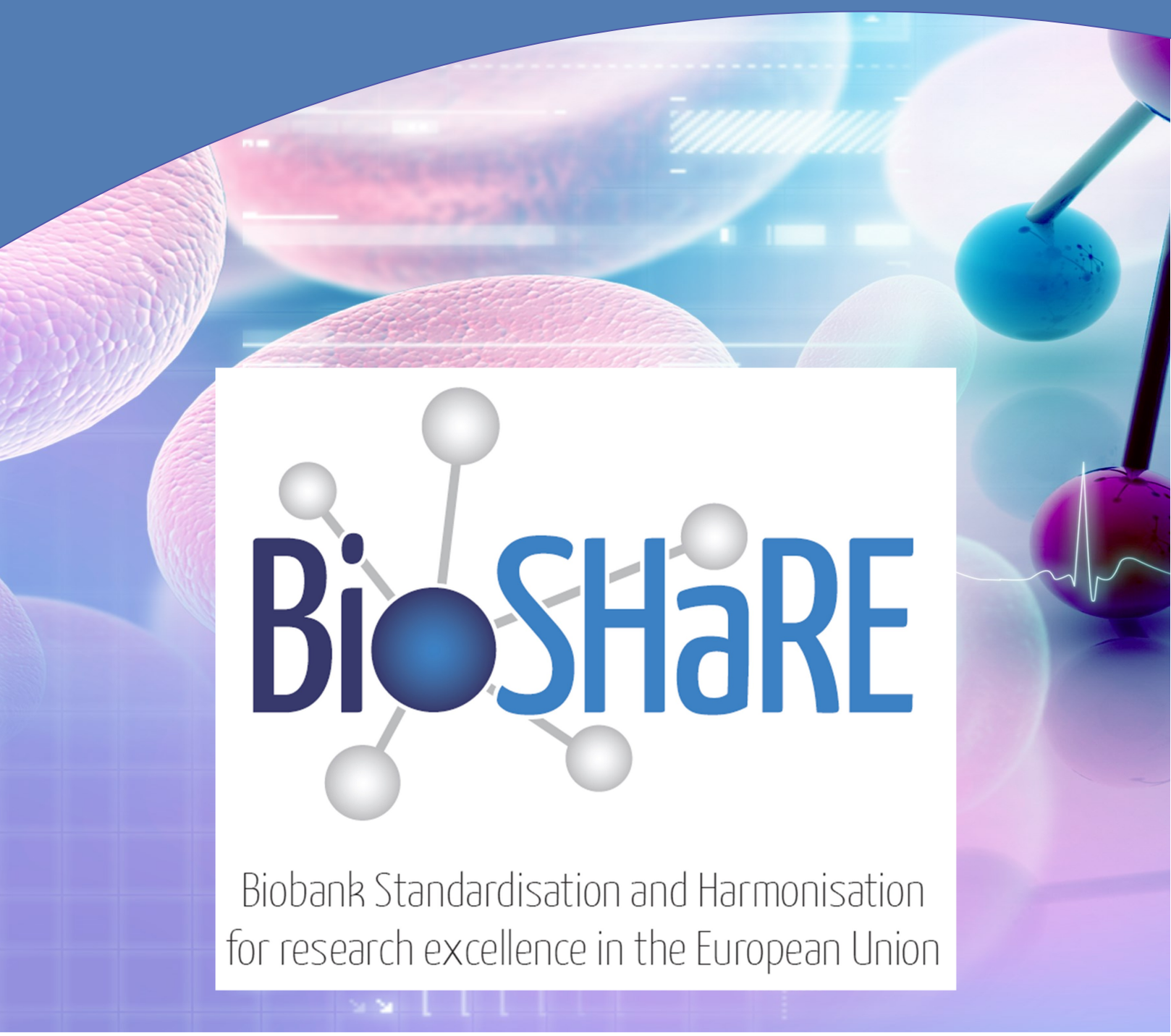
Biobank Standardisation and Harmonisation for Research Excellence in the European Union

FINAL PUBLISHABLE SUMMARY



BiOSHaRE

Biobank Standardisation and Harmonisation for research excellence in the European Union

www.bioshare.eu

TABLE OF CONTENTS

1. Executive Summary

A large number of biobanks and cohort studies have accumulated vast amounts of data and samples. These resources represent a major capital investment and constitute an important resource for understanding the interactions among genetic make-up, modifiable and non-modifiable risk factors, the onset of disease and healthy aging. Many of these interactions require large sample sizes to unravel, and thus ask for multiple datasets to be combined. Standardising and harmonising the data across datasets is a challenging process but essential to enable individual level pooling of data. This allows the use of existing data for scientific discovery.

Biobank Standardisation and Harmonisation for Research Excellence in the European Union (BioSHaRE-EU) is a European FP7 project funded from 2010 to 2015. BioSHaRE produced tools and methods for data harmonisation and standardisation, data sharing and analysis across multiple biobanks and databases. It is an ongoing consortium of leading population-based cohort studies, with international researchers from diverse domains of biobanking science, including epidemiologists, statisticians, software developers and ELSI experts.

BioSHaRE generated three types of foreground outcome: **tools**, **data** and **knowledge**. Software **tools** are developed for cohort studies and related databases to manage, present, catalogue, secure and share their data, facilitating a wider use and re-use of their data. The same tools are useful for the scientific community to explore and use the available data in an efficient and secure manner. In addition, broadly applicable guidance and recommendations have been made for sample handling and Ethical, legal and social implications (ELSI) of biobank research and sharing of data.

Data of 13 cohorts has been harmonized and analysed in scientific studies, which resulted in harmonised datasets, harmonisation algorithms and new data that enriched the participating cohort studies.

The scientific studies performed in BioSHaRE have produced **knowledge** about how to measure and harmonise specific lifestyle and exposure risk factors and health outcomes. BioSHaRE also produced knowledge about metabolically healthy obesity and about the effect of noise and air pollution exposure on health outcomes through a series of studies on these topics.

To maximize the reach of BioSHaRE-EU within the biobanking community, and to disseminate information regarding project activities and results to the larger community, a wide range of **dissemination activities** were employed, including expert meetings, presentations, publications, newsletters, tutorials, movies and development of a public website.

Exploitation of the tools, data and knowledge produced in BioSHaRE is done in a scientific context. Even though commercial exploitation is not foreseen, the potential and actual use and application of the foreground is extensive and has high value for the stakeholders, mainly for further development and research activities.

2. Summary description of project context and objectives

2.1 Project Context

Biobank Standardisation and Harmonisation for Research Excellence in the European Union (BioSHaRE-EU) is a European FP7 project funded from 2010 to 2015. BioSHaRE produced tools and methods for data harmonisation and standardisation, data sharing and analysis across multiple biobanks and databases.

BioSHaRE demonstrated the efficiency of the tools it developed by applying them in epidemiological research projects: the Core Projects. These are the Healthy Obese Project (HOP), the Environmental Determinants of Health Project or Environmental Core Project (ECP), the Metabolomics Project, the Social Core Project, the BioSHaRE -IT Project and the Statistics Core Project. This combination of tool development and application in the project greatly improved the tools and enhanced their user-friendliness.

Development of the tools has been and continues to be a joint effort, supported by BioSHaRE in conjunction with other sources of resources and funds. Most of the software tools were in development before BioSHaRE was initiated and will continue to be developed and improved in the context of new projects and initiatives. BioSHaRE built directly on extensive precursor work in the field of harmonisation and standardisation based on the now well-established DataSHaPER (Data Schema and Harmonisation Platform for Epidemiological Research) paradigm that has been developed jointly by a number of our partner projects including P3G, PHOEBE and GENECURE. In developing standardised approaches and infrastructures for bioinformatics BioSHaRE built upon and extended the work in bioinformatics and IT that has already been undertaken by GenomEUtwin (FP5), GEN2PHEN (FP7), BBMRI (ESFRI), OBiBa and P3G.

Through close collaborations with and participation of BioSHaRE investigators in many related international projects and initiatives, the BioSHaRE tools and methods are applied, exploited, further improved, and made accessible for the scientific and biobanking communities as well as for wider society. These projects and initiatives include but are not limited to BBMRI-ERIC, BBMRI-LPC, InterConnect, Maelstrom Research, The Global Alliance for Genomics and Health, CORBEL.

BioSHaRE activities were organized through nine work packages with WP1 for project management, WP2, 5, 6 and 7 for phenotype harmonisation and standardisation - implying phenotype in the broadest sense of the word, to include biomedical phenotype, environment, and socio-economic factors, WP 3 and 4 to develop bioinformatics tools, WP8 to support partnerships/dissemination, WP9 for implementation/roll-out in exemplar projects – guided professionally by informed ELSI work across all work packages. This arrangement is illustrated in Figure 1. The Work Packages and leaders are listed in Table 1.

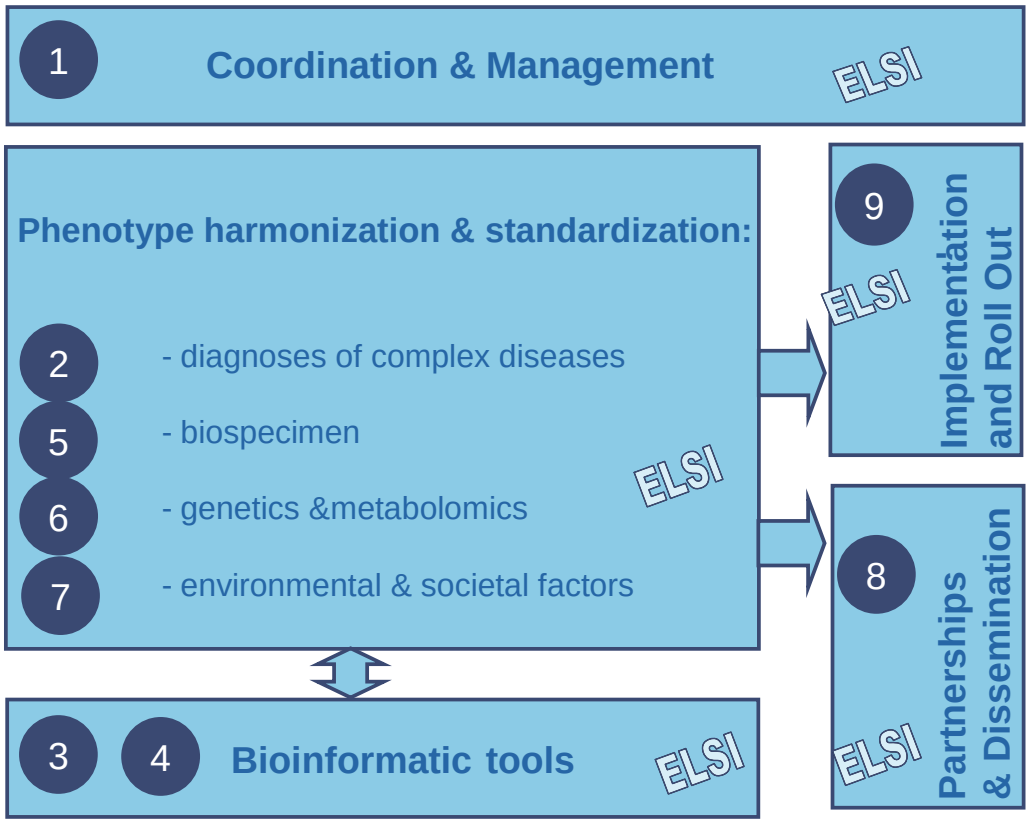


Figure 1 Conceptual overview of the BioSHaRE-EU project

Table 1 List of BioSHaRE work packages and leaders

WP1	Coordination and management	Ronald Stolk
WP2	Data repository and epidemiological/ clinical harmonization	Morris Swertz
Wp3	Epidemiology and Biostatistics for Biobank Harmonization	Paul Burton
WP4	Bioinformatics standardization/ harmonization for optimised information management	Anthony Brookes
WP5	Biospecimen harmonization/standardization	Melanie Waldenberger
WP6	Metabolomic and genetic risk factors for clustering of complex diseases	Markus Perola
WP7	Societal and environmental risk factors for complex diseases	Kristian Hveem
WP8	Strategic integration and coordination with major biobanking initiatives, partnerships and dissemination	Jennifer Harris
WP9	Implementation & Roll-out	Samuli Ripatti
	Ethical, legal and social issues	Bartha Knoppers

2.2 Overall objectives

The general objective of BioSHaRE-EU is to develop harmonized measures and standardized computing infrastructures enabling the effective pooling of data to investigate common complex diseases.

The overall objectives of BioSHaRE are:

1. To improve the assessment and classification of multivariate phenotypes associated with complex diseases, including environmental and life style exposures;
2. To enable interoperability of databases of both phenotype and genotype data;
3. To develop evidence-based standards for harmonised quality management and quality control during the collection, transport, processing, storage and retrieval of human biospecimens;
4. To develop effective strategies for optimising the correlation and integration of existing and novel data;
5. To maximize the sharing and exchange of information between population cohorts and clinical research centres/biobanks across Europe;
6. To build on pre-existing achievements (where available) and coordinate its activities with similar international efforts;
7. To consider ethical, social and legal aspects, as well as the relevance to public health.

2.3 BioSHaRE Partners

BioSHaRE has brought together 17 leading research organizations across Europe and Canada listed in Table 2. The project was led by Professor Ronald Stolk, an internationally established researcher in clinical epidemiology and director "Research Data & Biobanking" for the University Medical Center Groningen.

Table 2 List of BioSHaRE partners, respective countries and participant contacts

No	Participant	Country	Participant contact
1	University Medical Centre Groningen (UMCG)	The Netherlands	Ronald Stolk – coordinator Morris Swertz, WP2 leader
2	University of Leicester (ULEIC)	United Kingdom	Tony Brookes, WP4 leader
3	Norwegian Institute of Public Health (NIPH)	Norway	Jennifer Harris, WP8 leader
4	University of Helsinki, Institute for Molecular Medicine Finland (FIMM)	Finland	Markus Perola, WP6 leader Samuli Ripatti, WP9 leader
5	Helmholtz Zentrum München (HMGU)	Germany	Melanie Waldenberger, WP5 leader

6	Norwegian University of Science and Technology (NTNU)	Norway	Kristian Hveem, WP7 leader
7	Karolinska Institutet (KI)	Sweden	Jan-Eric Litton
8	Institut National de la Sante et de la Recherche Médicale (INSERM)	France	Anne Cambon-Thomson
9	University of Manchester (UNIMAN)	United Kingdom	Martin Yuille
10	Legal Pathway (LP)	The Netherlands	Jasper Bovenberg
11	McGill University (McGill)	Canada	Bartha Maria Knoppers
12	Medical University of Graz (MedUG)	Austria	Kurt Zatloukal
13	Public Population Project in Genomics (P3G)	Canada	Isabel Fortier
14	Ontario Institute For Cancer Research (OICR)	Canada	Vincent Ferretti
15	University of Oxford (OXF)	United Kingdom	Jane Kaye
16	Imperial College London (ICL)	United Kingdom	Paul Elliott
17	University of Bristol (UB)	United Kingdom	Paul Burton, WP3 leader

3. Main S&T results/foregrounds

3.1 Tools

In order to enable harmonisation and standardisation of data in biobanks and to facilitate data sharing and pooling across multiple biobanks, BioSHaRE developed tools and methods to assist database owners and researchers covering all processes from data discovery and search to analysis and publication.

Specifically, BioSHaRE offers tools and methods for:

1. Data description, presentation and search;
2. Data harmonisation across databases;
3. Data analysis across databases;
4. Contributor recognition;
5. Standardisation of sample handling;
6. Ethical, Legal and Social Implications (ELSI).

3.1.1 Data description, presentation and search

Biobanks and other research databases share common characteristics and have similar requirements for optimal use. The content they house should be readily identifiable to researchers and searching and discovery of data should be easily performed. Although some of the BioSHaRE data description, presentation and search tools were developed for genomic databases or biobanks/ cohort studies, the majority of our tools can be applied to a wide variety of research applications, and are complementary in application and utility. This broad application and the fact that they are open access give our tools value to researchers beyond these initial target users.

Tool	Description	Keywords
Café Variome	Platform for searching genomic data and meta-data	Genotype-phenotype, data discovery, data sharing, software, rare disease, matchmaking, biobanking, query-by-method APOI
OmicsConnect	Presentation of and access to different types of genomics data	Genomics, eDAS server, Dalliance browser, authentication
Mica	Create web portals for individual epidemiological studies or for study consortia	Study catalogue, variable catalogue, web portal, data presentation, data access, data search
MOLGENIS/ Observ-EMX	Portal for management, exploration, integration and analysis of scientific data with the focus on genomics and biobanking	Biobanks, genomics, data integration, data annotation, catalogue, genome browser

Café Variome

Description and purpose

Café Variome is a highly flexible data discovery platform suitable for use with genomic data and/or phenotype data in settings such as diagnostic networks, disease consortia, biobanks and research communities. It enables users to search for the existence rather than the substance of datasets, and as part of this offers a complete suite of data discovery capabilities, focused on the data rather than metadata. Following data discovery, the system also facilitates controlled data sharing.

'Café Variome Central' aims to consolidate all publicly available genetic variants into one discovery portal through which to announce, discover and acquire a comprehensive listing of observed neutral and disease-causing gene variants. It employs publicly available web services to gather and make searchable a set of pointers to records of interest, to help users discover the existence of variant data and direct them to the original data sources where the data may be examined in full.

Designed for

- Database owners – single or in networks, collaborations: to make their content more effectively discoverable and optionally shareable
- Biobanks - to make their biosamples, subjects and datasets better advertised and thereby accessible

- Researchers, clinicians - individual or in networks, collaborations: to make their biosamples, subjects and datasets better advertised and thereby accessible.

Use

- Genotype-phenotype data discovery and sharing
- Cohort subject / Patient discovery
- Operate as a standalone tool and/or in federated and/or in hub and spoke arrangements
- Ability to edit, add and remove any data field or attribute of interest
- Full support for local or standard ontologies
- A simple Google-like search box and a powerful query builder interface
- Rich administration interfaces
- Report of matched record counts are reported

Status and access

The software is in production as version 1.0 software, available presently for collaborative applications: http://www.cafevariome.org/

Café Variome can be installed stand-alone, or federated to allow searching across in-stances while the data remains at the source.

Café Variome requires the following components to be installed on the host server: Apache webserver, PHP, MySQL.

For Windows users the following all-in-one WAMP solutions are available:

http://bitnami.com/stack/wamp

http://www.apachefriends.org/en/xampp.html

We offer a fully hosted service to collaborators where we install a private copy of the Café Variome to allow users to trial the software or use in full production. The user has the option, at any stage, to transfer the platform to his own server.

Developed by

Café Variome is developed and hosted by ULEIC with substantial technical contribution by UMCG. Ethics and privacy considerations have been informed by P3G and other ELSI experts. Current collaborations with external partners include PhenoSystems SA, Belgium. Café Variome is funded by GEN2PHEN, BioSHaRE and by the IMI projects EMIF and EPAD.

Current applications

Café Variome is used in two IMI projects and is being used for diagnostic lab data sharing in the Netherlands, and considered for the same role in Sweden and Denmark. It is part of the GA4GH MatchMaker Exchange project. Commercial diagnostic lab software has been connected to the tool. International rare disease networks are using or testing the software to support their work.

Contact

Professor Anthony Brookes
University of Leicester, UK
ajb97@leicester.ac.uk

OmicsConnect

Description and purpose

OmicsConnect, underpinned by an 'extended DAS' (eDAS) protocol for data transfer, enables genomics data of many types to be fed into a genome browser tool from diverse sources, for collaborative visualisation. Each data provider controls which aspects of their data are seen in what way by which other users.

DAS is an Extensible Markup Language (XML) communication protocol that allows a single client (e.g. a genome browser) to integrate information from multiple DAS servers dispersed around the world to present a unified view of data. The eDAS system brings many new advantages; the data are controlled by the content providers and can be modified, restricted and updated as required and the data are shared in a way that makes it easy for the end user to get information about specific regions, genes or markers without having to download and process entire datasets.

Designed for

Database owners: presentation of and access to genomic data

Researcher: explore and mine complex genomics data

Use

- Local or remote solution for viewing and sharing data
- Customised version of the Dalliance Genome browser
- Allow public and private sharing of 'Omics' data by authentication
- Customize how data appears (via style sheets)
- Easy to setup and use
- User accounts and permissions to control the flow and view of data

Status and access

The latest version of OmicsConnect is available for use under standard terms of academic collaboration: http://omicsconnect.org.

The tool is currently being improved for better adaptability and faster performance.

No special infrastructure or facilities are required. One physical computer/server or virtual private server (VPS). No special platform or license is required. OmicsConnect can run in any up-to-date Linux distribution. 4 CPU's and 8GB RAM are recommended. Disk requirements will depend on the size of studies uploaded to OmicsConnect. Very basic knowledge of Python and server management skills are required to setup the OmicsConnect system

Developed by

OmicsConnect is developed and hosted by ULEIC. UMCG collaborated on the design and testing of the system.

Current applications

OmicsConnect has been tested and is customized for groups at ULEIC and UMCG.

OmicsConnect is currently built into the biobanking informatics infrastructure of the Molecular Medicine Research Center (MMRC) Biobank within the University of Cyprus, and the Cyprus Institute of Neurology and Genetics.

Contact

Professor Anthony Brookes

University of Leicester, UK

ajb97@leicester.ac.uk

Mica

Description and purpose

Mica is a software application developed to create web portals for individual epidemiological studies or for study consortia. Features supported by Mica include a standardised study catalogue, study-specific and harmonised variable data dictionary browsers, online data access request forms, and communication tools (e.g. forums, events, news).
When used in conjunction with the Opal software, Mica also allows authenticated users (i.e. with username and password) to perform distributed queries on the content of study databases hosted on remote servers, and retrieve summary statistics of that content.

Designed for

Database owner - biobank, other epidemiological study, consortium: to present and give access to data, to create a web portal, disseminate information about a network of studies or about individual studies
Researchers - individual or in consortium: to search and query the data

Use

- Create a website for an individual study or a consortium
- Create a study catalogue or registry
- Create a searchable data catalogue documenting data collected by individual studies or networks of studies
- Enable structured forms and workflows for data access management

Status and access

Mica is a Java-based, cross-platform, client-server application and comes along with the following two clients: the administrators' user interface and a content management system (Drupal) used to render the catalogue content on the study or consortium.
Mica is freely available for download at www.obiba.org and is provided under the GPL3 open source license.
All study and/or consortia website, data portal, or data access platform developed using the Mica software must exhibit the Mica logo and version number in the footer of its home page.
In addition, the Mica logo must link to the Maelstrom Research website at www.maelstrom-research.org.
Further, when appropriate, Mica should be mentioned in manuscripts, presentations, or other works made public and include a web link to the Maelstrom Research website (www.maelstrom-research.org).

Developed by

Mica is part of the Maelstrom Research suite of tools. Mica development is supported by BioSHaRE, Québec's Ministère de l'Économie, Innovation et Exportation, the Canadian Partnership Against Cancer, and the National Institutes of Health funded Integrative Analysis of Longitudinal Studies of Aging (IALSA) project.

Current applications

Mica is used in BioSHaRE to catalogue key characteristics of the participating biobanks, allow search and retrieve summary statistics of harmonised databases, and to disseminate BioSHaRE

activities (public website). Mica is used by multiple cohort studies and projects including CLSA, CPTP, IALSA, BBMRI-LPC, the International Network of Twin Registries, and InterConnect.

Contact

Dr. Vincent Ferretti
Ontario Institute for Cancer Research, Canada
vincent.ferretti@oicr.on.ca

MOLGENIS / Observ-EMX

Description and purpose

MOLGENIS is a portal for the management, exploration, integration and analysis of scientific data, with a focus on genomics and biobanking. In BioSHaRE the platform is adapted by moving from generation-time to run-time configuration, allowing the users to upload complete data structures (EMX entity model extensions), including a reference data model (Observ-OM), and including spread sheet and VCF upload format, data explorer, genome browser, REST/R-project APIs, visualization and annotation tools.

Observ-EMX is a data model based on the Observe-OM/TAB data model co-developed in BioSHaRE. It is a flexible data model that can be implemented by research consortia and biobanks to cope with different and changing data types generated by new and existing techniques and technologies, including genomic variation, Next Generation Sequencing (NGS), exome sequencing, GWAS, Phenotypic observations, lab processes, sample tracking and model organism data.

Observ-EMX has been implemented in the MOLGENIS toolkit and used in the OmicsConnect software package.

Designed for

Database owners – biobanks, individual research studies and study consortia, multi-omics and genetics studies, patient registries: to manage, annotate, present and share their data

Researchers - consortia: for data harmonisation & integration (BiobankConnect)

Researchers – individual: data search and analysis

Use

- Data management: modelling & capture, scriptable data management using R-project, R or python
- Data upload using spread sheets, TSV, CSV, VCF
- Data integration via genome browser
- Data annotation using public database/tools like 1KG, GoNL, CADD, etc
- Data search: flexible data explorer to display and filter data, large search indices
- Data access: set users, groups, permissions
- Data harmonisation & pooling using SORTA and BiobankConnect
- Customizable menu structure

Status and access

MOLGENIS software is production-ready and freely available for download as open source under license LGPLv3. See http://molgenis.org for general descriptions.

The software is built on industry standards like Maven, MySQL, SpringMVC, GitHub, Bootstrap,

Java 8 and ElasticSearch.
Interested users can download the code from http://github.com/molgenis/molgenis and compile themselves. Alternatively, users can download a WAR deploy file from public maven repositories, e.g. http://mvnrepository.com/artifact/org.molgenis/molgenis-app. Finally, UMC Groningen also provides hosting services for users who want to rent MOLGENIS as a service.
Installation instructions are available at http://github.com/molgenis/molgenis. Required software is Java, Tomcat, and Mysql. MOLGENIS typically runs on a standard Linux web server but it can also be run on Windows or Mac. The software is standard and can be deployed by most system administrators or Java developers.

Developed by

MOLGENIS is an international open source project coordinated by UMCG. BioSHaRE has greatly advanced the development of the core modular database, and specifically enabled the development of BiobankConnect and SORTA, which has been complemented by funding from of BBMRI-NL, BioMedBridges, RD-connect and other projects.

Current applications

MOLGENIS is currently used in more than 25 installations, including the LifeLines data request catalogue, BBMRI-NL national biobank catalogue, several rare disease patient registries and multi-omics projects.

Contact

Dr. Morris Swertz
University Medical Center Groningen, Netherlands
m.a.swertz@gmail.com

3.1.2 Data Harmonisation Across Databases

"Standardisation and harmonisation describe a corpus of practices intended to allow interoperability of data and sample collections along a continuum from absolutely uniform collection to unfettered local variation in collection. Standardisation includes practices (standards) for prospectively implementing uniform processes for collection, storage and transformation of samples and data. Harmonisation includes practices which enable the pooling of data from multiple cohorts/biobanks at a level of precision that is scientifically adequate, yet accommodates the existing heterogeneity of those collections. Harmonisation also includes practices whereby prospective agreement is made to collect data in such a way as to directly enable pooled analysis" (BioSHaRE Consensus position on the distinction between standardisation and harmonisation, 2012).
A variety of tools and methods are developed in BioSHaRE for retrospective and prospective harmonisation, facilitating full valorisation of the database contents for the scientific community.

Tool	Description	Keywords
BiobankConnect	Ontologies for variables classification index	Biobanks, data mapping, data harmonisation, data integration, data search
DataSchema	Template for the retrospective harmonisation	Data harmonisation, variable template,

	process by defining the common format measures to be derived using study data	common format
EnviroSHaPER	Noise modelling tool	Noise exposure, geographic information systems (GIS), CNOSSOS-EU, LAeq, road traffic
Opal	Management of study data enabling data harmonisation and data integration across biobanks/ cohort studies	Data storage, data management, data harmonisation, DataSHIELD
SORTA	System for Ontology-based Re-coding and Technical Annotation of biomedical phenotype data	Data harmonisation, data annotation, data recoding, ontology
Vortext/Spá	System for literature based discovery	Text mining, PDFs, literature based discovery, machine learning

BiobankConnect

Description and purpose

To effectively pool data across biobanks, researchers must search thousands of available data items and harmonise differences in terminology, data collection, and structure. To minimize these arduous and time-consuming tasks, we have developed BiobankConnect, a tool which catalogues available data items per biobank, and then semi-automatically searches for desired data items. BiobankConnect provides an easy user interface to significantly speed-up the harmonisation of biobanks by automating a considerable part of the work. This is achieved through
1) annotation of the desired data items with ontology terms using the BioPortal ontology service;
2) automatic expansion of the semantics of these items by adding synonyms and subclass information using OntoCAT;
3) automatic search of all available items for these expanded terms using Lucene lexical matching;
4) review of candidate items sorted by matching score, from which users can select the final mappings.

Designed for

Researchers - individuals, single research studies and study consortia: to find matching variables across databases/ biobanks data dictionaries

Use

Data mapping: find the mappings for research variables across biobanks

Status and access

BiobankConnect is freely available for download as a MOLGENIS open source application at http://www.github.com/molgenis.
A new version of BiobankConnect is currently in development.

Developed by

BiobankConnect is developed by the UMCG, solely funded by BioSHaRE, and incorporated in MOLGENIS. Maelstrom Research provided the data for the validation of this tool.

Current applications

BiobankConnect was evaluated using human curated matches from BioSHaRE, searching for 32 desired data elements in 7461 available elements from six biobanks.

Contact

Dr. Morris Swertz
University Medical Center Groningen, Netherlands
m.a.swertz@gmail.com

DataSchema

Description and purpose

DataSchemas incorporate and document sets of core variables targeted for harmonisation. They act as templates for the retrospective harmonisation process by defining the common format measures to be derived using data from participating studies. In order to allow multiple studies to participate in a collaborative endeavour while ensuring validity of the scientific output, the development of a DataSchema requires a balance between uniformity (e.g. exact same question wording and data collection procedures) and acceptance of certain level of heterogeneity across studies (e.g. slightly different wording or procedures).

Designed for

Investigators and consortia representatives aiming to harmonise data across a group of studies

Use

- Document the harmonised definition and format of a set of variables
- Generate a set of common format variables

Status and access

Opal and Mica software have both been developed by Maelstrom Research to facilitate data harmonisation using DataSchemas.
Once a DataSchema is defined, the Opal software application is used to manage study-specific and harmonised datasets, as well as to develop and implement data processing algorithms.
Moreover, a view of the DataSchema and harmonisation potential across studies is typically made available through a Mica-powered website.

Developed by

The development of Dataschemas is one of the key activities/ services of the Maelstrom Research data harmonisation methodology. All users interested in applying this harmonisation approach are encouraged to contact Maelstrom Research prior to initiating the harmonisation process.

Current applications

In BioSHaRE, two DataSchemas were developed: the Healthy Obese Project DataSchema and the Environmental determinants of health DataSchema. In order to answer a range of different research questions in these projects, researchers involved selected and defined 100 variables and 76 variables, respectively, that were included in the DataSchemas.

Contact

Dr. Isabel Fortier
Research Institute of the McGill University Health Centre, Canada
isabel.fortier@mail.mcgill.ca

EnviroSHaPER - CNOSSOS-EU Road Noise Model

Description and purpose

The CNOSSOS-EU (Common Noise Assessment Methods in Europe) model provides a common noise modelling framework for Europe, enabling harmonisation and comparison of noise from road, rail, industrial and aircraft sources for different regions across Europe.
The model was adjusted for use in the in BioSHaRE Environmental determinants of health project to handle low resolution data sets that are widely available with European-wide coverage. This allowed for harmonised and comparable measures of road-traffic noise exposure to be assigned to participants across BioSHaRE cohorts.
The EnviroSHaPER comprises an open source, free software tool, with a user friendly interface to enable users to more easily apply this complex model and assign noise exposures to their cohort/biobank data.

Designed for

Database owner - cohort/biobank: estimate noise exposure for individual participants
Researcher – individual or consortium: obtain harmonised noise exposures across databases/ studies

Use

- Provide road traffic noise exposure predictions at point locations (usually 1m in front of building facades)
- Harmonisation of noise exposure estimates across datasets

Status and access

EnviroSHaPER is currently available as a beta version and available under the conditions of the Apache License v2 (www.apache.org/licenses/LICENSE-2.0). The tool is available on request from: www.sahsu.org/content/data-download.
Windows systems and .NET framework 4.5 is required. The open-source database PostgreSQL with the PostGIS extension must be pre-installed. Some knowledge of GIS and access to GIS desktop software may be beneficial depending on the user's data requirements.
Actual use of the EnviroSHaPER and CNOSSOS-EU model requires geocoding of the participant locations and the availability of GIS layers as input (road geography, traffic flows, and land cover). European land cover data is freely available from the European Environment Agency via CORINE; World-wide road geography is freely available from OpenStreetMap; Traffic count data is available via relevant national agencies.

Developed by

The CNOSSOS-EU model has been developed by the EU JRC-IHCP (Joint Research Centre - Institute for Health and Consumer Protection).
The BioSHaRE noise model was developed at ICL based on the guidelines outlined in the CNOSSOS-EU framework.

Current applications

The noise model has been used to assign road-traffic noise exposure estimates to the EPIC Oxford, UK Biobank, HUNT and LifeLines cohorts within BioSHaRE's Environmental determinants of health project.

Contact

Dr. Susan Hodgson
Imperial College London, UK
susan.hodgson@imperial.ac.uk

Opal

Description and purpose

Opal is a software application to manage study data, and includes a feature enabling data harmonisation and data integration across studies. As such, Opal supports the development and implementation of processing algorithms required to transform study-specific data into a common harmonised format. Moreover, when connected to a Mica web interface, Opal allows users to seamlessly and securely search distributed datasets across several Opal instances.

Designed for

Database owners - Individual research studies/ biobanks and research study consortia: to manage and present data, to harmonise data, to give access to data in a federated database setting.

Use

- Data storage and management
- Data harmonisation and curation through data processing algorithms
- Data search and query in study data and data dictionaries
- Data analysis: generate descriptive statistics and produce reports

Status and access

Opal is freely available for download at www.obiba.org and is provided under the GPL3 open source licence. All studies or networks of studies using the Opal software for data storage, data management or data harmonisation must mention Opal in manuscripts, presentations, or other works made public and include a web link to the Maelstrom Research website (www.maelstrom-research.org). When using Opal to implement data processing algorithms to harmonise or clean data, basic knowledge of the JavaScript programming language is required.
Opal is a Java-based application, so it should run on any platform for which a Java Virtual Machine is provided. Detailed installation and configuration instructions are available at www.obiba.org.

Developed by

Opal development was initiated by OICR and is part of the Maelstrom Research suite of tools. Opal development is supported by BioSHaRE, Québec's Ministère de l'Économie, Innovation et Exportation, the Canadian Partnership Against Cancer, and the National Institutes of Health funded Integrative Analysis of Longitudinal Studies of Aging (IALSA) project.

Current applications

Opal software has been used in the BioSHaRE Healthy Obese and Environmental determinants of health projects to store the data used for combined analyses, develop and implement processing algorithms transforming study data into format, and create a federated infrastructure that allows researchers to jointly analyse harmonised data.

Contact

Dr. Vincent Ferretti
Ontario Institute for Cancer Research, Canada
vincent.ferretti@oicr.on.ca

SORTA - System for Ontology-based Re-coding and Technical Annotation of biomedical phenotype data

Description and purpose

There is an urgent need to standardise the semantics of biomedical data values, such as phenotype codes, to enable comparative and integrative analyses. However, it is unlikely that all studies will use the same data collection protocols because of their different contexts and purposes. As a result, retrospective standardisation is often required, which involves matching of original (unstructured or locally coded) data to widely shared code or ontology systems such as SNOMED, ICD-10, and HPO. This data curation process is usually a time-consuming process performed by a human expert.

To help mechanize this process, we have developed SORTA, a computer-aided system for rapid encoding of free text or locally coded values to a formal coding systems or ontology. SORTA matches target coding systems (uploaded in Excel or ontology format (OWL/OBO)) to original data values (uploaded in semi-colon delimited format) and semi-automatically shortlists candidate codes for each data value algorithms.

Designed for

Database owners and researchers: to code free text or locally coded values to formal coding systems or ontology.

Use

- Recoding/coding free text or locally coded values with standard terminologies

Status and access

The tool is in production and is freely available as an online service at:
https://molgenis19.target.rug.nl/
SORTA is part of open source project MOLGENIS, which can be found at:
https://github.com/ChaoPang/molgenis/releases

Developed by

SORTA is developed by the UMCG, solely funded by BioSHaRE, and incorporated in MOLGENIS.

Current applications

SORTA has been used to recode physical activity data within the LifeLines cohort study and will be used to recode LifeLines job-related data.

Contact

Dr. Morris Swertz
University Medical Center Groningen, Netherlands
m.a.swertz@gmail.com

Vortext/Spá

Description and purpose

Unstructured PDF documents remain the main vehicle for dissemination of scientific findings. Those interested in gathering and assimilating data must therefore manually peruse published articles and extract from these the elements of interest.

Machine learning provides a potential means of mitigating this burden by automating extraction. We present a web-based tool called Vortext/Spá that accepts an article as input and provides an automatically visually annotated rendering of this article as output. More generally, it provides a framework for visualizing predictions, both at the document and sentence level, for full-text PDFs.

Designed for

Researchers – individual or consortia: to facilitate evidence synthesis from literature

Use

- Visualize predictions from Machine Learning pipelines on full text PDF
- Extract relevant sentences and meta-data based on machine learning
- Manage literature in a web based system

Status and access

Vortext/Spá is in production and available online. More information can be found on http://vortext.systems.
Open Source code (GPLv3) for Vortext/Spá is available on GitHub: https://github.com/vortext.
Vortext/Spá is built upon a combination of Clojure, Python, NodeJS and R languages. The software requires extensive knowledge of software development methodologies and Linux servers to operate. Development of new and novel annotation pipeline requires knowledge of Natural Language Processing and Machine Learning.

Developed by

Vortext/Spá is a web-based tool developed by the UMCG Genomics Coordination Centre in collaboration with Vortext Systems (international IT consulting).
The tool was originally named Spá (Kuiper at al., 2014), but it was renamed when Vortext Systems was established as a consulting entity in collaboration with US and UK researchers to facilitate the ongoing development of novel methods and software.

Current applications

A pilot study of Vortext/Spá was conducted to extract elements of interest for Individual Participant Data (IPD) meta-analysis, building on existing work for identifying risk of bias in randomized clinical trials. Methods are being developed for automatically identifying HGVS entities from GWAS and rare disease literature.

Contact

Dr. Morris Swertz
University Medical Center Groningen, Netherlands
m.a.swertz@gmail.com

3.1.3 Data analysis across databases

Conventional methods to physically pool individual participant data can raise ethical, legal and regulatory questions, and introduce data governance issues. These can become particularly complex when researchers are operating in a global context. The primary aim of BioSHaRE is to facilitate data analyses across multiple databases: our analytic solutions are tailored specifically to multiple cohort studies, and are designed in such a way as to minimize these issues while

maintaining data security and increasing scientific power.
The tools for data analysis developed in BioSHaRE are designed for biobanks and cohort studies to accurately estimate sample size and power, and to allow for more flexible and secure data analysis between cohort studies.

Tool	Description	Keywords
DataSHIELD	Data Aggregation Through Anonymous Summary-statistics from Harmonised Individual levEL Databases. Enables central federated analyses on multiple datasets without physical data pooling	Data pooling, data analysis, federated analysis, sensitive data, governance, intellectual property
ESPRESSO	Estimating Sample-size and Power in R by Exploring Simulated Study Outcomes	Statistical power, sample size, association studies, measurement errors

DataSHIELD - Data Aggregation Through Anonymous Summary-statistics from Harmonised Individual-levEL Databases

Description and purpose

DataSHIELD was born of the requirement in the biomedical and social sciences to co-analyse individual patient data (micro data) from different sources, without disclosing identity or sensitive information. Under DataSHIELD, raw data never leave the data provider and no micro data or disclosive information can be seen by the researcher. The analysis is taken to the data – not the data to the analysis. It provides a flexible, modular, open-source solution ideally placed to serve a broad user and development community and to circumvent barriers related to ethical-legal restrictions, intellectual property and physical size of the data as a limiting factor.

Designed for

Database owners - biobanks, other studies: to allow analyses of individual level data while respecting ethical, legal and IP issues,
Researchers - consortia: to share and analyse data in a consortium or between multiple studies without actual data pooling.

Use

Applied to a single site

- Create a “secure data enclave” in which data can be analysed but not seen, to collaborate in consortium-based analyses without revealing source data.
- Provide a “secure data enclave” to hold potentially sensitive data, created using record linkage, thereby making them accessible for secondary analysis.
- Provide a post-publication platform that enables the data underpinning all of the analyses in a paper to be made publically available for extended analysis (including confirmation) without data being released into the public domain.
- Provide a publically accessible web-portal that enables researchers to undertake simple preliminary univariate and bivariate analysis of data before application for full access to those data.

Applied to multiple sites

- Co-analysis of individual-level data or study level meta-analysis from multiple studies

Status and access

All DataSHIELD packages are open source and in beta-testing. New packages, methodology and functions are also under development and will be tested and released into packages in due course. Full information and access to DataSHIELD is available at http://www.datashield.ac.uk including access to the DataSHIELD wiki (http://www.datashield.ac.uk/wiki) that contains all technical documentation and tutorials to install and use DataSHIELD.
DataSHIELD Client Software:
Runs in linux, Mac and Windows
Requires R and/or R Studio and the DataSHIELD client packages
Requires basic knowledge of epidemiological analyses / medical statistics methodology
Requires experience analysing data in R

Developed by

The following partners are involved in the ongoing development of DataSHIELD:
BioSHaRE partners UB (Data to Knowledge Research Group), OICR (including Obiba), McGill (including Maelstrom Research), NIPH, UMCG, ULEIC and external partner Eindhoven University of Technology, Netherlands.

Current applications

DataSHIELD is used for secure data analyses in BioSHaRE within the Healthy Obese Project and Environmental determinants of health projects. DataSHIELD will be used in InterConnect and other recently initiated projects.

Contact

Professor Paul Burton
University of Bristol, UK
Paul.Burton@bristol.ac.uk

ESPRESSO - Estimating Sample-size and Power in R by Exploring Simulated Study Outcomes

Description and purpose

Very large studies are required to provide sufficiently big sample sizes to adequately power association analyses. This can be an expensive undertaking and it is important that an accurate sample size is identified. For more realistic sample size calculation and power analysis, the impact of unmeasured aetiological determinants and the quality of measurement of both outcome and explanatory variables should be taken into account.
ESPRESSO is a tool that, unlike conventional methods that use closed-form solutions, allows for uncertainties around outcome and exposure measurements to be taken in to account in statistical power and sample size calculations.

Designed for

Databases owners - Individual research studies/ biobanks: to estimate statistical power given sample size or calculate sample size needed to answer a research question.
Scientific reviewers and funding bodies that want to verify the statistical power calculations put forward by researchers in their grant applications.

Use

- Given a set or target sample size, ESPRESSO allows one to estimate the statistical power

which can be achieved with that sample size.

- Given a target/desired statistical power, ESPRESSO allows one to calculate the sample size required to achieve that level of power.
- ESPRESSO can also be used to evaluate the effect of measurement errors on the statistical power of a study and its implication for the sample size e.g. the increase in sample size required to encompass the detrimental effect of the error(s).

Status and access

ESPRESSO is open source and is available for online calculations and downloads at http://espresso-research.org. The code is freely available at https://github.com/ESPRESSO-research.

To use the R version of the tool, the R development environment is required. R is open source and can be downloaded for free at http://cran.r-project.org/.

Developed by

ESPRESSO was developed by the Data to Knowledge (D2K) research group, initially at Department of Health Sciences, ULEIC and then subsequently at the School of Social and Community Medicine, UB. The Newcastle University School of Computing Science contributed to the development of the web interface for ESPRESSO.

Current applications

ESPRESSO has been used in two published analyses, to evaluate the impact of pre-analytic variation in analytes from the UK Biobank on the power of association studies and to assess the statistical power of the Canadian Partnership for Tomorrow project given its ultimate sample size. ESPRESSO has also been used by researchers outside of the BioSHaRE project.

Contact

Professor Paul Burton
University of Bristol, UK
Paul.Burton@bristol.ac.uk

3.1.4 Contributor recognition

Recognition of participation and contribution is not only important to individual researchers: biobanks and research subjects as well as researchers all need to be identified unambiguously for their involvement in biobank research, website content, databases, and data elements within databases. The establishment of globally unique digital ID systems is one method to achieve this goal. Through its development of two key tools, BioSHaRE has contributed to the global movement for contributor recognition of researchers and bioresources.

Tool	Description	Keywords
BRIF	Bioresource Research Impact Factor	Biobanks, bioresource, CoBRA, data sharing, impact factor
ORCID	Open Researcher and Contributor ID	Online identity, online identification, credentials, single-sign-on, federation, contribution, recognition

BRIF - Bioresource Research Impact Factor

Description and purpose

The BRIF is a collective international initiative to build a framework for recognising and measuring the use of bioresources for research. It targets 4 main objectives that are currently ongoing:
1) fostering the assignment of a unique and persistent identifier to the bioresource by an independent international institution or body,
2) the construction of the BRIF algorithm on the basis of a number of agreed parameters for the follow-up of the use of bioresources,
3) the modification of editorial guidelines in order to coherently integrate the citation and acknowledgement of the bioresources used in scientific articles, and
4) the assessment of incentives for bioresource access and sharing policies.
Recently, members of the journal editors subgroup published the CoBRA guideline, a standardised citation scheme specific to bioresources.

Designed for

Database owners - all type of bioresources: to measure quantitative use, valorisation success, enable traceability, acknowledge effort, show impact
Researchers: to acknowledge bioresource and track its use, impact and results
Editors: to develop guideline for acknowledging bioresources
Funding bodies: to check use, impact and results of bioresource

Use

- Recognition of the use of bioresources for research
- Measure quantitative use, valorisation success, enable traceability, acknowledge effort, show impact
- Acknowledge bioresource and track use, impact and results
- Aide in guideline development for acknowledging bioresources

Status and access

Although at this time only cohorts participating in BioSHaRE can request a BRIF number, any cohort can be acknowledged in a standardised manner by implementing the CoBRA guideline.

Developed by

In 2010 an international working group was created by members of several European projects. BioSHaRE partner INSERM is actively involved in the development and implementation of BRIF.

Current applications

The BRIF is piloted in BioSHaRE. Each cohort that participates in BioSHaRE requested a unique identifier (BRIF number) that is used throughout all BioSHaRE publications.

Contact

Dr. Anne Cambon-Thomsen
Institut National de la Santé et de la Recherche Medicale, France
anne.cambon-thomsen@univ-tlse3.fr

ORCID - Open Researcher and Contributor ID

Description and purpose

ORCID is an open, non-profit, community-driven effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers.
ORCID provides two core functions:
a registry to obtain a unique identifier and manage a record of activities, and
APIs that support system-to-system communication and authentication.

Designed for

Researchers, research funders, and organizations: to link research to individual researchers, funding sources, and organizations
Professional associations: to track research activity across multiple membership database sources
Publishers: to streamline manuscript admissions

Use

- Unambiguous identification of researchers
- Online authentication of researchers
- Contributor recognition of researcher output such as datasets, equipment, articles, media stories, citations, experiments, patents, and notebooks
- Knowledge discovery via linking of researcher outputs
- Tracking and acknowledgement of researcher output
- Streamline workflows

Status and access

ORCID makes its code available under an open source license, and will post an annual public data file under a CC0 waiver for free download at http://orcid.org/
For organizations there is a free public API or a member API.
The Public API can be used for signing in or retrieving a user's ORCID identifier, retrieve public data from a user's ORCID record and to search public ORCID registry data.
The member API can be used to link organization records to ORCID identifiers, to update ORCID records, to receive updates from ORCID, and to register their employees and students for ORCID identifiers.
Using the Public API requires a set of credentials consisting of a Client ID and a Client Secret.
ORCID Public and Member APIs are RESTful, and use OAuth 2.0, a well-established, standard protocol for user-based permissions.

Developed by

ULEIC is involved in the technical design of ORCIDs, partly by being on their Technical Work group, and also by drafting their first technical specification for their IT system. BioSHaRE together with GEN2PHEN lobbied in 2013 for a no-cost "limited membership" option and lower-cost tiers and/or no-cost waiver option for full membership, to facilitate integration by smaller, non-profit organizations expected to be light users of the service.

Current applications

Use of ORCIDs to control access to BioSHaRE information/content and BioSHaRE rights/privilege

management, for BioSHaRE partners only.
Including and/or linking ORCIDs as part of dissemination actions and reports, to facilitate external contribution tracking during and subsequent to the lifetime of BioSHaRE.

Contact

Professor Anthony Brookes
University of Leicester, UK
ajb97@leicester.ac.uk

3.1.5 Standardisation of sample handling

BioSHaRE has developed several tools to facilitate the standardisation of sample handling, such as standard operating procedures, recommendations and reports. These tools have been developed with input from extensive literature review, expert opinion, best practices of cohorts participating in BioSHaRE, and the conduct of additional scientific research within BioSHaRE.

Tool	Description
Standard Operating Procedures for the handling of liquid biosamples	Blood withdrawal Blood processing Urine withdrawal, processing and storage Shipping of liquid biosamples
Reports and recommendations for storage and analyses of data and samples	1. Evidence-based minimal standards on: Pre-analytical techniques for epigenetics Pre-analytical techniques for metabolomics Quality standards for OMIC-analysis of blood samples
	2. Temperature effects of preparing and thawing samples on different analysis techniques.
	3. Harmonisation and standardisation of inflammatory biomarkers.
	4. Trace element analysis
	5. Recommendations for utilization of omical data and/or identified patterns in disease prevention, diagnosis and treatment

SOPs for the Handling of Liquid Biosamples

Description and purpose

Standard operating procedures (SOPs) have been developed as templates to be adapted to the needs of the individual biobank or laboratory for the following procedures:

- Blood withdrawal
- Blood processing
- Urine withdrawal, processing and storage
- Shipping of liquid biosamples

These SOPs were developed based upon existing SOPs from large biobanks (KORA, UMGC, HUNT, German National Cohort) under the leadership of HMGU.

Info

The SOPS are available for download from the BioSHaRE website at www.bioshare.eu (Deliverable 5.2).

Contact

Dr. Gabriele Anton
Helmholtz Zentrum München, Germany
gabriele.anton@helmholtz-muenchen.de

Recommendations for Storage and Analyses of Data and Samples - Evidence-based minimal standards

Description and purpose

Evidence for pre-analytical procedures has been obtained from the literature and own experiments. The documents covers the areas blood sampling and blood processing as well as the important downstream applications in the omics field, namely genomics, epigenomics, transcriptomics, metabolomics and proteomics. Evidence from the literature is summarized and suggestions for harmonized quality assurance and documentation are made. Current state of the art preanalytical techniques for metabolomics and epigenetics have also been identified and summarized.

Evidence-based minimal standards are made on:

- Pre-analytical techniques for epigenetics
- Pre-analytical techniques for metabolomics
- Quality standards for omic analysis of blood samples

Info

The recommendations are available for download from the BioSHaRE website: www.bioshare.eu (Deliverable 5.1),

Contact

Dr. Gabriele Anton
Helmholtz Zentrum München, Germany
gabriele.anton@helmholtz-muenchen.de

Recommendations for Storage and Analyses of Data and Samples - Temperature effects of preparing and thawing samples on different analysis techniques

Description and purpose

The effect of up to four freeze-thaw cycles and of different pre-storage handling conditions on metabolomics parameters in serum was assessed by a targeted metabolomics approach. Based on changes in lysophosphatidylcholines phosphatidylcholines and amino acid concentrations, we present a measure that is able to distinguish between ‘good’ and ‘bad’ pre-analytical sample quality in our study.

Info

The results are described in Deliverable 5.3 at www.bioshare.eu and published in Anton et al. 2015.

Contact

Dr. Gabriele Anton
Helmholtz Zentrum München, Germany
gabriele.anton@helmholtz-muenchen.de

Recommendations for Storage and Analyses of Data and Samples - Harmonization and standardisation of inflammatory biomarkers

Description and purpose

This project had the following aims:
1. To study the potential effects of complex diseases on stability and standardisation of biosamples.
2. To assess differences for fresh and frozen samples for different analytic techniques.
3. To perform analysis in different sample sources (serum, EDTA-plasma, heparin-plasma) for quality / interchangeability.
The project yielded important results regarding the effect of disease state, stability of inflammatory markers, and comparability between assays used by the different biobanks. It also revealed pitfalls and risk factors in doing such a complex project with valuable material.
The analyses were performed on samples from 3 large biobanks KORA, LifeLines and HUNT, under the lead of UMCG.

Info

The results of this project will be made publicly available in 2015. (Deliverable 5.4)

Contact

Professor Bruce Wolffenbuttel
University Medical Center Groningen, The Netherlands
bwo@umcg.nl

Recommendations for Storage and Analyses of Data and Samples - Trace element analysis

Description and purpose

This report presents and discusses important principles of collecting, analysing and reporting on blood samples used for trace element analysis using high resolution inductively coupled mass spectrometry (HR-ICP-MS). It contains a SOP for Trace element analyses in whole blood by ICP-MS specified for the HUNT biobank in Norway.

Info

This report is available for download from the BioSHaRE website at www.bioshare.eu (Milestone 50)

Contact

Professor Kristian Hveem
Norwegian University of Science and Technology, Norway
kristian.hveem@ntnu.no

Recommendations for Storage and Analyses of Data and Samples - Recommendations for utilization of omical data and/or identified patterns in disease prevention, diagnosis and

treatment
Description and purpose
This white paper is intended for people with only little prior experience of omical data analysis and provides advice regarding general aspects in the utilization of omical data including phenotypic harmonization and data quality, as well as more detailed guidance about the use of genome-wide genotypic data, transcriptomics, metabolomics and epigenomics.
Info
This paper is available for download from the BioSHaRE website at www.bioshare.eu (Deliverable 6.2).
Contact
Professor Markus Perola Institute for Molecular Medicine, Finland markus.perola@thl.fi

3.1.6 Ethical, Legal and Social Implications (ELSI)

BioSHaRE has assembled a team of experts in ethical, legal and social implications (ELSI) to deal specifically with the ELSI issues arising in the BioSHaRE project. This expert team has paid particular attention to the international sharing of samples and data, geo-referencing and use of environmental risk and personal life-style data, and societal impact and stakeholders' engagement.
The result is a comprehensive set of guidelines to assist researchers with data use, sharing, and access within large international collaborative frameworks. Chief among these is the Framework for responsible sharing of genomic and health-related data.

Tool	Description
ELSI data and sample sharing tools	1. Legal requirements for sharing of data and samples across biobanks based in Germany, Finland, the Netherlands, Norway and in the UK
	2. Data access policies of different biobanks in Germany, Finland, the Netherlands, Norway and in the UK
	3. ELSI issues and solutions related to federated analysis using DataSHIELD
Framework for responsible sharing of genomic and health-related data	
ELSI guidance on geo-referencing and use of environmental risk and personal life-style data in European biobanks	
ECOUTER - Employing COnceptUal schema for governance, policy and Translational Engagement in Research	

ELSI tools for data and sample sharing - Legal requirements for sharing of data and samples across biobanks based in Germany, Finland, the Netherlands, Norway and in the UK
Description and purpose

Analysis of existing national and international laws that govern the cross-border transfer of human tissue and associated data for research purposes. It focuses on the laws in six jurisdictions – Germany, the Netherlands, Norway, Sweden, Finland and the United Kingdom – in accordance with the location of the biobanks that participate in the BioSHaRE project.

Info

This analysis was led by OXF and the results are described in BioSHaRE Deliverables 2.5 and 9.3. These are available for download at www.bioshare.eu.

Contact

Professor Jane Kaye
University of Oxford, UK
jane.kaye@law.ox.ac.uk

ELSI tools for data and sample sharing - Data access policies of different biobanks in Germany, Finland, the Netherlands, Norway and in the UK

Description and purpose

The overview is conducted on the basis of the information publicly available on the relevant websites, and specified for the various stages required for access, that is i) registration, ii) application (a. preliminary application, b. main application); iii) submission of preliminary approval and material transfer agreement.
Biobanks included: BioSHaRE biobanks: Finrisk, UK Biobank, KORA, HUNT, LifeLines, EPIC cohort, Morgam project, the German National Cohort, and Health 2000, WTCCC1, Generation Scotland, Bristol's Alspac, MalariaGen

Info

The analysis was led by OXF. These results are available for download at www.bioshare.eu (Deliverable 2.4).

Contact

Professor Jane Kaye
University of Oxford, UK
jane.kaye@law.ox.ac.uk

ELSI tools for data and sample sharing - ELSI issues and solutions related to federated analysis using DataSHIELD

Description and purpose

An ethico-legal analysis was conducted at ULEIC that examined each step of the DataSHIELD process from the perspective of UK case law, regulations, and guidance. In order to facilitate a similar analysis for other countries/ jurisdictions, a 'DataSHIELD Legal Analysis Template' is being made.
In a follow-up analysis by NIPH, ethics-related data-sharing concerns of Institutional Review Boards, ethics experts, international research consortia and research participants were identified through a literature search and systematically examined at a multidisciplinary workshop to determine whether DataSHIELD proposes mechanisms which can address these concerns.

Info

The results have been published as Wallace et al, 2014, and Budin-Ljøsne et al, 2014.

Contact

Dr. Susan Wallace
University of Leicester, UK
sew40@leicester.ac.uk

Framework for responsible sharing of genomic and health-related data

Description and purpose

The Framework for responsible sharing of genomic and health-related data aims to accelerate progress in human health by helping to establish a common framework of harmonized approaches to enable effective and responsible sharing of genomic and clinical data, and to catalyse data sharing projects that drive and demonstrate the value of data sharing.

The Framework is centered on four "Foundational Principles": 1) Respect Individuals, Families and Communities; 2) Advance Research and Scientific Knowledge; 3) Promote Health, Wellbeing and the Fair Distribution of Benefits; and 4) Foster Trust, Integrity and Reciprocity. These Foundational Principles are further elaborated by 10 "Core Elements": transparency; accountability; engagement; data quality and security; privacy, data protection and confidentiality; risk-benefit analysis; recognition and attribution; sustainability; education and training; and accessibility and dissemination.

Designed for

Researchers: data generation, sharing and use.
Research participants: consent to data sharing.
Publishers, funding agencies: determine data sharing plans.
Authorities, institutes: develop policy for data release.
Industry: balance commercial interests, respect for data donors, and the benefits of data sharing.

Use

- Principled and practical framework for international data sharing, collaboration and good governance in genomics research.
- Protect and promote the welfare, rights, and interests of individuals and participants in genomic and health-related data sharing.
- Complement laws and regulations on privacy and personal data protection, as well as policies and codes of conduct for the ethical governance of research.
- Tool for the evaluation of responsible research by research ethics committees and data access committees.
- Provide overarching principles to be respected in developing legally-binding tools such as data access agreements.

Status and access

The Framework for responsible sharing of genomic and health-related data is available online: http://www.thehugojournal.com/content/pdf/s11568-014-0003-1.pdf

Developed by

BioSHaRE partner McGill University has led the development of the Framework, engaged the international collaboration, published the Framework, and led its translation in 10 languages (as of June 30, 2015).

Current applications
Elaborating on the general principles and guidance offered in the Framework, the Global Alliance for Genomics and Health creates policies that provide specific guidance on particular issues, the first of which is a consent policy.
Contact
Professor Bartha Knoppers McGill University, Canada Bartha.Knoppers@mcgill.ca

ELSI guidance on geo-referencing

Description and purpose
This works entails the review of the European and the Dutch legal framework for data protection in relation to geographic information systems (GIS) in general and GIS for biobanks in particular, with the aim to ensure that the BioSHaRE GIS Toolkit is compliant with said laws. Geo-referencing of biobank participants' addresses and calculating the related exposures to for instance air pollution and noise poses a number of ELSI issues. Has the individual cohort member been informed about the enhancement of his phenotype by this kind of data? Is the enhancement subject to IRB-approval? What rules govern the access to the enhanced phenotypes? Who owns the resultant data? Do the study of GIS-Data and the enrichment of biobank data therewith, pose any (increased) risk of 'spatial point' or 'geo-location' discrimination of biobank participants when seeking access to public or private services?
Designed for
Database owners ; to assess the risk-benefit of linking GIS data Researchers (individual or consortia): to understand the consequences of using GIS-data, to manage oversight where needed
Use
• To guide researchers in the use of GIS and environmental exposure data, from an ethico-legal perspective • To address potential for study subject risk and discrimination from use of geographical data
Status and access
Deliverable 7.1 provides a preliminary set of answers to these questions based on an in depth analysis of the legal framework. The resulting publications propose a model for oversight of population biobank-GIS research.
Developed by
This review was performed by BioSHaRE partner LP in collaboration with BioSHaRE partners of ICL and McGill and external party Swiss Tropical and Public Health Institute Basel, Switzerland.
Current applications
In BioSHaRE participant data has been geo-referenced in the Environmental determinants of health project. The legal review is used to further develop guidance for biobanks on geo-referencing and privacy impact.
Contact
Mr. Jasper Bovenberg

Legal Pathways, The Netherlands
jabovenberg@xs4all.nl

ECOUTER - Employing COnceptUal schema for governance, policy and Translational Engagement in Research

Description and purpose

ECOUTER is both a tool and a methodology for stakeholder engagement. ECOUTER software uses mind mapping and the existing evidence base to examine questions of interest with and within a defined stakeholder community.
An ECOUTER begins with an initial question and links to key items from the existing evidence base. Participants are invited to respond and contribute ideas and links in a mind map.
Mind-mapped discussions are then analysed to generate a conceptual framework of the phenomenon or issue considered. The results are further discussed with the participants or, where participation is fleeting and anonymous, with participants from similar stakeholder communities. The conceptual schema(s) and feedback iterations can form the basis of recommendations for research, governance, practice and/or policy.

Designed for

Although designed initially to facilitate engagement among biobank research participants, patients and the public, ECOUTER can be used to discuss almost any topic and involve stakeholders from almost any community.

Use

- ECOUTER uses a self-generating online forum to discuss a central question using mind mapping software and online access to external information ('evidence')
- Conducting an ECOUTER exercise involves the following steps:
- Engagement and knowledge exchange ('mind mapping')
- Analysis/synthesis
- Concept and recommendation development
- Feedback and refinement
- ECOUTER brings engagement to the stakeholder instead of taking the stakeholder to the engagement: especially important for those who are geographically isolated or resource poor.

Status and access

ECOUTER uses an open source web tool, Mind42, as a forum for online discussion. Mind42 is accessible by anyone with a computer, tablet or smartphone and an Internet connection.
Experience with previous pilots of ECOUTER has demonstrated a series of technical issues that will need to be resolved before the potential of the tool can be fully realised.
Documentation has been made available by D2K to support the running of an ECOUTER event and to aid participants in their use of Mind42: https://wikis.bris.ac.uk/display/ECOUT
No other technical requirements or expertise are needed to participate in an ECOUTER exercise using the Mind42 website.

Developed by

ECOUTER was originally conceived and developed further by the Data2Knowledge (D2K) research group at the University of Bristol, under the BioSHaRE project, with cofunding from Wellcome

Trust and the Medical Research Council (UK).
Current applications
Several ECOUTER sessions have been held on issues related to biobanking: What are the ethical, legal and social issues related to trust in data linkage? Your medical records: handover or hands off? An ECOUTER at the BioSHaRE conference will discuss the results of an evaluation of BioSHaRE tools and methods to develop recommendations for their further use and development.
Contact
Professor Madeleine Murtagh University of Bristol, UK madeleine.murtagh@bristol.ac.uk

3.3 Data

The harmonisation process carried out in BioSHaRE generated new data and data tools including:

1. A catalogue with metadata of the participating biobanks
2. Dataschemas with the target variables
3. Harmonisation algorithms
4. Harmonised variables

The data associated with points 1 and 2 as well as the harmonisation algorithms under point 3 are openly available and currently available through the BioSHaRE website. Access to this information will be transferred to the Maelstrom Research website after the project ends. The actual harmonised data (point 4), produced from recalculating existing data held by the biobank participants, are stored on servers at the respective biobanks. Access to these data is subject to the procedures of the individual biobanks. BioSHaRE plans did not encompass establishing a central database with the research data and there is no opportunity for outside users to access the BioSHaRE datasets as a whole.

3.4 Knowledge

BioSHaRE produced knowledge related to the software tools, ELSI tools and standards for sample handling described in section 3.2. In addition BioSHaRE produced knowledge from the so called Core Projects: scientific projects that apply and combine tools developed in the work packages. Six Core Projects were initiated over the course of the BioSHaRE:

1. Healthy Obese Project (HOP)
2. Environmental Determinants of Health Project or Environmental Core Project (ECP)
3. Metabolomics Project
4. The Social Core Project
5. BioSHaRE - IT Project
6. Statistics Core Project

The HOP core project examined what percentage of the obese population is metabolically healthy, and why they stay healthy. In particular, lifestyle factors like smoking, physical activity, nutrition, and genetic information were investigated. The ECP studied the effect of environmental exposures, specifically to air pollution and traffic noise, on the health of the population. The Metabolomics project investigated whether and how the results of Nuclear Magnetic Resonance and Mass Spectrometry techniques can be compared. The Social Core Project examined the social and epistemic implications of tools and methodologies for data sharing, and biobank standardisation and harmonisation. The BioSHaRE-IT core project is closely related to the tools development in building an open source IT infrastructure for database federation, secure data harmonization and sharing among biobanks. The statistics core project developed statistical methodology to harmonize longitudinal data and use the current and new biobanks to illustrate the methodology.

Data harmonisation and federated analyses was performed in the HOP and ECP Core Projects using data from 13 cohorts, as listed in Table 3 below.

Table 3 Overview of the studies participating in the BioSHaRE HOP and ECP Core Projects

Study name	Country	Number of participants for analyses	HOP	ECP
Cartagene	Canada	7829	X	
Cooperative health research in the Region of Augsburg (KORA)	Germany	3080	X	
Cooperative Health Research in South Tyrol Study (CHRIS)	Italy	1583	X	
Cork and Kerry Diabetes and Heart Disease Study Phase II – Mitchelstown cohort	Ireland	2048	X	
EPIC-Oxford	United Kingdom	57446		X
FINRISK	Finland	5024	X	
LifeLines Cohort Study & Biobank	Netherlands	90920	X	X
Microisolates in South Tyrol Study (MICROS)	Italy	1060	X	
National Child Development Study (1958 Birth Cohort)	United Kingdom	7210	X	
Nord-Trøndelag Health Study (HUNT)	Norway	78968	X	X
Prevention of REnal and Vascular ENd-stage Disease (PREVEND)	Netherlands	8592	X	
Study of Health in Pomerania (SHIP)	Germany	4308	X	
UK Biobank (UKB)	United Kingdom	502656		X
Total participants		**773499**	**210 6223**	**740 594**

The research in the core project has driven the development of tools. As a result the tools have been directly used in the practice of scientific research. The core projects also supported the distribution of the tools into the scientific community, since researchers immediately were aware of the benefits using these tools.

User feedback of the BioSHaRE investigators has partly been monitored by social scientists assigned in the project and has been used directly in the project to further develop the tools. In addition a formal evaluation has been done of the pilot implementation of BioSHaRE tools and methods: the EnMESHD study.

The knowledge on the harmonisation has resulted in DataSchemas for the BioSHaRE Healthy Obese Project and the Environmental Determinants of Health projects, and are described in Deliverable 7.7 "Scientific report on harmonization of life habits/behaviours, physical and social environment, and socio-economic status variables"

The other scientific knowledge, as well as the recommendations, best practices and standards resulting from the project are reported in deliverables and peer-reviewed publications. BioSHaRE partners published nearly 100 scientific papers describing the foreground produced in BioSHaRE.

4. The potential impact and the main dissemination activities and exploitation of results

4.1 Potential impact- Contribution to the impacts foreseen in the Work Programme

BioSHaRE-EU directly contributes to the impacts foreseen in the Work Programme Health 2010: Harmonisation of phenotyping and biosampling for human large-scale research biobanks.

The impact of BioSHaRE-EU is achieved through its main objectives:

1. To improve the assessment and classification of multivariate phenotypes associated with complex diseases, including environmental and life style exposures;
2. To enable interoperability of databases of both phenotype and genotype data;
3. To develop evidence-based standards for harmonised quality management and quality control during the collection, transport, processing, storage and retrieval of human biospecimens;
4. To develop effective strategies for optimising the correlation and integration of existing and novel data;
5. To maximize the sharing and exchange of information between population cohorts and clinical research centres/biobanks across Europe;
6. To build on pre-existing achievements (where available) and coordinate its activities with similar international efforts;
7. To consider ethical, social and legal aspects, as well as the relevance to public health.

Figure 2 illustrates the contribution of each of the work packages to these objectives.

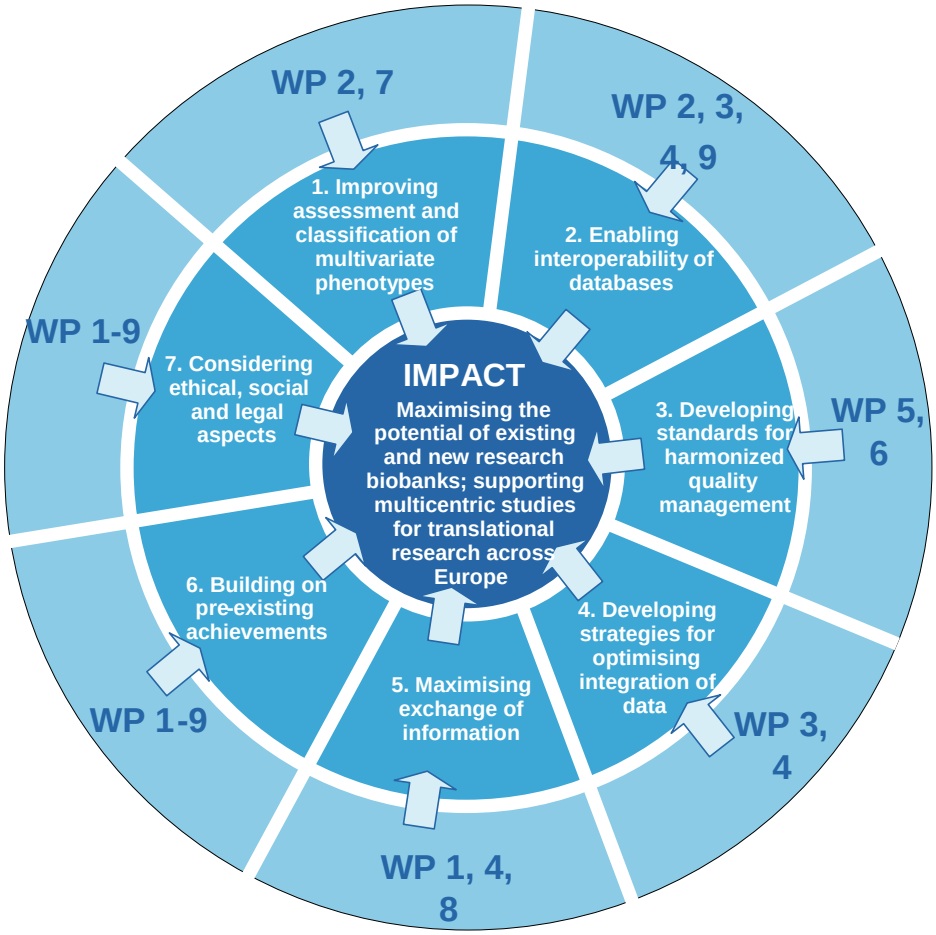


Figure 2 BioSHaRE-EU Work Packages and expected impact

(1) Improve the assessment and classification of multivariate phenotypes associated with complex diseases, including environmental and life style exposures.

Harmonisation of risk factors and health outcomes is one of the key objectives of BioSHaRE and has been achieved by the development of a number of tools and by harmonising variables in the Healthy Obese Project (HOP) and the Environmental Determinants of Health Project or Environmental Core Project (ECP).

Specifically, the software tools and methods Opal (Deliverable 2.2) , Mica (Deliverable 1.2 and 2.2), DataSHaPER (Deliverable 2.2), DataSchemas (Deliverables 2.2 and 2.3), EnviroSHaPER (Deliverable 7.4), BiobankConnect (Deliverable 2.2) , SORTA and Vortext/Spá have been developed to facilitate data harmonisation across multiple databases.

The actual harmonisation in the HOP and ECP resulted in 100 and 76 variables in the HOP and ECP DataSchemas, respectively. These variables cover a wide range of lifestyle and behavioural constructs, anthropometric and biochemical measures, socio-demographic information, environmental exposures, and health outcomes. The complete list of variables is given in Deliverable

7.7 and harmonisation steps and outcome are described for physical activity, health-related quality of life, nutritional habits, socio-economic status, somatic symptoms and environmental exposures.

(2) Enable interoperability of databases of both phenotype and genotype data.

Work packages 2 and 4 developed object models and standard data exchange formats for core molecular, phenotype, and environmental data specifically tailored to the needs of biobanking and compatible with other related data models. This work was incorporated in a number of software tools, notably Café Variome, OmicsConnect (Deliverable 4.1) and MOLGENIS / Observ-EMX (Deliverables 4.2 and 4.4).

(3) Develop evidence-based standards for harmonised quality management and quality control during the collection, transport, processing, storage and retrieval of human biospecimens.

A number of Standard Operating Procedures and recommendations have been developed in BioSHaRE, specifically Recommendations for Storage and Analyses of Data and Samples - Evidence-based minimal standards (Deliverable 5.1), SOPs for handling of liquid biosamples (Deliverable 5.2), and reports on the temperature effects of preparing and thawing samples on different analysis techniques (Deliverable 5.3) Harmonization and standardisation of inflammatory biomarkers (Deliverable 5.4) and Trace element analysis (Milestone 50).

(4) Develop effective strategies for optimising the correlation and integration of existing and novel data.

BioSHaRE developed statistical methods for meta-analysis of large scale studies, and bioinformatics tools for studying complex diseases. The tools include Café Variome, OmicsConnect (Deliverable 4.1) and MOLGENIS / Observ-EMX (Deliverables 4.2 and 4.4). Statistical methods have been developed in the Statistics Core Project and have been implemented in DataSHIELD for use in federated analyses (Deliverables 3.5 and 3.6).

(5) Maximizing the sharing and exchange of information between population cohorts and clinical research centres/biobanks across Europe.

The mission of BioSHaRE is to ensure the development of harmonized measures and standardized computing infrastructures to facilitate data sharing and analysis across multiple biobanks and related databases. The majority of the software tools contribute to achieving this objective: DataSHIELD, Opal, Mica, DataSHaPER, DataSchemas, Café Variome, OmicsConnect, MOLGENIS / Observ-EMX

(6) Build on pre-existing achievements (where available) and coordinate its activities with similar international efforts.

Development of the tools has been and continues to be a joint effort, supported by BioSHaRE in conjunction with other sources of resources and funds. Most of the software tools were in development before BioSHaRE was initiated. Many of them are integrated into new projects and initiatives and they will continue to be developed and improved in those contexts and with new sources of funding.

BioSHaRE built directly on extensive precursor work in the field of harmonisation and standardisation based on the now well-established DataSHaPER (Data Schema and Harmonisation Platform for Epidemiological Research) paradigm that has been developed jointly by a number of our partner projects including P^3G, PHOEBE and GENECURE. In developing standardised approaches and infrastructures for bioinformatics BioSHaRE built upon and extended the work in bioinformatics and IT that has already been undertaken by GenomEUtwin (FP5), GEN2PHEN (FP7), BBMRI (ESFRI), OBiBa and P^3G.

Interactions with the scientific community occurred through all partners and work packages, but in particular by Work Package 8 “Strategic integration and coordination with major biobanking initiatives, partnerships and dissemination”. This WP performed horizon scanning activities and interfaced with initiatives internationally through a series of international biobanking summits to provide a forum for the community to discuss the most current issues going forward.

Through close collaborations with and participation of BioSHaRE investigators in many related international projects and initiatives, the BioSHaRE tools and methods are applied, exploited, further improved, and made accessible for the scientific and biobanking communities as well as for wider society. These projects and initiatives include but are not limited to BBMRI-ERIC, BBMRI-LPC, InterConnect, Maelstrom Research, The Global Alliance for Genomics and Health, CORBEL.

(7) Consider ethical, social and legal aspects, as well as the relevance to public health.

The BioSHaRE ELSI stream has worked with all work packages to identify ELSI issues and to develop standards for large scaled harmonized data pooling and use of data from biobanks. Specifically, BioSHaRE developed ELSI tools for data and sample sharing (Deliverables 2.4, 2.5 and 5.5), a Framework for responsible sharing of genomic and health-related data (Deliverable 3.4), ELSI guidance on geo-referencing and use of environmental risk and personal life-style data in European biobanks (Deliverable 7.1) and a tool for governance, policy and Translational Engagement in Research (ECOUTER).

Table 4 Overview of BioSHaRE objectives and relevant tools

Expected impact/ aim	Tools
1. Improve the assessment and classification of multivariate phenotypes associated with complex diseases, including environmental and life style exposures	BiobankConnect Opal, Mica, DataSHaPER DataSchemas EnviroSHaPER SORTA Vortext/Spá
2. Enable interoperability of databases of both phenotype and genotype data;	Café Variome OmicsConnect MOLGENIS / Observ-EMX
3. Develop evidence-based standards for harmonised quality management and quality control during the collection, transport, processing, storage and retrieval of human biospecimens;	SOPs for the Handling of Liquid Biosamples Recommendations for Storage and Analyses of Data and Samples - Evidence-based minimal standards Temperature effects of preparing and thawing samples on different analysis techniques Harmonization and standardisation of inflammatory biomarkers Trace element analysis
4. Develop effective strategies for optimising the correlation and integration of existing and novel data;	Café Variome OmicsConnect MOLGENIS / Observ-EMX
5. Maximizing the sharing and exchange of information between population cohorts and clinical research centres/biobanks across Europe;	DataSHIELD Opal, Mica, DataSHaPER DataSchemas Café Variome OmicsConnect MOLGENIS / Observ-EMX
6. Build on pre-existing achievements (where available) and coordinate its activities with similar international efforts;	BRIF, ORCID Organisation of meetings WP8 etc.
7. Consider ethical, social and legal aspects, as well as the relevance to public health.	ELSI tools for data and sample sharing Framework for responsible sharing of genomic and health-related data ELSI guidance on geo-referencing and use of environmental risk and personal life-style data in European biobanks ECOUTER

4.2 Potential impact for the scientific community

Combining biobank data harmonisation with a federated approach to data analysis ("the BioSHaRE approach") yields a number of benefits for epidemiological research. These include the use and re-use of harmonized data from different studies, improved data quality through a rigorous approach to harmonization, increased statistical power, and improved analytical flexibility allowing researcher to do combined analyses in real-time and at their convenience. This new and unique approach requires careful central governance and organization to facilitate the activities of biobanks and researchers and to maximize gain from those efforts.

Using the BioSHaRE approach has scientific, economic and other impact/ benefits for biobank and cohort researchers:

1. **Scientific Benefits**

- Allows the **massive pooling of data sources** to answer a single scientific question. This yields extremely large sample sizes with greater statistical power than standard approaches, minimizes false positive results and ensures reliability of data findings. It also allows researchers to study smaller sub-populations of interest.
- **Improves the generalizability of results** by facilitating the examination of the consistency of evidence across broad and diverse study populations.
- Helps to **ensure the validity of comparative research**.

2. **Economic Benefits**

- **Provides long-term returns on a single initial investment**: through generating a permanently-accessible harmonised dataset, variables can be analysed and re-analysed for future research. New harmonisation can be conducted in the already-established infrastructure. After initial investments, maintenance costs are low and federated data analysis tools are freely available.
- Helps to **promote efficient use and re-use of existing research resources and infrastructures**. Encourages more efficient secondary usage of existing data. This maximizes the usage of existing cohorts, and better justifies future investment by institutions and funding agencies.

3. **Other Benefits**

- Provides **increased opportunities for collaborative multi-centre research** to address existing and emerging questions about health and disease development. Increases researchers' collaborative networks, raises visibility through increased publication activity.
- **Raises the research/researcher profile** through increased publication, broader collaborative networks (BRIF, publications), and expansion of possible research topics.
- **Reduces common privacy and security risks** often encountered in the transfer of individual-level data, through exchange of summarized data statistics. Individual data does not leave the local study site.

4.3 Socio-economic impact and the wider societal implications of the project so far

BioSHaRE has been acting on the forefront of ethical, legal and social issues related to biobanking and data sharing, privacy, consent and trust. Through activities such as organizing international conferences and workshops, and through designing ELSI guidelines and notably the Framework for responsible sharing of genomic and health-related data, BioSHaRE has turned the spotlight on the need to create safe, secure and reliable methods and frameworks for sharing data across research groups for the collective benefit of all researchers, clinicians and EU citizens. The social and epistemic implications of the BioSHaRE tools and methodologies have been extensively studied and fed back into the development of the tools by the Social Core Project.

In a broader sense, BioSHaRE embodies the future of collaborative cohort research in Europe. Existing data and sample collections and resources that were initially fragmented and disconnected can, using BioSHaRE's unique collection of tools and methodologies, be linked and repurposed for new research questions. In this way the BioSHaRE tools directly support the "data life cycle" and "Open Access" policies of the EU. Further, this can be accomplished without transnational data transfer or infringing upon the privacy of the European citizens who participate. In this sense BioSHaRE is greater than the sum of its parts: individuals who participate in especially small/specialized cohorts, have the opportunity to contribute to new, relevant and ground-breaking research pursuits through BioSHaRE's approach, in ways that would not have been possible as stand-alone resources. Re-use of these investments is cost-effective and efficient. It introduces a new, sustainable approach to research that changes the cultural framework of European researchers. And all of these are in addition to the scientific findings that these resources generate, such as the metabolic consequences of different types of obesity and the additional risk of air pollution on specific health indicators. These translate directly into prevention strategies and treatment innovations.

4.4 Main dissemination activities

BioSHaRE scientists have actively disseminated research results to both scientific and non-specialist audience. This has been exemplified by the extensive list of dissemination activities reported in A2 List of Dissemination Activities. The BioSHaRE project website (www.bioshare.eu) has served as central tool to widely disseminate BioSHaRE main research outputs and activities to BioSHaRE's three main target audiences: the scientific community, policy makers and the general public.

The **scientific community** is the most important audience for which all of BioSHaRE's output is relevant. Potential user groups in the scientific community are the database custodians (biobanks, cohort studies, other non- biobank databases) and researchers (individual or in collaborations).

Interactions with the scientific community occurred through all partners and work packages, but in particular by Work Package 8 “Strategic integration and coordination with major biobanking initiatives, partnerships and dissemination”. This WP performed horizon scanning activities and interfaced with initiatives internationally through a series of international biobanking summits to provide a forum for the community to discuss the most current issues going forward. WP8 also worked to ensure European-wide implementation of the tools and services developed in BioSHaRE to foster wider strategic scientific integration. This has required continual interfacing and communication with projects nationally and internationally.

Targeted **policy makers** include users of new knowledge within the Commission such as representatives of the Directorate-General Research and the Health and Consumer Protection Directorate General. Targeted policy makers outside the Commission include relevant authorities from the Member States (Health Ministries and Pubic Health Institutes), as well as opinion leaders in the relevant fields.

A large number of meetings and workshops have been organized by BioSHaRE partners in collaboration with other organizations and opinion leaders to discuss and define critical issues and build consensus on multiple topics facing international biobanking. These include issues related to ELSI, harmonisation, data directives and regulatory frameworks for sample use, cloud computing and ICT solutions for biobanking. Our input has been brought to the attention of policy makers though reports, statements, published papers and personal communication.

The **general public** has been informed through the BioSHaRE project web site to disseminate the basic concept of BioSHaRE. Through local dissemination efforts in partner institutions, BioSHaRE partners also disseminated their relevant research results to more layman audience through interviews by media and public lectures. BioSHaRE also produced audio-visual tools to introduce the project concept, tools, services and main results to the wider audience. As described below, the information videos have been widely distributed through web-based video sharing channels, such as the BioSHaRE You tube channel to help inform the general public.

Two major dissemination activities were organized during the final year of BioSHaRE: a Catalogue of tools and methods for data sharing was compiled and distributed in a printed version and as an online version. It has been widely distributed and features an overview of the diverse tools and services for harmonization developed in BioSHaRE. It was also handed-out at the other major dissemination activity this year, the BioSHaRE Tool Roll-Out conference entitled “LATEST TOOLS and SERVICES for DATA SHARING”, held July 28th, 2015 in Milan, Italy. This conference was attended by leading scientists, projects and major initiatives in biobanking. The registration list included 156 participants representing 22 countries.

4.5 Exploitation of results

Exploitation of the tools, data and knowledge produced in BioSHaRE is done in a scientific context. The tools developed in BioSHaRE are made available open source with no intention to become profitable and commercially exploited. The data that were produced has enriched the participating biobanks. These in turn can and will exploit the data by providing it to investigators for further research. The knowledge produced in BioSHaRE has resulted in a large number of publications that are mostly open access. As such, the foreground produced in BioSHaRE is considered to contribute to the general advancement of knowledge. Even though commercial exploitation is not foreseen, the potential and actual use and application of the foreground is extensive and has high value for the stakeholders, mainly for further development and research activities.

Development of the tools has been and continues to be a joint effort, supported by BioSHaRE in conjunction with other sources of resources and funds. Most of the software tools were in development before BioSHaRE was initiated. Many of them are integrated into new projects and initiatives and they will continue to be developed and improved in those contexts and with new sources of funding.

BioSHaRE built directly on extensive precursor work in the field of harmonisation and standardisation based on the now well-established DataSHaPER (Data Schema and Harmonisation Platform for Epidemiological Research) paradigm that has been developed jointly by a number of our partner projects including P^3G, PHOEBE and GENECURE. In developing standardised approaches and infrastructures for bioinformatics BioSHaRE built upon and extended the work in bioinformatics and IT that has already been undertaken by GenomEUtwin (FP5), GEN2PHEN (FP7), BBMRI (ESFRI), OBiBa and P^3G.

To date, a number of the BioSHaRE partners and closely linked organisations are exploiting the BioSHaRE foreground by offering services to researchers and consortia including the pan-European BBMRI-ERIC.

5. Address of the project public website and contact details

Website: http://www.bioshare.eu
Email: bioshare@umcg.nl

Contact details:

Professor Ronald Stolk
Corporate staff for Research Policy
University Medical Center Groningen
PO Box 30.001 | 9700 RB Groningen | The Netherlands
tel +31 50 36 11879
email r.p.stolk@umcg.nl