B.I.R.O.

Best Information through Regional Outcomes

A Public Health Project funded by the European Commission, DG-SANCO 2005

WP 5:

Privacy Impact Assessment - Step 4

“Privacy Impact Assessment Report”

The B.I.R.O. PIA Team

April 2009

Table of Contents

The P.I.A. Team

Privacy Facilitators:

- Dr. Concetta Tania Di Iorio, Legal Consultant, SeRectrix, Pescara, ITALY
- Dr. Fabrizio Carinci, Health System Research**,** SeRectrix, Pescara, ITALY

P.I.A. Team Members:

- Dr. Valentina Baglioni, Dipartimento di Medicina Interna, Università di Perugia (UNIPG), ITALY
- Dr. Scott Cunningham, Division of Medicine & Therapeutics, University of Dundee (UNIDUND), SCOTLAND
- Dr. Peter Beck, Institute of Medical Technologies and Health Management, Joanneum Research (JOANNEUM), Graz, AUSTRIA
- Dr. Sven Skeie, Department of Medicine, Section of Endocrinology, University of Bergen (UNIBERG), NORWAY
- Dr. Simion Pruna. Institute of Diabetes, Nutrition and Metabolic Diseases “N. Paulescu” (PAULESCU), Bucharest, ROMANIA
- Prof. Joseph Azzopardi, Department of Medicine, Universitat ta Malta (UNIMALT), La Valletta, MALTA
- Dr. Vivie Traynor, Department of Health Promotion, Ministry of Health (CYPRUS), Lefkosia, CYPRUS

1. Executive Summary

The B.I.R.O. Information System involves the use of sensitive-medical data collected through diabetes registries within national boundaries and further processed for public health studies at the international level.

Privacy impact assessment (PIA) is a systematic and flexible process for evaluating a proposal/project in terms of its impact upon privacy. As required by PIA methodology, it has been specifically adapted to the BIRO context.

Privacy Impact Assessment objectives were:

- the provision of a definitive description and analysis of privacy risks, applicable privacy legislation and mitigation strategies adopted in the implementation and management of the BIRO Information System,
- the identification of a general methodology supporting collaborative networks of regional disease registers and the routine evaluation of health information systems.

The entire process envisaged four consecutive steps: Preliminary Privacy Impact Assessment, Data Flow Analysis, Privacy Analysis and Final PIA Report.

The Preliminary PIA was conducted by a multidisciplinary team carrying out a systematic review of privacy literature, followed by a general discussion on the data flow.

The Data Flow Analysis focused on the description and in depth analysis of some alternative B.I.R.O. architectures identified in the first step. A Delphi consensus procedure was undertaken to define the best alternative through the production of the following materials:

- data flow tables, including all possible scenarios for the collection, use and disclosure of personal information/data, with related options,
- information flow questionnaire, used to assign marks for each scenario/option,
- overall consensus table, ranking all alternative architectures, scenarios and options.

The Privacy Analysis covered any privacy issue arising in the transfer of data from the local centres to the central database. Potential privacy risks were identified and thoroughly analysed through a summary table indicating mitigation strategies to be implemented. The level of risk was classified according to an ordinal scale of intensity.

The present PIA Report compiles the results from all phases using a structured format.

The Preliminary analysis led to the identification of three main candidate architectures with differing levels of data sharing: “individual patient data, de-identified through a pseudonym”; “aggregation by group of patients, with Centre’s identifiers available in de-identified form, securely encrypted”; and “Aggregation by Region”.

In the context of the PIA Step 2, Data Flow Analysis, the second architecture was selected as the best solution in terms of privacy protection, information content, scientific soundness and feasibility. The Privacy analysis performed a detailed assessment of the various aspects involved in the adoption of the final BIRO architecture.

The transfer of information occurring in the BIRO system, based upon the exchange of de-identified data and targeted mitigation strategies, identifies a low level of privacy risk.

According to the BIRO architecture, participating centres apply procedures for data anonymisation before any transfer to the BIRO central database is made. The central server processes aggregate records solely for statistical and scientific purposes. Ex Recital 26 of the EU Data Protection Directive*,* anonymisation allows personal data processing without consent, placing anonymous data outside the scope of the data protection principles therein contained. The processing of anonymous data is therefore to be considered legitimate.

The BIRO system processes only statistical objects that are stored as aggregate table into flat text comma delimited files. Hence, there is no possibility, according to the state of the art, to identify a patient, either directly or indirectly, with a reasonable effort.

Clinical Centres also receive privacy protection through the use of a pseudonym for Centres IDs, in combination with publication mechanisms that avoid identification from third parties (e.g. pseudonym combined with number of patients expressed only in percentage).

Aggregate data are processed by the local database engine and sent to the central statistical engine through an “ad hoc” communication software ensuring secure information exchange and compliance with security requirements enshrined in EU and international data protection norms. Considering that data are rendered anonymous by local BIRO centres and transmission occurs in a secure environment, the further processing at the level of the global statistical engine cannot pose any privacy risk, either directly or indirectly.

Trans-border data flow envisaged in BIRO is legally viable according to the EU legislation.

Publication of project results is performed to avoid any direct/indirect identification of data subjects and/or local centres.

Privacy impact assessment shows that the selected BIRO architecture fulfils privacy protection requirements by addressing and resolving broad privacy concerns from different angles. The architecture of the system flexibly affords the best privacy protection in the construction of an efficient model for the continuous production of European diabetes reports. The methodology identified can be usefully applied in other fields of health information, particularly where disease registers are involved as primary units for data collection and statistical analysis.

2. Introduction

2.1 Report Objectives

The present Report exhaustively documents the privacy impact assessment process occurred in the development and implementation of the B.I.R.O. Health Information System.

The initial part focuses, following a general introduction to the B.I.R.O. project, on the notion and definition of privacy impact assessment in order to allow for the reader to fully understand the various aspects of the process undertaken.

The various steps of the PIA process are then described, including aspects of the PIA methodology and materials and tools developed and applied "*in concreto*".

In addiction, the Report provides a full assessment of all possible privacy risks that might occur in the construction, implementation and further maintenance of the B.I.R.O. Health Information System.

The adoption of mitigation strategies, privacy enhancing technologies and security mechanisms are also herein documented.

2.2 The B.I.R.O. Project

"Best Information Through Regional Outcomes" (BIRO) is a three years public health project started in 2005, funded under the EC Public Health Programme 2003-2008[1]. The project is coordinated by the University of Perugia, Italy and includes, as partners, the University of Dundee (Scotland), Joanneum Research (Austria), University of Bergen (Norway), Paulescu Institute (Romania), University of Malta (Malta), Cyprus Ministry of Health (Cyprus). Other collaborating institutions include Serectrix (Italy), NOKLUS (Norway) and Telemedica (Romania).

The general objective of BIRO is to build a common European infrastructure for standardized information exchange in diabetes care, to monitor, update and disseminate evidence on the application and clinical effectiveness of best practice guidelines on a regular basis.

The general objective is pursued through the realization of several work-packages, allowing the identification of target parameters and indicators; definition of a common dataset and a data dictionary supported by an appropriate schema for its representation; development of a report template, associated database and statistical engines required to deploy its content in both printed and web format; validation of a secure protocol for international communication and shared data analysis; construction of a web portal to test the dissemination of European estimates on a routine basis.

The technology associated to the construction of the system is centered on the definition of the "Shared Evidence-based Diabetes Information System" (SEDIS), whose general architecture is based on the application of two consecutive data processing steps.

At the basic level, a general version of the system runs in each single register ("*local SEDIS*") to produce initial estimates that are valid for the local population. All partners in the network, using the same standardized procedures, repeat the process at their best convenience. All regional estimates are sent towards a central server that compiles all "partial" results into a global report that is valid for the European level.

The functionality of the basic level of the system is ensured by three fundamental elements.

The first is the *concept and data dictionary* (CDD), storing all common definitions adopted to collect and exchange data across the network. The CDD represents the evidence-based component in the model chain.

The second, the *report* template, is located at the opposite end of the chain, and it determines the selection of data procedures and statistical methods required to estimate all results for the health report.

The third component is represented by the core engines ("*database engine*" and "*statistical engine*"), which operate on the local databases and are only accessible by local administrators. The engines deliver statistical "*objects*" (*tables, parameters, graphs*) that are then amalgamated by central components.

The overall model (*global SEDIS)* directly follows from the local implementation: once statistical objects are available from each register, they are sent to the server using a secure transmission.
The level of aggregation chosen for each object is a trade-off among formal agreement, legislation, ethical values and practical limits, all aspects that are properly investigated in the framework of the BIRO project.
The general design of the BIRO project has been progressively implemented through the definition of candidate architectures submitted to a formal evaluation process coordinated by the Privacy Impact Assessment (PIA).

2.3 Privacy Impact Assessment: General Features

There is no unique definition of PIA in the literature. It has been defined as a "process whereby a conscious and systematic effort is made to assess the privacy impacts of options that may be open in regard to a proposal. An alternative definition might be that a PIA is an assessment of any actual or potential effects that the activity or proposal may have on individual privacy and the ways in which any adverse effects may be mitigated."[2]

Moreover, PIA is usually conceived as a "protean document in the sense that it is likely to continue to evolve over time with the continued development of a particular system."[3]

Hence, there is a general consensus that a PIA is not just an end-product or a statement or practice. PIA is better conceived as a *process* rather than an outcome, which should be open-ended and regularised throughout the life-cycle of a programme/project.

With regard to different jurisdictions that have employed PIAs as structured means to assess privacy risks in government/private programs or projects, the following definitions are of utmost significance, since they highlight a bulk of common features: PIA has been defined as an "assessment of actual or potential effects on privacy, and how they can be mitigated" (Australia), "a systematic process for evaluating a proposal in terms of its impact upon privacy" (New Zealand), a "framework to ensure that privacy is considered throughout the design or re-design of a programme…[and to] identify the extent to which it complies with all appropriate statutes. This is done to "mitigate privacy risks and promote fully informed policy" (Canada), an analysis of how information in identifiable form is collected, stored, protected, shared and managed…[to] ensure that system owners and developers have consciously incorporated privacy protection throughout the entire life cycle of a system (USA)[4].

According to the above definitions, PIAs should be designed to:

- conduct a prospective identification of privacy issues or risks before systems and programmes are put in place, or modified
- assess the impacts in terms broader than those of legal compliance
- be process rather than output oriented
- be systematic.

Legal compliance is, therefore, only one of the several criteria that need to be addressed in a larger process of risk assessment. Those larger questions include the "moral and ethical issues posed by whatever is being proposed"[5]. Many projects might be technically compliant with law, but may raise significant concerns, even resistance, in certain societies or among certain publics.

The broader significance of PIAs and its increasing importance in tackling privacy issues in both public and private sectors has been demonstrated by an exhaustive study/survey recently conducted by the English government, namely: "Privacy Impact Assessments: International Study of Their Application and Effects"[6]. Some of the study conclusions are hereafter reported:

- PIAs are a good idea and are increasingly recognised as such by privacy commissioners, government agencies, private corporations and privacy advocates. They help to address the increasing concerns about privacy within advanced industrial societies.

- PIAs have been spreading around the advanced industrial world as a result of: legislative requirements; policy guidance by central government agencies; recommendations by privacy and data protection commissioners; and recognition by organisations that PIAs can expose and mitigate privacy risks, avoid adverse publicity, save money, develop an organisational culture sensitive to privacy, build trust and assist with legal compliance.

- The early experience has been evaluated in several jurisdictions, and lessons are being drawn about the most valuable ways to encourage their completion. In this respect, the decision by the ICO to embark on this initiative for the UK is very timely, and in the context of the European Union, pioneering.

- To be valuable, PIAs need to offer a prospective identification of privacy risks *before* systems and programmes are put in place. In every jurisdiction, PIA processes have been designed to be prospective.

- Many exercises which are called PIAs are, however, little more than legal compliance checks. To be meaningful, PIAs have to consider privacy risks in a wider framework, which takes into account the broader set of community values and expectations about privacy.

- PIAs are more than the end-product or statement. They refer to an entire process of assessment of privacy risks.

- PIAs are only valuable if they have, and are perceived to have, the potential to alter proposed initiatives in order to mitigate privacy risks. Where they are conducted in a mechanical fashion for the purposes of satisfying a legislative or bureaucratic requirement, they are often regarded as exercises in legitimation rather than risk assessment.

- PIA processes vary across a number of dimensions: the levels of prescription, the application, the circumstances that might trigger PIAs, the breadth of the PIA exercise, the agents who conduct PIAs, the timing, the process or review and approval and the level of public accountability and transparency.

- There is no simple formula for the conduct of a PIA. Each PIA should be dictated by the specific institutional, technological, and programmatic context of the initiative in question. A mechanical “checklist” alone does not capture the broader social, political and ethical implications of many initiatives. Any PIA requires judgment.

- Therefore the scope and depth of the PIA needs to be sensitive to a number or crucial variables: the size of the organisation; the sensitivity of the personal data; the forms of risk; the intrusiveness of the technology. A PIA screening process is commonly used to determine whether a PIA is required, and if so, the form it should take.

3. PIA Design and Application in the Context of B.I.R.O.

The Privacy Impact Assessment (PIA) of the BIRO project aims at providing a balanced approach that allows to realize the best, most privacy protective solution for the B.I.R.O. Information System and to easily demonstrate that the very best possible solution has been delivered in terms of privacy protection, information content and technical complexity (feasibility).

According to project specifications and needs, the entire process was broken down in four steps:

- Step 1 - Preliminary PIA
- Step 2 - Data Flow Analysis
- Step 3 - Privacy Analysis and
- Step 4: PIA Report

3.1 Preliminary Privacy Impact assessment

The rationale for conducting a Preliminary Privacy Impact Assessment (PIA), instead of proceeding directly to the first step of a full PIA (project initiation/need assessment), resided in the fact that the BIRO project was yet at an early design stage and lacked of sufficient information on the data flow.

In particular, the available information did not allow the identification of all the types and volumes of personal information that were to be collected, used and disclosed. Consequently, it would have been difficult to identify with precision the legislative and policy framework of the BIRO Information System and, therefore, to determine which aspects of the project were likely to involve privacy risks.

First task of PIA Step 1 was the nomination of a PIA Facilitator (PF), specialized in international law, public health and ethics, and the formation of a PIA Team (PT), including one representative from each partner of the BIRO Consortium, whose duty was to actively collaborate with the PF to carry out all tasks involved in the separate steps.

The process started by drawing a draft BIRO Information Diagram describing, at a very general level, how the federated centres/regions would link to the Shared European Diabetes Information System (SEDIS).

Figure 1 documents the architecture of the general BIRO infrastructure, along with the flow of information throughout the system and the physical/logical separation of personal information/data.

A *legislative review* was initially conducted, using systematic keywords on major search engines (Box 1), to extract relevant papers and highlight the most relevant legislative framework for BIRO and consequently provide a basic *evaluation* of the potential privacy risks associated to the creation of a Shared Information System as designed by the original project specifications.

The *legislative review* conducted in the first step of the PIA identified the major privacy implications in the use of the BIRO system.

As required by PIA's general methodology, different alternative architectures of the BIRO Information System were drafted to allow the selection of the best privacy protective architecture.

Based on a comprehensive report of the first step, distributed to all partners, and the BIRO Information Diagram independently produced on the basis of technological matters, the Consortium identified three alternative architectures for the development of the BIRO Heath Information System, envisaging different levels of data sharing.

The first alternative required the transmission, from the single centres to the Central Database, located in Perugia, of "*individual patient data de-identified through a pseudonym*". In this case a need to specify secure patient's identity encryption algorithms and privacy protective technologies for securing the data transfer was considered crucial for implementation.

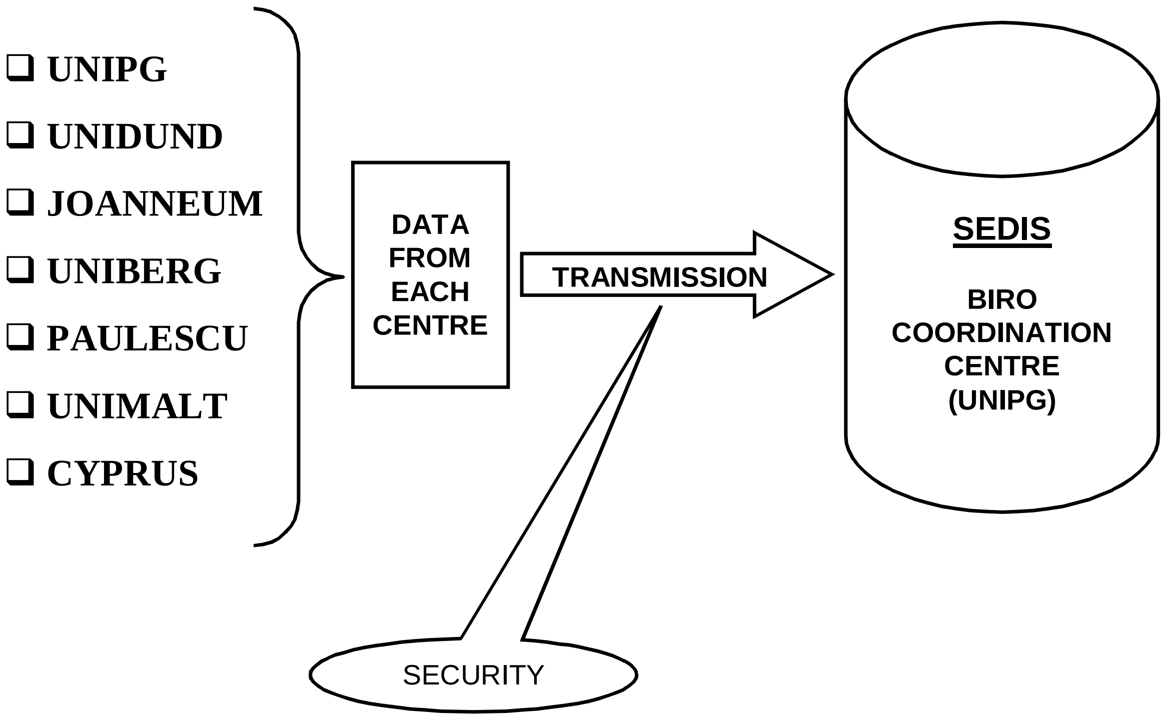
The second alternative architecture envisaged a data sharing occurring through an *"aggregation by group of patients, with Centre's IDs available but de-identified"*. It was pointed out that the use of aggregated data would require the specification of secure encryption algorithms for Centre's identity and privacy protective technologies for securing the data transfer.

The third alternative was based on an *"aggregation by Region".* A need to specify optimised data aggregation in order to impede reverse engineering was considered, in addition to the adoption of Privacy protective technologies for securing the data transfer.

The preliminary phase also considered the possibility of implementing privacy enhancing technologies and security solutions in each alternative architecture.

Special attention was dedicated to the security requirements to be implemented in the Central SEDIS, which was to be located in Perugia, Italy.

Fig. 1: Draft BIRO System Diagram



***Box 1 - Systematic Search of the literature: methods and selection criteria*:**

1st search: Ovid Medline (R) 1966 to Present with Daily Update

Search Criteria: (privacy AND ((registr* OR register) OR (health information system*) OR (health database*)))

Limits: human AND English Language AND yr = 2001-2006

Results available = 64

Core articles were identified after exclusion of papers focussing on:

- importance of diseases registries to enhance quality of care
- impact of non-European privacy laws on research
- genetic discrimination
- patient recruitment strategies

After the above selection, 12 articles have been identified as fully relevant.

2nd search: Law Journals

Search engines of the following journals have been selected for their focus on privacy and health information:

- European Journal of Health Law
- Privacy Law and Law Reporter

Results for the search:

(privacy AND ((registr* OR register) OR (health information system*) OR (health database*)))

Limits: years = 2001-2006

Results available: 11 articles

Core articles were identified according to previous exclusion criteria: 2 relevant articles were found.

Combined Results and Core Articles

All articles included in search 1 and 2 have been fully read. A core set of 12 articles has been revised by all partners to ensure compliance of BIRO with privacy requirements and a correct implementation of the PIA.

3.2 Data Flow Analysis

3.2.1 Objectives

The general objective of PIA Step 2, *data flow analysis,* was to describe and analyse the information flow occurring through the BIRO system in order to ultimately identify the best privacy protective BIRO architecture.
Specific objectives of the *data flow analysis* were:

- to develop a detailed description and analysis of BIRO data flow
- to describe and in-depth analyse the BIRO system alternatives, selected in PIA Step 1
- to identify the best privacy enhancing system architecture for BIRO

In order to document the BIRO data flow, the following activities were carried out:

- description and analysis of the BIRO Health Information System architecture through a *diagram*
- description of the information flow involved in the project through
  - identifying clusters of personal information/data involved in BIRO System
  - developing detailed *data flow tables of the BIRO selected alternatives*
- provision of an ad hoc *information flow questionnaire*, developed on the basis of the data flow tables
- ranking of the candidate architectures through the assignment of mark to each option on the basis of standard criteria involving privacy, information content and technical complexity.

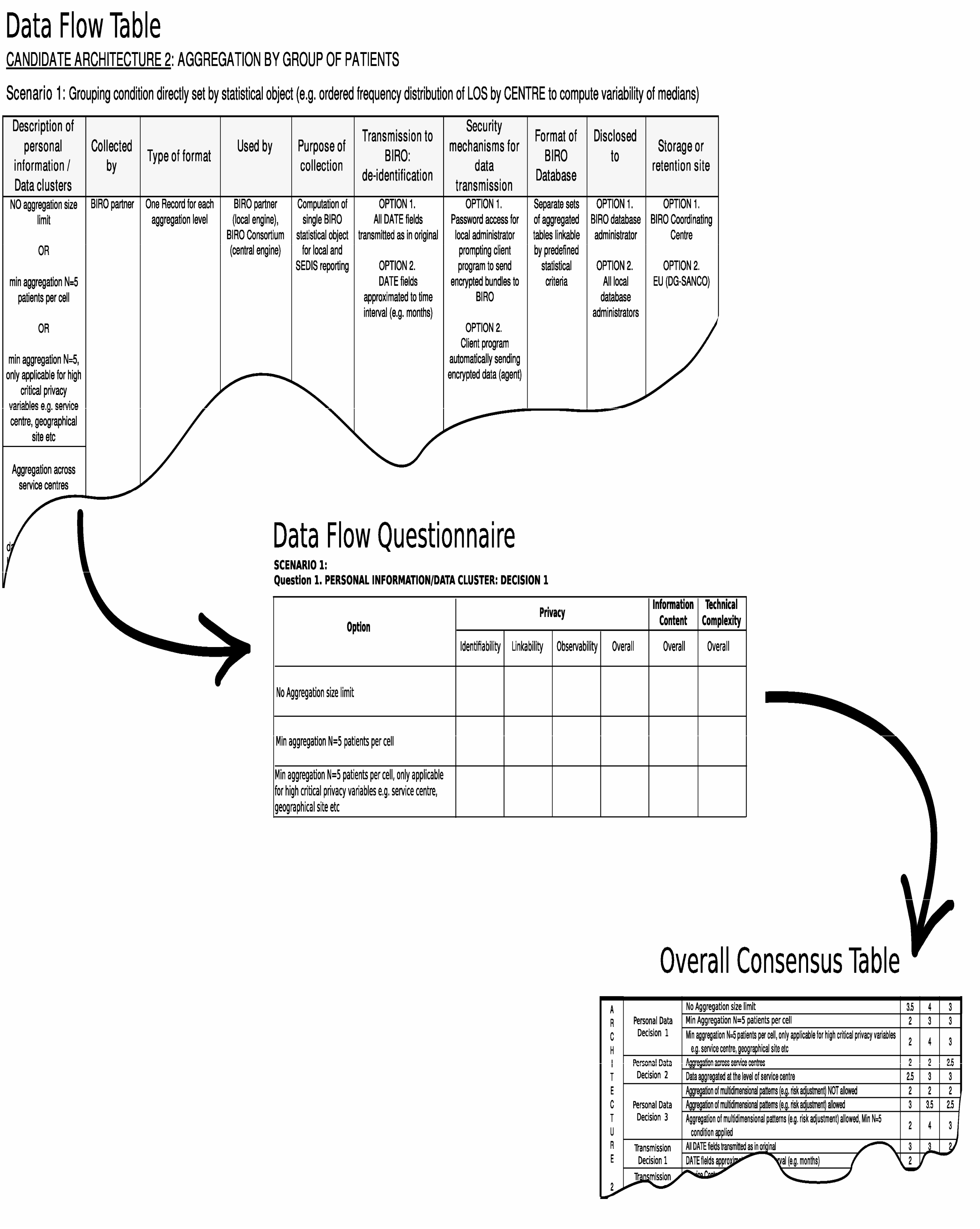
3.2.2 Materials and Methods

The *data flow analysis* included a detailed description and an in-depth analysis of the BIRO architecture as well as of the data flow occurring for each of the candidate alternatives. The identification of the best privacy enhancing information system architecture was carried out through a *Delphi Consensus Procedure* aiming at ranking the separate alternatives via scores assigned to each dimension involved.
The definition of the best alternative required three basic elements:

- a scheme to highlight relevant dimensions, with a number of possible options: data flow tables (DFT);
- a questionnaire to assist scoring for each dimension/option on the basis of the level of compliance to relevant principles, legislation and public concerns about privacy (IFQ);
- an overall consensus table (OCT).

Materials were assembled using the procedure described in figure 2.

Figure 2: Consensus Procedure



Data Flow Table

CANDIDATE ARCHITECTURE 2: AGGREGATION BY GROUP OF PATIENTS

Scenario 1: Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by CENTRE to compute variability of medians)

Description of personal information / Data clusters	Collected by	Type of format	Used by	Purpose of collection	Transmission to BIRO: de-identification	Security mechanisms for data transmission	Format of BIRO Database	Disclosed to	Storage or retention site
NO aggregation size limit OR min aggregation N=5 patients per cell OR min aggregation N=5, only applicable for high critical privacy variables e.g. service centre, geographical site etc	BIRO partner	One Record for each aggregation level	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO OPTION 2. Client program automatically sending encrypted data (agent)	Separate sets of aggregated tables linkable by predefined statistical criteria	OPTION 1. BIRO database administrator OPTION 2. All local database administrators	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)
Aggregation across service centres									

Data Flow Questionnaire

SCENARIO 1:
Question 1. PERSONAL INFORMATION/DATA CLUSTER: DECISION 1

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
No Aggregation size limit						
Min aggregation N=5 patients per cell						
Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc						

Overall Consensus Table

A R C H I T E C T U R E 2	Personal Data Decision 1	No Aggregation size limit	3.5	4	3
		Min Aggregation N=5 patients per cell	2	3	3
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	2	4	3
	Personal Data Decision 2	Aggregation across service centres	2	2	2.5
		Data aggregated at the level of service centre	2.5	3	3
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	2	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	3	3.5	2.5
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	2	4	3
	Transmission Decision 1	All DATE fields transmitted as in original	3	3	2
		DATE fields approxim rval (e.g. months)	2		
	Transmission				

The BIRO *Data Flow Tables* (Tables 1-3) were specifically developed to describe in detail the dynamics involved in both data collection and information exchange procedures. A specific data flow table for each selected alternative was constructed to describe all personal data elements associated with the proposed systems, and aspects of the collection, use and disclosure of personal information/data that would help building a list of few, essential options available in this context. The tables were revised by all components of the PT and finally approved. The content of the data flow tables have been then used as source of information for the definition of the *Information Flow Questionnaire* (Appendix 1).

The questionnaire provided a series of scenarios, broken down into separate sub-options, for each of which marks were assigned on the basis of a set of three essential criteria: privacy, information content for diabetes, and technical complexity (feasibility).

Each option of an alternative was given a composite indicator, based on the sum of three dimensions. All scores ranged 0-5 (not applicable to very high level).

The score on privacy was based on three separate criteria: "identifiability, linkability and observability"[7] (Figure 3).

Identifiability was intended to be a measure of the degree to which information is personally identifiable. The identity measurement has been considered as taking place on a continuum, from full anonymity (the state of being without name) to full verinymity (being truly named). The goal to be pursued was to decrease as much as possible the amount of identity elements in the BIRO system.

The minimalist design approach was therefore employed in the project. Since identity data were not required for an efficient running of the BIRO information system, they were removed from the architectural equation. Many tools employing reversible and non-reversible pseudonymity are actually available for this purpose.

Linkability was conceived as a measure of the degree to which data elements are linkable to the true name of the data subject, where unlinkability meant that different records cannot be linked together and related to a specific personal identity. In this regard, complex interrelations have been taken into account, considering that record linkage can be subtle, as it may be organized and/or made possible in different ways.

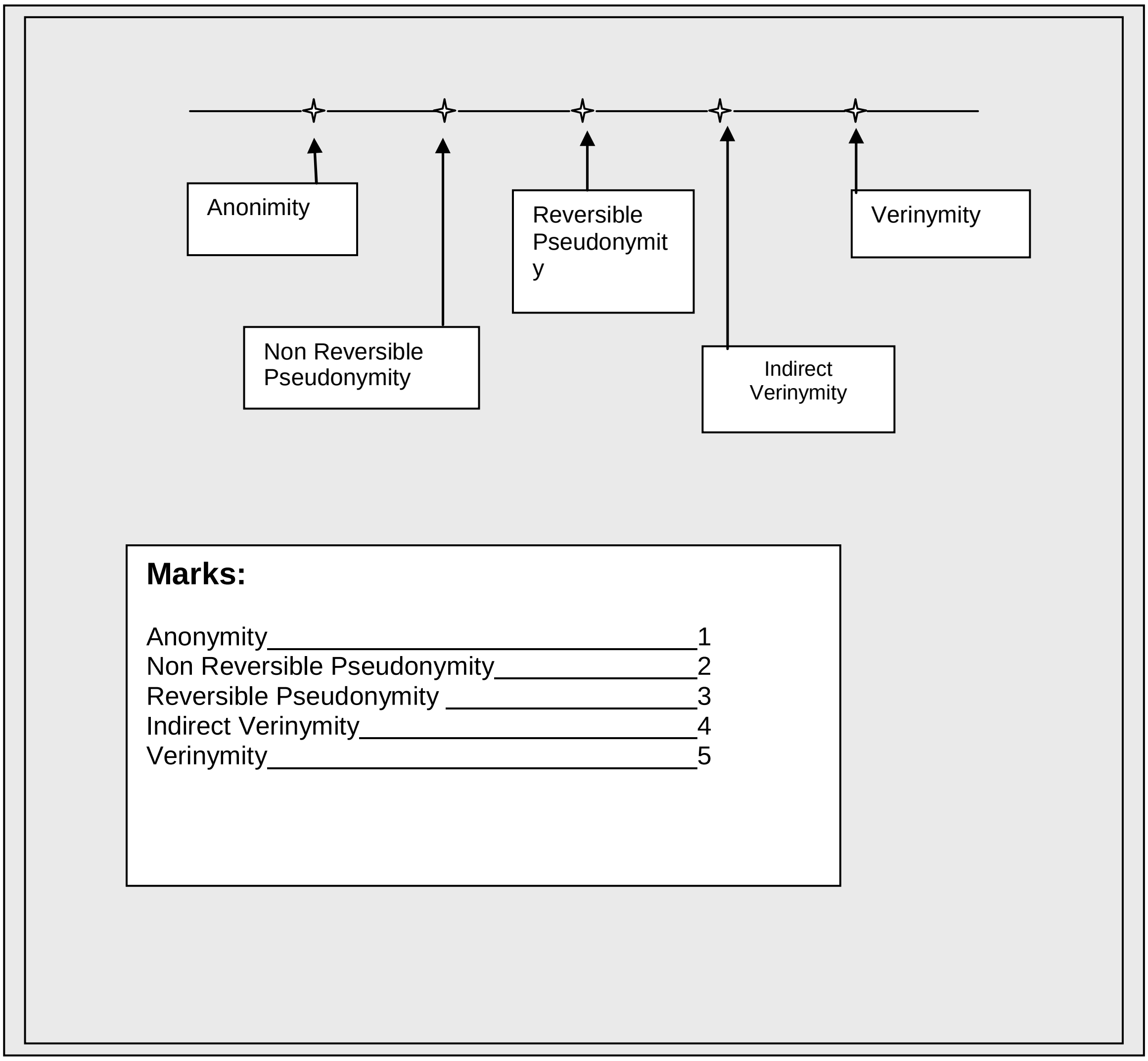
Observability was defined as a measure of the degree to which identity or linkability are affected by the use of a system. It considers, in fact, any other factor relative to data processing (time, location, data contents) that can potentially affect the degree of identity and/or linkability: an effect modifiers.

The overall privacy score for each questionnaire item was agreed by all partners to be obtained as the average of the three privacy dimensions.

Score for the *Information content* criterion was based on a single score providing a value for the level of information provided by the specific scenario/option in terms of relevance and level of evidence for diabetes.

The *technical complexity* criterion involved a single score related to the feasibility of the specific scenario/option.

Figure 3: Identifiability Metrics



Marks:

Anonymity	1
Non Reversible Pseudonymity	2
Reversible Pseudonymity	3
Indirect Verinymity	4
Verinymity	5

TABLE 1: DATA FLOW TABLES

CANDIDATE ARCHITECTURE 1: INDIVIDUAL PATIENT DATA

Description of personal information / data clusters	***Collected by***	***Type of format***	***Used by***	***Purpose of collection***	***Transmission to BIRO: de-identification***	***Security mechanisms for data transmission***	***Format of BIRO Database***	***Disclosed to***	***Storage and retention site***
SCENARIO 1: Health Service Medical Record[1]	Clinical Centres, Coordinating Centre[2]	OPTION 1 Longitudinal data collection OPTION 2 Multiple measurements averaged over time interval3	Local Health Authority, Coordinating Centre	Disease Management Program	Pseudonym used for data linkage[4], multiple measurements per patient OPTION 1. Centre IDs retained OPTION 2. Centre IDs de-identified[5]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO[6] OPTION 2. Client program automatically sending encrypted data (agent)[7]	OPTION 1. Full information on all medical records OPTION 2. Averaged over time[8]	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[9]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[10]
SCENARIO 2: Administrative Data Service Episode[11]	Local Health Authority[12]		Local Health Authority	Policy and Planning					
SCENARIO 3: Epidemiological measurement of multiple individual characteristics[13]	Research Organization[14]		Research Centre	Epidemiological Study					
SCENARIO 4.1: Health Service Medical Record + Administrative Data Service Episode	Population-based Regional/ National Diabetes Register[15]		Local Health Authority, Research Centre, Regional/National Government	Disease Management, Policy and Planning, Research					
SCENARIO 4.2: 4.1 + Epidemiological measurement of multiple individual characteristics									

TABLE 2: DATA FLOW TABLES

CANDIDATE ARCHITECTURE 2: AGGREGATION BY GROUP OF PATIENTS

Scenario 1: Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by CENTRE to compute variability of medians)[16]

Description of personal information / Data clusters	***Collected by***	***Type of format***	***Used by***	***Purpose of collection***	***Transmission to BIRO: de-identification***	***Security mechanisms for data transmission***	***Format of BIRO Database***	***Disclosed to***	***Storage or retention site***
NO aggregation size limit OR min aggregation N=5 patients per cell[17] OR min aggregation N=5, only applicable for high critical privacy variables e.g. service centre, geographical site etc[18]	BIRO partner	One Record for each aggregation level	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting[19]	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)[20]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO OPTION 2. Client program automatically sending encrypted data (agent)	Separate sets of aggregated tables linkable by predefined statistical criteria	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[21]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[22]
Aggregation across service centres[23] OR data aggregated at the level of Service Centre									
Aggregation of Multidimensional patterns (e.g. risk adjustment) NOT allowed[24] OR generally allowed[25] OR allowed with min N=5 condition applied[26]									

TABLE 3: DATA FLOW TABLES

CANDIDATE ARCHITECTURE 3: AGGREGATION BY REGION

Scenario 1: Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by REGION)[1]

Description of personal information / Data clusters	***Collected by***	***Type of format***	***Used by***	***Purpose of collection***	***Transmission to BIRO: de-identification***	***Security mechanisms for data transmission***	***Format of BIRO Database***	***Disclosed to***	***Storage or retention site***
Aggregation without restrictions OR with restrictions applied on specific stratification criteria (e.g. geographical variable, centres etc)	BIRO partner	One Record for each aggregation level by REGION	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)[2]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO OPTION 2. Client program automatically sending encrypted data (agent)	Separate sets of aggregated tables linkable by predefined statistical criteria	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[3]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[4]
Geographical mapping available[5] OR Unavailable									
Variability of Centres' Outcomes Available[6] OR Unavailable									
Aggregation by multidimensional patterns (e.g. risk adjustment) NOT allowed OR allowed without restrictions applied on specific stratification criteria OR allowed with restrictions applied on specific stratification criteria[7]									

EXPLANATORY NOTES TO TABLES 1-3

1. Data collected during medical examinations according to a structured procedure within a health service framework e.g. disease management program, systematically organized by means of an electronic database

2. Clinical centres may be coordinated by a local institution in the framework of a structured program e.g. disease management

3. For simplicity, data relative to the same subject can be amalgamated over a period of time in various ways. For instance, one may just retain the last measurement of Hba1c or compute the average of different measurements over n months. All other original data for the same variable are not retained. The process is systematically repeated, and the individual record updated or a new individual record appended to the previous for each new time interval.

4. Individual identifier is replaced by a unique, fake identifier created via an algorithm applied by the local database administrator.

5. Same process applied to de-idetified the individual subject is used for clinical centres. Other characteristics that can lead to identify any centre can be blinded, e.g. absolute frequencies are not retained and only percentages are sent to the BIRO central engine

6. Database administrator may decide when to send structured encrypted data bundles to the BIRO server, using ad hoc client software.

7. The client program automatically sends data packets to the BIRO central engine, based on a routine that activates according to a schedule agreed by the database administrator.

8. Information on individual data may be stored averaged over a predetermined time interval

9. Privileges to access pooled data may be extended to all local BIRO database administrators.

10.European Commission may be in charge of the maintenance of the permanent BIRO Central server

11.Data originated by administrative data flows e.g. hospital discharges, pharmaceutical, mortality data etc.

12. Local government ruling collection of administrative data. In the framework of the present document, a region is intended as a geographical area or even a cluster of geographical areas characterized by homogeneous criteria for data collection. For instance, Tayside may be recognised as a specific region. However, Scotland applies the same basic set of definitions for data collection, so the BIRO Consortium may even consider the wider geographical area as a single region.

13. Clinical, demographic and socio-economic characteristics of subjects studied in a epidemiological investigation

14. Institution conducting the epidemiological investigation

15. Typically, a regional population-based register involves linkage of different data flows, including general administrative data and medical records more targeted at the diabetes population.

16. Aggregated tables strictly relate to the construction of a statistical quantity. For this reason we can also call them as “statistical objects”, as each table is required to apply a particular statistical procedure. For instance, computing the average may only require the total sum of a specific variable, e.g. Length of Stay (LOS), plus the total number of observations related to that sum. A “bundled” table including both entities is a statistical object that can lead to the actual statistical parameter in a subsequent step (central server), where the formula AvLOS=Total (LOS)/n(OBS) is applied. The step is not always so immediate. To compute the median LOS, one requires the entire frequency distribution of LOS at each site/region, i.e. n(OBS) for each level of LOS. The median for all sites/regions is computed from the sum of all frequency distributions collected.

17. Small groups of subjects may lead to the identification of subjects/centres/regions etc. For instance the number of subjects aged 90+ or living in a specific geographical area may be so small and well known that all characteristics stored in tables may be indirectly linked to the specific individual/centre.

18. Since the criterion may be too strict for all variables included in the database, it may be only applied to specific characteristics that are more sensitive to privacy issues.

19. Tables can be used either to carry out reports for the individual region and/or to compute overall results for the BIRO collaboration

20. Dates pose a specific threat to privacy, as it can be very unlikely that same service or individual characteristic occurs at the same time for different individuals. Therefore it can be an option to approximate dates by weeks or months.

21. Privileges to access pooled data may be extended to all local BIRO database administrators.

22. European Commission may be in charge of the maintenance of the permanent BIRO Central server

23. Publication/exchange of tables stratified by health service centre - as in the case of league tables of performance indicators - is a specific condition affecting “institutional privacy” towards which policy makers can be particularly sensitive. A sharp decision in this regard may involve the restriction to publish all results without using centres as a specific level of aggregation.

24. Risk adjustment techniques may work even without exchanging individual data using different solutions (e.g. pooling multidimensional patterns in logistic regression). However, patterns may lead to very fine stratifications that can pose threats to privacy via indirect identification (low frequencies in specific cells of crosstabulations).

25. Risk adjustment techniques may work even without exchanging individual data using different solutions (e.g. pooling multidimensional patterns in logistic regression). However, patterns may lead to very fine stratifications that can pose threats to privacy via indirect identification (low frequencies in specific cells of crosstabulations).

26. Min N condition may provide a solution to control privacy in sparse cells

27. Aggregated tables strictly relate to the construction of a statistical quantity. For this reason we can also call them as “statistical objects”, as each table is required to apply a particular statistical procedure. For instance, computing the average may only require the total sum of a specific variable, e.g. Length of Stay (LOS), plus the total number of observations related to that sum. A “bundled” table including both entities is a statistical object that can lead to the actual statistical parameter in a subsequent step (central server), where the formula AvLOS=Total (LOS)/n(OBS) is applied. The step is not always so immediate. To compute the median LOS, one requires the entire frequency distribution of LOS at each site/region, i.e. n(OBS) for each level of LOS. The median for all sites/regions is computed from the sum of all frequency distributions collected.

28. Dates pose a specific threat to privacy, as it can be very unlikely that same service or individual characteristic occurs at the same time for different individuals. Therefore it can be an option to approximate dates by weeks or months.

29. Privileges to access pooled data may be extended to all local BIRO database administrators.

30. European Commission may be in charge of the maintenance of the permanent BIRO Central server

31. Geographical characteristics can be highly informative and useful for both epidemiological and policy purposes, but they are prone to privacy issues, as they can link to both the individual and the health service centre.

32. Even though centres’ tables are not made available, one may choose to exchange/publish overall variability of target indicators across centres. For instance, range of performance indicators, or standard deviations. However, these can disclose elements of performance across the region that policy makers may regard as jeopardising institutional privacy.

33. At the level of region, min N=5 may not be considered relevant, so other criteria may be applied.

3.2.3. Results: Selection of the Best Alternative

The Delphi procedure was used to reach consent among the BIRO project partners over the selection of the most privacy protective Information System Architecture, which was one of the main objectives of the BIRO PIA. The use of the Delphi methodology in the context of privacy impact assessment procedures is certainly innovative. However, the PIA Team agreed that the Delphi procedure represented the most scientifically sound methodology to fulfil the above objective and, at present, there is not in the literature any adverse indication of using it in the context of PIAs.

The questionnaire was distributed by email to all members of the PT to initiate a *modified Delphi procedure,* including two phases: in the first, each member of the PT assigned marks independently from remote. In a second phase, the panel met to carry out an interactive consensus process, chaired by the PF, aimed at converging towards the best architecture.

The Delphi consensus session took place in Cyprus during the 2nd BIRO Investigator Meeting (23-25 May 2007). Initial scores provided independently by members of the PT were collected and discussed in order to reach an agreement on common criteria.

The selection process involved value judgements over different options for each criterion, requiring specific expertise. For each case, relevant experts explained the content and meaning of the option, motivating their marks. Member of the panel were also given the opportunity to make questions, allowing a completely informed consensus process to be finalized.

The Delphi panel finally assigned marks for all options, as reported in the Overall Consensus Table (TABLE 4).

The selected mix of best scoring options allowed the identification of the best BIRO System architecture, classified as *"Aggregation by group of patients"* (Table 5), where grouping conditions are directly linked to the construction of the particular statistical object required to deliver the overall diabetes report.

Figure 4 shows the resulting B.I.R.O. system architecture, whose criteria were duly taken into account for implementation.

Statistical properties (e.g. those of the arithmetic mean, percentiles, etc.) were exploited to transmit target objects in separate bundles over the network, so that international reports avoid many potential risks and restrictions imposed by privacy legislation, with no exchange of individual records.

Specialized communication software has been developed to securely transmit statistical objects as encrypted compressed folders containing comma-delimited text files (.csv).

Security has been addressed comprehensively according to ISO/OSI 7498-2. For authentication, digital certificates trusted by a common certification authority were exchanged and installed in sender and receiver. Access control was configured such that only trusted identities were authorized to connect to services. Security mechanisms were also implemented through the use of encryption techniques. Data integrity as well as non-repudiation were provided by digital signatures.

Web services were selected as the core technology for communication for their compliance with standards set by the open World Wide Web consortium: SOAP (Simple Object Access Protocol) for messaging, HTTP (Hypertext Transfer Protocol) for Internet transport and XML (eXtensible Markup Language) together with its security extensions XMLenc (encryption) and XMLsig (digital signatures). Apache Axis 2, together with Apache Rampart provided by Java 2 Enterprise Edition, were chosen for pilot development and configuration of sending and receiving applications.

Encryption and digital signatures were applied on two layers. Firstly, transport layer security using HTTPS, i.e. HTTP protocol together with SSL (Secure Sockets Layer), was used to protect the entire data stream exchanged between sender and receiver. Secondly, on the data layer, individual chunks of data were encrypted and digitally signed, giving the application full control over further utilization, storage and processing of digital signatures and other security related information.

The whole B.I.R.O. process is controlled by integrated software linking the different modules through a simple graphical user interface (GUI). A ”local” module is used to allow users exporting local data to XML files, to add them to a local database, and to produce local reports and statistical objects for the central B.I.R.O. System. A “central” module is used by the server administrator to load statistical objects received from partial analyses in the form of csv files, and to run the overall analysis for the global B.I.R.O. report.

The B.I.R.O. architecture requires for the Central Engine to be managed by a unique administrator, ensuring compliance with all national and international security rules in the maintenance of the server, as specified in the Preliminary PIA Report[8].

Results are stored in a server database that will be connected to a web portal in charge of delivering online results to the masses, bundled with proper data definitions and methodological references.

Table 4: Overall Consensus Table

A.	Category	Option	P.	I.C.	T.C.
ARCHITECTURE 1	SCENARIO 1	One record for each service episode, centre IDs retained	5	5	3
		One record for each service episode, Centre IDs De-Identified	4	4	3
		Multiple measurements averaged over time interval, centre IDs retained	4.5	4	3
		Multiple measurements averaged over time interval, Centre IDs De-Identified	4	3	3
	SCENARIO 2	Population-based longitudinal records, linked across administrative datasets, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	5	5	3
		Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	4	4	3
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	3	3
	SCENARIO 3	Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients	3	3	3
	SCENARIO 4.1	Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	5	5	3
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	4	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	3	3
	SCENARIO 4.2	Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage overmultiple datasets, all relational structure sent to BIRO	5	5	3
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	4.5	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4.5	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	4	3
ARCHITECTURE 2	Personal Data Decision 1	No Aggregation Size Limit	3.5	4	3
		Min aggregation N=5 patients per cell	2	3	3
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**2**	**4**	**3**
	Personal Data Decision 2	Aggregation across service centres	2	2	2.5
		Data aggregated at the level of service centre	**2.5**	**3**	**3**
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	2	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	3	3.5	2.5
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	**2**	**4**	**3**
	Transmission Decision 1	All DATE fields transmitted as in original	3	3	2
		DATE fields approximated to time interval (e.g. months)	**2**	**3**	**2**
	Transmission Decision 2	Service Centre ID transmitted	3.5	3	2
		Pseudonym used for service centre	**2**	**2.5**	**2**
ARCHITECTURE 3	Personal Data Decision 1	NO restrictions on specific stratification criteria (e.g. geographical variable, centres, etc)	2	1	1
		Restrictions applied on specific stratification criteria (e.g. geographical variable, centres, etc)	1	1	2
	Personal Data Decision 2	Geographical mapping available	2	3	2
		Geographical mapping unavailable	1	1	1
	Personal Data Decision 3	Variability of centres outcomes available	2	3	3
		Variability of centres outcomes unavailable	1	1	1
	Personal Data Decision 4	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	1	1	1
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITHOUT restrictions applied on specific stratification criteria	3	3	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITH restrictions applied on specific stratification criteria	2	2	3
ALL	Security	**Password access for local administrator prompting client program to send encrypted bundles to BIRO**	**2**	**0**	**2**
		Client program automatically sending encrypted data (agent)	1	0	4
	Format	Full information on all medical records	4	5	3
		Averaged over time	**2**	**3**	**2**
	Disclosure	**BIRO database administrator**	**1**	**0**	**1**
		All local database administrators / registry managers	3	0	2
	Storage/Retention	**BIRO Coordinating Centre**	**2**	**0**	**2**
		EU/DG-SANCO	1	0*	3

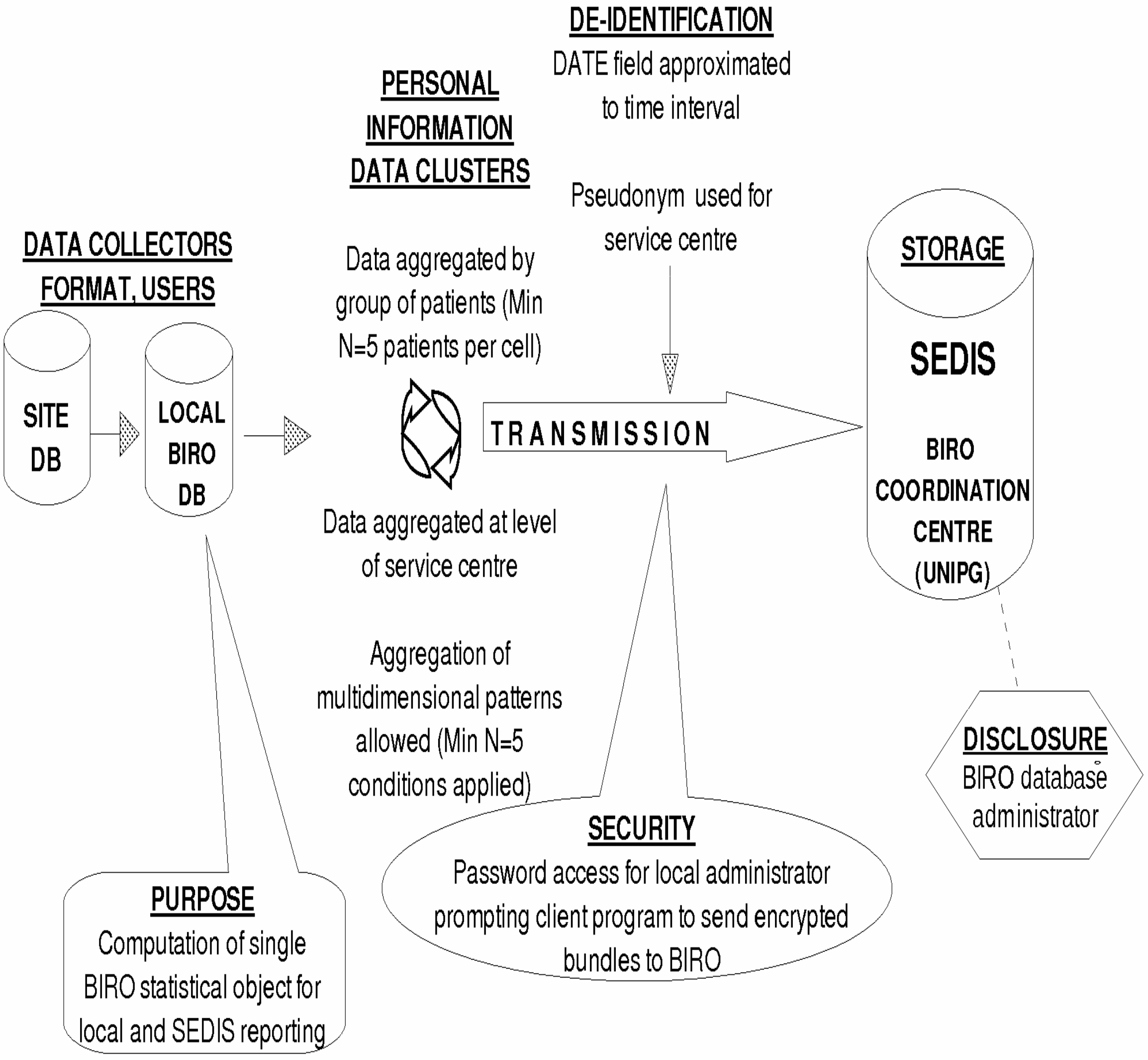
TABLE 5: BIRO DATA FLOW TABLE

BIRO ARCHITECTURE: AGGREGATION BY GROUP OF PATIENTS

Grouping condition directly set by statistical object (e.g. ordered frequency distribution of LOS by CENTRE to compute variability of medians)

Description of personal information / Data clusters	***Collected by***	***Type of format***	***Used by***	***Purpose of collection***	***Transmission to BIRO: de-identification***	***Security mechanisms for data transmission***	***Format of BIRO Database***	***Disclosed to***	***Storage or retention site***
Aggregation by group of patients: min aggregation N=5, only applicable for high critical privacy variables e.g. service centre, geographical site etc	BIRO partner	One Record for each aggregation level	BIRO partner (local engine), BIRO Consortium (central engine)	Computation of single BIRO statistical object for local and SEDIS reporting	DATE fields approximated to time interval (e.g. months) Pseudonym used for service centre	. Password access for local administrator prompting client program to send encrypted bundles to BIRO	Separate sets of aggregated tables linkable by predefined statistical criteria	BIRO database administrator	BIRO Coordinating Centre
Data aggregated at the level of Service Centre									
Aggregation of Multidimensional patterns (e.g. risk adjustment) allowed with min N=5 condition applied									

FIGURE 4: B.I.R.O. Architecture


DE-IDENTIFICATION
DATE field approximated to time interval
Pseudonym used for service centre
PERSONAL INFORMATION DATA CLUSTERS
Data aggregated by group of patients (Min N=5 patients per cell)
DATA COLLECTORS FORMAT, USERS
SITE DB
LOCAL BIRO DB
TRANSMISSION
Data aggregated at level of service centre
Aggregation of multidimensional patterns allowed (Min N=5 conditions applied)
STORAGE
SEDIS
BIRO COORDINATION CENTRE (UNIPG)
DISCLOSURE
BIRO database administrator
SECURITY
Password access for local administrator prompting client program to send encrypted bundles to BIRO
PURPOSE
Computation of single BIRO statistical object for local and SEDIS reporting

3.2.4 Engineering of the Best Alternative

The selected BIRO Architecture is defined by three consecutive steps, logically organized in two different parts: *local* and *global* (Fig. 4).

The *local* part of BIRO Architecture includes the set of software tools required by each collaborating centre to undertake two basic operations:

1) produce a standardized BIRO local report; and
2) transmit data to the BIRO server for the production of the global report.

Step 1 involves ***client data processing and statistical analysis****.*

A BIRO "*Adaptor"* is used to establish a connection to the local database and export data from any format used by the local diabetes register to the standardized format complying with the agreed specifications of the BIRO common dataset. Standardized specifications (XML Schema) have been specifically developed to implement common BIRO definitions into a uniformly defined database that allows using and pooling data from different centres using the same common format. A "*Metadata Dictionary*" has been realized in XML to incorporate many concepts and derive new variables from the original ones into the BIRO dataset.

The flat text file (XML export) is thus produced by each centre using Java and JDBC driver. This operation needs some basic pre-processing of local data to comply with basic requirements (one record for each individual subject, i.e. the so-called "Merge Table"). A configuration file is needed by the BIRO Adaptor to apply specific options to the relevant driver (this operation will be further simplified using a user friendly visual application).

The BIRO "*Database Manager*" reads the XML files and stores data into a local (Postgres) database that is used to organize local data in an optimal way, so that they could be automatically processed by the statistical engine. The Java language and tools e.g. Castor and Hibernate are used for the scope. A configuration file is needed for the scope.

The BIRO "*Statistical Engine*" connects to the local BIRO Postgres database and runs statistical functions creating "*statistical objects"*. A statistical object is defined as "an element of a distributed information system that carries essential data in the form of embedded, partial aggregate components, required to compute a summary measure or relevant parameter for the whole population from multiple sites". The definition is central to the functioning of BIRO, as it allows using pre-determined datasets as basic elements of statistical analysis run on top of aggregate data to product individual centre reports, and transmitted over the network for the production of global reports. This solution allows bypassing many possible risks and restrictions imposed by privacy legislation, as defined by the best architecture, avoiding the exchange of individual records.

Basically, statistical objects are tables that contain statistical aggregations of local data (arithmetic mean, percentile, variance, linear and logistic regression, bar plot data, histogram data, box pot data, etc), stored as flat text comma delimited files (CSV). Statistical objects are organized according to a dictionary that includes basic components of frequency tables, measures of location, measures of dispersion, graphical elements, regression, and standardization. Criteria agreed by the Delphi panel for the definition of the best architecture are duly taken into account in the specifications of statistical objects.

A *report template* has been developed to precisely define all outputs produced by the statistical engine. The same structure is used to automate the production of both the individual centres and the global BIRO reports. This feature is optimal as it allows using the same set of basic statistical functions for multiple, repeated applications.

The statistical engine connects to the local database using the open source statistical R software with proper Postgres drivers. According to the specifications given by the report template and the associated relevant definitions of the statistical objects, it processes the database to deliver statistical objects in the form of small CSV datasets, further processed to produce individual centre outputs and complete local reports in pdf and html formats, using the Latex software.

A compressed CSV folder is created to include all statistical objects produced by each run of the local reporting system, classified by date and centre id. This operation completes step 1) of the local engine.

*Step 2 involves **data transmission**.*

A dedicated communication software has been developed to securely transmit the CSV folder including statistical objects between the local and the Central BIRO system, as described in paragraph 3.2.3.

The *central* part of BIRO Architecture includes the set of software tools required by the BIRO server to undertake Step 3: **global statistical analysis.**

*Step 3 involves several operations including **database processing and statistical analysis**.*

At the central level individual data are no longer required and the BIRO system only deals with aggregate data, so all database specifications require to meta-data mainly referred to the concept of statistical objects.

A specialised application (BIRO CSV Importer) has been developed in Java to read CSV files embedding statistical objects and store them as separate tables of the Central BIRO database. As for the Adaptor and Database Manager, a configuration file is required to allocate proper options.

Same statistical objects, transmitted by separate centres, are appended to the same table to form a global collection of local aggregate data.

The BIRO Central Engine is specifically developed to load and organize all central aggregate data, and perform some basic data processing. In particular, elementary Postgres functions are used to compute a "*cumulative component*" for each statistical object as a pooled estimate of multiple "*local*" statistical objects.

Advanced statistical analysis is performed by R functions that are also included in the Central Engine. Cumulative components of statistical objects are processed to deliver all elements of the global report required to deliver the same template used for the local analysis, populated with results that now refer to the whole universe of BIRO collaborative centres.

Outputs of the Central Engine include a complete pdf report (as defined in the template), an html report (following specifications in the web portal), and CSV data, all produced using R and Latex software.

The final section of software development involves integration of the BIRO architecture into unique common software.

The BIRO process will be triggered by a simple "*local*" user friendly (GUI) application that will allow the user to:

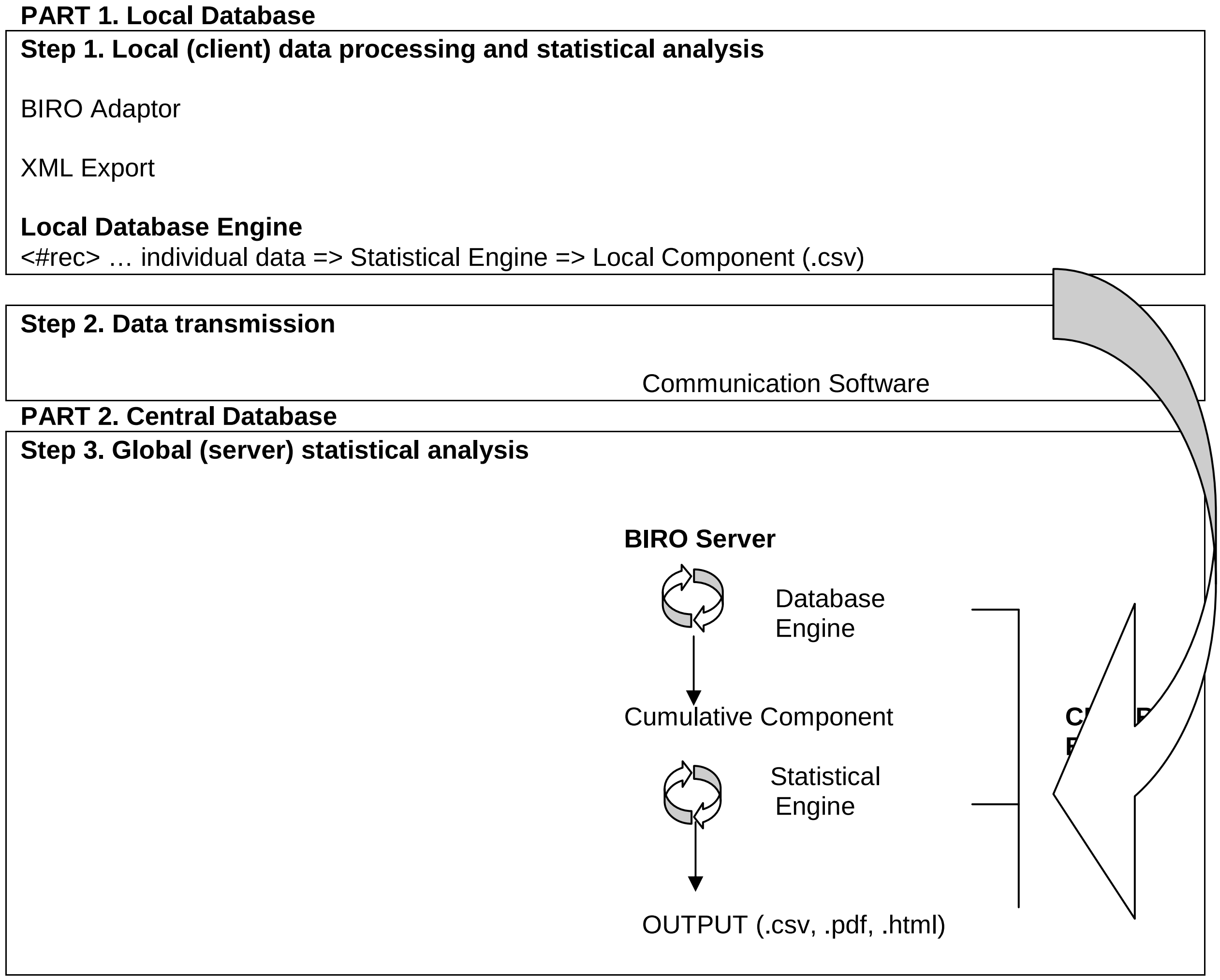
- export local data stored into a local database to XML files running the BIRO Adaptor
- import XML files to the local database using the BIRO Database Manager
- produce the local statistical report
- send the local statistical objects to the Central BIRO System

A "central" GUI application will allow the user to:

- import statistical objects stored as csv files
- run the global statistical analysis
- produce the global BIRO report

The BIRO architecture will require for the Central Engine to be managed by the BIRO coordinator, which would evidently ensure compliance with all national and international security rules for the maintenance of the server, as specified in PIA Report Step 1.

Figure 4: BIRO Software Engineering



PART 1. Local Database

Step 1. Local (client) data processing and statistical analysis

BIRO Adaptor

XML Export

Local Database Engine
<#rec> … individual data => Statistical Engine => Local Component (.csv)

Step 2. Data transmission

Communication Software

PART 2. Central Database

Step 3. Global (server) statistical analysis

BIRO Server

Database Engine

Cumulative Component

Statistical Engine

OUTPUT (.csv, .pdf, .html)

3.3 Privacy Analysis

3.3.1 Legislative Framework

Of all the human rights in the international catalogue, the **right to privacy** is perhaps the most difficult to define[9].

Definitions of privacy vary widely according to contexts and environments. Nevertheless, privacy is usually seen as the way of drawing the line of how far a society can intrude into a person's private life.

Privacy has been defined as the "right to be left alone"[10]; or as "the right of the individual to be protected against intrusion into his personal life or affairs, or those of his family, by direct physical means or by publication of information[11].

Although there is a lack of a single definition of privacy, it is a right generally recognized around the world and crystallised in many international instruments.

The 1948 *Universal Declaration of Human Rights* was the first international binding instrument to recognise privacy as a human right, specifically protecting territorial and communication's privacy[12]. Article 12 states: *"No one should be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks on his honour or reputation. Everyone has the right to the protection of the law against such interferences or attacks".*

In addition, numerous international human rights treaties specifically recognize privacy as a right. The International Covenant on Civil and Political Rights (ICCPR – art. 17)[13]; the UN Convention on Migrant Workers (Article 14)[14], and the UN Convention on Protection of the Child (Article 16)[15] adopt the same language. On the regional level, various treaties make these rights legally enforceable.

For instance, Article 8 of the *European Convention for the Protection of Human Rights and Fundamental Freedoms* (1950)[16] states that *"Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except as in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health of morals, or for the protection of the rights and freedoms of others".*

The Convention created the European Commission of Human Rights and the European Court of Human Rights to oversee enforcement. Both have been active in the enforcement of privacy rights, and have consistently viewed Article 8's protections expansively and interpreted the restrictions narrowly[17].

The Court has reviewed Member States' laws and imposed sanctions on numerous countries[18]; and has also reviewed cases of individuals' access to their personal information in government files to ensure that adequate procedures exist[19]. In the evolution of data protection, the interest in the right of privacy increased in the 1960s and 1970s with the advent of information technology.

The surveillance potential of powerful computer systems has increased the demand for specific rules governing the collection and handling of personal information.

Two crucial international instruments in the evolution of data protection are the Council of Europe's (1981) Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data[20], and the Organization for Economic Cooperation and Development's (OECD) Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data[21], which set out specific rules covering the handling of electronic data.

These rules describe personal information as data that have accorded protection at every step: from collection to storage and dissemination.

As a matter of fact, the above-mentioned agreements have had a profound effect on the enactment of laws around the world. Nearly thirty countries have signed the COE Convention; and the OECD guidelines have been widely used in national legislations, even outside the OECD member countries. The development of privacy protection in the EU took a step forward with the Council of Europe Convention on Human rights and Biomedicine (Oviedo 1997), which reinforced the principles that everyone is entitled to the right to privacy and confidentiality of personal medical data and the right to be informed about his/her health[22].

Finally, the *Charter of Fundamental rights of the European Union* (2000/C 364/01) [23] specifically provides protection of personal data. Art 8 states: "*Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority".*

The Charter of Fundamental Rights has been fully incorporated in the *European Constitution* (forming its part II)[24], signed in Rome on the 29th of October 2004. Although the Parliament, the Council and the Commission solemnly proclaimed the Charter on the 8th of December 2000, the Charter was not part of the Union's Treaties and therefore it had no binding legal force. The Constitution thus achieved a major breakthrough, which allows the Union to have its own catalogue of rights, binding for all European countries and enforceable through the Court of Justice, which will in fact ensure that the Charter will be adhered to.

It is worth noting that the content of the Charter is broader than that of the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)**,** signed in Rome on 4 November 1950 and ratified by all the Member States of the Union.

Whereas the ECHR is limited to civil and political rights, the Charter of Fundamental Rights covers other areas such as the right to good administration, the social rights of workers, the protection of personal data and bioethics.

Finally, The *Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research* (2005)[25] further reinforced the duty of confidentiality in the handling of personal information in health research and reaffirmed the obligation to treat them according to the rules relating to the protection of private life.

In line with all the aforementioned instruments, the EU has adopted a privacy legislative model that embraces comprehensive laws. The model is based on a general and abstract law that governs all aspects of the handling of personal information: from collection to use and dissemination, by both the public and private sectors.

The 1995 Data Protection Directive (95/46/EC)[26] set up a common level of privacy among European countries, ensuring compliance through the establishment of a regulatory body.

The Directive not only reinforced current data protection laws, but also established a range of new rights and basic principles, namely: the right to know where the data originated, the right to have inaccurate data rectified, a right of recourse in the event of unlawful processing, and the right to withhold permission to use data in some circumstances. The Directive contains strengthened protections over the use of sensitive data.

Art 7 of the Directive establishes a set of criteria of "legitimate processing". Processing, in order to be legitimate, has to take place: either with the unambiguous consent of the data subject, or where this is necessary for the performance of a contract with the data subject, for compliance with a legal obligation, or for the performance of a government task, just to mention a few examples.

More stringent conditions apply to the processing of special categories of *sensitive data*, such as medical data. Here, the processing of sensitive data is considered, in principle, not legitimate and Member States has to prohibit their processing, unless special conditions verify.

According to art. 8, the processing of sensitive data is allowed when:

- the data subject has given his explicit consent to the processing of those data, or
- processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or

- processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
- processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or
- the processing relates to data which are manifestly made public by the data subject or is necessary for the establishment, exercise or defence of legal claims.

Importantly, the prohibition of Article 8 (1) shall, according to Article 8 (3), also not apply where the data are required: for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject under national law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of secrecy.

Moreover, Member States may, according to Article 8 (4), for reasons of substantial public interest, lay down exemptions, in addition to those laid down, either by national law or by decision of the supervisory authority.

Art. 8(3) is extremely important for the health sector, since it justifies the collection, use, and processing of health data, for the specified purposes, without the patient's consent.

Although the free and informed consent will be necessary if, for instance, those data would be further used for research purposes or any other secondary use. The reference to professional secrecy contained in art. 8 (3) is crucial for obtaining a more effective protection of privacy in the handling of sensitive health data.

Although the issues surrounding the confidentiality of health data are not fully dealt with in the Directive, the referral to the obligation of confidentiality in the Directive represents a step forward towards an eventual harmonization of European legislations. At least, it imposes to Member States, in a binding form, the *duty of confidentiality* to any person involved in the processing of personal sensitive data, such as health data.

The duty of confidentiality has its origins in the duty of professional secrecy incumbent on health professionals either through a law or code of conduct. The principle of confidentiality of medical information, derived by the Hippocratic Oath, can be considered one of the oldest principles applying to data protection. Although privacy and confidentiality are conceptually distinct, they are strictly interrelated and need to be consistently implemented among European countries in order to enhance the protection of privacy when sensitive data are involved: as a matter of fact, confidentiality could rather be conceived as a means to protect the right to privacy.

In order to conduct scientific research without falling under the binding rules of the Directive, data should be rendered *anonymous*. Recital 26 of the EU Data Protection Directive in fact states that "principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable".

Recital 26 thus places outside the scope of the Directive the discipline of data processed for research purposes when both direct and indirect identification is avoided. Direct identification should be interpreted as identification from the data itself and indirect identification as identification from the data itself matched with any other data or means that are reasonably likely to be used, such as an identification number or to one or more factors specific to the subject's physical, physiological, mental, economic, cultural or social identity[27]. For instance, coded and encrypted data are not considered anonymous "per se". If decoding or de-encrypt techniques are still possible without an unreasonable effort. In this circumstance, data shall be still subjected to the Directive rules[28].

Importantly, the 1995 Directive imposes an obligation on member states to ensure that the personal information relating to European citizens has the same level of protection when it is exported to, and processed in, countries outside the EU. As a result, countries refusing to adopt adequate privacy protections may find themselves unable to conduct certain types of information flows with Europe, particularly if they involve sensitive data.

In line with the EU Data Protection Directive, the Council of Europe enacted, in 1997, a Recommendation on the Protection of Medical Data: **Council of Europe Recommendation No. R (97) 5** [29]. The Recommendation acknowledges that medical data requires even more protection than other non-sensitive personal data, reaffirming that the respect of rights and fundamental freedoms, and in particular of the right to privacy has to be guaranteed during the collection and processing of medical data.

For these reasons, Principle 3.2 recalls the requirement in Article 6 of the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) for appropriate safeguards in the law, in so far as the various stages of collection and processing of medical data are concerned.

According to the Recommendation, the processing of medical data is, in principle, prohibited, unless appropriate safeguards are provided by domestic law.

One of such safeguards is that only health-care professionals, bound by rules of confidentiality, should collect and process medical data, or where necessary persons acting on behalf of health-care professionals, as long as such persons are subject to the same or equivalent rules.

Since the definition of health professional may vary across different countries, the recommendation provides for the possibility that personnel not directly responsible for health care may collect and process medical data; but only on the condition that this category of professionals must abide by confidentiality rules comparable with those imposed on health-care professionals, or that domestic law provides for appropriate safeguards which are as efficient as confidentiality rules, that is, they are efficient enough to guarantee the respect of privacy of the data subject. The Recommendation in fact strengthens the duty of confidentiality within European countries.

Once again, with a view to the sensitive nature of medical data, Principle 4.1 recalls the provisions in Article 5 of the Convention: the collection and processing of medical data must be fair and lawful, and for specific purposes only.

The principle of fair collection is made more explicit in Principle 4.2: medical data must, in normal conditions, be obtained from the data subject himself/herself. This principle therefore concerns the "disclosure" of these data by the data subject himself/herself, and not "communication" of medical data by a third party (for example, the doctor).

Principle 4.3 lays down the rules governing the collection or processing of medical data. The latter may be collected or processed: if it is provided for by law, there is a contractual obligation to do so, if this is necessary for the establishment of a legal claim or if the data subject has given his/her consent. Principle 4.3 does not constitute a derogation from Principle 3.2, but sets conditions for the legitimacy of the collection or processing.

Medical data may also be collected from the data subject or from other sources if this is provided for by the law for one of the purposes set out in Principle 4.3(a): for public health reasons, the prevention of a real danger or the suppression of a specific criminal offence, or another important public interest.

Furthermore, medical data may be collected and processed if permitted by law for the purposes set out in *Principle 4.3 (b): for preventive medical purposes or for diagnostic or therapeutic purposes (in this case data may also be processed for the management of medical service operating in the interest of the patient), or to safeguard the vital interests of a data subject, or with a view to respecting specific contractual obligations, or with a view to the establishment, exercise or defence of a legal claim.* Thus, Principle 4.3 (b) reaffirms the rules set forth in the EU Data Protection Directive.

In accordance with principle 4.3 (c), medical data may also be collected and processed if the data subject has given his/her consent for one or more purposes in so far as domestic law does not provide otherwise.

Medical data may therefore be collected without consent, if the law provides for this, "for the purposes of" (that is, in the interest of) public health; this purpose is in line with the derogation for reasons of public safety in Article 9 of the Convention. It should also be noted that the words "in the interest of public health" include the management of health services.

One of the means to ensure that medical data are obtained and processed fairly and lawfully is to inform the data subject, whose data are collected, of a number of elements (information to be given to the data subject). These elements are listed in Principle 5.1.

It is obvious that such provision of information is indispensable when the data subject is required to give his/her "informed" consent. But even in cases where his/her consent is not required - that is, when the collection and processing of medical data follow an obligation under the law or under a contract, are provided for or authorised by law, or when the consent requirement is dispensed with - the recommendation provides that the data subject is entitled to relevant information.

Although Principle 5.1 should be interpreted strictly, two kinds of derogation are admitted.

First of all, Principle 5.6 allows for derogations to be made for certain reasons of public interest, for protection of the data subject or a third person, or in medical emergencies.

Secondly, information on the various elements listed in the principle has to be supplied only in so far as it is relevant.

Principle 5.1 identifies the following elements on which the data subject must be informed:

- the existence of a file containing his/her medical data and the type of data collected or to be collected;
- the purpose or purposes for which they are or will be processed;
- where applicable, the individuals or bodies from whom they are or will be collected
- the persons or bodies to whom and the purposes for which they may be communicated
- the possibility, if any, for the data subject to refuse his consent, to withdraw it and the consequences of such withdrawal;
- the identity of the controller and of his/her representative, if any, as well as the conditions under which the rights of access and of rectification may be exercised.

One of the conditions on which medical data may be collected and processed is that the data subject has given his/her consent**,** in so far as he/she is capable of doing so. As these data are regarded as sensitive data, Principle 6.1 requires that the consent be "free, express and informed". Consent is "informed" if the data subject is informed in particular of the purposes involved and the identity of the data controller. Consent is "free" if the data subject has the possibility to refuse his/her consent, to withdraw it or to modify the terms and conditions of consent. Consent can be expressed orally or in writings.

However, under certain conditions, medical data could be processed without the data subject's free, express and informed consent. These conditions are listed exhaustively in the recommendation.

As regards the collection of medical data in the course of a consultation or treatment for preventive, diagnostic or therapeutic purposes by a doctor, and which the data subject has freely chosen, the consent of the patient may not need to be expressed if the data were indeed to be processed only for the provision of care to the patient. This is also valid for processing medical data in the context of the management of a medical service operating in his/her interest.

The recommendation reaffirms the right of access: every person has to be enabled to have access to his/her medical data, either directly or through a health-care professional. Importantly, art. 8 (1) of the Recommendation states that the information must be provided to patients “in understandable form”. Access to medical data may be refused, limited or delayed only if the law provides for this.

The data subject has also the right to rectification: patients may ask for rectification of erroneous data concerning him/her and, in case of refusal, he/she has to be able to appeal.

In general, medical data shall be kept no longer than necessary to achieve the purpose for which they were collected and processed (conservation).

Although the Recommendation does not refer to it explicitly, the requirement in Article 5 of the Convention that personal data undergoing automatic processing should be adequate, relevant and not excessive applies equally to medical research. It means that only the data necessary for the purposes of such research should be used.

The primary means of protecting medical data to be used for scientific research purposes, are to make them anonymous. For this reason, researchers as well as public authorities concerned should develop anonymisation techniques, which should be continuously updated and kept efficient. The nature or objectives of certain research projects sometimes make it impossible to use anonymous data. In such cases, under Principle 12.2, personal data may be used if the purposes of the research project are legitimate and one of the listed conditions is fulfilled.

Firstly, personal data may be used for medical research if the data subject has been duly informed of the research project - or at least if the information requirements have been respected - and has given his/her consent for that particular project, or, at least, for the purposes of medical research.

Secondly, in the case of a legally incapacitated person, this consent must have been given in accordance with Principle 6.4, and the research project must have a connection with the medical condition or disease of the data subject (sub-paragraph b). This is provided to avoid that consent given on behalf of a legally incapacitated person might be motivated by material interests.

Thirdly, cases may arise where the data subject cannot be found or where for other reasons it is apparently impossible to obtain consent from the data subject himself/herself (for example, in the case of an epidemic). When in such cases the interests of the research project are such that they justify the consent requirement to be waived - for example in the case of an important public interest - and unless the data subject has explicitly refused any disclosure, then the authorisation to use personal data may be given by the body or bodies designated by domestic law and competent in the area of personal data. Such authorisation should, however, not be given globally, but case-by-case; moreover, the medical data should be used only for the medical research project defined by that body, and not for another project of the same nature (sub-paragraph c).

The authorisation, by the designated body, of communication of medical data for the purposes of a medical research project also depends on other factors implicit in the spirit of the recommendation in the present principle, or explicitly set out in other principles:

- the existence of alternative methods for the research envisaged;
- the relevance of an important public interest of the aim of the research, for example in the field of epidemiology, of drug control or of the clinical evaluation of medicines;
- the security measures envisaged to protect privacy;
- the necessity of interfering in the privacy of the data subject.

Under sub-paragraph (c), it would not be necessary to make the reasonable efforts in all cases; the person in charge must, however, consider whether with reasonable efforts it would be practicable to contact all data subjects. If this seems possible, then the efforts must be made. Furthermore, it was understood that to seek the consent of the data subject for medical research would be an unreasonable demand for the research institute, and would rather be the responsibility of the person or body envisaging disclosure of medical data.

According to article 12 (3), subject to complementary provisions determined by domestic law, health-care professionals entitled to carry out their own medical research are allowed to use the medical data which they hold, as long as the data subject has been informed of this possibility and has not objected.

Finally, personal data used for scientific research must not be published in a form that enables the data subjects to be identified, unless they have given their consent for the publication and publication is permitted by domestic law.

Analysing the above legislation and regulations, some considerations could be made.

In all EU and International legislative Instruments, the right to privacy is not considered an absolute right. It has in fact to be weighed against other matters that benefit societies. All the exemption to the prohibition of processing operations that involve personal data relative to health care and health research constitute clear examples of the non-absolute nature of the right to privacy.

Therefore, it could be inferred that the protection of privacy is conceived as value that should not unnecessarily jeopardize health research. The interest of societies in enhancing the health of populations is in fact strictly related to the possibility of conducting appropriate research in the health sector and the availability of personal data is fundamental for this purpose.

Considering that privacy protection and health research might conflict on the increasing demand of researchers to access data in identifiable form, appropriate methodologies and techniques should be implemented. PIAs are a valuable means to address this issue, providing a balanced trade-off between privacy protection and the efficient and effective conduction of research projects and programmes.

3.3.2 Privacy Protection in the Context of the BIRO Engineering

The B.I.R.O. project involves the processing of medical records collected through diabetes registries at national/regional level, to be further processed for benchmarking and public health monitoring at the international level.

The privacy analysis herein undertaken covers the identification of privacy issues that might arise in the transfer of data from the collaborating centres to the central B.I.R.O. database.

However, the way data are processed within the collaborating centres is not irrelevant to the project since a not legitimate handling of personal information in those centres could render not legitimate, at the same time, the further data processing for research and scientific purposes, which is a crucial component in the B.I.R.O. project.

In general, the data processing occurring at the local level should be subject to Art.8 (par. 3) of the EU Directive: each center in fact collects information related to an identified or identifiable natural person for the purpose of setting up diabetes registries. Hence, data could be considered collected and processed for purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health care services.

According to the EU Data Protection Directive, consent from the data subject may not be required in this case. The norm constitutes an exemption to the general prohibition of processing sensitive data, which is set forth by art. 8 of the Directive.

In this case, the exemption is justified by the need to protect the competing interests of society to a better health care, although domestic laws may provide more stringent rules.

The further processing of these data, other than caring for the patient and managing health services, would instead not be covered by the exemptions of Art.8 (Par. 3): in other words, consent would be required for any secondary use of those data.

Importantly, each centre of the BIRO consortium provides for the anonymisation of data before transferring them to the BIRO central database, where they are processed for statistical and scientific purposes (see figure 4: BIRO Software Engineering).

According to EU Data Protection Directive, the processing carried out to render data anonymous is to be considered as processing of personal data; hence, it is subject to data protection requirements.

Thus, the way data are rendered anonymous is central to determine if true anonymisation is actually envisaged in the BIRO System, according to the state of the art.

Ex Recital 26 of the Data Protection Directive, anonymisation allows the processing of personal data without consent, placing anonymous data outside the scope of the data protection principles therein contained.

Anonymisation could be therefore seen as a means to determine the boundaries of privacy protection principles. When data is truly anonymous, the interest of the data subject to maintain his/her data private and confidential is in fact protected “ipso iure”; hence, the processing should be considered legitimate.

Data is rendered anonymous, according to Directive, only if “the data subject is no longer identifiable”. The Directive specifies that an “identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”. The same Recital specifies that, in order to determine whether a person is identifiable, “account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person”.

Consequently, when the data subject could be identified with reasonable means directly from the data itself or indirectly through the combination of other means, data cannot be considered anonymous and, therefore, fall under the Directive principles, including the need to gain expressed consent from the data subject.

Data could be instead considered anonymous when “it would be reasonably impossible for the researcher or any other person to re-identify the data”[30].

The identification of the data subject through “reasonable means” is a vague concept that involves a value judgement. However, the reference to the state of the art in decoding and/or other similar techniques is usually considered decisive in valuing whether data is truly anonymous or not. For instance, coded and encrypted data are to be considered anonymous for the purpose of the EU Data Protection Directive if data cannot be decoded and de-encrypted with a reasonable effort.

In the context of BIRO, the local centres will use pseudonyms for patients IDs and data will be then stripped of their identifier and aggregated: at least n. 5 patients per cell are to be used. As a matter of fact, the BIRO System processes statistical objects, which basically are tables that contain statistical aggregations of local data (arithmetic mean, percentile, variance, linear and logistic regression, bar plot data, histogram data, box pot data, etc), stored as flat text comma delimited files (CSV).

Hence, there is no possibility, according to the state of the art, to identify, either directly or indirectly, a patient through a reasonable effort.

Although the privacy of legal persons, such as the BIRO Centres, does not receive protection within EU and International legislation, the PIA Team acknowledged that the availability of Centres’ IDs could pose broader privacy concerns. Project’s results could reveal information about participating Centres that might jeopardize their reputation. Hence, this factor could not positively impact on data sharing and eventually discourage participation in the project.

Moreover, when dealing with very small Centres, even doctors or patients could be indirectly identified, if specific information is disclosed together with Centres’ IDs.

In consideration of the above concerns, Centres’ IDs have been protected through the use of a pseudonym, together with to a reporting system based on percentages rather than on absolute numbers. Accordingly, the size of single Centres would be hidden, avoiding their indirect identification by third parties.

Although personal data is rendered truly anonymous and there is no need to justify the processing of those data without obtaining patients’ consent, the further processing of personal data for statistical or scientific research purposes is generally considered, even within the EU Directive, compatible with the purposes for which the data have previously being collected, provided that Member States provide appropriate safeguards[31].

This principle is indirectly expressed, among the others, in the provision of art. 11, par. 2 of the EU Directive.

While art. 10 and 11 impose the data controller, as a general rule, to give some kind of information to the data subject (for instance: the right to know the identity of the controller, the purpose of the processing and any further information), Paragraph 2 of art. 11 exempts the data controller from providing such information when the processing is performed for statistical or scientific research purposes, if the provision of such information proves impossible or would involve a disproportionate effort.

The case of BIRO would fall within the scope of the latter case. Considering its very large sample size, the effort to provide information to patients should herein be easily considered disproportionate. Consequently, the information to be provided to data subject could be waived by the single centres, unless domestic law provided differently, even if the kind of processing would be considered as falling under the EU Data Protection Directive rules.

The exemptions provided by the Directive are also in line with the principles contained in the Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data (1981), which envisages the possibility of restricting the exercise of the data subject's rights with regard to data processing operations which pose no risk (art. 9, par. 3). Examples of no or minimal risk operations are therein considered, in particular, the use of data for statistical work, in so far as those data is presented in aggregate form and stripped of their identifiers, as in the case of BIRO. Similarly, scientific research is included in this category.

The aggregated data, in the form of statistical objects, once processed through the local database engine, are to be sent to the central statistical engine, which will perform global analysis.

A dedicated communication software has been developed to ensuring a secure data and information exchange transmission between the regional information systems and the central SEDIS: statistical objects are transmitted as encrypted compressed folders containing comma-delimited text files (see paragraph 3.2.3 of the present report).

Considering the security mechanisms implemented in the BIRO system, it can be asserted that the security requirements enshrined in EU and international data protection norms and regulations are fully fulfilled, considering the actual state of the art.

According to the BIRO data flow and architecture, statistical analysis will be then performed at global level. Considering that data have been rendered anonymous by local BIRO centres and transmitted to SEDIS in a secure environment, the further processing performed by the global statistical engine cannot pose any privacy risk either directly or indirectly.

The last issue that could be considered in the privacy analysis of the BIRO project is relative to the transborder data flow. In fact, data is to be sent to a central database, which is located outside the single national boundaries, except for the Italian partner (Coordinator).

The BIRO System, as already demonstrated, processes only anonymous data; therefore, privacy rules should not bound its implementation.

Nevertheless, the free flow of information, regardless of frontiers, is a principle enshrined in Article 10 of the European Human Rights Convention. Accordingly, art 12 of the Convention on the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) and art. 25 of the EU Data Protection Directive (1995) discipline the transfer of data from one country to another.

The main rule contained in art 12 (paragraph 2) of the Convention (1981) is that, in principle, obstacles to transborder data follows are not permitted between Contracting States in the form of prohibitions or special authorisations of data transfers. The rationale for this provision is that all Contracting States, having subscribed to the same common core of data protection provisions, offer a certain minimum level of privacy protection.

The Council of Europe Recommendation on the Protection of Medical Data, resembles the Convention and establishes that the transborder flow of medical data to a state which has ratified the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, and which disposes of legislation which provides at least equivalent protection of medical data, should not be subjected to special conditions concerning the protection of privacy.

Where the protection of medical data can be considered to be in line with the principle of equivalent protection laid down in the Convention, no restriction should be placed on the trans-border flow of medical data to a state which has not ratified the convention, but which has legal provisions which ensure protection in accordance with the principles of that Convention and the Recommendation.

Unless otherwise provided for by domestic law, the transborder flow of medical data to a state which does not ensure protection in accordance with the Convention and with this Recommendation, should not as a rule occur, unless necessary measures, including those of a contractual nature, to respect the principles of the convention and this recommendation, have been taken, and the data subject has the possibility to object to the transfer; or the data subject has given his consent.

According to the EU Directive, the cross border flow of personal data is allowed only when an adequate level of privacy protection is envisaged in the countries involved in the processing operations.

Following the same reasoning applied to the interpretation of the Convention (1981), countries that have implemented the Directive are automatically allowed to transborder data flows: complying with the Directive ensures, “ipso iure”, an adequate level of protection.

The Centres involved in the BIRO project belong to European countries that have fully implemented the EU Data Protection Directive, and ratified the Convention; hence, an adequate level of privacy protection is fully guaranteed across the countries involved. This means that the exchange of data envisaged in the project would be any way legally viable, according to EU and international legislation, considering the system architecture and composition of the BIRO Consortium.

Finally, publication of project results is performed in a form that does not enable not only the data subjects but also local Centres to be identified.

4. Privacy risks and Mitigation Strategies

The potential privacy risks envisaged in the BIRO project could be summarized as follow:

- Data cannot be considered truly anonymous
- Data transmission from local to central database cannot be considered secure
- Performance of global analysis based on non-truly anonymous data could indirectly reveal patients’ identities; for instance through the publication of results.
- Access to central server may be hacked and reversibly used to access individual local server and break into personal information stored in computerized registries

The Potential privacy risks have been analysed through a summary table, which allows to estimating the better privacy protective alternative in the processing of data.

The level of risk has been classified as follow:

- Low: There is a possibility that the risk will materialize but there are mitigating factors
- Moderate: There is a strong possibility that the risk will materialize if no corrective measures are taken
- High: There is a near certainty that the risk will materialize if no corrective measures are taken

Anonymization is a crucial factor in the development and implementation of the BIRO project. In order to carry out research on anonymous data outside the application of the Data Protection Directive, data have to be acquired for legitimate purposes, from authorized controllers, local B.I.R.O. Centres, who had already anonymised the data irreversibly: the data subject re-identification through a reasonable effort is impeded before the transfer of data from the local Centres to the Central Database (SEDIS).

Different elements of anonymisation had to be then verified:

- data controller authority to collect and process those data
- purposes of processing
- efficiency of the anonymisation process, according to the state of the art

The local B.I.R.O. Centres collect and process health data according to different national legislations, which grant them authority to collect and process data through diabetes registries and/or databases.

The king of processing performed by local Centres is legitimate according to art. 8 (3) of the EU Data Protection Directive: each center in fact collects information related to an identified or identifiable natural person for the purpose of setting up diabetes registries. Hence, data are to be considered collected and processed for purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of health care services; which is one of the purposes considered legitimate for the collection of sensitive data ex art. 8 of the Directive.

The anonymisation techniques used and implemented in BIRO guarantee an irreversible anonymisation. The B.I.R.O. centres, in fact, send only aggregate records to the central server (figure 4). For the most sensitive variables, aggregated records are not transmitted if groups contain less than five patients. Statistical objects are sent as tables stored in compressed bundles of flat text comma delimited files (CSV). Hence, there is no possibility, either directly or indirectly, that a patient could be ever identified with a “reasonable effort”.

In broader terms, the privacy of clinical centres has also been considered in the project. The relative privacy risk has been mitigated through the use of pseudonyms for Centre IDs and a reporting system of project results that shows information in percentage rather than in absolute numbers; thus, it does not reveal, for instance, the size of local Centres, impeding their indirect identification.

Security of transmission

Aggregated statistical objects are sent to the central statistical engine to carry out global analysis.

A dedicated communication software has been specifically developed to ensure secure information exchange between the regional systems and the central SEDIS (see paragraph 3.2.3).

Global analysis

Global reporting does not pose any direct or indirect risk to privacy, as anonymous data sent by B.I.RO. Centres is transmitted to SEDIS in a secure environment, and further processed in aggregate form.

Access to central server

Relative to the access, security mechanisms are implemented using standard procedures at the strictest level. Once the application will be completely tested, it will be possible to conduct experiments to check the level of security using different hacking techniques. This task will be performed through updates of the BIRO system.

In conclusion, the BIRO Information System processes only de-identified data. Hence, the level of risk can be considered, in most of the cases described, low.

As highlighted in the privacy summary table (Table 6), efficient mitigation strategies have been implemented in the context of BIRO. Consequently, the aforementioned potential privacy risk could be considered fully avoided and/or removed from the system architecture by design.

Table 6. B.I.R.O. Privacy Summary Table

Element	Nature of risks	Level of risks			Comments	Mitigating Mechanisms
		Low	Medium	High		
Individual data: Pseudonym used for patients' IDs + Data is Aggregated (N=5 patient per cell) + statistical objects	Individual privacy	X			Pose an indirect risk to individual's privacy	Non-Reversible De-identification
Pseudonym used for Centres IDs	Non-Individual Privacy		X		Pose an indirect risk to Centres' privacy	Reversible De-identification + Reporting System : percentage
Data Transmission	Security Measures	X			Pose an indirect risk to individual's privacy	Encryption
Access to the BIRO network	Security Measures		X		Pose an indirect risk to individual's privacy	Secure applications Hacking tests
Global Statistical Analysis	Individual privacy + Non-Individual Privacy + Security Measures	X			Pose an indirect risk to individual's privacy and centres privacy	Non-reversible de-identification + Encryption

5. Conclusions

The B.I.R.O. project aims at implementing an international health information system linking data sourced by different diabetes registries.
The present Privacy Impact Assessment shows that the BIRO architecture fulfils privacy protection requirements, addresses and resolves any privacy risk identified, and tackles broader privacy concerns from different angles.
Advancements should also foresee conditions beyond the usual boundaries of personal involvement, e.g. professional and institutional integrity in the conduct of health research.
The architecture of the B.I.R.O. system flexibly affords the best privacy protection in the construction of an efficient model for the continuous production of European reports.
The privacy impact assessment method developed and applied in B.I.R.O. may represent a general tool that can be used to design trans-border health information systems.

APPENDIX 1 : PIA QUESTIONNAIRE

INSTRUCTIONS :

- **EACH MEMBER OF THE PIA TEAM SHALL PROVIDE MARKS FOR EACH QUESTION/DECISION, OPTION, CRITERION**

- **SCALE OF THE MARKS:**

  0 = not applicable
  1 = very low
  2 = low
  3 = sufficient
  4 = high
  5 = very high

- **CRITERIONS:**

Privacy Criterion 1: Identifiability

- Measures the degree to which information is personally identifiable
- The Identity measurement takes place on a continuum, from full anonymity (the state of being without name) to full verinymity (being truly named)
- Goal of the Privacy Impact Assessment (PIA) Team is always to decrease the amount of identity in the BIRO system
- A minimalist design approach should be employed and if identity data is not required, it should be intentionally removed from the architectural equation
- Many tools employing reversible and non-reversible pseudonymity are available for this purpose

Privacy Criterion 2: Linkability

- Measures the degree to which data elements are linkable to the true name of the data subject
- Unlinkability means that different records cannot be linked together and related to a specific personal identity.
- Complex interrelations need to be taken into account: record linkage can be subtle, as it may be organized and/or made possible in different ways

Privacy Criterion 3: Observability

- Measures the degree to which identity or linkability may be impacted from the use of a system
- It considers any other factor relative to data processing (time, location, data contents) that can potentially affect the degree of identity and/or linkability (effect modifiers)

Information Content

- Single score providing an overall mark for the level of information provided by the specific scenario/option in terms of relevance and level of evidence for diabetes

Technical Complexity

- Single score providing an overall mark for the feasibility of the specific scenario/option

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 1: Health service Medical Record

- ❑ **collected by: Clinical Centres, Coordinating Centre**
- ❑ **used by: Local Health Authority, Coordinating Centre**
- ❑ **purpose: Disease Management Program**

Option	**Privacy**				**Information Content**	**Technical Complexity**
	Identifiability	**Linkability**	**Observability**	**Overall**	**Overall**	**Overall**
One record for each service episode, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained						
One record for each service episode, Centre IDs De-Identified						
Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained						
Multiple measurements averaged over time interval, Centre IDs De-Identified						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 2:

Administrative Data Service Episode

- □ **collected by Local Health Authority**
- □ **used by Local Health Authority**
- □ **purpose Policy and Planning**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Population-based longitudinal records, linked across administrative datasets, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained						
Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified						
Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained						
Multiple measurements averaged over time interval, Centre IDs De-Identified						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 3:

Epidemiological measurement of multiple individual characteristics

- ❑ **collected by Research Organization**
- ❑ **used by Research Centres**
- ❑ **purpose Epidemiological study**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients						
Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 4.1:

Health service medical record + administrative data service episode

- □ **collected by Population-based regional/national diabetes register**
- □ **used by Local Health Authority, Research Centre, Regional/National Government**
- □ **purpose Disease management, policy and planning, research**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO						
Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified						
Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO						
Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 4.2:

Health service medical record + administrative data service episode + Epidemiological measurement of multiple individual characteristics

- □ **collected by Population-based regional/national diabetes register**
- □ **used by Local Health Authority, Research Centre, Regional/National Government**
- □ **purpose Disease management, policy and planning, research**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage overmultiple datasets, all relational structure sent to BIRO						
Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified						
Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO						
Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 2 : AGGREGATION BY GROUP OF PATIENTS

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)**

SCENARIO 1:

Grouping condition directly set by statistical object
(e.g. ordered frequency distribution of LOS by CENTRE to compute variability of medians)

- □ **collected by BIRO partner**
- □ **type of format One Record for each Aggregation Level**
- □ **used by BIRO partner (local engine), BIRO Consortium (central engine)**
- □ **purpose of collection (computation of single statistical object for local and SEDIS reporting)**

***Question 1.* PERSONAL INFORMATION/DATA CLUSTER: DECISION 1**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
No Aggregation Size Limit						
Min aggregation N=5 patients per cell						
Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 2 : AGGREGATION BY GROUP OF PATIENTS

***Question 1.* PERSONAL INFORMATION/DATA CLUSTER: DECISION 2**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Aggregation across service centres						
Data aggregated at the level of service centre						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 2 : AGGREGATION BY GROUP OF PATIENTS

Question 1. **PERSONAL INFORMATION/DATA CLUSTER: DECISION 3**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed						
Aggregation of multidimensional patterns (e.g. risk adjustment) allowed						
Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 2 : AGGREGATION BY GROUP OF PATIENTS

***Question 1.* TRANSMISSION: DECISION 1**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
All DATE fields transmitted as in original						
DATE fields approximated to time interval (e.g. months)						

Comments:

DATA FLOW QUESTIONNAIRE

Question 1. **TRANSMISSION: DECISION 2**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	*Overall*	Overall	Overall
Service Centre ID transmitted						
Pseudonym used for service centre						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 3 : AGGREGATION BY REGION

***Question 1*: Personal information/Data clusters, Collected by, Type of Format, Used by, Purpose of collection and Transmission**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

SCENARIO 1:

Grouping condition directly set by statistical object
(e.g. ordered frequency distribution of LOS by REGION)

- □ **collected by BIRO partner**
- □ **type of format One Record for each Aggregation Level by REGION**
- □ **used by BIRO partner (local engine), BIRO Consortium (central engine)**
- □ **purpose of collection (computation of single statistical object for local and SEDIS reporting)**

PERSONAL INFORMATION/DATA CLUSTER: DECISION 1

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
NO restrictions on specific stratification criteria (e.g. geographical variable, centres, etc)						
restrictions applied on specific stratification criteria (e.g. geographical variable, centres, etc)						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 3 : AGGREGATION BY REGION

Question 1. PERSONAL INFORMATION/DATA CLUSTER: DECISION 2

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	*Overall*	Overall	Overall
Geographical mapping available						
Geographical mapping unavailable						

Comments:

DATA FLOW QUESTIONNAIRE

Question 1. PERSONAL INFORMATION/DATA CLUSTER: DECISION 3

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	*Overall*	Overall	Overall
Variability of centres outcomes available						
Variability of centres outcomes unavailable						

Comments:

DATA FLOW QUESTIONNAIRE

Question 1. **PERSONAL INFORMATION/DATA CLUSTER: DECISION 3**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed						
Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITHOUT restrictions applied on specific stratification criteria						
Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITH restrictions applied on specific stratification criteria						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE: ALL

***Question 2*: Security Mechanisms**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Password access for local administrator prompting client program to send encrypted bundles to BIRO						
Client program automatically sending encrypted data (agent)						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA

***Question 3*: Format of BIRO database**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Full information on all medical records						
Averaged over time						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE: ALL

***Question 4*: Disclosure**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
BIRO database administrator						
All local database administrators / registry managers						

Comments:

DATA FLOW QUESTIONNAIRE

CANDIDATE ARCHITECTURE: ALL

***Question 5*: Storage and Retention Site**
Assign to each data flow item a mark (0-5) for each scoring dimension (privacy, information content, technical complexity)

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
BIRO Coordinating Centre						
EU/DG-SANCO						

Comments

References

[1] The B.I.R.O. project official website; available at: http://www.biro-project.eu/index.html

[2] Blair Stewart, “Privacy Impact Assessments,” *Privacy Law and Policy Reporter (1996), vol. 39* at: www.austlii.edu.au/au/journals/PLPR/1996/39.htm

[3] David H. Flaherty, “Privacy Impact Assessments: An Essential Tool for Data Protection,” in S. Perrin, H. Black, D.H. Flaherty and T. M. Rankin *The Personal Information Protection and Electronic Documents Act* (Toronto: Irwin Law, 2001), p. 272

[4] ICO (2007a) 'Privacy Impact Assessments: International Study of their Application and Effects' Information Commissioner's Office, Wilmslow, I.K., December 2007; available at: http://www.ico.gov.uk/Home/about_us/research/data_protection.aspx

[5] Flaherty, Privacy Impact Assessments, p. 266

[6] ICO (2007a) 'Privacy Impact Assessments: International Study of their Application and Effects' Information Commissioner's Office, Wilmslow, I.K., December 2007; available at: http://www.ico.gov.uk/Home/about_us/research/data_protection.aspx

[7] Peter Hope-Tindall, Privacy Impact Assessment – Obligation or Opportunity: The Choice is Ours! 2000-2002 – data Privacy Partners Ltd. Prepared for CSE ITS Conference – Ottawa, Ontario – May 16th, 2002

[8] The B.I.R.O. Privacy Impact Assessment Team, Preliminary Privacy Impact Assessment Report, Perugia 2006; available at: http://www.biro-project.eu/index.html

[9] James Michael, Privacy and Human Rights 1. (UNESCO) 1994; available at: http://webjcli.ncl.ac.uk/articles1/davies1.html

[10] Samuel Warren, Louis Brandeis. The Right to Privacy. Harvard Law Review 1890; 4:193–220

[11] (Chairman) David Calcutt QC. Report of the Committee on Privacy and Related Matters. London: Cmnd. 11027, 1990

[12] Universal Declaration of Human Rights, adopted and proclaimed by General Assembly resolution 217 A (III) of December 10, 1948; available at http://www.un.org/Overview/rights.html

[13] International Covenant on Civil and Political Rights, adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of December 16, 1966, entry into force March 23rd 1976; available at http://www.unhchr.ch/html/menu3/b/a_ccpr.htm

[14] International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families, adopted by General Assembly resolution 45/158 of December 18, 1990; available at: http://www.unhchr.ch/html/menu3/b/m_mwctoc.htm

[15] Convention on the Rights of the Child, adopted and opened for signature, ratification and accession by General Assembly resolution 44/25 of November 20, 1989, entry into force September 2, 1990; available at: http://www.unhchr.ch/html/menu3/b/k2crc.htm

[16] Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms, (ETS no: 005) open for signature November 4, 1950, entry into force September 3, 1950; available at: http://conventions.coe.int/Treaty/EN/cadreprincipal.htm

[17] Strossen Nadine, Recent United States and Intl. Judicial Protection of Individual Rights: A comparative Legal Process Analysis and Proposed Synthesis. Hastings Law Journal 1990; 41: 805

[18] European Court of Human Rights, Case of Klass and Others: Judgement of 6 September 1978, Series A No. 28 (1979). Malone v. Commissioner of Police, Series A82 (1984);

available at: http://hudoc.echr.coe.int/hudoc/ViewRoot.asp?Item=5&Action=Html&X=1007095902&Notice=0&Noticemode=&RelatedMode=0;

[19] European Court of Human Rights, Leander v. Sweden, series A No 116 (1987); available at: http://hudoc.echr.coe.int/hudoc/ViewRoot.asp?Item=0&Action=Html&X=1007101431&Notice=0&Noticemode=&RelatedMode=0

[20] Convention for the Protection of Individuals with regard to the Automatic Processing of Personal Data Convention. Strasbourg, 1981; available at: http://www.coe.fr/eng/legaltxt/108e.htm

[21] OECD, Guidelines Governing the Protection of Privacy and Transborder Data Flows of Personal Data. Paris, 1981; available at: http://www.oecd.org/dsti/sti/it/secur/prod/PRIV-EN.HTM

[22] Council of Europe Convention on Human rights and Biomedicine (Oviedo 1997); available at: http://conventions.coe.int/Treaty/EN/Treaties/Html/164.htm

[23] Charter of Fundamental Rights of the European Union (2000/C 364/01); available at: http://ec.europa.eu/justice_home/unit/charte/index_en.html

[24] Official Journal of the European Union C 310 Volume 47 of 16 December 2004; available at: http://eur-lex.europa.eu/JOHtml.do?uri=OJ:C:2004:310:SOM:EN:HTML

[25] Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. Strasbourg, 25.I.2005; available at: http://conventions.coe.int/Treaty/EN/Treaties/Html/195.htm

[26] Directive 95/46/EC; available at: http://ec.europa.eu/justice_home/fsj/privacy/law/index_en.htm

[27] Roy McClelland at all. European Standards on Confidentiality and Privacy in Healthcare, EuroSOCAP Project (2003-2006); available at: http://www.eurosocap.org/Downloads/European-Standards-on-Confidentiality-and-Privacy-in-Healthcare.pdf

[28] Roy McClelland at all. European Standards on Confidentiality and Privacy in Healthcare, EuroSOCAP Project (2003-2006); available at: http://www.eurosocap.org/Downloads/European-Standards-on-Confidentiality-and-Privacy-in-Healthcare.pdf

[29] Council of Europe Recommendation No R (97) 5; available at: http://www.coe.int/t/e/legal_affairs/legal_co-operation/data_protection/documents/international_legal_instruments/Rec(97)5_EN.pdf

[30] European Commission Contract 30-CE-0041734/00-55, European Health Management Association, Legally eHealth, 2006, D.2(PUB) v. 8

[31] European Commission Contract 30-CE-0041734/00-55, European Health Management Association, Legally eHealth, 2006, D.2(PUB) v. 8, p. 17