

방송통신정책연구 14-진흥-029

과학기술의 SW활용 현황분석 및 협력방안 연구

(A study on current state and collaboration measures of
software utilization in science and technology research)

김석원/임춘성/이현승/공한경

2014. 10

연구기관 : 소프트웨어정책연구소



미래창조과학부
Ministry of Science, ICT and
Future Planning

이 보고서는 2014년도 미래창조과학부 방송통신발전기금 방송통신 정책연구사업의 연구결과로서 보고서 내용은 연구자의 견해이며, 미래창조과학부의 공식입장과 다를 수 있습니다.

제 출 문

미래창조과학부 장관 귀하

본 보고서를 『과학기술의 SW활용 현황분석 및 협력방안 연구』의 연구결과보고서로 제출합니다.

2014년 10월

연구기관 : 소프트웨어정책연구소

총괄책임자 : 김석원

참여연구원: 임춘성

이현승

궁한경

목 차

제1장 서론	1
제1절 과학연구와 소프트웨어	1
제2절 연구의 필요성 및 목적	5
제2장 해외의 정책 체계 및 사례	8
제1절 미국 과학기술연구의 SW활용 정책	8
1. 미국의 과학기술 정책 조정 체계	8
2. 사이버인프라 추진 체계	9
3. 선진 사이버인프라 프로그램 (Advanced Cyberinfrastructure)	13
4. 개방형 과학 그리드(Open Science Grid)	21
제2절 유럽 과학기술연구의 SW활용 정책	22
1. 영국의 과학기술 정책 조정 체계	22
2. 영국의 e-Science(e-infrastructure) 프로그램	24
3. 유럽연합 현황	28
제3절 과학기술연구의 소프트웨어 활용 사례	29
1. 입자 물리학 - Higgs 입자 실험을 위한 아틀라스와 구글 컴퓨트 엔진 사례	30
2. Folding@home 프로젝트	34
3. FoldIt	39
4. EteRNA	41
제3장 국내 연구 현황	44
제1절 국내 연구 사례	44
1. 자문회의를 통한 사례 조사	44
2. 전문가 인터뷰를 통한 사례 조사	52
3. 공개 포럼	57

제2절 소프트웨어 활용 사례	58
제3절 국내 현황 조사 결과의 분석	66
제4장 문제 분석	69
제1절 문제의 도출	69
1. 학제간 공동 연구에 대한 연구 사례	69
2. 국내 연구 상황의 문제	71
제2절 해결 방안 검토	73
1. 공동 연구	73
2. 인력 양성	74
3. 정보 교류	75
4. 공개 소프트웨어	75
5. 개선 전략	76
제3절 국가 연구개발의 사업화 현황 조사	78
제4절 현행 연구결과 공개 제도	84
제5절 과학기술 빅데이터 공동 활용 종합계획	85
제6절 고려 사항	87
1. 기본 원칙	87
2. 공개 대상	88
제5장 소프트웨어 산출물 공개 제도	89
제1절 제도 추진 근거	89
제2절 제도의 개요	92
제3절 과제 제안 단계	94
제4절 소프트웨어 공개 단계	95
1. 공개 소프트웨어의 구성	95
2. 기존 오픈소스 공유 서비스 이용	97
3. 국내 공개 사이트 운영	102
4. 공개SW 라이선스	104

5. 라이선스의 결정	111
제5절 사업화 단계	111
제6절 평가 단계	113
제7절 법제도 고려 사항	115
1. 개요	115
2. 제안 및 선정 과정(협약 포함)	116
3. 보고 및 결과 평가	117
4. 공개 및 성과 관리	117
5. 추적 평가	118
6. 장기적 제언 - 대상 규범의 법제화	118
제8절 신규 지원 조직의 기능	118
제6장 결론	120
참 고 문 헌	123
부록. 해외 소프트웨어 활용 사례	129
제1절. 미국의 분야별 SW활용 사례	129
1. 환경 과학 (Environmental Science)	129
2. 생명 정보학 - BIRN 협력 사례	135
3. 입자 물리학 - Higgs 입자 실험을 위한 아틀라스와 구글 컴퓨트 엔진	142
4. 응집 물질 물리학 - 응집 물질 물리학을 위한 과학 클라우드 컴퓨팅 플랫폼	146
5. 네트워크 과학 - CINET (A cyberinfrastructure for network science)	153
6. 통합 계산 재료 공학 - 미국 마그네슘 프론트엔드 3개국 파일럿 프로그램	157
7. 사회 정보학 - 웹 2.0 기반 사회 정보학 데이터 그리드(SIDGrid)	161
제2절. 유럽 각 분야별 SW활용 사례	164
1. 신경과학분야 - CARMEN	164
2. 천문학 - AstroGrid Virtual Observatory	166
3. 지구과학 - GENESI-DEC	169

4. 구조생물학 - WeNMR	172
5. 과학 작업 흐름 관리 - Taverna	177
6. 금석학 - VRE SDM	180
7. 인류학 및 사회과학 - CLARIN	183
8. 인문 예술 분야 - DARIAH	185
제3절 시민 과학의 SW활용 사례	188
1. Folding@home 프로젝트	195
2. Foldit	199
3. EteRNA	202

표 목 차

〈표 2-1〉 캠퍼스 사이버인프라 프로그램 영역	14
〈표 2-2〉 데이터 인프라 구성 요소 프로그램 영역	15
〈표 2-3〉 국제 연구 네트워크 연결 프로그램 영역	16
〈표 2-4〉 사이버 인프라 소프트웨어 개발 프로그램 영역	18
〈표 2-5〉 지속 가능한 혁신을 위한 소프트웨어 기반구조 세부 분야	19
〈표 2-6〉 연구 위원회 별 주요 시범 사업	25
〈표 3-1〉 자문회의 연구사례 요약	44
〈표 3-2〉 전문가 인터뷰 요약	52
〈표 4-1〉 학제간 연구에 대한 연구 결과	70
〈표 4-2〉 기술확산 및 사업화 성과 지표	83
〈표 5-1〉 오픈소스 저장소 기능 비교	101
〈표 5-2〉 오픈소스 저장소 인기도 비교	101
〈표 5-3〉 공개SW 라이선스 Top 20	108
〈표 5-4〉 공개 소프트웨어 라이선스 비교	110
〈표 1〉 데이터 설명	130
〈표 2〉 데스크탑 머신과 애저 인스턴스 용량	133
〈표 3〉 재투영 실행 통계 - MOD04L2	133
〈표 4〉 재투영 실행 통계 - MOD06L2	133
〈표 5〉 재투영 실행 통계 - MOD11L2.005	134
〈표 6〉 1500개의 작업을 처리하는데 든 처리 시간(단위:시간)	134
〈표 7〉 사이버 인프라 적용 전후 효과 요약	135
〈표 8〉 BIRN 참여 기관	136
〈표 9〉 사이버 인프라 적용 전후 효과 요약	141
〈표 10〉 사이버 인프라 적용 전후 효과 요약	146
〈표 11〉 사이버 인프라 적용 전후 효과 요약	153
〈표 12〉 사이버 인프라 적용 전후 효과 요약	157

<표 13> 사이버 인프라 적용 전후 효과 요약161

<표 14> 사이버 인프라 적용 전후 효과 요약164

<표 15> WeNMR에서 제공하는 NMR 분석 서비스173

<표 16> WeNMR에서 제공하는 SASX 분석 서비스174

<표 17> 진행 중인 Citizen Science의 예시189

<표 18> Folding@home 전후 효과 요약199

<표 19> Foldit 전후 효과 요약202

<표 20> EteRNA 전후 효과 요약204

그림 목 차

[그림 1-1] 과학연구 패러다임의 변화	2
[그림 1-2] 여러 단계의 생물 모형	2
[그림 2-1] 미국의 과학기술정책 조정체계	9
[그림 2-2] 사이버인프라 추진 체계	10
[그림 2-3] 소프트웨어 인프라 투자 영역	11
[그림 2-4] 사이버인프라(파란색)의 일반적 구조	12
[그림 2-5] 국립 과학 재단 추진 체계	13
[그림 2-6] 영국 과학기술 행정 체계	23
[그림 2-7] FP6에서 도입된 e-Infrastructure의 구조	29
[그림 2-8] LHC와 지하에 묻힌 ATLAS 탐지기	31
[그림 2-9] PanDA 워크로드 관리 시스템	32
[그림 2-10] XRoot와 PROOF 클러스터	33
[그림 2-11] 전송 속도	34
[그림 2-12] 단백질 폴딩 전 후 모습	35
[그림 2-13] Folding@home 성능	38
[그림 2-14] Foldit의 단백질 폴딩 화면	39
[그림 2-15] EteRNA 워크플로우	42
[그림 3-1] 차세대 시퀀싱 기술이 무어의 법칙을 추월	59
[그림 3-2] 차세대 시퀀싱에서 생산되는 데이터 크기	59
[그림 3-3] 생물 연구실의 소프트웨어 활용 사례	60
[그림 3-4] 시스템 생물학의 정보 표현 사례	61
[그림 3-5] 생물계와 폰노이만 모형의 유사점	62
[그림 3-6] 영상처리를 이용한 심장혈류 계산	63
[그림 3-7] 미래의 환자 맞춤형 진료	64
[그림 3-8] Worldwide LHC Computing grid	65
[그림 4-1] 문제 대응 방안	78

[그림 4-2] 국가 연구개발 예산	79
[그림 4-3] 부처별 국가연구개발사업 투자 추이, 2012-2013	80
[그림 4-4] 2012년 항목별 연구성과 현황	81
[그림 4-5] 연도별 사업화 건수	81
[그림 4-6] 국가연구개발사업의 기술료 징수액과 징수건수	82
[그림 4-7] 소프트웨어 자산뱅크 검색 화면	85
[그림 4-8] 과학기술 빅데이터 목표 개념도	86
[그림 5-1] 과제 수행 단계별 개선 방안	93
[그림 5-2] 소스포지 서비스	98
[그림 5-3] 깃허브 사이트	100
[그림 5-4] 저장소 인기의 변화	102
[그림 5-5] CCPForge 공유사이트 메인 화면	103
[그림 1] 재투영 작업을 위한 작업 스케줄링 방식	131
[그림 2] 파이프라인 단계에 따른 시스템 구성도	132
[그림 3] 병렬 처리 속도의 향상	134
[그림 4] 데이터 수집 및 처리 흐름도	138
[그림 5] BIRN 중개자 활동의 예시	139
[그림 6] 글로버스 온라인 파일 공유 시스템 예시	140
[그림 7] BIRN을 활용한 쥐 스마트 아틀라스 이미징 예시	141
[그림 8] LHC와 지하에 묻힌 ATLAS 탐지기	142
[그림 9] PanDA 워크로드 관리 시스템	144
[그림 10] XRoot와 PROOF 클러스터	145
[그림 11] 전송 속도	146
[그림 12] 모델 측정 및 구현	148
[그림 13] 아키텍처 및 GUI화면	149
[그림 14] FEFF9 계산의 고속화	150
[그림 15] WIEN2k 계산의 고속화	151
[그림 16] WIEN2k 계산의 고속화	152

[그림 17] WIEN2k 계산의 고속화	153
[그림 18] 고수준 CINET 구성요소 및 상호작용 뷰	155
[그림 19] Brightkite 체크인	156
[그림 20] ICME 포털의 아키텍처	158
[그림 21] TO BE 아키텍처	159
[그림 22] 저장소의 재료 요약 뷰	159
[그림 23] 포털 구성 요소의 통합	160
[그림 24] SIDGrid 과학 게이트웨이 기본 구조	162
[그림 25] OpenSocial 컨테이너 스택	163
[그림 26] CARMEN 구조	166
[그림 27] VODesktop의 사용예시	168
[그림 28] GENESI-DEC 프로젝트 구축 방법	170
[그림 29] GENESI-DEC에서 사용할 수 있는 인프라	171
[그림 30] WeNMR중 MDD MNR 서비스 화면,	176
[그림 31] Taverna의 작업 흐름 예시	178
[그림 32] Taverna 구조	179
[그림 33] Taverna Workbench 구현 모습	180
[그림 34] VRE SDM 프로토타입	182
[그림 35] VRE SDM을 통해 문자를 해독하는 과정	183
[그림 36] CLARIN Component 메타데이터 프레임워크	185
[그림 37] DARIAH 기구 구성	186
[그림 38] DARIAH를 통해 EHRI의 e-인프라 구축	188
[그림 39] 단백질 폴딩 전 후 모습	195
[그림 40] Folding@home 성능	198
[그림 41] Foldit의 단백질 폴딩 화면	200
[그림 42] Eterna 워크플로우	203

요 약 문

1. 과학기술의 SW활용 현황분석 및 협력방안 연구

2. 연구 목적 및 필요성

□ 배경

- 과학기술연구자와 소프트웨어 전문가간의 협업에서 좋은 결과가 나오지 않는 현실
 - 서로 다른 전문분야간의 설명과 이해의 어려움
 - 한 쪽이 해결할 수 없는 문제를 공동으로 해결하는 연구보다는 한쪽이 요구사항을 정해주고 소프트웨어 전문가는 그대로 개발만 해주는 형태
 - 소프트웨어 개발 방법론이 발전하여 타 전문가가 시작은 쉽게 할 수 있는 환경이 됨. 그러나 전문적 수준으로 개발하려면 여전히 전문가가 필요함
- 국가 과학기술 연구결과로 나오는 소프트웨어와 데이터가 널리 활용되지 못함. 현재 연구결과는 공개나 공유하기 어렵고, 사업화도 안 된 상태로 실험실에서 사장되는 경우가 많음. 공개를 하고 싶어도 제도적 문제로 공개를 못하는 연구자도 있음.
- 지속적인 개선, 변신, 적응이 소프트웨어 가치의 핵심이어서 활용하지 않고 보관만 하고 있는 것이 가치를 보전해주지 못 함

□ 필요성

- 과학기술 연구에서 소프트웨어 분야의 최신 기술을 적극적으로 활용하는 것이 필수적인 요소가 되어 가고 있음
- 국내에서는 협업보다는 해당 분야의 연구원이 소프트웨어를 배워서 자

체 개발하거나 규격을 정의하여 위탁하는 형태가 대부분임

- 빠르게 변화하는 소프트웨어 분야의 신기술을 적용하여 과학기술 연구를 촉진하는데 한계

□ 목표

- 과학기술 연구에 있어서 소프트웨어 활용과 협업 활성화를 위한 정책 개발
- 신기술 활용을 촉진하고 전문가의 협업을 활성화할 수 있는 방안 개발

3. 연구의 구성 및 범위

□ 사례 조사

- 과학기술 연구에서 소프트웨어 기술을 적극 활용하는 해외 사례 조사
- 정부지원 과학기술 연구과제에서 소프트웨어 활용 사례 조사
- 산업계에서 과학 분야에 소프트웨어가 활용되는 현황 조사

□ 과학기술 연구에서 소프트웨어 활용이 미흡한 원인 파악

- 소프트웨어 전문가 활용이 필요한 분야, 애로사항 등을 파악
- 종합적 관점에서 문제점 도출

□ 소프트웨어 활용을 통해 시너지를 창출할 수 있는 정책 및 활성화 방안 마련

- 과학기술 연구에 소프트웨어 전문성의 적용을 확대할 수 있는 전략, 방안, 제도, 정책 등을 종합적으로 제시

4. 연구 내용 및 결과

□ 해외 현황

○ 과학기술연구에서 소프트웨어는 필수 도구임

- 2020 Computing, Nature, 2006. 2020년의 과학자는 컴퓨터전문가 수준의 지식이 필요하게 될 것이고, 수학과 과학이 서로 협력하여 발전했듯이 과학과 컴퓨팅이 협력하고 영감을 주고받으며 발전하는 시대가 될 것임
- Jim Gray, The Fourth Paradigm, 2007. 경험적 과학, 이론적 과학에서 계산적 과학, 데이터 중심 연구로 패러다임이 이동함. 시뮬레이션 데이터와 실험 데이터의 폭증에 의해 생성되는 대용량 데이터를 소프트웨어로 분석하여 연구하게 됨.

○ 미국 연방과학기술 행정체계

- 대통령산하 과학기술정책국(OSTP)이 정책을 세우고 NSF가 연구기관의 과제를 관리
- NSF를 통해 첨단 사이버기반구조 프로그램을 추진하여 과학기술연구에서 소프트웨어 기술의 활용 지원. 사이버기반구조 프로그램은 과학기술연구에서 필요한 고성능컴퓨팅, 데이터 관리, 실험, 측정 등의 워크플로우, 데이터 시각화, 공동작업 서비스 등의 요소와 이것의 기반 기술인 네트워크/가상화/소프트웨어 스택 등의 연구, 계산과학 전수를 위한 과학자 지원 프로그램 등으로 구성

○ 영국 과학기술 행정체계

- 수상 산하 기업혁신기술부에서 각종 연구위원회를 관리
- 여러 위원회의 공동 참여로 정책을 결정하며 실행은 연구위원회에서 추진
- e-infrastructure 프로그램은 과학기술에서 정보기술의 활용을 촉진하는 기반 구조를 제공하는 프로그램. 공학 및 자연과학 연구위원회가 이 프로그램을 총괄하여 추진하며 프로그램에 관계된 7개 연구위원회를 대표

○ EU e-Infrastructure 프로그램

- 2002년부터 FP6, FP7 등의 기반기술 구축 프로그램을 추진하여 과학연구의 소프트웨어 활용을 지원

○ 해외의 연구개발 결과 공개 장려 문화

- 미국은 정보자유법(Freedom of Information Act)에 의해 연방정부에서 지원

한 과제에서 개발된 소프트웨어의 소스코드와 데이터 공개를 장려함. 또한 2013년 2월 백악관 과학기술정책실에서 전 연방부처에 정부투자 과학기술 연구개발 결과와 데이터를 일반에 공개하도록 지시하는 등 연방예산에 의한 연구결과의 공개를 중요시함.

- 영국은 Ccpforge 등의 소스코드 및 데이터 공유 및 공동개발 커뮤니티 사이트를 정부에서 개발한 후 운영함. 영국과 유럽도 연구 개발자들이 부담 없이 공개하고 협력할 수 있는 환경을 제공하고 있음

○ 해외 사례의 시사점

- 소프트웨어는 모든 과학기술 연구개발의 기반기술이 되었음
- 해외 정부지원 정책은 전략적으로 과학기술 연구에서 컴퓨팅을 활용할 수 있는 기반기술 환경의 구축 지원과 분야별 개별 지원
- 연구에서 파생된 소프트웨어 등의 산출물의 공개 문화가 연구 자세의 기본으로 확립되어 있으며 연구 데이터를 공유하려는 정책이 추진되고 있는 중

□ 국내 현황

○ 과학기술연구개발 정책결정 기구

- 국가과학기술심의회에서 정책 심의 및 조정하며, 운영위원회에서 분야별 실무검토를 담당하나, 과학기술연구를 위해 소프트웨어 기술 활용을 촉진하기 위한 위원회는 없음

○ 국내연구현황은 20개 기관의 과학과 소프트웨어 전문가 42명을 대상으로 자문위원회 및 전문가인터뷰를 진행하여 조사

○ 국내 전문가 조사 결과

- 연구에서 필요한 소프트웨어는 주로 해외의 공개소프트웨어를 활용하거나 자체 인력이 있는 경우 부분적으로 자체 개발하여 활용
- 소프트웨어의 중요성에 대해서는 모두 인지하고 있음
- 협업의 필요성은 이해하고 있으나 협업이 자신의 전문 분야 연구에 비해 노력이 많이 들고 결과도 불확실하여 실적에 쫓기는 일부 전문가는 부담이 큰 일로 인식

○ 인력 필요성

- 학제간 경험을 가진 인력의 필요성은 모두 공감하나 양성방법에 있어서 정책적 조치의 효과는 기대가 낮음. 필요한 인력은 전문성과 동시에 타 전공 분야도 이해하고 있어야 하므로 융합학과 등의 교육을 통해 양성되기 어려움.

○ 연구결과의 공개에는 대부분 긍정적

- 대부분 찬성이며 나머지도 조건부 찬성 의견임 (사업화 대상은 제외, 핵심 기술 제외, 연구자의 자율권 보장 등)

○ 공개 방식에 대한 의견

- 강제적보다는 자발적, 기계적보다는 선택적 방안을 선호
- 공개를 위한 오버헤드에 대한 해소 방안, 핵심 노하우의 보호를 위한 방안이 함께 제시되길 바람

○ 국내에서 한국과학기술정보연구원을 중심으로 협업연구를 통해 과학연구 데이터 처리시간을 획기적으로 개선한 사례가 있음. 이 결과를 바탕으로 미래부에서 과학연구 빅데이터 공동 활용 종합 계획을 수립하여 추진 중

□ 학제간 연구에 대한 기존 연구

- 가장 큰 저해 인자는 기존 연구 평가체계와의 부조화임. 학제간 연구는 초기 습득기간이 길기 때문에 논문생산성이 낮음. 따라서 논문 숫자를 채워야 승진하는 학교의 평가 및 승진체계에서는 연구원이 적극적으로 나서지 않음
- 학제간 연구는 이론적인 연구가 아니라 현실적인 문제를 해결해야 하는 경향이 있음.

□ 국내의 협업장애 요인

- 전문성을 갖춘 융합인력의 부족. 융합학과를 졸업한 어중간한 전문성을 가진 인력이 아니라 한 분야의 전문성을 확보하고 다른 분야에 경험을 쌓은 인력의 부족

- 긴 숙련 기간. 전문가가 다른 분야와 정상적인 협업을 하기 위해서는 보통 1년에서 3년 정도의 기간이 필요함. 이런 경험을 쌓은 인력이 부족하고 인력이 배출될 환경도 갖춰져 있지 않음
- 전문가 참여 동기 부족. 연구 성과에 대해 협업에 참여한 모두가 만족하기 어려운 평가 체계. 투입한 시간과 노력에 비해 위험도가 높음.
- 정보의 국지성. 다른 분야의 인력, 연구 내용, 문제에 대한 공유 기회 부족. 과제와 별도로 사전 연구할 수 있는 기반이 미흡

□ 개선 방향

- 공동 연구를 촉진하는 방안은 현재 정부에서 추진하고 있는 융합 과제 정책의 효과를 기대하고 새로운 과제를 제안하지 않는 방향으로 함
- 인력 부족은 대학 교육 정책에 대한 대안을 제시할 필요는 없다고 판단함. 필요 인력의 규모가 크지 않고 석사급 이상의 전문성이 있는 인력이 필요함. 전문가가 관심을 기울일 수 있는 환경의 지원이 필요.
- 연구내용의 공유 장치, 서로 다른 분야의 전문가 간에 정보를 교류할 수 있는 기회, 연구내용을 서로 참고할 수 있는 개방된 문화의 보급이 필요.
- 따라서 공동 연구는 융합과제 정책의 효과를 기대하여 추이를 지켜보는 것이 좋고 인력 부족이나 공유 문화를 위해서는 개방된 연구 문화를 장려하는 제도와 환경을 만드는 쪽으로 추진

□ 국가 연구개발사업의 사업화 성과 조사

- 연구 문화를 개방적으로 만들기 위해서는 정부투자 연구개발사업의 결과가 사업화로 이어지는 사례에 대한 성과 조사 필요. 사업화로 이어지는 경우가 많고 성과가 중요한 경우에 개방적 연구 문화 장려 제도가 일부 연구에만 국한되거나 전혀 효과가 없을 수 있음.
- 2012년 사업화 건수는 과거에 비해 급증하여 15,000건에 달함. 반면 기술료 징수건수와 징수액은 과거와 비슷하거나 약간 증가한 수준임. 이는 사업화 건수로 보고된 내용이 아직 신뢰할 수 없는 일시적 현상일

수 있다는 의미임.

- 2012년 총 연구비는 16조원이나 기술료 징수율은 1.6%에 불과.
- 국내 연구 성과는 일반적으로 기술적 성공률이 평균 90%를 상회하나 사업화 성공률은 약 20% 수준이며 대학과 연구소의 성공률은 4.4% 정도임. 이는 선진국의 사업화 성공률인 영국 70.7%, 미국 69.3%, 일본 54.1%에 비해 낮은 수준임.
- 이는 연구개발 사업의 성과지표가 논문, 특허가 중시되고, 사업화는 성과 평가가 어렵기 때문임. 사업화의 성공 여부로 연구자의 성실성을 평가하기는 현실적으로 어려움.
- 미래창조과학부에서는 꾸준히 성과평가 개선 대책을 추진하고 있으므로 향후 이 대책에 대한 성과를 지켜볼 필요는 있음

□ 거시적 접근 필요

- 연구자간의 협업 촉진을 위한 단기적이고 직접적 대책은 정부의 융합 과제, 연구 성과평가 개선 대책 등 충분히 진행되고 있다고 판단됨.
- 연구 결과의 공개를 장려하여 연구 결과가 다른 연구자에게 도움을 주고 정보가 쉽게 공유될 수 있는 개방적 연구 문화를 만드는 것이 과학 기술연구자와 소프트웨어 전문가의 협업을 활성화하는 기본적이고 가장 시급한 대책으로 판단됨. 연구 결과가 공개되더라도 연구자의 권리는 논문, 특허, 저작권 등으로 보호가 가능함.
- 소프트웨어와 데이터의 공개를 통한 연구 결과의 재활용을 촉진하기 위해 공개에 적극적인 연구자를 위한 기반 환경과 제도를 제공하면 국가적인 반복 개발 비용을 절감하고, 피어 리뷰에 따른 소프트웨어와 데이터의 품질 검증 및 개선이 되고, 연구 결과의 빠른 확산에 의한 국가적 연구 역량을 제고할 수 있음.
- 데이터의 공개 및 공유 제도는 과학기술 빅데이터 공동 활용 종합계획에 포함되어 있으므로 이 방안을 따르기로 함.
- 결국 과제에서 규정할 내용은 연구에서 생산된 소프트웨어의 공개에 대한 방안이 됨. 여기서 소프트웨어란 연구과제 분류상 소프트웨어 분

야에서 수행하고 있는 과제의 공식 산출물 소프트웨어뿐만 아니라 전 연구 분야에서 연구를 수행하며 생산한 부수적 산출물 소프트웨어 프로그램을 모두 포함함. 또한 단순히 프로그램 코드만이 아니라 일반적으로 공개 소프트웨어에서 따르는 관련 문서, 샘플데이터, 샘플프로그램 등 제3자가 제대로 활용할 수 있도록 제공해야하는 일체의 구성물을 모두 의미함.

□ 단계별 개선 방안

○ 연구제안단계

- 과제 제안 시 제안서에 소프트웨어 산출물 명기

(제안서) 공식, 비공식 소프트웨어 산출물 명기

(제안요청서) 연구원이 공개를 고려한 연구 계획 수립을 할 수 있도록 공지

- 제안서 심사 단계에서 공개에 적극적인 연구자의 기여를 심사에 반영함. 여기에는 이전에 공개한 소프트웨어의 평가, 활용도, 완성도가 포함됨.

- 협약서에 공개 조건 반영

○ 과제 수행 및 종료 단계에서의 소프트웨어 공개 방식

- 일반 공개사이트 이용. Github, SourceForge 등의 유명 사이트에 공개하고 공유 링크를 제출

- 이와 별도로 국가운영 공개사이트를 정부 지원에 의해 개발하고 공유 및 커뮤니티 플랫폼으로 활용. 자체 개발의 이유는 사용자 통제와 일정기간 공개를 유예하는 경우를 위해 관리가 가능한 공개사이트가 필요함.

- 국가운영 사이트의 개발 및 운영을 위한 조직이 필요함. 이 조직은 사업화를 위한 공개 유예 소프트웨어 관리, 지속적인 개선, 유지보수, 이용 활동의 감시 및 분석 등을 수행함. 소프트웨어 공개에 익숙하지 않은 사용자를 위해 기존 공개 소프트웨어 라이선스에 대한 가이드라인을 제공함. 국가운영 사이트와 일반 공유사이트에 공개된 연구 산출물 소프트웨어에 대해 활용도, 개선 활동, 커뮤니티 참여도 등의 데이터를 추적. 데이터 축적은 데이터 수집 소프트웨어를 개발하여 최대한 자동화하여 운영.

○ 사업화를 추진하여 공개를 이루는 경우

사업화를 추진하는 과제의 경우에는 종료 후에도 공개를 유예할 수 있음.

- 기업을 통한 사업화 추진의 경우 과제 종료 시 소프트웨어 산출물을 제출하나 공개는 유예하고, 기술이전 혹은 사업화 계획의 제출을 의무화함. 기업의 사업화 계획 의사를 1년(예시)이내 확정하여 문서로 제출함. 정해진 시한(예. 3년)이 지나고 성과가 없으면 제출된 소프트웨어를 공개함.
- 개발 연구원 창업 시 경제적, 기술적 지원을 통해 사업화를 지원함. 현행 창업 지원 프로그램 등을 활용함. 유예 기간(예. 3년) 동안 기업 가치를 높여 신규 투자를 유치하거나 일정 규모 이상의 매출 실적을 올리는 것이 목표임.

○ 평가 단계

- 모든 과제는 과제를 수행하며 공개한 경우를 제외하고 소프트웨어 산출물을 과제 종료 시 과제 공개사이트에 게시 및 1차 평가. 연구 진행 중 공개하여 활용하는 것을 권장하며, 사업화 과제 산출물은 공개를 유예함.
- 공개된 내용에 대한 평가는 공개 자료 활용도, 개선을 통한 완성도, 커뮤니티 참여도 기준으로 수행

□ 법제도 현황 및 개선 사항

○ 2014년 1월 ICT R&D 관리규정 개정으로 근거 마련

- 정보통신 방송 연구개발 관리규정 제36조 11항 “수행기관의 장은 소프트웨어 관련 연구개발 사업 수행 결과 중 기술실시계약이 체결되지 않는 등 활용이 미흡한 기술의 확산을 위해 공개소프트웨어로 배포하는 것이 바람직하다고 판단되는 경우에는 사업의 결과물을 공개소프트웨어로 배포할 수 있다.”
- 제40조(공개소프트웨어의 특례) 공개소프트웨어 방식으로 추진하는 과제에 대한 추진방식, 기술료 감면 등 규정

○ 연구결과를 공개소프트웨어화 하는 근거가 되긴 하지만 개선이 필요한 부분이 있음

- 제40조 1항은 미래부장관이 공개소프트웨어 개발방식으로 추진하도록 할 수 있으며 이 경우 전담기관이 타당성조사를 해야 한다고 규정

- 대안으로는 공개소프트웨어로 추진해서는 안 되는 과제를 지정하는 방식이 바람직하며 공개를 안 하는 경우에 타당성조사를 하도록 규정하는 편이 공개를 더욱 장려하는 방향임

- 교육부의 “교육부 소관 이공분야 연구개발사업 처리 규정”도 위와 같이 변경 필요
- 전담기관인 연구재단 등은 미래창조과학부 및 교육부의 규정을 인용하여 활용 중이므로 별도로 개정이 필요한 사항이 없음

□ 신규로 지원되어야하는 업무 기능

- 공개되는 데이터, 소프트웨어 등의 평가, 관리, 추적
 - 활용 상황 감시 및 평가, 활용 확산을 위한 인프라 구축 및 운영, 일반 사용자가 이용하기 쉽도록 인터페이스, 검색 등 꾸준한 개선
- 연구개발 및 지원 서비스
 - 과제의 소프트웨어 개발 지원 - 공개를 위한 표준, 컨설팅, 개발
 - 공개된 산출물의 활용 확산 - 교육 프로그램, 산출물의 서비스 화 등
 - 관리, 운영뿐 아니라 전문 융합인력 양성을 위해 자체 연구 개발 활동도 필요함. 연구 주제는 과학기술연구의 기반구조 연구 개발, Computational Science 연구, 지식의 전파와 공유 서비스, 과학연구와의 협업을 통한 인력 개발 등.
- 정기 학회, 워크샵, 커뮤니티 운영 등 정보 교환 창구 확대 및 강화
- 이 업무를 위해 신규 조직을 구성하는 것 보다는 기존 관련 조직을 활용하여 이 역할을 담당하도록 하는 것이 바람직함
 - 한국과학기술정보연구원(KISTI)는 수퍼컴을 운영하고 있으며 빅데이터 과제 등 유사 업무에 경험이 있음
 - 정보통신산업진흥원(NIPA) 산하 소프트웨어공학센터는 현재 SW자산뱅크 등을 운영하고 있으므로 업무에 관련성이 있음

5. 정책적 활용 내용

- 국가 과학기술연구과제의 제안, 심사, 수행, 종료, 사업화 단계에 걸쳐 연구 산출물 중 소프트웨어 산출물(공식, 비공식 모두 포함)을 공개소프트웨어로 하는 것을 장려하는 개방적 연구 문화 조성 정책으로 활용
- 점진적으로 공개소프트웨어로 등록된 산출물의 활용도 평가 방식을 개선하고 많이 활용되는 사례에 대한 보상을 늘려서 연구자의 자발적 참여가 일반화되도록 지속적인 지원 정책 유지

6. 기대효과

- 연구개발 산출물의 소프트웨어와 데이터 공개를 통해 연구결과의 재활용을 촉진하고 소프트웨어 전문가와의 공동 작업을 유도함
- 재활용 촉진에 의해 국가적으로는 반복 개발 비용이 절약되며 연구자에게는 소프트웨어 전문가와 정보 교류 기회를 확대하게 됨
- 소프트웨어 전문가와의 공동 작업은 공식적인 공동 연구 개발일 수도 있으나 공개 소프트웨어를 매개로 한 온라인상의 공동 작업이 더 많은 비율일 것으로 예상함. 소프트웨어 전문가가 공개된 코드를 개선하거나 의견을 제시하는 등 공개된 소프트웨어의 품질이 개선되고 연구자는 소프트웨어에 대한 이해가 심화될 수 있음.
- 이와 같은 개방적 연구 문화 및 정보 교환 기회 확대를 통해 과학기술 연구자와 소프트웨어 전문가 간에 문제에 대한 이해가 깊어지며 함께 연구하는 기반이 갖춰질 수 있음

SUMMARY

1. A study on current state and collaboration measures of software utilization in science and technology research

2. Objective and Importance of Research

Modern scientific research requires software as an indispensable tool. Domestic research activities do not facilitate software engineers in its full swing. Some researchers just learn basics of computer programming and do it themselves. Other researchers do try collaborative research with computer scientists, but successful collaboration results are hard to find.

Science and technology are getting more and more important for national competitiveness. Therefore we have to find the measures to promote utilization of software in science and technology research to keep competitive down the road.

3. Contents and Scope of the Research

Foreign cases and domestic status are investigated. Foreign cases include success cases using software to enhance research performance and the government programs developing software infrastructure to support scientific research. For domestic research, advisory committee of scientists and engineers was formed to discuss the problem in the domains. We also conducted interviews with researchers of government funded projects. From the findings and a series of advisory conferences we drew up plans to promote collaboration by enhancing open research culture.

4. Research Results

Software technology acts as an infrastructure in modern scientific research. The governments of the developed countries have been supporting many projects to provide cyber-infrastructure that helps researchers utilize connectivity, high performance computation, large data management, visualization and collaboration. It should be noted the culture to share software research tools and scientific data with other researchers lies under the projects.

Our circumstances are slightly different. Most experts contributed to this project complained Korean researchers tend to keep software tools and research data that they made inside small group. Lack of good interdisciplinary researchers, the problem of academic reward system, and the need for more opportunities to share information are also pointed out. But the key to promote collaborative research between scientists and software experts begins with the change of mindset toward the value of software and data.

We located three problems that should be addressed. Promotion of collaborative research projects is the easiest and direct way to stimulate researchers. As the government has already been supporting many convergence projects, we will see how influential they may be in the coming years. The next is the interdisciplinary human resources. The convergence department program is not the right fit. What is needed is the experts with at least a master's degree in a discipline diving into new area. The last one is to cultivate the open culture that scientists and experts share information, knowledge, software and data. The second and the last problem can be alleviated by the spread of openness in the research culture.

The success rate of commercialization in government-funded research projects is as low as 4.4% for the projects by the universities and the research organizations. It goes up to 20% for those by the corporations, which is still quite low. Therefore we can

take new measures to improve openness and make an exception for the successful commercialized cases without losing the generality.

The software outcomes of the government-funded R&D projects are encouraged to register as an open source. The principal investigator declares what software will open to public in the proposal. He should make it public before the project closing. The quality and the popularity of the open sources are traced and consulted in every evaluation process. Those for commercialization have rights to ask for delay before registration. (ex. 3 years) If it fails, the software is also made public.

We expect the measures can lead openness to prevail gradually. It will, in turn, facilitate interdisciplinary collaborative research between software experts and others by expanding learning opportunities about the problems in scientific research to a great extent.

5. Policy Suggestions for Practical Use

Small changes should be made in the forms of RFPs, proposals and final reports to specify software deliverables. The process is also changed accordingly.

Current R&D management regulations still require Minister' s approval to make software deliverables open to the public. Requiring the approval to get an exemption from making public may expedite the change of mindset.

6. Expectations

Open software policy will promote the reuse of R&D deliverables and enhance the quality of software. Common code can be reused, and expand the information sharing. Through the process of software quality enhancement, scientists can understand the software more and software experts can learn scientific domain problem.

CONTENTS

Chapter 1. Introduction

Chapter 2. Oversea cases of software utilization in science and technology research

Chapter 3. Domestic research status

Chapter 4. Problem analysis

Chapter 5. Open source policy for software deliverables

Chapter 6. Conclusions

제1장 서론

제1절 과학연구와 소프트웨어

컴퓨터과학의 노벨상이라 할 수 있는 튜링상을 수상한 마이크로소프트 연구소의 테크니컬 펠로우 짐 그레이 박사는 2007년 1월 11일 캘리포니아 마운틴뷰에서 열린 국가연구위원회에서 현대 과학연구 패러다임의 변화에 대해 설명하며 컴퓨터과학이 과학연구의 필수도구가 되었다고 주장하였다.¹⁾ 과학연구는 고대 그리스시대부터 경험적 지식으로 자연현상을 설명하려는 시도로 시작되었다. 뉴튼이 경험적 지식 대신 수학적 법칙으로 설명하면서 지난 수백 년 동안 자연현상을 수식으로 표현하고 이 수식을 해석하는 이론적 연구로 발전했고 수식이 점점 복잡해지면서 해석적으로 풀기 어려운 자연현상을 시뮬레이션을 통하여 검증하는 계산적 연구방식이 이론적 연구와 나란히 발전해 왔다. 오늘날 과학기술 연구는 계측장비의 발달과 시뮬레이션 데이터의 급증에 따라 다뤄야 할 데이터가 급속도로 증가하여 과학 연구의 중요한 활동이 막대한 데이터를 의미 있는 정보로 해석하는 데이터 중심 연구 방법으로 바뀌었다.

기가바이트를 넘어서 테라바이트 급의 데이터를 분석하는 과학기술 연구에서는 이제 소프트웨어 분야의 최신 기술을 적극적으로 활용하는 것이 생존을 위한 필수적 요소가 되어 가고 있다. 예를 들어 화학 연구에서는 분자 수준의 구조, 반응, 기타 특성을 실험 대신 컴퓨터의 계산을 이용해서 예측한다. 이 방법은 실험보다 쉽고 안전하면서도 경제적으로 다양한 실험을 할 수 있다. 생물학에서는 소프트웨어를 이용하여 생물계의 모형을 원자적, 신진대사적, 세포적, 병리학적 수준 등 다양한 복잡도로 만들 수 있다. 이 모형은 단백질의 3차원 접힘 문제, RNA 구조 유전체학 등 계산량이 많은 문제에 적용된다.

1) Hey, Tony et al. (2009), pp. xix-xxxiii.

[그림 1-2] 과학연구 패러다임의 변화

• **경험적 과학**

- 자연현상을 직관으로 설명

• **이론적 과학**

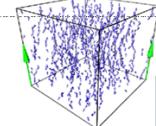
- 케플러, 뉴턴 이후 모델과 일반화로 설명

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3} \rho - \frac{kc^2}{a^2}$$



• **계산적 과학 (분기)**

- 이론적 모델이 복잡해지며 해석적으로 풀기가 어려워져서 시뮬레이션으로 검증



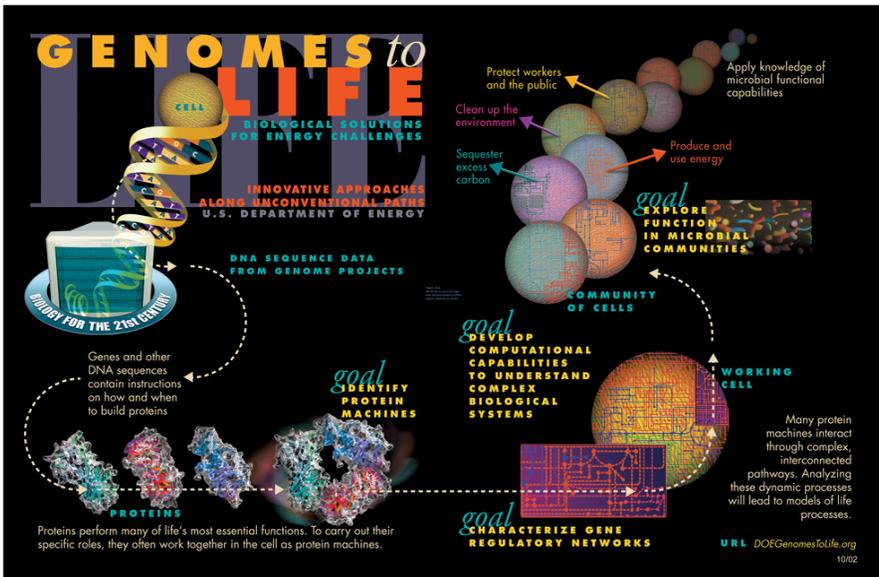
• **데이터 중심 연구**

- 시뮬레이션 데이터와 실험 데이터의 폭증에 의해 소프트웨어로 처리된 데이터를 보고 연구



자료: Hey, Tony et al. (2009)에서 발췌한 것을 번역

[그림 1-3] 여러 단계의 생물 모형



자료: <http://genomics.energy.gov/> 의 Genomics Image Gallery, 2014년 10월 검색.

소프트웨어는 각각의 연구에도 필수적 요소지만 공동 연구 수행을 위한 기반구조로서도 활용되고 있다. 인터넷, 분산 컴퓨팅과 같은 기술을 이용하여 연구개발의 공간적 제약이나 성능의 한계를 극복할 수 있는 전 세계적인 기반 구조로 소프트웨어를 활용하는 사례가 늘어나고 있다. 미국 국립가상천문대(US National Virtual Observatory)는 우주 관측 데이터와 이 데이터에 대한 분석 프로그램을 온라인으로 이용할 수 있도록 하여 누구나 사용할 수 있고 연구할 수 있는 온라인 가상 실험 환경을 공유하고 있다.²⁾ SETI@Home, Folding@Home 프로그램은 우주 관측이나 단백질 접힘 문제 등 막대한 컴퓨터 자원을 활용해야 하는 연구에 대해서 자발적인 참가자가 개인 컴퓨터의 유휴 시간 동안 데이터를 분석하는 글로벌 공개 분산 컴퓨팅 환경이다.³⁾⁴⁾ 또한 Eterna와 FoldIt은 복잡한 데이터 처리 문제를 인터넷을 통한 공개경쟁이나 게임 형식으로 구성하여 참가자의 재미와 의욕을 불러일으키면서 문제 해결을 시도하는 사례이다.⁵⁾⁶⁾

이런 변화의 중심에는 너무나 많아진 데이터가 있다. 데이터폭주(data deluge)라고도 부르는 이 상황은 컴퓨터를 쓰지 않고는 시뮬레이션도, 실험도, 검증도 할 수 없게 만들었다. 유럽입자물리연구소(CERN, European Organization for Nuclear Research)에서 만든 대형 강입자 충돌기(LHC: Large Hadron Collider)에서는 1년에 수십 페타바이트의 데이터를 쏟아내고 있으며, 한 사람의 유전자를 분석하기 위해 전체 게놈 시퀀싱(WGS: Whole Genome Sequencing)을 하면 수 테라바이트의 데이터가 추출된다.⁷⁾ 하드디스크 가격이 내려가서 이런 데이터를 저장하는 부담은 적다고 해도 가설을 세우기 위해 분석하고 실험을 통해 검증하려면 컴퓨터를 이용하지 않고는 불가능한 일이다. 과학자는 이 데이터를 분류하고, 정렬하고, 요약하여 사람의 인지능력으로 다룰 수 있는 수준으로 크기를 줄이는 고단한 과정을 거친 후에야 데이터를 분석하는 본질적 연구를 할 수 있다. 다차원을 지원하는 데이터큐브나 통계분석을 위한 R언어, 분야별로 공유되고 있는 공개

2) <http://www.virtualobservatory.org>
 3) <http://en.wikipedia.org/wiki/SETI@home>
 4) <http://folding.stanford.edu/>
 5) <http://eterna.cmu.edu/web/>
 6) <http://fold.it/portal/>
 7) <http://home.web.cern.ch/about/computing>

소프트웨어 등 여러 소프트웨어 도구의 도움을 받더라도 연구 시간의 절반 이상은 컴퓨터를 다루고 있을 것이다. 어떤 연구팀은 연구 시간의 70% 이상을 컴퓨터와 함께 한다고 한다.

데이터가 넘쳐나는 것은 계측기술의 발전과 더불어 더 작아지고 강력해진 프로세서와 센서기술의 발달에 기인한다. 생물학계에서는 차세대 시퀀싱(NGS: Next Generation Sequencing) 방식을 개발하여 DNA 순서 추출 시간을 대폭 줄였고 이에 따라 단위시간 당 추출되는 데이터양은 늘어나게 되었다.⁸⁾ 또한 신용카드 크기의 컴퓨터에 센서와 무선통신 기능을 포함하는 모트(mote)라는 무선센서컴퓨터로는 자연계의 현상을 사람이 직접 샘플을 채취하여 연구하는 대신 많은 모트를 퍼뜨려놓고 여기서 계측되는 데이터만 수집하여 연구하는 방식의 가능성을 열었다. 존스홉킨스 대학의 생태학자는 모트를 토양에 뿌려놓고 토양생태계를 연구하는 시도를 하고 있고, 영국에서는 빙하생태계를 연구하기 위해 빙하에 센서를 심어 놓기도 한다. 궁극적으로는 이런 무선센서컴퓨터가 지구 곳곳에 무수히 설치되고 자율적으로 동작하면서 측정된 데이터를 보내게 될 것으로 과학계는 예상하고 있다. 이렇게 되면 지질학, 생태학, 해양학 등 자연환경을 다루는 과학 분야의 연구 방식이 바뀐다. 연구를 위해 현장에 나가서 샘플을 채취하여 분석하는 대신 지속적으로 수집되는 실시간 데이터를 인터넷으로 접근하여 분석하는 것이 주된 활동이 될 것이다. 지구 자체가 온라인 데이터베이스가 되는 셈이다.⁹⁾

과학 연구의 패러다임이 바뀌고 있지만 공짜점심은 아니다. 연구에 소프트웨어의 비중이 늘어나는 것 자체가 과학계와 컴퓨터과학계의 공동 노력이 필요한 학제간 연구(interdisciplinary research)이다. 문제는 학제간 연구가 학제내 연구에 비해서 어려운 점이 훨씬 많다는데 있다. 우선 학제간 교류를 위한 인력의 공급이 어렵다. 한 분야의 전문가는 전통적 교육과정을 거치고 직장에서 경험을 쌓아서 전문가가 되는 길을 따르는데 비해 두 가지 이상의 분야에 대해 충분한 이해를 갖춘 인력은 예외적 경로를 따른 경우에만 배출된다. 이런 인력의 부족은 연구의 제약을 가져온다.

연구 효율성도 중요한 장벽이다. 학제간 연구는 이론보다 현실적인 문제에 초점을 맞춘 문제해결형 과제가 되는 경향이 있어서 학제내 연구에 비해 논문이나 특허를 내기

8) http://en.wikipedia.org/wiki/DNA_sequencing

9) Butler, 2006, pp.402-405

어렵고 논문과 특허의 수로 평가하는 전통적인 연구 체계에서는 연구자가 쉽게 뛰어들기 어려운 분야이다. 정말로 필요한 결과는 과학자의 경험과 소프트웨어 전문가의 경험이 조화를 이루어야 한다는 전제에는 수궁하면서도 연구자 개인 입장에서 적극적으로 뛰어들기에는 쉽지 않은 이유이다. 이런 환경을 개선하기 위해 선진국에서는 현대 연구 환경을 최대한 지원할 수 있도록 국가적 연구개발 체계를 구성하여 과학기술 연구를 지원하고 있다.

미국은 정부 주도로 사이버인프라를 설정하여 전략 과학 분야 혁신 인프라 이니셔티브를 잇달아 발표하고, 국립과학재단(NSF, National Science Foundation) 사이버인프라 위원회에서 “고성능 컴퓨팅, 데이터 분석 및 가시화, 분산 커뮤니티의 가상 조직, 교육 및 인력 개발”의 네 부분으로 나누어 추진하고 있다.¹⁰⁾ 영국도 2001년부터 e-Science 프로그램을 연구위원회(Research Council)와 무역산업부(Department of Trade and Industry)가 운영하였으며, 제1단계(2001~2004년)와 제2단계(2003~2006년)에 걸쳐 2억1천3백만 파운드를 지원하여 분야별 시범사업을 진행했다.¹¹⁾ 유럽연합은 R&D 프로그램인 FP6(6th Framework Programme for Research and Technological Development)와 FP7(7th Framework Programme for Research and Technological Development)의 일환으로 e-Infrastructure 연구를 지원하였다. 현재는 FP7을 잇는 Horizon 2020 프로그램을 통해 지속적으로 e-infrastructure 분야를 지원할 예정이다.¹²⁾¹³⁾

제2절 연구의 필요성 및 목적

과학기술 연구에서 소프트웨어의 비중이 커지면서 소프트웨어 전문가가 과학기술 연구자와 협력하여 연구의 효율성을 높일 수 있도록 정책적 지원을 하고 있는 선진국에 비해, 국내에서는 주로 과학기술 연구기관의 연구원이 소프트웨어 기술을 직접 습득하여 활용하거나 전문성이 부족한 개발자에게 위탁하여 개발하는 것이 일반적이다. 이런 연구문화에서는 빠른 속도로 발전하고 있는 소프트웨어 분야의 신기술을 적용하여 과학

10) http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504947&org=ACI&from=home

11) Atkins, 2010, p. 5

12) http://cordis.europa.eu/fp7/home_en.html

13) <http://ec.europa.eu/programmes/horizon2020/>

기술 연구를 촉진하는데 한계가 있다. 또한 과학기술 연구자와 소프트웨어 전문가 사이에도 소프트웨어에 대한 관점이 달라서 전문가가 참여하는 협업의 경우에도 효과적인 수행이 어렵다.

따라서 국가 과학기술 연구개발사업의 추진 체계를 현대의 상황에 맞도록 개선하고 소프트웨어를 충분히 활용할 수 있는 연구 문화를 지원하는 정책연구를 통하여 양쪽 분야 전문가 사이에 소통을 활성화하는 방안이 제시되고 교류의 물꼬를 여는 정책이 발굴되면 두 분야 모두에게 윈-윈이 되는 결과를 기대할 수 있을 것이다.

본 연구에서는 과학기술 연구에서 소프트웨어 기술을 적극 활용하기 위해 정부에서 어떤 지원을 하고 있고 어떤 사례가 있는지에 대한 해외 사례 조사 결과를 제시한다. 해외 사례는 미국, 영국, 유럽연합 등 과학기술연구에서 소프트웨어 기술을 적극적으로 활용하고 있는 국가의 정부 정책 결정 체계와 정부 주도 과제에 대해 조사한다. 그리고 이 과제를 통해서 어떻게 과학기술연구가 추진되고 있는지 사례별로 알아본다.

과학기술연구자와 소프트웨어 전문가의 협업에 관한 문제는 학제간 공동 연구(interdisciplinary collaborative research)라는 분야의 한 부분이라고 볼 수 있다. 일반적인 학제간 연구에 대한 기존 연구 결과를 조사하면 여기서 다루는 문제에 대한 근본적 성격을 이해할 수 있다. 이를 위해 문헌조사를 통해 기존 학제간 공동 연구에 대한 연구 결과를 조사한다.

국내 현황은 여러 연구 분야의 전문가를 대상으로 자문회의, 세미나, 인터뷰 등을 통해 현실적인 문제와 애로사항을 파악한다. 과제의 특성상 전문가는 과학기술의 여러 분야를 포괄해서 구성하며 각 전문분야에서 소프트웨어가 어떻게 활용되고 있는지 개선이 필요한 부분은 무엇인지 조사한다. 자문회의는 자문위원의 연구 사례에 대한 세미나를 통해 과학기술연구에서 소프트웨어가 차지하는 역할과 중요성을 심도 있게 조사하고, 전문가 인터뷰는 현재 소프트웨어를 활용하는 융합 연구 과제를 수행하는 연구조직을 방문하여 연구책임자와 실무 연구자를 대상으로 애로사항과 기대 등에 대한 조사를 진행한다. 또한 공개된 토론회를 통해 과제에 관련되지 않은 외부 인사의 의견을 듣는 기회도 가진다.

이를 바탕으로 자문회의에서의 토의와 연구를 통해 국내 과학기술 연구에서 소프트웨어 활용을 개선할 수 있는 전략을 수립하고 제도적 적용 방안을 제시한다. 전략적 방침

은 단기적인 사업보다는 문제가 되는 사회적 이슈, 기술적 이슈, 제도적 이슈 등을 고려한 포괄적이고 장기적인 방안을 고안하는 것에 중점을 둔다. 도출되는 방안은 단계적 적용을 염두에 두고 실행 방법을 강구하며 관련 법제도를 조사하여 보완할 내용을 제시한다.

전문가간의 협업이라는 문제의 특성상 사회적, 문화적 변화가 필요하며 이런 변화는 쉽게 일어나지 않는다. 이런 점을 고려하여 지속적이고 점진적이더라도 국내 연구 문화를 개선할 수 있는 방안을 마련하는 것이 중요하다.

제2장 해외의 정책 체계 및 사례

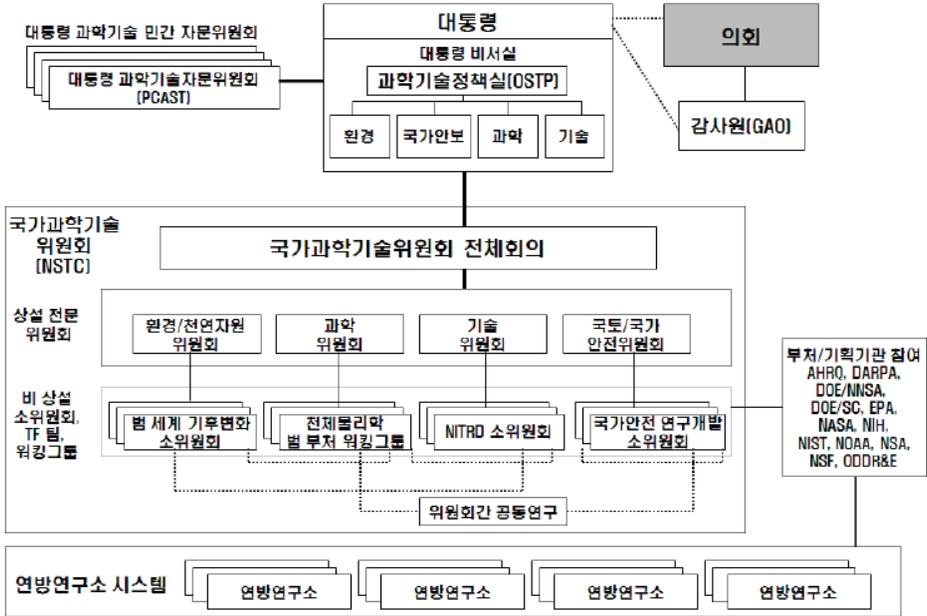
제1절 미국 과학기술연구의 SW활용 정책

1. 미국의 과학기술 정책 조정 체계

미국은 대통령실 산하의 국가과학기술위원회(NSTC, National Science and Technology Council)가 국가적 목표를 설정한다. 그 사무국으로 연방정부 차원의 과학기술정책 수립, 조정 및 예산 조정을 수행하는 과학기술정책실(OSTP, Office of Science and Technology Policy)이 대통령 비서실에 설치되어 있고, 과학기술 관련 이슈 및 역할에 대해 대통령에게 자문하고 민간기업의 참여와 투자를 독려하는 대통령 과학기술 자문위원회(PCAST, President's Council of Advisors on Science and Technology)가 있다. (성지은, 2010)

미국은 분산형 R&D 체계를 유지하면서도 국가우선순위 대상 선정과 범부처 R&D 프로그램을 통해 부처 미션 재조정 및 부처 간 협력을 강화하기 위해 국가과학기술위원회, 과학기술정책실, 대통령 과학기술 자문위원회 등의 대통령 직속 기관을 활용하여 균형을 유지하고 있다. 예산 편성은 관리예산실(OMB, Office of Management and Budget)의 담당이나 국가과학기술위원회와 산하 위원회의 의견이 예산에 충분히 반영되도록 과학기술정책실을 통해 긴밀히 협의하여 조정한다. 국가과학기술위원회는 예산에 대한 의견은 제시하지 않고 37개 분야별 소위원회(Subcommittee) 및 관계부처합동 실무그룹(Interagency Working Group)을 중심으로 전략적 수준의 투자우선순위 제시에 주력한다.(한국과학기술기획평가원, 2009; 성지은, 2010.)

[그림 2-1] 미국의 과학기술정책 조정체계



자료: 성지은, 세계 주요국의 과학기술혁신정책 조정체계와 최고 조정기구 비교분석, 2010.

2. 사이버인프라 추진 체계

미국은 이미 과학 분야 혁신을 위하여 사이버인프라(Cyberinfrastructure)를 생명/의과학, 나노 과학, 재료 과학, 사회 과학, 교육, 물리학, 화학, 환경 과학 등 전략적 기초/응용과학 분야에 폭넓게 사용하고 있다.

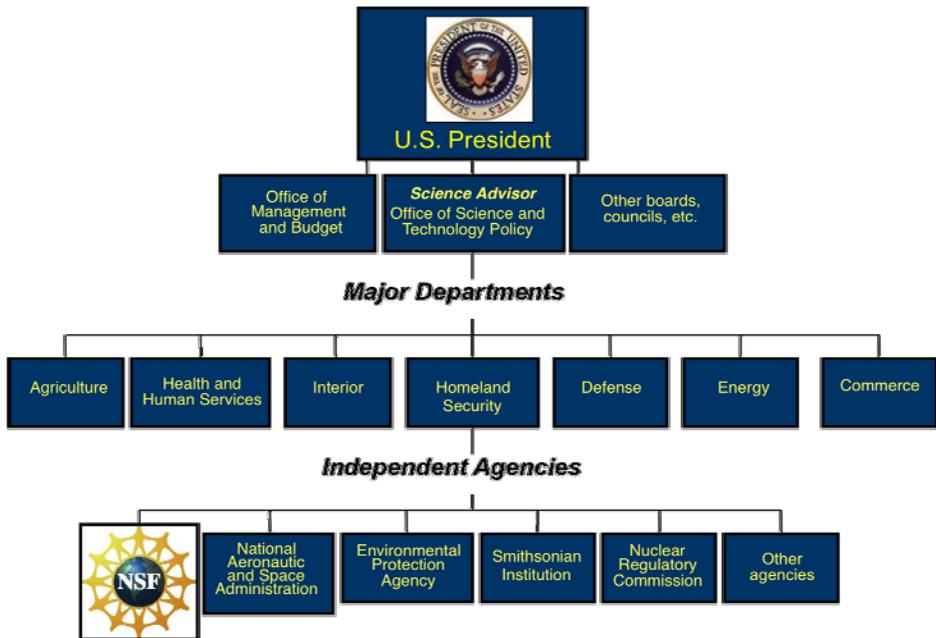
본래 1990년대 앨 고어에 의해 “국가 정보 인프라” 라는 용어가 일반화된 이후 1998년 대통령 정책 지침: 핵심 인프라 보안 부분(PDD-63)¹⁾에 대한 일반 대중 대상 발표에서 “사이버인프라” 라는 용어로 발전되었다.

이후 정부(국가과학기술위원회, 과학기술정책실 주도 전략 과학 분야 혁신 인프라 이니셔티브가 잇달아 발표되고 있으며, 국립과학재단(NSF, National Science Foundation)

1) The White House. Presidential Decision Directives 1998

사이버인프라 위원회에서 “고성능 컴퓨팅, 데이터 분석 및 가시화, 분산 커뮤니티의 가상 조직, 교육 및 인력 개발” 의 네 부분으로 나누어 21세기를 위한 사이버인프라 비전 (Cyberinfrastructure Vision for 21st Century Discovery²⁾)을 2007년 발표했다.

[그림 2-2] 사이버인프라 추진 체계



출처: Greer_ESIP_Jan2007_briefing.ppt³⁾

이어서 2012년에는 사이버인프라 프레임워크 비전문서(Cyberinfrastructure Framework for 21st Century Science and Engineering⁴⁾)를 발표하고 5개년 계획을 수행중이며, 현재는 선진 사이버인프라 프로그램(ACI, Advanced Cyberinfrastructure Program)의 일환으

2) National Science Foundation, Cyberinfrastructure Vision For 21st Century Discovery, March 2007. <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>

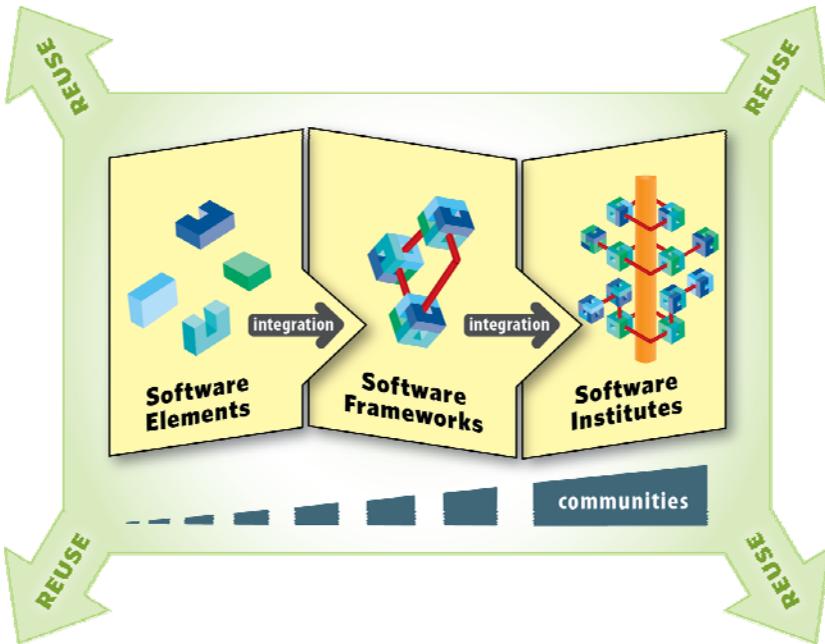
3) http://wiki.esipfed.org/index.php/File:Greer_ESIP_Jan2007_briefing.ppt

4) National Science Foundation. Cyberinfrastructure Framework for 21st Century Science and Engineering. 2012

로 11개 부문의 프로그램⁵⁾이 추진 중이다.

소프트웨어 인프라 투자는 크게 소프트웨어 구성 요소, 소프트웨어 프레임워크, 소프트웨어 허브로 작용하는 소프트웨어 기관으로 나뉘어 각각 투자된다. 이에 따라, 다양한 연구 기관과 커뮤니티의 필요성에 맞게 대규모 정부 지원(국립과학재단, 미 에너지부 선진 과학 컴퓨팅 연구 분과 등) 사이버인프라 프로젝트가 혼재하고 있다.

[그림 2-3] 소프트웨어 인프라 투자 영역

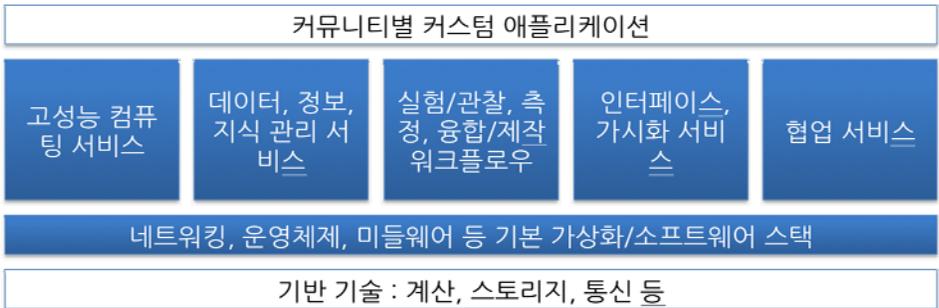


자료: Implementation of NSF CIF21 Software Vision (SW-Vision)⁶⁾

5) National Science Foundation. Division of Advanced Cyberinfrastructure Programs.
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504947&org=ACI&from=home
6) http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817&org=ACI&from=home

이러한 사이버인프라 구축 관련 프로그램의 종합적인 목표는 포괄적이고 안전하며 지속가능한 과학기술연구와 교육을 촉진하는 통합 사이버인프라 구조와 데이터 계산 중심 과학기술의 새로운 기능을 제공하여, 연구자가 선진 사이버인프라를 활용해서 새로운 통찰력을 얻고 좀 더 광범위하고 복잡한 도전 과제를 해결할 수 있도록 하는 것이다. 이를 위해 선진 사이버인프라는 개방성, 통합성, 확장성, 지속가능성이 요구되며, 장기적인 관점의 투자가 필요하다. 사이버인프라와 관련한 표준은 존재하지 않으나 주로 다음과 같은 구조로 구성된다.

[그림 2-4] 사이버인프라(파란색)의 일반적 구조



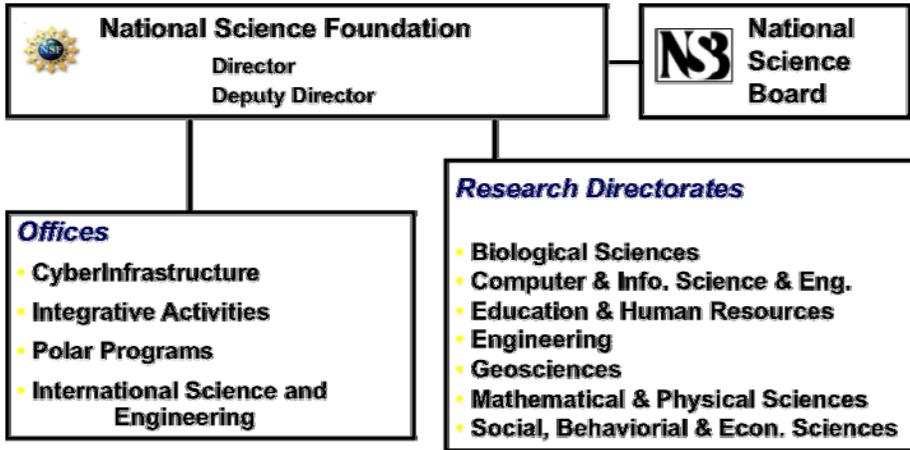
출처: National Science Foundation, 2003

[그림 2-4]의 일반적 구조에서 볼 수 있듯이 사이버인프라 기여 형태에 따라 데이터 공유와 통합, 연구 네트워크 형성과 협업 증진, 내, 외부 컴퓨팅 인프라의 활용, 인터페이스, 투명성 및 가시화 증진, 연구 워크플로우 개선 등으로 나누어 볼 수 있다.

3. 선진 사이버인프라 프로그램 (Advanced Cyberinfrastructure)

국립 과학 재단의 전체적인 추진 체계는 [그림 2-5]와 같다.

[그림 2-5] 국립 과학 재단 추진 체계



출처: Greer_ESIP_Jan2007_briefing.ppt

국립 과학 재단의 선진 사이버인프라 사업부는 과학 기술의 발전과 변화에 필수적인 첨단 사이버인프라 자원, 도구 및 서비스의 개발, 도입 및 제공을 조정하고 지원한다. 또한, 사이버인프라의 미래 역량 및 기능을 확장하는 미래 지향적 연구와 교육을 지원한다. 개발자, 연구원, 직원 및 사용자에게 걸친 숙련된 인력과 역동적인 기술적 생태계를 육성함으로써 고도 연산 및 데이터 처리, 네트워크에 의존하는 모든 분야에 걸쳐 과학자 및 기술자의 커뮤니티 성장을 뒷받침한다.

선진 사이버인프라 사업부는 슈퍼컴퓨터, 대규모 데이터 저장소 및 디지털화된 과학 데이터 관리 시스템, 소프트웨어 스위트, 도구, 라이브러리 및 프로그래밍 환경, 디지털 네트워크를 포함한 광범위한 사이버인프라 기술을 지원한다. 뿐만 아니라, 시스템의 효과적인 관리, 지속적인 유지 보수, 생산적인 활용이 가능하도록 다양한 교육 프로그램과 학술 교류 및 가상 조직을 지원한다.

선진 사이버인프라 프로그램에는 11개의 분과가 있다.

□ 캠퍼스 사이버인프라 (Campus Cyberinfrastructure - Infrastructure, Innovation and Engineering(CC*IE))

- 캠퍼스 사이버인프라 프로그램은 계산과학(Computational Science)의 컴퓨터 네트워크 및 시스템의 연구를 돕기 위한 광범위한 데이터 전송을 지원하기 위해 캠퍼스 수준의 개선과 재설계에 투자하는 프로그램이다. 과학 분야의 활용과 분산 연구 프로젝트의 성과, 안정성, 예측 가능성을 높이기 위해 네트워크 통합에 관련한 활동도 지원하고 있다.

<표 2-1> 캠퍼스 사이버인프라 프로그램 영역

프로그램 영역	설명
캠퍼스와 연구자를 위한 데이터 기반 네트워킹 인프라(Data Driven Networking Infrastructure for Campus and Researcher Award)	최대 2년간 최대 50만 불 지원
소규모 연구 기관을 위한 네트워크 설계 및 구현(Network Design and Implementation for Small Institutions Award)	최대 2년간 최대 35만 불 지원
네트워크 통합 및 응용 혁신(Network Integration and Applied Innovation Award)	최대 2년간 최대 100만 불 지원
ID 및 접근 관리 통합(Identity and Access Management Integration Award)	최대 2년간 최대 30만 불 지원
캠퍼스 CI 설계(Campus CI Engineer Award)	최대 2년간 최대 40만 불 지원
고급 네트워킹을 통한 지역적 조정 및 파트너십(Regional Coordination and Partnership in Advanced Networking Award)	최대 2년간 최대 15만 불 지원

출처: Division of Advanced Cyberinfrastructure Programs (ACI)⁷⁾

□ 데이터 인프라 구성 요소 (Data Infrastructure Building Block, DIBBs)

- 21세기 과학 및 공학 기술을 위한 사이버인프라 프레임워크(CIF21, Cyberinfrastructure Framework for 21st century)에 대한 국립과학재단의 비전은 통합된 확장성 있고 지속가능한 사이버인프라를 과학 및 공학 분야의 혁신을 위해 필수적인 것으로 간주한다. 데이터 인프라 구성

7) http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504947&org=ACI&from=home

요소(DIBBs) 프로그램은 사이버인프라 프레임워크의 중요한 부분이다. 데이터 인프라 구성 요소 프로그램은 데이터 기반 질의와 관련한 다양한 영역의 학제 및 공동 연구를 가속화하기 위해 강력하며 공유된 데이터 중심의 사이버인프라 개발을 장려하고 있다.

- 사이버인프라 구성 요소가 연구원의 데이터 요구사항을 만족하도록 보장하기 위해 사이버인프라 전문 지식과 각 분야 현업 연구자들의 협업을 요구한다. 모든 관련 활동은 학제 및 공동 연구에서 발생하는 데이터 문제를 해결해야 한다.

<표 2-2> 데이터 인프라 구성 요소 프로그램 영역

프로그램 영역	설명
파일럿 데모(Pilot Demonstration Awards)	한 연구 분야 내의 광범위한 요구사항 해결
조기 구현(Early Implementation Awards)	여러 연구 분야 내의 광범위한 요구사항 해결

자료: Division of Advanced Cyberinfrastructure Programs (ACI)

□ 고성능 컴퓨팅 시스템 도입(HPC System Acquisition: Continuing the Building of a More Inclusive Computing Environment for Science and Engineering)

- 21세기 과학 및 공학 기술을 위한 사이버인프라 프레임워크(CIF21)에 대한 국립 과학 재단의 비전은 오늘날 과학 및 교육 공동체가 직면하고 있는 가장 복잡한 문제와 이슈를 해결하기 위해 필요한 고도의 계산이나 데이터 기반 역량을 갖추 수 있도록 하는 것에 초점을 두고 있다. 이를 위해 모든 분야의 과학 및 공학 기술 연구자들이 혜택을 볼 수 있도록 다양한 수준의 종합적이고 혁신적인 인프라 내 고도 계산 역량을 갖추 필요가 있다. 그동안 세계적 수준의 고성능 컴퓨팅 환경을 배치하고 지원함으로써 페타 단위 규모의 역량을 갖추게 하는 데 집중

하였다면, 최근에는 다양한 분야의 과학 프로세스를 포함하는 복잡하고 동적이며 다양한 워크플로우의 통합에 초점을 두고 있다. 일부 애플리케이션은 고도의 데이터 기반 연산이 필요하다.

□ 국제 연구 네트워크 연결(International Research Network Connections, IRNC)

- 국제 연구 네트워크 연결 프로그램은 국립 과학 재단의 연구 커뮤니티가 참여하는 국제 과학 및 공학 연구 및 교육 협력에 필요한 고성능 네트워크를 지원한다. 이 프로그램에서 투자하는 고성능 네트워크는 과학 및 공학 연구와 교육 애플리케이션을 지원하기 위한 것으로 미국의 연구 및 교육 커뮤니티가 최첨단 국제 네트워크 서비스나 늘어가는 협업 및 데이터 서비스에 쉽게 접근할 수 있도록 한다. 확장된 국제 네트워크를 통해 이용할 수 있는 연구 및 생산 네트워크 서비스가 늘어나게 되어 국내 연구 네트워크 서비스를 보완하게 된다.

<표 2-3> 국제 연구 네트워크 연결 프로그램 영역

프로그램 영역
미국 연구 네트워크를 전 세계 연구 네트워크와 연결하여 기존 국제 네트워크를 개선
공개 네트워크 연결지점에 대한 미국의 기반구조 및 혁신을 지원
연구 및 교육 네트워크 운영 센터를 위한 혁신 및 최첨단 역량 함양 지원
국제 네트워크 경로 간 선진 네트워크 측정 역량과 서비스의 개발, 응용 및 활용 활성화
선진 네트워크 기술에 대한 연구 커뮤니티의 훈련 및 네트워크 성능 개선을 조정하고 지원

출처: Division of Advanced Cyberinfrastructure Programs (ACI)

- 사이버인프라를 활용한 고도 연산이 필요한 과학 분야의 국립 과학 재단 펠로우(NSF Fellowships for Transformative Computational Science using Cyberinfrastructure)
 - 학문의 경계를 넘나드는 중요한 신 성장 분야인 계산과학 (Computational Science)의 미래 지도자를 양성하기 위해 이 분야 박사 후 과정을 지속하는 데 관심이 있는 신규 박사 취득자인 과학자 및 엔지니어를 지원한다. 성공적인 펠로우는 예를 들면 자신의 분야에서 혁명적인 진보를 이루기 위해 사이버인프라를 활용할 수 있고, 또는 혁신적인 연산 방법을 가능하게 하기 위해 사이버인프라 기반 기술을 배포할 수 있다. 펠로우는 10%~25%의 시간을 교육 활동에 할애해야 한다는 조건이 있으며, 제안 마감은 2014년 11월 14일로 현재 제안 단계에 있는 프로그램이다.

- 페타 단위 규모의 컴퓨팅 자원 할당(Petascale Computing Resource Allocations)
 - 2013년 국립 과학 재단이 새로 투자한 페타 급 컴퓨팅 시스템인 푸른 바다(Blue Water)가 일리노이 주립 대학에 설치되었다. 이 프로젝트 및 시스템의 목적은 다양한 분야에서 훨씬 더 규모가 크고 복잡한 연구 도전과제들을 해결할 수 있도록 고도의 연산 역량을 제공하여 과학 및 공학 분야에 새로운 지평을 여는 것이다. 이 입찰의 목적은 푸른 바다 시스템의 자원을 할당받을 수 있는 연구 그룹을 섭외하는 것으로 제안자는 페타 급 컴퓨팅 자원을 필요로 하는 과학 및 공학 과제를 제시해야 한다. 제안 마감은 2014년 11월 14일로, 현재 제안 단계에 있는 프로그램이다.

- 사이버인프라 소프트웨어 개발 (Software Development for Cyberinfrastructure)
 - 사이버인프라 소프트웨어 개발의 목적은 광범위한 과학 및 공학 애플리케이션이 혜택을 받을 수 있도록 재사용 가능하고 확장 가능한 소프트웨어 구성 요소와 시스템을 개발하고 배치하는 것이다. 이 프로그램은 두 가지 영역을 지원한다.

〈표 2-4〉 사이버인프라 소프트웨어 개발 프로그램 영역

프로그램 영역

엔드투엔드 고성능 컴퓨터 네트워킹

사이버 보안

출처: Division of Advanced Cyberinfrastructure Programs (ACI)

- 지원되는 활동에는 다양한 분야의 연구 및 교육에 있어 개발, 테스트, 실험적 배치 및 시범 사용이 모두 포함된다. 중요한 점은 인프라의 연구에서 인프라 역량으로 초점이 변경되었다는 것이다. 이 프로그램은 데이터, 도구 및 컴퓨팅, 스토리지 자원의 공유를 통해 연구와 교육 활동의 협업을 촉진하고 과학적 생산성을 증진하는 모든 소프트웨어 활동에 투자하며, 오픈 소스 소프트웨어를 통해 개발해야 한다는 요건이 있다. 제안은 마감되었다.
- 지속 가능한 혁신을 위한 소프트웨어 기반구조(Software Infrastructure for Sustained Innovation)
 - 소프트웨어는 계산, 실험 및 이론의 중요한 원동력으로서 새로운 연산 인프라의 핵심 구성 요소이다. 과학적 발견 및 혁신이 점점 더 정교한 소프트웨어의 개발에 의해 근본적으로 새로운 경로를 따라 진행되고 있다. 소프트웨어는 직접적으로 과학적 생산성을 향상시키고 연구원의 역량을 증진시킬 책임을 진다. 국립 과학 재단은 과학 발전을 위해 이러한 핵심 원동력을 양성, 가속 및 지속 가능하게 하기 위해 소프트웨어 인프라 프로그램을 설립하였으며, 이는 연구 및 교육의 혁신을 사이버인프라의 중요한 부분인 지속 가능한 소프트웨어 자원으로 변환하는 것이다. 다음 세대의 연구자와 미래 사이버인프라 발명가의 교육에 기여할 수 있는 견고하고 안정성 있으며 사용 가능하고 지속 가능한 소프트웨어 인프라에 투자한다.

〈표 2-5〉 지속 가능한 혁신을 위한 소프트웨어 기반구조 세부 분야

프로그램 영역	설명
과학 소프트웨어 요소 (Scientific Software Elements)	하나 이상의 전략적 과학 및 공학 분야의 진보를 위한 요구 사항을 해결할 수 있도록 견고한 소프트웨어 구성 요소를 생성 및 배치할 소규모 그룹에 부여
과학 소프트웨어 통합 (Scientific Software Integration)	다양한 분야에 걸친 공통의 연구 도전 과제를 해결하기 위한 공통 소프트웨어 기반구조를 개발하고 활용할 좀 더 큰 규모의 그룹에 부여
과학 소프트웨어 혁신 연구 기관 (Scientific Software Innovation Institutes)	대규모 연구 커뮤니티를 위한 소프트웨어 기반구조 및 기술의 장기적 허브를 설립

출처: Division of Advanced Cyberinfrastructure Programs (ACI)

- SSE 제안 기간은 2015년 1월 19일부터 2월 2일까지, SSI 제안 기간은 2015년 6월 12일부터 6월 26일 까지 이다.
- 사이버인프라를 위한 전략 기술 (Strategic Technologies for Cyberinfrastructure)
 - 사이버인프라를 위한 전략 기술 프로그램의 주요 목적은 사회 전체에 걸쳐 광범위한 영향을 미치는 첨단 과학 및 공학 연구와 교육을 실현하기 위한 실험적/혁신적인 하드웨어나 소프트웨어 시스템, 또는 기타 고유한 사이버인프라 활동을 모두 지원하기 위한 것이다. 이 프로그램은 사이버인프라 프레임워크 형성 작업(CIF21)의 일환으로, 제안자들은 제안자의 현재 시스템이 혁신적이고 확장성 있으며 활용성이 강력하고 사용성이 높은 사이버인프라로 진화할 수 있는 잠재성이 있음을 보여 줄 수 있어야 한다.

- 실험적 시스템은 고성능 컴퓨팅, 대규모 데이터 집약적 컴퓨팅, 시각화, 분석 및 혁신적인 네트워킹 분야의 하드웨어 및 소프트웨어를 포함할 수 있다. 산학 협력 관계를 포함하는 프로젝트는 프로그램 관리자와 상의 후 제안서를 제출한다.
- 사회 기술적 시스템으로서의 가상 조직 (Virtual Organization as Sociotechnical Systems)
 - 사회 기술적 시스템으로서의 가상 조직 프로그램은 근본적인 사회, 조직 및 설계 과학 연구, 특히 가상 조직을 개발하는 방법이나 어떤 요건에서 가상 조직이 과학적, 공학적, 또는 교육적 생산 및 혁신을 활성화하고 진보시킬 수 있는지에 관한 이해를 증진하기 위해 고안되었다. 분석의 수준은 개인, 그룹, 조직 및 제도적 장치 등을 포함할 수 있으며, 이에 국한되지 않는다. 또한 인류학, 복잡성 과학, 컴퓨터 및 정보 과학, 의사 결정 및 관리 과학, 경제학, 공학, 조직 이론, 조직 행동, 산업 심리학, 행정학, 정치학 및 사회학 등의 분야를 포함하며, 연구 방법은 광범위한 정성적, 정량적 분석 방법을 모두 포함할 수 있다.

4. 개방형 과학 그리드(Open Science Grid)⁸⁾

개방형 과학 그리드는 자원 공급자 및 과학 연구 기관에 공통 서비스 및 지원을 제공하는 컨소시엄으로 높은 처리량을 보이는 연산 서비스의 분산된 패브릭을 활용한다. 개방형 과학 그리드는 자원을 소유하고 있지는 않지만 소프트웨어 및 서비스를 사용자와 자원 제공자에게 제공하여, 자원의 공유를 촉진한다. 개방형 과학 그리드는 미국 에너지부와 국립 과학 재단이 공동 투자하고 있다.

개방형 과학 그리드는 다음과 같은 과학 분야를 지원한다.

- 고 에너지 물리학 : CMS와 ATLAS
- 나노과학 : nanoHUB
- 구조 생물학 : SGrid
- 커뮤니티 가상 조직 (다양한 과학 분야) : 참여 중

개방형 과학 그리드는 주로 어떠한 과학 문제를 각각 독립적으로 수행될 수 있는 수많은 독립 작업으로 분할하여 해결하는 데 사용된다. 개방형 과학 그리드에서 수행되는 가장 성공적인 공유 애플리케이션은 다음과 같은 특징을 갖고 있다.

- x86 혹은 x86_64 아키텍처의 리눅스 애플리케이션
- 메시지 패싱을 요구하지 않는 멀티 스레딩 혹은 싱글 스레딩
- 1~24시간의 짧은 런타임
- 예측되지 않은 재시작 및 종료를 처리할 수 있음
- 라이선스가 필요한 서버에 접속하지 않는 소프트웨어로 구성

8) 개방형 과학 그리드 개요, <http://www.opensciencegrid.org/about/>

- 과학 문제는 여러 독립 작업으로 구성된 워크플로우로 설명 가능함
- 적은 수의 큰 작업 대신 매우 많은 수의 작은 작업을 실행

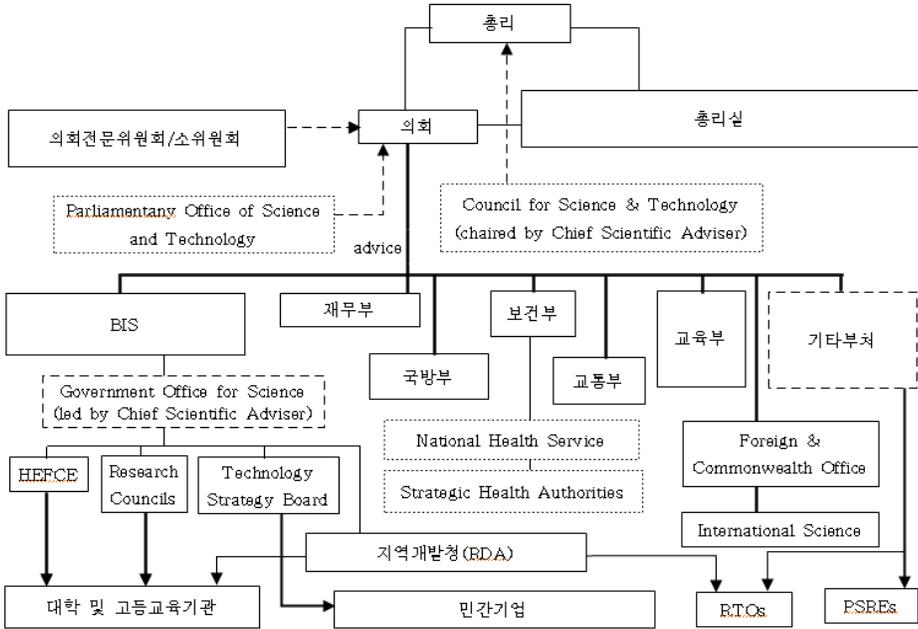
제2절 유럽 과학기술연구의 SW활용 정책

1. 영국의 과학기술 정책 조정 체계

영국은 산업계 학계 정부부처가 참여하여 필요에 따라 정책을 설정하고 조정하는 다원주의 체계이다. 과학기술분야는 브라운 정부 출범 이후('07년 6월) 혁신대학기능부(Department of Innovation, Universities and Skills)를 신설하여 과학기술혁신과 고등교육 기능을 통합하고, 사업·기업·규제개혁부(DBERR, Department for Business, Enterprise and Regulatory Reform)를 신설하여 규제개혁을 통한 기업 친화적 환경 구축을 추진하였다. 이어서 2009년에는 이 두 부서를 통합하여 기업혁신기술부(BIS, Department for Business, Innovation and Skills)를 신설했다. 이것은 연구개발에 기반을 두고 산업계의 혁신을 추진하겠다는 의지를 표명한 것이다.(한국과학기술기획평가원, 2009; 성지은, 2009; 홍형득, 2014)

BIS에 통합된 혁신대학기능부의 주요 업무는 세계수준의 연구기반 유지와 발전 및 고등교육의 질적 향상 및 과학기술분야 인력 양성이며 과학기술, 대학 및 과학기술관련 정책의 부처 간 조정 업무도 담당하고 있다. 산하부서인 과학청(GOS, Government Office for Science)은 정부 수석 과학 고문(GCSA, Government Chief Scientific Adviser)을 책임자로 하여, 부처 간 과학기술 정책 조정, 예측 및 과학기술위원회(CST, Council for Science & Technology)의 사무국 역할을 수행하고, 실제 과학기술연구를 주관하는 각종 연구위원회(Research Councils)와 기술전략회의(TSB, Technology Strategy Board) 등의 과학예산 배분 정책을 담당한다.

[그림 2-6] 영국 과학기술 행정 체계



자료: 홍형득, 영국의 과학기술행정체제와 투자구조의 변화와 특징분석, 2014.

영국의 연구위원회는 다음과 같은 특징을 가지고 있다. 첫째, 영국정부는 연구기관의 연구에 관한 자율권을 보장하여 주는 장치로써 정부와 연구기관 사이에 연구회를 두고 있으며, 연구기관을 정부의 간섭과 통제로부터 차단하고 보호하는 원칙을 홀데인원칙(Haldane Principle)이라는 이름으로 정착시켰다. 둘째, 연구위원회는 정부를 대신하여 연구기관을 관리할 뿐만 아니라, 예산당국으로부터 산하 연구기관을 위한 예산을 총액예산(Block Grant)으로 받아서 연구기관에 배분한다. 셋째, 산하 연구기관에 대하여 예산, 전략기획, 연구단위의 구조조정, 등 실질적인 집행권을 행사한다. 연구위원회와 연구기관의 각각의 권한과 책임은 역사적으로 발전하여 오면서 분명하게 정착하였고, 통제와 간섭은 명시적(관행적이 아닌) 규칙의 범위 안에서만 행해진다. 넷째, 연구위원회는 산하 연구기관과 대학의 연구역량을 접목하는 기능을 하고 있다. (홍형득, 2014)

2. 영국의 e-Science(e-infrastructure) 프로그램

영국은 1999년 영국과학기술청(UK Office of Science and Technology)의 John Taylor가 e-Science를 주장한 이후 지속적으로 과학 연구에 컴퓨터 기술을 도입하고 있다.⁹⁾ e-Science는 과학 연구에서 컴퓨터를 통해 실험 데이터의 수집, 정리, 분석과 장기적인 데이터의 저장과 효율적인 접근을 이뤄냄으로써 과학적 발견을 촉진하는 것을 의미한다.

영국은 2001년부터 e-Science 프로그램을 연구위원회와 무역산업부(Department of Trade and Industry)가 운영하였다. 제1단계(2001~2004년)와 제2단계(2003~2006년)에 걸쳐 2억1천3백만 파운드가 지원되었다. 다음과 같은 과학위원회를 통해 각각의 분야에서 시범사업을 진행하였다.

- 생명공학 연구위원회 (BBSRC, Biotechnology and Biological Sciences Research Council)
- 경제·사회 연구위원회 (ESRC, Economic and Social Research Council)
- 공학 및 자연과학 연구위원회 (EPSRC, Engineering and Physical Sciences Research Council)
- 의학 연구위원회 (MRC, Medical Research Council)
- 환경과학 연구위원회 (NERC, Natural Environment Research Council)
- 대형연구시설운영 연구위원회 (CLRC, Central Laboratory of the Research Councils)
- 입자물리학 및 천문학 연구위원회 (PPARC, Particle Physics and Astronomy Research Council)

9) <http://en.wikipedia.org/wiki/E-Science>

1, 2단계에서는 공학 및 자연과학 연구위원회가 연구위원회를 대표하여 e-Science 프로그램의 연구 관리를 맡아 진행하였다. 제3단계(2006-2011)에서는 공학 및 자연과학 연구위원회가 e-Science 프로그램을 대표하는 것에서 나아가 연구의 관리와 책임을 맡는 것으로 프로그램 구조가 변화하였으며, 현재는 공학 및 자연과학 연구위원회가 e-Science 프로그램을 이어 e-Infrastructure라는 프로그램으로 진행 중이다.

e-Science 프로그램 진행 기간 동안 영국 정부와 산업계뿐만 아니라 유럽 연합과, 대학, JISC(Joint Information Systems Committee)로부터 지원을 받았다. 또한 사업 기간 동안 글래스고와 에든버러 대학에 National e-Science 센터가 설립되었으며, 유럽 전역에 26개의 e-Science 지원 센터가 설립되었다.

2009년 연구위원회의 보고서에 따르면 e-Science 프로그램을 통하여 공유할 수 있는 소프트웨어의 표준과 온톨로지를 설립하는데 성공적이었다고 평가하였다.¹⁰⁾¹¹⁾ 특히 생명과학과 약학, 재료, 에너지 분야에서 효과적이었다고 밝혔다. 2001년부터 2009년까지 e-Science 프로그램을 통해 966개의 유명 저널 투고, 2054개의 컨퍼런스 투고, 181개의 소프트웨어 제작이 이루어졌으며, 관련 분야의 저널을 합치면 11,210개의 논문이 투고되었다. 또한 보고서에서는 이 프로젝트를 통해 과학자와 컴퓨터 과학자 사이의 커뮤니티가 형성된 것은 큰 성과라고 밝혔다.

10) Atkins. RCUK Review of e-Science: Building a UK Foundation for the Transformative Enhancement of Research and Innovation. 2010

11) 공유된 개념화(shared conceptualization)에 대한 정형화되고 명시적인 명세(formal and explicit specification). 온톨로지는 단어와 관계들로 구성된 일종의 사전이며, 그 속에는 특정 도메인에 관련된 단어들에 계층적으로 표현되어 있고, 추가적으로 이를 확장할 수 있는 추론 규칙이 포함되어 있어, 웹 기반의 지식 처리나 응용 프로그램 사이의 지식 공유, 재사용 등이 가능토록 되어 있다. 온톨로지는 시맨틱 웹 응용의 가장 중심적 개념으로서, 이를 표현하기 위해 스키마와 구문 구조 등을 정의한 언어가 온톨로지 언어(ontology language)이며, 현재 DSML+OIL, OWL, Ontolingun 등이 있다.

<표 2-6> 연구 위원회 별 주요 시범 사업

연구위원회	시범 사업명
공학 및 자연과학 연구위원회	CombEchem
	DAME
	Discovery Net
	GEODISE
	myGrid
	RealityGrid
	GOLD
	Integrative biology
	PMSEG (MESSAGE)
	CARMEN
	NanoCMOS
입자물리학 및 천문학 연구위원회	GridPP
	AstroGrid
생명공학 연구위원회	Biomolecular Grid
	Proteome Annotation Pipeline
	High-Throughput Structural Biology
	Global Biodiversity
의학 연구위원회	Biology of Ageing
	Sequence and Structure Data
	Molecular Genetics

연구위원회	시범 사업명
	Cancer Management
	Clinical e-Science Framework
	Neuroinformatics Modeling Tools
환경과학 연구위원회	Climateprediction.com
	Oceanographic Grid
	Molecular Environmental Grid
	NERC DataGrid

또한 영국은 고등교육기금위원회(Higher HEFCE, Higher Education Funding Council for England)등의 정부 기관으로부터 지원을 받는 비정부 공공기관으로 JISC(Joint Information System Committee)를 운영하고 있다. JISC는 1993년 고등교육에서 ICT를 교육하고, 연구에서 ICT를 활용하는데 도움을 주기 위해 설립되었다. JISC는 e-infrastructure 분야의 R&D를 지원하고 있다. Super JANET 5, Integrated Information Environment, Virtual Research Environments, Digital Repositories, Core Middleware Infrastructure and Technology Development, Semantic Grid and Autonomic Computing, Shared Services 등이 JISC의 지원을 받아 연구되고 있으며, 연구위원회와의 협력을 통해 e-infrastructure를 지원하기도 한다.

3. 유럽연합 현황

유럽연합은 R&D 프로그램인 FP6(6th Framework Programme for Research and Technological Development)와 FP7(7th Framework Programme for Research and Technological Development)의 일환으로 e-Infrastructure 연구를 지원하였다. 현재는 FP7을 잇는 Horizon 2020 프로그램을 통해 지속적으로 e-infrastructure 분야를 지원할 예정이다.¹²⁾

FP6는 2002년부터 2006년까지 진행한 프로그램으로 유럽 연구 집중 및 통합 영역(Focusing and Integrating European Research), 유럽 연구 구조화 영역(Structuring the European Research Area), 유럽 연구 기반 강화 영역(Strengthening the Foundations of European Research Area)의 3가지 방향으로 구성되어있으며, 이중 유럽 연구 구조화 영역(Structuring the European Research Area)의 연구 인프라 프로그램의 일환으로 e-infrastructure 프로그램을 진행하였다.¹³⁾ [그림 2-7]에서 보듯이 FP6 프로그램에서 범 유럽지역연구망(GÉANT: European Multi-gigabit Computer Network for Research and Education)과 같은 연구자 네트워크와 그리드 인프라를 이용하여 e-infrastructure를 구성하였다.

FP7은 2007년부터 2013년까지 진행한 프로그램으로 협업(Cooperation), 아이디어(Ideas), 사람(People), 역량(Capacities), 핵연구(Nuclear Research)의 5가지 주제로 진행되었으며, 이중 역량(Capacities) 분야에서 연구 인프라 프로그램의 일환으로 e-Infrastructure를 진행하였다. 이는 ICT 기반의 데이터와 실험 장비들을 원격으로 제어하여 ‘in-silico(가상실험)’ 이 가능한 환경과 동시에 연구자들 간의 가상의 커뮤니티를 설립하는 것을 목적으로 한다.¹⁴⁾

유럽연합은 2002년 유럽 연구 인프라 전략 포럼(ESFRI: European Strategy Forum on Research Infrastructures)을 설립하였다. 포럼은 유럽연합 국가들 간의 연구 인프라에 관련된 협약을 지원하고 정책적 결정을 도와 전략적으로 연구 인프라를 구축하는 것을

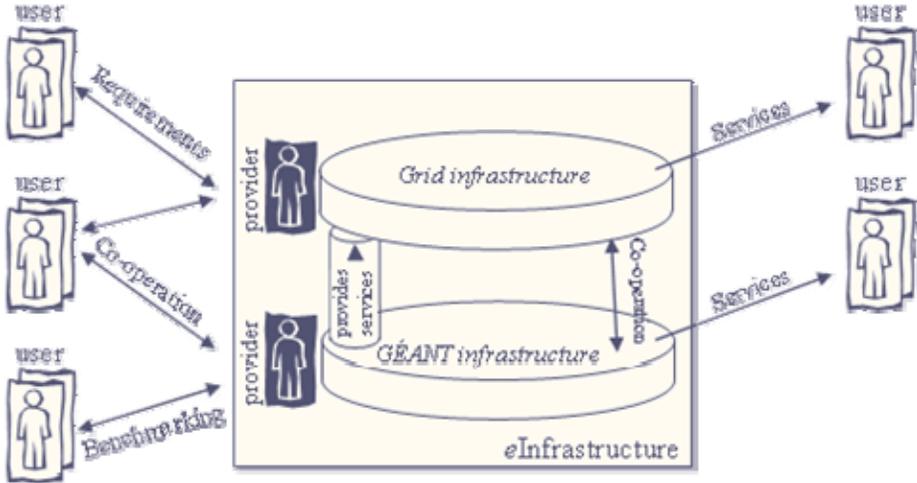
12) <http://ec.europa.eu/programmes/horizon2020/>

13) <http://cordis.europa.eu/fp6/activities.htm>

14) http://cordis.europa.eu/fp7/home_en.html

목표로 한다.¹⁵⁾

[그림 2-7] FP6에서 도입된 e-Infrastructure의 구조



출처: <http://cordis.europa.eu/ist/rn/ri-cnd/e-infrastructures.htm>

유럽 연구 인프라 전략 포럼은 2006년과 2008년, 2010년에 전략 로드맵을 발표하여 사회과학, 생물학, 환경, 에너지, 의료, 재료, 물리학 등의 분야에서 필요한 인프라를 파악하고 유럽 국가들의 지원을 통해 구축하고 있다. 또한 2009년부터는 유럽 연구 인프라 컨소시엄(ERIC: European Research Infrastructure Consortium)을 설립하여 범국가적 인프라 구축에 필요한 법적인 프레임워크를 만들었다.

제3절 과학기술연구의 소프트웨어 활용 사례

해외 과학기술연구에서 소프트웨어를 활용한 대표적 사례 몇 가지를 소개한다. 보다 많은 사례는 부록에 첨부하였다.

15) http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

1. 입자 물리학(Particle Physics) - Higgs 입자 실험을 위한 아틀라스와 구글 컴퓨트 엔진 사례¹⁶⁾

□ 산업 소개

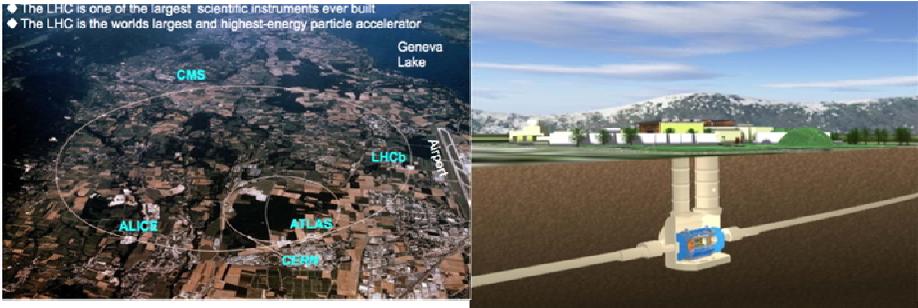
- 입자 물리학은 세상을 구성하는 가장 기본적인 입자를 다루는 분야로 물질을 구성하는 기본적인 입자가 무엇이고, 입자 사이의 상호작용을 규명하여 자연현상의 본질에 대해 탐구하는 학문이다. 고에너지 물리학이라 불리기도 한다.

□ 개요

- 유럽 입자 물리학 연구소(CERN)는 스위스 제네바와 프랑스 사이의 국경지대에 위치한 세계 최대의 입자 물리학 연구소이다. 원래 명칭은 유럽 원자핵 공동 연구소(Conseil Européen pour la Recherche Nucléaire)였고, 이를 따라서 CERN이라 불린다.
- CERN은 설립 초기부터 입자 가속기 등을 이용해, 고 에너지 물리학 연구에 많은 기여를 하였다. 또한 물리학자들의 문헌 검색 및 제휴를 위하여 고안된 HTML과 월드 와이드 웹의 발상지로도 유명하다.

16) Google I/O 2013. <https://developers.google.com/events/io/sessions/333315382>

[그림 2-8] LHC와 지하에 묻힌 ATLAS 탐지기



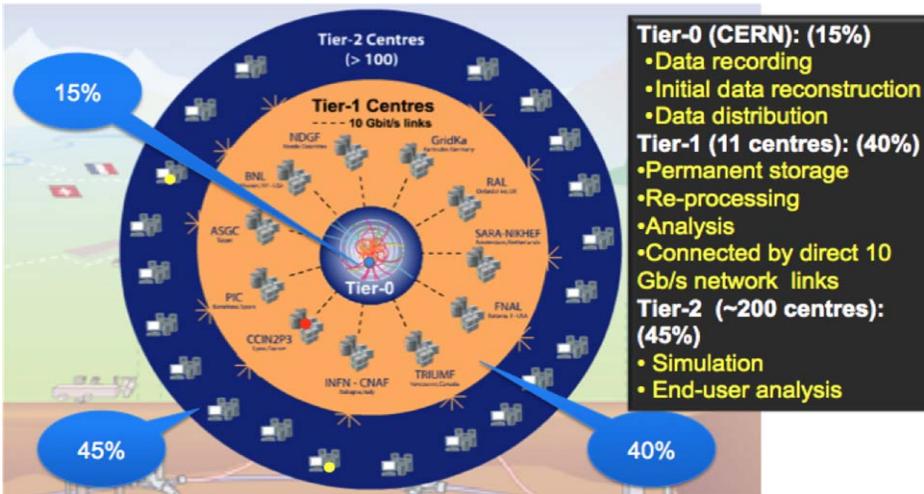
자료: <https://developers.google.com/events/io/sessions/333315382>

- 그 중에서도, CERN LHC의 ATLAS 실험에는 38개 국가의 3000명의 과학자, 1200명의 대학원생이 참여 중이다. ATLAS 실험이 해결해야 하는 주요 현안은 다음과 같다.
 - 입자 물리학의 고에너지 프론티어를 탐색
 - 신생 물리학 연구
 - 힉스 입자와 그 특성
 - 표준 모형 이상의 물리학 (SUSY, Dark Matter, extra dimensions, Dark Energy 등)
 - 표준 모형 파라미터의 정확도 측정
- 이러한 실험을 수행하기 위해 해결해야 했던 도전 과제는 다음과 같다.
 - ATLAS는 2013년 기준 140PB 데이터와 O(100)분산 컴퓨팅 센터를 전 세계적으로 다루고 있기 때문에 운영, 관리적 어려움이 있음
 - 매년 40 PB 데이터가 생성
 - 40 여개 이상의 국가에서 온 과학자가 협업해야 함

□ 연구 방법

- PanDA 워크로드 관리 시스템으로 매년 O(100)사이트의, O(100000) 코어, O(1000000000)의 작업을 처리하고 O(1000) 사용자를 관리

[그림 2-9] PanDA 워크로드 관리 시스템

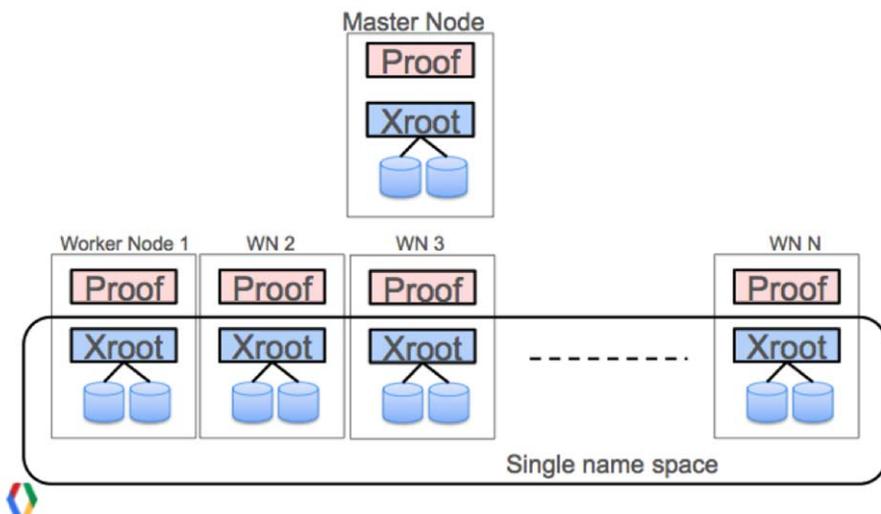


자료: <https://developers.google.com/events/io/sessions/333315382>

- 구글 컴퓨트 엔진 트라이얼 기간에 ATLAS 일부 작업을 8주간 시험
 - 고성능 분석 클러스터(PROOF)
 - 클라우드 스토리지 및 데이터 관리
 - XRoot와 클라우드 스토리지 간 집계 및 상호작용
 - 몬테카를로 시뮬레이션을 위한 PanDA 큐 활용

[그림 2-10] XRoot와 PROOF 클러스터

PROOF/Xrootd Cluster

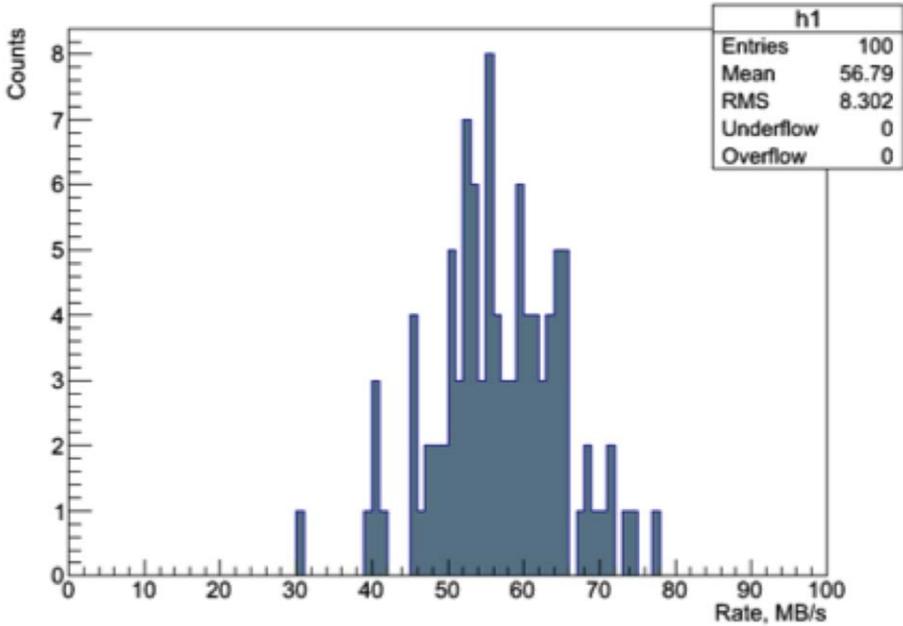


자료: <https://developers.google.com/events/io/sessions/333315382>

□ 활용 효과

- 클라우드에서의 데이터 전송 및 분산 환경 실험을 통해 새로운 인프라 확보 가능성을 타진하였으며, 데이터 접근을 용이하게 하기 위해 지연 시간, 하드웨어 요구 사항, 인건비를 최소화하고 확장성을 최대화하는 방안을 모색할 수 있었다.
 - 영구 디스크 스토리지 클러스터를 활용하여 더 나은 확장성 및 피크 성능을 확보
 - RAID를 통해 더 나은 성능을 확보
 - 공용 네트워크를 사용하더라도 XRoot연합과 GCE 스토리지 클러스터 간 57 Mbps의 전송속도를 보여줌

[그림 2-11] 전송 속도



자료: <https://developers.google.com/events/io/sessions/333315382>

2. Folding@home 프로젝트

□ 산업 소개

생명의공학(Biomedical Science and Engineering)은 의학과 공학 분야 간에 학문적 방법론, 개념, 기술, 기기 등을 상호 교환하고 이용함으로써 궁극적으로 두 학문 분야의 발전을 도모하는 학문이다.

□ 개요

Folding@home은 자발적 참여자의 컴퓨터 유휴 시간을 이용하여 많은 계산이 필요한 단백질 폴딩¹⁷⁾, 약물 설계, 기타 유형의 분자 역학을 시뮬레이션하는 분산 컴퓨팅 프로젝트이다. 이 프로젝트의 목적은 알츠하이머, 헌팅턴, 파킨슨 질병을 비롯한 다양한 암을 치료하기 위한 것이다. 이를 위해 단백질 폴딩 메커니즘을 연구하여 단백질의 최종 3차원 입체 구조를 파악하고, 잘못된 폴딩의 원인을 파악하고, 이를 약물 설계에 응용하기 위한 노력을 하고 있다.

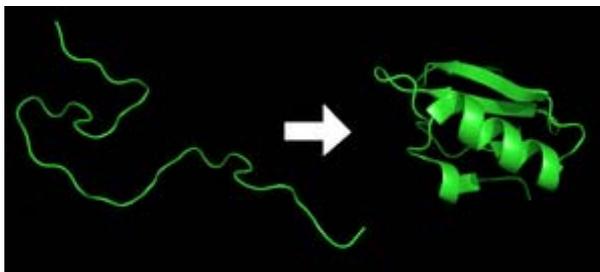
이 프로젝트는 자발적 참여자의 컴퓨터에 소프트웨어를 설치하고 컴퓨터가 일을 하지 않을 때의 유휴 처리 리소스를 활용한다. 2014년 7월 현재 179237명이 참여하고 있다. 이 프로젝트는 스탠포드 판데 실험실에서 비제이 판테의 지휘 하에 개발 및 운영되고 있으며 전 세계에 걸쳐 다양한 과학 기관 및 연구소(노르트담 Izaguirre 랩, 버지니아 주립대학, 스톡홀름 대학, 콜로라도 주립대학, HKUST, CSULB, 템플 대학, 크로아티아 생명 과학 지중해 연구소 등)에 공유되었고, 다양한 상업적 파트너(인텔, 구글, 소니, ATI, NVIDIA, 스톤하퍼 등)와 커뮤니티 봉사자들의 참여로 진행되고 있다.

□ 구현 방법

이 프로젝트는 자원 봉사자들이 소유한 개인용 컴퓨터의 유휴 처리 리소스를 활용하

-
- 17) 단백질 분자를 구성하는 펩티드 사슬이 고유의 2차구조의 배치순서를 유지하며 공간적 배치를 통해 3차원 구조를 형성하는 것. 다음 그림은 단백질 폴딩의 전후를 나타내는 그림이다.

[그림 2-12] 단백질 폴딩 전 후 모습



<http://en.wikipedia.org/wiki/Folding@home>

여 분산 컴퓨팅 구조를 구성한다. Folding@home의 소프트웨어는 사용자 관점에서 작업 단위와 코어 및 클라이언트로 구성된다.

○ 작업 단위

작업 단위는 클라이언트가 처리해야 하는 단백질 데이터이다. 작업 단위는 마르코프 상태 모형 내 상태들 간의 시뮬레이션 중 한 부분이다. 자원 봉사자는 이 작업 단위를 다운로드 받아 단백질 폴딩을 실험 및 완료하고 Folding@home 서버에 다시 업로드하며, 이 과정이 완료되면 자원 봉사자가 크레디트를 얻는다. 이 순환 주기는 자동으로 반복된다. 모든 작업 단위는 작업 기한이 있으며 기한이 지난 경우 자원 봉사자는 크레디트를 얻을 수 없으며 자동으로 다른 참가자에게 재 할당된다. 이러한 합리적인 작업 기한이 있기 때문에 전반적인 시뮬레이션 과정이 정상적으로 수행될 수 있고, 최소 시스템 요구 사항이 펜티엄 3 450MHz CPU(스트리밍 SIMD 확장 유닛 포함) 정도로 낮다. 고성능 클라이언트에 대한 작업 단위는 단일 프로세서에 대한 기한보다 짧아, 신속하게 시뮬레이션을 완료할 수 있도록 한다.

○ 코어

“FahCores” 혹은 약칭으로 “코어(Cores)”라 불리는 특화 분자 역학 프로그램은 작업 단위의 연산을 수행하는 백그라운드 프로세스이다. 코어의 대부분은 분자 역학 소프트웨어 패키지 중 가장 빠르고 유명한 GROMACS에 기반하고 있는데, 이 GROMACS는 수동 최적화된 어셈블리 코드로 대부분 이루어져 있고 하드웨어 최적화가 포함되어 있다. GROMACS는 오픈소스이지만 데이터의 유효성을 보장하기 위해 폐쇄소스 라이선스(closed-source license)를 사용한다. 코어는 클라이언트와는 별개로 자동으로 업데이트 되는 프로그램이다. 코어는 주기적으로 연산 체크포인트를 생성하며, 중단된 경우 해당 체크포인트부터 작업을 다시 시작할 수 있다.

○ 클라이언트

Folding@home 자원 봉사자는 각각의 개인용 컴퓨터에 클라이언트 프로그램을 설치한다. 클라이언트는 백그라운드의 다른 소프트웨어 구성 요소를 관리하며,

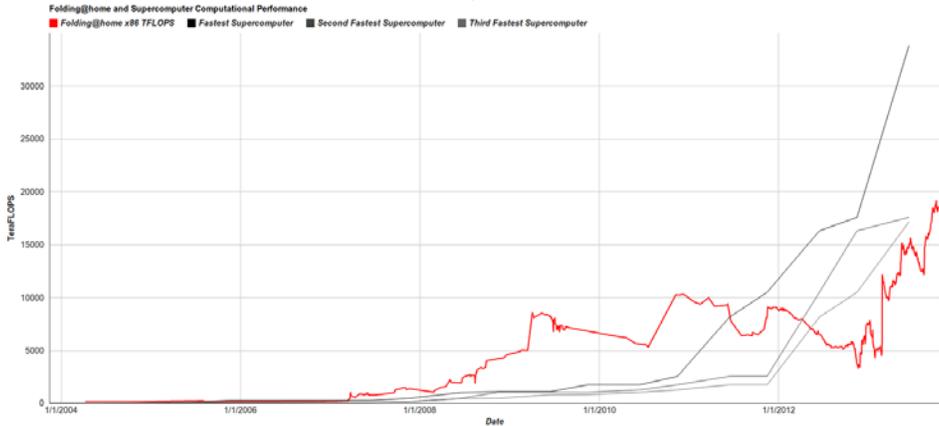
사용자는 클라이언트와 상호작용한다. 클라이언트를 통해 사용자는 폴딩 과정을 중지할 수 있고, 이벤트 로그를 열람할 수 있으며, 작업 진행 상황을 체크할 수 있고, 개인별 통계를 볼 수 있다. 이 클라이언트는 개인용 컴퓨터의 유휴 처리 리소스를 활용할 수 있도록 매우 낮은 우선순위로 백그라운드에서 계속 실행되어, 정상적인 컴퓨터 사용에 영향을 주지 않는다. 최대 CPU 사용량이 클라이언트를 통해 조정될 수 있고, Folding@home 서버에서 작업 단위를 받아오고 적절한 코어를 다운로드할 수 있다. 이 클라이언트는 단일 프로세서, 멀티 코어 프로세서 혹은 GPU에 맞춰 조정될 수 있다. 클라이언트 GUI는 오픈소스이지만 클라이언트 자체는 보안 및 과학적 무결성을 이유로 폐쇄소스 라이선스 하에 있다. 클라이언트와 Folding@home 데이터 서버 간 통신은 2048비트 디지털 서명을 통해 보호된다.

□ 활용 효과

2014년 7월 현재 총 179,237명이 참여하여 39,022 테라 플롭(FLOP, 1초에 38페타, 즉 3.8경희의 부동 소수점 연산)의 컴퓨팅 파워를 확보하였고, 이미 2007년 6월부터 2011년 6월 사이에 Folding@home(빨강)은 가장 빠른 500개의 슈퍼컴퓨터(검정)의 성능을 넘어선 바 있다. 동시간대 seti@home은 681 테라 플롭의 컴퓨팅 파워를 보여주고 있다.¹⁸⁾

18) Boinc Stats, <http://boincstats.com/en/stats/0/project/detail/overview>

[그림 2-13] Folding@home 성능



자료:

http://commons.wikimedia.org/wiki/File:Folding@home_and_Supercomputer_Computational_Performance.png

이러한 Folding@home을 활용하여 최근 홍콩과학기술대(HKUST, Hong Kong University of Science & Technology)에서 후이 후양의 그룹은 hIAPP(human islet amyloid polypeptide, amylin¹⁹)의 잘못된 폴딩을 조사하기 위해 대규모 분자 역학 시뮬레이션을 수행했다. 유형 II 당뇨병 환자의 95%가 hIAPP의 잘못된 폴딩을 보이고 있다. 후이 후양의 그룹은 hIAPP의 준 구조적 상태를 확인하고 그들 사이의 전환 역학을 파악함으로써 유형 II 당뇨병 치료에 한 걸음 더 다가갈 수 있었으며, 이와 같은 결과는 2013년 JACS(미국 화학회 저널)에 게재되었다.²⁰

19) 췌장의 β 세포에서 합성·분비하는 천연성 펩티드호르몬. 인슐린과 같이 당대사를 조절한다. 수용체의 해석이나 생리적 뜻의 해명이 현재 진행되고 있다
아밀린저해제는 비만인슐린 저항성을 수반하는 고혈압증이나 인슐린 비의존성 당뇨병 등의 치료제가 될 가능성이 있다.

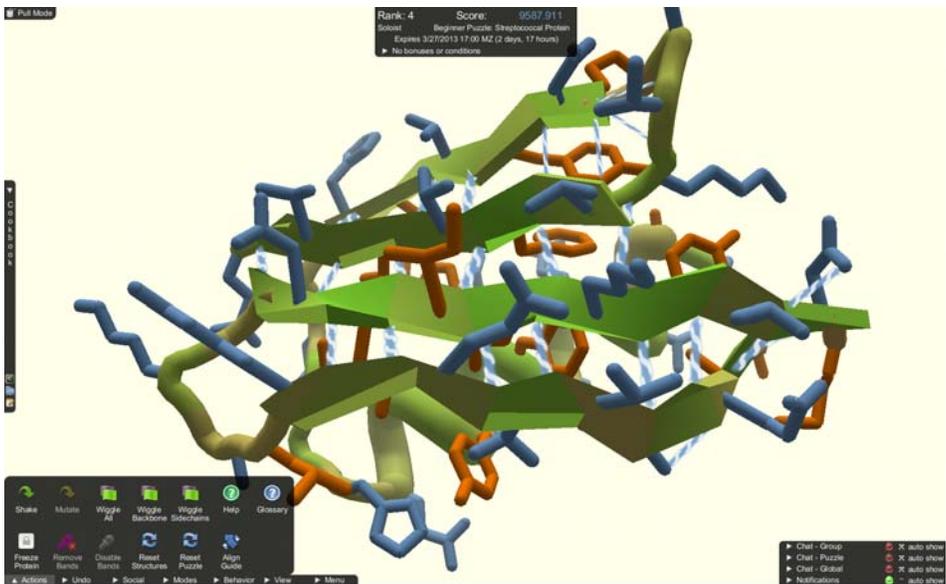
20) Qiao, et al. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation, 2013

3. Foldit

□ 개요

Foldit은 단백질 폴딩을 위한 온라인 퍼즐 비디오 게임이다. 이 게임은 실험적인 연구 프로젝트의 일부로서 워싱턴 대학 생화학 학과와 동 대학 게임 과학 센터가 공동으로 개발하고 있다. 이 게임의 목표는 게임 내에서 제공되는 다양한 도구를 활용해서, 상기 선택된 단백질의 구조를 가능한 많이 파악하여 폴딩하는 것이다. 최고의 점수를 받는 솔루션은 연구자들이 분석한 결과 현실 세계에서 관련 단백질에 적용할 수 있는 기본 구조 구성이 실제로 존재하는 경우이다. 과학자들은 이러한 솔루션을 다양한 대상 질병을 근절하고, 생물학적 혁신을 만들어내는 등 현실 세계의 문제들을 해결하는 데 사용할 수 있다.

[그림 2-14] Foldit의 단백질 폴딩 화면



자료: <http://fold.it/portal/info/about>

□ 구현 방법

FoldIt은 단백질 폴딩 구조를 빨리 찾아내기 위해서 크라우드소싱과 분산 컴퓨팅 방식을 활용한다. 특히 크라우드소싱과 커뮤니티 협업이 강조된다. 또한, 가상 상호작용(Virtual Interaction)과 게임화(Gamification)가 추가되어 독특하고 혁신적인 프로젝트 환경이 갖춰지게 되었다.

○ 가상 상호작용

FoldIt은 인간의 뇌가 갖고 있는 자연적인 3차원 입체 패턴 매칭과 공간 추론 능력을 활용하여 단백질 구조 예측 문제를 해결하려는 시도를 하고 있다. 현재의 퍼즐은 이미 구조에 대한 이해를 마친 단백질들을 기반으로 하고 있으며, 사람들이 직관적으로 이러한 퍼즐을 푸는 방식을 분석하여, 기존 단백질 폴딩 소프트웨어가 사용하는 알고리즘을 개선하고자 한다.

FoldIt은 간단한 단백질 구조에 대한 튜토리얼 시리즈를 제공하고 실제 단백질을 기반으로 한 퍼즐 집합을 정기적으로 갱신한다. 그래픽 인터페이스를 기반으로 다양한 도구들과 단백질 구조가 표시된다.

○ 게임화

FoldIt은 과학적 도구를 구축하는 데 있어 게임화 방식을 사용하여 일반 대중들에게 더 쉽게 다가갈 수 있도록 하였다. 그에 따라 과학적 교육 기반이 없는 사람도 사용할 수 있었고, 따라서 더 많은 대중을 참여시킬 수 있었다. 각 퍼즐의 구조가 바뀔 때마다 단백질이 얼마나 잘 폴딩되었는지를 분석하여 점수가 계산되고 가장 높은 점수의 리스트가 함께 제공된다. 사용자들은 그룹을 만들거나 참여할 수 있으며 서로 솔루션을 공유할 수 있다. 그룹의 최고 점수 리스트도 따로 관리된다.

□ 활용 효과

2011년 FoldIt 사용자가 에이즈를 일으키는 원숭이 바이러스의 일종인 메이슨-화이자 원숭이 바이러스(M-PMV)의 레트로 바이러스 단백질 분해 효소의 결정 구조를 해독하는데 도움을 주었다. 퍼즐을 푸는데 주어진 시간은 3주였지만 사용자들은 단 10일 만

에 정확한 입체 3D 모형을 완성하였다. 해당 효소의 구조를 구성하는 방식에 대한 연구 문제는 지난 15년 동안 과학자들이 풀지 못했던 난제였다.²¹⁾

2012년 1월에 미국 과학 잡지 Scientific American은 FoldIt 사용자들이 첫 번째로 크라운드소싱을 통해 단백질 구조를 재설계했다고 보도했다. 해당 단백질은 합성 화학에서 널리 사용되는 딜스-알더 반응을 촉진하는 효소이다. 시애틀에 있는 워싱턴 대학의 게임 과학 센터에서 데이비드 베이커 등의 팀이 처음부터 이 효소를 설계하였으나 개선이 필요했고, 사용자들이 13개의 아미노산을 추가하여 효소를 재설계하였고 그에 따라 활동량이 18배나 늘어났다.²²⁾

4. EteRNA²³⁾

□ 개요

EteRNA는 미국 국립과학재단이 지원하여 카네기멜론 대학교와 스탠포드 대학교에서 개발한 퍼즐 게임으로, 사용자가 RNA 분자 구조 퍼즐을 풀게 하여 RNA의 분자 구조를 알아내는 것이 목적이다.

RNA는 오탄당의 일종인 리보스를 기반으로 뉴클레오타이드를 이루는 핵산의 종류로 DNA의 일부가 전사되어 만들어 진다. RNA는 단백질을 합성하는 과정에 작용하며, 아데닌(A), 구아닌(G), 우라실(U), 시토신(C)의 핵염기가 연결되어 있는 형태이다. RNA의 종류에는 리보솜을 구성하는 rRNA, DNA의 유전정보를 옮기는 mRNA, 단백질의 기본 구성단위인 아미노산을 운반하는 tRNA 등이 있다.

RNA는 DNA의 이중나선 구조처럼 폴딩되며, 아데닌이 우라실에, 시노신이 구아닌과 만나 결합하는 경향이 있다. 따라서 RNA가 합성될 때 RNA의 모양은 일직선인 경우가 없으며, 3차원 모양을 가지게 된다.

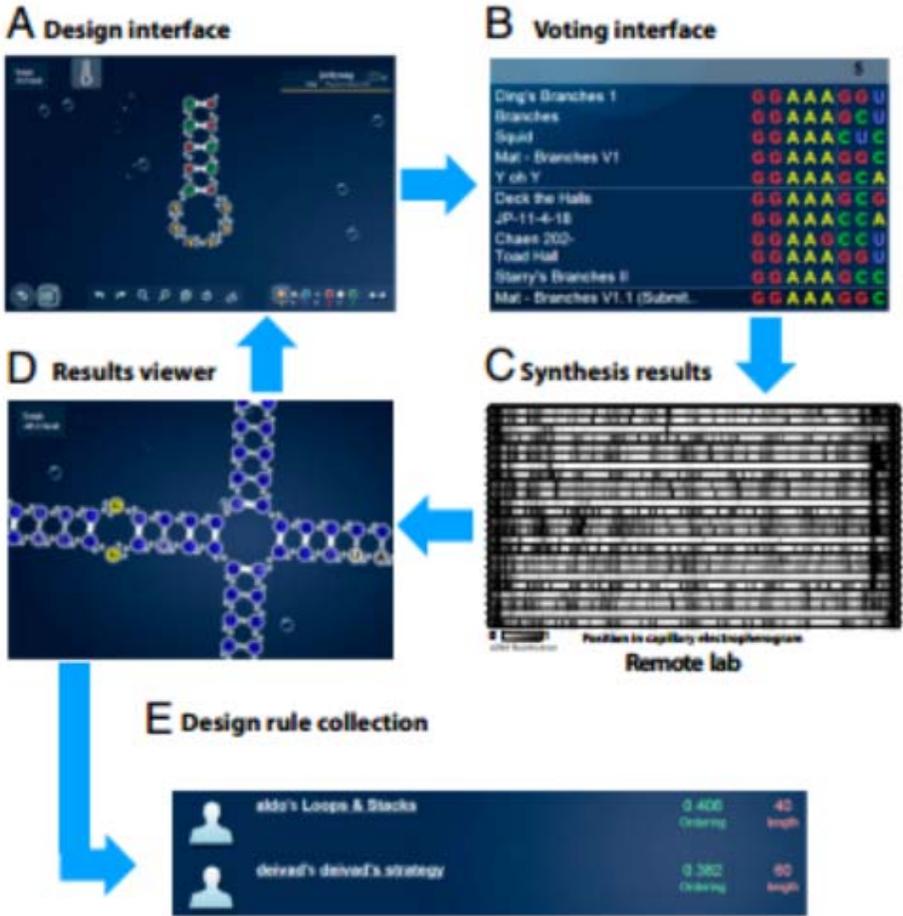
과학자들은 아직 RNA의 모든 역할을 밝혀내지 못했으며, rRNA, mRAN, tRNA 이외에도 shRNA, piRNA, snRMA 다양한 RNA가 밝혀지고 있다.

21) http://www.nature.com/nsmb/journal/v18/n10/full/nsmb.2119.html?WT.ec_id=NSMB-201110

22) <http://www.scientificamerican.com/article/victory-for-crowdsourced-biomolecule2/>

23) <http://eterna.cmu.edu/web/>

[그림 2-15] EteRNA 워크플로우



자료: Lee, et al. 2014

EteRNA는 사용자에게 목표하는 모양으로 퍼즐을 풀 것을 제시한다. 사용자에게 제시하는 RNA모형들은 이미 다른 소프트웨어 기술로 형태가 입증된 모양이다. 이는 사용자가 해당 모형의 구조에 대해 이해하고 더 나은 구조를 찾는 데 도움을 준다. 이 과정을 마치면 실제 RNA 설계 문제가 주어지게 되며, RNA가 제대로 폴딩되는지 검증하여, 계산 모델과 실제 모델의 차이를 없애줄 수 있을 것이라 기대한다.

□ 구현 방법

EteRNA의 워크플로우는 [그림 2-15]와 같다. 사용자가 시퀀스 디자인 인터페이스를 통해 목표된 RNA 구조를 만들면 사용자간의 투표를 통해 8개의 시퀀스가 실제로 합성되고 가능성이 입증된다. 실험의 결과는 온라인에 게시되며 사용자들은 이를 바탕으로 새로운 가설을 만들고 다시 디자인 인터페이스를 통해 폴딩될 수 있는 RNA 구조를 만드는 것을 반복한다.

□ 활용 효과

2011년 8월에는 26,000명의 사용자가 이 게임에 참여하여 306개의 RNA 디자인을 만들어 냈고 실제로(in vitro) RNA를 만들었다.²⁴⁾ 2013년에는 총 37,000명이 RNA디자인을 만들어 냈고, 이렇게 만들어진 RNA는 연구실에서 새로운 RNA모형을 제작하는데 큰 도움을 주었다. 또한 머신 러닝을 통해 새로운 RNA 디자인 알고리즘을 만들어 낼 수 있었고 이는 이전의 알고리즘 보다 더욱 좋은 성능을 내었다. 2014년 1월에는 EteRNA의 결과가 미국국립과학원회보(PNAS: Proceedings of the National Academy of Sciences of the United States of America) 저널에 게재되었으며, 저자에 EteRNA 사용자가 포함되었다.²⁵⁾

24) <http://en.wikipedia.org/wiki/EteRNA>

25) Lee, et al. 2013

제3장 국내 연구 현황

제1절 국내 연구 사례

국내의 연구 현황을 조사하기 위해 자문위원회를 구성하여 연구사례와 소프트웨어 활용 현황을 조사하였다. 과학기술분야와 컴퓨터과학분야를 고르게 배분하여 선정하였으며 산학연을 모두 참여시켜 구성하였다. 또한 조사 범위를 넓히기 위해 미래창조과학부에서 지원하는 과제 중 소프트웨어와 협업 성격이 있는 과제를 수행하는 기관과 고등과학원 등을 방문하여 전문가 인터뷰를 진행하였다. 과제 선정은 미래부의 과학기술 과제를 관리하는 연구재단을 통하여 소개받은 24개 과제 중 인터뷰를 수락한 10개 과제에 참여하는 연구원을 대상으로 하였다.

1. 자문회의를 통한 사례 조사

자문위원회는 총 10회의 세미나와 회의를 개최하였다. 자문위원이 서로 다른 분야의 전문가이고 연구 주제가 여러 전문 분야를 피상적으로라도 이해해야 하는 것이므로 소프트웨어 활용에 관한 분야별 사례를 조사하는 것이 필요하다. 정기 자문회의에서는 자문위원이 직접 수행한 연구사례를 발표하고 토의하여 분야별로 소프트웨어의 활용 정도와 중요성, 국내 연구 환경의 취약점, 개선 방안에 대해서 의견을 나눴다.

<표 3-1> 자문회의 연구사례 요약

차수	분야	주제	요약
1		과제소개	과제소개, 세미나 일정 결정, 자문위원 소개 및 의견 토의
2	CS	위성정보기반 기후변화 연구에서 빅데이터 고성능화를 통한 성능	파일럿 과제로서 세 연구소가 공동으로 대용량 위성정보 데이터 공유 플랫폼을 개발. 34시간 30분 걸리던 처리 시간이

차수	분야	주제	요약
		개선 사례	2시간 53분으로 약 1,200% 정도의 성능 향상. 파일단위로 관리하던 40TB의 전처리 데이터를 다차원배열의 클러스터링으로 구축하고 Hadoop으로 분산 처리.
	물리	물리학 연구에서 소프트웨어 활용 사례와 영국 Edinburgh 대학의 기초과학과 CS 협동 연구 지원 환경	해외에서 했던 연성물질 연구의 경우 통계, 대용량 데이터, 시뮬레이션, 병렬처리 기술이 이론 연구와 밀접하게 연관되어 있으며 연구자가 직접 컴퓨팅 환경을 활용하여 프로그래밍을 해서 연구를 진행해야 함 영국 Edinburgh 대학의 LUDWIG 시스템은 이런 연구를 수행하기 위해 구축된 시스템이며, 과학연구를 위한 대규모 소프트웨어 개발 관리를 지원하기 위한 CCP(Collaborative Computational Projects, http://ccpforge.cse.rl.ac.uk)에 속해 있음 Edinburgh 대학의 Soft Matter Physics Group은 이론물리학자, 실험물리학자, 컴퓨팅전문가 등이 속해 있으며 그룹 내 협업과 연구결과에 대한 공유가 자유로워 연구 효율성을 극대화함
3	생물	차세대 시퀀싱 기법을 이용한 질병 기작 연구에서의 소프트웨어 활용 사례	차세대 염기서열분석(Next-generation sequencing, NGS)은 유전자 분석 가격을 낮춰서 의료서비스의 패러다임을 바꾸는 파괴적 기술임. 테라바이트급의 많은 데이터가 생성되기 때문에 컴퓨터 활용이 필수. 분석용 소프트웨어로는 공개 소프트웨어가 많이 있음. 계산, 통계, 스크립팅 등 여러 프로그램을 조합하여 연구. 연구시간의 70%정도는 컴퓨터를 사용하는 연구.
4	의공학	연구플랫폼으로서의 의료영상시스템	의료영상에 이미지 처리 소프트웨어 기술을 접목하여 진단 정확성을 높일 수 있는 여러 기술 소개. 수술실에서 외부 정보와 외부 전문가 의견을 반영하여 실시간으로 판단할 수 있는 Surgical Imaging Informatics 소개. 의료용 계측장비의 결과를 활용하여 새로운 기능의 소프트웨어를 개발하는 것이 투자

차수	분야	주제	요약
			<p>대비 연구 효율이 높고 잠재력도 큼.</p> <p>1세대 2D PACS, 2세대 3D PACS를 넘어서는 3세대 Intelligent Enterprise Image Management System 개발 사례. 의료 영상 소프트웨어, 병원관리 시스템에 통합, 영상공유, 분야별로 특화된 시각화, 클라우드/모바일/SaaS 등 최신 기술 적용. 이런 과제는 의료와 기술을 함께 공부한 소수의 전문가 그룹에서 기획하고 분야별 전문가를 참여시켜 만들어지는 것이 일반적인 사례. 의료분야에서도 이런 첨단 연구 과제를 수행하려는 동기는 충분히 있음.</p>
5	의공학	지능형 영상진단 및 치료지원 시스템	
5	CS	국가 연구개발성과와 이공계 교육의 연계를 위한 첨단 사이언스 교육 허브	<p>국가 사이버인프라(수퍼컴퓨터, 가시화) 기반의 연구결과를 고등 교육에 활용하는 방안 연구. 미국 nanoHUB 사업을 벤치마킹. 기존 연구 결과가 있으면서 교육이 가능한 후보 분야 5개를 선정 (열유체공학, 구조/동역학, 계산화학, 계산물리, 공학설계)하여 시뮬레이션 소프트웨어를 교육에 활용. 연구와 교육 연계를 위한 플랫폼, 포털, 인프라, 커뮤니티 등의 환경 구축.</p> <p>2011년부터 시작하여 현재 진행 중 현재 시뮬레이션 SW 158종, 콘텐츠 223종 서비스 중</p>
6	생물정보	분자 바이오 빅데이터	<p>분자 바이오 빅데이터 연구 분야는 세포 전체의 생물학적 사건에 대한 모든 데이터를 측정하여 분석하기 때문에 데이터가 많아서 컴퓨터기술이 연구에 필수. 가령 암 연구라면 DNA, 유전체, 단백질, RNA 등 개별 요소뿐 아니라 요소간 연결 구조와 그 안의 생물학적 메커니즘 등을 종합적으로 고려한 연구가 필요함. 따라서 다뤄야하는 데이터가 폭증.</p> <p>바이오 빅데이터의 특징은 수집에 비용이 많이 들고, 같은 개체라도 환경, 건강상태에 따라 다른 결과가 나오기 때문에 수집해야 할 양이 많고, 분석 결과도 빅데이터가 됨.</p>

차수	분야	주제	요약
7	생물정보	천연물의약품 개발을 위한 시스템 바이오인포매틱스	<p>또한 데이터는 비싸지만 분석능력자는 부족. 대용량 컴퓨터, 스토리지, 데이터전송관리, 초고속 데이터마이닝, 데이터베이스, 시각화, 연구 협업 시스템 등이 연구의 필수 요소임.</p> <p>천연물은 복합약물임. 관심 증가 중. 표준화 문제 - 화학물질의 조성이 상황에 따라 달라지기 때문에 정확한 표준을 만드는 것이 어려움.</p> <p>복합 약리 분석 - 두 개 이상의 물질이어서 분석이 어려움. 각 요소마다 투여량을 구간으로 나누고, 모든 요소의 조합에 대해서 모든 투여량의 조합을 실험해야 함. combinatorial explosion 문제이며, 현실적으로 실험 자체가 불가능하여 컴퓨터 모델링 및 시뮬레이션을 통한 수밖에 없음</p>
	생물	바이오 빅데이터	<p>차세대시퀀싱 기술에 의해 염기서열을 값싸고 빠르게 추출할 수 있게 되었으나 반대로 처리할 데이터는 급증하게 되었음. (보통 한 사람당 1.5TB). 이 빅데이터를 분석하는 방법을 찾는 것이 생물정보학의 과제임. 주로 스트링 매칭, 조립, 검색, 시각화, 데이터 해석 및 추론, 불안전 매칭, 데이터 관리, 네트워킹, 압축 및 인덱싱 등 컴퓨터과학에서 다뤄온 문제. 많은 컴퓨팅 자원이 필요하고 생물학과 공동 연구 필요. 우리나라의 생물정보학 논문 비율은 낮은 편임. 국내에서는 연구를 통해 개발된 알고리즘은 연구실 내에서만 사용. 공유를 위한 소프트웨어 전문화에 노력을 들일 인프라와 리소스 부족. 전 직장이던 미국 연구소에서는 전문 소프트웨어 개발 담당자가 있어서 이 부분을 처리했음.</p>
	CS	정부의 공개소프트웨어 활성화 방안	<p>정부의 공개소프트웨어 활성화 정책 소개 - 소프트웨어의 기반 기술적 성격을 고려 기반성이 있는 것은 공개하는 방향으로 추진.</p> <p>- 특정 운영체제에 편중되지 않는 컴퓨팅 환경 조성, 국가 공개SW R&D 확대, 고급 공개SW 개발자 양성, 정부 및 공공부문의</p>

차수	분야	주제	요약
			공개SW 도입 확대 등을 통한 민관 협력 생태계 조성 - 국가 공개SW R&D는 개발 초기부터 공개SW 개발방식을 따르는 것을 의미 - 기존 결과물 중에서 공개SW로 전환하는 것도 추진 중. - 연구관리 규정은 변경되었음. 연구관리 지침은 NIPA에서 작성 중.
8	CS	중간 진도 발표	해외 사례 조사 국내 현황 조사 중간 결과 향후 추진 방향
9	CS	정부의 빅데이터 공동 활용 종합 계획 사례 및 정책 토의	정부에서 추진하는 과학연구 데이터 공동 활용 정책 토의 현황 조사를 통해 도출된 문제점 토의
10	CS	제안 정책 초안 토의	과학기술연구의 소프트웨어 산출물 공개를 통한 활성화 방안 토의

자문회의를 통해 도출된 문제를 정리하면 다음과 같다. 편의상 과학기술연구자를 도메인 전문가라고 표현하여 소프트웨어 전문가와 구분하였다.

□ 도메인 전문가와 소프트웨어 전문가 협업의 어려움

두 분야를 모두 이해하는 연구 인력이 부족하다. 이런 연구자는 단순히 대학교 융합 학과의 졸업생으로 채워질 수 있는 인력이 아니다. 한 분야에서 전문성을 확보한 인력이 다른 분야에 뛰어 들어 일정 수준의 전문성을 키운 수준을 의미한다. 따라서 1년 이상의 공동 연구 기간이 필요하다. 바이오 분야의 예를 들면 바이오 분석에서 나오는 빅데이터를 다뤄야 할 그룹은 극히 소수이므로 국가적인 센터정도만 갖추고 연구 결과를 공유하면 해결될 수 있으나 더 어려운 문제는 알고리즘을 함께 연구할 사람이다. 생물 학자와 함께 현상을 분석하고 해결책을 찾아낼 컴퓨터 전문가가 없다. 생물학 분야에서는 컴퓨터과학과 통계전문가가 필요한데 바이오는 인건비 등이 적어서 직접 채용하여 인력을 키워내기도 어렵다. 한마디로 아이디어나 문제는 많으나 사람이 없다.

한편으로는 공동연구에서 소프트웨어 연구자의 기여를 인정받기 어렵다는 주장도 있

다. 대부분의 연구를 학교나 연구소처럼 논문 중심의 평가 제도를 운영하는 조직에 소속된 연구자가 수행하기 때문에 논문 성과가 중요한 요건이다. 그러나 과학기술분야와의 공동연구에서는 공동으로 논문 작성 시 교신저자 배정 문제 등 연구 평가 제도와와의 상충점이 있다. 도메인 전문가와 소프트웨어 전문가가 공동 연구를 하는 경우에 해당 도메인에서는 소프트웨어를 적용하는 것이 논문으로서의 가치가 있으나 컴퓨터과학분야에서는 중요한 업적이 아닌 경우가 많다. 가령 과학 분야의 빅데이터를 시각화하여 웹에서 표현하는 공동과제를 수행하는 경우에 컴퓨터과학에서 이미 검증된 방법을 특정 도메인에 적용하여 과제를 진행하는 것이 일반적이기 때문에 컴퓨터과학 학술지에서는 가치를 인정받기 어렵다. 반면 과학 분야에서는 관심을 끄는 시도가 될 수 있다.

앞의 시각화 사례에서 주목할 점은 주도메인 전문가와 소프트웨어 전문가간의 교류가 깊은 수준으로 내려가지 않아도 된다는 점이다. 도메인 전문가가 도메인 지식을 자세히 설명하지 않고 요구사항 위주의 사양을 제공하고 소프트웨어 전문가도 깊은 원리를 이해할 필요 없이 주어진 요구사항을 만족하는 소프트웨어를 개발하는 것을 목표로 하게 된다. 즉 공동 연구를 하는 연구자 사이에 서로의 분야에 대해서 깊은 이해를 통해 이전에 해결하지 못 했던 문제를 학제간 협력을 통해 해결하는 단계까지 진행하지 않게 된다. 과학계 단순화한 사례이지만 실제 많은 융합 과제가 이런 방식으로 진행되고 있음을 융합 과제를 진행하는 연구자와의 전문가 인터뷰에서도 발견할 수 있었다. 공동 연구의 초창기에는 소프트웨어 전문가가 도메인 전문가에게 제공하는 가치가 연구를 도와주는 도구의 형태인 경우가 많기 때문에 이런 현상이 계속 발생하는 것으로 보인다.

□ 국내 연구자는 연구 결과에 대한 공유의식이 낮음

국내 연구자는 연구 결과로 생산된 소프트웨어나 실험데이터를 공개하는 비율이 낮다. 특히 해외에서 귀국한 연구원은 문화의 차이를 실감하고 있다. 과학연구 분야에서는 연구를 위해 개발한 소프트웨어나 데이터의 공개가 일반적이고 연구기관에는 이것을 지원하기 위한 소프트웨어 전문가 조직이 있는 경우가 많다. 소프트웨어와 데이터를 공개하면 동료검토를 통한 연구내용의 검증이나 동료연구자의 데이터 재활용을 통한 국가적 차원의 연구 효율성을 높일 수 있기 때문에 해외에서는 공개 문화가 기본으로 자리 잡

고 있는데 반해 국내에서는 아직 소극적인 수준이다.

전문가의 의견은 이런 상황을 타개하기 위해서 소프트웨어와 데이터를 공개하는 연구자에 대한 동기부여가 필요하다는 것이다. 해외에서는 공개소프트웨어로 등록하고 명성을 얻어 학문적 성취까지 일구는 사례가 종종 있으나 국내 연구자는 공개를 하면 손해라는 인식이 있는 것으로 추정된다. 공개를 하면 손해라는 인식 대신 공개를 안 하면 의심하는 풍토가 국가적으로 바람직한 방향이다.

공유문화는 쉽사리 만들어지지 않는다. 한 공동연구 사례에서는 소프트웨어 전문가가 도메인 연구자와 의사소통이 편해지기까지 6개월에서 1년 정도의 시간이 필요했다. 이것은 소프트웨어전문가와 도메인전문가 양쪽에 상당한 부담이 되는 일이므로 이 정책과제에서 참여자에 대한 적절한 보상 방안도 함께 고려되어야 한다.

소프트웨어 공개를 꺼리는 또 다른 이유는 연구개발에서 개발된 소프트웨어의 품질이 떨어지기 때문이기도 하다. 과학연구에서 소프트웨어는 연구를 위한 도구이기 때문에 소프트웨어의 품질을 높이기 위한 노력을 많이 하고 있지 않다. 소프트웨어의 공개가 일상화된다면 다른 사람이 보고 사용하게 될 것이므로 품질을 높이려는 노력을 할 것이라고 예상할 수 있다. 또한 공개된 소스를 소프트웨어 전문가가 개선할 기회도 줄 수 있다.

□ 과학연구 데이터 공유 방안

소프트웨어와는 별도로 과학연구에서 생성된 실험 데이터의 공유에 대한 필요성도 지적되었다. 이 부분은 국내에서 선행 연구가 이미 진행되었고 현재 미래창조과학부에서 정책으로 추진 중이다.(한국과학기술정보연구원, 2012. 미래창조과학부, 2013)

□ 연구성과지표의 개선이 필요

국가 연구개발과제의 성과평가는 3P(Paper, Patent, Product)가 핵심 요소이다. 연구의 부산물인 소프트웨어와 데이터의 공유를 통한 기여는 과제와 연구자의 평가에 영향을 주지 않으므로 연구자로서는 소극적인 태도를 취하게 된다. 전반적으로 국가연구평가 기준이 기존의 양적 평가에서 질적 평가로 바뀌고 있어서 논문수보다 인용횟수를 중시

하고 자체인용을 제외하는 등 개선하고 있는 추세이다. 공개된 소프트웨어와 데이터에 대한 활용도를 평가하여 연구기관과 연구원의 평가에 활용하면 적극적으로 참여하게 될 것이다.

□ 무조건적인 공개의 부작용에 대한 방안

국익적인 차원에서 연구에서 개발된 소프트웨어나 연구데이터를 전 세계에 완전공개를 하는 것이 문제가 있는가도 토의가 되었으나 적어도 과학연구에서는 그런 부분이 없다는 것이 일반적이다. 해외에서는 소프트웨어와 연구데이터 공개가 연구문화로 정착되어 있고 오히려 공개가 되지 않은 것은 신뢰성이 떨어진다고 평가하고 있다. 연구결과와 공유는 오히려 국내 연구자의 교류를 증진시켜 국내 연구 수준을 높이는데 기여할 것이다.

공학 분야의 경우에는 문제의 소지가 있을 수도 있다. 이를 보완하기 위해 공개할 소프트웨어와 데이터를 제출은 하되 유예기간을 두는 것은 고려할 필요가 있다. 특허의 경우 특허 출원 후 약 1~1.5년이 경과한 후 일반에게 공개하여 이의신청을 받는 것을 감안하여 이와 유사한 정도의 유예기간을 두는 것은 문제없을 것으로 보인다.

과제의 목적이 소프트웨어이고 과제종료 후에 상품화를 하려고 하는 경우에도 공개에 대한 유예기간이 있으면 이 소프트웨어를 개발한 연구팀이나 기술이전을 받은 업체에 대한 우선권을 충분히 배려한 것이다. 현대 소프트웨어 산업에서 1년여의 우선권을 가지고도 충분한 사업화 결과를 내지 못했다면 앞으로도 이 조직에서는 사업화 가능성이 거의 없다고 봐야 하며 오히려 공개를 통해 다른 사람에게 새로운 활용 기회를 주는 것이 효율적이다.

공개한다고 모든 정보를 공개할 필요는 없다. 핵심 기술은 숨긴 채 공개할 수 있다. 바이오정보학에서 사용되는 공개소프트웨어는 많이 있으며 각각 소프트웨어에 규정된 라이선스에 따라 사용하고 GPL(GNU Public Licence) 라이선스 코드처럼 수정한 내용을 공개해야 하는 경우에는 소스를 릴리즈하고 있다. 해외에서도 공개소프트웨어로 등록은 하지만 상용화 가능성이 있는 부분은 숨기는 편이다. 공개하는 연구자가 알아서 할 문제이다.

2. 전문가 인터뷰를 통한 사례 조사

전문가 인터뷰는 연구기관을 직접 방문하여 과제에 참여하는 연구원을 대상으로 실시하였다. 과학기술 연구에서 소프트웨어 전문가와 협업의 필요성을 이해하고 있는 자문 위원회의 내용과는 달리 과제별로 의견차이가 많았다.

<표 3-2> 전문가 인터뷰 요약

차수	분야	인원	연구 분야	요약
1	CS	2	클라우드, 디지털콘텐츠	의사소통의 문제 어려움. 연구 결과를 논문화하기 어려움.
2	CS	1	그래픽스, UI/UX	공동 작업의 의사소통 이슈가 큼. 도메인 전문가의 요구사항이 너무 단순. 논문거리가 안 됨. 서로가 원하는 범위가 너무 다름. 명백한 문제는 쉽게 찾지만 그 이상으로 가기 어려움. 커뮤니티가 도움이 됨.
3	물리	3	천체물리, 바이오	직접 코딩함. 연구실내 클러스터도 직접 구축. 국내에 사례가 별로 없어서 업체에서 해결 안 됨. 데이터는 테라바이트 규모를 다룸. 연구 결과는 웹서비스 형태로 공개. 공동 인력 부족이 문제임. 연구소 자체적으로 소프트웨어 개발자를 고용하여 테크니션으로 활용. 계산과학을 인정해 주는 풍토 필요. 협업을 깊이 들어가면 연구주체가 많음. 예를 들어 바이오정보학은 해외에서는 컴퓨터과학자가 주류가 되었음. 국가에서 지원하는 공동연구센터 고려.
4	CS	4	정보 보안	과제가 막 시작하여 아직 문제는 없음. 같은 분야에서도 협업은 어려움. 각자 깊이 들어가 있기 때문에 상대를 이해하기 어렵고 주종관계의 업무가 됨. 현 평가제도는 협업을 통해 좋은 논문을 한 편 쓰는 것보다 중간급 논문 여러 편을 쓰는 것이 유리.

차수	분야	인원	연구 분야	요약
5	CS	2	소프트웨어 품질 향상, 지능형 교통 시스템	인접한 실험실과의 협업이고 오픈랩 형식으로 협업과제가 수행되어 비교적 수월했음. 서로 업무가 잘 구분한 후 진행했음. 새로운 아이디어를 제시하는 환경은 아니었음.
6	우주항공	3	우주항공, 무인항공기	소프트웨어 활용 비중은 50% 이상. 해외의 공개된 데이터를 이용하고 있음. 연구실에 소프트웨어 잘 하는 연구원이 도움을 주고 있음. 문제는 개발 커뮤니티 등을 활용하여 자체 해결함. 과제가 더 진행되면 공동연구기관과의 의사소통 문제가 예상됨.
7	기계	4	우주항공, 정밀 자세제어, 추력기 설계	주로 컴퓨터로 연구함. 장치의 설계와 시뮬레이션, 결과 분석 등. 직접 코딩이 필요 없는 경우가 많고 자체 개발은 연구에 특화된 경우에 진행. 아직 소프트웨어를 적극적으로 활용하는 단계가 아닌 연구실.
8	우주항공	4	우주항공, 인공위성	소프트웨어 인력을 채용하여 육성 중. 핵심 기술이므로 외주나 공동연구로 할 수 없음. 소프트웨어 전문가가 들어와서 4.5년 정도 지나면 익숙해 짐. 범용성, 공공성이 있는 소프트웨어는 공개하지만 특수한 목적의 소프트웨어는 공개하지 않음. 데이터는 공개에 문제없음.
9	전자	2	IC 설계, 내방사선	컴퓨터를 시뮬레이션에 주로 사용하고 상용툴 이용 중. customization을 위해 코딩 필요. 연구원이 정보올림피아드에 나갈 정도로 소프트웨어 경험이 많은 사례. 소프트웨어 문제는 자체 해결. 커뮤니티 등을 통해 정보 교류. 데이터 공유는 불확실.
10	의학	2	바이오, 가정 의학	기술은 전문 업체에서 담당. 동일 분야의 경험이 있는 업체임. 연구 내용이 의학분야에서 이미 잘 정의되어 있고 이것을 소프트웨어 업체에서 개발하여 시험적용해서 실증하는 과제. 의공학 전문가가 두 분야의 중재자 역할을 함. 의사소통의 문제가 어려움.

차수	분야	인원	연구 분야	요약
				문서로 이해가 안 되면 회의로 합의하는 방식으로 진행하고 있음. 데이터 공개는 개인정보를 제외하면 가능하고 소프트웨어는 업체와 조율 필요.
11	의학	4	바이오, 가정의학	10번 사례와 유사한 연구. 연구책임자가 도메인 전문가인 상태에서 소프트웨어 플랫폼을 외부 기술업체가 추진하여 의사소통의 문제, 용어의 혼선 등이 발생. 산출물의 공유에는 찬성이고 추가로 들어가는 노력에 대한 고려 필요.

□ 연구 결과에 대한 시각차이

소프트웨어 전문가는 도메인전문가와와의 협업을 통해 연구 업적인 논문의 작성이 어렵다는 의견이 다수였으나, 소수의견으로 연구자 개인 성향의 문제로 보는 견해, 협업 성격의 연구에서는 논문을 기대하지 않는다는 견해 등도 있었다. 이 문제에서는 인터뷰 대상자의 소속기관의 성격, 연구자의 입장에 따른 차이가 가장 크게 나타났다. 특히 도메인 전문가는 이 문제를 전혀 제기하지 않았고 모두 소프트웨어 전문가로부터 제기되었다. 앞서 자문회의에서도 지적되었듯이 국내 공동 연구의 수준이 소프트웨어 기술을 이용해서 도메인에서 풀 수 없는 문제를 푸는 것보다 도메인 연구자가 제시하는 요구사항을 개발하는 수준이기 때문에 발생하는 문제로 해석된다.

학교에 있는 경우에 논문 압박이 강하기 때문에 더욱 불만이 많다. 테뉴어 전에는 장기 연구를 하기 어렵다. 현 제도에서는 중간급 논문을 많이 쓰는 것이 좋은 논문 한 편 쓰는 것 보다 낫다. 정부의 성과평가 제도가 인용횟수를 고려하는 등 질적인 평가를 도입하는 방향으로 개선되고 있다고 하지만 현실은 여전히 SCI 논문 갯수이며 인용횟수는 참고사항일 뿐이다.

협업을 하면 성과가 나오기까지의 기간이 훨씬 길고 오히려 논문을 쓰기 어렵다는 것이 장벽이다. 각자가 깊이 있는 연구를 하고 있기 때문에 협력이 어렵다는 주장도 있었다. 연구소 연구원은 이런 문제에 대해서는 관대한 편이며 이것도 논문에 대한 압박이

덜하기 때문으로 보인다. 이런 상황은 다음 장에서 제시되는 해외의 학제간 협업의 문제에 대한 연구 결과와도 일치한다.(Qin, Lancaster, Allen, 1997; Carayol & Thi, 2004; van Rijnsoever & Hessels, 2011)

□ 소통 문제

공통적으로 소통 문제가 제기되었다. 이것은 서로 다른 전문 분야가 함께 일을 하는 것이므로 당연한 결과이다. 소통의 제한으로 서로 명백한 문제까지는 협업이 진행되지 만 깊이 들어가서 공동으로 찾아야하는 문제를 발견하기 어렵다.

중재자 역할을 할 수 있는 인력이 매우 부족하다. 이런 인력의 확보는 연구팀의 장기적인 투자, 우연 등에 의지하고 있었다. 이런 인력이 없는 경우에는 연구가 도메인 내의 문제를 대상으로 하고 있다. 모든 인터뷰 대상자는 이런 인력을 확보하고 싶어 한다. 그러나 대학의 경우에는 전담 연구원 채용이 부담되어 학생을 활용하게 되므로 소프트웨어에 경험이 많은 학생이 들어오면 연구가 진행되고 졸업하면 수준이 낮아지는 등 지속적인 경험 축적이 어렵다.

소통 문제를 완화하는데 커뮤니티가 효과적인 방법일 것이라는 의견이 많았다. 현재 국내 커뮤니티를 활용하는 연구자의 경우에 커뮤니티의 용도는 당면한 기술적인 문제에 대한 직접적인 도움을 기대하는 경우가 많기 때문에 자생적으로 활성화되기 어려울 것으로 보인다. 국내 커뮤니티라는 것도 소프트웨어 업체가 자신의 제품을 지원하기 위해 구성하고 제품 지원의 일환으로 답변을 주는 형식이 많았다. 소프트웨어 개발에서 stackoverflow.com 의 역할을 할 수 있는 정도로 규모가 크고 사용이 편리한 커뮤니티가 되면 효과가 있을 것으로 보인다.

커뮤니티와 더불어 넓은 분야에 걸쳐서 공유와 토론의 장을 만드는 것이 도움이 될 것이다. 전문가들은 학회 참석과 비슷한 이유로 이런 행사에 참석하여 아이디어를 발굴하는 기회로 활용할 수 있다.

□ 인력 양성

물리나 생물학 연구의 경우에 커다란 계산능력과 데이터 처리 능력이 필요하기 때문에 업체의 지원에 기대기 어려운 경우가 많고, 결국 직접 개발하여 사용하고 있는 상황이다. 이런 연구자사이에서는 도메인 지식을 어느 정도 갖춘 소프트웨어 전문가로 구성된 전문 계산 조직에 대한 필요성이 아주 높다.

인력 양성을 하려면 과학연구에서 기술지원하는 조직을 연구조직으로 인정하는 문화가 필요하다. 국내에서는 이것을 서비스나 행정 조직으로 취급하기 때문에 연구소를 만드기가 어렵지만 해외에서는 그렇지 않다는 것이 도메인 연구자의 의견이다. 기존의 이론, 실험 분야에 덧붙여 계산 분야를 인정해 주어야 한다는 의견도 있었다. 전통적 과학 연구의 결과는 논문이더라도 계산 쪽은 학술대회 발표, 최적화 결과 공개, 소스 코드 등을 업적으로 인정해주는 제도적 뒷받침이 필요하다. 또한 이런 공동 연구를 하는 연구 센터는, 적어도 초기에는, 개별적으로 운영하기 어려우니 국가가 운영하는 것이 좋을 것이다.

융합 인력양성에 필요한 기간은 1년에서 5년 정도까지 분야별로, 기대수준별로 다양하지만 기본적으로 1년 이상이 걸린다는 것에는 공감한다. 대학의 경우에는 전담 연구원을 채용하기 어려운 상황이므로 학생의 분포에 따라 영향을 받는다. 소프트웨어를 잘 하는 학생이 들어오면 소프트웨어 개발까지 하다가 이런 학생이 없거나 졸업하면 기존 소프트웨어를 이용만 해서 연구하는 형식이다. 연구기관의 경우에는 자체적으로 필요한 인력을 확보하여 대응하는 것이 일반적 상황이다.

□ 연구 결과 공개 및 공유에 대한 문제

공학계열의 연구에서는 소스 공개가 어려운 경우가 많다고 한다. 과제 협약시 공개 범위를 선택하게 되어있으며 주로 실험결과물 정도를 공개하고 있다. 연구는 논문 위주이며 논문에 데이터가 일부 포함된다.

인터뷰의 상대를 보면 연구책임자가 참석했던 인터뷰에서는 공개에 대해 긍정적이었고, 박사과정 학생이 중심이 된 인터뷰에서 부정적인 의견이 많았다. 이것은 학생이 자기 권한을 넘어가는 것에 대한 의견에 소극적으로 대응한 것으로 보인다.

반면 학생연구자의 경우에도 자신의 연구에 해외의 공개 소프트웨어를 활용하는 사례는 많았다. 국내 오픈 소스는 많지 않기 때문에 활용사례로 찾기 어려웠다.

데이터와 소프트웨어 공개 수준은 도메인별로 많은 차이를 보였다. 기초 연구의 경우에는 적극적인 경향이고 공학쪽은 비교적 소극적이었다. 공학쪽 연구는 주로 MATLAB, LabView 등 상용소프트웨어를 사용하고 업체의 지원에 의존하는 경향이 높았다. 이런 환경에서는 자체적인 소프트웨어 개발도 업체에서 지원하는 환경에 따라 갈 수밖에 없다. 예를 들어 연구에 모델 개념을 도입하고 싶어도 자체적으로 환경을 개발하기는 어렵고 업체에서 모델기반 개발 방식을 제공하기 시작하면 그때부터 따라 가는 등의 사례가 있었다. 이런 연구 환경에서는 상용 소프트웨어로 연구를 거의 진행할 수 있기 때문에 그 이상의 성과를 위해 소프트웨어 전문가와 협업을 할 필요는 별로 느끼지 못하고 있었다.

공개되는 소프트웨어에 대한 인증 조직의 필요성을 제시한 경우도 있었다. 소프트웨어를 모아 놓기만 해서 되는 일이 아니라 활용을 도와주는 환경이나 개발 등이 필요하다.

3. 공개 포럼

지난 10월 14일 과제의 연구내용을 종합하여 공개 포럼을 개최하였다. 포럼은 소프트웨어정책연구소의 정기 포럼의 일환으로 열렸으며 “과학기술연구의 소프트웨어 활용 개선 방안”을 주제로 연구소 김석원 책임연구원과 한국과학기술정보연구원 이상환 실장의 발제, IITP 김두현 CP와 고등과학원 주기형 박사의 패널 토의로 진행되었다.

주요 내용은 정부의 과학기술연구과제에서 생산되는 공식, 비공식 산출물 중 소프트웨어와 과학 데이터를 공개하는 방안에 대한 논의였다. 과학데이터 공유 방안은 미래부에서 추진 중인 방안에 대한 소개였고 소프트웨어 공개 방안은 새로 추진할 정책에 대한 의견을 들어보는 자리였다.

기초 연구에 대해서는 공개에 찬성의견이 다수였으나 일부 신중한 의견도 있었다. 이 경우에도 공개에 반대하는 것이 아니라 공개 소프트웨어의 선택이 좋다는 논리의 확산이 먼저라는 지적이었다. 또한 사업화가 전제되는 경우에는 신중해야 한다는 의견도 있었다.

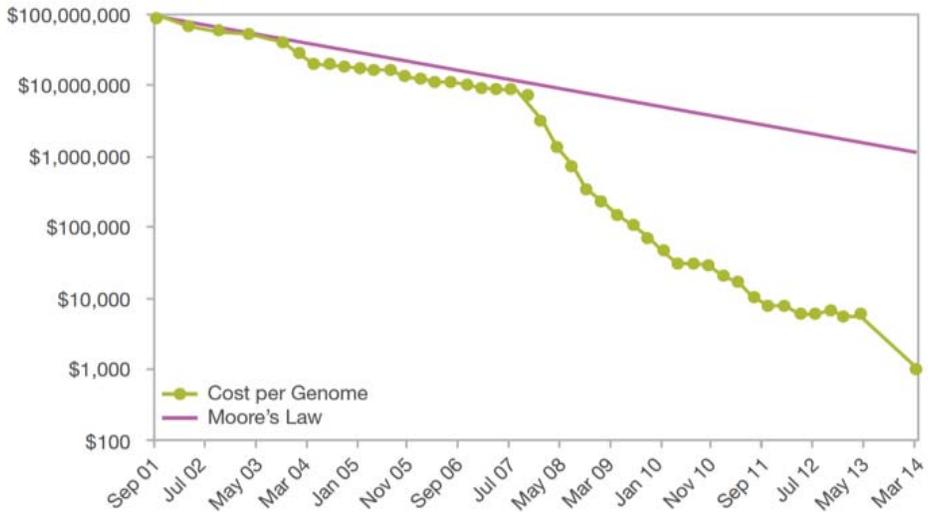
제2절 소프트웨어 활용 사례

생물 연구에서 소프트웨어가 차지하는 역할을 소개하기 위해 크게 가장 기본적인 데이터를 다루는 게놈 서열 분석과 상위 레벨에서 종합적인 구조를 연구하는 시스템 생물학의 관점으로 나누어 알아보기로 한다.

차세대 시퀀싱 기술의 개발에 의해 게놈 서열 추출 속도가 획기적으로 빨라지면서 게놈 서열 분석을 통한 생명현상과 질병에 대한 연구가 활발해졌다. 추출 속도가 빨라지면서 자연스럽게 따라오게 되는 문제는 분석할 데이터가 급격히 늘어난 것이다. 한 사람의 게놈 서열은 2014년 현재 \$1,000에 3일이면 구할 수 있다. 일반적인 차세대 시퀀싱 데이터는 전체 게놈 시퀀싱을 할 경우에 한 사람당 1.5테라바이트 정도가 된다. 이 데이터는 게놈 서열 정보이므로 게놈의 기본 단위인 A, G, C, T로 구성된 매우 긴 문자열이 주축이다. 1.5테라바이트 하드디스크에 AGCT 문자로만 구성된 텍스트 파일이 꽉 채워져 있는 것을 연상하면 된다.

게놈 서열 분석은 이 데이터에 대해서 스트링 매칭 연산을 하는 것이 기본이다. 이미 알려진 스트링이 어디에 있는지 찾는 매칭 연산은 컴퓨터과학에서 많이 사용되는 기술이며 많은 연구가 되어 있다. 게놈 서열에서 다른 점은 스트링 매칭이 완전히 매칭하는 경우만이 아니라 변형된 스트링도 찾아야 하는 것이고 이 변형에 대한 정보는 생물학적인 지식이 필요하다. 가령 스트링의 일부가 반복되어 나타나는 경우도 있고, 일부가 누락되는 경우도 있다. 또한 데이터를 시각화하여 편하게 볼 수 있도록 해야 하며 데이터 저장, 전송, 압축, 인덱싱 등도 중요한 분석 기능이다.

[그림 3-1] 차세대 시퀀싱 기술이 무어의 법칙을 추월



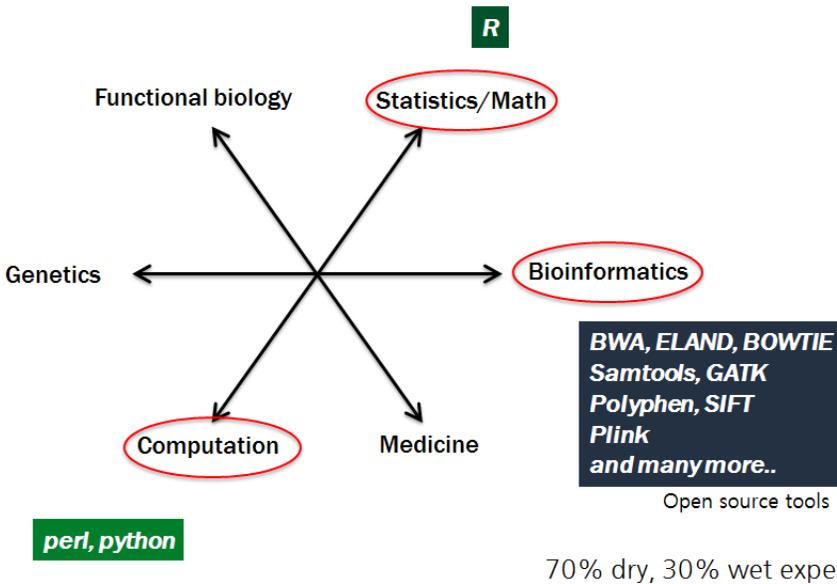
자료: <http://ieglabletters.blogspot.kr/2014/01/1000-genome.html>

[그림 3-2] 차세대 시퀀싱에서 생산되는 데이터 크기

Type	Per sample		
	WGS(70X)	WES(100X)	RNA-seq
Fastq (.gz)	90G	3G	3G
BWA BAM	300G	8G	8G
GATK BAM	300G	8G	8G
Pileup	400G	10G	
Pindel	1.3G	200M	
VarScan	86.2M	20M	
CNV & SV	200M	45M	
Others	300G	8G	8G
Total Size	~1.5 Tb	~40G	~27G

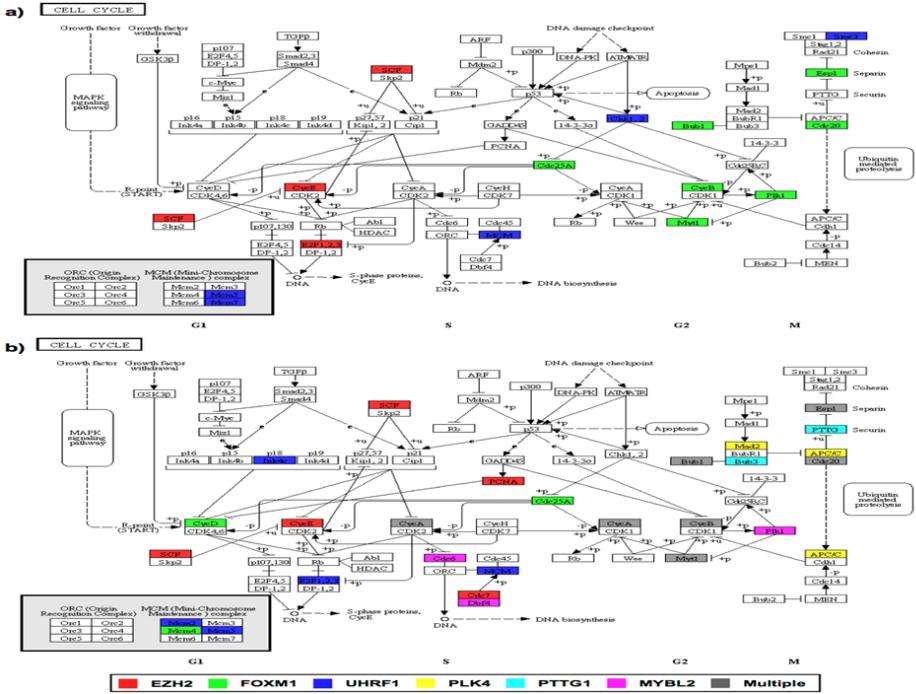
이런 분석을 하기 위해서는 기존의 소프트웨어 프로그램을 그대로 이용하는 경우도 있고 데이터 분석을 위한 소프트웨어 프로그래밍을 하여 연구 할 필요도 있다. 국내 연구 환경에서는 주로 해외의 공개소프트웨어를 많이 사용하고 추가로 필요한 것은 직접 프로그래밍을 해서 연구를 하고 있다. [그림 3-3]은 국내 한 연구실의 사례이다. 이 연구실에서는 전체 실험 시간의 70%를 드라이 실험, 즉 컴퓨터를 사용해서 데이터 분석하는 일에 쓰고, 30% 정도만 실제 시료를 다루는 데 사용하고 있다고 한다. 이것은 생물 연구가 얼마나 소프트웨어에 의존하고 있는지를 잘 보여주는 사례이다.

[그림 3-3] 생물 연구실의 소프트웨어 활용 사례



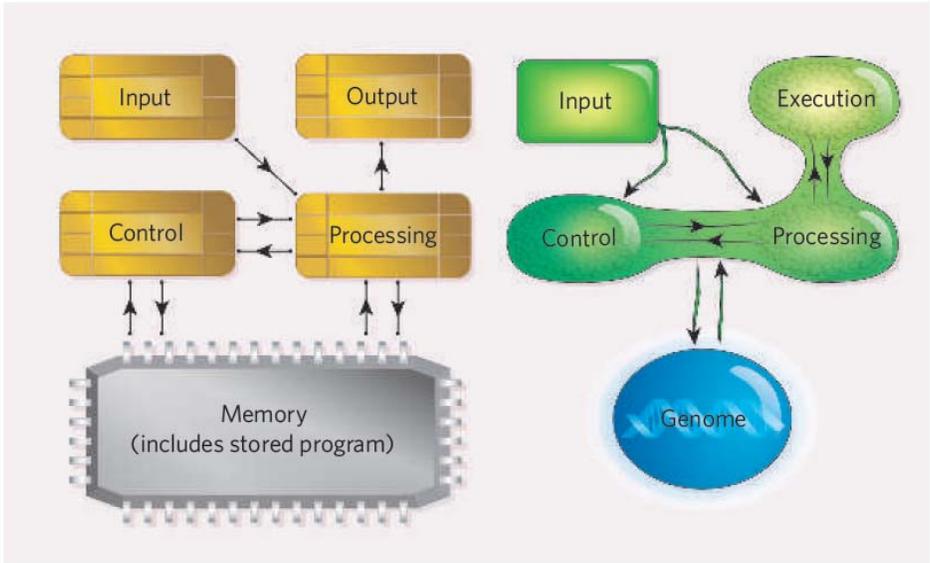
게놈 서열 분석과 같이 기초적인 데이터를 분석하여 연구하는 것이 아니라 게놈보다 상위 단계의 RNA, 단백질, 세포 단위의 각종 생물학적 사건과 인과관계를 다루는 시스템 생물학에서는 소프트웨어가 더욱 필요하게 된다. 여기에서는 생물학 정보의 경로, 정보의 전달, 단백질 상호작용 등 구조적 정보와 제어 정보를 모두 고려하여 세포가 동작하는 방식을 분석하기 때문에 더욱 복잡한 연산이 필요하고 컴퓨터식 논리가 활용될 여지가 많다.

[그림 3-4] 시스템 생물학의 정보 표현 사례



시스템 생물학에서는 스트링 매칭 연산이 아니라 컴퓨터 과학의 프로그램 설계 기법, 프로그램 언어, 정보 표현 방법 등 종합적인 지식이 적용되어야 한다. 이 분야에는 개별 연구를 위해서도 컴퓨터 지식이 필요하고 종합적인 현상을 과학적 기준에 따라 정확히 표현하는데도 필요하다. 세포 또는 그 이상의 단계에서 발생하는 생물학적 현상을 현재 생물학 연구에서는 자연언어로 풀어쓸 수밖에 없다. 수학을 이용하여 이론을 명확하게 표현하는 것이 중요한 과학 분야에서는 이런 방식이 만족스럽지 않다. 그러나 생물학 현상을 자세히 분석하면 컴퓨터의 기본 모형인 폰노이만 모형과 유사한 부분이 있다. 생물학계에서는 이런 유사점을 이용하여 생물 현상의 표현을 컴퓨터과학계와 공동으로 개발하려는 요구도 있다.(Brent and Bruck, 2006)

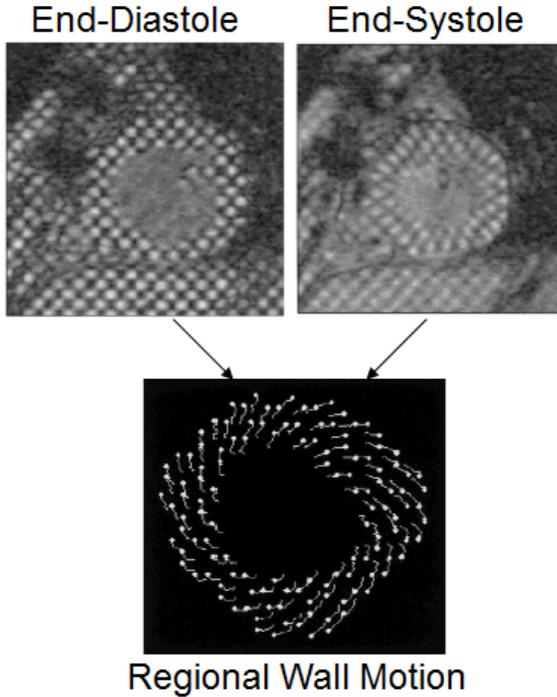
[그림 3-5] 생물계와 폰노이만 모형의 유사점



자료: Brent & Bruck, 2006

의학에서 많이 쓰이고 있는 CT와 MRI의 영상데이터를 관리하고 분석하는데 컴퓨터를 사용하고 있다는 것은 이미 알고 있다. 이 영상데이터를 분석하고 시각화하는 영상처리 기술을 발전시켜 보다 적절한 치료를 할 수 있도록 하는 연구가 활발하게 진행되고 있다. 가령 MRI로 찍은 여러 장의 영상을 분석하면 신체 장기에 대한 정보뿐 아니라 시간에 따른 혈류 특성에 대한 동적인 정보도 얻을 수 있다. 이 기술을 이용하면 심장의 혈류 상태, 척추 디스크의 돌출 현상 등을 보다 자세히 조사할 수 있고 적절한 조치를 취할 수 있다.

[그림 3-6] 영상처리를 이용한 심장혈류 계산



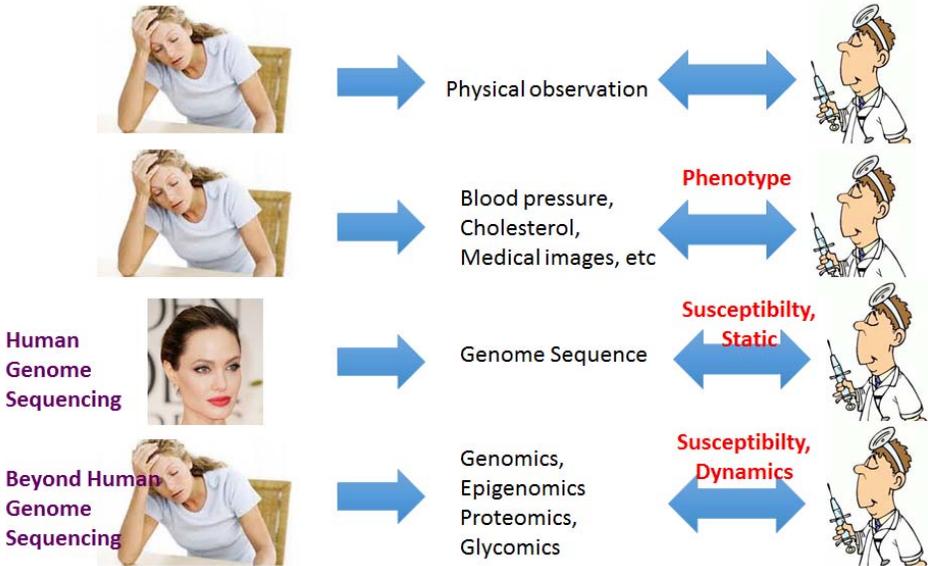
최신 생물학 연구의 경향을 보면 미래의 진단은 환자 맞춤형으로 발전할 것으로 예상된다. 현재는 환자의 혈액 검사와 영상 검사 등을 통해 일률적인 방법으로 진단하여 치료하는 것이 일반적이다. 환자의 게놈 서열을 분석하여 미래의 발병 인자를 고려하여 위험한 질병은 미리 예방하는 진료 방식이 최근 화제가 된 안젤리나 졸리의 유방암 예방 수술로 인해 알려지기 시작했다.²⁶⁾ 미래에는 선천적으로 유전된 게놈 서열만이 아니라 후천적 요인에 의해 개인적으로 달라진 생물학적 현상도 함께 고려하여 예방하는 단계로 발전하게 될 것으로 보고 있다. 여기에는 유전체학(genomics), 후성유전체학(epigenomics), 단백질체학(proteomics), 글리코믹스(glycomics) 등 소위 오믹스 데이터의 분석 기술이 중요한 역할을 하고 있고 이 데이터를 분석할 수 있는 소프트웨어가 필수

26) <http://www.bbc.com/news/health-29251765>

요소이다.

[그림 3-7] 미래의 환자 맞춤형 진료

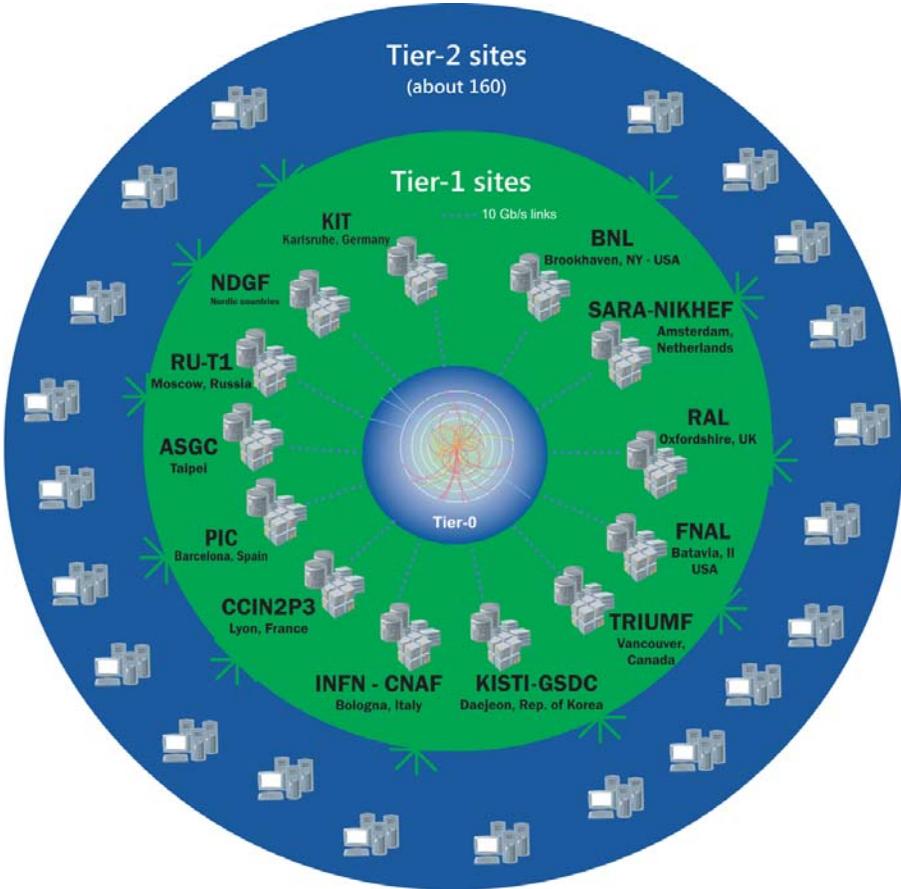
Personalized Medicine vs. Data



물리학 분야는 다뤄야하는 데이터의 규모가 생물학을 넘어선다. 입자물리학에서는 개별 입자의 행동을 수학으로 계산하기 어렵기 때문에 시뮬레이션을 하여 연구를 진행하며 이 과정에서 많은 데이터를 계산하고 처리해야 한다. 프랑스와 스위스에 걸쳐 구축된 대형 강입자 충돌기는 양자역학실험을 위해 세워진 장치이며 지하 100m, 둘레가 27km에 달하는 거대한 실험 설비이다. 여기에서는 매년 수십 페타바이트의 실험 데이터가 생산된다. 이 데이터의 분석, 저장, 관리를 위해 40여개국 170여 기관이 참여하는 Worldwide LHC Computing grid를 운영하여 공동으로 분산, 병렬처리를 하고 있다.²⁷⁾ 한국과학기술정보연구원(KISTI)의 Global Science Data-hub Center도 2014년부터 Tier1 기관으로 승인되어 연구에 동참하고 있다.

27) <http://wlcg-public.web.cern.ch/>

[그림 3-8] Worldwide LHC Computing grid



자료: <http://wlcg-public.web.cern.ch/tier-centres>

전문가 인터뷰의 3번 인터뷰의 대상인 국내 물리학 연구실의 경우에도 연구를 위해 강력한 컴퓨터 시스템을 구축해 놓고 있다. 6400코어의 클러스터 컴퓨터를 연구소 내에 구축하여 사용하고 있으며 우주 연구를 위한 시뮬레이션에는 한국과학기술정보연구원의 슈퍼컴퓨터를 이용하여 실험하고 있다. 연구를 위한 클러스터 컴퓨터를 구축하는데 국내 사례가 없어서 물리학자가 관리용 소프트웨어를 공개 소프트웨어를 수정하여 개발하

였다고 한다. 연구실 내에 소수의 소프트웨어 전문가를 고용해서 일을 나눠하고 있으나 지금도 연구를 위한 소프트웨어 개발을 직접 하고 있고 일부 결과는 홈페이지에서 웹서비스 형태로 공개하여 다른 연구자가 이용할 수 있도록 하고 있다. 자문회의 2번째 세미나에서 물리학 사례를 발표한 자문위원도 연구에서 프로그래밍은 핵심이기 때문에 직접 클러스터 컴퓨터나 슈퍼컴퓨터의 프로그래밍을 하여 연구를 하고 있다고 발표했다. 과학연구에서 핵심 이론에 대한 연구는 프로그래밍이 필수이기 때문에 연구자가 프로그래밍 지식을 갖추는 것은 기초 교양이 되었다.

제3절 국내 현황 조사 결과의 분석

국내 환경 조사 내용을 종합하면 연구에 있어서 소프트웨어의 중요성에 대해서는 모두 이해하고 있다고 판단된다. 자문회의와 전문가 인터뷰에 포함된 모든 분야에서 소프트웨어가 연구 활동의 상당 부분을 차지하고 있고 연구자도 동의한다.

협업의 필요성에 대해서도 대부분 필요하다고 인정한다. 다만 입장에 따라 시각의 차이가 있었다. 컴퓨터를 써야 하는데 내가 모르는 부분을 채워 줄 이상적인 전문가를 바라는 경우도 있었고, 소프트웨어의 잠재력을 이해하기보다 일반적으로 알려진 활용 방식을 자신의 연구 분야에 도입하려는 형태의 협업을 그리는 경우, 있으면 좋지만 당장의 연구가 급하니까 그냥 연구를 진행하겠다는 경우 등 여러 가지 상황이 제기되었다.

협업의 필요성에 대해 낮게 평가한 경우도 자세히 들여다보면 입장에 따른 영향을 받은 것으로 보인다. 자체적으로 전문성을 가진 인력을 보유한 상태인 경우에는 굳이 협업까지 할 필요가 없다는 의견을 낸 사례가 있고, 논문을 쓰는 것이 시급한 경우에도 협업보다는 도메인내의 연구에 집중하여 논문 쓰는 것에 집중해야 한다는 사례도 있었다. 특히 공학쪽은 연구원이 소프트웨어를 개발한 경험이 있는 경우가 많아서 필요한 것은 자체적으로 해결하는 경향이 있었다.

학제간 연구 경험을 가진 인력의 필요성은 모두 공감한다. 다만 확보 방법에 있어서 대학교육을 통한 정책적 조치는 큰 효과를 기대하지 않는다. 융합학과에 많은 기대를

하지 않으며 소프트웨어 연구자가 공동 연구를 통해 지식을 습득하는 것이 자연스러운 방법이다. 이런 방식은 오랜 기간이 걸리기는 하지만 연구의 수준을 높이기 위해서는 필수적인 과정으로 보고 있다.

한편으로는 정부가 지원하는 연구조직에서 학제간 연구자와 계산과학자를 확보해 주기를 희망하는 사례도 있었다. 컴퓨터로 과학기술의 고성능 계산 처리를 하는 것은 이미 하나의 전문 분야로 자리 잡고 있으며 연구에도 도움을 줄 수 있다. 정부가 지원하는 계산과학연구소가 생긴다면 과학연구 분야의 소프트웨어 전문성 부분을 채워 줄 수 있을 것이고 이 연구소의 연구원이 학제간 경험을 가진 연구자로 성장할 수 있을 것이다.²⁸⁾²⁹⁾

연구결과로 생산된 소프트웨어와 연구 데이터의 공개에는 대부분 긍정적 입장이다. 대부분의 전문가는 방법과 환경의 문제일 뿐 원칙 자체에는 반대하지 않았다. 반대의 경우에도 조건부 찬성 의견이었다. 가령 연구 참여기업에서 사업화하려는 대상은 공개 하기가 곤란하다거나, 논문에는 드러나지 않는 연구의 자세한 전문지식이 소프트웨어에 드러나기 때문에 핵심기술을 제외할 수 있도록 하자거나, 연구자가 공개 여부를 판단할 수 있게 해달라거나 하는 등의 조건이다.

공개 방식에 대해서는 강제적보다는 자발적으로 하는 제도, 기계적으로 모든 것을 공개하는 것이 아니라 선택적으로 할 수 있는 방안을 선호했다. 공개를 위해 소프트웨어와 데이터를 정리하는데 인력과 시간이 들어가게 되니 현재 진행 중인 연구에 적용할 경우에는 오버헤드에 대한 해소 방안이 함께 제시되길 바라는 의견도 있었다.

정보를 공유할 수 있는 학제간 커뮤니티 조직과 정기적인 행사에 대한 요구는 공통적이었다. 협업을 활성화하는 최소한의 장치로 이런 모임이 있고 정기적으로 워크숍이나 학술대회 등을 개최한다면 서로간의 연구 분야에 대한 관심을 갖는데 도움이 될 것이다. 연구자의 흥미를 끌고 자발적인 참여를 촉발한다는 측면에서 중요한 활동이다. 국내

28) University of Tsukuba. Center for Computational Sciences.

<http://www.ccs.tsukuba.ac.jp/eng/>

29) <http://www.journals.elsevier.com/journal-of-computational-science/>

에서 성공적으로 운영되는 HCI 학회도 이런 학제간 성격을 가지고 있는데 전산학, 디자인, 인문학 전문가의 교류를 일으키는 중요한 역할을 해 오고 있다.³⁰⁾

해외 사례 조사에서 보듯이 선진국은 국가적 차원에서 과학기술 연구를 활성화 할 수 있도록 컴퓨터과학 기술을 적용하고 함께 연구하고 새로운 분야를 개척할 수 있는 길을 열어주는 장기적인 계획을 세워 추진하고 있다. 서로 다른 전문분야간의 협업이 힘든 일이고 시간도 많이 필요하지만 현대의 연구에서는 꼭 필요하다는 것을 이해하고 있다.

반면 우리나라에서는 정보통신 분야에서의 융합사업은 몇 년 전부터 추진되어 오고 있으나 과학기술연구에서의 소프트웨어 활용을 지원하기 위한 정책은 미흡한 상황이다. 이것은 정부 부처가 다른 것에도 원인이 있을 것이고 과학기술 분야와 소프트웨어 분야간의 교류의 부족에도 원인이 있을 것이다.

30) <http://hcikorea.sql.co.kr/2014/>

제4장 문제 분석

제1절 문제의 도출

1. 학제간 공동 연구에 대한 연구 사례

학제간 공동 연구 (Collaborative Interdisciplinary Research)의 특성에 대해서는 해외에서도 꾸준히 연구되어 왔다. 기존의 연구는 본 과제의 관심인 소프트웨어 전문가와 과학기술 전문가와의 학제간 연구가 아니라 일반적인 학제간 연구를 다루고 있다. 분야별로 양상이 다를 수는 있겠으나 국내 전문가 의견을 뒷받침하는 결과도 있으므로 살펴볼 필요가 있다.

Qin, Lancaster, Allen(1997)은 문헌 조사와 설문 조사를 통해 학제간 공동 연구의 특징을 발표했다. 그의 발견에는 분야에 따라 학제간 연구의 정도가 다르며 생물, 지구과학, 의학은 높은 반면 수학은 낮다는 것이 있다. 또 발표된 논문에서 다른 분야 연구자와 공동 저자가 되는 비율이 높지 않다고 했다. 이것은 학교의 평가체계가 피어리뷰, 테뉴어, 승진 등을 포함하여 단일 분야의 연구를 위한 체계로 되어 있기 때문이다. Carayol과 Thi(2004)의 연구에서는 산업계와 공동 연구를 하는 경우에 산업계에서 이론이 아니라 문제 해결형 결과를 요구하기 때문에 학제간 공동 연구를 추진하는 경향이 높다고 발표했다. Qin 등(1997)과 마찬가지로 학교의 평가체제는 학제간 공동 연구를 방해하는 수준에 가깝다고 주장하기도 했다. 학교에서는 주어진 기간 안에 많은 논문을 발표하는 것을 요구하기 때문에 연구기간이 오래 걸리고 연구결과가 논문을 쓸 수 있는 수준이 아닐 위험성이 있는 학제간 공동 연구는 가급적 피하는 경향이 있다는 얘기다. 최근에도 van Rijnsoever(2011)가 유사한 주장의 논문을 발표했다. 전략적 과제는 학제간 연구가 진행되는 경향이 있지만 일반적으로는 경력 개발에 손해가 되는 경향이 있어서 꺼려한다는 것이다.

1990년대부터 지금까지 이런 경향이 꾸준히 유지되는 것은 학제간 공동 연구의 어려움을 보여주는 것과 동시에 학계의 평가 체계가 문제 해결을 위한 노력은 경시하는 구조라는 것도 보여준다. 생물학자와 컴퓨터 공학자가 공동 연구를 하여 DNA의 구조를 잘 보여주고 생물학자의 연구를 도와주는 프로그램을 만드는 경우에 이런 기능을 하는 최초의 프로그램은 논문도 될 수 있고 명성도 얻을 수 있겠지만 그 이후에 개발되는 프로그램은 논문이 되기도 어렵고 학교에서 인정받기도 어렵다.

<표 4-1> 학제간 연구에 대한 연구 결과

저자 및 연도	연구 방법	결과 요약
J. Qin et al. 1997	문헌 통계 분석과 설문	분야별로 학제간 연구 비중 차이 (생물, 지구과학, 의학은 높고, 수학은 낮음) 학교의 평가체계가 단일 학문의 연구에 유리함
N. Carayol et al. 2004	연구원의 논문 통계 분석	산업계와의 계약에 의한 연구에서 많이 나타남 (학제간 연구는 이론 탐구가 아니라 문제 해결의 성격을 가짐) 영년직 후에 학제간 연구를 하는 경향이 약간 있음 분야별로 차이가 큼 보상체계가 많지 않음 (평가체계)
F. J. van Rijnsoever et al. 2011	연구원 설문 조사	경력개발에 부정적임 (학교평가체계와 맞지 않음) 전략적 과제와 연관되는 경향

해외 연구과 국내 전문가 의견 중 가장 눈에 띄는 공통점은 기존 연구 평가체계와의 부조화이다. 대학, 연구소 등의 체계는 어디서나 논문이 핵심 성과 지표이다. 모든 학제간 연구는 이론적 연구보다 문제해결 과제의 성격이 강하므로 논문생산성이 낮다. 좋은 논문이나 혹은 현실에 얼마나 기여했느냐는 이차적 문제이고 단위 시간당 출간 수로 보았을 때의 비교이다. 학교의 승진체계는 논문 숫자를 채워야 승진하는 것이 기본이므로 성과에 쫓기는 연구자는 학제간 연구를 하는 것은 위험한 선택이다. 더구나 초기 습득 기간이 긴 것에 대한 고려도 없고 투자기간이 긴 것에 대한 보상도 없다.

시간이 지나면서 이런 상황이 해소될 것이라는 기대도 크게 믿을 바가 못 되는 것 같다. 1997년 Qin 등의 논문과 2011년 van Rijnsoever의 논문은 연구 대상과 국가, 범위가 다르다는 것을 고려하더라도 지난 15년간 큰 변화가 없었다는 것을 추측하게 한다.

또 하나 눈에 띄는 것은 학제간 연구가 이론연구가 아닌 문제해결 성향이 있으므로 산업계와의 공동 연구나 전략적 과제와 연관되는 경향이 있다는 점이다. 국내 상황에 대입해 보면 정부 주도로 융합연구를 추진하는 것이 이것과 상관이 있어 보인다. 연구자 입장에서 과제 수행이 중요한 요건이면 전략적으로 융합과제에 참여하여 연구할 동기가 될 수 있다.

2. 국내 연구 상황의 문제

국내 현황 조사 결과에서 우선 드러나는 문제는 두 분야를 이어주는 융합 인력이 절대적으로 부족하다는 점이다. 도메인 전문가는 당연히 자신의 분야에 대한 용어를 이해하면서 필요한 연구를 수행할 수 있는 소프트웨어 전문가가 있으면 좋겠다고 생각한다. 이런 연구자가 있다면 의사전달이 훨씬 쉽고 빨라질 것이고 연구가 신속하게 진행될 수 있을 것이다. 그러나 현실에서는 이런 인력이 많지 않기 때문에 차선책을 택한다. 그 중 하나는 도메인 전문가의 연구팀 내에서 소프트웨어 개발 경험이 있는 사람을 찾아내거나 고용하는 것이다. 이 인력은 해당 도메인 연구자이면서 컴퓨터에 관심을 가지고 취미든 부전공이든 개인적인 노력에 의해 컴퓨터 프로그래밍의 경험을 쌓게 된 사람이다. 이런 사람이 없거나 고용할 여건이 안 되면 직접 소프트웨어를 공부해서 필요한 코딩을 하거나, 그냥 공개소프트웨어나 상용 소프트웨어에서 제공되는 기능만 이용하는 것이다. 소프트웨어의 확장성과 융통성을 생각하면 이런 연구 환경은 궁극적으로 전체 연구개발 수준을 떨어뜨리는 요인이 될 것이다.

인력 부족과 관련된 또 하나의 요소는 필요한 수준의 전문가가 되기까지 긴 숙련 기간이 필요한 점이다. 전문가의 경험에 따르면 소프트웨어 전문가가 각 도메인 전문가와 공동 연구를 하는 경우에 1년에서 3년 정도의 기간이 지나야 비로소 어느 정도 연구를 진행할 수 있는 수준이 된다고 한다. 1년은 공동 연구를 할 때 서로 편하게 일을 하게 되는 기간이고 소속기관이 전문가를 고용하여 육성하는 경우에는 3-4년을 말하고 있다. 이것은 공동 연구와 직접 고용한 경우의 기대 수준의 차이라고 보인다.

1년이라는 수치는 오현석, 배형준, 김도연(2012)이 수행한 과학기술분야 융합연구자 입문 과정에 관한 연구 논문에서도 보고된다. 이 논문에서 융합연구자라고 표현된 연구자와의 인터뷰에서 자신의 분야를 떠나 새로운 분야로 옮겨서 익숙해 질 때까지 “1년”이란 표현이 자주 언급되었다고 한다. “한 1년 정도는 눈물이 나요”, “짐을 싸들고 적진으로”, “힘난한 인생의 시작” 등의 표현은 국내 연구진이 원하는 인력이 얼마나 어려운 과정을 겪게 되는지 보여주는 사례라 하겠다.

이렇게 긴 시간이 필요한 이유는 분야별 전문 용어를 이해해야 하고, 같은 용어를 다르게 쓰는 경우도 있고, 관점과 문화의 차이도 이해해야하기 때문이다. 공동 연구를 하는 연구자가 처음 부딪치는 장애는 용어의 이해이다. 어느 분야건 전문가 그룹에서는 약자를 많이 쓰기 때문에 용어의 습득이 필요하다. 그리고 연구 방법, 관점, 문화 등을 이해하기 위해 새로운 전문 분야에 대해 깊이 공부하고 공동 연구자와 많은 대화를 해야 한다. 이런 초기의 높은 장벽이 전문 인력의 부족을 일으킨 요인 중 하나이다.

전문가의 참여 동기가 부족한 점도 중요한 문제이다. 국내 연구의 조사에서 학제간 공동 연구를 할 때 연구 성과에 대한 만족도가 다른 경우가 여러 번 지적되었다. 학제간 공동 연구는 초기의 긴 숙련기간이 소요된다. 컴퓨터과학 분야에서 학제간 연구를 하면 초기 습득 기간 동안 컴퓨터과학에서 개발된 기술을 과학기술 분야에 적용하는 과정을 하며 새 분야를 익히게 되는데 이 과정의 연구 성과가 충분히 인정받지 못한다. 논문 중심의 평가 체제에서 연구하는 연구자는 현행 평가 제도와 일치하지 않는 연구를 장기적인 관점에서 추진하기 어렵다. 양적 평가에서 질적 평가로 점차 개선하는 정책이 추진되고 있으나 연구자에게 파급되기는 아직 이르다. 더구나 이런 과정을 거친 후에도 결과를 낼 수 있다는 확신이 부족하기 때문에 성과에 쫓기는 위치에서 시간과 노력에 비해 위험도가 높은 연구 방향을 택하지 않는 경향이 생긴다. (미래창조과학부, 2013d)

마지막으로 국내 연구 환경은 다른 분야의 인력, 연구내용, 문제에 대한 공유 기회가 부족하다. 과학기술 연구자와 소프트웨어 연구자가 교류할 수 있는 학술대회, 워크샵, 연구회 등의 오프라인 행사와 온라인 커뮤니티 등이 활성화되어 서로의 연구 내용과 관심사를 알게 되는 것이 협력의 시작이 된다. 이런 행사와 환경이 자리를 잡고 자료가 축적되면 연구자가 사전에 새 분야에 대해 조사하고 사고의 폭을 넓히는 기반이 될 수 있다.

제2절 해결 방안 검토

앞서 살펴본 내용을 종합하면 과학기술 연구에서 소프트웨어 활용을 활성화하기 위해서는 다음의 문제를 해결해야 한다. 공동 연구를 할 기회를 연구자에게 제공하여 초기 습득기간을 넘길 수 있는 동기를 주고, 이를 통해 두 분야에 익숙한 전문가를 배출하고, 전문가 집단 간에 정보 교류의 장을 만들어 줘야 한다.

1. 공동 연구

공동 연구를 활성화하는 것은 현 정부에서 융합 과제를 강조하고 추진하고 있다. 현실적으로 논문위주 평가제도의 제약이 여전하므로 연구자에게 동기부여를 하려면 정책적으로 연구 과제를 추진해야 한다. 정부에서는 국가중점과학기술 전략로드맵 2014에서 ICT 융합 신산업 창출을 5대 분야 중 하나로 선정하여 융합 과제를 지속적으로 육성하는 전략을 추진하고 있다.³¹⁾ 나머지 4대 분야에 대해서도 세부 내용에서 융합 기술을 포함하고 있다. 또한 2015년 정부연구개발투자 방향 및 기준에서 기초연구 및 융합연구 확대와 ICT·SW 기반 산업융합을 지속적으로 강화하는 등 융합 과제 정책을 꾸준히 펴고 있다.³²⁾ 학계의 평가 체계가 쉽게 바뀌지 않는다는 점을 감안하면 정책적으로 융합 과제를 추진하는 것이 단기적으로 효과를 기대해 볼 수 있는 방안이다.

주의할 점은 전문가 인터뷰에서도 드러난 바 있듯이 융합 과제가 도메인과 ICT의 물리적 결합에 머무르는 것을 경계해야 한다. 어느 분야건 ICT와의 융합을 생각할 때 처음 떠오르는 것은 물리적 결합이다. 건물에 네트워크를 설치하고 가정에 태블릿을 나눠주는 것은 건설과 ICT의 융합이라고 부를 수는 있으나 피상적이고 물리적인 결합이다. 본 과제에서 해결하려고 하는 것은 소프트웨어와의 화학적 결합을 할 수 있는 공동 연구 기반이다. 이것은 소프트웨어를 이용하여 과학기술 연구에서 그 동안 해결하지 못했던 문제를 해결하거나 새로운 문제를 찾아내서 도전하는 등의 심도있는 연구를 하는 것을 말한다. 이것을 오현석 등(2012)은 “문제를 보는 관점의 변화” 라고 표현하고 있다.

31) 미래창조과학부, 2014e

32) 미래창조과학부, 2014d

이렇게 되려면 전문가가 상대 분야를 깊이 이해하게 되어야 한다.

아직 정부의 융합과제에 대해 직접적인 결과나 파급 효과에 대해 판단하기는 어렵다. 이 형태의 연구가 잘 진행되면 공동 연구가 활성화되는 효과가 있을 것이다. 정부에서 적극적으로 추진하는 전략이므로 본 연구에서 새로운 형식의 공동 연구 과제를 제안하는 것보다 연구자에게 동기부여가 되고 결과적으로 좋은 융합 사례가 나오는 것을 기대하는 것이 바람직하며 좀 더 지켜보고 평가하는 것이 적절하다.

2. 인력 양성

인력 부족의 문제도 중요하다. 이 문제의 특징은 다수의 평이한 인력이 아니라 소수의 전문성있는 인력이 필요하다는 점이다. 소수가 어느 정도를 의미하느냐는 불확실하지만 대학에서 교육을 통해 양산해 내는 수준의 인력 공급이 필요하지는 않다는 의미로 이해할 수 있다. 전문성있는 인력이 필요한 것은 한 분야의 전문성을 충분히 쌓은 후에 새로운 분야에 도전하는 연구자라는 의미이다. 오현석 등(2012)의 연구에서도 같은 발견을 하고 있다.

“이는 융합교육이 전문성이 낮은 상태에서도 창의적 성과를 창출할 수 있게 하는 방법지팡이가 될 수는 없으며, 융합은 전문분야의 타 분야를 아우르는 지식과 통찰의 축적이라는 관점 아래에서만 제대로 수행될 수 있음을 보여주는 것이다.”³³⁾

이런 점을 종합한다면 소프트웨어 전문가로 경력을 쌓은 연구자가 과학기술 분야를 공부하거나 공동 연구를 수년간 수행하여 전문성을 쌓거나, 과학기술 연구자가 컴퓨터 과학을 공부하여 전문가 수준으로 실력을 쌓아야 필요한 수준에 오를 수 있다는 의미이다.

인력 부족은 교육과정의 개선보다는 전문가사이의 협력 환경을 구축하여 점진적인 해결을 추진하는 것이 바람직하다. 교육과정의 개선은 시간이 오래 걸리는 일이다. 자문회의에서의 결론도 한 분야에서 일정 수준의 전문성을 갖춘 후에 다른 분야로 넓혀가는 방식으로 개발된 인재가 필요하다는 것이다.

33) 오현석 등, 과학기술분야 융합연구자의 융합연구 입문과 과정에 관한 연구, 2012, p. 328.

3. 정보 교류

서로 다른 분야 간의 정보 교류를 위한 기반의 부족도 중요한 문제다. 연구자간의 교류가 어려운 환경에서는 서로의 분야에서 풀어야 하는 문제에 대해서 알 수 없고 자신이 어떻게 기여할 수 있는지 아이디어를 내는 것도 불가능하다. 연구 내용이 공유되고 연구자가 정보를 주고받을 수 있는 현실 혹은 사이버 상의 공간, 행사, 활동이 제공되어야 한다. 이와 더불어 이런 환경을 충분히 활용할 수 있도록 연구에 대해 개방된 문화가 연구원 사회에 자리 잡아야 한다.

인프라의 부족은 정책적으로 지원이 필요한 분야이다. 가장 기초적인 지원은 다른 분야의 연구자간에 정보가 유통될 수 있는 기회를 주는 것이다. 정기 학회, 워크샵, 포럼 등을 통해 서로 정보를 나누게 되면 이것이 동기가 되어 공동 연구나 협력으로 발전할 수 있는 기회가 될 수 있다.

4. 공개 소프트웨어

해외 연구 문화와 국내 조사 결과를 비교하면 두드러진 차이는 소프트웨어의 공유에 있는 것으로 보인다. 해외 연구소에서는 소프트웨어의 중요성을 인지하여 전문 개발자가 고용되어 있고 이런 직업이 과학에 익숙한 소프트웨어 전문가를 배출하게 되는 배경이다. 소프트웨어의 중요성을 이해하려면 사용하는 소프트웨어를 공개하고 개선된 기능의 효과를 확인하는 기회를 가져야 한다. 해외에서 귀국한 연구자가 개방적인 태도인 것도 이것과 관련이 된 것으로 볼 수 있다. 그렇다면 현 상황에서는 소프트웨어 전문가와의 협업을 활성화하는 해결 방안의 시작은 공개에 있을 것으로 보인다.

연구 내용과 결과를 공개하고 공유하는 문화는 장기적 관점에서 중요한 요소이다. 우리나라는 해외에 비해 연구결과를 공개하는 것에 인색하다. 연구 결과로 나온 소프트웨어나 데이터를 공개하여 다른 연구자에게 도움을 주고 동시에 자신의 연구결과를 검증받는 것이 해외에서는 일반적인 관례로 되어 있고 정부에서도 장려하고 있다. 국내의 환경은 최소한의 공개만 하는 형편이다.

소프트웨어의 공개와 공유는 연구개발분야만의 현상이 아니다. 소스코드의 공개가 기술을 유출하는 것이 아니라 오히려 가치를 키운다는 문화의 확산에 따라 산업계에서도

회사의 근간을 이루는 것으로 보이는 소프트웨어를 공개하는 회사가 늘고 있다. 구글은 전사적으로 하나의 코드베이스만을 운영하며 모든 개발자가 대부분의 코드에 접근할 수 있다.³⁴⁾ (Schmidt와 Rosenberg, 2014) 페이스북과 넷플릭스도 대부분의 코드를 공개소프트웨어로 개발하고 있으며 심지어 보안이 까다로운 애플도 공개소프트웨어를 지원하고 있다.³⁵⁾³⁶⁾³⁷⁾

최상위 소프트웨어 업체에서 공개소프트웨어를 포용하는 전략을 펴는 이유는 소프트웨어 개발환경의 발달과 신기술 개발 주기의 단축으로 소스코드 자체의 가치는 빠른 시간에 하락한다는 것은 이해하고 있기 때문이다. 소프트웨어 기반 서비스와 제품은 끊임 없는 개선과 혁신만이 가치를 키울 수 있다. 닫힌 문화에서의 개선과 혁신은 개방된 환경보다 느릴 수밖에 없기 때문에 일류 소프트웨어 기업이 공개소프트웨어를 지원하는 정책을 택하고 있다.

공개소프트웨어의 품질에 대한 논쟁도 오랜 기간 이어져 왔다. 공개 소프트웨어가 반드시 품질이 좋다고 주장할 수는 없을지 몰라도 비공개 상용 소프트웨어보다 품질이 떨어지는 것도 아니라는 것은 최근의 통계 조사에서 드러났다. Coverity는 소프트웨어의 품질 및 보안 테스트 서비스를 제공하는 회사이며 공개와 비공개 소프트웨어를 포함하여 2013년 현재 1500개 이상의 프로젝트를 수행하였다. 이 회사에서 발간한 Coverity Scan: 2013 Open Source Report 에서 공개소프트웨어의 품질이 비공개 소프트웨어의 품질보다 높아졌음을 보고했다.³⁸⁾

5. 개선 전략

소프트웨어 전문가와의 협업을 활성화하기 위해 정부 과제를 추진하는 것은 단기적 대처 방안이다. 유사한 과제가 융합과제 형식으로 이미 진행 중이므로 중복이기도 하다. 전문가인터뷰의 결과에 따르면 사업화 성격이 강조되는 융합과제에서 정해진 연구기간

34) <http://googletesting.blogspot.kr/2011/06/testing-at-speed-and-scale-of-google.html>

35) <https://code.facebook.com/projects/>

36) <http://netflix.github.io/#repo>

37) <https://opensource.apple.com/>

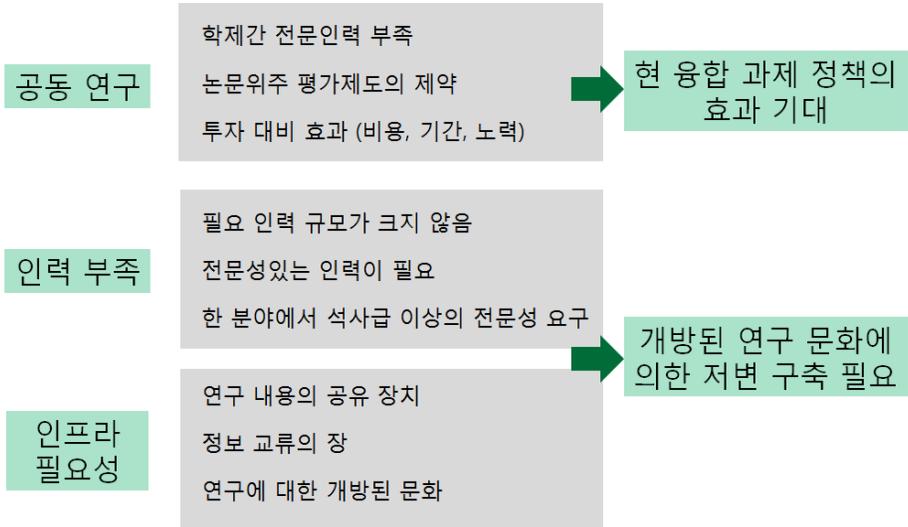
38) <http://www.coverity.com/press-releases/coverity-scan-report-finds-open-source-software-quality-outpaces-proprietary-code-for-the-first-time/>

내에 결과를 내려면 성과를 내는데 시간이 걸리는 공동 연구의 수준으로 깊이 내려가는 것이 어려울 수 있다. 그렇지만 본 과제에서의 소프트웨어 융합은 기반기술의 역할로서 화학적 결합, 즉 각 전문가가 서로의 분야를 깊이 이해하는 수준이 필요한 연구이다. 그러므로 융합 과제의 결과가 시간이 지나며 깊이를 더할 수 있도록 지원하면서 다른 한편으로는 보다 장기적인 대책을 마련해야 한다.

인력문제는 교육적 방법이 해답이 아니다. 수적으로는 교육 기관에서 배출하는 많은 인력이 필요한 것이 아니라 소수의 전문적 지식을 가진 전문가 필요하다. 또 이미 많은 융합학과가 있으므로 그곳에서 배출되는 인력 중 일부가 융합 인력으로 성장하는 것을 기대해 볼 수 있다. 그와 함께 전문가의 참여를 이끌어 낼 수 있도록 기반환경을 만들어 주는 것이 낫다.

종합적으로는 단기적 정책보다 장기적인 영향을 고려하는 전략적 접근이 필요하다. 장기적으로 본다면 협업의 활성화를 위해서는 국가 연구개발 과제 결과의 공유 문화를 확산하는 것이 필요하다. 과학기술연구 결과로 생성되는 소프트웨어와 데이터의 공개를 장려하면 도메인전문가는 공개되는 산출물을 잘 만들어야 하는 압력을 받고 소프트웨어 전문가는 공개된 소프트웨어를 접하면서 기여할 수 있는 분야를 찾기가 용이해 질 수 있다. 글로벌 공개소프트웨어 사이트에서와 마찬가지로 공개된 산출물의 품질은 사용자가 사용평을 남기면서 자연스러운 평가체제가 만들어 진다. 이를 통해 연구자가 자발적으로 협업에 참여하도록 하고, 소프트웨어 산출물의 품질 기대 수준을 높여서 연구과정 내에서 협업의 필요성이 제기되도록 유도하고, 전문가간에 정보의 교환이 쉽게 이루어 지는 환경을 조성할 수 있다.

[그림 4-1] 문제 대응 방안



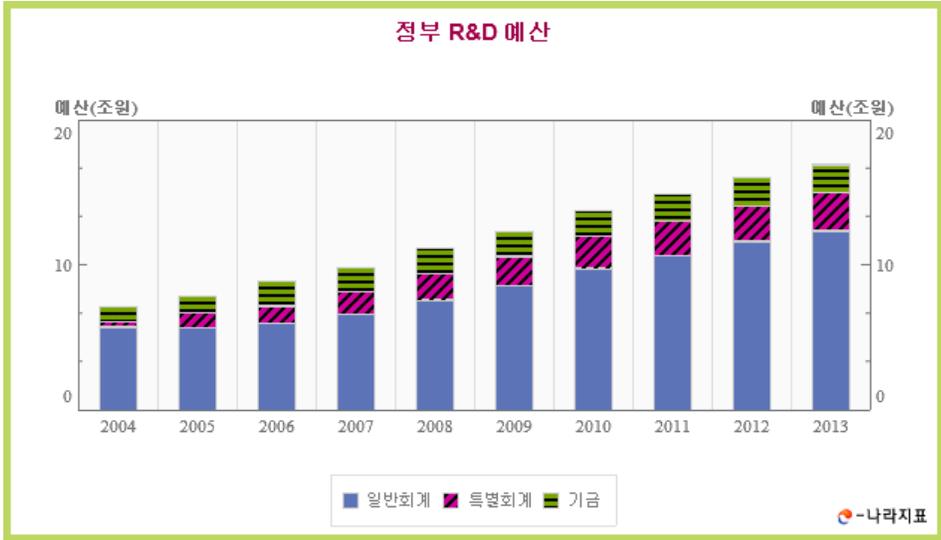
협업을 활성화하기 위해 연구 과정에서 생산된 소프트웨어와 데이터를 공개하는 정책을 추진하기 전에 검토해야 할 내용이 연구개발 결과의 사업화 현황이다. 연구결과를 사업화하는 것은 결과를 공개하는 것과 상충될 가능성이 있기 때문에 현황을 살펴보고 정책의 방향과 범위를 정하는 것이 순서다.

제3절 국가 연구개발의 사업화 현황 조사

우리나라의 한 해 국가 연구개발 예산의 규모는 꾸준히 늘어왔으며 2012년에는 16조 원, 2013년에는 16.9조원을 투자했고, 2014년에는 17.7조원이 투자될 것으로 예상되고 있다.³⁹⁾ (미래창조과학부, 2014b년도 정부연구개발투자 방향 및 기준(안), 2013b; 미래창조과학부, 2013년도 국가연구개발사업 조사·분석 결과(안), 2014b)

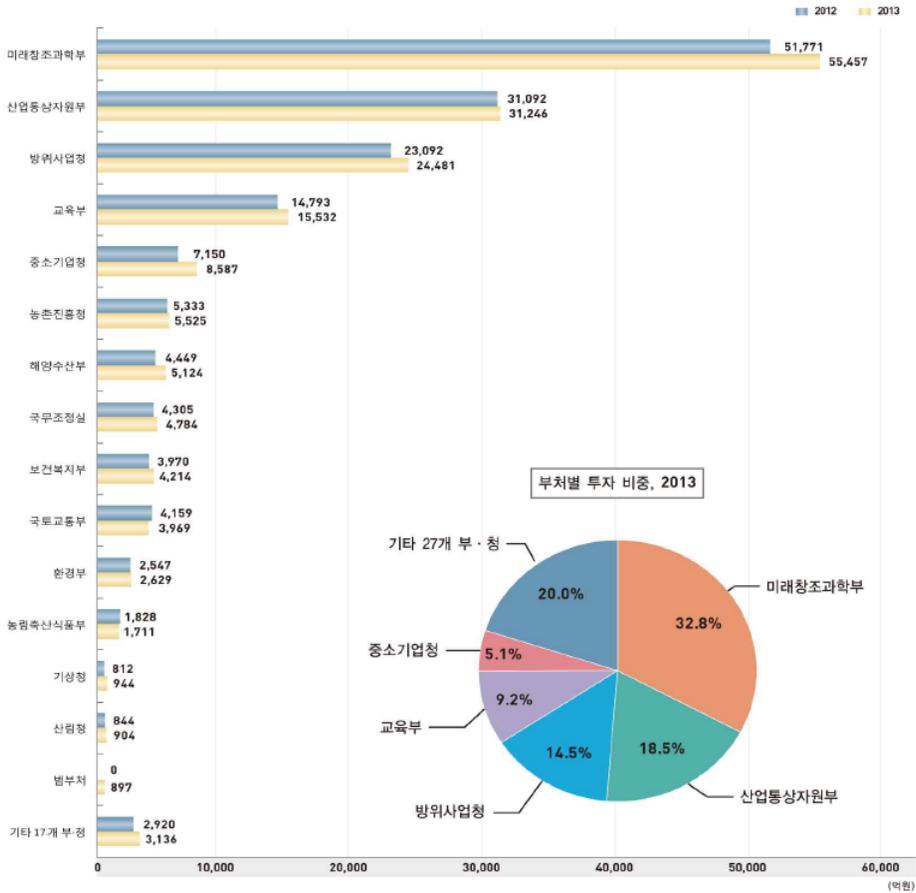
39) <http://www.index.go.kr>

[그림 4-2] 국가 연구개발 예산



자료: www.index.go.kr

[그림 4-3] 부처별 국가연구개발사업 투자 추이, 2012-2013



자료: 미래창조과학부, 2013년도 국가연구개발사업 조사·분석 결과(안), 2014b

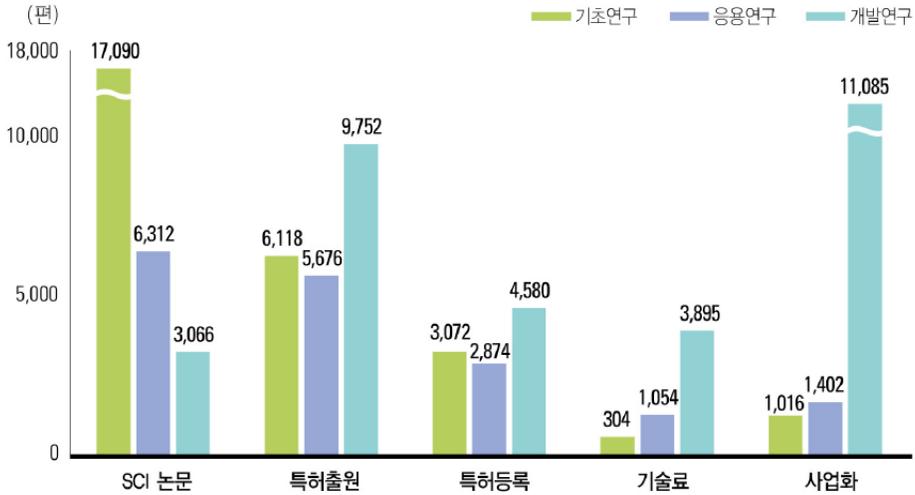
2012년도 국가연구개발사업 성과분석 보고서에 따르면 2012년 사업화 건수는 전해에 비해 세 배 가까이 늘었다.⁴⁰⁾ (2.84배) ([그림 4-5] 연도별 사업화 건수) 이것을 보면 연구개발 결과의 사업화가 활발하게 일어나고 있다고 볼 수도 있다. 제2차 연구성과 관리·활용 기본계획이 성과를 내기 시작했다고 해석할 수도 있다.⁴¹⁾ 이것을 좀 더 확인

40) 미래창조과학부·한국과학기술기획평가원, 2012년도 국가연구개발사업 성과분석 보고서, 2013

41) 국가과학기술심의회, 제2차 연구성과 관리·활용 기본계획(안), 2011

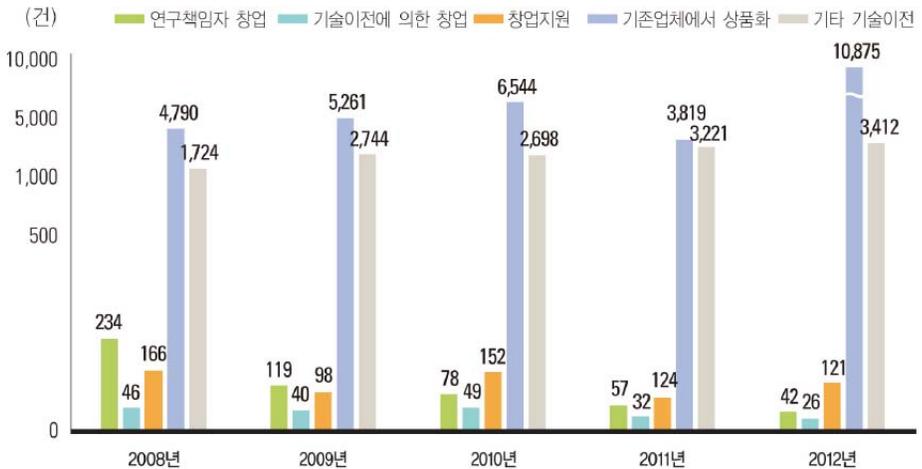
하기 위해서는 실제 기술료 징수액을 보아야 한다.

[그림 4-4] 2012년 항목별 연구성과 현황



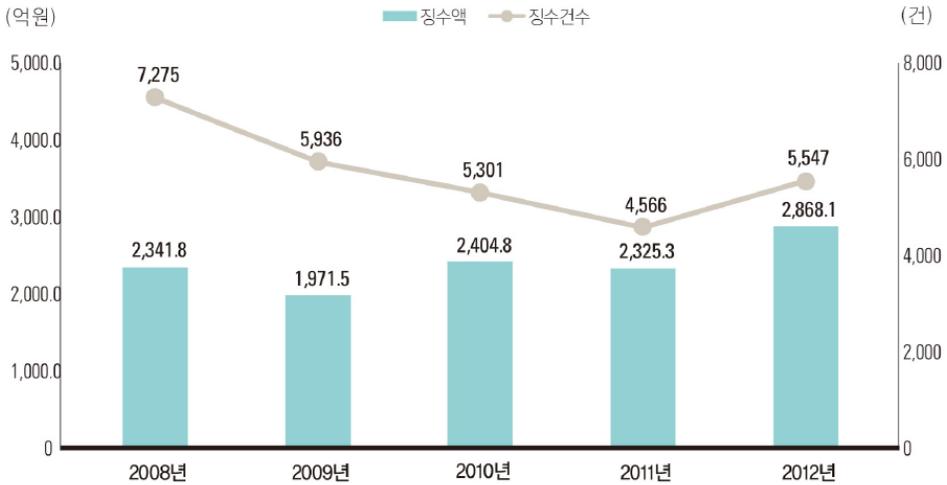
자료: 미래창조과학부 · 한국과학기술기획평가원, 2012년도 국가연구개발사업 성과분석 보고서, 2013

[그림 4-5] 연도별 사업화 건수



자료: 상동

[그림 4-6] 국가연구개발사업의 기술료 징수액과 징수건수



자료: 상동

[그림 4-6] 국가연구개발사업의 기술료 징수액과 징수건수를 보면 2012년에 기술료 징수액과 징수건수가 늘어나긴 했지만 23% 늘어난데 그쳐서 사업화 건수의 증가에는 크게 못 미친다. 물론 사업화 건수와 기술료 징수에는 시차가 있다. 2012년에 보고된 사업화 건수가 기술료 징수액에 반영되려면 1, 2년정도 시간이 필요할 수 있다. 이 결과가 성공적인 매출로 이어져 기술료 징수액에 반영되는지는 추적조사를 통해 확인하게 될 것이다.

한편 국회예산정책처에서 발행한 국가 연구개발 투자의 성과 측정 방법 연구 보고서에서는 우리나라의 사업화 성공률이 선진국에 비해 지나치게 낮다고 지적하고 있다.⁴²⁾ 연구개발의 기술적 성공률은 평균 90%를 상회하는데 사업화 성공률은 약 20% 수준이며, 대학과 연구소는 4.4%에 불과한 것으로 지적하고 있다. 평균 20%라는 것은 선진국인 영국의 70.7%, 미국의 69.3%, 일본의 54.1%에 비해 지나치게 낮은 비율이다. 또한 미래창조과학부에서 국가과학기술심의회에 보고한 내용에 따르면 “공공연은 19만건 이상의 기술을 보유하고 있으나, 아직까지 15.4만건 이상의 R&D결과물이 사업화되지 못하

42) 국회예산정책처, 국가 연구개발 투자의 성과 측정 방법 연구, 2013

고 휴면상태(‘12년)’이다. (미래창조과학부, 10호 제5차 기술이전 및 사업화 촉진계획 (안), 2014a)

국가 연구개발 사업의 성과지표를 보면 사업화가 잘 안 되는 이유의 한 측면을 볼 수 있다. 미래창조과학부와 한국연구재단이 발간한 2013 미래창조과학부 주요 연구개발사업 성과분석보고서를 보면 연구결과의 성과지표는 다음과 같다.(미래창조과학부·한국연구재단, 2013 미래창조과학부 주요 연구개발사업 성과분석보고서, 2013)

- 논문: 학술지, 학술대회, 저역서
- 특허
- 기술확산(기술이전, 기술지도, 기술평가), 사업화, 기술실시계약
- 인력활용 및 양성
- 기타: 국제학술대회 초청강연, 공동/위탁 연구추진

이 중 기술확산, 사업화, 기술실시계약에 대한 성과지표가 <표 4-2> 기술확산 및 사업화 성과 지표이다.

<표 4-2> 기술확산 및 사업화 성과 지표

2012년 성과	기술확산			사업화		2012년 계속과제
	기술이전	기술지도	기술평가	사업화 완료	사업화 추진중	
기초연구사업	2	655	0	7	46	5258
원천기술개발사업	0	61	0	3	45	1322
원자력기술개발사업	0	181	0	0	15	574
거대과학사업	0	11	0	N/A	N/A	141
과학기술국제화	0	115	0	N/A	N/A	367

자료: 미래창조과학부·한국연구재단, 2013 미래창조과학부 주요 연구개발사업 성과분석보고서, 2013

이 자료를 보면 기술 확산을 위한 활동의 대부분은 기술지도이다. 기술지도는 컨설팅과 강연, 세미나 등의 활동이 포함된다. 컨설팅, 강연, 세미나는 기술의 확산을 위해 중요한 활동이기는 하지만 사업화에 직접적으로 영향을 주기는 어렵다.

정부에서는 2013년부터 연구개발의 성과평가 개선 대책을 꾸준히 추진하면서 성과를 기대하고 있다. 이런 노력은 정부가 당연히 해야 하는 활동이고 일시적이 아니라 꾸준히 집요하게 추진해야 한다. 반면 지금까지 살펴 본 지표에 의하면 성과에 대해서는 아직 판단이 이른 상태이다. 따라서 연구결과를 사업화와 연결하기 위해서라는 이유로 굳이 연구결과를 적극적으로 공개하는 정책을 펴지 못 할 이유가 없다. 오히려 연구 결과를 적극적으로 공개하고 공유함으로써 다른 연구자나 기업체에 의해 사업화가 활성화될 수도 있다. 정부에서는 연구결과를 사업화로 연결시키려는 노력과 함께 연구결과를 공개하는 것을 장려하는 정책을 병행하여 공유생태계를 활성화할 필요가 있다.

제4절 현행 연구결과 공개 제도

현행 제도상으로 국가연구개발 결과는 국가과학기술지식정보서비스(National Science & Technology Information Service)에 등록하여 과제 정보를 공개하도록 되어 있다. 공식산출물은 저작권위원회에 연구산출물로 등록한다. 또한 소프트웨어 기술 및 제품은 소프트웨어공학센터에서 운영하는 SW자산뱅크에 등록하여 다른 사람에게 정보를 제공할 수 있게 되어 있다.

그러나 NTIS는 통계 위주의 데이터 제공 서비스로 구성되어 개별 과제에 대한 정보가 불충분하다. SW자산뱅크의 정보도 그림에서 보는 바와 같이 연구계획서나 보고서의 일부 내용을 보여주는 수준이며 소프트웨어에 대한 자세한 설명과 기능은 부족하다. 무엇보다 모든 내용이 공식산출물을 대상으로 하고 있기 때문에 연구 진행 중에 개발되는 도구나 유틸리티 등의 부수적 산출물에 대해서는 정보를 얻을 방법이 없다.

소프트웨어는 국가 연구개발 모든 분야에 걸쳐서 활용되고 있기 때문에 오히려 부수적 소프트웨어 산출물이 다른 연구자에게 도움이 될 여지가 많다. 생물학에서 유전자 분석을 위한 매칭 유틸리티나 데이터 시각화 도구, 물리학 연구의 시뮬레이션을 도와주

는 데이터 처리 유틸리티와 네트워크 프로그램 등 연구의 직접적 목표가 아니면서도 다른 연구자에게 도움이 되는 부수적 소프트웨어 산출물은 종류와 범위가 다양하다.

[그림 4-7] 소프트웨어 자산 بانک 검색 화면

The screenshot shows the 'SOFTWARE ASSET BANK' interface. At the top, there's a search bar and navigation options like '관심정보 등록', '인쇄하기', and '다운로드'. The main content area is titled 'AUTOSAR 기반 차량제어기 SW플랫폼 및 개발도구 상용화 개발'. It includes a central diagram with a core 'R3.1 개발도구 경쟁력 요소 (핵심)' surrounded by several key areas: '도구 사용 편의성 제공을 위한 Wizard 기능', '통합 도구 체인', 'Dafacto 진장 표준데이터 포맷 변환', and '진행상황 Progress 뷰'. Each area has a list of bullet points describing its features and benefits. A right sidebar offers additional actions like '요약보기', '분류체계 상세보기', '연관기술 상세보기', '첨부파일 상세보기', and '문의하기 바로가기'. At the bottom, the '구성 기술' section lists 'AUTOSAR R3.1 플랫폼 (통신 및 시스템 기능)' and 'Network Management Interface, Communication Manager'.

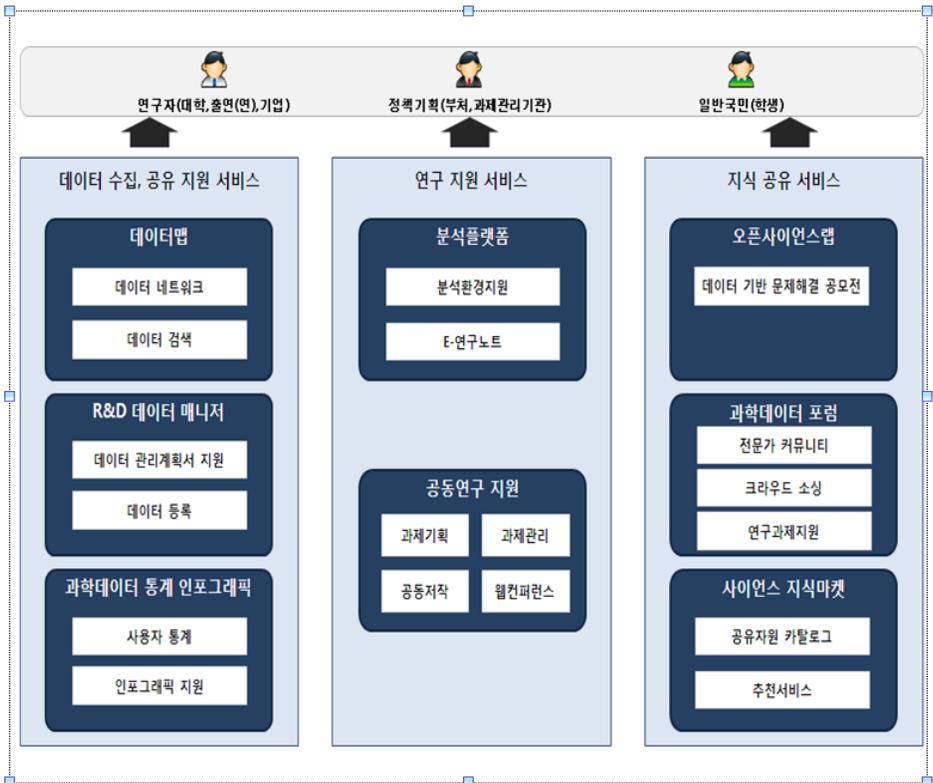
자료: <https://www.swbank.kr/>

현재의 형식적이고 불편한 연구결과 공유 제도를 개선하여 공식, 비공식 소프트웨어 산출물을 자유롭게 공개하고 공유할 수 있는 기반 환경과 연구 문화를 제공하는 것이 과학기술연구에서 소프트웨어의 품질을 높일 수 있는 기본적인 요건이다.

제5절 과학기술 빅데이터 공동 활용 종합계획

연구개발 결과의 공개를 추진하는 대상은 과학기술 연구 데이터와 연구과정에서 생산된 소프트웨어가 된다. 이 중 연구 데이터는 이미 미래창조과학부에서 제도화하여 추진하고 있다.(미래창조과학부, 과학기술 빅데이터 활용 종합계획, 2013.11.19) 이 계획에 따르면 과학기술연구에서 산출된 데이터를 수집하고 공동활용 플랫폼을 구축하여 연구개발 품질을 높이며 공개·개방 문화를 활성화하는 것으로 되어 있다.

[그림 4-8] 과학기술 빅데이터 목표 개념도



자료: 미래창조과학부, 과학기술 빅데이터 활용 종합계획(안), 2013c

이 제도는 본 연구에서 추진하는 소프트웨어 공개 정책과 완전히 같은 목표를 가지는 제도이며 과학데이터에 집중하여 제도화하고 있으므로 이 방안을 활용하는 것이 효과적이다.

제6절 고려 사항

1. 기본 원칙

본 연구에서 추진하는 방안은 공개를 장려하는 연구 환경 구현이 목적이다. 소프트웨어를 공개한 후에 다른 연구자에 의해 널리 활용되는 산출물의 가치를 인정하도록 한다. 공개후의 품질 개선, 공동 개발, 기술 전파 등의 활동에 대한 중요성도 인정하여 소프트웨어 공개가 단순한 절차에 그치지 않도록 하는 것이 중요하다.

공개 범위는 연구자가 판단하여 정하도록 한다. 공개를 장려한다고 모든 것을 공개소프트웨어로 만들라는 것은 오히려 역효과가 날 수 있다. 연구자의 분야나 연구 특성에 따른 차이가 분명히 존재하고 공개가 어려운 내용이 있을 수 있다. 공개 소프트웨어 내용은 연구자가 판단하도록 하되 적극적인 활동을 하는 연구자에게 인센티브가 주어지는 체계를 구축해야 한다.

사업화에 실패한 산출물의 공개를 의무화하는 것도 중요한 원칙이다. 급변하는 소프트웨어 환경의 특성상 일단 사업화가 실패하면 같은 코드로 다시 사업화를 시도하기 어렵다. 아이디어는 새로 조명을 받을 수 있더라도 코드는 상당 부분 새로 개발해야 한다. 심지어 사업화에 성공한다고 하더라도 연구 단계에서 개발한 코드를 새로 작성하는 경우가 흔하다. 연구에서 개념이 옳은 것을 검증했으니 사업화를 위해 제대로 동작하는 상용 품질의 코드로 개선하는 것이다. 그러므로 일정 기간 유예를 주어 사업화를 시도했으나 실패한 소프트웨어는 공개를 통해 다른 사람이 활용할 기회를 주는 것이 국가연구개발제도 취지에 합당하다.

기본 원칙 외에 제도를 추진하면서 과도기에 고려할 사항이 있다. 소프트웨어 공개를 위해서는 적지 않은 노력이 들게 된다. 이미 진행 중인 연구과제에서는 이 부분에 대한 고려가 연구 계획에 없었기 때문에 과도기 동안은 공개를 위한 추가 작업에 대한 인정

과 지원을 해 줘야 한다. 단순하게는 소프트웨어를 간단한 설명과 함께 제출하도록 하고 운영 조직에서 정리하여 공개하는 방법이 있다. 운영 조직도 공개되는 소프트웨어의 다양성을 파악하고 관리 지침을 만들기 위해서는 있는 그대로의 소프트웨어를 수집하여 분석할 필요가 있다. 과도기를 이용하여 지침을 준비하도록 일단 무조건 수집하는 것이 좋을 것이다.

제도가 시행된 후에는 과제 제안 시 제안요청서에 공개 규정을 명시하여 연구자가 계획 수립에 반영하도록 유도한다.

2. 공개 대상

과학기술 연구 데이터는 앞서 제시한 과학기술 빅데이터 공동활용 종합계획을 따르기로 하고 소프트웨어에 대해서 설명한다. 연구과제에서 개발되는 공식, 비공식 소프트웨어의 범위는 매우 다양할 것이고 각자의 사정도 천차만별일 것이므로 엄밀한 규칙을 정하는 것은 의미가 없다. 또한 본 제도의 취지도 소프트웨어의 공유를 통해 품질에 대한 의식을 높이고 반복을 줄이는 것이기 때문에 연구자가 규정에 따라 최소한의 참여를 하게 되는 것보다 적극적으로 결과를 공유하는 연구 환경이 만들어지는 것이 더 바람직하다.

소프트웨어는 공식산출물일 수도 있고 연구에 사용되는 도구이거나 부수적 산출물일 수도 있다. 공식산출물은 과제의 목적이 소프트웨어 개발인 경우이다. 이 경우에 소프트웨어 개발 과제이면서 즉시 사업화 하지 않는 과제는 공개하도록 한다. 부수적 산출물은 예를 들어 시뮬레이션 프로그램이나 도구, 상용 프로그램의 커스텀 프로그램, 데이터를 처리하는 유틸리티, 데이터를 시각화하는 프로그램 등이 있을 수 있다.

예상컨데 공식 산출물보다 부수적 소프트웨어가 활발히 이용될 것이다. 공식 산출물은 보통 크기가 크고 명확한 목적이 있기 때문에 범용성이 떨어진다. 오히려 연구 과정을 통해 발생한 부수적 소프트웨어가 도구의 성격을 가지고 여러 사람에게 도움이 될 것이다. 그러므로 부수적 소프트웨어의 비중을 낮게 간주해서는 안 된다.

제5장 소프트웨어 산출물 공개 제도

제1절 제도 추진 근거

국가연구개발 과제에서 생산된 소프트웨어를 공개하여 함께 사용하도록 한다는 원칙에는 이견이 있을 수 없다. 국가연구개발과제는 국가 전체의 경쟁력을 높이기 위해 투자하는 사업이며 결과를 공개하여 함께 활용하는 것은 원 취지에 잘 맞는다. 사업화를 전제로 하는 과제의 문제도 일정기간 동안 예외로 처리하면 될 것이다. 과거의 실적을 보면 사업화 성공률과 기여도가 낮기 때문에 공개 정책이 사업화에 미치는 영향은 미미할 것이다. 현재 정부에서 추진 중인 성과평과 개선 대책은 연구결과를 사업화까지 연계하도록 지원하는 대책이며, 이런 단기적 성과를 기대하는 정책과 병행하여 연구 산출물을 공개하고 함께 이용하여 생태계를 활성화하는 거시적 정책이 필요하다.

미국은 정보자유법(FOIA: Freedom of Information Act)에 의해 연방정부에서 지원한 과제에서 개발된 소프트웨어의 소스코드와 데이터를 공개하여 다른 사람이 활용할 수 있도록 공유하는 것을 권장하고 있다.(Fischer, 2013) 구글이나 wikipedia에서 검색해 보면 얼마나 많은 소프트웨어가 공개되고 있는지 볼 수 있다.⁴³⁾ 영국도 정부투자로 과학기술연구자가 소프트웨어 소스를 공개하고 공동으로 연구할 수 있는 플랫폼을 개발하여 운영하고 있다.⁴⁴⁾

이런 결정의 배경에는 연구과정에서 나오는 소프트웨어나 데이터는 특정 연구 집단이 배타적으로 소유하는 것보다 연구자간의 공개를 통해 공동의 노력으로 꾸준히 개선하고 검증하는 것이 필요하다는 생각이 깔려있다.

연구개발과정에서 생성되는 공식 산출물이나 부수적 산출물인 소프트웨어를 공개소프트웨어로 공개하여 얻을 수 있는 이점은 여러 가지가 있다

43) http://en.wikipedia.org/wiki/List_of_open-source_bioinformatics_software

44) <http://ccpforge.cse.rl.ac.uk/gf/>

□ 소프트웨어와 데이터의 공개를 통한 연구 결과의 재 활용 촉진

연구 결과가 연구자의 책상에서 사장되지 않고 다른 연구자에게 도움을 줄 수 있다. 이것이 직접적으로 다른 연구자의 연구에 사용될 수 없는 수준이라 하더라도 이 분야를 공부하려는 연구자에게 교재가 될 수 있다. 논문을 읽는 것보다 그 논문을 구현한 소프트웨어를 보는 것이 내용에 대해 더 자세히 이해할 수 있다. 소프트웨어에는 문서로 정리한 것보다 상세한 정보가 들어있기 때문이다. 이것은 장기적으로 국내 연구 수준을 높이는 기반 환경의 역할을 할 수 있을 것이다.

□ 공개에 적극적인 연구자를 위한 기반 환경 및 제도 제공

산출물의 공개에 적극적인 연구자는 주변에서 많이 볼 수 있다. 현행 제도가 공개를 불편하게 만들어서 꺼려하는 것뿐이지 편리한 환경이 만들어지면 많은 연구자가 산출물을 공개하고 다른 연구자와 공유하게 될 것이다.

지금도 연구자 나름대로 공개를 위한 활동을 하는 분이 많다. 웹서비스 형태로 공개하여 다른 사람이 연구결과를 시험해 볼 수 있도록 하는 경우도 있고, 실험용 소프트웨어와 데이터를 공개사이트에 올려서 검증도 받으면서 활용을 촉진하는 경우도 있다.

이런 활동이 일반적인 연구 활동으로 간주되지 않고 개인적인 활동에서 머무는 것은 우리나라의 연구제도가 공개를 장려하지 않도록 추진되고 있기 때문이다. 이번 기회에 공개소프트웨어 문화를 확산하며 법제도의 정비도 함께 추진하면 개방적 태도를 가진 연구자에게 더 적극적 참여를 북돋울 수 있는 지원체제를 갖추게 될 것이다.

□ 국가적인 반복 개발 비용 절감

공개되는 소프트웨어는 커다란 시스템이 아니라 실험을 위한 유틸리티나 도구가 포함된다. 이런 것을 각 연구자가 새로 개발하여 쓸 것이 아니라 공개소프트웨어를 활용한다면 반복개발을 피할 수 있어 국가 전체적으로는 개발 비용을 절감할 수 있다. 더구나 공개소프트웨어를 사용할 뿐 아니라 개선하는데 참여한다면 공개된 소프트웨어의 품질

을 높이는 데에도 기여할 수 있고 그 영향은 모든 사용자에게 미칠 것이다.

만약 연구자의 입맛에 딱 맞는 소프트웨어가 없다고 하더라도 유사한 코드를 참조하여 자신이 필요한 기능으로 개선하여 전체적인 연구 시간을 효율적으로 쓸 수 있다.

□ 피어리뷰에 따른 소프트웨어와 데이터의 품질 검증 및 개선

소프트웨어가 공개됨으로써 다른 연구원으로부터 연구결과에 대한 객관적인 평가를 받을 수 있다. 소프트웨어의 오류나 성능이 개선될 여지가 있는지 여부를 공개소프트웨어 사용자로부터 검증받을 수 있고 품질개선을 공동 작업으로 추진할 수도 있다. 소프트웨어는 최초 버전이 마지막 버전이 되는 사례는 거의 없다. 꾸준한 오류의 수정과 기능의 개선으로 좋은 소프트웨어가 만들어지고 많은 사람들이 사용하게 된다. 연구결과의 공개를 장려함으로써 이런 활동의 토대를 제공할 수 있다.

특히 다른 분야의 연구자로부터 의견을 들을 수 있는 기회가 된다. 공개된 소프트웨어가 과학연구에 활용되는 것이라도 이 소프트웨어를 보고 소프트웨어 전문가가 개선할 기회도 생기고 이 분야에 대한 이해를 높이는 역할도 기대할 수 있다. 과학자와 소프트웨어 전문가가 정보를 교류하는 것이 쉽지 않은 현실에서 공개된 정보를 통한 교류는 학제간 협력의 장으로 발전할 수 있다.

□ 연구 결과의 빠른 확산에 의한 국가적 연구 역량 제고

소프트웨어의 공개를 연구가 끝난 후가 아니라 진행 중에도 하도록 권장하여 연구 결과를 빠르게 확산하고 다른 연구자가 활용할 수 있게 된다. 과제의 핵심이 되는 부분은 연구가 끝날 때까지 공개할 수 없는 경우도 있겠으나 부수적 산출물인 경우에는 연구를 진행하며 공개소프트웨어로 공개할 수 있다. 연구진행중에도 다른 연구자로부터 문제점이나 개선사항에 대한 의견을 들을 수 있을 뿐 아니라 다른 연구자가 개선한 내용을 사용하게 될 수도 있다. 이것이 국가 전체적으로는 연구의 수준을 높이는 결과를 가져올 것이다.

연구 결과를 공개소프트웨어로 만들자는 것이 연구자의 권리를 빼앗자는 의미는 아니다. 공개되는 내용이 학술적인 가치가 있다면 논문으로 출간한 후에 공개하면 된다. 해외 학술지의 경우에 논문을 제출했을 때 실험에 사용한 소프트웨어와 데이터를 공개하도록 요구하는 것이 특별한 사례는 아니다. 이론적 논문이 아니라 컴퓨터 시뮬레이션이나 컴퓨터 계산을 통해 결과를 유도했을 경우에는 소프트웨어와 데이터를 공개하여 논지를 검증받는 것이 자연스러운 절차이다.

사업적 가치가 있는 경우에는 특허 등록을 하여 권익을 보호할 수 있다. 또한 핵심적인 부분을 보호할 필요가 있을 때는 그 부분을 숨기고 공개할 수도 있다. 이것은 이미 많은 스타트업 회사가 하는 것과 같은 방법이다. 핵심 매개변수의 값을 단순화하여 공개하거나 핵심 기능을 간단한 기능으로 대체한 소스를 공개하는 등 여러 방법으로 사업화를 위한 대비를 할 수 있다.

공개된 소프트웨어에 대한 저작권을 주장하고 싶은 경우에는 그에 맞는 라이선싱을 부과하여 보호할 수 있다. 현행 규정상 공식산출물은 주관기관의 소유이기 때문이다.(미래창조과학부, 정보통신·방송 연구개발 관리규정, 2014f; 교육부, 교육부 소관 이공분야 연구개발사업 처리규정, 2013) 미래창조과학부의 경우에 2014년초에 관련 규정이 개정되어 과제를 시작할 때 장관의 승인을 받으면 산출물을 공개 소프트웨어로 만들 수 있는 법적 근거는 마련되었다. (미래창조과학부, 정보통신·방송 연구개발 관리규정 제36조 11항, 제40조)

제2절 제도의 개요

소프트웨어 산출물 공개 제도는 제안 단계에서 연구를 통해 생산되어 공개할 소프트웨어를 연구책임자가 과제 제안서에 기입하여 제안하도록 한다. 공식 산출물과 비공식 산출물을 모두 고려하여 과제에서 생산되어 공개 소프트웨어로 등록할 모든 대상을 기입한다. 소프트웨어를 공개하려면 자원과 노력이 들어가게 되므로 연구책임자는 공개에 필요한 인력, 비용, 기간을 고려하여 제안서를 작성해야 한다.

제안 심사에서는 제안서 내용과 더불어 공개소프트웨어의 내용을 고려하여 과제를 선정한다. 이 단계에서 공개 소프트웨어에 적극적인 연구책임자가 가점을 얻도록 하거나

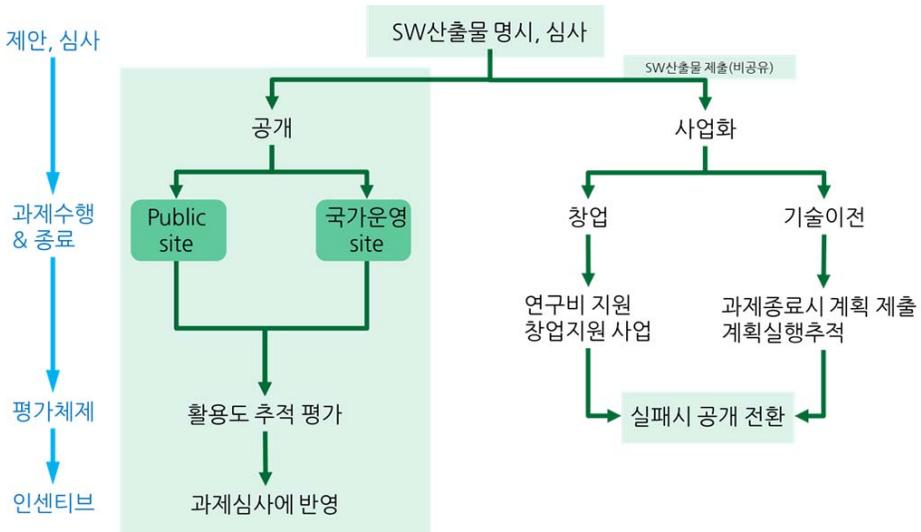
적극적이지 않은 쪽에 감점을 하거나 하여 공개가 장려되도록 한다.

과제 종료 시에는 제안대로 이행했는지 확인하여 과제평가에 반영한다. 소프트웨어의 공개는 반드시 과제 종료할 때 할 필요는 없고 연구를 수행하면서 활용하는 것이 더 장려된다.

과제 종료 이후에는 공개된 소프트웨어의 활용도를 추적 평가하여 사용이 활발한 소프트웨어에 대한 보상을 한다.

사업화를 목표로 하는 경우에는 과제 종료 후 소프트웨어를 제출하지만 공개는 유예한다. 유예 기간이 지나고 사업화가 실패한 경우에는 보관된 소프트웨어를 공개로 전환한다.

[그림 5-1] 과제 수행 단계별 개선 방안



제3절 과제 제안 단계

연구책임자는 정부과제 제안서에 공개할 소프트웨어를 명기하여 제출한다.

공개 대상 소프트웨어의 대상은 다음과 같다.

○ 공식 산출물

- 애플리케이션, 플랫폼, 프레임워크
- 공식 산출물의 임베디드 애플리케이션/소프트웨어

○ 연구도구 및 부수적 소프트웨어

- 공식 산출물이 아니더라도 연구 과정 중에 개발하게 되는 각종 소프트웨어 도구 및 요소
- 시뮬레이션 프로그램/툴, 상용프로그램의 커스텀 프로그램
- 데이터 프로세싱 & 유틸리티
- 시각화 & 사용자인터페이스 프로그램
- 기타 연구개발 과정에서 개발되는 소프트웨어

정책의 취지가 공개된 소프트웨어의 숫자를 늘리는데 있는 것이 아니라 연구 결과의 공개를 통해 공개소프트웨어를 활성화하는 것이므로 공개될 소프트웨어를 판단하는 것은 연구책임자의 재량이다. 책임자는 자신의 연구 중 공개해도 될 소프트웨어를 직접 판단하면 된다. 현실적으로도 가장 잘 이해하고 있는 사람은 연구책임자라고 할 수 있다.

공개 범위는 반드시 제안서에 명시한 내용에 그칠 필요가 없다. 연구를 진행하며 공개소프트웨어로 만들 수 있는 프로그램이 나오면 공개할 수 있도록 한다.

연구책임자는 공개에 필요한 인력, 비용, 기간을 고려하여 제안서를 작성해야 한다. 과제 관리부서에서는 연구원이 공개를 고려한 연구 계획 수립을 할 수 있도록 제안요청서에 명확히 공지하여야 한다.

명시된 내용은 제안서 심사 단계에서 고려하여 선정하도록 한다. 연구결과를 적극적으로 공개하여 활용하려는 연구책임자를 우대하여 공개 소프트웨어 문화가 활성화되도록 한다.

시간이 지나면서 이전에 공개한 소프트웨어에 대한 사용자의 평가, 다운로드와 커멘트로 나타나는 활용도, 사용자 의견과 소프트웨어의 개정 상황 등 공개된 소프트웨어의 완성도를 객관적으로 평가할 수 있는 자료가 쌓이면 이 자료를 제안서 심사단계에서 평가요소로 포함하여 심사할 수 있다.

제안서에 소프트웨어 공개를 명시한 경우에는 늦어도 과제 수행 및 완료 단계에서 개발된 소프트웨어를 공개해야 한다.

제4절 소프트웨어 공개 단계

1. 공개 소프트웨어의 구성

□ 제3자가 공개된 내용을 참고하여 빌드하고 사용할 수 있도록 하는 일체의 구성품을 포함한다

공개 소프트웨어 이용자가 이 소스를 찾고 활용할 수 있도록 하기 위해서는 단순히 소스코드만 공개해서는 안 된다. 소스코드의 코딩스타일이나 모듈화가 어느 수준은 되어 있는 것이 기본이겠지만 어찌 보면 이것이 가장 중요한 것은 아니다. 오히려 이용자가 이 소스를 다운받기 전에 어떤 일을 하는 것인지 정확히 알 수 있고, 다운로드해서 실행코드를 만드는 절차를 명확히 규정하고, 코드의 시험을 쉽게 해 볼 수 있는 자료를 제공하는 것이 더 중요하다. 공개 소프트웨어를 사용하려는 사람이 너무 많은 시간을 들여 코드를 시험해야 한다면 차라리 다른 소프트웨어를 찾는 것이 더 효율적이기 때문이다. 인터넷에 공개소프트웨어를 위한 가이드라인은 여러 가지가 있으며 기본적인 내용은 N. Zakas, Starting an Open Source Project(2013)을 참고하면 된다.

가장 기본적인 내용은 이 소프트웨어가 하는 기능과 빌드 방법에 대한 간단한 소개글이다. 입력은 어떻게 받고 결과는 어떻게 나오는지를 명확히 표현해야 한다. 빌드를 위한 운영체제, 컴파일러 종류 및 버전과 환경, 실행 방법, 기타 환경 설정 등 사용자가 막힘없이 소프트웨어를 빌드하고 실행해 볼 수 있는 명확한 지침을 명시한다.

실행을 위한 샘플 데이터와 샘플 프로그램을 제공하면 더 많은 도움을 줄 수 있다. 공개된 프로그램이 독립적으로 실행되는 것이면 샘플 데이터를 제공하는 것이 큰 도움이 된다. 샘플 데이터를 보고 문서에 빠진 입력 조건이나 제약에 대해서 사용자가 이해할 수 있고 데이터를 부분적으로 수정해 가면서 프로그램을 이해할 수 있다. 공개된 코드가 다른 프로그램에서 호출하는 라이브러리 성격이면 용례를 보여주는 샘플 프로그램을 적어도 하나 이상 제공해야 한다. 다른 사용자는 샘플 프로그램을 참조해서 이 라이브러리를 사용하는 법을 이해하게 된다.

소프트웨어의 일부에 연구자가 핵심 지식을 포함하고 있어서 공개를 꺼리는 경우에는 그 부분을 일반적인 코드로 대체하여 공개하는 것이 가능하다.

핵심 알고리즘이 구현된 부분을 공개하지 않으려면 그 부분을 대신하는 코드로 변경하여 공개할 수 있다. 대체되는 코드는 기능을 대리하는 코드이기 때문에 다른 부분과의 인터페이스 등은 원 코드와 정확하게 일치해야 한다. 정부가 지원한 연구의 결과로서가 아니라 그 이전부터 쌓아 온 기술이나 노하우를 공개하고 싶지 않은 경우에도 연구자는 소프트웨어의 특정 부분을 보호하고 공개할 수 있다.

공개에 필요한 구성은 프로그램마다 다를 수 있으나 일반적인 구성은 다음과 같다.

- 소프트웨어 프로그램
- 샘플 데이터와 샘플 프로그램
- 관련 문서

원칙적으로 제3자가 공개된 내용을 참고하여 스스로 실행코드를 빌드하고 사용할 수 있게 하는 일체의 구성품을 제공한다.

공개된 전체 소스는 다른 사람이 컴파일하여 사용할 수 있어야 하기 때문에 보호할 부분을 모듈화하여 구성하는 것이 좋다. 이 모듈과의 인터페이스를 잘 정의하고 문서도 잘 정리하여 다른 사용자가 해당 부분을 직접 구현하여 전체 프로그램을 사용할 수 있어야 한다. 보호된 모듈에는 원래의 보호되는 코드 대신 샘플 코드를 대신 포함하여 사용자가 쉽게 전체 프로그램을 실행할 수 있도록 구성해야 한다.

□ 인터페이스 문서

인터페이스는 소프트웨어 개발에서 확장성과 재사용성을 높이기 위해 매우 중요한 정의이다. 공개된 소프트웨어의 인터페이스 정의를 정확히 하고 문서를 명확하게 작성해서 배포하는 것은 샘플 프로그램만큼 중요한 요소이다.

2. 기존 오픈소스 공유 서비스 이용

공개 소프트웨어를 활성화하는 것은 새로운 사이트를 만들어서 국내 활용 위주로 추진하는 것보다 기존의 공유 사이트를 이용하는 것이 적절하다. 일반 공유 사이트를 이용하여 공개한 경우에 연구자는 사이트에 올려서 공개한 소프트웨어의 링크를 제출하면 공개소프트웨어를 개발했다는 근거가 된다. 기존의 공유 사이트는 버전 관리 등을 비롯해서 오랫동안 개선된 좋은 기능을 제공한다. 가장 좋은 점은 전세계 사용자에게 공개된다는 것이다. 유명한 공유 서비스를 이용하면서 널리 알려질 기회도 얻게 되는 것이다.

오픈소스 저장소 서비스를 제공하는 사이트는 SourceForge, GitHub, RubyForge, Tigris.org, BountySource, Launchpad, BerliOS, JavaForge, GNU Savannah, Gitorious 등이 있다. 이 중 가장 널리 이용되는 공유 사이트인 깃허브(GitHub)와 소스포지(SourceForge)를 여기에 소개한다. 다른 사이트도 유사한 개념으로 이해하면 된다. 두 사이트의 활용에 대한 상세한 내용은 인터넷에서 찾아볼 수 있다.⁴⁵⁾

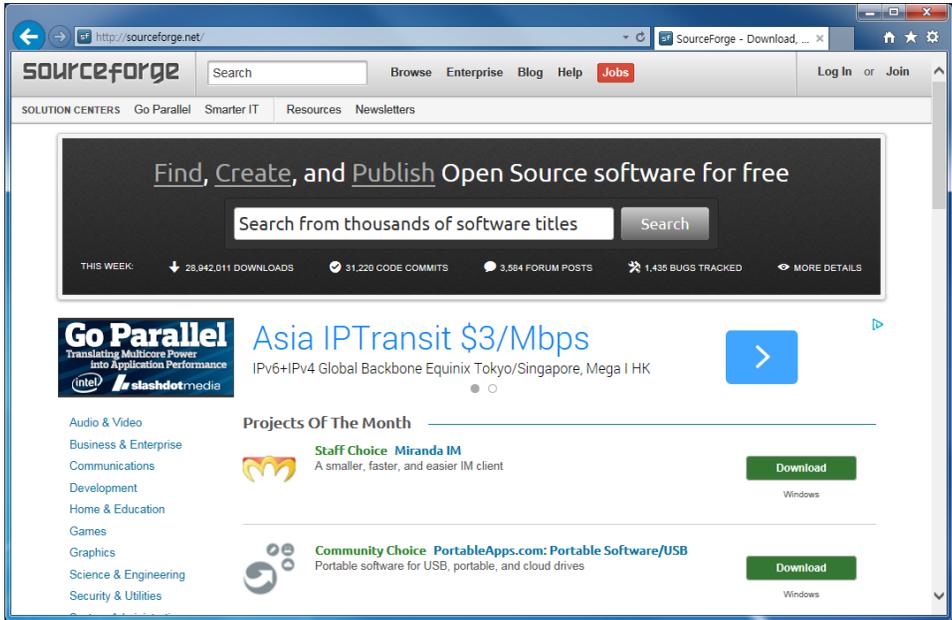
45) 소스포지는 <http://vulcan9.tistory.com/81> , <http://yesarang.tistory.com/100> ,

○ SourceForge

소스포지(SourceForge.net)는 소프트웨어 개발자들을 위한 오픈 소스 소프트웨어 개발관리를 위한 웹사이트이다. 중앙에 집중된 개발관리 시스템으로 소스 코드 저장소로서 동작한다. 현재 지크넷이 소스포지를 운영하고 있다. 버전관리 시스템인 Subversion(SVN)을 오픈소스 호스팅 기반으로 사용한다. 소스포지는 오픈 소스 소프트웨어 개발자들에게 호스팅 서비스와 개발 툴을 무료로 지원하여 개발자 커뮤니티에서 인기가 높으며 많은 오픈 소스 프로젝트가 이 사이트에서 호스팅 되고 있다. 2011년 7월 기준 프로젝트 수는 무려 300,000개 이상이고 등록 사용자 수는 200만명에 달한다.

<http://www.mangosystem.com/?p=343> 등 깃허브는 <http://blog.outsider.ne.kr/865> , <http://blog.hjf.pe.kr/37> , <http://riniblog.egloos.com/viewer/1024993> 등의 많은 안내문서가 인터넷에 올라와 있다.

[그림 5-2] 소스포지 서비스



자료: <http://sourceforge.net>

○ GitHub

깃허브는 깃(git)이라는 소프트웨어 버전관리 도구를 이용하는 소프트웨어를 위한 공동작업 서비스이다. 유상으로 사용할 수도 있으며 오픈소스인 경우에는 무상으로 사용할 수 있다. 2011년에는 가장 인기있는 오픈 소스 저장소로 꼽혔다. 깃허브에서는 버전관리와 소스 공유를 한꺼번에 처리할 수 있으며 부수적인 부담이 적다. GitHub는 영리적인 서비스와 오픈소스를 위한 무상 서비스를 모두 제공한다. 2011년의 조사에서는 가장 인기있는 오픈 소스 코드 저장소로 꼽혔다.

- 소스포지와 Google Code를 빠르게 앞서고 있는 오픈소스 프로젝트 근거지
- 단순한 소스 호스팅 뿐 아니라 개발자들끼리의 SNS로 기능하고 있음
- 버전관리시스템인 깃(Git)을 소스 호스팅 도구로 사용

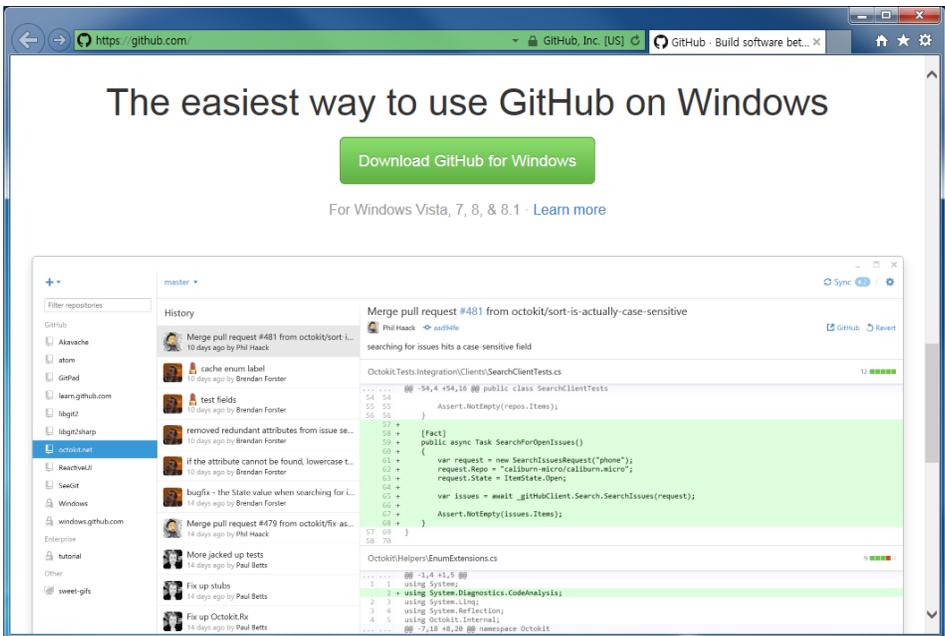
- 코드 생산자가 아니라 소비자로 활동하기에도 좋다

것은 2006년경 BitKeeper라는 리눅스 커널 개발에 쓰던 분산형 패치 도구에 대한 대안으로 리눅스 토발즈가 직접 개발한 분산형 소스 컨트롤(Source Control Management) 시스템이다.

- Offline으로 동작 가능하며 거의 모든 기능의 반응 속도가 기존 버전 관리 툴을 압도

- GitHub를 쓰기 위해 Git을 사용하는 사람이 늘고 있음 (GitHub가 활성화되었음을 의미)

[그림 5-3] 깃허브 사이트



자료: <https://github.com/>

<표 5-1> 오픈소스 저장소 기능 비교

기능	GitHub	Google Code	SourceForge
Code review	○	○	○
Bug tracking	○	○	○
Web hosting	○	×	○
Wiki	○	○	○
Translation system	×	×	×
Mailing List	×	×	○
Forum	×	×	○
Personal branch	○	○	○
Private branch	○	×	○
Announce	○	×	○
Build system	3rd-party	×	×
Team	○	×	○
Release Binaries	○	○	○

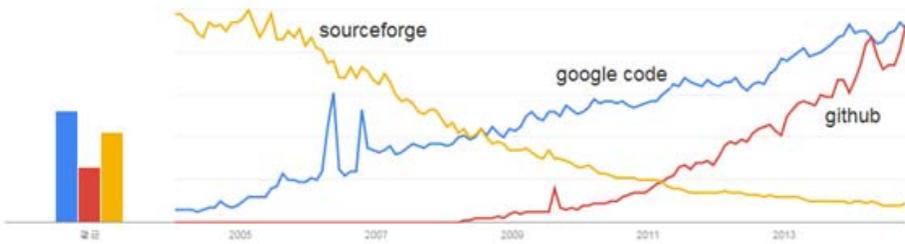
자료: http://en.wikipedia.org/wiki/Comparison_of_open-source_software_hosting_facilities

<표 5-2> 오픈소스 저장소 인기도 비교

Name	GitHub	Google Code	SourceForge
Users	6,700,000	Unknown	3,400,000+
Projects	1,100,000+	250,000+	324,000
Prominent projects	Ruby on Rails, jQuery, Moodle, node.js, Spring Framework, PHP	Inferno, Android, Chromium	LAME, MinGW, Poedit, 7-Zip, Audacity, EMule, FileZilla, phpMyAdmin
Alexa rank (수가 작을수록 더 유명)	134 (2014/08/22)	-	177 (2014/07/03)

자료: http://en.wikipedia.org/wiki/Comparison_of_open-source_software_hosting_facilities

[그림 5-4] 저장소 인기의 변화



자료: Google Trend, 검색어 기준

3. 국내 공개 사이트 운영

기존 서비스에 익숙하지 않거나 국내 위주의 공개를 원하는 경우에는 국내에서 운영되는 공개소프트웨어 사이트를 개발하고 여기에 올릴 수 있도록 운영한다. 공개 소프트웨어는 굳이 사이트를 따로 운영하는 것이 바람직하지 않지만 국가에서 관리하는 사이트를 운영할 필요성이 몇 가지 있다.

○ 간편한 등록

소프트웨어에 익숙지 않은 연구자는 버전 관리 등의 용어도 생소하고 공유 사이트도 낯설 수 있다. 이런 경우를 위해 쉽게 등록할 수 있는 국내 사이트가 필요하다.

○ 공개를 조건부로 보류

사업화 과제의 경우에 소프트웨어를 사이트에 등록은 하지만 공개는 사업화 여부에 따라 결정할 수 있도록 보류할 수 있어야 한다.

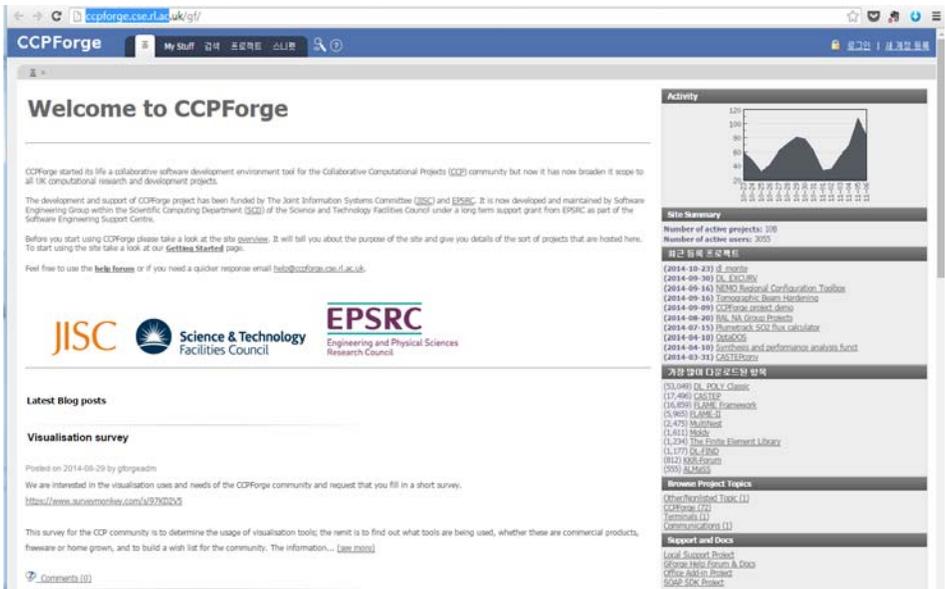
○ 사용자 제한

공개는 했으나 국내 사용자만을 대상으로 하고 싶은 경우에는 국내 공개 사이트를 이용하도록 한다. 중요한 점은 국내 사용자라는 것이 완전한 보안을 의미하는 것은 아니라는 점이다. 내용상 해외에 유출되면 안 되는 코드는 공개 소

프웨어로 등록하면 안 된다. 여기에서 사용자를 제한한다는 것은 일차적으로 사용자 등록 절차를 요구하여 무분별한 접근을 회피한다는 정도를 의미한다.

국내 공개사이트의 형태는 영국의 ccpforge 같은 모델이 될 것이다.⁴⁶⁾ 이 사이트는 영국 정부의 지원으로 개발하고 현재는 일반 공유 및 커뮤니티 플랫폼으로 활용되고 있다. 이 사이트가 일반 공유 사이트와 다른 점은 사용자 통제를 좀 더 강화한 것이다. 사이트 활동에 도움이 될 경력을 가진 연구자인지 확인하고 아이디를 발급해 주는 단계가 있다. 국내 사이트도 국내 연구자만 대상으로 할지 내국인 일반을 대상으로 할지 검토할 필요가 있지만 어느 정도 사용자에 대한 관리가 필요하다고 볼 수 있다.

[그림 5-5] CCPForge 공유사이트 메인 화면



자료: <http://ccpforge.cse.rl.ac.uk/gf/>

46) <http://ccpforge.cse.rl.ac.uk/gf/>

국내 사이트가 운영되어야 하기 때문에 개발 및 운영을 위한 조직이 필요하다. 이 조직은 단순히 사이트를 운영하는 조직이 아니라 공개소프트웨어의 지속적인 개선, 유지 보수, 활용 감시 및 분석을 활성화하는 미션을 가진 개발 조직으로 구성하는 것이 바람직하다.

또한 연구자의 커뮤니티 참여도 등의 데이터 축적을 할 수 있어 향후 과제 심사를 위한 정보 제공 소스로서의 역할을 할 수 있다.

4. 공개SW 라이선스

공개SW 라이선스에 대한 최신 정보는 정보통신산업진흥원(2014)에 요약되어 있다. 여기에서는 주요 라이선스의 개요를 소개한다.

□ GPL(GNU Public License) 2.0

- **GPL 2.0**은 현재 가장 많은 공개SW가 채택하고 있는 라이선스로 자유 소프트웨어재단(**Free Software Foundation, FSF**)에서 만들고 배포하였다. 공개SW 라이선스 중에서 의무사항이 매우 강력한 라이선스이다. 대표적인 특징은 이 라이선스를 갖는 소스코드를 사용하여 프로그램을 만들면 그 프로그램의 소스코드를 모두 공개해야 한다는 점이다.
- **GPL 2.0**의 주요 의무사항은 다음과 같다.
 - 복제와 배포가 이루어 질 때는 본 허가서와 프로그램에 대한 보증이 제공되지 않는다는 사실에 대해서 언급되었던 모든 내용을 그대로 유지시켜야 하며, 영문판 GPL 라이선스를 함께 제공해야 한다.
 - 파일을 개작할 때는 파일을 개작한 사실, 내용 및 그 날짜 등을 파일 안에 명시해야 한다.
 - SW를 수정하거나 새로운 소프트웨어를 링크(static과 dynamic linking 모두)시키는 경우 GPL에 의해 소스코드를 제공해야 한다.
 - 프로그램의 일부를 본 허가서와 배포 기준이 다른 자유 프로그램과 함께 결합하고자 할 경우에는 해당 프로그램의 저작자로부터 서면 승인을 받아야 한다.

- 자신의 특허를 구현한 프로그램을 GPL로 배포하는 경우에는 그 프로그램을 GPL 조건에 따라 이용하는 이용자에게 특허에 대한 사용료를 받을 수 없으며, 제3자의 특허를 구현한 프로그램인 경우에는 그 특허권자가 GPL 조건에 따라 이용하는 프로그램 이용자에 대하여 특허사용료를 받지 않을 때에만 그 프로그램을 GPL로 배포하는 것이 가능하다.

□ GPL(GNU Public License) 3.0

- 기본적인 내용은 **GPL 2.0**과 비슷하지만 **GPL 3.0**에서는 **DRM** 관련 내용, 소프트웨어 특허문제, 양립성 문제 등이 추가되었다.
- 주요 내용은 다음과 같다.
 - GPL 3.0의 소스코드를 사용자 제품에 포함시키거나 혹은 그와 함께 배포하는 경우에는 해당 소스에 설치정보(installation information)를 함께 제공해야 한다.
 - ‘설치정보’란 소프트웨어를 수정하여 해당 제품에 설치하고 실행하는데 필요한 방법(methods), 절차(procedures), 인증키(authorization keys) 혹은 여타 정보 모두를 의미한다.
 - 다만 소프트웨어가 롬(ROM)에 설치된 경우처럼 해당 제품의 제조업체나 여타 제 3자도 수정된 코드를 제품에 설치할 수 없는 경우에는 설치정보를 제공하지 않아도 된다.
 - 특허와 관련해서 원래의 소스코드를 개선하여 배포한 기여자의 경우 자신이 기여한 부분에 대해서는 비차별적이고 특허 사용료가 없다는 내용의 라이선스를 제공해야 한다.
 - 특허와 관련해서 라이선스 등으로부터 특허소송이 제기되는 경우 소송을 제기한 날에 특허소송을 제기한 라이선스의 공개SW 라이선스는 종료된다.
 - 소프트웨어를 작성하고자 할 경우 기존에 만들어진 코드를 재사용하거나 결합하는 경우가 많은데 결합되는 각 코드의 라이선스가 상호 상충되는 경우 같이 사용할 수 없다.
- 저작권 표시 등 공개의 범위는 **GPL 2.0**과 같다.

□ LGPL(GNU Lesser General Public License) 2.1

- GPL이 사용자에게 너무 강한 제약을 주어 활용이 줄어들자 FSF에서 공개SW의 사용을 장려하기 위해 전략적으로 사용자의 의무를 완화한 라이선스를 정의했다. 상용 라이브러리와 동일한 기능을 제공하는 공개 SW 라이브러리에 GPL과 같은 엄격한 라이선스를 적용하게 되면 라이브러리를 사용하는 SW의 소스코드를 공개해야하기 때문에 상용SW 개발자들은 공개SW 라이브러리의 사용을 꺼려할 것이다.
- 오히려 이미 널리 사용되고 있는 상용라이브러리와 동일한 기능을 제공하는 공개SW 라이브러리를 LGPL로 배포하여 원 프로그램의 소스코드는 공개하지 않고, 이에 사용된 해당 공개SW 라이브러리의 소스코드만 공개하게 함으로써 공개SW 라이브러리의 사용을 장려하고 사실상의 표준으로 유도하는 한편 관련된 다른 공개SW를 보다 더 많이 사용할 수 있도록 하겠다는 것이다.
- LGPL은 'Library General Public License'였으나, 'Library'란 단어가 라이선스 이름에 포함되어 개발자들이 모든 라이브러리를 위한 라이선스로 오인하는 경향이 있었다. 이러한 오해는 일부 한정된 라이브러리에 대해서만 LGPL을 사용하려는 것이 FSF의 의도와는 맞지 않아 LGPL 버전 2.1에서는 이러한 오인을 방지하기 위하여 'Library'를 'Lesser'로 수정하였다.
- 주요 내용
 - 소프트웨어를 배포하는 경우 저작권 표시, 보증 책임이 없다는 표시 및 LGPL에 의해 배포된다는 사실을 명시해야 한다.
 - LGPL 라이브러리의 일부를 수정하는 경우 수정한 라이브러리의 소스코드 공개해야 한다.
 - LGPL 라이브러리에 응용프로그램을 링크시킬 (static과 dynamic linking 모두) 경우 해당 응용프로그램의 소스코드를 공개할 필요가 없다.
 - 다만 사용자가 라이브러리 수정 후 동일한 실행 파일을 생성할 수 있도록 Static Linking시에는 응용프로그램의 Object Code를 제공해야 한다.

- 특허의 경우 GPL과 동일하다.

□ BSD(Berkeley Software Distribution) License

○ SW의 소스코드를 공개하지 않아도 되는 대표적인 공개SW 라이선스

-- BSD 라이선스의 허용범위가 넓은 이유는 BSD 라이선스로 배포되는 프로젝트가 미국정부에서 제공한 재원으로 운영되었기 때문이며, 이는 SW에 대한 대가를 미국 국민의 세금으로 미리 지불했기 때문에 사람들에게 그들이 원하는 방식으로 소프트웨어를 사용하거나 만들 수 있도록 허가한 것이다.

- 따라서 BSD 라이선스의 소스코드를 이용하여 새로운 프로그램을 개발해도 새로운 프로그램의 소스코드를 공개하지 않고 BSD가 아닌 다른 라이선스를 적용하여 판매할 수 있다.

- SW를 배포하는 경우 저작권표시, 보증 책임이 없다는 내용을 표시해야 한다.

- 수정 프로그램에 대한 소스코드의 공개를 요구하지 않기 때문에 상용SW에 무제한 사용이 가능하다.

○ 저작권 표시 등 의무사항

- 소스코드를 재배포하기 위해서는 반드시 아래와 같은 저작권표시와 문구 그리고 면책조항을 유지하여야만 한다.

- Copyright (c) <연도>, <소유권자> All rights reserved.

- 예) Copyright (C) 2001, White Corporation All rights reserved.

- 소스코드의 재배포는 위의 저작권표시와 여기 나열된 조건들, 그리고 아래의 보증부인 고지를 포함해야 한다.

- 바이너리형식으로 재배포 할 때는 위의 저작권표시와 여기 나열된 조건들 그리고 아래의 보증부인 고지를 배포할 때 제공되는 문서 및 기타자료에 포함해야 한다.

- 사전에 서면으로 허가를 받지 않는 한, <기관>의 이름이나 기여자의 이름이 본 SW에서 추출한 제품을 보증하거나 홍보하는데 사용되어서는 안 된다.

- BSD 라이선스의 경우 의무사항만 준수한다면 소스코드를 공개하지 않아도 된다.

□ Apache License 2.0

- 아파치 라이선스는 아파치 웹서버를 포함한 아파치재단(**ASF: Apache Software Foundation**)의 모든 **SW**에 적용되는 라이선스로 **BSD** 라이선스와 비슷하여 소스코드 공개 등의 의무가 발생하지 않는다.
 - “Apache” 라는 이름에 대한 상표권을 침해하지 않아야한다는 조항이 명시적으로 들어가 있고 특허권에 관한 내용이 포함되어 BSD 라이선스보다는 좀 더 법적으로 완결된 내용을 담고 있다.
 - 특히 아파치 라이선스 2.0은 특허를 주장할 수 없다는 조항이 삽입되어 있어 GPL 2.0으로 배포되는 코드와 결합하는 것이 어렵다는 문제가 있었는데 GPL 3.0에서는 이 문제를 해결하여 아파치 라이선스로 배포되는 코드가 GPL 3.0으로 배포되는 코드와 결합하는 것이 가능해졌다.
 - 소프트웨어를 배포하는 경우 저작권 표시, 보증 책임이 없다는 내용을 표시해야 한다.
- 수정 프로그램에 대한 소스코드의 공개를 요구하지 않기 때문에 상용 **SW**에 무제한 사용이 가능하다.

<표 5-3> 공개SW 라이선스 Top 20

순위	라이선스 명	점유율
1	GNU General Public License (GPL) 2.0	26%
2	MIT License	19%
3	Apache License 2.0	16%
4	GNU General Public License (GPL) 3.0	11%
5	BSD License 2.0 (3-clause, New or Revised) License	7%
6	Artistic License (Perl)	5%
7	GNU Lesser General Public License (LGPL) 2.1	5%
8	GNU Lesser General Public License (LGPL) 3.0	2%
9	http://www.opensource.org/licenses/ms-pl	2%

순위	라이선스 명	점유율
10	Eclipse Public License (EPL)	2%
11	Code Project Open License 1.02	1%
12	Mozilla Public License (MPL) 1.1	< 1%
13	Simplified BSD License (BSD)	< 1%
14	Common Development and Distribution License (CDDL)	< 1%
15	Microsoft Reciprocal License	< 1%
16	GNU Affero General Public License v3 or later	< 1%
17	Sun GPL With Classpath Exception v2.0	< 1%
18	CDDL-1.1	< 1%
19	zlib/libpng License	< 1%
20	Common Public License (CPL)	< 1%

자료: <http://osrc.blackducksoftware.com/data/licenses/>, 미래창조과학부·정보통신산업진흥원 2014에서 재인용

<표 5-4> 공개 소프트웨어 라이선스 비교

라이선스의 주요내용	GPL v2	GPL v3	LGPL v2	LGPL v3	MIT	Apache 2.0	Eclipse Public License	BSD 3-Clause
복제, 배포, 수정의 권한허용	○	○	○	○	○	○	○	○
배포시 라이선스 사본첨부	○	○	○	○	○	○	○	47)
저작권고지/ Attribution고지 유지	○	○	○	○	○	○	○	○
배포시 소스코드 제공의무와 범위	전체 코드	전체 코드	2차적 저작물	2차적 저작물			모듈 단위	
조합저작물 작성 및 타라이선스 배포허용 ⁴⁸⁾	조건부		○	○	조건부	○	○	조건부
수정시 수정내용 고지		○	○	○		○	○	
명시적 특허 라이선스의 허용		○		○		○	○	
이름, 상표, 상호에 대한 사용제한						○		○
보증의 부인	○	○	○	○	○	○	○	○
책임의 제한	○	○	○	○	○	○	○	○

자료: 한국저작권위원회 <http://www.ojis.or.kr/oss/license/compareGuide.do>

47) 빈칸은 해당 라이선스에 명시적으로 언급이 없음을 의미한다. 그러나 언급이 없더라도 묵시적으로 허용하거나, 금지하는 것으로 해석할 수 있음.

48) 조합저작물(larger work)이란 라이선스 적용 코드 전체나 그 일부를 본 라이선스의 적용을 받지 않는 코드와 결합한 저작물을 의미한다.

5. 라이선스의 결정

라이선스는 연구자가 선택하도록 한다. 기존의 공개 소프트웨어 사이트에 올리는 것은 가장 권장할 일이다. 이 경우에는 공개 소프트웨어 사이트의 방침을 따르면 된다.

기존 사이트에 올리지 않는 소프트웨어는 정부에서 지원하는 공개소프트웨어 저장소에 등록하도록 한다. 현재 저작권협회에 등록하는 것, 소프트웨어 자산으로 등록하는 것 등을 원스탑으로 처리할 수 있는 방안을 마련해야 한다.

제5절 사업화 단계

정부에서는 국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률(약칭: 기술이전법)에 근거하여 정부연구개발 결과의 사업화를 추진하고 있다. 이 법에서는 구체적으로 정부에서 기술이전·사업화 촉진계획을 수립·시행하여야 한다고 규정하고 있으며 현재 5차 촉진계획이 시행중이다.(국가법령정보센터, 2014. 국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률; 미래창조과학부, 제5차 기술이전 및 사업화 촉진 계획, 2014.4.23)

정부과제의 사업화 성과는 두 가지로 구분된다. 첫 번째는 기술 확산이며 두 번째는 사업화를 통한 기술실시계약이다. 기술 확산은 기술이전, 기술지도, 기술평가 등으로 구분하여 성과관리를 하고 있다. 이 중 사업화와 직접적으로 연관된 내용은 기술이전이며 기술지도와 기술평가는 연구내용의 발표, 컨설팅 등의 활동이다.

연구과제에 기업체가 참여하여 개발한 경우에 해당 기업체가 사업화까지 추진할 수 있다. 기업은 사업화를 위한 기초 연구개발투자를 국가로부터 지원받은 것이지만 사업으로 이어지는 계획을 가진 시스템을 미리 공개하는 것은 원하지 않을 것이다. 마찬가지로 연구에 직접 참가하지 않은 기업에 기술이전을 통해 사업화의 기회를 제공할 수도 있다. 또는 연구에 참여한 연구원이 결과물을 이용하여 창업을 하려고 하는 경우도 있다. 이런 경우에도 확실히 소프트웨어를 공개하라고 요구하는 것은 바람직하지 않다. 이런 모든 사례를 사업화라는 하나의 용어로 규정하면 사업화를 계획하는 과제에 대해서는 일정 기간 산출물 공개의 유예기간을 주는 것이 필요하다.

사업화를 추진하는 경우에는 과제 종료시 소프트웨어 제출은 공개의 경우와 동일하게 취급한다. 제안서에 명시했던 소프트웨어 중 아직 공개소프트웨어로 공개하지 않은 것에 대해 공개소프트웨어로 등록하는 수준의 정리를 거쳐서 정부 공개소프트웨어 사이트에 등록한다. 이 산출물은 사업화가 추진되는 기간동안 비공개로 보관되며 정해진 기간 동안 사업화 성과가 없으면 공개소프트웨어로 전환된다.

연구원이 창업하여 사업화하려는 경우에도 규칙은 동일하게 적용한다. 다만 연구원 창업의 경우에는 연구비 지원과 창업 지원 프로그램 연계를 사업화 진행을 위해 필요한 연구개발 및 사업 비용을 지원하는 규정이 필요하다. 이것은 제4차, 제5차 기술이전·사업화 촉진계획에서 지속적으로 추진해 온 정책이므로 본 보고서에서 따로 다루지는 않는다.

연구과제의 결과물인 소프트웨어를 사업화한다면 보통 3년 정도가 유효기간일 것이다. 정부의 사업화 촉진계획에서도 R&D후 3년 이내의 사업화를 중점 지원하고 있으므로 3년이 지나고 사업화가 성공하지 못 한 과제의 결과를 공개소프트웨어로 하는 것은 무리가 없다. 일 년에 여러 번 버전이 바뀌는 현대 소프트웨어의 수명주기를 고려하더라도 3년간 사업화가 안 된 소프트웨어는 공개하는 것이 국가 연구개발의 취지를 살리는 길이다. 이를 추진하기 위해 우선 사업 종료가 되는 과제에서 기술이전이나 사업화 계획을 종료시점에 제출하도록 해야 한다. 그리고 다른 기업에 기술이전을 추진하는 경우에는 종료 후 1년 이내에 기술이전을 받을 기업의 의향서를 받아야 한다. 이런 조치는 연구자가 소프트웨어를 공개하지 않고 사업화한다고 했을 때 연구자의 판단을 신중하게 하여 실제 구체적인 사업화 아이디어가 있는 경우에만 사업화 과정으로 선언하도록 한다.

혼동을 피하기 위해 공개소프트웨어의 대상과 사업화 대상을 다시 설명하면 연구책임자가 사업화한다고 결정한 것은 꼭 공식산출물이 아니라 부수적인 산출물을 모두 포함한 것이다. 예를 들어 공식산출물을 활용하여 사업화를 추진하면서 연구과정 중에 개발한 소프트웨어를 공개하는 것도 가능하다. 일반적으로 연구결과를 사업화하면서 산출물의 전부 혹은 일부를 공개소프트웨어로 등록하는 것이 가능하며 사업화가 공개를 미루는 핑계로 악용되어선 안 될 것이다.

종료 시점에 사업화 계획을 제출하지 않거나, 기술이전을 추진하면서 1년 이내에 기업의향서를 받지 못하거나, 사업화를 추진하며 3년 이내에 가시적 성과가 없을 경우에는 제출된 소프트웨어를 공개로 전환한다.

사업화가 성공한 경우에는 해당 기업의 의향에 따라 제출한 소프트웨어의 공개여부가 결정될 것이다. 사업화의 성공은 현행 제도상으로는 매출을 일으키고 기술료를 납부하는 것이 구체적 지표가 된다. 이 외에도 유예 기간(예. 3년) 동안 기업 가치를 높여 신규 투자를 유치하는 것도 지표가 될 수 있다.

사업화에 대해서는 다양한 경우가 발생할 수 있으므로 원칙을 정하고 각 사례별로 원칙을 따르는 판단을 해야 할 것이다.

- 사업화와 소프트웨어 공개는 병행할 수 있다

- 어떤 것을 공개소프트웨어로 할 것인지 사업화할 것인지는 연구자가 판단

- 소프트웨어 공개 정책은 연구 결과를 널리 퍼뜨리는 것이지 사업화를 독촉하는 정책이 아니다

제6절 평가 단계

과제 종료 시점에 제안서에 명시했던 공개소프트웨어가 공개되었는지 확인하여 과제 종료 평가에 반영한다. 공개된 소프트웨어는 등록할 때 산출물의 품질, 문서 품질, 샘플 코드 등을 평가하기 때문에 이미 공개되어 있으면 공개소프트웨어로서의 최소한의 품질을 만족한 것으로 볼 수 있다. 공개가 유예된 소프트웨어는 종료 전에 연구책임자가 제출하여 보관하도록 한다. 이때에도 공개 소프트웨어와 동일한 품질 검증을 거친다.

공개된 소프트웨어는 활용도에 따라 정기적으로 평가한다. 평가의 요소로는 다운로드 건수, 지속적인 갱신을 통한 완성도 개선, 이용자 평가, 공개 사이트를 중심으로 이루어

지는 커뮤니티 참여도 등을 고려한다. 이용자의 평가와 커뮤니티 참여도는 주관적인 정성적 판단이고, 다운로드건수나 갱신 횟수 등은 정량적 판단이기는 하지만 의도적으로 조작될 여지가 있다. 그러므로 평가의 결과는 참고용으로 활용되는 것이 바람직하며 순위를 매기는 용도로 사용되면 왜곡된 결과를 초래할 수 있다.

기술이전을 추진하는 과제는 1년 이내 기업의 기술이전 합의가 달성되었는지의 여부를 1차로 평가한다. 연구 참여 기관이 사업화를 추진하는 과제와 외부 기업이 기술이전 합의를 한 과제는 연구 종료 후 3년 이내에 사업화 성과를 평가하여 사업화 성공 여부를 판단한다. 사업화가 실패한 과제는 제출한 소프트웨어를 공개한다.

연구원 창업의 경우에 3년 이내 투자 유치 혹은 사업화 성공이 평가 기준이 된다. 이 때에도 성과를 달성하지 못하면 비공개로 등록된 소프트웨어를 공개소프트웨어로 전환한다.

사업화 성공의 기준 중 투자유치는 연구원 창업의 경우뿐 아니라 사업화의 경우에도 적용할 수 있다. 다만 일반 중소기업 이상의 기업체는 연구과제와 상관없이 투자유치가 진행될 수 있으므로 대상에서 제외하는 것이 바람직하다. 이 부분은 사례별로 평가하는 것이 좋을 것이므로 해당 기업에서 투자 유치와 연구 결과가 관련된다는 점을 해명하도록 하는 것이 필요하다.

본 방안은 연구자가 과제 제안 시 공개할 소프트웨어를 명시하고 여기에 들어가는 인력, 비용, 일정까지 고려하여 계획을 세우고 제안하도록 하고 있다. 이것이 채택되어 정책에 반영되는 시점에 이미 진행 중인 과제의 경우에는 이런 내용이 반영되지 않았을 것이므로 원칙대로 품질을 지켜서 공개하기 어려울 수 있을 것이다.

그러므로 과도기적으로는 공개를 위해 추가적인 노력이 많이 들어간다고 주장하는 연구자는 따로 많은 노력을 들이지 않고 코드, 데이터, 샘플프로그램, 관련 문서 등 있는 그대로 공개하도록 요구한다.

공개사이트를 운영하는 쪽에서도 제출하는 내용의 형식에 대한 가이드라인을 제공한 후에 정상적인 공개를 요구하는 것이 옳을 것이다. 과도기 동안 연구자가 제출한 코드와 문서를 운영측이 직접 검토하고 시험하면서 연구자에게 요구할 가이드라인을 현실화할 기회를 가질 수 있다.

제7절 법제도 고려 사항

1. 개요

- 국가연구개발사업의 관리 등에 관한 규정(이하 ‘R&D 관리규정’ 이라 함)은 과학기술기본법에 근거(제11조 내지 제12조의5)하고 있음
 - R&D 관리규정에 맞게 국가 연구개발사업 관련 관리규정이나 처리규정을 제정하며, 다른 규정의 상위 법령(대통령령)으로서 지위를 점하고 있음
- 정보통신·방송 연구개발 관리규정(이하 ‘ICT 관리규정’ 이라 함)에서는 정보통신 및 방송 분야 R&D만을 그 대상으로 하며 미래부 고시형태로 제정되어 있음
 - ICT 관리규정은 구체적으로 공개소프트웨어의 활성화 방안에 대해 규정하고 있음
 - 제36조 11항 “수행기관의 장은 소프트웨어 관련 연구개발 사업 수행 결과 중 기술실시계약이 체결되지 않는 등 활용이 미흡한 기술의 확산을 위해 공개소프트웨어로 배포하는 것이 바람직하다고 판단되는 경우에는 사업의 결과물을 공개소프트웨어로 배포할 수 있다.”
 - 제40조(공개소프트웨어의 특례) 공개소프트웨어 방식으로 추진하는 과제에 대한 추진방식, 기술료 감면 등 규정
 - ICT 관리규정은 연구결과를 공개소프트웨어화 하는 근거가 되긴 하지만 개선이 필요한 부분이 있음
 - 제40조 1항 “장관은 소프트웨어 연구개발에 관하여 제16조에 의한 과제기획을 실시하는 경우 연구개발의 효율성을 높이거나 사업성과의 활용을 촉진하기 위하여 공개소프트웨어로 배포하는 것이 바람직하다고 판단되는 연구개발 사업에 대해서는 공개소프트웨어 개발방식으로 추진하도록 할 수 있다. 이 경우 전담기관의 장은 공개소프트웨어 타당성 분석을 실시하여야 한다.”

- 제40조 1항은 미래부장관이 공개소프트웨어 개발방식으로 추진하도록 할 수 있으며 이 경우 전담기관이 타당성조사를 해야 한다고 규정
- 대안으로는 공개소프트웨어로 추진해서는 안 되는 과제를 지정하는 방식이 바람직하며 공개를 안 하는 경우에 타당성조사를 하도록 규정하는 편이 공개를 더욱 장려하는 방향임

□ 각 부처별로는 연구개발사업의 관리나 처리를 위해 훈령 형태로 처리규정을 갖고 있음

- 과학기술 R&D에 관하여는 대표적으로 ‘교육부 소관 이공분야 연구개발사업 처리규정’ 및 ‘미래창조과학부 소관 과학기술분야 연구개발사업 처리규정’을 들 수 있음
- 양 규정에서는 ‘연구재단’을 양 규정의 전담기관으로 지정하여 운영하고 있으며, 개별적으로 연구개발특구진흥재단 및 한국산업기술진흥협회(이상 미래부 처리규정), 한국산업기술진흥협회 및 한국과학기술단체총연합회(이상 교육부 처리규정) 등을 전담기관으로 지정하고 있으며, 주로 교육부 또는 미래부 소관 법령의 R&D를 대상으로 함
- 위에서 언급된 관리규정 및 처리규정을 이하에서는 ‘대상 규정’이라고 함
- 연구재단 등 현 전담기관은 미래창조과학부의 규정에서 지정된 내용을 따르고 있으며 자체적으로 별도의 규정을 사용하고 있지 않음
- 따라서 ICT 관리규정의 변경이 전담기관의 관리규정에도 영향을 줌

2. 제안 및 선정 과정(협약 포함)

□ 기술 수요조사, 과제 공고·신청 단계에서 각종 서류에 R&D 소프트웨어 성과물(공식, 비공식)의 활용방안을 구체적으로 기재토록 필요

- 대상 규정에서는 제안 등의 과정에서 R&D 성과물의 활용방안에 대해 기재토록 하고 있으나, 양식을 개선하여 공개 소프트웨어화 방안 등을 구체적으로 명시하도록 함

- 참고로, IITP의 RFP에서는 과제특징으로서 정책지정(), SW자산뱅크등록대상(), 기술사업화적용(), 경쟁기획(), 후불형(), 표준화연계(), 기술료징수() 등 해당항목을 체크하도록 하고 있음
- 과제 제안서에 공개할 소프트웨어 산출물 리스트를 제시하는 항목을 추가하도록 제안서 양식 변경
- 공식, 비공식 소프트웨어 산출물을 공개하면 과제 선정 시 인센티브가 부여됨을 RFP와 제안서 양식에 명시하도록 관리 규정과 양식 변경
- 선정에 따른 협약서에서도 이를 구체적으로 명기하여야, R&D 종료 후 결과보고 및 결과평가에서도 평가할 수 있음

3. 보고 및 결과 평가

- 연구개발 결과보고서에 최종보고서에 포함되어야할 사항은 ‘연구개발 결과의 활용계획’에 관한 사항으로, 이때 협약에 명시한 소프트웨어 공개의 성과에 대해 기술하도록 함
 - 실제 성과물이 협약에 체결한 내용대로 도출되고, 공개소프트웨어 방식으로 배포된 경우에는 평가에 가산점 부여

4. 공개 및 성과 관리

- 결과물의 공개 및 성과 관리에 있어서 공개소프트웨어를 적극적으로 활용할 수 있도록 관련 법제도의 체계화에 대한 검토
 - 현재는 연구개발결과의 공개 시에 SW의 경우 저작권위원회, SW자산뱅크 등에 등록 또는 기탁토록 되어 있음. 공개된 소프트웨어가 공식 산출물인 경우에는 이 절차와 병행하여 운영
 - 공개된 소프트웨어의 활용 현황을 연구 성과 추적 조사에 반영하도록 절차 개선 필요

5. 추적 평가

- 추적 평가시에도 공개소프트웨어의 활용도에 대한 평가를 통해서, 다른 R&D 과제의 심사 시에 가점을 부여할 수 있도록 함

6. 장기적 제언 - 대상 규범의 법제화

- 현행 R&D 관리규정은 과학기술기본법에 근거한 대통령령으로 제정되었으나, 국가R&D 사업의 중요성 및 산출물을 효율적으로 이용하기 위해서는 일반 법률로서 제정하는 것이 바람직하다고 봄
- 국가R&D의 한 축인 연구결과물에 대한 평가를 담당하고 있는 「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률」(연구성과평가법)과의 균형을 위해서라도 별도 법률로서 규정하여야 할 것임
 - R&D 관리 규정 등 관련 규정 및 과학기술기본법 등 R&D 관련 규정을 체계화하여 정리하여, 「(가칭)연구개발 사업관리 및 이용활성화에 관한 법률」 제정
 - 법제정을 다른 관련 법률이나 법령의 효력이 제한되는 것이 아닌 상호 보완적인 관계가 유지되도록 함
- SW에 특화된 부분에 대해서는 SW산업진흥법에서 SW R&D 결과물의 활용을 위한 규정을 두거나, 또는 ICT특별법에서 공개SW 및 R&D 결과물의 활용을 위한 근거를 두는 것도 필요

제8절 신규 지원 조직의 기능

국내 공개 사이트를 운영하는 것 외에도 기능적으로 신규로 필요한 내용이 있다. 공개되는 데이터, 소프트웨어 등의 평가, 관리, 추적 기능이 제공되어야 한다. 소프트웨어의 활용 상황을 감시하고 평가하는 업무, 소프트웨어의 활용을 확산하기 위한 인프라 구축 및 운영, 사용자의 소프트웨어에 대한 접근성을 향상시키기 위한 서버와 서비스의

개선 등 소프트웨어 등록자나 사용자에게 편리한 서비스를 제공하는 역할이 필요하다.

이 조직을 단순한 서비스 운영조직이 아니라 연구개발조직으로 키운다면 도메인 전문가가 아쉬워했던 과학기술 연구개발에 대한 소프트웨어 기술 지원 서비스도 이 조직에서 점차 역량을 키워가며 지원할 수 있다. 단기적으로는 이 조직에서 공개에 어려움을 겪는 연구개발 과제에 대해 소프트웨어 개발을 지원하면서 공개를 위한 표준 정의, 컨설팅, 개발 지원을 할 수 있다. 나아가 공개된 산출물의 활용 확산을 위해 과학소프트웨어 교육 프로그램, 산출물의 서비스화 등을 할 수 있다. 고급 인력을 모은다면 과학기술연구의 기반구조를 개발하는 연구 개발 조직으로 확장할 수 있다.

정보 교류를 활성화하기 위한 활동도 이 조직이 할 수 있는 내용이다. 과학자와 공학자가 모이는 정기 학회와 워크숍을 주관하고 정보 교환 커뮤니티 운영 등 정보 교환 창구를 신설, 확대 및 강화하는 업무를 이 조직이 할 수 있다.

제6장 결론

현대 과학기술 연구에서 소프트웨어는 필수적인 도구가 되었다. 시뮬레이션 데이터와 측정데이터의 규모가 커지면서 컴퓨터를 이용한 데이터 처리가 없이는 사람이 분석할 수 없는 것이 현실이다. 선진국에서는 10여 년 전부터 이 문제를 인식하고 과학자와 컴퓨터과학자의 공동 연구가 중요하다는 것을 전파하고 연구 환경을 개선해 왔다. 미국, 영국, EU 등의 정부에서도 국력의 기초인 과학기술의 발전을 지속하기 위해 과학연구를 위한 사이버 기반구조 구축 프로젝트를 수행해 오고 있다. 사이버 기반구조에는 고성능 컴퓨팅, 데이터 관리, 데이터 시각화, 공동작업 서비스, 네트워크, 가상화, 소프트웨어 스택 등이 포함된다.

한편 과학연구 문화적 측면에서는 연구용으로 개발한 소프트웨어를 공개소프트웨어로 등록하여 다른 연구자가 이용하고 개선할 수 있도록 하는 것이 일반적이다. 최근에는 과학연구 데이터를 공개하도록 정부에서 나서는 상황이다. 유럽에 있는 대형 강입자 충돌기의 경우에 1년에 30페타바이트의 데이터가 추출되기 때문에 전 세계의 과학자가 공동으로 분석에 참여하고 있다. 사이버 기반구조는 이런 문화를 배경으로 하고 있다.

국내의 연구현황은 과학연구에서 소프트웨어의 중요성에 대한 문제의식은 공유하고 있으나 실행하기가 쉽지 않은 여건이다. 소프트웨어 분야에서 충분한 전문성을 갖춘 인력이 과학연구에 공동으로 참여하여 1년 이상 경험을 쌓은 융합인력이 필요한데 이런 사례는 매우 드물다. 공동으로 깊이 있는 연구를 할 수 있는 연구 환경도 부족하다. 학계의 평가제도는 논문위주의 단일 체계 중심인데 반해 융합인력은 새로운 분야에 적응하는 초기 수년간 실적이 나오기 어렵기 때문에 논문실적에 쫓기는 상황에서 과학연구에 공동으로 참여하기 어렵다. 그리고 과학자와 소프트웨어 전문가가 정보를 교환할 기회도 거의 없는 형편이다.

우리나라 정부에서는 융합과제를 적극적으로 추진하고 있으나 주로 사업화를 위한 방향으로 추진되고 있다. 본 연구에서 인터뷰한 연구자 중에도 융합과제를 수행하는 분이 많았는데 실제로 두 분야가 깊이 교류할 수 있는 과제가 아니라 과학자가 기획한 애플리케이션을 소프트웨어 전문가가 구현하는 식의 협력이 많았다. 이런 피상적 협력관계

로는 과학연구에 필요한 소프트웨어 도구를 제공하기는 어려워진다. 융합과제 자체는 많이 추진되고 있는데 협력의 수준이 피상적이 되는 것은 정부에서 과제를 추진한다고 문제가 해결되는 것이 아니라는 것을 의미한다.

과학자와 소프트웨어 전문가의 협력 수준을 개선하려면 미시적 조치가 아니라 연구 문화를 바꾸는 거시적 방안이 필요하다. 이를 위해 본 연구에서는 국가 과학기술 연구의 결과로 생산되는 소프트웨어를 공개소프트웨어로 등록하도록 하는 제도를 제안한다. 일견 단순한 제도로 보일 수도 있으나 이것을 통해 몇 가지 중요한 변화를 가져올 수 있다.

우선 과학자는 과제를 통해 작성하는 소프트웨어의 품질을 높이기 위해 노력해야 한다. 소프트웨어 공개는 단순한 일이 아니며 공개 후에 좋은 평을 들으려면 전문적 지식이 필요한 경우도 많다. 과학자가 소프트웨어에 대한 공부를 더 하거나 외부 전문가와 공동 작업을 하게 된다. 공개된 소프트웨어는 다른 사람이 수정해서 개선할 수 있고 이 사람은 소프트웨어에 전문적 지식을 가진 사람일 수도 있다. 또 사용자의 버그 신고나 불만 의견에 따라 소프트웨어가 계속 진화할 수 있다.

소프트웨어 전문가는 이런 소프트웨어의 코드를 보면서 과학연구의 중요한 문제를 이해할 수 있게 된다. 과학연구에서 무엇이 필요한지 이해할 수 있다. 일반적으로 논문에 포함되는 내용보다 소프트웨어에 더 자세한 내용이 들어있다. 공개된 코드를 통해 융합 인력의 첫 단계를 시작하기 쉬워지게 되는 것이다. 또한 서로 정보교환을 할 수 있는 창구도 공개소프트웨어 사이트를 통해 생기게 된다.

소프트웨어 산출물을 공개 소프트웨어로 등록하는 제도가 지속적으로 유지되고 제도에 적극적으로 참여하는 연구자에 대한 보상이 꾸준히 이어진다면 국내 연구 문화는 지금보다 한 단계 성숙해 질 것이라고 생각된다. 연구자의 수준이 더 정확히 드러날 것이고 의미있는 연구를 찾아내기도 쉬워질 것이다. 한 연구팀에서 개발한 연구 도구 소프트웨어를 다른 연구팀이 활용할 수 있으므로 국가적으로 중복개발을 피할 수 있고 소프트웨어의 품질이 꾸준히 높아지면 원 저작자도 혜택을 누릴 수 있다.

국가 연구개발 과제를 공개소프트웨어로 추진하는 것은 미래창조과학부에서도 추진해 오던 일이고 올해 1월에 관리규정을 개정하여 장관이 승인하면 공개소프트웨어로 추진할 수 있게 되었다. 현 제도에서도 소프트웨어 산출물을 공개하는 것은 가능해 진 것이

다. 다만 제도의 효과를 높이기 위해서는 공개하지 않을 과제만 장관승인을 받는 식으로 보완하는 것이 좋을 것이다.

과학자와 소프트웨어 전문가의 협력은 학제간 공동 연구이고 선진국에서도 쉬운 일은 아니다. 지난 20년간 학제간 공동 연구에 대해 발표된 논문이 유사한 문제를 지적하고 있는 것은 이 문제가 쉽지 않다는 것을 단적으로 보여준다. 소프트웨어 산출물을 공개하는 것만으로 이 문제가 완전히 해결되지 않을 것은 명백하다. 그러나 이 한 걸음이 과학자와 소프트웨어 전문가의 거리를 좁혀서 협력이 더 쉬워질 것이라는 것은 확실하다.

참 고 문 헌

국내 문헌

- 과학기술정책연구원 (2009), 통합적 혁신정책을 위한 정책조정방식 설계, 서울: 동연연구소. 정책연구 2009-07.
- 교육부 (2013), 교육부 소관 이공분야 연구개발사업 처리규정.
- 국가과학기술위원회 (2011), 제2차 연구성과 관리·활용 기본계획(안).
- 국가법령정보센터 (2014), 국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률, <http://www.law.go.kr>.
- 국회예산정책처 (2013), 국가 연구개발 투자의 성과 측정 방법 연구.
- 김해도 (2010), 국가연구개발사업의 기술료제도 변천과정 고찰과 현행 제도의 문제점 연구. 지식재산연구 5(3). pp.87-120.
- 미래창조과학부 (2013a), 2012년도 국가연구개발사업 성과분석 결과(안).
- 미래창조과학부 (2013d), 국가연구개발 성과평가 개선 종합대책(안).
- 미래창조과학부 (2013b), 2014년도 정부연구개발투자 방향 및 기준(안).
- 미래창조과학부 (2013c), 과학기술 빅데이터 활용 종합계획(안).
- 미래창조과학부 (2014c). 2014년도 정부연구개발 투자 방향 및 관리제도.
- 미래창조과학부 (2014d), 2015년도 정부연구개발투자 방향 및 기준(안).
- 미래창조과학부 (2014a), 10호 제5차 기술이전 및 사업화 촉진계획(안).
- 미래창조과학부 (2014e), 국가중점과학기술 전략로드맵(안).
- 미래창조과학부 (2014b), 2013년도 국가연구개발사업 조사·분석 결과(안).
- 미래창조과학부 (2014f), 정보통신·방송 연구개발 관리규정.
- 미래창조과학부·한국과학기술기획평가원 (2013), 2012년도 국가연구개발사업 성과분석 보고서.
- 미래창조과학부·한국연구재단 (2013). 2013 미래창조과학부 주요 연구개발사업 성과분석보고서.

- 성지은 (2010). 세계 주요국의 과학기술혁신정책 조정체계와 최고 조정기구 비교분석.
과학기술정책연구원, STEPI Working Paper Series, WP 2010-06.
- 오현석 · 배형준 · 김도연 (2012), “과학기술분야 융합연구자의 융합연구 입문과 과정에
관한 연구” , 아시아교육연구, 13(4), pp. 297-335.
- 이정찬 (2014), 아베노믹스와 일본의 과학기술혁신정책. 동향과 이슈, 제11호.
과학기술정책연구원.
- 정보통신산업진흥원 (2014), 공개SW 라이선스 가이드. 미래창조과학부.
- 충북대학교 (2012),. 국가과학데이터의 효율적 관리 및 활용을 위한 법제도 기본연구,
대전: 한국과학기술정보연구원, K-12-SG-102-01P-1.
- 한국과학기술기획평가원 (2009), 선진국의 과학기술관련 종합조정체계 및 주요 정책동향
분석.
- 한국과학기술기획평가원 (2013), 2012년도 국가연구개발사업 조사 · 분석 보고서
-통계표-, 경기: 미래창조과학부, 11-1191000-000001-10.
- 한국과학기술원 (2011), SW R&D체계 개편방안 연구, 서울: 지식경제부.
- 한국과학기술정보연구원 (2011), 국가과학데이터 공유 · 융합 체제 구축에 관한 연구,
대전: 동연구소, ISBN: 978-89-6211-727-1-93400.
- 한국생산기술연구원 (2013),. 융합활성화를 위한 참여기관 협업방안 연구.
- 홍성주 (2012), 과학기술기본계획의 추이 분석과 시사점 : 최근 10여 년간 한국과
일본의 과학기술기본계획을 중심으로. STEPI Insight 2012. 3. 1. 제89호.
과학기술정책연구원.
- 홍형득 (2014), 영국의 과학기술행정체제와 투자구조의 변화와 특징분석.
한국자치행정학보 28(1). pp.95-115.

해외 문헌

- Abdelhamid, S.E., Alo, R., Arifuzzaman, S. M., et al. (2012). “Cinet: A
cyberinfrastructure for network science.” In: 8th IEEE International
Conference on eScience, Chicago, October 8-12, 2012.

- Atkins, D. (2010). *RCUK Review of e-Science: Building a UK Foundation for the Transformative Enhancement of Research and Innovation*. Swindon: UK, Research Councils UK.
- Boinc Stats. "Detailed stats SETI@Home."
<http://boincstats.com/en/stats/0/project/detail/overview>
- Brent, Roger & Bruck, Jehoshua (2006). "Can computers help to explain biology?" *NATURE* 440(23), pp.416~417.
- Butler, D. (2006). "Everything, Everywhere." *Nature*, 440, March 23, pp.402-405.
- Carayol N. and T. U. N. Thi. (2005). "Why do Academic Scientists Engage in Interdisciplinary Research?" *Research Evaluation* 14(1) pp.70-79.
- De La Flor G., Jirotko, M., Luff, P., Pybus, J. & Kirkham, R. (2010). "Transforming scholarly practice: Embedding technological interventions to support the collaborative analysis of ancient texts." *Computer Supported Cooperative Work (CSCW)*, 19(3-4), pp.309-334.
- Division of Advanced Cyberinfrastructure Programs.
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504947&org=ACI&from=home.
- Fischer, Eric A. (2013). Public Access to Data from Federally Funded Research: Provisions in OMB Circular A-110. Congressional Research Service, R42983.
- Google I/O 2013. (2013). "Cloud Computing and High-Energy Particle Physics: How ATLAS Experiment at CERN Uses Google Compute Engine in the Search for New Physics at LHC."
- Haupt, Tomasz. (2010). "Cyberinfrastructure for Integrated Computational Material Engineering. Minerals," *Metals and Materials Society/AIME*, 420 Commonwealth Dr., P. O. Box 430 Warrendale PA 15086 USA.[np]. pp.14-18.
- Hey, T., Tansley, S., and Tolle, K. (eds.). (2009). "Jim Gray on eScience: A Transformed Scientific Method." in *The Fourth Paradigm*, ed. By T. Hey, S. Tansley, and K. Tolle, (pp. xix-xxxiii). Redmond, WA: Microsoft Research.
- Jorissen, K., F. D. VILA, and J. J. REHR. (2012). "A high performance scientific

- cloud computing environment for materials simulations,” *Computer Physics Communications*, 183(9), pp.1911-1919.
- Lee, J. Wipapat Kladwang, Minjae Lee, Daniel Cantu, Martin Azizyan, Hanjoo Kim, Alex Limpaecher, Sungroh Yoon, Adrien Treuille, Rhiju Das, & EteRNA Participants. (2014). “RNA design rules from a massive open laboratory,” *Proceedings of the National Academy of Sciences*, 111(6), pp.2122-2127.
- Li, J., Agarwal, D., Humphrey, M., van Ingen, C., Jackson, K., & Ryu, Y. (2010). “eScience in the cloud: A MODIS Satellite Data Reprojection and Reduction Pipeline in the Windows Azure Platform.” *2010 IEEE Parallel & Distributed Processing Symposium*.
- Missier, P., Soiland-Reyes, S., Owen, S., Tan, W., Nenadic, A., Dunlop, I., Williams, A., Oinn, T. & Goble, C. (2010). Taverna, reloaded. In *Scientific and Statistical Database Management* (pp. 471-481). Springer Berlin Heidelberg. “Taverna, reloaded,” In *Scientific and Statistical Database Management*. Berlin: Springer. January, pp.471-481.
- National Science Foundation. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. January.
- National Science Foundation. (2007). *Cyberinfrastructure Vision For 21st Century Discovery*, March.
- National Science Foundation (2012). *Cyberinfrastructure Framework for 21st Century Science and Engineering*, May.
- National Science Foundation (2014). “Division of Advanced Cyberinfrastructure Programs.”
http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504947&org=ACI&from=home.
- Olson, J. S., Ellisman, M., James, M., Grethe, J. S. & Puetz, M. (2008). “The Biomedical Informatics Research Network,” in *Scientific Collaboration On The Internet*, ed. By Gary M. Olson, Ann Zimmerman, and Nathan Bos, Cambridge:

- MA: The MIT Press.
- Qiao, Qin, Bowman, Gregory R., and Huang, Xuhui (2013). “Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation,” *Journal of the American Chemical Society* 135, no. 43, pp.16092–16101.
- Qin, J., F. W. Lancaster and Bryce Allen (1997). “Types and Levels of Collaboration in Interdisciplinary Research in the Sciences,” *Journal of the American Society for Information Science* 48(10) pp 893–916.
- Schmidt, E. & Rosenberg, J. (2014). *How Google Works*. New York, NY: Grand Central Publishing.
- van Rijnsoever, Frank and Laurens K. Hessels. (2011). “Factors associated with disciplinary and interdisciplinary research collaboration,” *Research Policy* 40(2011) pp 463–472.
- Watson, P., Hiden, H., & Woodman, S. (2010). “e-Science Central for CARMEN: science as a service.” *Concurrency and computation: Practice and Experience*, 22(17), pp.2369–2380.
- The White House. (1998). Presidential Decision Directives, May 22, <http://fas.org/irp/offdocs/pdd/pdd-63.htm>
- Wu, W., Uram, T. & Papka, M. E. (2009). “Web 2.0-based social informatics data grid.” In: *Proceedings of the 5th Grid Computing Environments Workshop*. ACM, p.6.
- Zakas, N. (2013). “Starting An Open-Source Project.” <http://www.smashingmagazine.com/2013/01/03/starting-an-open-source-project/> .

부록. 해외 소프트웨어 활용 사례

제1절. 미국의 분야별 SW활용 사례

1. 환경 과학 (Environmental Science)⁴⁹⁾

□ 산업 소개

환경 과학이란 인간이 살고 있는 환경오염의 감소·방지 등 공해문제를 중심으로 한 여러 환경문제를 과학적으로 구명하려는 응용과학의 한 분야이다.

□ 개요

MODIS (Moderate Resolution Imaging Spectroradiometer) 위성 데이터는 매 1~2일마다 테라 및 아쿠아 위성으로부터 생성되는 토지, 해양, 하층부 대기의 전세계 동태와 프로세스에 대한 이해를 돕기 위한 데이터로, 이러한 MODIS 데이터는 대지기반의 센서로부터도 수집된다. 미국 마이크로소프트 대외 연구측과 로렌스 버클리 국립 연구소, 버지니아 주립대학 컴퓨터 과학과와 UC 버클리 환경 과학, 정책 및 관리 과가 합동으로 추진한 MODIS 데이터 통합 및 파이프라인 재구성 프로젝트는 다음과 같은 주요 현안을 중심으로 진행되었다.

- 입력 **MODIS** 데이터를 타임프레임과 해상도에 맞춘, 규격화된 지질학적 포맷의 데이터로 재투영
- 대지 기반의 센서와 위성으로부터 얻은 **MODIS** 데이터를 통합하여 신규 환경 과학 데이터를 도출하는 기술 개발
- 윈도우 애저 플랫폼 **API** 및 사용 방식을 공개하고, 사이버 인프라의 활용 방안에 대하여 제언

49) LI, J. et al. 2010.

○ 대규모 데이터 프로세싱과 저장에 초점

MODIS 데이터는 투영 타입에 따라 두 가지(Swath, Sinusoidal)로 나눌 수 있으며 대기 및 토지 정보를 다루고, 585K의 원본 데이터에서 총 3.9TB의 재투영 데이터가 생성된다.

<표 1> 데이터 설명

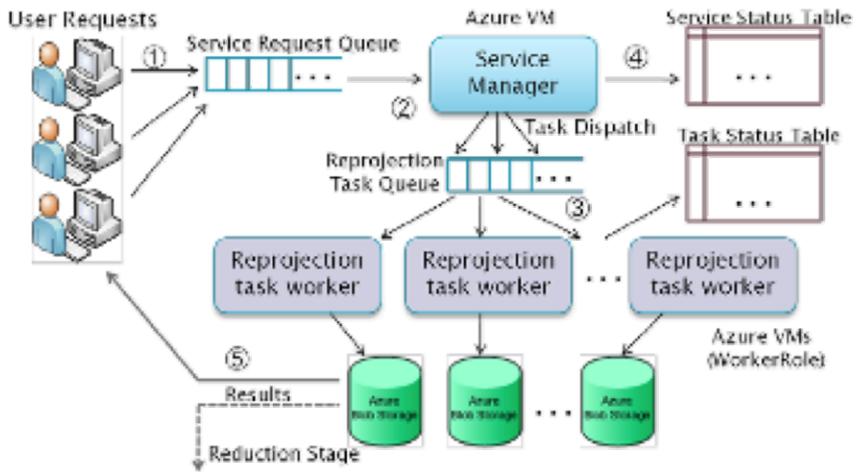
구분	MO(Y)D04 L2	MO(Y)D05 L2	MO(Y)D06L2	MO(Y)D07 L2	MCD12Q1. 005
데이터 그룹	대기	대기	대기	대기	대지
프로젝트 종류	Swath	Swath	Swath	Swath	Sinusoidal
#소스 데이터 파일	65K	130K	130K	130K	150
데이터 크기	85GB	500GB	2TB	850GB	13GB
구분	MCD15A2. 005	MCD43B3. 005	MO(Y)D11.I. 2.005	MO(Y)D13 A2.005	
데이터 그룹	대기	대기	대기	대기	
프로젝트 종류	Sinusoidal	Sinusoidal	Sinusoidal	Swath	총계
#소스 데이터 파일	7K	7K	113K	3K	585K
데이터 크기	13GB	110GB	337GB	40GB	3.9TB

이전에는 이러한 데이터 수집, 변환, 관리의 과정이 시간 소모적이고 반복적이며 오류가 많이 발생했으며, US 대륙의 10년치 데이터를 프로세싱하는 데 있어 수만 시간의 CPU 시간이 필요했다.

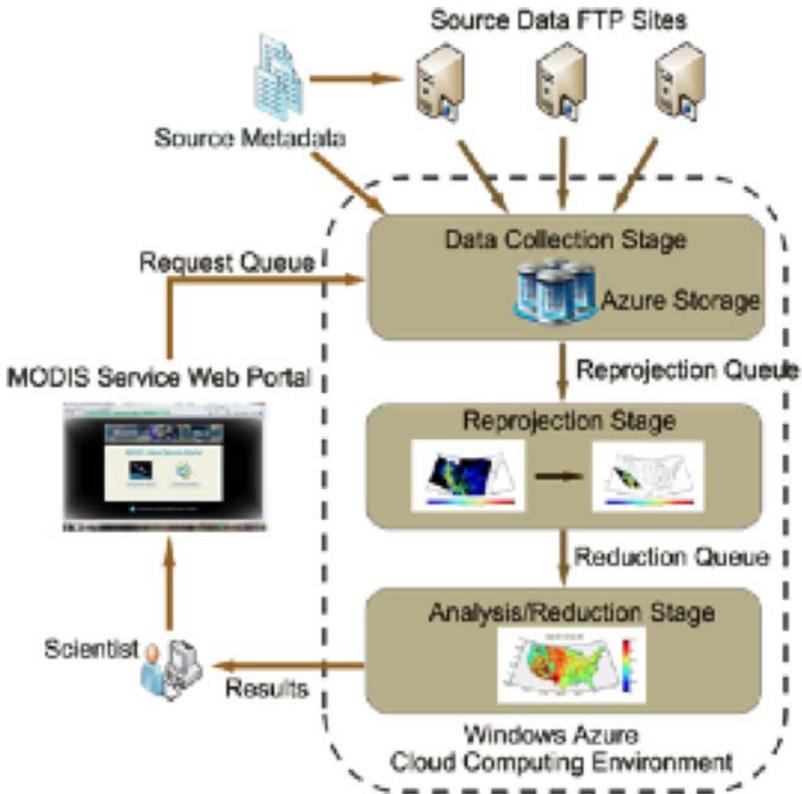
□ 연구 방법

데이터 저장에는 윈도우 애저 블랍 스토리지를 사용하였고, 데이터 처리 및 분석에는 .NET 기반 프레임워크를 활용한 커스텀 애플리케이션을 사용하였다. 파이프라인 단계는 “데이터 수집 -> 재투영 -> 분석 및 유의미한 데이터 도출” 로 이어진다.

[그림 1] 재투영 작업을 위한 작업 스케줄링 방식



[그림 2] 파이프라인 단계에 따른 시스템 구성도



□ 활용 효과

데스크탑보다 저사양의 중급 애저 인스턴스를 활용하여 단일 처리 및 병렬 처리/확장 수준을 파악해 보면 다음과 같다.

단일 애저 인스턴스는 데스크탑보다 저사양인데도 불구하고, 해당 중급 애저 인스턴스 150개를 활용하여 기존보다 90배 이상 빠르게 처리할 수 있었다.

<표 2> 데스크탑 머신과 애저 인스턴스 용량

	데스크탑	애저 인스턴스
용량	CPU: 인텔 코어2듀오 E6850@3.0GHZ 메모리:4GB 하드 디스크: 1TB SATA 네트워크: 1Gbps 이더넷 OS: 윈도우 7 RC 빌드7100(32비트)	CpU: 1.5~1.7GHZ X64 동급 프로세서 메모리:2GB 로컬 스토리지: 250GB 네트워크: 100Mbps OS:윈도우 2008 서버 x64(64비트)

<표 3> 재투영 실행 통계 - MOD04L2

	평균 Tw	평균 Tc	평균 Tr
데스크탑	39.1s	9.2s	29.9s
애저 인스턴스	82.7s	14.9s	67.7s
성능 비율	2.1	1.6	2.3
SD(성능 비율)	0.21	0.79	0.13

<표 4> 재투영 실행 통계 - MOD06L2

	평균 Tw	평균 Tc	평균 Tr
데스크탑	169.5s	65.8s	103.7s
애저 인스턴스	267.5s	52.6s	214.9s
성능 비율	1.6	0.80	2.07
SD(성능 비율)	0.13	0.31	0.07

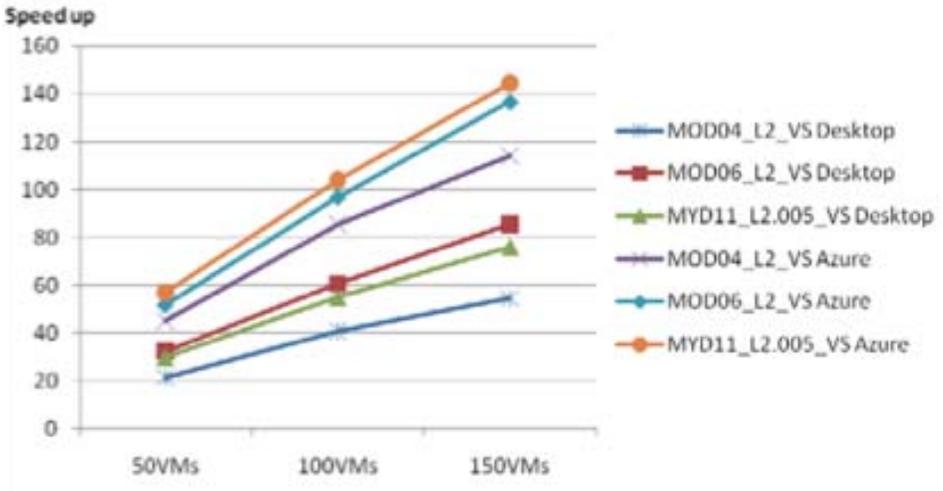
<표 5> 재투영 실행 통계 - MOD11L2.005

	평균 Tw	평균 Tc	평균 Tr
데스크탑	80.3s	14.9s	65.4s
에저 인스턴스	151.5s	15.3s	136.1s
성능 비율	1.9	0.10	2.1
SD(성능 비율)	0.39	0.73	0.40

<표 6> 1500개의 작업을 처리하는데 든 처리 시간(단위:시간)

	MOD04_L2	MOD06_L2	MYD11_1.2.005
데스크탑	16.29	72.62	33.45
단일 인스턴스	34.21	116.19	63.56
50 인스턴스	0.76	2.25	1.12
100 인스턴스	0.40	1.20	0.61
150 인스턴스	0.30	0.85	0.44

[그림 3] 병렬 처리 속도의 향상



〈표 7〉 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
<p>데이터 수집, 변환, 관리의 과정이 시간 소모적이고 반복적이며 오류가 많이 발생.</p> <p>US 대륙의 10년치 데이터를 프로세싱하는 데 있어 수만 시간의 CPU 시간이 필요하였으나 데스크탑을 활용</p>	<p>데스크탑보다 90배 이상 빠르게 처리할 수 있었음</p>

2. 생명 정보학 - BIRN 협력 사례⁵⁰⁾

□ 산업 소개

생명 현상과 의학 연구에 컴퓨터와 소프트웨어를 이용하는 것으로, BIRN을 이용한 협력 분야는 크게 다음 세 가지를 포함한다.

- 뇌 영상 구성
- 인간의 신경 장애 관련 유전 정보학
- 관련 동물 모델 실험

□ 개요

BIRN(Biomedical Informatics Research Network)은 미국 국립 일반 의과학 연구소(NIGMS)에서 관리하는 생명과학 분야의 대규모 데이터 및 컴퓨팅 자원의 처리를 다루는 분산 인프라로, 다음과 같은 기관에서 참여하였다.

50) Olson, J. S., et al., 2008.

〈표 8〉 BIRN 참여 기관

	참여자수	기관
Function BIRN	186	캘리포니아 주립 대학(샌디에고), 듀크 대학, 캘리포니아 주립 대학(LA), 브링엄 여성 병원(BWH), 메사추세츠 종합 병원(MGH), 캘리포니아 주립 대학(알바인), 스탠포드 대학, 미네소타 주립 대학, 아이오와 주립 대학, 뉴 멕시코 주립 대학, 노스캐롤라이나 주립 대학
Morphometry BIRN	153	캘리포니아 주립 대학(샌디에고), 듀크 대학, 캘리포니아 주립 대학(LA), 브링엄 여성 병원(BWH), 메사추세츠 종합 병원(MGH), 존스 홉킨스 대학, 캘리포니아 주립 대학(알바인), 워싱턴 대학(세인트루이스)
Mouse BIRN	73	캘리포니아 주립 대학(샌디에고), 듀크 대학, 캘리포니아 주립 대학(LA), 캘리포니아 기술 연구소, 테네시 주립 대학(멤피스)
BIRN-CC	33	캘리포니아 주립 대학(샌디에고)

○ BIRN 프로젝트는 위 참여 기관 분류에서도 볼 수 있듯이, 크게 다음 네 가지로 구성된다.

– Function BIRN

조현증(정신분열증)의 원인과 치료에 대한 이해를 심화하기 위한 다중 기능성 자기 공명(MR) 도구 개발에 초점을 둔다.

– Brain Morphometry BIRN

기억 장애나 우울증과 같은 증상의구조적 차이와 관련한 뇌 구조의 차이를

분석하기 위해 교정 및 해부학적 분석 툴을 개발한다.

– Mouse BIRN

다중 경화증, 정신분열증, 파킨슨 병, 주의력 결핍 과잉 행동 장애, 뚜렛 증후군, 뇌 암과 같은 사람의 질병의 쥐 모형에 초점. 이러한 신경 장애를 이해하는 데 필요한분자 정보와 해부학적 영상을 포함하는 다양한 차원의 데이터를 집계한다.

– BIRN-CC

공동 기술 인프라를 지원하고, 데이터 및 분석 도구의 공유를 가능하게 하며, 이러한 리소스에 접근하기 위한 직관적인 웹 포털을 제공하는 등 BIRN 협력 그룹의 공동의 기술 이슈를 해결하는데 초점을 둔다.

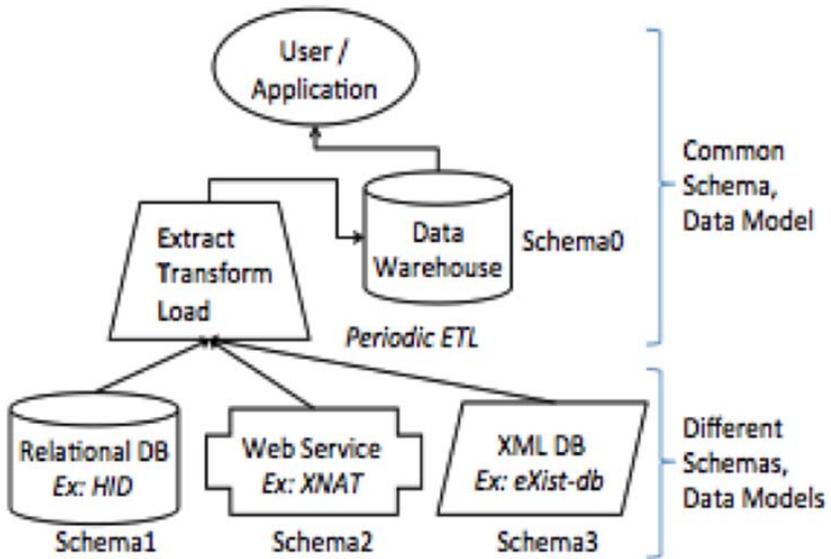
○ **BIRN** 프로젝트가 해결해야 했던 도전 과제는 다음과 같다.

- 데이터 식별을 통한 메타 데이터 생성
- 다양한 이기종 데이터 소스의 통합
- 데이터 이관 및 공유를 용이하게 해야 함
- 처리 및 분석의 커스텀화

□ 연구 방법

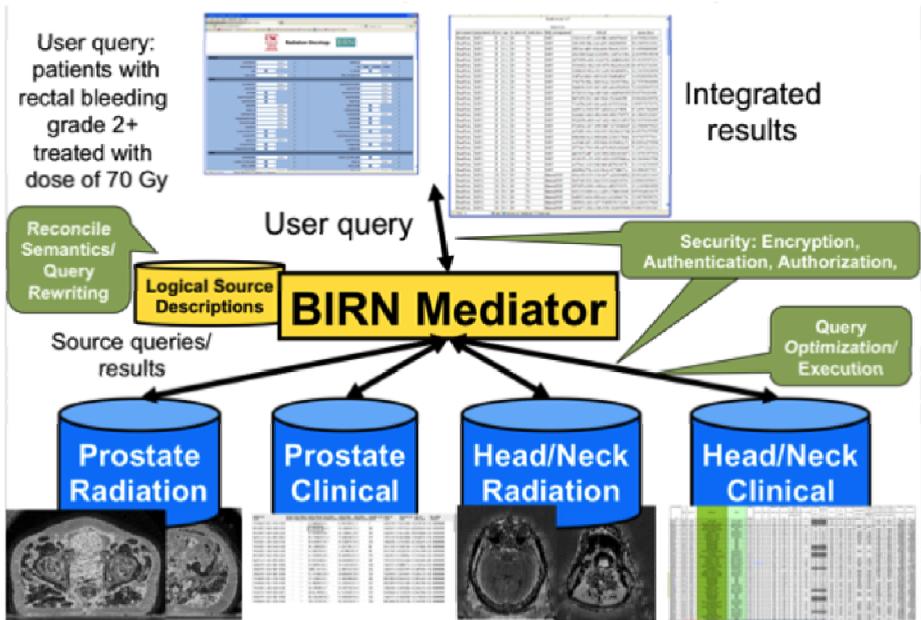
- 공동의 스키마로 데이터를 변환하여 하나 이상의 데이터웨어하우스에 로드

[그림 4] 데이터 수집 및 처리 흐름도



- 중개자가 각 독립 소스의 데이터에 대한 쿼리를 번역하고, 데이터와 로컬 스키마는 그대로 보존

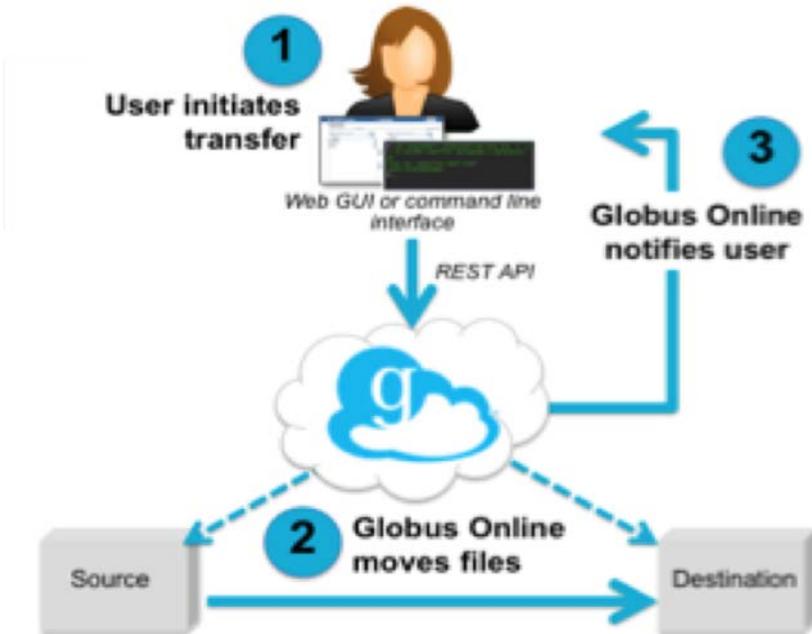
[그림 5] BIRN 중개자 활동의 예시



자료: <http://www.birncommunity.org/resources/data/>

- 연구자들이 문서를 쉽게 공유할 수 있는 온라인 파일 공유 시스템 구축

[그림 6] 글로버스 온라인 파일 공유 시스템 예시

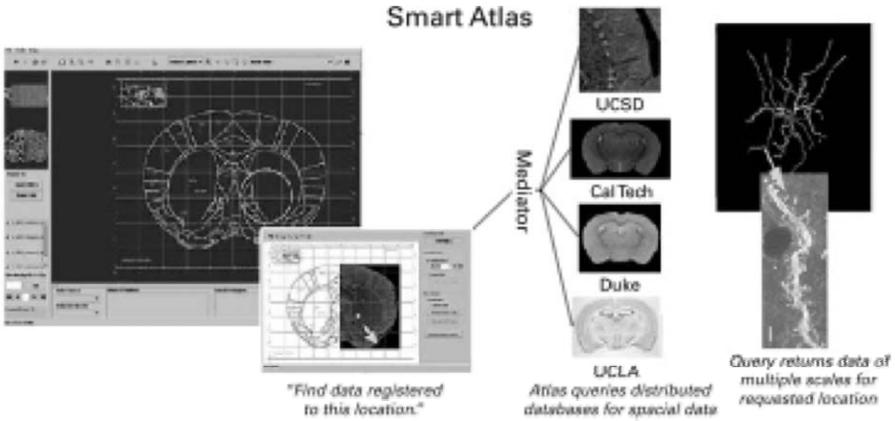


자료: <http://www.birncommunity.org/resources/data/>

□ 활용 효과

- 2007년 6월 기준 96개 이상의 논문이 게시되었고, 1500만개 이상의 파일(16테라바이트 이상)이 데이터 그리드에 존재하고 있다.

[그림 7] BIRN을 활용한 쥐 스마트 아틀라스 이미징 예시



자료: <http://www.birncommunity.org/resources/data/>

- 400개 이상의 계정을 관리 중이며, 데이터 그리드의 파일이 1800만 번 공유되었고, 이후 계속적으로 개발되어 13년 글로벌스 온라인을 제공하고, 파이프라인을 개선 중이다.

<표 9> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
이기종 데이터베이스의 혼재 및 데이터의 분산, 파일공유가 어려움	데이터 통합 및 파일 공유가 쉬워졌으며 처리 및 분석은 커스텀화 되었음

3. 입자 물리학 - Higgs 입자 실험을 위한 아틀라스와 구글 컴퓨터 엔진⁵¹⁾

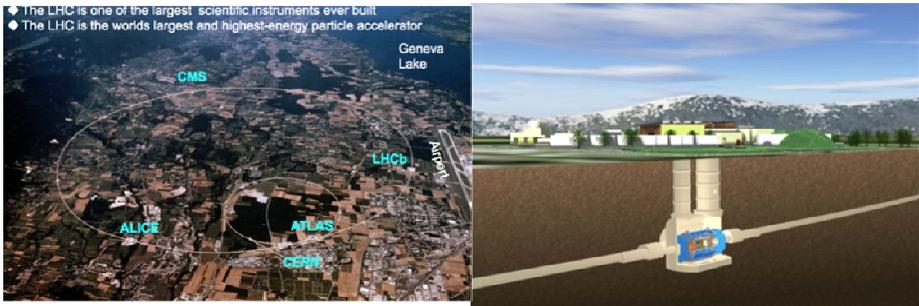
□ 산업 소개

- 입자 물리학은 세상을 구성하는 가장 기본적인 입자를 다루는 분야로 물질을 구성하는 기본적인 입자가 무엇이고, 입자 사이의 상호작용을 규명하여 자연현상의 본질에 대해 탐구하는 학문이다. 고에너지 물리학이라 불리기도 한다.

□ 개요

- 유럽 입자 물리학 연구소(CERN)은 스위스 제네바와 프랑스 사이의 국경지대에 위치한 세계 최대의 입자 물리학 연구소이다. 원래 명칭은 유럽 원자핵 공동 연구소(Conseil Européen pour la Recherche Nucléaire)였고, 이를 따라서 CERN이라 불린다.
- CERN은 설립 초기부터 입자 가속기 등을 이용해, 고 에너지 물리학 연구에 많은 기여를 하였다. 또한 물리학자들의 문헌 검색 및 제휴를 위하여 고안된 HTML과 월드 와이드 웹의 발상지로도 유명하다.

[그림 8] LHC와 지하에 묻힌 ATLAS 탐지기



자료: <https://developers.google.com/events/io/sessions/333315382>

51) Google I/O 2013, 2013.

- 그 중에서도, CERN LHC의 ATLAS 실험에는 38개 국가의 3000명의 과학자, 1200명의 대학원생이 참여 중이다. ATLAS 실험이 해결해야 하는 주요 현안은 다음과 같다.

- 입자 물리학의 고에너지 프론티어를 탐색

- 신생 물리학 연구

- 힉스 입자와 그 특성

- 표준 모형 이상의 물리학 (SUSY, Dark Matter, extra dimensions, Dark Energy 등)

- 표준 모형 파라미터의 정확도 측정

- 이러한 실험을 수행하기 위해 해결해야 했던 도전 과제는 다음과 같다.

- ATLAS는 2013년 기준 140PB 데이터와 O(100)분산 컴퓨팅 센터를 전세계적으로 다루고 있기 때문에 운영, 관리적 어려움이 있음

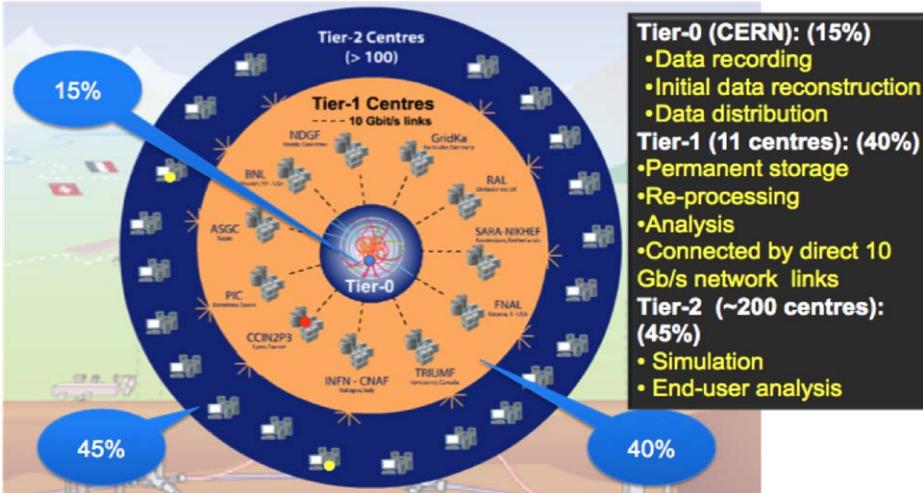
- 매년 40 PB 데이터가 생성

- 40 여개 이상의 국가에서 온 과학자가 협업해야 함

□ 연구 방법

- PanDA 워크로드 관리 시스템으로 매년 O(100)사이트의, O(100000) 코어, O(100000000)의 작업을 처리하고 O(1000) 사용자를 관리

[그림 9] PanDA 워크로드 관리 시스템

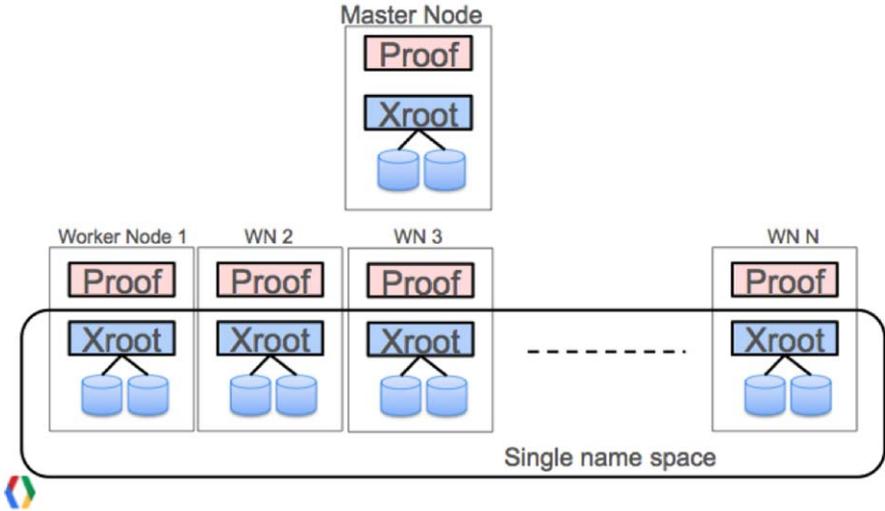


자료: <https://developers.google.com/events/io/sessions/333315382>

- 구글 컴퓨트 엔진 트라이얼 기간에 **ATLAS** 일부 작업을 8주간 시험
 - 고성능 분석 클러스터(PROOF)
 - 클라우드 스토리지 및 데이터 관리
 - XRoot와 클라우드 스토리지 간 집계 및 상호작용
 - 몬테 카를로 시뮬레이션을 위한 PanDA 큐 활용

[그림 10] XRoot와 PROOF 클러스터

PROOF/Xrootd Cluster



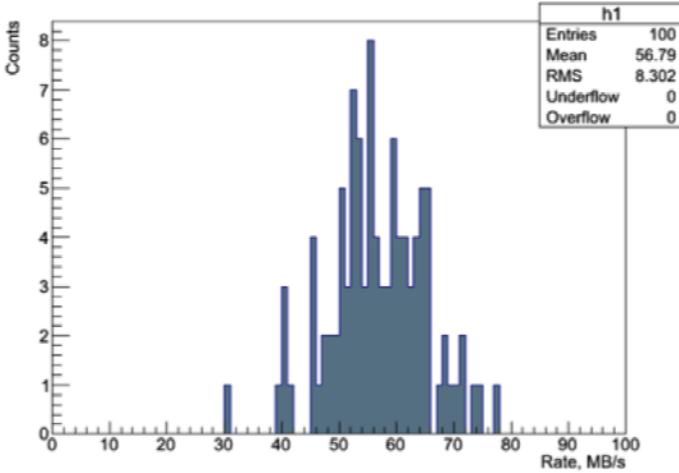
자료: <https://developers.google.com/events/io/sessions/333315382>

□ 활용 효과

○ 클라우드에서의 데이터 전송 및 분산 환경 실험을 통해 새로운 인프라 확보 가능성을 타진하였으며, 데이터 접근을 용이하게 하기 위해 지연 시간, 하드웨어 요구 사항, 인건비를 최소화하고 확장성을 최대화하는 방안을 모색할 수 있었다.

- 영구 디스크 스토리지 클러스터를 활용하여 더 나은 확장성 및 피크 성능을 확보
- RAID를 통해 더 나은 성능을 확보
- 공용 네트워크를 사용하더라도 XRoot연합과 GCE 스토리지클러스터 간 57 메가비트/s의 전송속도를 보여줌

[그림 11] 전송 속도



자료: <https://developers.google.com/events/io/sessions/333315382>

<표 10> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
개방형 과학 그리드 및 자체 시스템	퍼블릭 클라우드 적용 테스트를 통해 복잡해지는 요구사항에 알맞은 새로운 인프라 확보 가능성을 타진

4. 응집 물질 물리학 - 응집 물질 물리학을 위한 과학 클라우드 컴퓨팅 플랫폼⁵²⁾

□ 산업 소개

- 응집 물질 물리학은 물질의 거시적인 특성을 다루는 물리학의 분야다. 응집 물질 물리학에서는 특히 계 내부의 구성요소가 극히 많고 그 사이의 상호작용이 강할 경우 언제나 나타나는 "응집"상태가 관심의 대상

52) Jorissen et al., 2012

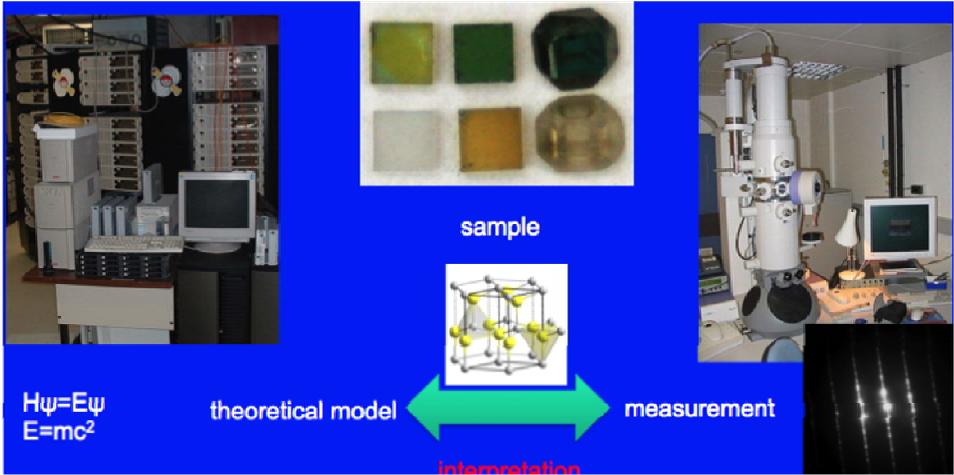
이다. 응집 상태의 가장 익숙한 예로는 원자 간의 전기력에 의해 생성되는 고체와 액체를 들 수 있다. 더욱 특이한 응집상태에는 극히 낮은 온도 하에서 몇몇 원자계의 경우에 나타나는 초유체와 보스-아인슈타인 응축물, 특정 물질에서 전도 전자가 갖는 초전도 상태, 원자 격자에서 스핀의 강자성 및 반강자성 상태 등이 포함된다.

- 응집 물질 물리학은 현대 물리학 분야 중 가장 크다. 미국의 물리학자 중 약 3분의 1이 이 분야에 속한다고 한다. 역사적으로 보면 응집물질 물리학은 고체 물리학에서 나왔는데, 고체 물리학은 이제 응집물질 물리학의 주요 하위분야로 간주되고 있다. 응집물질 물리학이라는 용어는 필립 앤더슨이 1967년 기준에 "고체 이론"이었던 그의 연구 그룹 이름을 바꿀 때 만들어 낸 것이다. 1978년 미국 물리학회이 고체 물리학 분과는 응집물질 물리학 분과로 개명되었다.
- 이 분야를 응집물질 물리학이라고 부르게 된 원인 중 하나는 고체 연구를 위해 개발된 개념과 기술 중 많은 부분이 유체계에 적용되기 때문이다. 예를 들어 전기 도체내의 전도 전자들은 원자들로 이루어진 유체와 같은 특성을 지닌 양자유체를 이룬다. 사실 전자들이 응축하여 흠여짐 없이 흐를 수 있는 새로운 유체가 되는 초전도 현상은 저온에서 ^3He 가 초유체로 되는 현상과 매우 비슷하다. 또한, 응집물질 물리학은 물성물리학 분야와 매우 깊은 연관이 있다. 그리하여, 응집 물질 물리학자들은 이른바 "나노기술" 즉, 리처드 파인만에 의해 처음 예견되었던 원자정도 크기의 기계 개발을 가능케 하는 기술을 연구하고 있다

□ 개요

- NSF(National Science Foundation)에서 펀딩하고 워싱턴 대학교에서 주도한 이 과학 클라우드 컴퓨팅 플랫폼 프로젝트는 다음과 같은 주요 현안을 해결해야 했다.
 - 이론적 모델을 평가하기 위한 정량적 측정의 해석/구현에 HPC(High Performance Computing)가 필요
 - 향상된 코드를 실행하는데 클러스터가 필요
 - 향상된 코드를 사용하기가 어려움

[그림 12] 모델 측정 및 구현

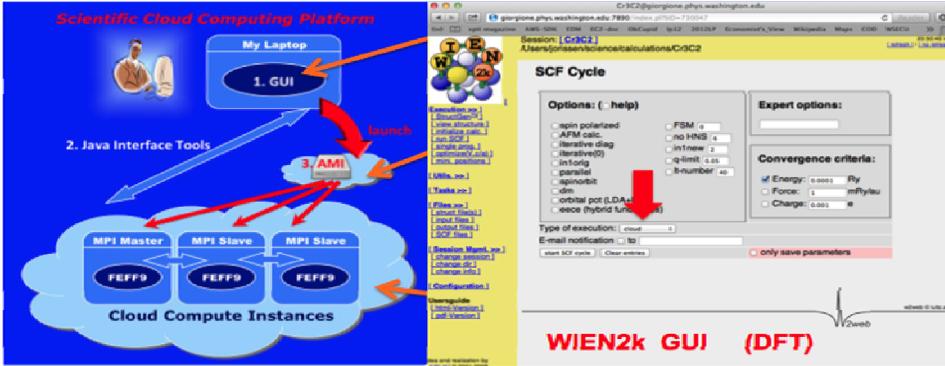


자료: Jorissen et al., 2012

□ 연구 방법

- 워크플로를 단순화시키는 인터페이스(GUI, APP 형식)
- 미리 설치된 이미지를 런치하여 활용
- 필요할 때마다 필요한 만큼 클러스터를 구성하여 사용 후 폐기

[그림 13] 아키텍처 및 GUI화면



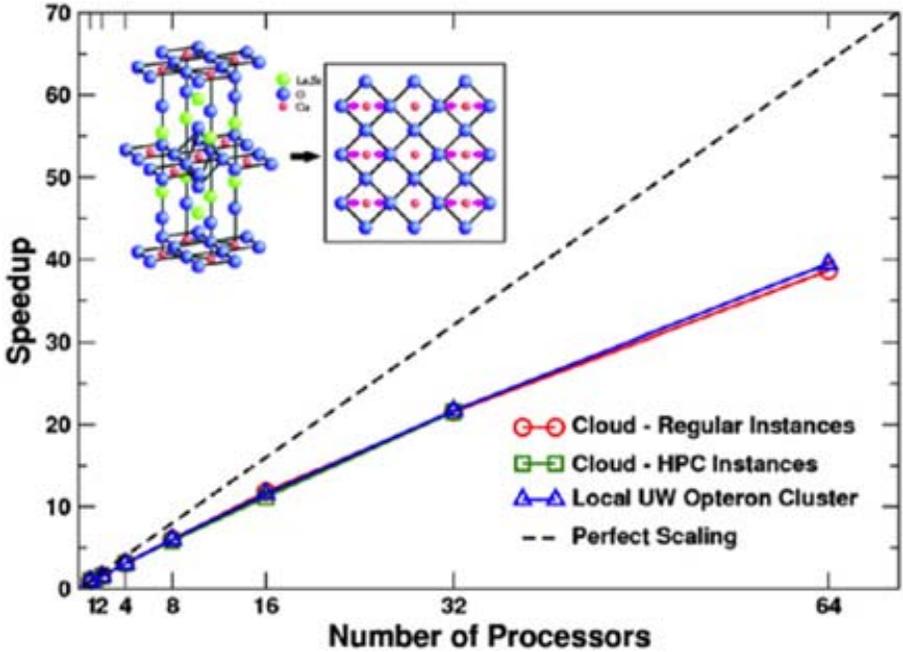
자료: Jorissen et al., 2012

□ 활용 효과

○ 느슨한 결합 프로세스

- LSCO XAS의 FEFF9 계산의 고속화가 가능하였다. 대각선은 완벽한 확장성을 나타내므로 N 프로세서가 있을 때 1개의 프로세서보다 N배 빠름을 나타낸다. 세 개의 플랫폼의 성능을 보여주는 그래프로, 로컬 UW 오프온 클러스터와, 일반/고성능 가상 클라우드 클러스터의 성능을 각각 보여준다. [그림 14]의 모형은 LSCO 고온 초전도체의 구조를 나타낸다.

[그림 14] FEFF9 계산의 고속화

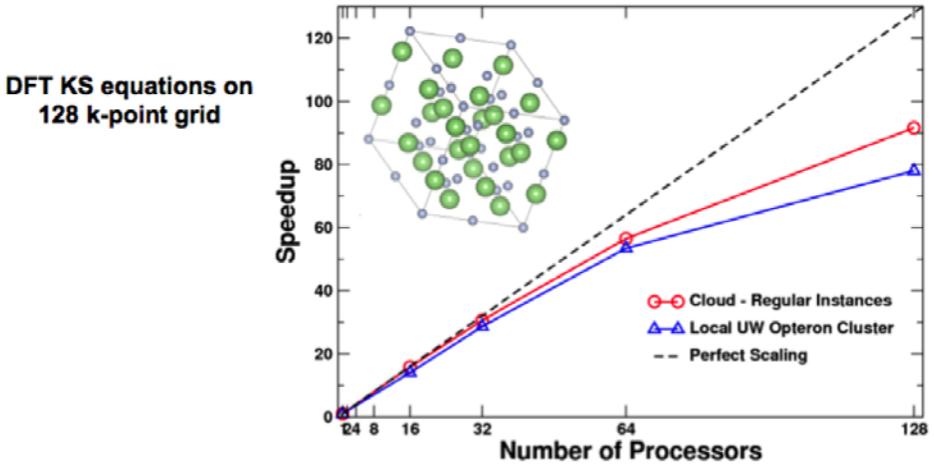


자료: Jorissen et al., 2012

○ 밀착 결합 프로세스

- [그림 14]는 32-원자 GaN 셀의 구조를 나타내는 그림으로, 밀착 결합 프로세스를 계산하는 데 쓰이는 WIEN2k 계산을 고속화할 수 있었다. 로컬 인피니밴드를 활용한 UW 옵테론 클러스터보다 일반 공용 네트워크를 활용한 클라우드 일반 클러스터가 더 확장성이 좋았다.

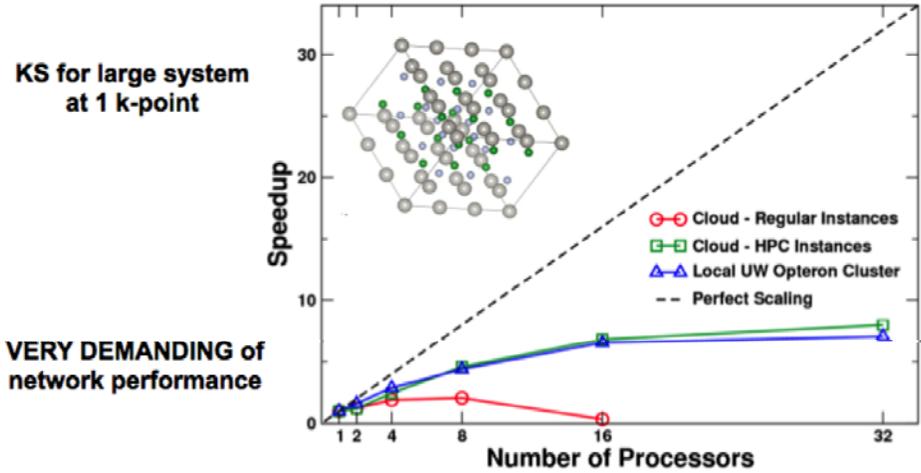
[그림 15] WIEN2k 계산의 고속화



자료: Jorissen et al., 2012

- [그림 15]은 64-원자 GaN 셀의 구조를 나타내는 그림으로, 밀착 결합 프로세스를 계산하는 데 쓰이는 WIEN2k 계산을 고속화할 수 있었다. 로컬 인피니밴드를 활용한 UW 옵테론 클러스터보다 일반 공용 네트워크를 활용한 클라우드 HPC 클러스터가 더 확장성이 좋았다. 이는 해밀토니안 행렬 계산을 포함한다.

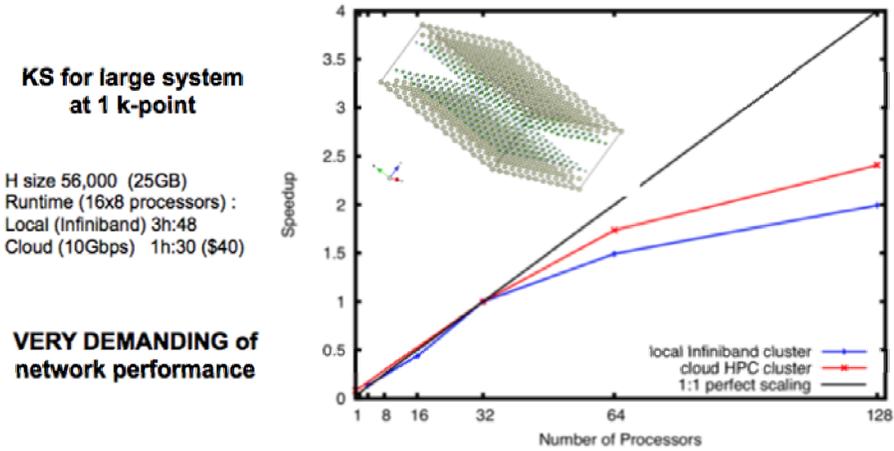
[그림 16] WIEN2k 계산의 고속화



자료: Jorissen et al., 2012

- [그림 16]은 1200-원자 Rh/BN slab 셀의 구조를 나타내는 그림으로, 밀착 결합 프로세스를 계산하는 데 쓰이는 WIEN2k 계산을 고속화할 수 있었다. 로컬 인피니밴드를 활용한 UW 옵테론 클러스터보다 일반 공용 네트워크를 활용한 클라우드 HPC 클러스터가 더 확장성이 좋았다. 이는 해밀토니안 행렬 계산을 포함한다.

[그림 17] WIEN2k 계산의 고속화



자료: Jorissen et al., 2012

<표 11> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
마땅한 현실적 대안이 없음. 개선된 코드를 실행할 플랫폼이 없음.	클라우드를 통해 WIEN2k계산 등 개선된 코드를 실행시킬 수 있었으며 일반 공용 네트워크를 활용한 클라우드 클러스터 및 로컬 클러스터의 확장성 테스트할 수 있었음

5. 네트워크 과학 - CINET (A cyberinfrastructure for network science)⁵³⁾

□ 산업 소개

- 복잡한 네트워크 관련 연구를 포함한 네트워크 연구는 1990년대 이후 지금까지 급격한 속도로 확산되고 있다. 이는 그래프의 추상화, 이론,

53) Abdelhamid, et al., 2012

알고리즘, 에이전트 기반 모델링 시뮬레이션을 포함한다. 네트워크과학의 응용 분야는 생물학, 생태학, 세포 면역학, 사회 과학, 보건 과학, 경제, 컴퓨터 네트워크, 역학, 통계 물리학, 언어 진화학 등이있다. 네트워크 과학은 이러한 도메인 내에서 시스템 속성과 행동을 이해하는 데 유용하다. 따라서 그래프 기반의 계산을 수행할 수 있는 소프트웨어 및 인프라는 큰 가치가 있다.

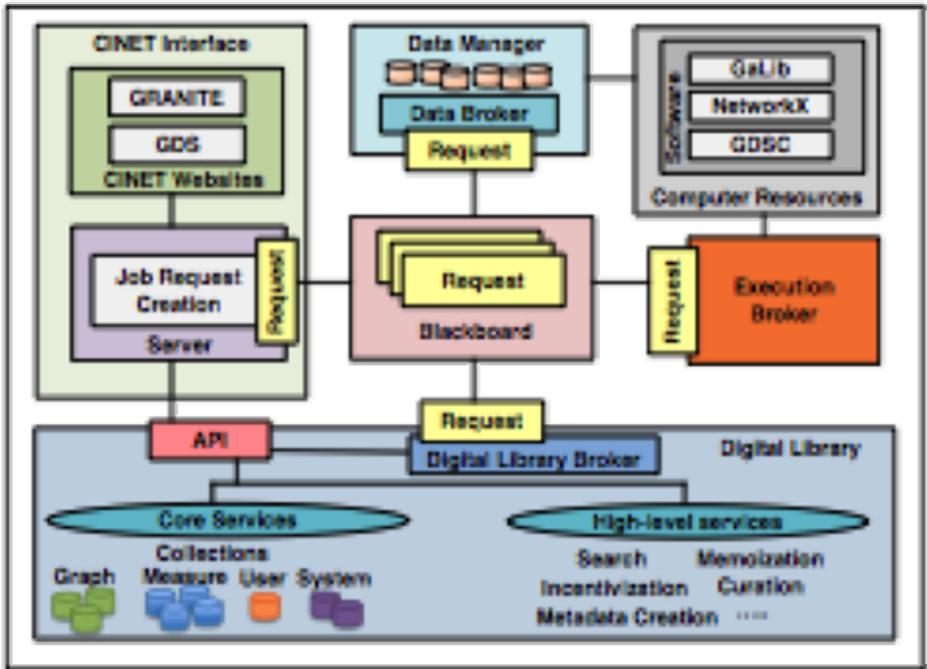
□ 개요

- 네트워크 과학을 위한 사이버인프라 프로젝트는 NSF의 펀딩 하에 시카고 아곤 국립 연구소, 버지니아텍 대학교, 휴스턴 대학교, 인디애나 대학교, 알바니 대학교 등에서 참여하였다.
- 이 프로젝트의 주요 현안은 네트워크 과학 연구를 위한 사이버 인프라를 제공하는 것으로 다음과 같은 도전 과제가 있었다
 - 연구 논문 및 다양한 결정론적 랜덤 네트워크 생성기를 참조하여 실제적 네트워크를 제공
 - 네트워크 연구와 특성화를 위해 많은 알고리즘 모듈과 측정 수단을 제공
 - 분산된 고성능 컴퓨팅 파워를 활용해 대규모 네트워크를 다루고 복잡한 연산을 할 수 있도록 함
 - 직관적 웹 인터페이스 제공

□ 연구 방법

- **Galib, Network X**와 같은 계산 엔진 및 그래프 라이브러리와 그래프 동적 시스템 계산기인 **GDSC**, 분산 워크플로우 관리, 시맨틱 웹 툴과 같은 구성 요소를 갖춘 **HPC as a Service**를 구현하였다.

[그림 18] 고수준 CINET 구성요소 및 상호작용 뷰



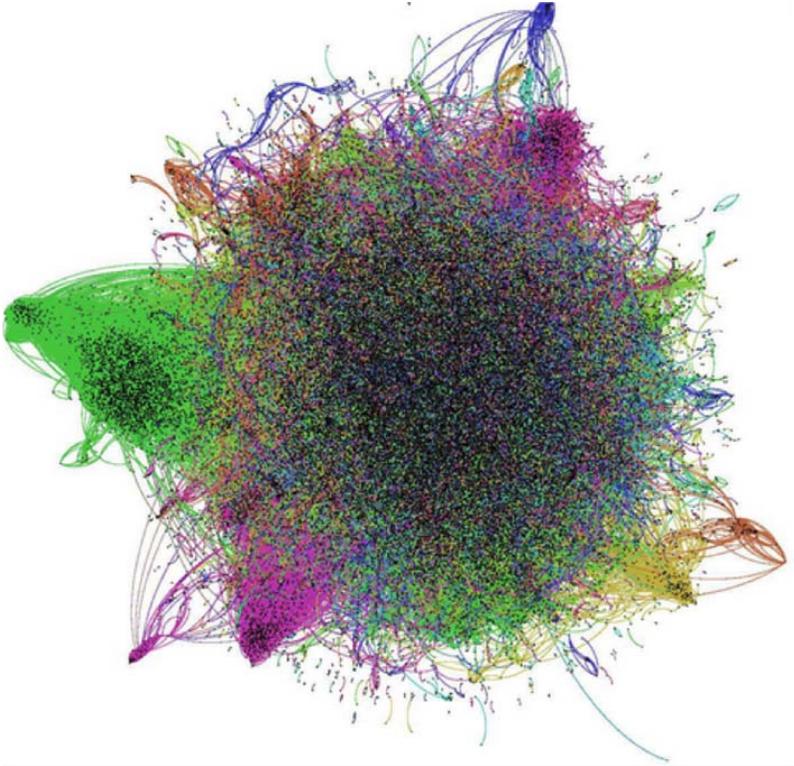
자료: Abdelhamid, et al., 2012

□ 활용 효과

- 버지니아텍에서 시스템을 운영하고 일반 대중에게 공개되었다
- 일부 시스템은 **Open Science Grid**⁵⁴⁾에 맵핑되는 중이다.

54) Open Science Grid 컨소시엄은 과학 연구를 위한 분산 컴퓨팅을 촉진하는 Open Science Grid를 관리하는 조직으로 2004년에 설립되었으며 서비스 및 자원 제공 업체, 대학 및 국립 연구소의 연구자 뿐 아니라 미 전역에 걸친 컴퓨팅 센터와 연계되어 있다. 미국 에너지 부와 NSF의 편당을 받았다.

[그림 19] Brightkite 체크인



자료: http://cinet.vbi.vt.edu/cinet_new/

- 다양한 네트워크 관련 프로젝트가 **CINET** 기반으로 진행되는 중이며, [그림 19]는 위치 기반 소셜 네트워킹 서비스인 **Brightkite**의 2008년~2010년간 4491143개의 체크인을 모아 가시화한 것이다. 기존에 친구 관계 네트워크가 공용 **API**를 통해 수집되었고, 이는 58228개의 노드와 214078개의 연결선으로 구성되었다. 네트워크는 본래 방향성이 있었으나 친구 관계가 양방향으로 형성된 경우 방향성이 없는 연결선으로 네트워크를 구성하였다.

<표 12> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
연구자들이 사용할 마땅한 현실적 통합 대안이 없음	일반 대중에게 네트워크 과학용 플랫폼을 공개

6. 통합 계산 재료 공학 - 미국 마그네슘 프론트엔드 3개국 파일럿 프로그램⁵⁵⁾

□ 산업 소개

- 신소재를 포함한 공업재료의 제조공정, 성질, 구조 상호간의 관계를 컴퓨팅을 이용하여 연구하는 학문

□ 개요

- 미시시과 주립 대학교에서 주도한 이 마그네슘 프론트엔드 3개국 파일럿 프로그램을 위한 프로젝트는 다음과 같은 주요 현안을 해결해야 했다.

- 실험 결과의 수집

- 온라인 모형 보정 틀을 활용하여 특정 재료 모형(Damage Fit, Multistage Fatigue Fit 등)의 재료 특성 도출

- 검증된 메타모형을 구축하는 데 있어 유한 요소 분석시 해당 재료 모형을 사용

- 재료 최적화 설계

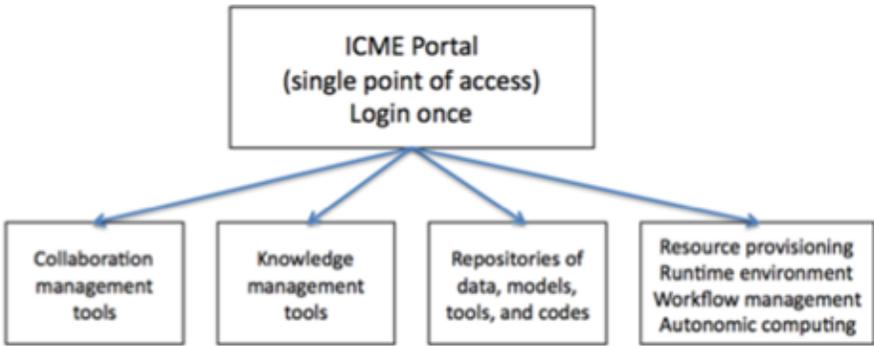
- 통합 계산 재료 공학을 위한 사이버인프라 구축

□ 연구 방법

- 협업 관리 및 지식 관리를 위한 Web 2.0 개념과 기술을 활용한 포털과 위키 구성

55) Haupt, 2010.

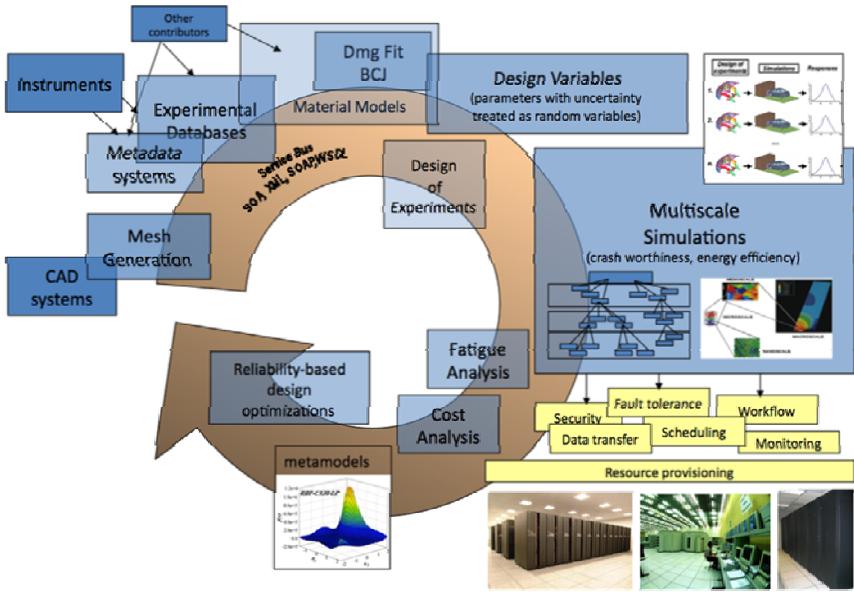
[그림 20] ICME 포털의 아키텍처



자료: Haupt, 2010

- 향후 보안, 구성, 수정 등이 자동화된 사이버 인프라 구축 예정

[그림 21] TO BE 아키텍처



자료: Haupt, 2010

- 재료 특성 저장소(RMP, Repository of Material Properties)는 세 개의 독립된 웹 애플리케이션(실험 데이터 저장소, 재료 상수 저장소, 온라인 모형 보정 툴)을 통합하는 것으로, 데이터 서비스와 계산 서비스로 구성된다.

[그림 22] 저장소의 재료 요약 뷰

Microstructure	Damage	Fatigue	Images	Experimental Data	Strain-Life
• AZ31-LD_ND-01_particle	• MyAZ31.1 • MyAZ31	none	• AZ31-LD-ND-01.TIF	• T1R4N4_A1_DMG.data • MSU • AL7075_MSF.data • AL7075_MSF.data • 205Q-495_MSF.data • MSU_T=298K_RATE=0.001.data.dat	none

자료: Haupt, 2010

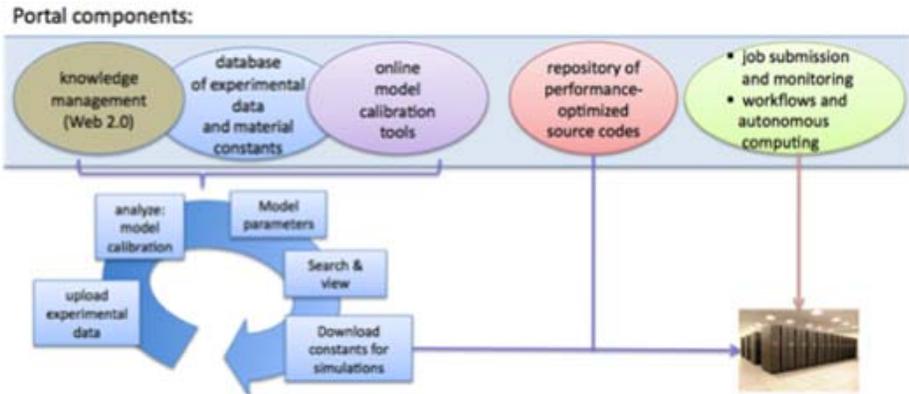
○ 데이터 서비스

- 메타데이터 서비스, 저장 서비스, 복제 위치 추적기를 총괄하며 전반적인 워크플로우는 메타데이터가 생성되고, 파일이 저장 서비스에 업로드되며, 메타 데이터의 위치와 저장된 파일의 위치가 URI로 전송되어 복제 위치 추적기에서 맵핑된다.

○ 계산 서비스

- 모형 보정 툴은 MATLAB 애플리케이션으로 구현되었고 이를 웹 환경에서 구동하기 위한 서비스가 계산 서비스이다. 세 개의 MATLAB 애플리케이션 (테이지 모형, 이미지 분석기, Multistage Fatigue Fit) 인스턴스가 원격으로 접속된다.

[그림 23] 포털 구성 요소의 통합



자료: Haupt, 2010

□ 활용 효과

- 통합 계산 재료 공학을 위한 사이버 인프라로서의 영역이 확대되는 중이다.

<표 13> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
마땅한 현실적 대안이 없음.	신규 사이버 인프라 기술 적용을 통한 데이터 및 계산 서비스 활용

7. 사회 정보학 - 웹 2.0 기반 사회 정보학 데이터 그리드(SIDGrid)⁵⁶⁾

□ 산업 소개

- 사회정보학은 정보기술의 인간적·사회적·조직적 측면에 관한 이론적·경험적 연구를 말한다. 정보기술이 특정 사회적 맥락 속에서 어떻게 설계 (design), 구현(implementation), 사용(use)되며, 그 효과는 무엇인지 설명하는 것으로, 사회 정보학에 대한 경험적 연구는 오늘날 컴퓨터화와 관련된 다양한 사회 문제들에 관해 유용한 통찰을 제공한다.

□ 개요

- NSF(National Science Foundation), OASCR(Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy)의 펀딩 하에 아르곤 국립 연구소와 시카고 주립 대학이 주도적으로 수행한 프로젝트로, 다음과 같은 주요 현안을 해결해야 했다.

□ 주요 현안

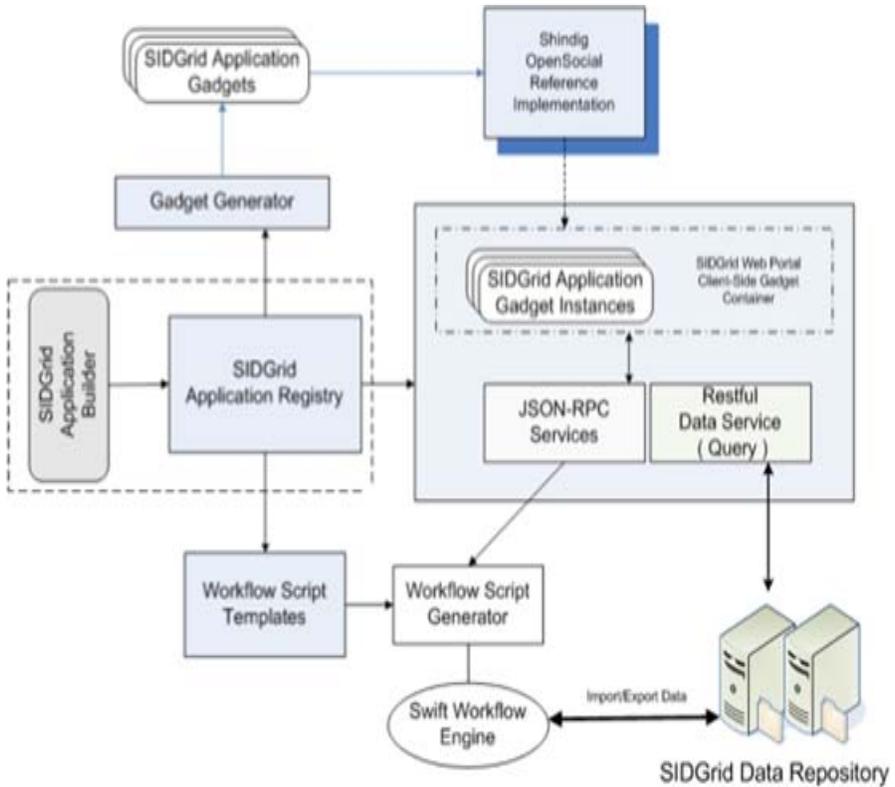
- 일반적인 사회, 행동 과학 연구 툴의 통합
- 데이터 관련 작업이 많은(data-intensive) 애플리케이션을 테라그리드 위에서 효율적으로 수행(주로 텍스트, 이미지, 오디오, 동영상, 및 센서 데이터를 모두 포괄)
- 사용하기 용이한 웹 인터페이스 제공

56) Wu, 2009.

□ 구현 방법

- 구글이 시작한 소셜 네트워킹 프레임워크인 **OpenSocial**을 활용하여 **Web 2.0**기반의 통합을 시도
- 스위프트 스크립트 언어, 테라 그리드 활용

[그림 24] SIDGrid 과학 게이트웨이 기본 구조

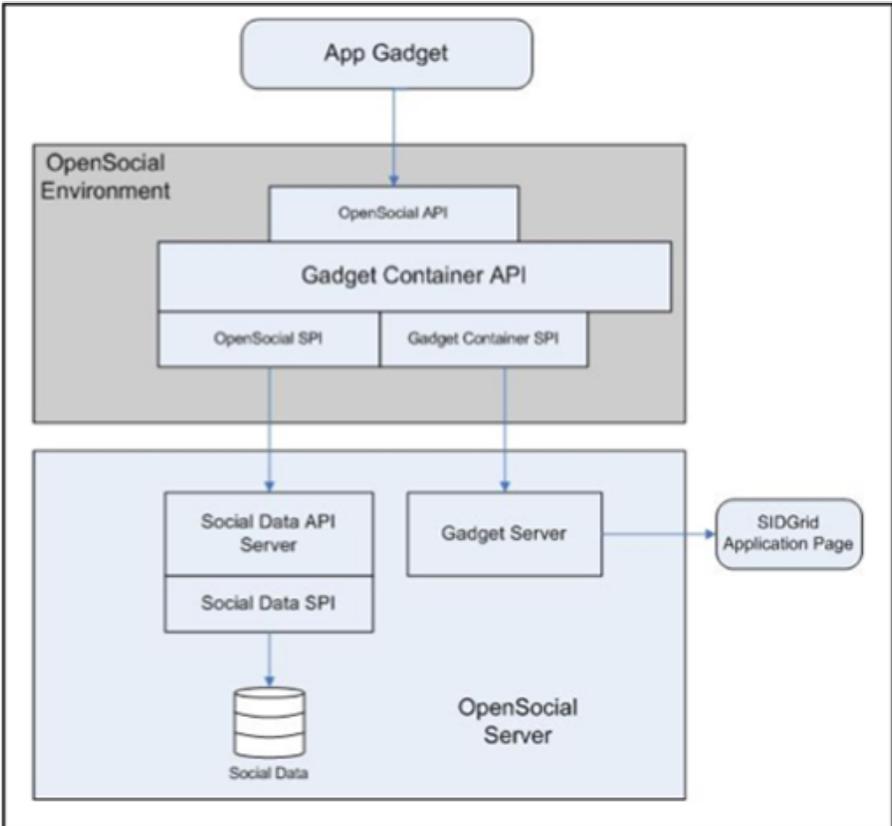


자료: Wu, 2009

- [그림 24]는 **SIDGrid**의 과학 게이트웨이의 기본 구조를 보여준다. **SIDGrid** 애플리케이션 프레임워크는 5개의 기능적 보조 시스템으로 구성된다. 애플리케이션 관리 모듈, 가젯 호스팅 모듈, **SIDGrid** 웹 서비

스 모듈, 워크플로우 실행 모듈, **SIDGrid** 데이터 관리 모듈이 그것이다.

[그림 25] OpenSocial 컨테이너 스택



자료: Wu, 2009

- 특히 **OpenSocial** 컨테이너 내 가젯 서버에서는 가젯 **XML** 파일을 파싱하여 **HTML/JavaScript** 코드와 사용자 코드, **OpenSocial** 지원코드로 렌더링한다. 가젯 서버는 가젯 간의 통신 서비스도 제공한다. **Social Data API** 서버는 외부 저장소의 소셜 데이터를 다룰 수 있도록 해 준다.

- 이러한 **Social Data API**를 활용하여 소셜 네트워킹 애플리케이션을 만들고, 이를 통해 **SIDGrid** 협업을 지원하였다.

□ 활용 효과

- 2009년 사회 과학자 및 행동 과학자들에게 공개되었다.

<표 14> 사이버 인프라 적용 전후 효과 요약

사이버 인프라 적용 전	사이버 인프라 적용 후
마땅한 현실적 대안이 없음.	사이버 인프라 신규 기술을 적용하여 소셜 네트워킹 애플리케이션 활용

제2절. 유럽 각 분야별 SW활용 사례

1. 신경과학분야 - CARMEN⁵⁷⁾

□ 산업 개요

신경과학(neuroscience)은 신경 시스템에 관한 과학으로 전통적으로는 생물학에 속하였으나, 현대에 와서는 화학, 생리학, 약학, 전산학 등을 포괄하는 광범위한 학문 분야이다.

□ CARMEN 개요

CARMEN(Code Analysis, Repository & Modeling for E-neuroscience)은 영국의 공학 및 자연과학 연구위원회(EPSRC: Engineering and Physical Sciences Research Council)의 e-Science의 시범 과제로써, 신경 생리학 분야의 가상 연구 환경을 구성하는 프로젝트이다. 뉴캐슬, 캠브리지 등 11개 대학에서 신경생리학자, 신경정보학자, 전산학자가 참여

57) Watson, et al., 2010.

하고 있다.

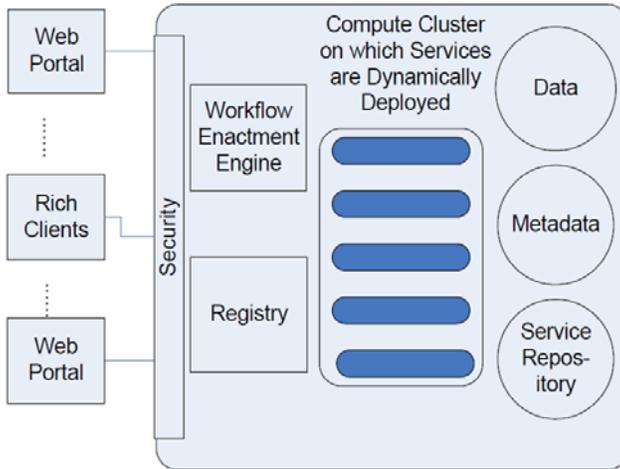
□ SW 활용의 필요성

중추 신경계에서 일어나는 감각 세포 사이의 네트워크에서 일어나는 일을 연구하는 것은 방대한 신경 데이터와 이를 분석할 수 있는 컴퓨팅 파워를 필요로 한다. 각각의 연구자들이 생산하고 분석하는 신경 데이터들은 서로 연관되어 있다. 따라서 데이터를 모아 공유하고 이를 분석할 수 있는 컴퓨팅 파워를 갖춘 연구 환경 구축이 필요했다.

□ SW 활용 방법

CARMEN은 클라우드 기반 연구 포털 서비스로, 데이터의 저장과 분석을 위한 도구 및 환경을 제공한다. CARMEN을 통해 연구자들은 다른 연구자들의 데이터를 검색하고 이용할 수 있으며, 개인의 연구를 CARMEN에 공유할 수 있다. 또한 연구자들은 CARMEN에서 제공하는 분석 도구를 이용하거나, 연구자의 분석 도구를 이용하여 웹상에서 데이터를 분석할 수 있다. 따라서 분석을 위해 테라바이트 단위의 데이터를 업로드 하거나 다운로드 받을 필요가 없으며, CARMEN의 컴퓨팅 파워를 이용하여 빠른 시간에 데이터 분석이 가능하다. 연구자는 CARMEN에 업로드한 데이터나 분석 도구를 다른 연구자들에게 공개 하는 시점을 결정할 수 있어, 자신의 연구를 다른 연구자들로부터 보호할 수 있다.

[그림 26] CARMEN 구조



자료: Watson, et al., 2010.

2. 천문학 - AstroGrid Virtual Observatory⁵⁸⁾

□ 산업 소개와 소프트웨어 활용 필요성

천문학은 우주의 구조, 천체의 현상, 다른 천체와의 관계 등을 연구하는 실증적 학문으로 세계 각지의 천문대에서 관측된 이미지, 스펙트라, 표 등의 다양한 형태로 된 방대한 양의 데이터를 통해 연구 한다. 따라서 이러한 방대한 양의 데이터에 쉽게 접근하고, 처리하여 분석할 수 있는 있는 컴퓨팅 환경이 필요 하였다. 특히 1990년대부터 2000년대에 들어서는 Chandra, XMM-Newton, Spitzer 등과 같은 고해상도 관측 기기들이 등장하면서 데이터가 많아지고 분석이 복잡해 졌다.

이를 위해 2002년에는 International Virtual Observatory Alliance(IVOA)가 설립되었다. IVOA는 아르헨티나, 중국, 프랑스, 독일, 헝가리, 인도 등의 20여개의 VO들이 참여하여

58) <http://www.astrogrid.org/>

천문학 데이터의 표준을 만들어 상호 연동 되도록 하고 있다. AstroGrid는 IVOA의 창립 멤버로써 현재는 AstroGrid가 설립에 기여한 Euro-VO(European Virtual Observatory)가 참여하고 있다.

□ AstroGrid 개요

AstroGrid는 2000년부터 2008년까지 영국에서 진행된 Virtual Observatory(VO) 구축 프로젝트이다. 프로젝트는 e-Science 프로그램의 일환으로 입자물리학 및 천문학 연구위원회(PPARC, Particle Physics and Astronomy Research Council)에서 지원받았다. 또한 유럽 연합의 다양한 VO 프로젝트의 파트너로써 지원을 받았다. AstroGrid의 구축 경험은 유럽 전역의 VO를 잇는 Euro-VO를 구축하는데 도움이 되었다.

구체적으로 AstroGrid는 데이터 센터에 기술적인 인프라 소프트웨어를 지원하고, 천문학자를 위한 파일 저장 공간과 데이터 분석 도구 등을 제작하는 것과 국제적으로 통용되는 프로토콜과 표준을 만드는데 협력하는 프로젝트이다.

□ AstroGrid 구성

AstroGrid는 Java 기반의 웹서비스로써, IVOA의 호환 기준을 적용하였다. AstroGrid는 크게 Server-side와 Client-side로 나뉘며, Server-side의 구성요소는 다음과 같다.

○ Registry

AstroGrid 서비스에서 실행가능 한 데이터와 어플리케이션의 메타데이터의 데이터베이스 리스트이다.

○ VOspace

연구자들이 데이터와 워크플로우를 공유하는 저장 공간이다.

○ Data Set Access

데이터 공급자에게 표준 VO 인터페이스를 통해 데이터를 퍼블리시 할 수 있도록 하는 소프트웨어이다.

○ Astro-Taverna

Taverna 워크플로우 시스템에서 사용하기 위한 구성요소이다.

○ Community

SSO(Single Sign On)을 위한 사용자 관리 소프트웨어이다.

Client-side의 구성은 다음과 같다

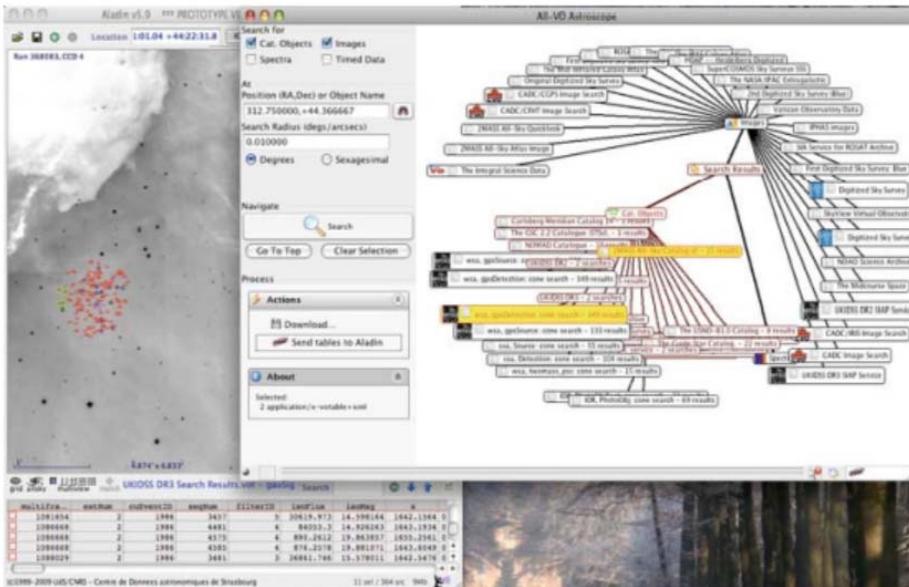
○ VODesktop

그림 34에서 볼 수 있듯이 VODesktop은 VO들을 사용하기 위한 사용자와의 인터페이스이다. 자료 중심으로 설계되어 찾고자 하는 자료를 가진 리소스를 찾고 리소스에서 데이터를 추출해 낸 후 데이터를 연구자의 컴퓨터에 다운로드 받거나 다른 분석 툴에 넘겨 분석을 수행할 수 있다.

○ Astro Runtime

AstroGrid VO 미들웨어로, 다양한 VO 서비스간의 표준 인터페이스를 제공한다.

[그림 27] VODesktop의 사용예시



IC5700 성운 주변의 AstroScope 쿼리를 실행한 모습

□ AstroGrid 활용 성과

AstroGrid를 통해 소프트웨어와 서비스와 더불어 122개의 저널과 컨퍼런스에 퍼블리시 되었으며, 21개의 국제 표준 문서가 만들어졌고, 6개의 유럽연합 design study 보고서가 제출되었다.

3. 지구과학 - GENESI-DEC⁵⁹⁾

□ GENESI-DEC 개요

GENESI-DEC(Ground European Network for Earth Science Inter-operations - Digital Earth Community)는 FP7의 지원을 받은 프로젝트로 Global Earth Observation System of System(GEOSS), International Society for Digital Earth와 같은 디지털 지구과학 이니셔티브의 주요 참여 기관들이 참여한 지구 과학 분야의 인프라 구축 프로젝트이다. 선행 프로젝트인 GENESI-DR을 이어 GENESI-DEC는 인프라 구축이 추가되었다. GENESI-DR은 EC(European Commission과 ESA(European Space Agency)의 주도로 시작하였다.

GENESI-DEC는 다양한 시스템에 있는 인프라를 연결하여 가상 환경에 지구를 표현하고, 연구자들에게 데이터와 분석 도구 및 서비스를 제공한다. 이로써 GENESIS-DEC는 연구자들에게 신뢰 있는 데이터의 제공과 함께, 데이터 편리한 접근과 분석이 가능하게 하였다.

□ 구축 방법

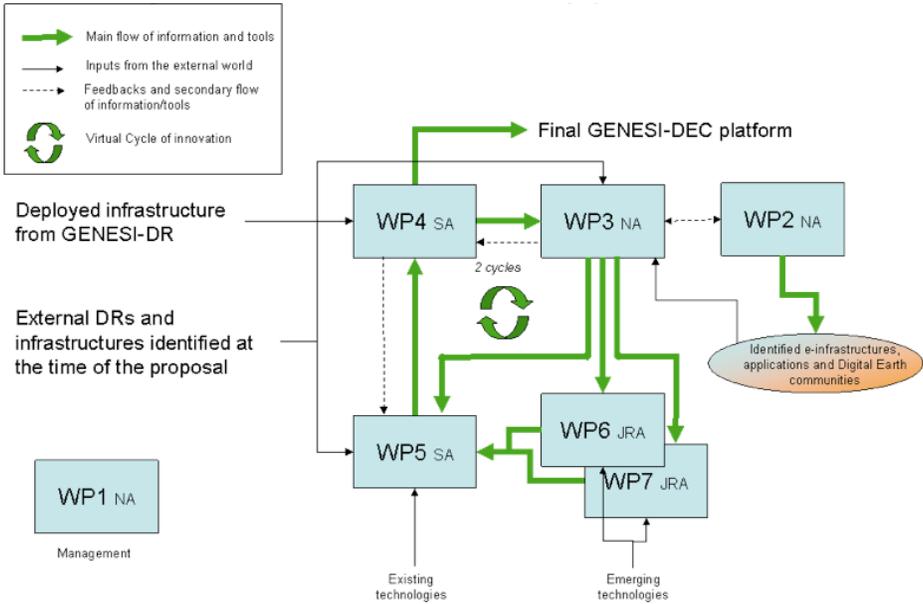
GENESI-DEC 구축에는 [그림 28]과 같은 작업 공정(WP: Working Process) 이 진행되었으며, 내부와 외부의 인프라를 GENESI-DEC로 통합하는 작업들을 반복적으로 진행하였다. WP1은 프로젝트 관리, WP2는 표준화 및 교육, WP3는 외부 인프라 네트워크 활

59) <http://www.genesi-dec.eu/project/>

동, WP4는 배치 및 운영, WP5는 구현과 테스트, WP6는 연구 활동의 통합, WP7은 호환성 및 보안 확보이다.

GENESI-DEC는 그림 36와 같이 해양 관찰, 비행기를 통한 대기 관찰, 국제적 중요 변화, 흑해 해수면 관찰, 공간 계획 분야의 연구자들이 필요한 인프라를 모아 상호 연동되도록 하였다.

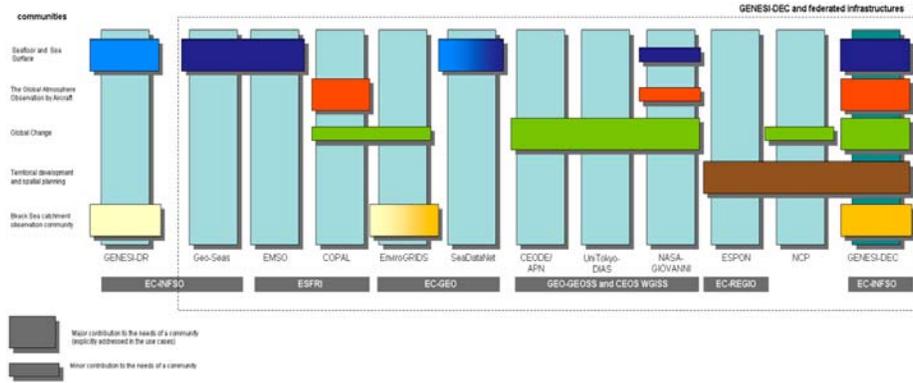
[그림 28] GENESI-DEC 프로젝트 구축 방법



자료: <http://www.genesi-dec.eu/project/>

[그림 29] GENESI-DEC에서 사용할 수 있는 인프라

Framework Programme 7 (2007-2013) Research infrastructures projects



자료: <http://www.genesi-dec.eu/project/>

4. 구조생물학 - WeNMR⁶⁰⁾

□ 핵자기 공명법(NMR Spectroscopy) 개요

핵자기 공명은 자기장 속에 놓인 원자핵이 특정 주파수의 전자기파에 의해 공명하는 현상을 통해 분자의 물리, 화학, 전기적 성질을 알아내는 방법으로 분자 분광법⁶¹⁾의 일종으로 사용되고 있다.

외부의 에너지에 의해 핵의 스핀 상태를 변화시킬 수 있으며, 화학물에 따라서 원자가 가지고 있는 핵의 스핀 강도가 다르기 때문에 같은 세기의 자기장을 걸어줘도 흡수하는 주파수가 다르다. 따라서 화학물이 흡수하는 주파수를 측정함으로써 그 화학물에 어떤 원자가 존재하는지 추론이 가능하다. 이외에도 X선을 이용하는 방법도 있으며 이를 SASX⁶²⁾(Small-angle Xray Scattering)이라고 한다.

□ WeNMR 개요

WeNMR은 자기공명법과 분자구조를 연구하는 생물학의 분야인 구조생물학(Structural biology)의 연구의 e-Infrastructure를 구축하는 프로젝트로 FP7 와 다국적 그리드 프로그램인 유럽 그리드 이니셔티브(European Grid Initiative)의 지원을 받았다. 2010년부터 진행된 프로젝트는 eNMR의 후속연구이다. eNMR서비스를 계승하여 그리드로 확장하였으며, 보완적인 연구와의 호환성을 높였다. 이로써 연구자들은 웹 어플리케이션을 통해 많은 컴퓨팅 파워가 필요한 작업들을 할 수 있게 되었다.

WeNMR은 구조생물학에서의 보완적 연구들을 Virtual research community로 통합하여, NMR과 SASX 데이터 분석과 구조 모델링에 필요한 계산적 접근이 가능하도록 가상 환경을 제공하고 있다.

60) <http://www.wenmr.eu>

61) 분자 분광법이란 물질에서 방출되거나 물질에 흡수되는 스펙트럼을 분석하여 물질을 식별하는 것으로, 핵자기공명법(NMR:Nuclear Magnetic Resonance)이외에도 적외선 분광법, 라만 분광법, X선 분광법, 원소 분광법 등이 있다.

62) 직사광의 둘레에 수도의 회절각 범위 안에 나타나는 산만산란을 작은각 산란(SAS: Small-angle scattering)라 하며 이에 X선을 이용하는 것은 소산각 X선 회절(SASX: Small-angle X-ray Scattering)이라고 한다. 이를 통해 어떤 원자가 존재하는지 추론이 가능하다.

□ WeNMR에서 제공하는 연구 서비스

WeNMR은 크게 NMR 서비스와 SASX서비스로 나뉜다. NMR서비스는 Processing, Assignment Analysis, Structure Calculation, Molecular Dynamics Simulation, Modeling과 그 외의 소프트웨어 툴로 총 25개의 다른 분석 도구를 하나의 시스템으로 제공하고 있으며 표 2와 같다. SASX 서비스는 Ab Initio Shape Determination, Modeling, Instrument Access 분야에서 8가지 분석 도구를 제공하고 있다.

<표 15> WeNMR에서 제공하는 NMR 분석 서비스

분류	서비스명
Processing Assignment	MDD NMR
	Auto Assign
	MARS
	UNIO
Analysis	TALOS+
	AsnisoFIT
	MaxOcc
	iCing
Structure Calculation	CS-ROSETTA
	CYANA
	UNIO
	Xplor-NIH
	CcpNmr WMS
Molecular Dynamics Simulation	AMBER
	GROMACS

분류	서비스명
Modelling	3D-DART
	HADDOCK
Various Software Tools	Format Converter
	SHIFT2X
	RCI
	Antechamber
	Preditor
	RCI
	UPLABEL
	SedNMR

자료: www.wenmr.eu

<표 16> WeNMR에서 제공하는 SAXS 분석 서비스

분류	서비스명
Ab Initio Shape Determination	DAMMIN
	DAMMIF
	GASBOR
	MONSA
Modelling	CRY SOL
	SASREF
	EOM
Instrument Access	EMBL' s SAXS beamline P12 remote ccess

자료: www.wenmr.eu

□ WeNMR 활용효과

WeNMR을 구조분자학에서 가장 큰 가상 연구 커뮤니티로 발전하였으며, 2007년부터 시작한 프로젝트는 2011년 현재 272개의 WeNMR 전용 CPU와 2.87TB의 데이터가 저장되어 있으며, 공유되는 곳에서는 10,000개의 CPU와 37TB의 데이터가 저장되어 있다. 1년동안 백만개가 넘는 작업들이 그리드를 통해 수행되고 있으며, CPU 효율성은 99%에 이른다.

[그림 30] WeNMR중 MDD MNR 서비스 화면,

MDD NMR

Input Files and Parameters

Files Upload

Archive/Zip/tar file with fid/ser files

nmrPipe script for initial FID conversion and FT of the direct dimension :

nmrPipe script to FT remaining dimensions in reconstruction after mdd has produced full matrix :

Choose the type of Schedule File

Bruker NUS
 Varian NLS

Choose the type of Processing

MDD
 sparseFT (this is DFT of sparse data, it is meant only for quick test / checking data / phases etc.)

Enter Processing parameters

Leftmost (high field) point of region of interest (ROI) (ppm)

ROI size (ppm) :

Recommended size of sub-region (ppm):

Number of iterations in mddnmr (30-50 for preliminary and 100-500 for final calculations) :

Default number of components in sub-region (10-30 for small-medium proteins) :

Calculation parameter Lambda (0.001-0.05 with higher value for high S/N) :

Memory allocation (Mb): 200 for small size of FID (<1 day of recording) and 500 for 4D (>2 days) :

Extra Parameters

More Parameters & Files

MDD는 NUS(Non-uniformly sampled) nD 스펙트럼을 Multi-Dimensional Decomposition Program을 사용하여 프로세스 하게 한다. 비어있는 데이터들을 FT 프로세싱을 통해 완벽한 데이터로 만든다.

5. 과학 작업 흐름 관리 - Taverna⁶³⁾

□ 산업 소개

작업 흐름 관리 시스템(Workflow Management System)은 과학의 연구에서 데이터의 수집, 처리, 계산과 같은 여러 단계에 걸쳐 있는 작업의 흐름을 디자인하고 만들어진 흐름과 같이 실행 할 수 있는 시스템이다. e-Science와 그리드 환경이 구축되면서 과학자들이 접근할 수 있는 지식과 데이터들이 많아졌기 때문에 데이터의 유형이나, 분석 도구 등을 한 번에 접근할 수 있는 작업 흐름 관리 시스템이 필요하게 되었다.

대표적인 작업 흐름 관리 시스템으로 Taverna, Discovery Net, Kepler 등이 있으며, 작업 흐름을 시각화된 도구로 디자인 할 수 있어서 과학자들은 모델화하고 결과를 분석하는데 더욱 집중할 수 있다.

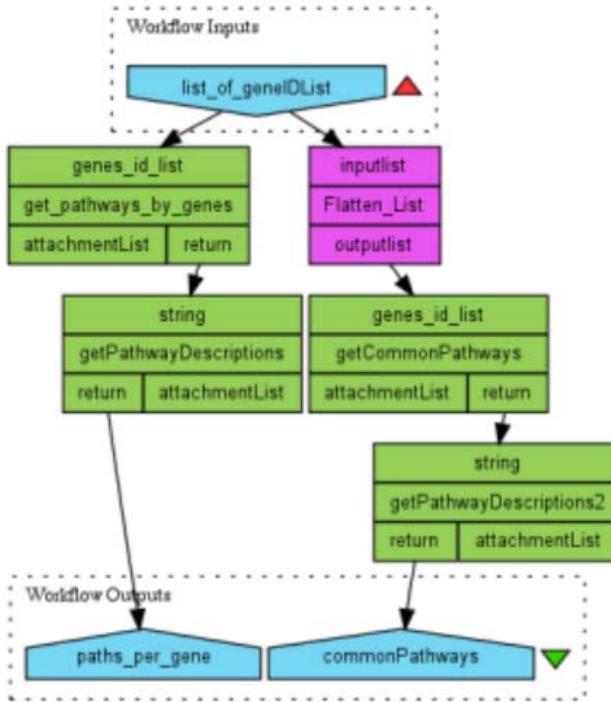
□ Taverna 개요

Taverna는 myGrid 연구팀에 속해 있는 맨체스터 대학교의 전산학과 Carole Goble 교수의 연구팀이 개발한 작업흐름 관리 시스템이다. myGrid 연구팀은 공학 및 자연과학 연구위원회, 생명공학 연구위원회, FP7, Jisc, IMI(Inovative Medicine Initiative) 등 의 지원을 받는 연구팀으로 맨체스터 대학교를 포함하여, 사우스햄튼, 옥스퍼드 대학이 참여하고 있다. Taverna를 포함하여, 단백질 시퀀스 분석 도구인 utopia, 세만틱 분석 도구인 RightField 등 생물학, 천문학, 화학, 등 다양한 분야의 e-Science 툴을 제작하고 있다.

Taverna는 생물정보학(Bioinformatics), 천문학, 생물다양성 등의 연구에서 작업 흐름을 관리하고 흐름과 같이 분석을 실행 할 수 있는 프로그램이다. Taverna를 사용한 작업의 흐름은 [그림 31]과 같이 나타난다. LGPL(Lesser General Public License)기반의 오픈 소스 프로그램으로 자바를 기반으로 하며, Taverna의 구조는 [그림 32]와 같다.

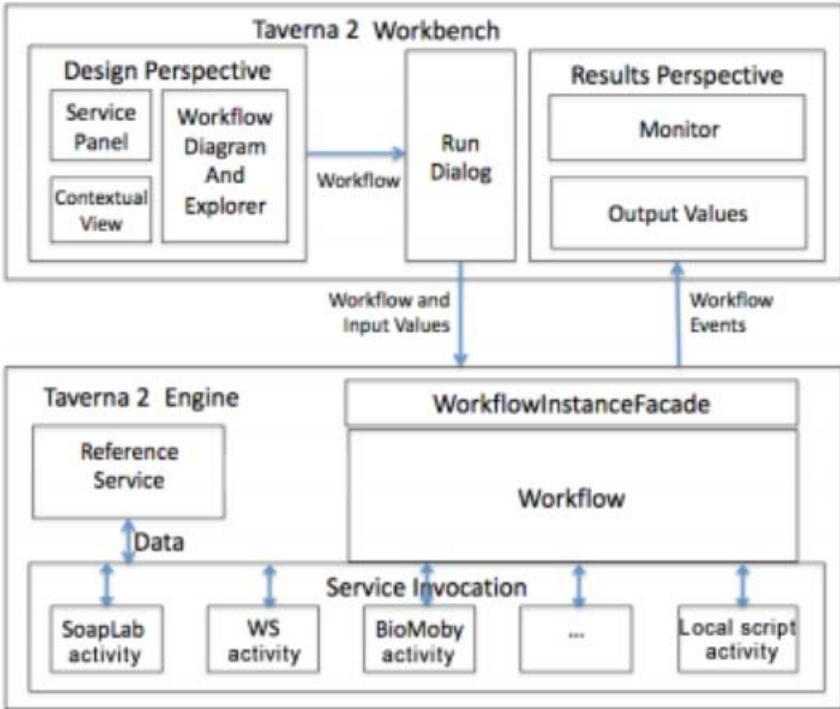
63) Missier, et al., 2010.

[그림 31] Taverna의 작업 흐름 예시



자료: Missier, et al., 2010

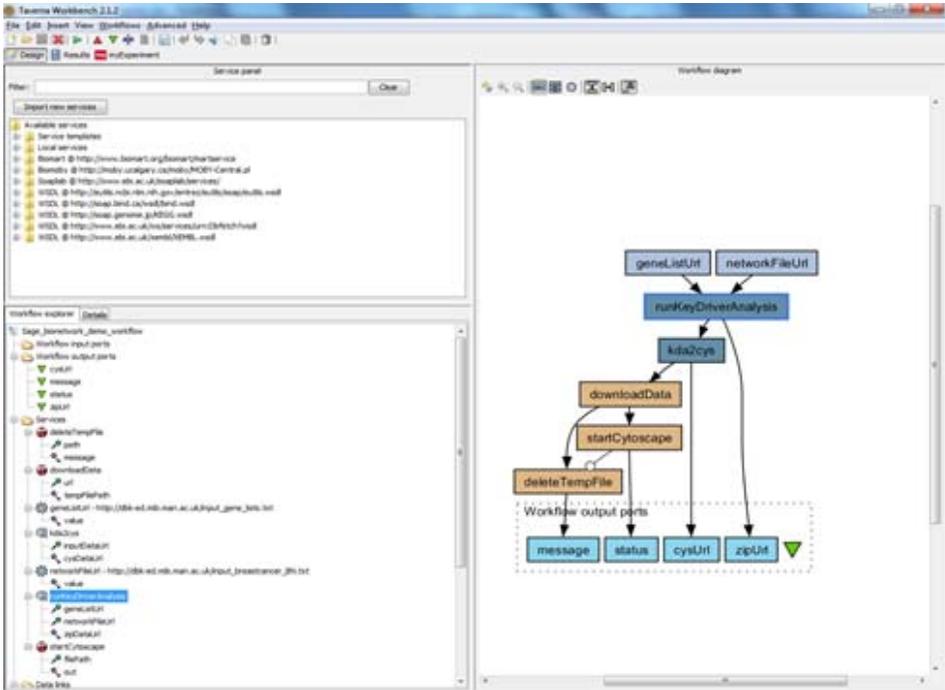
[그림 32] Taverna 구조



자료: Missier, et al., 2010

Taverna를 실행 화면은 [그림 33]과 같으며, 작업 흐름을 만들고 입력과 출력을 지정 하면 작업의 흐름대로 지정한 데이터 분석 서비스들에 지정한 데이터가 자동으로 입력 되고 여기서 나오는 결과를 볼 수 있다. 연구자들은 Catalogue of Life, BioSTIF, eFamily 등 곳곳에 흩어져 있는 데이터를 한곳에서 이용할 수 있으며, 데이터의 흐름을 시각화 된 도구를 이용하여 관리하고 분석할 수 있다. 여러 종류의 데이터의 호환성을 손쉽게 유지할 수 있으며 각각의 작업 흐름을 추적할 수 있다. 또한 다른 연구자가 디자인한 작업 흐름과 서비스 등을 공유할 수 있다.

[그림 33] Taverna Workbench 구현 모습



자료: <http://www.taverna.org.uk>

□ Taverna 활용 효과

Taverna는 2004년에 서비스를 시작한 이후 2014년 6월 9만 건이 넘는 다운로드가 이루어졌으며, 약 380 곳의 연구기관이 사용하고 있다

6. 금석학 - VRE SDM⁶⁴⁾

□ 금석학

금석이나 비석에 쓰여 있는 명문을 연구하는 학문으로, 고고학과 문헌학의 중간 학

64) De La Flor, et al., 2010.

문으로 역사학, 문학, 언어학 등의 보조과학이다.

□ VRE SDM 개요

VRE SDM(Virtual Research Environment for Humanities for the Study of Document Manuscripts)은 Jisc(Joint Information Systems Committee)의 후원으로 옥스퍼드 대학에서 진행하였으며, 금속과 비석에 쓰여져 있는 명문을 해독하는데 도움을 주는 SW를 개발하는 프로젝트이다.

□ 금속학에서 SW활용의 필요성

금속학 연구는 비석이나 금속에 쓰여 있는 글자를 적외선 투사, 그림자 분석 등을 통해할 수 있으며, 분석 방법에 따라 결과의 차이가 있기도 하다. 비석이나 금속에 쓰여 있는 문자는 가독성이 떨어지는 경우가 대부분이며, 시대에 따라 글자의 모양도 달라질 수 있다. 따라서 금속학자들은 문자를 분석하기 위해 다양한 분석 방법을 사용하면서, 기존의 데이터와 비교하고 금속학자들 간의 의견을 수렴할 수 있는 협력 작업이 필요하였다.

□ VRE SDM 구축 방법

VRE SDM은 금속이나 비석을 이미지화하여 저장하고, 밝기와 채도 등을 조정하여 다양한 방법으로 볼 수 있게 하였다. 또한 금속학자들 간의 협력을 강화하기 위하여 이미지를 공유할 수 있도록 하였다. 나아가 이미 구축되어 있는 데이터에서 분석하고자 하는 문자와 비슷한 문자들을 찾고 보여줌으로써, 금속학자들이 비슷한 문자를 찾기 위해 노력을 덜 수 있게 되었다. VRE SDM의 프로토타입은 [그림 34]에서 확인 할 수 있다.

□ VRE SDM 활용 효과

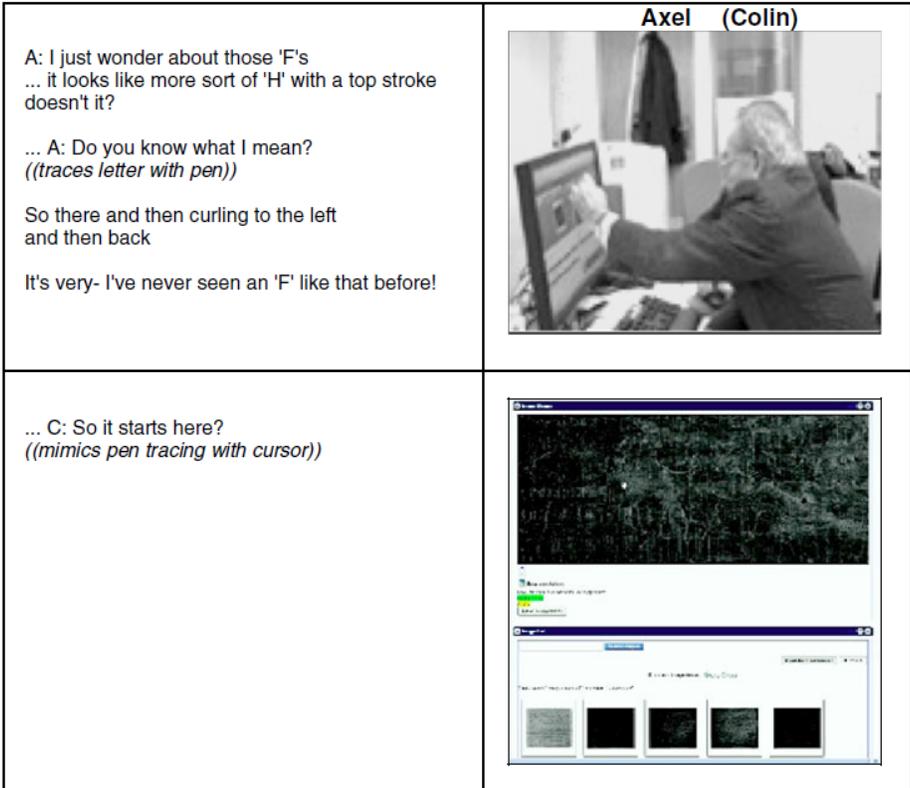
VRE SDM의 활용효과는 [그림 35]와 같다. 연구자들은 각자의 해석에 대해 불필요한 논의를 줄일 수 있게 되었고 이는 해석의 정확도를 높이는 결과를 가져왔다.

[그림 34] VRE SDM 프로토타입



자료: De La Flor, et al. 2010.

[그림 35] VRE SDM을 통해 문자를 해독하는 과정



자료: De La Flor, et al. 2010.

7. 인류학 및 사회과학 - CLARIN⁶⁵⁾

□ CLARIN 개요

CLARIN(Common Language Resources and Technology Infrastructure)은 유럽 연구 인프라 전략 포럼(ESFRI: European Strategy Forum on Research Infrastructures)에 선정된 연구 인프라이다. 연구자들이 언어와 관련된 인류학 및 사회학 데이터를 자유롭게 접근

65) <http://www.clarin.eu>

하고 향상된 분석 도구를 제공하는 것이 CLARIN의 목표이다.

CLARIN은 연구 기관이기 보다는 국제적인 연구 네트워크이며, 이 연구 네트워크는 CLARIN-ERIC(European Research Infrastructure Consortium)에 의해 관리 및 유지된다. CLARIN-ERIC은 오스트리아, 불가리, 체코, 덴마크, 에스토니아, 독일, 네덜란드, 폴란드가 참여하고 있다.

□ SW활용 방법

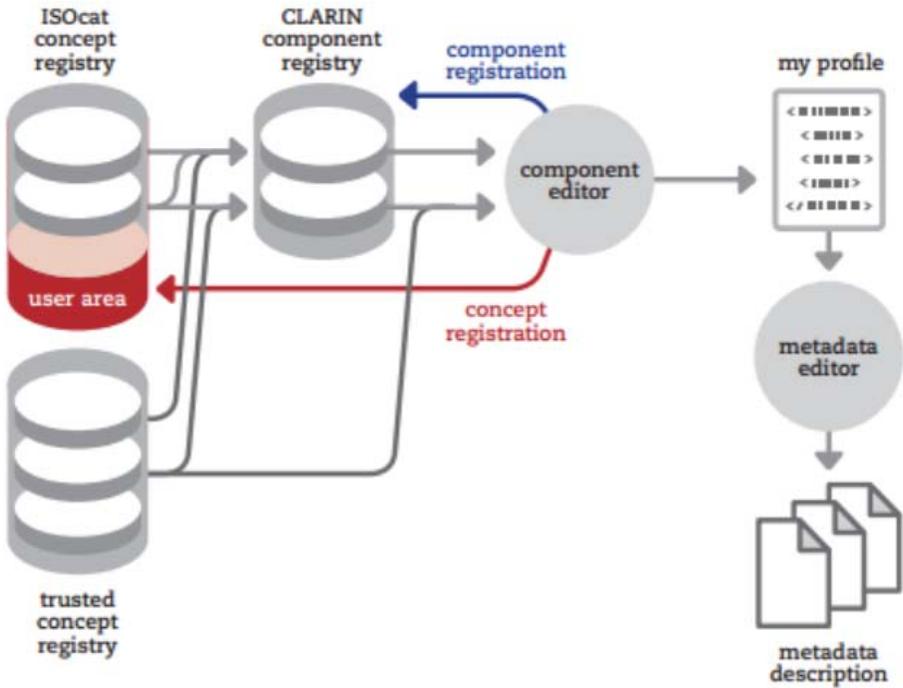
CLARIN은 연구자들이 원하는 언어학 자료를 쉽게 찾고, 이를 법적인 문제없이 이용할 수 있도록 한다. 메타 데이터를 기반으로 자료를 검색할 수 있으며, 여러 국가가 가지고 있는 언어학 자료의 시맨틱 호환성을 유지하면서 여러 메타 데이터를 함께 사용할 수 있는 통합 환경을 제공한다.

연구자들은 CLARIN에 데이터를 업로드 할 수도 있으며, 이 자료들은 CLARIN을 통해 분류되어 연구자들의 데이터 활용에 도움을 준다. 이로써 CLARIN은 개별 연구자들의 언어학 자료가 유실 되지 않도록 하며, 다른 연구자들이 이를 검색하여 새로운 연구에 활용할 수 있도록 한다.

또한 CLARIN은 Sentence Splitters, Tokenizer, Part-of-Speech Triggers, Morphological analyzer and lemmantizer, Syntax-parsers and chunker등 다양한 SW 분석도구를 제공한다. 이러한 분석은 CLARIN에 저장되어 있는 언어학 관련 자료들을 웹 기반에서 분석할 수 있어서 연구자들이 방대한 양의 데이터를 분석을 위해 다운로드 받지 않아도 된다.

CLARIN의 메타데이터 프레임워크는 [그림 36]과 같으며, 표준화된 스키마(Schema)를 요구하는 것이 아니라 ISOcat 데이터 카테고리 레지스트리를 사용하며 이를 기반으로 사용자가 직접 메타데이터를 만들고 수정할 수 있다. 검색은 ISOcat 레지스트리에서 다양한 프로파일을 기준으로 하게 된다.

[그림 36] CLARIN Component 메타데이터 프레임워크



자료: <http://www.clarin.eu>, CLARIN ShortGuide,

8. 인문 예술 분야 - DARIAH⁶⁶⁾

□ DARIAH 개요

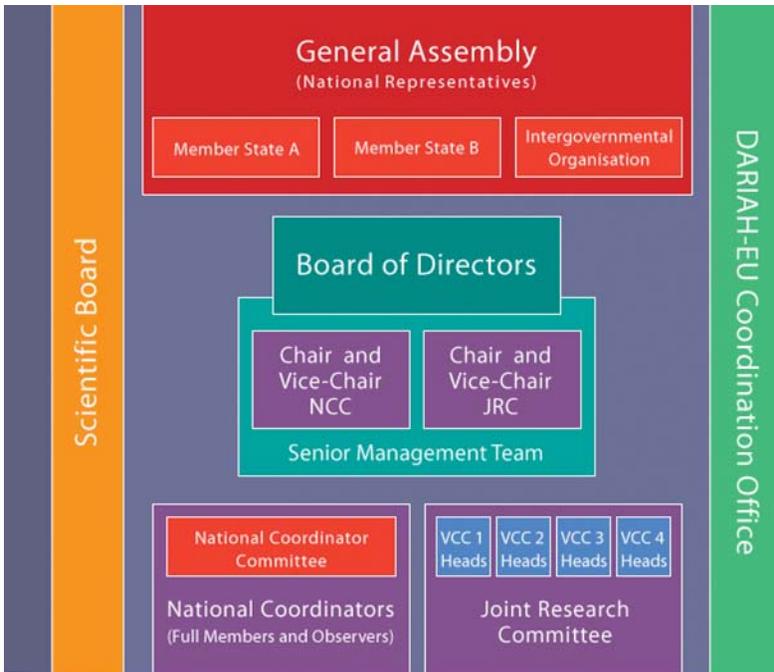
DARIAH(Digital Research Infrastructure for the Arts and Humanities)는 유럽 연구 인프라 전략 포럼(ESFRI: European Strategy Forum on Research Infrastructures)에 선정된 연구 인프라이다. DARIAH는 인문 예술 분야에서 ICT를 활용한 연구를 지원하며, 유지하는 역할을 한다.

66) <http://www.dariah.eu/>

□ DARIAH의 구성

DARIAH는 연구 기관이기 보다는 DARIAH-ERIC(European Research Infrastructure Consortium)에 의해 유지되는 연구 네트워크이다. 이 네트워크는 오스트리아, 벨기에, 프랑스, 독일, 그리스 등 15개 국가로 구성된 총회(General Assembly), 이사회, 코디네이터 등의 기구들로 구성되며 [그림 37]과 같다. DARIAH의 조직의 Virtual Competency Centre(VCC)는 e-인프라, 연구 및 교육, 학문적 연대 등의 목표를 가지고 있으며, 연구 인프라 구축은 VCC의 하부 조직으로 운영된다.

[그림 37] DARIAH 기구 구성



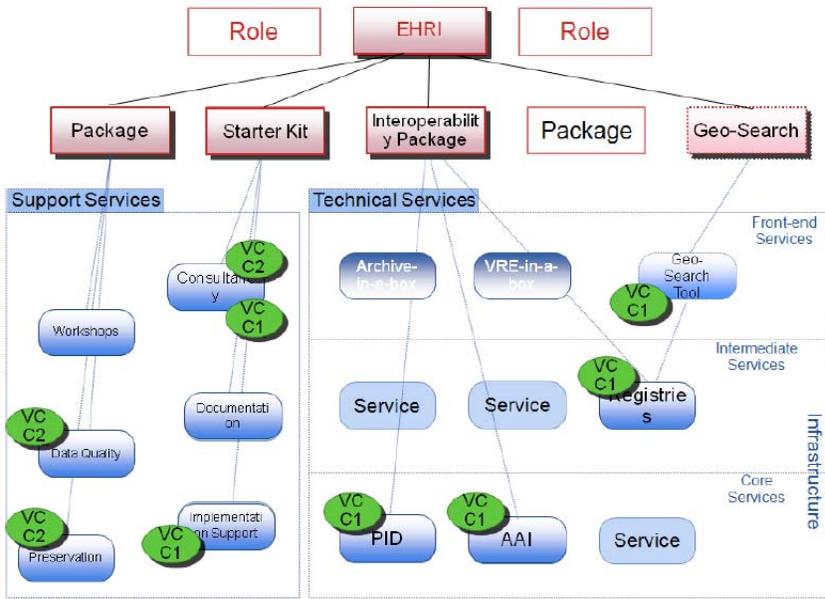
자료: <https://dariah.eu/about/organisation.html>

□ SW 활용 방법

DARIAH는 인류예술 분야 개별 연구자들이 생산한 데이터와 분석도구 등을 공유할 수 있도록 도움을 준다. DARIAH는 PIDs(Persistent Identifiers)를 통해 데이터를 저장하고 공유할 수 있도록 하며, 메타 데이터를 통한 레퍼런싱 기능 등을 제공한다.

DARIAH는 유럽 홀로코스트 연구 인프라 (EHRI: European Holocaust Research Infrastructure) 와 유럽 공동 디지털 아카이브 인프라(CENDARI: Collaborative European Digital Archive Infrastructure)를 구축하였다. EHRI는 유럽에서 일어난 학살과 관련된 연구를 진행하는데 있어, 연구자간 협력을 위한 데이터 저장소를 구축하였으며, 이로써 데이터의 신뢰성 확보하였으며, 또한 학살과 관련해 사회적으로 민감한 자료들이 외부로 유출되지 않도록 방지하였다. DARIAH는 CENDARI와 함께 중세와 현대 유럽 사회의 현존하는 역사 연구 자료들을 하나로 모았으며 이를 정리하여 연구자들이 쉽게 찾아 볼 수 있도록 하였다.

[그림 38] DARIAH를 통해 EHRI의 e-인프라 구축



자료: DARIAH ERIC - Technical and Scientific Description, 2011.

http://www.huma-num.fr/sites/default/files/dariah_scientific-technical-description.pdf

제3절 시민 과학의 SW활용 사례

Citizen Science는 Crowd science, Crowd-sourced science, civic science, networked science 등 다양한 이름으로 알려져 있으며, 과학자가 아닌 비전문가나 아마추어가 데이터를 모으고 분석하여 과학적 분석을 하는 것을 말한다.

소프트웨어 기술의 발전으로 시민간의 협업이 쉬워지고 데이터의 공유가 일반화 되면서 Citizen Science에 더 많은 기회가 생기고 있다. 나아가 일반 대중들이 발견한 과학적 지식을 과학자가 검증하거나, 과학자가 확보해 놓은 데이터 등을 일반 대중에 공유하여 일반 대중이 과학적 발견에 도움을 주는 경우도 늘어가고 있다. 시민들이 과학 기

술을 직접적으로 연구하는 것을 넘어 간단한 게임(EyeWire, FoldIt, EteRNA)을 통해 새로운 과학적 지식을 발견하는 것을 돕기도 한다.

<표 17> 진행 중인 Citizen Science의 예시

프로젝트명	분야	지원	시작일
AgeGuess	생물학	Center for Research and Interdisciplinarity, University of Southern Denmark	2012
Agent Exoplanet	천문학	Las Cumbres Observatory Global Telescope Network	
Amphibian Migrations and Road Crossings	동물학, 생태학	Department of Environmental Conservation, NY, USA	
Asteroid Mappers: Vesta	행성천문학	CosmoQuest.org	2011
Audubon Christmas Bird Count	조류학	National Audubon Society	1900
B.C. Cetacean Sightings Network	고래학	Vancouver Aquarium, Fisheries and Oceans Canada	1999
BirdSleuth K-12	조류학	Cornell Lab of Ornithology	2004
Budburst	생물계절학	National Science Foundation	2007
BugGuide	곤충학	Iowa State University	2003
Bumble Bee Watch	곤충학	Xerces Society, University of Ottawa, Wildlife Preservation Canada, Natural History Museum, London, BeeSpotter, Montréal Insectarium	2014

프로젝트명	분야	지원	시작일
Canadian National Wildlife Disease Surveillance	야생 동물, 건강	Canadian Wildlife Health Cooperative	1992
Celebrate Urban Birds	조류학	Cornell Lab of Ornithology	
Cicada Watch: Brood II	곤충학	National Geographic Society	2013
Clumpy	생물학, 식물학	University of Exeter	2012
Condor Watch	조류학	Zooniverse collaboration	2014
Cooperative Observer Program	기상학, 기후학	National Weather Service	1890
eBird	조류학	Cornell Lab of Ornithology, National Audubon Society	2002
eButterfly	곤충학	University of Ottawa, Oregon State University, Vermont Center for Ecostudies	2010
EDDMapS	생물학, 식물학, 곤충학	University of Georgia Center for Invasive Species and Ecosystem Health	2005
eShark	해양 생태, 해양 관리, 관광	Crowd-sourced	2005
EyeWire	신경 과학	MIT, Gatsby Charitable Foundation, NIH, HHMI, Max Planck Society	2012

프로젝트명	분야	지원	시작일
Flying ant survey	생물학, 곤충학, 야생동물	Society of Biology, University of Gloucestershire,	2012
Fraxinus	유전학	Sainsbury Laboratory, BBSRC, DEFRA	2013
FrogWatch USATM	보존, 생태학	Association of Zoos and Aquariums	1998
Galaxy Zoo	천문학	Zooniverse collaboration	2007
Garden Wildlife Health	보존, 수의학	Zoological Society of London, British Trust for Ornithology, Froglife, Royal Society for the Protection of Birds	2013
Great Backyard Bird Count	조류학	National Audubon Society, Cornell Lab of Ornithology	
Genographic Project	집단 유전학	National Geographic Society, IBM	2005
Hazelnut Project, The	식물학, 농업 경제학	Hybrid Hazelnut Consortium (Arbor Day Foundation, University of Nebraska-Lincoln, Rutgers and Oregon State)	2000
Herbarium@home	식물학	Botanical Society of Britain and Ireland	2006
Herpetological Education and Research Project	파충류 학	HERP Board of Trustees	2007

프로젝트명	분야	지원	시작일
Hummingbirds at Home	조류학	National Audubon Society	2013
IceWatchUSA[6]	생물 기후학	Nature Abounds	
iNaturalist	식물학, 곤 충학, 생태 학, 조류학 , 생물 계절 학, 동물학	iNaturalist	2008
iSeahorse	어류 학, 생태학 , 해양 생물학, 해 양 보존, 동물 학	Project Seahorse, Guylian Belgian Chocolate, iNaturalist, John G. Shedd Aquarium, The University of British Columbia, Zoological Society of London	2013
Loss of the Night app	천문학, 생 태학, 원격 탐사	Federal Ministry of Education and Research (Germany), Freie Universität Berlin, Leibniz Institute of Freshwater Ecology and Inland Fisheries	2013
Monarch Health	곤충학, 생 태학	University of Georgia	
Monarch Larva Monitoring Project	곤충학, 생 태학	University of Minnesota	
Monarch Watch	곤충학, 생 태학	University of Kansas	1992

프로젝트명	분야	지원	시작일
Moon Mappers	행성 천문학	CosmoQuest.org	2011
Nature's Calendar	생물 기후학	Woodland Trust	
Nature's Notebook	생물 기후학	USA National PhenologyNetwork	2009
NatureWatch	생태학, 생 물 계절학	Nature Canada	
NatureWatch NZ	식물학, 곤 충학, 생태 학, 조류학 , 생물 계절 학, 동물학	New Zealand Bio-Diversity Recording Network Trust	2006
Neighborhood Nestwatch	조류학	Migratory Bird Center, National Zoological Park, SI	
NestWatch	조류학	Cornell Lab of Ornithology	2008
North American Field Herping Association	파충류 학	North American field herping community	2007
Notes from Nature	식물학, 곤 충학, 조류 학	Zooniverse collaboration	
Old Weather	기후학	Zooniverse collaboration	
Orca Game	Bioacoustic s	Orchive	2013

프로젝트명	분야	지원	시작일
Planet Mappers: Mercury	행성 천문학	CosmoQuest.org	2011
Portland Urban Coyote Project	생태학, 지리학, 동물학	Portland State University	2011
Quantum Moves	양자 물리학, 양자 역학	CODER, Aarhus University	2013
Radio Galaxy Zoo	천문학	Zooniverse collaboration	2013
Snowtweets	기상학	University of Waterloo	
Species Observations System	식물학, 곤충학, 생태학, 조류학, 생물계절학, 동물학	Norwegian Biodiversity Information Centre, Norwegian Biodiversity Network	2008
Target Asteroids!	행성 천문학	OSIRIS-REx asteroid sample return mission	2012
Target NEOs!	행성 천문학	The Astronomical League	2012
UBiome	미생물학	uBiome	
VerbCorner	언어학	Massachusetts Institute of Technology	2013
WildlifeLog.org	야생 생물	WildlifeLog Community Interest Company	2014

프로젝트명	분야	지원	시작일
Stardust@Home	천문학	University of California, Berkeley	2006
Yellowhammer Dialects	조류학	Charles University in Prague	2013

자료: http://en.wikipedia.org/wiki/List_of_citizen_science_projects

1. Folding@home 프로젝트

□ 산업 소개

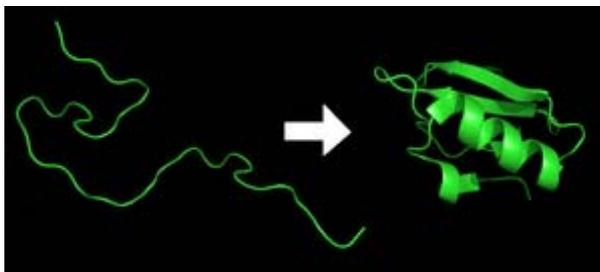
생명의공학은 의학과 공학 분야 간에 학문적 방법론, 개념, 기술, 기기 등을 상호 교환, 응용함으로써 궁극적으로 두 학문 분야의 발전을 도모하는 학문이다.

□ 개요

Folding@home은 단백질 폴딩⁶⁷⁾, 고도의 계산이 필요한 약물 설계나 기타 다른 유형의 분자 역학을 시뮬레이션하는 질병 연구를 위한 분산 컴퓨팅 프로젝트이다. 이 프로

67) 단백질 분자를 구성하는 펩티드 사슬이 고유의 2차구조의 배치순서로 공간적 배치를 통해 고차구조를 형성하는 것.

[그림 39] 단백질 폴딩 전 후 모습



<http://en.wikipedia.org/wiki/Folding@home>

젝트의 목적은 알츠하이머, 헌팅턴, 파킨슨 질병을 비롯, 다양한 암을 치료하기 위한 것이다. 이를 위해, 단백질 폴딩의 메커니즘을 연구하여 단백질의 최종 3차원 입체 구조를 파악하고 잘못된 폴딩의 원인을 파악하고, 이를 약물 설계에 응용하기 위한 노력을 하고 있다.

이 프로젝트는 자신의 시스템에 해당 소프트웨어를 설치한 자원 봉사자들이 소유한 개인용 컴퓨터의 유휴 처리 리소스를 활용한다. 2014년 7월 현재 179237명이 참여하고 있다. 이 프로젝트는 스탠포드 판데 실험실에서 비제이 판데의 지휘 하에 개발 및 운영되고 있으며 전 세계에 걸쳐 다양한 과학 기관 및 연구소(노틀담 Izaguirre 랩, 버지니아 주립대학, 스톡홀름 대학, 콜로라도 주립대학, HKUST, CSULB, 템플 대학, 크로아티아 생명 과학 지중해 연구소 등)에 공유되었고, 다양한 상업적 파트너(인텔, 구글, 소니, ATI, NVIDIA, 스톤하퍼 등)와 커뮤니티 봉사자들의 참여로 진행되고 있다.

□ 구현 방법

이 프로젝트는 자원 봉사자들이 소유한 개인용 컴퓨터의 유휴 처리 리소스를 활용하여 분산 컴퓨팅 구조를 구성한다. Folding@home의 소프트웨어는 사용자 관점에서 작업 단위와 코어 및 클라이언트로 구성된다.

○ 작업 단위

작업 단위는 클라이언트가 처리해야 하는 단백질 데이터이다. 작업 단위는 마르코프 상태 모형 내 상태들 간의 시뮬레이션 중 한 부분이다. 자원 봉사자는 이 작업 단위를 다운로드 받아 단백질 폴딩을 실험 및 완료하고 Folding@home 서버에 다시 업로드하며, 이 과정이 완료되면 자원 봉사자가 크레딧을 얻는다. 이 순환 주기는 자동으로 반복된다. 모든 작업 단위는 작업 기한이 있으며 기한이 지난 경우 자원 봉사자는 크레딧을 얻을 수 없으며 자동으로 다른 참가자에게 재할당된다. 이러한 합리적인 작업 기한이 있기 때문에 전반적인 시뮬레이션 과정이 정상적으로 수행될 수 있고, 최소 시스템 요구 사항이 펜티엄 3 450MHz CPU(스트리밍 SIMD 확장 유닛 포함) 정도로 낮다. 고성능 클라이언트에 대한 작업 단위는 단일 프로세서에 대한 기한보다 짧아, 신속하게 시뮬레이션을 완료할 수 있도록 한다.

○ 코어

“FahCores” 혹은 약칭으로 “코어(Cores)”라 불리는 특화 분자 역학 프로그램은 작업 단위의 연산을 수행하는 백그라운드 프로세스이다. 코어의 대부분은 분자 역학 소프트웨어 패키지 중 가장 빠르고 유명한 GROMACS에 기반하고 있는데, 이 GROMACS는 수동 최적화된 어셈블리 코드로 대부분 이루어져 있고 하드웨어 최적화가 포함되어 있다. GROMACS는 오픈소스이지만 데이터의 유효성을 보장하기 위해 폐쇄 소스 라이선스를 사용한다. 코어는 클라이언트와는 별개로 자동으로 업데이트 되는 프로그램이다. 코어는 주기적으로 연산 체크포인트를 생성하며, 중단된 경우 해당 체크포인트부터 작업을 다시 시작할 수 있다.

○ 클라이언트

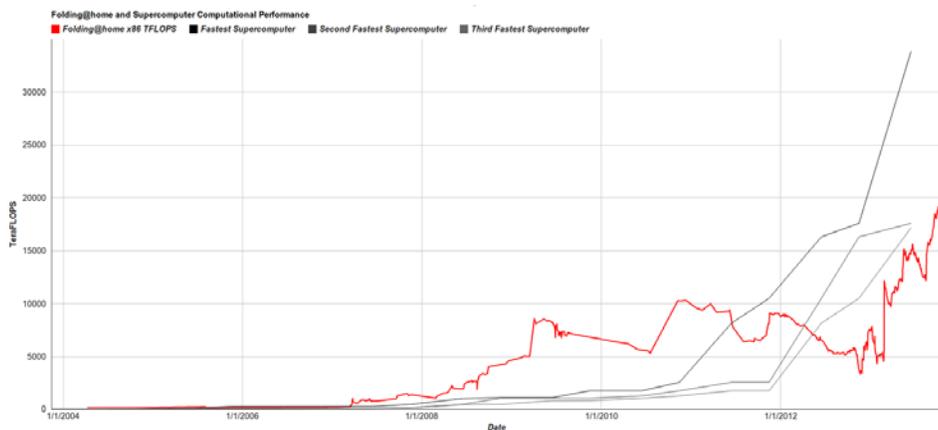
Folding@home 자원 봉사자는 각각의 개인용 컴퓨터에 클라이언트 프로그램을 설치한다. 클라이언트는 백그라운드의 다른 소프트웨어 구성 요소를 관리하며, 사용자는 클라이언트와 상호작용한다. 클라이언트를 통해 사용자는 폴딩 과정을 중지할 수 있고, 이벤트 로그를 열람할 수 있으며, 작업 진행 상황을 체크할 수 있고, 개인별 통계를 볼 수 있다. 이 클라이언트는 개인용 컴퓨터의 유휴 처리 리소스를 활용할 수 있도록 매우 낮은 우선 순위로 백그라운드에서 계속 실행되어, 정상적인 컴퓨터 사용이 영향을 받지 않는다. 최대 CPU 사용량이 클라이언트를 통해 조정될 수 있고, Folding@home 서버에서 작업 단위를 받아오고 적절한 코어를 다운로드할 수 있다. 이 클라이언트는 단일 프로세서, 멀티 코어 프로세서 혹은 GPU에 맞춰 조정될 수 있다. 클라이언트 GUI는 오픈 소스이지만 클라이언트 자체는 보안 및 과학적 무결성을 이유로 폐쇄 소스 라이선스 하에 있다. 클라이언트와 Folding@home 데이터 서버 간 통신은 2048비트 디지털 서명을 통해 보호된다.

□ 활용 효과

2014년 7월 현재 총 179237명이 참여하여 39022 테라 플롭(FLOP, 1초에 38페타, 즉 38000조회의 부동 소수점 연산)의 컴퓨팅 파워를 확보하였고, 이미 2007년 6월부터 2011

년 6월 사이에 Folding@home(빨강)은 가장 빠른 500개의 슈퍼컴퓨터(검정)의 성능을 넘어선 바 있다. 동시간대 seti@home은 681 테라 플롭의 컴퓨팅 파워를 보여주고 있다.⁶⁸⁾

[그림 40] Folding@home 성능



이러한 Folding@home을 활용하여 최근 HKUST에서 후이 후양의 그룹은 hIAPP(human islet amyloid polypeptide, amylin⁶⁹⁾)의 잘못된 폴딩을 조사하기 위해 대규모 분자 역학 시뮬레이션을 수행했다. 유형 II 당뇨병 환자의 95%가 hIAPP의 잘못된 폴딩을 보이고 있다. 후이 후양의 그룹은 hIAPP의 준 구조적 상태를 확인하고 그들 사이의 전환 역학을 파악함으로써 유형 II 당뇨병 치료에 한 걸음 더 다가갈 수 있었으며, 이와 같은 결과는 2013년 JACS(미국 화학회 저널)에 게재되었다.⁷⁰⁾

68) 보닉 통계 자료, <http://boincstats.com/en/stats/0/project/detail/overview>

69) 췌장의 β 세포에서 합성·분비하는 천연성 펩티드호르몬. 인슐린과 같이 당대사를 조절한다. 수용체의 해석이나 생리적 뜻의 해명이 현재 진행되고 있다
아밀린저해제는 비만인슐린 저항성을 수반하는 고혈압증이나 인슐린 비의존성 당뇨병 등의 치료제가 될 가능성이 있다.

70) Qiao, et al. 2013.

<표 18> Folding@home 전후 효과 요약

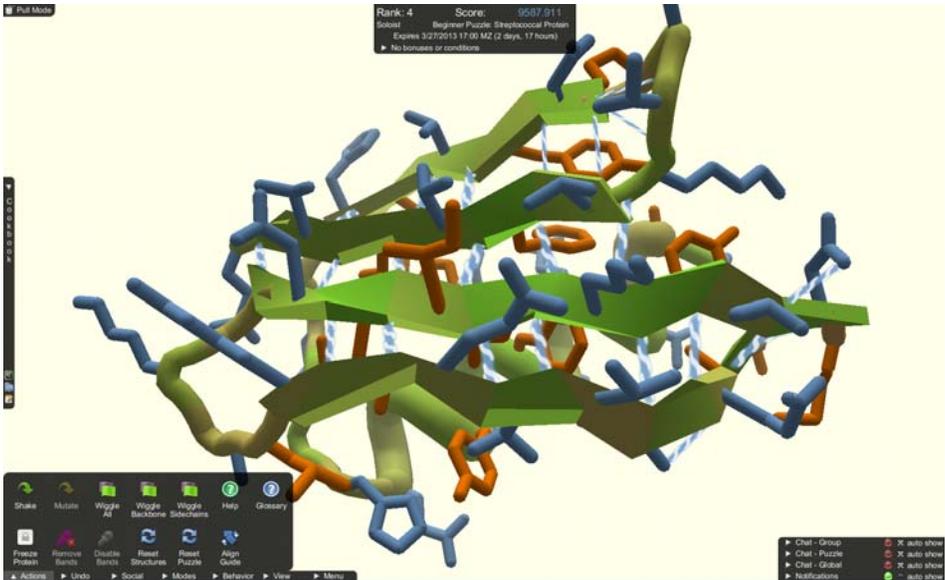
시민 과학 적용 전	시민 과학 적용 후
슈퍼컴퓨터를 할당 받아야만 동태적 분자 구조를 시뮬레이션 할 수 있었으며 그 외 마땅한 현실적 대안이 없음	SETI@home처럼 자원 봉사자 컴퓨터의 유휴 자원과 노력을 활용함으로써 질병 연구에 한 걸음 더 다가갈 수 있었음

2. Foldit

□ 개요

Foldit은 단백질 폴딩을 위한 온라인 퍼즐 비디오 게임이다. 이 게임은 실험적인 연구 프로젝트의 일부로서, 워싱턴 대학 생화학 학과와 동 대학 게임 과학 센터가 공동으로 개발하고 있다. 이 게임의 목표는 게임 내에서 제공되는 다양한 도구를 활용해서, 상기 선택된 단백질의 구조를 가능한 한 많이 파악하여 폴딩하는 것이다. 최고의 점수를 받는 솔루션은 연구자들이 분석한 결과 현실 세계에서 관련 단백질에 적용할 수 있는 기본 구조 구성이 실제로 존재하는 경우이다. 과학자들은 이러한 솔루션을 다양한 대상 질병을 근절하고, 생물학적 혁신을 만들어내는 등 현실 세계의 문제들을 해결하는 데 사용할 수 있다.

[그림 41] Foldit의 단백질 폴딩 화면



자료: <http://fold.it/portal/info/about>

□ 구현 방법

Foldit은 단백질 폴딩 구조를 빨리 찾아내기 위해서 크라우드소싱과 분산 컴퓨팅 방식을 활용한다. 특히 크라우드 소싱과 커뮤니티 협업이 강조된다. 또한, 가상 상호작용(Virtual Interaction)과 게임화(Gamification)가 추가되어 독특하고 혁신적인 프로젝트 환경이 갖춰지게 되었다.

○ 가상 상호작용

Foldit은 인간의 뇌가 갖고 있는 자연적인 3차원 입체 패턴 매칭과 공간 추론 능력을 활용하여 단백질 구조 예측 문제를 해결하려는 시도를 하고 있다. 현재의 퍼즐은 이미 구조에 대한 이해를 마친 단백질들을 기반으로 하고 있으며, 사람들이 직관적으로 이러한 퍼즐을 푸는 방식을 분석하여, 기존 단백질 폴딩 소프트웨어가 사용하는 알고리즘을 개선하고자 한다.

Foldit은 사용자가 활용할 있는 간단한 단백질과 같은 구조에 대한 튜토리얼 시

리지를 제공하고 실제 단백질을 기반으로 한 퍼즐 집합을 정기적으로 갱신한다. 그래픽 인터페이스를 기반으로 다양한 도구들과 단백질 구조가 표시된다.

○ 게임화

Foldit은 과학적 도구를 구축하는 데 있어 게임화 방식을 사용하여 일반 대중들에게 더 쉽게 다가갈 수 있도록 하였다. 그에 따라 과학적 교육 기반이 없는 사람들도 사용할 수 있었고, 따라서 더 많은 대중을 참여시킬 수 있었다. 각 퍼즐의 구조가 바뀔 때마다 단백질이 얼마나 잘 풀렸는지 분석하여 점수가 계산되고 가장 높은 점수의 리스트가 함께 제공된다. 사용자들은 그룹을 만들거나 참여할 수 있으며 서로 솔루션을 공유할 수 있다. 그룹의 최고 점수 리스트도 따로 관리된다.

□ 활용 효과

2011년 Foldit 사용자가 에이즈를 일으키는 원숭이 바이러스의 일종인 메이슨-화이자 원숭이 바이러스(M-PMV)의 레트로 바이러스 단백질 분해 효소의 결정 구조를 해독하는 데 도움을 주었다. 퍼즐을 푸는 데 주어진 시간은 3주였지만 사용자들은 단 10일 만에 정확한 입체 3D 모형을 완성하였다. 해당 효소의 구조를 구성하는 방식에 대한 연구 문제는 지난 15년 동안 과학자들이 풀지 못했던 난제였다.

2012년 1월에 미국 과학잡지 Scientific American은 Foldit 사용자가 첫 번째로 크라우드 소싱을 통해 단백질 구조를 재설계했다고 보도했다. 해당 단백질은 널리 합성 화학에서 사용되는 딜스-알더 반응을 촉진하는 효소이다. 시애틀에 있는 워싱턴 대학의 게임 과학 센터에서 데이비드 베이커 등의 팀이 처음부터 이 효소를 설계하였으나 개선이 필요했고, 사용자가 13개의 아미노산을 추가하여 효소를 재설계하였고 그에 따라 활동량이 18배나 늘어났다.

<표 19> Foldit 전후 효과 요약

시민 과학 적용 전	시민 과학 적용 후
슈퍼컴퓨터를 할당 받아야만 동태적 분자 구조를 시뮬레이션 할 수 있었으며 그 외 마땅한 현실적 대안이 없음	자원 봉사자 컴퓨터의 유휴 자원과 노력을 활용함으로써 질병 연구에 한 걸음 더 다가갈 수 있었음

3. EteRNA

□ 개요

EteRNA는 National Science Foundation이 지원하여 카네기멜론 대학교와 스탠포드 대학교에서 개발한 퍼즐 게임으로, 사용자가 RNA 분자 구조 퍼즐을 풀게 하여 RNA의 분자 구조를 알아내려고 한다.

RNA는 오탄당의 일종인 리보스를 기반으로 뉴클레오타이드를 이루는 핵산의 종류로 DNA의 일부가 전사되어 만들어 진다. RNA는 단백질을 합성하는 과정에 작용하며, 아데닌(A), 구아닌(G), 우라실(U), 시토신(C)의 핵염기가 연결되어 있는 형태이다. RNA의 종류에는 리보솜을 구성하는 rRNA, DNA의 유전정보를 옮기는 mRNA, 단백질의 기본 구성단위인 아미노산을 운반하는 tRNA 등이 있다.

RNA는 DNA의 이중나선 구조처럼 폴딩되며, 아데닌이 우라실에, 시노신이 구아닌과 만나 결합하는 경향이 있다. 따라서 RNA가 합성될 때 RNA의 모양은 일직선인 경우가 없으며, 3차원 모양을 가지게 된다.

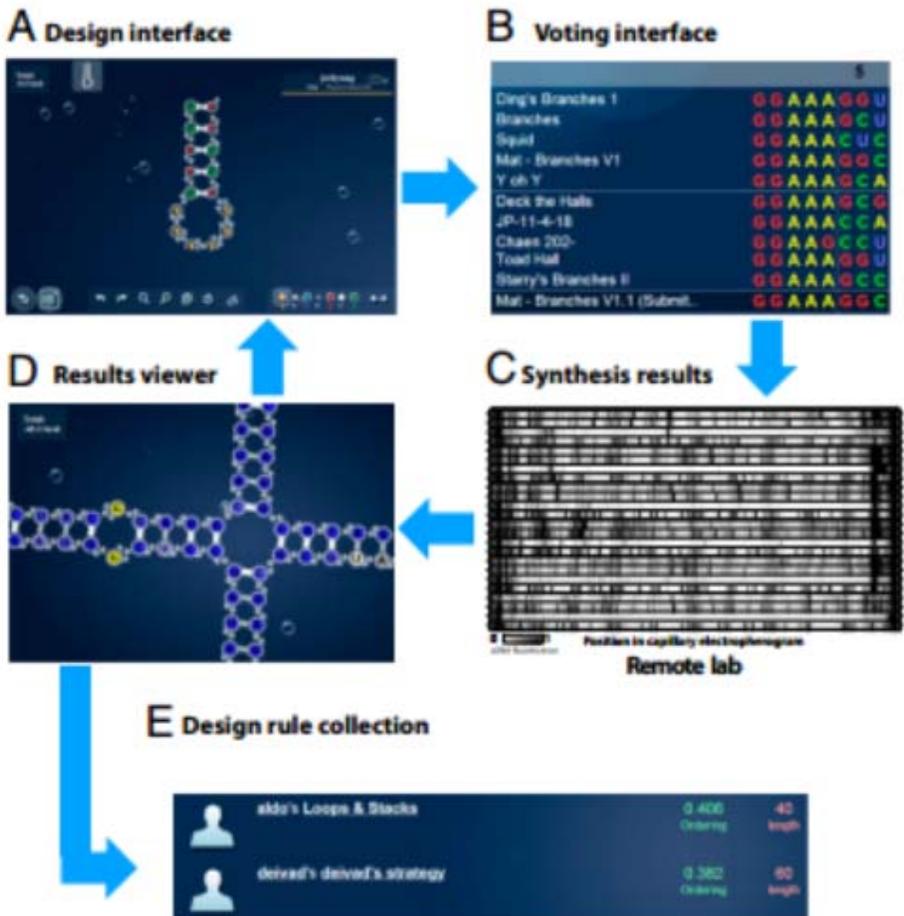
과학자들인 RNA의 모든 역할을 밝혀내지 못했으며, rRNA, mRAN, tRNA 이외에도 shRNA, piRNA, snRMA 다양한 RNA가 밝혀지고 있다.

EtRNA는 사용자에게 목표하는 모양으로 퍼즐을 풀 것을 제시한다. 사용자에게 제시하는 RNA모형들은 이미 다른 SW기술로 형태가 입증된 모양이다. 이는 사용자가 해당 모형의 구조에 대해 이해하게 하고 더 나은 구조를 찾는데 사용자가 도움을 주는 방식이다. 이 과정을 마치면 실제 RNA 설계 문제가 주어지게 되며, RNA가 제대로 폴딩되는 지 검증하여, 계산 모델과 실제 모델의 차이를 없애줄 수 있을 것이라 기대한다.

□ 구현 방법

EteRNA의 워크플로우는 [그림 42]와 같다. 사용자는 시퀀스 디자인 인터페이스를 통해 목표된 RNA 구조를 만들면 사용자간의 투표를 통해 8개의 시퀀스가 실제로 합성되고 가능성이 입증된다. 실험의 결과는 온라인에 게시되며 사용자들은 이를 바탕으로 새로운 가설을 만들고 다시 디자인 인터페이스를 통해 폴딩될 수 있는 RNA 구조를 만드는 것을 반복한다.

[그림 42] EteRNA 워크플로우



자료: Lee et al. 2014.

□ 활용 효과

2011년 8월에는 26,000명의 사용자가 이 게임에 참여하여 306개의 RNA 디자인을 만들어 냈고 실제로(in vitro) RNA를 만들었다. 2013년에는 총 37,000명이 RNA디자인을 만들어 냈고, 이렇게 만들어진 RNA는 연구실에서 새로운 RNA모형을 제작하는데 큰 도움을 주었다. 또한 머신 러닝을 통해 새로운 RNA 디자인 알고리즘을 만들어 낼 수 있었고 이는 이전의 알고리즘 보다 더욱 좋은 성능을 내었다. 2014년 1월에는 EteRNA의 결과가 미국국립과학원회보(PNAS: Proceedings of the National Academy of Sciences of the United States of America) 저널에 게재되었으며, 저자에 EteRNA 사용자가 포함되었다

<표 20> EteRNA 전후 효과 요약

시민 과학 적용 전	시민 과학 적용 후
슈퍼컴퓨터를 할당 받아야만 동태적 분자 구조를 시뮬레이션 할 수 있었으며 그 외 마땅한 현실적 대안이 없음	자원 봉사자 컴퓨터의 유휴 자원과 노력을 활용함으로써 질병 연구에 한 걸음 더 다가갈 수 있었음

● 저 자 소 개 ●

김 석 원

·KAIST 전산학과 박사
·현 소프트웨어정책연구소 책임연구원

임 춘 성

·College of Medicine 박사
·현 소프트웨어정책연구소 책임연구원

이 현 승

·서울대 컴퓨터공학과 석사
·아주대 로스쿨
·현 소프트웨어정책연구소 선임연구원

궁 한 경

·뉴저지 공대 전산공학과 석사
·현 소프트웨어정책연구소 초빙연구원

방송통신정책연구 14-진흥-029

과학기술의 SW활용 현황분석 및 협력방안 연구

2014년 11월 26일 인쇄

2014년 12월 1일 발행

발행인 미래창조과학부 장관

발행처 미래창조과학부

경기도 과천시 관문로 47 정부과천청사

TEL: 02-2110-2824

Homepage: www.msip.go.kr
