

Developing a European Cancer Information System: a proposal from the European Partnership for Action Against Cancer (EPAAC)

EPAAC-Work Package 9 “Information and Data”

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 the whole of institutions, persons, procedures, and resources dealing with cancer information and data, and coordinated to provide the knowledge necessary to optimize cancer control activities. ECIS should unify the coordination of the entire process of data centralization, quality control, management, analysis, and diffusion. It should allow a 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 much efficiently as possible, the ECIS activities should be implemented by putting together 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 able, if well coordinated, to carry out the main tasks required.

The Joint Research Centre, the European Commission in-house science service, should most appropriately take on the task of hosting and managing the central data repository. The European Network of Cancer Registries could play a crucial role in maintaining a strict connection between the ECIS activities and those of the participating cancer registries. The International Agency for Research of Cancer (IARC) should continue to exercise the function of accreditation of cancer registries, training, and provision of comparable cancer information at the global level. A network of European scientific institutions (such as those involved in EURO CARE or EURO COURSE) is the most appropriate organizational structure able to carry out in an efficient way the activities of data quality control, analysis, diffusion and dissemination.

Assuring a permanent source of funds is also necessary to evolve from the present variety of independent and fragmented initiatives to a sustained information system. If based on the concept of networking the already existing technical and scientific capacity, the development and maintenance costs of the cancer information system would be relatively low and sustainable.

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 oncological 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 at best, since the non optimal use of existing data has direct negative implications on public health. All these considerations point to the *need to build 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 EURO COURSE 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. First of all, detailed mortality data at regional level, as available from Eurostat, could provide valuable information of uniform quality which at present is not sufficiently known and adequately

utilized. Furthermore, it is particularly important to consider clinical databases containing detailed and updated information (not systematically available to population-based registries) on diagnosis, treatment and patients' follow-up, which however may be not representative of the whole population of cancer patients. Indicators measuring clinical effectiveness can be constructed connecting population and clinical based registries, aimed, for example, at evaluating the outcome of high costing drugs. Consensus-based lists of the most relevant indicators needed for public health and, more specifically, for cancer control activities have been developed by the EU projects ECHI and EUROCHIP. In addition to those previously cited, these lists include indicators derived from population surveys (such as life style factors), census (education and deprivation), administrative sources (such as those related to health care organization). 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 EUROCCARE (for survival, prevalence and patterns of care) who centralize and analyze cancer registry data. Stage and treatment data are collected by many cancer registries across the EU for representative samples of patients in the framework of high resolution studies, with internationally agreed protocols. Good examples of linkage between population based and clinical registries are available from Nordic countries. The EUROCCOURSE 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 portal. 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 not still fully deployed to produce a proportionally great advance of knowledge in the field of cancer epidemiology. This is mainly due to the lack of central coordination at the European level, a condition necessary, for instance, for the development of a public use database. As it was outlined before, the network of European cancer registries could offer a huge opportunity for public health research. However, a survey of relevant literature shows an impressive 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. This is mainly due to the lack of central coordination at the European level and to the existence in the US the public database of cancer patients' individual records (the SEER Research Data File), which is accessible to all research institutions under the signed agreement that its use would be limited to research purposes. Only individual level data are fully adequate for research activity which, on the contrary, is heavily limited by data tabulated according to pre-defined variables and categories. To set up and manage in Europe an epidemiological database equal to the SEER in the US is one of the priorities indicated by the Eurocan + Plus project.

2 Steps toward a European cancer information system.

The sum of the above listed ongoing activities cannot be considered an information system, that is intended as the whole of institutions, persons, procedures, and resources dealing with cancer information and data, and coordinated to provide the necessary knowledge to optimize cancer control activities. We identify four conditions necessary to reach this objective.

2.1 *Centralization.* One of the aims of EPAAC Work Package 9 (WP9) is to build a unified platform for the diffusion of cancer registry based indicators (incidence, mortality, survival, prevalence and high resolution clinical data) with common definitions regarding cancer entities, geographical areas, reference periods, etc. A cancer information system should move a step ahead, unifying the coordination of the entire process of data centralization, quality control, management, analysis, and diffusion. A unique repository of individual patients' data continuously provided by the European cancer registries will be the starting point of the system. High resolution data collected on a sampling basis will be also structurally included into the data base. Gradually, the database should be enlarged including or linking data from clinical registries and other relevant datasets to allow common analysis. In this way, the database would be used for reporting, websites updating, and data provision for research. The necessary human, scientific, and technical resources will be deployed to assure its managing and optimal use.

2.2 *Open access.* The value of the data is directly related to the extent to which they are used. Unavailable and unused data are of no value. Single institutions are unlikely to be able to exploit the full potential of the data they produce. Furthermore, cancer data should be as much as possible analysed comparatively across countries and regions, and therefore mostly outside the place where they are collected. This requires a regulated but open access to data by the scientific community and other stakeholders. Data collection is costly and usually financed with public money. Limiting

access to the data strongly hampers the efficient use and, in some cases, the use *tout court* of the data itself. For the above reasons, and with the only limitations necessary to avoid disclosure of sensitive individual data, data included in a European cancer information system should be public.

2.3 *Legal basis.* The development of an information system is not limited to simply improving the connection between and the access to different datasets. Preliminary to such technological aspect, is to overcome obstacles and barriers that make difficult the data collection across different countries and their widest diffusion. Some of these difficulties may be of legal type, mainly due to privacy regulations. Others are less stringent, as the reluctance of providers to share data, or the distrust on their correct use outside the place in which they are produced. An information system at European level should join a recognized scientific authority, an explicit commitment from national authorities, and a well defined legal status enabling it to receive and redistribute the relevant data and information.

2.4 *Regular funding.* Information systems must be planned to run as long as the knowledge they provide is expected to be useful. In the case of cancer, this implies the need of planning over a time scale of decades. During its activity period, an information system has to be kept updated with respect to evolving information needs and data availability. Its funding on a project-by-project basis is therefore inefficient and inappropriate. Uncertainty on the financial resources available in the future makes it impossible to plan the development of a system. Unexpected shortage of funds can cancel ongoing procedures, activities, and entire parts of the information system structure, with the consequent waste of already spent resources and energy. Assuring a permanent source of funds is therefore necessary to evolve from the present variety of independent and fragmented initiatives to a coordinated information system.

3 Functions and structure.

The network should be structured to address at least six different functions, each one involving one or more partners. The technical core of the ECIS structure is represented in Figure 1

3.1 *Governance.* Access to information is a key need for effective action. Therefore, decisions about which data will be included, the detail in which they should be analyzed, from whom and how they could be accessed, have direct implications on the effective running of cancer control activities. The information system governing committee should therefore be established with the clear political mandate of taking into account the needs and the points of view of the various actors

involved in cancer control activities (citizens, policy makers, oncologists, patients, data providers, researchers). Its tasks will include the definition of the general programme of activity, the decisions on the allocation of resources, the provision of ethical and legal guidelines for data management, the preparation of a protocol for data access and diffusion, and an oversight of the network activities.

3.2 *Coordination.* An information system has to assure the contribution of many components and their efficient execution of different but interrelated tasks, such as definitions of protocols for data collection and communication, data collection, database organization and updating, data analysis and diffusion, etc. Since the proposed information system is based on a network of different institutions, a technical coordination body is necessary. This role could be taken by a technical committee under the supervision of the governing committee. Its main task is to lay out a working programme adequate for overtaking the objectives stated by the governing committee, and to monitor its accomplishment. The work programme will include the distribution among partners of the various tasks and the definition of the corresponding timetable. The technical committee will also evaluate the scientific and technical feasibility of upgrading the information system by adding new data, introducing innovative data analyses, or improving the channels for diffusion and dissemination of the information.

3.3 *Gateway.* Receiving the data files from data providers, and particularly from cancer registries, is a critical task, from which depends the ability of the entire information system to actually provide relevant, effective and timely information. The data should be collected at the most detailed level as possible and reasonable. Cancer registries data will be provided in individual anonymized form. Information allowing personal data linkage between different datasets will not be included, leaving this task (only if strictly necessary) to the national or local data collection level. The structure aimed to act as a data gateway should maintain individual contacts with all data providers (such as the 200 cancer registries in Europe), run an automated data transfer tool, carry out a careful analysis of data validity and quality, and interact with data providers. An automated data uploading procedure will be used, such as the cancer registry data gateway implemented at IARC by EURO COURSE, or the data uploading system prepared by EURO CARE. The data must be validated through sound quality control procedures, to remove possible errors and clarify uncertainties regarding the submitted data files. Quality is a crucial task for a database to be used

simultaneously for current reporting, cancer control planning, and research purposes. Quality controls cannot be defined once forever, but must evolve according to the evolving information needs. For instance, the analysis of cure rates have recently focused the 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. In experimental sciences, instruments' calibration and measurement error estimation are normally considered inherent tasks of research activity. The same holds in observational epidemiology, and quality control of the data should be done together with research teams involved in the analyses. On the basis of previous experiences in the gathering of cancer registries data at the European level, IARC, ENCR, and EURO CARE will collaborate to implement a comprehensive set of procedures for quality control.

3.4 *Data repository.* A unique place should be in charge of organizing, maintaining, and protecting the information system database. The data repository component will receive from the gateway component the individual and the aggregated data files, already cleaned and validated, providing the informative content of the system. A preliminary list of possible data sets to be included is:

- Individual data from population based cancer registries
- Individual data from clinical registers
- WHO-IARC-Eurostat mortality data (for all causes and cancer)
- Population counts
- Socio economic data aggregated at national and cancer registry level
- Risk factors prevalence aggregated data (tobacco, alcohol, diet)
- Area specific data on health care system resources (hospitals, RT machines, doctors)

The different data sets will be harmonized through a relational database organization. The data repository will be responsible of the correct storing of all data to avoid losses and undue access. This assured, the database will be made accessible to the partners in charge of the analysis and diffusion tasks.

3.5 *Data analysis.* The efficient analysis of a comprehensive European cancer database and its optimal use is complex. Regarding cancer registry data, it is necessary to deal with their large variability in terms of 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, stage and treatment information, availability of cause of death. 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, 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 is a large experience in Europe both on the methodological grounds and on the application of the most advanced methods of data analysis. A non exhaustive list of important and informative topics to be addressed is:

- Incidence and mortality analysis
- Survival analysis
- Prevalence analysis
- High Resolution studies
- National estimates for countries with partial CR coverage
- Time trends
- Cancer burden forecasts
- Joint analyses of CR data with data from other sources (socio-economic, demographic, clinical, health care services, etc.)

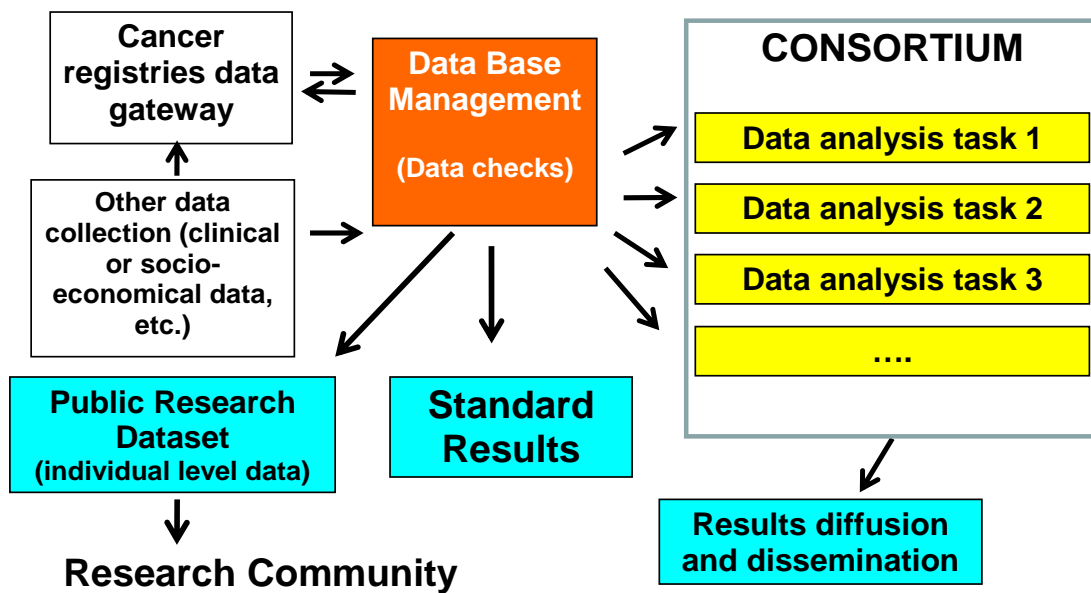
Data analysis will be taken in charge by a group of partners to be identified among Institutions with a recognized leading role in one or more of the above topics. Each partner will have access to the central database, according to pre defined protocols, to systematically carry out a specific analysis on the entire European dataset. Partners will also interact with the gateway component to define the appropriate quality controls of the data relevant to their specific task. Partners will interact with the diffusion component to deliver a planned set of statistics and indicators.

3.6 *Diffusion and dissemination.* Organization and provision of the output information, obtained by the analysis of the data, is the final goal of the entire information system. A dedicated diffusion and dissemination component will systematically receive the set of cancer indicators from the data analysis partners. These will be diffused through all the available channels: publications, press, leaflets, and web based tools. Much activity is already being carried out for the design of websites aimed to the diffusion of data relevant for cancer control at European level. Relevant data on cancer risk factors, demographic and socio-economic variables, and health care services are available through linkage with the HEIDI platform and the OECD database. The EURO CARE

website is providing survival data in an aggregated format for all the registries participating to the project. The European Cancer Observatory (ECO) was designed to provide all the basic epidemiological indicators derived from cancer registries data, as well as WHO mortality data, with different levels of detail according to different needs of potential users. Further, ECO can be easily implemented to address all the topics listed in paragraph 3.5. The diffusion and dissemination component will use as much as possible the work already done in terms of website design, protocols of data presentation, and possibly it will also use already developed software. Its objective is to make available the ECIS information through a dedicated official website under the EU health portal. The website will allow easy connection between data derived from cancer registries and from other sources. It will improve the information available by including new indicators provided by the data analysis groups (such as, for example: incidence and survival by stage and treatment, time trends indicators, survivorship, cure and appropriateness of care). Experts in technical design of websites, in communication techniques and in epidemiology will be involved in order to provide such wide range of information with the necessary scientific rigor but also with the greatest possible clarity.

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 huge advancement with respect to the data now available from EU projects, which are accessible in tabulated form. Data privacy regulations are different across EU and in many EU countries they are more restrictive than in the USA. Thus, the variables to be included in such dataset and their level of detail should be carefully designed to avoid the possibility of disclosing individual patients' data. Rules for assuring confidentiality will be developed in accordance with the EU policies that are currently under discussion at the EU parliament and will be hopefully soon released. Notwithstanding the adoption of these regulations, some countries or registries could be unable or unwilling to participate with their data. A pilot version of the database will be implemented on a voluntary basis. Providing that a critical mass will be reached, the spread of this instrument will encourage all the European registries to participate. The dataset will be periodically updated with respect to time period and to new adhering registries.

Figure 1. Technical structure of the proposed European Cancer Information System (ECIS), as described in Paragraphs 3.3-3.6



4 Implementation

The roadmap to a future cancer information system depends on various factors, such as the available resources, the timetable, and mostly its institutional configuration. In order to build as much efficiently as possible on the existing resources, the ECIS activities could be implemented by putting together the resources and experiences from European institutions already involved in cancer information and data diffusion, most of which have already developed knowledge, experience and instruments able, if well coordinated, to carry out the main tasks required. This kind of approach presently appears the most feasible one, since its implementation requires relatively limited resources and time.

A European Commission institution should most appropriately take on the task of hosting and managing the central data repository. As the European Commission's in-house science service, the Joint Research Centre (JRC) has been identified as best suited for the role. JRC is independent of all national, private and commercial interests and has a proven track record (since 1957) in the harmonization and standardization of scientific/technical processes and systems. It will coordinate the implementation of ECIS in full collaboration with all the major stakeholders to leverage maximum impact and build on the foundations already laid in earlier projects. In

particular, it will support the governance and technical coordination processes and will take on the responsibility of releasing the official cancer statistics in liaison and agreement with the stakeholder community.

The European Network of Cancer Registries (ENCR) could play a crucial role in maintaining a strict connection between the ECIS activities and those of the participating cancer registries. ENCR is a scientific association and does not have dedicated staff and autonomous work capacity. However, as representative of European registries, and carrier of their scientific, technical and organizational views, ENCR will be included in the management of ECIS.

A Consortium of European scientific institution is the most appropriate organizational structure able to carry out in an efficient way the activities of data analysis, diffusion and dissemination. Various mechanisms (such as Joint Actions, Calls for Tenders or competitive Calls) can be considered for the identification of the partners of the Consortium. It should be an open Consortium, initially involving excellence research groups (such as those involved in EURO CARE or EURO COURSE). In perspective, it will be aimed at progressively spread the acquired know-how and the innovation capacity to young scientists and technicians from all European member states. Particularly important will be the increasing involvement of new Member States and of the less affluent areas in the EU. Mechanisms such as residential courses, visiting fellows and exchanges of personnel proved to be effective in the framework of ENCR and could be used to this purpose.

The International Agency for Research of Cancer (IARC) exercises the function of accreditation of cancer registries worldwide and of permanent training of their personnel. Further, the Agency provides comparable cancer information at global level. It is necessary that IARC will continue these activities in collaboration with ECIS, in order to maintain the quality of European cancer registries at highest international levels, and to continue assuring the full comparability between European and world cancer statistics. Research groups at IARC could be part of the Consortium according to their specific research expertise. The terms of the collaboration of IARC to the ECIS will be defined in the framework of the present regulations.

If based on the concept of networking the already existing technical and scientific capacity, the costs of developing and maintaining the cancer information system would be relatively low and sustainable. Its funding should not necessarily subtract resources to the normal EU funding programmes for research and innovation programmes. To avoid this, direct funding from Member States can be considered. If a majority of Member States could be involved, a modest amount of money would be needed from each of them.