

Task 7.1 Advancing registries data through integration with administrative data sources

Mapping data sources available for linkage with registry data

Author(s):	Lead authors: Roberta De Angelis, Silvia Rossi, Giovanna Tagliabue Co-authors: Andrea Tittarelli, Giulio Barigelletti, Milena Sant, Silvia Francisci, Annalisa Trama, Pamela Minicozzi, Daniela Pierannunzio, Alice Bernasconi, Camilla Amati
Version:	2.0
Date:	28. 11. 2019

Contents

Abbreviations	3
Executive summary	3
1 Task 7.1	5
1.1 Background	5
1.2 General aim	6
1.3 Methods: the survey questionnaire	6
1.4 Results	9
1.4.1 Survey management	9
1.4.2 Characteristics of participating cancer registries	9
1.4.3 Legal and privacy conditions	10
1.4.4 Electronic data sources accessible to the CRs	12
1.4.5 Access to patients' diseases other than cancer	15
1.4.6 Data sources completeness and relevance to reconstruct pathways	16
1.4.7 Feasibility of transmission to iPAAC	17
1.4.8 File formats of the accessible data sources	17
1.4.9 Desired data sources	19
1.5 Discussion	19
1.5.1 Increasing and heterogenous use of electronic data sources	19
1.5.2 Data sources availability	20
1.5.3 Privacy and data protection	21
1.6 Concluding remarks	21
1.7 References	23

This report arises from the Innovative Partnership for Action Against Cancer Joint Action, which has received funding from the European Union through the Consumers, Health, Agriculture and Food Executive Agency of the European Commission, in the framework of the Health Programme 2014-2020. The content of this report represents the views of the author/s only and is his/her/their sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains. The authors are not responsible for any further and future use of the report by third parties and third-party translations.

Abbreviations

EU	European Union
iPAAC	Innovative Partnership for Action Against Cancer
EPAAC	European Partnership for Action Against Cancer
CANCON	Cancer Control Joint Action
CRs	Population-based cancer registries
MS	Member States

Executive summary

The aim of Task 7.1 is to map health administrative data sources accessible for individual linkage with data from cancer registries (CRs) in Europe. Task 7.1. investigated the availability of information on clinical patients' pathways, on additional determinantssuch as socio-economic status, and on direct costs, taking into account ownership, data quality and standardisation, legislative background.

Task 7.1 informed the pilot data collection of Tasks 7.2-7.4, it allowed to assess the registries' possibility to participate and to design the protocols for data collection and analysis. Protocols were tailored according to the availability of key data sources, to their structure and legal conditions for data sharing.

A survey on available electronic data sources for individual linkage was conducted among the European registries involved in iPAAC WP7. The list of electronic data sources considered entails: hospital discharges, private hospital/clinical facilities, hospital oncology registries, pathology laboratories, outpatient reports, autopsy reports, medical records, general practitioners databases, death certificates/mortality, hematology laboratories, radiotherapy departments, screening programmes, hospices, mandatory notification, health insurance, other hospital records, biological data banks, health population registry, health survey, disability registry. Access to socio-economical data sources on financial assistance, social benefits, employment, marital status, offspring birth characteristics were also explored.

Overall, 27 population-based CRs from 14 different countries replied to the questionnaire (out of 33 contacted registries). These were: Norway, Netherlands, Belgium, Crete, Malta, Portugal, Italy, Spain, Bulgaria, Croatia, Czech Republic, Poland, Slovenia, and Serbia..

The most commonly used data sources resulted to be hospital discharges (96%), death certificates (89%) and pathology laboratories records (85%). The use of other data sources considered in the survey appears to be more heterogeneous. The number of CRs accessing drug prescriptions, biobanks and sources regarding socio-economic status is very limited.

One of the major limitations documented by the survey is the lack of legal mandate to access information on diseases other than cancer. This limits the possibility to investigate comorbid conditions and late effects of treatments in cancer survivorship. Another important limitation regards the fact that screening programmes data cannot be linked to individual patients' data in about 50% of cases.

Health administrative data sources available for linkage to European cancer registries are not homogeneously accessible and used. This situation reflects heterogeneous health care systems, data owners, legal frameworks and socio-economic conditions.

A significant proportion of registries, however, incorporates these data sources in their routine activity and for research purposes. Part of these sources are sufficiently standardised in terms of coding classification and data structure and can be considered valid for deriving comparable indicators on cancer care (e.g. hospital claims).

1 Task 7.1

1.1 Background

Cancer is the second leading cause of death for the European population. For years, the European Union (EU) has been promoting cancer control programs aiming to reduce cancer incidence and mortality (European Partnership for Action Against Cancer – EPAAC, 2009–2014), and the delivery of evidence-based recommendations for cancer control and care aiming to reduce inequalities in cancer across the EU countries (the EU Cancer Control Joint Action – CANCON, 2014-2017).

The EU also supports a data-based cancer policy. Reliable, comparable, high-quality data and indicators on cancer are essential to improve prevention and control programmes across the EU.

Population-based cancer registries (CRs) are a fundamental source of objective cancer data, and are thus indispensable for the evaluation of the cancer burden, and to design effective cancer control plans.

Besides registries, healthcare systems rely on other relevant data sources: hospital discharge records, clinical registries, administrative and research data. As recommended by previous European Joint Actions, a better integration of all the above health data flows is needed in order to cover the whole clinical pathway, and to measure the performance of cancer care services.

In the last twenty or thirty years, registries are progressively shifting from the use of paper clinical records to the use of electronic files produced for administrative or care purposes, such as claims records or pathology report files. The increasing use of electronic files allowed CRs not only to improve the registration process, but also to start studying different aspects of oncological care, such as hospital trajectories of cancer patients or cancer cost profiles (Forsea AM, 2016).

Many studies have been performed in Europe and worldwide to address the issue of completeness and accuracy of this kind of electronic files. The quality of these sources has revealed to be dependent on the region and on the year of compilation. It has been more and more important in this setting to understand European heterogeneity in the use of electronic files, the availability of these files to cancer registries, the information reported on the files and the quality of the reported items.

1.2 General aim

One of the objectives of iPAAC WP7 is to advance cancer information at the population level by piloting the integration of population CRs data with other electronic data sources to derive key indicators on the overall patients' trajectory. In the pilot studies envisaged in Tasks 7.2-7.4, three different domains are explored in selected European countries or regions, these are: cancer care pathways and quality of care, cancer care costs, and long-term sequelae in Adolescents and Young Adult (AYAs) cancer survivors.

In order to achieve such targets, it was decided to have the status-of-the art on data sources available and used by CRs as a core activity of Task 7.1 .

The aim of Task 7.1 is therefore a mapping exercise on the clinical/administrative sources of information eligible for individual linkage with cancer registry data across the EU, taking into account ownership, data quality and standardisation, and the different legislative background in the Member States. Scopes of the investigation are the availability of information on the whole cancer patient trajectory or clinical pathway (risk profile including information on screening programmes, comorbid conditions, patterns of care, progression and prognosis, late/adverse effects, second tumours), and information on additional determinants, such as socio-economic status, and on direct costs associated to cancer care pathways..

Task 7.1 is horizontal and informs pilot data collection performed in Tasks 7.2-7.4. Its activity is articulated in two main goals:

- census of data sources useful and accessible for electronic linkage to the European cancer registries;
- supporting Tasks 2-4 in assessing the feasibility of registry participation and in defining the protocols for data integration. Indeed, such protocols are driven by preliminary information about the key data sources used, the structure and the legal conditions regarding data sharing , for each CR..

1.3 Methods: the survey questionnaire

To achieve Task 7.1 aims, a survey on electronic data sources accessible to cancer registries was conducted among registries of the European countries involved in iPAAC WP7.

A questionnaire (Annex 1) was developed in the second and third trimesters of 2018 integrating the different information needs emerged from discussions within WP7 team. The survey questionnaire was organized in seven different sections related to:

1. CR features;
2. Legal and privacy conditions;
3. Operational modes,
4. Data sources description;
5. Suggestions;
6. Accessible data sources file formats;
7. Supporting information.

More in detail the different sections were organised as follows.

1. CR features

In this section the registry supplied general information about its organization, the area covered and relevant contacts.

2. Legal and privacy conditions

This section concerned the legal framework of cancer registration, such as the presence or not of country-specific legal restrictions for linking patients' records with data sources not routinely used for registration, the possibility to access to complete information on patients' health conditions (or only to that directly related to cancer), and information regarding personal data sharing conditions. Besides, information on the level of data anonymisation was explored, considering that individual linkage of different data sources is a pre-requisite for pilots 7.2-4.

3. Operational mode

This section collects information on the period of diagnosis covered by the CR, the end-of-follow-up date, cancer sites (all or only specific cancer entities), and registration mode (active, passive or mixed) of cancer cases.

4. Data sources description

This represents the core section of the questionnaire in which information on electronic data sources used for registration, or potentially accessible to the CR, is collected. For each data source listed, the registry indicates: the period of availability, if it is limited to cancer disease or

whether it includes also other diseases affecting cancer patients, the level of completeness of the source and its relevance for the registry's work (in case the source is routinely used for registration, the more the source is used, the higher the relevance measured), the possibility to perform record linkage and to share linked data with external researchers for iPAAC studies. The file format for each single data source is also requested and the specific instructions on how to provide them are given in the subsequent Section 6. The list of electronic data sources considered includes: hospital discharges, private hospital/clinical facilities, hospital oncology registries, pathology laboratories (pathology lab), outpatient reports, autopsy reports, medical records, general practitioners databases, death certificates/mortality, hematology laboratories, radiotherapy departments, screening programmes, hospices, mandatory notification, health insurance, other hospital records, biological data banks, health population registry, health survey, disability registry. Moreover, in accordance with the aims of WP7 pilot studies, the accessibility of socio-economical data sources gathering information on financial assistance, social benefits, employment, marital status, characteristics of the offspring birth is also explored. There is also the possibility to insert additional sources not present in the list, but used or accessible to the registry. Finally, the last part of this section is devoted to the desired sources and gives the possibility to list data sources the registries consider useful, although not available at the moment for reasons not depending on the registry's will, for example for legal or privacy impediments, matters of cost, format difficult to treat.

5. *Suggestions*

This is a free space where each cancer registry can supply additional information and details to support answers or suggestions to improve its compliance to the project.

6. *Accessible data sources file format*

This section describes how to provide the file format of each electronic data source available for linkage, in order to obtain standardised information as much as possible. These file formats are essential to understand the availability and degree of standardization of key cross-country variables and to assess the feasibility of using each specific data source in WP7 pilot studies. An example of file format is supplied (Annex 2).

7. Support.

This last section includes the contacts information of the working group managing the survey in order to get support in case of problems or doubts. Indications on how to fill in and send the questionnaire are also reported here.

The questionnaire was sent to all European CRs participating to tasks 7.2-7.4 of iPAAC WP7. The questionnaire was implemented with Adobe Acrobat, using a special PDF module designed for the automatic acquisition in electronic format of the data. So all data sent in pdf format were easily transferred into an Excel worksheet.

All analyses were then conducted using Microsoft Excel features to group, order or select data, to calculate sums, average values, percentages and so on.

1.4 Results

1.4.1 Survey management

The questionnaire was sent by mail to a total of 33 CRs who expressed interest in participating in the different WP7 pilot studies. To increase participation, several reminders were sent out. The first mail was sent in October 2018, with 5 registries responding. At the end of November 2018 a new mail was sent and 14 CRs sent back the compiled questionnaires. After a third message (February 2019), other 6 registries answered. Finally, a last reminder was sent at the beginning of July 2019, and between mid-July and August two more registries replied. Overall, 27 CRs (82%) participated to the survey.

1.4.2 Characteristics of participating cancer registries

The study involves 27 European population-based CRs from 14 different countries (**Table 1.1**). Among the 27 participating registries, 3 (11%) belong to Northern and Central Europe (National CRs of Norway, Netherlands, and Belgium), 18 (67%) belong to Southern Europe (Crete, Malta, North Region of Portugal, 9 Italian and 6 Spanish CRs), and 6 (22%) are located in Eastern Europe (National CRs of Bulgaria, Croatia, Czech Republic, Poland, Slovenia, and the Central Serbia CR).

For Poland, two different organizations participated to the survey: Lower Silesian cancer registry and the National Health Fund. The Lower Silesian CR does not have access to administrative or clinical data, thus data linkage with external electronic data sources does not

result to be possible. The National Health Fund has access to other data sources and it is able to link its data to the National Cancer Registry. Only information related to the National Health Fund are reported here.

Cancer registries started the registration activity of cancer cases in different periods thus the time extension of available databases is varying. In Slovenia and Norway, cancer registration started back in the fifties. Eight CRs were established in the period between the end of the 70s and the end of the 80s, other eight during the 90s, and nine CRs started their activity from 2000 on.

Almost all CRs reported the date of last life-status ascertainment performed on their incident cases (follow up date): 37% in 2018, 33% in 2017, 19% in 2016, 4% in 2015.

Two CRs (7%) did not indicate a precise date, but they declared that cases are followed up only for five years after incidence date or that their follow-up is up to date.

As for the registration mode, the wide majority operates in mixed mode (67%), whereas 26% works with active mode, and only 7% with fully passive mode.

1.4.3 Legal and privacy conditions

A crucial issue CRs have to deal with during their activity is the protection of personal data. The situation became more difficult with the new European Regulation on Data Protection (GDPR) entered into force since May 2018 and imposing several rules for safe personal data treatment.

In the questionnaire, the legal conditions for accessing, linking, and sharing external datasets (not routinely used for registration) were explored twice, first in general and then more specifically for each reported data source.

In the majority of cases (20 out of 27, 74%) the registries declared to be authorized to link individually administrative sources not routinely used to assess cancer incidence (**Table 1.2**). As for data sources not legally accessible by default, 64% of the registries (14 out of 22) declared there are specific procedures to obtain linkage permission from the competent authorities.

In this first list of questions, the conditions related to personal data sharing were also investigated. Overall, 19 out of 26 registries reported to be allowed to share anonymised patients' data for international studies, such as those envisaged in iPAAC. A possibility to activate procedures to obtain legal permission to share anonymous linked data was documented by 6 out of the 7 registries reporting data sharing limitations.

Table 1.1: List of cancer registries answering the survey

#	Country	Cancer registry	Incidence data		Last follow up date	Registration mode
			from	to		
1	Belgium	Belgian Cancer Registry	2004	2016	2018	passive
2	Bulgaria	Bulgarian National Cancer Registry	1980	Up to now	2018	mixed
3	Croatia	Croatian National Cancer Registry	2001	2016	2017	mixed
4	Czech Republic	Czech National Cancer Registry (CNCR)	1977	2016	2018	mixed
5	Greece	Cancer Registry of Crete (CRC)	2013	2017	2018	mixed
6	Italy	Cancer Registry of Siracusa Province	1999	2015	2018	mixed
7	Italy	Friuli Venezia Giulia Cancer Registry	1995	2013	2017	mixed
8	Italy	Integrated Cancer Registry of Catania-Messina-Enna	2003	2015	2018	mixed
9	Italy	Napoli-3 SUD Cancer Registry	1996	2016	2018	active
10	Italy	Palermo Province Cancer Registry	2003	2015	2017	active
11	Italy	Puglia Cancer Registry (network of 6 local CRs)	minimum year range 2003-2014	maximum year range 2012-2015	2017	mixed
12	Italy	Reggio Emilia Cancer Registry	1996	2017	2017	active
13	Italy	Trapani-Agrigento Cancer Registry	2002	2013	2016	active
14	Italy	Veneto Cancer Registry	1987	2010, 2013	2017	active
15	Malta	Malta National Cancer Registry	1993	2017	Up to date	mixed
16	Netherlands	Netherlands Cancer Registry	1989	2017	2018	active
17	Norway	Cancer Registry of Norway	1953	2017	2017	passive
18	Poland	National Health Fund, on behalf of National CR	2010	2019	2019	passive
19	Portugal	North Region cancer registry (RORENO)	1988	2012	2017	mixed
20	Serbia	Cancer Registry for Central Serbia	1999	2015	2016	mixed
21	Slovenia	Cancer Registry of Republic of Slovenia	1950	2015	2018	mixed
22	Spain	Basque Country Cancer Registry	1986	2015	2016	mixed
23	Spain	Castellon Cancer Registry	2004	2015	2016	mixed
24	Spain	Cuenca Cancer Registry	1993	2012	2015	mixed
25	Spain	Girona Cancer Registry	1994	2015	2018	mixed
26	Spain	Granada Cancer Registry	1985	2014	2016	mixed
27	Spain	Murcia Cancer Registry	1983	2015	5 years	active

Table 1.2: Questions concerning privacy and answers of CRs

Questions	Yes	No
2a) Are you legally permitted to link the registry personal record data with health or administrative data sources not routinely used for registration?	20	7
2b) For the data sources that you are not permitted to link, are there specific procedures to obtain permission by the competent authorities?	14	8
2c) Are you legally permitted to provide case data linked to other data sources in anonymized format for international studies (e.g. iPAAC)?	21	6
2d) If in 2c) you answered No, is it possible to activate a procedure to provide the linked anonymous data to iPAAC?		6 yes

1.4.4 Electronic data sources accessible to the CRs

The availability of clinical or administrative data sources is quite varying among the CRs and although some registries started registration activity in the 1950s, the use of electronic data sources is more recent.

The most common used and better known data source is represented by hospital discharge records: 26 out of 27 CRs (96%) use them to detect incident cases residing in their catchment area. The second and third most used data sources the registries use are death certificates (89%) and pathology laboratory records (85%) (**Table 1.3**). The above data sources, together with medical reports, represent the essential basis of information to generate cancer incidence data and they are used by almost all interviewed European registries.

The use of other data sources considered in the survey is more heterogeneous. Screening programs' databases are available since the beginning of the nineties and are used by 56% of the registries. Information derived from private hospital/clinical facilities by 52% of registries, outpatient reports by 48%, autopsy records by 44, general practitioners databases and hematology laboratories data are used by 37% of the registries. Medical records are consulted by 33% of the CRs and radiotherapy departments data by 30%. Hospice databases, supplying information on the treatment of terminal patients, are used only by 11% of responding CRs.

Only three CRs (11%) have the possibility to access also to biological data banks, three (11%) to health population registries and three (11%) to invalidity registry databases.

Only two CRs (8%) have access to territorial pharmacy prescriptions and in-hospital prescriptions databases.

The number of CRs accessing sources regarding socio-economic status, employment data, financial assistance and social benefits is very limited.

Time range of availability changes depending on the source considered. The most widely used sources are also the most updated (hospital discharges, death certificates, pathology laboratories, outpatient reports, radiotherapy departments); hematology laboratories, hospices and invalidity registries data are updated to 2018.

Table 1.3: Electronic Data sources accessed by the CRs

Data Source description		CRs accessing the data source		Permission to share linked data	
		N (a)	Overall proportion %	Yes (b)	Proportion b/a %
I.	Hospital discharges	26	96%	17	65%
IX.	Death certificates /mortality	24	89%	17	71%
IV.	Pathology laboratories	23	85%	13	57%
XII.	Screening programmes	15	56%	14	93%
II.	Private hospital / Clinical facilities	14	52%	11	79%
V.	Outpatient reports	13	48%	7	54%
VIII.	General practitioners	12	44%	8	67%
VI.	Autopsy reports	10	37%	8	80%
X.	Haematology laboratories	10	37%	9	90%
III.	Hospital oncology registries	9	33%	5	56%
VII.	Medical records	9	33%	3	33%
XI.	Radiotherapy departments	8	30%	5	63%
XIV.	Mandatory notification	7	26%	4	57%
XV.	Health insurance	4	15%	1	25%
XIII.	Hospices	3	11%	2	67%
XIX.	Health Survey	3	11%	1	33%
XVI.	Other hospital records	3	11%	2	67%
XVII.	Biological data banks	3	11%	2	67%
XVIII.	Health population registry	3	11%	1	33%
XX.	Invalidity register	3	11%	0	0%
XXII.	Financial assistance	2	7%	1	50%
XXIV.	Employment	2	7%	1	50%
XXI.	Education	1	4%	1	100%
XXV.	Marital Status	1	4%	1	100%
XXVI.	Characteristics of the birth of the offspring	1	4%	1	100%
XXIII.	Social benefit	0	0%	0	-
XXVII.	Other 1: Ticket exemptions	4	15%	2	50%
XXVIII.	Other 2: Drug prescriptions	2	7%	1	50%

1.4.5 Access to patients' diseases other than cancer

Information on comorbidity conditions and late/adverse effects of therapies is necessary to accurately describe the complete pathway of cancer patients. However, some registries have access by law only to data strictly needed to detect cancer cases, and access to information on other diseases occurring in these patients is forbidden. Such legal restrictions may limit the possibility to reconstruct the whole pattern of care, therefore a specific question was included in the questionnaire.

From the survey's results (**Table 1.4**) it emerges that, among registries using hospital discharges, only half (46%) can access discharge records related to other pathologies affecting cancer patients during their life course. Similarly, only 54% of registries have access to death certificates for all diseases, while the rest can only access oncological causes of death.

Other sources could potentially provide information on co-morbidities are outpatient reports or general practitioners' databases, but they are rarely used by CRs. Among the registries using outpatient reports or general practitioners' databases, a minority have access to all pathologies, i.e. 31% and 25% respectively..

Table 1.4: Available information by type of disease for main data sources (number and % proportion of registries accessing information)

Data source	All pathologies	Only cancer	No information
Hospital discharges	12 (46%)	14 (54%)	0
Death certificates	13 (54%)	10 (42%)	1
Pathology laboratories	4 (17%)	19 (83%)	0
Screening programmes	0 (0%)	15 (100%)	0
Private hospital / Clinical facilities	3 (21%)	11 (79%)	0
Outpatient reports	4 (31%)	9 (69%)	0
General practitioners	3 (25%)	8 (67%)	1
Autopsy reports	2 (20%)	7 (70%)	1
Hematology laboratories	1 (10%)	9 (90%)	0

1.4.6 Data sources completeness and relevance to reconstruct pathways

For each accessed data source, registries were asked to indicate the level of completeness (proportion of cases caught by the source), its relevance for the contribution to the definition of cases and its importance to reconstruct the complete therapeutic pathway of patients (**Table 1.5**).

Considering the six most frequently used data sources, hospital discharges were reported to have the highest completeness level on average (83%), followed by death certificates (79%), pathology laboratories (77%), screening programs data (73%), private hospitals (73%) and outpatient reports (48%).

The highest relevance for the definition of cases is attributed to pathology laboratory records, used for the histological verification of cases, with an average value of 62%, then to hospital discharges (37%) and to outpatient records (32%).

To reconstruct the therapeutic pathway faced by cancer patients during the course of the disease, hospital discharges were reported to supply the majority of useful information (69%), followed by death certificates (54%), pathology laboratory records (42%) and outpatient records (27%).

Table 1.5: Relevance, completeness and importance to reconstruct care pathways for main data sources, % proportions

Data sources	Completeness % (average)	Relevance % (average)	Pathway % (proportion)
Hospital discharges	83	37	69
Death certificates	79	30	54
Pathology laboratories	77	62	42
Private hospitals	73	25	31
Screening programmes	73	20	23
Outpatient reports	48	32	27

1.4.7 Feasibility of transmission to iPAAC

Data centralization in WP7 pilot studies has the clear advantage of ensuring the application of common procedures for quality checks, selection criteria and data analysis. In this logic, once having linked patients records in their database with the relevant accessible external data sources, cancer registries participating in the pilot studies, are expected to transmit the integrated dataset to the iPAAC WP7 team for the analyses. To assess the actual feasibility of centralised data analyses, the legal conditions for sharing personal data were investigated in the questionnaire.

A maximum of 17 CRs (63%) declared to be allowed to transmit data to iPAAC WP7 analysis team and among the most accessible data sources the hospital discharges were available for sharing in 65% of cases, death certificates (71%), screening programmes data (93%) and pathology laboratories records (57%) (**Table 1.3**).

In some countries, data can be shared upon request of specific permissions. Whenever the procedures to ask permissions to share personal data proved to be complex, time consuming or not compatible with the workplan, a decentralized data analysis approach was considered as a valid alternative option to facilitate registries' participation in Tasks 2-4 pilots.

1.4.8 File formats of the accessible data sources

The survey asked the registries to also send the file format of each available source. A detailed description of all single items is necessary so as to understand the possibility of using the data to derive comparable indicators across Europe in Tasks 2-4 pilots.

Out of 27 CRs, 17 (63%) supplied the file formats of data sources. As for the rest, three registries (11%) did not supply the structure due to the complexity of some file formats, whereas 8 out of 27 (30%) could not supply file formats for other reasons. **Table 1.6** gives a description of the file formats provided by the 17 CRs. File format of Hospital discharges is the most frequently provided and the registries sent the structure for more than one data source (from a minimum of 2 to a maximum of 12 for Veneto CR in Italy).

Table 1.6: File formats of accessible sources provided by the cancer registries

#	Country	Cancer registry	Sources with file format
1	Belgium	Belgian Cancer Registry	Pathology lab, Hospital oncology registry
2	Croatia	Croatian National Cancer Registry	Hospital Discharges, Pathology lab, Mortality
3	Greece	Cancer Registry of Crete (CRC)	Hospital discharges, Hospital oncology registry, Pathology lab, Mortality, Hematology databases, Radiotherapy databases, Financial insurance
4	Italy	Friuli Venezia Giulia Cancer Registry	Hospital discharges, Health population registry
5	Italy	Integrated Cancer Registry of Catania-Messina-Enna	Hospital discharges, Private hospital facilities, Pathology lab, Outpatients reports, Medical records, Mortality, C Flux, M Flux, General practitioners, Ticket exemptions database
6	Italy	Napoli-3 SUD Cancer Registry	Hospital discharges, Mortality, Pathology lab, Drug prescriptions database, In-hospital drug prescriptions, Outpatient reports
7	Italy	Palermo Province Cancer Registry	Hospital discharges, Mortality
8	Italy	Reggio Emilia Cancer Registry	Hospital discharges, Mortality, Pathology lab
9	Italy	Cancer Registry of Siracusa Province	Hospital discharges, Private hospital facilities, Pathology lab, Outpatients reports, Medical records, Mortality, C Flux, M Flux, Screening database, General practitioners, Ticket exemptions database, Invalidity registry
10	Italy	Trapani-Agrigento Cancer Registry	Hospital discharges, Mortality, General practitioners
11	Italy	Veneto Cancer Registry	Hospital discharges, Pathology lab, Autopsy reports, Mortality, Health population registry, Outpatients records, Hospices, Home-based-long-term care, Emergency room, In-hospital drug prescriptions, Drug prescriptions database, Ticket exemptions database, Residential care
12	Malta	Malta National Cancer Registry	Hospital oncology registry, Pathology lab, Autopsy reports, Mortality, Hematology reports, Radiotherapy reports, Screening reports, Mandatory notification
13	Portugal	North Region cancer registry (RORENO)	Hospital discharges, Private hospital facilities, Hospital oncology registry, Pathology lab
14	Spain	Castellon Cancer Registry	Hospital discharges, Private hospital facilities, Pathology lab, Outpatients reports, Mortality, Other hospital records
15	Spain	Cuenca Cancer Registry	Hospital discharges, Private hospital facilities, Hospital oncology registry, Pathology lab, Outpatients reports, General practitioners, Mortality, Hematology database, Screening database,
16	Spain	Girona Cancer Registry	CR Operative manual including file format description
17	Spain	Granada Cancer Registry	Hospital discharges, Private hospital facilities, Pathology lab, Autopsy reports, Medical records, General practitioners, Mortality, Hematology database

1.4.9 Desired data sources

In the survey, the registries could also indicate which data sources they would like to access to for improving data collection (desired data sources). The reasons preventing access to such data sources were also requested.

Out of 27-responding CRs, 12 (44%) reported interest to access other data sources: most of them (9) reported more than one desired source and 3 reported at least one desired source.

Among the desired sources are: home-based long-term care for cancer patients, regional databases on cancer patients, radiotherapy or radiology departments data, professional diseases database; in-hospital and territorial drug prescriptions databases; historical archives of municipal registers, and the screening programs databases.

1.5 Discussion

1.5.1 Increasing and heterogeneous use of electronic data sources

Cancer registries play a crucial role in cancer surveillance being a fundamental source of objective and high quality cancer data. Population-based cancer registries collect systematically all cancer cases diagnosed among residents in their catchment area, thereby providing indispensable data to determine cancer burden and to design cancer control plans (*Steliarova-Foucher E et al, 2017*).

Cancer registries work under the authority of an institution which is usually public, responsible for the organization and financial support of the registry (data owner). Moreover, their function depends on the health system and on the socio-economical and legislative context of the country or region where they operate. The quality of a CR depends essentially on the completeness and the validity of data, such as the methodology used for data collection: active or passive. Comparative analyses of CR output data constantly reveal a gradient in data quality and complexity across Europe, with Nordic CRs regularly reporting the highest performance, and multiple setbacks recorded in Eastern European countries. (*Forsea AM, 2016*).

Cancer registries, thanks to the use of computerised databases loaned to epidemiology from the clinical-administrative context, have the availability of a large amount of clinical data in their routine activity (from hospital discharges, pathology and medical reports to outpatient and general practitioner records). Although the quality and availability of these additional data sources is heterogeneous across Europe (*EUROCOURSE, 2013*), their use is easier and more

practical compared to classical medical reports and they are gradually becoming a valid resource for quality of care assessment.

Another important aspect to underline is that computerised sources are available almost everywhere with common classifications and similar structures, so they can be used for international studies.

The survey also pointed out a significant heterogeneity of use or availability of sources, and of legal landscapes among different European registries. The compliance to the survey was rather good (82% out of 33 invited registries replied) and higher than in previously reported surveys, though addressed to a wider number of registries (Siesling S, 2015).

1.5.2 Data sources availability

The survey documents that the best known and most frequently used data sources are hospital discharges (96%), as well as pathology laboratories (85%) and death certificates (89%), whereas the use of other data sources is less frequent and more varying across countries. Similar results were obtained in the EUROCOURSE study where 93% of CRs reported to collect data from hospital records and discharge diagnoses and 78% from death certificates (Coebergh WW et al. 2011). These three flows represent the fundamental data sources analysed by the CRs to assess cancer burden and outcome.

Other databases are now available, though not all CRs can access them. Disability registries or fee exemption data are useful to supply information on cancer related pathologies, though with some limitations (*Mangone L et al. 2015*).

Drug prescriptions databases constitute a valid support to identify “lost” cancer cases (*Mercadante S, 2001*) and specific subtypes (*Schultz NM et al. 2018*), but were indicated by few CRs as accessible or used data sources.

The access of CRs to information sources with non-health content (employment, social benefit, education) is even more limited, maybe because they are not used for routine activity but only for particular studies (*Lillini R, 2019*).

Behind the use of a computerised database there is a great effort of standardization and data quality checks, so that not all CRs have enough trained personnel and economic resources to manage these processes.

Lastly, the survey highlighted that few registries are still using sources on paper (particularly those of Eastern Europe) and thus they were not eligible for the iPAAC pilot studies focusing on the advantages of electronic integration of the current registries dataset.

1.5.3 Privacy and data protection

One limit concerning the administrative sources is that most CRs only work using a subset of the information related to oncological patients and do not access complete information on all diseases, in compliance with the criterion of risk minimalization introduced by the privacy legislation (GDPR – General Data Protection Regulation). Adaptation to personal data protection legislation is becoming one of the major barrier for registries' functioning. In the large EURO COURSE project, 20-35% of cancer registries reported legal-related obstacles to cancer registration in Europe. In particular, different country-specific barriers were reported to the linkage with other health-related databases essential for cancer registration, e.g. vital status, hospital discharges or causes of death databases (*Coebergh JW et al, 2015*).

The fragmented and heterogeneous legal context at EU level hinders data sharing, cross-borders collaborative research, and important EU initiatives.

1.6 Concluding remarks

Task 7.1 survey was preparatory to pilot studies foreseen in Tasks 7.2-7.4. The main purpose was to provide preliminary information i) to assess the feasibility of registries participation to pilots and ii) to design protocols for data collection and analysis tailored to the specific registries' operating conditions.

The survey was also an occasion to assess to what extent electronic data sources are routinely available for an optimised and harmonised use by the cancer registries in Europe.

The results confirmed a quite heterogeneous landscape. Health administrative data sources available for linkage to the European cancer registries are not homogeneously accessible and used. This situation reflects a background of heterogeneity in the health care systems, data owners, legal frameworks and socio-economic conditions across countries.

There is however a significant portion of registries incorporating these data sources in routine activity and for research purposes. Part of these sources are sufficiently standardised in terms

of coding classification and data structure to be considered valid for deriving comparable indicators on cancer care (for instance hospital claims).

Data sources regarding socio-economical conditions were reported to be the least frequently available, although they are increasingly needed to explore the multi-dimension domain of cancer survivorship. Even the linkage to screening programmes data cannot be given for granted.

One of the major limitations documented by the survey is the lack of legal mandate to access information on diseases other than cancer. In some countries, remodulating the institutional mandate of cancer registries could help coniugate the compliance to data protection regulation and the need to expand the scope of registration (from epidemiology to quality of care assessment).

The potential to link population-based datased to the increasing number of computerised demographic and health information systems is huge and there is wide room for improvement. A more integrated and efficient use of the available information systems can be extremely beneficial for research on health services, for the patients and the society at large (*Pukkala E, 2011*).

1.7 References

- *Steliarova-Foucher E et al. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol. 2017 Jun; 18(6): 719-73*
- *Forsea AM. Cancer registries in Europe-going forward is the only option. Ecancermedalscience. 2016. Review.*
- *Siesling S et al.. Uses of cancer registries for public health and clinical research in Europe: Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010-2012. Eur J Cancer. 2015 Jun; 51(9): 1039-49*
- *Eurocourse. Analysis of funding and research of cancer registries in Europe In Vol Deliverable 1.3. Eindhoven The Netherlands: EUROCOURSE. 2013*
- *Coebergh WW et al. Cancer Registries in Europe - Serving cancer surveillance and research 1988 – 2012 – 2020. EUROCOURSE. 2011*
- *Mangone L et al. Evaluation of the cancer co-pay fee exemption data source (048 code) to estimate cancer incidence. Epidemiol Prev. 2015 Jul-Aug; 39(4): 226-33).*
- *Mercadante S, Fulfaro F, Casuccio A. Pattern of drug use by advanced cancer patients followed at home. J Palliat Care. 2001 Spring; 17(1): 37-40.*
- *Schultz NM et al. Treatment Duration, Healthcare Resource Utilization, and Costs Among Chemotherapy-Naïve Patients with Metastatic Castration-Resistant Prostate Cancer Treated with Enzalutamide or Abiraterone Acetate: A Retrospective Claims Analysis. Adv Ther. 2018 Oct; 35(10): 1639-1655.*
- *Lillini R, Vercelli M. The local Socio-Economic Health Deprivation Index: methods and results. J Prev Med Hyg. 2019 Feb 28; 59(4), Suppl 2.*
- *Coebergh JW et al. EUROCOURSE recipe for cancer surveillance by visible population-based cancer registries in Europe: From roots to fruits.Eur J Cancer. 2015 Jun; 51(9): 1050-63)*
- *Pukkala E. Biobanks and registers in epidemiologic research on cancer. Methods Mol Biol. 2011; 675: 127-64*