

## ***Scientific report***

*on the implementation of the project in January - December 2019*

Project Title: **Next-Generation Non-Hodgkin Lymphoma Registry with an Ontology-Based Approach and Data Fusion**

Phase 2: **Implementation of a population NHL registry in the NV region of Romania using the complete NHL registration solution.**

**Overall objective of the project.** Development of a Non-Hodgkin Lymphoma (NHL) recording method based on combined ontology and data fusion techniques for integrating patient data with administrative and demographic data from multiple sources to create an associated dataset which could be queried in new ways.

**Specific Objective 1:** Develop a complete NHL recording solution capable of providing multisource data integration, semantic consistency, and data integrity.

To achieve this, we conducted the following activities included in the Work Plan Phase 2:

Activity 2.1. Development of the beta version of the software solution (database and user interface)

Activity 2.4. Administrative activities for register implementation

Activity 2.5. Data call for the registration of patients with Non-Hodgkin's Lymphoma in the North-West region of Romania - phase I

Activity 2.7. Organize a series of training seminars focused on cancer registration, especially in the case of patients with Non-Hodgkin's Lymphoma.

The deliverables provided for this stage:

- New generation software solution, beta, for recording non-Hodgkin's lymphomas;
- Report on the work of the training session.

We have implemented a semantic web platform that allows users to formulate user-defined incremental queries that will be duplicated, in the second phase of development, with a graphical interface. The semantic web can be viewed as a next-generation web, in which information receives a well-defined meaning, allowing better computers and people to work in cooperation. Ontologies are the standard knowledge representation mechanism for the Semantic Web. Technologies such as Web Ontology Language (OWL) for ontology construction allow for a formal representation of the domain. Clinical data is currently stored predominantly in relational databases. For transforming relational data into semantic formats, we followed 3 steps: (1) defining mapping rules between database fields and ontology; (2) OWL data generation; and (3) importing OWL data into the semantic database.

The ontology contains a total of 247 classes, 13 properties and 18 object properties, with 974 logical axioms. Ontology covers the following classes:

· *The patient* is a person with any type of non-Hodgkin's lymphoma. Properties: sex, date of birth, diagnosis, therapies and disease evolution.

· *The patient's condition* represents the patient's health status at a given time. Properties: WHO performance status at the time of assessment.

· *Diagnosis* is the diagnosis of the patient at some point. The classification system is the superclass of all classes and is based on the systems of classification of malignant lymphopathies: WHO 2016 classification, WHO-EORTC classification of cutaneous lymphomas, ICD-O-3, ICD-10, grading, Ann Arbor staging system, molecular markers and genetics.

· *Therapy* represents the therapies administered to a patient at a given time. Different types of therapy have been modelled in the ontology as subclasses of the Therapy class. Properties: medication, start date and end date.

· *The course of the disease* represents the course of the patient's disease for a diagnosis at a given time. Different types, such as complete remission, progression, recurrence and others have been modelled in ontology as subclasses of the disease class. Properties: patient conditions and date.

The classes and subclasses were modelled on the user interface through a chained combo-box system and conditioned panels.

The platform has been tested internally, and from December 2019 will be available at:

<https://www.nextgenrel.ro>

Within the administrative activities of implementing the registry, we have taken steps to register the **nextgenrel.ro** Internet domain, as well as SSL certification.

In December 2019 the first series of training sessions with specialists who will be involved in the use of the Registry will be organized. The Agenda proposed for this training session is:

- 2016 revision of the WHO classification of lymphoid neoplasms - Coding rules for topography, morphology, behaviour and grade
- Coding stage - specific principles and aspects
- Using NextGenReL ontology-based software solution
- Open discussions

I intend that these training sessions will continue after the project is completed, thus ensuring the sustainability of the project.

Also, after the training session, the first data call will be launched for the registration of patients with Non-Hodgkin's Lymphoma in the North-West region of Romania.

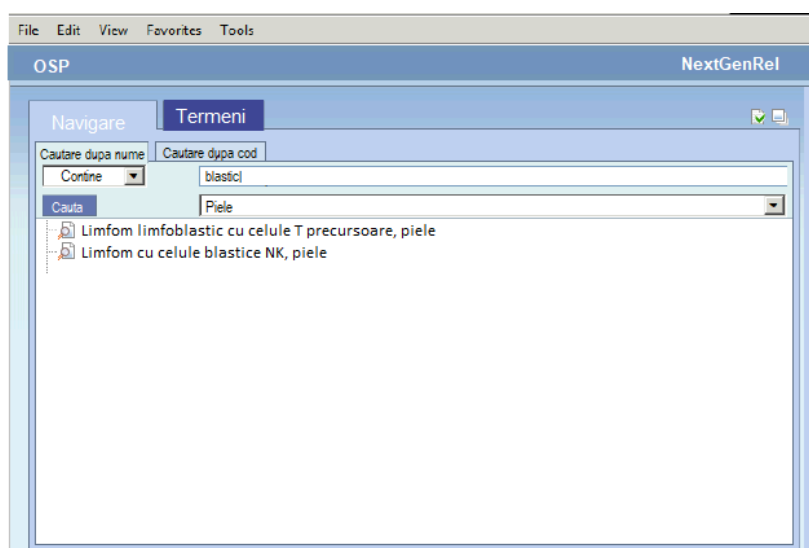
## Specific objective 2. Dissemination

Activity 2.2. Elaboration of an article for publication in a peer-reviewed ISI-indexed journal

Activity 2.6. Statistical analysis of data collected from the NW region of Romania

Primary cutaneous lymphomas are a heterogeneous group of T and B cell lymphomas that are present in the skin with no evidence of extra cutaneous disease at the time of diagnosis. Consensus classification of the World Health Organization and the European Organisation for Research and Treatment of Cancer (WHO-EORTC) from 2005 it has served as a gold standard for the diagnosis and classification of these conditions. In September 2018, an updated version of the WHO-EORTC classification was published in the fourth edition of *WHO Classification of Skin Tumours Blue Book*. Due to the fact that skin lymphomas represent a distinct subgroup that poses the biggest problems from the point of view of a lymphoma registry, I decided to study in particular this type of lymphoma in the context of testing the new generation software solution for recording non-Hodgkin lymphomas.

Primary cutaneous lymphomas (CL) generate diagnostic, therapeutic and registration challenges due to their rarity, heterogeneity and unique characteristics. So far, only a few population studies have been conducted and there is no population study focusing on CL in Romania. Our aim was to report incidence patterns for specific subtypes in the North-West region of Romania by translating unstructured records from the North-West Regional Cancer Registry (RRCNV) into the semantic platform of the Non-Hodgkin Lymphoma Registry. We also designed and performed an Ontology Based Search Procedure (OSP) for the identification of primary cutaneous lymphoma in pathology reports. By this way I managed to identify cases that had not been registered in the RRCNV and as such improve the quality indicators of the registry.



CL cases were classified according to the WHO-EORTC consensus classification of 2005, and where possible, some cases were re-evaluated and classified according to the 2018 update.

The cases of primary cutaneous lymphoma identified in the North-West region from 2008 to 2013 were included in the analysis. The results are summarized in Table 1.

<b>Table 1.</b>	<b>No.</b>	<b>Frequency, %</b>	<b>Cell type, %</b>	<b>M/F ASR Ratio</b>	<b>USA SEER<sup>1</sup> %, ASRR</b>	<b>Western Europe<sup>2*</sup>, %</b>
<b>Total</b> M: ASR = 0.24 F: ASR = 0.16	<b>48</b>			<b>1.51</b>	<b>1.72</b>	
<b>Cutaneous T-cell lymphoma</b> M: ASR = 0.18 F: ASR = 0.12	<b>34</b>	<b>70.8</b>		<b>1.49</b>	<b>71.5 1.70</b>	<b>75 - 80</b>
<b>Indolent clinical behavior</b>						
<b>Mycosis fungoides</b>	<b>9</b>	<b>18.8</b>	26.5		<b>38.3</b>	<b>39</b>
<b>CD30+ T-cell lymphoproliferative disorders of the skin</b>	<b>5</b>	<b>10.4</b>	14.8		<b>10.2</b>	<b>20</b>
<b>Primary cutaneous anaplastic large cell lymphoma</b>	<b>4</b>	<b>8.3</b>				<b>8</b>
<b>Lymphomatoid papulosis</b>	<b>1</b>	<b>2.1</b>				<b>12</b>
<b>Primary cutaneous acral CD81 T-cell lymphoma</b>	<b>1</b>	<b>2.1</b>	2.9		-	<b>&lt;1</b>
<b>Aggressive clinical behavior</b>						
<b>Sézary syndrome</b>	<b>2</b>	<b>4.2</b>	5.8		<b>0.8</b>	<b>2</b>
<b>Primary cutaneous peripheral T-cell lymphoma, unspecified</b>	<b>16</b>	<b>33.3</b>	47.1		<b>20.8</b>	<b>8</b>
<b>Precursor T-cell lymphoblastic lymphoma Blastic NK-cell lymphoma</b>	<b>1</b>	<b>2.1</b>	2.9		<b>0.2</b>	<b>-</b>
<b>Cutaneous B-cell lymphoma</b> M: ASR = 0.06 F: ASR = 0.04	<b>14</b>	<b>29.2</b>		<b>1.57</b>	<b>28.5 1.76</b>	<b>20 - 25</b>
<b>Indolent clinical behavior</b>						
<b>Primary cutaneous marginal zone B-cell lymphoma</b>	<b>5</b>	<b>10.4</b>	35.8		<b>7.1</b>	<b>9</b>
<b>Primary cutaneous follicle center lymphoma</b>	<b>1</b>	<b>2.1</b>	7.1		<b>8.5</b>	<b>12</b>
<b>Intermediate clinical behavior</b>						
<b>Primary cutaneous diffuse large B-cell lymphoma</b>	<b>6</b>	<b>12.5</b>	42.9		<b>11.4</b>	<b>5</b>
<b>Small lymphocytic lymphoma</b>	<b>1</b>	<b>2.1</b>	7.1		<b>0.8</b>	<b>-</b>
<b>Cutaneous Burkitt lymphoma</b>	<b>1</b>	<b>2.1</b>	7.1		<b>0.1</b>	<b>-</b>

1 Bradford PT, Devesa SS, Anderson WF, Toro JR. Cutaneous lymphoma incidence patterns in the United States: a population-based study of 3884 cases. *Blood*. 2009;113(21):5064–5073.

2 Willemze, R., Cerroni, L., Kempf, W., Berti, E., Faccchetti, F., Swerdlow, S. H., & Jaffe, E. S. The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood*, 2019 133: 1703-1714.

\*Based on data included in Dutch and Austrian cutaneous lymphoma registries between 2002 and 2017.

The specific incidence by age groups (per 100,000 persons-years) by sex, in the North-West region, between 2008 and 2013 is represented in figure 1 for primary skin lymphomas with T cells, respectively in figure 2 for primary skin lymphomas with B cells:

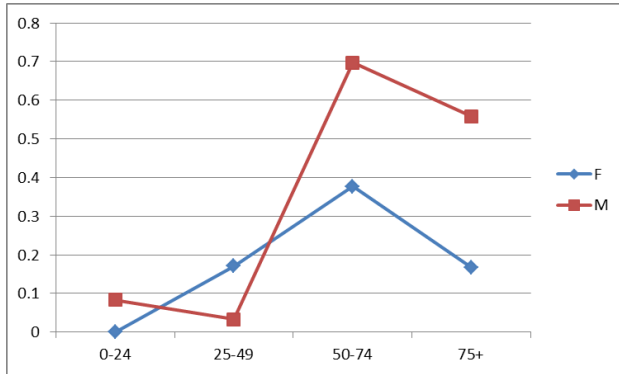


Figure 1.

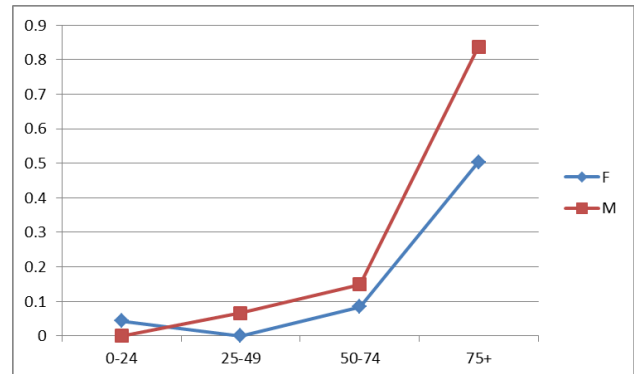


Figure 2.

There is a slightly higher incidence of cutaneous lymphoma in men from the north-western region of Romania. (ASR 1.51 rate). We also note a higher incidence rate of B-cell lymphoma in older men (over 75 years).

We note a slightly different frequency of the main subtypes of cutaneous lymphoma compared to those reported in the western world, with a higher frequency of intermediate and aggressive lymphoma subtypes. This could explain a lower relative survival rate at 5 years.

I have disseminated these results by presenting an E-Poster at the International Conference **EORTC Cutaneous Lymphoma Meeting 2019 "Insights in research and patient care"**, held in Athens, Greece, from 26.09.2019 to 28.09.2019. Presentation title: "**Primary cutaneous lymphoma patterns in the North-Western region of Romania**". The abstract was published in a dedicated volume of the **European Journal of Cancer (Volume 119: S23)**, ISI-indexed peer-reviewed journal with **Impact Factor 6.680**. DOI: 10.1016/S0959-8049(19)30575-1

For a thorough analysis it is necessary a re-evaluation of the cases, a standardization of the histopathological reports and for the future the registration of cases in the Non-Hodgkin Lymphoma Registry developed within this project.

A more in-depth analysis of the 48 cases of cutaneous lymphoma will be constituted in an article that we are preparing for publication in a peer-reviewed ISI-indexed journal.

### **Specific objective 3. Management**

#### **Activity 2.3. Monitoring the project and preparing the required reports**

As part of this activity, we continued to follow the coordinated and coherent development of the whole project until the stage objectives were met with the budget allocated and in accordance with the monitoring plan regarding the integration of all the activities of project implementation, control of risks and costs, internal communication and ensuring results. We considered updating the project website and preparing the necessary documentation for the purchase of a laptop, provided in the project budget, which I will use for conducting training sessions with the specialists who will be involved in using the Non-Hodgkin Lymphoma Registry. The website address of the project is:

<http://nwcanportal.iocn.ro/NextGenRel/Default.aspx>

Project Director, Dr. Bogdan FETICA

Mentor, Prof.Dr. Alexandru IRIMIE

### **References**

1. Zhang, H., Guo, Y., Li, Q. et al. An ontology-guided semantic data integration framework to support integrative data analysis of cancer survival. *BMC Med Inform Decis Mak* 18, 41 (2018).
2. AAIAbdulsalam, A. K., Garvin, J. H., Redd, A., Carter, M. E., Sweeny, C., & Meystre, S. M. (2018). Automated Extraction and Classification of Cancer Stage Mentions from Unstructured Text Fields in a Central Cancer Registry. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2017*, 16–25.
3. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*. 2012;13(6):395.
4. Esteban-Gil A, Fernández-Breis JT, Boeker M. Analysis and visualization of disease courses in a semantically-enabled cancer registry. *Journal of Biomedical Semantics*. 2017;8(1):46.
5. Quintanilla-Martinez, L. The 2016 updated WHO classification of lymphoid neoplasias. *Hematological Oncology*. 2017; 35( S1): 37– 45.
6. Rein Willemze, Lorenzo Cerroni, Werner Kempf, Emilio Berti, Fabio Facchetti, Steven H. Swerdlow, Elaine S. Jaffe; The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood* 2019; 133 (16): 1703–1714.