

International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining

Citation for published version (APA):

Roelofs, E., Dekker, A., Meldolesi, E., van Stiphout, R. G. P. M., Valentini, V., & Lambin, P. (2014). International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining. *Radiotherapy and Oncology*, 110(2), 370-374.
<https://doi.org/10.1016/j.radonc.2013.11.001>

Document status and date:

Published: 01/02/2014

DOI:

[10.1016/j.radonc.2013.11.001](https://doi.org/10.1016/j.radonc.2013.11.001)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

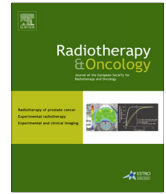
www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Data sharing

International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining



Erik Roelofs^{a,1}, André Dekker^{a,*,1}, Elisa Meldolesi^b, Ruud G.P.M. van Stiphout^a, Vincenzo Valentini^{b,1}, Philippe Lambin^{a,1}

^aDepartment of Radiation Oncology (MAASTRO), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center (MUMC+), The Netherlands;

^bDepartment of Radiation Oncology, Policlinico Universitario Agostino Gemelli, Rome, Italy

ARTICLE INFO

Article history:

Available online 3 December 2013

Keywords:

Medical informatics
Knowledge engineering
Machine learning
Data mining
Open source software

ABSTRACT

Extensive, multifactorial data sharing is a crucial prerequisite for current and future (radiotherapy) research. However, the cost, time and effort to achieve this are often a roadblock. We present an open-source based data-sharing infrastructure between two radiotherapy departments, allowing seamless exchange of de-identified, automatically translated clinical and biomedical treatment data.

© 2013 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 110 (2014) 370–374

Sharing data across institutions is required for multi-institutional radiotherapy research [1]. Besides exchanging data for specific research projects, there is a recognized need to establish a culture of data pooling both within the radiotherapy [2] and the broader cancer community [3]. For the transition from population based treatment options (where “one size fits all”) toward personalized medicine we are increasingly depending on decision support systems that require large heterogeneous datasets [4–8]. Randomized controlled trials hardly offer such data with only 3% of adult cancer patients included in trials [9–11]. However, aggregating routinely collected real-time biomedical patient data and innovative “rapid-learning” research techniques allow us to use the knowledge of the masses for the benefit of the individual [3,12,13].

Medical informatics driven research, for instance in the field of predictive modeling, requires a large amount of data to provide sufficient statistical power to act as acceptable decision supporting tools. Furthermore, another substantial amount of data is needed for validation of the models, preferably by external datasets.

This brings up some stringent and challenging demands on the quality as well as the quantity of the data. Data of inferior quality do not improve by pooling it with other data. It can actually worsen the value of good datasets. Work by the Quality Assurance Review Center (QARC; <http://www.qarc.org>) and a review by the

EORTC show the importance of proper Quality Assurance (QA) programs in collaborative efforts and the long history thereof in the field of Radiotherapy [14–16].

Another challenge for data-sharing initiatives in the field of biomedical research is that the investigated data are often multifactorial, comprising of laboratory data, diagnostic and clinical imaging and treatment outcome data, among others. Combining these data securely can be troublesome, even when sharing between departments within the same institution, let alone when between institutions, especially international ones.

Furthermore, in many research projects, dedicated data management staff are required to translate and copy data into trial-specific case report forms and/or dedicated IT staff are needed to de-identify DICOM images or build databases that are suitable for machine learning and data mining techniques. However, without dedicated staff, the sheer amount of time it takes to collect, de-identify and share data often is a roadblock to participation in clinical research. With many research projects not or underfunded, especially in the initiation phase, one requires existing staff to balance other duties with these research requests. This causes the process of data sharing to take a long time, despite the cooperation and willingness of everyone involved.

In this technical report, we describe one way to quickly build a low cost, infrastructure that makes sharing of data easier between two institutions wishing to work together, but having different IT systems. This infrastructure was implemented to share data from the Policlinico Universitario Agostino Gemelli in Rome, Italy (Gemelli) to the MAASTRO Clinic in Maastricht, the Netherlands (MAASTRO) to facilitate research projects such as the Thunder clinical trial (NCT00969657, <http://ClinicalTrials.gov>) and “knowledge

* Corresponding author. Address: Department of Radiation Oncology (MAASTRO), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Center (MUMC+), Dr. Tanslaan 12, 6229 ET Maastricht, The Netherlands.

E-mail address: andre.dekker@maastro.nl (A. Dekker).

¹ These authors contributed equally to this work.

engineering” research using data-mining and machine learning techniques to develop predictive models for various cancer sites (<http://www.predictcancer.org>).

Material and methods

Clinical data sources

In general, radiotherapy research requires various types of information:

- Clinical data (e.g. demographics, TNM-stage, date of diagnosis, histopathology, etc.).
- Diagnostic imaging data (e.g. diagnostic and follow-up PET, CT and MR imaging).
- Radiotherapy treatment planning data (e.g. delineation, planning-CT, dose matrix, beam setup, prescribed dose and fractions).
- Radiotherapy treatment delivery data (e.g. cone beam CTs, Orthogonal EPID imaging, delivered fractions).
- Non-radiotherapy treatment data (e.g. surgery, chemotherapy).
- Outcome data (e.g. survival, local control, toxicity).

Typically, in a radiotherapy department, this information is scattered across a number of data sources from a variety of vendors, which do not necessarily share the same patient identification number. In the case of Gemelli, the data sources were as given in Table 1.

Data model

For the research database (DB), a patient-centric data model was designed that would enable queries for both medical data and the existence of imaging data (Fig. 1). A simple data model was deliberately chosen to allow easy identification of core disease characteristics but with most information in the form of lists of performed procedures as well as performed imaging study and series.

De-identification

A coding scheme was employed in which a secure database is maintained that holds the link (Key) between a unique random patient identification code (ID) and all directly identifying data (ID's as used in the clinical data sources, name, birth date etc.). This Key is maintained by local hospital personnel and only accessible from within the firewall of the hospital. In the research DB, the patient is only identified by the research ID. Rather than applying

an irreversible anonymization method to the patient data, we used “coding” or “pseudo-anonymization” to enable extending the datasets with additional information at a later stage, which would otherwise be impossible to do.

Some data elements were not de-identified as they were considered to be important for the research while they only carry a small risk of identifying a patient. The de-identification scheme was reviewed and approved by the local ethics authority. The elements that were kept were: exact dates of various procedures (including treatments), exact dates of diagnosis & death, DICOM UIDs and CT and MR imaging of the head.

Clinical terms and translation

To convert Italian to English standard terms, SNOMED Clinical Terms [17] were used as the dictionary. SNOMED CT is considered as the most comprehensive multilingual medical terminology in the world. A separate database was maintained in which local terms were mapped to the SNOMED-CT concepts and both the preferred term and the concept ID were stored.

Research hardware & software

On a research workstation (Windows 7 64-bit, Intel Xeon, 2.53 GHz, 4 GB RAM) the following software was installed: SQL Server 2008 (free Express version, Microsoft, Redmond, WA); Clear Canvas Image Server and Workstation (both free and open source, Clear Canvas, Toronto, Canada); DCMTK DICOM toolkit (free and open source, Offis, Oldenburg, Germany); RSNA Clinical Trial Processor (CTP) (free and open source, RSNA, Oakbrook, IL) and Matlab Compiler Runtime (MCR) engine (free, Mathworks, Natick, MA).

Clear Canvas Image Server is a PACS and was installed with a temporary partition holding identifiable DICOM headers and a research partition holding only de-identified DICOM objects. The Clear Canvas Workstation was used for DICOM import of the Nuclear Medicine department's optical disks. SQL Server 2008 was used to host the mentioned databases as well as the database of the Clear Canvas Image Server.

Finally, for data synchronization a variety of SQL scripts was designed and run through the command line interface. The MCR engine was used to run compiled custom Matlab code in which the DCMTK toolkit was called. The CTP package was used to build a de-identification pipeline (DICOM import → de-identification → DICOM Export) and a file export pipeline (DICOM Import → File Export to shared directory). CTP allows customizable de-identification settings through a web interface or by directly editing an XML configuration file.

Table 1
Data sources at Gemelli.

Data type	Data format	Database	Department	Name, Vendor
Clinical Outcome Non-RT treatment	Text	SQL database	Multiple	Spider, Opengraph (local development)
Diagnostic imaging	DICOM-CT DICOM-MR DICOM-PT	DICOM server	Radiology	Careview, Kodak
	DICOM-PT	Optical Disks	Nuclear Medicine	PET workstation, Philips
RT treatment planning	DICOM-CT DICOM-RTDOSE DICOM-RTIMAGE DICOM-RTPLAN DICOM-RTSTRUCT	DICOM server	Radiotherapy	Aria, Varian
RT treatment delivery	Text DICOM-RTRECORD DICOM-RTIMAGE	Sybase Database DICOM server	Radiotherapy	Aria, Varian

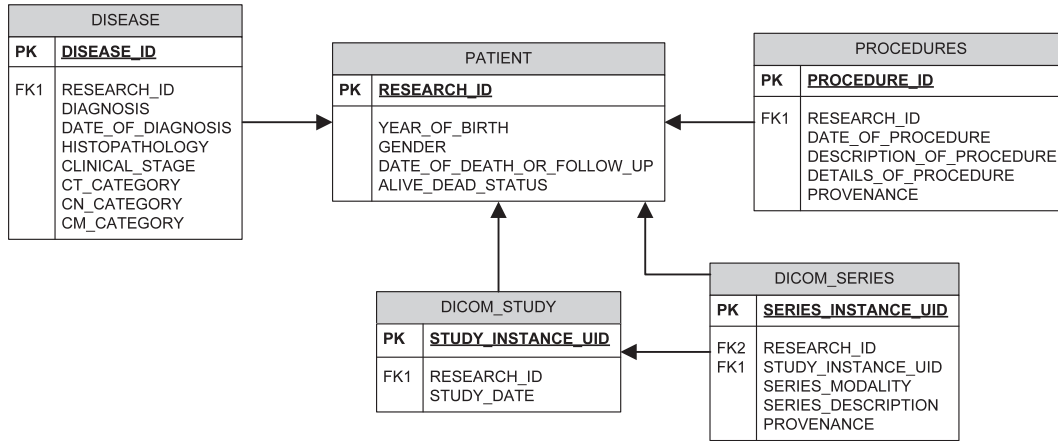


Fig. 1. Data model of the research database (ResearchDB). (PK = Primary Key; FK = Foreign Key).

Synchronization mechanism – text data

The name, birth date, hospital identification number and identification numbers for all data sources were inserted using an SQL query into the secure Key DB for all patients with lung or rectum cancer that had no entry yet. Upon insertion, the patient received a unique research ID, which is a random, non-sequential 8-digit number. Note that the Key DB is a permanent database and that patients will keep the new ID assigned to them.

A second SQL query mapped local terms for demographics, disease characteristics, procedures and clinical findings to a SNOMED term using the dictionary and inserted this information into the Patient, Disease, Procedure and Clinical Finding tables of the research DB.

The above two SQL queries were scheduled to run nightly as to minimize interference with the clinical work, although testing did not indicate any effect of these queries on the clinical process.

Synchronization mechanism – DICOM data

In compiled Matlab code, a query on the Key DB resulted in a list of patient identification numbers for the DICOM servers. Using DCMTK commands, a so-called DICOM C-FIND request was then sent to each DICOM server, which requested the unique identifiers for all series for a given patient being of the CT, PT, MR, RTPLAN, RTSTRUCT, RTDOSE, RTRECORD or RTIMAGE modality. In Matlab, these DICOM data were compared to the existing data in the research PACS. When the data was not yet available, the Matlab code

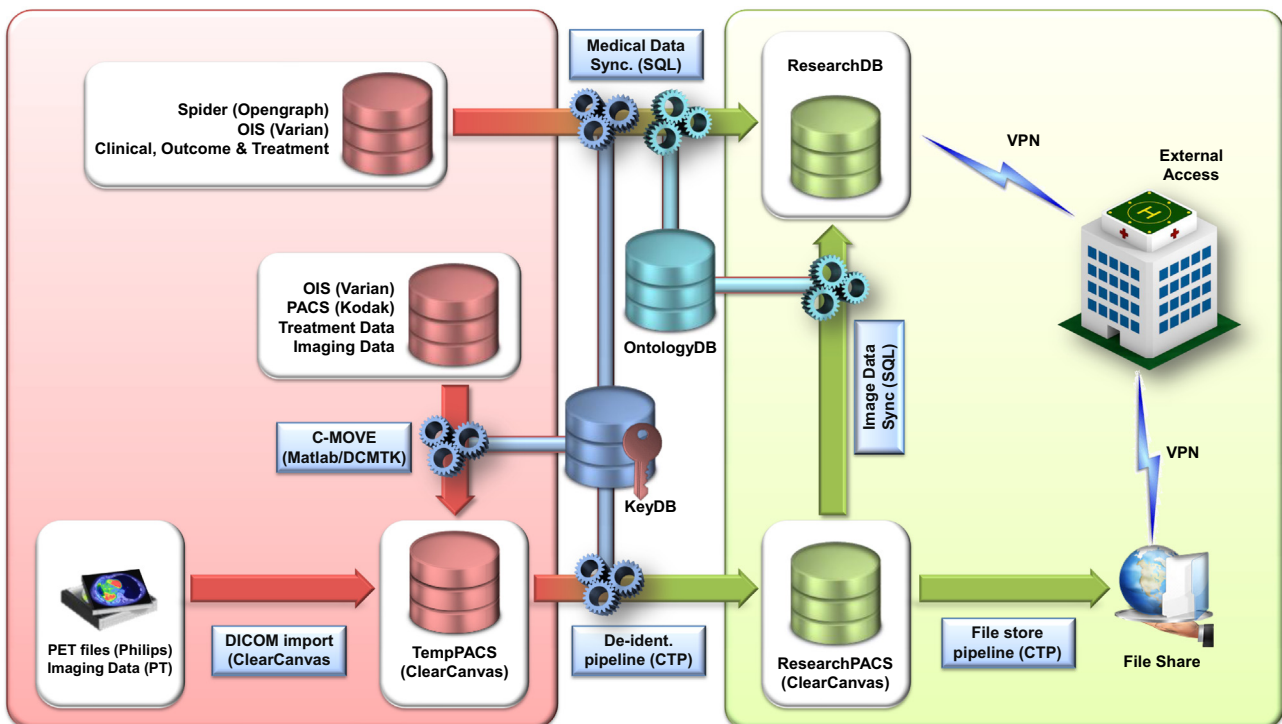


Fig. 2. Overview of data sources, flow and external access. Terms as mentioned in the table below:

Term	Meaning
CTP	Clinical Trial Processor
KeyDB	Secure key database linking the random research patient identifier (ResearchID) and original patient data
OntologyDB	Database storing preferred terms and the concept IDs after mapping local terms to the SNOMED CT concepts
ResearchDB	Research database holding medical data and imaging meta data
ResearchID	Random patient identification code
ResearchPACS	Research PACS partition holding only de-identified DICOM objects
SeriesInstanceUID	Unique series identifier for all images in a series for a given patient
SNOMED-CT	Systematized Nomenclature of Medicine – Clinical Terms
TempPACS	Temporary PACS partition holding identifiable DICOM headers
UID	Unique Identifier

issued a DICOM C-MOVE request to transfer data from the DICOM Server to the private temporary partition of the research PACS.

In a similar manner, a separate Matlab executable compared DICOM data between the temporary and research PACS. For new datasets, a C-MOVE request was issued to transfer these from the temporary PACS to the CTP de-identification pipeline, de-identified the data and stored them in the research PACS. The data synchronization process is depicted in Fig. 2.

Data retrieval

To securely transfer data from the research DB and PACS to a remote location, first a VPN connection was established. VPN was chosen, as this is the *de facto* standard to create a so-called “tunnel” (a secured one-to-one connection via the public Internet). An SQL query was then issued to the research DB (and only this DB). For the DICOM data, direct retrieval was not possible due to security concerns. Instead, a C-MOVE request had to be sent (e.g. using DCMTK commands) to send images from the research PACS to a CTP file-share pipeline. This pipeline temporarily stored the requested images in a shared location from where the images could be copied.

Legal and ethical

A collaboration and data transfer agreement was signed which describes the type of data, the permitted use and the protection of the data. This agreement was submitted to and approved by the local ethics authority. An example agreement can be found online in the supplemental data of this article at (<https://www.cancerdata.org/10.1016/j.radonc.2013.11.001>).

Results and discussion

In this report, we have described one way to build a data-sharing infrastructure quickly and at low cost, which makes sharing of data easier between two institutions wishing to collaborate and exchange data. Starting with contractual arrangements on the use and protection of the data, it is possible to build a data infrastructure that uses free or open source products and that can address common problems such as data de-identification, the need to share DICOM RT objects and the need for a common language/ontology.

The data-sharing infrastructure was implemented in two months (mid 2010). As of July 2012 already, the research DB holds about 1500 lung, 1500 rectum, 1000 head & neck, 300 cervix and 200 pancreas cancer patients and the research PACS holds about 13,500 de-identified DICOM studies using a total of 3 TB storage.

All new rectum and lung cancer patients are added automatically to the research DB and PACS through the nightly synchronization, with a projected yearly increase of 500–600 patients. With a Good Clinical Practice (GCP) compliant [18] data warehouse

solution that was previously reported in a unicentric setting, these databases could easily provide high quality dose–volume data, for instance [19]. This would significantly reduce the amount of time needed by data managers to support existing or new clinical trials, especially in a multicentric setting.

Unexpected problems were that patients were known by different IDs in the various information systems. One reason being that department-specific vendors pose different restrictions (e.g. no leading zero's, or limit on the number of digits) so that the central hospital ID could not be applied. Furthermore, information from external hospitals was often identified by the external hospital patient ID. When personal identification was achievable, for instance when patients brought images on DVD from another hospital, the data were combined, but linking based on name or birth date alone was not done. The level of detail when removing identifiable information or the decision to maintain this is a matter of the local ethics committee and should be considered per study cohort.

The infrastructure has been used to pool two existing datasets from both institutes [20] and is actively used in the Thunder trial (NCT00969657, <http://ClinicalTrials.gov>) in which multiple PET/CT scans before, during and after concurrent chemo-radiation are used to predict the likelihood of a complete pathological response in rectum cancer patients [21]. A second NIH funded trial using the infrastructure is the “Radiomics” [22] trial (NCT01302626) in which advanced analysis of CT and PET images of lung cancer patients are correlated with genetic expression and outcome.

Another active application of the data-sharing infrastructure is “knowledge engineering” research. In this type of research, large datasets are required in combination with machine learning techniques to learn predictive models for outcome in cancer patients such as we described before [7,23,24]. To prove the accuracy of the model and prevent erroneous conclusions due to over-fitting external validation against independent datasets is crucial. The reported framework facilitated such validation for published predictive models in rectum cancer [5,20].

Finally, the infrastructure is used for the initiation phase of prospective clinical trials to test hypotheses on historical data or to estimate the number of patients that might be included in a new trial given a certain patient selection. Examples showing the power of the infrastructure are (a) outcome and number of patients older than 75 treated radically with limited stage small lung cancer, (b) patients with a PET/CT scan before and during a course of radiotherapy of lung cancer or (c) occurrences of dry mouth in a cohort of 1000 patients being treated with conformal versus intensity-modulated radiotherapy.

Our infrastructure has shown to solve the issue of sharing large amounts of medical and imaging data for research purposes. The automatic transfer from the clinical data flow to a multicentric research environment facilitates daily aggregation of information and offers valuable data for mining and validation. The basic infrastructure is easily extended with more tumor sites or additional relevant procedures. These extensions require modifications of

the nightly synchronization and mapping table, which only take minutes. In a manual data sharing method, adding a data element or disease site would require dedicated staff going through all (paper) charts again twice according to GCP guidelines.

The presented method should be seen as an *ad hoc* data sharing infrastructure and an initial step to collaborate with two institutions. Previously, we described the MISTIR framework for *in silico* clinical trials in Radiotherapy, which enables multiple institutions to securely exchange research data under a strict protocol and QA program [25]. One example of a QA measure we applied for MISTIR as well as for the Thunder trial is to calibrate the standardized uptake values (SUV) from the PET images to enable reliable SUV-driven segmentation among different institutions [26].

In a joint initiative from the Center for Translational Molecular Medicine (CTMM), the Dutch Cancer Society, the Dutch Heart Foundation, the Netherlands Federation of University Medical Centers, the Netherlands Bioinformatics Centre, the String of Pearls Initiative (PSI) and the Netherlands eScience Center (NLeSC), among others, a medical informatics project called Translational Research IT (TraIT) was started to support the transition of highly promising medical research of twenty-one CTMM projects into clinical care (<http://www.ctmm-traits.nl>). Within the TraIT program, we have launched the national BioMedical Imaging Archive (<http://www.bmia.nl>) to store and exchange clinical imaging data using the open-source grid-based caBIG™ tools offered by the NCI, NIH [27–29]. Furthermore, TraIT has set up a national instance of OpenClinica (OpenClinica LLC, Waltham, USA) to support electronic data capture for large inter-institutional trials among most of the university medical centers (<http://www.openclinica.nl>). Efforts are currently undertaken to offer direct image access from the electronic case report forms.

We hope that prototyping projects such as MISTIR and the one described here and extensions such as TraIT and the advanced federated EuroCAT computer network (<http://www.eurocat.info>) with participating institutions in Belgium, Germany and The Netherlands will contribute to global initiatives that offer data-sharing and research capabilities for the entire oncology sector.

Acknowledgements

Authors acknowledge financial support from the QuIC-ConCePT project, which is partly funded by EFPI A companies and the Innovative Medicine Initiative Joint Undertaking (IMI JU) under Grant Agreement No. 115151. Authors also acknowledge financial support from the National Institutes of Health (NIH-USA U01 CA 143062-01, Radiomics of NSCLC), the CTMM framework (AIRFORCE project, Grant 030-103), EU 6th and 7th framework program (METOXIA, EURECA, ARTFORCE), euroCAT (IVA Interreg), NGI Pre-Seed Grant (n° 93612005), Kankeronderzoekfonds Limburg from the Health Foundation Limburg and the Dutch Cancer Society (KWF UM 2008-4210, KWF UM 2011-5020).

Appendix A. Appendix: Supplemental Data

To support the open source and open data concept, we published supplementary data online at (<https://www.cancerdata.org/10.1016/j.radonc.2013.11.001>).

References

- [1] Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *The Lancet* 2011;377:1633–5.
- [2] Deasy JO, Bentzen SM, Jackson A, et al. Improving normal tissue complication probability models: the need to adopt a “Data-Pooling” culture. *Int J Radiat Oncol Biol Phys* 2010;76:S151–4.

- [3] Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol* 2010;28:4268–74.
- [4] Lambin P, Petit SF, Aerts HJWL, et al. The ESTRO Breur Lecture 2009. From population to voxel-based radiotherapy: exploiting intra-tumour and intra-organ heterogeneity for advanced treatment of non-small cell lung cancer. *Radiother Oncol* 2010;96:145–52.
- [5] Valentini V, van Stiphout RGPM, Lammering G, et al. Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the Basis of European Randomized Clinical Trials. *J Clin Oncol* 2011;29:3163–72.
- [6] Egelmeier AGTM, Velazquez ER, de Jong JMA, et al. Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiother Oncol* 2011;100:108–15.
- [7] Dehing-Oberije C, De Ruyscher D, Petit S, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010;97:455–61.
- [8] Valentini V, Lambin P, Myerson RJ. Is it time for tailored treatment of rectal cancer? From prescribing by consensus to prescribing by numbers. *Radiother Oncol* 2012;102:1–3.
- [9] Movsas B, Moughan J, Owen J, et al. Who enrolls onto clinical oncology trials? A radiation patterns of care study analysis. *Int J Radiat Oncol Biol Phys* 2007;68:1145–50.
- [10] Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA* 2004;291:2720–6.
- [11] Grand MM, O'Brien PC. Obstacles to participation in randomised cancer clinical trials: a systematic review of the literature. *J Med Imaging Radiat Oncol* 2012;56:31–9.
- [12] Lambin P, van Stiphout RGPM, Starmans MHW, et al. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat Rev Clin Oncol* 2012;10:27–40.
- [13] Lambin P, Roelofs E, Reymen B, et al. “Rapid learning health care in oncology” – an approach towards decision support systems enabling customised radiotherapy. *Radiother Oncol* 2013;109:159–64.
- [14] Peters LJ, O'Sullivan B, Giralt 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.
- [15] Weber DC, Poortmans PMP, Hurkmans CW, et al. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. *Radiother Oncol* 2011;100:150–6.
- [16] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: evidence-based medicine in radiation therapy. *Radiother Oncol* 2012;105:4–8.
- [17] International Health Terminology Standards Development Organization. SNOMED Clinical Terms (SNOMED CT) n.d.
- [18] European Medicines Agency – Inspections – Good-clinical-practice compliance, 2013.
- [19] Roelofs E, Persoon L, Nijsten S, et al. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. *Radiother Oncol* 2013;108:174–9.
- [20] Van Stiphout RGPM, Lammering G, Buijsen J, et al. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging. *Radiother Oncol* 2011;98:126–33.
- [21] Janssen MHM, Öllers MC, Riedl RG, et al. Accurate prediction of pathological rectal tumor response after two weeks of preoperative radiochemotherapy using 18F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging. *Int J Radiat Oncol Biol Phys* 2010;77:392–9.
- [22] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012;48:441–6.
- [23] Dehing-Oberije C, Yu S, De Ruyscher D, et al. Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2009;74:355–62.
- [24] Dehing-Oberije C, Aerts H, Yu S, et al. Development and validation of a prognostic model using blood biomarker information for prediction of survival of non-small-cell lung cancer patients treated with combined chemotherapy and radiation or radiotherapy alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys* 2011;81:360–8.
- [25] Roelofs E, Persoon L, Qamhiyeh S, et al. Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies. *Radiother Oncol* 2010;97:567–71.
- [26] Öllers M, Bosmans G, van Baardwijk A, et al. The integration of PET-CT scans from different hospitals into radiotherapy treatment planning. *Radiother Oncol* 2008;87:142–6.
- [27] Langella S, Oster S, Hastings S, et al. The Cancer Biomedical Informatics Grid (caBIG™) Security Infrastructure. *AMIA Annu Symp Proc* 2007;2007:433.
- [28] Oster S, Langella S, Hastings S, et al. caGrid 1.0: an enterprise grid infrastructure for biomedical research. *J Am Med Inform Assoc* 2007;15:138–49.
- [29] Buetow KH. An infrastructure for interconnecting research institutions. *Drug Discover Today* 2009;14:605–10.