Maastricht University

Clinical data science in Radiotherapy

Citation for published version (APA):

van Soest, J. P. A. (2018). *Clinical data science in Radiotherapy: data extraction and quantitative analysis*. Datawyse / Universitaire Pers Maastricht. https://doi.org/10.26481/dis.20181128js

Document status and date:
Published: 01/01/2018

DOI:
10.26481/dis.20181128js

Document Version:
Publisher's PDF, also known as Version of record

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.

Download date: 25 Oct. 2020

Valorization addendum

The activities related to valorization can belong to several (overlapping) themes: (a) Knowledge dissemination, (b) Societal exploitation and (c) Economic exploitation. In this chapter, we will bring this thesis into perspective of valorization, categorized into these three themes.

Knowledge dissemination

Next to scientific publications, we can perform knowledge dissemination among (academic) peers by different methods. Software products developed in this thesis are in general online available, and ready to try for other researchers. These software products are:

- The Radiation Oncology Ontology (ROO): developed as part of **Chapter 6**, with public examples available (https://github.com/RadiationOncologyOntology/ROO).
- The Semantic DICOM (SeDI) ontology: developed as part of **Chapter 7** to extract metadata from DICOM images, and make the complete metadata available for querying (http://bioportal.bioontology.org/ontologies/SEDI)
- The DataFAIRifier: an open-source platform for data extraction from clinical databases and DICOM images. This platform is based on the insights from **Chapters 4-8** (https://github.com/maastroclinic/DataFAIRifier).
- The PyTaskManager: an open-source infrastructure for distributed learning, and as being used in **Chapter 13** (https://github.com/PersonalHealthTrain /PyTaskManager).
- The SparqlExampleR repository: a codebase showing how to develop distributed learning applications in the statistical programming language R. Currently this application retrieves univariate distributions for all variables available in a cohort, in relation to **Chapter 5** (https://github.com/DistributedRapidLearning/SparqlExampleR)

Although distribution of software applications is a logical consequence of scientific publications, we also need tutorials and trainings to teach peers how to use these software components. One of these examples is a pre-meeting course organized and conducted at the Practical Big Data Workshop 2018 in Ann Arbor, Michigan, United states. During this one-day course, we trained participants how to use the DataFAIRifier to convert clinical and DICOM data (typically used within radiotherapy) into a format useful for scientific research. At the end of course, we were able to perform a distributed learning run among 8 participants. The contents of this workshop are available at https://github.com/jvsoest/PBDW2018_hackathon, whereas the distributed learning was performed using the PyTaskManager, in combination with the code publicly available to perform distributed learning of a support vector model (SVM; a specific type of machine learning model see https://github.com/DistributedRapidLearning /DistributedSVM/tree/feature/docker_dl).

Societal exploitation

Translational research

Although knowledge dissemination is valuable to drive the academic fields of clinical data science and radiotherapy forward, it does not exempt us from applying these techniques in clinical practice. This is also known by the "Valley of Death" of medical innovation, where many software products are developed in research, however never make it to clinical practice. One of the reasons for this "Valley of Death" is the growing divide between researchers (without clinical activities) and clinical practice (without research activities). An umbrella term to perform research to investigate tools for clinical practice is called translational research. In this thesis, we attempted to address translational research two times, by performing clinical evaluations of research (software) products to support clinical practice. For example, Chapter 12 evaluates the time reduction of clinical experts to delineate organs and tissues in medical images, resulting in more time for other tasks. In this example, the direct affected stakeholder is the clinical expert, however indirect the patient and society as well. By saving time on delineations, more time would become available for other critical tasks in the chain of tasks needed for patient treatment.

Societal education

Next to the practical and translational aspect, tools for distributed learning can be a back-end infrastructure to feed decision aids with the most accurate numbers for treatment outcomes. For example, based on the specific patient characteristics, we developed a dashboard to compare current patient characteristics to previous treated patients in multiple hospitals. When results are presented in an understandable representation in terms of treatment (adverse) effects, clinicians can use this dashboard to retrieve a prognosis for a new patient.

Furthermore, various clinical decision aids are in development to educate patients and relatives about the disease at hand, and what treatment options are available, including their consequences (for example, see http://www.beslissamen.nl). This information is currently filled by information from scientific articles, however could in the future be regularly updated with information from a network of hospitals participating in a distributed learning effort.

Societal expectation management

Next to the societal exploitation, we need to consider expectation management of the society as well. Information Technology has changed our society in the last 15-20 years, and it is expected to work for medicine as well. Unfortunately, we are not (yet) able to transfer all patient data between healthcare institutes. This is mainly due to the nature of medical data. Medical images, produced by a scanner and software, are transferable between institutes. This is due to the available DICOM standard (although with variations in implementations), and the little interference possible or needed by humans. Unfortunately, such a standard does not exist on a detailed level for (structured) electronic medical record data, or departmental information systems (such as an oncology/radiotherapy information system). Every hospital and department can decide how to structure their own data, resulting in incompatible data structures and terminologies. Many initiatives are attempting to address this issue; however, these initiatives are complex and need to consider stakeholders on the executive, medical and technical level, both for internal and transmural alignment.

Economical exploitation

This thesis can be related to four different economical exploitations. Varying customers from hospitals, research institutes to other businesses as well.

First, the clinical evaluation in **Chapter 12** was conducted to test the performance of a deep learning contouring tool, developed by a commercial company (MIRADA Medical Ltd., United Kingdom). After this evaluation, this company has brought the product to market. Hence, hospitals can purchase this software, and therefore contributes to the economical theme of valorization.

Second, we co-developed the software to extract metadata from medical (DICOM) images with a commercial company (Sohard Software GmbH, Germany). Enabling us to query the complete metadata at once for a specific patient as described in **Chapter 7**. This tool is further used in research projects and is marketed as a business-to-business software product, which can be implemented by hospitals, or providers of hospital software (e.g. a Picture Archiving and Communication System software vendor) to enable more advanced questions on imaging metadata.

Third, we developed algorithms (see https://github.com/DistributedRapidLearning) to run on a distributed learning infrastructure developed by an international commercial company (Varian Medical Systems). This infrastructure is publicly available as a free-for-use infrastructure, however requires a “first right to negotiate” to buy/license intellectual property (IP) on the knowledge gathered by researchers using this infrastructure. Hence, it can be an economical exploitation for universities, to finance further scientific research. At the same time, the company (being one of the largest radiotherapy device manufacturers) can catalyze the use of the gained IP, however with the risk of vendor lock-in.

Fourth and maybe foremost, this thesis contributed to the start of a spin-off company (Medical Data Works B.V.), focusing on support, maintenance and hosting of medical (research) software. Next to the knowledge dissemination described above, healthcare institutes want support contracts for software, to guarantee uptime and operational continuity. Universities in general do not have a core-business of delivering support and maintenance for developed software. This would impact the academic output over time (as researchers are slowed down by the time needed for support/maintenance) and would add a risk of liability for this software. Hence, in this spin-off company we are performing these tasks, relieving the academic institute from this non-scientific burden.