Building the European database of cancer registry data

Eva Steliarova-Foucher
International Agency for Research on Cancer
Lyon, France

History

- Population-based cancer registries in Europe since 1950s
- Organised in a Network since 1989 (EU funds)
- Currently about 200 members
- Recommendations, courses, consultancies
- EUROCIM
  - database
European cancer database - rationale

- Need for an authoritative source of European cancer data
  - Standardised
  - Comparable
  - Complete
  - Informative
- To be used for:
  - Benchmarking and monitoring of cancer burden
  - Planning public health resources
  - Evaluation of strategic decision
  - Research
  - Improvement of the cancer management outcome

European cancer database – the added value

- Conformity to international standards
- Standardised and comparable output
- Easy access
- Sharing resources
- Combining expertise
- Exponentiating knowledge
European cancer database – the barriers

- Multi-lingual Europe
- Differences in organisation of public health system
- Variations in registration procedures
- Variable infrastructures used by the registries
- Use of local ‘standards’
- Data confidentiality requirements
- Registries reluctance to provide the data

Europe against Cancer: Optimisation of the Use of Registries for Scientific Excellence in research

www.eurocourse.com

1 April 2009 – 31 March 2012
ERA-NET Seventh Framework Programme of the DG Research of the EC

A major project of the ENCR
EUROCOURSE – database-related deliverables

- D 4.1 Public access dataset for routine situation analyses
- D 4.2 Research dataset with access restricted to the registries and researchers
- D 4.5 Development of a European Cancer Observatory: the formation of a comprehensive programme of work on cancer intelligence

European cancer database – construction

1. Database design
2. Call for data
3. Data submission
4. Central data processing
5. Data validation
6. Consolidation of the common database
7. Dissemination
1. Database design

- Brand new database
- Stored on SQL Server
- Complex structure with multiple relationships:
  - Registries
  - Individual collaborators
  - Studies
  - Data calls
  - Files
  - Surveys
  - Processes
1. Database design

- Automation
  - Data submission/reception
  - Process monitoring
  - Archiving
- Efficiency
  - Less staff time
  - Faster processing
- Reduction of error
  - Transcription
  - Process follow-up
- Increased security
  - Password protected access
  - Less human intervention
- Versatility and sustainability
  - Inclusive of on-going and future activities

2. Call for data

- Working group to prepare a call
- 21st May 2010: ENCR call for data
  - Invitation by a letter to the cancer registries
  - Invitation by an e-mail
  - Invitation through the ENCR web-site
- On-line electronic submission
  - First use of the portal
  - Successful transfer of all files so far through portal
  - No other modes of submission used
- Deadline: 30th June 2010
- Alignment with the EUROCARE call
2. Call for data

- Cancer cases file (45 variables)
  - Person: IDs, sex, age, dates
  - Tumour: ICD-O-3, stage, treatment
  - Follow-up information
- Mortality data (sub-national registries)
- Life-table files
- Population data
- Coding schemes

3. Data submission - Registries portal

- Secure log-in using HTTPS / SSL
- Equifax high-grade 128 bit encryption
- Registry-specific log-in based on newly devised coding standards
4. Central data processing
Sorting & viewing of ENCR data at IARC

- Descriptive emails are sent to CIN staff when files are uploaded
  - This occurs at the same time as the registry receives their confirmation
- Files are conveniently named, stored and categorised for easy access
- History and details of files uploaded can be viewed by CIN staff through the new collaborators database (built in Access - VB).
- Reports of the status

4. Central data processing
Call for data
Registry
Registries portal
Standard folders on a dedicated server
Preliminary checks, formatting
Collate
Errors
Unlikely
Queries
Queries
DEPeds
Stata routines
Peer review?
Standard dataset
Analyses
Common database
Dissemination
### Status of submissions on 10 June 2011
cancer files

<table>
<thead>
<tr>
<th>Registration Type</th>
<th>Received</th>
<th>Expected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registries submitting</td>
<td>99</td>
<td>57</td>
<td>156</td>
</tr>
<tr>
<td>Including 'superregistries'</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Including regional registries</td>
<td>46</td>
<td>15</td>
<td>61</td>
</tr>
<tr>
<td>Including specialised</td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Including childhood</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>Total registries</td>
<td>95</td>
<td>56</td>
<td>151</td>
</tr>
<tr>
<td>Total general registries</td>
<td>87</td>
<td>51</td>
<td>138</td>
</tr>
</tbody>
</table>

### Status of submissions on 10 June 2011
Cancer records

<table>
<thead>
<tr>
<th>Category</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records submitted</td>
<td>18,413,386</td>
<td>100</td>
</tr>
<tr>
<td>Records processed</td>
<td>15,962,054</td>
<td>86.7</td>
</tr>
<tr>
<td>Records excluded</td>
<td>53,940</td>
<td>0.3</td>
</tr>
<tr>
<td>Records included in summary statistics</td>
<td>15,908,114</td>
<td>86.4</td>
</tr>
</tbody>
</table>
### Status on 10 June

<table>
<thead>
<tr>
<th>Country</th>
<th>Submitted National</th>
<th>Submitted Regional</th>
<th>Processed %</th>
<th>Expected National</th>
<th>Expected Regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>2</td>
<td>50</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belarus</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Croatia</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estonia</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>10</td>
<td>50</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gibraltar</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iceland</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>28</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>1</td>
<td>1</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Portugal</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>2</td>
<td>50</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serbia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Slovakia</td>
<td>1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>1</td>
<td></td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>6</td>
<td>67</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switzerland</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkey</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>1</td>
<td>11</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Unspecified sites
C26, C39, C48, C76, C80
Unspecified histologies
M-8000 – M8005

Percentage of records with missing information
7. Dissemination – proposal of a WG

- Aggregated data
  - 181 cancer entities by sex, 5-year age groups and the 5-year period as defined as CI5
- Individual records with selected variables
  - Geographical area
  - Tumour sequence number
  - Sex at birth
  - Month and year of incidence
  - Age in years
  - ICDO3 Topography
  - ICDO3 Morphology
  - ICDO3 Behaviour
  - Basis of diagnosis
  - Grade
  - Laterality
  - Vital status at last contact
  - Month and year of end of follow-up
  - Duration of survival in days
  - Cause of death
  - ICD edition coding the cause of death
- Individual records with all variables
7. Dissemination – online search and download

- Search Common Database
  - Presentation of incidence, mortality and survival data in various tables or graphs.
  - The tables can be exported to a text file or saved as a PDF file.

- Search Cancer Records
  - Location of case
    - Country
  - Registry
    - Registry number
  - About the patient
    - Ethnic group
    - Sex
    - Age
    - Cause of death

- About the tumour
  - Year of incidence
  - TCGA Topography
  - TCGA Morphology
  - TCGA Staining
  - TCGA Grade

- Search Database

7. Dissemination – online search and download

- Search Database

15/16
### 7. Dissemination – data sharing

Registries survey based on 91 respondents

<table>
<thead>
<tr>
<th>Controlled mechanism</th>
<th>Grouped data</th>
<th>Individual records</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENCR</td>
<td>70.3%</td>
<td>71.4%</td>
</tr>
<tr>
<td>CR</td>
<td>63.7%</td>
<td>50.5%</td>
</tr>
<tr>
<td>Anyone</td>
<td>41.8%</td>
<td>34.1%</td>
</tr>
<tr>
<td>ENCR password</td>
<td>48.4%</td>
<td>36.3%</td>
</tr>
<tr>
<td>CR password</td>
<td>44.0%</td>
<td>28.6%</td>
</tr>
<tr>
<td>Anyone password</td>
<td>27.5%</td>
<td>18.7%</td>
</tr>
<tr>
<td>Anyone no limits</td>
<td>27.5%</td>
<td>6.6%</td>
</tr>
<tr>
<td>Other</td>
<td>7.7%</td>
<td>16.5%</td>
</tr>
</tbody>
</table>

### Conclusions

- Construction of the European database in progress
- Access to the data under discussion
- Communication with the registries planned
- Output is being developed
- Good compliance by the registries
- Need to work on data comparability
- Review of the Call for data