**efpia***
**European Federation of Pharmaceutical Industries and Associations**
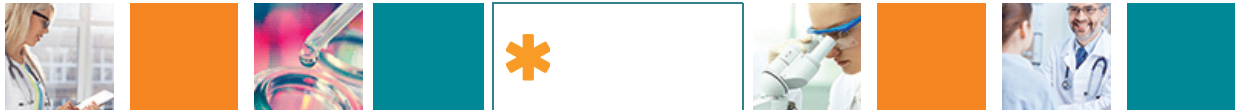
Draft ☐   Final ✳

# The oncology data landscape in Europe: report

Calypso Montouchet, Michael Thomas, James Anderson, Sam Foster

## Purpose

Health data has a significant role to play in supporting the discovery and development of new therapies. It also helps understand where therapies can deliver most value for patients and the health systems that support them.

With the advent of new technologies, health data is increasingly available from a wide range of sources, but challenges remain in a complex and fragmented European landscape. This is particularly true for oncology, where data sources struggle to keep pace with the increasing speed of innovation and new treatment paradigms.

In February 2018, EFPIA commissioned IQVIA and A.T. Kearney to review the landscape for oncology data in Europe, in order to gain better visibility of the situation and identify opportunities to improve the collection, analysis and use of this data.

The purpose of this document is to:
- Present the findings of this research.
- Raise awareness and understanding of the European health data landscape, overall and for oncology specifically.
- Discuss challenges and opportunities to improve the collection, analysis and use of health data.
- Discuss recommendations for improved use of health data, leading to better patient outcomes.

## Table of contents

The following research has been conducted by A.T. Kearney and IQVIA, and does not constitute an EFPIA position on health data in oncology.

# 1. An Opportunity-Rich Environment for Innovation in Oncology

With the advent of new technologies, **health data is increasingly available** from a wide range of sources. This explosion of information offers new opportunities but also highlights old challenges, such as the fragmentation of the European data landscape, quality and methodological limitations, and concerns around data privacy and security, to name but a few. Given the potential of data to improve population outcomes and health system sustainability, these challenges must be overcome.

**This is particularly critical in oncology**, which has seen unprecedented innovation in recent years: treatment paradigms are shifting from tumour- to mutation- and biomarker-based, gene editing is now within the realm of the possible, but data sources and standards are struggling to keep pace. Moreover, certain aspects of oncology – such as the reliance on genetic information and biomarkers, narrow patient populations, and the growing promise of combination therapy – add layers to the already complex European data landscape.

Over the course of several round tables with patient representatives, payers, health technology assessors and industry members, the **European Federation of Pharmaceutical Industries and Associations (EFPIA) identified the need to get a better understanding of the European oncology health data landscape**, outline the main challenges and identify opportunities for improvement. In February 2018, EFPIA commissioned IQVIA and A.T. Kearney to review the landscape for oncology data in Europe, in order to gain better visibility of the situation and identify opportunities to improve the collection, analysis and use of this data. This research is based on a comprehensive literature review of recent publications, reports, news articles, as well as more than 50 interviews with patient representatives, oncologists, data source owners, academics, regulators, health technology assessors, payers, tech innovators and pharmaceutical companies.

This report summarises the research and presents several avenues for collaboration to drive lasting changes in health data, eventually leading to improved access to innovation for patients. It is structured as follows:

- **The Changing Role of Health Data in Oncology** – outlines the types and uses of oncology health data, as well as the drivers for its adoption.
- **Diversity and Fragmentation of European Health Data** – presents an overview of health data approaches across select European countries, as well as detailed insight into select data sources and European initiatives.
- **Ongoing Challenges in Oncology Health Data** – lists the main barriers to health data collection, analysis and use in oncology, including examples of where these occur and potential solutions to overcome them.
- **Next Steps for a Better European Health Data Landscape** – outlines possible interventions to drive improvement in the quality, availability and use of health data in oncology.

# 2. The Changing Role of Health Data in Oncology

Over the last decade, **the understanding of disease biology and care, especially in oncology, has greatly increased**. Cancers are increasingly being understood and characterised at a molecular level rather than at a traditional histological level – fifty years ago, blood cancers were categorised into leukaemia and lymphoma; today, 40 unique leukaemia types and 50 unique lymphoma types can be differentiated[1].

Targeted therapies and immunotherapies use the molecular aberrations of a cancer cell, the cancer environment or cancer-fighting immune cells. Patients have benefited from these innovations: data show that more people are living longer, better quality lives following cancer diagnosis[2]. New technologies (e.g. CRISPR gene editing, CAR T-cell therapy) will continue to push the frontier and challenge the ways in which cancer is approached.

In line with this innovation in oncology, the **health data collected on cancer patients, their disease and treatment modalities is evolving rapidly**. Randomised controlled trials (RCTs) remain the gold standard for the European Medicines Agency (EMA), health technology assessment (HTA) agencies and payers, but they have several limitations in oncology:

- **Representativeness** – current oncology treatment regimens involve multiple lines of treatment and combination therapies, which cannot be reproduced in a controlled setting at sufficient scale to be statistically meaningful, and trial populations tend to be selected based on their physical wellbeing and thus are younger and healthier than real-world patients.
- **Timeliness** – highly-innovative therapies are increasingly approved through accelerated or adaptive pathways, with limited time to run clinical trials and the need for continuous information post-launch (including, but not limited to, after conditional approval).
- **Quality** – clinical trials are often conducted for the main indication(s) of a new treatment, such that the data on potential uses of a treatment outside its authorised indications will not be of the same quality as data developed in a RCT programme for those indications, and measures of efficacy may vary from those preferred by regulators (e.g. progression-free survival [PFS] compared to overall survival [OS]).
- **Ethics** – one-arm trials may represent the only ethically-appropriate approach for patients with rare cancers (in the absence of standard of care or where the hypothesised benefit from the trial drug is superior to the potential comparator), requiring the use of historical controls from registries and other non-RCT sources to fully understand comparative effectiveness.

A growing number of **health stakeholders are therefore turning to real-world data (RWD)** to supplement RCTs, both in and beyond oncology. Definitions of RWD vary, but a commonly-accepted view in Europe is that RWD constitutes "longitudinal patient level data captured in the routine management of patients that which can be repurposed to study the impact of healthcare interventions". This includes: (1) **electronic health record** (EHR) data on patient symptoms, referrals, prescriptions and treatment outcomes (including patient-reported outcomes [PROs]); (2) **claims data** on service usage, insurance and other administrative hospital data; (3) **-omics data** (e.g. genomics, proteomic) individuals and associated biomarker data; (4) **pharmaceutical data** such as pharmacovigilance (i.e. medicines safety); (5) **social media and web data**, for example from patient forums; (6) data from **mobile apps, wearables and sensors**; and (7) additional information from **ad hoc sources** (e.g. geospatial health data, information on well-being, socio-economic status or behaviour)[3].

This vast, diverse amount of **RWD can be used for numerous purposes** (Figure 1). Different health decision-makers tend to focus on specific applications: for example, governments and policy-makers typically try to achieve a better understanding of the healthcare context and treatment patterns to improve the quality of care and overall resource allocation. Healthcare professionals (HCPs) consider treatment patterns, the real-world clinical value of drugs and patient outcomes, in order to prescribe the most appropriate treatment for individual patients based on their characteristics and response to the drug. Pharmaceutical, medical device and biotech companies use RWD to inform R&D decisioning and conduct trials more efficiently, and to support discussions with health authorities and fulfil post-approval requirements. Recent years have seen the growing intent in the use of RWD to consider the socio-

economic value of economic interventions, enable innovative pricing mechanisms and provide better insight into the patient experience of their disease, treatment and overall wellbeing. These stakeholders are eager to use RWD opportunities to achieve real shared benefits, and this will require increased acceptance of RWD by decision-makers.

For all these health stakeholders and society at large, **RWD has delivered significant value and will continue doing so as its usage increases**. For example, in Italy, IBM and the National Cancer Institute of Milan launched a project to use genomics and analytics technology to improve the treatment of rare tumours, sarcomas and cancers of the head and neck, leading to more personalised care and better patient outcomes[4]. In Hungary, the government has launched a national health app to enable the good management of chronic diseases through lifestyle and treatment advice, based on uploaded patient data. This has helped detect inefficiencies in charging and reimbursement for cancer therapy, adapt cancer patient pathways accordingly, and increase the detection of liver metastases by 50% in 18 months[3].

**Figure 1. Applications and benefits of oncology RWD**

| Application | | Sample benefits | |
|---|---|---|---|
| R&D enablement | To support identification of promising compounds, investigation of the genome & smarter clinical trials | EHR4CR | The **EHR4CR** initiative enables more precise recruitment, retention & site-selection strategies via better patient-level data |
| Healthcare context | To understand the context of the disease & patient populations (e.g. population, biomarkers/ genetic characteristics & unmet need) | | In Italy, **IBM & the National Cancer Institute of Milan** use genomics to improve the treatment cancers, leading to personalised care & better outcomes |
| Treatment patterns | To understand real-world usage of anti-cancer treatments, including by patient group, line of therapy & geography | | In Hungary, a **national health app** has been used to detect inefficiencies in charging & reimbursement for cancer therapy, leading to pathway adjustment & increased detection |
| Real-world clinical value | To measure the delivery of cancer interventions' clinical promise in a real-world setting (including outcomes & safety, quality assurance, etc.) | | In the US, the FDA granted accelerated access to avelumab based on an open-label, single-arm study supported by RWD in metastatic Merkel cell tumour |
| Socio-econ value | To measure the value of cancer interventions beyond that provided to patients & health systems (inc. lost employment, absenteeism…) | | In Sweden, the societal & humanistic value of new drugs is considered as part of the **health technology assessment** process |
| Pricing enablement | To provide a mechanism for flexible pricing, based on use, indication and/ or outcomes | | In Italy, MEAs established from 2006-2008, mostly for oncology drugs, showed that they helped decrease the time to market by 75% (from 343 to 84 days) |
| Patient perspective | To offer insight into QoL(inc. PROs), covering aspects of care beyond clinical outcomes | patientslikeme | The **PatientsLikeMe epilepsy portal** allows better involvement of patients in clinical trial processes, facilitating research that responds to patient needs |

Looking further into the topic, **there are strong pressures to pursue the use of RWD** and several trends will increase the availability of this information[10]. With ageing populations and rising healthcare costs, private and public payers are exploring ways to ensure long-term financial sustainability. Digital technologies are being tested to decrease costs and outcomes-based payment models to manage the influx of innovative drugs, generating increasing amounts of data available on patients and the treatments they receive. Regulatory agencies are increasingly accepting and even requesting RWD, to document safety or support effectiveness data. With the development of accelerated and adaptive pathways that recognise RWD as critical to measure new treatments' value, the availability and quality of health data developed for this purpose is likely to increase

In addition, patients are increasingly at the centre of their own care, expanding their reach not only as data generators but also as data consumers. Growing demands for transparency and value realisation from their data are likely to improve accountability in the health data landscape. Beyond mobile health ("mHealth") which is increasingly used across Europe, new technologies such as artificial intelligence,

machine learning and blockchain have the potential to revolutionise healthcare data. Although still at a pioneering stage, these show promise in their potential to accelerate data collection, improve quality and foster transparency.

However, **there are also a number of opposing trends that may limit or delay the use of relevant RWD for oncology**. Although recent years have seen the emergence of pan-European efforts, including public-private partnerships (PPPs), including under the Innovative Medicines Initiative (IMI), national and regional preoccupations are challenging the cohesiveness of Europe, its member states, and the numerous data sources and initiatives that exist across these. In May 2018, the General Data Protection Regulation (GDPR) came into effect, with the aim of harmonising data privacy laws, of protecting and of empowering EU citizens. This will push Member State governments to re-consider and companies to re-shape the way in which organisations approach data privacy but could also increase the administrative burden and fragmentation across Europe.

# 3. Diversity and Fragmentation of European Health Data

European healthcare systems reflect centuries of political decisions, economic challenges and cultural mindsets to name but a few. There are national-led systems (e.g. UK, France, Belgium) and regional ones (e.g. Spain, Italy). There are tax-funded systems (e.g. UK, Nordics, Spain) and insurance-based models (e.g. Germany, France, Belgium, Netherlands). **These differences predictably affect policies regarding health data** (Figure 2) [5,6,7]: Nordic countries have had national datasets since the 1970's and their national EHR strategies allow linkage across various settings of care, while France and Belgium started developing national EHRs in the mid 2000's. Some regional health systems like Spain are making efforts to define common standards across regions, while others like Italy continue to have fragmented data with regional disparities. In Germany, health data is typically owned by statutory health insurance providers for which this constitutes commercial advantage, and there are no plans to deploy national systems.

**The European landscape is characterised by a multiplicity of data sources**. There are over 1,100 oncology data sources across Europe. France, Germany, Spain, Italy and the UK have the largest number of cancer data sources, but the Nordics, Austria, Switzerland, Slovenia and Estonia have the highest concentration of sources per capita[8] – none of which is a gauge of quality (Figures 3 and 4). Three quarters of oncology data sources in Europe are either standalone or partnership academic registries, while the remainder is made up of large-scale clinical registries, administrative data and claims, facilitated networks, and electronic medical records (Figure 5). Almost 80% of these sources cover several cancers, typically alongside other therapy areas, requiring a broad range of inputs that may not be suitable for all analyses. The most common single cancer sources are for breast, prostate, colorectal and lung cancers, and leukaemia.

> *"The information we need is out there, it's in the heads of the clinicians, the notes, the electronic health records, the specialty medical systems. The issue is that it is atomised, we need to understand all of those different pieces of information pulled together."*[9]

**The vast majority of data sources come with multiple challenges when considering their value in supporting insight generation**[9]. Access to data sources is difficult across all types, with particular difficulty with academic registries where governance and legal process may not be as well defined as for other sources. Once access is agreed, issues often remain around the quality and consistency of datasets; where this is improved, it is often to the detriment of data breadth and/ or data capture latency. Some improvement can be found in facilitated networks where a lead organisation is able to provide improved governance, consistency and efficiency to sites within the network.

Although many of these data sources have been set up to answer specific queries, a significant share of these are established at the national or regional levels, by official institutions and governments, for specific applications. However, despite their role in informing research or healthcare decision-making, **most of these data sources only barely meet their intended applications** (Figure 6). European oncology data sources tend to be variable in their ability to shed light on healthcare context, treatment patterns or the real-world clinical value of drugs, with the exception of facilitated networks and claims which provide good information on treatment patterns. Facilitated networks, EHRs and claims also provide decent information to enable pricing arrangements. For all other current applications of health data, especially to highlight the patient perspective or socio-economic value of drugs, current data sources fail to provide sufficient information to derive systematic, meaningful insight.

**Figure 2. Health data landscape in Europe**

| Maturity: | Leading | Emerging | Lagging |
| --- | --- | --- | --- |

| Country | Comments |
| --- | --- |
| (Belgium) | eHealth platform introduced in 2008; data sharing limited to regional level; limited legislation on use of health data |
| (France) | Country-wide EHR system in place; initiative underway to enable a shared oncology database; lacking standards & data quality |
| (Germany) | EHRs owned by SHIs; poor linkage due to strict privacy rules; lack of standards |
| (Italy) | Regional EHR systems; lack of national eHealth and/or oncology plan; several managed-entry agreements in place for new oncology drugs |
| (Netherlands) | Gaps in a national EHR plan (but being solved); widespread use of EHRs; limited sharing across healthcare centres or quality standards |
| (Poland) | Mandatory EHRs; plans to introduce a national patient account & ID system; legal issues around access |
| (Spain) | Regional EHRs despite national strategy; limited data sharing; lack of legal procedures that hinders widespread access |
| (Sweden) | National EHR strategy that allows linkage across health centres & databases using a patient ID; clear & well-understood patient consent |
| (Switzerland) | National plan for EHRs but regional disparities; limited country-wide sharing; ad hoc access approval, with few process standards |
| (United Kingdom) | Widespread EHR adoption; independent body to establish national cancer databases; well-developed data quality & linkage across datasets |

*SHI = social health insurer; source: Cancer Atlas; WHO. 'Global eHealth survey' (2015); European Commission. 'Overview of national laws on EHR' (2013); OECD. 'Strengthening Health Info Infrastructure' (2015); Eurobarometer surveys on 'Digital health literacy' and 'Data protection'; Taylor Wessing. 'Global data protection guide', access Mar 2018; OECD "Health Data Governance"; External interviews; A.T. Kearney analysis*

**Recognising the limitations of data sources, a number of initiatives – activities with a clearly defined purpose that utilise one or more data sources to achieve their objectives – have arisen across Europe.** These typically have one or more of the following aims (Figure 7)[9]:

- **Collect data that is not yet being collected** – this is typically done via novel approaches and technologies, for example with the 100,000 Genomes Project which is sequencing genomes from NHS patients with rare diseases and cancer to embed genomic medicine into clinical pathways
- **Standardise the ways in which data is collected** – this ensures that datasets are more comparable and can be linked to generate optimal insight; for example, the International Consortium for Health Outcomes Measurement (ICHOM) was launched in 2012 to support standardised measuring and reporting.
- **Improve access to existing datasets or allow their interrogation** – this can for example be done by enabling localised querying (e.g. InSite), or using simulated datasets based on patient characteristics to fully safeguard patient privacy (e.g. Simulacrum).

- **Collate existing datasets into a central repository** – for example, in the UK, the Haematological Malignancy Research Networks supports patient follow-up by incorporating Hospital Episode Statistics (HES), cancer registry and national administrative datasets.
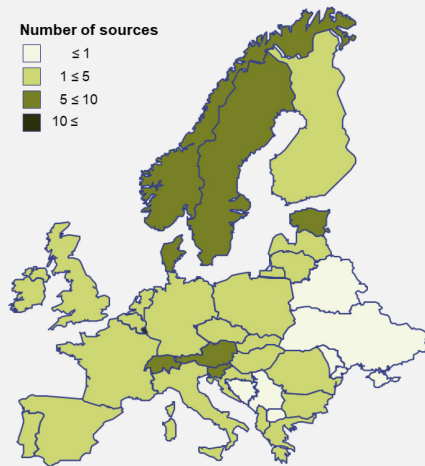
## INSIGHT BOX: ONCOLOGY DATA SOURCES IN EUROPE

**Figure 3. Distribution of known oncology data sources across Europe (absolute)**

Number of sources
- ≤ 10
- 10 ≤ 50
- 50 ≤ 100
- 100 ≤
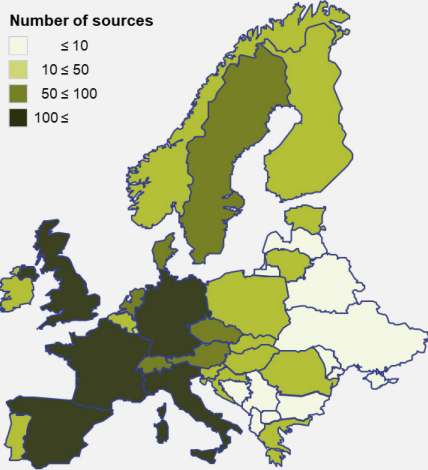
**Figure 4. Distribution of known oncology data sources across Europe, per capita (millions)**

Number of sources
- ≤ 1
- 1 ≤ 5
- 5 ≤ 10
- 10 ≤

**Figure 5. Distribution of data sources, by archetype**



**Large scale clinical registries**
Good source of valuable clinical data for high numbers of patients. Significant political will and investment required to expand beyond current scope. Access usually restricted to medico-scientific purposes though well defined

**Admin & claims**
Narrow focus that will always be limited in terms of data provision even if quality is higher than others; access is often well defined and protocolised

**EMR-linked databases**
Improved access to valuable data but requires investment in infrastructure and clinician buy-in; often more mature within primary care

**Research database**
The most common data source archetype but severely limited in value and scope. Access often possible for protocolised studies though funding is limited and can become a barrier to collaboration.

**Facilitated network**
Ability to bring the right data to the right people but requires time to develop before insight generation begins

60% *(Standalone)*
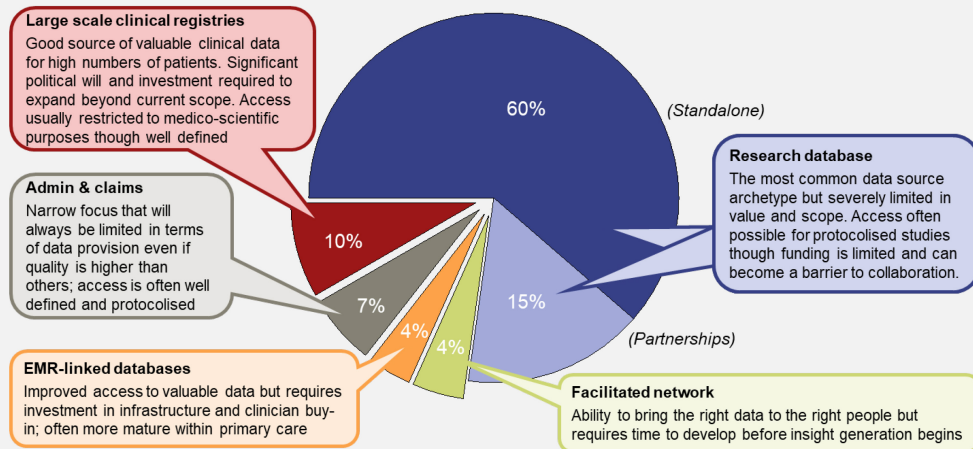10%
7%
4%
4%
15%
*(Partnerships)*

## Figure 6. Common characteristics of sources and ability to support applications, by archetype

| | | Research database (standalone) | Research database (partnerships) | Facilitated networks | EMR-linked source | Admin/claims | Large clinical registries |
|---|---|---|---|---|---|---|---|
| **Characteristics** | Access to source | Difficult | Variable | Good | Variable | Variable | Variable |
| | Funding (amount) | Insufficient | Sufficient | Sufficient | Sufficient | Sufficient | Sufficient |
| | Funding (duration) | Insufficient | Sufficient | Sufficient | Secure | Secure | Sufficient |
| | Coverage | Narrow | Moderate | Moderate | Narrow/ Mod. | Broad | Broad |
| | Depth of data variables | Moderate | Deep | Moderate | Moderate | Limited | Limited |
| | Quality of data | Poor | Moderate | Good | Moderate | Good | Moderate |
| | Latency | Moderate | Poor | Good | Moderate | Good | Poor |
| **Use Cases** | R&D enablement | Poor | Poor | Poor | Variable | Poor | Poor |
| | Healthcare context | Variable | Variable | Variable | Variable | Variable | Variable |
| | Treatment patterns | Variable | Variable | Good | Variable | Good | Variable |
| | Real-world clinical value | Variable | Variable | Variable | Variable | Poor | Variable |
| | Socio-econ. Value | Poor | Poor | Poor | Poor | Poor | Poor |
| | Pricing enablement | Poor | Poor | Variable | Variable | Variable | Poor |
| | Patient perspective | Poor | Poor | Poor | Poor | Poor | Poor |

### INSIGHT BOX: KEY DATA INITIATIVES IN EUROPE

## Figure 7. Types of data initiatives in Europe

| Improve Access | Improve Collation | Standardise Data | Collect New Data Types |
|---|---|---|---|
| *Aims to improve access to existing datasets or allow their interrogation* | *Aims to incorporate existing datasets into a central repository* | *Aims to standardise the ways in which data is collected so that datasets re comparable* | *Aims to collect data that does not yet exist, often via novel approaches* |
| • BD4BO<br>• CODE<br>• GOBDA<br>• HemoBase<br>• IMI Harmony<br>• INSITE<br>• PHEDRA<br>• POI<br>• Simulacrum | • Cancer Core Europe<br>• ECIBC<br>• ECIS<br>• EUROCARE<br>• HMRN<br>• ENCR<br>• EUCAN<br>• EUSOMA<br>• Greater Manchester Cancer<br>• IMI Protect<br>• Innovative Pricing Solutions<br>• I-O Optimise<br>• REAL Oncology<br>• Sarcoma BCB | • EHDN<br>• GA4GH<br>• GEKID<br>• FRANCIM<br>• Health Informatics Collaborative<br>• ICHOM<br>• OMOP Oncology | • 100,000 Genomes Project<br>• AURORA<br>• EUROSTAT<br>• CRISP<br>• IRONMAN<br>• OWise<br>• My Clinical Outcomes<br>• SCAN-B<br>• Universal Cancer Databank<br>• WEB-RADR |

## Figure 8. Use cases and objectives of initiatives in Europe

| | Better Access | | | | Better Collation of Existing Data | | | | | | | | | | Collecting New Data | | | | | | Std. | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Main focus / Additional** | BD4BO | INSITE | REAL Oncology | Simulacrum | CODE | ECIBC | ECIS | EUROCARE | GMO | HMRN | IMI PROTECT | I-O Optimise | REAL | UCD | CRISP | GE/100,000 Genome Project | IRONMAN | OWise | My Clinical Outcomes | WEB-RADR | GA4GH | OMOP |
| R&D workbench | ★★ | ★★ | | ★ | | ★ | ★ | | | ★★ | | ★ | | ★★ | | ★★ | | ★★ | | | ★ | ★★ |
| Healthcare context | ★★ | ★★ | ★★ | ★★ | | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | | ★ | ★★ | ★★ |
| Treatment patterns | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★ | ★★ | ★★ |
| Real-world clinical value | ★★ | ★★ | ★★ | ★★ | ★ | ★★ | ★★ | ★★ | | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ | ★★ |
| Socio-econ. Value | ★★ | | ★ | ★ | | ★★ | ★ | | | ★★ | | | | | ★★ | ★★ | | ★★ | ★ | | ★ | |
| Pricing enablement | ★★ | | ★ | ★ | ★★ | | | | | ★★ | | | | | | | | ★ | ★ | | ★ | |
| Patient perspective | ★★ | | | | | ★★ | ★★ | | ★★ | ★★ | ★★ | ★ | ★ | | ★★ | | | ★★ | ★★ | ★★ | ★ | |

**Of 40 data initiatives surveyed across Europe, 35% have a focus on improving collation, and 25% have a focus on improving access (Figure 7)**[9]. 38% have more than one objective, highlighting the need for comprehensive approaches to improve the health data landscape. Public or grant support, including European Commission financing and/or governance, is a source for 70% of initiatives. Although a therapy area focus is frequent, 30% of initiatives are disease-agnostic, 25% address cancer-specific issues overall, 23% cover more than one tumour type and 23% focus exclusively on one cancer type. These initiatives have brought much improvement to the European oncology data landscape, e.g. by promoting the use of the Observational Medical Outcomes Partnership (OMOP) model – a common data model enabling the comparison of data collected in different formats.

Similarly, the IMI is the world's largest life sciences PPP and has funded efforts such as Electronic Health Records for Clinical Research (EHR4CR), GetReal, the European Medical Informatics Framework (EMIF) and Big Data for Better Outcome (BD4BO), each of which supports further disease-agnostic and disease-specific efforts.

These initiatives address some, but not all, of the data source gaps. Over the course of more than 20 interviews conducted with data source and initiative owners, **more than a third reported fragmentation, governance, the lack of manpower and necessary data skills as problematic**[9]. A slightly lower proportion were preoccupied by the availability and time frame of funding, by the quality, coverage and granularity of data, and by access to data and its timeliness. In line with these ongoing issues, initiatives help meet demand for some applications of health data (i.e. healthcare context, treatment patterns and real-world clinical value), but supply remains low for less traditional applications such as socio-economic value, pricing and R&D enablement, and patient perspective (Figure 9).

*"There isn't even data sharing across the street, let alone across provinces and countries."*[9]

*"The biggest barrier is getting the right people with the right skills."*[9]

**Figure 9. Match between supply and demand for health data, by application**
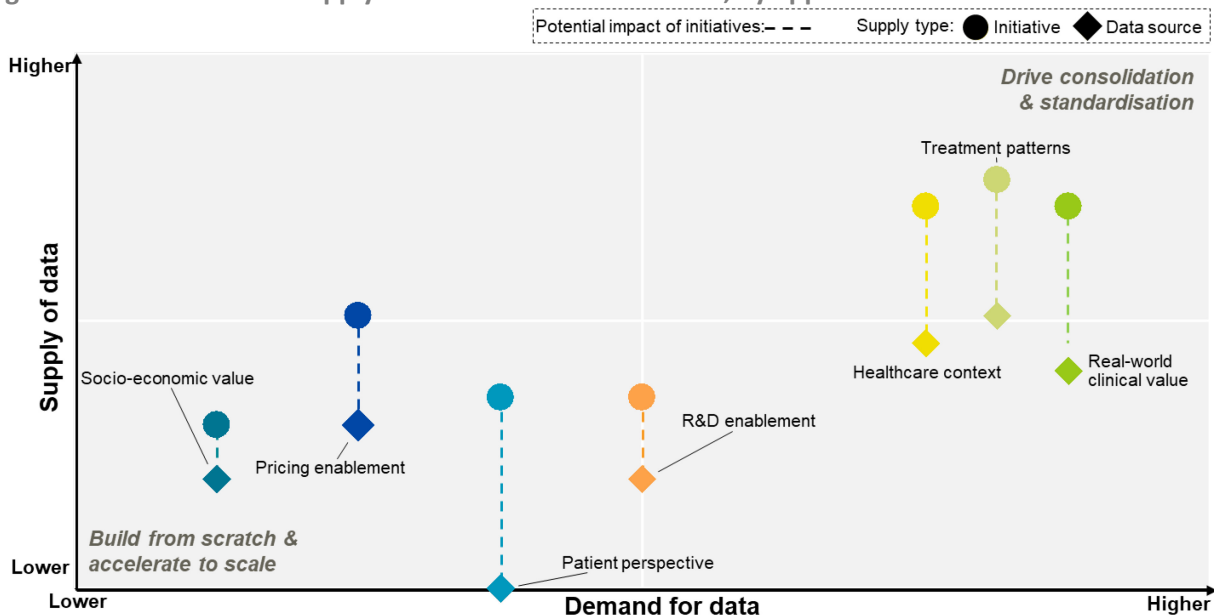


# 4. Ongoing Challenges in Oncology Health Data

The lack of demand for certain essential applications of health data (e.g. patient perspective) and the inability of data sources and initiatives to fully meet demand for others is underpinned by a number of challenges. These are not unique to oncology but may be particularly problematic in this therapy area for reasons including, but not limited to, the importance of genetics in determining treatment success, small patient populations and stigma associated with cancer. **Some issues are inherent to the data itself and the structure of the systems in which it is hosted, while others lie in the processes to collect and access it, the technologies in use and the stakeholders involved in health data** (Figure 10). The next section explores some of these issues.

**Figure 10. Current challenges with oncology health data in Europe**

| | |
|---|---|
| **Data** – requiring broader, deeper & interoperable datasets | • Most current data sources do not collect all the relevant information that is useful in oncology (e.g. biomarkers, ECOG scores, surrogate endpoints, etc.)<br>• Much of the data used to manage cancer is unstructured and coded in different languages<br>• Data quality varies across datasets, due in part to insufficient quality control mechanisms |
| **Structure** – needing a stable, open and supportive environment for data | • Few cancer plans and country policies explicitly target health data<br>• Member states have the ability to legislate locally for health, genetic and biometric data<br>• Funding for health data tends to be short-term and come with individual interests<br>• Linkage is key to enrich decision-making but remains limited for legal, societal and technical reasons |
| **Process** – progressively scaling up to world-class, transparent processes | • Access requirements differ widely across datasets<br>• Patient consent is not optimised and can both delay and limit the availability of data<br>• The timeliness of data availability and access is a significant problem<br>• Significant resources must be expanded to protect patient data and privacy |
| **Technology** – enabling solutions that were previously difficult | • The lack of interoperability between systems limits the ability to link different sources of health data<br>• Software and platforms are rarely user-friendly, limiting the ability to collect enough data<br>• Existing technologies may be outdated or likely to become so given new processing requirements |
| **People** – building skills & mindsets for involvement & sharing | • There is a lack of awareness and misconceptions undermine the full potential of health data<br>• Patient concerns around their privacy remain strong<br>• Healthcare providers remain concerned about the use and sharing of their patients' data<br>• There is a lack of qualified individuals to undertake the increasingly complex task of collecting data<br>• Vested interests and stakeholders' own agendas hinder collaboration in private and public settings |

## Data – requiring broader, deeper and interoperable datasets

**Cancer is a complex disease**, in which an improved understanding of patients' genotype and the impact on disease progression have been playing a growing role. Where only a few years ago treatment was based on the location (e.g. breast, lung, colon) and type of the tumour (e.g. sarcoma, leukaemia), drug regimens are now prescribed based on genetic mutations.

- Most current data sources – established years ago – do not collect patients' DNA, nor the **additional information that is useful in oncology** such as biomarkers, Eastern Cooperative Oncology Group (ECOG) performance scores and surrogate endpoints (e.g. PFS)[10].
- Recent years have seen some **initiatives**, such as the UK's 100,000 Genomes Project or US-based Flatiron Health, develop to collect this information for various cancer types, but these initiatives are in early stages.

**Much of the valuable information used in cancer management is unstructured** – it is not stored as coded inputs, but collected as notes, voice recordings, scans, histological stains, etc. that are virtually impossible to compare systematically across datasets using conventional technologies.

- These use **different languages** across countries and different coding standards (e.g. DICOM, WADO, HL7 are used across Europe[5]), if at all.
- **Machine learning** could help bridge this gap and there are some efforts to align coding standards – for example, the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) sets standards for genomic biomarkers as part of topic E15[11].

**Like unstructured data, structured data can be coded in different ways**, limiting the ability to compare datasets across sources and countries.

- France, Germany and Austria use **ICD-10** for diagnosis, Denmark and Finland also use **ICPC** and **ICPC2**, and Belgium and the UK use ICD-10 but also **SNOMED**[5].

- The World Health Organisation (WHO) is publishing guidelines to encourage wider adoption of the International Classification of Disease (ICD) **coding framework**[5].
- **Federated querying** can pool comparable available data in different formats, and **open, cloud application programming interfaces** (API) like that launched by Google in 2018 can help manage medical datasets in multiple standards[12].

**Further limitations of data collection protocols further limit the relevance of data collected**, though there are ongoing efforts to overcome this.

- The exact **data differs** across and within countries, with only some Member states (e.g. Belgium, Poland, Spain) defining minimum datasets for their public databases5.
- In some cases, the data collected lacks the requisite **granularity**: for example, registries tend to collect only first-line treatment[9].
- In the US, the FDA developed a **guideline on biomarkers** in 2005, and its ICH E15 version defines coding standards for biomarkers[11].

**Data quality varies across datasets**, due to incomplete coverage (i.e. when data is not or partially entered for patients) or inaccuracies (i.e. when data is erroneously entered).

- **Current quality control mechanisms are not sufficient to address this**: there is often unclear responsibility for quality assurance between the EU and its Member states, and many countries do not have specific legislation on data quality (e.g. Bulgaria, Estonia, Greece) or mandatory quality audits of EHRs (e.g. France, Germany, Netherlands, Sweden)[5].
- Some countries have recognised this as an issue, with the NHS England recommending a **clinical coding audit** to take place every 12 months, and countries like Iceland and Estonia implementing EHR audits[5].
- In Belgium, the government has **incentivised healthcare providers** with almost €900 per head to subscribe to an EHR system with decision aids and categories to help enter the right data[4].

> *Improvements in overall data granularity, quality and interconnectivity are particularly necessary in oncology due to the increasingly stratified nature of the disease. Guidelines and new technologies are paving the way to improving this, but further alignment and understanding of the use of health data to improve care, incentives to build this data and scaling ability are still required.*

## Structure – needing a stable, open and supportive environment for data

**The lack of clear political will and direction across European, national and regional health structures has led to the development of numerous, uncoordinated sources** – these emerge to fit individual, ad hoc needs, often lacking clear standards and duplicating work done elsewhere (due to the lack of visibility of many ongoing efforts).

- There are no European cancer plans, and most **national cancer** plans barely touch upon health data, though some – like the French Plan Cancer 2014-2019 – do[13].
- In **regional health systems**, individual cancer plans and data policies vary widely, if they are even defined, leading to disparities in health data use across regions: for example, Lombardy is leading the way in Italy, and Catalunya and Andalusia in Spain)[10]
- Even with the GDPR aiming to harmonise data privacy laws across Europe, Member states retain the **ability to legislate locally for health, genetic and biometric data**, which is unlikely to lead to European alignment[22].
- **EU-wide efforts**, such as the European Patient Smart Open Services (epSOS, launched from 2008-2014) or Cancon (2014-2017, co-funded by the EU Health Programme to improve the

quality of cancer control, including cancer data, across Europe) have been launched; some can be short-lived and add to an already complex landscape, while others like epSOS become more widespread and underpin new best practice[14,15.]

Health data infrastructure is not available across all settings of care and European countries, and where it is, it often needs to be updated, requiring significant resources to update and maintain.

- **Funding** continues to be an issue for many data sources, initiatives and innovations, which struggle to attract and retain neutral, long-term funding – 14% of healthcare providers see funding as the main eHealth challenge they face, reaching ~30% in Ireland, Austria and the UK[16].
- This is primarily due to the **short-term nature of investment**, with funds typically running out after a set number of years, uncertainty around which entity then steps in, and possible loss of the longitudinal value of data if the initiative ends: for example, funding for EUROCARE stopped in 2018 and the researchers are writing up the final manuscript[9].
- Investment often comes from **several sources**, each with their own objectives – this is the case for EHR funding for the NHS England, which is provided by the Integrated Digital Care Fund, Nursing Technology Fund, the NHS Innovation Scheme and Vanguard sites[17].
- **Processes** to obtain funding may be unclear and complex, requiring time or even a government partner to understand how granting agencies evaluate proposals[5].
- Some centrally-funded initiatives **cannot apply for external funding**[9].

At a more granular level, **the data infrastructure is such that linkage is essential to enrich insights gathered from data but remains limited for societal and technical reasons**. Indeed, many are concerned about the ability to identify patients by connecting sufficient information about them, possibly leading to stigmatisation by peers or profiling by insurers. As a result, many countries (e.g. Germany, Portugal, Norway) report limited linkage of databases[5].

- This can be due in part to the **lack of single identifying numbers** to link relevant data, or to their limited scope: in France, identifying numbers used by hospitals can vary across hospitals and are different from those used for medical insurance[5].
- In addition, in many countries **linkage must be allowed on a case-by-case basis** by a dedicated authority (e.g. the Privacy Commission in Belgium), or by national or local legislation (e.g. the law in the state of Bremen in Germany authorises linkage)[5].
- Nonetheless, many countries successfully use single identifying numbers, providing comprehensive insight into patients' health that extends beyond traditional data: the Swedish 'personnummer' is unique in being used for purposes as diverse as tax, social welfare, health care, living conditions and education[5].
- In 2014, the **eIDAS regulation** was approved by EU co-legislators to ensure that people and businesses can use their own national electronic identification schemes to access public services in countries where EUid are available, and creates an European internal market for time stamps and other means of authentication[18].

*Moving forward, a new approach is required that focuses on providing sustained funding and enabling the linkage of datasets to provide a 360 view of disease, treatment dynamics and the patient experience, while respecting the need for confidentiality. Local legislation and infrastructure may limit the ability to do this, but examples of where this has been successful can challenge established mindsets and open up systems to the use of RWD.*

## Process – progressively scaling up to world-class, transparent processes

**Processes to collect, use and share data often require clear justification to protect patients**. In the context of the GDPR, this rationale has been more clearly explored but can still pose challenges.

- The **use of RWD without patients' consent** is justified on medical care and public health grounds, but how this is interpreted locally and which data users may qualify for these uses will vary, possibly leading to fines[22].
- **Secondary use of data** (i.e. use of data not explicitly collected for this use) is typically acceptable in the context of scientific research, but is currently regulated locally: some countries may accept other objectives for secondary use, such as Belgium where data can be used for historical and statistical analyses, but these are not the norm[6].
- Many data initiatives feel that they **sufficiently cover these grounds,** and should therefore be able to continue using their data, but only time will tell whether that is true and the data remains available[9].
- As part of its Public Sector Information proposal, the European Commission aimed to increase the availability of data by bringing new types of public and publicly funded data into the scope of the Directive, and provide that data already available in "open access" research data repositories should be re-usable for commercial and non-commercial purposes[19].

**Processes surrounding health data are often unclear, time-consuming and restrictive**.

- **Access requirements** differ widely across datasets, limiting stakeholders' ability to readily consult multiple sources by re-using similar materials and applications, and various bodies may be required to approve access – in Germany, 16 regional data protection agencies review data access protocols[10,9].
- Although most European databases are accessible to academics upon request, some data sources only grant access via **third parties** or offer **limited access to private entities**: in France, the Programme de Médicalisation des Systèmes d'Information (PMSI) hospital claims databases can only be accessed via a neutral third-party provider[10].
- The **more stakeholders** are involved in an initiative, the more cumbersome the process; different governing members may also be more conservative than others within a single initiative[9].
- In some cases, requirements and processes are so complex that initiatives had to **stop using data** due to changes in third-party access requirements[9].

**Patient consent remains at the core of data collection, but it is not optimised and can both delay and limit the availability of data**.

- Only 13 of the 28 EU countries have **specific rules regulating patients' consent** for EHRs, while frameworks and best practice tend to remain local[6].
- **Consent processes differ widely**, are often unclear for patients and can be quite complex: consent forms for research can range from three to 30 pages, with an average readability suitable for a college graduate[20].
- **Opt-in consent management solutions**, **user-friendly videos** and other tools have been used to facilitate the consent process, allowing patients to have a better view of the data they offer and how this is used for different applications; in the Nordics and Belgium, non-sensitive identifiable personal data can be made available to researchers without prior consent for research purposes, accelerating access[4].
- The GDPR sets out to delineate **stronger and clearer conditions for consent**, but if not done properly, this could also increase the amount of information that patients are required to consider, understand and agree to.

Across European countries and data sources, **the timeliness of data availability and access is a significant problem – it can take up to four years to obtain data**, even though cancer evolves on a daily basis and time is of the essence.

- Contract signing, scientific and ethical approval are often slow, requiring multiple parties' review.
- Depending on databases' and studies' set-up, scope and software used, data can take months to be collected and cleaned up, let alone accessed by those having followed the right process: for example, one national cancer registry's ethics and access agreements currently take more than six months, with data that is more than 18 months old due to requirements for consolidation and learning[9].
- To enable optimal decision-making in oncology, timely information is critical – to this effect, initiatives like the Collaboration for Oncology Data in Europe (CODE) aim to provide real-time data for treatment decision-making and new payment models.

Beyond ethics, consent, analysis and access processes, **significant resources must be expanded to protect patient data and privacy**.

- Patient data can be **de-identified**, but this is not infallible; full anonymisation may be necessary but can be challenging, requiring multi-stage de-identification with clear governance and controls approved by relevant authorities[9].
- **Data aggregation** helps overcome some privacy concerns but may not provide the granularity of information required for some decisions.
- If accepted by decision-makers, **simulated data** based on real patient characteristics could provide a way through and inform data-querying models and decision-making, though this would be limited in scope to hypothesis generation and methods validation in the near future – Simulacrum, a collaboration between Public Health England, Health Data Insights, IQVIA and AstraZeneca, provides simulated data modelled from the Cancer Analysis System.
- **New technologies** can also foster better data privacy, but remain in their early stages: for example, blockchain (a list of data blocks secured by complex codes and accessed via a transparent ledger) can support a clear audit trail across a secure platform with decentralised ownership; a number of start-ups are already operating in this space: in 2017, Estonia's eHealth Authority signed a deal with Guardtime to secure the health records of its citizens[21].

In line with the GDPR, **data protection must increasingly be built into data initiatives** in order to avoid severe penalties.

- This includes, for example, appointing **data protection officers**, conducting data **protection and impact assessments**, and more systematic **breach reporting**[22].
- These steps to anonymise and protect patient data therefore require **time and funding** to employ the right personnel and follow robust processes.
- As a result, **fines** for data breaches or failure to comply with the law are becoming more commonplace: the GDPR allows for data processors and controllers to be fined up to €20 million or 4% of total annual worldwide turnover for GDPR breaches[22].
- Overall, stakeholders feel that the **balance between bureaucracy and deriving insights** from data is not adequate[9] – this is unlikely to improve given the new GDPR requirements, but GDPR will raise the bar and the quality of data sources that are able to comply.

> *Processes are already complex and burdensome, and likely to become so given the understandable GDPR push for transparency and better use of health data. Better planning and systematic efforts to put patients at the centre of data collection and use in a user-friendly way, most likely using new technologies, represent the best options to simplify this environment.*

**INSIGHT BOX: GDPR[23]**

The GDPR is the new legal framework in the EU that aims to:
• Harmonise data privacy laws across Europe
• Protect and empower all EU citizens
• Reshape the way organisations across the region approach data privacy

The regulation came into effect on 24th May 2016, but took full effect on May 25th 2018, replacing the Data Protection Directive 95/46/EC. It establishes minimum mandatory requirements across the EU but provides some ability for Member States to legislate locally on certain discrete matters, including the use of health data.

The GDPR applies across four main areas, entailing:

| | | |
|---|---|---|
| **A** Clarification of data definition & rationale for use | 1 | Clear justification needed to process health data |
| | 2 | Restriction of automated decision-making, including profiling |
| | 3 | Definition of more types of health data as sensitive (inc. genetic & biometric) |
| **B** Expanded monitoring & liability | 4 | Increase of codes of conduct & certifications |
| | 5 | Application of GDPR to more stakeholders |
| | 6 | Stronger data protection agencies |
| **C** Strengthened individual rights & consent | 7 | Clarification of individual rights for data subjects (to access, to rectification, to data portability) |
| | 8 | Clarification of individual rights for data subjects (to be forgotten, to restrict processing, to object) |
| | 9 | Additional info. required to explain context for use ("transparency & fair processing") |
| | 10 | More stringent definition of consent |
| **D** Processing accountability & compliance mechanisms | 11 | Qualified compliance framework & derogations for scientific research |
| | 12 | Stronger data protection & impact assessments |
| | 13 | Mandatory data breach reporting |
| | 14 | Mandatory appointment of data protection officers |
| | 15 | Accountability & increased reporting of processing |
| | 16 | Higher threshold for anonymization |

Several actions can be taken to ensure GDPR compliance and to limit potentially negative impacts of the new regulation on oncology health data:

| Internal | Review & adjust | • **Review & adjust processes** to ensure compliance & best practice<br>• **Dedicate resources to support new requirements** (e.g. individual rights, data protection impact assessments, documentation of reporting, etc.) |
|---|---|---|
| | Consult & discuss | • **Work with local politicians & regulators at the European & international levels**, to establish clarity around specific terms & limit risk / fines while maximising ability to collect & use data<br>• **Set clear criteria & ongoing documentation** for interpretable elements (e.g. "scientific research", "disproportionate effort", etc.)<br>• **Establish ongoing consultation with local & European data protection agencies** to test feasibility & ensure research can continue<br>• **Obtain legal advice on an ongoing basis** for data initiatives to ensure compliance |
| | Inform & train | • **Issue joint information statement / Q&A on the GDPR** to explain its content & impact on health data<br>• **Collect case studies of health data on patient outcomes & GDPR impact on data collection**, to sensitize all relevant stakeholders to the importance of health data (inc. impact on historical data)<br>• **Develop standards, templates & trainings** for DPIAs & DPOs |
| External | Collaborate & partner | • **Collaborate with all relevant stakeholders to develop health industry code** of conduct<br>• **Partner with patient associations** to ensure their interests are respected & supported in applications of the law & derogations at the national level, & to develop consent / information forms that best address legal & patient requirements<br>• **Have EU & non-EU processors collaborate** to enable best practice sharing & ensure compliance with new rules<br>• **Investigate new technologies** for anonymization & analysis to protect patient data & privacy |

*This does not constitute legal advice. Legal counsel should be sought to ensure GDPR compliance.*

## Technology – enabling solutions that were previously difficult

Technology provides unprecedented opportunities to generate more, higher-quality health data, and share it with all the relevant parties, but barriers remain well beyond oncology. **The lack of interoperability between systems limits the ability to link different sources of health data**.

- In the UK, there are more than 100 **commercial suppliers** of EHR software, let alone for other sources for health data[17].
- In France and Spain, most hospitals **develop their own, fit-for-purpose software** with limited intent to connect with other databases[10].
- Across the EU, only 13 countries have set up **specific rules on interoperability** (e.g. Austria, Belgium), and only six for cross-border interoperability (e.g. Spain)[6].
- In 2018, the European Commission announced the development of technical specifications for a European EHR exchange format, to enable the European data space[24].
- Numerous initiatives such as O-Wise and the Haematological Malignancy Research Network (HMRN) have been established to **enable linkage**, but these are still not the norm[9].

**Software and platforms for health data are rarely user-friendly, limiting the ability to collect sufficient high-quality data**.

- Out of 38 papers on EMR implementation, seven listed **ease of use** as a main barrier[25].
- Interviews with oncologists in France, Italy and Spain emphasized the **complexity, low user experience, and high requirement for manual processing** across their hospitals' EMR systems[10].
- To bypass this, successful data sources and initiatives often employ **dedicated technicians** for data entry[10].
- Recent years have also seen improvements in the quality ratings for US-based EMR interfaces and visual appeal, suggesting a democratisation of these software and a growing focus on the users' experience[26]; this could potentially increase adoption in Europe, as well.

**Existing technologies may be outdated or likely to become so in view of growing analytical and processing requirements**.

- Some adjustments will be required – between 100 million to two billion human genomes could be sequenced by 2025, **requiring 2-40 exabytes of storage capacity** and processing that is six orders of magnitude faster than is possible today.
- Technology is already stepping in to meet this gap: for example, **cloud computing solutions** can be used for large-scale analysis and storage of health data, enabling continuous coordination of patient-care and seamless integration with health systems; **machine learning** can help automate part of the data entry process.
- Awareness, understanding and mastery of these technologies currently remains limited to more advanced **IT and digital companies**, highlighting the need for increased availability, remuneration and training in these skills.
- Updating infrastructure is costly but can be preferable to a complete system overhaul that could jeopardise hospitals', cancer centres and general practices' continuous flow of data – the **balance between 'building from scratch' and incrementally upgrading** will depend on the systems, capabilities and requirements of each data stakeholder[10].

> *GDPR will raise the bar in terms of the quality and management of health data, with requirements likely driving professionalisation of data collection activities. Larger datasets will emerge from this, bolstered by new technologies that must be embraced, and their use accelerated.*

## People – building skills and mindsets for involvement and sharing

**There is a clear lack of qualified people to undertake the increasingly complex and comprehensive task of collecting and analysing data.**

- Although some initiatives provide specific training for employees, the **lack of data scientists** and professionals qualified in new technologies (e.g. analytics, machine learning) is the top issue for 7% of European healthcare providers[16], and is strongest in the public sector.
- Many healthcare professionals have **limited digital literacy or training in data collection**, in addition to having numerous other responsibilities – in the UK, poor staff engagement and training led to the Cambridge University Hospital Trust reverting to paper records after an attempt to roll out 2.1 million EMRs in 2014[17].
- More simply, **limited manpower** can also be a challenge: data collection and analysis are resource-intensive, and availability of staff can limit data initiatives' ability to scale up.
- Dedicated courses and degrees on analytics, data sciences and digital health skills are emerging across most European universities, with companies like IBM setting up partnerships with universities and funding for more data science training[27].

**As health data becomes more of a business, vested interests and stakeholders' own agendas may sometimes hinder collaboration across both private and public settings.** Data sources and initiatives, which also spent a lot of time and effort collecting and cleaning up data, can also be reluctant to share it.

- Private pharmaceutical, medical device, biotech and/or technology companies protect their **commercial interests**, but health insurers – including state-owned ones like Germany's statutory health insurances – and other publicly-funded entities also limit the potential use of their data.
- Although the GetReal melanoma case study was funded by EFPIA, EMA, the UK's National Institute for Health and Care Excellence (NICE) and the Dutch National Health Care Institute (ZIN), some participating registries restricted access to enable their PhD students to **publish their theses** with that information[9,10].
- In the Netherlands, the Dutch Upper GI Cancer Group has established a process to enable sharing with any who ask for data: a **committee reviews applications** to access their data, whose members can oppose access, but this rarely happens and the data is readily shared[10].
- Recognising the value of this data, pharmaceutical companies are increasingly **partnering with and acquiring data sources**, as demonstrated by Roche's acquisition of FlatironHealth, an oncology-focused EMR company[28].

Beyond commercial interests, **healthcare providers remain concerned about the use and sharing of their patients' data**, and may not be aware of ongoing data initiatives or mechanisms in place to protect patient confidentiality.

- In the UK, the NHS's care.data scheme – designed to unify patients' care across general practices and hospitals into one central database – was postponed and subsequently **cancelled** due to physicians' opposition to privacy and consent issues[29].
- In France, the Dossier Médical Partagé – an initiative to ensure every French patient has a medical record – reached 400 thousand records within two years, well below the objective of 500 thousand records within one year; this was primarily due to the lack of awareness or campaigns geared towards physicians[4].
- Recognising lack of health professional engagement as a barrier, several governments are **partnering with trusted entities** and collaborating more closely with physicians to foster better buy-in in data initiatives: the Belgian government is collaborating with Custodix, a trusted third party EMR vendor with a strong reputation for data hosting and transfer, thereby inspiring trust and reducing resistance towards collection of health data[30].

**Patient concerns around their privacy remain strong, particularly given recent scandals and data breaches**.

- Only 38% of EU patients believe that healthcare providers offer **effective data security**, and many fear that their data could be used for profiling by insurers (e.g. to increase premiums)[31,33].
- As a result, numerous efforts to collect patient data meet **continued opposition** or have failed: the introduction of health e-cards was delayed over 15 years in Germany[32].
- In the Netherlands, a publicly-funded initiative to that aimed to build a national EHR system to facilitate patient-level information exchange between care providers failed due to opposition from patient groups around data **privacy issues** during information exchange[4].

Until the public and patients specifically can fully envision the **benefits of collecting and using health data**, and be assured that it will be well protected, the health data landscape will continue to have mixed perceptions within what must be its most important supporter group. European citizens and patients are also taking a growing role in the generation and utilisation of health data, which provides unique opportunities to understand real-life health events, choices and perspectives.

- **Lack of awareness and misconceptions undermine the full potential of health data:** even in healthcare, many individuals cannot readily point to the benefits of health data[10].
- Patients are rightfully concerned by the **use of their data** where there is no clear public benefit and solely commercial motivation, but 60% of UK patients would rather grant access to their data to commercial entities than miss out on benefits deriving from their innovation[9,33].
- Recent **research and campaigns** such as #datasaveslives are attempting to fill that gap, but given the omnipresence of data across healthcare systems, much more could be done to quantify its impact on patients and institutions[3,34].
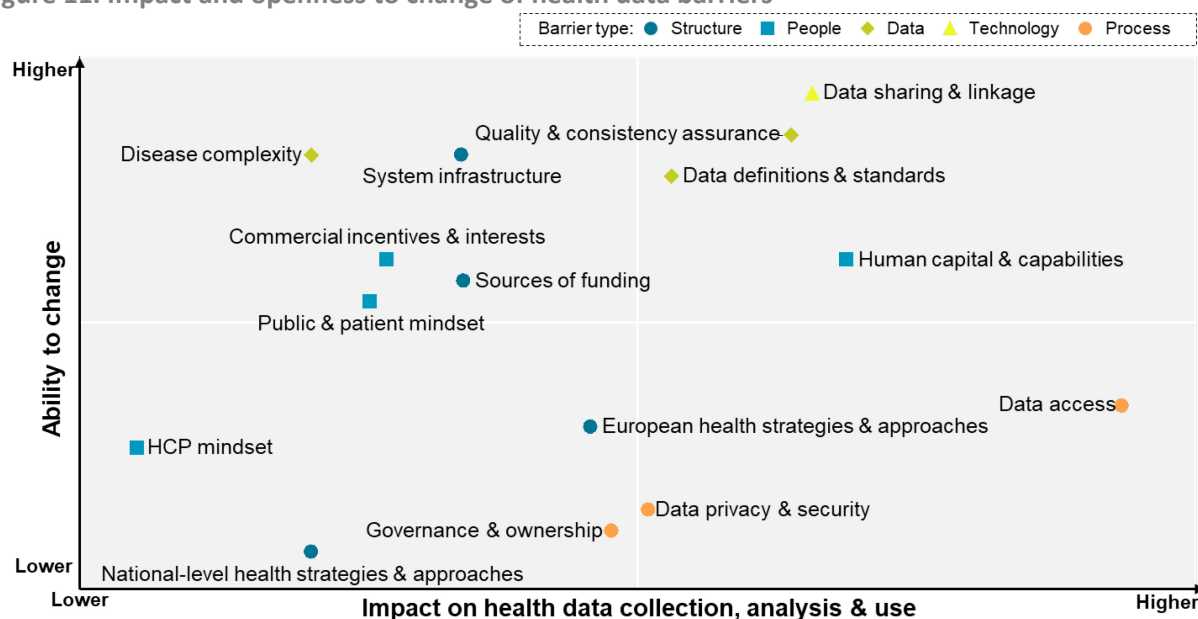
> *The lack of data science skills and ability to retain talent in healthcare currently limits the quality of RWD available, which is particularly challenging for oncology given the complexity of the disease. Mentalities are changing, however, and with better information and communication around health data, cancer communities can play a greater role in sharing data and ensuring patients benefit.*

## Conclusion

Although many of these challenges are significant, several (e.g. technologies, mindsets) have seen positive changes in recent years and have been overcome in various countries. Others remain that may require limited effort to improve rapidly, for example around establishing quality assurance or the right data infrastructures (Figure 11). Still others, such as data access, data privacy and security, and European health strategies, are anchored in processes or legislation and may take years to evolve. On the whole, however, **there is greatest potential in tackling the challenges that are strongest but also more open to change, such as data definitions, standards, sharing and linkage, and building capabilities** – this will require long-term, forward-looking collaboration across all stakeholders who stand to benefit from health data.

**Figure 11. Impact and openness to change of health data barriers**



Barrier type: ● Structure  ■ People  ◆ Data  ▲ Technology  ● Process

Higher

Ability to change

▲ Data sharing & linkage

Disease complexity ◆
Quality & consistency assurance ●
● System infrastructure
◆ Data definitions & standards

Commercial incentives & interests ■
■ Human capital & capabilities
■ ● Sources of funding
Public & patient mindset

Data access ●

■ HCP mindset
● European health strategies & approaches

● Data privacy & security

Governance & ownership ●
Lower
● National-level health strategies & approaches

Lower                                                                 Higher
**Impact on health data collection, analysis & use**

# 5. Next Steps for a Better European Health Data Landscape

A recent study analysing 230 policies on data sharing showed that although there was some overlap on restricted policy themes (e.g. privacy, autonomy), there was no alignment on their importance or how they should be addressed[35]. This fragmentation is characteristic of most aspects of health data across Europe. **To significantly and permanently impact the health data landscape, solutions must be developed jointly by patients, public and private entities**. Different collaborations between these stakeholder groups will be required to meet different challenges. Initially, awareness of the importance of health data must be developed to drive support. Standards and infrastructure can then be built to ensure a robust environment in which the data can be collected, analysed and used. Lastly, the right skills and behaviours must be fostered across and beyond settings of care, to realise optimal value from oncology health data.

## Build awareness

Given the limited awareness and misconceptions around the current status of health data in Europe, **a first step to strengthen the position of health data will be to ensure that citizens, patients and all healthcare stakeholders buy into the benefits of sharing and using oncology data**. Through information campaigns and communication, Europeans can get a better sense of the role and implications of health data. **This will pave the way for recognition of data science as a core health skill**, as governments and education systems continue in their evolution to maximise the use of digital and analytics. National and local efforts will be required to ensure that the value of data science reaches all healthcare settings, and is disseminated beyond private companies, into the public space.

At a more specialised level, **health stakeholders will need to strengthen their understanding of the rapidly-evolving requirements of oncology**. This means having a full, up-to-date and flexible understanding of the technologies that can enhance health data – expanding beyond the general knowledge that technology can solve many problems, into the practical, solution-oriented testing of technologies in oncology health data. In addition, **with the growth of indication-specific and**

**combination therapies and cancer, innovative pricing models will increasingly become a necessity**. Though these are currently poorly-understood and can be burdensome for those implementing them, these outcomes-based agreements have a role to play in supporting value-based healthcare systems. They must therefore be tested and rolled out more extensively than they currently are, in order to support the sustainability of innovation and increasingly complex care.

## Develop standards

Since the emergence of "real-world data" as a term more than ten years ago, practical experience and expert opinion have repeatedly alluded to the lack of standards. Although the situation has improved thanks to various initiatives, some issues remain**: it is therefore critical to define clear guidelines and best practice to work with health data**, including anonymisation and privacy protocols, minimum datasets and linkage (in particular for oncology

> *"What would be very helpful is them to be open about the success and failures of those initiatives. There are [I imagine] a lot of initiatives that perish and there needs to be some collective responsibility for those initiatives perishing. Because, there are a lot of people trying to do good things, the money runs out, the energy and enthusiasm runs out, and, we end up with essentially, the fruit dying on the vine."[9]*

data), quality control mechanisms, access protocols, and many others. No single, one-size-fits-all solution can be provided, but a comprehensive view of the different options and the circumstances in which they work will go a long way in improving data handling and providing clarity around what is done. On that basis, **a data quality accreditation framework could be used to support the implementation of best practice and the recognition of data sources that offer high-quality data** in a timely and compliant manner. Co-developed by all health stakeholders, this could eventually accelerate approval processes for accredited stakeholders who can demonstrate robust collection and/or usage processes.

**As patients continue to take ownership of their health data, stakeholders must do all in their power to foster the transparency and ease-of-use of patient consent processes**. Without this, not only are both patients and providers burdened, but the range and quality of potentially life-altering data will suffer. Moreover, **patient-reported outcomes and other measures providing an insightful view of patients' experience of cancer (e.g. surrogate endpoints) must be more strongly defined** and their acceptance supported. In order to ensure their relevance and measurable impact on patient outcomes, **both PROs and socio-economic measures of benefit should be tested in different countries, cancers and settings of care, then harmonised and established as standards**. Much will be needed to ensure that these measures reflect the reality of patient care and health systems, thereby enabling the right health and non-health data to be collected.

## Build infrastructure

The right structures and environment must be established to support the collection and use of oncology RWD. **This entails establishing a consistent approach to govern, fund, manage and scale healthcare data projects**, across countries and health stakeholders. **Part of this will stem from having aligned EU and national grant that foster high standards and collaboration around health data**, as well as enabling data sources to scale with the right resources, capabilities and protocols. Another critical factor will be the interpretation of GDPR, which on health, genetic and biometric data remains subject to local legislation: **stakeholders and policy-makers should work together to support local GDPR implementation that supports the use of health data** for patients' and society's benefit, instead of restricting the availability, quality and/or handling of data.

Given the numerous efforts to improve cancer management and initiatives dedicated specifically to oncology data, **there is a strong rationale to enable the collaboration of cancer experts across and within countries** (e.g. via European Reference Networks supported by the European Commission, one of which is being considered for oncology[24]). Currently much expertise around oncology health data remains unshared, both due to a lack of dedicated forums but also because of limited visibility of dedicated interventions. Hence, beyond websites and conferences, new spaces – physical or virtual – could be established for these experts to disseminate best practice. In addition, **an open-source, up-to-date, comprehensive catalogue on oncology RWD sources in Europe would provide much-required transparency on the landscape**: in the first instance, this could include the location, participating entities, cancer types and endpoints covered, access requirements and key contacts. Eventually, this could evolve to incorporate fuller assessments of data quality, timeliness, ease of access and collaboration, and other metrics. Many stakeholders currently have partial views of these data sources, but an all-encompassing and regularly-update catalogue could facilitate communication and overall alignment.

Beyond supporting structures, dedicated data structures can also open up important discussion and pioneer new ways of approaching health data. Specifically, **patients should be supported in owning, sharing and benefiting from their data, for example via data platforms**. There are precedents for such platforms, such as the Swedish 118 records and a more recent initiative, the Universal Cancer Databank, where oncology patients can donate their information. Similarly, **a platform for raw, consolidated but de-identified data to be collated and interrogated could provide meaningful insights to all stakeholders**, while protecting data privacy – linkage and sharing would be enabled within agreed boundaries, to derive the most from this data.

### Develop skills

Underpinning most evolutions are skills – capability gaps can truly enable or hinder some of the greatest innovations, and have been a rate-limiting factor for health data. **It is therefore critical to develop key data skills across industries and sectors of care**, to ensure that patients, healthcare providers, innovators and scientists across Europe are equipped to make the most of health data. This could be accomplished via dedicated campaigns, educational programmes, online education and apprenticeships. **This, as well as increased process automation and the development of appropriate incentives, could also facilitate the collection of complete, high-quality data by healthcare professionals and other stakeholders responsible for entering data**. Joint efforts should also be considered to support the preparation of regulatory-compliant RWD through early advice processes and clear guidelines. This could be led by EMA and innovators, to ensure that practices around health data improve and have a measure impact on health decision-making in Europe.

### Conclusion

Recent years have seen incredible advances in the creation, collection, analysis and use of oncology health data across Europe – as well as better decision-making and the improved patient outcomes that ensue. Although much is already being done to improve this landscape and overcome some of the persisting barriers, many efforts lack scale or are insufficiently targeting the more systemic issues around the data itself, processes, technology, skills and mindsets. For example, there is a need for stakeholders to agree on how data can be used to enhance and go beyond current treatments, such as in innovative approaches to tracking treatment use that can change reimbursement approaches. Only by working together to build awareness and infrastructure, to foster alignment across policies and to develop standards and skills, can health data truly be used to its full potential and transform the way in which cancer care is delivered.

# 6. Acknowledgements

# 7. References

1   PhRMA, A Decade of Innovation in Cancer, 2016
2   Bengt Jönsson et al., Comparator Report on Patient Access to Cancer Medicines in Europe Revisited, 2016
3   RAND, Understanding value in health data ecosystems: a review of current evidence and ways forward, 2014
4   RAND, Health and healthcare: assessing the real-world data policy landscape in Europe, 2014
5   OECD, Strengthening Health Info Infrastructure, 2015
6   European Commission, Overview of national laws on electronic health records, 2013
7   WHO, Global eHealth survey, 2015
8   IQVIA, RWD catalogue, accessed April 2018
9   IQVIA, 21 expert interviews and desk research on health data sources and initiatives, conducted April 2018
10  A.T. Kearney, 37 expert interviews and desk research on barriers and trends in health data, conducted April 2018
11  NCBI, Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine, 2013
12  Digital Health, Google launches cloud healthcare API, available from https://www.digitalhealth.net /
13  Institut National du Cancer, Plan cancer 2014-2019 : priorités et objectifs, 2015
14  ePSOS website, accessed April 2018
15  Cancon website, accessed April 2018
16  HiMSS, Annual eHealth survey, 2017
17  House of Parliament, Electronic health records, 2016
18  European Commission, Trust Services and Electronic identification (eID), available from https://ec.europa.eu/digital-single-market/en/trust-services-and-eid
19  European Commission, European legislation on the re-use of public sector information, available from: https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information
20  Larson Elain et al., Reading Level and Length of Written Research Consent Forms, 2015
21  Guardtime, Estonian  eHealth Authority Partners with Guardtime to Accelerate Transparency and Auditability in Health Care, available from www.guardtime.com
22  European Commission, GDPR portal, available from https://www.eugdpr.org/

23   Hogan Lovells, The Final GDPR Text and What It Will Mean for Health Data, 2016; Noerr, Taking advantage of patient data – an outlook on the upcoming General Data Protection Regulation, 2017; BHBIA, 'GDPR Quick Guide' and 'GDPR Legal Grounds for Data Processing', 2017; IGA, The GDPR: What's New'

24   European Commission, Communication on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society, 25 April 2018

25   Gesulga Jaillah Mae et al., Barriers to Electronic Health Record System Implementation and Information Systems Resources: A Structured Review, 2017

26   Healthcare IT News, 2015 Satisfaction Survey results, available from: http://www.healthcareitnews.com

27   Lee Gardner, IBM and Universities Team Up to Close a 'Big Data' Skills Gap, available from: www.chronicle.com

28   Roche, 'Roche to acquire Flatiron Health to accelerate industry-wide development and delivery of breakthrough medicines for patients with cancer', 2018

29   NHS England, care.data website, accessed May 2018

30   The European Institute for innovation through health data, 2016

31   European Commission, Special Eurobarometer 431: Data Protection, 2014

32   Stafford Ned, Germany is set to introduce e-health cards by 2018, 2015

33   Wellcome, 'Public attitudes to commercial access to health data', 2016

34   Health e-Research Centre, #datasaveslives, available at: https://www.herc.ac.uk

35   Alessandro Blasimee et al., Data-sharing for precision medicine: policy lessons and future directions, 2018