European Commission
Directorate C
Public Health and Risk Assessment
Health & Consumer Protection Directorate general

GRANT AGREEMENT
n° 2010 22 02

Developing a European Cancer Information System: a proposal by the European Partnership for Action Against Cancer (EPAAC)
EPAAC-Work Package 9 “Cancer Information and Data”

WP9 DELIVERABLE n. 08
February 2014

Authored by
The EPAAC WP-9 Coordinating team
Milena Sant, Riccardo Capocaccia, Roberta De Angelis, Gemma Gatta, Camilla Amati
and the ECIS document Editorial Board: David Forman, Stefano Rosso, Annalisa Trama, Sabine Siesling, René Otter, Ulrik Ringborg, Francesco De Lorenzo

In collaboration with
EC - DG SANCO
EC - JRC
Italian Ministry of Health

Corresponding address:
Milena Sant, MD
Analytical Epidemiology and Health Impact Unit
Department of Preventive and Predictive Medicine
Fondazione IRCCS Istituto Nazionale dei Tumori
Via Venezian 1, 20133, Milan, Italy
Tel.: +39-02-23903519-2902-2867
Fax: +39-02-23903528
milena.sant@istitutotumori.mi.it
TABLE OF CONTENTS

Foreword ................................................................................................................. page 3
Executive Summary ............................................................................................... page 4
1. Background ........................................................................................................ page 6
2. Steps toward a European Cancer System ......................................................... page 10
3. Possible functional solutions .......................................................................... page 15
4. Conclusions ........................................................................................................ page 21
Foreword

The European Partnership for Action Against Cancer (EPAAC) is a Joint Action co-financed by the European Commission aimed at improving the coordination at the European level of the national health policies on cancer control. The EPAAC involves 38 Associated Partners among Health Authorities, Research Institutes and Patients Associations. The EPAAC addresses all aspects of cancer control activities from health promotion and prevention to health care and a specific task regards cancer information and data (Work Package 9). Main tasks of the EPAAC Work Package 9 (WP9) are to provide an analysis of the current status of cancer information in Europe and to propose, accordingly, a strategic reorganization of the data and information flow, aimed at better supporting cancer control and cancer research activities.

The following document presents a WP9 proposal for the development of a European Cancer Information System (ECIS). The proposal was discussed in several meetings by a dedicated WP9 group of experts, has been submitted for consultation to all EPAAC Partners, and now incorporates most of the comment received.

This document mainly regards the rationale and the requirements of ECIS. Considerations on governance, institutional configuration, and funding mechanisms of ECIS are deliberately not taken into account at this stage.
Executive Summary

Population based information on cancer is fundamental for cancer control activities as well as for health care research and, thanks to the existence of registries and to a long tradition of epidemiological research, is much more available than for most other diseases. In addition, a very wide and high level research activity related to cancer information and data is ongoing in Europe. This considerable amount of experience and resources is not yet fully deployed to produce a proportionally great advance of knowledge in the field of cancer epidemiology. The EPAAC WP9 is developing proposals for the best use of all these resources, overcoming the present fragmentation and duplication of efforts.

To this aim, the need is recognized to start developing the future European cancer information system (ECIS), intended as a public health and research infrastructure functionally connecting all institutions, people, procedures and resources producing meaningful information from cancer data and working within a common framework of concepts, methods, structures and technical standards. It harmonizes the data produced by all these stakeholders and makes the information derived accessible to users under agreed conditions and regulations, providing as much knowledge as possible to facilitate interpretation of the dynamics of cancer in populations.

The repository of cancer registries’ data should be the main component of the system. Gradually, the functionality of the system should be extended to include or link other relevant datasets to allow common analyses, such as detailed clinical information on population based patients samples (collected by the so called High Resolution studies), detailed mortality data at regional level, data from HES/HIS population surveys, from censuses, from administrative sources, etc. Important in this regard would also be the linkage with clinical registries. ECIS should facilitate as much as possible the correlation and the cross analysis between different types of information by implementing a unique and user-friendly interface to synchronised different datasets.

ECIS should take as a major task the continuous improvement of data quality and standardization. Check procedures for high quality data, such as those from cancer registries, will be kept updated according to advances in scientific knowledge and changing information requirements. For other datasets of lower quality level, ECIS activity will be aimed at defining acceptability criteria, developing protocols for comparative analyses, and providing recommendations/guidelines for future quality improvement. Experts (particularly those involved in research activities) should be closely involved in quality analysis.

Within ECIS activities, a comprehensive plan of analysis should be defined for the provision of the main outcomes that will be systematically and periodically released. The plan will identify and agree the priorities; build appropriate flow charts and timetables to coordinate the activities of data collection, quality checks and statistical analysis; identify the subjects taking over the different tasks. ECIS should finally allow appropriate functionality for the public access to the data, and also pursue a regulated but open access at individual level data by the scientific community and other stakeholders. In order to work as efficiently as possible, the ECIS activities should be implemented by pooling existing
resources and experiences from European Institutions that are already involved in cancer information and data diffusion, most of which have already developed knowledge, skill and instruments, and able – if appropriately coordinated – to carry out the main tasks required. The Joint Research Centre (JRC), the European Commission in-house science service, has been identified to establish, and support the ECIS, and to coordinate its further development. JRC is expected to work in close collaboration with all the major stakeholders, including the European Network of Cancer Registries (ENCR), the International Agency for Research of Cancer (IARC), other networks of European scientific institutions (such as those involved in EUROCONTROL) to define the best effective options on all the major ECIS functions, as data collection, quality control, statistical analysis, diffusion and dissemination of cancer information.

**European Cancer Information System (ECIS) in brief**

ECIS is defined as a public health and research infrastructure functionally connecting all institutions, people, procedures and resources producing meaningful information from cancer data and working within a common framework of concepts, methods, structures and technical standards. Its main characteristics are:

- One access point for data retrieval.
- Network-based data analysis and data quality control.
- Operating as a research infrastructure for the cancer community at large.
- Funded on a sustained basis.
1. **Background**

1.1 Within its population of 500 million citizens (more than 600 million when considering Acceding, Candidate, and EFTA countries), Europe presents a huge variability in all aspects related to cancer control. These range from different cancer-related life styles, to environmental risks, available resources, cancer care organization (for example: implementation of comprehensive cancer centres, multidisciplinary teams for oncolgical care, public/private mix, in/outpatient rate), and cancer plans.

1.2 This variability offers a unique framework for cancer research and its application in health care activities. It gives the possibility to study comparatively the results of different cancer control policies in different economic, social, and environmental situations. Learning from differences is essential to build a coordinated European policy against cancer, and to improve efficacy and effectiveness of the undertaken actions. Availability of, and access to high quality cancer information, and in particular epidemiological data, play a crucial role in health care research.

1.3 Cancer has a huge impact on the population health (it is the second cause of death in Europe), and requires a large amount of resources in public health, technology and research. Moreover, cancer is an extremely complex disease, requiring a significant amount of detailed information which must be studied in depth. Population based information is much more widely available for cancer than for most other diseases, due to the existence of registries and to a long tradition of epidemiological research. It is important that such information is used to its fullest capacity, since the non-optimal use of existing data has direct negative implications on public health. All these considerations point to the need to have a cancer-specific information system in Europe.

1.4 Population-based registries provide, in the field of cancer, an added value as comparable disease-specific sources of data is not available for any other major disease. The use of cancer registry data is involved in all phases of cancer control activities, from aetiology and prevention, to early diagnosis, care and rehabilitation of cancer patients, and in planning and evaluation of health care services. The EUROCOURSE project has provided a detailed analysis of the potential role of
population-based registries in cancer information. Population-based cancer registries should provide the core component of a cancer information system. Several other relevant and potentially available data can be added or linked in order to implement a comprehensive information system. Consensus-based lists of the most relevant indicators needed for public health and, more specifically, for cancer control activities have been developed by EU funded projects such as ECHI and EUROCHIP. Presently, not all the listed indicators are available in Europe, and those available are not always of sufficient quality to be meaningfully used in comparative analyses. In Annex 1, a list of these indicators is reported together with their present degree of availability.

1.5 An extraordinarily wide spectrum of activities related to cancer information and data is ongoing in Europe and provide the necessary components for the development of a cancer information system. Europe-wide databases of cancer indicators are maintained by IARC (for incidence and mortality) and EUROCARE (for survival, prevalence and patterns of care) who centralize and analyze cancer registry data. Special population based studies, aimed at analyzing patterns of cancer care and related outcome (the so called High Resolution studies), are carried out with the participation of many cancer registries across the EU by collecting, for representative samples of patients and with internationally agreed protocols, detailed data on stage, staging procedures and treatments. Good examples of linkage between population based and clinical registries are available from Nordic countries and in the Netherlands. The EUROCOURSE Project has developed a gateway for automated cancer registry data centralization and has improved the European Cancer Observatory (ECO) website for the diffusion of cancer registry based indicators. General health related data, necessary for an appropriate interpretation of cancer indicators, are organized within the EU health websites. General and health specific economic data are collected in the OECD database. Finally, the European scientific community is in the forefront of methodological research in population-based epidemiology and public health, from analysis and projection of incidence and mortality trends, to survival analysis, prevalence estimation, planning and conduction of high resolution studies, and for the study of social and economic inequalities on health.

1.6 This considerable amount of experience and resources is still not fully deployed to produce the proportionate advance possible of knowledge in the field of cancer epidemiology. Several structural deficiencies which restrain the potential progress exist, as well as lack of adequate initiatives in
removing barriers between production and use of cancer data. The main obstacle is the lack of a central coordination necessary for transferring among countries good national best practice, and for harmonizing research projects at the European level – in particular, providing them with the required means for continuity and sustainability.

Furthermore, there is a major gap between the potential demand of data from the cancer research community and their actual accessibility. A survey of relevant literature shows a significant gap between the number of research papers based on the US cancer registries' data with respect to those based on data from the network of European cancer registries. One major contributory factor is due to the existence of a US public database holding individual records of cancer patients (the SEER Research Data File), which is accessible to all research institutions under a signed agreement limiting its use to research purposes only. Its worldwide impact is demonstrated by the 4,500 peer-reviewed articles that have been published in indexed scientific journals since the year 2000, generating more 130,000 overall citations (600 of these papers were published in 2012). Over the same period, the number of research papers integrating and jointly analyzing CR data from the network of European CRs does not exceed 500 (60 in 2012).

The SEER experience cannot be immediately applied to Europe, since SEER covers a fraction of one jurisdiction and US federal law regulates SEER activities, whereas the European Union covers 28 countries with different organization of cancer registration activity. Moreover the EU Member States and still present somehow different legislation on data protection (but in perspective a common law should enter in force in the next years). Finally, the 28 EU Member States present an extremely wide variation in funding for cancer information systems, but nothing as well-funded as SEER.

The differences in resources and jurisdictions do not mean, of course, that the same objective is not affordable or achievable in Europe, but that a higher degree of harmonization and accessibility to cancer information should first become a priority in the European agenda. In Europe, no easier alternative way exists to access individual cancer patients data, other than individually contacting and asking permission to each single registry. European level research based on cancer registries data is presently only affordable for large international Institutions or well established networks.
Academic research is by fact hampered. Most of academic works on cancer epidemiology are actually based on national level or on US-SEER data.

Many valuable datasets, collected by and stored in European public Institutions are not accessible, and even their existence is not known to the cancer community and to the general public. For example, this is generally the case for High Resolution data and for clinical registers. A simple mapping of what exists, in terms of content and possible access ways, would be of great help for potentially interested users. Risk factors data from population surveys, screening activity data, information on availability and provision of health care services, macroeconomic indicators, mortality data and other demographic indicators, are all examples of existing and available information not always consistent, each other and with cancer registry based indicators, regarding the referred populations, geographical units, and time periods of. As a consequence, such data can be correlated with difficulty with incidence and survival indicators.

In conclusion, a substantial progress has to be done to move from the sum of presently ongoing activities in cancer information to the establishment of a true information system, able to the optimize their use and the enhance their impact on public health.
2. **Steps toward a European cancer information system.**

Information system is here intended as an infrastructure aimed to bring together and disseminate a broad range of information about cancer in Europe, to facilitate research and, in a broad sense, to reduce cancer burden and benefit cancer patients. In order to fulfil this definition, the future European Cancer Information System (ECIS) should encompass the coordination of the entire process of data organization, quality control, management, analysis, and diffusion. Points 2.1 - 2.5 of this section address the main activities that should be included in the underlying objectives of ECIS and points 2.6 - 2.7 stipulate two fundamental needs of ECIS to ensure its viability.

2.1. Data collection and organization. This is the core activity of an information system. The repository of individual patients’ data, currently collected by population based European cancer registries and allowing estimates to be made of main cancer burden indicators (incidence, survival, prevalence), will be the first and most important component of the system. Detailed clinical information on population based patients samples, as can be provided by periodical High Resolution studies will be also structurally included into the database. Gradually, the functionality of the system will be extended to include or link other relevant datasets to allow common analyses. Important in this regard are clinical based registries containing detailed and updated information (not systematically available to population-based registries) on diagnosis, treatment and patient follow-up (notwithstanding that such data may be not representative of the whole population of cancer patients). Moreover, detailed mortality data at regional level (available from Eurostat) provide valuable information of consistent quality which at present is neither sufficiently known nor adequately utilized. In addition, data from HES/HIS population surveys (such as life style factors), census (education and deprivation), administrative sources (such as those related to health care organization) could be linked by the system. The datasets included in ECIS will be organized to make them efficiently available to different users, according to their specific permissions and credentials.

2.2. Data quality and standardization. Good quality of data is a crucial requirement for the system since data will be used contemporary for reporting, cancer control planning, and research purposes. This requires developing appropriate procedures for checking completeness of data and cleaning possible errors, as well as for coding and transmission of the data. Standardization of data
coding, error checking criteria, information completeness, and data processing is essential to assure exchangeability of the data and to carry out meaningful comparative analyses. ECIS should keep the continuous improvement of both quality and data standardization as a main objective.

For many years, European cancer registries’ data on incidence and survival have been submitted to systematic checks for their quality and standardization by ENCR, EUROCARE, and IARC. However, quality controls cannot be defined once forever, but must evolve according to advances in scientific knowledge and changing information requirements. As an example, the analysis of cure rates has focused attention on the completeness of long-term follow-up in cancer registries, an issue not fully perceived when the same data were used only for incidence or five-year survival estimation. As a further example, the need for including stage as a descriptive variable in incidence, survival and prevalence analysis will require the definition and application of specific quality checks for that purpose.

Other datasets of greatest potential relevance for ECIS are far from ideal in terms of data quality level and degree of standardization. Clinical variables, such as those describing treatment, recovery or relapse, rehabilitation, quality of life, etc, are less frequently collected on a population basis, and their quality and standardization at the European level is in large part unknown. Also, the quality of existing information on health care resources (doctors, radiotherapy and imaging equipment) is heterogeneous and not well known across countries. For these datasets, a dedicated task should be undertaken to: (a) define acceptability criteria; (b) to develop protocols allowing comparative analyses; and (c) to provide recommendations/guidelines for future quality improvement.

Since this work is strictly connected to the use of the data, experts (particularly those involved in research activities) should be closely involved in quality analysis.

2.3. Linkage between different datasets. Linking and correlating different types of information, (e.g. data on cancer incidence with those on risk factor distribution across populations, or provision of health-care services with survival indicators) is of profound importance. ECIS should facilitate as much as possible these kinds of analyses. A unique and user-friendly interface should be implemented permitting access to synchronised different datasets and avoiding the need for time
consuming searches among many heterogeneous, scattered, and not always reliable data sources. This does not necessarily imply duplication of the data if virtual connections can be established whenever appropriate. In many cases, the joint use of different datasets could be further facilitated by simply harmonizing definitions of variables (for example using the most uniform and updated disease classifications, or adopting uniform geographical areas) and measurement scales (age classes, or time periods).

2.4. Unified plan for data analysis. The efficient analysis of a comprehensive European cancer database and its optimal use is a complex undertaking and requires a large amount of analytical work, particularly in view of the large variability of data characteristics in the different cancer registries (for instance regarding national population coverage, year of starting registration, year of last updating, completeness of incidence and follow-up data, reporting delay, completeness and quality of diagnostic, tumour staging and treatment information, availability of cause of death etc.). Also the analysis of data not derived from cancer registries will require a careful consideration of the definitions and coding systems used in the different countries, as well as the management of missing data, the analysis of collinearities, and so on. Efficient management of these problems often requires advanced techniques, such as statistical modelling, and ad hoc development of new methods. There exists however wide experience in Europe both on the methodological grounds and on the application of the most advanced statistical methods of data analysis. A non exhaustive list of important and informative topics to be addressed is:

2.4.1. Incidence and mortality analysis
2.4.2. Survival analysis
2.4.3. Prevalence analysis
2.4.4. High Resolution studies
2.4.5. National estimates for countries with partial CR coverage
2.4.6. Time trend analyses of cancer indicators
2.4.7. Cancer burden forecasts
2.4.8. Joint analyses of CR data with data from other sources (socio-economic, demographic, clinical, health care services, etc.)
Independently on how and by whom these tasks will be carried out, ECIS coordination should in any case develop a single comprehensive plan of analysis for the main outcomes that will be systematically and periodically released. The plan will: identify and agree the priorities; build appropriate flow charts and timetables to coordinate the activities of data collection, quality checks and statistical analysis; identify the subjects taking over the different tasks. Such plan is necessary to optimize the use of the available resources avoiding duplication of efforts, while assuring appropriate coverage of the agreed priority topics.

2.5. Dissemination, diffusion, and data release. Organization and provision of results obtained from the analysis of the data is the final goal of the entire information system. ECIS has to provide for at least three different levels of output information:

- A set of core pre-calculated cancer indicators presented mainly in graphical format, not necessarily very detailed in terms of cancer sites and of descriptive variables (such as age, geographical area, etc.), and accompanied by explanation and interpretation notes. This type of output will be mainly directed to the general public, as well as to policy makers and to the media.

- A wider set of pre-calculated indicators, presented with the highest possible level of detail, and accompanied with systematic information (metadata) on the data sources, as well as the quality and comparability of data, the methods of estimation, and the underlying assumptions. This output is intended to be used by the professional disciplines (doctors, nurses, paramedical staff, health care providers and managers, etc.) that need precise information, but do not have the time, possibility, or resources to carry out the necessary analyses by themselves.

- Quality checked individual records. These are the only data that are fully adequate for in-depth research activities and which do not suffer from the severe limits of data tabulated according to pre-defined variables and categories.

Within ECIS, close interaction is required with the main target user categories in order to ensure the matching of data provision to the associated data needs. The outputs of ECIS will be disseminated via many different channels: general and specialised publications, policy and press communications, information leaflets, and web-based tools.
2.6. Legal basis. The most difficult task in developing the ECIS database is not that of technically improving the organization of and the access to a large amount of data and information. Prior to putting on suitable technological solutions, it is necessary to address a number of obstacles and barriers that will impede data sharing in the widest sense. Some of these difficulties may be of a legal nature due to data privacy regulations. Others concern more the reluctance or distrust of providers themselves to share the data outside the place in which they are produced. These difficulties will be much easier tackled when ECIS could join a recognized scientific authority to an explicit commitment from national authorities. It is necessary to establish the legal and institutional basis of ECIS (based on the consensual decision of single countries) enabling it to receive and redistribute the relevant data and information.

2.7. Regular funding. ECIS should provide decisive contributions both to a better understanding of the determinants and the burden of cancer, and for planning and evaluating health-policy interventions. It is clear that it must be sustainable over many years/decades in order to capture and utilise the correct information. Cancer is not expected to be overcome in a few years, thus ECIS must be planned to run over a time scale of several decades. During its time of operation, ECIS will have to be kept updated with respect to evolving information needs and data availability. Its development and maintenance on a short-term project-by-project basis is therefore untenable. Any uncertainty in the future financial resources available will severely compromise the commitment to its initial development. Unexpected shortage of funds will serve not only to cancel on-going developments but could render the entire information system useless, with the consequent waste of resources and energy already spent. Assuring a sustained source of funds is therefore necessary to progress from the status quo characterised by independent, fragmented, and uncoordinated initiatives towards a coordinated, sustainable, and comprehensive information system.
3. **Possible functional solutions**

The characteristics and the functions required of ECIS, and listed in the previous section can be reached with different organizational options. In this document possible ways to address points 2.1 to 2.5 will be considered. Points 2.6 and 2.7 mainly pertain to policy decisions and fall outside the scope of the EPAAC.

3.1. **Database management.**

The data flowing into ECIS need of course to be organized conforming to a unique and coherent structure. This does not necessarily imply building a single centralized data repository. Use of web-based remote connection to distributed data repositories can be a possible alternative. These two scenarios are considered in more detail below:

- **Centralized database.** The data providers (e.g. cancer registries) will prepare the data files to be included in ECIS according to pre-specified protocols. All the files will be collected and physically stored in a single location. The central database will be managed by a single structure that will be identified as the central data management of ECIS. All the subsequent changes to the existing data will be reported in the central database. The data analysis tasks will be carried out by accessing specific parts of the central database in a "read-only" way and with user-tailored access rights. The central data management of ECIS will be responsible for the data safety and data protection against unauthorized access.

- **Distributed database.** The data providers (e.g. cancer registries) will prepare the data files to be included in ECIS according to pre-specified protocols. However, the original data will remain at the data centres of the institutions that collect them, placed in a storage area accessible from the central data management function of ECIS. Errors in – and changes to – the data will be communicated to the data providers, who will be responsible for modifying the files placed in the accessible area accordingly. The data analysis will be carried out via the ECIS central data management that will automatically access the required data sets from providers of the requested data, collate the data files in a unique data file and make it accessible to the user, in a "read-only" way and with user-tailored access rights. The data providers and the central data management of
ECIS will be both responsible for the data safety and for their protection against unauthorized access.

The first option is simpler and more straightforward to implement than the second option. A further advantage is that a centralised management of files assures a more uniform maintenance of the files released, both in terms of data cleaning and data updating. With a distributed database, it is likely that more quality controls are necessary to ensure homogeneity of the variables, thus requiring resources at the local level. The second option however has the advantage of keeping more strict local control on the use of the data. It does not prevent building special datasets conforming to the ECIS protocols, but avoids their formal submission to a third party.

3.2. Quality controls.

Quality controls necessarily require a close interaction between the data providers and the coordination centre: the former to assure the correspondence between the coded data and the original information, the latter to oversee the overall consistency of all the data contributed to the ECIS database by the different providers. This activity can be organized in at least two different ways, according to whether the data checks are carried out at the local (data provider) or the central level. They are described below taking as an example the quality controls for the individual patients’ data coming from cancer registries.

- **Local data checks.** This solution implies that the procedures for testing the internal consistency (for example between cancer site and morphology, or between date of diagnosis and date of death) are carried out locally at the registry level. Records with errors or possible inconsistencies will be immediately checked by the registry against the original data source (e.g. pathology report or clinical records) and corrected in the registry database before submission, if some error is detected.

- **Centralized checks.** In this case, the checks are systematically carried out at the central level on the files submitted by the registries. Records with errors or unlikely combinations are sent back to the registries to be checked against the original data sources and corrected or confirmed according to the case. The corrections are then reported in the central database.
The two procedures described above should, in principle, be equivalent and lead to the same final dataset if perfectly definite criteria for the correctness of the data could be used. Whereas this is generally the case for true and trivial errors (for instance the date of death cannot be prior to the date of birth), it is not always possible to ensure for less definite warning conditions (such as unlikely topography-morphology combinations) the plausibility of which depends on several non clear-cut conditions (such as age of the patient, microscopic verification, frequency of that specific condition in a given registry in comparison with other registries, etc.). In practice, some influence of subjective judgment is unavoidable. The local checks procedure is more practical and less time consuming. The centralized procedure better ensures the consistency and the standardization of the whole database between data providers.

The most efficient approach could be a combination of the two levels of quality checks.

A first set of automated data checks can be envisaged at the local level, for trivial errors requiring a univocal correction, the detection of which implies the need of correction in order for the record to be accepted (e.g. wrong sequence of dates, incompatible gender-organ association). Centralised data controls could be envisaged for detecting less clear error conditions, which possibly need to be examined comparatively with other registries or with other records within the same registry.
3.3. Data analysis.

As outlined in point 2.4, the systematic analyses within ECIS aimed at providing indicators, specific reports, and data descriptors on a regular basis require some form of centralized planning. This could be achieved via several forms of organization:

- Centralized analyses. According to this option, the analyses will be carried out by an epidemiological and statistical team working at the ECIS coordinating centre. The team will have an unrestricted access to the ECIS database, and will use a set of agreed data analysis procedures to calculate annually the set of indicators necessary to update the ECIS website, to draft periodic reports, and to provide the more general description of the latest data. This option is particularly appropriate for the provision of the core and more consolidated indicators (for example: point estimates of incidence, prevalence, survival and mortality).

- Network-based analyses. This option contemplates that the analyses fall under the responsibility of a group of partners organized into a Data Analysis Network. The Network will be considered as part of the ECIS. Its members will be identified among Institutions with a recognized, established, and leading role in population based cancer epidemiology. Each partner will have access to the central database to carry out a specific set of analyses on the entire European dataset and to deliver the planned set of statistics and indicators. Partners will also interact with ECIS to define the appropriate quality controls of the data relevant to their specific tasks. This organization is appropriate for systematic and periodic, but more complex analyses, requiring the use of statistical modelling, of other advanced methods, or the development of new methods.

- Ad-hoc analyses. It is also possible to entrust specific analyses to external partners, not formally involved in ECIS. This form of collaboration is more appropriate for well delimited tasks, to be carried out in response to specific requests, e.g. for public health or research purposes, such as experimental analysis of new datasets included in ECIS, or reports on very specific topics.

The three types of ECIS data analysis organizational modalities are not meant to be considered mutually exclusive. They can be combined where necessary to implement any given programme of data analysis in the most appropriate and effective way.
3.4. Data release.

Further to performing systematic analyses and regularly providing cancer indicators, ECIS will be a key epidemiological research infrastructure for the European Research Area. As such, ECIS will have to make data available at the highest level of detail possible. Patients’ data as collected by cancer registries (but also high-resolution data, or clinical registry data if accessible from ECIS) is optimally useful for research if access to the individual patients’ record level is possible. A limitation to the full release of the data arises from privacy regulations in European countries. The data to be included in research datasets will have to be consistent with the privacy protection laws in all the contributing countries. Rules for assuring confidentiality will be developed in accordance with the EU legislation. Three main different mechanisms for data release can be envisaged.

- Release under consent from data providers. This is the current procedure for providing individual data from European cancer registries’ databases (as EUROCare and EUROCIM) to research groups for specific aims. A short protocol is submitted and circulated among cancer registries. Only the data from registries that approve the protocol and explicitly give their assent are included in the released dataset. Due to the high number of data providers involved, this procedure requires a very efficient organization and a considerable amount of bureaucratic work to be completed in a reasonable time without an unacceptable number of missing responses. It also requires extra work for the contributing registries, which may become a problem if a high number of requests (as desirable) are presented and have to be evaluated. Another disadvantage is the production of ad-hoc datasets for each specific request, which may hamper the reproducibility of research results.

- Release after approval by a central committee. Also in this case, the team interested in the data have to prepare and submit a study protocol of the research project. The protocol is evaluated by a committee that is delegated by the registries to verify the appropriateness and the quality of the requests under some general and pre-defined criteria. The positive evaluation by the committee is the sufficient condition for the delivery of the requested dataset. This procedure is more effective (in time, work load and resources) than asking consent from each data provider.
• Free public use database. A public use dataset (similar to that provided by the US-SEER) will be created and put online on the public domain. The availability of individual records for users driven analyses will constitute a major advancement with respect to the data now available from EU projects, which are accessible only in tabulated form. Data privacy regulations are different across the EU and in many EU countries they are more restrictive that in the USA. Thus, the variables to be included in such a dataset and their level of detail should be carefully designed to avoid the possibility of disclosing individual patients’ data. Notwithstanding, some countries or registries may be unable or unwilling to participate with their data. In this case, it should be possible for the provider to refuse public data access. In order to test practical solutions and the feasibility of an open access database, a pilot version of the database could be implemented on a voluntary basis. Providing that a critical mass will be reached, the spread of this instrument may encourage all the European registries to participate. The dataset could be periodically updated regarding time period and inclusion new adhering registries.

The three outlined procedures are of course schematic and can be combined in several ways. They do not differ from each other with respect to data privacy and protection issues, since all of them foresee the delivery of individual patients’ data. They rather provide different practical solutions with different trade-offs between an increasing level of openness and distribution of the data and a decreasing control of providers on the use of their own data.
4. Conclusions

The roadmap to a future cancer information system depends on various factors, such as the available resources, the time lines, and mostly its institutional configuration. These factors will also influence the choice among the different technical options, presented in this document, for the development of ECIS. Choosing among these options is therefore outside the scope of EPAAC. However, as part of a multidisciplinary and multinational partnership, EPAAC WP9 is strongly in favour of those solutions capable of promoting the widest participation, among different specialties, countries and institutions, to perform the anticipated ECIS activities and to access ECIS data and information. We believe that, in order to build as efficiently as possible on the existing resources, the ECIS activities could be implemented by pooling the resources and experiences of European institutions already involved in cancer information and data diffusion, most of which have already developed the necessary underlying knowledge, experience and instruments to carry out the main tasks required by an effective ECIS. We also believe that the value of the data (the collection of which is costly and usually financed with public money) is directly related to the extent to which the data are used. Limiting access strongly hampers the effective use and, in some cases, the use *tout court* of the data itself. For these reasons, and with the only limitations necessary to avoid disclosure of sensitive individual information, the data included in ECIS should be openly accessible by the scientific community and other stakeholders, under appropriate regulations.

The Joint Research Centre (JRC), the European Commission in-house science service, has been identified to establish, and support the ECIS, and to coordinate its further development. JRC is expected to work in close collaboration with all the major stakeholders, including the European Network of Cancer Registries (ENCR), the International Agency for Research of Cancer (IARC), other networks of European scientific institutions (such as those involved in EUROCARE) to define the best effective options on all the major ECIS functions, as data quality control, statistical analysis, diffusion and dissemination of cancer information, etc.