Joint Action on Networks of Expertise

Deliverable number D14.1

Recommendations to integrate local IT infrastructures for European cancer networking

Funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them

DOCUMENT INFORMATION

Author(s)	**Eivind Hovig, Larson Hogstrom**
Deliverable lead partner	Oslo University Hospital
Contributing partner(s)	
Work Package	TTF 3
Task	1-4
Deliverable type	Final report
Due delivery date	30/09/2024
Actual delivery date	30/09/2024
Dissemination level	Public
Annex version	2

Deliverable 14.1 is part of the Joint Action "JANE", GA 101075328, which has received co-funding from the European Union under the EU4Health programme

REVISION HISTORY

Version	Date of Issue	Author(s)	Brief description of changes
0.1	30/09/2024	OUS	Final draft

LIST OF ABBREVIATIONS

Abbreviation	Definition
TTF3	JANE Task Force Three
AI	Artificial Intelligence
IT	Information technology
WP	JANE work package
JANE	Joint Action Network of Expertise
GDPR	General Data Protection Regulation
EU	European Union

RECIPIENTS OF THIS DOCUMENT

This document is addressed to the whole JANE consortium. It is an official deliverable for the project and shall be delivered to the European Commission and appointed experts.

Table of contents

Executive summary

The Transversal Task Force Three (TTF3) working group was gathered to evaluate the current state of IT and AI capabilities and future opportunities within the context of European cancer networking. The group worked to map European related initiatives, formally gathered data from a significant number of partner institutions to understand current IT capabilities and worked to establish a path towards AI expertise mapping for JANE2 efforts.

Introduction

Transversal Task Force Three of JANE had four primary tasks related to technological capabilities:

1. Recruiting expertise on AI and IT resources.
2. Mapping partners' infrastructures to interact within the European space.
3. Mapping available options for computation.
4. Mapping cancer data and AI interoperability expertise.

By mapping and evaluating these and related themes, the TTF3 group sought to strengthen EU cancer research requiring IT resources and support other JANE WPs including WP9 on cancer genomics and WP10 on emerging technologies.

Methodology

Beginning in 2023, the task force leaders worked first to gather a working group and establish key partners. We utilized the following methodology to drive towards our aims and provide recommendations:

1. Expert discussion and WP collaboration
2. The construction of a European Cancer IT and Data Sharing Survey which resulted in data from 75 cancer care and research institutions across Europe from 23 different countries.
3. Mapping of EU initiatives and current collaborations

This section will discuss each of these methodologies and activities in turn.

1) Expert discussion and WP collaboration were conducted. We successfully onboarded 12 active TTF3 group members who are AI, data, or IT experts. We have worked with these members to grow a list of additional skilled academics and professionals working within the European Cancer AI community. The TTF3 team has been collaborating to discuss and map innovative solutions that would streamline AI research and applications within the JANE network. We have communicated with additional work packages including WP9 and WP10 to discuss required expertise moving forward into

JANE2 efforts. We are also connected to the full ELIXIR Europe network, including their established Cancer Data Community, as well as several EOSC infrastructure projects, and other relevant infrastructure projects.

This group mapped cancer data and AI interoperability expertise. The TTF3 made strides in evaluating IT systems for enhanced interoperability, with a particular focus on genomic and imaging data (efforts here include WP9 collaboration). Federated data sharing in cancer research allows institutions to analyze and share insights from patient data without centralizing or transferring sensitive information. This approach aims to preserve patient privacy and maintain data security, minimizing risks of breaches. A framework for the secure sharing and analysis of cancer-related datasets has been discussed including procedures and best practices for federated data networks. *Our efforts have earmarked federated data solutions as being of especially high interest for JANE and JANE2 moving forward*. The TTF3 group sees the value of these approaches for fostering collaborative research efforts within the EU.

2) Completion of Data Sharing and IT survey of European cancer research and care institutions.

The aim of the survey was to understand the baseline of current data sharing efforts and practices across many European institutions including regional hospitals, national hospitals, and academic medical centers. By better understanding the current baseline of data sharing practices, researchers and policy experts are better positioned to take realistic steps towards infrastructure improvements. The context of this survey is that continued systematic improvements to cancer care and research within Europe require better data sharing practices, but these face hurdles around legal, ethical, and technological issues.

Early in our TTF3 expert discussions, we identified the need to improve EU cancer data sharing capabilities to enable AI opportunities and accelerate the pace of biomedical and clinical research on cancer. Although policy improvements and technical frameworks have garnered substantial support in this area, the road to improving European data sharing capabilities is riddled with challenges that are unique to needs, capabilities, and resources of individual member states.

To understand current data sharing capabilities and resources across institutions who will be impacted by JANE and JANE2 efforts, we conducted a survey. We aggregated results from 75 cancer care and research institutions across Europe from 23 different countries. Results spanned different institution types from well-resourced international research hubs to regional cancer care centers. In addition to gathering information about institution type, we gathered responses to 16 questions that covered the extent of data sharing, technological capabilities around data structure and storage, and current

challenges faced by respondents in regard to data sharing for cancer research purposes. We gathered responses from one or two individuals at each institution. Respondents were affiliated across a diverse set of departments such as medical oncology, IT, radiotherapy, and management and administration.

Of the institutions surveyed, 73% and 71% participate in clinical trials and primary cancer care respectively. Similarly, most institutions contribute to basic research (60%) and translational research (67%). There were 27% of institutions who provide IT service partitions. About half of the institutions are academic medical centers (48%). In the context of providing care, 40% were primarily research institutions while 20% were regional hospitals.

Detailed results of the survey regarding data sharing are discussed in the next section of the report.

3) Mapping of EU initiatives and current collaborations. One outstanding challenge related to EU cancer networking is how to align the large number of initiatives across Europe that improve or utilize IT and cancer data. These will touch on data generation and sharing. During our task force work, we engaged with leaders from several initiatives. The ELIXIR Cancer Data Community is a bottom-up-organized community centered on cancer data from a research perspective, where the connection between useful infrastructure and timely research questions is a critical focus (Nikolski et al. 2024). This is organized within the realm of ELIXIR Europe, which is the European research infrastructure for life science data and is highlighted as a European landmark by the European Strategy Forum for Research Infrastructures, ESFRI. ELIXIR drives the Genomic Data Infrastructure (GDI) project. The purpose is to enable access to genomic and related phenotypic and clinical data across Europe, both population genomic data and for diseases, including cancer. This project is building infrastructure to achieve data sharing across countries, based on principles of open science. One main purpose of this undertaking is to serve the 1+ million genomes (1+MG) initiative. The ambition of this project is to provide solutions for realizing the sharing of clinical, phenotypic and genomic data across countries for cancer, rare diseases, complex diseases, and infectious diseases. One of the various working groups within the European 1+MG initiative, is the cancer-related working group. It coordinates cancer-related topics and provides relevant guidance in the context of data sharing (Riba et al. 2024). The European Open Science Cloud structure has funded the EOSC4Cancer project, which is endeavoring to drive infrastructure functions based on cancer pilots within colorectal cancer. This includes data harmonization, software for analysis, including novel data types for dynamic inclusion into larger workflow systems, as well as data sharing. EUCAIM is located in the cancer image field, developing federated solutions geared towards reaping the benefits of AI in cancer diagnostics and treatment. Another aspect is that of public-private partnerships, where the BiGPicture Imaging Consortium is an example of data images for

common benefit. Many of these initiatives are funded through the EC to drive development of useful infrastructure to meet the needs for analysis across large datasets in effective ways within research. This is addressed both in a generic disease-agnostic way, but also with the cancer domain increasingly being a driver, through the Cancer Mission, with its four objectives of: 1) Understanding cancer, 2) Prevention and early detection, 3) Diagnosis and treatment, 3) Quality of life for patients and their families. The ambitions are to some degree aligned and also to some degree organized bottom-up based on research competence.

With the above approaches, we believe we have been able to grasp fundamentals of the current picture of the field. Interactions between these actors and the impact on cancer networking are discussed below.

State of play and mapping the future

Some of the main lessons learnt include the very variable maturation level of the institutions that will be required to effectively share data, the complex setting of laws and regulations, and their interpretation in the EU, including each country and institution, the lack of a robust computational platform for use for shared data, both for federation in general, and for AI in particular. Although elements of solutions exist, significant work will still be required. This will also impact the speed with which networks of expertise will be able to move forward while timely addressing the challenges. Below, we discuss in some more detail our findings.

Data Survey Results highlight a need for improved technology, legal, and administrative support across Europe. Our survey results point to addressable gaps in infrastructure and legal context to improve scientific and clinical data sharing. We asked respondents what they perceived to be the largest non-technical hurdle related to sharing cancer data for research. The most common responses related to general legal challenges (24%) or General Data Protection Regulation (GDPR) and privacy compliance (30%). While understandable in the light of different laws between countries, these perceived challenges are likely exacerbated by different GDPR interpretations and compliance approaches. Other common non-technical issues that were reported were related to resourcing (cost or staffing) (15%), and administrative complexity (13%). In this area, European Cancer healthcare networking could work to address these by disseminating administrative best practices or providing resources that could operate on an international scale.

Similarly, there are large (but addressable) technical hurdles that hinder data sharing for cancer research and networking. The main reported technical hurdles related to data sharing included

interoperability issues (35%), security and privacy protocols (17%), storage space or connection speed (14%), and/or technical staffing (7%).

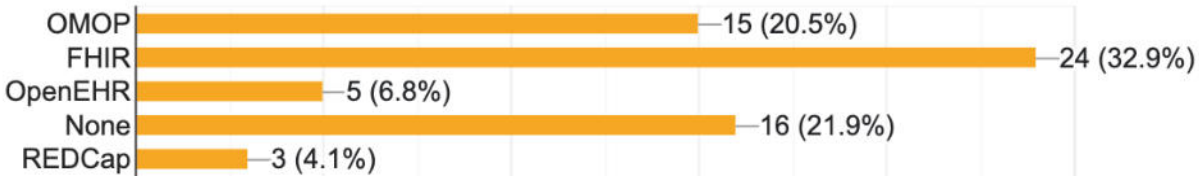

Figure 1. Frequency of utilization of the top 4 most common data models for clinical and EHR data. Institutions can select multiple data types if applicable.

For clinical data storage, we found utilization of diverse data structures and protocols. Common data models that standardize the format and content of healthcare data were utilized for data sharing at most institutions. Although a number of institutions utilized Fast Healthcare Interoperability Resources (FHIR) (33%) or the Observational Medical Outcomes Partnership (OMOP) (21%), there was no clear winner suggesting that substantial gaps currently exist in capabilities for providing interoperable transfers of clinical data (see Figure 1). The need for and importance of improved healthcare record standardization cannot be overstated. Efforts including clinical care improvements, epidemiological, and basic research initiatives will rely on these records for tracking patient treatments and outcomes. European Health networking efforts need to conduct further research and engage input from stakeholders in this area. Here, networking activities may play an important role in making more explicit recommendations and enforcing standards as engagements mature.

Developments in the field

The collective ongoing efforts can be considered preludes and pilots to the upcoming European Health Data Space, which carries the promise to empower individuals to "*take control of their health data and facilitate the exchange of data for the delivery of healthcare across the EU (primary use of data)*", to "*foster a genuine single market for electronic health record systems*" and to "*provide a consistent, trustworthy, and efficient system for reusing health data for research, innovation, policy-making, and regulatory activities*". This is a very ambitious undertaking, taking into consideration the complex European landscape, viewed at several levels. One aspect is that standardization is still very far from being achieved at the level of cancer hospitals, which is the main source of such data. Standards exist but are only now becoming relevant from many institutions. This is true both for clinical/phenotypic data, clinical trial data, as well as for many types of molecular data. Similarly, to realize the power of AI on large datasets, there is a need for data federation, due to the complex legal setting in Europe. The General

Data Protection Regulation, GDPR, applies, but efficient and reliable sharing of data is hampered by both national laws and regulations on top of this, as well as the institutional interpretation of this landscape. The consequence is that technical development in many instances requires the use of synthetic data, as real world data are not forthcoming due to legal and ethical restrictions. To realize the potential of federation, there is also a need to establish federated computing solutions, which are currently mostly missing. A number of additional requirements exist, including authentication and authorization infrastructure (AAI), data discovery and access options.

Recommendations seen from a networking perspective

- From ongoing efforts to realize federated analysis across member states, it is becoming increasingly clear that the implementation of the GDPR is heterogeneous, both between member states and even between and within institutions. Similarly, the ethical framework related to patient consent for secondary use of their data is also subject to different interpretation and granularity of implementation. This applies to both the data, but also to the design of relevant infrastructure. The consequence will often be more complex solutions, and with less built-in security. The regulatory setting would benefit from improvement in order to foster networking across member states.
- While development of infrastructure through European research funding is a good thing, it is not entirely clear whether all lessons learnt from ongoing initiatives are systematically registered, and it may be the case that overlapping activities occur. Better alignment between funding actors, both nationally and on the European level would be welcome. In this context, the 1+MG initiative is a very good lead in engaging both the governmental level and the scientific level from each country. This type of approach should be fostered to more quickly unleash the potential for large-scale European cancer data. Networks would be a useful provider of input in this context.
- Currently, the incentives of the healthcare system may be improved. Given that the healthcare system is under strong pressure to deliver health to patients, there is little room for improvement within the system to invest in activities that may not provide immediate benefit. Thus, incentives should be developed, potentially from the governmental level of the member states, in order to achieve the necessary infrastructure improvements required to realize a true data federation environment across Europe.
- Given the significant variation in maturation that we found through our IT survey, it seems sensible to argue for a **federated data sharing pilot** in the context of JANE2 networking, focused on early-adopters. It would be highly advisable to leave room for both high- and low-

resource institutions and member states, as it is likely to learn valuable lessons from this existing complexity.

- EU cancer networking should **serve as a catalyst for data aggregation and statistical modeling efforts** between member states. An important part of a pilot would be to mobilize electronic health data in standardized ways. These data require systemic support from each institution and will be a critical step in realizing the potential of AI, e.g. for epidemiological research. A potential pilot could be within the JANE WP9 domain, with carefully selected research questions to demonstrate the power of federated and multilayered molecular and clinical data. EU cancer networking should be set up with **shared resource provision for administrative compliance and dissemination of best practices on compliance.** Legal resources should be part of such activities, as no currently existing project alone is able to significantly impact the legal situation. Further, there is little impetus to comply with existing data and analytical standards.
- The European Health Data Space (EHDS) was approved by the European Parliament in April 2024. The actual usefulness of this creation will critically depend on both the regulatory landscape, the governance and the functionality provided. The concepts described involve all relevant actors, including patients, clinicians and researchers. It will be important to register all lessons learnt from ongoing initiatives when realizing the EHDS, in order to prevent monolithic systems with rigidity towards change, and rather develop lean solutions built on standardized open protocols. One should plan also for the inclusion of networking aspects in the creation of the EHDS.

A paper entitled, "*The landscape of European cancer data sharing: analyses of international survey results*," is currently being drafted with publication expected for early 2025.

References

Macha Nikolski, Eivind Hovig, Fatima Al-Shahrour, Niklas Blomberg, Serena Scollen, Alfonso Valencia, and Gary Saunders. 2024. "Roadmap for a European Cancer Data Management and Precision Medicine Infrastructure." *Nature Cancer* 5 (3): 367–72.

Michela Riba, Cinzia Sala, Aedin C. Culhane, Åsmund Flobak, Attila Patocs, Kjetil Boye, Karla Plevova, et al. 2024. "The 1+Million Genomes Minimal Dataset for Cancer." *Nature Genetics* 56 (5): 733–36.

Appendix 1 - List of survey questions

Question number	Question
1	Country
2	Institution Name
3	Department Name
4	Employment Title
5	Institution's involvement with cancer data
6	Institution type
7	Institutions delivers patient care (select all that apply)
8	What type of data is currently collected at your institution (select all that apply)
9	Which of these data types is currently be shared with external collaborators for research purposes? (select all that apply)
10	If Yes to any of the above, who was this data shared with?
11	Apart for GDPR, what legal restrictions are in affect for your institution related to data sharing?
12	What data standards or platforms are being used for clinical information storage or sharing?
13	If you are not sharing data at this time, do you have any plans to do it in the future?
14	If yes to above, please briefly describe the intended data sharing plans related to cancer research
15	What do you perceive as the main non-technical hurdle for sharing data?
16	What computational environment are you use to process and store genomic data?
17	What do you perceive as the main technical hurdle for sharing data?
18	Are you using commercial resources for data processing and storage? Or using open source software?
19	Do you provide genomic information to molecular tumor boards (MTBs)?
20	How is data from molecular tumor boards structured and stored?
21	Currently using standardized data formats for imaging or genomic data?
22	Are you relying on national or international service providers for sharing of data?
23	Would you be open to participating in future surveys or case-study interviews?

Appendix Table 1. A list of questions provided to survey participants.