

## 다가오는 웹 3.0시대의 바이오인포메틱스

● 허 보 경 (한국생명공학연구원 국가생물자원정보관리센터)

### 1. 서론

우리는 웹에 버전의 번호를 붙여 시간의 흐름에 따른 웹 발전을 단계적으로 평가해 왔다. 웹 1.0시대에는 ‘사용자에게 정보를 보내는 것’에 중점을 두었으나, 웹 2.0시대에는 ‘웹을 플랫폼으로서 이용하여 협업과 커뮤니티의 공간으로 만드는 것’에 중점을 두어 사용자의 참여, 풍부한 사용자 체험, 데이터의 중요성, 웹 서비스의 활용에 의해 유연하게 결합되는 웹을 지향한다.

이 콘셉트들을 훌륭히 실현시키고 있는 기업이나 사이트의 성공 예를 살펴보자.

구글 : 웹 2.0의 콘셉트를 활용해 이 정도로 성공을 거둔 기업은 구글 밖에 없다. 구글의 비즈니스 모델

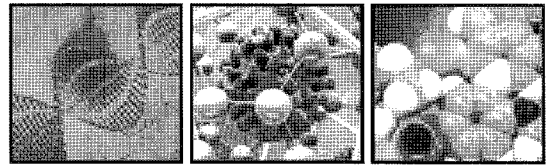
은 ‘누구나 액세스 할 수 있는 방대한 양의 웹 사이트와 관련된 데이터’에 근거하고 있다. 사용자는 구글을 통해 정보를 간단히 찾아낼 수 있을 뿐만 아니라 임의대로 다양하게 사용할 수 있다. 게다가 구글에서 제공하는 수많은 서비스는 웹 서비스 경유로 이용될 수 있다.

위키피디아(Wikipedia) : 위키피디아는 온라인상의 협업을 대표하는 예. 사용자는 위키피디아상에서 데이터를 간단하게 추가하고 편집할 수 있다. 위키피디아의 콘텐츠를 이용하는 방대한 사용자들은 스스로 적절한 기사를 확실하게 게재하고 있다.

웹 3.0의 시대는 곧 도래할 것인가? 웹 2.0의 콘셉트가 우리에게 아주 친숙해진 후 자연스럽게 다음엔 무엇이 나타날까에 대한 의문이 생긴다.

표 1 e-Science의 특성

기존의 연구환경	e-Science 연구환경
<p>기존의 연구환경에서 연구자들은 초단장비를 사용하기 위해 장비가 있는 지역을 직접 찾아가거나, 필요한 거대 데이터를 직접 저장장치에 담아오거나, 적합한 협업 인력과 직접 만나 연구개발을 함</p>	<p>e-Science 연구환경이 구축되면 연구자들은 시공의 제약없이 언제, 어디서나 원하는 연구개발을 수행할 수 있게 됨. 전 세계에 산재해 있는 슈퍼 컴퓨터, 입자가속기, 초고속시퀀서 등 고가의 연구장비를 공동으로 사용하고, 계층 데이터나 고화질 영상같은 대용량 정보를 실시간으로 주고 받으며, 전세계의 연구자들과 마치 한 공간에 있는 것처럼 연구할 수 있는 가상 공간을 제공함</p>



그럼 다음 버전은 어떠한 변화를 초래할까. 아마 주요 콘셉트는 ‘테크놀로지를 더 이용하기 쉽게 만들고 상시 진행 및 상시 참여의 개념을 추구하게 될 것이다.

그렇다면 웹 3.0 시대의 생명정보학의 모습은 어떤 모습일까?

## 2. 본 론

웹 3.0시대의 생명정보학의 키워드는 ‘대량화, 자동화, 연계화, 고속화’로 대변될 것으로 사료된다. 이러한 개념의 대표적인 프레임워크로 e-Science[1~15]를 들 수 있는데, 이것은 첨단 IT기술과 고성능연구망을 이용하여 사이버 상에서 자유롭게 고성능 컴퓨터 및 첨단연구장비, 대용량 데이터, 연구인력 등을 동시에 활용함으로써 연구생산성을 획기적으로 향상시켜주는 첨단 사이버 연구환경이다(표 1).

다가올 개인유전체 시대의 거대문제 해결을 위해서는 대용량 생명정보를 초고속으로 분석하고, 원격지에 있는 데이터를 공유할 수 있는 시스템을 요구한다.

이러한 e-Science는 참여, 개방, 공유를 통해 연구개발 시간과 비용을 수십, 또는 수백배 이상 줄여주고, 연구성과의 수준은 급격히 높여준다. 연간 수천 억원 이상의 예산절감 효과는 물론 신약개발의 경우, 개발기간은 절반 이상, 비용은 최대 100분에 1까지도 절감할 수 있게 해준다.

국의 선진국들은 막대한 자금을 투자하여 관련 인프라 및 소프트웨어를 개발하고 있다(표 2).

우리나라의 경우, 국가 과학경쟁력 7위, 기술경쟁력 6위(2007년 스위스 국제경영개발원 발표)라는 뛰어난

어난 국가 과학기술경쟁력을 갖추고 있으며, 세계 최초의 환형 국제과학기술협업연구망인 글로리아드(GLORIAD)의 핵심멤버로 참여할 만큼 기발기술 수준도 세계적인 수준이다.

세계최고의 수준의 IT 인력과 기술력을 보유하고 있어 e-Science 연구환경 조기구축이 가능하며, 국가 과학기술 인프라 연동 기술, 시스템 통합 구현기술, 각종 소프트웨어 개발 역량 역시 뛰어나다. 한국과학기술정보연구원(KISTI)은 국가차원의 고성능연구망, 2008년까지 280TFlops 급 슈퍼컴퓨팅 4호기 도입, 미국 과학재단(NSF) 수준의 응용 소프트웨어 보유, 첨단 DB 등 e-Science 연구환경을 구축할 수 있는 뛰어난 연구역량을 가지고 있다.

지난 2004년 대통령과학기술자문회의의 수요조사에 따르면 설문 대상 연구자들이 96%가 협업연구를 희망하고 있고, 91%가 연구자원의 공동활용이 필요하다고 생각하며, 97%가 e-Science 연구환경 구축을 요구하고 있는 것으로 나타났다.

국내의 경우, e-Science 기반의 시범사업을 통해 몇 가지 성공한 사례는 다음과 같다.

### 2.1 분자시물레이션 e-Science 연구환경 시범구축

대용량 데이터가 요구되는 바이오, 나노분야 연구를 사이버 상에서 진행할 수 있도록 e-Science 기반의 분자시물레이션 작업환경을 구축하고, 시물레이션 데이터 베이스를 구축해 연구자들이 보다 빠른 시간 안에 보다 높은 효율로 연구할 수 있는 환경을 구축하였다. 2007년 3월 태국에서 개최된 PRAGMA(Pacific Rim of Applications and Grid Middleware Assembly) 12차 회의를 통해 미국이 주도하는 조류독감 연구에 공식적으로 참여함으로써 선진국들과 동등한 위치에서 e-Science

표 2 선진국의 e-Science 연구현황

국가	내용
유럽 연합	○e-Science 기반의 EMBL-EBI 중심의 유럽생명정보 네트워크 구축 - EGEE 중심의 유럽 Bio Grid 프로젝트 수행
미국	○TeraGrid의 Life-Science Gateway 중심으로 Bio 전문가들에게 Bio 정보 및 대규모 슈퍼컴 자원 제공 - 일반적으로 모든 사용자에게 제공되지 않았던 데이터, 어플리케이션, 워크플로우를 통하여 통합 제공함
일본	○국가 생물자원센터 및 생물다양성센터간의 국가 생명자원 통합 관리센터 운영 - 오사카 슈퍼컴센터, JSRR 등이 참여하는 BioGrid 프로젝트 수행

# 기획특집

기반 협업을 할 수 있게 되었다.

## 2.2 HG2C를 위한 서비스지향구조 기반의 e-Science 환경 시범구축

유전자, 단백질, 화합물 정보와 관련된 생물정보학, 화학정보학 프로그램을 이용하여 인간유전자와 관련이 있는 활성화합물을 탐색하는 가상실험 시스템(HG2C)을 구축하고, 이를 e-Science 연구환경에서 활용할 수 있도록 함으로써 연구자들이 보다 빠르고 정확한 연구성과를 도출할 수 있도록 하였다. 국내 최초로 유전자, 단백질, 화합물 등의 상이한 정보를 저장하고 그 연관관계를 도출할 수 있는 DB 표준안을 제안하였다. 또한 신약반응을 모사하는데 쓰이는 단백질-화합물 반응 방식에 대한 3,000쌍 이상의 결과를 도출하였다.

e-Science의 핵심요소는 다양한 분야 연구자들의 요구를 기반으로 재사용이 가능하고, 확장성이 있는 SOA(Service Oriented Architecture)의 기반기술, Scientific workbench, 관련 미들웨어 기술 등이 있다.

이러한 핵심요소 중에서 현재 바이오 e-Science에서 가장 부각되고 있는 이슈는 Scientific workbench 소프트웨어를 이용하여 워크프로우(workflow)형태로 원하는 분석 과정(파이프라인)을 구성하여 수행할 수 있는 가상의 협업공간을 만드는 것에 초점을 두고 있다.

예를 들어, EBI(European Bioinformatics Institute)의

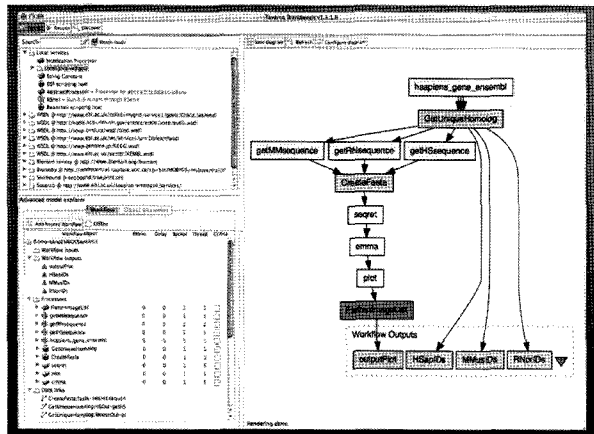


그림 1 Taverna workbench 시스템

경우, Taverna(그림. 1)라는 Scientific workbench 소프트웨어를 개발하여 사용자가 원하는 분석파이프라인을 구성한 후, myexperiment 라는 가상협업 웹사이트(www.myexperiment.org)를 운영하여 생물정보와 관련된 노하우 공유 및 협업이 가능하도록 하였다.

이러한 협업소프트웨어들은 사용자로 하여금 정보 또는 노하우의 소비자(consumer)인 동시에 공급자(provider)인 prosumer의 역할을 수행할 수 있는 수단을 제공하게 할 뿐 아니라, 데이터그리드를 통해 연구에 관련된 대용량 데이터를 저장/공유하고, 컴퓨팅 그리드를 통한 고성능 전산자원의 공동활용으로 분석파이프라인을 실행한다.

또한, 가상 연구협업 커뮤니티를 통한 새로운 연구에 대한 아이디어 창출 및 복잡한 문제의 공동해결 수단을 제공한다(그림 2). 각 실험실의 실험 프로토콜, 연구노트 등을 블로그(blog)할 수 있는 도구를 제공한다(그림 3).

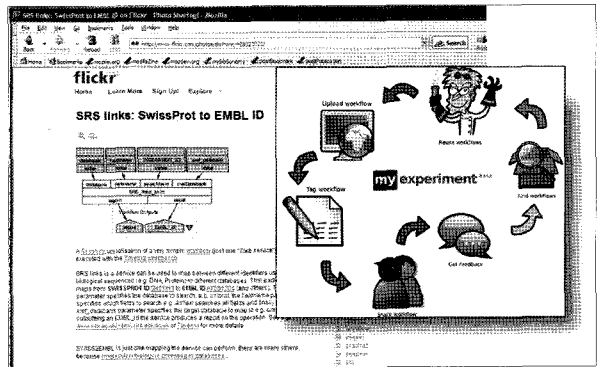


그림 2 myexperiment를 통한 가상 협업공간 제공

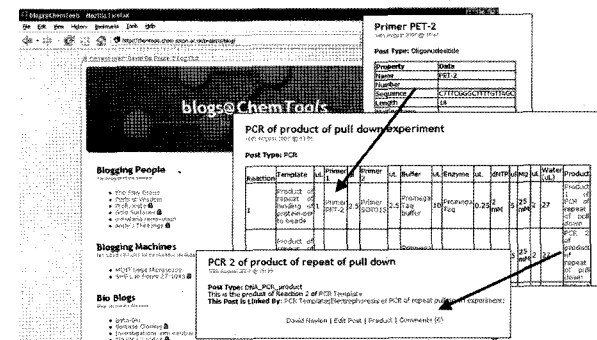


그림 3 실험실의 관련 정보의 블로그 기능 제공

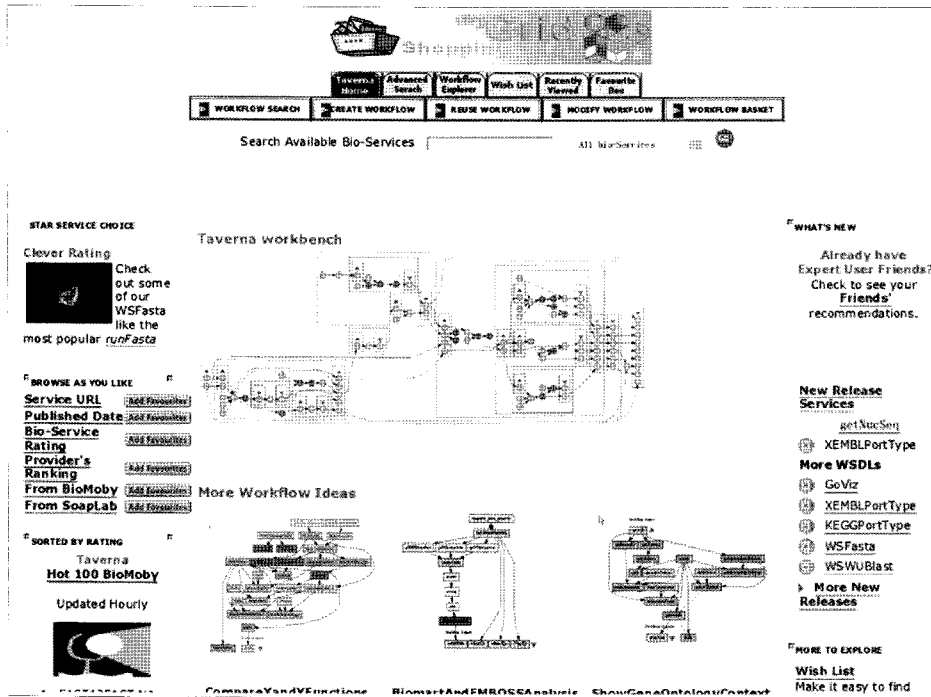


그림 4 One-stop 쇼핑개념의 생물정보학 도구 제공

또한, 사용자가 원하는 분석파이프라인을 one-stop 쇼핑형식으로 가져가서 수정한 후, 다시 myexperiment 사이트에 올려 공유하므로 사용자가 prosumer로서의 역할을 수행할 수 있도록 해준다(그림 4). 즉, 이러한 협업 시스템의 활용으로 생명정보학의 모든 관련 분석도들이 하나의 서비스 개념으로 제공되므로 다가올 웹 3.0시대의 가장 중요한 핵심요소가 될 것으로 사료된다.

그리고 사용자가 연구결과를 논문에 투고하는 경우, myexperiment와 같은 사이트에 관련 분석 워크플로우를 공유하여, 다른 사람이 쉽게 연구의 노하우 및 결과를 공유하고 검증하는 것이 가능하므로 Black-box science를 지양할 수 있어 연구의 신뢰성을 확보할 수 있다. 즉, 연구협업을 통한 광범위한 연구커뮤니티를 형성하여 마치 레고블록을 올리듯이 거대한 생명정보 서비스 커뮤니티를 구축할 수 있어 생명정보 관련 연구 분야에 일대 혁신을 불러 올 것으로 기대된다(그림 5).

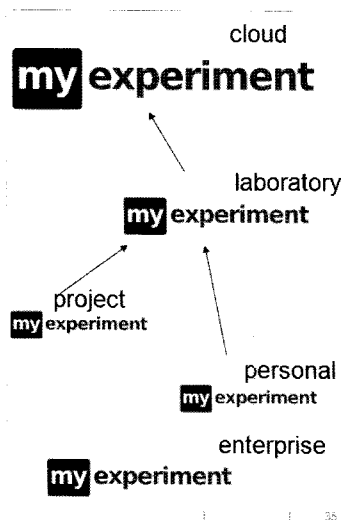


그림 5 생명정보 서비스 커뮤니티

우리나라도 하루 빨리 e-Science 공동협업 환경을 구축하여, 개인적인 생명정보 분석 노하우를 과감하게 공개·활용하여 me-Science에서 we-Science로의 패러다임 전환을 신속하게 달성해야 한다. 즉, 다가올 웹 3.0시대의 시대정신에 맞게 개인 또는 산·학·연 연구센터들이 생산 또는 보유하고 있는 장비, 데이터 및 분석파이프라인의 공유·활용을 통해 더욱 효율적이고 생산적인 연구체계를 구축할 수 있다는 인식의 전환이 필요하다고 본다.

웹 3.0시대의 생명정보학은 인터넷의 공개된 방대한 정보 및 노하우를 혼합(mash-up)하여 원하는 결과

# 기획특집

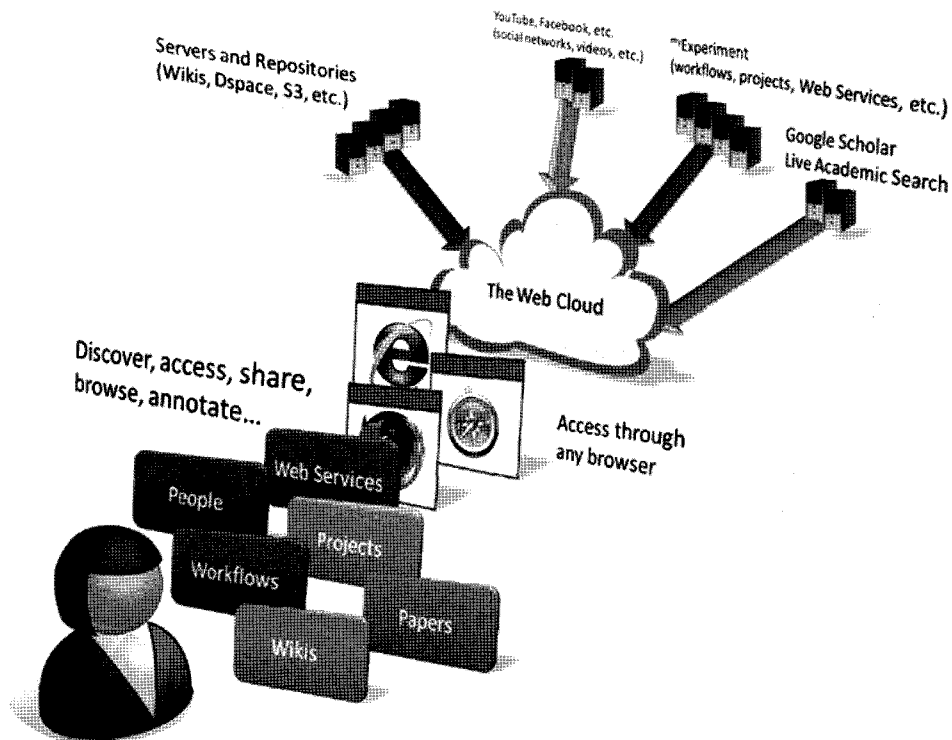


그림 6 mash-up 기술

를 초단 시간에 해결할 수 있는 방향으로 나갈 것으로 예측된다. 여기서, mash-up이란 평범한 술들을 비율에 맞추어 잘 섞으면 새로운 맛의 카테일이 만들어지는 것처럼, 평범한 웹서비스나 기술 몇 개를 섞어서 새로운 웹서비스를 만드는 기술이다(그림 6).

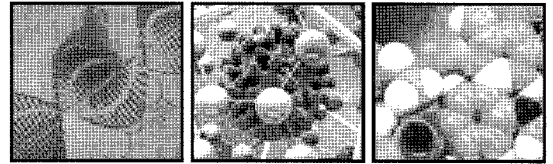
### 3. 결 론

웹 3.0시대의 생명정보학의 기본 정신은 「Cooperate, Don't Control」이다. 해외 선진국과 같이 우리나라도 정보 공유, 노하우 공유를 통해 국가차원의 생명정보 분석 역량을 한 단계 업그레이드 시켜야 한다. 다행히도 현재 진행 중인 「생물자원 확보·관리 및 활용에 관한 마스터플랜(안)」에도 웹 3.0시대를 대비한 Bio-E-Science 망의 구축 및 국가 e-Science망과의 연계에 대한 구체적인 내용을 담고 있어 매우 고무적인 일이라고 사료된다. 공유·연계·협력·기여라는 웹 3.0시대의 정신을 이어받아 우리나라가 미래 바이오 지식산업 강국에 필요한

초석을 마련할 수 있기를 기대해 본다.

### 참고문헌

- [1] J. Almond and D. Snelling. UNICORE: Secure and Uniform Access to Distributed Resources via the World Wide Web. White Paper, October 1998,
- [2] I. Altintas et al. A Framework for the Design and Reuse of Grid Workflows, International Workshop on Scientific Applications on Grid Computing (SAG'04), LNCS 3458, Springer, 2005
- [3] F. Berman et al. The GrADS Project: Software Support for High-Level Grid Application Development. International Journal of High Performance Computing Applications(JHPCA), 15(4):327-344, SAGE Publications Inc., London, UK, Winter 2001.
- [4] J. Cao et al. GridFlow:Workflow Management for Grid Computing. In 3rd International Symposium on



- Cluster Computing and the Grid (CCGrid), Tokyo, Japan, IEEE CS Press, Los Alamitos, CA, USA, May 12-15, 2003.
- [5] E. Deelman, J. Blythe, Y. Gil, and C. Kesselman. Workflow Management in GriPhyN. The Grid Resource Management, Kluwer, Netherlands, 2003.
- [6] E. Deelman et al. Mapping Abstract Complex Workflows onto Grid Environments. Journal of Grid Computing, 1:25-39, Kluwer Academic Publishers, Netherlands, 2003.
- [7] T. Fahringer et al. Truong. ASKALON: a tool set for cluster and Grid computing. Concurrency and Computation: Practice and Experience, 17:143-169, Wiley InterScience, 2005.
- [8] D. Fernández-Baca. Allocating Modules to Processors in a Distributed System. IEEE Transactions on Software Engineering, 15(11): 1427-1436, November 1989.
- [9] I. Foster and C. Kesselman (editors), The Grid: Blueprint for a Future Computing Infrastructure, Morgan Kaufmann Publishers, USA, 1999.
- [10] A. Geppert, M. Kradolfer, and D. Tombros. Market-based Workflow Management. International Journal of Cooperative Information Systems, World Scientific Publishing Co., NJ, USA, 1998.
- [11] 변옥환, 국가 e-Science 사업 추진성과 및 주요성과, 2007.
- [12] 한국기초과학지원연구원, e-Science 구축방안 연구, 2003.
- [13] DTI(Department of Trade & Industry). e-Science: Building A Global Grid, 2002.
- [14] NSF. Cyberinfrastructure Vision for 21st Century Discovery, 2007. NSF. "Revolutionizing Science and Engineering Through Cyberinfrastructure", A Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, 2003. 1.
- [15] The National e-Science Centre and the UK e-Science Programme, [http://umbriel.dcs.gla.ac.uk/NeSC/general/talks/mc\\_commtechnesc2002.pdf](http://umbriel.dcs.gla.ac.uk/NeSC/general/talks/mc_commtechnesc2002.pdf)