



Data exchange in radiotherapy

Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets



Tomas Skripcak^{a,*}, Claus Belka^b, Walter Bosch^c, Carsten Brink^{d,ae}, Thomas Brunner^e, Volker Budach^f, Daniel Büttner^a, Jürgen Debus^g, Andre Dekker^h, Cai Grauⁱ, Sarah Gulliford^j, Coen Hurkmans^{k,l,m}, Uwe Justⁿ, Mechthild Krause^{a,n,o,p}, Philippe Lambin^h, Johannes A. Langendijk^q, Rolf Lewensohn^r, Armin Lühr^{a,o}, Philippe Maingon^{l,r}, Michele Masucci^s, Maximilian Niyazi^b, Philip Poortmans^t, Monique Simon^a, Heinz Schmidberger^u, Emiliano Spezi^v, Martin Stuschke^w, Vincenzo Valentini^x, Marcel Verheij^y, Gillian Whitfield^z, Björn Zackrisson^{aa}, Daniel Zips^{ab,ac,ad}, Michael Baumann^{a,n,o,p}

^a German Cancer Consortium (DKTK) Dresden and German Cancer Research Center (DKFZ) Heidelberg; ^b German Cancer Consortium (DKTK) Munich and German Cancer Research Center (DKFZ) Heidelberg, Germany; ^c Dept. of Radiation Oncology, Washington University, St. Louis, MO, USA; ^d Laboratory of Radiation Physics, Odense University Hospital, Odense, Denmark; ^e German Cancer Consortium (DKTK) Freiburg and German Cancer Research Center (DKFZ) Heidelberg; ^f German Cancer Consortium (DKTK) Berlin, German Cancer Research Center (DKFZ) Heidelberg; ^g German Cancer Consortium (DKTK) Heidelberg and German Cancer Research Center (DKFZ) Heidelberg, Germany; ^h Dept. of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands; ⁱ CIRRO Department of Oncology, Aarhus University Hospital, Aarhus, Denmark; ^j Joint Department of Physics, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK; ^k Dept. of Radiation Oncology, Catharina Hospital, Eindhoven, The Netherlands; ^l EORTC-Radiation Oncology Group; ^m EORTC-Global Clinical Trial QART Harmonisation Group, Brussels, Belgium; ⁿ Dept. of Radiation Oncology, Medical Faculty and University Hospital Carl Gustav Carus, Technische Universität Dresden; ^o OncoRay – National Center for Radiation Research in Oncology, Medical Faculty and University Hospital Carl Gustav Carus, Technische Universität Dresden and Helmholtz-Zentrum Dresden-Rossendorf; ^p Institute of Radiooncology, Helmholtz-Zentrum Dresden-Rossendorf, Germany; ^q University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ^r Department of Radiation Oncology, Centre Georges – François Leclerc, Dijon Cedex, France; ^s EurocanPlatform and Karolinska Institutet Stockholm, Sweden; ^t Department of Radiation Oncology, Radboud University Medical Centre, Nijmegen, The Netherlands; ^u ARO-Speaker, University Medical Center Mainz, Germany; ^v Dept. of Medical Physics, Velindre Cancer Centre, Cardiff, UK; ^w German Cancer Consortium (DKTK) Essen/Düsseldorf and German Cancer Research Center (DKFZ) Heidelberg, Germany; ^x Radiation Oncology Department GEMELLI-ART, Università Cattolica S. Cuore, Rome, Italy; ^y Department of Radiation Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands; ^z The Christie NHS Foundation Trust and the University of Manchester, Manchester, UK; ^{aa} Umea University, Department of Radiation Sciences, Oncology, Umea, Sweden; ^{ab} German Cancer Consortium (DKTK) Tübingen and German Cancer Research Center (DKFZ) Heidelberg; ^{ac} Radiation Oncology, Medical Faculty and University Hospital Tübingen, Eberhard Karls University Tübingen, Germany; ^{ad} Chair Clinical Committee, ESTRO; ^{ae} Institute of Clinical Research, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Article history:

Received 11 September 2014
Received in revised form 1 October 2014
Accepted 2 October 2014
Available online 28 October 2014

Keywords:

Data pooling
Interoperability
Data exchange
Large scale studies
Public data
Radiotherapy

ABSTRACT

Disconnected cancer research data management and lack of information exchange about planned and ongoing research are complicating the utilisation of internationally collected medical information for improving cancer patient care. Rapidly collecting/pooling data can accelerate translational research in radiation therapy and oncology. The exchange of study data is one of the fundamental principles behind data aggregation and data mining. The possibilities of reproducing the original study results, performing further analyses on existing research data to generate new hypotheses or developing computational models to support medical decisions (e.g. risk/benefit analysis of treatment options) represent just a fraction of the potential benefits of medical data-pooling. Distributed machine learning and knowledge exchange from federated databases can be considered as one beyond other attractive approaches for knowledge generation within “Big Data”. Data interoperability between research institutions should be the major concern behind a wider collaboration. Information captured in electronic patient records (EPRs) and study case report forms (eCRFs), linked together with medical imaging and treatment planning data, are deemed to be fundamental elements for large multi-centre studies in the field of radiation therapy and oncology. To fully utilise the captured medical information, the study data have to be more than just an electronic version of a traditional (un-modifiable) paper CRF. Challenges that have to be addressed are data interoperability, utilisation of standards, data quality and privacy concerns, data ownership, rights to publish, data pooling architecture and storage. This paper discusses a framework for conceptual packages of ideas focused on a strategic development for international research data exchange in the field of radiation therapy and oncology.

© 2014 Elsevier Ireland Ltd. Radiotherapy and Oncology 113 (2014) 303–309 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

* Corresponding author at: Dept. of Radiation Oncology, University Hospital C.G. Carus, Technische Universität Dresden, Fetscherstr. 74, 01307 Dresden, Germany.
E-mail address: t.skripcak@dkfz.de (T. Skripcak).

Background and rationale

Clinical and pre-clinical radiotherapy study data represent one of the most valuable assets for academic radiation therapy and oncology research institutions. Rapidly pooling research data via the process of data exchange has become beneficial and a necessary requirement for conducting large multi-centre radiotherapy studies [1]. Resulting data pools represent the primary input for generation of medical knowledge bases with a broad range of applications, including predictive models for decision support systems based on clinical data [2] and discovery of prognostic features in radiomics [3]. Predictive model research has potential to not only improve quality-of-life but also increase survival, for example by using isotoxic strategies [4]. Fig. 1 depicts the process of an application-specific knowledge discovery from large scale multi-centre data pools.

Integrated radiotherapy research data (originating from multiple data sources) represent a powerful research tool to evaluate dose, volume and time parameterised responses in tumours and normal tissues. Such data are fundamental for generating novel multivariable prediction models for tumour control probability (TCP) and normal tissue complication probability (NTCP). These prediction models can be translated into innovative studies on personalised radiotherapy, e.g. for biologically based intensity modulated dose distributions which may reduce the risk of treatment toxicity or increase the probability of local tumour control. As such they can also be used to inform and involve patients in treatment decisions through shared decision making [5]. Reliable estimates of treatment consequences are a prerequisite for discussing patients' preferences and for assessing their personal trade-off between the risks and benefits of treatment options. Conversely, data on patient values and preferences can also be added to the database to incorporate the patients' perspectives.

The data also are extremely useful for comparative analyses of treatment approaches, e.g. particles vs. photons or different treatment combinations [6,7], and have the potential to decrease health care costs with a more rational use of expensive medical

technology [8]. By linking them to investigations on tissues of the corresponding patients, they may also provide a backbone for the identification and validation of (imaging) biomarkers for radiation oncology. Sharing research data can accelerate the process of medical quality assurance, including checks for consistent contouring, dose (re-)planning and protocol adherence in prospective radiotherapeutic studies. Finally, sharing research data may speed up the adoption of research results into day to day clinical practice.

It is the concern of translational research informatics to provide an appropriate software solution for managing integrated research datasets, enabling the broader collaboration of research institutions.

On 26th November 2013 a workshop organised by the German Cancer Consortium (DKTK) and EurocanPlatform was hosted in Dresden, Germany to examine radiotherapy-specific IT solutions developed within Europe. Existing projects within the European Society for Radiotherapy and Oncology (ESTRO) and several regional, national and international initiatives were presented. The workshop resulted in two important conclusions. Firstly, the presented platforms, as diverse as they are, focus on the same set of problems mostly on an institutional level with few examples on a national and international dimension. Secondly, a strong interest was stated in setting up a collaborative effort to accelerate and harmonise the ongoing data collection activities and to promote open access to radiotherapy research datasets.

The main goal of this paper is to initiate the development of a radiotherapy-specific data exchange strategy preventing disconnected institutional level solutions and move towards international data interoperability. This can be achieved by the implementation of well-chosen concepts, without the need for unnecessary reinventions.

The following major challenges that currently hamper effective collaboration and data exchange efforts were identified:

- Interoperability between clinical IT solutions: systems differ in their acceptance/support of internationally standardised protocols, formats and semantics.
- Maturity of radiotherapy information standards: incomplete development of radiotherapy specific data element dictionaries, controlled vocabularies and ontologies.
- Uniformity of data collection: data are collected using different scoring systems (e.g. scoring of radiation-induced toxicity) and at different time points, which may render data merging complicated or even impossible.
- Data completeness: data are often represented without sufficient meta-data, causing the risk of information loss after exchange.
- Data quality: the quality of collected information can vary from project to project and from institution to institution, making it necessary to establish quality assurance work-flows.
- Data bias: difference in practice, protocols and equipment may cause a systematic difference between data from different institutes.
- Patient privacy: the protection of privacy and the relation to informed consent as well as secondary use of research data have to be considered seriously, also in view of the very different interpretation and application of confidentiality and privacy rules and laws between different countries, different states of one country and sometimes even between different ethical committees.
- Open source data: in disciplines like genetics there is the tradition to rely on published public repositories data. This is not the case with most of the clinical disciplines.

These challenges impede the realisation of large scale multi-centre exchange of medical data and leads to unnecessarily high costs. It is unrealistic to expect an immediate and conclusive solu-

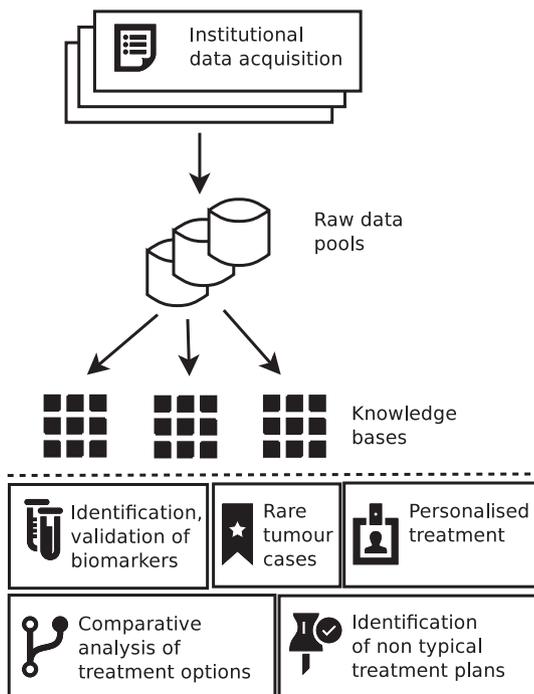


Fig. 1. Large scale multi-centre studies produce raw data pools, which can be used to generate application-specific prediction models or knowledge bases.

tion for the harmonisation of currently used IT research platforms. However, without the efforts of interested researchers, their institutions and radiotherapy organisations, the goal of research data interoperability will remain a continuing challenge and risk to fade away in future plans for setting up studies. The recent innovations in clinical data standardisation [9] together with the European Commission's data protection reform in progress [10] suggest that now it is the ideal point of time to start to analyse and to establish the necessary processes for multi-institutional data exchange. It will require sincere engagement but may result in great benefit to clinical as well as translational cancer research. In future, the interactive data bases might even be used for personalised medicine by means of generating predictions on outcome for individual patients based on analyses of their patient-tumour- and treatment-related data, which would facilitate treatment choice, either by physicians or through shared decision making. Additionally, this initiative could be of great importance from a health economic perspective, by enabling evaluation of efficacy and cost-benefit of different approaches, such as new technologies and/or new combined modality treatments.

Radiotherapy data management

For successful creation of an international data exchange strategy it is necessary to understand the core principles in radiotherapy data management. This section explains why aggregating radiotherapy research datasets is a non-trivial task. It provides details about different types of data pooling system architectures and shows the importance of clinical data and metadata standardisation. In addition to technical concepts it describes the role of information technology in study quality assurance. Data protection issues are addressed taking into account current developments of protection laws in the EU. This section ends with a summary of data pooling and sharing initiatives as well as software platforms as basic for forming an initiative to unify radiotherapy data exchange processes.

Working with radiotherapy research data

Information that is necessary to conduct research in the domain of radiation therapy and oncology is present in various modalities and scattered within diverse information systems. Table 1 provides an overview of possible radiotherapy research data types with their common information management systems [11].

These data sources need to be queried to provide complex datasets for comprehensive data analyses, as depicted in Supplementary Fig. 1. An international effort to promote the interoperable exchange of DICOM images and treatment planning data has been undertaken jointly by clinicians and equipment manufacturers through one of the Integrating Healthcare Enterprise (IHE) initiative [12] profiles. IHE profiles sit on top of existing standards and define detailed rules/workflows for linking medical information systems within an institution.

Gaining in-house clinical IT systems interoperability is important especially with respect to the convenient creation of locally anonymised/pseudonymised datasets. These are managed by institutional data warehouse, which provides a universal access to aggregated research data that are afterwards discoverable under chosen semantic model (ontology). This is why research data warehouses present important components for multi-centre and multi-study data collection and analysis.

Data pooling architectures

The pooling architecture defines how the data are processed, shared, stored and used in a specified system. It is possible to dif-

Table 1
Radiotherapy research data types within their common IT systems.

Information type	Data examples	IT system
Baseline clinical data	Demographics (including co-morbidity and family history), TNM-stage, date of diagnosis, histopathology	HIS, TDS
Diagnostic imaging data	Diagnostic CT, MR and PET imaging	PACS
Radiotherapy treatment planning data	Delineation/structure sets, planning-CT, dose matrix, beam set-up, prescribed dose and fractions	PACS, RIS
Radiotherapy treatment delivery data	Cone beam CTs, orthogonal EPID imaging, delivered fractions	PACS, RIS
Non-radiotherapy treatment data	Surgery, chemotherapy	HIS, TDS
Outcome data	Survival, local control, distant failure, toxicity (including patient reported outcomes), quality of life	EDC, TDS
Follow-up imaging data	Follow-up CT, MR and PET imaging	PACS
Biological data	Sample storage, shipping, tracing and lab results	LIMS
Additional study conduct data	Study design, protocol, eligibility criteria	EDC, CTMS

ferentiate between the following major classes of data pooling models pictured in Fig. 2:

1. Centralised model: giving priority to full control over data, which are logically located in a centralised repository. There is no direct communication between institutions and all processes happen in a central system (e.g. push/ pull transactions, auditing). This leads to a simple architecture, however it raises several questions to be solved, including data privacy and anonymisation, independent access-control to data, Intellectual Property (IP) rights to publish and the security risk of data accumulated in one place. Advantages are that the data are centralised, stored in a virtual storage (cloud data repository), and updating of individual data is straightforward (depending on the to-be-agreed-upon protocols).
2. Decentralised model: prioritises separation of data through institution's autonomous data repositories. Sharing is project-based via direct communication of two or more institutions without any mediator usually as export/ import jobs. Infrastructure information that is necessary to technically enable data exchange is distributed to each location. Data can be stored redundantly (after exchange). One of the challenges is the required interactivity for updating of the federated data whenever information is added; the risk exists that several versions of merged data exist depending on the dates when the data exchange took place.
3. Hybrid model: tries to take the best from centralised and decentralised models. The data exchange is again realised via direct communication of two or more participating institutions. To simplify this communication the central server is used to store infrastructure information necessary for data exchange. The central server can also hold the data model, controlled terminologies and other necessary meta-data to enable the data interoperability within decentralised data exchange.

Given the heterogeneity of currently used IT platforms, decentralised and hybrid approaches should be considered as preferred architectures behind a new international data exchange strategy. Technologically even these solutions could be configured to automate the export/ exchange/ update data processes and thus hide the complexity of the systems and provide a swift and interactive user experience.

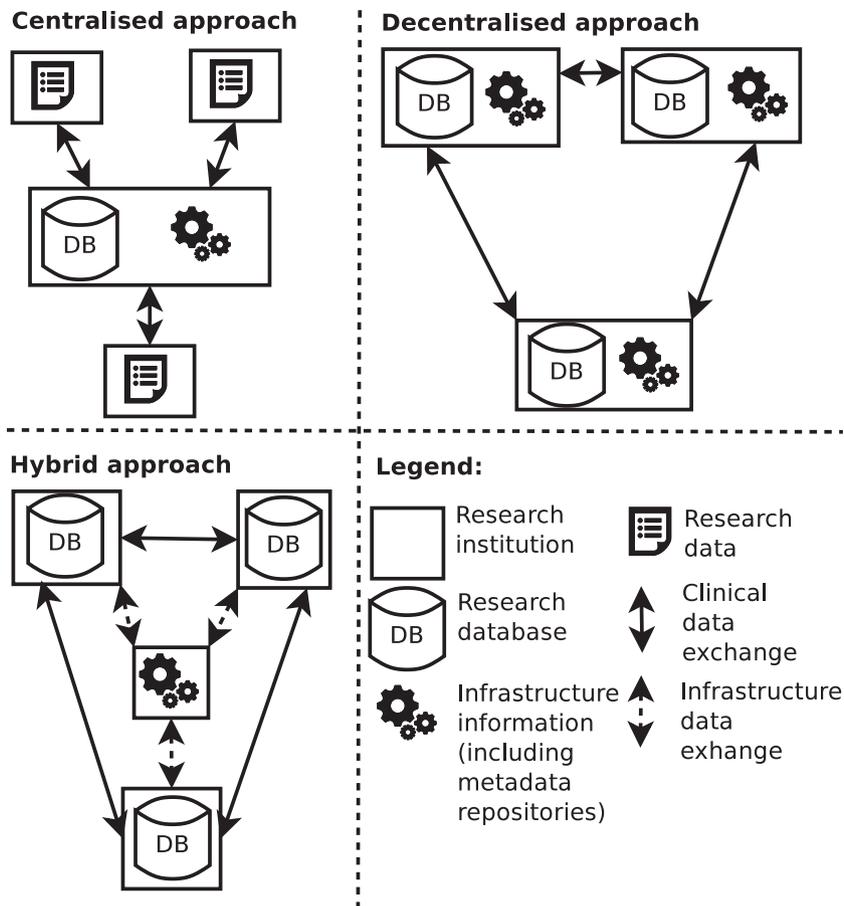


Fig. 2. Schematic drawings of centralised, decentralised and hybrid data pooling models. A centralised approach depends on a central data repository. A decentralised solution consists of a network of sibling repository nodes. A hybrid approach combines a network of decentralised repository nodes with a central infrastructural database.

Additionally in a situation, where locally collected data for legal/ethical reasons cannot be shared with partners, distributed solutions provide more possibilities for advanced data analysis such as the exchange of medical knowledge/models from locally aggregated research datasets. The final hypothesis is then derived from several local models reported by participating institutions. This concept is known as “distributed learning” and its successful application is presented in [13]. This principle can even be taken one step further by setting up an “online learning” environment where the master (merged) knowledge model continuously updates (improves) as more and more patient data are available for analysis.

Fundamental elements for data interoperability

Data interoperability is the key element for a useful data exchange strategy. It can be described as the system’s ability to read and understand information produced by another system. Internationally developed standards are the starting point for achieving interoperability. However, in a real world scenario, the application of standards does not work as a plug-and-play solution. It requires a complex multi-stage process which will make interoperability possible. First of all, data interoperability consists of two main sub-principles:

- **Syntactic interoperability:** focuses on establishing common data formats and exchange protocols. In other words syntactic interoperability is unifying write/ read information processes.

- **Semantic interoperability:** focuses on the proper interpretation of the information. It ensures that the meaning of information is not lost or changed during the data exchange process. This way it makes the information reliable and understandable.

The structure of information in a clinical research domain can be represented by a hierarchical pyramid as depicted in [Supplementary Fig. 2](#). It consists of several layers, where each has its place within the process of making data interoperable.

At the bottom, a standardisation of medical terms leads to the creation of controlled vocabularies, where terms describing medical context are defined. To avoid national (linguistic) names, they often need to have a code representation. When the relationship between defined terms is also captured (e.g. simple parent–child relationship or more complex self-defined relations) the resulting concept is called ontology.

In the middle, there is a formalisation of descriptive information about data fields collected within e.g. CRFs. Data fields collect information represented as basic data type or medical terms from controlled terminology. They also contain meta-data information that is necessary for the data acquisition process (e.g. required value, standardised questions, etc.)

The upper level of the information hierarchy is represented via a concept called the information model. Within the information model, data are composed to form complex data types representing clinical domain real world entities (e.g. study subject, study protocol, etc.).

Achieving this level of information consistency requires substantial efforts however, it would bring a lot of advantages:

- The model completely defines the clinical study process.
- It ensures data and metadata integrity during data exchange.
- It is time-resistant for long term storage, update and usage.

In reality there are multiple implementations of these core contents, because the understanding and perception of information differs within medical domain areas (different point of view of healthcare, clinical research or biology experts). This sustains the need for harmonising and linking activities to allow transparent utilisation of multiple medical information models (see [Supplementary Fig. 3](#)). One example of such initiative is the UML based Biomedical Research Integrated Domain Group (BRIDG) model [14] harmonising Clinical Data Interchange Standards Consortium (CDISC) [15], Health Level 7 (HL7) [16], the U.S. Food and Drug Administration (FDA) and the U.S. National Cancer Institute (NCI) activities. However in view of current technological developments the utilisation of semantic web technologies (also known as “Linked Data”) seems to be a more flexible option. The biggest advantage of semantic web is frictionless linkage of information across multiple information models (semantic web uses ontological representation of information). Leading information model providers like CDISC or NCI, which understand the needs of clinical informatics practitioners, are trying to publish their standards in representations suitable for semantic web (using W3C Resource Description Framework RDF specification).

A strategy for the development, selection and utilisation of standards has to be considered carefully for the purpose of a successful data exchange. Application of standards on a post-facto basis is difficult, time consuming and prone to systematic as well as random errors. That is why it is important to establish up-front defined data collection elements for broader institutional collaboration.

Quality assurance

Information technology can support the process of quality assurance (QA) for collected radiotherapy data. It can be used for “real-time” assessment of treatment plans by peer-review or trial centre [17]. The utilisation of standards will lead to a higher consistency of prospectively collected data but does not automatically improve data quality. Therefore, the quality control lies within the responsibility of designated QA personnel. Correct usage of technologies for study design and conduct should help ensure a certain level of data quality. Information systems can e.g. guide QA according to standardised –to be designed– procedures that define the scope and rules of automatic validation and verification (e.g. subject cannot die before birth, automatic body mass index calculation, etc.). Obviously the automation of QA depends on the level of agreement on definition of such QA procedures. Data pooling fed with low- quality data may lead to big datasets, but their practical usability will be very limited. One way to improve data quality is via establishment of “umbrella protocols” with CRF standardisation, which can be defined and published in vendor independent human and machine readable formats.

Ethics and regulations

Current EU data protection laws do not harmonise rules for health related data processing [10]. The conditions for data utilisation for research differ across countries and sometimes even within regions of one country. This fragmentation causes major problems for international scientific collaborations in medical research. In addition, the interpretation and therefore application of the same rules might lead to a varying conduct at an even smaller scale (ethics committees, hospital).

The European Commission’s planned new data protection reform represents a draft for a Regulation which will replace the existing Data Protection Directive (95/46/EC) and associated Member State legislation [18]. If approved, this reform might bring many benefits. Most importantly it will [10]:

- Unify current EU data protection rules into one regulation,
- Establish one supervised regulation authority for data protection,
- Be applicable to institutions within as well as outside the EU (under the provisions stipulated in Article 3),
- Take direct effect in all EU Member States (the reforms will not have to be transposed) [18].

The original draft Regulation included a requirement for specific and explicit consent for the use and storage of personal data, but provided an exemption for research, subject to certain safeguards in Article 83. The European Parliament’s amendments to Articles 81 (Processing of personal data concerning health) and 83 (Processing for historical, statistical and scientific research purposes) substantially reduce the scope of this research exemption. This means, if implemented in the current version, that the use of personal data in research without specific consent would be prohibited or become impossible in practice [19], which is indeed a major issue for cancer research and even for quality assurance. In practice each patient (or after death his/her relatives), even when the patient has signed general informed consent for scientific evaluation of the data, would need to be re-consented each time a new scientific project is started. On the other hand once a specific informed consent is present, it will be valid throughout the entire EU. Another provision is that the “right to be forgotten” [20] does not apply to scientific research sectors [10].

Because of the possible threat to performing (clinical) research in the EU, a position paper in *Annals of Oncology* [21] was published authored by the European Society for Medical Oncology (ESMO), endorsed by the European Cancer Organisation (ECCO), the European Cancer Patient Coalition (ECPC), the European Middle Eastern & African Society for Biopreservation and Biobanking (ESBB), the European Organisation for Research and Treatment of Cancer (EORTC), the EurocanPlatform, the European Society of Paediatric Oncology (SIOPE), the European Society for Radiotherapy & Oncology (ESTRO), the European Society of Surgical Oncology (ESSO) and the Association of European Cancer Leagues (ECL).

An alternative for international data pooling would be complete anonymisation of patient data. Anonymised patient data are not within the material scope of the Regulation. The problem lies in the fact that anonymised patient data may not be possible as more data elements are being shared on an individual patient.

Data pooling and sharing initiatives

Research and development across the field of medical and clinical informatics is very active including several ongoing data collection and exchange initiatives. Some of these initiatives are providing open access to their deliverables in the form of application platforms, terminologies, guidelines and collected data. It is wise to consider them and if possible to build on them to leverage the experience and already invested resources. [Supplementary Table 1](#) provides a partial overview of software systems used for research data and metadata management [22,23]. Some of these platforms are released as freeware or under open source licenses which make them affordable for all academic research institutions. The selection of existing initiatives and research databases in the field of radiation therapy and oncology is presented in [Supplementary Table 2](#) [24–29]. When defining a common data exchange

strategy in radiotherapy it is necessary to consider involvement of existing initiatives in order to gain broader acceptance.

The next steps for data exchange in radiotherapy

From the information summarised above, it is apparent that it is timely now to initiate broader collaboration for radiotherapy data exchange. Present information technology innovations offer advanced methods for establishing data interoperability and for accelerating the data pooling process. However, the commitment of cancer research institutions is necessary to trigger and harbour the activities that will lead to a formal definition of the data exchange strategy.

This paper would like to promote an agile solution for establishment of a standard data exchange. Agile work is characterised via continuous iterative delivery and validation by prototypes. An international “dummy run” can be set up as a test case/prototype for evaluating the robustness of the data exchange strategy and is foreseen between several of the partners that participated in the workshop. It can also be used as a testing case for each participating institution to prove whether the exchange strategy

specification criteria have been met. A simplified working plan is depicted in Fig. 3.

The commitment and engagement of professionals is mandatory to establish the collaboration model between clinicians, physicists, IT, legal/ethics personnel of the participating institutions (including several other specialisations if available like mathematicians, statisticians etc.). Within the interdisciplinary and inter-institutional cooperation, the medical group professionals will design the set-up of a dummy study for radiotherapy. Initially, it should be minimalistic (with relatively small numbers of patient records) but complete regarding data types necessary for research in radiotherapy. The medical group also has to harmonise data collection elements. A good start can be e.g. utilisation of CRF harmonisation activities from NCI [30] and/or by using the validated terminology proposed by the Global Harmonisation Group and/or using the Linked Data principle. An IT group will analyse currently used IT platforms to find common characteristics across the institutions and define technological solutions that should be proposed for the data exchange strategy. The output should cover aspects like data formats and communication protocols. A legal/ethics group will summarise the pre-conditions necessary to exchange and pool clinical data in compliance with national laws.

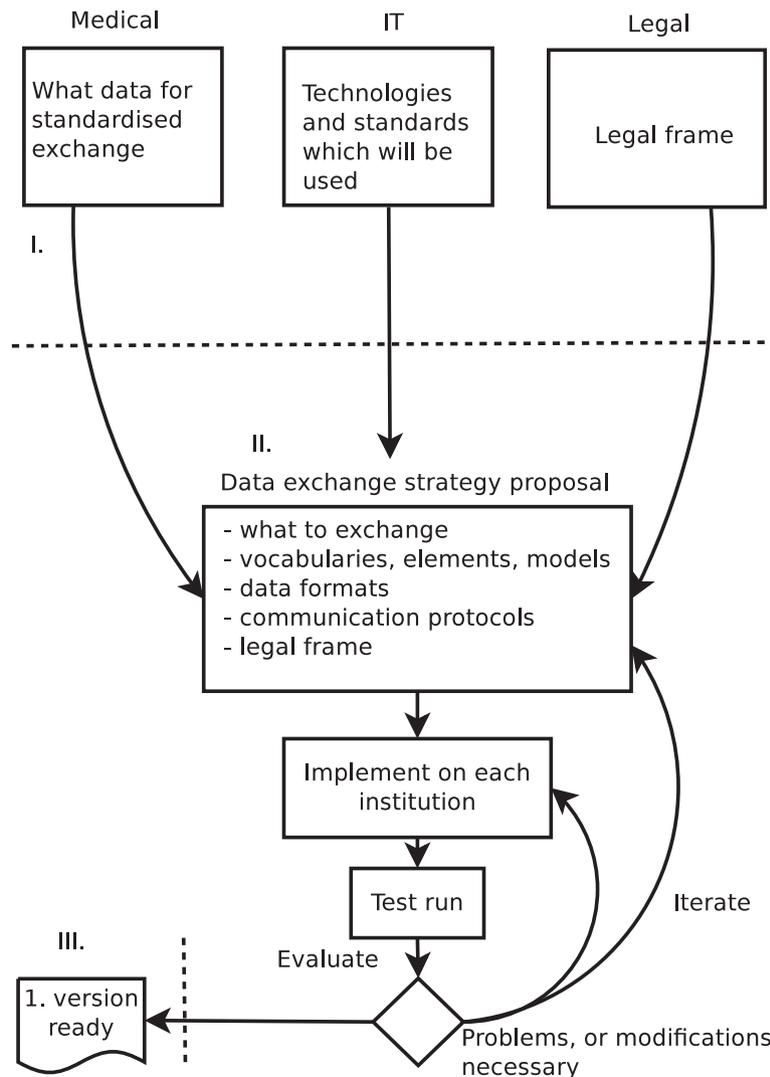


Fig. 3. Simplified working scheme for the creation of a data exchange strategy. The first step is formation of working groups that will prepare a draft strategy. The next step is the implementation of the proposed strategy by participating institutions followed by a dummy run. Finally, the data exchange strategy is officially released with all documentation and guidelines.

The outcome of this preliminary step will be processed into the first proposal for a data exchange strategy. This standard document will formally define the data sharing process with exact specification of data elements, their coding, storage data format and exchange communication protocols. It will formally define and describe import/export scenarios. Implementation of these within the institution's environment will allow the institution to participate in a test run of the dummy study.

After completion of the test, each institution should have one big pooled dataset locally at its site.

The deliverables resulting from the development of the data exchange strategy (in the form of software, documentation, guidelines, data etc.) will be hosted and openly available for all participating institutions. As soon as the first version of the strategy is created a strategy maintenance process will be established to keep strategy elements (e.g. data element repositories) up to date.

With the first aggregated big data pools are in place, an initiative dedicated to data knowledge extraction and biomedical modelling will start in order to develop dedicated decision support tools. Furthermore establishing open public access to data published under DOI data (approach very successfully used in genetics) will make reusing a research data straightforward and as such will stimulate research in radiation oncology.

Conclusion

Creating a robust and usable radiotherapy specific data exchange strategy is challenging but feasible. It requires investments and full commitment of participating institutions. However such a strategy is a fundamental prerequisite to enable multi-centric pooling of cancer research data into common well understandable and reusable datasets. This process will allow seamless collaboration on large-scale international studies and computer-aided analysis of the large amount of high quality clinical research data and will be the basis for rapid knowledge generation in the field of radiotherapy.

The data exchange strategy should be thought of as an evolutionary process where the baseline for collaboration could be exchange of standardised study protocols, data element definitions and clinical or study data together with imaging and treatment plans and rendering open public datasets. The complexity can be gradually increased over time e.g. by allowing information from local knowledge bases to be part of the exchange processes.

Conflict of interest

None declared.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.radonc.2014.10.001>.

References

- [1] Roelofs E, Persoon L, Nijsten S, Wiessler W, Dekker A, Lambin P. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. *Radiother Oncol* 2013;108:174–9.
- [2] Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol* 2012;10:27–40.
- [3] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5.
- [4] Reymen B, van Baardwijk A, Wanders R, Borger J, Dingemans A-MC, Bootsma G, et al. Long-term survival of stage T4N0-1 and single station IIIA-N2 NSCLC patients treated with definitive chemo-radiotherapy using individualised isotoxic accelerated radiotherapy (INDAR). *Radiother Oncol* 2014;110:482–7.
- [5] Stacey D, Légaré F, Col NF, Bennett CL, Barry MJ, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev* 2014;1. CD001431.
- [6] Roelofs E, Engelsman M, Rasch C, Persoon L, Qamhiyeh S, de Ruyscher D, et al. Results of a multicentric in silico clinical trial (ROCO): comparing radiotherapy with photons and protons for non-small cell lung cancer. *J Thorac Oncol* 2012;7:165–76.
- [7] Roelofs E, Persoon L, Qamhiyeh S, Verhaegen F, De Ruyscher D, Scholz M, et al. Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies. *Radiother Oncol* 2010;97:567–71.
- [8] Langendijk JA, Lambin P, De Ruyscher D, Widdler J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.
- [9] CDISC Share. <https://www.cdisc.org/cdisc-share>; 2014.
- [10] Data protection day 2014: full speed on eu data protection reform. https://www.europa.eu/rapid/press-release_MEMO-14-60_en.htm.
- [11] Roelofs E, Dekker A, Meldolesi E, van Stiphout RGPM, Valentini V, Lambin P. International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining. *Radiother Oncol* 2014;110:370–4.
- [12] Integrating the Healthcare Enterprise. <https://www.ihe.net>; 2014.
- [13] Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, et al. Rapid learning health care in oncology – an approach towards decision support systems enabling customised radiotherapy. *Radiother Oncol* 2013;109:159–64.
- [14] Biomedical Research Integrated Domain Group. <https://bridgmodel.nci.nih.gov>; 2014.
- [15] de Montjoe J. Introducing the CDISC Standards: New Efficiencies for Medical Research. CDISC. <https://www.cdisc.org>; 2009.
- [16] Health Level Seven International. <https://www.hl7.org>; 2014.
- [17] Martin J, Frantzis J, Chung P, Langah I, Crain M, Cornes D, et al. Prostate radiotherapy clinical trial quality assurance: how real should real time review be? (A TROC-OCOG Intergroup Project). *Radiother Oncol* 2013;107:333–8.
- [18] Data protection legislation. <https://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Personal-information/Data-protection-legislation/index.htm>; 2014.
- [19] Protecting health and scientific research in the Data Protection Regulation (2012/0011(COD)): position of non-commercial research organisations and academics. http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/WTP055584.pdf; 2014.
- [20] Weber R. The right to be forgotten more than a Pandora's Box? *jipitec n.d.*; 2014.
- [21] Casali PG. Risks of the new EU Data protection regulation: an ESMO position paper endorsed by the European oncology community on behalf of the European Society for Medical Oncology (ESMO) Switzerland. *Ann Oncol* 2014;25:1458–61.
- [22] Breil B, Kenneweg J, Fritz F, Bruland P, Doods D, Trinczek B, et al. Multilingual medical data models in ODM format: a novel form-based approach to semantic interoperability between routine healthcare and clinical research. *Appl Clin Inform* 2012;3:276–89.
- [23] Marcus DS, Olsen TR, Ramaratnam M, Buckner RL. The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 2007;5:11–34.
- [24] Westberg J, Krogh S, Brink C, Vogelius IR. A DICOM based radiotherapy plan database for research collaboration and reporting. *J Phys: Conf Ser* 2014;489:012100.
- [25] Baumann M, Höltscher T, Begg AC. Towards genetic prediction of radiation responses: ESTRO's GENEPI project. *Radiother Oncol* 2003;69:121–5.
- [26] De Ruyscher D, Severin D, Barnes E, Baumann M, Bristow R, Grégoire V, et al. First report on the patient database for the identification of the genetic pathways involved in patients over-reacting to radiotherapy: GENEPI-II. *Radiother Oncol* 2010;97:36–9.
- [27] Melidis C, Bosch WR, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials – Global Harmonisation Group. *Radiother Oncol* 2014;111:327–9.
- [28] Efstathiou JA, Nassif DS, McNutt TR, Bogardus CB, Bosch W, Carlin J, et al. Practice-based evidence to evidence-based practice: building the National Radiation Oncology Registry. *J Oncol Pract* 2013;9:e90–5.
- [29] Meldolesi E, van Soest J, Dinapoli N, Dekker A, Damiani A, Gambacorta MA, et al. An umbrella protocol for standardized data collection (SDC) in rectal cancer: a prospective uniform naming and procedure convention to support personalized medicine. *Radiother Oncol* 2014;5.
- [30] CRF Harmonization and Standardization. <https://wiki.nci.nih.gov/display/CRF/CRF+Harmonization+and+Standardization>; 2014.