Practical realization of Data Security and Genome Database

Erkki Leego
Head of IT Department
Estonian Genome Center,
University of Tartu
Biobanks

• Biobank
  – An organized collection of human biological material and associated information stored for one or more research purposes

• Population biobank
  – The collection has a population basis
  – To supply biological materials or data derived therefrom for multiple future research projects
  – It contains biological materials and associated personal data
Background of the Estonian Biobank

- Research institute of the University of Tartu
- Longitudinal, prospective, population based biobank, established in 2000
- 52,000 gene donors recruited
- 5% of the adult population – 18 years and older
- Supported directly by the government
- The project is conducted according to the Estonian Human Genes Research Act
Data “infrastructure” for research

- Good quality data collection
  - Biomedical data (DNA, plasma, WBC)
  - Phenotype
  - Genealogy
  - Epidemiology data
  - Genotypes
- Open to various research projects
- Well defined procedures for data release
- Usable next 30 years
Questionnaire of EGCUT

- Personal data
  - Place of birth
  - Place of residence
  - Nationality
  - Education
  - Occupation

- Genealogy
  - Parents
  - Children
  - Siblings
  - Grandparents

- Health behavior
  - Smoking and alcohol
  - Personality inventory NEO-PI-3
  - Physical activity EQ-5D

- Diseases
  - Diagnosis ICD – 10
  - Treatment ATC
  - Psychiatry module M.I.N.I. and SSP

- Health behavior
  - Nutrition
  - Questions about diabetes
  - Questions about cardiovascular diseases

- Objective data
  - Height & weight
  - Blood pressure
  - Pulse
  - Handedness
  - Waist
  - Hip

- Chronotype questionnaire MCTQ

- Nutrition

- Health self-assessment
Diagnoses in the database
(50155 participants)

- Certain Infectious And Parasitic Diseases A00-B99 (64245)
- Neoplasms C00-D48 (8928)
- Diseases Of Blood And Blood-Forming Organs And Certain... (3106)
- Endocrine, Nutritional And Metabolic Diseases E00-E90 (12677)
- Mental, Behavioural Disorders F00-F99 (10766)
- Diseases Of The Nervous System G00-G99 (11833)
- Diseases Of The Eye And Adnexa H00-H59 (19918)
- Diseases Of The Ear And Mastoid Process H60-H95 (13833)
- Diseases Of The Circulatory System I00-I99 (40209)
- Diseases Of The Respiratory System J00-J99 (60778)
- Diseases Of The Digestive System K00-K93 (39490)
- Diseases Of The Skin And Subcutaneous Tissue L00-L99 (12585)
- Diseases Of The Musculoskeletal System And Connective Tissue... (37322)
- Diseases Of The Genitourinary System N00-N99 (21170)
- Pregnancy, Childbirth And The Puerperium O00-O99 (3352)
- Certain Conditions Originating In The Perinatal Period P00-P96 (134)
- Congenital Malformations, Deformations, And Chromosomal... (1405)
- Symptoms, Signs And Abnormal Clinical And Laboratory Findings... (3202)
- Injury, Poisoning And Certain Other Consequences Of External... (5710)
- Factors Influencing Health Status And Contact With Health Services... (1610)
- External Causes Of Morbidity And Mortality V01-Y98 (619)

- In average a participant has reported 7.8 different diagnoses
- 98.3% of participants have reported at least one disease
Education: the Estonian population vs. EGCUT
(50155 participants)

- Scientific degree
- Higher education
- Professional secondary education
- Secondary education
- Basic education
- Elementary education
- No elementary education
- Education unknown

%
Realization and data security
Network of recruiters

- Primary care providers (PCPs)
- Network of 640 recruiters
  - 454 GPs (56% of all GPs in Estonia)
  - 186 senior nurses and nurses
- 30h training
  - genetics, biobanking, data protection, ethics & law
Data and sample collection

Filled questionnaires
Personal data, health data, genealogy data.
Consent. Transportation code.
Information encrypted.

Consent (paper form).
Biological samples.

Coding center

Consent forms. Questionnaires.

Communication server

National Digital Health Record DB & registries

Memory stick

Health data and identification data

Operative database

Phenotype database

Data collector IS

High security area

Coding center IS

Laboratory IS

Cryo Preservation – MAPI

Courier

Consent (paper form), Biological samples.

Receptionist

New barcode

Scientists

Access to database

Eesti geenivaramu
tartu ülikool
Coding challenges

EXTERNAL DATABASES
- data collecting

BIOSAMPLES DATA
- logistics
- management

HEALTH DATA
- data collecting

PERSONAL-, GENEALOGY-, HEALTH DATA
- Formatting for integration, deidentification

PHENOTYPES
- integration, quality control, analysis

GENOTYPES
- raw data, quality control, analysis

BIOSAMPLES DATA
- analysis

CODE RELATIONS
- DVL

DATA RELEASE
- genotypes, phenotypes, biosamples, analysis

FEEDBACK
- PERSONALIZED MEDICINE

FOLLOW UP
- coordination

PROJECT-BASED RESEARCH
- genotypes, phenotypes, biosamples, analysis
Data updates

• Follow-up and re-examining participant
• National registries
  – Citizen Registry
  – Estonian Causes of Death Registry
  – Estonian Cancer Registry
  – Estonian Tuberculosis Registry
• Hospitals
• National Health Insurance Fund
• Estonian National Health Information System
Human Genes Research Act
Enforced 08.01.2001

• HGRA regulates
  – scientific research on human genetics
  – establishment and maintenance of the biobank
  – use of genetic information (informed consent)
  – legislation is forbidding third party access to the database (police, employers, insurance companies etc.)

• HGRA protects
  – confidentiality of the gene donor
  – public from the misuse of the genetic information
  – gene donor from the genetic discrimination

• HGRA allows re-contacting and collection of health data from other registries

• Gene donors have the right to get feedback on their genetic information
  • feedback should be accompanied by clinical counseling
Public opinion very positive and awareness high
EGCUT servers and storage
June 2012

- 15 servers
- Datastorage 1,1 PB harddrives
- 1,2 PB system storage tape library

**EGCUT Network**

- Datastorage with servers 1,1 PB HDD
- Analyser server 24 Core, 48GB RAM (Clone off-site)
- Illumina Genome Analyser
- Hamilton ASM
  - ASM Store
  - ASM Server
- BC/MAX
- MAPI

**High Performance Computing Centre**

- Computing power
  - 3100 processor core
  - 6,2 TB RAM

- **EGCUT servers and storage**
  - June 2012
  - 15 servers
  - Datastorage 1,1 PB harddrives
  - 1,2 PB system storage tape library
  - Infrastructure Servers
    - 24 Core, 192GB RAM, 1,76TB HDD
  - Database Collator
    - Phenotype Database Laboratory IS
  - Operative Database
  - Coding Centre IS
  - System Storage Tape Library
    - 1,2PB storage
  - Coding Centre IS
  - Backup Server
    - 16 Core
    - 1,2TB HDD
    - 24GB RAM

**HPC Cluster**

- "Aurumasin"
  - 52 node,
  - 512 Core,
  - 2,3TB RAM,
  - 50TB HDD
- "Vedur"
  - 80 node,
  - 2560 Core,
  - 3,75TB RAM,
  - 110TB HDD

Serverite ülevaade 2012-05-30a RHansson
Tools and enablers of information exchange

• One universal national identification code
  – Registries and databases use to uniquely identify persons
• National PKI infrastructure
• The Estonian ID card
  – Smartcard with two digital certificates
• National Data Exchange Layer X-Road
• Obligatory national data security framework
• High public acceptance and trust
  – No public incidents or misuses (10 years)
The Estonian ID card

• The ID card is a **mandatory** ID document for all Estonian residents from the age of 15
• Enables secure digital authentication and signing
• A digital signature has the same legal consequences as a hand-written signature
• Does not have any additional information
  – No bank account, no health information etc.
• Active cards: **1 192 102** (08.02.2013)
  – Estonian Population **1 286 540** (01.01.2013)
  – Estonia has been issuing ID cards from January 1st 2002
Public Key Infrastructure PKI

• PKI or the public key infrastructure enables secure digital authentication and signing
• The infrastructure also allows forwarding data by using an encrypting key pair: a public encryption key and a private decryption key
• In Estonia, this technology is used in relation with electronic identity (ID card, mobile ID, digital ID)
• Certification Service and Time-stamping Service provider is non-governmetal
Data Exchange Layer X-Road

- Technical and organisational environment, which enables secure Internet-based data exchange between the state’s information systems
5 main principles of security of Estonian E-health system

1. A secure authentication of all users (ID-card)
2. A maximum accountability (transparency)
   - All action will leave an unchangeable (and unremovable) secure trail
3. Coding of personal data
   - Separating of personal data from medical data
4. Encrypted database
   - Allows to remove the confidentiality risk from the technical administrators
5. Effective monitoring tool
   - All actions are monitored and corresponding countermeasures are applied
Personal control as security measure

• People have easy and universal access with ID-card
  – No need for usernames or other access methods
• People can see what data is available about them
• People can see who has been accessing their data
• People can give legal commands online (ID-card)
  – For example: Person can close doctors access to his/hers EHR data in Patient Portal
National data security framework

• Legislation
  – Public Information Act
  – Personal Data Protection Act
  – Electronic Communication Act

• Three-level IT baseline security framework ISKE
  – Government Regulation, obligatory for Public sector

• Supervision
  – Data Protection Inspectorate defends citizens constitutional rights
  – Estonian Information Systems’s Authority inspects the security of the information systems of state and local government agencies and providers of vital services
  – Computer Emergency Response Team (CERT) Estonia
Conclusions

• Estonia has great potential to implement state level personalized medicine solutions
  – Genetic research with 5% of population genetic and continuously updated phenotype information
  – Nation wide Health Information Exchange platform
  – 10 years of experience of national level e-services (PKI, X-Road, ID-card, security framework)
  – High level public trust and acceptance
Thank you!

• Additional information:
  – Erkki Leego, erkki.leego@ut.ee
  – http://www.geenivaramu.ee/en/ (Estonian Genome Center, University of Tartu)
  – http://www.e-tervis.ee/ (Estonian E-Health Foundation)