미래 농업·농촌을 위한 농업·농촌데이터 현황 분석 및 제언

이 지 민
서울대학교/연구조교수
habi1004@snu.ac.kr

1. 데이터의 중요성

과거와 달리 데이터의 수집이 용이해지고 데이터의 양 또한 급속도로 증가하고 있다. 다양한 기계, 장치, 센터 등을 통해 방대한 양의 데이터가 생성되고 컴퓨터 기술 발전으로 데이터 저장 및 분석에 소요되는 시간과 비용이 확연히 줄어들고 분석 및 처리 기술 다양화되어 활용 분야도 확대되었다. 유발하라리는 그의 저서에서 '호모데우스'에서 인간이 축적된 데이터와 알고리즘을 통해 결정하는 '데이터이즘'이라는 새로운 철학의 시대가 올 것이라 예측하였다.

지난 2016년 1월 다보스 포럼(WEF; World Economic Forum)에서는 제4차 산업혁명이 제시되었으며, 「The Future of Jobs(WEF, 2016)」에서는 과학기술측면에서 '모바일 인터넷', '클라우드 기술', '빅 데이터', '사물 인터넷(IoT)' 및 '인공지능(A.I.) 등이 4차 산업혁명의 주요 변화 동인이 될 것으로 예측되었다. 특히 4차 산업혁명의 동인으로 제시되는 A.I, 클라우드 등은 데이터를 기반으로 하는 기술로, 데이터는 4차 산업혁명의 뿌리라고 인식되고 있다. 과거에도 데이터는 중요한 역할을 하였으나, 데이터 양의 증가 및 처리의 용이성, 활용분야의 확대 등을 통해 그 중요성은 더욱 증대되고 있다.

국가 정책 측면에서도 데이터는 중요한 역할을 한다. 근거중심 정책의 기반으로 현상의 진단, 문제 도출, 해결방안 제시, 정책의 결과 확인 각 단계에서 활용될 수 있다. 이러한 데이터를 사람의 건강과 관련하여 살펴보면, 우리나라는 국민의 건강증진을 위해 건강검진을 실시하고 있는데, 건강검진을 통해 아픈 곳이 있는지 진단을 하고 이를 통해 예방적 효과를 갖는다. 질병을 발견하였을 시에는 치료를 진행하며 치료 중에서도 지속적으로 혈액검사, 관련 검사를 통해 치료효과를 확인 및 치료방법을 개선해 나가는 방법을 사용한다. 데이터는 국가단위, 지역정책 단위에서뿐만 아니라 개개인의 의사결정에도 활용될 수 있다. 농작물 생장과 관련해서는 생육정보 및 환경데이터를 통해 진단하고 조절을 통해 생산량을 증대시킬 수 있을 것이다(그림 1).

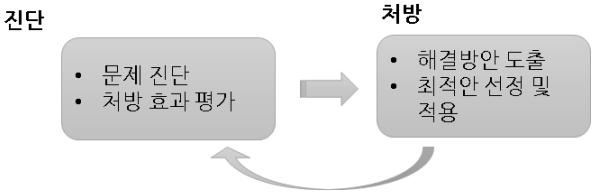

그림 1. 진단과 처방의 관계(출처: 김대종 외(2016)의 그림 참고하여 수정)

기후변화와 4차 산업혁명이라는 변화에 대응해야하는 현 시점에서 데이터는 중요한 자원의 역할을 한다. 본 원고에서는 농업농촌데이터 현황을 살펴보고 앞으로 빅데이터를 포함한 다양한 농업농촌데이터 구축 및 활용에 대한 제언을 제시하고자 한다.

2. 농업농촌 데이터 현황

농업농촌데이터는 농업생산 및 농촌공간과 연관된 다양한 데이터가 존재한다. 본 원고에서는 전통적인 데이터 형식인 통계자료와 빅데이터 현황 및 공간정보시스템으로 제공되고 있는 정보들을 살펴보고자 하였다.

2.1 국가통계

국가통계란 국가통계 작성기관이 정부 정책의 수립·평가 또는 경제·사회 현상의 연구·분석 등에 활용 목적으로, 산업·물가·인구·주택·문화·환경 등 특정의 집단이나 대상 등에 관하여 작성하는 수량적 정보를 말한다. 국가통계는 정부나 사회의 의사 결정에 기초 자료로 사용되는 사회공공재로 신뢰성이 중요하다. 국가 통계 중 농업농촌 관련 통계를 살펴보면, 기관별로는 통계청, 농림축산식품부, 농촌진흥청, 산림청, 농어촌공사, 축산물품질평가원, 농림수산식품교육문화정보원에 의한 44개 항목의 통계가 존재함을 알 수 있다(표 1). 그러나 이들 통계들은 각기 다른 조사주기, 조사체계, 공표범

표 1. 승인통계 중 농업농촌 관련 통계

기관명	조사명
통계청	가축동향조사, 귀농어·귀촌인통계, 농가경제조사, 농기판매및구입가격조사, 농림어업조사, 농림어업총조사, 농업면적조사, 농작물생산조사, 농축산물생산비조사, 어가경제조사, 어류양식동향조사, 어업생산동향조사
농림축산식품부	가공식품소비자태도조사, 과실류가공현황, 기능성양잠산업현황, 농기계보유현황, 농림업생산지수, 농업법인조사, 도축검사보고, 말산업실태조사, 배합사료생산실적및원료사용실적, 시설채소온실현황및생산실적, 식품산업원료소비실태조사, 여성농업인실태조사, 외식산업경기전망지수, 우유및유제품생산소비상황, 특용작물생산실적, 화훼류재배현황
농촌진흥청	농산물소득조사, 농어업인복지실태조사, 농업기계이용실태조사, 농업인의업무상질병및손상조사
산림청	국가산림자원조사, 목재수급통계, 목재이용실태조사, 산림기본통계, 산불통계, 임가경제조사, 임산물생산비조사, 임산물생산조사, 임업경영실태조사
농어촌공사	농업생산기반정비통계조사
축산물품질평가원	축산물등급판정통계
농림수산식품교육문화정보원	농업법인정보화수준및활용도조사

위를 갖고 있다. 국가적인 차원에서 농업 농촌현황을 파악하기엔 용이하나 총조사 자료 이외의 통계들은 소지역 및 세부적인 데이터로 활용하기에는 한계를 가진다.

한편 통계청에서 조사 집계되는 통계들은 데이터로써 의미와 함께 지표로써 의미를 갖고 있음을 알 수 있으며, 보고통계도 다수를 차지하고 있었다. 통계자료 또한 4차 산업혁명의 흐름에 맞추어 행정자료, 위성영상 등 외부 자료를 활용하는 방향이 변화하고 있으며, 최근 통계청은 새로운 통계작성이 필요할 경우 사전에 행정자료를 활용해 통계 작성이 가능한 지 여부를 판단하도록 통계법을 개정한 바 있다.

2.2 농림축산식품분야 빅데이터[1)]

농림업분야 빅데이터는 여러 기관 및 출처로부터 다양하게 생성되고 있다. 생산 분야의 스마트팜 데이터, 유통단계에서 도매시장가격 데이터, 소매단계에서 POS데이터, 소비단계에서 가계지출 데이터 등을 빅데이터 분석에 활용할 수 있을 것이다. 데이터 생성 기관 및 형태별로는 정부기관(농림축산식품부, 축산물품질평가원, 농촌진흥청, 산림청, 농협중앙회, 농림수산식품교육문화정보원 등)에서 생성한 정형데이터, 민간부문에서 생성된 정형데이터(농협 출하 및 판매 데이터, 스마트팜 데이터 등), 비정형 데이터(웹 데이터, 이미지 데이터 등)로 구분할 수 있다. 이 중 농업경영체정보, 소고기이력제정보, 농수축산물

표 2. 농림분야 빅데이터 종류(출처: 김경필 외(2016) 25페이지 표 2-6 일부)

구분	빅데이터	보유기관	공표주기
정부기관정형데이터	농업경영체정보	농림축산식품부	–
	축산농장정보	농림축산검역본부 가축위생방역지원본부	–
	축산차량이동정보	농림축산검역본부 가축위생방역지원본부	1일
	쇠고기 이력정보	축산물품질평가원	실시간
	돼지고기 이력정보	축산물품질평가원	실시간
	축산물 등급판정정보	축산물품질평가원	1일
	수입쇠고기 이력정보	농림축산검역본부	1일
	농축수산물 도매시장 상세 경락가격	농림수산식품교육문화정보원	1개월
	농수축산물 품목별 상세조사가격	농림수산식품교육문화정보원	1개월
	농산물 이력추적 정보	국립농산물품질관리원	실시간
	수입농산물 검사정보	국립농산물품질관리원	1년
	농업기상정보	농촌진흥청	실시간
	소비자패널정보	농촌진흥청	1년
민간데이터	농협 출하 및 판매데이터	지역농협, 농협 하나로마트	–
	농가데이터	개별농가	–
비정형데이터	웹데이터	포털사이트	–

* 규모(Volume), 빠른 생성 속도(Velocity), 다양성(Variety) 중 1개 이상 속성을 갖는 경우 빅데이터로 선정됨

1) 김경필, 구자춘, 안현진, 한정훈, 2016, 농림업분야 빅데이터 활용 제고방안, 한국농촌경제연구원 보고서를 요약정리하였음

경락정보, 소비자패널정보 등이 대표적이다(표 2).

이 중 가장 포괄적인 농업 빅데이터인 농업경영체정보는 농업경영체 등록제를 기반으로 형성된 데이터이다. 농업경영체 등록제는 2008년 6월부터 실시 2009년 12월 31일까지 일괄등록 기간을 가진 이후, 2010년 1월부터 상시 관리체계로 전환되었다. 2009년 「농어업경영체 육성 및 지원에 관한 법률」을 제정, 융자·보조금을 지원받고자 하는 농업경영체는 의무적으로 경영정보 등록하도록 강제화하였기에 정책대상 농가에 대한 전수조사를 실시하여 구축된 데이터베이스로 볼 수 있다(김한호 외, 2015). 농업경영체 자료는 행정자료와 구축된 특성을 가지며 농업정책 개발 및 성과측정에 활용될 수 있다(한석호 외, 2016).

이 외 다양한 빅데이터가 생성되고 있으나 정부주도로 체계적으로 저장 관리되지 못하는 데이터도 존재할 것으로 생각된다. 또한 농축산물 생산에 관한 데이터가 편중되어 있어 다각적 측면에서 데이터 발굴이 필요할 것이다.

2.3 공간정보시스템

1) 농지정보시스템(농지포탈)

농지정보시스템은 농지의 생애주기 관리를 위해 농지전용 통합DB를 구축하고 농지원부, 농지조서 등 농지 관련정보, 토지이용, 부동산, 지역, 농업 관련 정보를 함께 연계하였다. 농지공간포털에서 제공되고 있는 공간정보는 다음 표 3과 같으며, 농업생산 빅데이터의 기본이 될 수 있는 데이터라 할 수 있다.

2) 토양환경정보시스템(흙토람)

토양환경정보시스템은 「친환경 농어업 육성 및 유기식품 등의 관리지원에 관한 법률」에 기반을 두어 구축되었다. 작물별 토양적성도, 농경지화학성, 토양특성, 정밀농업기후도, 생물상

표 3. 농지공간포털 공간정보 현황(출처: https://njy.mafra.go.kr)

구분		축척	출처	갱신주기
기본도	연속지적도	1:5000	국가공간정보통합체계	반기
	행정경계	1:5000	국가공간정보통합체계	반기
	지형도	1:5000	국가공간정보통합체계	1년
	도곽선	1:5000	국가공간정보통합체계	1년
	항공영상	1:5000	농지정보시스템	2년
용도지역	농업진흥지역	1:5000	국가공간정보통합체계	반기
	영농여건불리농지	1:5000	국가공간정보통합체계	1년
	국토이용계획	1:5000	국가공간정보통합체계	1년
농지관리	농지전용(농식품부)	1:5000	농지정보시스템	실시간
	농지전용(시군구)	1:5000	농지정보시스템	1년
생산기반	일반경지정리	1:25000	한국농어촌공사	갱신없음
	대구획경지정리	1:25000	한국농어촌공사	5년
	간이경지정리	1:25000	한국농어촌공사	갱신없음
	대상제외지	1:25000	한국농어촌공사	갱신없음
	배수개선	1:25000	한국농어촌공사	5년

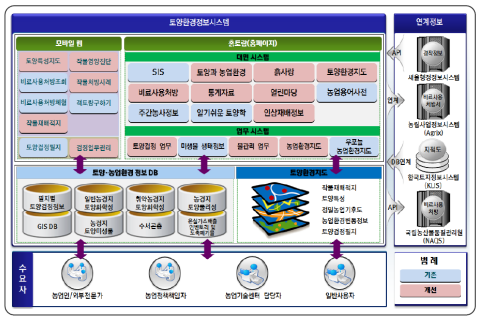

그림 2. 토양환경정보시스템 목표시스템(출처: 심재헌 외, 2016)

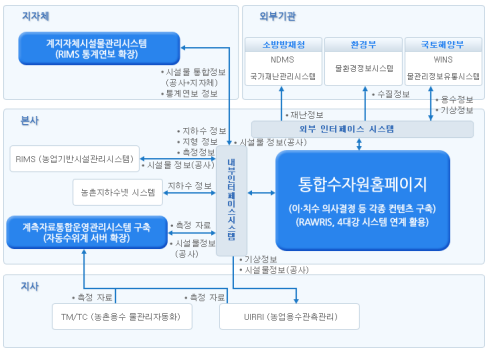

그림 3. RAWRIS시스템 구성도(출처: https://rawris.ekr.or.kr)

분포, 농업환경변동정보 등을 인터넷을 통하여 제공함으로써 영농인, 정책담당자, 내·외부 연구자, 일반 국민 등 다양한 수요자들이 쉽게 활용할 수 있도록 보급하고, 농업환경의 보전, 농산물 안전 생산의 전국적인 기반을 구축함을 목적으로 하고 있다(그림 2).

3) 농촌용수종합정보시스템(RAWRIS)

농촌용수종합정보 시스템은 농어촌 용수를 종합 관리할 수 있도록 '농업기반 시설', '재해정보', '수자원' 등의 다양한 정보를 통합 관리하는 시스템으로, 농촌용수 관리의 최적화를 목적으로 한다(그림 3).

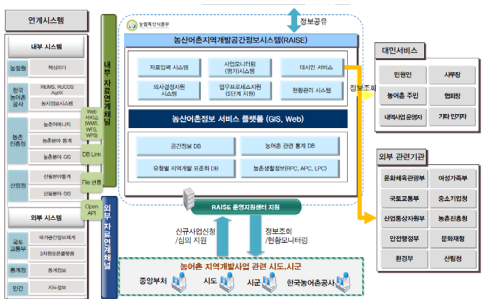

그림 4. 목표 시스템 구성도(출처: http://www.raise.go.kr)

4) 농산어촌지역개발 공간정보시스템(RAISE)

농산어촌지역개발 공간정보시스템은 일반농산어촌 지역개발사업 및 농산어촌 활용정보 DB를 구축하고 지역개발 사업의 효율적인 관리 및 기획을 지원한다. 제공되는 정보로는 지역개발사업 현황정보 및 농촌어메니티 자원 정보, 체험마을, 지역개발 우수사례 등이 있으며(그림 4), 농촌공간과 관련된 시스템으로 사업관리 및 지원 이외 농촌빅데이터로의 확대가 필요하다고 생각된다.

5) 스마트팜맵

정부는 농업경영체DB, 직불금 신청 정보와 GPS활용 위치정보를 연계·통합한 농경지 전자지도(일명, 스마트팜맵)를 구축함으로써 과학적인 직불제 이행점검 추진하고 있다(농림축산식품부, 2016). 농림수산식품교육문화정보원에서 구축한 스마트팜맵은 항공영상, 위성영상 등을 활용하여 경지구획도를 구축하였으며, 직불제 등록정보, 농업경영체등록정보, 흙토람 토양정보, 통계청, 농식품부 유관기관의 통계정보를 연계하여 59종의 공간정보를 제공한다(심재헌 외, 2016).

스마트팜맵은 대표적인 농업빅데이터인 농업

경영체DB와 농지 공간정보, 토양정보를 연계한 시스템으로 각기 구축된 데이터베이스 및 공간정보시스템을 일정부분 통합된 시스템이라 볼 수 있다.

3. 4차 산업혁명시대 농업농촌 데이터를 위한 제언

1) 새로운 데이터 확보

데이터는 4차 산업혁명의 근간으로 지금부터라도 데이터의 지속적인 축적 및 관리를 필요로 한다. 데이터는 하나의 자원으로 그 중요성이 증대되고 있으며, 새로운 농업 농촌 데이터의 확보가 필요하다. 이에 기존에 발생하고 있었으나 버려진 데이터에 대해서도 그 유용성을 살펴보아야 할 것이다.

○ 농업 데이터

4차 산업혁명시대는 초연결 및 초지능화의 시대이다. 농업분야도 다양한 과학기술과의 연계를 위해서는 가장 기본적으로 데이터의 확보가 필요하다. 농업 빅데이터로는 환경 정보, 생육 정보, 기상 정보, 농산물 유통정보, 농산물 생산량 정보, 농산물 가격 정보 등 다양한 데이터의 축적을 필요로 한다.

○ 농촌 데이터

농업빅데이터로는 농업경영체 DB가 존재하나 농촌빅데이터는 부재한 상황이다. 데이터는 데이터를 만드는 주체의 의도와 관점이 반영되므로 농촌정책 측면에서 필요로 한 데이터 수집이 필요로 할 것이다.

○ 데이터가 갖추어야 할 요건

앞서 통계조사 항목에서 살펴보았듯 데이터는 그 자체로 하나의 지표로서 역할을 하기도 한다. 지표는 대표성(representativeness), 신뢰성(reliability), 지속성(continuity,시계열적 분석이 가능, 주기적으로 자료갱신), 가용성(availability, 자료의 확보 가능)이라는 기본원칙을 지켜야 한다. 새롭게 조사 구축되는 데이터 역시 이러한 요건을 갖추어야 할 것이다.

○ 새로운 방식에 대한 고려

최근 통계청은 지역통계의 중요성 증대 및 조사 통계의 한계로 행정자료를 활용한 보고통계 생산 및 활용방안이 고려되고 있다. 농업 농촌 정책 및 사업 시행을 통해 수집된 다양한 행정자료가 농업농촌빅데이터의 원천이 될 수 있을 것이다. 또한 데이터 축적을 위해서 IoT 센서의 활용, 클라우드 기반 데이터 수집 시스템 등이 활용 될 수 있을 것이다.

농가단위, 지역단위가 아닌 새로운 형태의 플로우 데이터(혹은 네트워크 데이터)에 대한 고려도 필요한다. 플로우 데이터란 네트워크로 이동하는 사람, 사물, 정보, 자금 등의 흐름에 대한 시간과 위치정보를 가진 데이터를 의미한다. 대표적인 사례로는 개개인의 필수품이 된 신용카드와 모바일폰을 통해 얻을 수 있는 유동인구, 이동패턴, 소비에 관한 데이터가 존재한다. 현재까지는 데이터 구득 및 가공에 비용이 높고 분석결과에 대한 해석이 용이하지 않으나 앞으로 고려되어야 할 데이터 중 하나이다.

2) 데이터 분석기술을 통해 맞춤형 지식 제공

데이터는 이해할 수 있을 경우에만 가치가 생

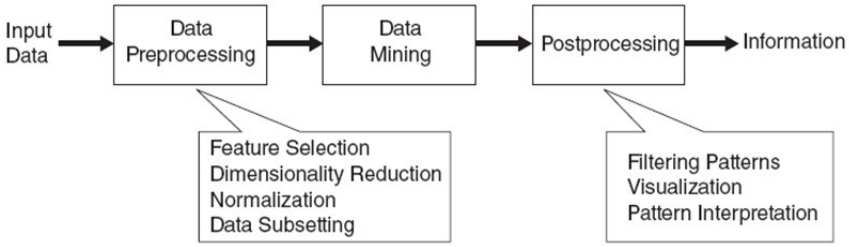

그림 5. 데이터 마이닝 프로세스(그림 출처 : Introduction to Data Mining(Pang Nin Tan et al., 2005))

성되며, 그 이전에는 임의적인 숫자에 불과하다. 데이터가 새로운 가치를 갖게 되는 것은 방대한 양의 데이터가 아니라 데이터를 수집, 저장, 분석하고 변환하는 기술에 있다. 이러한 과정을 통해 새로운 지식을 발견하고 새로운 관계를 찾고 예측할 수 있기 때문이다. 즉, 데이터마이닝[2] 단계에서 전처리와 분석을 거쳐 지식이나 지혜 단계에 이르러야 데이터는 비로소 의미를 갖게 된다. 농업농촌빅데이터도 데이터 확보뿐만 아니라 분석기술, 활용방안에 대한 충분한 연구가 필요할 것이다.

기존 데이터마이닝에서도 초기 데이터 전처리의 비중이 높았으나, 빅데이터는 방대한 양으로 인하여 원시데이터에서 유용한 데이터를 뽑아내기 위해서는 전체 과정의 80%의 노력이 필요할 수도 있다고 한다(그림 5). 따라서 빅데이터뿐만 아니라 스몰데이터도 적극 활용해야 할 것이다.

또한 데이터의 제공 단계에서는 맞춤형 지식이 제공되어야 한다. 즉, 데이터는 수요기관의 수요자 맞춤형 지식을 전달하여야 하며, 이러한 지식전달을 위한 지능정보 서비스를 개발하여야 할 것이다.

3) 위치기반 공간정보데이터베이스

앞서 살펴본 바와 같이 농업농촌데이터는 다양한 공간정보시스템을 통해 서비스되고 있다. 이는 공간정보와 연계하여 데이터를 제공 및 시각화해야 정보의 활용성이 높아지기 때문이다. 또한 위치정보 기반의 데이터베이스는 공간을 기반으로 다양한 자료의 통합 및 연계, 가공이 가능하여 이를 통한 공간적 의사결정을 위한 분석이 가능하다. 따라서 농업농촌 빅데이터는 위치기반 정보를 포함한 공간빅데이터로 구축되어야 할 것이다.

또한 데이터가 점적 위치정보나 일정한 경계를 갖지 않는 집계자료인 경우 행정구역단위로 생성될 경우 시계열적인 지속성을 확보하기 어려우므로 이러한 데이터의 경우에는 변화하는 행정구역이 아닌 격자단위 기반의 공간단위를 고려하여야 할 것이다.

4) 데이터 간의 통합 및 연계 중요

스마트 팜 맵의 경우, 농업경영체정보, 토양정보 등 다양한 정보가 통합된 시스템으로 구성되었다. 이처럼 다양한 데이터를 연계, 통합하여 제공될 경우 농업농촌에 대한 정확한 분석이

2) 데이터마이닝 (Data mining)은 대량의 데이터로부터 의미 있는 정보를 발견하는 것

가능하며 정책 개발 및 관리 등 활용 측면에서 효율이 더 높아질 것이다. 이러한 데이터의 통합 및 연계를 위해서는 각 데이터에 대한 충분한 이해가 필요하며, 서로 다른 기준에 의해 구축된 데이터 통합방안에 대한 모색 또한 필요하다. 한 예를 들어 농업경영체정보의 농가수는 농어업총조사 농가수가 다르게 나타나는데 이는 농업경영체에 등록되지 않은 농가수도 있으나 두 데이터의 조사기준이 다르기 때문이다. 또한 농업경영체 작물 정보를 지적도와 통합할 경우 하나의 지번에 여러 개의 작물 정보가 연결되는데 이는 밭작물의 경우 하나의 필지에 여러 작물이 재배되기 때문이다. 이와 같이 각 데이터는 정확하고 명확한 메타데이터가 존재해야하며, 여러 데이터를 연계 통합하여 사용할 경우 각 데이터에 대한 이해가 필수적이다.

사 사

본 원고는 농촌진흥청의 ICT융합 한국형 스마트팜 핵심기반기술개발 사업(과제번호:PJ012104, 과제명: 생산환경 빅데이터를 활용한 농경지이용 활성화방안 연구)의 연구 결과를 바탕으로 작성된 것임.

참고문헌

1. Pang Nin Tan, Michael Steinbach and Vipin Kumar, 2005, Introduction to Data Mining, Pearson.
2. 김경필, 구자춘, 안현진, 한정훈, 2016, 농림업 분야 빅데이터 활용 제고방안, 한국농촌경제연구원.
3. 김대종, 황명화, 정문섭, 2016, 국토진단을 위한 공개GIS SW 분석기술 동향 및 활용방안, 국토연구원.
4. 김한호, 이태호, 남대희, 우아미, 민진경, 2015, 농업경영체 등록정보를 활용한 농가 유형별 특성 분석 및 DB 구축 개선 방안, 농림축산식품부 보고서.
5. 농림축산식품부, 2013, 스마트 팜 맵 구축 방안 수립 최종보고서.
6. 농림축산식품부, 2016, “정부3.0 생활화로 농식품분야 일하는 방식 확 바꾼다.” 농림축산식품부 보도자료 2016.9.14.
7. 심재헌, 정도채, 유은영, 이정해, 2016, 농업 농촌 공간정보시스템 현황 및 표준화 방안, 한국농촌경제연구원.
8. 심재헌, 정도채, 유은영, 이정해, 2016, 농업 농촌 정책 지원을 위한 통합 공간정보인프라 구축 기초 연구(2/2차년도), 한국농촌경제연구원.
9. 한석호, 김수석, 채광석, 유찬희, 남경수, 정호연, 2016, 맞춤형 농정을 위한 농업경영체 DB분석 및 활용 연구, 한국농촌경제연구원.
10. 농산어촌지역개발 공간정보시스템 http://www.raise.go.kr
11. 농지포털 https://njy.mafra.go.kr/map/mapMain.do
12. 농촌용수종합정보시스템 https://rawris.ekr.or.kr
13. 흙토람 https://soil.rda.go.kr