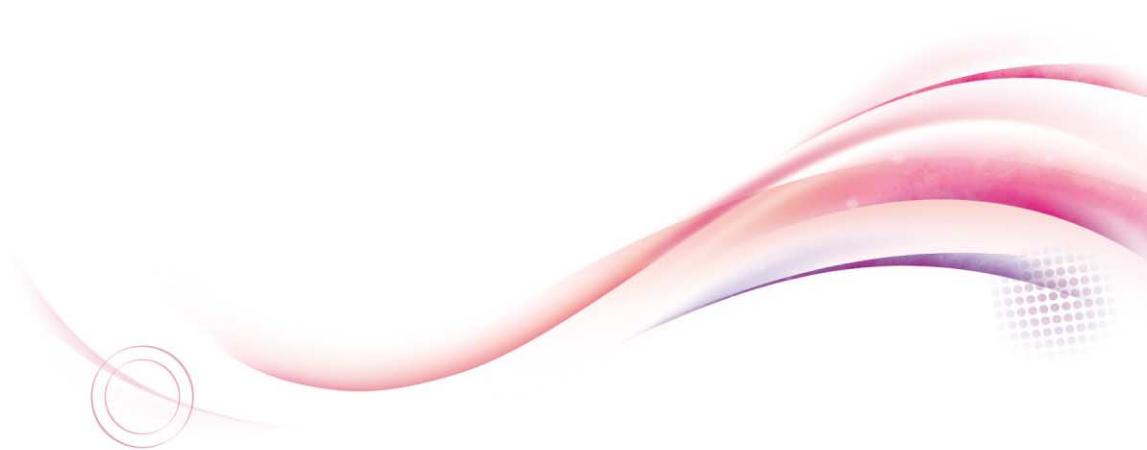


# Big Data 글로벌 10대 선진 사례

빅데이터로  
세상을  
리드하다



## Big Data 글로벌 10대 선진 사례

빅데이터로  
세상을  
리드하다



# Big Data 글로벌 10대 선진 사례

빅데이터로 세상을 리드하다

## Contents



### 빅데이터로 꿈꾸는 행복한 사회

- 1** 미국 국세청, 탈세 방지 시스템 통한  
국가 재정 강화 • 6
- 2** 일본, 센서데이터를 활용한 지능형  
교통안내 시스템 • 9
- 3** 한국석유공사, 국내 유가 예보 서비  
스를 통한 비즈니스 최적화 • 12



## 빅데이터를 통한 건강한 사회

- 4 미국 국립보건원, 유전자 데이터 공유를  
통한 질병치료체계 마련 • 18
- 5 미국 국립보건원, Pillbox 프로젝트를  
통한 의료개혁 • 22
- 6 건강보험회사 웰포인트(WellPoint),  
슈퍼컴퓨터를 활용한 효율적인  
환자치료 • 26
- 7 구글, 검색어 분석을 통한 독감  
예보 서비스 제공 • 30



## 빅데이터로 희망하는 안전한 사회

- 8 싱가포르, 국가위험관리시스템(RAHS)을  
통한 국가안전관리 • 34
- 9 FBI, 유전자 색인 시스템 활용한  
단시간 범인 검거 체계 마련 • 37
- 10 샌프란시스코, 범죄 예방 시스템으로  
안전 지역사회 구축 • 40

# 빅데이터로 꿈꾸는 행복한 사회

1. 미국 국세청, 탈세 방지 시스템 통한 국가 재정 강화
2. 일본, 센서데이터를 활용한 지능형 교통안내 시스템
3. 한국석유공사, 국내 유가 예보 서비스를 통한 비즈니스 최적화

## 맞춤형 생활공감 서비스 제공



취업 • 미래유망 직업 분석 및 진로개발 가이드

금융 • 직장인 생애 첫 재무설계

금융 • 상황별 교통정보 흐름 및 예측 서비스, 위치기반 라이프로깅 플랫폼



육아 • 보육시설 운영 현황 모니터링 시스템

물가 • 장바구니 물가예측 시스템

자기계발 • 위김맘 자기계발 프로그램 서비스



일자리 및 재취업 • 실버 멘토링 네트워크 지원시스템

레저 • 실버래크레이션 및 레저정보서비스 제공

# 미국 국세청, 탈세 방지 시스템 통한 국가 재정 강화

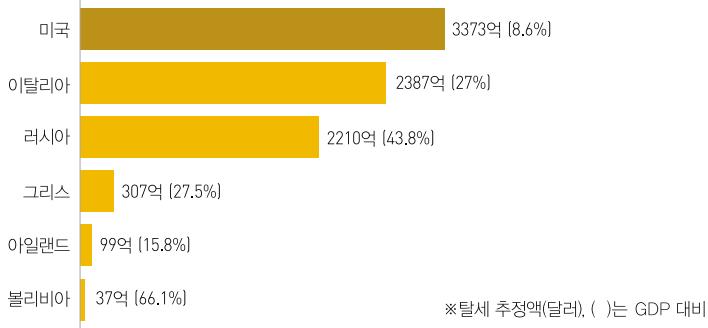
## 빅데이터 분석을 활용하여 탈세 및 사기 범죄 예방 시스템 구축

- 사기 방지 솔루션, 소셜 네트워크 분석, 데이터 통합 및 마이닝 등 활용
- 세금 누락 및 불필요한 세금 환급 절감의 효과 발생

## 추진목적 및 배경

- ▣ 탈세 및 사기로 인한 국가의 재정 위기 가능성 증가
  - 글로벌 금융위기로 인한 재정 위기의 확산은 개인과 기업의 탈세에 따른 낭비성 재정 지출 문제 발생
    - 세계은행은 탈세 및 세금 사기와 관련된 지구촌 지하경제의 규모가 전체 GDP의 18%에 이를 것으로 전망
      - 이탈리아는 GDP 대비 22.3%, 스페인은 19.3%, 포르투갈은 19.2%, 그리스는 25.1%에 이르는 것으로 추산
    - OECD는 지하경제에 흘러다니는 탈세를 조세망으로 끌어들여 지속 가능한 경제 발전을 이루어야 한다고 경고
  - 전 세계적으로 탈세 금액이 증가
    - 미국의 탈세 금액은 2010년을 기준으로 저소득층 의료보장 총액을 초과
    - 볼리비아의 경우, 자국 경제에서 지하경제가 차지하는 비중은 66.1%이며, GDP의 13.3%만이 세수로 집계

국가별 탈세 규모 및 액수

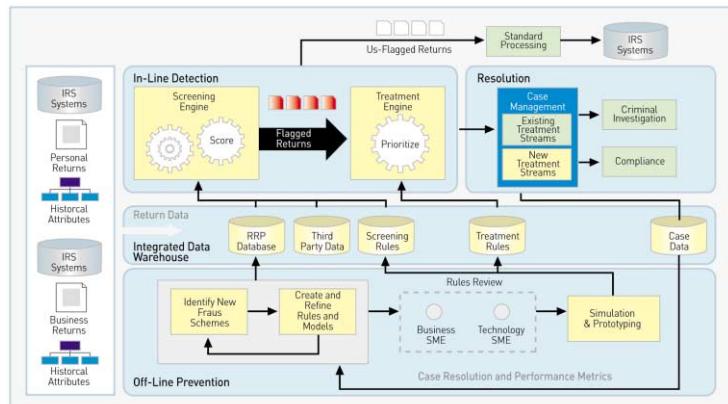


자료 : 포린풀리시

## 추진내용

- ▣ 대용량의 데이터와 다양한 기술을 결합한 탈세 및 사기 범죄 예방 시스템 구축
  - 정부기관 사기 방지 솔루션
    - 방대한 자료로부터 이상 징후를 찾아내고 예측 모델링을 통해 과거의 행동 정보를 분석하여 사기 패턴과 유사한 행동 검출
  - 소셜 네트워크 분석을 통한 범죄 네트워크 밭굴
    - 계좌, 주소, 전화번호, 납세자 간의 연관관계 분석 실시
    - 페이스북이나 트위터를 통해 범죄자와 관련된 소셜 네트워크를 분석하여 범죄자 집단에 대한 감시 시스템 마련
  - 다양한 데이터 분석을 통한 지능형 감시 시스템 구축
    - 데이터베이스와 데이터 웨어하우스를 통합해 대용량 데이터를 효율적으로 활용하고 프로세스를 통해 지능형 데이터 분석 능력을 지원할 수 있는 시스템 구축

## RRP(Return Review Program) 시스템 구조



자료 : SAS Korea

## 효과 및 전망

- 통합형 탈세 및 정부사기 방지 시스템을 통해 연간 3,450억 달러에 달하는 세금 누락 및 불필요한 세금 환급 절감
- 과학적 데이터를 근거로 탈세 조사를 수행함으로써 탈세자 수의 감축 등 우수성과 발생
- 과거 데이터 분석을 통해 향후 발생할 수 있는 사기 범죄 및 탈세 관련 사건을 미연에 방지

## 정책적 시사점

- 사기, 재정낭비, 부당 지출과 관련된 여러 첨단 기술을 접목시켜 다양한 정부 사업에 응용 가능
- 의료보험 및 복지프로그램 관련 사업에도 동일한 기술 응용 가능

# 2

## 일본, 센서데이터를 활용한 지능형 교통안내 시스템

### 실시간으로 GPS 데이터를 분석하여 최적의 교통 정보를 사용자에게 전달하는 서비스

- 교통상황과 관련된 데이터를 종합 분석하여 실시간으로 출발지에서 목적지까지의 최적 경로를 안내
- 택시 및 정보 제공에 동의한 내비게이터 사용자로부터 얻어진 교통 정보를 이용
- 사용자에게 최적의 교통상황 및 경로를 안내함으로써 에너지 효율을 높이고 교통 체증 감소 효과

### 추진목적 및 배경

#### ▣ 지능형 교통 정보 시스템의 발전

- 일본 ITS(Comprehensive Plan for ITS in Japan, '96)의 표준 체계 발표
  - 일본 건설성, 통산성, 운수성, 우정성, 경찰청의 5개성 정부차원에서 지능형 교통 정보 시스템 마련
  - ITS 사업의 아키텍처와 표준화 계획을 제시하여 일본 현실에 맞는 구축사업으로 발전
  - 2000년부터 2015년 이후까지 4단계로 구분하여 개발과 구축을 위한 장기적 전략을 제시

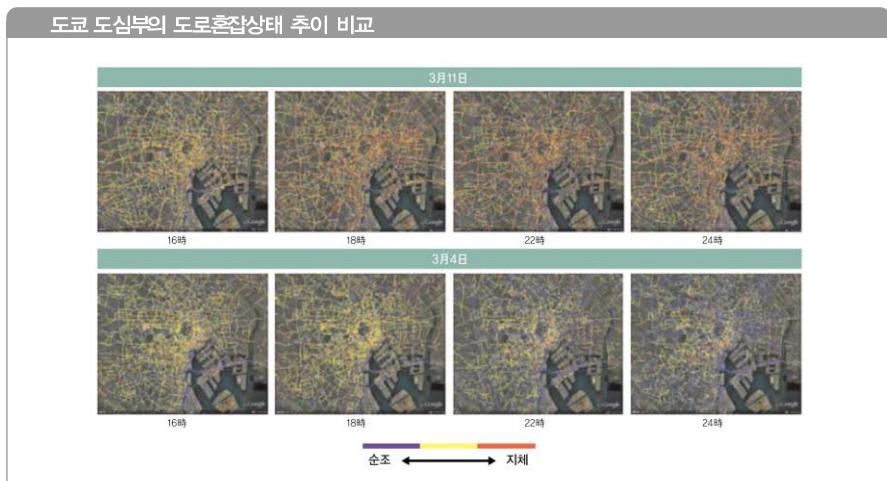
#### ■ 일본내 내비게이터를 이용하는 사용자의 폭발적인 증가

- 일본의 자동차 내비게이션 시장은 2008년 기준 연간 317만 대 규모로 2007년에 비해 0.6% 발전하였지만, 포터블 내비게이션인 PND(Personal Navigation Device)\*가 차지하는 비중은 급속히 증가 추세

\* PND : 액정화면이 4인치 정도로 작고 지도데이터를 본체 플래시 메모리나 외부 메모리에 보존하는 간이형 내비게이션

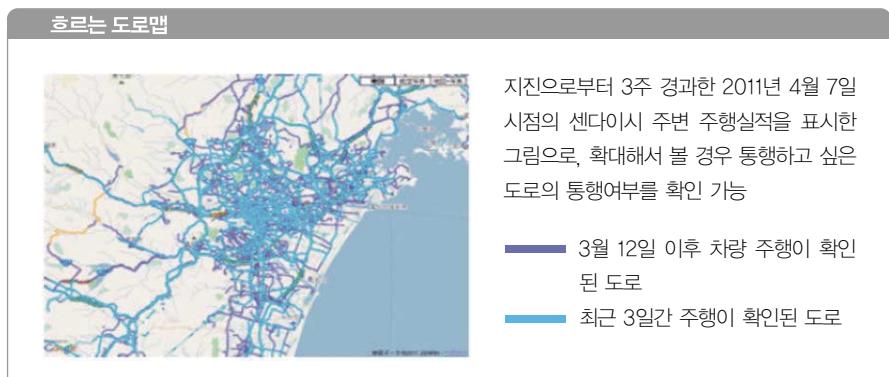
## 추진내용

- ❖ GPS로부터 자동차의 주행 스피드를 계산하여 교통 정보 수집
  - 일본 노무라연구소는 스마트폰형 내비게이션 서비스인 '전력안내! 내비'를 활용하여 2011년 일본 대지진시 도로교통 체증 피해 최소화
    - 교통 체증 감소 효과 및 구조차량을 위한 피해지의 실제 도로 교통 상황 안내
  - 일본 전역 지정도시의 택시 약 11,000여 대와 데이터 제공에 동의한 사용자로부터 실시간으로 교통정보 수집
  - 수집된 교통정보를 바탕으로 실시간으로 최적의 교통 안내 서비스 제공
- ❖ UTIS(Ubiqlink Traffic Information System)을 통한 독자적 도로교통정보망 구축
  - GPS 데이터에서 자동차의 주행스피드를 계산하여 도로 교통정보를 예측한 후 사용자의 스마트폰으로 송신
  - 도로 체증이 발생할 경우 최상의 빠른 길을 재검색하여 출발지에서 목적지까지의 최적 경로 안내



자료 : 노무라연구소, IT Solutions Frontier, Vol.29 No.4, 2012.2

- 노무라연구소는 UTIS를 활용하여 구조차량 및 지원자원 수송 차량에게 피해자의 실제 도로교통상황을 안내하는 ‘흐르는 도로맵’ 무상 제공
  - 시간경과에 따라 확대되는 통행가능 도로 정보를 인터넷상에 제공하는 시스템으로 정보는 하루에 6번 갱신



자료 : 노무라연구소, IT Solutions Frontier, Vol.29 No.4, 2012.2

## 효과 및 전망

- 다양한 사용자에 의해 취득된 정보를 바탕으로 한 실시간 교통 정보를 공유함으로써 최적의 교통 안내 서비스 가능
- 교통체증으로 인한 불필요한 에너지 낭비 방지를 통해 에너지 효율 증대

# 3

## 한국석유공사, 국내 유가 예보 서비스를 통한 비즈니스 최적화

### 급격한 유가변동에 대응하고 고유가에 따른 소비자 부담 감소를 위한 서비스

- 유가에 직간접적으로 영향을 미치는 여러 변수를 이용하여 유가정보 예측 서비스 제공
- 유가 변동에 대한 유기적인 대처와 대책마련 가능

### 추진목적 및 배경

#### ▣ 국제 정세에 따른 국제 유가의 변동

##### ■ 금융 위기에 따른 국제 유가의 급격한 변동

- 최근 유로 재정위기의 확대, 세계 경기의 재침체 우려로 인해 국제 유가의 변동 폭이 매우 크게 변화
- 글로벌 금융위기 속에서 세계경기 침체에 따른 원유 수요는 감소하였으나, 신흥국을 중심으로 원유 수요가 지속적으로 증가하고 있는 추세

#### ▣ 유가변동의 원인 및 변수의 다양성

##### ■ 선진국에 비해 높은 성장률을 보이는 비 OECD 국가를 중심으로 세계 원유 수요의 증가

- 2008년 세계경제 성장률은 2.8%에 불구하였으나 신흥국은 7%대의 높은 성장률을 달성하고 있으며, 세계 원유 수요에서 점차 비중 확대
- 세계 원유 소비 중 비 OECD 국가의 비중은 2007년 40.5%에서 47.2%로 지속적인 증가 추세

## ■ 금융환경 변화에 의한 유가변동

- 금융 위기의 여파로 투자자들의 위험회피 성향이 확대되며, 달러화의 강세로 인한 상품 시장의 투자자금 이탈

## 추진내용

### ▣ 국내 주유소 유가 가격 데이터 수집

- 한국석유공사는 2011년 말, 데이터 분석 전문기업 SAS와 협력하여 유가예보 시스템 개발
  - 고유가에 따른 소비자 부담을 감소하기 위해 유가의 단기 미래 가격을 예측하여 제공하는 오피넷 시스템 구축
- ‘오피넷’ 웹서비스를 통해 국내 1,300여 개의 주유소로부터 수집된 휘발유 가격 정보 제공
  - 부가통신 사업자(VAN)와 함께 주유소 카드 단말기 결제 시스템을 통해 하루에 6차례 수집
  - 사업자가 관련 사업을 하기 위해서는 관련 정보를 석유공사에 제공하도록 석유 및 석유 대체연료사업법 개정(2009년)

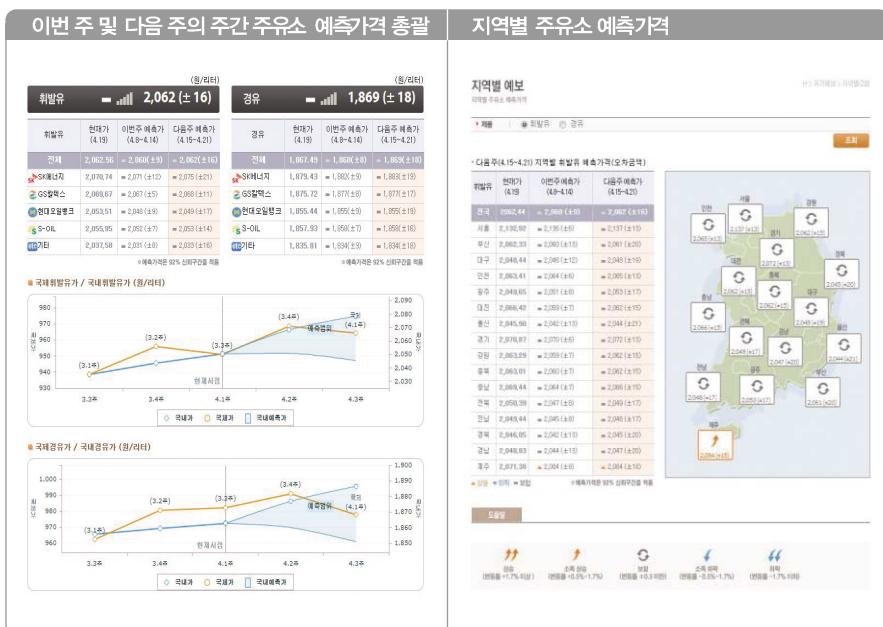
### 오피넷의 가격정보 및 유가예보 서비스



자료 : [www.opinet.co.kr](http://www.opinet.co.kr)

## ▣ 빅데이터로 예측한 1주일 뒤의 유가와 실제 유가 비교 가능

- 국제유가를 기반으로 국내 정유사와 주유소의 판매가격을 추정하는 예측모델 개발
  - 소비자들이 지역별, 상표별 및 5가지 변화 단계(상승, 소폭 상승, 보합, 소폭 하락, 하락)로 구분되는 시각화 자료 확인 가능
- 국내 유가예보 서비스를 통해 소비자들은 주유소별 휘발유 및 경유의 예상 판매가격을 확인한 후 직접 주유소를 선택하여 유류 구입 가능



자료 : Opinet 유가정보서비스

## ▣ 국제 유가변동에 따른 유가 예측

- 유가변동에 영향을 주는 다양한 변수를 고려한 국제·국내 유가 및 예측
  - 2년 동안 싱가포르 시장에서 발표하는 국제유가 및 시점별 환율 조사
  - 국제 유가변동이 국내 시장에서 유가에 영향을 미치는 시간이 7일에서 14일정도 소요되는 점을 이용하여 국내 유가변동 예측
  - 국제 유가 정세에는 정치적인 상황 및 기후, 환율과 같은 다양한 변수를 고려

## 효과 및 전망

### ■ 사용자 및 차량 중심으로 최저가의 유가 서비스를 제공 가능

- 사용자의 위치와 가장 가까운 곳 또는 가장 저렴한 유가 정보를 제공함으로써 시간과 에너지 감소 가능
- 유가 예측을 통해 휘발유 사재기 등의 부작용 방지
- 인근 주유소들 간의 유가정보를 공유함으로써 건전한 경쟁을 이루어 유가를 낮출 수 있는 환경 마련

### ■ 국제 유가에 민감한 국내 물가 안정에 기여

- 국제 유가 가격에 따라 국내 물가 또한 매우 큰 영향을 받기 때문에 국제 유가를 예측 함으로써 국내 물가를 포함한 국내 경제 전망에도 긍정적 기여 가능

### ■ 국제 유가에 영향을 미치는 다양한 변수의 과학적 분석 필요

- 현재 국제·국내 유가에 예측을 위한 변수는 매우 제한적이며 정치적인 문제에 대해서는 주관적인 판단 필요
- 과거의 데이터를 분석하여 유가변동에 영향을 미치는 다양한 변수를 포함한 과학적 분석 과정 필요

## 정책적 시사점

- 주유소에서 사용된 사용자 정보를 이용하여 다른 사용자에게 실시간으로 수집된 유가 정보를 제공함으로써 에너지 효율 증대
- 과학적 근거를 바탕으로 한 국내 경제 전망 및 계획 수립에 기여

# 빅데이터를 통한 건강한 사회

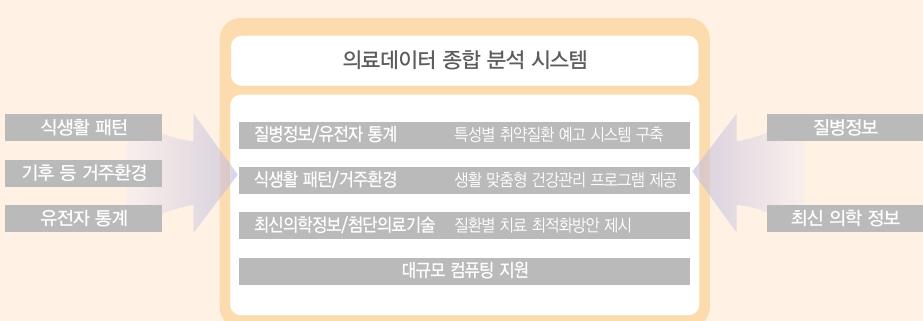


1. 미국 국립보건원, 유전자 데이터 공유를 통한 질병치료체계 마련
2. 미국 국립보건원, Pillbox 프로젝트를 통한 의료개혁
3. 건강보험회사 웰포인트(WellPoint), 슈퍼컴퓨터를 활용한 효율적인 환자치료
4. 구글, 검색어 분석을 통한 독감예보 서비스 제공

## 평생 질병 예측 및 건강관리 프로그램 제공

### 연구·의료기관의 의료데이터 공유 및 활용

#### 의료데이터 종합 분석 시스템



20대  
개인별 맞춤형 운동 프로그램 수립 지원



40대  
가족단위 건강관리 계획 수립,  
가족 의료비 절감



60대  
알츠하이默 등 노인성 질환 조기 예방 및 효과적 재활 치료

## 미국 국립보건원, 유전자 데이터 공유를 통한 질병치료체계 마련

다양한 질병을 연구하기 위해 유전자 데이터를 공유 및 분석할 수 있는 시스템 마련

- 주요 관리 대상에 해당하는 질병에 대한 관리 및 예측
- 1,700명의 유전자 정보를 아마존 클라우드에 저장하여 누구나 접근하여 데이터 이용 가능

### 추진목적 및 배경

#### ▣ 유전자 분석을 통한 질병 치료

- 인종별, 국가별 특성에 따른 유전자 특성 정보를 분석하여 질병 치료에 활용
  - 인간의 35,000개의 유전자에 따라 같은 질환이라도 증상이 다르게 나타나며, 약물 반응도 다르게 발생
  - 미국에서는 한 해 220만 명 이상이 약물 부작용에 의해 입원
  - 약물 부작용은 미국인 사망원인 4위일 정도로 심각

#### ▣ 1000 게놈 프로젝트를 이용한 인간의 다양성 분석

- 인간 유전체의 다양성을 알아내기 위한 국제협력 프로젝트
  - 개인의 유전체는 유전자의 기능이 무력화된 250개에서 300개의 기능을 잃은 돌연변이에 의해 발생
  - 발견된 돌연변이들 중 50%는 향후 개인별 유전체를 연구할 때 큰 도움이 될 것으로 전망

## 추진내용

- ▣ 2012년 3월, 오바마 대통령은 ‘빅데이터 연구개발 이니셔티브(Big Data Initiative)’ 발표
- 오바마 대통령은 2억 달러 이상을 투입하여 빅데이터 기술 개발을 추진할 예정이며, 빅데이터가 현재 미국이 당면한 현안 등을 해결할 수 있을 것으로 기대
- 이에 국립보건원(National Institute of Health, NIH), 국방부(Department of Defense, DOD), 국립과학재단(National Science Foundation, NSF), 에너지부 (Department of Energy, DOE) 등 주요 연방기관이 참여할 예정

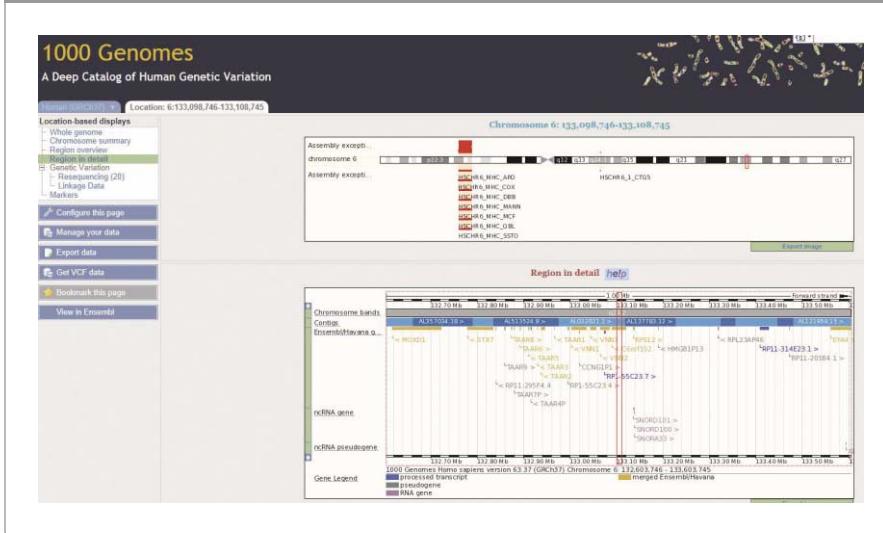
### | 참고 | 국립보건원의 빅데이터 프로그램 주요 내용

- ▣ 국립암연구소(National Cancer Institute, NCI)
  - 암과 관련된 영상 데이터의 공유 서비스를 추진하고, 대량의 유전체 염기서열을 분석하여 2014년 까지 페터바이트에 달하는 미가공 데이터 수집 예정
- ▣ 국립심장·폐·혈액연구소(National Heart, Lung and Blood Institute, NHLBI)
  - 사용자의 부담을 최소화하기 위한 협업추진 및 안전한 데이터를 저장·통합·분석하는 자원을 제공하여 심혈관계 연구를 위한 데이터 분석 툴 공유 추진
- ▣ 국립생의학영상 및 생체공학연구소(National Institute of Biomedical Imaging and Bioengineering, NIBIB)
  - 상호운영가능하고 일관된 정보를 제공하기 위한 데이터베이스 구축을 통해 나노재료의 제작 및 특성화에 대한 표준화·데이터 정보 제공
  - 환자 주도의 의료용 영상자료 공유를 위한 인터넷 기반 네트워크 구축 방안 모색
  - 유전자, 행동패턴, 인구, 기타 데이터 등을 미국 전역의 수집 사이트를 통해 수집하고, 데이터를 익명화시켜 전 세계 사용자들이 다양한 형태로 다운로드 가능하도록 지원

- ▣ 미국 국립보건원(NIH), 기업 및 기관들과의 파트너쉽을 통해 200TB의 유전자 데이터 확보
- 국립보건원과 75개 기업 및 기관들이 파트너십을 통해 공동으로 진행한 1000 유전체 프로젝트(1000 Genomes Project)의 일환으로 200TB의 유전자 정보 확보

- 1000 유전체 프로젝트는 과학 연구를 위해 전세계에서 2,662명의 유전자 정보를 저장하고, 질병 연구를 위해 1% 이상의 빈도를 나타내는 유전적 다양성을 분석
  - 국립보건원은 빅데이터 연구개발 이니셔티브 발표에 따라 1000 유전체 프로젝트를 아마존 웹서비스로 이전하여 저장
    - 연구원들은 유전자 데이터를 무료로 사용할 수 있으며, 사용한 만큼의 컴퓨팅 서비스에 대한 비용만 지불 가능

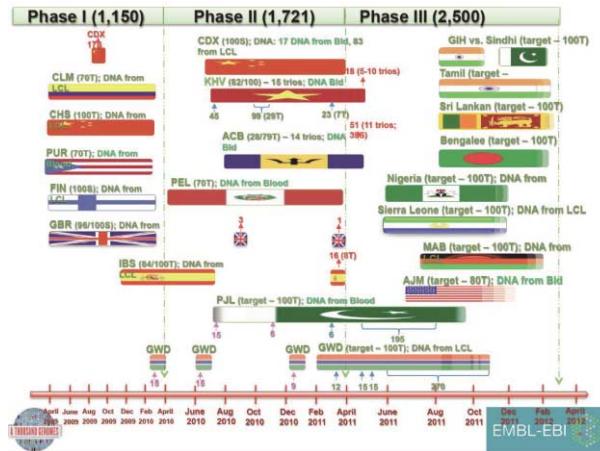
1000 Genomes Project 데이터 검색



자료 : The 1000 Genomes Tutorial The Website and Browser, 2012.2.17

- 아마존 웹 서비스(AWS)는 1,700명의 유전자 정보를 아마존 클라우드에 저장
  - 아마존 웹 서비스에 유전자 정보를 넣으면 아마존 웹 서비스 서버에 저장되며, 1,700개 이상의 유전자 정보는 익명의 개인들로부터 취합
  - 북동유럽에서 조상들이 미국으로 이주하고 현재 유타주에 거주하는 사람, 멤버에 사는 중국계, 로스앤젤레스에 사는 멕시코계, 미국 남서부에 거주하는 아프리카계 등 전세계 인구 표본에서 데이터 취합

## 아마존 웹서비스의 유전자 분석



자료 : The 1000 Genomes Tutorial A Brief History of Data and Analysis, 2012.2.17

## 효과 및 전망

- 유전자 정보를 공유함으로써 새로운 질병에 대한 빠른 진단 서비스 제공 가능
- 난치병 및 불치병과 관련된 유전자 정보를 공유하고 분석함으로써 새로운 치료제 개발 가능성 제시
- 최신 IT 기술을 결합함으로써 기존에 치료하기 힘들거나 오래 걸리는 질병에 대한 치료 확률을 높일 수 있는 기회 마련

# 5

## 미국 국립보건원, Pillbox 프로젝트를 통한 의료개혁

미국 국립보건원 (National Laboratory of Medicine)이 운영하는 약 검색 사이트

- 주요 관리 대상에 해당하는 질병에 대한 관리 및 예측
- 약 검색 서비스를 통해 얻어진 다양한 사용자의 질병에 대한 통계 데이터 활용
- 주요 질병의 분포 및 추세를 예측함으로써 국가차원의 조기 대응이 가능

### 추진목적 및 배경

#### ◆ 미국 건강보험개혁안 (AHA :Affordable Health Care)

- 의료보험 혜택 수혜율의 확대
  - 일정 소득 수준 이상인 시민이나 영주권자의 건강보험 가입 의무화
  - 빈곤 계층에 대한 의료보험 지원금 보조를 통해 건강보험 수혜율을 최대 95%로 증가시키는 것을 목표로 설정

#### ◆ 오바마 대통령은 의료기관, 환자, 정부, 의료보험회사를 통합하여 효율적으로 운영하기 위한 Health 2.0 제안

- 의료 비용 절감을 위해 IT 기술 적극 도입
  - 오바마 정권에서는 온라인 헬스케어뿐만 아니라 웹 2.0 기술을 결합한 온라인 웹 기술을 적극적으로 활용
  - 민간 웹 서비스에서 다양한 아이디어를 차용하여 매우 직관적인 인터페이스를 활용함으로서 누구나 손쉽게 접근할 수 있는 시스템 마련

## ■ 환자, 의료기관 및 보험회사간의 유기적 연결을 통한 정책 마련 및 비용 절감

- 환자, 의료기관, 보험수가와 관련된 기본 데이터의 수집, 저장, 통합, 안정과 같은 기본적인 업무 수행
- 환자와 병원을 연결해주기 위한 소셜 네트워크와 클라우딩 기능을 결합한 서비스로 발전
- 사용자가 문의하는 약에 대한 정확한 정보를 제공함으로서 의약품 오남용을 막기 위한 시스템으로 필박스(Pillbox) 모델 제시

## 추진내용

### ▣ 미국 국립 의료원 (National Laboratory of Medicine)에서 제공하는 의약품 정보 서비스

#### ■ 사용자가 요구하는 다양한 약에 대한 정보 제공

- 사용자가 사용 중인 약에 대한 정보가 불분명할 때 필박스를 통해 약에 대한 정확한 정보 확인
- 알약에 새겨진 글자, 번호, 색깔, 모양, 크기 등 간단한 약에 대한 설명만으로도 정확한 약의 효능 및 정보 제공 가능

#### ■ 알약의 제조사와 사용자간의 유기적인 쌍방향 상호작용을 통해 약에 대한 정보 제공

- 새로운 약을 새로 개발한 제약회사는 자유롭게 새로운 약에 대한 정보를 직접 입력할 수 있는 데이터시스템 제공
- 사용자 또한 직접 약에 대한 정보를 공유함으로써 사용자간의 유기적인 정보 공유 가능

#### ■ 남녀노소 쉽게 검색할 수 있는 시스템 제공

- 약에 대한 정보를 필요로 하는 사람의 대부분이 노년층이 많다는 점을 고려하여 누구나 쉽게 약을 검색할 수 있는 유저 인터페이스(UI) 제공
- 사용설명을 비디오로 제공함으로써 인터넷 환경에 취약한 계층도 사이트 접근이 용이하도록 설계

The screenshot shows the Pillbox website homepage. At the top, there's a logo for 'Pillbox' with 'BETA' above it, followed by the tagline 'rapid identification, reliable information'. To the right, there's a logo for 'United States National Library of Medicine National Institutes of Health' with the NLM logo. Below the header, a navigation bar includes 'Home', 'About', 'FAQ', 'API', and 'Feedback'. The main content area has a blue background with the title 'Two ways to identify an unknown pill'. It features two screenshots: one for 'Quick Search' showing a grid of various pills and capsules, and another for 'Advanced Search' showing a more detailed search interface. Below each screenshot, there's a brief description. On the left, under 'What?', it says 'Pillbox enables rapid identification of unknown solid-dosage medications (tablets/capsules) based on physical characteristics and high-resolution images.' and 'Once a medication is identified, Pillbox provides links to drug information and drug labels.' On the right, under 'How?', it says 'Combining data derived from drug labels submitted to FDA by drug manufacturers and distributors and NLM's RxNorm, Pillbox is an identification and reference system for oral solid-dosage medications.' and 'Pillbox's data and search engine are also available through an API.' At the bottom of the page, there's a footer with links to 'National Library of Medicine', 'National Institutes of Health', 'U.S. Dept. of Health & Human Services', 'Copyright', and 'Accessibility'. A small note at the bottom center says 'Last updated: February 24, 2012'.

## 효과 및 전망

### ■ 약의 정체 및 기능을 확인하는 비용 절감

- 한 해 동안 미국 국립보건원에 접수되는 알약의 기능이나 유효기간을 문의하는 민원 수는 100만 건 이상
- 평균 한 건당 알약의 기능 및 유효기간을 확인하는데 필요한 비용은 약 50달러였으나, 필박스 서비스를 이용함으로써 연간 5,000만 달리의 비용 절감

### ■ 빅데이터를 이용한 의약품 사용에 대한 정보 제공 가능

- 사용자가 검색한 약의 정보를 통해 의약품 사용과 관련된 지도 및 그래프를 입체적으로 작성하여 현재 유행하고 있는 질병의 발생 장소 및 전염속도에 대한 분석 가능
- 사용자에 의해 만들어진 데이터를 통해 질병의 전염속도, 인적대응 방법, 방제인원에 대한 효율적인 방제대책 마련 가능
- 현재 필박스에서 제공하는 시스템만으로는 사용자의 의약품 검색 의도를 구분할 수 없으므로 정확한 질병에 대한 대책 마련을 위해서는 검색의도에 대한 정보 추가 필요

## 정책적 시사점

### ■ 환자, 의료기관, 보험회사간의 유기적인 질병관리 시스템 개발

- 실시간으로 약의 판매 현황 및 진찰 기록을 관리함으로써 의료 진단 시스템의 투명한 관리 기능
- 환자가 복용하고 있는 약에 대한 정보를 순차적으로 기록하여 약의 오남용 방지

### ■ 환자에 대한 의료정보에 대한 보안 필요

- 현재 필박스에서는 사용자의 개인정보를 필요로 하지 않지만 정확한 질병에 대한 예측을 위해서는 사용자의 개인정보 수집이 필요한 경우가 발생하므로, 개인의 의료정보가 유출되지 않도록 정보보안 유의

# 6

## 건강보험회사 웰포인트(WellPoint), 슈퍼컴퓨터를 활용한 효율적인 환자치료

### 의료진의 진단과 환자 치료에 슈퍼컴퓨터를 사용하는 시스템

- 임상실험 결과 및 베스트 프랙티스 같은 과거 사례를 분석하여 진단을 및 치료를 향상
- 환자의 증상, 환자 면담결과, 진단 연구 등의 진료 내역 데이터 활용
- 환자에게 적절한 치료법을 제시하고 최신 정보를 과학적인 방법으로 제시 가능

### 추진목적 및 배경

#### ▣ IT 융합기술이 의료 시장의 기술 혁신 주도

- 고령화 사회로 진입함에 따라 의료비용이 급증하고, 병원의 정보화를 통해 효율화를 극대화 하는 디지털병원에 대한 관심 증가
- 대규모로 축적되는 의료 정보의 효율성을 추구하는 Health 2.0과 같은 혁신적인 의료 IT 융합 기술에 대한 연구 진행
- 첨단 기술간의 접목을 통하여 과학 기술적 한계를 극복하고 의료 산업 발전에 혁신적인 기여 가능
  - 국내에서는 통신망을 이용하여 의료진 간에 의료영상 등 의료정보를 주고 받으며 협진 하는 원격진료 시스템 등장
  - 각종 초소형 센서의 개발을 통해 언제 어디서나 건강과 질병에 관련된 정보를 손쉽게 측정 가능

### ■ 고령화 사회에서의 효율적인 의료 서비스의 필요성 대두

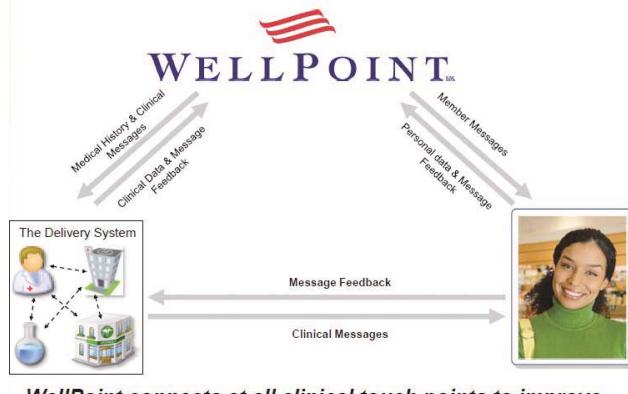
- 국내 고령인구에 의한 의료비는 2009년 이미 전체 의료비의 30.5%에 이르며 선진국은 전체 의료비의 40~50%를 차지
- 노인 의료비 중 70~80%가 당뇨병, 고혈압, 심장질환 등의 만성 질환에 의한 것으로 의료비의 부담은 GDP의 6% 이상
- 만성질환은 유헬스(u-Health)를 통해 효율적으로 관리함으로써 완화할 수 있으며, 전체 의료비의 7.2%인 1.5조원의 의료비 절감효과 발생

## 추진내용

- ▣ IBM과 미국 의료보험사인 웰포인트(WellPoint)는 의사와 다른 의료진들이 진단과 환자 치료에 사용할 수 있는 애플리케이션 제공
- ▣ IBM의 웨일스 솔루션을 도입해 건강보험 자료와 회사에 등록된 3,420만 명에 대한 환자 정보를 통합 분석하여 이를 기초로 복잡한 의학적 치료법 검색
- 다양한 환자의 의료 기록 수집을 통해 데이터 분석
  - 환자의 증상, 환자 면담 결과, 진단 연구 등 진료 내역에 대한 모든 정보를 저장
  - 8코어 프로세서의 IBM 서버를 통해 모든 사례를 고려하여 최적의 진단 및 환자 치료 가이드라인 제시
  - 2억 페이지에 해당하는 자료를 검색 분석하여 3초 안에 결과 제시 가능

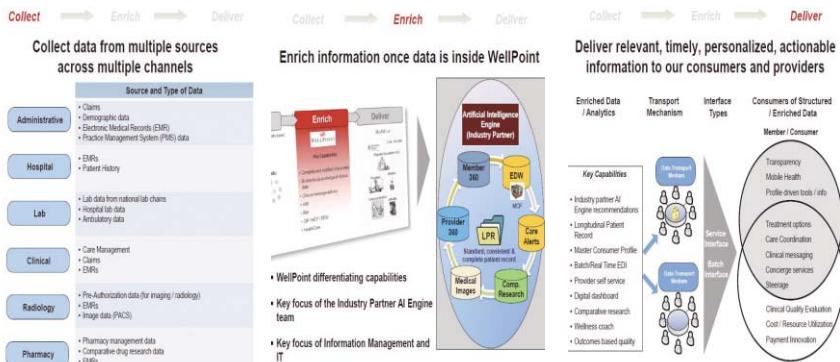
## 웰포인트의 의료정보 제공 체계

### High Level Flow of Health Information



자료 : Transforming the Information Infrastructure: Build, Manage, Optimize, Computerworld, 2011.

### 웰포인트의 데이터 수집(collect), 강화(enrich), 전달(deliver) 프로세스



자료 : Transforming the Information Infrastructure: Build, Manage, Optimize, Computerworld, 2011.

## 효과 및 전망

- 환자의 상황에 맞는 가장 최선의 치료 방법 제시 가능
  - 빠르게 변화하는 진단 및 치료 방법을 서로 공유하여 환자의 진료 치료 만족도 증대
  - 불필요한 치료 및 진료를 줄여 환자 및 의료보험 회사의 불필요한 진료비 낭비 방지
  - 고령화 사회로 접어들면서 노인들의 만성 질환을 체계적으로 관리하여 고령층에 대한 효과적인 진료 서비스 제시

## 정책적 시사점

- 빠르게 변화하는 의료 환경에서의 진단 및 치료 방법을 환자, 의료보험회사 및 의사가 공유함으로써 환자에게 최적의 치료 서비스 제공 가능
- IT 기술을 응용한 의료 기기의 보급과 네트워크 환경의 개선으로 원격 치료가 가능해져 환자 중심으로 의료 시장 개편 가능

# 7

## 구글, 검색어 분석을 통한 독감예보 서비스 제공

### '감기'와 관련된 검색어 분석을 통한 독감 예보 시스템 제공

- 구글 검색 사이트에 사용자가 남긴 검색어의 빈도를 조사, 독감 환자의 분포 및 확산 정보 제공
- 사용자의 다양한 검색어를 분석하여 다시 사용자들에게 유의미한 정보로 피드백 가능

### 추진목적 및 배경

- ▣ 네트워크 환경 개선에 따라 검색엔진 사이트를 통한 검색어 입력 빈도가 증가
  - 무선랜(WiFi) 이용지역 확대, 스마트폰 등 스마트 디바이스의 빠른 확산 등 편리한 인터넷 사용 환경이 마련
  - 인터넷을 통한 정보 검색이 보다 용이해지며, 사용자들의 현재 상태와 이에 따른 후속 상황을 예측할 수 있는 정보들이 실시간으로 누적
- 
- ▣ 검색 정보에 사용자의 현재 상태나 상황에 관한 내용이 담겨있으므로, 실시간으로 무수히 많이 누적되는 정보 속에서 사회적인 변화나 흐름을 파악하는 것이 가능

### 추진 내용

- ▣ 구글 홈페이지에서 독감, 인플루엔자 등 독감과 관련된 검색어 쿼리의 빈도를 조사, '구글 독감 동향(Google Flu Trends)'이라는 독감 확산 조기 경보체계 마련

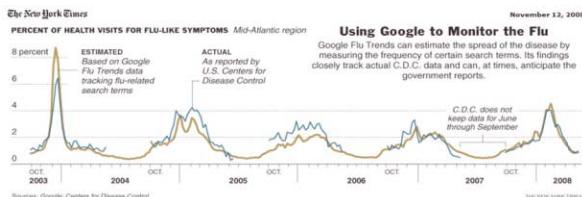
- 미국 보건 당국보다 한발 앞서 시간 및 지역별 독감 유행 정보를 제공
- 미국 질병통제 예방센터의 데이터와 비교 결과, 검색 빈도 및 실제 독감증세를 보인 환자 수자 사이에 매우 밀접한 상관관계가 있는 것을 확인

### 구글 독감 동향(Google Flu Trends)



자료 : <http://www.google.org/flutrends/>

### 구글 독감 동향 및 실제 확산의 비교 결과



자료 : 뉴욕타임즈, 2008. 11. 12

## 효과 및 전망

- 다양한 사용자의 검색어를 분석, 사용자에게 다시 유의미한 데이터로 제공 가능
  - 세계 곳곳에서 다양한 사용자의 검색을 통해 정치, 문화, 경제, 스포츠와 관련된 사용자의 관심 분야와 내용이 매우 빠르고 정확하게 전달

## 정책적 시사점

- 실시간 사용자의 의도를 분석할 수 있는 기술 개발로 더 빠르고 정확한 통계의 확보 가능

# 빅데이터로 희망하는 안전한 사회

1. 싱가포르, 국가위험관리시스템(RAHS)을 통한 국가안전관리
2. FBI, 유전자 색인 시스템 활용한 단시간 범인 검거 체계 마련
3. 샌프란시스코, 범죄 예방 시스템으로 안전 지역사회 구축

## 선제적 위험예측 및 국민안심서비스 마련



# 8

## 싱가포르, 국가위험관리시스템 (RAHS)을 통한 국가안전관리

국가적 차원의 위험요인과 기회요인을 선제적으로 파악, 대응 방안을 수립하기 위한 시스템

- 미래에 영향을 미칠 수 있는 잠재적 위험요소와 불확실성 요소들을 탐색하여 이머징 이슈를 분석
- 빅데이터를 활용, 시뮬레이션 및 시나리오 기법 등을 통해 국가위험관리 계획의 수립 및 추진
- 국가와 국민의 위험 요소를 사전에 예측하고 대응 방안을 모색하기 위한 과학적 시스템 마련

### 추진목적 및 배경

#### ▣ 빈번한 테러 발생 및 국가적 재난, 전염병 등의 확산

- 9.11 이후 국제 사회는 테러와의 전쟁을 전개해 오고 있으나 테러의 공포로부터 벗어나지 못하고 있는 상황
  - 국내 여러 계층간, 국가간, 또는 종교적인 갈등에 의한 테러가 전개
  - 민간인을 대상으로 한 무차별적 테러로 인해 국민들의 불안감 증대

- 재난, 전염병 등 국민의 안전과 건강을 위협하는 요소에 대한 사전 감지 및 예방

- 환경안보, 해상안보, 사이버안보, 식량안보 등 다양한 재난 상황을 시뮬레이션
  - 상황별 대응조치 마련 및 위기시 즉각적으로 대응하는 지원체계 수립

#### ▣ 글로벌 금융위기로 인한 전 세계적 파급효과

- 유로존 국가의 위기가 전 세계 금융시장으로 확대 되는 동시에 실물 경제로의 전이가 본격화
  - PIIGS 국가의 GDP 대비 재정 적자 비율과 국가채무비율이 2007년 서브프라임 사태 이후 악화되면서 이를 국가의 신용위험이 상승
  - 유로존 국가 및 금융기관 간의 상호 연계성 심화로 인한 재정 위기의 빠른 확산

## 추진내용

- ❖ 싱가포르의 국가위험관리시스템(RAHS: Risk Assessment and Horizon Scanning)을 통해 질병, 금융 위기 등 모든 국가적 위험을 수집 및 분석
- 국가에 위험을 미칠 수 있는 다양한 데이터를 수집하여 분석
  - 불확실한 미래의 대비를 위하여 2004년부터 빅데이터를 기반으로 한 위험 관리 계획 추진
  - 수집된 정보는 시뮬레이션, 시나리오 기법을 통해 분석하여 사전 위험 예측 및 대응방안 모색에 활용

**RAHS 솔루션 센터의 분석추진단계**

PHASE	OBJECTIVE	ICON
ENVIRONMENTAL SCANNING	Identify early warning signals and trends to identify and monitor potential disasters and emerging issues.	
ISSUES TO INDICATORS	Identify key indicators process is a method for translating Emerging Change Issues to Indicators or high-priority issues for monitoring. This phase also identifies how an integrated method balances big picture issues with specific details.	
SENTIMENT ANALYSIS	Identify the most salient web mentions to augment the index in discovering strength of open source information.	
NARRATIVE CAPTURE	Analyze approach to better understand the environment by identifying positions, sentiments and possible risks within political, cultural and social contexts.	
SKAN TO TREND	To identify strong and trending signals from emerging issues through tracking emerging themes & relevant news.	
EMERGING STRATEGIC ISSUES	Identify, prioritize and project emerging issues through their impact on the progression of global supply chain.	
SCENARIOS TO STRATEGIES	Develop scenarios using known socio-economic and political changes to generate and reassess changes.	
DATA FUSION & ANALYSIS	Using data fusion, it is possible to bring together large amounts of data from multiple sources and systems.	
QUANTITATIVE MODELLING	Quantitative modeling is an approach analysis can use to better understand the dynamics of complex systems and confirm assumptions to assess the impact of changes of the system over time.	

자료 : [www.hsc.gov.sg](http://www.hsc.gov.sg)

## 효과 및 전망

- 국가 및 국민의 위험 요소를 파악, 대비함으로써 국민의 생명과 재산을 보호
- 수많은 데이터와 변수를 동시에 고려하는 전천후 국가 위험 관리 체계로 발전



자료 : 선진국의 데이터기반 국가미래전략 추진현황과  
시사점, 한국정보화진흥원, 2012. 4.

자료 : 新 가치창출 엔진, 빅데이터의 새로운 가능성과  
대응 전략, 한국정보화진흥원, 2011. 12.

## 정책적 시사점

- 정부가 보유한 다양한 정보를 체계적으로 관리 및 분석함으로써 효율적인 국가 위험 관리 시스템으로 발전
  - 국가의 위험 요소는 국내적 요소 외에 국가간 발생 요소들이 더 중요하므로 다양한 변수를 고려하는 것이 필수

## FBI, 유전자 색인 시스템 활용한 단시간 범인 검거 체계 마련

유전자 색인 시스템을 통해 유전자 분석표를 대조함으로써 사건을 해결

- 과거 범죄자의 유전자 데이터를 기반으로 범죄자 색출
- 유죄 판결을 받은 혐의자들의 혈액, 정액 및 기타 법의학적 증거에서 추출된 유전자 분석표 데이터 이용
- 과거 범죄자들의 유전자 데이터 분석을 통해 빠른 시간 내에 범죄 용의자 추적이 가능

### 추진목적 및 배경

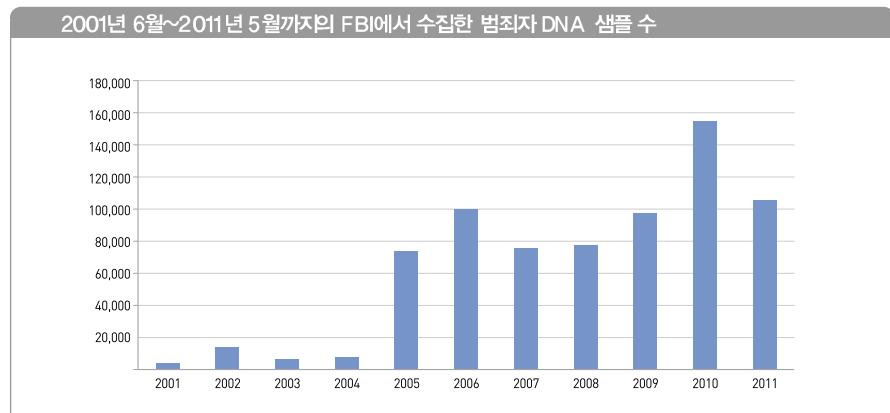
#### ▣ 유전자 데이터를 활용한 범죄 해결 방안

- 범죄자 유전자 데이터베이스를 통해 흉악 범죄자 검거를 위한 과학적 수사 기반 마련
  - 유전자 감식 결과 도출된 고유한 패턴을 데이터베이스에 저장, 검색할 수 있도록 해당 법안 개정
  - 구축된 유전자 데이터베이스를 기초로 입력 대상, 활용범위 등을 지속적으로 확장시켜 효율적인 범죄자 감식 시스템 마련

### 추진 내용

#### ▣ 유전자 정보은행 CODIS(Combined DNA Index System) 구축

- CODIS에는 미제 사건 용의자 및 실종자에 대한 DNA 정보 1만 3,000건을 포함한 12만 명의 범죄자 DNA 정보 저장
  - FBI는 매년 2,200만 명의 DNA 샘플을 추가하여 범죄 수사에 활용

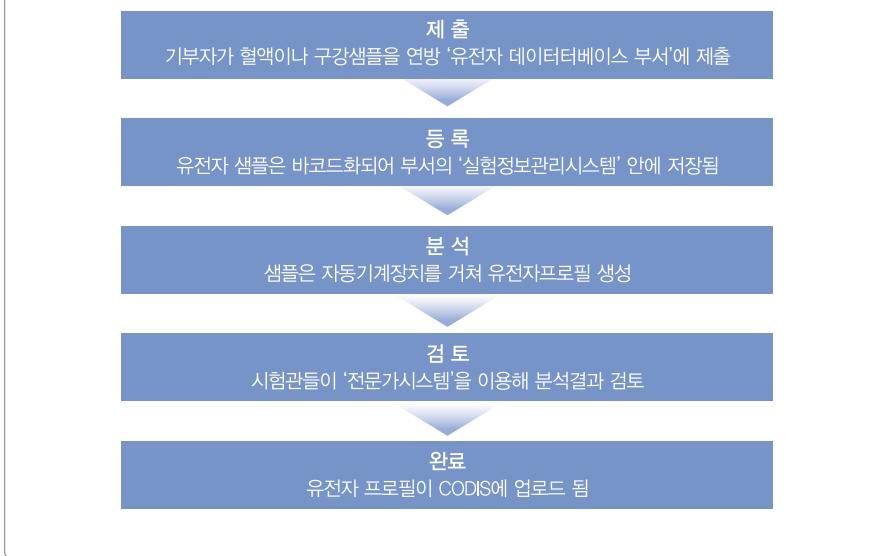


자료 : U.S. Department of Justice Office of the Inspector General Audit Division, 2011.9

- CODIS은 50개 모든 주와 연방 정부가 수집한 확정 판결을 받은 범죄자들과 일부 체포자들 에게서 추출한 DNA 분석표로 구성
  - 약 350만 개의 DNA 분석표가 내장

**FBI CODIS 웹사이트 : [www.fbi.gov/about-us/lab/codis](http://www.fbi.gov/about-us/lab/codis)**

## FBI의 DNA 데이터분석 과정



자료 : The FBI's Federal DNA Database Unit

## 효과 및 전망

- 범죄 사건 해결의 획기적 성과 달성
  - 내장된 DNA 분석정보를 활용하여 2007년 45,400건의 범인 DNA 적중도를 달성
  - 1시간 내에 범인 DNA 분석을 위한 주정부 데이터베이스 연계 및 빅데이터 실시간 분석 솔루션 확보
- 과거 범죄자들의 유전자 정보를 데이터베이스화하여 빠르고 과학적인 수사가 가능
  - 잘못된 용의자가 피해를 보는 경우를 최소화
- 민감하고 소중한 개인정보의 오남용 우려 존재
  - 개인 식별 유전자 정보는 법적으로 보호 받아야 하는 매우 민감한 정보 중의 하나이므로 신원 확인 이외의 목적으로 사용할 수 없도록 하는 제도적 장치 필요

# 10

## 샌프란시스코, 범죄 예방 시스템으로 안전 지역사회 구축

### 범죄 발생 지역 및 발생 시각을 예측하여 범죄를 미연에 방지하기 위한 시스템

- 범죄 발생 지역 및 시각을 예측하여 필요한 곳에 경찰 인력을 배치, 범죄를 예방
- 과거 발생한 범죄 패턴을 분석하여 후속 범죄 가능성 예측
- 과거 데이터에서 범죄자 행동을 분석하여 사건 예방을 위한 해법 제시

### 추진목적 및 배경

#### ▣ 경제 악화로 인한 흉악 범죄의 증가

- 글로벌 경제 위기에 따라 흉악 범죄가 증가하고 있는 추세
  - 경제 위기로 위한 양극화가 심화되면서 서민을 상대로 한 범죄가 빠르게 증가하는 추세
- 사건 발생 빈도에 비해 경찰 인력의 부족
  - 수많은 사건의 해결과 사전 방지를 위해서는 다수의 경찰 인력이 요구
  - 효율적인 범죄 예방을 위해 사건 발생 장소의 예측과 이에 따른 인력의 집중 배치 필요

#### ▣ 동일한 장소에서의 범죄 재발 가능성

- 빈집, 빈차량이 등의 범죄는 사건 발생 현장 또는 그 인근에서 제2, 제3의 범행이 재발할 가능성 높후

## 추진 내용

- ❖ 과거 범죄 기록 분석을 통한 효율적 경찰 인력 배치
- 과거의 범죄 데이터를 분석하여 범죄 정보 업데이트
  - 과거 8년 동안 범죄가 발생했던 지역과 유형을 세밀하게 분석하여 후속 범죄 가능성을 예측함으로써 범죄를 사전 예보하는 방식을 이용
  - 과거 범죄에 대한 통계 정보를 제공하는 것과 달리 새로운 범죄 가능성 정보를 제공



자료 : [www.crimemapping.com](http://www.crimemapping.com)

## 효과 및 전망

- 획기적인 범죄 예방 및 감소
  - 6개월간의 테스트 결과, 예보 정확도가 71%에 달하였으며 범죄가 예보된 10곳 중 7곳에서 실제 사건이 발생
- 효율적인 경찰 인력 배치 및 순찰 시스템으로 발전
  - 제한된 경찰 인력으로 광범위한 영역을 순찰하고 범죄를 예방하는 시스템으로 발전 함으로써 안전한 사회 구현에 이바지

## 정책적 시사점

- 과거 범죄자 및 범죄 유형을 소셜 네트워크를 통해 지속적으로 관찰함으로써 그와 관련된 조직 및 범죄에 대한 예방 가능

- 
1. 본 보고서는 정보통신진흥기금으로 수행한 정보통신연구개발사업의 연구 결과입니다.
  2. 본 보고서의 내용은 한국정보화진흥원의 공식 견해와 다를 수 있습니다.
  3. 본 보고서 내용에 대해 무단전재를 금하며, 가공 · 인용할 때는 반드시 「한국정보화진흥원」이라고 출처를 밝혀 주시기 바랍니다.

작성 및 문의 : 한국정보화진흥원 빅데이터 전략연구센터

윤미영(yoonmy@nia.or.kr, 02-2131-0325)

권정운(jekwon@nia.or.kr, 02-2131-0734)

자문 : 윤상민 교수(연세대학교 미래융합기술연구소,

sangmin.yoon@yonsei.ac.kr)

---



## Big Data 글로벌 10대 선진 사례

빅데이터로  
세상을 리드하다



한국정보화진흥원

(100-775) 서울시 종구 청계천로 14 한국정보화진흥원 빅데이터 전략연구센터

TEL 02-2131-0114 FAX 02-2131-0109

[www.nia.or.kr](http://www.nia.or.kr) [www.bigdataforum.or.kr](http://www.bigdataforum.or.kr)