용방안 및 기대효과

빅데이터 분석 플랫폼은 단일화 채널이 아니며 공공기관들과의 연계성 미비, 데이터 표준체계 부족 등의 이슈를 가지 보여주고 있다. 빅데이터의 링크드 데이터 연계 구축 플랫폼을 통한 끊김 없는 확장된 데이터 중심의 분석플랫폼을 제안한다. 소프트웨어공학 설계기법인 프로덕트라인 공법을 적용한 LOD 적용 플랫폼에서 등록 발행된 링크드 데이터를 재사용하고 데이터 자산을 응용, 활용하여 공공 빅데이터 가치 확산에 기여할 수 있다. 더불어 소프트웨어 아키텍처 모델링 언어인 UML을 적용하여 원활한 분석 플랫폼 개발 소통과 지식 공유를 돕는다.

## 목 차 | Contents

---

|                            |    |
|----------------------------|----|
| 제1장 서론                     | 1  |
| 제2장 해외 공공 빅데이터             | 3  |
| 1. 해외 공공 빅데이터 활용 현황 (영국)   | 3  |
| 2. 해외 공공 빅데이터 활용 현황 (싱가포르) | 3  |
| 3. 해외 공공 빅데이터 활용 현황 (미국)   | 5  |
| 제3장 사업내용                   | 7  |
| 1. EU(영국) 공공 빅데이터 인사이트     | 7  |
| 가. EU(영국) 공공 빅데이터 분석 플랫폼   | 7  |
| 나. EU(영국) 공공 빅데이터 분석 기술 현황 | 11 |
| 다. EU 공공 빅데이터 분석 활용 사례     | 19 |
| 2. 싱가포르 공공 빅데이터 인사이트       | 27 |
| 가. 싱가포르 공공 빅데이터 분석 플랫폼     | 27 |
| 나. 싱가포르 공공 빅데이터 분석 기술 현황   | 28 |
| 다. 싱가포르 공공 빅데이터 분석 활용 사례   | 31 |
| 3. 미국 공공 빅데이터 인사이트         | 34 |
| 가. 미국 공공 빅데이터 분석 플랫폼       | 34 |
| 나. 미국 공공 빅데이터 분석 기술 현황     | 36 |
| 다. 미국 공공 빅데이터 분석 활용 사례     | 39 |

---

## 목 차 | Contents

---

|                                       |        |
|---------------------------------------|--------|
| 4. 대한민국 공공 개방 데이터 분석 플랫폼 현재 (AS-IS)   | 4      |
| 가. 대한민국 공공 빅데이터 분석 플랫폼                | 44     |
| 나. 대한민국 공공 데이터 분석 기술 현황               | 48     |
| 다. 대한민국 공공 빅데이터 분석 활용 사례              | 50     |
| 5. 국내 공공 빅데이터 분석 플랫폼 구축 마스터 플랜(TO-BE) | 4      |
| 가. 현재 국내 공공 오픈 데이터 플랫폼 진단             | 54     |
| 나. 미래 국내 공공 오픈 데이터 플랫폼 제안 (TO-BE)     | 5      |
| <br>제4장 사업 목표 달성도 및 국내 파급효과           | <br>62 |
| <br>제5장 사업 결과의 활용                     | <br>63 |
| <br>참고문헌                              | <br>64 |

---







## 제1장 서론

빅데이터의 의미는 기존 저장 용량을 초과하는 데이터 이상으로 일반적인 도구나 방법으로 수집, 저장, 검색, 분석 등을 하기 어려운 정형 또는 비정형 데이터 세트로 양 (Volume), 속도 (Velocity), 다양성 (Variety) 로 정의되고 있다. 따라서 빅데이터 특성으로는 다음과 같이 분류될 수 있다.

- 1) 기존의 데이터 처리에 요구되는 즉각적인 처리속도와는 달리, 즉각적인 의사결정이 상대적으로 덜 요구된다.
- 2) 다양한 데이터 소스, 복잡한 로직 처리, 대용량 데이터 처리 등으로 인해 처리 복잡도가 매우 높으며, 이를 해결하기 위해 통상적으로 분산 처리 기술이 필요하다.
- 3) 데이터 수집 및 분석을 장기간에 걸쳐 수행해야 하므로 기존 방법과 비교해 처리해야 할 데이터양은 방대하다.
- 4) 소셜 미디어 데이터, 로그 파일, 클릭스트림 데이터, 콜 센터 로그, 통신 CDR 로그 등 비정형 데이터 파일의 비중이 매우 높다. 처리의 복잡성을 증대시키는 요인이기도 하다.
- 5) 처리/분석 유연성이 높다: 잘 정의된 데이터 모델/상관관계/절차 등이 없어, 기존 데이터 처리방법에 비해 처리/분석의 유연성이 높은 편이다. 또한, 새롭고 다양한 처리방법의 수용을 위해, 유연성이 기본적으로 보장돼야 한다.
- 6) 대용량 및 복잡한 처리를 특징으로 하고 있어, 동시에 처리가 필요한 데이터양은 낮다.

따라서 (준)실시간 처리가 보장되어야 하는 데이터 분석에는 적합하지 않다. 컴퓨터, 전자제품, 정보통신 분야에서는 이미 빅데이터 활용을 통한 생산성 증대 및 개선 효과가 나타나고 있으며 공공 분야를 포함한 다른 산업으로도 빅데이터의 경제적 효과에 대한 가능성이 현실로 드러나고 있다.

빅데이터는 지식정보사회에서 천연자원을 대체하는 자원으로써 공공 영역의 정책 활용 방법론 이상의 전략도구로 자리잡고 있다. 국내에서는 공공데이터포털(daa.go.kr)의 개방되고 있는 공공 데이터 수량은 전년 대비 10배수가 넘고 있으며 활용도 또한 대기오염정보, 대중교통정보, 생활기상 정보 등의 실시간 현황을 보여주는 데이터뿐만 아니라 유동인구데이터베이스, 교통사고통계, 주요상권현황 등의 공간, 통계 집계 성의 군집 데이터에 대한 관심이 높아지고 있다.

한국 정부는 2014년까지 10개, 2017년까지 100개 공공데이터의 개방 표준을 확립하고 특정 소프트웨어 보유와 무관한 오픈 포맷으로의 공공 데이터 변환 율을 전체 데이터 개방에서의 과반수 이상을 차지할 수 있도록 진행할 계획에 있다. 공공데이터 개방의 원칙은 자유롭게 이용할 수 있는 권리를 보장하며 누구나 평등하게 접근할 수 있으며 데이터를 활용한 영리적 추구에 대한 보장을 뜻한다.

2011년 10월 31일부터 시행된 공공데이터의 제공 및 이용 활성화 법에 따라 전략적인 개방 추진과 이용자 중심의 개방이 확대되어야 함이 정책 기조의 중요 방향이나 기존의 공공데이터 규격의 표준화, 개인 프라이버시 보호 장치 마련, 공공데이터 품질 개선 등의 공공데이터 마스터플랜 수립에 해결되어야 할 선제 이슈들이 남아 있다. 공공데이터가 개방 데이터로써 빅데이터 특성을 보유하게 되는 시점에서 중요한 과제 중 최상위 이슈는 정제된 데이터 수집 변환 가공을 전제로 한 핵심 가치를 어떻게 도출 할 수 있을 지의 분석 기술 연구와 전략적 방향에 근거한다.

## 제2장 해외 공공 빅데이터

### 1. 해외 공공 빅데이터 활용 현황 (영국)

영국은 빅데이터를 분석 활용한 주요 이슈들 중 국민들의 비만 문제에 주목하고 있다. 빅데이터 분석을 통한 미래예측 프로젝트에서 2050년까지 영국 남성의 60%, 여성의 50% 이상이 비만 신체를 보유하게 되며 연해 85조원이 비만 문제 해결을 위해 소비될 것이라는 전망이 제기되었다.

정부는 비만 문제를 국가 사회적 이슈로 간주했고 이에 대응코자 사전 해결책 마련을 위해 공조 체제를 구축했다. 또한 영국의 지리학적 위치 조건으로 인해 해안 침식이 지속적으로 진행되고 있어 이에 대한 위험 노출을 빅데이터를 통해 예측했으며 홍수 피해 재난에 대해 미리 준비할 수 있도록 기후, 지역, 생태계 등의 관점에서의 다양한 방법론을 통한 과학적 정책 수립을 구성 중에 있다.

### 2. 해외 공공 빅데이터 활용 현황 (싱가포르)

빅데이터를 범 부처차원의 클라우드와 데이터를 기반으로 통합적 중장기 전략 수립 위한 국가미래전략센터로써 싱가포르는 지식정보 과학적 분석 미래전략시스템을 구축, 운영하고 있다.

싱가포르 RAHS(Risk Assessment Horizontal Scanning)프로그램은 빅데이터 기반 국가안보 위협 요소 평가와 주변 환경 탐지를 위해 싱가포르 국가안보조정 사무국(NSCS)에서 담당, 2004년부터 운영되어 왔는데, 테러, 전염병, 금융위기 등 국가적 차원의 위협, 기회 요인을 식별, 대응키 위한 목표를 지니고 있다.

RAHS는 미래 위협에 대한 평가와 기회 요소들을 분석, 예측함으로써 싱가포르 정부의 장기 정책 수립 전략에 유효한 역할을 하고 있으므로 평가된다. 환경탐색

(Horizon Scanning) 부문은 데이터 수집 및 분석 프레임워크 통한 미래 예측에 따른 전략 수립 과정을 제시하고 있다. 공공 데이터 수집 분석을 통한 이슈 분석, 심층 정보 분석 서비스, 기술 동향 서비스, 정책 분석 연계 서비스 등 공공 서비스를 확장 개발해 가고 있다.

1. Target Focus (목표 정의), 2. Environment Scanning (관련 분야 거시적 미시적 환경 분석), 3. Sense Making (변수, 이슈 분석 발견), 4. Developing Possible Futures (가능 미래 시나리오 개발), 5. Designing Strategies (미래 시나리오 전략 도출), 6. Monitoring (변수, 이슈에 대한 지속적인 모니터링 분석)

단계들로 이뤄진 프레임워크는 모든 국가적 위협을 수집 및 분석하여 위협을 선제적으로 관리토록 지원한다. 수집된 위협 정보는 시뮬레이션, 시나리오 기법을 통해 분석, 주요함으로 확인된 사전 위협을 예측하고 해당 위기에 대응하는 전략 수립으로 진행하고 있다. 싱가포르의 총리실 산하의 공공서비스부(Public Service Division, PSD)와 국가안보조정사무국(National Security Coordination Secretariat, NSCS)에서 국가 미래 준비를 담당하고 있다.

기본적으로 영국의 미래전략청을 벤치마킹한 싱가포르의 NSCS는 2004년부터 국가 안전을 위협하는 요소에 대한 평가와 주변 환경 변화를 탐지해 새로운 기회를 발굴하는 RAHS(Risk Assessment Horizon Scanning) 프로그램을 운영하고 있다. RAHS는 싱가포르의 거시/미시적 환경 탐색을 통해 획득한 데이터를 분석해 국가적으로 영향을 미칠 수 있는 잠재적 위협 요소와 불확실성 요소를 탐색해 변화 이슈를 분석한다.

2007년 데이터 분석 실험센터(RAHS Experimentation Centre, REC)를 창설해 새로운 미래연구 방법론과 기술들을 개발, 탐구하고 타국 연구소와 긴밀한 협력 관계를 유지하고 있다. 싱가포르 정부는 이를 통해 해상 상황 인식 프로젝트를

추진해 해상 테러, 해안 침투 등 해안 영역에서의 안전을 확보했다. 또한 조류 독감이 유행할 당시 독감이 싱가포르에 들어온다는 가상 시나리오를 연구하고 대응 방안을 마련한 바 있다.



그림 1. RAHS Programme3

### 3. 해외 공공 빅데이터 활용 현황 (미국)

미국 과학기술정책자문위원회는 2010년 모든 연방정부는 빅데이터 전략 수립이 필요하다고 제시했으며, 이에 오바마 정부는 2012년 3월 빅데이터 연구 개발 수립과 추진에 2억 달러 이상을 투자하는 빅데이터 연구 개발 이니셔티브를 발표했다.

국립과학재단(NSF), 국립보건원(NIH), 국방부(DoD), 고등방위연구계획국(DARPA), 에너지부(DoE), 지질조사원(USGS) 등이 참여한 빅데이터 연구 개발 이니셔티브는 의료, 교육, 지구과학 및 국방 분야 등 빅데이터 활용 효과가 뛰어난 분야의 기관들이 우선적으로 참여하고 있다.

특히 미 국립보건원은 세계 최대의 인류 유전자 변형 데이터 셋을 아마존 웹 서비스 클라우드에서 무료로 공개하고 있다. 이 데이터는 전세계 1,000개 유전체 프로젝트에서 생성된 것으로, 200TB에 달한다. 해당 데이터로부터 지식을 추론

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

하고 데이터 분석 제공 활용 프로그램을 개발, 독려하고자 버클리대 컴퓨팅 원정대 프로젝트에 1,000만 달러의 기금을 투입하였다.

미국 정부는 'Data.gov' 단일 플랫폼을 통해 국가 및 공공부문의 총 65만여개의 데이터세트 공개 및 주요 API를 배포하여 민간 사업자에서 공공정보를 활용한 비즈니스 모델을 구축할 수 있도록 지원한다. 지역 도시들 또한 도시 내부 공공 서비스 개선과 경제 활성화를 위해 공공 데이터 공유와 활용에 주안점을 둔 분석 인프라 구축 및 지원 제도 수립에 활발한 활동을 전개하고 있다. 별개로 민간 파트너십 협력자로서 구글은 'Public Data Explorer' 공개를 통한 오픈 데이터 서비스 비즈니스 모델을 수행 중으로 독감 유행 예측 서비스와 같은 공공 서비스로서의 소스 제공 및 벤치마킹 가이드가 되고 있다.



## 제3장 사업내용

### 1. EU(영국) 공공 빅데이터 인사이트

#### 가. EU(영국) 공공 빅데이터 분석 플랫폼

##### 1) EU 데이터 개방 전략 (Open Data Strategy)

- EU의 데이터 개방 전략 (Open Data Strategy)은 데이터 개방을 통한 접근 뿐만 아니라 재사용 권한을 보장함으로써 데이터 활용을 통한 상업적 서비스 개발을 독려하고 있다. 공공정보 개방포털(PublicData.eu)에서 EU 회원 27개국에서 생산되는 공공정보와 데이터를 의무적으로 공개, 활용토록 하고 있다.

##### 2) 영국 오픈 데이터 정책 (Open Data Policy)

- 영국 정부는 2010년부터 링크드 데이터의 제창자 팀 버너스리(Time Bernerslee)를 주축으로 한 데이터 포털(data.gov.uk) 개설하였다. 해당 플랫폼은 사용자 친화적 인터페이스 레이아웃과 데이터 가시화를 부가하여 영국 정부 서비스 플랫폼으로의 통합 발전 중에 있다. 또한 오픈 공공데이터 공개를 통한 경제적 효과를 고려하여 ivOGL(Open Government License) 도입을 통해 민간 사회에서의 활용을 독려해 왔다.

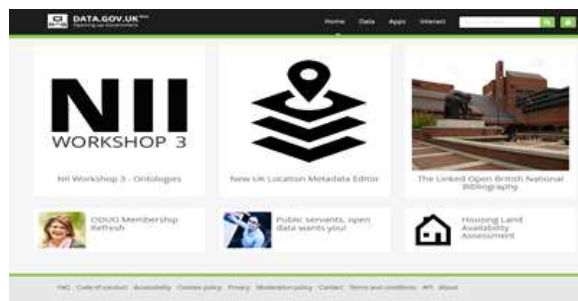


그림 2. 영국 공공데이터 공개 플랫폼 (www.data.gov.uk)

- 영국의 싱크 탱크 그룹 팔러시 익스체인지 (Policy Exchange)가 발간한 보

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

고서 (출처. 2012. 6. 3) 에 따르면 시민 여권, 운전정보, 세금 환급, SNS 이용 등에 따른 빅데이터 활용 정도에 따라 연간 최대 330억 파운드까지의 예산 절감이 가능한 것으로 판단하고 있으며 인구 센서스와 같은 기존 정책 시행에 대한 비용 절감 또한 빅데이터를 통해 이뤄질 수 있음으로 보았다.

- 그림 2.에서와 같이 EU데이터 지표에서도 공유한 바와 같이 영국 정부는 2017년까지 빅데이터 전문가 양성240% 증대를 목표로 한 영국 정보공개법 (Freedom of Information Act 2000)은 공공기관의 모든 소유 기록정보에 대해 국민의 접근법을 보장하며 예외 항목과 공공기관의 의무를 명시한다.
- 2012년에는 기업 혁신 숙련 부를 포함한 17개 부처가 ‘오픈 데이터 전략을 정립하고 의료, 교육, 기상 부문 등의 데이터를 2015년까지 순차적으로 공개할 예정에 있다. 민간에서의 공공 데이터 활용을 촉진하고자 공개 참여 커뮤니티를 운영해오고 있다.

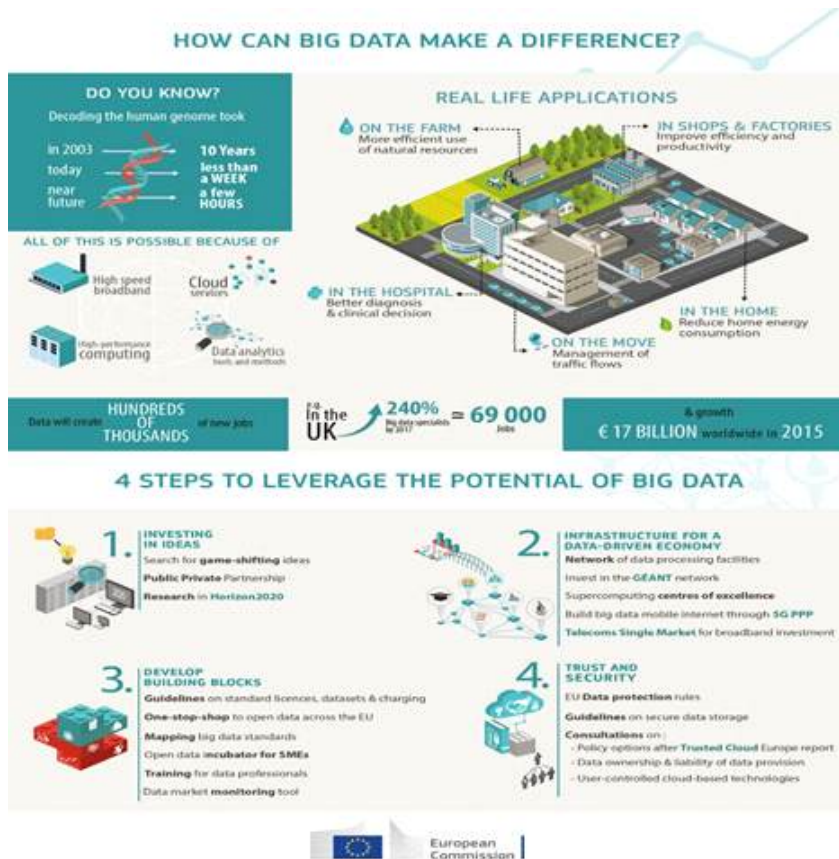


그림 3. EU 빅데이터 청사진 도식

### 3) 영국 호라이즌 스캐닝 센터 (Horizon Scanning Center)

- 데이터 기반 정책 분석 서비스를 통한 정부의 혁신적 미래 정책 개발을 위해 2005년 설립된 영국 호라이즌 스캐닝 센터는 정치, 경제, 사회, 환경, 기술 분야별로 미래 연구 프로젝트를 진행하고 있다. 해당 연구 결과들은 정부 부처들로 적극적으로 반영되고 있으며 관련 전문가 그룹들간의 아이디어 공유 및 실천을 위한 네트워크를 구성하여 정기적으로 FAN Club (Futures Analysts Network Club) 운영 및 보고서를 발간한다. 다양한 연구 활동 중 대표적인 역할은 델타 스캔(Delta Scan) 및 시그마 스캔(Sigma Scan)을 이용하여 데이터를 수집, 분석하는 것이다. 예측 방법론 24 개를 활용한 5단

계 전략 분석 방법 [2]으로 전략개발, 정책, 호라이즌 스캐닝, 전문가 참여, 예측 단계를 통해 미래 예측 전략을 수립, 정책 반영으로 운영하고 있다.

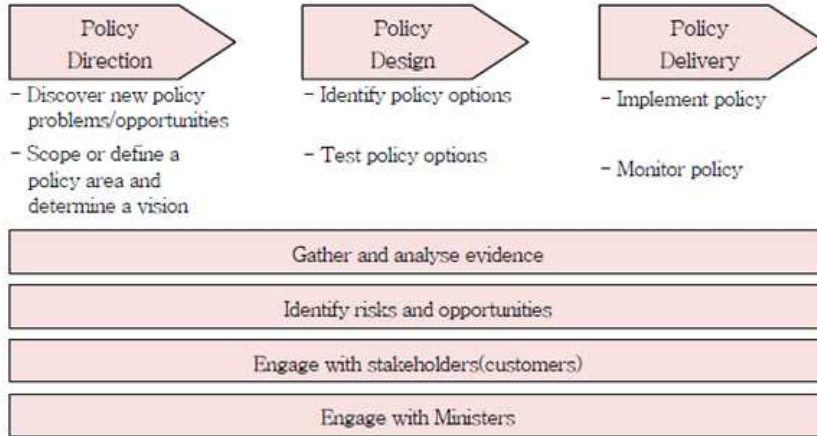


그림 4. 영국 Horizon 정책 수립 단계 (www.foresight.gov.uk)

#### 4) 호라이즌 스캐닝(Horizon Scanning) 단계

- 호라이즌 스캐닝 단계에는 미래예측 범위 지정(Scoping), 정보 수집(Gather information), 신호탐색(Spot Signals), 경향분석(Watch Trends), 이해(Make Sense), 반응동의(Agree the Response)가 있는데, 미래예측 방법론은 전문성과 기간 별로 분류되어 적용되게 된다. 예를 들어 델파이(Delphi) 방법은 전문가와의 협의를 통해 미래를 예측하는 것으로 단기 정책 적용에 해당되며 전문성 수준은 높으므로 호라이즌 스캐닝 에서의 신호 탐색 단계에서 이를 사용하여 결과를 파악, 분석할 수 있다. 최근 공개된 “Future Toolkit[3]”에서 정책 수립 프로젝트 사이클에서 미래 예측기법에 대한 적용 방안이 제시되었다. ROAMEF 정책 사이클은 끊임없는 단계 (요구사항 확인-목적 정의-평가-모니터링 수행-평가-피드백) 반복을 통해 지속적으로 완성되어 가며 각 단계별 적용되는 미래예측 분석 기술들은 개요, 수행방법, 적용 시점, 예측 결과 등으로 분류되어 소개된다.

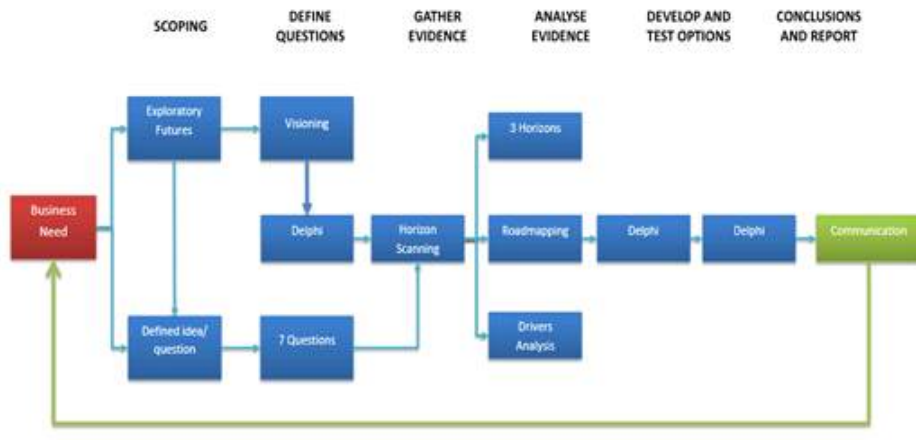


그림 5. RAHS 프로젝트 라이프사이클의 미래예측 기술

## 나. EU(영국) 공공 빅데이터 분석 기술 현황

### 1) 미래 도구 분석 주요 기법 소개

#### 가) 호라이즌 스캐닝 (Horizon Scanning)

- > 호라이즌 스캐닝은 근거 기반으로 이슈를 분석 평가하여 이를 우선순위화, 의사결정을 지원하는 방법으로 프로젝트 초기에 일반적으로 활용된다. 해당 구현에 있어 STEEPLE (Society, Technology, Environment, Economics, Politics, Legal, Ethics) 프레임워크가 추천되는데 단기, 중기, 장기로 구분, 관련 선택 요소들에 따른 이슈 매트릭스로 구성하여 평가 분석할 수 있다. 경우에 따라 군사, 인구 통계와 같은 요소들을 추가할 수 있으며 삭제 가능하다. 평가 이전에 비용, 지리적 요건 등을 통해 구별하여 이슈들을 우선순위 할 수 있으며 분석 결과 표현 시에 시간 별 대의 이슈 중요도 표현 방법인 호라이즌 3 사용이 유용하다. 호라이즌 스캐닝은 미래 이슈들을 탐색하는 데에 필요한 근거와 이슈에 대응하는 관리 방안들을 제시한다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

표 1. STEEPLE 템플릿 예시

|             | 0-2 Years | 2-5 Years | 5+ Years |
|-------------|-----------|-----------|----------|
| Society     |           |           |          |
| Technology  |           |           |          |
| Environment |           |           |          |
| Economics   |           |           |          |
| Politics    |           |           |          |
| Legal       |           |           |          |
| Ethical     |           |           |          |

나) 로드매핑 (Roadmapping)

- 로드매핑은 미래 예측되는 중요 이슈 선정 뒤 영향을 줄 수 있는 가능성 있는 이벤트들의 시간적 경로와 예상 방향을 그림으로써 전문가들의 협업을 통해 프로젝트 초기 단계에 진행된다. 로드매핑은 근거 기반 모델로 화이트 보드, 워크샵 등을 통해 구성되게 되며 시나리오 프로세스에 유용한 도구가 될 수 있다.

다) 시나리오 (Scenarios)

- 미래에 이뤄질 가능성이 높은 복수 시나리오를 구현하여 전개 과정을 추정하는 시나리오 기법은 미래에서의 실현 가능성 있는 잠재 개발들에 있어서 장기적 의사결정 또는 정책을 반영토록 하고 있다. 수개월 간에 이뤄지는 여러 워크샵 운용을 통해 이해관계자들과 연결, 미래 나리오 구성을 진행한다. 동인 발견부터 총체적 시나리오 수렴까지 전제된 워크샵 활동에서 중간 선택 개념이 없는 2X2 매트릭스, 비전문가들의 부가적인 조사/인터뷰, 시나리오 테스트 활용 등을 통해 시나리오 기법은 의사결정을 위한 전략적이고 효과적인 도구로 인식된다.

## 2) 호라이즌 스캐닝 시스템

### 가) 호라이즌 스캐닝 (Horizon Scanning) 분석 구성 요소

- 호라이즌 스캐닝 시스템은 목적과 예상 결과를 기술하며 관련 지원과 참여자들간의 조직 운영 구조에 대해 구분하고 있다. 호라이즌 스캐닝 시스템의 기능은 전담 조직 구성에 영향을 주며 모니터링, 이슈 발견과 준비 운영에 변수로 작용한다. 미래 예측 영향을 측정하기 위한 분석 지표 구성을 위해서 경우 수에 따라 시스템 운영 구조와 진행 방식을 결정하는 것이다.

|                               |                       | Choices for Parameter Values                            |   |   |                                  |  |
|-------------------------------|-----------------------|---|---|---|----------------------------------|--|
| HS System Function Parameters | Purpose               | Scan for all national policy fields                     | Single-purpose scan for all national security fields    | HS specific to national security strategy framework                     |                                  |  |
|                               | General method        | Active  | Passive   |   |                                  |  |
|                               | Reporting frequency   | Bi-yearly major report                                  | Yearly report with additional short reports             | Frequent short reports  | Major report timed to meet needs | One-off effort                         |
|                               | Sponsor               | Ministerial   | Secretary-level   | Deputy Secretary-level  | Below Deputy Secretary           | Board including non-government members |
|                               | Release policy        | All reports released publicly                           | Selected (declassified?) reports made public            | Reports released only to a closed network                               | All product classified           |  |
| HS System Form Parameters     | Type of product       | Synthesised scan reports for the entire area of concern | Multi-subject reports (combine two or more topic areas) | Single-topic reports (usually on a specific but small area of interest) |                                  |  |
|                               | Scanning participants | Intelligence only                                       | All government only                                     | Government and non-government participants                              |                                  |  |
|                               | Team structure        | 'Centre of Excellence'                                  | Dedicated team  | Coordinating core   | Individual coordinator           | Virtual team (network)                 |
|                               | Outreach              | Public conference                                       | Closed network conference and seminars                  | Classified conference and seminars                                      | No conference or seminars        | Briefings for key meetings             |
|                               | Aftercare             | Dedicated team  | Follow-up by individuals                                | Nil   | Pull by users                    |  |

### 3) LOD (Linking Open Data) 연계 플랫폼

- ▶ LOD는 W3C 표준기술을 활용한 데이터 공개 공유 기법인 링크드 데이터 (Linked Data)와 재이용이 가능한 데이터 개방 전략을 의미하는 오픈 데이터(Open Data)의 합성어로서 시맨틱 중심 데이터 웹(Data Web)을 지향하며 RDF(Resource Data Format), OWL(Ontology Web Language)를 통해 표현된다. 링크드 데이터는 웹에서의 데이터 자원을 노출(expose), 공유(share), 연결(connect)하는 기법으로 시맨틱 웹<sup>1)</sup> 환경에서의 명시적 의미(explicit meaning)로 연결된 구조화된 데이터를 뜻한다.

링크드 데이터에 대한 의미적 연결을 위해서는 도메인을 설정하고 데이터 간의 논리적 관계, 즉 온톨로지를 정의하여 표준 형식의 데이터로 표현되어 REST(Representational State Transfer) 프로토콜 기반 URI(Uniform Resource Identifier)로의 식별 체계를 통해 웹에서 공유, 연결된다. 즉, 데이터 간의 의미 있는 연결을 생성하므로 웹에서의 데이터를 공유, 연결, 재사용하는 가치에 부합, 데이터 개방과 활용에 관련된 국가별 오픈 데이터 정책 추진에 범용적으로 적용되는 추세에 있다. 민간 부문에서의 DBPedia는 전세계 사용자 참여 구조의 데이터 백과사전 플랫폼인 위키피디아에 링크드 데이터 기술을 적용한 개방형 데이터 플랫폼으로 데이터웹 형태의 실시간 데이터 관리를 지원한다.

---

1) 기계가 정보의 의미를 이해하고 처리하는 정보의 공간:

[http://ko.wikipedia.org/wiki/%EC%8B%9C%EB%A7%A8%ED%8B%B1\\_%EC%9B%B9](http://ko.wikipedia.org/wiki/%EC%8B%9C%EB%A7%A8%ED%8B%B1_%EC%9B%B9)



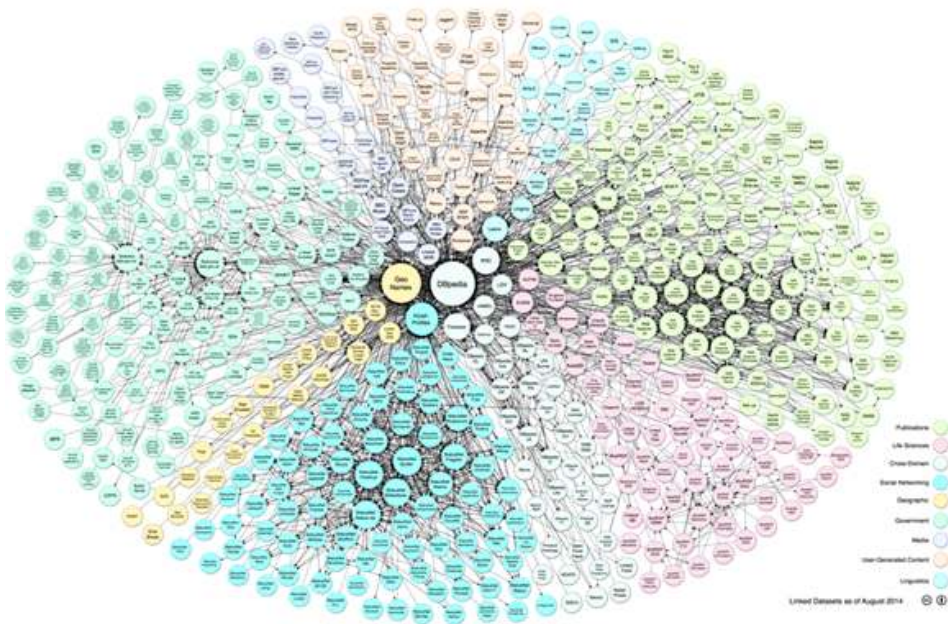


그림 6. 2014년 8월 공개 LOD 클라우드 다이어그램  
( <http://www.lod-cloud.net> )

#### ▶ 링크드 공공 데이터 분석 플랫폼

영국 정부의 data.gov.uk를 중심으로 한 데이터 개방 및 활용을 위한 링크드 데이터 플랫폼 운영은 초기 www.창시자 팀 버너스리 주도로 정보의 의미적 연결을 지원하는 LOD 확산과 함께 오픈 데이터 품질, 재이용성 평가로 5스타(★) 전개 구조를 제안, LOD 형식 개방 적용을 진행해 왔다. LOD 개방 도메인 선정, 로드맵 수립 및 데이터 온톨로지 구축으로 LOD 인식을 넓혔고 현재 LOD 생태계 구축을 위해 온톨로지 간 연계 플랫폼 구축 및 애플리케이션 활용 방안 계획 수립 및 관련 글로벌 표준 정책 참여와 전략 기반 마련을 점진적으로 진행하고 있다. 영국 정부는 오픈 데이터 공개를 위해 ‘5스타 오픈 데이터’ 전개 구조를 적용, 데이터의 가독성과 재이용성을 고려하여 제공할 수 있도록 최소 ‘3 스타’ 인증을 요구하고 있다. ‘5스타 오픈데이터’ 구조의 데이터 개방 형식은 사람과

기계가 읽을 수 있는 데이터 변환을 토대로 이를 LOD화하여 개방, 배포, 공유와 재활용을 통한 지원을 목표로 한다.

1 스타는 특정 포맷에 관계없이 보기, 출력, 저장, 입력, 공유 가능하며 출판 용이성과 공개 편의성을 제공한다. 2 스타는 구조화된 데이터로 소프트웨어 통한 처리와 포맷 변환이 가능하며 역시 출판 용이성을 보여준다. 3 스타는 비독점 포맷을 사용하며 특정 소프트웨어 종속 없이 데이터 처리가 가능하며 데이터 추출 위한 변환기 또는 플러그인이 필요할 수 있다. 4 스타는 링크 연결이 가능하며 URI 동일 시에 데이터 결합이 가능하다. 더불어 데이터 접근에 대해 최적화(로드 밸런싱, 캐싱) 및 타 공급자의 데이터와 상호 데이터 링크 연결이 될 수 있음을 뜻한다. 끝으로 5 스타는 문맥 연결로 데이터끼리 이어지며 모든 데이터 처리가 가능하며 데이터 구조 정제, 데이터 가치를 증대할 수 있다.

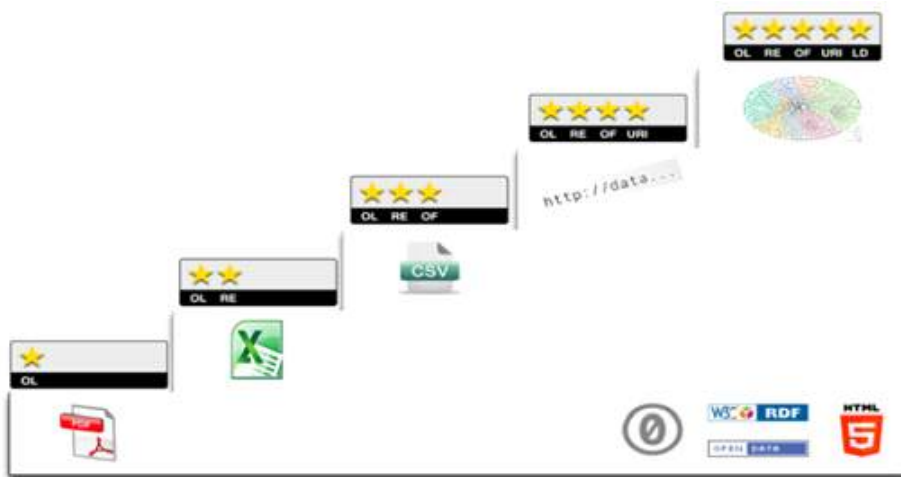


그림. 7. 스타 오픈데이터' 구조 데이터 개방 형식 ( <http://5stardata.info> )

정부는 데이터 개방을 위해 개방 데이터 위원회 (Open Data Institute) 의 개방 데이터 인증 (Open Data Certificates)을 요구하며 개방 데이터 정책은 하기와 규칙 사항으로 이를 뒷받침 한다. 현재 [data.gov.uk](http://data.gov.uk)는

CKAN과 같은 오픈소스 데이터 개방 플랫폼을 기반으로 소스 코드 공개와 더불어 누구나 동일한 데이터를 활용할 수 있도록 데이터 분석 시각화 및 메타 데이터 등을 포함한 정보 카탈로그를 개선해가고 있다.

▷ CKAN (Comprehensive Knowledge Archive Network) 플랫폼

비영리 기관인 OKF(Open Knowledge Foundation)에서 개발한 공공 데이터 개방 오픈 소스 플랫폼으로 영국뿐만 아니라 해외 국가의 공공 부문 오픈 데이터 개방과 재사용 활용을 지원해오며 메타데이터 검색 카탈로그, 데이터 저장소 및 관리 솔루션을 탑재하고 있다. 기존, 신규 데이터를 개방 데이터 표준 포맷인 RDF화 하여 공개 또는 CKAN에 등록 시 LOD 클라우드에 연결, 공개자 노트 생성이 되면서 저장소에 보관된다. LOD는 URI로 식별되며 ‘5스타 오픈 데이터’ 구조에 따라 평가된 웹 데이터 저장소에서 SPARQL Endpoint를 통한 동적 검색, 질의, 수정, 다운로드 등 글로벌 데이터베이스처럼 자유롭게 활용할 수 있게 하며 데이터 통합을 용이하게 해준다. 또한 내부, 외부 데이터 셋에서 정형화된 메타데이터 카탈로그를 생성하고 구글 분석도구(Google Analytics)와 연동하여 데이터 셋 관련 통계 데이터와 함께 분석 리포트를 보여준다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

표 3. 공공데이터

| 번호 | 설명  |
|----|---|
| 1  | 공개 데이터 정책과 사례들은 데이터 배포 시점과 형식을 포함하여 데이터 사용을 원하는 대중과 비즈니스에서 도출되어야 한다.                  |
| 2  | 공공 데이터는 재사용 가능 해야 하며 기계가 읽을 수 있는 형식이어야 한다.  |
| 3  | 공공 데이터는 상업적 활용을 포함하여 자유로이 재활용할 수 있는 오픈 라이선스 하에 배포된다.                                  |
| 4  | 공공 데이터는 (data.go.uk) 온라인 단일 채널을 통해 쉽게 찾을 수 있어야 한다.                                    |
| 5  | 공공 데이터는 개방 기준 및 W3C 관련 규정에 맞게 출판된다.   |
| 6  | 동일한 주제에 대한 타 부서들간의 공공 데이터는 표준 포맷과 함께 동일한 내용이어야 한다.                                    |
| 7  | 정부 소유 사이트에서의 공공 데이터는 재활용 가능 형식으로 출판된다.  |
| 8  | 공공 데이터는 시기 적절한 정제된 내용이어야 한다.  |
| 9  | 배포 데이터는 링크드 데이터 형식을 포함한 개방 표준 포맷에 적합해야 한다.  |
| 10 | 공공 데이터는 적법한 범위 내에 사용 가능하다.  |
| 11 | 공공 데이터는 애플리케이션 또는 별도의 등록, 사용자 정보 없이도 접근이 가능하다.  |
| 12 | 공공 기관들은 공공 데이터의 재사용-용에 대해 적극적으로 권장해야 한다.  |
| 13 | 공공 기관들은 데이터 소장과 관리에 충실히 이행한다.   |
| 14 | 공공 기관들은 데이터 셋과 관련된 메타데이터를 출판해야 하며 이는 단일 채널에서 이뤄지며 데이터 포맷 출처와 의미에 대한 설명 또한 함께 공개해야 한다. |

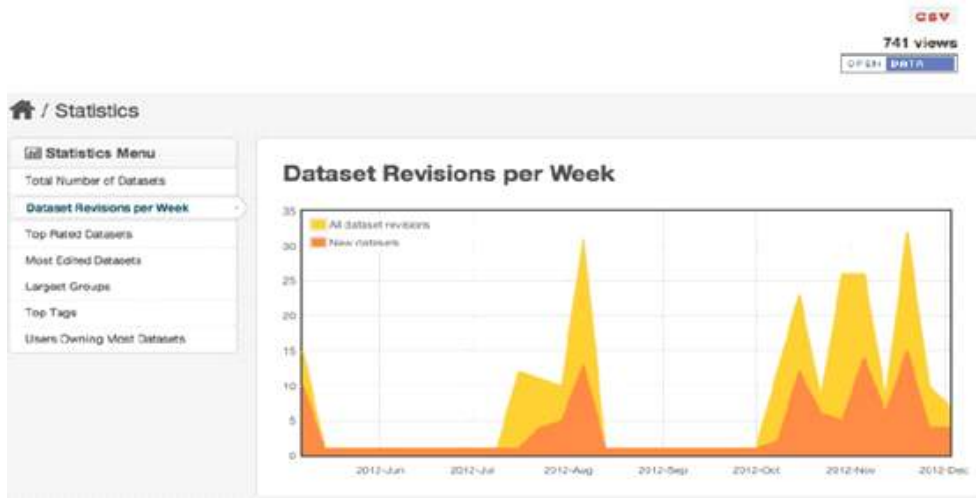


그림 8. CKAN 분석 통계 리포트 예제 (<http://theodi.org> )

CMS(Contents Management System)와의 연동 및 검색, 분석, 저장 처리 등의 적용 오픈 소스들을 GitHub라는 세계적 개발 커뮤니티에 배포, 개인 개발자들이 자원하여 소스들을 업그레이드하고 있다. 범용 CMS솔루션 중의 하나인 드루팔(Drupal) 기반 CKAN 버전인 DKAN은 API 호환성 보장, 손쉬운 설치, 카탈로그 저장, 데이터 맵핑 분석 / 시각화 지원, 커뮤니티 관리 등의 여러 장점들을 제공하나 파이썬(Python) 프로그래밍 언어로 이루어진 CKAN의 복잡한 자체 개발 환경, 소프트웨어 종속 관리 프로세스 등의 한계점이 있다.

## 다. EU 공공 빅데이터 분석 활용 사례

### 1) 소개

#### 가) 호라이즌 스캐닝 (Horizon Scanning) 활용 사례

- > 2008년부터 2009년까지 영국이 속한 EU에서 유럽 2020 전략 수립을 구현하는 과정에서 JRC-IPTS(EC Joint Research Centre, Institute for Prospective Technological Studies)와 BEPA(Bureau of European Policy Advisers)에서는 호라이즌 스캐닝 분석법을 적용하여 2025년까지

## > 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

예측할 수 있는 종합적인 미래 이슈 확인 분석을 통해 포트폴리오 모델링 스크리닝 프로세스 (RPM)를 제안하였다. 인구통계, 경제, 환경, 연구, 정부와 사회, 국방 안보의 상위 6개 분류화를 진행하였고 외/내부의 전문가들의 개방적인 참여로 수집된 129개 리포트들은 초기 그룹내부의 참여자 개인판단에 선별되고 양방향 집단 토의를 통해 중간 보고서 형태가 나오게 되었다. 별도로 트렌드와 와일드카드 즉, 예치기 못한 이슈간의 연결 관계에 대해서도 전문가들의 암묵적 지식을 체계적으로 정리할 수 있게 했다. 이러한 프로세스를 통해 370여개의 분류화된 추적 가능한 이슈들이 확인 기술되었다.

### > 이슈 평가

RPM 프레임워크를 통해 (Liesio et al. 2007; Konnola et al. 2007) 개별 이슈들은 가중치, 빈도 순으로 점수가 측정되는데 상대 평가를 위해 3가지 분석기법이 활용되었다. 대다수로부터 적절함과 가능성으로 관련성의 평가 가중치로 측정되는 평균 기준 분석법, 특정 이슈에 대한 평가 차이 변동성을 중점으로 한 변동 기준 분석법, 불가능하나 타당하며 영향력이 높은 이슈들에 대한 희소 사건 기반 분석법으로 나뉘어진다. 다음으로 EC의 이사회 대표진, 6개 상위 영역에서의 각 국제 전문가들 등으로 추가 그룹이 결성되었고 RPM 분석에서 진단된 각 분석을 통해 도출된 상위 이슈들에 대한 재평가 및 3개 이상의 이슈들을 결합 교차 시 위기로 촉발될 가능성을 우선하여 의견을 수렴하였다. 표 2는 인구통계, 환경, 국방 안보 주제에서 공통적으로 진단된 상위 교차 이슈들을 보여준다.

표 2. 인구통계, 환경, 국방 안보 부문 상위 교차 이슈  
( [http://foresight.jrc.ec.europa.eu/survey\\_issues.pdf](http://foresight.jrc.ec.europa.eu/survey_issues.pdf) )

| Area No.  | Issue code | Key words from issue description (optional)  |
|---|------------|--|
| Save natural resources (water, food) to prevent conflicts over their scarcity and other impacts such as migration |            |  |
| 1   | DI04       | Massive migration due to climate change  |
| 3   | ENV03      | Global under-pricing and overconsumption of water                                  |
| 3   | ENV68      | Global decline of freshwater availability leading to an increase in water scarcity |
| 3   | ENV70      | Global decline in biodiversity and loss of ecosystems services                     |
| 6   | DS13       | Attacks on infrastructure facilities   |
| 6   | DS15       | A major war by 2020  |
| 6   | DS81       | Pervasive sensors for real-time surveillance widely diffused                       |

#### 나) 공공 링크드 데이터 적용 서비스 사례

- 영국 정부는 기본적으로 공공 데이터 활용을 통한 경제적 부가가치를 높이 고자 '5스타 오픈 데이터' 기반 공공 데이터 LOD 프로젝트를 추진하여 관련 표준 제정 및 공개 범위를 넓혀 왔다. EU 11개국과 한국 기관(KAIST)에 서 참여, 2014년까지 진행되고 있는 LOD2 프로젝트는 상호 연계된 데이터 로부터 지식을 생성하기 위한 지능형 정보 관리(Intelligent Information Management for Creating Knowledge out of Interlinked Data)를 목표로 한다. 데이터 출판자와 사용자의 LOD 사용 촉진, 개방 데이터 품질 관리, 관계형 데이터에서 링크드 데이터로의 전환을 위해 데이터 생명 주기에 따른 각종 연구 과제들로 이루어져 있다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

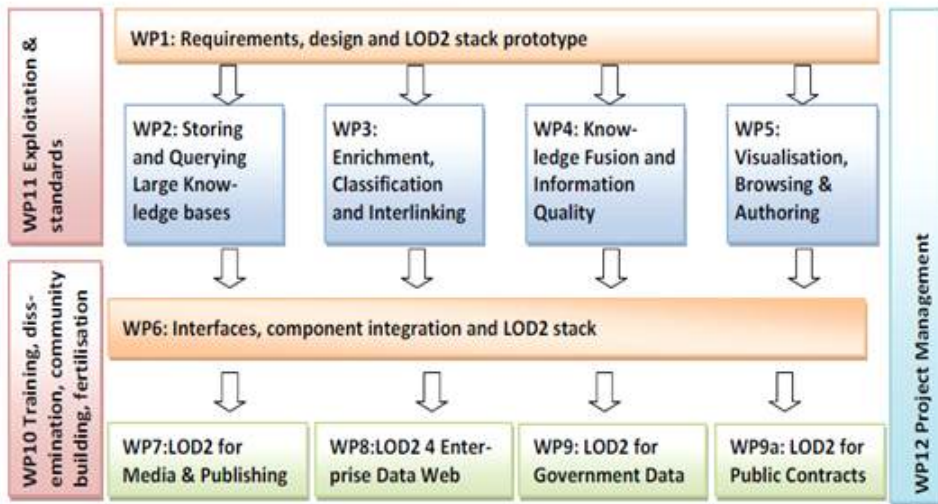


그림 9. LOD2 프로젝트 범위 ( <http://lod2.edu> )

- > 프로젝트는 총 WP12 모듈로 구성되어 있는데 WP1은 요구사항 설계 및 LOD2 스택 프로토타이핑을 진행하며 WP2는 RDF 데이터 대형 저장소 개발, WP3은 링크드 데이터의 상위 온톨로지와의 연결 및 데이터 간 결합 지원 및 데이터 추출/보완/복원 기술 개발, WP4는 링크드 데이터 품질 평가와 맵핑 기반 데이터 통합, W5는 링크드 데이터 탐색, 시각화 및 저작 도구 개발에 대한 프로젝트 연구 주제를 포함한다. 이어 WP6에서는 W7~W9a에 걸친 유스 케이스를 기반한 LOD2 스택 개발을 진행하는데 사용자 인터페이스, 스택 API 구성요소들, 평가, 문서화, 튜토리얼로 구현된다. WP7에서는 데이터 출판자를 위한 LOD2, 외/내부 기업용 링크드 데이터 웹의 활용을 위한 연구를 의미하는 WP8, 개방 정부 데이터 유스 케이스 연구인 WP9, 별개로 공공 정부계약 표준 온톨로지 제안 WP9a, 오픈 소스 소프트웨어 공개, 링크드 데이터 활용 교육, 마케팅 및 표준 제정을 뜻하는 W10~W11, 순환적인 프로젝트 지속 관리 W12로써 프로젝트 주기와 세부활동 상위 항목을 정의하고 있다.



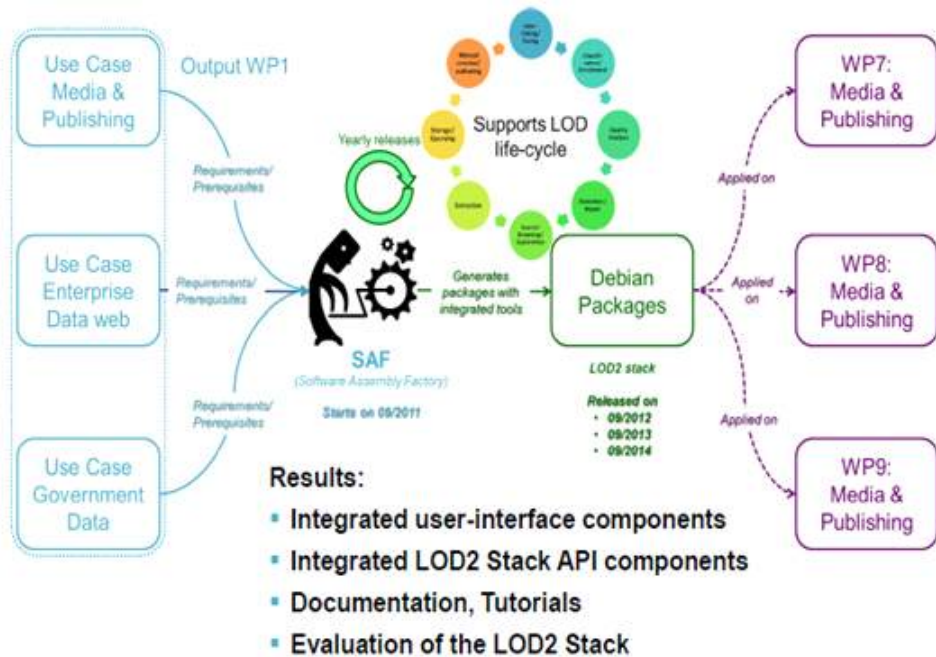


그림 10. LOD2 프로젝트 W6 단계 ( <http://lod2.edu> )

‣ Data.go.uk/apps 에서는 공공 링크드 데이터를 분석 활용, 앱 개발을 장려, 성공 사례들을 보여준다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

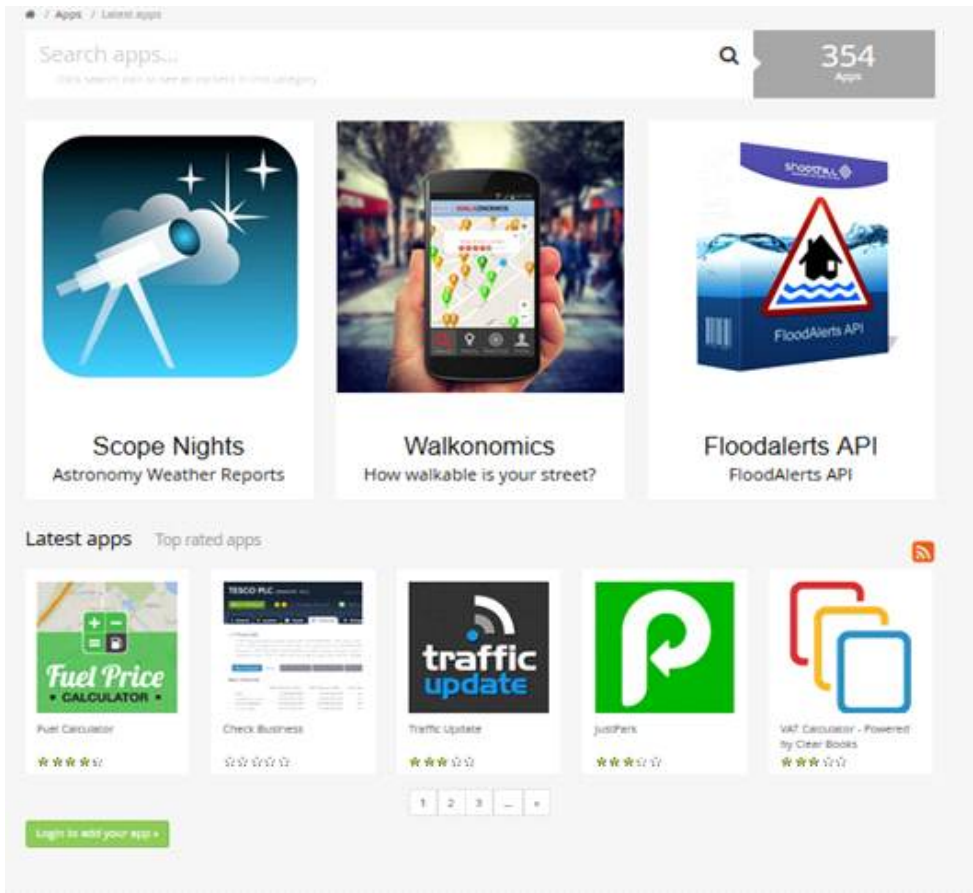


그림 11. 공공 데이터 활용 애플리케이션 목록 (<http://data.gov.uk/apps>)

- Scope Nights 모바일 애플리케이션은 천체 망원경으로 하늘을 관찰하기 적당한 순간을 찾기 위해 3시간 단위로 영국 기상청에서 제공하는 데이터 포인트 서비스, 공개 데이터를 통해 해가 지기 전 청정한 밤 하늘을 볼 수 있는 시간대를 예측할 수 있게 해준다. 기상청 데이터 포인트 서비스에서는 강우량, 구름, 온도 등의 실제 관측 데이터 기반 예측 모델을 생성한 뒤 재사용가능 한 데이터 포맷 처리한 부문은 [data.go.uk](http://data.go.uk), 정보 카탈로그로 전달한다.

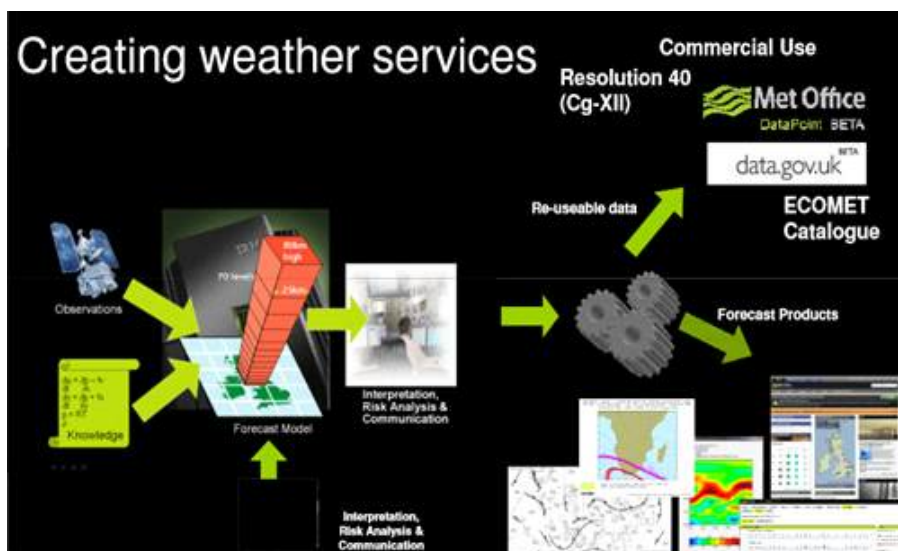


그림 12. 기상 예측 데이터 처리 전개도 (<http://www.agi.org.uk/> )

- > Walkonomics는 커뮤니티 맵핑을 통해 걷고 있는 거리가 위생, 접근성, 디자인, 치안 등에 대한 사람들의 평가로 등급이 매겨져 거리 지도를 채워가는 방식의 애플리케이션이다. 공공 오픈 링크드 데이터와 클라우드 소싱이 결합된 모델인데 우편번호를 통해 거리를 찾아볼 수 있게 하여 실제 사람들의 의견을 수렴하여 거리 관리 정책에 활용할 수 있다. Floodalerts API는 강 또는 바다에서 범람을 사전에 예측하고 지속적인 관측소 모니터링을 위해 제공된다. 15분마다 업데이트 되는 환경청에서 제공하는 공공 데이터에 연결이 가능하며 범람 경보 정보, 위험 영역 표시 등에 대해 웹으로 표시 가능한 파일 포맷으로 변환시킬 수 있다.
- > 영국이 소속된 EU R&D 범국가적 프로젝트들에도 대규모 데이터 분석을 통한 시뮬레이션 모델, 저장 인프라 구축 등에 있어 공공 데이터 분석 플랫폼을 통한 효율성을 높이하고자 도입 구축 시나리오 개발 또는 적용 계획, 파일럿 테스트를 진행하고 있다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

표 3. 공공 프로젝트

| 프로젝트 명                        | R&D 개발 목표                            |
|-------------------------------|--------------------------------------|
| Arcomem<br>arcomem.eu         | 소셜 웹의 지능형 아카이빙 : 클라우드 소싱             |
| CHESS<br>chessexperience.eu   | 박물관의 맞춤형 스토리텔링 가이드                   |
| CONFES<br>confes.gr           | 리소스 최적화와 유연한 서비스 프로비저닝이 가능한 융합형 네트워크 |
| EFG1914<br>project.efg1914.eu | 1차 세계대전과 관련된 영화/그 외 미디어 매체의 디지털화 복원  |
| HBP<br>humanbrainproject.eu   | 인류 뇌 프로젝트                            |
| iMarine<br>i-marine.eu        | 해양 자원과 어장 관리를 위한 데이터 인프라             |
| INTER-ACT<br>eu-interact.org  | 북극 지형 연구 및 모니터링을 위한 국제 네트워크          |
| MERMAID<br>mermaidproject.eu  | 파동/풍력 기반 오프쇼어 에너지 플랫폼                |
| OpenAIRE<br>openaire.eu       | EU 오픈 액세스 플랫폼                        |
| Optique<br>optique-project.eu | 온톨로지 기반 고가용 병렬 데이터 플로우 처리            |

## 2. 싱가포르 공공 빅데이터 인사이트

### 가. 싱가포르 공공 빅데이터 분석 플랫폼

싱가포르 정부 미래 연구 센터 RAHS 프로그램

- 싱가포르 정부는 2004년 7월 대내/외적 국가 안보 전략 계획을 재구성하면서 미래의 위험요소와 이슈를 조기 발견, 대응할 수 있도록 위험 평가(Risk Assessment) 및 미래이슈분석(Horizon Scanning)에 집중한 RAHS (Risk Assessment and Horizon Scanning) 프로그램을 발족하였다. 이를 수행하기 위한 기관으로 HSC(Horizon Scanning Center)와 RAHS Experimentation center(REC)를 설립, 국가 안보 조정 사무국 산하로 배치, 운영하고 있다. 온라인 정형 또는 비정형 데이터에 대해 광범위한 정보 수집, 분석, 제공, 모델링, 모니터링을 걸쳐 SKAN-to-Trend Process, Policy Game 등의 고유한 분석방법을 활용한 정성, 정량적 측정 결과에 대해 가독성을 높인 방법으로 표현하여 국가 정책의 의사결정을 지원하고 있다.



그림 13. Structure of policy coordination in the NSCS, (Source: own picture based on Singapore National Security Coordination Secretariat  
: <http://app.nscs.gov.sg/public/content.aspx?sid=138> )

가) 호라이즌 스캐닝 센터 (Horizon Scanning Centre) 시스템

- ▶ 기존 영국의 호라이즌 스캐닝 센터를 벤치마킹하여 세워진 호라이즌 스캐닝 센터는 서비스 지향 아키텍처 (Service Oriented Architecture) 기반하에 정보 공유와 프로세스 처리가 이루어지므로 각기 다른 부서들 간의 데이터 교환, 웹 서비스로 활용할 수 있다. 해당 시스템은 상호연동 커뮤니케이션, 처리분석 저장소, 시각화를 특징으로 구축되었으며 더불어 데이터 구조화 서비스는 일관적인 매트릭스와 연계된 시스템 맵 구축 및 형태소 분석까지 가능하게 한다. 시스템적 사고와 복잡성 분석과 같은 개념들까지 포함하는 데 사용자가 손쉽게 쓸 수 있도록 편의성을 제공하는 것이 현재 해결해야 할 과제중의 하나이다.
- ▶ 시스템은 크게 개방/폐쇄적 네트워크로 나누어 활용하고 있으며 RAHS를 독립적 포털 시스템으로 두는 대신 개별 포털은 RAHS 결과를 확장하는 방식으로 운영된다. 첫째, 싱가포르 대학들은 RAHS 시스템을 사용하여 호라이즌 스캐닝 개념과 기법들을 익히고 이에 대해 피드백을 줌으로써 RAHS의 정치, 사회, 경제와 같은 영역들에 대한 연구 모델을 세울 수 있다. 둘째, 확장 프로그램들은 지식 발전 기여를 위해 특정 영역에서의 민간 전문가들과의 협업을 통해 발전된다. 셋째, OneMap(하나의 지도) 개념 하에 모든 정부 부처들의 지형공간 정보를 연계 사회 서비스와 맵핑, 대중과 함께 공유하면서 피드백을 통한 플랫폼 개선, 공공 데이터 API 이용에 따른 부가가치 서비스 창출을 유도한다.

## 나. 싱가포르 공공 빅데이터 분석 기술 현황

### RAHS 분석 모형 기법

- ▶ RAHS 프로그램은 미래 예측 기법인 시나리오 플래닝(Scenario Planning)을 보완하기 위해 데이터 이슈 수집 후 이를 분석 지원하기 위한 방법론과 도구를 개발시켜 왔다. RAHS 프로그램은 데이터 이슈간의 상호 연관성과 미래적 함의를 도출 하고자 버전 업그레이드를 통해 시스템을 개선해 왔으며 이 중 핵심 기능들로 시나리오 빌더와 랭킹 빌더가 있다. 시스템 구성 요소

보다는 전체적인 시각과 상호 연계성을 중점으로 둔 시스템적 사고 구조의 시나리오 빌더는 시스템 변수들간의 상호 영향과 관계성을 보여준다.

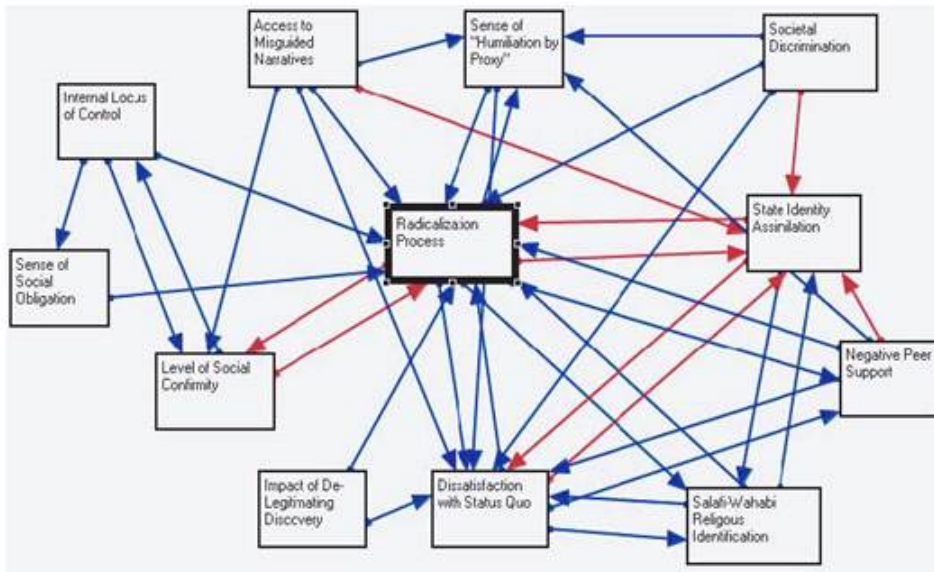


그림 14. RAHS 시스템 모델 (자기급진화 프로세스) : 파란색 화살표는 같은 방향의 관계를 의미하며 빨간색 화살표는 정반대 방향의 관계를 표시한다.

- 랭킹 빌더는 랭킹 수직 구조 트리에서 사용자 정의 상황을 평가하는 플랫폼으로 명시된 규칙에 기반하여 노드들이 생성되고 가중치가 매겨진다. 속성으로 특정 이슈에 대한 조기 경고 지표를 나타내며 현장 데이터들은 가중치 알고리즘을 따라 해석, 평가, 인덱스 되면서 최종 인덱스로 높은 가중치를 보인 데이터가 전문가 그룹의 견해에 따라 일종의 경고 신호로 판단될 수 있다.
- RAHS 시스템은 여러 형태의 분석 기술 기법들을 적용하고 있는데 대표적인 방법론들로 Search (검색), Clustering(군집화), Duplicate Detection(중복 발견), Summarization(요약), Keyword Analysis(키워드 분석), Entity Analysis(개체 분석), Entity Network(개체 네트워크), Timeline Analysis(시

계열 분석), Question and Answering(질의 응답)을 들 수 있다. 먼저 Search는 글자 형태 구조와 패턴에 기반하여 검색 키워드를 확장하는 방식으로 Boolean 연산자를 활용할 수 있다. Clustering은 메타데이터와 같은 주제를 추출하는 방법으로 유사 주제들로 그룹화 된 문서들에 한해 이슈 별로 분류가 가능하다.

‣ Duplicate Detection은 단어들간의 상호 연관성에 기반하여 중복된 문서들을 찾을 수 있는 방법으로 “유사성 한계점” 사용자 정의 매개변수를 활용한다. Summarization은 여러 프로세스에서 사용되는데 많은 정황과 사실들을 분석해야 하는 시점에서 중요 문장들을 정리, 미리 보기를 제공한다. 다음으로 Keyword Analysis는 데이터에 명시된 키워드들을 정의하여, 문제를 해결하기 위한 전제로 구조화시킨다. Entity Analysis는 메타데이터 등의 개체들의 빈도 및 시간 별 활동 분석 시각화를 통해 이슈 변수들간의 관계를 증명할 수 있다. Entity Network 또한 개체들간의 관계를 네트워크화하여 노드와 링크로 연결, 표현함으로써 보다 수월한 개체분석이 이뤄진다.

‣ Timeline Analysis는 사용자 정의 키워드와 시간 순서대로 이벤트들이 나열되어 주기적 현상 분석 또는 일정 흐름을 찾는 데에 단서를 보여준다. 끝으로 Question and Answering은 사실 기반 질문 응답 모듈을 보유한 시스템에서 질문 시도를 통해 적절한 문장을 보유한 데이터, 개체를 찾아 내는 단계 경로를 경험할 수 있다. 아래 그림은 “Middleland Freedom Fighter”를 찾기 위해 시스템 자동화 또는 수동으로 자동 개체 추출, 자동 질의 응답, 요약, 횡단면 분석, 개체 네트워크 분석 등을 통해 이슈의 단서가 될 수 있는 데이터를 추려서 문제를 풀어가는 과정을 설명하고 있다.



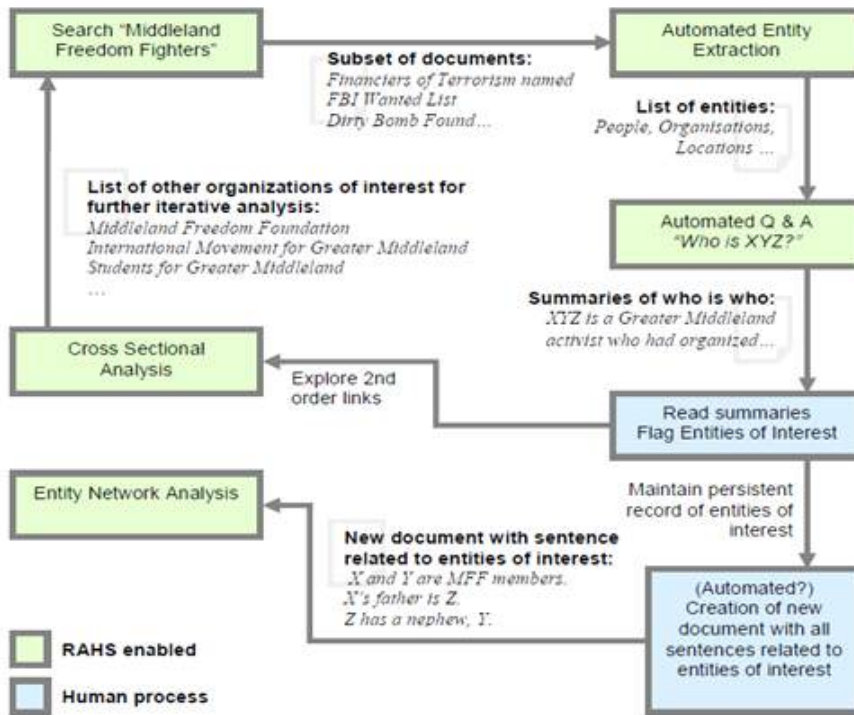


그림 15. 지능적 분석 프로세스를 통한 RAHS 프로세스 개념 설명

(<http://www.mindef.gov.sg/> 싱가포르 국방부)

## 다. 싱가포르 공공 빅데이터 분석 활용 사례

### 1) 미래 이슈 시뮬레이션

- > RAHS 시스템은 여러 빅데이터 분석 기술 기법들을 적용하고 있는데 각 정부 부처에서는 다양한 형태의 적용을 통해 공공 정책, 운영 또는 예측 시스템을 개선하고 있다. 보건성인 경우 조기 잠재 전염병 확인을 위한 감시 시스템을 운영 중인데 WHO를 비롯한 세계 관련 기관들과의 협조도 중요하나 호라이즌 스캐닝을 통해 예측 가능한 이슈들을 진단하여 할 수 있기 때문에 효과성이 증대하였다. 국가 해양 신호 탐색 센터는 전조 위험을 탐지하기 위해 실시간 위협 평가를 수행하며 사이버 보안 센터 또한 RAHS를 활용하여 상황인식

및 기존과 예측 사이버 보안 이슈 연계성을 진단하여 이에 대비책을 조정 관리한다. 별도로 운영중인 사이버 모니터링(Watch) 센터는 전체적인 정부 시스템에 대한 위협을 사전 파악하고 비인가 접근 변경사항들을 추적하며 대응할 수 있도록 RAHS를 사용하고 있다. RAHS는 앞선 정보 기술 플랫폼을 부처, 연구 계에 공개하여 빅데이터 분석을 통한 유기적인 미래 예측과 사전 대응을 위한 협력 시스템 구성, 조직화에 주요 역할을 담당하고 있다.

## 2) 공공 빅데이터 활용 민간 서비스 현황

- ▶ Data.gov.sg에서는 공공 데이터를 카탈로그로 공개하며 개발 활용이 가능하도록 API 또한 오픈 소스로 공개 중에 있다. 해당 포털에서 60여 정부 부처 및 기관에서는 8800 데이터 세트를 공개하였으며 (2014년 4월 기준) 경진대회, 이벤트 홍보를 통해 오픈 데이터 활용을 지원하고 있다. 싱가포르 교통국 데이터 공개 지원으로 스마트 폰 GPS 기능을 통해 인근 버스 정류소를 찾을 수 있는 “SG Buses” 앱이 대표적인 사례로 꼽힌다.

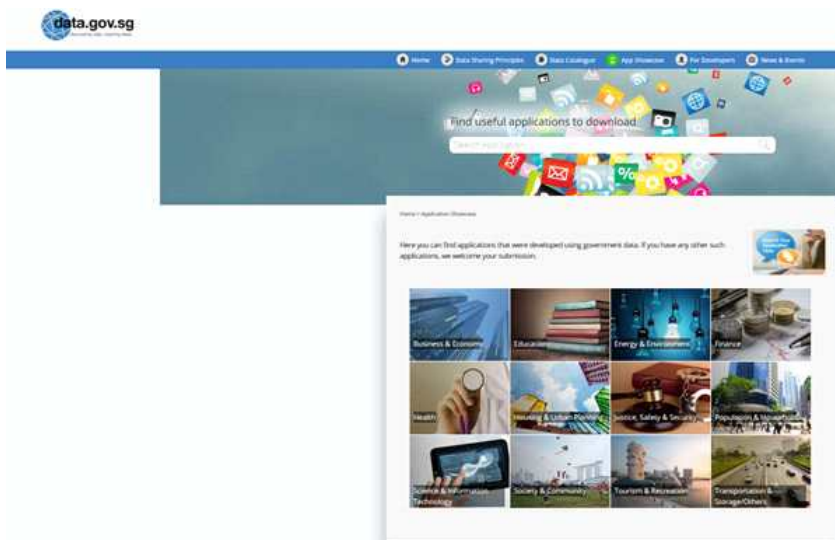


그림 16. 싱가포르 공공 데이터 활용 애플리케이션 showcase  
(<http://data.gov.sg> 싱가포르 데이터 포털)

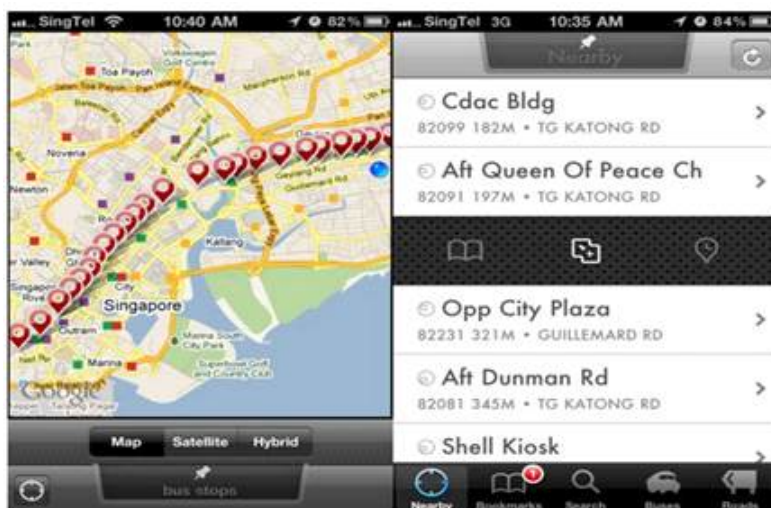


그림 17. 싱가포르 공공 데이터 활용 애플리케이션 “SG Buses”

- > 미국과 유럽, 기타 공공 데이터 플랫폼을 선도하는 국가들과 비교할 때 공공 데이터 공개 수준과 양이 많은 편은 아니며 정제된 포맷으로 제공되는 레벨이 평균 보다 높지 않다. 그러나 환경 탐색으로써 얻은 빅데이터를 분석하여 국가에의 위협 요소를 판별 축적해온 해온 RAHS 운영 노하우와 성숙된 분석 기술 역량, 새로운 미래예측 연구개발 인프라가 있으므로 개방 활용의 데이터 공개 플랫폼을 데이터 표준화, 실시간 데이터 수집 및 양질의 데이터 등록 발행까지 이르는 관리 단계로 확장하게 되면 개방 데이터의 가치 발전과 함께 시민사회와 소통하는 데이터 중심 정부 서비스 플랫폼 구현이 어렵지 않게 이뤄질 수 있는 가능성을 보유한다.
- > 2014년 10월 싱가포르 정보통신 개발국에서는 데이터 분석기반 스마트 국가를 비전으로 민간 데이터 통합관리를 통한 데이터 생태계 활성화를 위해 DaaS(Data as-a-Service) 플랫폼 구축 시범사업을 추진하기로 한 상황이다. 해당 시범사업을 통해 오픈 데이터 포털 플랫폼인 CKAN을 기반한 데이터셋을 기반한 저장소를 구성하며 민간 데이터 제공 사업자들에게 클라우드 서비스, 데이터 품질 매트릭스 (DQM: Data Quality Matrix)를 제공하게 된다.

### 3. 미국 공공 빅데이터 인사이트

#### 가. 미국 공공 빅데이터 분석 플랫폼

##### 소크라타 (Socrata)

- ▶ 미국은 공공 데이터 개방을 위해 제도적 기반을 개방 정부 계획 (Open Government Initiative) 하에 세웠고 'data.gov' 데이터 포털을 통해 다양한 부문의 정부 데이터를 개방해 왔다. 클라우드 기반의 개방형 데이터 플랫폼을 지향하며 세금이 사용된 데이터 결과물에 대해 오픈 라이선스를 적용하였다. 또한 데이터 검색, 소셜 네트워크 공유 등의 온라인 접근성을 높이며 활용성을 다각화하기 위해 데이터 분석 결과에 대한 정밀성과 가시성을 개선하기 위해 노력해 왔다. 2013년 백악관 추진 'Presidential Innovation Fellowship Program' 중 하나로 채택된 개방 데이터 계획 (Open Data Initiative)은 공공 데이터를 활용한 일자리 창출과 비즈니스 가치 구현이라는 모토로 진행되고 있다.
- ▶ 소크라타는 공공 데이터의 개방 및 공유와 연관된 분석서비스를 제공하는 플랫폼으로 사용자는 위치기반 기술로 이뤄진 직관적 인터페이스로 사용하게 되며 별도의 인증 절차 없이 지역 기반 공공 데이터의 시각화된 분석 결과를 확인할 수 있고 오픈 API를 통해 데이터를 손쉽게 생성할 수 있다. 소크라타 플랫폼은 사이트 활동, 단계별 관여 지표 등을 나타내는 분석 API를 보유하며 성능 관리 도구와 같은 애드온(Add-on) 서비스 구성이 가능하다. 토큰 형태의 가입 유료기반 SaaS(Software-as-a-Service) 클라우드 서비스로 SODA(Socrata Open Data API) 기술 표준을 구현하므로 데이터 생애 주기에 따라 자동으로 데이터 이력을 관리하게 되며 운영자는 데이터 기관 간의 연계성을 위한 별도의 작업을 진행하지 않아도 된다. Data.gov 포함, 샌프란시스코, 뉴욕, 시애틀, 시카고 등의 정부 공공기관에 구축 지원되고 있는 해당 플랫폼은 개방 데이터 성숙도 모델을 제시하며 공공 기관 보유 정보를 공개하여 왔는데 영국 정부의 '5스타 오픈데이터' 전개구조와는 다르게

데이터간의 연결 활용을 위한 기준 대신 데이터 자원 관리에 주안점을 두고 있음을 알 수 있다.

- 레벨 1은 산재된 스프레드시트와 PDF 형태의 문서 데이터, 비표준화된 웹 애플리케이션과 같은 상태이며 레벨 2는 다운 가능한 문서 형식, 수동식 데이터 출판 형태 등이 해당된다. 레벨 3으로 올라가면 차트와 지도와 같이 기본적인 데이터 가시화가 가능한 데이터 형태로, 소셜 공유 등이 있으며 레벨 4에서는 사용자 맞춤형 웹 경험, 애플리케이션, 매쉬업, 대시보드, 데이터 기반 스토리텔링, 자동화된 데이터 출판, 모바일 경험 디자인 등이 부합된다. 성숙 레벨 5단계에 도달한 데이터 기준은 사실 기반 결정과 성능에 충족된 데이터, Yelp와 구글과 같은 내재된 데이터 경험치수, 클라우드 소셜 데이터, 플러그애플레이 애플리케이션, 실시간 데이터 생태계 및 협업 허브 등을 들 수 있다. 미국 정부는 공공 데이터 개방에 대해 활용 가능성이 높은 에너지, 교육, 재정, 공공안전, 글로벌 협력 분야를 우선 선정, 점진적인 개방 정책을 추진하고 있다.

## Open Data Maturity Model

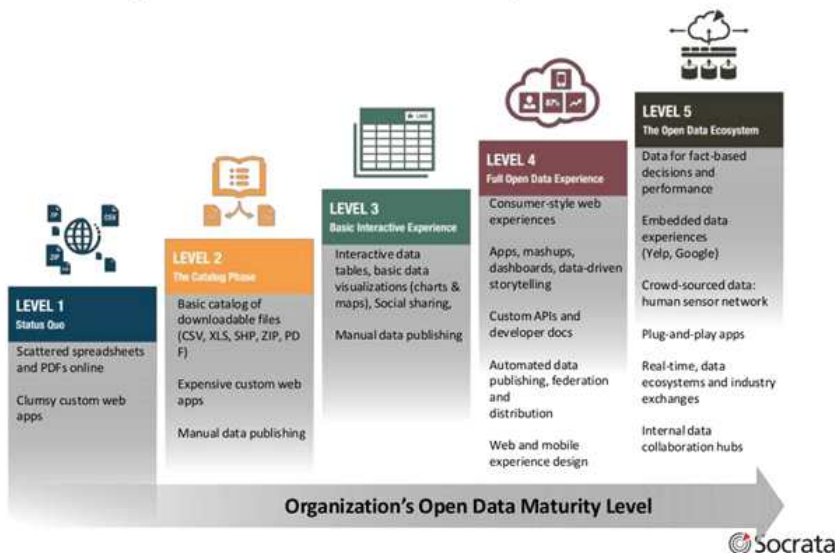


그림 18. 개방 데이터 성숙도 모델 (www.socrata.com)

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

- > 별도로 2013년 미국 정부는 영국에서 태동된CKAN과 OGPL 결합 플랫폼을 진행하기로 결정했는데OGPL(Open Government Platform)은 미국과 인도 정부가 공동 협력 개발 중인 오픈 데이터 플랫폼이며data.gov의 백엔드 시스템으로 데이터를 외부와 내부로 상호 연결해주며 데이터 관리와 데이터 포털의 종합 기능을 지원하는 오픈 소스 형태로 되어 있다. 즉, 미국 지역뿐만 아니라 공공 데이터 오픈 플랫폼을 추진하는 다른 나라, 해외 지역에서도 손쉽게 사용자 접근과 사용에 있어 플랫폼 선 학습에 대한 어려움 없이 통합적이고 유기적인 서비스로 구성할 수 있다. 각기 다른 사례와 현황에서 데이터를 공유, 연결하고 활용하게 되고 시스템 간의 데이터 교환 또한 자유롭게 되면서 새롭고 가치 있는 데이터 활용 서비스 개발을 독려하고 있다.

## 나. 미국 공공 빅데이터 분석 기술 현황

### 1) 분석 기술 역량화

- > 2013년 빅데이터 R&D 추진 계획을 발표한 정부 당국은 세계 1순위로 빅데이터 분석 과 관련한 특허, 사업화 기술들을 민간 사업자 협력을 통해 실현하고 있다. 공공 개방 데이터 플랫폼으로 선정된 소크라타 역시 링크드 데이터 플랫폼 구현을 통해 개체, 컨텐츠, 관계 별 분석 기법들을 적용하여 사용자들의 가독성과 편의성을 높이는 표현방식을 채택하고 있다.

| Object               | Entity  | Contents                           | Relationship                       |
|----------------------|---|------------------------------------|------------------------------------|
| Focus                | Association among Properties of Entities      | Meanings contained in the Entities | Structural Positions and Diffusion |
| Analytic Methodology | Statistics<br>Machine Learning<br>Data Mining | NLP<br>Text Mining                 | Network Analytics<br>Graph Mining  |

그림 19. 빅데이터 형식별 분석기법

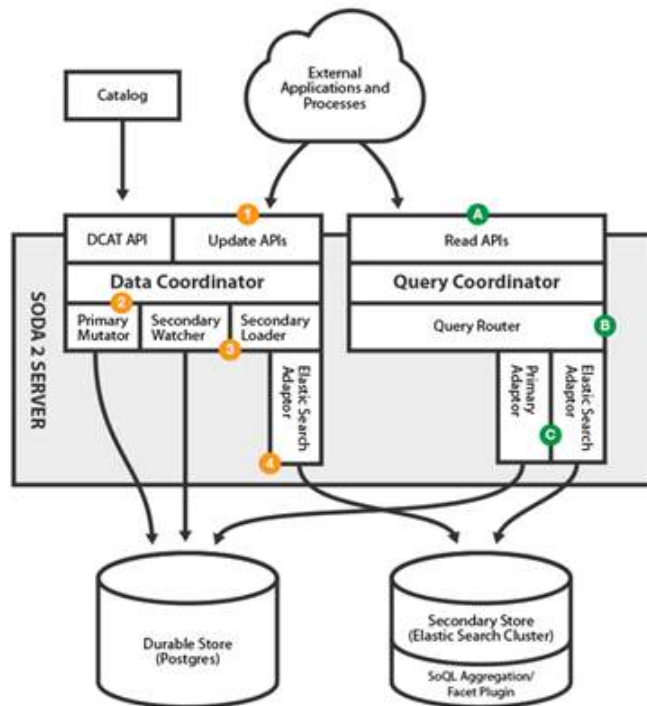


그림 20. 소크라타 데이터 플랫폼 공개 시스템 아키텍처 (www.socrata.com)

## 2) 예측 치안 모델 (Predictive Policing)

- 미국 정부 당국은 빅데이터 분석 관리를 위한 총체적 접근 방법 들을 장기 계획을 통해 R&D 연구과제를 진행해왔다. 범죄 예측을 위한 치안 모델은 사건 발생 패턴과 함께 익숙한 지역에서 벗어나지 않으려는 범죄자들의 행동 경향 등의 관련한 변수들의 상관관계를 분석, 범죄 통계 데이터를 정제화하여 장소, 시간, 범죄 유형에 대해 분석 알고리즘으로써 보다 정확한 예측을 할 수 있다. 2011년 샌타크루즈경찰국(SCPD)으로 시범도입 된 예측프로그램은 인권 보호 법률과 상충되는 부분이 있으나 범죄 산출율이 높은 연결지점인 레드박스(152.4m<sup>2</sup>)크기, 범죄유형 선택, 실시간 갱신 또는 주기 조절 등의 용이성으로 효율적인 인력 운용과 효과적인 범죄 예측 대응으로 호응을 얻고 있다.



> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링



그림 21. 예측 치안 모델

> 해당 알고리즘은 패턴 발견 알고리즘인 지진 발생 후의 여진 예측 모델에서 비롯된 분석 기법으로 과거, 현재 데이터들을 병합하여 분석하는데 범죄 유형 목록에서 개인별 요소로 여진 예측 분석 수확모델을 적용, 일대일 비교 방식으로 진행된다.

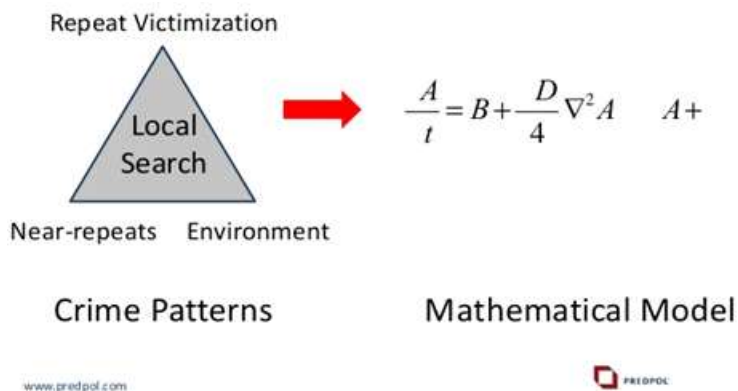


그림 22. 예측치안모델 분석알고리즘 (www.predpol.com)



## 다. 미국 공공 빅데이터 분석 활용 사례

### 1) 빅데이터 활용 공공 서비스 개선 사례

- > 미국 주 정부는 공공 기관간의 데이터를 연결 통합하여 행정 서비스 효율화 및 시민을 위한 정책 개선 제고에 사용하고 있다. 국토 안보국은 개인정보, 시민자유권, 법률 감독 기관들의 대표자들과 함께 데이터 스튜어드 (data steward), 주기적으로 데이터를 사용 관리하는 담당자를 지정하여 개별 데이터베이스의 데이터 그룹, 속성, 권한 등을 검토하였고 메타태그, 참조 정보를 업데이트, 규칙 분류를 진행해 오고 있다. 데이터베이스 속성은 크게 3가지로 중요 개인 정보, 부가 개인 정보, 점점 정보로 나뉘어지며 점점 정보는 당국의 관찰과 인터뷰를 통해 생성되기 때문에 데이터 접근 제한 설정에 영향을 준다. 국토 안보국은 기관 위협 감지를 위해 시스템 사용 이력 파악과 예측, 이상탐지 분석을 진행할 수 있다.
- > 국립보건연구원은 정부의 빅데이터 R&D 개발 투자 초기에 1,700 명으로부터 수집한 DNA 시퀀스를 토대로 인간 유전체 해석을 위해 아마존 웹서비스 (AWS)에 공개한 바 있다. 2014년 생물의학 빅데이터 분석 및 이용 촉진을 위한 전략 개발 프로젝트로 6억 5,600만 달러 기금을 조성, 2020년까지 투자할 계획을 밝힌 연구원은 수 천명 이상의 역학 데이터가 필요한 생물 의학 연구에서 빅데이터는 새로운 통찰력을 제시하며 인류의 건강과 관련된 질병과 같은 요소 예측 분석과 사전 위험에 대비한 진단, 치료, 예방 을 향상하는 데에 효율적인 도구로 판단하고 있다.
- > 공공 보건 부문의 활용 사례로 알려진 Pillbox는 보건부와 국립보건원의 데이터 제공을 통해 국립의학도서관에서 제공하는 서비스로 유전자 데이터 공유를 통해 질병 진료 체계를 수립하며 의약품의 상세 정보를 소비자에게 전달한다. 제약 회사가 정보 제공 사업자로 활동하며 사용자들 간 개인별 특성, 진료 이력, 유전자 정보 등을 공유하면서 비교, 분석 할 수 있다. 누구나 쉽게 참여할 수 있고 의약품 정보 검색과 개인의 질병 데이터를 이용한 분석

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

모델 수립 예측이 가능하다. 공공 빅데이터의 예측 플랫폼으로써 다양한 사용자 질병 통계 데이터와 의약품 설명 데이터베이스가 확보되어 특정 질병 분포와 추세에 대해 예측할 수 있게 되었고 민원 서비스 관리 비용 절감 및 관련한 공공 서비스 개선 효과를 가져왔다.

그림 23. Pillbox 홈페이지 (<http://pillbox.nlm.nih.gov>)

- 지역 도시에서도 빅데이터를 활용한 사례들이 늘고 있는데, 뉴욕시인 경우, 공공 개방 데이터를 동적으로 변환시켜 실시간으로 업데이트를 하고 있으며 주요 데이터들을 RDF 포맷화하였으며 지도와 함께 매쉬업(Mash-up) 시각화 서비스를 제공한다. 모바일 애플리케이션을 통해서도 동일한 링크드 데이터를 공유 확산하기 쉽도록 표준 API를 공개하고 있다.

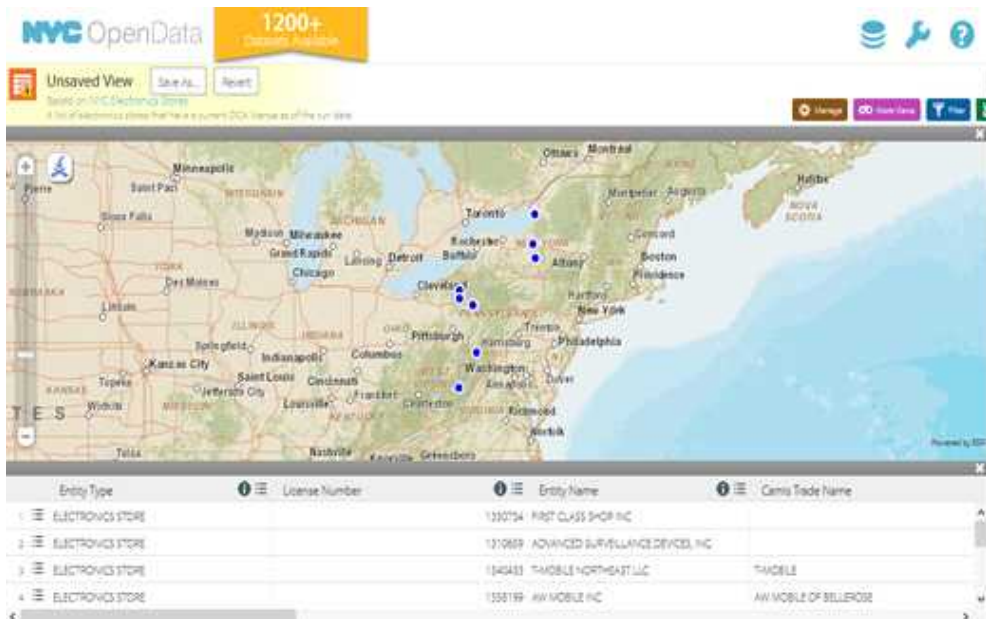


그림 24. NYC 전자매장 지도 매쉬업 (<https://data.cityofnewyork.us/>)

- > 공공 오픈 데이터 개방과 더불어 범죄, 화재 예방, 탈세 방지 등의 공공 서비스 개선을 위해 공공 데이터 활용을 활발히 추진하고 있다. 일례로 화재 예측 시나리오를 통한 건물들의 예방 서비스를 시행하고자 교통위반, 범죄, 사업자 고유 번호 등의 발생 가능 요소 60여가지를 추출, 33만 개의 빌딩들에 대해 화재 가능성에 대한 가중치 점수를 책정하는 알고리즘을 수립하였다. 매주 소방관들과 경찰관들이 방문하여 측정한 실사데이터를 정기적으로 입력하였고 알고리즘이 적용된 데이터 마이닝을 통해 2400여개의 카테고리 결과를 분류하여 정기적으로 화재 발생 가능성이 높은 건물들을 추려내어 퇴소 조치를 취한 결과 건축물 화재율을 줄일 수 있었다.

## 2) 공공 빅데이터 활용 민간 서비스 현황

- ▶ 미국에서는 공공 데이터 분석을 기반으로 한 민간 서비스들이 늘어나고 있는 추세이다. 2013년 다국적 농업기업인 몬산토로 합병된 클라이밋 코퍼레이션 (Climate Corporation)은 250만 개 위치에 대한 국립기상서비스의 실시간 기상 데이터, 14 테라바이트에 달하는 농무성의 토양 형태 정보, 지난 60여 년간 수확량 공공데이터를 수집하여 10조 단위의 날씨 예측 시뮬레이션 모델을 생성, 농가를 위한 날씨 보험과 위험 분석 서비스를 제공한다. 그 외에 토양, 물, 열에 대한 작물들의 반응과 관련한 비즈니스 모델을 구상하고 있는데 상업적 예측 모델 서비스 개발에 대해 무상으로 공공 데이터를 활용하고 있다.



그림 25. 클라이밋 코퍼레이션 홈페이지 (<http://www.climate.com>)

- ▶ 2013년 12월 뉴욕대 Governance Lab에서 조사한 공공 오픈 데이터를 활용한 제품 또는 서비스를 공급하는 민간 기관 500곳 목록이 'Open Data 500' 프로젝트 명으로 사이트([www.opendata500.com](http://www.opendata500.com))에 공개되었다. Zillow.com과

같은 인구통계, 지역정보, 부동산 정보 등을 조합한 부동산 거래 서비스와 같이 공공 데이터의 경제적인 가치를 높이는 비즈니스 성공 모델이 실현되고 있으나 1차로 조사한 바에 따르면 필요로 하는 데이터셋의 품질 문제 및 현실 데이터와의 괴리가 있음에 여러 불만들이 제기되었다고 한다. 그럼에도 불구하고 공공 개방 데이터의 수요가 증가할 수 밖에 없는 이유는 사용자의 요구에 따른 좀더 유용한 나은 품질과 표준을 지닌 데이터로 변모하게 됨을 확신하고 있기 때문으로 리서치 팀에서는 설명하고 있다.

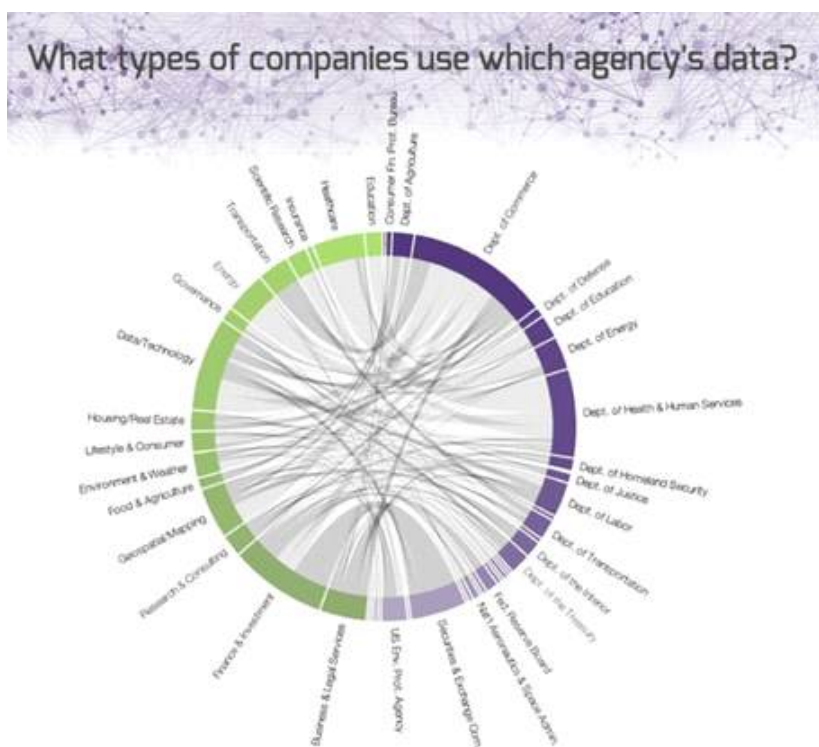


그림 26. 공공 개방 데이터 사용 희망 기관 산업별 도해 (www.opendata500.com)

## 4. 대한민국 공공 개방 데이터 분석 플랫폼 현재 (AS-IS)

### 가. 대한민국 공공 빅데이터 분석 플랫폼

#### 1) 대한민국 전자정부와 정부 3.0

- ▶ 1990년 대 이후 국가 경쟁력 향상을 위해 정부에서는 정보화 촉진 비전 전략 수립 아래 Korea 기본계획까지 5차례에 걸쳐 정보화 사업을 추진 한 결과 국제 전기통신 연합 (ITU) 디지털 기회지수 (1위), 국제 연합(UN) 전자정부 성숙수준 (6위), 전자정부 모델 해외수출 달성, JAVA플랫폼 기반의 표준 개발 프레임워크 정립 및 공통 컴포넌트 47종 CMMI 인증 등의 괄목할만한 성과를 이루었다. 그러나 전자정부 메타데이터 표준 미정립, 다수 사업간의 연계성 부족, 주요사업 성과관리 조정 누락, 이전 사업 평가반영 효용성 저하 등의 문제점들을 표출하였고 이후 2013년 공공 데이터의 적극적 개방을 통한 신뢰 정부 구현 및 신 성장동력을 창출하자는 비전을 담은 정부 3.0 전략이 나오게 되었다.
- ▶ 공공 데이터의 제공 및 이용 활성화에 대한 법률 제정, 공공 데이터 전략위원회 및 분쟁조정위원회 출범으로 공공 데이터 개방이 가속화되었고 2014년부터 2017년까지의 변화 핵심 성과지표를 목표로 클라우드 기반의 범정부 개방 데이터베이스가 구축되어 개방형 플랫폼인 공공데이터포털(data.go.kr), 국가공간정보 유통시스템(www.nisc.go.kr)에서 제공되고 있다.



그림 27. 대한민국 정부 오픈 데이터 포털 (<http://www.data.go.kr>)

고가치, 고수요 데이터를 우선적으로 개방하며 민간 주도 개방 로드맵 수립, 빅데이터 데이터 개방체계 마련 및 질적 관리 표준체계 구축으로써 공공 데이터 포털을 통해 단일화로 통합제공을 목표로 하고 있다. 2016년까지 데이터 품질 인증제 도입, 2017년까지 자유로운 가공과 융합이 가능한 오픈 포맷 데이터비중을 전체 공공데이터의 50%까지 확대 예정에 있다. 또한 공정경쟁의 데이터 생태계 조성을 위해 공공데이터의 민간활용을 위해 지원 협력 사업들을 추진하고 있다.

| 구 분                | '14년         | '17년    |
|--------------------|--------------|---------|
| 데이터 개방 순위(ODB 순위)  | 12위('13년 평가) | 5위      |
| 국가중점 개방 데이터(대용량DB) | 1개(시도·새울)    | 30개 이상  |
| 오픈 포맷 데이터 비중       | 8.8%         | 50% 이상  |
| 데이터셋 표준 개발         | 30개          | 100개    |
| 활용 서비스 수           | 300여개        | 2,000여개 |

그림 28. 2014년 공공데이터 개방 발전전략, 공공데이터전략위원회, 2014.9



> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

- > 2014년 공공데이터 개방 발전전략, 2014년 발표된 공공 데이터 개방 발전 전략 보고서에서의 추진 결과를 통해 2013년 대비 공개 데이터의 활용도가 데이터셋770%, OpenAPI는 140%까지 상승했으며 공공 데이터 활용 민간 부문 서비스 개발 사례건 또한 전년 대비 6.5배 증가하였음을 볼 수 있다.

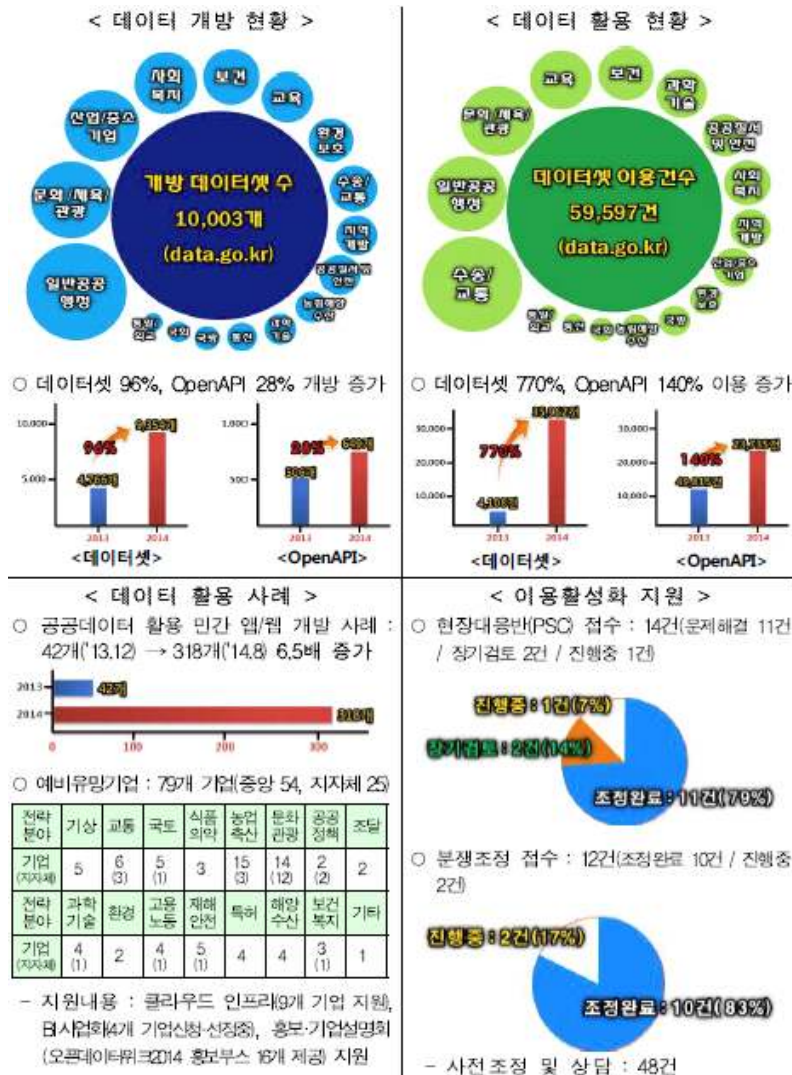


그림 29. 2014년 공공데이터 개방 발전전략, 공공데이터전략위원회, 2014.9



## 2) 대한민국 공공데이터 플랫폼

- > 대한민국 정부는 공공 개방 빅데이터 활용을 위한 자산 구축과 링크드 데이터 응용을 위해 포털 플랫폼의 아키텍처를 데이터 플랫폼 오픈소스로 전면 전환하는 방안을 검토 중에 있으나 현재까지는 공공 데이터 포털과 지방 자치 단체들간의 데이터 연계 및 선도 모델 육성, 법제도 개선, 교육가이드라인 수립, 담당체계 정립 등의 데이터 표준화를 통한 상호운용성과 개방성에 초점을 맞추고 있다.
- > 데이터의 일관적인 연결성과 데이터 재사용을 위한 온톨로지 기반 링크드 데이터 기반 구축은 이미 영국, 미국, EU 등에서 정부 공공 데이터 오픈플랫폼과 연계하여 추진, 발전단계에 있는 데에 반해 국내에서는 2012년 공공DB피디아 명의 LOD 시범사업(한국정보화진흥원)과 개별적인 공공기관에서 제한적으로 진행되고 있다. 공공DB피디아(lod.data.go.kr)는 상위 분류로 교육, 통계, 공공시설, 문화, 부동산, 행정으로 나누어 공공 기관에서 제공한 데이터 셋을 기반하여 명세화, 모형화, LOD생성, 데이터 및 메타데이터 발행 순서로 링크드 데이터를 발행 서비스하고 있으나 현재저 적은 제공 데이터 크기와 비정형 데이터 지원 미비, 특히 제공기관들의 자체 데이터 온톨로지 부재 등으로 전문적이고 체계적인 온톨로지 기반의 링크드 데이터 구축이라고 하기는 어렵다.
- > 또한 데이터 제공 참여 공공기관들과의 연계 가능성을 고려한 링크드 데이터 구축 사업 가이드라인이 누락되었고 데이터 연계 책임 운영 체계 비표준화, 연계를 위한 동기화 API(Syndication API)만을 제공하고 있으므로 공공DB피디아 시스템 아키텍처 플랫폼을 갖추지 않은 상황에 있는 기관들은 궁극적 링크드 데이터 구축 연계 플랫폼을 구현하기는 사실상 불가능한 실정이다. 또한 데이터 재사용을 위한 개별 데이터셋의 URI 생성을 위해 언론, 방송 분야에서는 데이터 영역별 공공기관이 정해져 있지 않은 상황에서 접근이 쉬운 위키피디아를 중심으로 한 URI 생성 방법을 채택하고 있다.

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

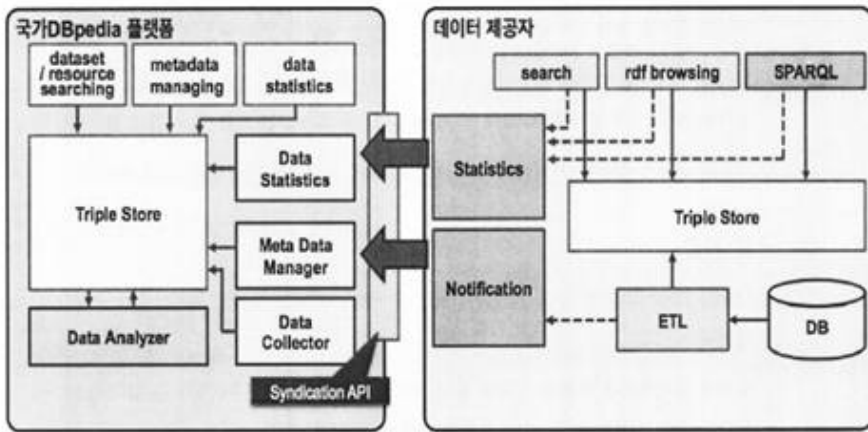


그림 30. 공공 DB피디아 시스템 아키텍처 (한국정보화진흥원, 2012)

## 나. 대한민국 공공 데이터 분석 기술 현황

- 2013년 빅데이터 특허 출원 동향을 살펴 보면 한국은 미국 다음으로 많은 빅데이터 기술에 대한 특허 출원을 보유하고 있으나 대부분이 저장기술에 집중이 되어 있고, 데이터 분석 기술에 대한 출원은 상대적으로 저조한 편이다. 실시간 처리, 저장, 표현 기술 특허 출원율은 높은 편이나 데이터 분석 및 수집 기술은 낮게 나타나 이에 대한 기술 연구 개발이 시급한 것으로 보고된 바 있다<sup>2)</sup>. 물론 유럽과 일본에 비해 데이터 분석 기술 출원 현황은 높은 수치이나 실제로 출원 세부 내역 검토 시 분석 원천 기술 보다는 활용 서비스 또는 방법에 중점을 두고 있음으로 파악할 수 있다.

2) 2014 Electronics and Telecommunication Trends ETRI 발행 창조경제 시대의 기술 시장 전망 특집  
- 특허 분석을 통한 빅데이터 기술개발 동향

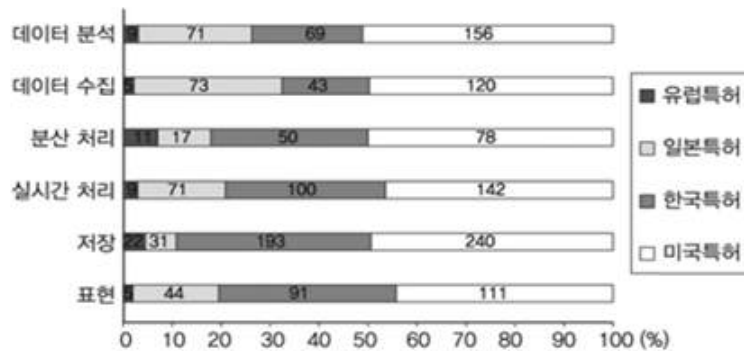


그림 31. 빅데이터 세부 기술별/국가별 특허출원 동향

| 대분류    | 소분류                                      | 검색개요(기술범위)   |
|--------|--|--|
| 데이터 수집 | 크롤링 (crawling)                           | 검색엔진의 로봇을 이용한 데이터 수집   |
|        | ETL(Extraction, Transformation, Loading) | 소스 데이터의 추출, 전송, 변환, 적재   |
| 저장     | NoSQL Databases                          | 비정형 데이터 관리, 분산 데이터베이스(Not only SQL)                              |
|        | Storage                                  | 빅데이터 저장, High Performance  |
|        | Servers                                  | 초경량 서버   |
| 실시간 처리 | In-Memory computing                      | In-Memory 플랫폼, In-Memory 메시징, In-Memory 데이터 관리 (DBMS, Data Grid) |
|        | 데이터 스트림 처리                               | DSMS, Storm, ESPER, S4, Hstreaming CEP(Complex Event Processing) |
| 분산 처리  | Cloud computing                          | 클라우드컴퓨팅, 분산처리  |
|        | Hadoop                                   | HDFS, MapReduce  |
| 데이터 분석 | NLP (Neuro Linguistic Programming)       | 자연어 처리   |
|        | Machine Learning                         | 기계학습을 통한 데이터의 패턴 발견  |
|        | Data Mining                              | 데이터마이닝   |
|        | Serialization                            | 데이터 간의 순서화   |
| 표현     | Visualization                            | 데이터를 도표나 그래픽으로 표현, 시각화, R, 시뮬레이션, 수치해석                           |
|        | Acquisition                              | 데이터의 획득 및 재해석  |

그림 32. 빅데이터 기술 분류 [출처: ETRI]

## 다. 대한민국 공공 빅데이터 분석 활용 사례

### 1) 빅데이터 활용 공공 서비스 개선 사례

- ▶ 2011년 국가정보화전략위원회의 ‘빅데이터를 활용한 스마트정부 구현안’에 따른 빅데이터 마스터 계획 수립 이후 미래 창조과학부는 2013년 교통, 창업지원, 보건 의료 등을 포함한 공공 빅데이터 활용 스마트 서비스 복수 시범사업을 진행했다. 이 중 하나인 서울시 심야버스 노선 최적화 지원사업은 서울시와 민간 통신회사와 공동협력으로 이뤄졌으며 유동 인구 밀집 구간을 선별, 서울시 전역을 지름 1km의 1252개 구간으로 분류한 뒤 심야 시간대에 수집된 약 30억건의 전화 통화 데이터와 심야 택시 승/하차 500만건을 분석하여 교통 수요 구간을 우선 순위로 정리하였다.
  
- ▶ 또한 교통 카드 이용량 등의 공공 데이터를 기반하여 요일, 시간, 노선 별 유동 인구의 가중치를 통한 패턴 분석 결과를 시각화하였고 이들을 결합하여 기존 심야버스 와 비교, 조정 추가하여 최적화된 노선 설계에 반영하였다. 2013년 9월 13일부터 11월 1일까지 올빼미 버스 9개 노선을 운영한 결과 야간 시간 동안 하루 평균 6천명 이상이 이용하였고 운행 기록 분석과 시민 의견을 수렴하여 배차 간격 조정, 운행 경로 조정, 서울 교통 포털사이트, 모바일 웹/애플리케이션, 버스정류소 도착안내단말기(BIT) 연동 등의 지속적인 개선활동으로 서울시 시행 정책들 중 서울시민이 뽑은 10대 뉴스 중 1위 (2013년 기준)에 선정되기도 하였다.
  
- ▶ 이 때 사용한 GIS 기반 분석소프트웨어는 공간 정보 분석을 담당하였으며 대용량 데이터 분산 처리 도구로 Hadoop 오픈소스가 활용되었다. 그러나 해당 사업에서 처리된 모든 데이터가 빅데이터는 아니었으므로 처리능력 보다는 로그, 비정형 데이터들의 가치를 찾아내기 위해 사용된 분석 알고리즘으로써 데이터 기반 최적화된 노선 결정을 이뤄낸 결과로 볼 수 있다.

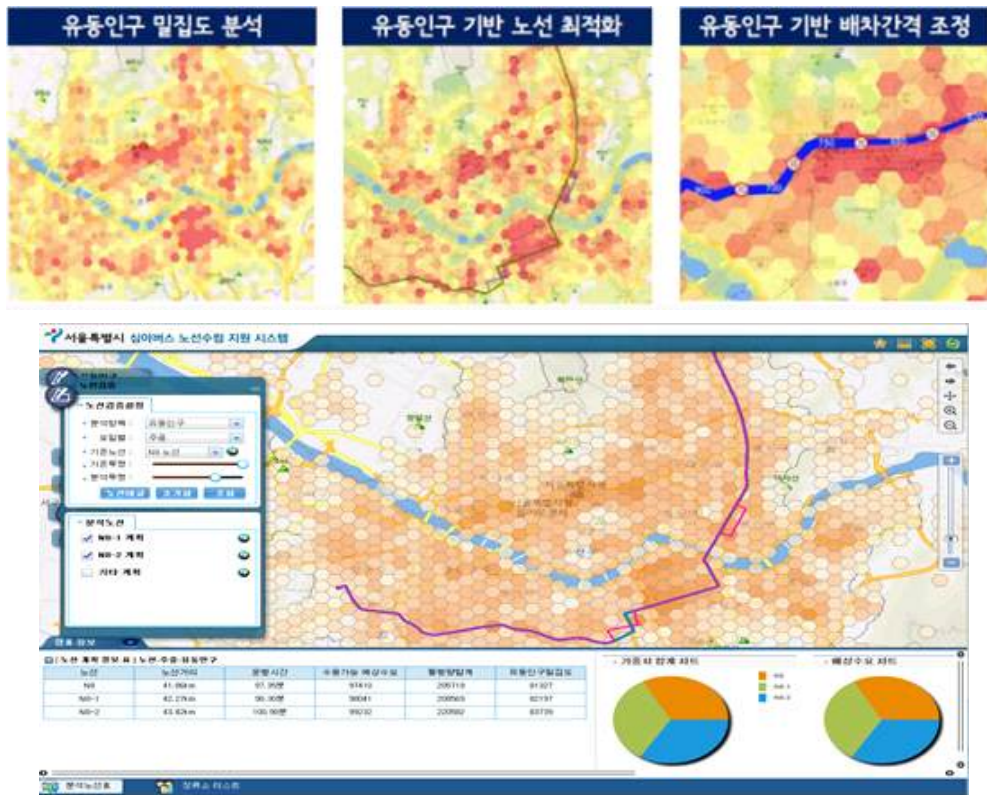


그림 33. ArcGIS 분석소프트웨어 활용 서울시 심야노선 버스 최적화 사례

## > 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

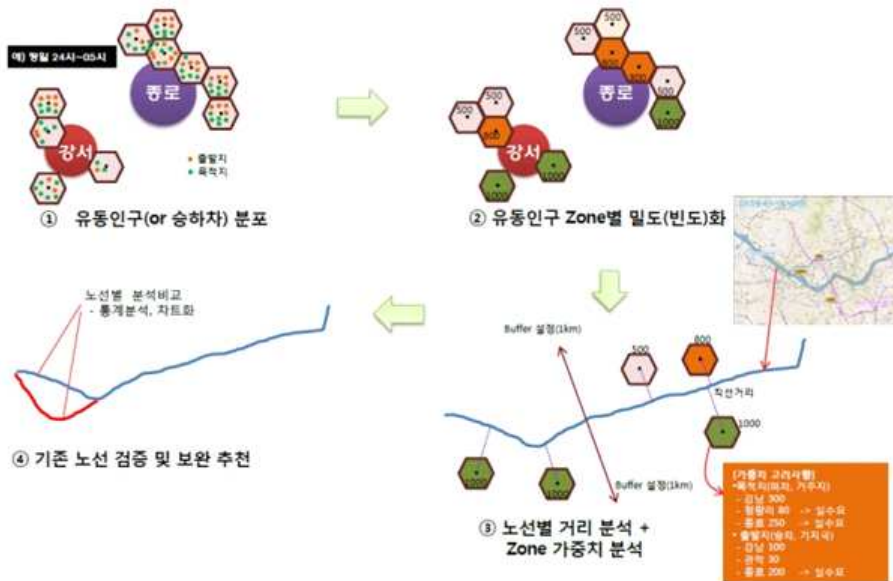


그림 34. 서울시 심야버스 노선 최적화 빅데이터 분석 적용 과정  
(<http://zeronova.kr/2013/08/07/seoul-bus-route-optimization/> )

## 2) 공공 빅데이터 활용 민간 서비스 현황

> 정부에서는 공공 데이터 활용 유망 기업 육성 및 데이터 활용 경진대회, 인증 제도 운영, 창업 기반을 지속적으로 확충하는 형태로 적극적인 지원을 진행하고 있다. 데이터 기반 정책 수립을 위한 인프라 구축, 사회적 현안 해결을 위해 빅데이터를 활용한 정책 수립, 공공 기관간의 공동 연구 협력 체계 마련, 정부 부처간의 상호 협력 등의 여러 활성화 지원책이 운영되고 있는데 이 중 2014년 빅데이터 분석 역량 강화 지원 전략이 인상적이다. 즉, 국내외 미래전략 기관들, 싱가포르 RAHS, 영국의 FORESIGHT 프로그램 기관 등을 초청, 협력방안을 논의하며 한국형 분석방법론 개발 계획 수립을 완료하고자 함이 목적이다. 국내 실정에 맞는 예측모델, 분석 시나리오, 기법을 연구 개발하는 조직을 별도로 운영, 데이터 분석을 통한 미래예측 및 전략 수립 과정 가이드 제작 및 보급계획에 있다.

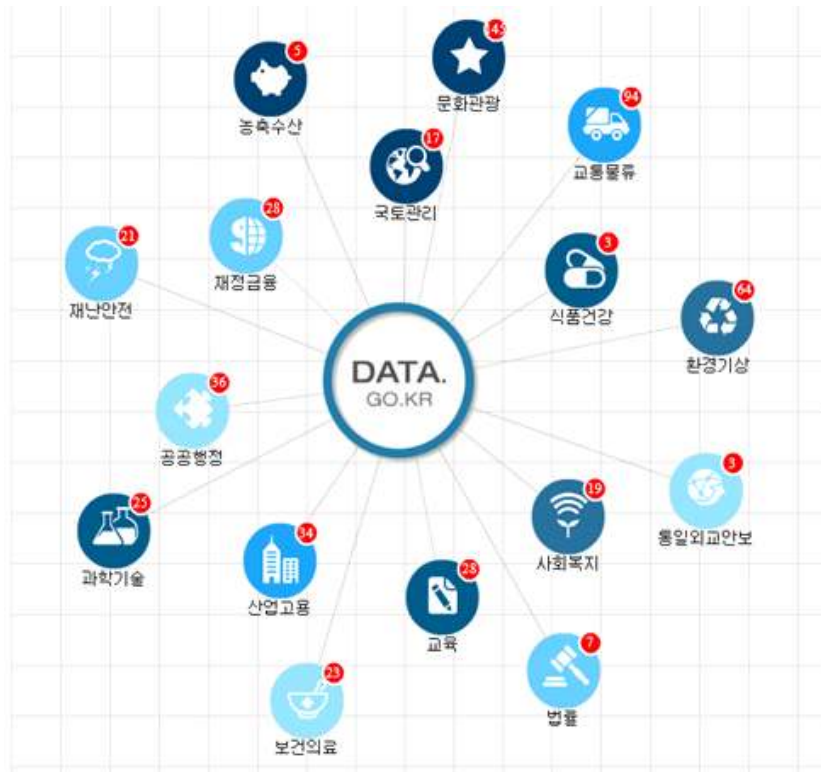


그림 35. 대한민국 오픈 공공 데이터 활용 사례 실시간 현황 (2014년 11월 20일 기준)  
(<http://www.data.go.kr>)

## 5. 국내 공공 빅데이터 분석 플랫폼 구축 마스터 플랜(TO-BE)

### 가. 현재 국내 공공 오픈 데이터 플랫폼 진단

- 공공데이터 포털은 데이터 공유자원과 오픈 API공유자원으로 분리하여 국가가 보유한 정보자원을 제공한다. 또한 활용지원센터 운영, 개발자 LAB 지원을 통해 개발자가 용이하게 공공정보를 활용하여 부가가치 서비스 개발이 용이한 환경을 제공함으로써 공공 데이터의 활성화 방안을 모색하고 있다. 추가로 각 기관에서 생성된 공공 데이터의 접근과 재사용이 가능하도록 링크드 데이터로 변환하여 공공 DB 피디아<sup>3)</sup>를 통해 제공 중에 있다 (lod.data.go.kr). 국가 DB 구축 사업에서 LOD 구축 사업을 통해 공공 지식 정보의 상당수가 링크드 데이터로 변환되었고 이 중 민간 수요 및 정보 간 연관성이 높은 데이터를 우선하여 공공 DB 피디아에 연결, 서비스를 제공하고 있다. 서울시의 경우 OKF의 CKAN 플랫폼을 참조 모델로 데이터 제공 방법을 모색 중이며 출연기관과 서울시의 자치구 데이터도 통합, 전체적으로 서울시 데이터 공개를 목표로 한다.
- 공공 서비스를 개발하기 위해 대규모 링크드 데이터 처리를 위한 MapReduce 기반의 분산 인덱싱 및 검색 기술 또한 개발 중이며 Open API를 활용하여 공공 빅데이터를 활용한 다양한 응용 서비스를 기획 개발하고 있다.

---

3) 공공 DB 피디아: 각 기관에서 생성된 공공정보를 링크드 데이터 형태로 변환하여 제공함으로써 공공 데이터의 접근이용 가공을 용이하게 만든 데이터 포털로 현재까지 시범서비스 중에 있다.





그림 36. 공공 링크드 데이터 기반 지식 플랫폼

#### 나. 미래 국내 공공 오픈 데이터 플랫폼 제안 (TO-BE)

- 정부에서는 공공 데이터 활용 유망 기업 육성 및 데이터 활용 경진대회, 인증 제도 운영, 창업 기반을 지속적으로 확충하는 형태로 적극적인 지원을 진행하고 있다. 데이터 기반 정책 수립을 위한 인프라 구축, 사회적 현안 해결을 위해 빅데이터를 활용한 정책 수립, 공공 기관간의 공동 연구 협력 체계 마련, 정부 부처간의 상호 협력 등의 여러 활성화 지원책이 운영되고 있는데 이 중 2014년 빅데이터 분석 역량 강화 지원 전략이 인상적이다.
- 즉, 국내외 미래전략 기관들, 싱가포르 RAHS, 영국의 FORESIGHT 프로그램 기관 등을 초청, 협력방안을 논의하며 한국형 분석방법론 개발 계획 수립을 완료하고자 함이 목적이다. 국내 실정에 맞는 예측모델, 분석 시나리오, 기법을 연구 개발하는 조직을 별도로 운영, 데이터 분석을 통한 미래 예측 및 전략 수립 과정 가이드 제작 및 보급계획에 있다. 그러나 온톨로지 기반의 LOD 체계 구축이 되지 않고, 공공 기관들간의 데이터 셋 URI 분담구조가 아닌 상황에서의 분석 기술 개발 및 플랫폼 구축은 장기적인 공공 개

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

방 빅데이터 기반 발전에 제약을 가져올 것이다. 따라서 UML을 기반한 프로덕트 라인 아키텍처 (Product Line Architecture)를 적용한 정부 (data.go.kr) 단일 채널 링크드 데이터 연계 플랫폼을 제안하고자 한다.

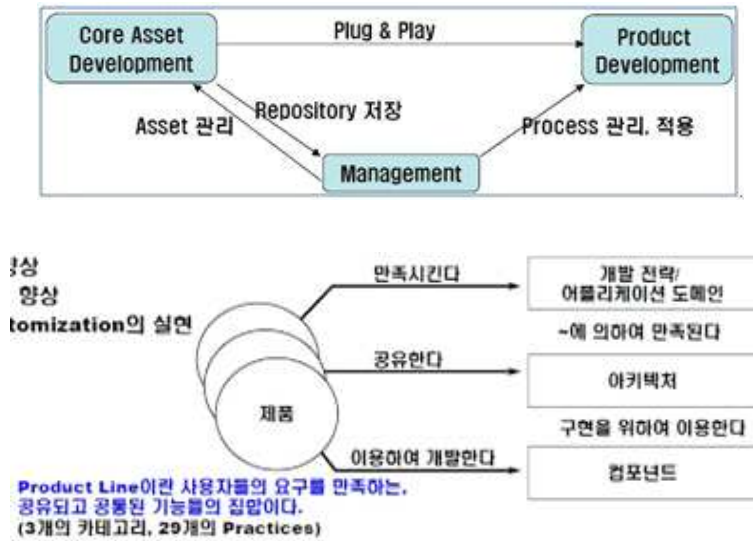


그림 37. 프로덕트 라인 구성도

표 4. 프로덕트 라인 구성요소

| 구분                                  | 내용  |
|-------------------------------------|---|
| Core Asset<br>(부품개발)<br>Development | <ul style="list-style-type: none"> <li>- Domain engineering / 핵심자산 개발과정</li> <li>- 플랫폼 개발, 공통부/가변부의 관리</li> <li>- 한 도메인에 속한 여러 어플리케이션들의 공통적인 Feature(기능)을 핵심 자산으로 모델링하고, 이를 재사용할 수 있도록 구체화하는 작업</li> <li>- 프로젝트 초반에는 해당 domain의 핵심적인 업무인 core process를 선행 개발하여, production을 위한 핵심전략 및 architecture의 prototyping 수행</li> </ul>                   |
| Product<br>Development              | <ul style="list-style-type: none"> <li>- Application Engineering</li> <li>- 제품 적용/개발, 제품 특화된 가변부의 적용</li> <li>- Core Asset을 각 어플리케이션의 요구사항에 맞게 Instance화하여 어플리케이션을 효과적으로 개발하는 과정</li> <li>- Core Asset에서 제공하지 않는 목표 어플리케이션의 일부 기능을 모델링하며 Instance화된 Core Asset과 통합하여 어플리케이션을 개발함</li> <li>- 안정적인 핵심 전략/방안의 정제과정을 위해서는 반드시 별도의 조직이 존재</li> </ul> |
| Management                          | <ul style="list-style-type: none"> <li>- 플랫폼 개발과 제품 적용/개발의 연동(프로세스/조직)</li> <li>- 성공적인 제품 계열 생산을 위해 기술적 측면과 관리적 측면에서 Core Asset개발과 제품 개발 활동을 관리</li> <li>- 기술적 관리: 형상관리, 프로세스 개선, 기술적 코칭</li> <li>- 조직적 관리: 조직관계, 운영, 조직계획, 반복 및 core asset과 Product development의 조율(orchestrates)</li> </ul>   |

- 프로덕트 라인이란 소프트웨어 공학에서 사용하는 의미로 특정 영역에 맞춰서 재사용할 단위인 핵심 자산을 미리 개발하고 실제 제품을 개발하는 것은 이미 재사용의 단위로써 만들어진 핵심 자산을 이용하여 여러 제품들을 만들어내자는 접근방법을 뜻한다. 즉 LOD 적용 플랫폼에서 등록 발행된 링크드 데이터를 재사용하고 데이터 자산을 응용, 활용하여 공공 빅데이터 가치를 확산시키는 의미와 부합될 수 있다. 첫째로 LOD를 구축하고, 개방·연계해 자유롭게 할 수 있는 지원 방안 마련 필요- LOD 성공의 관건은 서로 연결해서 이용을 촉진할 수 있는 데이터의 선별과 데이터 모델링이며, 이것

> 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

은 각 도메인의 전문지식을 보유한 전문가의 역할이 필요하다. 이 때 데이터 모델링은 기존 W3C 원칙을 개선한 UML(Unified Modeling Language)로 구현할 것을 권고한다. UML은 소프트웨어공학에서 사용되는 범용적 모델링 언어로써 플랫폼의 구조적 청사진을 시각화 하는 표준안을 제공할 수 있다. 가독성과 상호 운용성이 높은 UML은 온톨로지의 재사용성을 증대시킨다. 정부를 포함한 국민·기업·커뮤니티 등 다양한 주체의 참여를 유도하고, UML 기반 온톨로지 스키마를 통해 구축된 LOD에 적합한 데이터를 보유한 도메인에서 협력·지원을 받도록 한다.

- 누구나 자유롭게 LOD를 개방하고, 연계해서 활용해 창의적인 아이디어를 공유할 수 있도록 이러한 LOD 기반의 단일 공통 플랫폼 구축하는 것이 중요하다. LOD는 시작단계의 설계가 중요하며 LOD 데이터의 통합 지원과 관리 방안을 제시할 컨트롤 타워가 중심이 되어야 한다. 더불어 LOD 데이터 관리 센터는 데이터 연계의 적절성 점검, 기반 데이터의 연계 방법 제공, 온톨로지 설계의 적절성 확인, 개방된 데이터의 접근 경로 확인 등의 역할을 수행하게 된다.

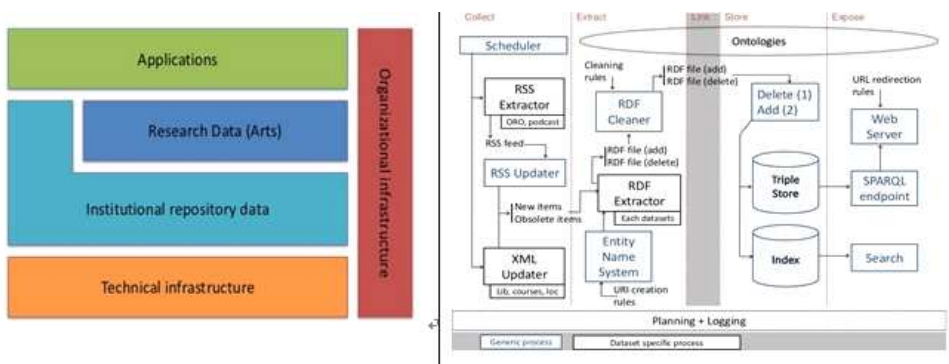


그림 38. 영국 Open University 링크드 데이터 스택 및 플랫폼 (data.open.ac.uk)

- 공공 빅데이터 개방과 활용을 목표로 한 링크드 데이터 플랫폼은 다음과 같은 고려 요건을 통해 UML 기반 프로덕트 라인 아키텍처 구현으로 적용하기를 권고한다.

## 1) 링크드 데이터 기반 공공 데이터 온톨로지 모델링

- > 공공 기관들은 각각의 도메인에서 적합한 데이터를 선정하여 링크드 데이터를 구축, 공유, 연결해야 하며 이를 통한 빅데이터 온톨로지를 구축해야 한다. 온톨로지란 도메인 내에서 공유되는 데이터들의 개념화 및 명세화된 규정을 의미한다. 클래스 분류, 속성 정의 등의 온톨로지 매핑 규칙을 사용한 연결화, 구조화, 맥락화에 초점을 둔 데이터 모델링을 진행하되 모델링 공유 템플릿을 통일화한다.

## 2) UML을 활용한 온톨로지 스키마 구성

- > RDF(Resource Description Framework)는 자원(Resources)을 속성(property)과 속성값(property value)으로 표현하는 링크드 데이터를 구성한 대표적인 모델로 W3C에서 정의한 메타데이터 기술과 호환을 위한 표준 규칙, 의미(semantics), 구조(structure), 구문(syntax)을 제공한다. UML을 활용한 온톨로지 스키마 구성을 위해서 서술 로직(Description Logic)을 적용한다. 서술 로직은 적용 예상 도메인의 술어 표현 지식 체계로써 특정 구문론 규칙을 보유한다. RDF, RDFS에 형식적 의미의 어휘를 추가하여 클래스와 속성에 대한 상속적 계층 구조 개념과 제약조건이 정의된 OWL (Web Ontology Language)인 경우 UML 구현이 보다 용이하다.

## > 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

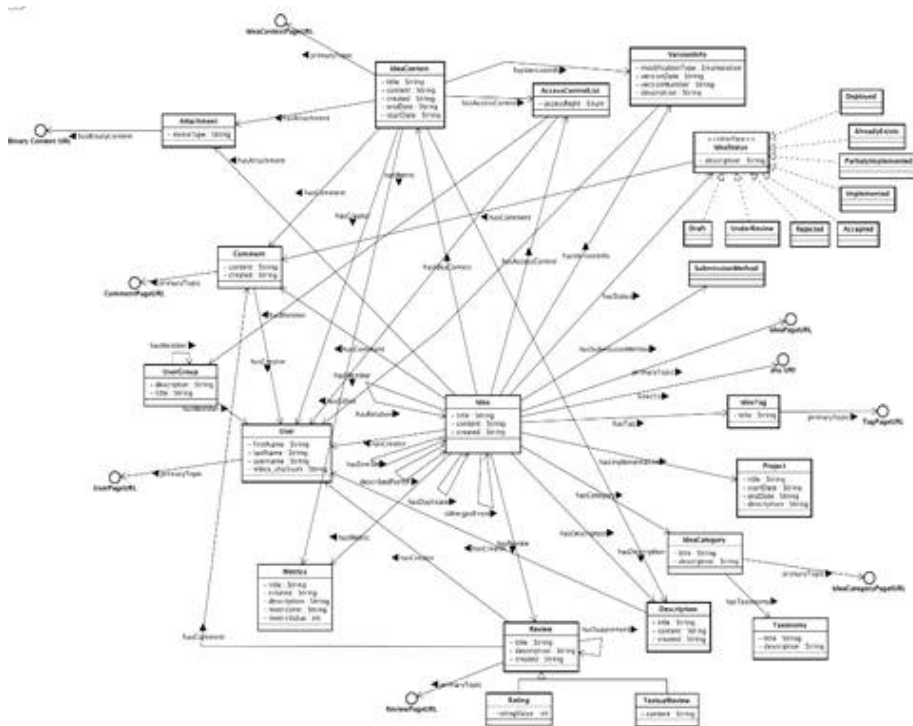


그림 39. GI2MO 온톨로지 UML 클래스 다이어그램 (<http://gi2mo.org>)

### 3) U프로덕트 라인 아키텍처 적용 LOD 아키텍처 관리 제안

링크드 데이터의 재사용을 위해 프로덕트 라인 아키텍처를 적용함으로써 상호 운용성의 특성을 지닌 LOD 아키텍처로 링크드 데이터 서비스가 구축될 수 있다. 링크드 데이터 생성, 변환, 관리에 따른 발행으로 진행되는 프로세스에서 프로덕트 라인 아키텍처의 C&V 모델을 참조하여 공통 피처와 가변 피처를 구분, 중복되는 데이터 셋의 메타데이터에 대해서 공통 피처로, 신규 데이터셋은 가변 피처로 분류, 공개 데이터 온톨로지 모델을 생성한다. 이를 실시간 또는 트리플 변환하여 저장, 운영 정책에 따라 관리하며 LOD 아키텍처 드라이버를 도출하여 중요도 데이터셋 분류화 및 변경 등의 히스토리 명세 내역을 관리함으로써 데이터셋의 생명 주기 관리 및 타 LOD와의 매쉬업 또는 협업 설계의 기반을 제공하는 기반을 제공하게 된다.



그림 40. 프로덕트 라인 아키텍처 적용 LOD 아키텍처 관리 제안

#### 4) 공공 기관간의 공유 협력 커뮤니케이션

- > 공공 부문간 공통된 플랫폼 연계 수립을 통해 데이터 공개 프로세스의 효율화와 표준 관리 체계 수립으로써의 장점을 얻을 수 있으나 이를 극대화하기 위해서는 공통으로 정보 공개에 대한 인식 전환과 공공성과 효율성의 균형을 지속할 수 있도록 협력을 지속할 수 있는 공공 기관들의 커뮤니케이션 조율이 중요하다고 볼 수 있다. 정보 공개를 통한 대중의 수혜를 제공하는 것에 중점을 두는 것이 아닌 정보 이용 권리에 응답한 개방과 재사용 방향을 제시해야 하며 도메인 별 공공 빅데이터 셋에 대한 링크드 데이터로의 전환 선택과 링크드 데이터 플랫폼의 체계적인 운영 프로세스를 수립, 상호 협의 하에 지속적인 개선을 추진할 수 있도록 인센티브적 정책마련이 필요하다.

## 제4장 사업 목표 달성도 및 국내 파급효과

본 사업은 해외 정부(영국 / 싱가포르 / 미국)의 빅데이터 분석 플랫폼 기술과 정책 동향을 진단하였으며 데이터 재사용성 활용에 초점을 둔 링크드 데이터 중심 플랫폼 방법론 제안에 중점을 두었다.

영국은 ‘5스타 오픈 데이터’ 기반 공공 데이터 LOD 프로젝트를 추진하면서 동시에 관련 표준 정립 및 활성화를 위한 법제 마련을 통해 즉시적 보다 점진적인 링크드 데이터 확충에 따른 분석 활용 인프라를 구축해 가고 있다.

싱가포르는 국가적 위기에 대처하는 과학적 분석 미래전략시스템을 구축, 운영하는 데에 빅데이터를 범 부처차원의 클라우드와 데이터를 기반으로 통합적 중장기 전략수립 위한 국가미래전략센터로써 지식정보 과학적 분석 미래전략시스템을 구축, 운영하고 있다.



## 제5장 사업 결과의 활용

정부에서는 공공 데이터 활용 유망 기업 육성 및 데이터 활용 경진대회, 인증 제도 운영, 창업 기반을 지속적으로 확충하는 형태로 적극적인 지원을 진행하고 있다. 데이터 기반 정책 수립을 위한 인프라 구축, 사회적 현안 해결을 위해 빅데이터를 활용한 정책 수립, 공공 기관간의 공동 연구 협력 체계 마련, 정부 부처간의 상호 협력 등의 여러 활성화 지원책이 운영되고 있는데 이 중 2014년 빅데이터 분석 역량 강화 지원 전략이 인상적이다.

즉, 국내외 미래전략 기관들, 싱가포르 RAHS, 영국의 FORESIGHT 프로그램 기관 등을 초청, 협력방안을 논의하며 한국형 분석방법론 개발 계획 수립을 완료하고자 함이 목적이다. 국내 실정에 맞는 예측모델, 분석 시나리오, 기법을 연구 개발하는 조직을 별도로 운영, 데이터 분석을 통한 미래예측 및 전략 수립 과정 가이드 제작 및 보급계획에 있다.

그러나 온톨로지 기반의 LOD 체계 구축이 되지 않고, 공공 기관들간의 데이터 셋 URI 분담구조가 아닌 상황에서의 분석 기술 개발 및 플랫폼 구축은 장기적인 공공 개방 빅데이터 기반 발전에 제약을 가져올 것이다. 따라서 UML을 기반한 프로덕트 라인 아키텍처 (Product Line Architecture)를 적용한 정부(data.go.kr) 단일 채널 링크드 데이터 연계 플랫폼을 제안하고자 한다.

## 참고문헌

- 1) Tim Berners-Lee, "Design Issues: Linked Data", 2006.  
<http://www.w3.org/DesignIssues/LinkedData.html>
- 2) Linking Open Data Project Wiki  
<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityPProjects/LinkingOpenData>
- 3) Irene Celino, "From research to business: the Web of Linked Data", 2009. 04  
<http://www.slideshare.net/iricelino/from-research-to-business-theweb-of-linked-data>
- 4) Tim Berners-Lee, "Putting Government Data online", 2009. 6.  
<http://www.w3.org/DesignIssues/GovData.html>
- 5) "Open Government: Berners-Lee and the UK to Show Obama How It's Done",  
posted by Marshall Kirkpatrick, June 10, 2009,  
[http://www.readwriteweb.com/archives/open\\_goverment\\_berners-lee\\_and\\_the\\_u\\_k\\_to\\_show\\_obam.php](http://www.readwriteweb.com/archives/open_goverment_berners-lee_and_the_u_k_to_show_obam.php)
- 6) "Moving Data.gov towards the Semantic Web", Posted by Paul Miller, August 10th, 2009, <http://blogs.zdnet.com/semantic-web/?p=319>
- 7) 스마트 국가 구현을 위한 빅데이터 마스터 플랜, 국가 과학기술위원회 외, 2012
- 8) 한국정보화진흥원, 국가지식 지능형연계체계 가이드라인. 서울: 한국정보화진흥원, 2012
- 9) 데이터 강국을 위한 국가정보화사업 추진방향, 김성태, 한국정보화진흥원, 2013
- 10) 빅데이터 시대: 효과적인 공공정보 개방을 위한 데이터 플랫폼 구축방향, 한국정보화진흥원, 빅데이터 전략연구센터, 2013
- 11) UK Government Futures Toolkit version 2014  
[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/328069/Futures\\_Toolkit\\_beta.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/328069/Futures_Toolkit_beta.pdf)

## 저자소개

이혜영 (現, 이탈리아, BricsKorea Co., Ltd. / 사업개발부)

## 주의사항

1. 본 보고서는 미래창조과학부가 시행한 기술정보사업 연구보고서입니다.
2. 본 보고서의 내용을 발표할 때에는 반드시 미래창조과학부가 시행한 기술정보사업의 연구 결과임을 밝혀야 합니다.
3. 국가 과학기술 기밀 유지에 필요한 내용은 대외적으로 발표 또는 공개해서는 안 됩니다.
4. 본 연구의 내용은 본 연구원의 공식적인 견해가 아닌 참여 저자들의 의견입니다.

### 빅데이터 분석 플랫폼 구축 마스터 플랜 모델링

|     |  |
|-----|--|
| 발행인 | 한선화  |
| 발행일 | 2015년 05월 20일 발행                                   |
| 발행처 | 한국과학기술정보연구원<br>대전시 유성구 대학로 245<br>Tel:042-869-1012 |
| 판매가 | 비매품  |









**KØSEN**

(<http://www.kosen21.org>)

대전광역시 유성구 대학로 245

Tel:042-869-1012