strong aya

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

Deliverable Report

WP5– Scientific coordination and project management

Deliverable D5.5

Publication on the blueprint of the national and the pan-European STRONG AYA ecosystems

Due date of deliverable: 30/09/2024

Actual submission date: 11/09/2025

Project:	STRONG-AYA
Lead Contributor	Dan Stark, Leonard Wee
Email	d.p.stark@leeds.ac.uk; leonard.wee@maastro.nl
Other Contributors	Oana Lindner, Simone Hanebaum, Nicola Hughes, Urska Kosir, Winette Van Der Graaf, Olga Husson
Emails	o.c.lindner@leeds.ac.uk; s.hanebaum@nki.nl; n.f.hughs@leeds.ac.uk; urska.kosir29@gmail.com; w.vd.graaf@nki.nl; o.husson@nki.nl

Due date	30 09 2024
Delivery date	11 09 2025
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V1.0	**11/09/2025**

Description:
Scientific paper

Publishable summary (max ½ page)

The STRONG-AYA consortium is addressing the challenge of using real world data to understand the clinical and patient-centred outcomes that are specific to cancer diagnosed aged 15-39. We will use this data to improve care, policy, and ultimately the outcomes themselves. A harmonised 'Core Outcome Set' is agreed, to define what should be measured. We are building a series of AYA cancer data ecosystems, at local, national and international levels, that will use federated learning methods to provide pre-specified analyses in real time, based upon fundamental ethical, scientific and technological principles. Here we describe the problem being addressed. Then, for others interested in AYA cancer to understand and join our initiative, we describe:

- The definitions and decisions we have made.
- The structure and components of our ecosystem
- The context of our ecosystem alongside similar others, and our approaches to challenges we face

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

4 Deliverable Introduction/context

4.1 Deliverable introduction

The following deliverable is the manuscript for the ecosystem blueprint for STRONG AYA. It will be submitted for publication in a shorter length but we wished to preserve its detail and full length via this deliverable.

5 Introduction

International healthcare systems aim to provide high-quality care and improve patient outcomes, ranging from clinical parameters to symptoms and optimized quality of life. Achieving these goals relies heavily on the accurate recording, monitoring, and evaluation of healthcare quality indicators (such as clinical effectiveness, patient safety, patient-centred and patient-reported outcomes - 'PROs'). This data increasingly arises from real-world data (RWD) collection and is then utilised within 'learning healthcare systems' as the evidence base for care transformation.

Cancer in Adolescents and Young Adults (AYAs) are either rare cancer types, such as sarcomas, lymphomas, and germ cell malignancies, or uncommon presentations of common cancers e.g. early-onset carcinomas. Therefore, generating practice-changing evidence supported by high quality data is particularly challenging. This challenge is exacerbated because indicators of care quality and PROs have been (up to now) collected sporadically, or non-systematically, and the most important outcomes to measure are lacking a consensus. This makes it harder to prioritize and implement the right healthcare improvements for the AYA cancer population from a strong evidence base.

5.1 AYAs specific cancer care needs

In our scope, we selected the existing European gold standard, defining AYAs with cancer as people receiving a cancer diagnosis between 15 and 39 years of age inclusive. First (for the individuals with cancer) there are the challenges of AYA receiving optimal cancer diagnosis, treatment and care. Worldwide, an estimated 1.2 million AYAs [1] were diagnosed with cancer in 2020, representing around 5% of all cancer diagnoses [2]. The incidence of AYA cancer is increasing over time [3]. AYA-specific diagnosis and treatment has recently improved, since this patient group became a focus of specific focus and research. All-cause and cancer-related mortality has fallen in high-income countries [4], so there over 85% of AYAs with cancer will survive ≥5 years [5,6] and some will live >50 years beyond their treatment[7].

If they receive the most effective cancer care (which should be theirs by right) then the individual AYAs with cancer join a rapidly growing cohort of cancer survivors who will spend the remainder of their lives with increased risks of physical, psychological, and social effects; cardiovascular disease, secondary malignancies, infertility, financial toxicity, and premature mortality [8-10]. These late effects are also long recognised as specific to the AYA age range [11].

Until recently, across many countries, healthcare professionals (HCP), policy-makers and other stakeholders possessed limited knowledge about this population's specific needs [12]. Sadly, cancer in AYAs is often diagnosed, treated and cared for by either paediatric (<15-18 years) or adult (>16) care specialists, when in fact AYAs are liminal for both those services (caring mainly for patients aged under 12 or over 70 respectively) and are a distinct group, with a specific set of priority challenges in clinical practice, requiring a specific focus to address [1, 2, 13-15]. AYAs' challenges pertain to their:

(1) Unique epidemiology - AYAs develop both paediatric- and adult-type tumours, requiring expertise from both clinical services [16].
(2) Distinct tumour biology - cancers in AYAs often exhibit age-specific therapeutically significant molecular features and pharmacological responses, compared to younger children or older adults with superficially similar situations[16].
(3) Delayed diagnosis - a lack of awareness about cancer in AYAs and silos in services contribute to diagnostic delays and therefore poorer outcomes [17,18].
(4) Limited access to clinical trials - AYAs participate in clinical trials at lower rates due to unavailability of tailored treatments and trial designs that do not meet their specific needs, resulting in slow improvement in outcomes [19].
(5) Personal Development - AYAs are undergoing life transitions (e.g. identity formation, autonomy, and career development) which intersect with their cancer, making both healthy development and best cancer treatment more difficult to achieve [13,20].
(6) Survival rate disparities - survival improvements for AYAs lagged behind those seen in other age groups due to the factors above, unless systemic gaps in care and research are addressed [21].

5.2 The requirement for AYA-specific cancer outcomes data

In services where they are liminal, AYAs with cancer struggle to find care that manages their cancer in keeping with their age [22]. Their unique support needs are often not adequately met by centralised paediatric services (and "family-centred" teams) nor dispersed adult oncology services (and "disease-centred" teams) [14,23,24]. Recently, AYA cancer services in some parts of the world have been accepted as a specific distinct part of a complete cancer service, and provided with AYA-specific cancer care[14]. Maintaining this recent momentum requires ongoing research improvements, to define the data required to understand this progress, and to continue to improve outcomes.

In 2016, the National Cancer Institute (NCI) AYA cancer progress review group report requested AYA caregivers to pool process-related and outcomes data across institutions and countries, to make large cohorts available for research [12,22]. Extending these recommendations, the AYA Working Group of the European Society for Medical Oncology (ESMO) and the European Society for Paediatric Oncology (SIOP Europe) concluded that finding rapid solutions to 'speak the same language' among AYA cancer professionals is essential to further improve outcomes for AYAs. Meanwhile widespread geographic variation in AYA cancer policy and care persist within and between nations, and equitable services and optimal care for AYAs have not yet been achieved [14].

Collection of all cancer outcomes data is currently not standardized, and is implemented unevenly across health services; some European health services routinely record and monitor clinical outcomes as well as PROs, but many do not. Commonly-used clinical trial endpoints (such as five-year overall survival) are important, but these do not reflect AYAs' specific needs, and neither do outcome measures designed for young children, older adults, or any site-specific cancer type. Where data is collected (with more and better data to be collected in future) the task of "pooling" highly granular patient data into large repositories runs into challenges due to differences in research governance structures, rules for data re-use, nuances and interpretations of patient privacy protection laws. Therefore, novel approaches to perform large-scale statistical analyses across multiple dispersed locations are needed, which can act in a complementary fashion to traditional data centralization into singular repositories.

5.3 The STRONG AYA initiative

The STRONG AYA initiative was created in 2022, to serve the holistic health needs of AYAs with cancer; a marginalized and under-served patient population. Therefore, the STRONG AYA consortium directly addresses gaps in AYA cancer care by aligning research methods, clinical research questions, and standardized outcome measurements (including AYA-specific PROs). This will enable healthcare systems to generate real world evidence of AYA patients' outcomes [25].

The STRONG-AYA Consortium is made up of leading patient advocacy organisations, healthcare providers and clinical academic research centres across Europe, who already lead existing research and clinical projects in AYA cancer. The consortium also builds upon successful smaller-scale initiatives in various regions, but adds wider engagement, data standardization, technical innovation, large cohort studies, and cross-border multi-institutional collaboration [14]. The results of the consortium's work can therefore inform clinical decision-making, service design and policy so that AYAs in all regions will be able to access specialized, optimized and equitable multi-disciplinary care in order to survive and thrive.

In terms of concrete objectives:

- The STRONG-AYA consortium has developed and disseminated a **standardized Core Outcome measurement Set**, ('COS'), to specify, measure and prioritize outcomes of specific relevance to AYA through clinical data and PROs, given specific enhanced AYA relevance by the structured consensus of AYA patients, clinicians, researchers and policy makers. We will now generate, collect and use this data in clinical practice.
- Second, we have set up a **federated data ecosystem** of dedicated people and their skills, clinical informatics, and data management resources, supported by **an ethical-legal governance framework**). Here we will collect, analyse and interpret the COS. STRONG-AYA chose to create five national data ecosystems that collect, collate and utilise clinical and patient-reported outcomes data in AYA cancer practice, and connect the 5 to create an international (Pan-European) data ecosystem.
- Third, we have implemented a **secure federated data analytics infrastructure,** i.e. a 'Personal Health Train ('PHT') that enables STRONG-AYA partners to perform large-scale data analyses over long distances without sharing any individual subject-level data, thus addressing some challenges associated with traditional data sharing. This has the benefit of preserving privacy and confidentiality, while making research results available in timely and transparent fashion. These analyses will demonstrate quantitatively how AYA-specific cancer outcomes vary and change over time in various demographic groups and clinical settings, to support sustainable, cost-effective, and value-driven cancer care, specific for AYAs.

This paper details the specific aims and objectives of the STRONG-AYA initiative, the principles and practices which underpin the local, national, and international data ecosystems, their development and implementation by the STRONG-AYA Consortium and show how a unified approach to data collection and analysis would leads to the envisioned transformation of AYA cancer care.

5.4 Definition of a data ecosystem in cancer care

Generally, a data ecosystem in cancer care consists of:

- **Communities of people**, which includes anyone involved in the provision and receipt of medical health care (healthcare providers, researchers, policy makers, lived experience co-researchers, patients, caregivers, etc.) who collaborate on a research topic. In STRONG-AYA, AYAs with cancer themselves co-design our work, to then be able to use and benefit, directly and as a group, from it.

- **Securely protected and purposefully defined personal data,** underpinned by an ethical, secure, privacy- and confidentiality-preserving technology platform, for quality assurance, analysis and interpretation [26-28]. Data may pertain to clinically measured outcomes, patient-reported outcomes and/or general population data.
- **Relevant (mainly digital) tools and resources** through which stakeholders can co-operate, view appropriate information and personally relevant outputs from the research, interact with research findings and communities, and identify additional clinically-relevant questions.

For the development of a robust international data ecosystem, first local and national ecosystems need to exist, and then be connected. We chose this approach because those local and national ecosystems enable the local champions to address implementation challenges, participate in shared analysis of their data, promote reports and innovative data analytical approaches, and disseminate research findings and conclusions. Ecosystem members also act as local care transformation leaders for future growth of additional local and national STRONG-AYA partners, a development we in STRONG-AYA welcome and anticipate, to bring their own local or national ecosystem into this endeavour eventually. Below we describe the local, national, and international data ecosystem defined within STRONG-AYA in more detail (see Figure 1 below).

6 Data ecosystems within STRONG-AYA (table 1)

6.1 Local structures

A local data ecosystem has stakeholders within a city, small community or healthcare facility. It focuses on addressing the needs of local stakeholders, such as residents, service providers, or local governments. It works with municipal data, such as provided by a local healthcare provider. It seeks insights which address neighbourhood-level patterns (of disease or inequality, for example), and may collaborate with other municipalities or regions more than they do with nations. Local systems are narrow by design, but are often able to provide much more granular and contextualized data as a result. In STRONG-AYA we have examples of local ecosystems, in Poland where there is one participating hospital with their electronic patient data, and in Italy where there is one city's cancer dataset.

6.2 Partly national structures

Partial national structures are defined by their broader processes of data collection, data management, data sharing and data stewardship, that take place in a harmonized manner from at least two locations within a country. They generally focus on shared learning between those locations, with the aim of improving service provision across a (limited) number of institutions.

The Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP) is such an example. This collects and harmonizes data from hospitals within Yorkshire- a population of around 5.4 million within England- contributing research to public health policies in this region that would only be partially generalizable to other UK regions. Patients' clinical information is registered since 1990 (e.g., site, size, stage, time of diagnosis, treatment, recurrence(s), second malignancies, mental health, comorbidity) and will be linked to existing questionnaire data from research studies and PROs used in clinical practice since 1995.

In the UK, participants in STRONG-AYA are working within the consortium on their journey from existing strong local ecosystems to a national AYA-onset cancer data ecosystem. AYA cancer services across the north

of England are working together to align retrospective and prospective data. The detailed, longstanding and meticulously curated Yorkshire population-based AYA-onset cancer data, encompassing 3 comprehensive cancer centres and 10 local cancer treatment environments [29] is coming together with a large data-rich world-leading NHS specialist cancer treatment and research intensive NHS trust (Christie Hospital).

6.3 National structures

In the Netherlands, there is a nationwide AYA research infrastructure focusing on incidence, risk factors and mechanisms behind medical and psychosocial long-term and late effects. This dynamic, observational cohort of AYA cancer patients consists of: A longitudinal component: Including people recruited around cancer diagnoses and followed over time (original *COMPRAYA* cohort) [30]; A cross-sectional component: Including people >5 years after cancer diagnosis at time of study participation (original *SURVAYA* cohort) [31]. The NCR is used to identify and select AYA cancer patients according to the inclusion criteria for both COMPRAYA and SURVAYA. Patients' clinical information will be available from the population-based NCR (e.g., site, size, stage, date of diagnosis, detailed treatment characteristics, recurrence(s), second malignancies and comorbidity) will be linked to questionnaire data from both studies on long-term and late effects. Data from the infrastructure is freely available for other researchers.

The UK also has (since 2005 at least) its own national research and patient engagement infrastructure, several national NGO and advocacy organizations embedded with clinical and research groups, some national AYA cancer outcomes data, and an age-specific AYA cancer service policy for patients aged 15-24 inclusive. STRONG-AYA are seeking to enhance the UK contribution with existing population-based but more superficial national data collected as part of routine public health monitoring, where it pertains to care in the current UK partners in STRONG-AYA. Within other national-level UK research activity are; plans for several Living Lab projects for enhanced inter-disciplinary clinical, researcher and service-user engagement that extend advocacy beyond problem definition to problem-solving.

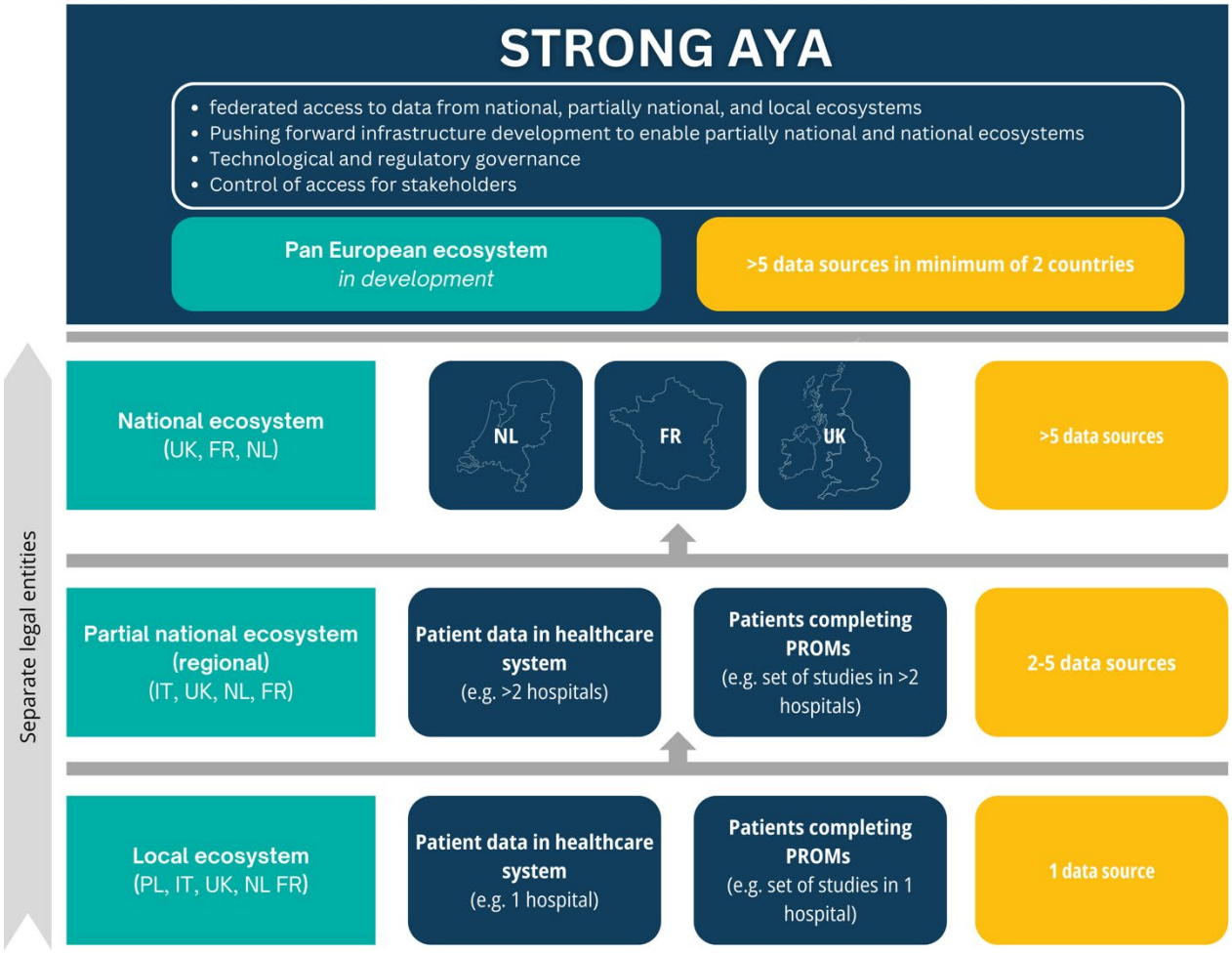
France has two hospitals serving their local city and beyond, providing regional data collection in STRONG-AYA currently, but working together to lay the groundwork for national reach. The data ecosystems at IGR and CLB will use CONSORE and RedCap data collection platforms respectively for PROs, which then allows patients to take an active role in their own care. Implementation of STRONG-AYA at IGR and CLB will demonstrate technical feasibility for data collection, preparation, federated learning, acceptability and impact.

6.4 International structures

The STRONG-AYA consortium is, in large part, already an international data ecosystem; its scope is multi-national, its leaders are active within global AYA-specific cancer epidemiology, health services research and policy, it collaborates with international NGO, professional and advocacy organisations, and it seeks to address international issues such as cross-border research using sensitive personal health data. The international ecosystem of STRONG-AYA will be built from the local and national ecosystems because, for the reasons described above, there is a rationale for strengthening the necessary local, part-national and national structures underlying the international collaboration. Rather, the desired path towards transformation is to align outcomes measures, data utilization and collaborative research from the existing structures. These ecosystems relate to each other, in our Strong-AYA data system (figure 1).

Table 1. Definitions of local, national and international data ecosystems

Aspects	Local data ecosystem	Partly national	National	International
Scope	City/Community-specific	Regional	Country-wide	Multi-country/Global
Stakeholders	Local agencies, institutions, residents	Universities, healthcare centre	National government, NGOs	Global institutions, multinational bodies
Data sources	Local data (e.g. municipal)	Regional registries, surveys	National registries, surveys	Cross-border research, global databases
Challenges addressed	Context-specific issues (e.g. neighbourhood-level health trends)	Regional issues (region or county-level public health)	National priorities (e.g. disease prevalence)	Global challenges (e.g. pandemics, climate change)
Collaboration	Local collaborations	Regional partnerships	Federal-state coordination	International coalitions
Number of data sources	One clinical or population-based data source	Two to 5 clinical or population-based data sources	More than 5 clinical or population data sources	National data sources in at least 2 countries
Ethical governance	Local governance within the specific hospital/clinical academic centre	Local and national governance (e.g. UK GDPR, HRA)	Local and national governance (e.g. UK GDPR, HRA)	Local, national, and international governance (e.g. UK GDPR, HRA, EU GDPR)
Primary purpose	Delivery of clinical care and monitoring of internal clinical processes	Structure of clinical care and monitoring between localities	Public health research for national policy	Public health research comparing nations
Example	Patient data on a specific hospital electronic patient record	Patient data in a specific registry (e.g. Yorkshire register) and general population data (e.g. UK Household survey)	Registers across multiple regions and/or several general population sources	The sum of registers and/or population-based data across multiple countries


STRONG AYA
federated access to data from national, partially national, and local ecosystems
Pushing forward infrastructure development to enable partially national and national ecosystems
Technological and regulatory governance
Control of access for stakeholders
Pan European ecosystem
in development
>5 data sources in minimum of 2 countries
Separate legal entities
National ecosystem
(UK, FR, NL)
NL
FR
UK
>5 data sources
Partial national ecosystem
(regional)
(IT, UK, NL, FR)
Patient data in healthcare system
(e.g. >2 hospitals)
Patients completing PROMs
(e.g. set of studies in >2 hospitals)
2-5 data sources
Local ecosystem
(PL, IT, UK, NL FR)
Patient data in healthcare system
(e.g. 1 hospital)
Patients completing PROMs
(e.g. set of studies in 1 hospital)
1 data source

7 Governance model, for the STRONG-AYA data ecosystem

The structure of a data ecosystem as well as its components are defined by a Governance Model. This model accounts for the expected and possible incentives and motives of the ecosystem partners, methods of access to the findings of analyses for current and future members, and procedures and processes for day-to-day operations, as well as compliance with national and international laws, governance and technical standards. It ensures alignment between these and the rights and responsibilities of stakeholders

A comprehensive governance model for STRONG AYA as a data ecosystem consists of:

1. **Its founding principles:** which should guide future decision-making and thresholds for defining criteria and rules (detailed in Section 3.1).
2. **Governance** of the data ecosystem, e.g. the legal and ethical frameworks for each partner institution for data collection (Section 3.2), informing participants and retaining flexibility including:
   a. **Sustainability** over time and the accession of new members;
   b. **Intellectual property sharing**; and,
   c. **Democratising of the benefits arising** from the research.
3. **The personnel/human resources needed:** with a structure for member rights and responsibilities, criteria for joining, leaving, maintaining an active presence despite change (Section 3.3).
4. **Data analytics and technical standards** (Section 3.4), such as data security, integrity, privacy and access, that keep pace with emerging threats and regulations. This includes minimum rules for access, data preparation and data standardization

7.1 The founding interpersonal and ethical principles of the STRONG-AYA data ecosystem

The STRONG-AYA Consortium is committed to transparency and specificity. There is an expectation that current and new members will follow and abide by the same founding principles. Furthermore, all consortium members are expected to be transparent in communicating their incentives and motives for the data ecosystem. We expect the incentives and motives of new members who become part of the ecosystem will align with (and refine further, by consensus) the principles of STRONG-AYA. This will help the functioning of the pan-European ecosystem by maintaining a current vision of the expectations and responsibilities of different users and levels.

Individuals within a data ecosystem, such as STRONG-AYA, may have one or several roles. A taxonomy for such roles has been proposed by Oliveira et al[36] Figure 2, below, describes those various roles, which may come into place initially in sequence, but then tend to be occupied by different individuals in a dynamic system over time.

Initial set up	**Keystone actor** •Founder	**Data owner**	**Data producer**	**Data provider** •Storer •Aggregator •Harmonizer •Publisher •Registrar (administrator) •Data maintainer

				•Data curator (quality and availability)

Figure 2: a broad taxonomy of potential roles within a data ecosystem, adapted from S. Oliveira, M.I., Barros Lima, et al., 2019.

The alignment of incentives and motives across these roles relies on a keen understanding of each individual and institution's capacity to achieve their goals through the joining of a pan-European data ecosystem. The ongoing incentives and motives for the existence of the STRONG-AYA Consortium and pan-European ecosystem are reflected in the principles outlined in more detail further, but are here summarized as:

1. Improvement in the provision of healthcare services:
    a. The development of value-based clinical practice;
    b. Accelerating the implementation of real-world insights to improve clinical practice.
2. Performing ethical research on large datasets:
    a. Encouraging international collaboration for research and clinical expertise;
    b. Improving and expanding upon the applications of available data in research;
    c. Understanding and leveraging institutional capacity.
3. Improving patient outcomes:
    a. Defining and identifying patient unmet needs – in research and clinical fields;
    b. Improving the diversity and reach of participants and beneficiaries from research and patient outcome measurement;
    c. Improving the utilisation of real-world health and healthcare data.

We believe it perfectly reasonable that different members may have a range of incentives for participating in the Consortium. Institutional incentives and motivations might also go beyond direct patient benefit, which naturally influences the purposes of data analysis that they conduct. These may encompass discovery (i.e. mining increasing volumes of data to help manage rare diseases), or accelerating implementation of existing initiatives (i.e. sharing administrative workload on diverse datasets), or encouraging sustainable international collaboration (i.e. tangible outcomes that foster collaborative practices).

Below we provide more detail on the foundational principles of STRONG-AYA, which drive the structure and functioning of the components as defined above.

7.1.1 Ethical, value-based healthcare

By broadening the research using novel privacy-safe approaches towards data analytics, STRONG-AYA will bring novel insights into AYA healthcare, encourage ethical re-use of scarce AYA research data, provide real-world evidence to all important AYA oncology decision-makers, and thus, increase the benefit for patients by making care more efficient.

7.1.2 Inclusive and focused on patient benefit

STRONG-AYA tools and activities are patient-centred, following the ethos of 'nothing about me without me'. These ethical principles, centred on the interests, preferences, and values of patients, are foundational to the Consortium, reiterated and reflected in the design, built and maintained through stakeholder

engagement and reflections by the Patient Advisory Board (PAB). The PAB is composed of young people with lived experience, composed based on Equity Diversity and Inclusion principles, actively seeking representation across various dimensions such as expertise, geography, and tumour type. The PAB provides ongoing support for the STRONG-AYA initiative by sharing knowledge, facilitating feedback with AYA cancer patient communities, and contributing to outreach, webinars, and other collaborative activities.

Importantly, any tools and actions within STRONG-AYA aim to complement and optimize face-to-face healthcare; patients are informed of the benefits and the limits of tools developed by the Consortium, and can customize their interactions with these tools. STRONG-AYA's resulting insights and tools need to be accessible, meaning patients:

- can easily and reliably access/retrieve their own STRONG-AYA health data (usually through their own hospitals' systems);
- see and better understand their own health data in the context of other, similar, AYAs;
- easily access information on the uses of their health data, and its' purpose, such that they should be able to easily enhance, grant or remove access to their health data depending upon their perspective on each 'case of use' of their data, and exercise this right freely;
- will have a voice in designing tools developed by STRONG-AYA, including promoting the principles of EDI such that the solutions remain accessible to all (including for example people with disabilities, neurodivergent populations, or different levels of digital literacy). They should be intuitive and easy to use; patients should have access to training to help them understand these tools, which should be implemented in routine care that includes human communication, space and time for additional questions and feedback.

Hence, STRONG-AYA will not only equip AYAs with lived experience with information to optimize their healthcare, but also offer an opportunity to build on their health literacy, and support shared and personalized decision-making between AYA and health care providers. This extends to involving patient groups in the conversations around the design, actions and operations within the ecosystem and the co-creation of the patient platform and methods of data visualizations, any design features, tutorials, and even appropriate warning or explanatory messages. 'Meaningful' patient participation creates actionable patient-centred and patient-informed insights. Co-creation with PAB and their outreach is expected to result, via local/national-level governance and ethical systems, in patients' views as guiding principles of the consortium.

7.1.3 Open and transparent science

STRONG-AYA abides by the principle that the research pursued and insights generated using the data (including publications, statistics, software, etc.) and its disseminations shall be accessible to all stakeholders and the general public. Therefore our methods should be transparent and accessible, and knowledge developed should be shared and actionable across the ecosystem. Stakeholders commit to all algorithms used for data analytics being openly accessible and constantly reviewable/critiqued. All publications related to STRONG-AYA will be open access under relevant local or EU funding. Any dissemination activities are openly available to members of the wider public as well as specialists, via a register maintained by the STRONG-AYA Project Coordinator.

While most insights and products will be accessible to members within the STRONG-AYA ecosystem, as well as the methods to produce these, the granularity to which the datasets can be accessed and what is queried

will depend on the type of stakeholder requesting access and their motives. Therefore, while all methods and disseminations are open access, the *individual level patient data will not be open access*, adhering to data protection, and privacy- and confidentiality-by-design principles.

For equity in data provision and data queries, and hence to reduce potential for misalignment resulting from power asymmetry, the STRONG-AYA data ecosystem will ensure:

- Data ownership and control will be held by local Principal Investigators and their teams.
- The actions herein are not a new topology of data control exercise by a single partner.
- Data analysis must still serve its named function if one partner conscionably opts out of a particular analysis or a specific use case (albeit in such cases on a likely reduced data sample size).

7.2 Governance in STRONG-AYA

STRONG-AYA partners believe that greater transparency about data ownership, collaborative analysis, and therefore clear data controllership promotes trust – among consortium partners and among stakeholders we disseminate results to. At local and national levels, STRONG-AYA governance is inclusive of local policy and legal agreements across the nations involved. Responsibility to adhere to these remains with the partners. Practically therefore, STRONG-AYA conforms to explicit levels of data anonymization; sufficient risk-guided population minimums must be met to permit analysis or re-analysis, to protect against unintended re-identification of individual data subjects (e.g. cohorts >10 as default, but may be adjusted with justification).

At an international and pan-European levels, STRONG-AYA uses a self-governance model which is open and inclusive across its operations. Specifically, our principles encourage a wide community of experts to join in the production of scientific insights, accepting transparency in data, inclusiveness, inter-operability, collaboration, longevity, and sustainability. STRONG-AYA's governance and legal principles align at the pan-European, local and national levels. It aligns with, among others, the European Ethical Principles for Digital Health (2022), national and international privacy laws - such as the General Data Protection Regulation (GDPR; https://gdpr-info.eu/), and applicable sub-legislations relating to lawfulness; fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation, integrity and confidentiality; and accountability.

Within this Governance Model STRONG-AYA acknowledges and welcomes the differences between institutions and countries in their capacity, capabilities, and limitations in what data can or cannot be used in our analyses. There is an expectation that not all countries and institutions will be able to fully meet all requirements all the time (e.g. technical, administrative, legislative, etc.) and there will be gaps in policies as well as pragmatics of data security, quality control, and management. The ethos of a collaborative pan-European ecosystem is that these gaps will impose less limits upon scientific learning within our data ecosystem. Therefore, a crucial task for STRONG-AYA to implement is the transparent communication and discussion of these potential differences and gaps, to allow room for their troubleshooting and secondary plans to achieve explicit goals.

In the case of prospective (novel) data collection (such as PROs based on the COS), each local partner needs to ensure that transparency is present in communication with participants. This means that all participants have provided and documented informed consent. In the case of retrospective (historical) data available and provided by partners, article 4 of GDPR applies, thereby the data collected by a partner must meet all three grounds allowing the processing of such data. Namely; the data can be processed if a) the participant offered

freely given, specific, and informed consent (Article 6 (1) (a)); b) processing is a legitimate interest exception such as there is a legitimate interest of controllers AND the legitimate interest cannot reasonably be achieved by using alternative means AND the interests and fundamental rights and freedom of the participant do not take precedence over the legitimate interests of the controller(s); c) the data is anonymized and therefore falls outside GDPR. As an example, a legitimate interest exception would be that of testing hypotheses for the purpose of public health.

The Consortium encompasses a Committee for Coordination (CfC), which was established by NKI (as the leading organisation) to oversee the pan-European ecosystem. The CfC has a critical role in ensuring our principles are adhered to. Comprising leaders of local ecosystems and patient advocates it defines the responsibilities of members, including the Patient Advisory Board, defines and implements criteria for newly joining members, encourages adherence to founding principles and communication among national ecosystems. It also ensures consistency, standardization and harmonization of data and promotes scalability to and within countries to guarantee the sustainability and longevity of the network of ecosystems.

The CfC as part of a defined organisational structure, has the responsibility to define and manage processes related to:

- existing member actions, responsibilities, and rights;
- new member access applications and rules of using of the data ecosystem;
- aligning the acceptable motives and incentives for accessing and using the ecosystem;
- defining and enforcing rules related to the submission of new use cases, utilising the existing ones and queries that can be run on the data that already exists in the ecosystem;
- ensuring all members actively contribute and abide to the required standards;
- resolving potential disputes and ambiguous situations;
- sharing success, e.g. intellectual property from the use of the data ecosystem.

As defined in our ecosystem goals (section 1.4), we have opted for a federated data ecosystem which preserves several ethical and governance objectives, without penalizing a relatively rapid pace of 'learning' (data analytics) that could be performed. It is important to clarify what is mean by 'federation' here.

7.2.1 Definition of federated learning

'Federated learning' refers to a group of computational methodologies where research questions or research problems ('use cases') that can be addressed mathematically and cooperatively, but without exchanging identifiable human subject-level data between any of the cooperating institutions (https://theodi.org/insights/explainers/what-is-federated-learning/). Our meaning of federation is NEITHER a cloud-based NOR a physically-located master data location, to which all partners must centralize their subject-level data to the 'owner' of that master repository on behalf of the consortium.

In contrast, our concept of federation embraces fully decentralized digital data repositories – two or more of such repositories, ideally located inside each care-giving institution's zone of control (where the individual-level healthcare data was collected in the first place e.g. a hospital). In such as topology, the descriptive statistics and epidemiological mathematical modelling must necessarily be performed on multiple local computing devices, co-located within each data repository, on the sample size (e.g. > 10 cases per variable) that is contained inside each institution.

Finally, the partial mathematical results (but most importantly NOT the subject-level data) from each institution are brought together on an encrypted telecommunications network, to generate the multi-institutional overall 'result'. For epidemiological models, this generally requires iterative and stepwise numerical integration of results, which converge on the mathematical 'answer'. Technical details on the implementation will be provided in Section 3.5 later.

The implementation of STRONG-AYA as a federated data ecosystem enables collaborative learning, and therefore aligns naturally with the following pan-European consortium aspirations:

- Data Autonomy – our federation approach allows local and national ecosystems to remain fully in control of their data at all times, and no one outside of the institution can see their respective individual-level data, while at the same time cooperative overall insights and collective answering of research questions remains fully supported.
- Data Solidarity – our AYA subjects altruistically gave consent to use their personal data to improve healthcare, but they do not want any negative repercussions to trace back to them leading to embarrassment, stigmatization, surveillance by authorities, or violation of privacy. By not exchanging any patient-level data, only statistical computations on multiple cohorts, this makes it extremely difficult for a malicious actor – either inside the consortium or outside – to re-identify any data subject even if there were successfully able to break the industry-grade telecommunications encryption.
- Open Science – since decentralized computations are not physically performed by human eyes and hands, it is necessary to protocolize and pre-specify the statistical analyses *a priori*, which includes the selection of cohorts and the endpoints before analyses. Every data analysis needs to be explicitly written co-operatively among partners, as computer code which will be open access, permanently auditable and necessarily reproducible on public software repositories (such as GitHub) even though the patient data remains secret.
- Diversity and Inclusivity – clinical and health care research benefit from learning as much as possible from the heterogeneity that is present in the real world. It is only by comparing and understanding differences in outcomes that the STRONG-AYA ecosystem and community improve the understanding the determinants of optimal value-based AYA cancer care. The ability to learn from sensitive personal data, without betraying the data subjects who have volunteered their data, helps us to address impacts of sample size, diversity and inclusivity when dealing with liminal health research subjects such as AYAs with cancer.

In STRONG-AYA, no external persons are given any access to information about AYAs within local institutional repositories. Since patient data stays locally where they are initially generated, this facilitates opt-out for some types of analyses and enhanced opt-in for certain other analyses, provided that patient engagement and the COS-guided data collection can support this level of detailed ongoing engagement, without requiring modification of a centralised database.

7.2.2 Legal governance supporting federated data ecosystems

A federated data ecosystem therefore requires, at minimum, a number of explicit agreements that establish: 1) the community of institutions and acknowledgement of their supplementary governance procedures – this is commonly known as a 'Collaborative Research' or equivalently 'Consortium' agreement. 2) Data protection impact assessments (DPIAs) that are resonant with interweaving of the local, national and international structural requirements. 3) A data access framework agreement that defines the principles of data controllership and permissible purposes for cooperative data processing in context of a federated

telecommunications system. 4) Infrastructure service agreement(s) with disinterested technology service provider(s) that are able to host the technical and computational components, but do not have any say whatsoever about research questions, or patient data or knowledge that arise due to consortium activities.

An example of agreement templates from previous federated learning projects, that have been used to develop the STRONG-AYA legal agreements, may be found openly accessible here: (https://www.medicaldataworks.nl/governance). In due course, the (redacted) STRONG AYA set of documentations will also be openly available here (www.strongaya.eu) as a template for future projects.

7.3 Personnel in STRONG-AYA

Local/national ecosystems consist of people with lived experience of AYA-onset cancer, personnel from multiple healthcare disciplines, including clinicians, allied health professionals, epidemiologists and individuals with data science skills. Each centre has its own responsible project and data managers. These, with the Principal Investigators, form the operational groups co-ordinating our local and partly national ecosystems. PPIE groups, charities and policy makers are integral to success, so each centre has clear plans and roadmaps for service user stakeholder engagement, as well as communication about the project with wider local healthcare professionals. We expect this to expand as other stakeholders, whose ethical principles align to those of STRONG-AYA, join our initiative.

7.4 Data analytics technology and technical standards in STRONG-AYA

In this section, we explore the technical nuances in more detail, specifically the key enabling technologies that makes a federated data ecosystem usable.

Federated learning ('FL') is an established and rapidly maturing methodology having uniquely high value in pan-European projects, especially when dealing with sensitive personal health data. The legally-binding implementation of the European Health Data Spaces (EHDS) follows a long tradition of federation projects in Europe, that rely on novel privacy-enhancing technologies – for example, secure multi-party computation, differential privacy and federated learning – as a means of enhancing multi-institutional and global cooperation, while protecting the rights and confidentiality of data subjects. European-level projects of similar scale and construction as STRONG AYA include: (i) FeatureCloud, which uses encryption and blockchain-based computations to protect individual identity. (ii) TRUMPET, which collectively learns from multi-site radiation dosimetry data in order to personalize radiotherapy treatment. (iii) AI4EOSC, which will provide computationally intensive federated Artificial Intelligence to researchers in the European Open Science Cloud. (iv) DIGIONE, which pioneers an EU-wide rapidly learning oncology system aiming to optimize treatment selection for cancer patients.

We have highlighted above the roles of local institutions, national bodies and international consortia in the formation of a federated data ecosystem. To connect these with an easily visualizable technology, we may consider: (i) the local institutional and national data repositories as stand-alone 'stations'. (ii) instead of moving personal patient data around, we will instead send out computer code that can perform statistical computations on our behalf, i.e. 'trains'. (ii) while the data is inherently protected by its own data stations, the integrity of the telecommunications connectivity between stations must be kept private to the consortium, and safeguarded against malicious actors trying to 'eavesdrop' on the statistical results or in

some way attempt to re-identify one of the data subjects. Such protections had to be woven into the digital and operational infrastructure of the ecosystem, which we would call 'tracks'. This gives us the easily accessible analogy of protocolized 'trains' travelling outwards to privately held data 'stations' to do statistical analysis, then bringing together the cohort-based analysis results (no individual data) using cybersecurity-assured 'tracks', i.e. an implementation of the "Personal Health Train" paradigm (https://www.dtls.nl/fair-data/personal-health-train/).

7.4.1 Autonomy of local and national data collection structures

Where local systems already collect Patient-reported outcome or Patient-reported experience data, this is utilised for STRONG-AYA data collection, and where there are not such local systems the expertise in the consortium supports its implementation. That is, each participating institution retains full control of its data repository, its data acquisition resources and its own data management policies, using insofar as possible the technical capabilities it has built up prior to forming STRONG-AYA. However, to support systematic and internationally harmonized collection of AYA data, the Consortium has now published and deployed the 'COS' for use at the local and national levels.

7.4.2 Technical implementation of data 'stations'

Whilst local and national actors fully control and entirely own their respective data, some intermediation technologies are required by each consortium partner to make this data simultaneously private yet also available for collaborative learning. Therefore STRONG-AYA develops, extends and deploys for all partners a set of free, open source and publicly accessible software tools for bridging institutional data formats with the consortium COS, despite the range of local and national systems [32].

In STRONG-AYA, we adhere to the Findable-Accessible-Interoperable-Reusable ('FAIR') data management principles [33]. This states two key principles that must be emphasized : (i) *FAIR data does NOT imply that data is openly accessible*, and (ii) that FAIR-ness constitutes data which is suitable – by way of 'metadata' annotations describing the actual data - that makes it possible for both humans (manually) and computer code (remotely) to interact with the data. Additionally, the FAIR principles states that not one single format or structure necessarily takes precedence over any other. For example, the OMOP Common Data Model ('CDM') is popular in many countries, but STRONG-AYA does not exclude any member's data solely on the basis that it did not conform to one 'CDM'. We include a plurality of data systems, structures and formats, already in use in local and national systems, some more and some less interoperable with each other. The key advantage of a 'metadata overlay' type of methodology is that, as a consortium, we will strive for a standardization of an optimum minimal level of data, and we can harmonize/inter-calibrate the rest, thus aiming at syntactic and semantic interoperability that will mature in parallel with the Consortium.

Whenever the Consortium needs to enhance the COS or update the definitions, it is only necessary to revise the metadata, not the actual patient data itself. Finally, multiple code dictionaries co-exist within the metadata, so the metadata implements the harmonization across terminologies and measurement scales, and it is not necessary to revise or re-code the raw data, to 'force' equivalencies, after it is collected. Federation also supports decentralized statistical model 'training' (e.g., regression, neural networks) without compromising data privacy and adhering to data minimization thresholds.

7.4.3 Statistical computing code as standalone 'trains'

All code for use by the Consortium is placed in a publicly readable repository, such that any person can openly examine the code and inspect it for potential vulnerabilities, and local data experts are required pragmatically to review that code as it is created. For consortium work, technology partners are included, and independently cross-verify that the code performs the mathematical function correctly, and does not contain any malicious action, before the specific version of the code in the repository is locked for use within the consortium. Only locked code is packaged into standalone, independent, and operating-system agnostic software applications (i.e., Docker® containers) which we refer to as 'trains'. The train thus encode the statistical analysis requested and desired by the Consortium partners acting jointly, for example, one train may contain a generalized regression model, or another train may have code to statistically summarized prevalence of different cancer types.

Thus development of the trains goes hand in hand with: (i) the specific use cases expressed by the Consortium, (ii) the agreement in advance of the Consortium about which data fields and what analysis needs to be performed, and (iii) how the final results of the computational actions are to be made available, and to whom.

7.4.4 Data analytics dashboards

The statistical results and epidemiological models coming from the STRONG-AYA data collections needs to be presented in a cognitively accessible way, to be disseminated to champions who can then implement transformation of AYA care in their local and national ecosystems. We take this to mean 'dashboards' (i.e. user-friendly data visualization web-browser based applications). These present the results of, or insights from, analyses completed upon the consortium data either on a scheduled (time-based), *ad hoc* (e.g. a specific one-off clinical hypothesis test) or on-demand (e.g. a user wishes to see a generalized cohort summary of PROs values across broad range of cancer types immediately, for good reason).

Therefore, *co-design and co-creation* – meaning clinicians, researchers, AYAs and policy makers as a cohesive whole – is fundamental to the way we report our results. The underlying technology is intended to be invisible and unobtrusive where possible, while the user focusses on gaining insight, communicating knowledge and applying figures/graphs created. The aim is to integrate AYA-specific aspects into, for example, research questions (from EuroCare enquiries), patient-driven questions (from AYA discussion fora) and for policy drafting (such as by ECO). The objective here is to make data insights, such as the COS accrued by our Consortium, accessible in a specific time-frame including 'in real time', and with granularity befitting the user and the question.

The choice of web-browser based application as a design principle is informed by the need to partly decouple the communication/dissemination aspects, which require capabilities in web design and visual user interaction, from the underlying infrastructure technology which handles the privacy, access controls, network cybersecurity and telecommunications protocols. The former is jointly a work in progress with consortium partners including YCE; the latter has been contracted in large part to a Medical Data Works B.V. that has specific expertise in Vantage6 open-source federated infrastructure.

7.4.5 Privacy enhancements in STRONG-AYA

A federated approach for AYA-onset cancer needs to address the challenges of co-ordinated working with small case numbers and yet provide timely, richer and more robust insights into care or treatment compared to single database or formal data transfer methods.

In STRONG-AYA our greatest shared responsibility is data security. This begins first at the local ecosystem level, where the detailed datasets are collected and stored. At all levels across STRONG-AYA, access and control protocols have been determined and developed to block certain information reaching specific stakeholders. For example (in future) a pharmaceutical or medical devices commercial partner may not be permitted to develop analyses involving specific drugs administered or special procedures carried out; however such information might be made accessible to clinicians and researchers who wish to develop prediction models. Well-developed access management from local levels shall underpin cooperation at an international level to support the beneficial use of the consortium infrastructure. The privacy enhancement in federated learning is underpinned by the unfeasibility of reverse-engineering a particular participants' identity from summaries, statistical models or other such combined analysis results based on groups of participants (typically presented in reports and publications).

Second, our data security must consider statistical analyses requiring frequencies of outcomes to answer research questions – such as the numbers of persons 'alive (or deceased)', 'diagnosis positive (or negative) or a time interval (such as '90 days since a certain event'). In special types of research questions, there is a risk that such an outcome may be characteristic to only one or a few patients – in which case the descriptive statistics or epidemiological model may no longer obscure the data values of individual subjects. OUr infrastructure is designed to block analyses that do not have sufficient cohort size. By default, as stated earlier, this number is arbitrarily set at 10, however institutions in certain use cases may have justification to increase or reduce this number, on a case-by-case risk-guided basis.

Finally, the development of the COS itself is based on the idea of data minimization, even though no person-level data is being shared. Structured information from the institutional collection of the COS data fields will be extracted by investigators within each institution, then populated into their respective STRONG-AYA data 'stations' controlled by their own institution and governed by their own data protection regulations. All data placed in these areas will be checked to ensure it is pseudonymized or anonymized (as consistent with the DPIA and data access agreements) to avoid the re-identification of individuals, and to remain compliant with GDPR and local regulations. Further, potentially traceable identifiable information which are not necessary for consortium analyses - such as birth dates and postcodes – are removed after extraction, and all absolute dates of clinical events will be replaced by time intervals, say number of days from diagnosis till recurrence of cancer, for example.

7.4.6 Challenges encountered in the federated analytics approach

While federated learning comes with certain attendant advantages in terms of data access and transparent analytics, there are additional considerations that must be acknowledged. Federated ecosystems may be able to process disparate, large, and complex datasets, to accelerate findings in a specific health field and thereby increase research power, detail, and reach. The federated data ecosystem concept acknowledges and respects that data collection workflows and data management practices will be specific locally, and already in use extensively among local project partners for numerous operational and research purposes. Datasets owned by project partners will be re-used for the unitary common goal of answering pre-defined clinical research questions, through initial priority research questions.

However, this autonomy and control over data at the local and national levels will also introduce personnel, expertise and resource allocation questions for the institutions that participate in the federated data ecosystem. In centralized designs, the partners may – if they wish – take a less active role by sending data to the principal investigator, thereby relying on computing resources, data cleaning, domain expertise and analysis skills available at the leading institution. The federated approach warrants that the skills and expertise are more dispersed among the partners, rather than being concentrated in one. There is therefore a strong and significant case for consortium-wide technical support, knowledge sharing, and deployment of customizable software tools to facilitate working in a federated fashion.

Additionally, there needs to be a fundamental willingness from the lead institution to share its expertise and support the maturation of research capacity (along with its attendant resources) more widely around the consortium, rather than keeping these in one location.

8 What can we learn from similar initiatives?

The Health Outcomes Observatory (H2O) project is a public-private partnership funded by the European Union. It aims to build an international ecosystem to incorporate health outcomes, including patient-reported outcomes in healthcare decision making across three diseases: diabetes, cancer and inflammatory bowel disease. There are several similarities with STRONG AYA. Both projects are interested in outcome sets, particularly patient-reported ones, and both use a federated data infrastructure and seek to use digital tools to involve patients in the data sharing pathway. They differ in terminology; H20 has favoured the term 'observatories' to describe its data infrastructure whereas we use the term 'ecosystem' and they vary in their governance structure. H20 has opted to create separate non-for-profit legal entities, co-led by a public and private enterprise representative in each country. These entities are currently operational in three of the four countries involved in the project: Austria, Spain and the Netherlands (Germany has not yet been set up according to public resources). A review of their website (https://health-outcomes-observatory.eu/) suggests that this legal entity route may allow for further grant acquisition later as they would be new entities applying for funding, but that a considerable amount of time and resources were allocated within the project to the bureaucracy to create these entities. Taking these steps at an early stage potentially limits buy-in from other parties as feasibility has not yet been proven and buy-in is based on the premise rather than the actual function. Furthermore, the national observatories may be national in their legal registration but face the challenge to produce nationwide data if they cannot on board multiple centres, data sources, or nation-wide studies in each country. The Dutch national observatory still only includes the Erasmus University Medical Centre as a source of data, and similarly only one medical centre is involved in Spain and in Austria and Germany, where one centre is contributing data but no national observatory has been established. No pan-European observatory has been formally established, although international aggregated data is being used and visualized via the H20 Insight Centre (https://public.tableau.com/app/profile/k.funk.tableau/viz/H2OInsightsCentre/H2OInsigtsCentre), a public facing dashboard. Patient recruitment also currently sits at under 1100 patients according to the Insight Centre statistics. By defining our ecosystems by data source and by technical functionality and with flexible pragmatism, STRONG AYA's ecosystems can perform earlier analysis on a larger number of patients to generate insights earlier.

Data for the Common Good (https://commons.cri.uchicago.edu/) is an example of a strong and effective data insights ecosystem, with related but distinct objectives and methods. Its focus is across diseases, but encompasses cancer. Its reach is intercontinental, with a particular emphasis on diseases affecting young children, though it features an AYA cancer working group. The initiative prioritises existing data (in particular,

tightly curated clinical trials datasets) over real-world data. The primary method for data co-analysis involves a data commons rather than a federated learning system. In this approach, data is explicitly re-coded according to consensus-based data dictionaries and formally exported to the common database, where members can access and analyse it, line by line, with appropriate permissions. Its work is well advanced, with 90 publications and over 44,000 cases of childhood cancer, integrating socio-economic data and producing decision-support tools for specific disease entities (more than work based upon age) from its existing large dataset.

The European Reference Networks (ERNs) are concerned with low-prevalence diseases so are a useful point of comparison with STRONG-AYA. ERNs were established by the European Union to 'bring together European hospital centres of expertise and reference to tackle rare, low prevalence and complex diseases and conditions requiring highly specialised healthcare' (https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en). Moreover, the building of data registries is a key component of the ERN mandate; given the challenges of this in their diseases of focus, ERNs inherently require pan-European data infrastructures to be built. One such example is the Blueberry project led by the Dutch Comprehensive Cancer Organisation (IKNL) and Istituto Nazionale dei Tumori in Italy. The project established a blueprint for the creation of a EURACAN (ERN for rare adult solid tumours) registry, and aims to identify challenges and ways forward to establish a pan-European registry, with sarcoma as its use case. It uses a federated data infrastructure very similar to the one STRONG-AYA proposes, to connect sarcoma registries in France, Italy, Austria, Norway and Spain. Key challenges identified by the project included governance, navigating GDPR, securing long term financial investment, especially for low-prevalence cancers where it is harder to attract private investment from the pharmaceutical industry and data harmonization.

9 Key challenges facing the STRONG AYA Consortium and other data ecosystems

9.1 Scalable, sustainable, and interoperable

The STRONG-AYA Consortium is committed to the principle of scalability of its resulting activities, tools, and pan-European ecosystem. This requires a pragmatic approach to the continuing autonomy of local AYA cancer investigators internationally, and an openness to the various ways they undertake their local research, their ever-changing ethical codes, and ways they may structure, encode and understand their data in future. STRONG-AYA can therefore remain translatable to other contexts - and promote sustainability - beyond the current funded lifetime of the research project. Challenges include the identification of resources, from further academic grants or users of the system for example, and the maintenance of expertise, data and a modest IT infrastructure.

9.2 Management of national and international differences, over time

Within the definition of the Governance Model with its structure, principles and alignment of incentives and motives, the Consortium acknowledges and welcomes the differences between institutions and countries in their capacity, capabilities, and limitations in what data can or cannot used for federated learning or analyses. The ethos of a collaborative pan-European ecosystem is that these gaps can be bridged by our working

practices. Before joining the Consortium, there is an expectation of local in-person evaluations and discussions with leadership, technical and legal teams to identify local capacity. Local evaluation of data availability and metadata structuring, to identify strengths and opportunities for development at each national ecosystem level is likely to be be necessary. These assessments need to be shared with the CfC as part of the new member application process. Interviews and audits to identify potential local knowledge gaps should prompt local discussions and incorporate at minimum the following categories:

A. Provision and availability of retrospective data: Is data available and can it be analysed?

**The questions below apply to retrospective data. Approvals in line with GDPR and local legal/ethical guidelines will be arranged by the consortium member.*

- What data is currently available, from previous studies or clinical practice?
- Did patients consent to their data being processed for public health research?
- Can the existing data be FAIR-ified?
- What is the local complaints procedure (i.e. immediacy of action, timeframes for reply etc)?

B. Data collection and consent norms for prospective data: How is data collected?

**The questions below apply to prospective studies. Appropriate Ethics Board approvals in line with GDPR and local legal and ethical guidelines will be expected to be in place in each location.*

- Do patients know why the data is being collected?
- Do patients agree to share the data?
- Do patients know why the data is being shared?
- Do patients understand how it is being shared?
- What is the local institutional process of communicating results back to patients (if necessary)?
- Does data collection follow pre-existing regulations and rules?
- What is the local complaints procedure (i.e. immediacy of action, timeframes for reply etc)?

C. Operational norms and standards: How does each institution operationalize the Consortium internally?

- From whom will you accept a data query?
- For what purpose will you use (or allow use) of the data?
- What results will you allow to be returned?
- Who adjudicates disputes or ambiguous situations locally, in regard to data collection and analyses (either from participants or researchers)?
- What is the threshold to entry for a new member of the federations?

D. Technical standards: What are the technical standards (i.e. How are the datasets managed to guarantee data security, data integrity, and patient privacy)?

- How do you ensure security of queries within the transfer process?
- How is patient privacy protected?
- Do you make your build standards transparent?
- What is the change management process?
- What level of interoperability needs to be reached (in terms of software, APIs, and datasets)?
- What are your mechanisms for ensuring data integrity?
- How is data standardization currently achieved?

E. Legal and ethical governance standards: What are the legal bases for consenting, collecting, and using patient data in research in each country and institution?

- Does a joining institution have consent and research process protocols and practices in place?
- Do these allow for insight sharing as part of a federated learning ecosystem (and to what extent)?

9.3 Management of wider stakeholder access – within and outside of the consortium

Our Governance Model describes the rules for managing data access across different user levels, such as various consortium members and non-members. All engaged stakeholders are assigned a 'user category' within the Consortium, which defines their 'user access' to information within the STRONG-AYA data ecosystem. Consistent with the model of the Personal Health Train described above, data stations are always under the control of the local institution, safeguarded by local data protection measures. No individual or organisation external to the member institution should ever have any rights or permissions to access line-by-line local data through STRONG-AYA. Different levels of user access build upon the 'tracks' and 'trains' analogy of our system. The following levels of access are defined:

(1) International Administrative: view and edit rights to the coded technical architecture of the international data ecosystem. This access is reserved to those who manage and maintain the data ecosystem, but those individuals cannot post use cases or run analyses.

(2) Local administrative: view and edit rights to the institutional data station. This user access is for local technical staff who can generate the necessary authentication keys of their data station, enable appropriate end-to-end encryption (i.e. data station-to-backend), and so require individual level access to the local data station for maintenance purposes. They will also not post use cases, epidemiological questions or statistical analyses.

(3) Full access: complete view rights of content but without ability to edit. This level will be reserved for core and enduring consortium members and their teams within institutional member organisations. They will be able to look up descriptive statistics, post use cases, pose epidemiological questions and formulate statistical hypotheses for validation by the consortium, which can be answered with the available data via the user interface. They will NOT have access to individual data stations nor can they create data analytic queries which have not yet been validated by the consortium members.

(4) Partial access: view rights relevant to approved stakeholders who are not core consortium members and new members. These members will have restricted access to descriptive statistics. For example, a commercial partner might be allowed restricted access to descriptive statistics but not to deeper epidemiological modelling. This would be an access level applicable to some cancer service users as well. This will be subject to *bona fides* assessments, Data protection and Infrastructure user agreements. For example a patient who has given consent to take part in a specific study to address a use case, or researcher in a specific institution with an institutional profile compatible with their involvement, may access more detail of analyses of specified use cases only. For example, pharmaceutical companies may not be permitted to develop queries/models involving drugs administered or procedures carried out without that request being considered by the Consortium as a whole. Data that is used for any such analyses may also be limited according to national ethical and regulatory frameworks. For instance, data on ethnicity cannot be utilised by law in France, so analyses of it may not be acceptable to be viewed by French participants.

(5) Observational access: view rights to insights produced by the Consortium, but without access to analysis results themselves is provided for the wider public, without access to deeper online interfaces. What the public is allowed to view will be pre-agreed by the Consortium members and is subject to local approvals.

For all types of users, at the local level and across STRONG-AYA, we will develop access and control protocols to block some information reaching some stakeholders. A special case within the management of user access is that of Data democracy. For equity in data provision and data queries and hence to reduce misalignment resulting from power democracy, the STRONG-AYA pan-European ecosystem ensures that:

- Data stewardship and control will be jointly held by Principal investigators (PIs) and their teams of investigators. Through this joint power over data, we ensure the context and subtlety related to data is maintained, and if data needs to be enriched, enhanced or requires additional management, this can be done in partnership within the local ecosystems, rather than it being the responsibility of a single person.
- All partners are equal in terms of suggesting a use case (if it is congruent with the foundational data usage principles and incentives of the consortium) and all partners will have the same “power” in terms of being able to run an analysis and test their hypotheses.

9.4 Engagement and ‘buy in’

9.4.1 Technological buy-in

Acceptability and utility to patients with lived experience of AYA-onset cancer make it more likely they will buy into technological systems; that are simple enough to use and provide information they value. The technical methods of the analysis is not the relevant factor here, if the system is well governed and so the data is secure and used to address important problems. This requires clear and open communication so that patients can understand what level of data protection the technology does offer, and where are the known limitations to what it can do.

Usability, comprehension and acceptability for healthcare staff requires an ongoing change from relying on resources and expertise of the central system, such as provided by the hospital, to become accustomed to obtaining the information needed more autonomously, through a running use case analysis.

Integration with other hospital / research systems is a challenge in STRONG-AYA, to utilize and set free the data siloed in various computerised systems and make the analyses available for clinical or research utilisation. Embedding the analyses, or links to the STRONG-AYA portal, within the Electronic Health Record or other frequently used resource will enhance its use. Availability of local computing resources and local telecommunications remains a challenge in many parts of the world where local budgets are exhausted by operational needs, and extra resources for data and data ecosystems needs to be purchased with research funding, until a self-sustaining income generation model from the data can come into long-term maturity.

Maintaining skill mix in the partner centres, particularly human technical expertise is a challenge for the future, within existing and new partners. In a centralized model of collaboration, the burden is carried by the PI and the member sites supply the raw or nearly raw data directly to the PI for all processing and analysis. In a federated data ecosystem, the data collection, data archival and statistical computing must be done locally within the site – since only processed statistical results are shared between members. This means that

it will NOT be sufficient to only have a clinician participate in a data ecosystem, but each clinician must be the 'quarterback' of a cohesive team: the clinician is a domain expert, but needs to be supported to lead a team of data managers, data collectors, epidemiologists, data scientists and IT professionals. Instead of relying on this to only be present at the PI institution, this competency, training, skill and support needs to be more evenly distributed as human technical expertise and not siloed into one single master institution.

9.4.2 Data hesitancy (patients, governments, IP and academic politics, registry redundancy fears, etc)

Data hesitancy from stakeholders is a challenge across the board when building any data ecosystem. Hospitals and their administrators are often concerned about the consequences, both legally and reputationally, of data leaks. Data has the potential to leak from any data collecting institution or – if the data is centralized – it can also leak from the PI institution. This is always the case for any clinical investigation, completely irrespective of whether the study design is centralized or distributed. This is known as "Leak at Source".

Data can leak when transferred between stakeholders – this is known as "Leak in Transit". In a federated data ecosystem, due diligence and safeguards are needed to protect against re-identifying a given data subject from the values or results of the analysis. No individual level data is in transit, therefore the risk here is interception of the cohort-based results and attempting to ***deduce or infer*** the information of individuals. Consortia making use of federated data ecosystems need to consider both leaks at source (as is the case for conventional data collection projects) and leaks in transit, and to devise suitable countermeasures for leaks in transit. Some examples that are used in STRONG AYA include message encryption, disinterested third party hosting of infrastructure, and improved statistical methods.

To assure against data security attacks, we have embedded and run cybersecurity audits demonstrating no currently known method to penetrate this federated data ecosystem from the outside. However malicious intent or unintentional mistakes can still occur inside the consortium, which potentially raises the issue of data leakage *among* ecosystem members. Hence data security is a shared obligation, and lawful purposes of data processing are shared responsibilities. Ways to address this challenge include a multi-pronged approach to data protection; whilst there is never a guarantee that data will be secured, the various steps that STRONG-AYA takes ensures that the risk is very minimal.

However, value needs to be communicated to patients at recruitment and where possible we need to give patients access to their data. Collaboration and building trust with colleagues and reiterating the value of traditional registries and showing how federated learning can optimize research and enable access to even larger datasets might help to convince more hesitant colleagues. After data safety concerns, as negligence or lack of expertise or lack of communication could drive data opt-out among patients, younger patients are often particularly keen to contribute their data but may hesitate or withdraw from studies if the impact of their data contribution is not clearly articulated, at both the macro level but also at the individuals level. The consortium is therefore considering the role for a solidarity-based data governance approach. In short, data solidarity consists of assessing both the value generated from any given data used for society, and the harm it is likely to cause. Once these balance out, assessing the overall public value is likely to derive from said data use (the tool used for that assessment is (https://pluto.univie.ac.at/). This extends beyond considering data from an individualistic perspective (as "personal" data that must be protected) and non-personal or anonymized data (that can be used freely). Public value approaches prioritize 3 specific pillars of data: (1) creating data use that is most likely to bring meaningful public value without creating undue risk to individuals. (2) Assessing then preventing harm from data use, or mitigating it when prevention is not feasible. (3) Ensuring that a fair portion of the benefits drawn from data use are fairly redistributed to each

of those who made such use possible [34]. Solidarity-based data governance strengthens collective control and ownership of data, ensuring that costs and benefits of digital practices are borne collectively and fairly [35]. Expertise from lived experience of AYA-onset cancer is supporting STRONG-AYA to promote data solidarity, and working to create informed contemporary data solidarity.

A challenge STRONG-AYA needs to contend with is the desire for patients to see their data and to experience direct impact of their provision of data in their care. Federated analysis is a potential disruptor in the space of traditional epidemiology, which historically saw large scale data cleaned, transferred and pooled centrally for analysis by the principal investigator. Federated analysis puts the onus on researchers to be more fastidious and systematized in their data collection, so data scientists can develop good algorithms to analyse the model in a federated infrastructure. It also means that analysis of the data can be conducted by any researcher within the consortium. As open science principles become more widely accepted, this may become less of a contentious issue, but for some researchers, this may be perceived as loss of 'primacy' and exerting control over an already-existing large dataset. There is also the attendant risk that the professional acumen, data stewardship and research resources that come with being a principal investigator, end up being diluted by dissemination among consortium members, therefore creating hesitancy to contribute data or expertise. Building a large-scale federated infrastructure of data may also raise concerns about possible redundancies among existing cancer registries, or encroachment into their traditional purview. Significant efforts therefore remain among the STRONG-AYA participants to incorporate institutional data sources as well as more traditional types of cancer registries, into a working federated data ecosystem.

9.5 Resources (digital, financial, personnel, expertise etc.)

Resources are, unsurprisingly, a challenge in building data ecosystems. Project-based or grant-based funding makes it difficult to fund ecosystems beyond the lifetime of a project without follow up grants which are far from guaranteed. Project based funding is also reliant upon evolving economic and political climates, regarding which initiatives receive funding priority at any one time.

This economic challenge can become compounded when the data ecosystem and data within it is not viewed as a good investment from commerce or industry, often the pharmaceutical industry in the context of cancer. Governmental funding may be possible in some instances but aging populations and limited resources in healthcare may mean that novel systems may not be of governmental priority. Data ecosystems also are potentially expensive, when they require people to clean data and regularly update software and/or hardware systems, which is highly skilled labour where capacity may be limited. In places where records are unstructured or not electronic in the first instance, there is a high level of labour required to convert data to electronic in a structured and systematic way. Healthcare professionals are often overworked with limited capacity and may not have the time or motivation to use, or learn to use, a new electronic system. Solutions to this include diversifying the portfolio of funding for data ecosystems and seeking governmental or more secure sources of project funding. This could include the implementation of a subscription service for some users, as the project matures, such as is used by the UK Cancer Registration Service, charging its' costs for academic data exports.

Providing technical support from elsewhere in the consortium to help guide local staff in building and operating the system is also helpful to rapidly overcome learning curves and hence maximise the use of existing funding. On-site visits from technological and information governance expert teams within the Consortium have been invaluable in assuaging local concerns and helping them build their local infrastructures from the ground up. Vantage6 software development is open-source and as such benefits

from a community of experts to help refine it further. Training programmes for healthcare professionals on the technical infrastructure and on the use of PROMs in research and clinical practice might also help generate greater buy-in and support for systems that use patient reported outcomes.

9.6 Longevity/sustainability

The STRONG-AYA approach is adaptable to several changes in technology – our use of Application Programming Interfaces 'APIs' for example is accepting of the inevitable new software implemented in healthcare centres.

The technology used in STRONG-AYA is itself scalable but is dependent on the resources of partners wishing to get involved and their financial and political constraints. Here the flexible approach of STRONG-AYA to overcome these differences will be paramount to longevity, maintaining inclusivity and quality of analysis.

The development the COS for AYAs with cancer enables STRONG-AYA to offer new opportunities for the collection of homogeneous prospective data, which will support new research and healthcare innovation. The agile and adaptable approach of FL allows flexibility in incorporating existing systems and practices in larger ecosystems, allowing for the accommodation of new local contexts and ecosystems. Tools which will enable and motivate stakeholders to join and work together are the websites, specific ways of set-up of local databases, portals for stakeholders to enter and access desired information, and finally the access to the expertise required to operate ecosystems at local and European levels. Finally, STRONG-AYA will remain responsive to emerging and unanticipated ethical issues by adhering to iteratively developed ethical guidance throughout the duration of the project, including patient-centred.

Close links, communication and collaboration with user groups and stakeholders are vital to ensuring that STRONG-AYA evolves with their needs and as new research questions arise in AYA. As a Consortium the inclusion of early career researchers across all disciplines will help ensure that ongoing success and longevity of STRONG-AYA.

10 Future actions and considerations

There are a range of future actions under consideration: Strong-AYA can grow and expand for example by

- Engaging other local or national data ecosystems about AYA-onset cancer.
- Incorporating other types of data e.g. genomics, pharmacological data
- Widen functions and use cases, and adapting where needed within our governance framework

At the heart of this are the young people with lived experience of cancer. The STRONG-AYA ecosystem primacy of continued patient involvement recognises that technical solutions are only one part of the advances. The next steps need to focus on scaling and sustaining meaningful PPIE and co-design. As those with lived experience learn with the project, they become better advocates and invaluable peers in the Consortium, contributing their expertise across various domains, from research and data analysis to marketing techniques. STRONG-AYA is uniquely positioned to empower AYAs with personalized, tailored information and opportunities to explore their own data, all the while enhancing digital health literacy—a skill crucial for the 21st-century. Through time, this can result in ever more seamless integration or research

into routine clinical care, so that participation benefits individuals, society, as well as medical communities. Achieving these goals requires respecting and expanding diversity of patient perspectives, working transparently, and prioritizing ethics, privacy, and security.

11 References

1. Group AaYAOPR: Closing the gap: Research and Care Imperaives for Adolescents and Young Adults with Cancer, in NIH (ed): No. 06-6067. Bethesda, MD, National Institute of Health, National Cancer Institute, and Livestrong Young Ault Alliance, 2006
2. Close AG, Dreyzin A, Miller KD, et al: Adolescent and young adult oncology-past, present, and future. CA Cancer J Clin 69:485-496, 2019
3. Gupta S, Harper A, Ruan Y, et al: International Trends in the Incidence of Cancer Among Adolescents and Young Adults. J Natl Cancer Inst 112:1105-1117, 2020
4. Bray F, Ferlay J, Soerjomataram I, et al: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394-424, 2018
5. Lewis DR, Seibel NL, Smith AW, et al: Adolescent and young adult cancer survival. J Natl Cancer Inst Monogr 2014:228-35, 2014
6. Trama A, Botta L, Foschi R, et al: Survival of European adolescents and young adults diagnosed with cancer in 2000-07: population-based data from EUROCARE-5. Lancet Oncol 17:896-906, 2016
7. Adams SC, Herman J, Lega IC, et al: Young Adult Cancer Survivorship: Recommendations for Patient Follow-up, Exercise Therapy, and Research. JNCI Cancer Spectr 5, 2021
8. Chao C, Bhatia S, Xu L, et al: Chronic Comorbidities Among Survivors of Adolescent and Young Adult Cancer. J Clin Oncol 38:3161-3174, 2020
9. Palmer JD, Hall MD, Mahajan A, et al: Radiotherapy and Late Effects. Pediatr Clin North Am 67:1051-1067, 2020
10. Ryder-Burbidge C, Diaz RL, Barr RD, et al: The Burden of Late Effects and Related Risk Factors in Adolescent and Young Adult Cancer Survivors: A Scoping Review. Cancers (Basel) 13, 2021
11. Woodward E, Jessop M, Glaser A, et al: Late effects in survivors of teenage and young adult cancer: does age matter? Ann Oncol 22:2561-2568, 2011
12. Fidler MM, Frobisher C, Hawkins MM, et al: Challenges and opportunities in the care of survivors of adolescent and young adult cancers. Pediatr Blood Cancer 66:e27668, 2019
13. Perez GK, Salsman JM, Fladeboe K, et al: Taboo Topics in Adolescent and Young Adult Oncology: Strategies for Managing Challenging but Important Conversations Central to Adolescent and Young Adult Cancer Survivorship. Am Soc Clin Oncol Educ Book 40:1-15, 2020
14. Ferrari A, Stark D, Peccatori FA, et al: Adolescents and young adults (AYA) with cancer: a position paper from the AYA Working Group of the European Society for Medical Oncology (ESMO) and the European Society for Paediatric Oncology (SIOPE). ESMO Open 6:100096, 2021
15. Collaborators GBDAYAC: The global burden of adolescent and young adult cancer in 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Oncol 23:27-52, 2022
16. Bleyer A, Barr R, Hayes-Lattin B, et al: The distinctive biology of cancer in adolescents and young adults. Nat Rev Cancer 8:288-98, 2008
17. Dommett RM, Redaniel MT, Stevens MC, et al: Features of cancer in teenagers and young adults in primary care: a population-based nested case-control study. Br J Cancer 108:2329-33, 2013
18. Fern LA, Birch R, Whelan J, et al: Why can't we improve the timeliness of cancer diagnosis in children, teenagers, and young adults? BMJ 347:f6493, 2013

19. Meeneghan MR, Wood WA: Challenges for cancer care delivery to adolescents and young adults: present and future. Acta Haematol 132:414-22, 2014
20. Levin NJ, Zebrack B, Cole SW: Psychosocial issues for adolescent and young adult cancer patients in a global context: A forward-looking approach. Pediatr Blood Cancer 66:e27789, 2019
21. Keegan THM, Abrahao R, Alvarez EM: Survival Trends Among Adolescents and Young Adults Diagnosed With Cancer in the United States: Comparisons With Children and Older Adults. J Clin Oncol 42:630-641, 2024
22. 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 122:988-99, 2016
23. Osborn M, Johnson R, Thompson K, et al: Models of care for adolescent and young adult cancer programs. Pediatr Blood Cancer 66:e27991, 2019
24. Fardell JE, Patterson P, Wakefield CE, et al: A Narrative Review of Models of Care for Adolescents and Young Adults with Cancer: Barriers and Recommendations. J Adolesc Young Adult Oncol 7:148-152, 2018
25. Das M: New report from STRONG-AYA on addressing needs of adolescents and young adults with cancer in Europe. Lancet Oncol 25:960, 2024
26. Shah A, Stewart AK, Kolacevski A, et al: Building a Rapid Learning Health Care System for Oncology: Why CancerLinQ Collects Identifiable Health Information to Achieve Its Vision. J Clin Oncol 34:756-63, 2016
27. Lambin P, Roelofs E, Reymen B, et al: 'Rapid Learning health care in oncology' - an approach towards decision support systems enabling customised radiotherapy'. Radiother Oncol 109:159-64, 2013
28. Price G, Mackay R, Aznar M, et al: Learning healthcare systems and rapid learning in radiation oncology: Where are we and where are we going? Radiother Oncol 164:183-195, 2021
29. Cromie KJ, Crump P, Hughes NF, et al: Data Resource Profile: Yorkshire Specialist Register of Cancer in Children and Young People (Yorkshire Register). Int J Epidemiol 52:e18-e26, 2023
30. Husson O, Ligtenberg MJL, van de Poll-Franse LV, et al: Comprehensive Assessment of Incidence, Risk Factors, and Mechanisms of Impaired Medical and Psychosocial Health Outcomes among Adolescents and Young Adults with Cancer: Protocol of the Prospective Observational COMPRAYA Cohort Study. Cancers (Basel) 13, 2021
31. Vlooswijk C, Poll-Franse LVV, Janssen SHM, et al: Recruiting Adolescent and Young Adult Cancer Survivors for Patient-Reported Outcome Research: Experiences and Sample Characteristics of the SURVAYA Study. Curr Oncol 29:5407-5425, 2022
32. Gouthamchand V. CA, Hoebers FJP., Wesseling FWR., Welch M., Kim S., Kazmierska J.,, Dekker A., Haibe-Kains B., van Soest J., Wee L.: Making head and neck cancer clinical data Findable-Accessible-Interoperable-Reusable to support multi-institutional collaboration and federated learning. BJR|Artificial Intelligence 1:ubae005, 2024
33. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018, 2016
34. Prainsack B, El-Sayed S: Beyond Individual Rights: How Data Solidarity Gives People Meaningful Control over Data. Am J Bioeth 23:36-39, 2023
35. Prainsack B, El-Sayed S, Forgo N, et al: Data solidarity: a blueprint for governing health futures. Lancet Digit Health 4:e773-e774, 2022

36. S. Oliveira, M.I., Barros Lima, G.d.F. & Farias Lóscio, B. Investigations into Data Ecosystems: a systematic mapping study. *Knowl Inf Syst* **61**, 589–630 (2019). https://doi.org/10.1007/s10115-018-1323-6