Clinical Oncology 37 (2025) 103657

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Danish and Swedish National Data Collections for Cancer — Solutions for Radiotherapy

C.E. Olsson [*†], S.L. Krogh [‡], M. Karlsson [§], J.G. Eriksen [¶‖], T. Björk-Eriksson [***], C. Grau [§§], D. Norman [††], B.V. Offersen [¶‖§§], T. Nyholm [§], J. Overgaard [¶], B. Zackrisson [§], C.R. Hansen [‡‡‡§§]

[*] *Regional Cancer Center West, Western Sweden Healthcare Region, Sweden*
[†] *Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg University, Sweden*
[‡] *Laboratory of Radiation Physics, Odense University Hospital, Denmark*
[§] *Department of Radiation Sciences, Umeå University, Sweden*
[¶] *Department of Experimental Clinical Oncology, Aarhus University Hospital, Denmark*
[‖] *Department of Oncology, Aarhus University Hospital, Denmark*
[**] *Department of Oncology, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg University, Sweden*
[††] *Regional Cancer Center North, Northern Sweden Healthcare Region, Sweden*
[‡‡] *Institute of Clinical Research, University of Southern Denmark, Denmark*
[§§] *Danish Center of Particle Therapy, Aarhus University Hospital, Denmark*

Abstract

Collecting large amounts of radiotherapy (RT) data from clinical systems is known to be a challenging task. Still, data collections outside the original RT systems are needed to follow-up on the quality of cancer care and to improve RT. This paper aims to describe how RT data is collected nationally in Denmark and Sweden for this purpose and gives an overview of the stored information in both countries' national data sources.

Although both countries have clinical national quality registries with broad coverage and completeness for many cancer diagnoses, some were initiated already in the seventies, and less than one in ten includes quantitative information on RT to a level of detail useful for more than basic descriptive statistics. Detailed RT data can, however, be found in Denmark's DICOM Collaboration (DcmCollab) database, initiated in 2009 and in Sweden's quality registry for RT launched in 2023 (SKvaRT). Denmark has collected raw DICOM data for all patients enrolled in clinical trials, with files being directly and automatically transferred to DcmCollab from the original data sources at each RT centre. Sweden collects aggregated RT data into SKvaRT for all patients undergoing RT in Sweden, with DICOM files being transferred and selected alpha-numeric variables forwarded via a local intermediate storage database (MIQA) at each hospital. In designing their respective solutions, both countries have faced similar challenges regarding which RT variables to collect and how to technically link clinical systems to their data repositories. General lessons about how flexibility currently is balanced with storage requirements and data standards are presented here together with future plans to harvest real-world RT data.

Keywords: Data collection; National database; Quality parameters; Radiotherapy; Radiotherapy quality assurance; Treatment plan data

Introduction

Radiotherapy (RT) is one of the most data-rich disciplines in modern healthcare. Although national registries for the cancer domain, including disease-specific national quality registries (NQRs), have been present for more than 50 years in both countries, the information about RT in these data collections is for many cancers still limited [1,2]. The lack of detailed RT information from large cohorts can be problematic for healthcare stakeholders and researchers, as there can be significant differences in results from randomised clinical trials and the impact on real-world patients [3]. In Denmark and Sweden, past and ongoing strategies have been devised to create central storage solutions for RT data to enable their use outside the clinical systems.

In Denmark, basic information on RT, like dose prescription, fractionation, RT start/stop dates, treatment

Author for correspondence: C. Olsson, Regionalt cancercentrum väst; Medicinaregatan 18 G; 413 90 Göteborg, Sweden.
E-mail address: caroline.olsson@vgregion.se (C.E. Olsson).

https://doi.org/10.1016/j.clon.2024.10.009
0936-6555/

modality/technique, etc. is available for some specific cancer sites in national databases like the Danish Head and Neck Cancer Database (DAHANCA) [4] and the Danish Breast Cancer Group (DBCG) [5]. These are created to store information about all cancer patients, including their treatment and outcomes and are almost 100% complete for Danish patients with such diagnoses (annual DAHANCA and DBCG reports [6,7]). Although RT data for large cohorts can be obtained from these, the collected information is too general to provide detailed new insights on RT effects. In addition, the collected data typically varies across sites. Similarly, as for some of the Danish NQRs, the Swedish cancer-diagnose-specific NQRs also include basic RT information for many common cancer diagnoses. Although the NQRs are primarily used for quality improvements and to monitor adherence to best-practise guidelines in both countries, they are also used for research. Many of the NQRs have full coverage (all relevant healthcare units participate) and approach full completeness (almost all of the intended target population are registered) [2,4,8–10]. Unfortunately, the collected RT data in NQRs are also typically too general for analyses of RT effects. More detailed information on RT can, however, be found in Denmark's DICOM Collaboration (DcmCollab) database that was initiated in 2009 [11] and in Sweden's quality registry for RT (SKvaRT) launched in 2023 *(in Swedish:* **S**venska **Kva**litetsregistret för **R**adio**T**erapi [12]). DcmCollab is a stand-alone solution that can be linked to the Danish NQRs for specific projects; SkvaRT is created on the same IT platform as the cancer-specific NQRs and linkage between them is technically easy.

Worldwide, there are few descriptions of initiatives collecting detailed information on RT into the same national data source or providing a national collaborative infrastructure and data-sharing policy between all RT clinics [13]. This paper aims to describe how RT data is collected in Denmark and Sweden and to give an overview of the stored information in both countries' national data sources as well as the future plans to harvest real-world RT data.

The Danish/Swedish National RT Data Collections

In Denmark and Sweden, a unique identification number is given to all citizens at birth. This identification number allows for data from different registries to be linked. The national cancer registries of Denmark and Sweden were initiated in the 1940s and the 1950s, respectively. As for the other Nordic countries, they cover total populations on the basis of the unique individual patient identification number and the legislated mandatory reporting of all newly diagnosed cancers [14,15]. The information comes from multiple sources like hospitals, primary care physicians, pathology laboratories, and death registries. The first NQR for cancer was launched in the 1970s in Denmark and in the 1980s in Sweden [2]. Today, there are 22 additional cancer-diagnose-specific registries, including sub-registries, in Denmark [16] and 36 in Sweden [17]. For most diagnoses, individual clinical information about disease characteristics, treatments, and overall cancer treatment outcomes are well documented.

Data collection of RT data has, in the past, mainly consisted of manual reporting of limited information on treatments for both Denmark and Sweden. Examples of such collected data in its more basic form include whether RT ha been given (yes or no) and the dates for delivery (period from first to last treatment day); in more detailed form, RT total prescribed doses to the tumour volume (target) and dose per fraction (in Grey [Gy]) could also be given. For many clinical trials, such information was sufficient at the time, with several randomised controlled studies having changed international clinical practice in RT, *e.g.* the Danish DAHANCA 6&7 trial reported in 2003 [18] and the Swedish Rectal Cancer trial reported in 1997 [19]. Studies like these primarily investigated effects by older conformal RT treatment techniques and managed with less detailed RT information than what now is needed to increase knowledge about effects after modern intensity or volumetric modulated treatment techniques. Newer Danish trials like Narlal 2, investigating iso-toxic dose escalation for advanced lung cancer [20] and DAHANCA 35, investigating the potential toxicity reduction of proton vs photon RT for H&N cancer [21,22] or the Swedish trial investigating ultra-hypofractionated versus convent ionally-fractionated RT for prostate cancer [23] need more detailed information regarding both tumour and non-tumour tissue dose deposits. The clinical IT systems used in RT are oncology information systems (OISs) and treatment planning systems (TPS). OIS and TPS data typically adhere to the Digital Imaging and COmmunication in Medicine (DICOM) standard, introduced in the 1990s for information handling and transmission of digital images in radiology but has been extended to other medical specialities, including RT [24]. In RT, imaging and treatment data for each patient are stored in four main types of DICOM objects with different characteristics: (1) DICOM CT, or DICOM MRI in modern MR-only workflows, contains the image scans used for treatment planning; (2) DICOM RT Structure Set defines the image segmentation or contouring of treatment volumes, *i.e.* tumour volumes to be targeted and non-tumour-tissue volumes to be avoided (organs at risk [OARs]); (3) DICOM RT Plan describes the technical settings of the treatment; (4) DICOM RT Dose defines the dose distribution across an image scan. All image objects and the dose object outline the patient's anatomy based on small volume elements (voxels), making it possible to illustrate tissue and dose distributions to the scale of mm^3.

History

Danish Perspective

In Denmark, the regional clinical quality development program (*in Danish*: Regionernes Kliniske KvalitetsudviklingsProgram [RKKP]) captures data from different clinical services, of which 22 are cancer-related. The patients are, by law, granted oncologic treatment within specific timespans and the parameters captured are

mainly linked to the timing of these treatments. In 16 of the 22 cancer quality databases (73%), RT-related parameters are captured; however, the parameters are all simple, with almost no information related to the actual treatment given (Table 1).

In 2009, a national Danish collaboration supported by The Lundbeck Foundation Center for International Research in Radiation Oncology (CIRRO) funded the startup of digital data collection from the TPS. In 2017, CIRRO was renamed the Danish Comprehensive Cancer Center for Radiation Therapy (DCCC RT [25]) and was then supported by the Danish Cancer Society. The design of the detailed information from the treatment plans to be stored was thoroughly discussed at national workshops since collecting RT data is difficult and time-consuming, as numerous items can be collected. It was soon realised that irrespective of how foreseeing and elaborate the data collection was going to be, there would always be a risk of missing important information. Retrospective data collection for new trials or incorporation of new knowledge would be needed. A way to solve this problem was to collect the raw data, which is the three-dimensional (3D) treatment plan for RT. The solution became the national DcmCollab (DICOM Collaboration) database where DICOM data from patients included in all Danish clinical trials involving RT nowadays are collected [11] (Figure 1).

The DICOM files transferred to DcmCollab are sent through the secure Danish Health Data Network (SDN) [26], a network dedicated to the data exchange of patient-sensitive information between Danish healthcare institutions. Modern TPS's can export the DICOM files needed to get a full treatment plan, DICOM CT, RT Structure Set, RT Plan, and RT Dose, facilitating the raw data collection. With the raw data collected, new detailed information can be generated without having to extract all patient files again. For instance, local naming conventions for targets and OARs can be mapped automatically per protocol or manually per patient. Dose-volume histograms (DVHs) can be sampled to the level of detail needed for a specific study since they are generated from the contours in the DICOM RT Structure Set file and information on dose in the associated DICOM RT Dose files.

Clinical trials and research projects like the DBCG RT Nation study [27] including RT, can use DcmCollab to collect detailed information, and depending on the trial, specific dose metrics can be extracted to validate the plan quality and ensure that the plans adhere to the trial protocol [28]. All this can be set up per protocol in DcmCollab. The access to the data is controlled by the system, where the user can have the right to see the data in a specific protocol or have the right to alter the data and protocol setting. These settings are specified for each user and each of the centres involved. Patients can be part of several protocols at the same time without having to duplicate the data.

Table 1
Overview of the number of RT variables in the Danish and Swedish cancer-diagnose-specific NQR

Registry/subregistry	Denmark	Sweden
Brain cancers[1]	10–50	-
Central Nervous System	-	10–50
Pituitary	-	>50
Breast cancer	10–50	10–50
Eye tumours	0	-
Gastric and esophageal cancer	<10	10–50
Gastrointestinal cancers		
Colorectum	<10	10–50
Anus	-	10–50
Gynaecologic cancers	<10	>50
Head and neck cancers	10–50	10–50
Hematologic malignancies		
Acute Lymphocytic Leukemia[2]	<10	<10
Acute Myeloid Leukemia[2]	<10	0
Chronic Lymphocytic Leukemia	0	0
Chronic Myeloid Leukemia	0	0
Myelodysplastic syndrome	<10	<10
Myeloproliferative disorders	0	0
Mastocytosis	-	0
Lymphoma	<10	0
Myeloma	0	<10
Hepatobiliary cancer[3]	-	<10
Kidney cancer	10–50	<10
Lung cancer[4]	<10	<10
Neuroendocrine abdominal tumours	-	0
Pancreatic cancer	<10	10–50
Pediatric cancers[a]	10–50	>50
Penile cancer	0	10–50
Prostate cancer	10–50	>50
Sarcoma	10–50	<10
Skin cancers		
Malignant melanoma	<10	0
Non-melanoma	0	-
Testis cancer	<10	10–50
Thyroid cancer[5]	-	10–50
Urinary bladder and urinary tract cancers	10–50	10–50

In Sweden, [1]also includes intradural and meningeal cancers, [3]also includes liver metastases, [4]also includes mesothelioma; In Denmark, [2]are included in the same acute leukemia database, [5]is included in the head and neck cancer registry [2].
Danish data were retrieved in September 2023 from https://www.rkkp.dk/and Swedish data were retrieved in September 2023 from https://cancercentrum.se/samverkan/using RT-specific search terms such as "RT", "Gy", "dose", "target", and "OAR" in the accessed variable lists. Registries with <10 variables typically register if RT has been used (yes/no), start and end dates of treatment, and total prescribed dose; those capturing 10–50 variables may also register if there are multiple targets or boost volumes, number of fractions, and treatment intent (curative/palliative); those capturing >50 variables often have additional details on dose per fraction, fractions per day/week, radiation quality, energy, treatment technique, and doses to pre-specified target volumes and organs at risk.

[a] There are seven Swedish ongoing pediatric protocols registered on INCA of which five include RT variables (Lymphoma/RADTOX/SALUB/VCTB/VSTB/: <10/>50/>50/>50/10–50 RT variables).

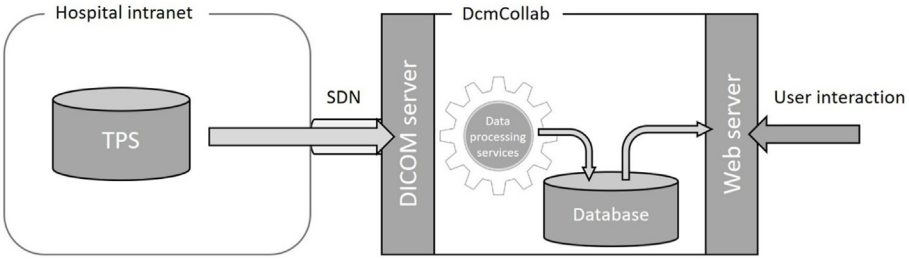

Fig 1. The Danish data source DcmCollab used for storage of detailed RT information.

Swedish Perspective

In Sweden, collection and quality control of the cancer registry and of the cancer-diagnose-specific NQRs are administered on the Information Network for Cancer Care (INCA) [29]. INCA is an IT platform jointly owned by the six regional cancer centres (RCC), appointed national competence centres by the Swedish government since 2009 to form an educational and supporting organisation for cancer care [9,17]. INCA provides functionality for data capture, processing, and output solutions. Of the 36 Swedish cancer-diagnose-specific NQRs on INCA, there are 26 (72%), which presently include RT variables of some sort [17] (Table 1).

The purpose of SKvaRT is to collect information about treatment volumes, OARs, and various dose and volume parameters for all patients undergoing RT in Sweden (Figure 2). SKvaRT is considered a "methodologic NQR", *i.e.* it complements the cancer-diagnose-specific NQRs with sufficiently detailed RT data to answer general questions about RT. Two earlier initiatives have paved the way for SKvaRT, one developing a standardised Swedish naming convention for RT [30] and the other developing a database solution for unified storage of RT data outside the clinical RT systems, the Medical Information Quality Archive (MIQA) [31].

MIQA acts as an intermediate storage database between the hospitals delivering RT in Sweden and SKvaRT [31]. It is installed as a local quality registry at each centre and linked to the centre's OIS and TPS. MIQA is designed, developed, and maintained at Norrland's University Hospital in collaboration with RCC in Northern Sweden. MIQA has so far been funded by these organisations and by external research funding as well as annual fees from most Swedish RT centres. Data following the DICOM standard (DICOM RT Structure Set, RT Plan, and RT Dose; images are optional) are transferred from the OIS to MIQA using export filters, customised to fit the prerequisites of each centre. In MIQA, the data are checked to ensure that the imported information is complete and that the structure volumes follow the nationally agreed-upon standardised naming convention for RT. When this is confirmed, the data are automatically forwarded to SKvaRT, which is located on the INCA platform together with the other NQRs in cancer. Data transfer between a hospital's MIQA and SKvaRT fulfils the same requirements of certificate and encryption protocols as data transfer to any other NQR on INCA [29]. In addition, an independent check of data consistency between the original data source and SKvaRT is performed as a centre reports data into SKvaRT for the first time [12].

The information stored for each patient in SKvaRT provides a condensed but representative version of the planned dose distribution used for the delivery of RT without being as comprehensive as the original information. In the OIS, the data storage requirement for the raw DICOM files needed to reconstruct the treatment plan for a patient can add up to several hundred MB or more. In contrast, the variables stored in SKvaRT contain alphabetical and

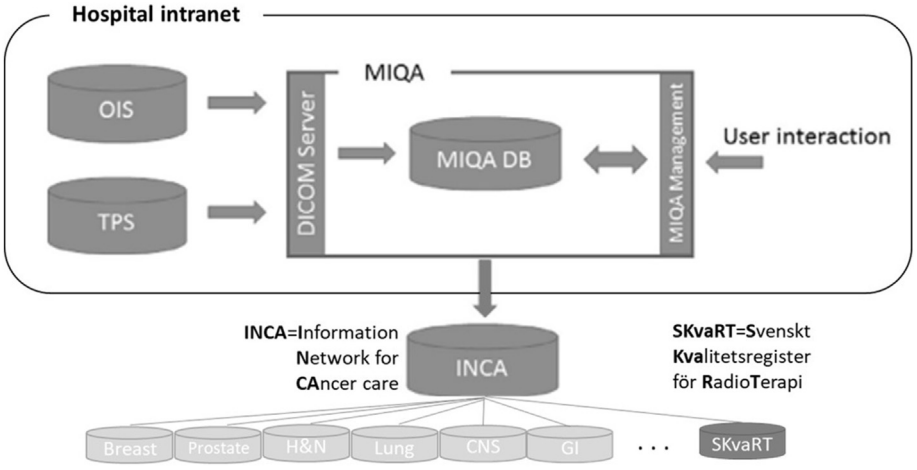

Fig 2. The Swedish data source SKvaRT used for storage of detailed RT information.

numerical values and stay in the range of a few kB. This is achieved by condensing the raw OIS data into a limited set of variables that together capture information about the overall treatment (*e.g.* treatment intent), ICDcode [32], each treatment fraction, (*e.g.* treatment technique), energy, and the structures defined in each treatment plan (*e.g.* name, type, target [33] or OAR doses, and volumes). A sparsely sampled DVH with 15 pre-determined variables represents doses to volumes of all structures defined in a patient's treatment plan (dose to 1/2/5/10/20/30/40/50/60/70/80/90/95/98/99% of the volume). Minimum, mean, and maximum doses are also stored in separate variables.

Current Status

Danish Perspective

Since 2010, all eight Danish centres have an established connection to DcmCollab. The data transfer is automatic and is realised through pre-defined export filters in the TPS at each centre. Currently, there are 58 ongoing Danish trials involving RT that export data to Dcmcollab and the database currently includes around 14,700 patients treated since 2001 (Figure 3).

Over the years, DcmCollab has moved from storing and presenting data on demand to making it possible to see dose distributions and contours in a web browser [11]. Here, the user can get a basic understanding of the submitted data for curation purposes.

DcmCollab provides an audit tool which can be used for quality assurance purposes. The feature has been used for contouring and treatment planning audits, where multiple, individually anonymised copies of a single dataset have been distributed to participating centres. This ensures correct handling of the DICOM information when the data is sent back to DcmCollab.

Adding an empty ROI in the struct file with a pre-defined trigger name can automatically link a treatment plan to a specific protocol. Likewise, it is possible to automatically forward the data to another treatment centre with the appropriate forward trigger. These features mean that the typical DcmCollab user does not have to log in to the system, as most tasks are automated.

All technical information can be extracted with the full RT plan available [34]. The typical features are DVH metrics or structure volumes. It is possible to perform simple statistical operations in DcmCollab, i.e. to calculate minimum, mean and maximum values across the patient cohort. Still, more advanced analyses like plan comparison or correlation analyses are performed outside of DcmCollab, with data needing to be exported from the DcmCollab interface to programming software like Phyton or MatLab.

Swedish Perspective

In September 2023, 16 of Sweden's 17 RT departments/clinics ha their own MIQA database (94%). Of these, 11/16 (69%) also fulfil the requirements to transfer data between MIQA and SKvaRT and have also completed or are undergoing the independent data consistency check as they begin to report data into the registry. Altogether, nine hospitals have until now reported data into SKvaRT of which six do this regularly. The registry currently includes around 20,200 patients treated since 2016 (when the standardised naming convention was nationally accepted [30]; Figure 3).

To implement MIQA, staff engagement from the hospital in question is needed. Medical physicists typically assist in creating export filters from local OISs and oversee that the MIQA system functions correctly. Some have created automatic data export solutions from their OIS to MIQA, while others manually trigger the export after the patient completes their treatment. For various reasons, it has taken

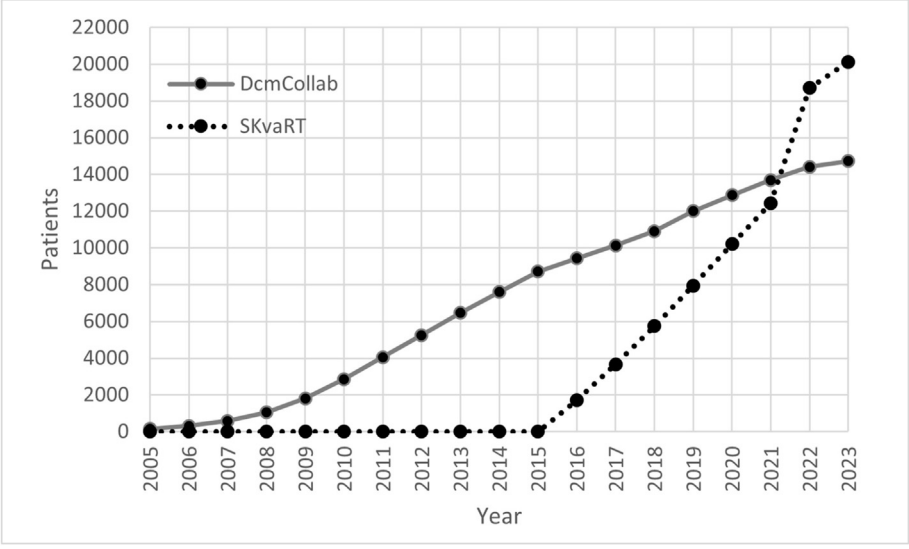

Fig 3. Number of patients accumulated over time in Denmark's RT data source DcmCollab and Sweden's RT data source SKvaRT (September 2023).

longer to get MIQA up and running at certain hospitals than others. This is partly explained by some IT departments being less flexible in supporting the implementation of non-standard software solutions and RT staff shortages, making initiating and engaging in MIQA-related activities difficult.

Results from SKvaRT are publicly available on the RCC website [12]. Similarly to other NQRs on the INCA platform, data overviews are visualised in tables, charts, and plots. Currently, results for the number of patients per region, treatment lengths and schedules for selected patient groups are presented. The information can be visualised by ICD code or by body region (CNS and brain, head and neck, breast, thorax, gastrointestinal, gynaecology, prostate, and genitourinary; up to 19 ICD codes per group). There is also one miscellaneous group, which includes 48 ICD codes. The presented treatment schedules for SKvaRT are based on the delivered total dose and have, in part, been identified based on recommendations in care programs and after input from expert radiation oncologists. There are ongoing discussions on how to present the more detailed structure doses in upcoming annual reports, it has not been decided how this level of detail should be communicated nor when the data will be merged with the cancer-diagnose-specific NQRs.

Discussion and Concluding Remarks

Most Danish and Swedish cancer-diagnose-specific NQRs include RT data; however, less than one in ten includes a level of detail useful to increase detailed knowledge about the effects of modern RT. In Denmark, DcmCollab is a national data source that was launched in 2009 to collect detailed RT information from all Danish clinical trials involving RT. SKvaRT, together with MIQA, is the corresponding Swedish data source solution since 2023 to collect detailed RT information from all patients undergoing RT in Sweden. The two countries have faced similar challenges regarding which RT variables to collect and how to technically link clinical systems to their respective registry. Given the two approaches to address these issues, one of the main differences between the Danish and Swedish solutions is the purpose behind reporting and using the collected data. The Danish policy promotes storing information for research purposes in DcmCollab but opens up possibilities to use them for quality insights. The Swedish policy promotes the opposite, with SKvaRT primarily being classified as a methodological NQR.

The patient data stored in DcmCollab are the original DICOM files, including images used for the clinical treatments, and require database space of TB magnitude. All information necessary to recreate the full treatment plan is present in one database and is accessed from the same web browser. The RT data will, therefore, be insensitive to the time at which they were exported from the original data source since, if needed, DVHs can be recalculated for new OAR segmentations. Specific target and OAR dose metrics can be extracted according to preference. This flexibility is one of the main strengths of the Danish solution and has, as one example, been used to pool data from two Danish trials to identify an age-dependent relationship between treated breast volume and tissue fibrosis normal tissue complication probability modelling in breast cancer patients [35]. DcmCollab allows for automatically mapping structures, which can be a potential drawback since this process can be cumbersome without protocol-specific naming. These naming conventions are in place for prospective trials, but in retrospective analysis, this is rarely the case. In contrast, the patient data stored in SKvaRT are mapped according to the national naming convention and represent a subset of pre-determined RT variables. Although the SKvaRT data demand less storage than the DcmCollab data, a potential drawback is that the SKvaRT information represents the point in time when data were exported from the original source. Even if the original DICOM files may be kept in MIQA after the information has been sent to SKvaRT, it is up to the local centre to include DICOM images, which limits the access and updates that can be performed on these data including possibilities to recalculate dose for new OAR segmentations. Comparing the variables in DcmCollab and SKvaRT with similar initiatives internationally proved to be difficult (a PubMed search on July 24th 2024 including various combinations of “radiotherapy”, “national”, “database”, “data source”, “registry”, “infrastructure”, “data collection”, and “data sharing” resulted in two hits, one concerning Norway by Åsli *et al.* from 2014 [13] and one concerning France by Chauvet *et al.* from 2009 [36] (*in French*). Of these, the Norwegian initiative, also mentioned in a more recent publication from 2021 by Fosså *et al.* [37], describes how RT data are annually delivered as electronic records from all Norwegian RT centres to the Cancer Registry of Norway (CRN) since 1997. The collected data include dates on RT initiation and end, region irradiated, treatment intention, total RT dose, and number of fractions. Thus, this content differs little from the limited information that can be retrieved from some of the NQRs in Denmark and Sweden.

Concerning their intended target populations, the completeness and accuracy of the registered data in DcmCollab are high and expected to be high in SKvaRT with time. The data quality in DcmCollab has been tested by the DBCG. The group has manually collected RT parameters for specific RT trials. This was compared to the automatically extracted data from the DICOM data transferred to DcmCollab. Brink *et al.* showed that the manually collected data had an error rate of approximately 10% [38]. As data accumulates in SKvaRT, the data quality is planned to be tested in a larger setting in line with how the current independent check of data consistency is done as a new centre reports data into SKvaRT for the first time [12]. However, estimating the completeness of SKvaRT is challenging since no obvious data source like the National Cancer Registry can be used for reference. The current plan is to use the original data sources at each hospital and obtain the number of patients treated at each hospital in question during a specified period. Whether this will be a feasible and sustainable solution is currently unknown. Still, as part of SKvaRT maturing as a registry and with confirmed

high-quality data, the hope is to climb the Swedish certification system for registries to meet requirements for yearly funding from authorities to increase possibilities for further developments [2]. As one part of achieving the highest certification level, a register is also required to have a plan to present its content on a meta-level according to the generic statistical information model (GSIM [39]) and to be part of the register utiliser tool (RUT) as created by the Swedish Medical Research Council to promote registry-based research [40].

For DcmCollab, the major challenge is to develop the user-friendliness. As the project has evolved over the last 10 years, the front end and back end need a makeover to move to a more modern software style. DcmCollab will evolve and modernise through a newly Novo-funded project, DESIRE: Data Science Research Infrastructure In Radiotherapy. The project will focus on automatic data access, harvesting, curation, and security. This will be done by building separate software modules, including bulk import and export, data outlier detection [41], federated learning [42], and artificial intelligence model sharing [34]. DESIRE will extend the national IT infrastructure dedicated to RT. Statistical packages for RT audits and plan comparison analyses will be developed so researchers across Denmark can access advanced analyses within DcmCollab. The federated learning module will make international collaboration much easier, as the Danish data do not have to leave the country, and the analysis method can travel between the databases and only retrieve the aggregated regression coefficients [43]. Potentially, federated learning can include clinical parameters from other databases without combining the data in one, also called vertical federated learning [44]. There are also efforts to make the DcmCollab database a clinical NQR, ensuring financial support and not relying on external funding. However, this would perhaps make the database more rigid, as new initiatives would have to be approved. Currently, the database can evolve dynamically; as the needs arise, they can be solved, and only the programming resources limit the development.

For SKvaRT, the major goal and future challenge is to provide RT data that matches the needs of care providers, decision-makers, and researchers to improve the quality of Swedish RT and the conditions for equity in cancer care. Another objective is to increase possibilities for data-driven decision-making initiatives and knowledge about RT effects. To this end, the Swedish RT community and other interested parties, such as the Swedish Radiation Safety Authority, need to be coordinated so that the SKvaRT variables will be tailored to the level of detail needed to match their interests and, when applicable, also acknowledge international initiatives on RT quality indicators [45]. In addition to the publicly available data, any reporting centre will be able to access its own data, similar to reporting centres of the diagnose-specific NQRs. The stored information will also be accessible to researchers with projects approved by the Swedish Ethical Review Authority. Another challenge is building transparency between SKvaRT and the existing NQRs to avoid data overlap between registries and identify blind spots. For instance, variables reflecting scores by frequently used toxicity grading systems in RT such as the CTCAE [46] are not included in any of the cancer-diagnose-specific NQRs but are warranted in the future if SKvaRT data are to *e.g.* illustrate relationships between dose and side effects.

In conclusion, Denmark and Sweden have made substantial progress in collecting and managing RT data, though each follows distinctly different approaches. Denmark's DcmCollab prioritises research, offering rich and flexible data, while Sweden's SKvaRT focuses on quality improvements, providing a more streamlined dataset. Both nations face the common challenge of balancing data detail with storage requirements. Looking ahead, Denmark's DESIRE project aims to modernise DcmCollab, enhancing data accessibility and security, while Sweden aims to tailor SKvaRT to meet the diverse needs of RT stakeholders. These efforts underscore the importance of comprehensive RT data in improving cancer treatment outcomes and promoting data-driven decision-making. Given Denmark and Sweden's unique national identification numbers and public healthcare systems, both countries have the advantage of accessing nationwide RT data and serve as examples of proactive data management in the field of RT. Although there currently are no plans for a joint RT database or collaborative infrastructure and data-sharing policy between Denmark and Sweden, a joint Scandinavian solution would preferably be based on federated learning to reduce legal issues. Such a data-sharing solution would increase the diversity of data and could potentially contribute more valuable insights to the global healthcare community than each isolated country.

Author Contributions

1. guarantor of integrity of the entire study

Olsson C (Swedish part) and Hansen CR (Danish part).

2. study concepts and design

Olsson C and Hansen CR (overall).
Karlsson M, Björk-Eriksson T, Norman D, and Nyholm T (Swedish part).
Krogh SL, Eriksen JG, Cai Grau, Offersen BV, and Overgaard J (Danish part).

3. literature research

Olsson C, Hansen CR (overall).

4. clinical studies

N/A.

5. experimental results/data analysis

N/A.

6. statistical analysis

N/A.

7. manuscript preparation

Olsson C, Hansen CR.

8. manuscript editing

Olsson C, Krogh SL, Karlsson M, Eriksen JG, Björk-Eriksson T, Cai Grau, Norman D, Offersen BV, Nyholm T, Overgaard J, and Hansen CR.

All authors made substantial contributions to (1) the conception or design of study, or acquisition of data, or interpretation of data, (2) drafting of the article or revising it critically for important intellectual content, and (3) the final approval of the version to be submitted.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgement

C. Grau and C.R. Hansen Supported by part of DESIRE - Data Science Infrastructure in Radiotherapy (Grant no. NNF22OC0077893). C. Olsson Supported by Jubileumsklinikens Cancer Research Foundation (Grant no. 2023:505).

References

[1] Pukkala E, Engholm G, Hojsgaard Schmidt LK, Storm H, Khan S, Lambe M, *et al*. Nordic Cancer Registries - an overview of their procedures and data comparability. *Acta Oncol* 2018; 57:440–455. https://doi.org/10.1080/0284186X.2017.1407039.

[2] Nilbert M, Thomsen LA, Winther Jensen J, Moller H, Borre M, Widenlou Nordmark A, *et al*. The power of empirical data; lessons from the clinical registry initiatives in Scandinavian cancer care. *Acta Oncol* 2020;59:1343–1356. https://doi.org/10.1080/0284186X.2020.1820573.

[3] Peters LJ, O'Sullivan B, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, *et al*. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol* 2010;28:2996–3001. https://doi.org/10.1200/JCO.2009.27.4498.

[4] Overgaard J, Jovanovic A, Godballe C, Grau Eriksen J. The Danish Head and Neck Cancer database. *Clin Epidemiol* 2016; 8:491–496. https://doi.org/10.2147/CLEP.S103591.

[5] Offersen BV, Ejlertsen B, Balslev E, Flyger H, Gerdes AM, Hansen MK, *et al*. Is DBCG abreast of new developments? *Acta Oncol* 2018;57:1–2. https://doi.org/10.1080/0284186X.2017.1408960.

[6] Group DAHANCA. *Årsrapport 2022 for den kliniske kvalitetsdatabase DAHANCA* 2022, https://www.sundhed.dk/content/cms/5/39205_dahanca_aarsrapport-2022-offentliggjort-version-20230623.pdf. [Accessed 9 August 2024].

[7] Gruppe DBCG. *DBCG Årsbereting 2022* 2022, https://www.dbcg.dk/images/PDF/Aarsberetning/DBCG%20%C3%A5rsrapport%20for%202022_FINAL.pdf. [Accessed 9 August 2024].

[8] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. *Eur J Cancer* 2009;45:747–755. https://doi.org/10.1016/j.ejca.2008.11.032.

[9] Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish Healthcare Quality Registries. *J Intern Med* 2015;277:94–136. https://doi.org/10.1111/joim.12303.

[10] Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. *Completeness Eur J Cancer* 2009;45:756–764. https://doi.org/10.1016/j.ejca.2008.11.033.

[11] Krogh SL, Brink C, Lorenzen EL, Samsoe E, Vogelius IR, Zukauskaite R, *et al*. A national repository of complete radiotherapy plans: design, Results, and experiences. *Acta Oncol* 2023:1–8. https://doi.org/10.1080/0284186X.2023.2270143.

[12] Olsson C, Karlsson M, (on behalf of the SKvaRT steering group). *Kvalitetssäkring av data i strålterapiregistret (SKvaRT) (Eng. Quality assurance of data in the Swedish quality register for radiation therapy)* 2023, https://cancercentrum.se/globalassets/vara-uppdrag/kunskapsstyrning/stralbehandling/kvalitetssakring_data_stral_2023.pdf. [Accessed 9 August 2024].

[13] Asli LM, Kvaloy SO, Jetne V, Myklebust TA, Levernes SG, Tveit KM, *et al*. Utilization of radiation therapy in Norway after the implementation of the national cancer plan–a national, population-based study. *Int J Radiat Oncol Biol Phys* 2014;90:707–714. https://doi.org/10.1016/j.ijrobp.2014.06.059.

[14] Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, *et al*. NORDCAN–a Nordic tool for cancer information, planning, quality control and research. *Acta Oncol* 2010;49: 725–736. https://doi.org/10.3109/02841861003782017.

[15] Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. *Acta Radiol Oncol* 1984;23:305–313. https://doi.org/10.3109/02841868409136026.

[16] *Regionernes Kliniske KvalitetsudviklingsProgram (RKKP)* 2024https://www.rkkp.dk/kvalitetsdatabaser/databaser/. [Accessed 9 August 2024].

[17] Regionala cancercentrum i samverkan. *(Eng. Regional cancer centres)* 2024, https://cancercentrum.se/samverkan/. [Accessed 9 August 2024].

[18] Overgaard J, Hansen HS, Specht L, Overgaard M, Grau C, Andersen E, *et al*. Five compared with six fractions per week of conventional radiotherapy of squamous-cell carcinoma of head and neck: DAHANCA 6 and 7 randomised controlled trial. *Lancet* 2003;362:933–940. https://doi.org/10.1016/s0140-6736(03)14361-9.

[19] Swedish Rectal Cancer T, Cedermark B, Dahlberg M, Glimelius B, Pahlman L, Rutqvist LE, *et al*. Improved survival with preoperative radiotherapy in resectable rectal cancer. *N Engl J Med* 1997;336:980–987. https://doi.org/10.1056/NEJM199704033361402.

[20] Hansen O. Novel approach to Radiotherapy in Locally Advanced Lung Cancer Concomitant Navelbine (NCT02354274), https://clinicaltrials.gov/study/NCT02354274?id=NCT02354274&rank=1;2015. [Accessed 9 August 2024].

[21] Friborg J, Jensen K, Eriksen JG, Samsoe E, Maare C, Farhadi M, *et al*. Considerations for study design in the DAHANCA 35 trial of protons versus photons for head and neck cancer. *Radiother Oncol* 2023:109958. https://doi.org/10.1016/j.radonc.2023.109958.

[22] Hansen CR, Jensen K, Smulders B, Holm AIS, Samsoe E, Nielsen MS, *et al*. Evaluation of decentralised model-based selection of head and neck cancer patients for a proton treatment study. DAHANCA 35. *Radiother Oncol* 2023:109812. https://doi.org/10.1016/j.radonc.2023.109812.

[23] Widmark A, Gunnlaugsson A, Beckman L, Thellenberg-Karlsson C, Hoyer M, Lagerlund M, *et al*. Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the HYPO-RT-PC randomised, non-inferiority, phase 3 trial. *Lancet* 2019;394:385–395. https://doi.org/10.1016/S0140-6736(19)31131-6.

[24] Law MY, Liu B. Informatics in radiology: DICOM-RT and its utilization in radiation therapy. *Radiographics* 2009;29:655–667. https://doi.org/10.1148/rg.293075172.

[25] Danish national research center for radiotherapy. *DCCC Stråleterapi* 2024, https://www.straaleterapi.dk/. [Accessed 9 August 2024].

[26] medcom. *SDN* 2024, https://medcom.dk/systemforvaltning/sundhedsdatanettet-sdn/. [Accessed 9 August 2024].

[27] Refsgaard L, Skarso ER, Ravkilde T, Nissen HD, Olsen M, Boye K, *et al*. End-to-end framework for automated collection of large multicentre radiotherapy datasets demonstrated in a Danish Breast Cancer Group cohort. *Phys Imaging Radiat Oncol* 2023;27:100485. https://doi.org/10.1016/j.phro.2023.100485.

[28] Lorenzen EL, Thomsen MS, Bechmann T, Berg M, Nielsen HM, Hansen HV, *et al*. Inter-center variations in contouring and planning compromises in the DBC Skagen trial. *Radiother Oncol* 2022;170:S822–S823.

[29] Regionala cancercentrum i samverkan. Om INCA (*Eng*. About INCA), https://cancercentrum.se/samverkan/vara-uppdrag/kunskapsstyrning/kvalitetsregister/stod-for-kvalitetsregister/om-inloggning/om-inca/. [Accessed 9 August 2024].

[30] Montelius A and the national reference group. *A standardized Swedish naming convention for radiation therapy* 2016, https://www.stralsakerhetsmyndigheten.se/contentassets/a7d743fd0e09417d97f6857f40683e35/201620-a-standardized-swedish-naming-convention-for-radiation-therapy. [Accessed 9 August 2024].

[31] Nyholm T, Olsson C, Agrup M, Bjork P, Bjork-Eriksson T, Gagliardi G, *et al*. A national approach for automated collection of standardized and population-based radiation therapy data in Sweden. *Radiother Oncol* 2016. https://doi.org/10.1016/j.radonc.2016.04.007.

[32] *International Statistical Classification of Diseases (ICD) and Related Health Problems* 2024https://icd.who.int/browse10/2019/en. [Accessed 9 August 2024].

[33] International Comission on Radiation Units and Measurements (ICRU). *Report 50, Prescribing, Recording, and Reporting Photon Beam Therapy* 1993, https://www.icru.org/report/prescribing-recording-and-reporting-photon-beam-therapy-report-50/. [Accessed 9 August 2024].

[34] Nielsen CP, Lorenzen EL, Jensen K, Sarup N, Brink C, Smulders B, *et al*. Consistency in contouring of organs at risk by artificial intelligence vs oncologists in head and neck cancer patients. *Acta Oncol* 2023:1–8. https://doi.org/10.1080/0284186X.2023.2256958.

[35] Thomsen MS, Alsner J, Nielsen HM, Jakobsen EH, Nielsen MH, Moller M, *et al*. Volume matters: Breast induration is associated with irradiated breast volume in the Danish Breast Cancer Group phase III randomized Partial Breast Irradiation trial. *Radiother Oncol* 2022;177:231–235. https://doi.org/10.1016/j.radonc.2022.09.024.

[36] Chauvet B, Bolla M, Alies-Patin A, Bara C, Bourguignon M, Chantome G, *et al*. [French radiotherapy database: results of a survey of French radiation oncology centers in 2007]. *Cancer Radiother* 2009;13:466–470. https://doi.org/10.1016/j.canrad.2009.07.010.

[37] Fossa SD, Aas K, Muller C, Jerm MB, Tandstad T, Lilleby W, *et al*. Definitive radiotherapy for prostate cancer in Norway 2006-2015: Temporal trends, performance and survival. *Radiother Oncol* 2021;155:33–41. https://doi.org/10.1016/j.radonc.2020.10.022.

[38] Brink C, Lorenzen EL, Krogh SL, Westberg J, Berg M, Jensen I, *et al*. DBCG hypo trial validation of radiotherapy parameters from a national data bank versus manual reporting. *Acta Oncol* 2018;57:107–112. https://doi.org/10.1080/0284186X.2017.1406140.

[39] Statswiki U. Generic Statistical Information Model (GSIM), https://statswiki.unece.org/display/gsim/Summary.

[40] Swedish Research Council. *RUT – Register Utiliser Tool* 2024, https://www.vr.se/english/mandates/research-infrastructure/find-research-infrastructure/list/2020-04-23-rut–register-utiliser-tool.html. [Accessed 9 August 2024].

[41] Brink C. Automatic validation of organ at risk delineations based on machine learning (PD-0824). *Radiother Oncol* 2021;161:S658–S659. https://doi.org/10.1016/S0167-8140(21)07103-6.

[42] Hansen CR, Price G, Field M, Sarup N, Zukauskaite R, Johansen J, *et al*. Open-source distributed learning validation for a larynx cancer survival model following radiotherapy. *Radiother Oncol* 2022;173:319–326. https://doi.org/10.1016/j.radonc.2022.06.009.

[43] Hansen CR, Price G, Field M, Sarup N, Zukauskaite R, Johansen J, *et al*. Larynx cancer survival model developed through open-source federated learning. *Radiother Oncol* 2022;176:179–186. https://doi.org/10.1016/j.radonc.2022.09.023.

[44] Yang Z,NWR. Improved Privacy-Preserving Bayesian Network Parameter Learning on Vertically Partitioned Data. Tokyo, Japan. In: *21st international Conference on data engineering workshops (ICDEW'05)* 2005. https://doi.org/10.1109/ICDE.2005.230. 1196-1196.

[45] Harden SV, Chiew KL, Millar J, Vinod SK. Quality indicators for radiation oncology. *J Med Imaging Radiat Oncol* 2022;66:249–257. https://doi.org/10.1111/1754-9485.13373.

[46] (NCI) NCI. *Common Terminology Criteria for Adverse Events (CTCAE)* 2024, https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm. [Accessed 9 August 2024].