strong aya

A new, interdisciplinary, multi-stakeholder European network to improve healthcare services, research and outcomes for Adolescents and Young Adults with cancer.

NETHERLANDS CANCER INSTITUTE
ANTONI VAN LEEUWENHOEK

EORTC
European Organisation for Research and Treatment of Cancer
The future of cancer therapy

YOUTH CANCER EUROPE

UNIVERSITY OF LEEDS

University of Southampton

NHS
The Leeds Teaching Hospitals
NHS Trust

Maastricht University

european cancer ORGANISATION

FFUND

CENTRE DE LUTTE CONTRE LE CANCER LEON BERARD

Narodowy Instytut Onkologii
im. Marii Skłodowskiej-Curie
Państwowy Instytut Badawczy

FONDAZIONE IRCCS
ISTITUTO NAZIONALE
DEI TUMORI

GUSTAVE ROUSSY
CANCER CAMPUS
GRAND PARIS

MANCHESTER 1824
The University of Manchester

IKNL integraal kankercentrum Nederland

Deliverable Report

WP3 – Infrastructure and interoperability

Deliverable D3.4

Development of core elements of stakeholder-centric reports

Due date of deliverable: 30/09/2024

Actual submission date: 31/10/2024

Project:	STRONG-AYA
Lead Contributor	Leonard Wee, UNIMAAS
Email	leonard.wee@maastro.nl
Other Contributors	Bart van Beusekom, IKNL Joshi Hogenboom, UNIMAAS Simone Hanebaum, NKI Alizee Diatchenko, CLB Urske Kocir, YCE
Emails	b.vanbeusekom@iknl.nl; joshi.hogenboom@maastro.nl; s.hanebaum@nki.nl; Alizee.DIATCHENKO@lyon.unicancer.fr; urska.kosir29@gmail.com

Due date	31 09 2024
Delivery date	31 10 2024
Deliverable type	R
Dissemination level	SEN

Description of Work	Version	Date
	V1.0	**31/10/2024**

Description:

These core elements will form the basis of the stakeholder-centric reports, developed over the course of the project in at least two rounds of iterative revision.

Summary

The STRONG AYA data ecosystem and federated learning infrastructure is a digital artefact for sharing insights from data. Technical and governance of the data ecosystem and federation technology is described elsewhere in other deliverables. This deliverable reports on the CORE ELEMENTS of how the insights gained from federated data analysis will be shared amongst approved users.

By reports, we mean the digital outputs of our technological infrastructure, executed on demand by the user, which principally consists of group statistics and/or epidemiological statistical modelling. The raw output, of itself, is not easily understandable and not cognitively easy to digest by a diverse and highly heterogeneous community of potential users (a.k.a. “stakeholders”), therefore the emphasis of the reports needs to be on DATA ANALYTIC VISUALIZATIONS, GRAPHICAL SUMMARIES, INTERACTIVE DASHBOARDS, and other similarly-related ways of disseminating insight derived from background processing of richly granular data.

By core elements, we mean to identify the modules and components that can be adaptably and flexibly chained together in a digital fashion (such as a website or a web-application) so that a given target user may, to the best degree that we are able, be given a user-customized way of understanding the analysis results. If we intend to do this in a highly sustainable manner with less labour-intensive effort, we need to design user-centric (stakeholder-centric) visualization concepts around a set of modular core elements.

This report therefore describes the consultative process of multi-stakeholder co-creation, to identify at an early stage of the project, what are the most basic core elements of digital reporting, that can be later assembled into easily digestible views of data analysis results that are fitted to the needs of different potential users. We then present the initial concept plans and mood board, to give an example of how certain elements can be brought together. While it is not possible at this stage to create mood boards to represent every conceivable type of user, we will point out how the software tools behind the digital concepts can be edited in the background, so that it would be possible to create “views” (reports) fitting to different groups of users (stakeholders).

1 Table of Contents

2 Definitions

STRONG AYA consortium members are referred to as following within this text:

1. **NKI-AVL** – Stichting het Nederlands Kanker Instituut – Antoni van Leeuwenhoek Ziekenhuis (NL)
2. **YCE** – Youth Cancer Europe (RO)
3. **INT** – Fondazione IRCCS Instituto Nazionale dei Tumori (IT)
4. **FFUND** – FFUND BV (NL)
5. **CLB** – Centre de Lutte Contre le Cancer Leon Berard (FR)
6. **ECO** – European Cancer Organisation (BE)
7. **UNIMAAS** – Universiteit Maastricht (NL)
8. **IKNL** – Stichting Integraal Kankercentrum Nederland (NL)
9. **EORTC** – European Organisation for Research and Treatment of Cancer AISBL (BE)
10. **IGR** – Institut Gustave Roussy (FR)
11. **MSCNRIO –** Narodowy Instytut Onkologii im. Marii Sklodowskiej-Curie – Panstwowy Instytut Badawczy (Marie Sklodowska-Curie National Research Institute of Oncology) (PL)
12. **UOM** – University of Manchester (UK)
13. **UOL** – University of Leeds (UK)
14. **LTHT** – Leeds Teaching Hospitals National Health Service Trust (UL)
15. **SOUTHAMPTON** – University of Southampton (UK)

- **Grant Agreement** (including its annexes and amendments): the agreement signed between the beneficiaries of the HORIZON Research and Innovations Actions (hereafter referred to as Horizon) and the European Health and Digital Executive Agency (hereafter referred to as HADEA) for the undertaking of the STRONG AYA project (Grant Agreement no. 101057482).
- **Beneficiary**: Signatories of the Grant Agreement
- **Associated Partner**: Entities which participate in the action but without the right to charge costs or claim contributions.
- **Project**: the sum of all activities carried out in the framework of the Grant Agreement.
- **Consortium**: the STRONG AYA consortium, including all the aforementioned partners.
- **Consortium Agreement**: The agreement made between STRONG AYA members for the implementation and execution of the action outlined in the Grant Agreement. The agreement shall not affect the parties’ obligations to HADEA on behalf of the European Union, and/or to one another arising from the Grant Agreement.

3 Abbreviations

Acronym/Abbreviation	Meaning
HCP	Health Care Provider
PRO	Patient Reported Outcome
PROM	Patient Reported Outcome Measure
COS	Core Outcome Set
WP	Work Package
WPL	Work Package Lead(s)
WP1	Work Package 1 (Development Core Outcome Set AYA with cancer & data collection)
WP2	Work Package 2 (Governance, Data Security and Ethics)
WP3	Work Package 3 (Infrastructure and Interoperability)
WP4	Work Package 4 (Operation of STRONG AYA ecosystems, stakeholder and patient involvement, dissemination, exploitation, communication)
WP5	Work Package 5 (Scientific coordination and project management)
KPI	Key Performance Indicator
OA	Open Access
PAB	Patient Advisory Board
EC	European Commission
HADEA	European Health and Digital Executive Agency
SC	Steering Committee
MT	Management Team
DMP	Data Management Plan

4 Project Introduction

Project background

To assist the reader with the background of this project, the following texts have been derived from standard documentation pertaining to the entirety of the STRONG AYA project :

Cancer at adolescent and young adult (AYA) age (15-39 at primary diagnosis) is rare. However, this rarity does not reflect the significant personal and societal costs of cancer in this population, as reflected in the potential years of life lost or saved, the decreased productivity and quality-of-life due to the impact of the disease during formative years, and the long-term complications or disabilities[1]*. AYAs with cancer form a unique group; they face age-specific issues (e.g. Infertility, unemployment, financial problems) and decreased quality of life due to cancer and its treatment. Unlike dedicated healthcare and trials for paediatric cancer patients, AYA-specific healthcare services are scarce and vary across Europe. AYAs who are at the core of society and the economy need access to age-appropriate and high-quality healthcare.*

The STRONG-AYA project aims to tackle the underrepresentation of AYA's experiences and outcomes when navigating the healthcare system and in clinical care by developing national infrastructures for outcome data management and clinical decision-making within a pan-European ecosystem and establishing communication feedback for AYAs with cancer and the healthcare systems. This will be key to improving healthcare services, research, outcomes and policies for AYAs and to ultimately better cancer care for this patient group. The project brings together an international multi-disciplinary consortium (academic/research, clinical, stakeholder and patient organisations) across seven European countries.

It has been identified that one of the most important contributions to AYA research would be to pool data (e.g. patient-reported outcomes, clinical and treatment data) across institutions and countries and create large cohorts for researchers to address the burden of cancer in AYA[2]*. There is at present a lack of data standardization, data interoperability and (prospective) collection of outcomes of relevance for AYAs with cancer across cancer care institutions across Europe.*

The STRONG-AYA project aims to tackle the underrepresentation of AYA's experiences and outcomes when navigating the healthcare system and in clinical care by developing national infrastructures for outcome data management and clinical decision-making within a pan-European ecosystem and establishing communication feedback for AYAs with cancer and the healthcare systems. This will be key to improving healthcare services, research, outcomes and policies for AYAs and to ultimately better cancer care for this patient group.

Building on previous initiatives, a STRONG-AYA data ecosystem will be set up for value-based care, research and policy for AYA with cancer by:

1. ***Developing a Core Outcome Set** (COS) specifically for AYAs with cancer, via a participative consensus process defining most important aspects for those directly affected by AYA cancer, including patients and healthcare professionals.*
2. ***Implementing the COS across several national European healthcare systems**. Data will be collected at local level and will then be included in a data integration platform. An overall ecosystem framework for data analytics and output will be created supporting federated analyses and the creation of reports across clinical and patient-reported data and making national repositories of (patient-reported) health data, available to individual patients, patient*

---

[1] Stoneham SJ. AYA survivorship: The next challenge. Cancer 2020; 126: 2116-2119.

[2] Smith AW, Seibel NL, Lewis DR et al. Next steps for adolescent and young adult oncology workshop: An update on progress and recommendations for the future. Cancer 2016; 122: 988-999.

*organisations, regulatory authorities, as well as the patients' health care providers to inform clinical decision-making. The five resulting **national ecosystems** will be connected to each other into **the pan-European ecosystem** using a federated approach also utilizing the overall ecosystem framework.*

3. ***Disseminating the COS to a wide range of local as well as pan-European stakeholders**, in particular by developing analytical tools to process and present patient outcome data and establish feedback loops that inform patients and clinicians.*

Implementing these project objectives are five work packages within the STRONG-AYA project:

- *WP1: Development Core Outcome Set AYA with cancer and Data Collection (Lead: SOUTHAMPTON)*
- *WP2: Governance, Data Security and Ethics (Lead: EORTC)*
- *WP3: Infrastructure and Interoperability (Lead: UNIMAAS)*
- *WP4: Operation of STRONG-AYA ecosystems, Stakeholder and Patient involvement, Dissemination, Exploitation, Communication (Lead: E.C.O.)*
- *WP5: Scientific Coordination and Project Management (Lead: NKI)*

STRONG-AYA will therefore enable AYA care and research to benefit from collection and pooling of patient-centered data and collaboration among all stakeholders: patients, healthcare professionals, scientists, and policymakers. More widely, the project will leverage the network of interested organisations and networks established under STRONG-AYA for long-term strengthened promotion of the necessary implementation of specialist AYA cancer services across Europe. This will ultimately bring novel insights into AYA cancer care, research and policy, contributing to the long-term improvement of outcomes for people with AYA cancer.

5 Deliverable introduction

The STRONG-AYA consortium believes that innovative and paradigm-breaking ways of accessing privacy-sensitive multi-institutional data, and facilitating sharing of knowledge across institutional data siloes is urgently needed to improve healthcare for this unique group of people. The key pillars of STRONG AYA are : (1) consensus (Delphi) development of a Core Outcomes Set i.e. COS; (2) federated learning infrastructure for collecting, managing and enriching AYA data among participating countries; (3) disseminating data-driven insights and data analysis tools at pan-European level to help improve patterns of care for AYAs everywhere.

A **federated data ecosystem paradigm** for data management and data analysis was chosen as a method to address fragmentation of data, to address adoption barriers associated with centralized patient data repositories, and to be more easily able to support scale-up with a large number of expected future partners/institutions. The technological cornerstone of the project is **federated analytics**, i.e. performing statistical data analyses and generating summary visualizations of data-analytic results but without transmitting any individual patient-level data outside of the data-owning institution. STRONG-AYA additionally enables ***federated learning*** such that prognostic and predictive (epidemiological) models can be developed without the need to transfer individual patient-level between participating institutions.

This deliverable report specifically focusses on the dissemination of the results of the data analysis and federated learning. However, the raw results of data processing for statistics and/or statistical modelling is not easily understandable to a wide and heterogenous community of potential users. Therefore the emphasis of the reports from data processing needs to be on DATA ANALYTIC VISUALIZATIONS, GRAPHICAL SUMMARIES, INTERACTIVE DASHBOARDS, and other similarly-related ways of disseminating insight derived from the digital computer-based processing of the underlying richly granular AYA personal data.

Furthermore, we know that a one-size fits all approach cannot be an appropriate solution, since we anticipate that each type of user/stakeholder might have different needs and different levels of engaging with the results of the data analysis. Therefore, the digital tools for reporting needs to be conceptualized as modular components, that can later be assembled easily in software, so that we can later on customized specific tools for specific communities of users.

5.1 Disclaimer

This document is intended as a living document. As requirements develop and more functionality becomes feasible, there shall be changes and updates. The reference copies of each version will be kept on the NKI STRONG-AYA consortium Sharepoint®.

5.2 Complementary documentation

This document is for the use of Partners of STRONG-AYA Project (reference number 101057482) as defined in the General Agreement (executed 18 May 2022) and the Consortium Agreement (executed on date 19 December 2022).

This document may be used complementarily with the STRONG-AYA business architecture (D2.1), The Code of Conduct (D2.2) and the overarching operational plan(s) for the STRONG AYA ecosystems both national and pan-European (D4.4, D4.6, D4.7), hence this report mainly addresses the technological components. This document may also to be used complementarily with the STRONG-AYA Data Management Plan (D3.1) and the STRONG-AYA technical architectural blueprint (D3.3).

6 Consultation and identification of preliminary user archetypes

The consultations for the core elements of reports were principally created through (i) intensive interviews with the STRONG AYA partners, representing a wide professional cross-section of AYA-treating physicians, clinical researchers, cancer survivors and technology experts. Secondly, a public stakeholder forum was organized and sponsored by consortium partner European Cancer Organization in September 2023, which was used as an advisory contact point with a wider group of potential users.

In these consultations, it was clear that we cannot represent every unique point of view, because our community of future users will be vast and highly heterogenous. The heterogeneity is in level of education, level of familiarity with using digital tools, level of personal engagement with the topic of this consortium, and most fundamentally heterogeneity in the objective/purpose of seeking information from data analysis in the first place.

We tried to synthesise the intelligence gleaned from the above focus settings, to come up with initially 5 user archetypes that could help parameterize the digital design work. We know these fictional archetypes drawn from the forementioned interviews and community consultation cannot ever be exhaustive, but we wanted to used these as “boundary markers” to help us parameterize the design space inside which we need to plan our initial work. To these 5, we have added one implicit “user” (marked * below) which is based inside the STRONG AYA consortium leadership and needs tools for operationalization of data analysis from within the consortium. The present archetypes are presented as follows in no particular order :

1) Recently-diagnosed AYA with cancer

2) Long-term cancer survivor

3) Specialist clinical nurse working in an AYA care ward
4) Senior clinician-leader attempting to set up/re-design specialist AYA-specific cancer care
5) Senior policy advisor to government or political organization
6) Research leader within the STRONG AYA consortium (*)

JAN, 19
STUDENT

- Photography student
- Handball
- Colon cancer
- Disbelief
- Avoidant
- Fear for future

“Concise info shown by healthcare practitioner.“

“Little interaction but the right information.”

Image: Generated with Adobe Firefly

RHYS, 28
RESTAURATEUR

- Entrepreneur
- Stable disease
- Brain tumour
- Monitor wellbeing
- Sporadic interaction
- Compare with peers

“Simple information in easy and relaxed format.”

“Noncommittal reminders and monitoring.”

“Insight into own history.”

Image: Generated with Adobe Firefly

LAZZARO, 38
NURSE

- Specialised nurse
- AYA-ward
- Digitally literate
- Correct information
- Help AYAs navigate
- Explain the situation

“Accessibility without technical knowledge.“

“Clear information on suggested topics.”

“Modifications to draw a more relevant picture.”

Image: Generated with Adobe Firefly

PAULINE, 53
POLICY

- Senior resident
- Business-case
- Set up AYA-ward
- Care-needs
- Specialism necessity
- Staffing requirements

“Accessibility without technical knowledge.“

“Clear information on suggested topics.”

“Narrative creating for specific settings.”

Image: Generated with Adobe Firefly

ADA, 64

POLICY

- Large political body
- Various portfolios
- Not so digitally literate
- Regional disparities
- Societal impact
- Policy-actionability

“Direct and quick numbers.“

“Simple and high-level information.”

Image: Generated with Adobe Firefly

> ***Internal stakeholder – research leader within the STRONG AYA project.*** *The principal concerns here are about (i) monitoring data completeness and data quality, (ii) consistency and inter-operability of the data and (iii) able to execute on-demand statistical modelling, including a range of statistical tools for data exploration and/or fitting of statistical models.*

6.1 Summary synthesis of consultations

Simple accessibility, with technological complexity placed “under the covers”

In all our consultations to date, one of the major impressions was to "hide the technological complexity", to which we wholly agree. The internal functioning of the technology infrastructure should not be so obvious or distracting to the user, even the expert users (such as the clinical lead or the research nurse). Feedback was strongly in favour of being able to get the information without having to do any programming or statistical manipulation by the users themselves. Pertinent comments related to this included statements such as "not digitally literate", "need simple and direct results" and "accessible without deep technical knowledge".

Self-navigable, layered information presentation with links to additional resources

The other major impression gleaned was a strong level of support for a simple, smooth but unthreatening digital interface. There was very little or no desire amongst our consulted co-creators for nudges, pushiness, or invasive reminders about their (cancer) condition. Some basic information was needed in a quickly-accessible way but only at the highest (i.e. most general overview), however certain types of users expressed some desire to have an option (only if they voluntarily choose to do so) to either drill deeper into the data visualization to explore sub-groups (mainly clinician and research oriented users) or to search for more details about where to get extra advice (mainly the patient-advocate and cancer survivor-oriented users).

Soft look and unthreatening presentation of easily understandable statistics/numbers

In terms of the information presentation itself, the expressed desire was for "pre-chewed" information (which we take to understand the results of pre-set data analysis and processing) rather than the raw numbers itself. The concern expressed was that either the analytic complexity itself would be too overwhelming for many users, or that the information presented would be quite confrontational and potentially "triggering".

There was only one community of users (researchers and senior clinicians) that expressed a desired to be able to ask for the results of statistical summaries in a downloadable tabular format (such as CSV) so that they could re-plot the same information by themselves as pie-charts/bar-charts/survival curves by themselves for the purpose of scientific publication and professional communication with their clinician-peers.

7 Core elements

1. In light of the initial consultations, we selected to aim at an implementation as a web-browser based interface at the present time. This might be accessed through a standard HTML-supporting web browser on a desktop computer, a laptop computer or a smartphone. The decision at the moment was to delay the implementation of a smartphone-only device-specific "app" until we had more time to mature and understand the users and use cases. This was not deemed a major deficiency at the early stage of the project, because the feedback we received suggested that apps-based notifications, intrusive reminders and social network-style of heavy engagement was not so high on the priority list. Therefore, one (or more) registered and controlled world wide web Unique Resource Locator (URL) will be a core requirement so that users have a landing page with their browser (see below **Fig.1**).

2. Access control and level of functionality must be linked to that user profile. Here, the feedback was guiding us towards implementing at least two very broad levels of access.

First, there might be an interested member of the public or a general user who does not want to have a registered account. It was strongly felt to support this level of utilization because, for example, the Global Cancer Repository (globocan), the Dutch National Cancer Registry and the European Cancer Organization have similar public-facing data dashboards about all cancer and all patients in general (see examples in appendix A).

The STRONG AYA work gives insights into a sub-demographic of cancer patients and cancer survivors (i.e. AYAs) for which specific data is both rare and extremely valuable. Therefore, at least the most basic features and some of the rudimentary highest-level data visualizations ought to be available publicly.

On the other hand, some of the more specialized functionalities - such as data exploration, hypothesis generation and statistical modelling – are certainly supported by the federated learning architecture, but we do not necessary want to open these to all and sundry. This means that, in addition to a general public landing page, there will be a restricted area with specialized functionalities that require user registration, user password and multi-factor authentication to access (see also **Fig.1**).

3. Registered user profiles were considered to be a useful element to preserve, for the case of repeat users and/or users with higher level data analysis privileges (see **Fig.2**). For the general no-account no-password general access visitor, we do not feel that such profiles are per se necessary, and for them we may be able to skip directly to topics-based modules about cancer care, treatment and outcomes.

4. Topics of interest to a given user or given user profile may be stored as topic modules. However, these topical modules will need to be co-developed and co-created with HCPs and AYA representatives. Two potential use cases have already featured prominently in our discussions with AYAs – information you may need during treatment or right after diagnosis, versus information that you may want to look up once you are off treatment or as a long-term cancer survivor (see **Fig. 3**).

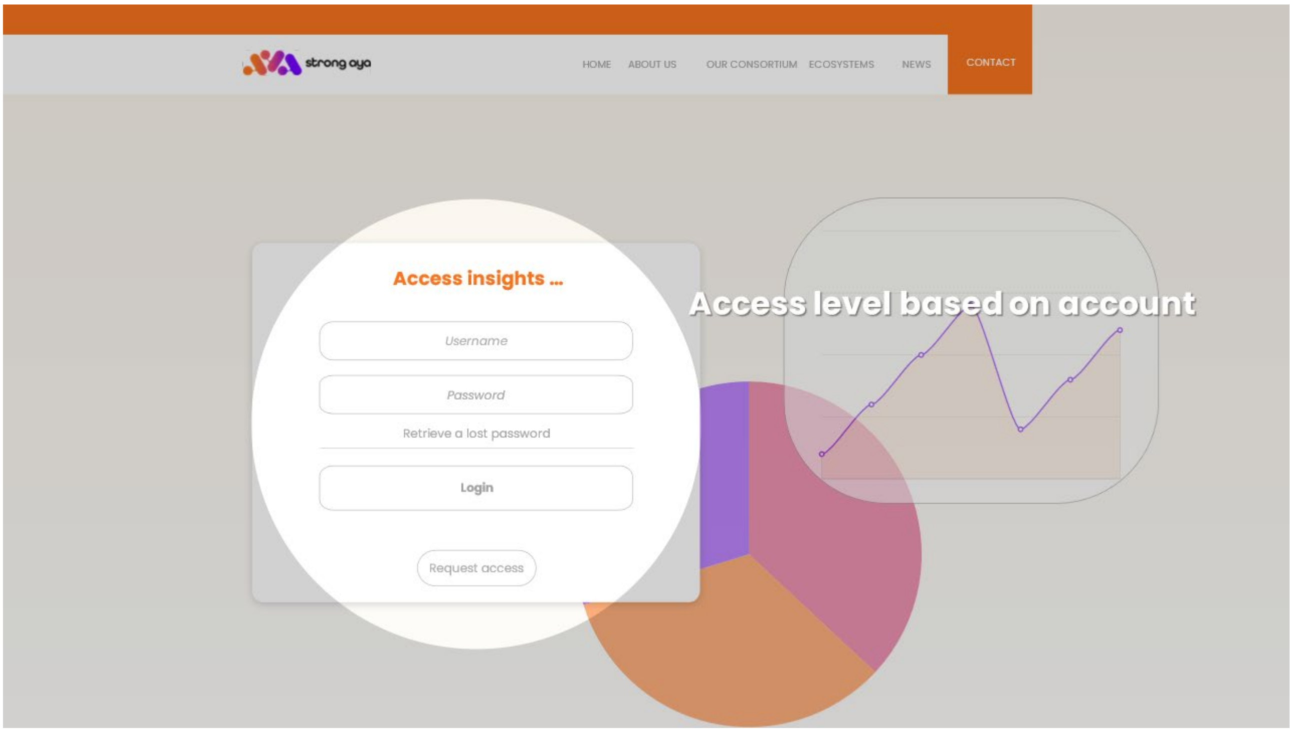
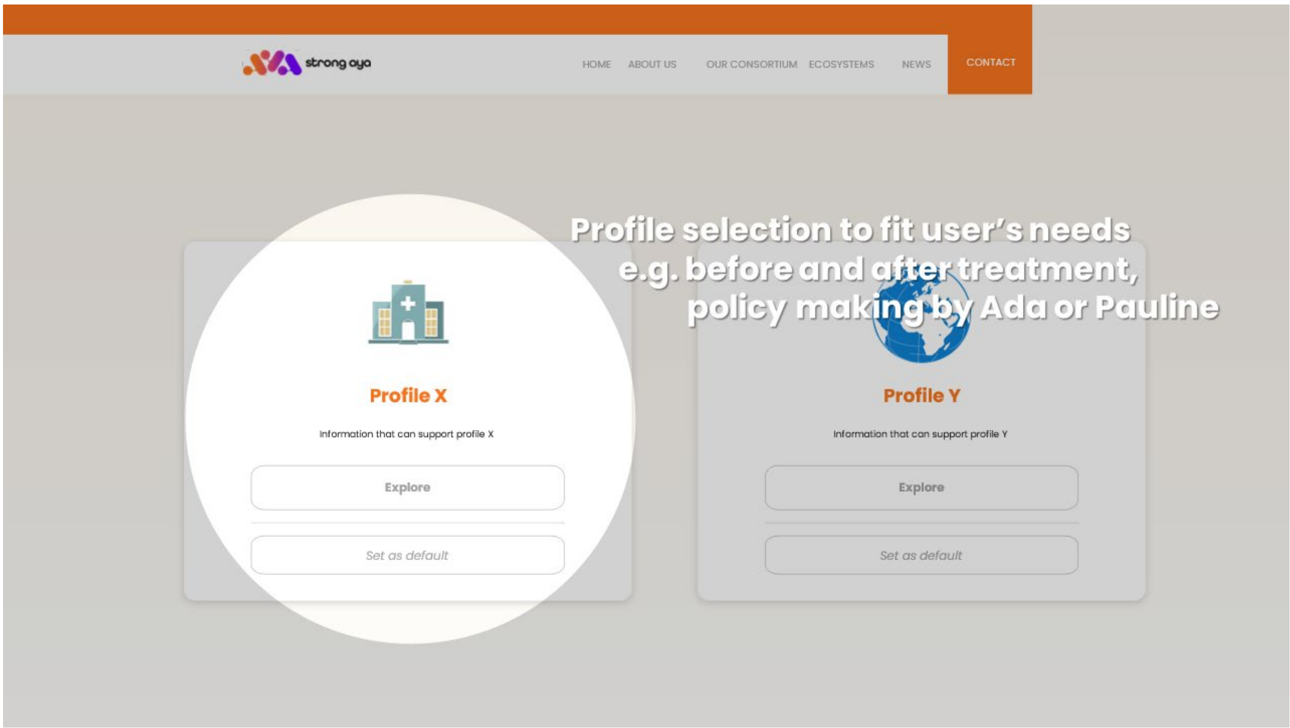


Fig.1 Illustrative concept of a standard HTML-enabled web browser landing page. The example shown here is for the restricted area with special access to data analytics and data exploratory functions, which would not be available to the random public visitor.



Profile selection to fit user's needs e.g. before and after treatment, policy making by Ada or Pauline

Information that can support profile X

Information that can support profile Y

Fig.2 For registered users where we wish to protect specific functionality such as filtering/searching, interactive data exploration and running statistic modelling, we require an element to save or define specific user profiles with their associated permissions.

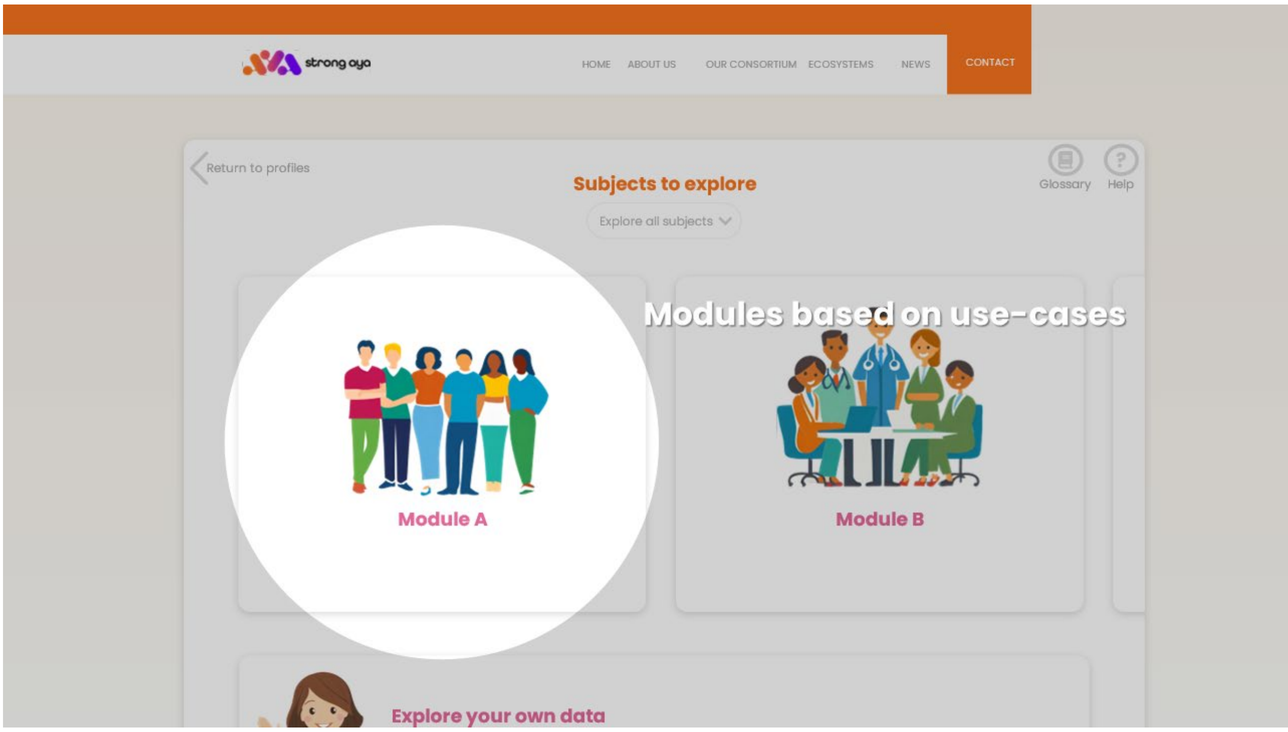


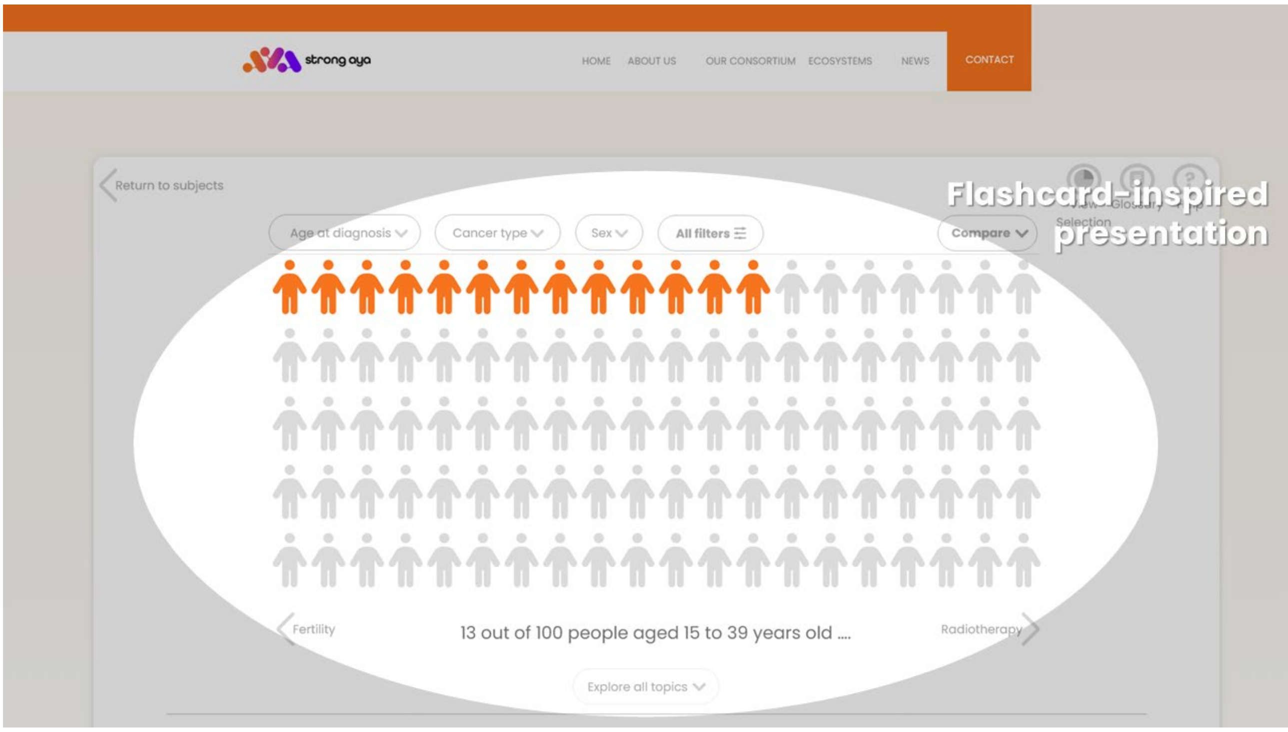
Fig. 3 Illustration of the concept of specific topics of interest to a user, which are hereby reserved as topical modules. The modules will need to be configured on use cases that are currently in discussion and/or emerging in the future of the project.

5. Within a given topic of interest, for example treatment options or quality of life outcomes, we were advised to adopt a "flashcard" infographic style of presentation. In keeping with well-established literature on communication of data and statistics to a non-specialized audience, it was agreed that percentages and ratios was of very limited practical help towards understanding.

Following best practices in the medical communication domain, and guidelines relating to the development of "decision aids" and similar communication/education tools, we selected to adopt a standard framework of being able to report an arbitrary statistical summary as "X out of 100 people …" where X is whichever statistic of interest for a given topic.

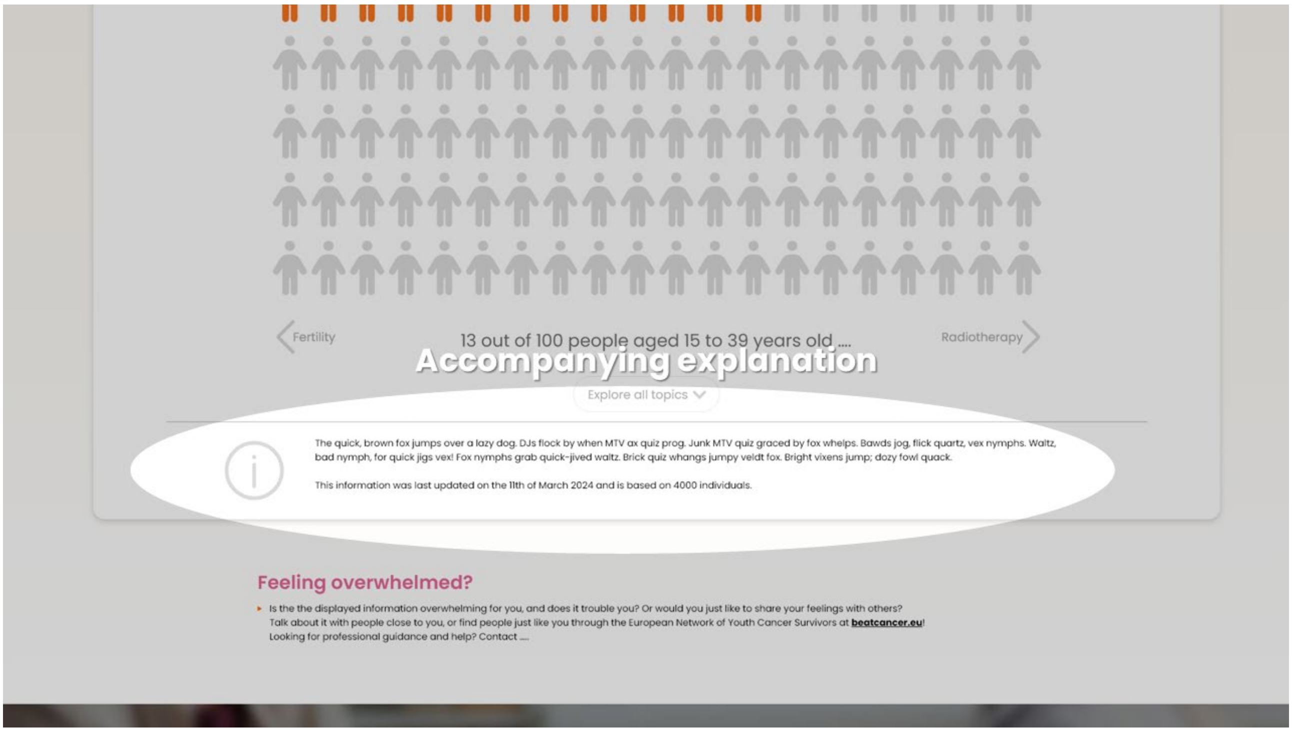
For a purely hypothetical instance, based on whatever statistical computation that is happening under the hood, we would then be able to report some raw numerical results of the background computation (e.g. 30%) as something like "30 out of 100 people aged between 15-39 years reported FATIGUE from 2 years up to 15 years after cancer treatment" (see **Fig. 4**). As can be seen in the aforementioned figure, it was advised to present statistical results in both visual graphical format using generic "human" figures in a large easily visible format (so as not to disadvantage persons with some sight impairments) as well as in large sized text font below the infographic (which can – in future iterations – also be connected to a digital text-to-speech tool to read out the information for the user).

Additionally, we were asked to try out a connotatively "neutral" colour theme with softer shades of saturation, avoiding certain colours that might be subconsciously misinterpreted (e.g. red = "bad", green = "good"). This was truly a subjective opinion, because colour schema that appeal to one person might not appeal to another. At this point of the process, we looked at the colours contained in the official STRONG AYA logo, and tried using some of those colours and shades appearing in the logo. On the other hand, superficial features such as the colour or font or font size of a core element (since this is a digital dissemination platform) would be very easy to re-configure at a later point of time.



13 out of 100 people aged 15 to 39 years old ....

Fig. 4 The infographic element of data visualization of proportions/ratio/frequencies, using large easily visible figures as well as the same information in large font text.



13 out of 100 people aged 15 to 39 years old ....

The quick, brown fox jumps over a lazy dog. DJs flock by when MTV ax quiz prog. Junk MTV quiz graced by fox whelps. Bawds jog, flick quartz, vex nymphs. Waltz, bad nymph, for quick jigs vex! Fox nymphs grab quick-jived waltz. Brick quiz whangs jumpy veldt fox. Bright vixens jump; dozy fowl quack.

This information was last updated on the 11th of March 2024 and is based on 4000 individuals.

Feeling overwhelmed?

- Is the the displayed information overwhelming for you, and does it trouble you? Or would you just like to share your feelings with others? Talk about it with people close to you, or find people just like you through the European Network of Youth Cancer Survivors at **beatcancer.eu**! Looking for professional guidance and help? Contact ....

Fig. 5 Accompanying information element describing how reliable or how qualified the statistical analysis might be. This may also include information about the data sources, or if the analysis has included some amount of de-identified real subjects and/or synthetically-generated data (for example, using artificial intelligence tools).

6. A further necessity is a supplementary information element to be attached to the infographic element, relating to how qualified, or how reliable, or how representative the data might be that was used to generate the aforementioned infographic. Feedback from HCPs, ethicists and patients also suggested that we should

be transparently open about what kind of data sources, or (as we may anticipate in future) whether some form of machine learning was used to generate or process the data.

7. The following element here is discussed but does not appear obviously to the user. This is in keeping with the request that the technology needs to work under the covers, but the user should not be expected to have specialized programming skills or technical skills. To this end, we have implemented the following functionality framework underneath the visible layers of the aforementioned elements.

First, when required, the use case selected by the user triggers a federated analysis through the Vantage6 software implementation that powers the technology architecture. This means that a pre-defined and pre-approved query is distributed to all contributing data institutions. The federated analysis runs across these participating data pools, and only the combined statistical result is transmitted back in order to be displayed in the infographic. In this manner, the statistics could potentially be updated live and in near real-time, however depending on the complexity or size of the result to be computed, will have an implicit latency of a few seconds in order to come up in the display.

Secondly, as defined in great detail in the technical architectural blueprint and DPIA, unknown and uncertified statistical algorithms will be blocked by the infrastructure. Allowed statistical analyses are certified and validated in a private statistical algorithm repository owned by the consortium, and only these analyses can be triggered by the visualization front-ends.

13 out of 100 people aged 15 to 39 years old ....

The quick, brown fox jumps over a lazy dog. DJs flock by when MTV ax quiz prog. Junk MTV quiz graced by fox whelps. Bawds jog, flick quartz, vex nymphs. Waltz, bad nymph, for quick jigs vex! Fox nymphs grab quick-jived waltz. Brick quiz whangs jumpy veldt fox. Bright vixens jump; dozy fowl quack.

This information was last updated on the 11th of March 2024 and is based on 4000 individuals.

Feeling overwhelmed?

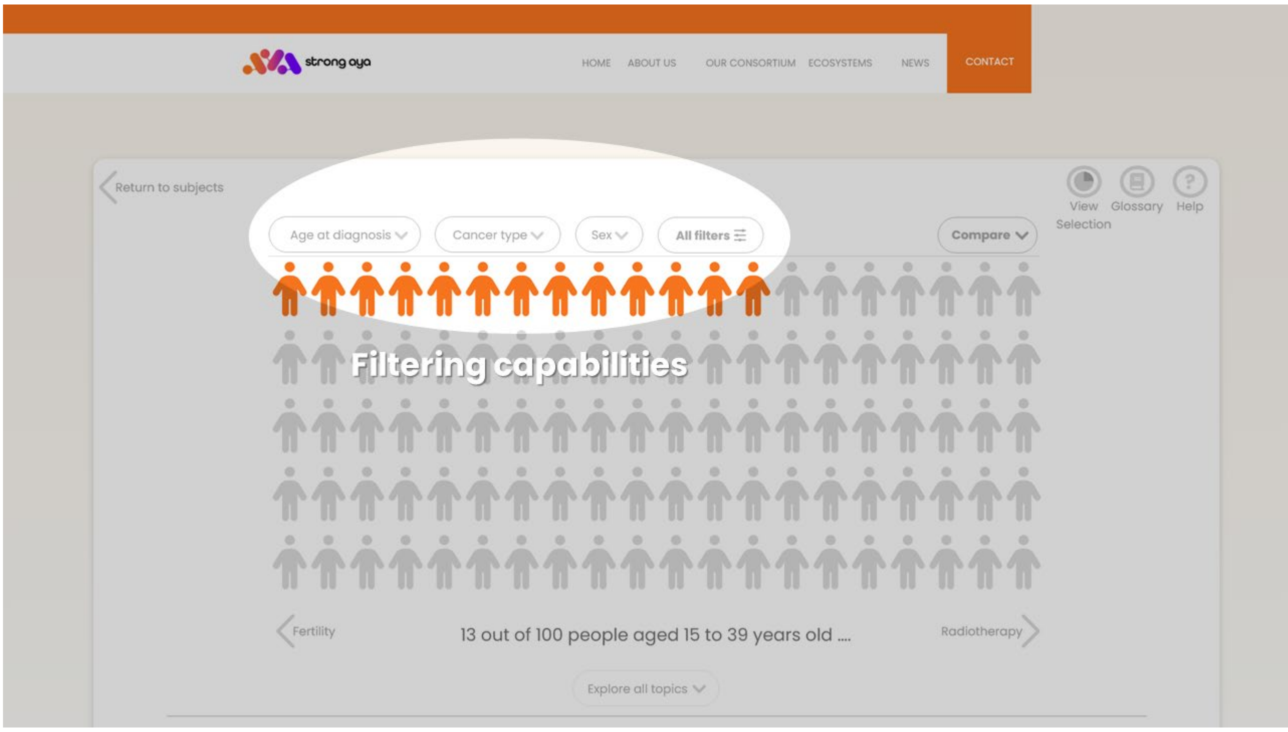
- Is the the displayed information overwhelming for you, and does it trouble you? Or would you just like to share your feelings with others? Talk about it with people close to you, or find people just like you through the European Network of Youth Cancer Survivors at **beatcancer.eu**! Looking for professional guidance and help? Contact .....

Fig. 6 References element, pointing to external resources of support, counselling, advice and clinical care if the user feels the need to look this kind of information up.

8. In keeping with the conceptual archetypes of potential users in the previous section, the expressed priority was for information to be correct, brief and accessible, but should not come across as pushy or intrusive or instructional. However, it was also felt that – in case the need for additional support might be triggered by reading about the presented information in the infographic, there had to be an element that allows us to direct the user – if they so wish – to seek out counselling, advice or other form of care (see **Fig. 6**).

This re-direction to additional external resources will be supported in the front end of our data visualizations, but only high quality information chosen specifically by HCPs and AYA experts will be incorporated here as additional sources of support (or linked information).

9. Layered-information presentation; across different user types, the guidance we received was that a pre-determined workflow or a forced sequence of steps might feel too controlling to the user. What we heard was, the data visualizations had to be digestible and accessible up-front at the top (general level) of the statistics, but that sometimes too much of information at one time was counterproductive. This is particularly the case where the statistics can be confronting or upsetting or triggering. We took this to mean that we needed to provide tooling element(s) in the visualization front ends that would allow the user – strictly voluntarily – to drill down deeper into the dataset.



13 out of 100 people aged 15 to 39 years old ....

Fig. 7 Illustration of the filtering and data exploratory elements that would allow users – subject to their levels of permissions and access privileges, to drill down to a certain degree within the infographic flashcard.

For the general public-access user, we would keep this functionality relatively basic and rudimentary and constrained to only high-level summaries. There could (subject to consensus within the project consortium) be the ability to drill into broad categories of cancer types (see **Fig. 7**), or to compare subgroups (by age deciles, for instance, or by major groupings of countries, for another example) (see **Fig. 8**).

For the specific users with higher levels of permission and exploratory privileges, we would still allow filtering and searching according to potentially more complex sets of parameter-based criteria, but still subject to the overarching protection mechanisms within the federation architecture to prevent filtering statistics down to a subset of too few individuals (in which case this would not be sufficiently obscuring of individual personal data values).

In keeping with Item 8 mentioned above, the technological design of the federated learning architecture is that user inputs may be able to trigger new federated analysis runs (which will happen under the covers), that will combine the partial statistics from multiple dispersed datasets and bring these together as the global aggregated statistic to be shown in the infographic. We anticipate that this will add some time latency of some seconds before the interactive dashboard display can update, but we also feel that the layered-information presentation and the drill-down functionality is worth the effort for the sake of the users.

Fig 8. Illustration of the subgroup comparison functionality that would allow users – subject to their levels of permissions and access privileges, to compare subgroups to a certain level of data granularity within the infographic flashcard. Not illustrated yet here, but the intention is then to show two infographic elements side by side, so that a visual and textual comparison of the two groups can be accessed on one flashcard style of presentation.

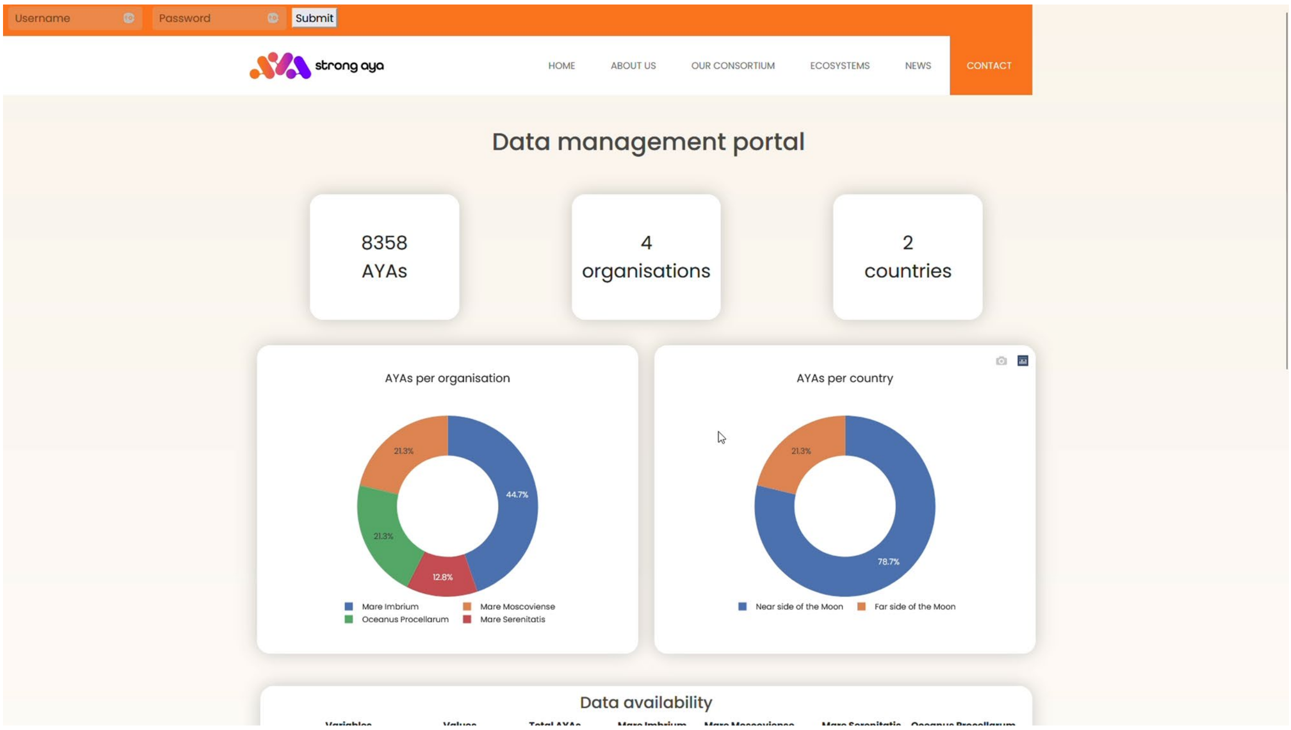
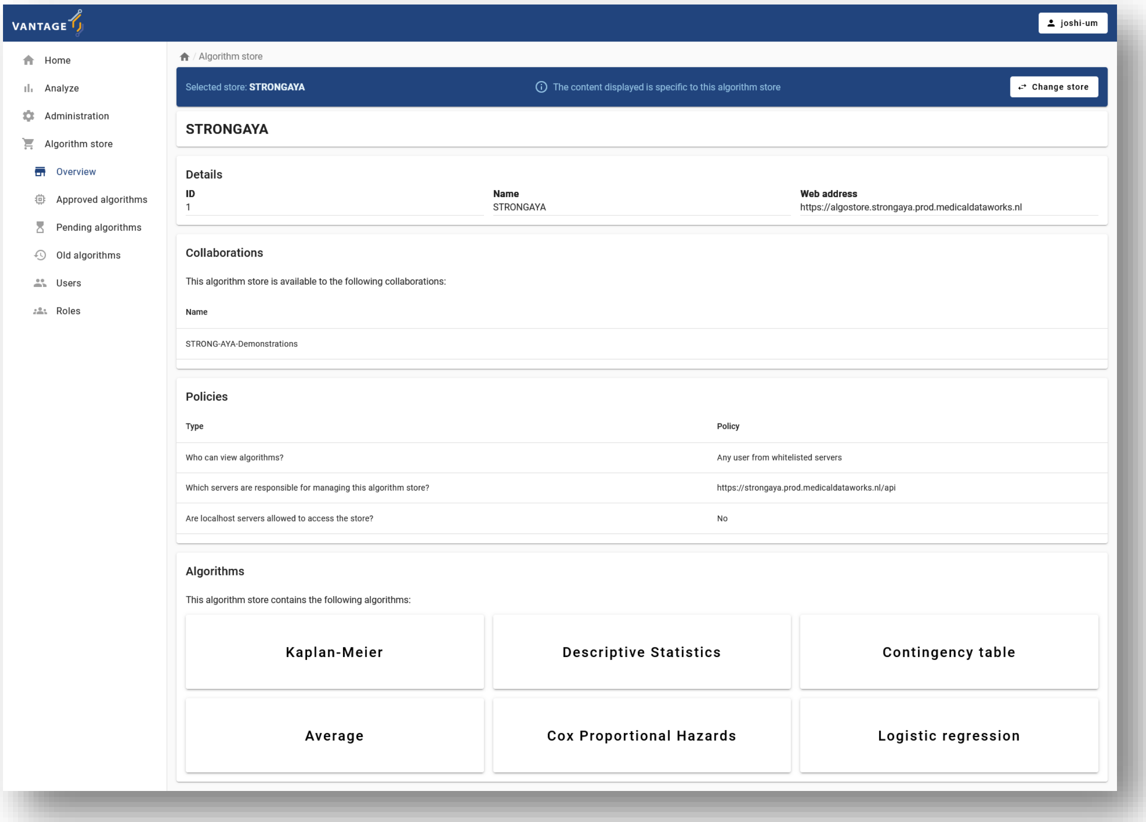


Fig 9. Conceptualization of a separate research-leader oriented view of the STRONG AYA consortium data across multiple participating institutions. The intention of the specially privileged view – which will not be widely accessible to external stakeholders – will allow leaders in the consortium to get an overview of data quality and explore subsets of data without betraying an individual's actual data values.

10. Data managers and research leaders portal view of consortium data; at the highest and most privileged level of functionality of exploring the datasets collected across the whole consortium, we have prepared a concept for a data management dashboard that would provide granular statistics of the consortium collection, but without exposing individual subjects' data values (see **Fig. 9**). This was inspired by work done internally by consortium partner CLB (see Appendix B), however we had to adapt it so that individual subjects' data values would remain obscured by the federation analytic infrastructure. For this sophisticated use of the platform, it is no longer possible to retain the infographic visualization style, therefore we revert to more researcher-oriented view based on traditional data visualization dashboards (such as the examples shown in Appendix A).

11. Lastly, for specially designated researchers within the consortium with specific permission to run statistical modelling in a federated fashion across multi-institutional datasets, we have rapidly advanced on building a researcher-centred "query-experiment-report" portal functionality directly into the Vantage6 infrastructure (see **Fig. 10**) in close cooperation with consortium technology partner IKNL. In this conception, there will be a guided workflow to assist the researcher in terms of (a) defining the parameters (b) selection of a suitable statistical framework (such as generalized regression or Cox proportional hazards, in the example shown in the figure) and (c) retrieving the statistical modelling result (i.e. model coefficients). Subject to further refinement of the use cases in consultation with the research members of STRONG AYA, we intend to also support graphical presentation of the results in the researcher portal (in the example in Fig. 10, it may be a contingency table or a Kaplan-Meier curve).



Selected store: **STRONGAYA**

The content displayed is specific to this algorithm store

STRONGAYA

Details

ID	Name	Web address
1	STRONGAYA	https://algostore.strongaya.prod.medicaldataworks.nl

Collaborations

This algorithm store is available to the following collaborations:

Name
STRONG-AYA-Demonstrations

Policies

Type	Policy
Who can view algorithms?	Any user from whitelisted servers
Which servers are responsible for managing this algorithm store?	https://strongaya.prod.medicaldataworks.nl/api
Are localhost servers allowed to access the store?	No

Algorithms

This algorithm store contains the following algorithms:

Kaplan-Meier	Descriptive Statistics	Contingency table
Average	Cox Proportional Hazards	Logistic regression

Fig 10. Preliminary prototype of the portal for federated statistical modelling to be made available to STRONG AYA members with suitable access permissions and profiles. A guided workflow is envisaged, and we have illustrated here an integration with certified and validated statistical packages (Cox, logistic regression, etc.) that is available to the researcher.

8 Additional elements

The following functional elements were included for completeness, as they provide supplementary functionalities requested by our focus groups which enhance the usability of the visualization portals, but do not affect the core functionality of the federated STRONG AYA data ecosystem.

Fig. 11 below illustrates, for the less digitally literate user, it would be helpful to have some accessibility-enhancing guidance about how to proceed with using the available functions.

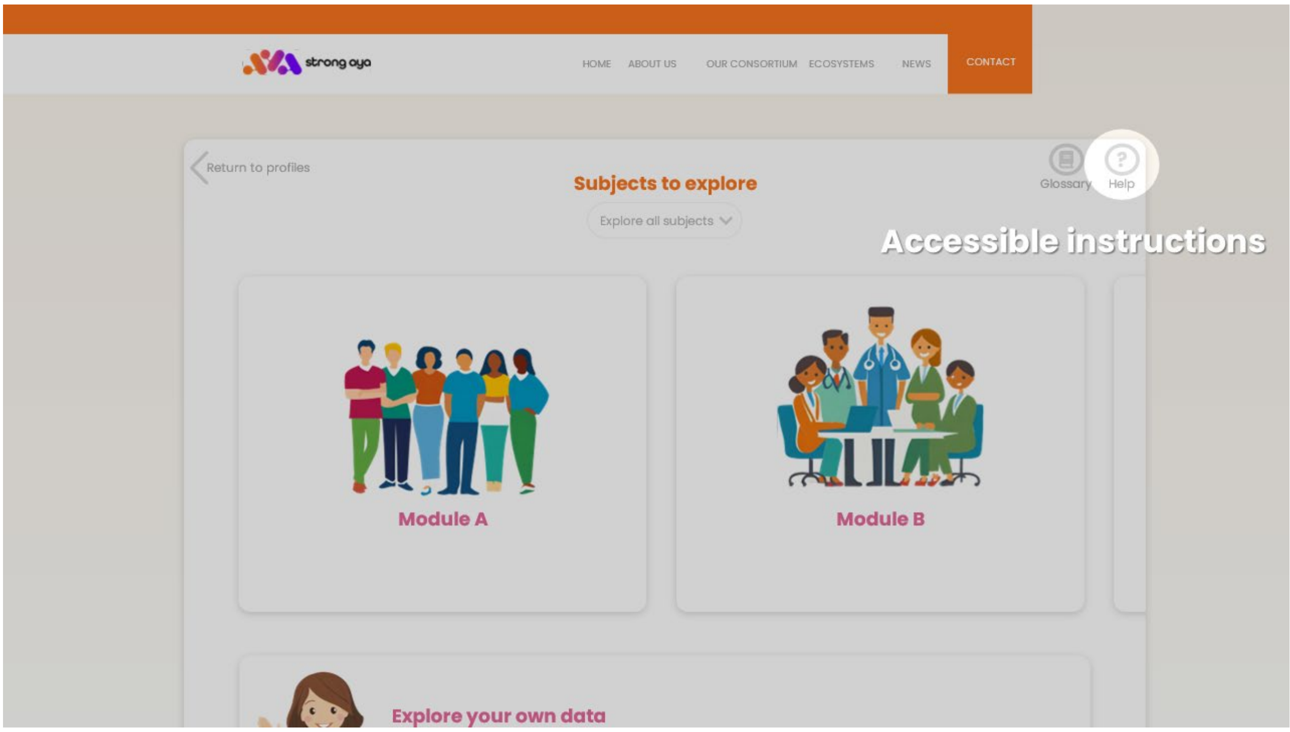
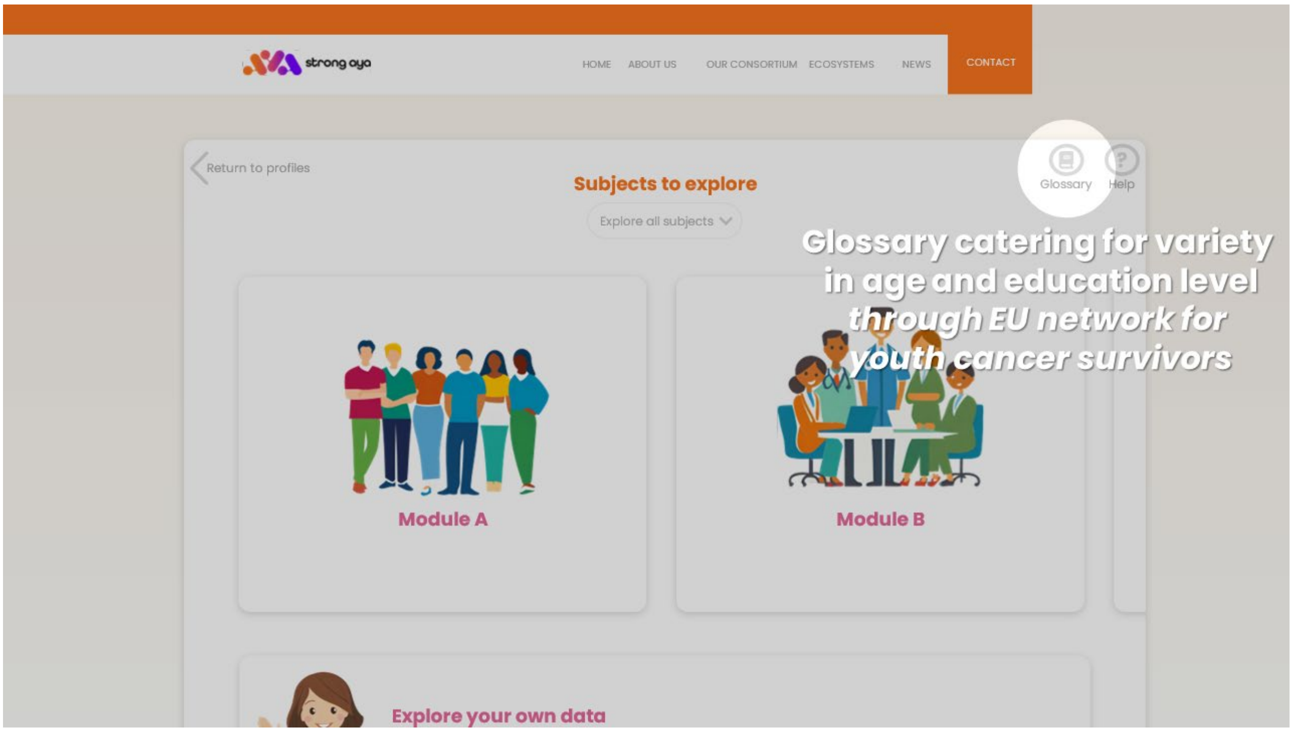


Fig. 11 – instructions for guiding the lesser digitally-literate user, about how to go about making use of the data visualization.

Similarly, for the less medically literate user, **Fig. 12** illustrates the accessibility to a glossary of terms in language that is more accessible to a target group of AYAs with cancer. This is actually sourced from consortium partner YCE, that has invested greatly into making medical terms understandable and improving patient-doctor communications about their diagnosis.

Lastly, in response to both HCP potential users and AYA potential users, it was felt that a functionality to export a picture of a certain infographic flashcard would aid in future recall of a conversation or consultation, plus had the extra benefit of being able to show the same information to friends, family and partners. The HCP also expressed a desire to be able to export the statistics contained in the infographic, so that they could re-plot it at a later time in a different form (e.g. piecharts, bar plots, etc.) for dissemination and communication to their professional colleagues. This is illustrated below in **Fig. 13**.



Glossary catering for variety in age and education level through EU network for youth cancer survivors

Fig. 12 – availability of a simple to understand glossary of terms, to aid in understanding of users in the AYA target age groups, that are expected to have widely varying levels of age, education and health literacy.

Variation in view

13 out of 100 people aged 15 to 39 years old ....

Fig. 13 – addition of a further element to switch the view, either as a picture card or with actual numerical results which can be re-plotted later on.

9 Conclusion of deliverable report

This deliverable report captures the core and supplementary elements of how results of a statistical data analysis would be presented to a wide and highly heterogeneous set of potential users. For the vast majority, the simple and non-threatening flashcard-style of infographic presentation would be preferable. However, our consultations with the focus groups also capture the needs of more sophisticated users who would like to dive deeper into the analysis, with the possibility (within safe limits based on use case) to filter into specific cancer types or age deciles, for example. Lastly, the most highly sophisticated and complex use cases are like to come from research leaders with specific and unique portal views for data quality management, statistical modelling and granular data exploration.

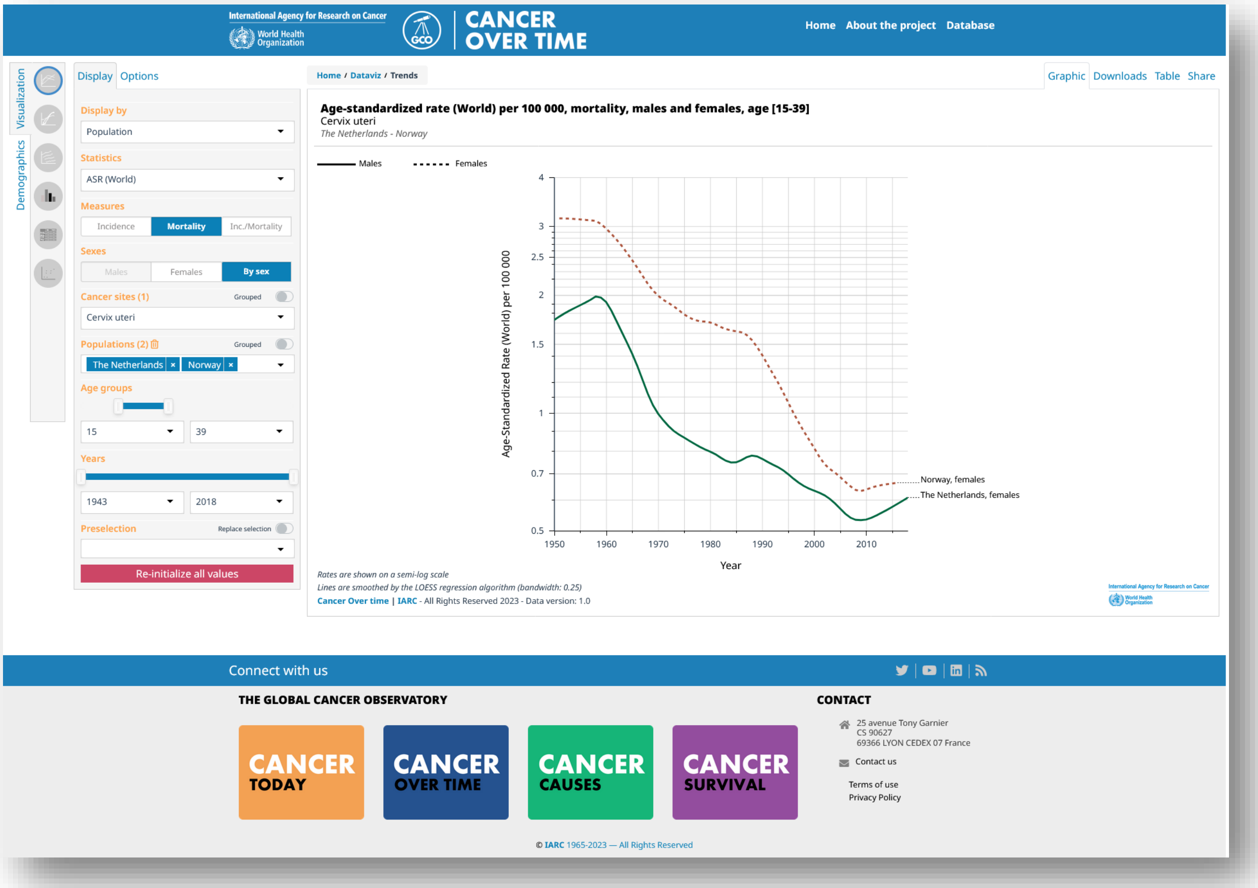
Given the vast diversity of use cases expected, we anticipate that there will be no single one-size-fits-all approach. Therefore, this deliverable report defines the elements that can be brought together – for users with different needs, levels of privileged use of data and levels of complexity required.

It needs to be recognized by the reader of the report that the elements described here needs to be modular, agile and adaptable – as the project matures and use cases evolves together with the consortium. However we believe the elements we have expressed here will remain generally universal and highly re-usable for other work.

Lastly, being implemented as a fully digital platform, we anticipate that by getting the basic architecture and core functionalities correct, we can easily adapt the more superficial aspects of the aforementioned elements easily within software / web graphic design tools.

Appendix A : Examples of publicly openly accessible dashboards

Example of GLOBOCAN – the global cancer observatory that pools high level general cancer data from population registries worldwide :



Display by: Population
Statistics: ASR (World)
Measures: Mortality
Sexes: By sex
Cancer sites (1): Cervix uteri
Populations (2): The Netherlands, Norway
Age groups: 15 – 39
Years: 1943 – 2018
Preselection

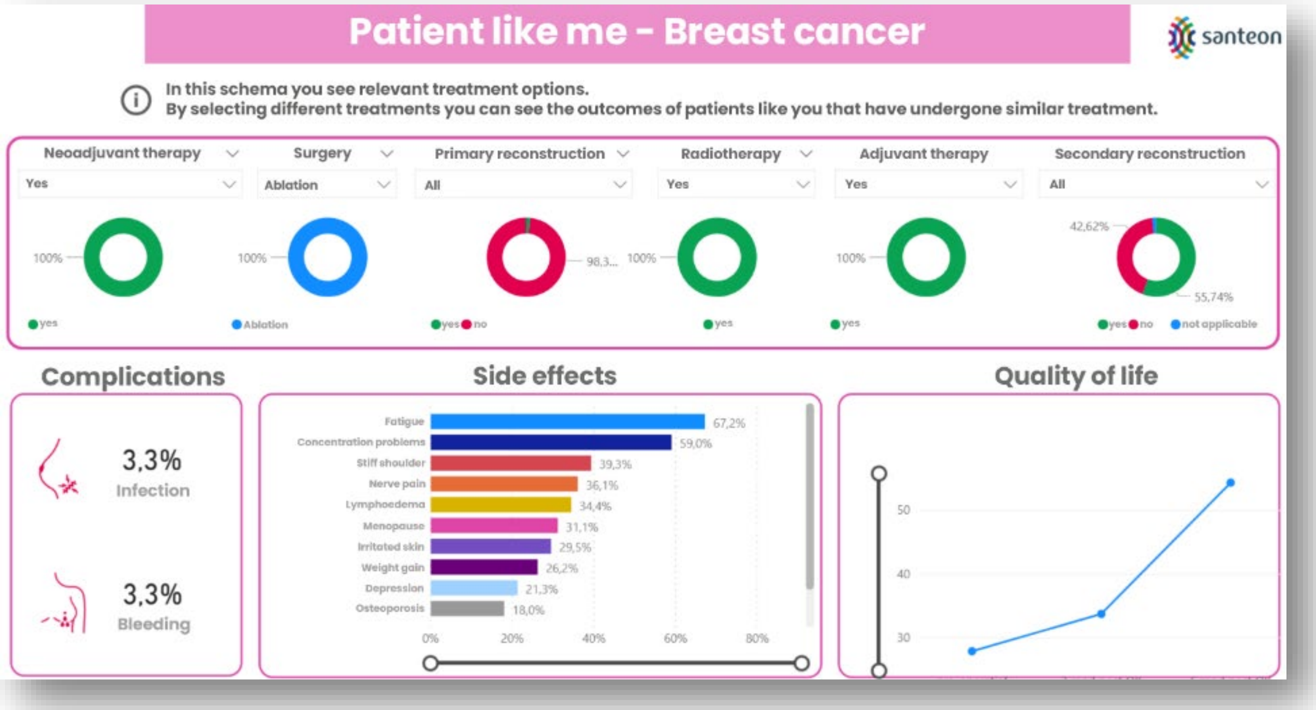
Age-standardized rate (World) per 100 000, mortality, males and females, age [15-39]
Cervix uteri
The Netherlands - Norway

Rates are shown on a semi-log scale
Lines are smoothed by the LOESS regression algorithm (bandwidth: 0.25)
Cancer Over time | IARC - All Rights Reserved 2023 - Data version: 1.0

25 avenue Tony Garnier
CS 90627
69366 LYON CEDEX 07 France

© IARC 1965-2023 — All Rights Reserved

Example of Patient Like Me – showing a representative of breast cancer treatment options, side effects and quality of life trends, but not specifically addressed towards AYAs :



Patient like me - Breast cancer

In this schema you see relevant treatment options.
By selecting different treatments you can see the outcomes of patients like you that have undergone similar treatment.

Neoadjuvant therapy: Yes

Surgery: Ablation

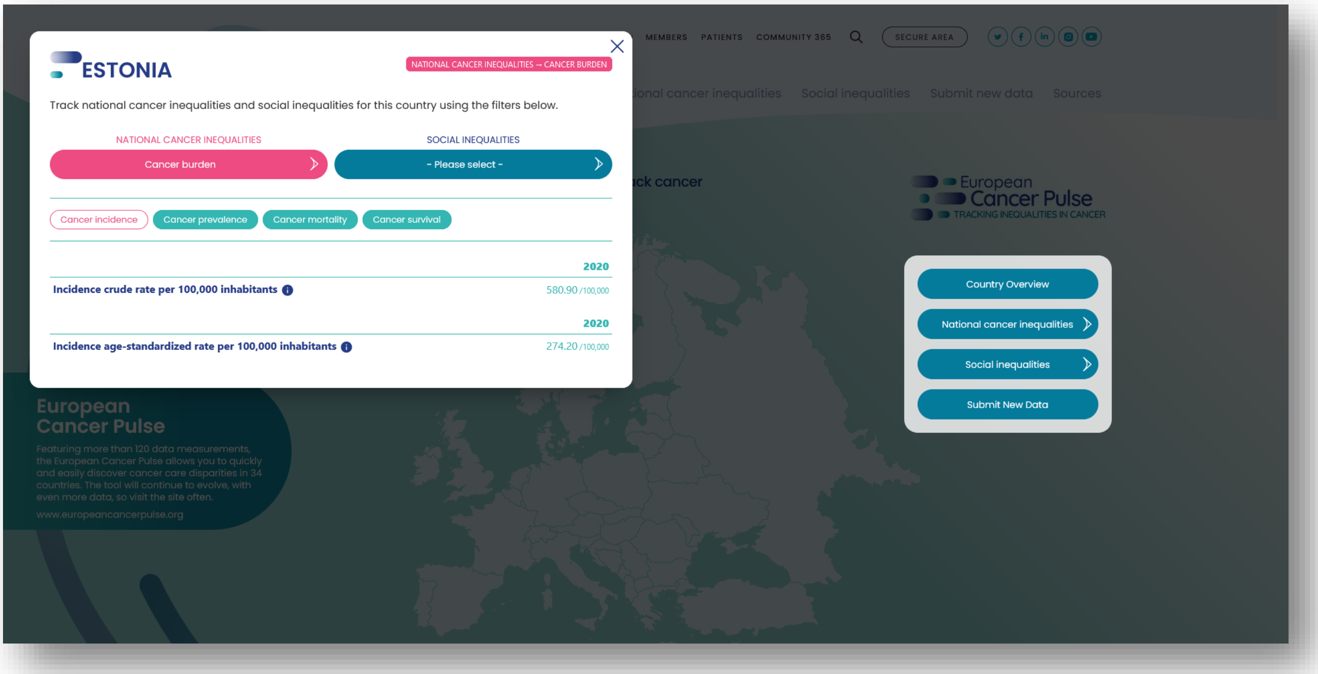
Primary reconstruction: All

Radiotherapy: Yes

Adjuvant therapy: Yes

Secondary reconstruction: All

Example of European Cancer Pulse – provided by the European Cancer Organization, showing top level statistics of care at country level, but the information can be many years old and not able to be refreshed at a rapid pace.



ESTONIA

NATIONAL CANCER INEQUALITIES → CANCER BURDEN

Track national cancer inequalities and social inequalities for this country using the filters below.

NATIONAL CANCER INEQUALITIES: Cancer burden

SOCIAL INEQUALITIES: - Please select -

	2020
Incidence crude rate per 100,000 inhabitants	580.90 /100,000
Incidence age-standardized rate per 100,000 inhabitants	274.20 /100,000

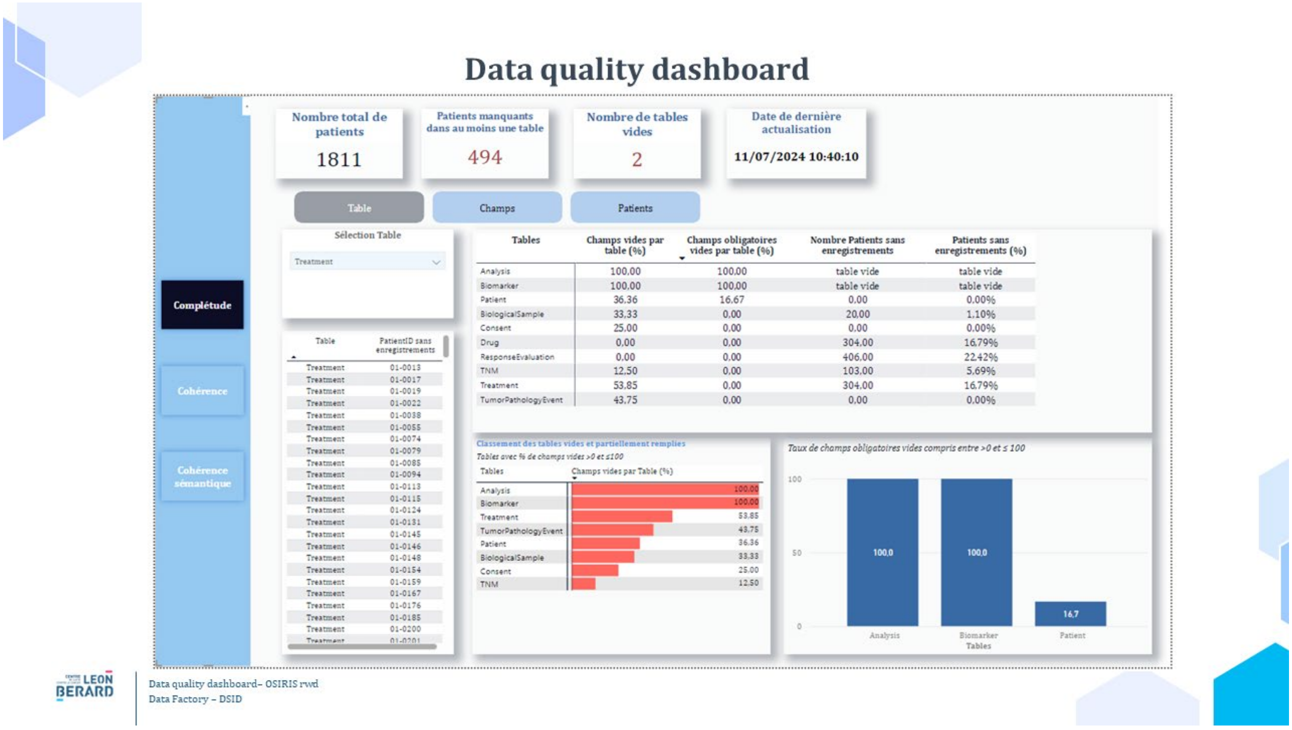
European Cancer Pulse

Featuring more than 120 data measurements, the European Cancer Pulse allows you to quickly and easily discover cancer care disparities in 34 countries. The tool will continue to evolve, with even more data, so visit the site often.

www.europeancancerpulse.org

Appendix B : Example intra-institutional data dashboard

Example of consortium partner CLB – internally facing data dashboard for quality control and data exploration :



Data quality dashboard

Nombre total de patients: 1811

Patients manquants dans au moins une table: 494

Nombre de tables vides: 2

Date de dernière actualisation: 11/07/2024 10:40:10

Sélection Table: Treatment

Table	PatientID sans enregistrements
Treatment	01-0013
Treatment	01-0017
Treatment	01-0019
Treatment	01-0022
Treatment	01-0038
Treatment	01-0055
Treatment	01-0074
Treatment	01-0079
Treatment	01-0085
Treatment	01-0094
Treatment	01-0113
Treatment	01-0115
Treatment	01-0124
Treatment	01-0131
Treatment	01-0145
Treatment	01-0146
Treatment	01-0148
Treatment	01-0154
Treatment	01-0159
Treatment	01-0167
Treatment	01-0176
Treatment	01-0185
Treatment	01-0200
Treatment	01-0201

Tables	Champs vides par table (%)	Champs obligatoires vides par table (%)	Nombre Patients sans enregistrements	Patients sans enregistrements (%)
Analysis	100,00	100,00	table vide	table vide
Biomarker	100,00	100,00	table vide	table vide
Patient	36,36	16,67	0,00	0,00%
BiologicalSample	33,33	0,00	20,00	1,10%
Consent	25,00	0,00	0,00	0,00%
Drug	0,00	0,00	304,00	16,79%
ResponseEvaluation	0,00	0,00	406,00	22,42%
TNM	12,50	0,00	103,00	5,69%
Treatment	53,85	0,00	304,00	16,79%
TumorPathologyEvent	43,75	0,00	0,00	0,00%

Classement des tables vides et partiellement remplies

Tables avec % de champs vides >0 et ≤100

Tables	Champs vides par Table (%)
Analysis	100,00
Biomarker	100,00
Treatment	53,85
TumorPathologyEvent	43,75
Patient	36,36
BiologicalSample	33,33
Consent	25,00
TNM	12,50

LEON BERARD

Data quality dashboard– OSIRIS rwd
Data Factory – DSID