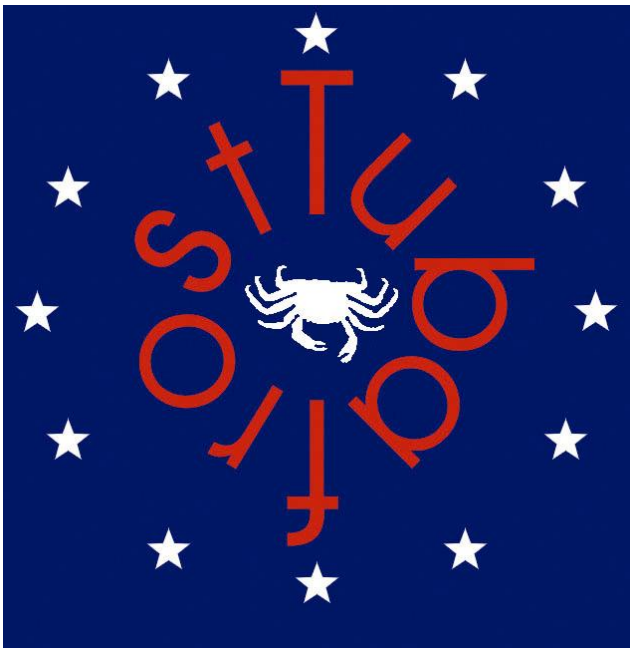
Quality of Life and Management of Living Resources

European Human Frozen Tumour Tissue Bank

TUBAFROST

QLRI-CT-2002-01551

Deliverable D 4.1



This report concerns the assessment of the local database systems used to store tissue information at each of the participating institutes and how the relevant tissue information will be entered into the central database system. It is part of Work Package 4: “Hard and software for co-ordinated tissue storage”.

1. Datasets and user requirements

A questionnaire was created and distributed to the participating institutes to access what datasets (including age, sex, place of residence, digital images, storage codes…) are required for local registration of the collected frozen tissues. These results would form the basis of the user requirements for the local database system. The list was split into two:

1. Minimum (required) dataset
2. Facultative dataset

These two lists will also form the basis of the central database system. The results of these questionnaires are reported in appendix 1.0

To summarize, the minimal[1] (required) dataset contained:

1. patient consent type
2. diagnosis (If applicable Grade, Stage Node status),
3. age,
4. sex,
5. tissue type
6. availability of normal and premalignant
7. tumor type and how the diagnosis was reached (WHO typing)

whereas the facultative[2] dataset contained:

1. histology Images plus info on staining,
2. demographic: country, country of availability of tissue
3. anamnestic data, past history (including SNOMED codes used for illnesses in patient's history e.g. diabetes)

[1] List(s) as agreed by the TuBaFrost consortium during the Aviano Meeting in May 2003
[2] List(s) as agreed by the TuBaFrost consortium during the Aviano Meeting in May 2003

4. concomitant disease (only cancer related)
5. secondary tumors
6. toxicity, acute, late
7. biochemical, (cyto)genetics and immunological data
8. salvage treatment
9. survival status
10. treatment
11. after statistical analysis: All data of survival: outcome, activity of medical treatment/tumor
12. time passed before freezing
13. clinical trial participation
    a. trial numbers
        i. trial sample number
        ii. status of trial (active/closed)
        iii. EORTC affiliated (yes/no)
    b. patient codes/identification

2. Assessment of the local database systems

The online questionnaire collected information about the type and content of database that each participating institute uses. The questionnaire and the results are reported in appendix 2.0

By combining the recommended datasets lists and the assessment of the local database systems we can see what items need to be added to the already existing local database systems in order to fulfill requirements for the minimal and facultative data set.

- **For the minimal (required) dataset:** Almost all (>90%) of participants collected data on tissue type and the diagnosis; 70% collected data on the age (birthyear/birthdate), patient's sex, site and status of primary tumor and availability of tissue samples; only 20% collected data on type of patient consent. Since the minimal (required) dataset required that all participating institutes need to collect information on diagnosis, age, sex, tissue type, availability of normal and pre-malignant: 30% of participants need to add sex of patient, age, and site and status of primary tumor; 80% of participants need to add type of patient consent.

- **For the facultative data:** only 40% collected data related to biochemical data and cyto(genetics) and survival status; only 30% collected data on current and past treatment, secondary tumors and immunological data; only 20% stored histological image data and toxicity data; only 10% stored data on country of residence.

3. Proposed process for Central Database System data-entry

By the end of the second TuBaFrost consortium meeting in Aviano (May 2003) there existed two options for getting tissue information, located on the local database systems, into the central database (which is to be developed):

1. Participants export batch data (in the form of text files) to be sent or uploaded to the Central database located at the EORTC Data Center
2. Participants input data (at least minimal required dataset) into the Central database directly online using the TuBaFrost Central Database Interface.

After discussion with the IT unit at EORTC they recommended the 2nd option (TuBaFrost Central Database Interface) due to the following reasons:

1. Exporting of batch data would require some work to filter and standardize (fields, codes, length and format) the data being sent and received before importing into the central database.
2. Importing of batch data, in the past, has generated many problems before due to incompatible values (wrong decimal places, dates etc.) and data types.
3. Unlike web-based and EORTC paper form data entry, immediate data validation is not guaranteed. User will not get the immediate alert that required fields are incomplete or wrong data format. This would mean that the administrator of the central database would have to go back to the original institute data manager to get the right data, which is time-consuming for both parties.
4. It would be easier to allocate the 2nd tissue (TuBaFrost) code immediately online while the participant is entering the data. This way the participants would not have to wait until a later date in order to acquire this 2nd tissue code (i.e. do not have to wait until batch data has be manually inputted at EORTC).

Using a standard computer with the minimum specification stated in section 4 (below), the user will start the web browser application and will go to the web address for the TuBaFrost Central Database System (contained on a server at the EORTC Data Center in Brussels).

The user will then be prompted for his/her allocated username and password. Once these have been verified by the system, the user will then be asked to complete the Online Tissue Information Form. The fields, to be completed, will correspond to the

minimum dataset (stated in appendix 1.0) along with fields to add additional facultative dataset (also stated in appendix 1.0)

The user will be able to open the relevant record from his/her local database system and copy and paste the relevant information from the local database system into the empty fields on the Online Tissue Information Form.

Once the user has successively completed the relevant fields he/she will click on submit form and the form will be sent to the TuBaFrost Central Database System where it will automatically checked and validated. If there are any problems (e.g. wrong data formats, incomplete required fields etc.) a message will be displayed and the user will be prompted to make the necessary corrections to the form. If there are no problems encountered the system will accept the form and a message will be displayed informing the user that the form has been successively created. He/she will also be given the allocated TuBaFrost Material code (second material code) for the sample entered which they can copy and paste into their local database system.

The allocated TuBaFrost Material code (second material code) will be composed of the following structure:

TF_Institution code_sequential (local) code

An Institution code will be allocated to each institute participating within the project and these codes will be stored at the EORTC.

The local inventory sequential code will be the code used by the participating institute. Since only 50% of the participating institutes use a sequential code for the material, these other 50% must make sure that they create an additional sequential coding system for their local databases. This sequential code must not relate to the pathology number or other identifiers.

After the tissue information form has been successively created online, an automated e-mail will be generated which will contain all the information entered into the form, and will be sent to the e-mail account of the user.

The user can then continue to enter a new Online Tissue Information form for his/her next frozen tissue sample.

4. Minimum local computer system demands

Due to the efficient solution that has been chosen by the TuBaFrost consortium to communicate with the central database, a nowadays standard computer (PC or Apple Mac) with a few adaptations will suffice. Each participate must have a dedicated computer, to access the local tissue bank database and connect to the central database system, with the following minimal specifications:

PC:

Pentium 4

1.6 GHz or higher

256 MB + Ram

80GB Hard Drive

56K modem internet connection or faster

Firewall security and virus protection installed

Browser: Internet explorer version 5.0 or higher (Netscape can be utilized if they conform to IE5.0 standards)

Windows 9x, NT4.0, 2000 and XP

17" monitor, true color, 1024 x 768 resolution

Apple Mac:

PowerPC G4 or G5 processor.

1.6 GHz or higher

256 MB + Ram

80GB Hard Drive

56K modem Internet connection or faster

Firewall security and virus protection installed

Browser: Safari 1.0, Mozilla or Internet Explorer 5.

MacOS 8.5 or MacOS X

17" monitor, true color, 1024 x 768 resolution

Of course the computer must also be on the institution's local network in order to access the local tissue bank database and the network's internet connection.

5. Security and Protection of the Local and Central Database Systems

The participant's computer network must have adequate security installed to prevent against viruses attacks and hackers. It is therefore instrumental to install an anti-virus protection program (update virus definitions daily) and a Firewall system.
The EORTC Data Center, where the central database system is to be developed and housed, has a Firewall system installed across its network and the virus definition files are regularly kept update by the EORTC IT unit who maintain the network. Along with usernames and passwords (allocated to each participant) the central database system also has a data encryption system installed so that data that is entered and submitted in the on-line form is encrypted whilst being transmitted to the central database system**.**

APPENDIX 1.0 - DATASETS

Questionnaire WP 4

Assessment of Information Technology requirements of the database system.

During a meeting held for WP4 we have encountered some items that needed feedback from the whole consortium. First we need to assess which items are needed for a search in the centralised database and therewith what items are needed in the dataset accompanying the tissues. Minimal means data that is required to describe a stored tissue specimen, whereas facultative is meant for data that will be very desirable for the better experiments, but is not necessary to describe all stored tissues.

In addition, we need to know if it is possible to set up a database link with local hospital information system to update clinical data of patient data. Although, we expect that most of the systems will be closed for these options due to the protection of privacy of the patients we would like to investigate the possibilities at the sites where it is allowed. These sites could serve as good examples for the closed sites.

Conclusions:

If all received suggestions are followed the following sets are obtained:

All data needed for search:

Minimal

- Diagnosis (If applicable Grade, Stage Node status),
- age,
- sex,
- Tissue type
- Availability of normal and premalignant
- Tumour type and how the diagnosis was reached (WHO typing)

Facultative

- Histology Images plus info on staining,
- Demographic: country, country of availability of tissue
- Anamnestic data, past history (Snomed)
- Concomitant disease (Only cancer related)
- Secondary tumors
- Toxicity, acute, late
- Biochemical, (cyto)genetics and immunological data
- Salvage treatment
- Survival status
- Treatment
- After analysis: All data of survival: outcome, activity of medical treatment/tumour
- Time passed before freezing

DATA SET:

Minimal data set:

- Patient consent type
- Diagnosis (If applicable Grade, Stage Node status),
- Age
- Sex,
- Tissue type
- Tumour type and how the diagnosis was reached (WHO typing)
- Availability of the tissue / specimen normal and premalignant

Facultative:

- Histology Images plus info on staining,
- Anamnestic data, past history (Snomed)
- Country of origin
- Concomitant disease
- Secondary tumors
- Toxicity, acute, late

- Biochemical, (cyto)genetics and immunological data
- Demographic: country, country of availability of tissue
- Salvage treatment
- Survival status
- Treatment

Link with local systems: A large variety of local systems have been mentioned. In all cases it appeared not possible to directly link.

APPENDIX 2.0 - ASSESSMENT OF LOCAL DATABASE SYSTEMS

1. Questionnaire Design

Name:

Institution

E-mail (please make sure this is correct):

1. Is your database system in-house developed or purchased?

- In-house developed
- Purchased

If in-house developed, please provide some details of the system (e.g. structure, development language (e.g. Visual Basic, C++, Delphi,…), platform (e.g. Windows, Mac) and operating system (e.g. Windows XP, Windows 2000....)

If purchased, what make/type of database do you have?

- Filemaker Pro
- Oracle
- KEA
- Sybase
- Microsoft Access
- Microsoft SQL server
- Not Applicable
- Other (please specify)

Comments:

2. What data items are collected regarding the tissue collected?

- Staining
- Type of material
- Data of sampling
- Fixation
- Histological Images
- Current location of sample

- Availability of tissue sample
- Other (specify)

3. How is tissue identified in the database?

- Randomized code
- Original biopsy number
- Sequential code
- Multiple codes
- Other (specify)

4. What basic clinical data is collected about patient?

- Date of birth or birth year
- Country of residence
- Diagnosis of patient (grade, stage, TNM, SNOMED, etc.)
- Sex
- Type of patient consent
- Survival status
- Progression/recurrence
- Current and past treatment or medical history
- Site, size and status (organ and tissue type) of primary tumor
- Secondary tumors
- Lymph node involvement
- Amount of material collected for patient
- Date of diagnosis
- Other (specify)

4a. If patient consent is collected for patient, please specify the type of patient consent collected.

- For clinical trials
- For defined tissue research
- For undefined (future) tissue research
- Other (specify)

5. What kind of laboratory data is collected?

- Biochemical
- (Cyto)genetics
- Toxicity
- Immunological
- Other (specify)

6. Please indicate other kinds of data which is collected in the database.

7. For the most important tumor types, are templates created for each different type of tissue information?

- Yes
- No

8. Is this tissue/pathology database linked to other databases in your hospital/institution?

- Yes
- No

9. Is this tissue/pathology database linked to external databases outside your hospital/institution?

- Yes
- No

10. Does your database system support importing and exporting of data?

- Yes
- No

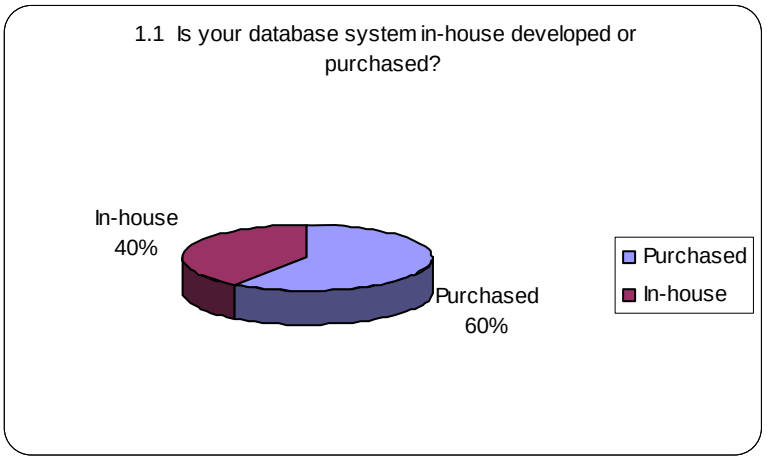
10a. If yes, what file format type(s) can it be exported as?

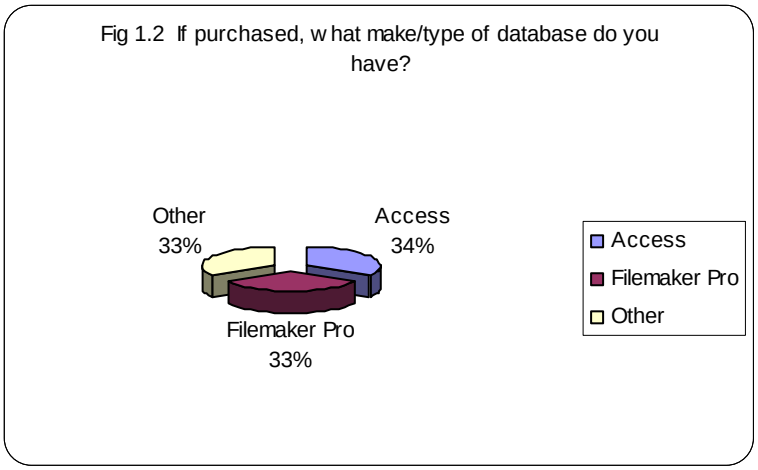
11. Comments

2. Results

Question 1: Is your database system in-house developed or purchased?

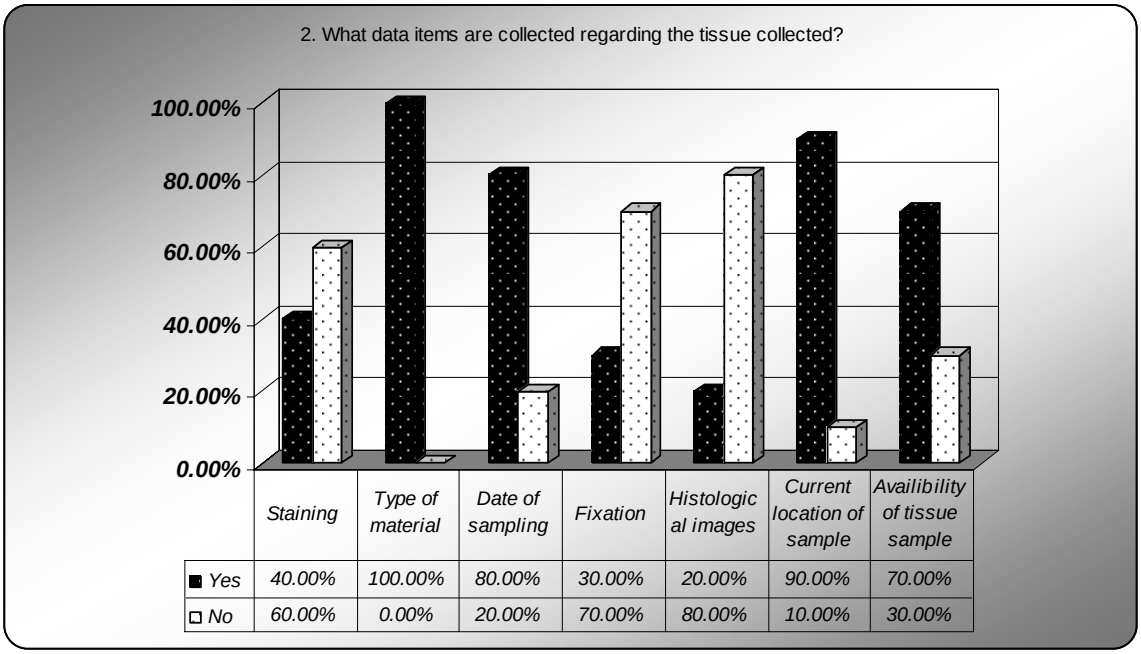
- 60% of the participants purchased their database system.
  - MS Access (33%) and Filemaker Pro (33%) are most often used
  - Other systems: PHP, MySQL, MS Excel
- 40% had their system in-house developed.
  - IGR uses Oracle and VB, Windows, Windows 2000
  - CNIO uses Java (language) and Linux (Red Hat) as operating system
  - The EORTC database is developed in Delphi for Windows 2000 (and above) and using Ms SQL server (Version 2000)
  - KU Leuven uses a Windows 2000 or XP platform. Peripheral hardware: Dell; central hardware: SUN, with a UNIX operating system. The database is SYBASE, the program is written in PROLOG.





Question 2: What data items are collected regarding the tissue collected[3]?

- All the participating hospitals/institutions collect data concerning the type of material.
- 90% collects data concerning the current location of the sample.
- Over 70 % collects data with regard to the date of the sampling and the availability of tissue sample.
- Data concerning staining is collected by 40% of the participants
- A minority collects histological images[4] and data concerning fixation.
- Other: Pathology ID number (Erasmus); biopsy vs. autopsy (CNIO)

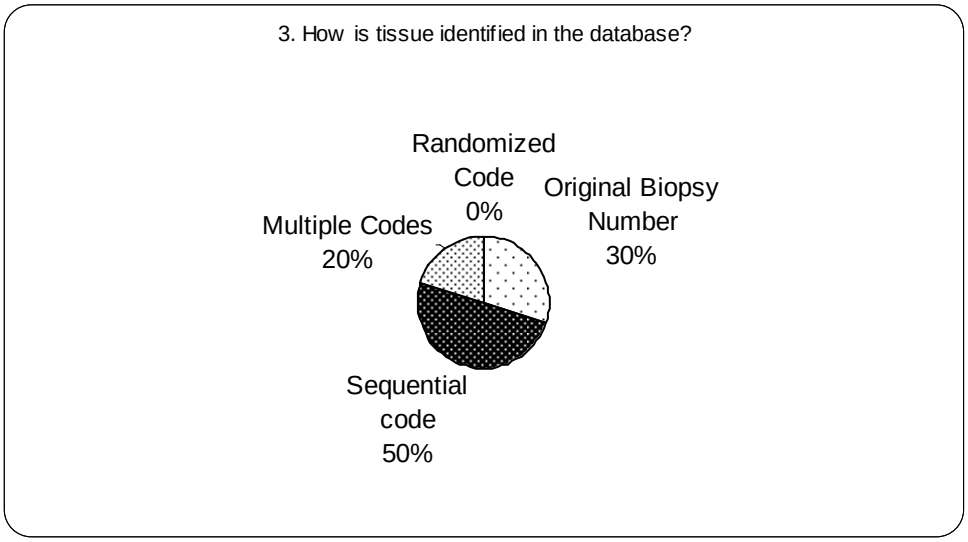


	Staining	Type of material	Date of sampling	Fixation	Histological images	Current location of sample	Availibility of tissue sample
Yes	40.00%	100.00%	80.00%	30.00%	20.00%	90.00%	70.00%
No	60.00%	0.00%	20.00%	70.00%	80.00%	10.00%	30.00%

[3] IGR: the database system is linked with the HIS clinical and follow-up data.
[4] CNIO does not yet collect histological images, but plans to do so in the future.

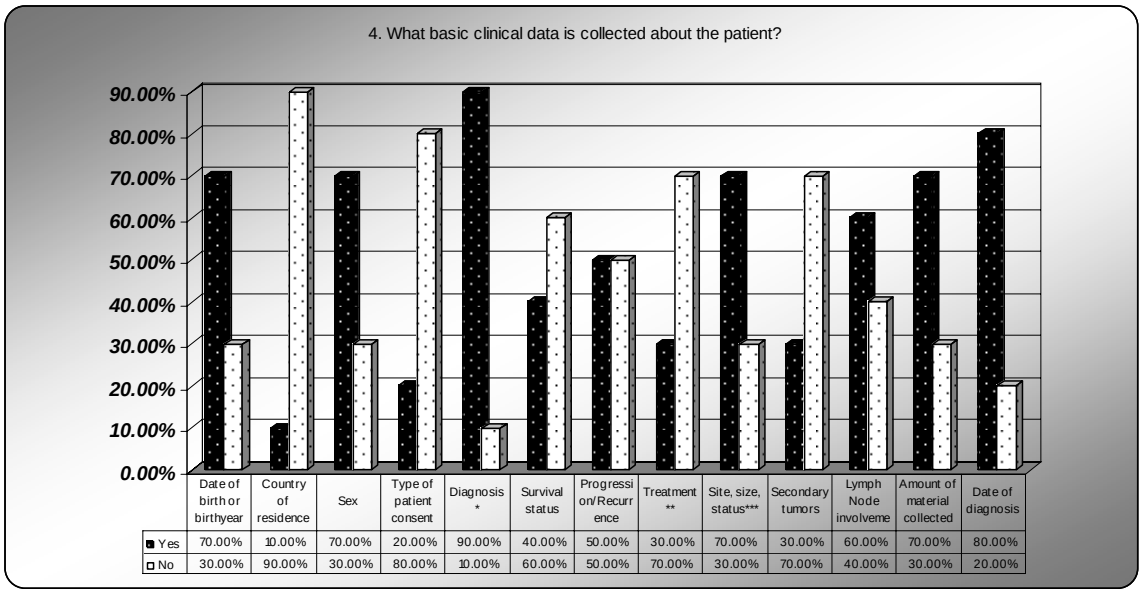
Question 3: How is tissue identified in the database?

- 50 % of the participants identify the tissue by using a sequential code.
- 30% of the participants use the original biopsy number.
- 20% use multiple codes.
- None of the participants use a randomized code.



Question 4: What basic clinical data is collected about the patient?

- Nearly every participating hospital/institution collects data concerning[5]:
  - Diagnosis of the patient (grade, stage, TNM, SNOMED, etc.)
  - The date of the diagnosis
- 70% collects data concerning:
  - Date of birth/birth year
  - Sex
  - Site, size and status of primary tumour
  - Amount of material collected
- 60 % collects data concerning lymph node involvement.
- About 50 % of the participants collects data concerning
  - Progression and recurrence
- Only a minority collects data about:
  - Secondary tumours
  - Survival status
  - Current and past treatment and medical history
- Only 2 participants collect patient consent.
- Only 1 participant collects data about the country of residence.
- Other: known infections + open field for other data (CNIO)[6]



	Date of birth or birthyear	Country of residence	Sex	Type of patient consent	Diagnosis *	Survival status	Progression/Recurrence	Treatment **	Site, size, status***	Secondary tumors	Lymph Node involveme	Amount of material collected	Date of diagnosis
Yes	70.00%	10.00%	70.00%	20.00%	90.00%	40.00%	50.00%	30.00%	70.00%	30.00%	60.00%	70.00%	80.00%
No	30.00%	90.00%	30.00%	80.00%	10.00%	60.00%	50.00%	70.00%	30.00%	70.00%	40.00%	30.00%	20.00%

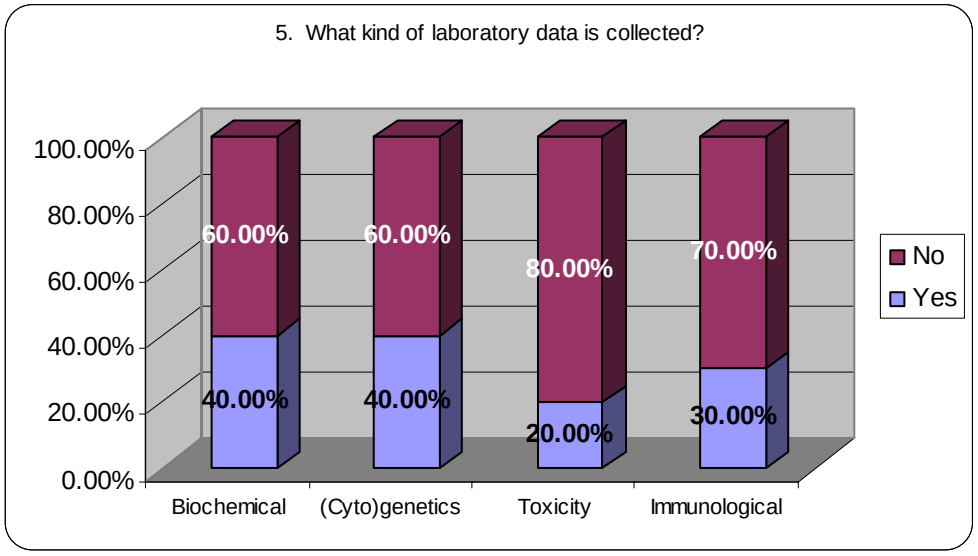
Question 4a. If patient consent is collected for patient, please specify the type of patient consent collected.

[5] The Dutch Cancer Institute collects clinical data separately.

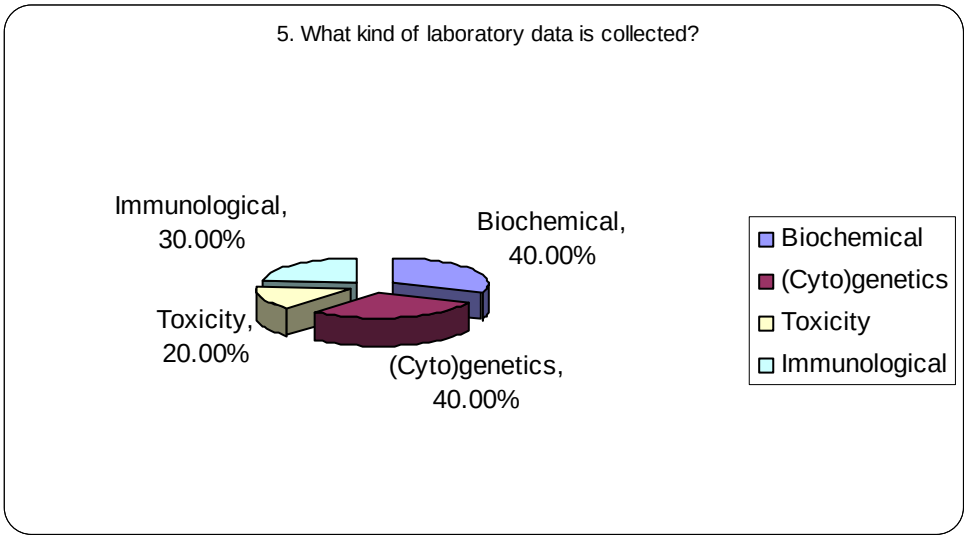
- EORTC: for clinical trials, for defined tissue research and for undefined (future) tissue research.
- IGR: for undefined (future) tissue research

Question 5: What kind of laboratory data is collected[7]?

- 40% collects biochemical and (cyto)genetics data
- 30% collects immunological data
- 20% collects data concerning toxicity
- Other:
  - Virological data (1 participant)
  - IGR: all laboratory data available
  - CNIO only collects pathological diagnosis



[6] CNIO only collects data at the time of diagnosis; no follow-up.
[7] Since the EORTC tissue database is linked to the clinical database at EORTC, the list of available clinical data is dependent on the individual protocols.

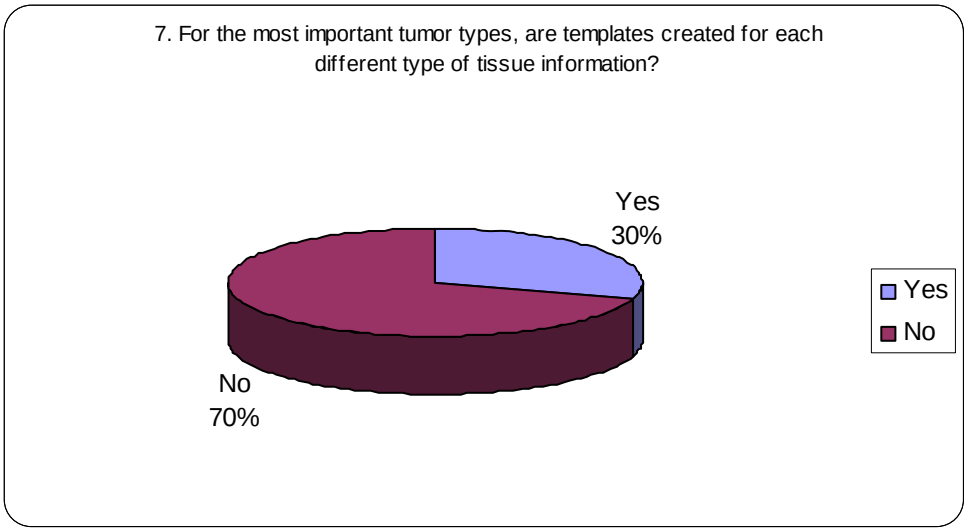


Question 6: Please indicate other kinds of data which is collected in the database

- Current registration of researchers to whom tissue samples (amount, type and reasons) have been given (CRO).
- Clinical trials status (IGR)
- Experimental results (NTRAC)
- EORTC: since the database is linked to the clinical database at EORTC, the list of available clinical data is dependent on the individual protocols
- AKW: Some people collecting samples of patients with inflammatory disease look for more clinical information.
- Screen shots CNIO

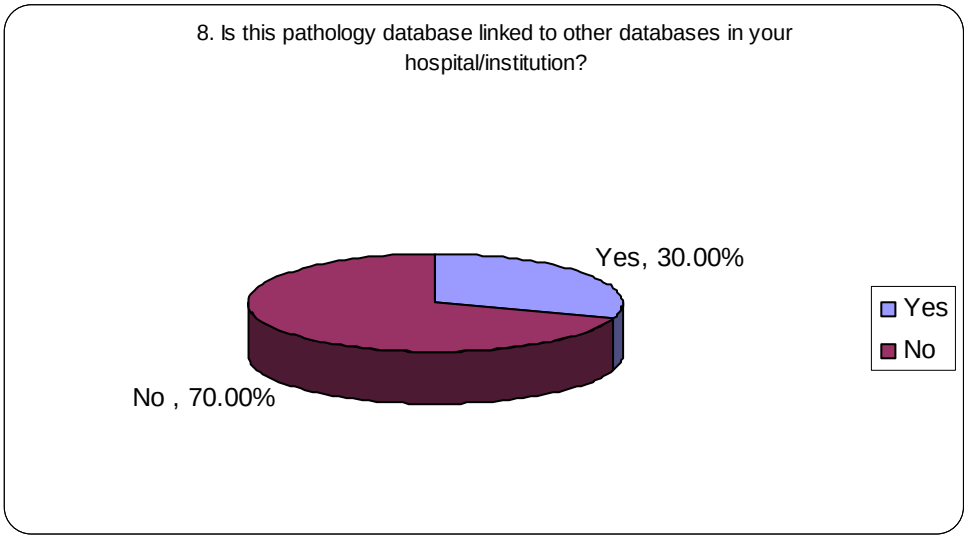
Question 7: For the most important tumour types, are templates created for each different type of tissue information?

- IGR, NTRAC and EORTC create templates for the most important tumour types



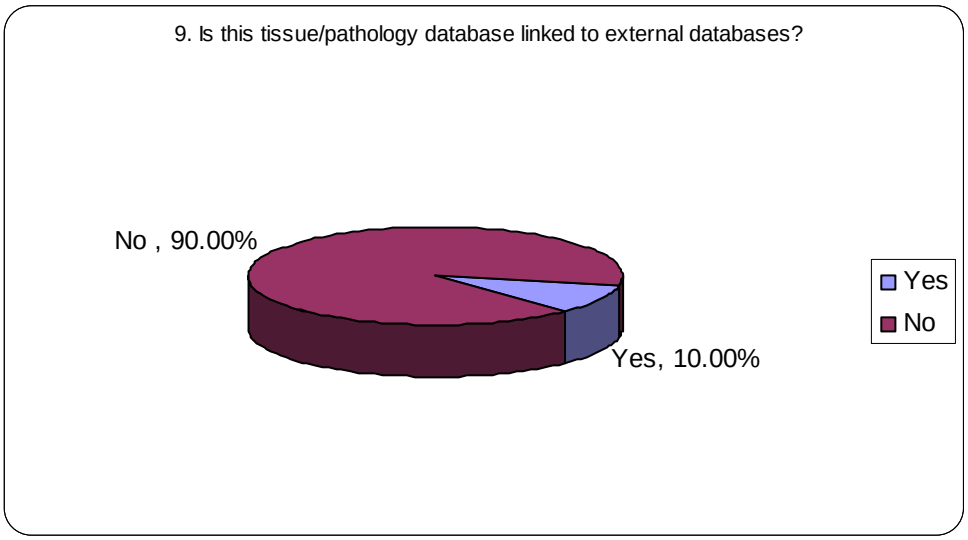
Question 8: Is this tissue/pathology database linked to other databases in your hospital/institution?

- The databases of IGR, KU Leuven and EORTC are linked to other databases within the hospital or institution.



Question 9. Is this tissue/pathology database linked to external databases outside your hospital/institution?

- Only CNIO links its tissue/pathology database to external databases.



Question 10: Does your database system support importing and exporting of data?

- All the database systems support importing and exporting of data.
- Format file types[8]:
  - Text files (Erasmus, NTRAC, CRO, IGR, CNIO)
  - Excel (IVO, CRO, IGR, NTRAC, AKW)
  - Access (EORTC, IGR, CRO)
  - DBF (NTRAC, IVO)
  - Filemaker (AKW[9])
  - SAS (EORTC)
  - SYLK, DIF, WKS, BASIC, Merge, HTML, ODBC (NTRAC)
  - Lotus, Paradox (IVO)

Question 11: Comments

- EORTC:
  - At the EORTC, paraffin blocks and glass slides are collected and stored. However, in addition, the EORTC Tumour Bank also collects information (e.g. location, codes, number, and material type) about frozen material, blood samples and other different material (NB only when consent has been given by the patient). An on-line search tool has been designed that allows users to search through the Virtual Tumour Bank (VTB) archive.

[8] IGR and CRO: standard formats
[9] AKW: No link between these systems and the hospital owned database system. A new hospital-developed database system is expected to work in one or two years.

- CNIO
    - The original biopsy number is only visible in the original hospital, not in the central database (CNIO). The sequential code is our working code of 8 numbers: Hospital AA + Year BB + Sequential number for each hospital CCCC.
    - Connection between associated hospitals and central office of the network (CNIO) by secure private telephone line (ISDN)
    - Special cooperative studies and clinical required special information, but in parallel, not through TB network.