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 2

Data Flow Analysis:

Ranking BIRO Architectures through the Data Flow Table and Information Flow Questionnaire

September 2007

(The BIRO P.I.A. Team)

The P.I.A. Team

Privacy Facilitators:

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

P.I.A. Team Members:

- Dr. Marco Orsini, 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

TABLE OF CONTENTS

1. Introduction

1.1. The BIRO project

The general objective of the BIRO project is *to build a common European infrastructure for standardized information exchange in diabetes care, for the purpose of monitoring, updating and disseminating evidence on the application and clinical effectiveness of best practice guidelines on a regular basis.*

The aforementioned general objective is being pursued through the realization of the following specific objectives:
- identification of a set of clinical guidelines based on the scientific literature
- selection of a European minimum dataset for international comparisons
- adoption of common health and quality indicators for routine monitoring of diabetes outcomes
- finalisation of a concept and data dictionary for information exchange and data processing
- definition of standardized statistical analyses, in the form of report templates.
- design and implementation of a relational data model
- design and implementation of statistical methods for the production of health reports
- validation of a secure protocol for international communication and shared data analysis
- customisation and development of specialized software to be deployed in the public domain
- linkage of the different components in a user-friendly reporting facility
- dissemination of all results through a web portal and a specialized publication

In order to fulfil these objectives, a coherent system, defined as "Shared Evidence-based Diabetes Information System", hereafter referred to as "SEDIS", is being built.

SEDIS represents an efficient and sustainable solution to perform the following tasks:

- analysis of longitudinal trends and average outcomes in a diabetic population
- identification of patterns of care and prevention consistently showing positive results
- identification of population strata and/or practices that do not show effective results
- verification of the application/applicability of best practice guidelines
- on-field testing of collaborative information systems in chronic diseases.

1.2 The Data Model

The SEDIS data model is divided into two parts: a static part, related to data collection (which hardly changes over time) and a more dynamic part, related to medical concepts (that is more susceptible to changes in the medical knowledge). Isolating the dynamic part will largely eliminate the risk of performing frequent updates in the software.

The complete SEDIS data cycle is based on the application of two consecutive data processing steps. As a matter of fact, the fundamental aspect of the system is to ensure its basic functionalities at the level of each single register ("*local SEDIS*"). The model is then generalized through its repeated application by all registers, supported by an overall step that compiles all "partial" results into a global report.

Statistical analysis and epidemiological modelling of a disease register require an in depth understanding of all aspects related to the characteristics stored in the database. The organization of biometric and socio-demographic information must be based on solid classification criteria, e.g. normal levels for glycated haemoglobin using different kits, or algorithms for the construction of an index of socio-economic status (SES). In these circumstances it is useful to keep all definitions stored in a *data dictionary* using a common format. If we include in such a "progressive diary" also more general clinical concepts, such as the list of tests recommended to patients with hypertension, over 65, with a high level of glycated haemoglobin (*guidelines)*, or a particular "severity score" (*comorbidity index*), then the result would be a "*concept and data dictionary*" (CDD).

The CDD in the context of a "*local SEDIS*" can be represented as a chain of steps logically intertwined (Fig. 1). The availability of a CDD is essential to compare different analysis, both geo-

graphically and longitudinally. The CDD is the evidence-based and first component in the model chain. It follows the definition of a minimum dataset and it needs to be regularly updated. At the opposite end of the chain is the final output of the system, i.e. a health system report. The content of the report is based on the initial specification of a template that influences the selection of data procedures and statistical methods ("*database engine"* and *"statistical engine"*).

The engines operate on top of the local databases that are not directly accessible by other partners. The reports will be composed through the amalgamation of statistical "*objects*" (*tables, parameters, graphs*) that are to be produced in turn by the joint application of the engines.

The definition of an overall model (*global SEDIS)* directly follows the local implementation (fig.2). Once the statistical objects are available for each register, these can be exchanged across the network using a secure format. The level of aggregation chosen for each object, according to the proposal, is a combination of:

a) formal agreement
b) legislation
c) practical limits

These conditions have been discussed within the Privacy Impact Assessment of SEDIS, Step 2, which includes a consensus procedure (Delphi Panel) to establish the type of information to be exchanged among BIRO partners.

A legislative review has been already conducted in the context of the Preliminary Privacy Impact Assessment (Step 1), and made available to partners in order to progress to the following steps.

2. Scope and state of the art of the Privacy Impact Assessment

The Privacy Impact Assessment (PIA) provides 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

A PIA process has been defined for the conduction of BIRO WP5, including four steps:
- ❑ Step 1: Preliminary PIA
- ❑ Step 2: Data Flow Analysis
- ❑ Step 3: Privacy Analysis
- ❑ Step 4: PIA Report

WP5 achievements at completion of Step 1 include:
- ❑ establishment of PIA Team
- ❑ conduction of a summary evaluation of potential privacy risks of the BIRO Information System
- ❑ definition of a checklist of key privacy requirements/criteria
- ❑ general description of candidate architectures for the BIRO Information System
- ❑ delivery of Preliminary PIA Report to the European Commission

3. Objectives of P.I.A. Step 2

The general objective of PIA Step 2 (*data flow analysis)* is 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* are:

- ❑ to develop a detailed description and analysis of BIRO data flow
- ❑ to describe and in-depth analysis of the BIRO information 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 have been 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
  - o identifying clusters of personal information/data involved in BIRO System
  - o 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.

4. Materials and Methods

4.1. Building the BIRO Health Information System Diagram

The BIRO Health Information System Architecture Diagram documents:

- ❑ The general BIRO infrastructure architecture
- ❑ The general flow of information through the system
- ❑ Any physical or logical separation of personal information/data and/or
- ❑ Security mechanisms that prevent improper access to personal information/data and/or
- ❑ Means to maintain any required separation

4.2. Definition of the Data Flow

The in dept description of the information flow involved in project involves the following activities:

- ❑ Identifying clusters of personal information/data involved in BIRO System
- ❑ Describing all personal data elements associated with the proposed system (example: a data cluster could be elements of patient identification e.g. name, country of birth, ethnicity, etc.)
- ❑ Developing detailed data flow tables
- ❑ Describing the collection, use and disclosure of personal information/data in the BIRO project
- ❑ Listing the different options available for data collection and exchange in each BIRO candidate architecture

The *data flow table* is a specific tool developed in order to describe in depth the dynamics involved in both data collection and information exchange procedures. Data flow tables have been used for each of the candidate architectures identified in PIA previous step. It includes details of personal information/data and how they are handled along the entire process: from collection, use, disclosure and to disposition.

The tables include information on:
- ❑ data sharing, data retention and data disposal
- ❑ source of data
- ❑ acquisition (direct, indirect)
- ❑ authority to collect
- ❑ use and purpose of collecting information (authority for use)
- ❑ disclosure and retention (security levels for information)
- ❑ how long information is retained for
- ❑ where it is retained

Scope of the data flow table is to highlight all major components that should be taken into account to rank the different BIRO alternative architectures (described in Step 1 of the PIA process) and, ultimately, to identify any privacy risk that the handling of data through the system might involve (privacy analysis).

3

4.3 Construction of the Information Flow Questionnaire

The *information flow questionnaire* is constructed by using the various individual components listed in the data flow tables. The various options have been discussed and grouped to specify the different solutions available for the definition of the final structure of the BIRO information system.

Scope of the questionnaire is to assign marks to each alternative (and, within the single alternative, to each sub-scenario and/or sub-option) in order to depict the best alternative for the BIRO information system.

The evaluation of each item is based on three different criteria:

- ❑ privacy protection
- ❑ information content
- ❑ technical complexity

The procedure to perform the questionnaire requirements includes the distribution of the questionnaire to the each PIA Team Member of the BIRO project, who fills in *independently the questionnaire and* returns it to the BIRO Coordinating Centre.

4.4. Ranking the different architectures

Once all the questionnaires have been returned to the BIRO Coordination Centre, duly filled in, the discussion has been re-opened at the Delphi consensus session, held in Cyprus during the 2nd BIRO Investigator Meeting (23-25 May 2007).

The candidate architectures, including all scenarios and sub-options, have been re-evaluated in the light of the individual questionnaire results, highlighting all critical aspects of the procedure. Finally, the alternatives have been ranked according to a mixture of identified criteria and discussion agreements.

5. Results

The following sections highlight the procedure and tools employed to fulfil all PIA Step 2 objectives. As already anticipated in the report, the resulting information has allowed the identification of the best privacy protective BIRO architecture.

5.1 BIRO Information System Diagram

This section presents the draft BIRO Information Diagram as it links the different connected centres/regions to the Shared European Diabetes Information System (SEDIS).

The following diagram (Figure1) has been revised and updated in the light of the discussion over the BIRO architecture and data flow.
Fig. 2 describes the BIRO architecture and software requirements.

```
UNIPG

UNIDUND

JOANNEUM                 DATA
                         FROM
                         EACH        TRANSMISSION            SEDIS
UNIBERG                  CENTRE
                                                      BIRO COORDINATION
                                                           CENTRE
PAULESCU                                                   (UNIPG)

UNIMALT

CYPRUS

                      SECURITY
```
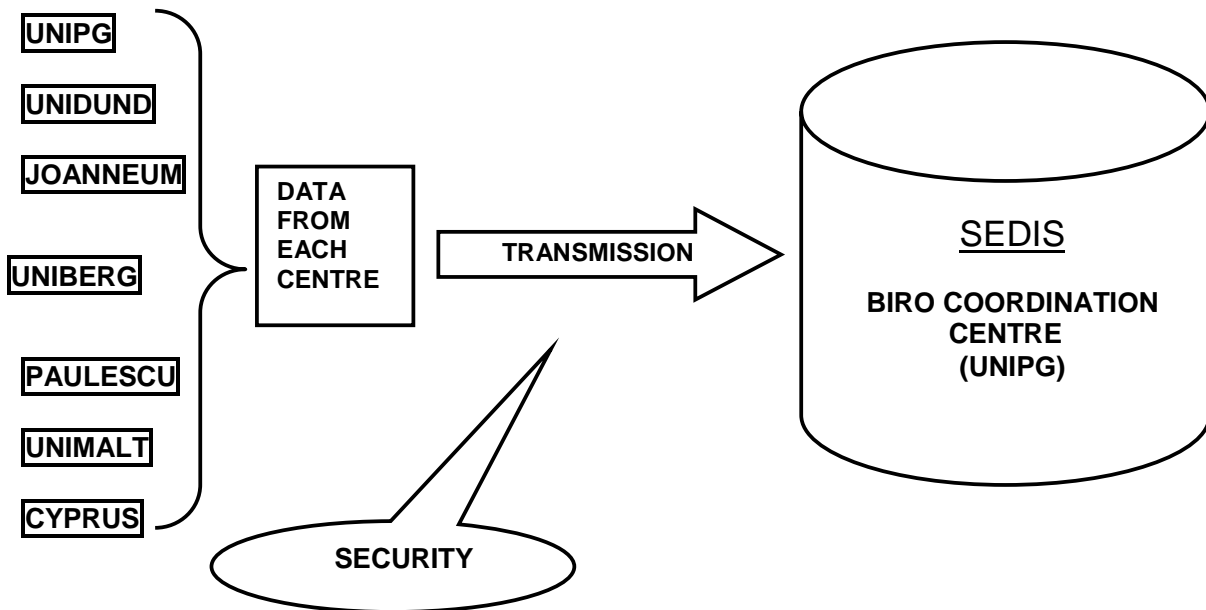
Fig 2. BIRO SYSTEM DIAGRAM
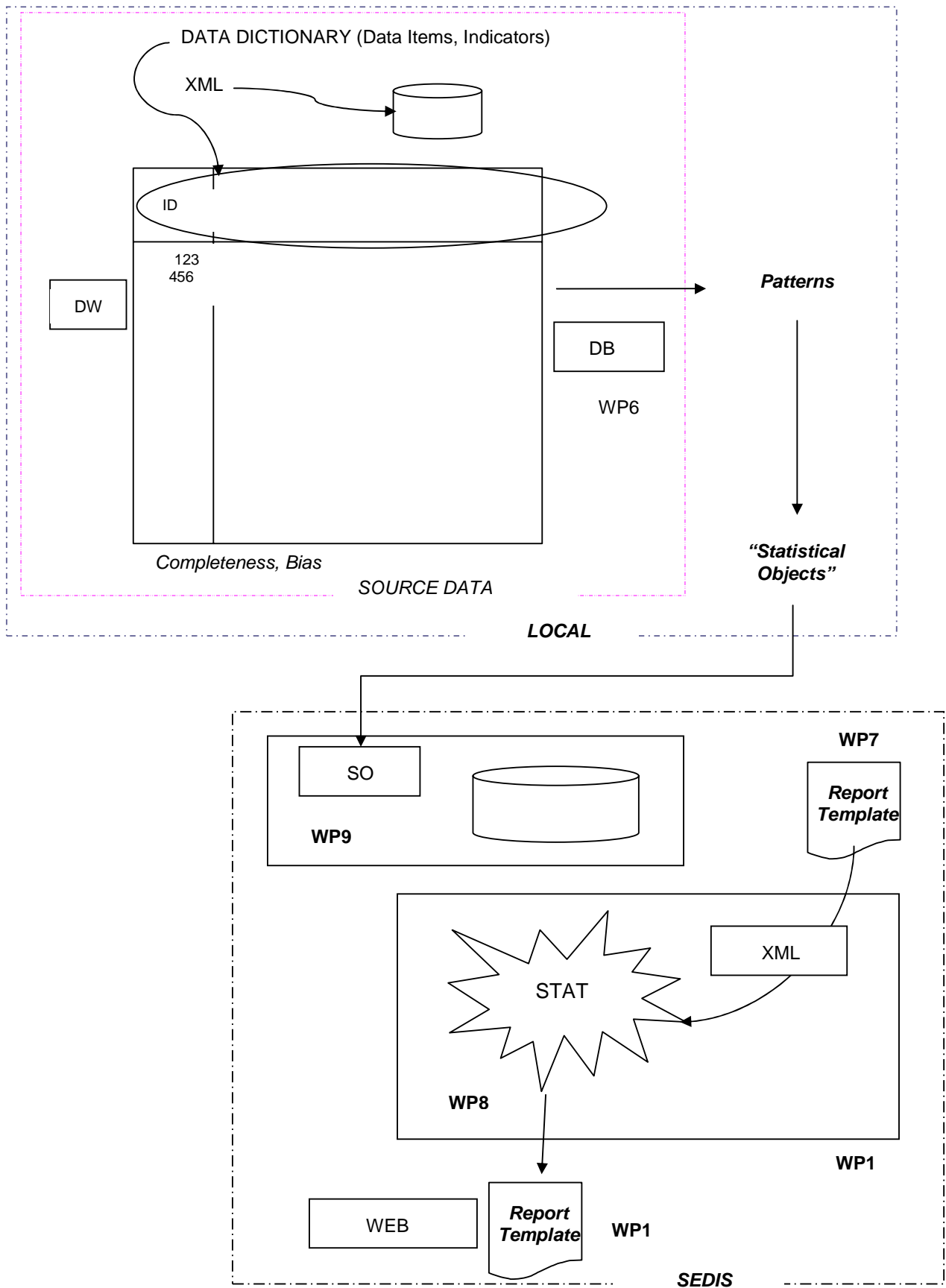
Fig.2 BIRO Architecture and Software Requirements

5.2 Data Flow Tables

The following section presents the Data Flow Tables relative to each architectural alternative of the BIRO information system, as selected in PIA Step 1.

The first Data Flow Table describes the flow of information through the BIRO information system in case individual data will be used, the second Data Flow Table deals with data aggregated by group of patients and the third presents data aggregated by region.

The contents of the data flow tables have been used to construct the Information Flow Questionnaire.

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[i]	Clinical Centres, Coordinating Centre[ii]	OPTION 1 Longitudinal data collection OPTION 2 Multiple measurements averaged over time interval[iii]	Local Health Authority, Coordinating Centre	Disease Management Program	Pseudonym used for data linkage[iv], multiple measurements per patient OPTION 1. Centre IDs retained OPTION 2. Centre IDs de-identified[v]	OPTION 1. Password access for local administrator prompting client program to send encrypted bundles to BIRO[vi] OPTION 2. Client program automatically sending encrypted data (agent)[vii]	OPTION 1. Full information on all medical records OPTION 2. Averaged over time[viii]	OPTION 1. BIRO database administrator OPTION 2. All local database administrators[ix]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[x]
SCENARIO 2: Administrative Data Service Episode[xi]	Local Health Authority[xii]		Local Health Authority	Policy and Planning					
SCENARIO 3: Epidemiological measurement of multiple individual characteristics[xiii]	Research Organization[xiv]		Research Centre	Epidemiological Study					
SCENARIO 4.1: Health Service Medical Record + Administrative Data Service Episode	Population-based Regional/ National Diabetes Register[xv]		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									

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)[xvi]

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[xvii] OR min aggregation N=5, only applicable for high critical privacy variables e.g. service centre, geographical site etc[xviii] Aggregation across service centres[xxiii] OR data aggregated at the level of Service Centre Aggregation of Multidimensional patterns (e.g. risk adjustment) NOT allowed[xxiv] OR generally allowed[xxv] OR allowed with min N=5 condition applied[xxvi]	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[xix]	OPTION 1. All DATE fields transmitted as in original OPTION 2. DATE fields approximated to time interval (e.g. months)[xx]	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[xxi]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[xxii]

9

CANDIDATE ARCHITECTURE 3: AGGREGATION BY REGION

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

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) Geographical mapping available[xxxi] OR Unavailable Variability of Centres' Outcomes Available[xxxii] 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[xxxiii]	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)[xxviii]	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[xxix]	OPTION 1. BIRO Coordinating Centre OPTION 2. EU (DG-SANCO)[xxx]

5.3 Information Flow Questionnaire

This section presents the resulting Information Flow Questionnaire preceded by relevant methodological issues.

Methodological issues in the construction of the questionnaire

The following issues have been considered when scoring items in the Information Flow Questionnaire:
- ❑ definitions
- ❑ identify major dimensions (scoring columns)
- ❑ agree metrics
- ❑ identify scoring dimensions

Scoring Dimensions

Consideration of privacy issues in the definition of a specific BIRO information system architecture should take into account various fundamental dimensions in order to allow the implementation of the best solution in terms of both privacy, quality of care and outcomes evaluation.

The impact of BIRO on privacy is therefore a trade-off between:
- ❑ higher levels of privacy protection
- ❑ relevance of information content in relation to target diabetes indicators
- ❑ minimal technical complexity

The applied scoring system therefore produces a composite indicator incorporating all of the above dimensions in order to objectively support a final decision on the candidate best architecture.

Scoring Dimension 1. Privacy

The score on privacy is based on three separate criteria:

- ❑ Identifiability
- ❑ Linkability
- ❑ Observability

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)
- ❑ The goal of the Privacy Architect and the PIA author is always to decrease the amount of identity in a given 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

Figure 3 describes how the degree of identity could be measured.

Criterion 2: Linkability

- ❑ Measures the degree to which data elements are linkable to the true name of the data subject

11

- ❑ 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

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)

Marks:

Anonymity_____1
Non Reversible Pseudonymity_____2
Reversible Pseudonymity _____3
Indirect Verinymity_____4
Verinymity_____5
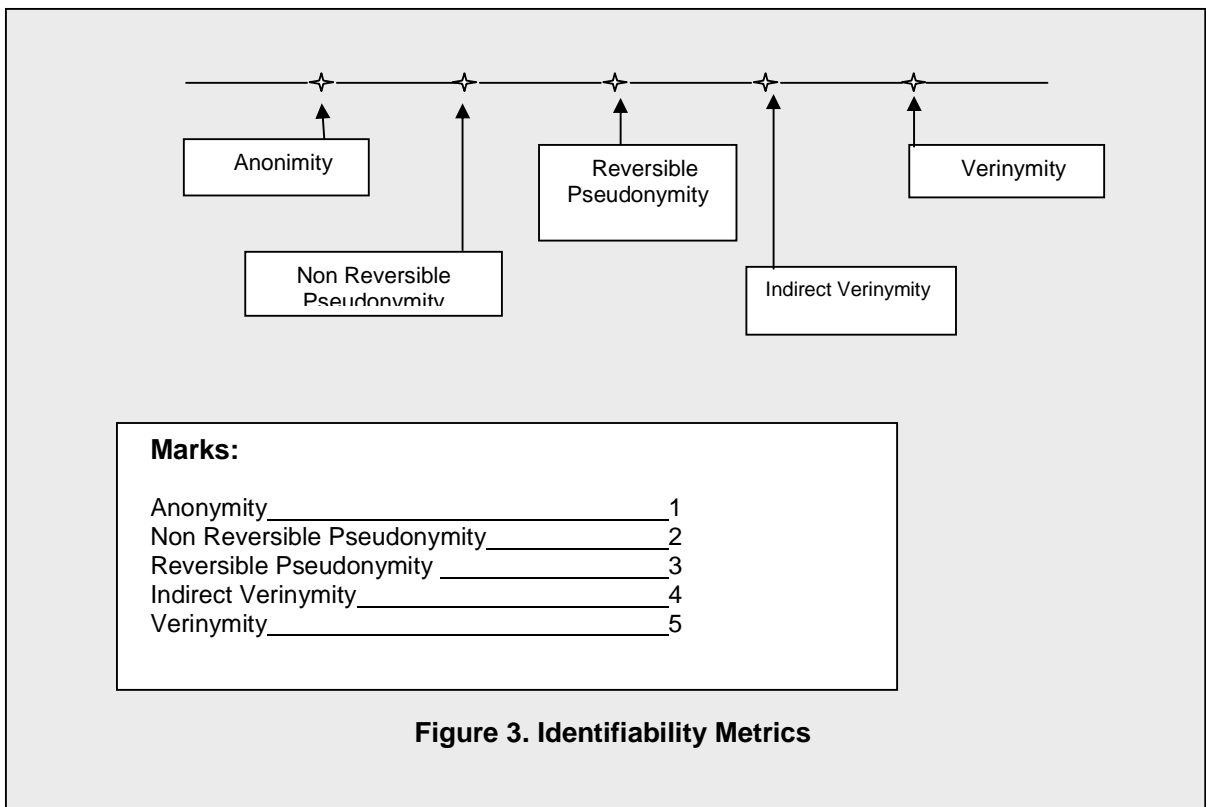
Figure 3. Identifiability Metrics

Although the proposed metrics do not produce truly objective measurements (standards are yet to be identified/developed), they represent the building blocks of a scoring system underpinning a fair comparison among different solutions and a means to minimize the degrees of identifiability, linkability and observability in the proposed system.

A single privacy score for each questionnaire item has been obtained calculating the average mark of each proposed criteria.

Scoring Dimension 2. Information Content

The information content criterion is based on a 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, using a scale of the marks that goes from 0 (= not applicable) to 5 (= very high level of information content).

Scoring Dimension 3. Technical Complexity

The technical complexity criterion involves a single score providing an overall mark for the feasibility of the specific scenario/option, using a scale of the marks that goes from 0 (= not applicable) to 5 (= very high level of technical complexity).

The following pages present the Information Flow Questionnaire.

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**
❑ **pseudonym used for data linkage, multiple measurements per patient**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
One record for each service episode, centre IDs retained						
One record for each service episode, Centre IDs De-Identified						
Multiple measurements averaged over time interval, centre IDs retained						
Multiple measurements averaged over time interval, Centre IDs De-Identified						

Comments:

14

DATA FLOW QUESTIONNAIRE

<u>**CANDIDATE ARCHITECTURE 1 : INDIVIDUAL PATIENT DATA**</u>

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)**

<u>**SCENARIO 2:**</u>

Administrative Data Service Episode
- ❑ **collected by Local Health Authority**
- ❑ **used by Local Health Authority**
- ❑ **purpose Policy and Planning**
- ❑ **pseudonym used for data linkage, multiple measurements per patients**

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

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal collection of clinical characteristics						
Multiple measurements averaged over time interval						

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**
❑ **pseudonym used for data linkage (over multiple datasets)**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal data collection across relational data-warehouse, 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, 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**
- ❑ **pseudonym used for data linkage (over multiple datasets)**

Option	Privacy				Information Content	Technical Complexity
	Identifiability	Linkability	Observability	Overall	Overall	Overall
Longitudinal data collection across relational data-warehouse, 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, 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 4

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

Comments:

DATA FLOW QUESTIONNAIRE

<u>**CANDIDATE ARCHITECTURE**</u>**: ALL**

Question 2. **SECURITY MECHANISMS: DECISION 1**

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

<u>CANDIDATE ARCHITECTURE 1</u> : INDIVIDUAL PATIENT DATA

Question 3. FORMAT OF BIRO DATABASE: DECISION 1

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: DECISION 1**

**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

<u>**CANDIDATE ARCHITECTURE**</u>**: ALL**

Question 5. **STORAGE AND RETENTION SITE: DECISION 1**

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

5. Panel Discussion

During the Delphi Consensus Panel, the general alternatives have been described in detail through the data flow tables, while a dedicated instrument (data flow questionnaire) has been used to assign marks to each alternative.

Partners agreed that the score for privacy protection should influence the overall score for each option available. Accordingly, the best solution has been identified among those ensuring a higher level of privacy protection, regardless of other criteria

All BIRO partners participated to the scoring exercise in different phases; however, only one vote per partner (represented by the partner's PIA Team Member) has been allowed. Each panelist's score has been anonymous.
A final phase summarized the results of the Consensus Panel into an agreed conclusion describing the best architecture.
All scores have been collected and included in the present PIA report (see next paragraph).

6. Results of the Delphi Consensus Panel

The first seven tables represent individual and anonymous mark assigned by all panelist to any BIRO architectural alternative.

The last table, instead, summarises the consensus marks, agreed after discussion at the Delphi consensus meeting (held in Cyprus).

PANELIST 1

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 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	4	3
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	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	4	2
		Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified	3	4	2
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	3	4	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	4	2
	SCENARIO 3	Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients	0	5	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients	0	4	4
	SCENARIO 4.1	Longitudinal data collection across relational data-warehouse, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	0	5	4
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	0	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	0	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	0	4	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	4
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	4	5	4
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4	5	4
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	5	4
A R C H I T E C T U R E 2	**Personal Data Decision 1**	No Aggregation Size Limit	2	4	3
		Min aggregation N=5 patients per cell	2	4	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	4
		Data aggregated at the level of service centre	3	1	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	2	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	2	2	3
	Transmission Decision 1	All DATE fields transmitted as in original	2	3	2
		DATE fields approximated to time interval (e.g. months)	2	2	1
	Transmission Decision 2	Service Centre ID transmitted	3	2	2
		Pseudonym used for service centre	3	3	1
A R C H I T E C T U R E 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)	0	1	2
	Personal Data Decision 2	Geographical mapping available	2	4	2
		Geographical mapping unavailable	1	1	1
	Personal Data Decision 3	Variability of centres outcomes available	1	1	3
		Variability of centres outcomes unavailable	1	2	2
	Personal Data Decision 4	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	2	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	1	1
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)	3	3	3
	Format	Full information on all medical records	3	5	4
		Averaged over time	2	3	2
	Disclosure	BIRO database administrator	1	0	0
		All local database administrators / registry managers	3	1	1
	Storage/Retention	BIRO Coordinating Centre	2	0	2
		EU/DG-SANCO	1	0	3

PANELIST 2

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 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	3	3
		Multiple measurements averaged over time interval, centre IDs retained	4	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	3	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	5	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	3	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	5	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	5	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	3	3
A R C H I T E C T U R E 2	**Personal Data Decision 1**	No Aggregation Size Limit	3	3	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	3	3
	Personal Data Decision 2	Aggregation across service centres	2	2	3
		Data aggregated at the level of service centre	3	3	3
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	2	2	3
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	3	3	3
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	2	2	3
	Transmission Decision 1	All DATE fields transmitted as in original	3	3	2
		DATE fields approximated to time interval (e.g. months)	1	2	2
	Transmission Decision 2	Service Centre ID transmitted	4	3	2
		Pseudonym used for service centre	3	2	2
A R C H I T E C T U R E 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)	2	1	2
	Personal Data Decision 2	Geographical mapping available	2	3	1
		Geographical mapping unavailable	2	1	1
	Personal Data Decision 3	Variability of centres outcomes available	3	4	3
		Variability of centres outcomes unavailable	2	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	2	3	3
		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	1	0	3
		Client program automatically sending encrypted data (agent)	2	0	4
	Format	Full information on all medical records	4	5	3
		Averaged over time	3	3	3
	Disclosure	BIRO database administrator	2	0	2
		All local database administrators / registry managers	4	0	2
	Storage/Retention	BIRO Coordinating Centre	2	0	1
		EU/DG-SANCO	1	0	2

PANELIST 3

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 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	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	4	3
		Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified	4	3	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	5	5	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	4	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	5	5	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	5	5	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	4	3
A R C H I T E C T U R E 2	Personal Data Decision 1	No Aggregation Size Limit	4	4	3
		Min aggregation N=5 patients per cell	2	4	3
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**3**	**4**	**3**
	Personal Data Decision 2	Aggregation across service centres	2	2	2
		Data aggregated at the level of service centre	**3**	**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	4	3
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	**2**	**4**	**4**
	Transmission Decision 1	All DATE fields transmitted as in original	4	5	3
		DATE fields approximated to time interval (e.g. months)	**2**	**3**	**2**
	Transmission Decision 2	Service Centre ID transmitted	5	3	2
		Pseudonym used for service centre	**2**	**2**	**2**
A R C H I T E C T U R E 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	2	2
		Geographical mapping unavailable	1	1	1
	Personal Data Decision 3	Variability of centres outcomes available	3	4	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	4	3	3
		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	**0**	**0**	**2**
		Client program automatically sending encrypted data (agent)	0	0	4
	Format	Full information on all medical records	3	3	2
		Averaged over time	**2**	**4**	**3**
	Disclosure	BIRO database administrator	**1**	**1**	**1**
		All local database administrators / registry managers	3	1	2
	Storage/Retention	BIRO Coordinating Centre	**2**	**4**	**2**
		EU/DG-SANCO	1	1	4

PANELIST 4

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 1	**SCENARIO 1**	One record for each service episode, centre IDs retained	5	3	1
		One record for each service episode, Centre IDs De-Identified	4	1	3
		Multiple measurements averaged over time interval, centre IDs retained	4	2	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	2	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	5	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	5	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	4	3
A R C H I T E C T U R E 2	**Personal Data Decision 1**	No Aggregation Size Limit	4	4	3
		Min aggregation N=5 patients per cell	3	3	3
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**3**	**3**	**3**
	Personal Data Decision 2	Aggregation across service centres	1	1	3
		Data aggregated at the level of service centre	**2**	**3**	**3**
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	3	2	3
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	3	3	3
		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	4	3	2
		Pseudonym used for service centre	**2**	**2**	**2**
A R C H I T E C T U R E 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	3
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITH restrictions applied on specific stratification criteria	2	2	2
ALL	**Security**	Password access for local administrator prompting client program to send encrypted bundles to BIRO	**1**	**2**	**1**
		Client program automatically sending encrypted data (agent)	2	2	2
	Format	Full information on all medical records	3	4	4
		Averaged over time	**2**	**2**	**3**
	Disclosure	BIRO database administrator	**1**	-	**1**
		All local database administrators / registry managers	3	-	2
	Storage/Retention	BIRO Coordinating Centre	-	-	-
		EU/DG-SANCO	-	-	-

PANELIST 5

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 1	**SCENARIO 1**	One record for each service episode, centre IDs retained	3	5	2
		One record for each service episode, Centre IDs De-Identified	3	4	3
		Multiple measurements averaged over time interval, centre IDs retained	2	3	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	1	2	3
	SCENARIO 2	Population-based longitudinal records, linked across administrative datasets, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	4	4	2
		Population-based longitudinal records, linked across administrative datasets, Centre IDs De-Identified	3	4	1
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	2	3	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	1	3	1
	SCENARIO 3	Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients	4	5	2
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients	3	4	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	4	5	4
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	3	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	3	4	3
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	3	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	4	5	3
		Longitudinal data collection across relational data-warehouse, Portion of relational structure sent / Centre IDs de-identified	3	4	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	3	4	4
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	3	4	4
A R C H I T E C T U R E 2	**Personal Data Decision 1**	No Aggregation Size Limit	4	4	3
		Min aggregation N=5 patients per cell	3	4	4
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**2**	**4**	**4**
	Personal Data Decision 2	Aggregation across service centres	3	4	2
		Data aggregated at the level of service centre	**4**	**4**	**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	2	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied	**1**	**2**	**2**
	Transmission Decision 1	All DATE fields transmitted as in original	4	3	3
		DATE fields approximated to time interval (e.g. months)	**3**	**3**	**3**
	Transmission Decision 2	Service Centre ID transmitted	3	3	3
		Pseudonym used for service centre	**2**	**3**	**3**
A R C H I T E C T U R E 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	3	3	2
		Geographical mapping unavailable	1	2	1
	Personal Data Decision 3	Variability of centres outcomes available	2	3	2
		Variability of centres outcomes unavailable	2	2	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**	**2**	**1**
		Client program automatically sending encrypted data (agent)	1	2	2
	Format	Full information on all medical records	3	5	2
		Averaged over time	**1**	**3**	**1**
	Disclosure	BIRO database administrator	**1**	**0**	**1**
		All local database administrators / registry managers	3	0	2
	Storage/Retention	BIRO Coordinating Centre	**1**	**0**	**1**
		EU/DG-SANCO	2	0	2

PANELIST 6

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

PANELIST 7

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 1	**SCENARIO 1**	One record for each service episode, centre IDs retained	4	4	3
		One record for each service episode, Centre IDs De-Identified	2	4	3
		Multiple measurements averaged over time interval, centre IDs retained	3	3	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	2	2	2
	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	2
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients, centre IDs retained	4	4	2
		Multiple measurements averaged over time interval, Centre IDs De-Identified	3	3	2
	SCENARIO 3	Longitudinal collection of clinical characteristics, Pseudonym used for data linkage, multiple measurements per patients	5	5	3
		Multiple measurements averaged over time interval, Pseudonym used for data linkage, multiple measurements per patients	4	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	5	5	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	4	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	5	4	4
		Multiple measurements averaged over time interval, Pseudonym used for data linkage over multiple datasets, all relational structure sent to BIRO	4	4	4
		Multiple measurements averaged over time interval, Portion of relational structure sent / Centre IDs de-identified	4	4	4
A R C H I T E C T U R E 2	**Personal Data Decision 1**	No Aggregation Size Limit	3	4	2
		Min aggregation N=5 patients per cell	1	3	2
		Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc	**2**	**4**	**2**
	Personal Data Decision 2	Aggregation across service centres	1	1	3
		Data aggregated at the level of service centre	**2**	**3**	**3**
	Personal Data Decision 3	Aggregation of multidimensional patterns (e.g. risk adjustment) NOT allowed	1	2	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed	2	4	3
		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	2	3	3
		DATE fields approximated to time interval (e.g. months)	**2**	**3**	**3**
	Transmission Decision 2	Service Centre ID transmitted	3	4	2
		Pseudonym used for service centre	**1**	**2**	**2**
A R C H I T E C T U R E 3	**Personal Data Decision 1**	NO restrictions on specific stratification criteria (e.g. geographical variable, centres, etc)	1	2	1
		Restrictions applied on specific stratification criteria (e.g. geographical variable, centres, etc)	1	2	2
	Personal Data Decision 2	Geographical mapping available	2	4	2
		Geographical mapping unavailable	1	1	2
	Personal Data Decision 3	Variability of centres outcomes available	2	3	2
		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	4	2
		Aggregation of multidimensional patterns (e.g. risk adjustment) allowed WITH restrictions applied on specific stratification criteria	2	3	3
ALL	**Security**	Password access for local administrator prompting client program to send encrypted bundles to BIRO	**2**	**0**	**1**
		Client program automatically sending encrypted data (agent)	1	0	3
	Format	Full information on all medical records	2	5	3
		Averaged over time	**2**	**3**	**2**
	Disclosure	BIRO database administrator	**1**	**0**	**0**
		All local database administrators / registry managers	3	0	0
	Storage/Retention	BIRO Coordinating Centre	**2**	**4**	**5**
		EU/DG-SANCO	1	1	2

OVERALL CONSENSUS TABLE

A.	Category	Option	P.	I.C.	T.C.
A R C H I T E C T U R E 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
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 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**
A R C H I T E C T U R E 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

41

7. CONCLUSION

The selected features of the BIRO Architectural structure, at conclusion of the Delphi consensus panel, can be summarized as follow:
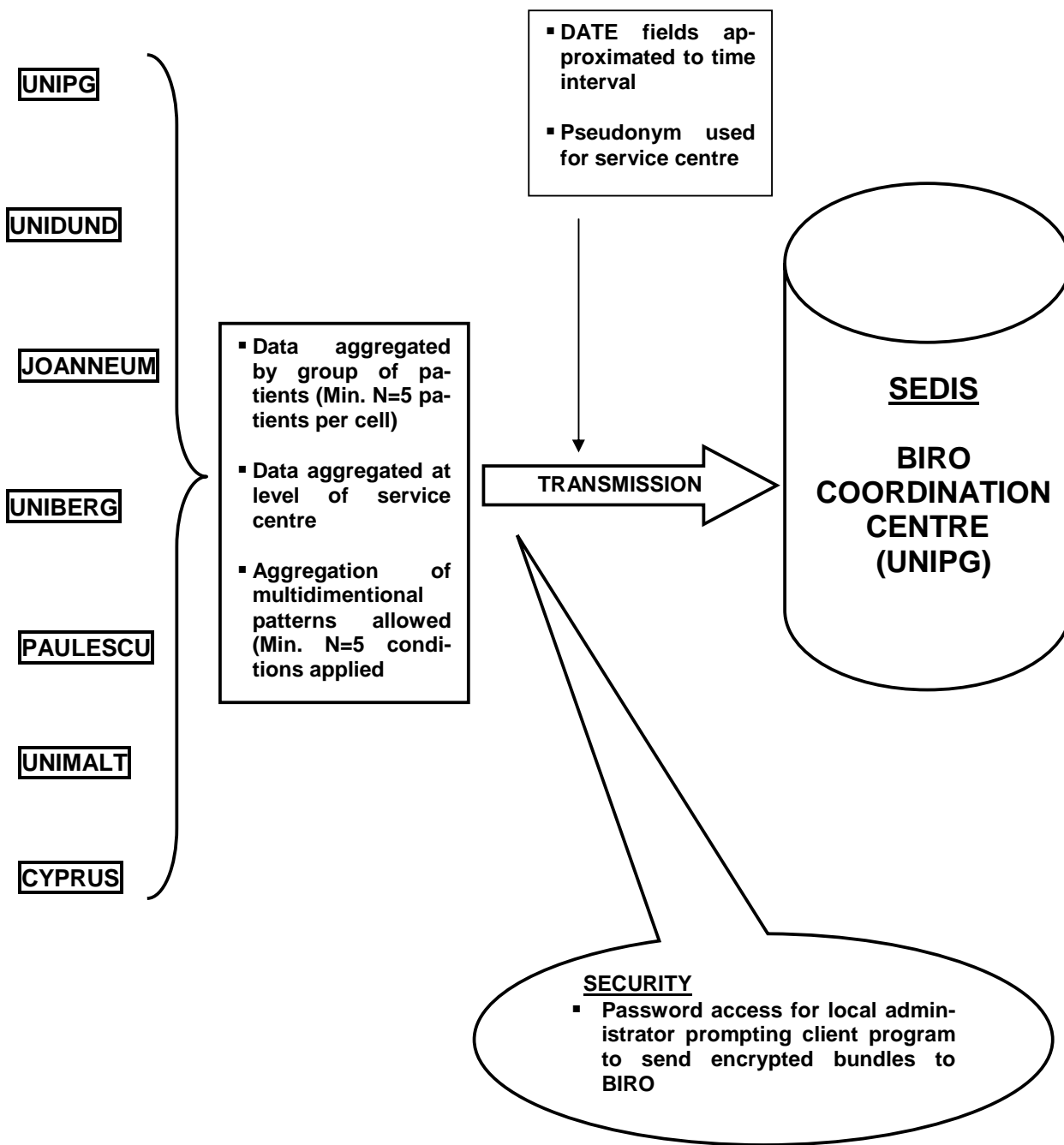
❑ **ARCHITECTURAL OPTION**:
  ▪ Aggregation by group of patients – **DATA FLOW TABLE N. 2** –

❑ **PERSONAL DATA:**
  ▪ Min aggregation N=5 patients per cell, only applicable for high critical privacy variables e.g. service centre, geographical site etc
  ▪ Data aggregated at the level of service centre
  ▪ Aggregation of multidimensional patterns (e.g. risk adjustment) allowed, Min N=5 condition applied

❑ **TRANSMISSION:**
  ▪ DATE fields approximated to time interval (e.g. months)
  ▪ Pseudonym used for service centre

❑ **SECURITY:**
  ▪ Password access for local administrator prompting client program to send encrypted bundles to BIRO

❑ **FORMAT:**
  ▪ Averaged over time

❑ **DISCLOSURE:**
  ▪ BIRO database administrator

❑ **STORAGE/RETENTION SITE:**
  ▪ BIRO Coordinating Centre

The above characteristics identify the selected best BIRO system architecture in terms of privacy protection, balanced with information content, scientific soundness and feasibility of the project in terms of technical complexity.

The following table describes the flow of information through the BIRO Information System, as agreed by all partners, and represents the final data flow table of the BIRO project and the final BIRO System Diagram.

Any potential privacy risk and mitigation strategies will be evaluated and analysed in PIA (Privacy Impact Assessment) Step 3: Privacy Analysis.

B.I.R.O. DIAGRAM

UNIPG

UNIDUND

JOANNEUM

UNIBERG

PAULESCU

UNIMALT

CYPRUS

- DATE fields approximated to time interval

- Pseudonym used for service centre

- Data aggregated by group of patients (Min. N=5 patients per cell)

- Data aggregated at level of service centre

- Aggregation of multidimentional patterns allowed (Min. N=5 conditions applied

TRANSMISSION

SEDIS

BIRO COORDINATION CENTRE (UNIPG)

SECURITY
- Password access for local administrator prompting client program to send encrypted bundles to BIRO

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 trans-mission	Format of BIRO Database	Dis-closed to	Storage or retention site
Aggregation by group of patients: min aggregation N=5, only appli-cable for high critical privacy variables e.g. ser-vice centre, geo-graphical site etc	BIRO partner	One Record for each aggregation level	BIRO partner (local engine), BIRO Consor-tium (central engine)	Computation of single BIRO statis-tical object for local and SEDIS re-porting	DATE fields ap-proximated to time interval (e.g. months) Pseudonym used for service centre	.Password access for local adminis-trator prompting client program to send encrypted bundles to BIRO	Separate sets of aggregated tables linkable by predefined statistical criteria	BIRO database administra-tor	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									

APPENDIX 1: DATA FLOW TABLES EXPLANATORY NOTES

i   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

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

iii   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.

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

v   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

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

vii   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.

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

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

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

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

xii   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.

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

xiv   Institution conducting the epidemiological investigation

xv   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.

xvi   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 Av-LOS=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.

xvii   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.

xviii   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.

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

xx   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.

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

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

xxiii   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.

xxiv   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).

xxv   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).

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

xxvii   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.

xxviii   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.

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

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

xxxi   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.

xxxii   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.

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