
**2015년 「통계생산 현대화 워크숍 (Workshop on
the modernization of statistical production, MSP)」
참가 결과 보고**

2015. 5.

목 차

I. 개 요

1. 회 의 명 3
2. 회의장소 및 일시 3
3. 회의주관 및 개최배경 3

II. 회의 프로그램 5

III. 주요활동

1. 「한국 통계청 빅데이터 활용 진척사항과 향후계획」 발표 ... 8
2. 각국 사례 수집 및 토의 참석 9
3. 주요내용 10

IV. 시사점 검토

1. CSPA 도입 시사점 20
2. 그 외 시사점 22

V. 결론 및 후속조치

1. 워크숍 결론 및 각국 통계청에의 요청사항 24
2. 한국 통계청에의 요청사항 25
3. 후속조치 25

1. 회의명

- 통계생산 현대화를 위한 워크숍
(Workshop on the modernization of statistical production, 약자: MSP)

2. 회의장소 및 일시

- 회의장소 : 제네바 UN본부
- 회의일시 : 2015. 4. 15.(수) ~ 4. 17.(금)

3. 회의주관 및 개최배경

- 회의주관 : HLG(High Level Group) modernization committee
- 개최목적
 - 급변하는 정보기술과 통계방법론의 발전사항을 반영하여 통계생산방식을 보다 현대화할 방법을 국제적 관점에서 논의
 - CSPA 등 국제표준 준수와 공유를 통한 국제협력방안 논의
- 개최배경
 - 1) '14년까지 매년 개최되던 MSIS 회의의 포맷을 변경하여 올해부터 당해 워크숍(MSP)으로 대체
 - 기존 1박2일 MSIS 회의에서 각 국간 충분한 토론을 끌어내지 못했다는 판단에 올해는 기간을 연장하고 토론 위주로 세션 구성
 - 참여자의 직급은 과장 및 사무관급(mid manager)으로 전문가회의 성격을 가지게 됨
 - 통계정보시스템의 관리적 측면(management-MSIS의 M)보다 현대

화(modernization-MSP의 M)를 강조하여 토론 주제를 다양화

2) 한편, 이번 워크숍은 기존에 매년 개최되던 HLG 워크숍을 대체하지 않음

- 기존에는 HLG modernization committee 주관으로는 매년 1회 대표 워크숍이 개최된 것이 전부이나, 올해부터는 다양한 주제로 워크숍 및 컨퍼런스를 개최
- HLG modernization committee 회의는 초청을 받은 기관에 한해 비공개로 개최

<2015년 HLG 워크숍 일정>

- Workshop on big data(3.9)
- Workshop on the modernization of statistical production(4.14~4.17)
- Workshop on statistical communication(4.27~4.29)
- Workshop on data collection(4.29~5.1)
- Workshop on international collaboration for standards-based modernization(5.5~5.7)
- Conference of European statistician seminar "modernization of statistical production and service, and managing efficiency"(6.16)
- Work session on statistical data editing(9.14~9.16)
- Work session on data confidentiality(10.5~10.7)

○ 참가국 및 기관 : 총 37개 기관 약 60명

- 각국 통계청: 호주, 아저바잔(Azerbaijan), 불가리아, 캐나다, 크로아티아, 에스토니아, 핀란드, 독일, 헝가리, 아일랜드, 이스라엘, 이탈리아, 몽골, 네덜란드, 뉴질랜드, 노르웨이, 폴란드, 한국, 러시아, 슬로바니아, 스페인, 스웨덴, 스위스, 마케도니아(구 유고슬라브), 터키, 우크라이나, 영국, 미국
- 국제기구: EEC, Eurostat, ILO, IMF, OECD, PARIS21, UNESCO, UNIDO, WTO

II

회의 프로그램

<1일차(4.15,수)>

시간	내용	발표자
09:30	워크숍 개막	
09:45	주제1 소개: 통계생산에 영향을 주는 비즈니스와 IT 변화	Rune Gloersen, Lorna Drennen
09:50	(슬로베니아) 데이터 저장에서 데이터 서비스까지 - 확장가능한 데이터 아키텍처 구축	Tomaz Speh
10:00	(스웨덴) 변화 관리	Jakob Engdahl
10:10	질의응답	
10:20	(USDA/NASS) 연방통계청에서 사례관리서비스의 구현	Lorna Drennen
10:30	(네덜란드) ESSnet ValiDat Foundation 시스템 소개	Olav ten Bosch
10:40	(Eurostat) 통계생산 현대화	
10:50	그룹 토의 세션 소개	
11:00	휴식	
11:30	그룹 토의	
12:30	점심	
14:00	키노트 연설: 통계생산현대화의 주요한 도전 패널 질의응답	Paradraig Dalton
15:30	휴식	
16:00	그룹 토의 결과발표	
16:30	주제1에 대한 전반적 토의	
17:15	토의내용 요약	
17:30	1일차 폐회	

<2일차(4.16,목)>

시간	내용	발표자
09:00	1일차 내용 정리	
09:15	주제2 소개: 통계생산현대화에서 EA의 역할	Rosemary McGrath, Joe Treacy
09:20	(슬로베니아) SURS를 이용한 통계생산현대화	Andreja Smukavec
09:30	(터키) DDI 기반 통계생산플랫폼 Harzemli	Ilker Guven
09:40	(Eurostat) CSPA 준수(compliant) 가이드라인 구현을 통해 일반적 비즈니스 수요 명세	Hubertus Cloodt
09:50	(네덜란드) CSPA 준수하기 (becoming CSPA compliant)	Robbert Rensen
10:00	휴식	
10:10	(뉴질랜드) EA 작업이 매일의 업무에 영향을 주려면 어떻게 해야 하는가?	Rosemary McGrath
10:20	(영국) 영국 통계청의 EA 프로젝트	Ben Humberstone
10:30	(뉴질랜드) 통계기관을 위한 방법론 아키텍처	Rosemary
10:40	그룹 토의 세션 소개	
10:45	휴식	
11:15	그룹 토의	
12:00	그룹 토의 결과발표	
12:30	점심	
13:30	CSPA sprint 결과 - 프레젠테이션과 토의	
15:00	전반적 토의와 주재자의 요약	
15:30	휴식	
16:00	주제4에 대한 패널토의: 이 분야에서의 국제협력의 비용 및 편익	
17:30	2일차 폐회	

<3일차(4.17,금)>

시간	내용	발표자
09:00	2일차 내용 정리	
09:15	주제3 소개: 통계현대화에의 기회를 제공하는 기술 및 방법론의 혁신	Marton Vucsan, Carlo Vaccari
09:20	Data strand (스페인) 체계화/통합된 방식으로 데이터 수집과 선택적 데이터 에디팅. INE의 추진 경험	David Salgado
09:30	(한국) 한국 통계청의 빅데이터 활용 진척상황과 향후계획	이선희
09:40	(UNECE) 빅데이터 프로젝트 Sandbox: 어디까지 진행되었는가?	Bruno Voisin, Niall Wilson
09:50	휴식	
10:00	Process strand (캐나다) 각국 통계기관의 기계학습 기술 도입 현황 (machine learning documentation initiative)	Claude Poirier
10:10	(이탈리아) CORE: CSPA 구현 시스템	Marco Silipo
10:20	(UNIDO) 공식통계와 조사방법론에서 R 사용	Valentin Todorov
10:30	그룹 토의 세션 소개	
10:40	휴식	
11:00	그룹 토의	
12:00	향후 제안 도출, 우선순위 설정	
12:30	모든 세션에 대한 일반적 결론 도출, 제안사항과 향후계획	
13:00	워크숍 종료/점심	

1. 「한국 통계청 빅데이터 활용 진척사항과 향후계획」 발표

- 주제3(통계현대화에 영향을 미치는 기술혁신)과 관련하여, 한국 통계청에서 수행한 과거 빅데이터 프로젝트 구축내용과 활용사례를 소개하고 빅데이터를 활용한 통계생산체계 구축을 위한 고려사항 및 향후계획을 제시
 - 통계청은 통계의 시의성에의 요구와 변화된 통계조사환경을 반영, 빅데이터를 통계생산 및 서비스에 활용하는 방안을 고민해 왔으며 2012년부터 매년 파일럿 프로젝트를 수행
 - 작년에는 GPS 위치정보를 사용하여 일일인구이동 분석시스템을 구현하였으며, 장기적인 빅데이터 활용 통계생산체계 구축을 위해 연구과제를 수행하였음
 - 빅데이터 활용 통계생산체계 구축을 위한 대표적인 이슈로는, ① 법적 이슈, ②비용효율성, ③사생활 노출, ④방법론, ⑤데이터 관리, ⑥IT 기술 활용 등이 있으며 그 밖에도 기존 GSBPM 기반의 통계생산 프로세스를 빅데이터를 다룰 수 있도록 수정하거나 통계방법론을 정립하는 작업이 선행되어야 함
- 우리 청의 빅데이터 프로젝트에 대해 많은 국가가 흥미를 보였으며, 특히 작년 일일인구이동 분석시스템에 많은 관심을 보이며 활용사례 확대 방안 등 다양한 의견을 제시
 - 일일인구이동분석시스템을 모바일시스템으로 제공할 것을 제안. 특히 아이들 교통안전을 위한 모바일 앱 개발 등의 활용사례 확대방안을 제시
 - 통계법 개정을 통해 다른 기관의 데이터를 가져올 근거를 만들면 빅데이터 프로젝트 추진이 수월할 것이라는 의견

- 빅데이터 만으로는 신뢰할 만한 결과를 얻을 수 없으므로 다른 데이터소스를 함께 이용하여야 한다는 의견. 관련하여 한국 통계청의 다중데이터소스 분석 방법론 질의
- 다른 기관의 빅데이터 프로젝트와 겹치는 문제가 발생할 수 있다는 질의
- 향후에도 온라인 CPI 생산과 같이 빅데이터 기반 지수를 생산할 계획이 있는지 질의
- 빅데이터 공통기반 관련 질의 및 sandbox 프로젝트 참여 요청

2. 각국 사례 수집 및 토의 참석

- 대부분의 일정은 토의 중심으로 진행되었으며, 토의 시간을 확보하기 위해 프레젠테이션은 1인당 10분으로 엄격하게 제한
- 프레젠테이션 및 토론에서는 각국의 EA 및 SOA 적용사례를 중심으로 통계현대화에 영향을 미치는 여러 요소를 종합적으로 논의
 - 특히 이번 워크숍의 주요목적이 CSPA의 진척상황을 보고 및 홍보하는 것에 있기 때문에 CSPA 논의에 많은 시간 할당
 - 그 외에도 DW, 모바일, 소셜미디어, 클라우드, 빅데이터, 기계학습(machine learning) 등 IT 신기술이 통계생산에 미칠 수 있는 영향에 대해 광범위하게 논의하였으며, IT 관점이 아닌 통계방법론 관점에서 통계현대화를 추구할 방법에 대해서도 논의
- 그룹 토의 세션은 전체 참가자를 4~6개의 그룹으로 나누어 각각의 주제에 대해 토의를 진행 후 결과를 종합하는 방식으로 진행하였으며, 한국은 그 중 클라우드 컴퓨팅.방법론 지식관리.전반적 기회 등 3개의 주제에 대해 토론 참가

3. 주요내용

1) 키노트 연설 및 CSPA sprint 추진내용 보고

○ (키노트) 통계현대화를 위한 도전

- 통계현대화를 위해 국제협력의 중요성을 강조하며 HLG MOS (HLG Modernisation)의 활동을 소개

.CSPA 구현, 빅데이터 플랫폼(sandbox), 표준(GAMSO 등), 모바일, 기계학습 등

.HLG 산출물에 대한 브랜딩에 힘쓰고 있으며 작년 뉴욕과 캔버라 회의를 계기로 '통계현대화커뮤니티(Statistical Modernization Community)'를 설립

.통계현대화커뮤니티 회원은 11월 연례회의에 참석하며(invitation only), 참가 의사를 밝히는 'declaration of intent' 및 CSPA, 빅데이터 등 표준 활동에 동참할 것을 서약하는 Manifest에 서명함

.또한 역량 공유 방안에 대한 아이디어가 있으며 UNECE와 함께 역량 레지스터를 '지식 브로커'라는 이름으로 개발 예정

- 그 밖에 현대화를 위한 주요 장애물을 소개하며 특히 조직문화 및 변화에의 저항 극복의 중요성을 강조

.통계생산의 현대화 과정에 IT 부서만이 아닌 통계생산부서(methodologist)를 참여시키는 것이 중요

.새로운 기술에 대응할 수 있는 역량을 가진 data scientist를 조직 차원에서 개발할 필요

○ (CSPA sprint) 호주 캔버라에서 지난 3월 개최한 CSPA sprint 회의의 결과 보고

- CSPA 프로젝트는 HLG 후원으로 3년째 진행중인 프로젝트로, SOA (service-oriented architecture) 개념을 채택한 EA 레벨의 국제표준

※ SOA(Service-oriented Architecture): 전체 시스템을 하나하나의 서비스 모듈의 합으로 정의하는 개념의 아키텍처. 각각의 서비스모듈은 재사용가능하게 기능단위로 설계되며 서비스모듈간의 결합은 표준인터페이스로 느슨하게 연결되어 필요에 따라 다른 모듈로 바꿔 끼우는 과정이 용이함.

.**(목적)** CSPA의 궁극적 목적은 각국 통계기관에서 개발된 CSPA 기준을 준수하는 서비스모듈을 한 곳에서 모아 서비스하는 것으로, 각국은 같은 기능을 가진 통계서비스모듈을 개발하지 않고 공유하여 사용할 수 있음

.**(현 상태)** 각 통계기관에서 자발적으로 제공하는 서비스모듈에 대해 HLG에서 매년 CSPA 준수 여부를 평가하여 HLG 위원회와 공유하고 있으나, 널리 공유하고 있지 않으며 체계적으로 우선순위를 두어 CSPA 준수 서비스모듈을 개발하고 있지 않음

.**(필요성)** ① 서비스모듈을 개발하는 데 드는 자원을 절약하는 것 외에도 ② 서비스 기반으로 업무체계를 전환함으로써 통계생산 부서와 IT 부서와의 소통의 향상도 기대할 수 있음

.**(주요 장벽)** CSPA 도입은 IT 부서의 업무뿐만 아니라 통계생산 부서의 업무에도 전반적인 변화를 불러올 수 있음(IT 부서는 서비스모듈을 개발하고 통계부서는 그 모듈을 요구/선별하여 조합하는 능력이 요구됨). 따라서 조직구성원의 저항이 있을 수 있음

.**(다른 표준과의 관계)** CSPA는 GSIM, GSBPM, DDI/SDMX를 기반으로 개발된 참조 아키텍처임

- 회의에는 총 9개 기관이 참가하였으며, 결과물로 포트폴리오 관리 틀·브로셔 및 프레젠테이션·페이퍼 및 성명문(Manifest) 작성

.특히 성명문은 각국 통계청이 CSPA 준수에 협조하겠다는 의지를 담은 문서로, 6월 기관장 회의에서 서명을 받을 예정

.앞으로도 계속 CSPA에 관심과 참여를 독려

2) 각국 통계청 및 국제기구 사례발표

가. CSPA 관련

○ (Eurostat) CSPA 준수 가이드라인 개발

- 현재 CSPA 서비스모듈 정의 방법을 가이드라인으로 개발 중(CSPA 활성화 1단계)

.가이드라인 제작 후 그에 따라 제작된 첫 번째 SOA 서비스를 이용, 개념을 증명(proof of concept)

.이후 각국 통계기관들의 일상적인 CSPA 서비스 이용을 위해 가버넌스 체계 구축(CSPA 카탈로그 등)

- CSPA 가이드라인은 개념적 정의, 논리적 명세, 물리적 구현의 3단계로 구성되어 있으며 CSPA 1.0 버전을 준수

○ (네덜란드) CSPA 준수하기

- CSPA의 공급자(서비스모듈 설계자), 수요자(이용자) 측면 접근
- CSPA 개발자는 다양한 분야의 전문성을 가져야 함(통계방법론, 비즈니스 분석, 소프트웨어 설계 등)
- CSPA 적용에 대한 보상은 없으며, CSPA의 장점을 보여주어 설득
- CSPA의 장점은 ‘컴포넌트 단위로 사고함으로써 모두가 같은 아이디어를 공유하고 같은 언어를 쓸 수 있다는 것’

○ (이탈리아) CSPA 구현시스템 CORE 소개

- 이탈리아에서 작년 개발완료한 CORE는 CSPA 준수 서비스모듈을 간단한 인터페이스로 조립하여 전체 통계프로세스에 따라 작동하는 시스템을 실시간 구축할 수 있는 플랫폼
- 각 서비스모듈은 아이콘으로 표시되고 사용자는 아이콘을 Drag & Drop 하여 간단하게 조립 후 바로 업무처리를 실시할 수 있음
- 이 시스템 도입으로 IT 뿐만 아니라 통계업무자의 업무내용도 변하게 되며 관련하여 이탈리아 통계청은 6월 조직개편 예정

○ (스웨덴) 통계생산아키텍처 변경 시 변화관리 사례 소개

- CSPA 도입을 위해서는 SOA 패러다임을 받아들여 시스템 구현 및 활용의 방식을 변경할 필요가 있음
- 이는 시스템, 업무 등 조직 내 전반적인 부분에서의 변화를 수반하므로 조직구성원들의 저항이 있게 됨. 이 저항을 극복하는 것이 CSPA 구현의 핵심적인 부분

나. IT 신기술 관련

○ (UNECE) 빅데이터 플랫폼 Sandbox

- 원격접속으로 데이터 처리가 가능한 빅데이터 플랫폼 구상
- 4 node 구성의 하드웨어와 하둡, 인메모리DB, R을 기반으로 한 소프트웨어 구성
- 현재는 하둡 기반의 작은 시스템에 불과하며, 향후 이용사례에 따라 발전방향을 설계. 통계생산환경 구성과 인공데이터(synthetic data)를 이용해 제3자와 협업을 강화하는 방안을 검토하고 있음

○ (캐나다) 기계학습 문서화계획

- 인공지능 처리방식 중 하나인 기계학습(machine learning) 기술을 통계생산에 활용하는 사례를 국가별로 수집하여 문서화
- 로지스틱 회귀분석, 지원벡터머신, 클러스터 분석 등의 기술이 주로 활용되며 자동코딩·에디팅·추론·무응답 추정 등의 분야에 적용
- 알려진 사례는 영국, 아일랜드, 독일, Eurostat, 포르투갈, 뉴질랜드, 미국, 이탈리아, 캐나다, 헝가리이며 한국은 아직 기계학습을 통계에 활용한 사례는 없음

다. 그 외

○ (Eurostat) 통계생산 현대화

- MSIS 2014에서 처음 소개했던 IT rationalization 계획(각국이 더

높은 IT 서비스 성숙도를 가지고 정보화 목표를 달성할 수 있게 하고자 하는 계획)의 진행상황을 보고

- IT 전략 관점, 비즈니스 프로세스 조화 관점, IT 인프라 관점, IT 포트폴리오 관점, IT 호스팅 서비스 관점 각각에 대해 진행 중

○ (네덜란드) ESSnet ValiDat Foundation

- 통계생산 시 논리적 검증(validation) 프로세스를 표준화하고 검증 규칙을 명세할 수 있는 표준언어 VTL(validation transformation language)를 개발해 국가간 공동활용할 수 있는 통계 내검시스템을 개발하고자 하는 프로젝트
- 독일이 주도하며 이탈리아, 리투아니아, 네덜란드, eurostat이 참여

○ (뉴질랜드) 통계기관을 위한 공통 방법론 아키텍처 개발사례

- 통계방법론과 관련 틀을 아키텍처로 관리함으로써 의사결정과 효율적인 투자를 지원

○ 그 외 시스템 구축사례 소개

- (슬로베니아) SOA 기반 통계시스템 구축사례(SURS)
- (USDA/NASS) 연방통계청 사례관리서비스(case management service) 구현사례
- (터키) DDI 기반 통계생산 플랫폼 구축사례(Harzemli)
- (영국) 영국통계청 EA 구축사례
- (스페인) 스페인통계청 EA 구축사례
- (UNIDO) 공식통계와 조사를 위해 R 사용

3) 분임토의 주제 및 주요결론

○ (Data warehouse) 최근 떠오르는 BI(Business Intelligence) 기술을 통계 생산에 적용하기 위한 방안은?

※ BI(Business Intelligence): 원시자료를 비즈니스 분석에 활용할 수 있는 유용하고 의미 있는 정보로 바꿀 수 있는 툴과 기술의 집합체를 의미하는 정보화용어. 리포팅, OLAP, 데이터마이닝, 프로세스마이닝, 텍스트마이닝, 복합이벤트프로세싱, BPM, 벤치마킹, 예측분석 등을 포괄

- 국제협력이 중요함. 각국 통계청에서 BI 솔루션 업체에 접촉할 때 개개의 기관 관점에서 협상하는 것보다 여러 국가가 연합하여 커뮤니티로서 협상하면 요구사항을 보다 강하게 관철할 수 있음
- 국제표준에 대한 인식을 강화할 필요가 있음(GSBPM/GSIM/CSPA)
- IT 파트와 통계생산부서(methodology)의 협력 강화

○ (Business/technology agility) 통계기관이 업무 민첩성(Business agility)을 획득하기 위한 방안은?

※ Business agility: 조직 및 시스템이 빠르게 변화에 적응하는 능력. 특히 시장/환경변화에 생산적/비용효율적으로 빠르게 적응하는 능력을 가리킴

- 사례학습과 워크숍을 많이 개최해 각 기관의 정보를 공유하자는 제안
- 국제적 레벨에서 민첩한 기업에 대한 사례를 공유할 필요가 있음
- 현대화를 위한 가이드라인(recipe)을 만들자는 제안
- LinkedIn 커뮤니티를 더 만들어 국제소통하자는 제안

○ (Mobile and Social media) 통계기관이 모바일/소셜미디어 혁명에 대응할 방안은?

- 모바일 데이터를 통계생산에 활용하는 사례 프로젝트를 진행
- 모바일 어플리케이션에 활용할 수 있는 오픈소스 CSPA-준수 서비스를 개발할 것(특히 데이터 수집과 배포 분야에)
- 모바일 데이터를 위한 표준 개발
- 각국 공식통계에의 모바일 데이터/기기 활용에 대한 설문조사 실시

○ (Cloud computing) 클라우드 기술에 대한 통계기관의 대응방안, 특히 IaaS, PaaS, SaaS 각각에 대한 각국사례를 토의

※ IaaS(Infrastructure as a Service): 인프라 레벨에의 클라우드 컴퓨팅. 서버 등 자원을 공유. 전산통합센터가 이에 해당

PaaS(Platform as a Service): 플랫폼 레벨에의 클라우드 컴퓨팅. 어플리케이션을 개발할 수 있는 플랫폼 제공. 빅데이터 공통기반이 이에 해당

SaaS(Software as a Service): 어플리케이션 레벨의 클라우드 컴퓨팅. 소프트웨어를 직접 개발하지 않고 사용료를 지불하고 인터넷으로 공유하여 쓰는 개념

- 타 산업에서 사례를 찾을 필요
- 통계산업에의 클라우드 컴퓨팅 적용의 현 상태와 관련하여 ABS에서 편찬한 논문이 있음. 이를 업데이트하여 주기 바람

○ (CSPA 도입 확산) 국제사회에서 CSPA 도입을 확산할 방안은?

- HLG 레벨에서 통계기관들과 협력하여 CSPA 홍보할 것
- CSPA 도입을 위해 통계기관들이 협력하여 사업자들과 협상 진행
- CSPA 관련하여 국제적 소통과 현재까지 진행상황에 대한 시각화를 강화할 것
- CSPA 계획에 대해 국제적 재정지원을 유지할 것
- CSPA를 준수하는 방법에 대해 다른 통계청 사례를 포함한 세부 가이드라인(cookbook) 마련
- CSPA 도입을 위해서는 기관장 지지가 필요하며, 해당 조직구성원을 설득하기 위한 정보를 HLG에서 제공할 필요가 있음

○ (통계방법론) 통계생산현대화를 위해 통계방법론이 수행해야 할 역할은?

- 앞으로 통계전문가와 IT 전문가가 함께 참여할 일이 더 많을 것
- 통계방법론 부서와 IT가 모두 참여할 수 있는 연구프로젝트 수행
- 공통의 비전을 가장 낮은 레벨에서 높은 레벨까지 창출할 것
- 통계방법론과 IT 간 소통의 효율성을 위해 6 sigma를 툴로 활용
- 통계방법론 분야에서도 GSBPM 같은 표준을 개발

- (전반적 기회) 통계정보화 분야에서 향후 5년 후에 다가 올 전략적 기회로는 무엇이 있을까?
 - 새로운 데이터 소스를 다룰 수 있는 방법과 기술 개발
 - GSBPM 관련하여 정규 미팅 개최
 - 오픈데이터의 기계 to 기계 전송을 다루는 표준 개발

- (데이터 입력처리) 통계기관이 다루는 데이터의 종류가 늘어나면서(빅데이터, 행정자료) 데이터 전처리의 중요성이 부각되고 있음. 문제와 해결방법은? 기계학습 기술을 사용한 데이터 전처리 자동화 방안은?
 - 이상치(outlier) 탐지와 관리를 위한 표준화된 접근 필요
 - 자료 매핑과 코딩, 레코드 링킹 접근법 개발 필요
 - 데이터 전처리의 우수사례를 문서화하여 공유
 - 자료 추정의 효과를 평가할 방법 개발

- (빅데이터) 통계청은 빅데이터를 활용할 준비가 되었는가? 한편 빅데이터 제공자가 통계청의 업무를 대체하지는 않을까?
 - 아직은 각국 통계청이 빅데이터를 활용할 준비가 되어 있지는 않으나 통계청 고유의 업무를 데이터 제공자가 대체하진 않을 것
 - 빅데이터를 다루는 새로운 방법론을 개발할 필요

- (기계학습) ① 인간과 소통할 수 있는 인지통계시스템을 만들 수 있을까? ② AI 기술을 통계생산에 이용할 수 있는 방안은? 시스템이 자동으로 인터넷에서 데이터를 가져오거나 패턴인식을 통한 데이터 클렌징 등
 - AI를 활용한 지능적 데이터 클렌징 기법 개발
 - 국가계정이나 유로그룹 레지스터와 같은 영역에 예측분석 적용 가능성을 검토해 볼 것
 - 캐나다 논문(기계학습 사례분석)을 더 상세하게 편찬·공유할 것

- 추정(imputation)의 자동화 기술 개발을 위한 경연대회 개최
- 가공 데이터 생성기(synthetic data generator)를 개발
- 고용·국가계정과 같은 대표통계 지원을 위해 데이터마이닝 기술 검토
- 데이터의 바다에서 메타데이터를 찾기 위한 인공지능시스템 개발
- 빅데이터 프로젝트에 기계학습 기능을 포함
- 과학계와 협업활동을 강화(위성 데이터 이용 등)

○ (CSPA의 발전) 국가통계청 간 CSPA 모듈을 효과적으로 주고받을 방법은? CSPA 모듈간 물리적 호환성을 강화할 확실한 방법은?

- CSPA 컴포넌트 저장소(CSPA 기준을 따라 제작된 서비스모듈을 공유하기 위한 포털)를 계속해서 개발, 특히 라이선스 문제를 포함한 관리 이슈를 다룰 수 있도록 해야 함
- 이탈리아와 Eurostat에서 해 온 일을 바탕으로 CSPA 서비스에 대한 구체화 필요

4) 패널토의: 통계분야에서 국제협력의 비용/편익

- 국제협력의 목적 중 하나는 비용 절감
 - 빅데이터 Sandbox가 좋은 예로, 자체 플랫폼을 구축할 여력이 없는 통계기관에 자원을 제공할 수 있음
 - 그러나 목적 달성을 위해서는 오랜 기간 투자가 필요하므로 관련 기간의 협조를 요청
- 새로운 분야가 도래했을 때 역량을 집중하여 앞서나갈 필요
 - 빅데이터와 같은 분야가 대표적인 예로, 지금이 국제협력의 시기
 - 국제협력에 있어 'why?' 대신 'why not?'의 자세가 필요
- 국제표준 도입의 중요성
 - 각 통계청에서 개별로 제시한 표준보다 국제협력의 결과로 제시된 표준이 통계기관에서 받아들여지기 더 쉬움
 - 개발도상국은 기존 시스템이 없으므로 표준을 전파하기 더 좋은 환경
 - 새로운 시스템을 개발하면서 국제표준을 도입하지 않는 것을 제제할 방안(CSPA에 기반하지 않는 시스템을 illegal로 해야 한다고 표현)
 - 단순히 틀만 제공하는 게 아니라 교육이 수반되어야 함
- UNECE에의 지원 요청
 - UNECE의 역할은 국제협력을 촉진하는 것이며 국제 투자 펀드를 관리하는 역할까지 확대되어야 함
 - 이 역할 수행을 위해 더 많은 자원이 필요

1. CSPA 도입 시사점

○ (CSPA 개념 이해) CSPA는 기관에서 SOA 개념을 도입하여 EA를 운용하고 있음을 전제로 하는 국제표준

※ SOA(Service-oriented Architecture): 전체 시스템을 하나하나의 서비스모듈의 합으로 정의하는 개념의 아키텍처. 각각의 서비스모듈은 재사용가능하게 기능단위로 설계되며 서비스모듈간의 결합은 표준인터페이스로 느슨하게 연결되어 필요에 따라 다른 모듈로 바뀌 끼우는 과정이 용이함.

- 기존의 정보시스템 아키텍처 하에서는 각 조사시스템이 조사의 모든 과정을 처음부터 끝까지 하나의 시스템에서 처리할 수 있도록 설계

- SOA 개념을 도입하면 조사시스템은 하나의 통으로 설계되는 대신 기능별로 쪼개어 각각의 기능을 수행하는 작은 서비스모듈의 합으로 설계됨

.만들어진 서비스모듈은 공통저장소(카탈로그)에 보관되며 나중에 다른 시스템 제작에 재사용되거나 다른 기관과 공유하여 이용·판매도 가능

.통계생산부서는 정보화부서에 전체 시스템 개발을 요청하는 대신 카탈로그에서 기존 서비스모듈을 조합해 새 시스템을 직접 제조 가능

.이 경우 정보부서의 주역할은 특정기능을 가진 서비스모듈을 보충해서 카탈로그에 올리는 것

- CSPA는 SOA 개념을 따르는 시스템을 설계할 때 서비스모듈 간 메시지를 주고받는 인터페이스와 중개자모듈의 설계방식을 정의하는 표준

○ (CSPA 도입 필요성) 장기적으로 비용절감 측면에서 CSPA 도입검토가 필요하며, SOA 도입을 통한 IT-통계생산부서간 소통 강화 측면 및 국제협력 강화 등의 장점이 있음

- CSPA의 주목적은 각국 통계기관간 서비스모듈을 국제 카탈로그로 공

- 유하여 통계시스템 신규구축 시 개발비를 절감하는 것이나, 한국은 아직 비용절감 목적의 SOA 도입필요성에 대한 충분한 공감대는 없음
- 한국은 정부3.0 패러다임 하에 정보화 시범사업 등 예산지원 프로그램이 많으며 정부차원에서 클라우드, 빅데이터 플랫폼 등을 구축하여 통계기관끼리의 정보화협력이 절실하지 않음
- 특히 정부기관에서 SOA 활용이 많지 않으며, 시스템 개발 시에도 오류 가능성 등 이유로 이전코드를 재활용하기보다는 처음부터 개발하는 것을 선호하는 편
 - 그러나 CSPA 도입으로 IT-통계생산부서간 서비스모듈에 대한 명확한 개념 공유로 청내 구성원간 소통이 강화되는 측면도 있음
- ‘컴포넌트 단위로 사고함으로써 모두가 같은 아이디어를 공유하고 같은 언어를 쓸 수 있다는 것이 가장 큰 장점이다’ - 네덜란드
- 국제협력 강화로 한국 통계청의 위상 제고 효과도 기대
 - 단기적으로는 SOA 방식의 시스템 구축에 익숙하지 않음으로써 구축비용이 상승할 수도 있으나 장기적으로는 서비스모듈 재사용과 공유로 비용절감 효과도 기대
- (조직구성원 저항) SOA는 서비스를 설계하는 패러다임이 바뀌는 것으로 타 표준 도입 시보다 조직구성원 저항이 클 것이 우려됨
 - GSBPM·GSIM 적용은 실사과 관여가 별로 없었으나 CSPA의 성공적 도입을 위해서는 실사과의 적극적인 참여는 물론 새로운 업무역량개발과 필요시 조직개편까지 해야 할 수도 있음
- 실사과 담당자는 자신이 필요한 구체적인 서비스모듈 기능을 정확히 명세할 수 있는 능력과 함께 카탈로그에서 필요한 서비스모듈을 찾아 조합할 수 있는 능력이 요구됨
- 정보국 담당자는 시스템 단위 설계가 아닌 서비스모듈 단위 설계 방식에 익숙해져야 하며, 새로운 서비스모듈 설계보다 기존모듈 재사용

- 에 우선순위를 두어 개발해야 함
- .CSPA 환경 하에서는 팀도 시스템 단위가 아닌 기능단위로 구성. 즉, 추정모듈팀, 내검모듈팀 등으로 구성이 되며 하나의 기능을 여러 팀에서 개발하지 않는 것이 원칙. 따라서 조직개편이 수반됨.
- 조직구성원 저항 완화를 위해 사전에 워크숍 등으로 CSPA 도입에 대한 공감대를 구성하고 교육을 통해 필요한 역량을 강화
- (그 밖의 고려사항) 정부3.0 차원에서 클라우드, 빅데이터 플랫폼 구축하듯이 SOA 구축계획이 없는지 고려하여 진행

2. 그 외 시사점

- (기계학습 기술) 통계조사시스템 선진화를 위해 AI 기술을 더 적극 도입할 필요
 - 많은 국가가 에디팅, 추정, 레코드 연결, 무응답 추정 등 여러 분야에서 기계학습 기술을 적용해 자동화하는 실험을 진행 중
- .우리청도 자동코딩, 원격탐사시스템 등 일부 자동화를 시도하고 있지만 매우 기본적인 기술만 적용하고 있음
 - 통계정보시스템 선도를 위해 우리 청도 보다 적극적이고 실험적으로 기계학습 기술을 도입하는 것을 검토
- (국제협력) HLG 쪽에서는 각국의 보다 적극적인 참여를 요청. 현재의 사례수집과 발표보다 더 나아가 국제 프로젝트에 참여할 것인지에 대한 검토 필요
 - HLG 쪽에서 각국에 참여를 요청하는 분야는 ① CSPA 또는 다른 표준 관련 틀 개발, ② 국제 표준(현재는 주로 CSPA) 명세 상세화, ③ AWG, TCC group에 멤버로 가입하여 활동 등이 있음
 - 국제표준을 제정하는 과정에 직접 참여하는 것이 한국통계청 위상

강화에는 도움이 되겠지만 관련하여 담당자 업무조정이 필요

- (워크숍 참가) 통계정보국 뿐만 아니라 실사과에서도 워크숍에 함께 참가하는 것이 실질적인 토의와 국제협력에 도움이 될 것으로 보임
 - 이번 워크숍 참가자의 절반 정도가 IT전문가가 아닌 통계전문가로, 통계수요자의 관점을 제시함으로써 보다 전문성 있는 토의를 진행
 - 특히 이번 워크숍은 통계생산시스템의 관리적 측면이 아닌 현대화를 위한 방법론적 측면까지(방법론 아키텍처 등) 함께 다루므로 통계방법론 전문가가 함께 참여하는 것이 적극적인 토론이 가능

1. 워크숍 결론 및 각국 통계청에의 요청사항

- (워크숍 결론) 향후에도 CSPA 및 국제협력 관련 홍보를 계속할 것이며 관련 프로젝트에 각국의 적극적인 참여를 부탁
 - 다음 HLG 워크숍은 6월의 통계컨퍼런스(CES)로 기관장급
 - 매년 10월마다 다음 해 프로젝트를 결정하는 회의 진행
 - (향후 워크숍 운영방안) MSP는 특별한 변동사항이 없는 한 연례 워크숍으로 운영
 - 다만, 워크숍 개최 횟수를 늘리고 수시로 만나(매월) 국제협력에 추진력을 얻자는 의견도 있었음
 - 자주 참석하기 어려운 국가는 화상회의를 통해 국제교류를 강화 하자는 결론
 - 워크숍 개최지는 기본적으로 제네바이나 다른 국가의 요청이 있는 경우 해당국에서 개최
 - (각국 통계청에의 요청) 국제표준을 준수하는 통계현대화커뮤니티 (statistical modernization community) 참여를 선언하는 성명문 (manifest)에 기관장이 서명을 할 수 있도록 내용 공유 요청
 - 6월 통계컨퍼런스에서 다시 한 번 내용을 보고할 계획이며, 성명문 서명을 위해 해당 컨퍼런스에 기관장급(chief statistician)을 초청할 계획
 - ※ Conference of European statistician seminar "modernization of statistical production and service, and managing efficiency"(6.16)
- .실무자 급에서도 미리 내용을 전달하여 CSPA 취지 공감 형성을 부탁
- 성명서에 서명을 한다고 해서 각국의 정보화계획 수립에 강제성을 가지지 않으며, ‘물에 한쪽발을 담그는 정도’의 의지표명임을 강조

As active participants within the Statistical Modernisation Community, committed to the success of that community, the organizations listed below:

1. Commit to:

a. Adopting community behaviours by:

i. Being willing to make compromises to take advantage of, and improve, existing services (where it is sensible to do so)

ii. Actively contributing to collaboration

iii. Behaving as a trusted partner

iv. Sharing intellectual property and code with the community where possible.

b. Adopting an industry standard architecture in accordance with the CSPA, including the Generic Statistical Information Model (GSIM) and the Generic Statistical Business Process Model (GSBPM).

c. Sharing information about investment plans and application portfolios, looking for opportunities to collaborate.

d. Identifying existing strong capabilities within each member organization, and seeking to make them available as CSPA compliant services with appropriate documentation.

e. Producing (or commissioning from third parties) new and enhanced CSPA compliant statistical services in a manner that enables sharing within the community

f. Prioritizing mechanisms that support and encourage community participation when sourcing statistical capabilities

2. Affirm that the community will respect the individual sovereignty and requirements of member organizations and groups

그림 1 CSPA manifest

2. 한국 통계청에의 요청사항

○ Sandbox 프로젝트 참여요청

- 우리청의 빅데이터 경험을 살려 국제 빅데이터 플랫폼 구축 프로젝트인 Sandbox에 참여해 달라는 요청

○ 빅데이터 공통기반 관련 자료 요청

- 한국의 빅데이터 공통기반 관련 영문자료 요청

3. 후속조치

○ Sandbox 프로젝트 참여요청: 해당과에 요청내용 전달

○ 빅데이터 공통기반 자료: 행안부에서 영문자료를 받아 송부