F1000Research

Check for updates

SOFTWARE TOOL ARTICLE

REVISED Radtools: R utilities for convenient extraction of medical image metadata [version 3; peer review: 2 approved]

Previously titled: Radtools: R utilities for smooth navigation of medical image data

Pamela H. Russell ⓘD, Debashis Ghosh ⓘD

Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO, 80045, USA
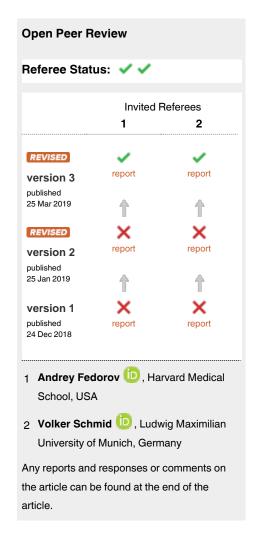
Abstract

The radiology community has adopted several widely used standards for medical image files, including the popular DICOM (Digital Imaging and Communication in Medicine) and NIfTI (Neuroimaging Informatics Technology Initiative) standards. These file formats include image intensities as well as potentially extensive metadata. The NIfTI standard specifies a particular set of header fields describing the image and minimal information about the scan. DICOM headers can include any of >4,000 available metadata attributes spanning a variety of topics. NIfTI files contain all slices for an image series, while DICOM files capture single slices and image series are typically organized into a directory. Each DICOM file contains metadata for the image series as well as the individual image slice.

The programming environment R is popular for data analysis due to its free and open code, active ecosystem of tools and users, and excellent system of contributed packages. Currently, many published radiological image analyses are performed with proprietary software or custom unpublished scripts. However, R is increasing in popularity in this area due to several packages for processing and analysis of image files. While these R packages handle image import and processing, no existing package makes image metadata conveniently accessible. Extracting image metadata, combining across slices, and converting to useful formats can be prohibitively cumbersome, especially for DICOM files.

We present radtools, an R package for convenient extraction of medical image metadata. Radtools provides simple functions to explore and return metadata in familiar R data structures. For convenience, radtools also includes wrappers of existing tools for extraction of pixel data and viewing of image slices. The package is freely available under the MIT license at GitHub and is easily installable from the Comprehensive R Archive Network.

Keywords
Medical imaging, DICOM, NIfTI, R package

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
REVISED version 3 published 25 Mar 2019	✔ report	✔ report
	↑	↑
REVISED version 2 published 25 Jan 2019	✘ report	✘ report
	↑	↑
version 1 published 24 Dec 2018	✘ report	✘ report

1 **Andrey Fedorov** ⓘD , Harvard Medical School, USA

2 **Volker Schmid** ⓘD , Ludwig Maximilian University of Munich, Germany

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the International Society for Computational Biology Community Journal gateway.

This article is included in the RPackage gateway.

Corresponding author: Pamela H. Russell (pamela.russell@ucdenver.edu)

Author roles: Russell PH: Conceptualization, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Ghosh D**: Conceptualization, Funding Acquisition, Writing – Review & Editing

Introduction

Medical image analysis often lies at the boundary of research and the clinic, presenting challenges in both domains. Institutional and privacy concerns can compete with the objective of open data for research purposes. In particular, it remains standard practice to perform analysis with proprietary software or unpublished scripts. Additionally, the majority of imaging studies do not make image data publically available due to patient privacy requirements. These complex challenges can present barriers for scientists working in the image analysis domain.

In recent years, a small but growing number of open source computational tools have been developed to process and analyze medical images, promoting sharing of code; some of the most widely adopted are described in 1–3. To address the issue of availability of public image data, our group previously developed TCIApathfinder[4], an open source R package to simplify access to the thousands of publicly available images in The

Cancer Imaging Archive[5]. Here, we present radtools[6], an open source R package that lowers barriers to image analysis by simplifying the extraction of image properties and complex header information. Although several excellent image processing and analysis packages exist for the R environment[2,7–10], none currently offers special functionality for convenient presentation of image metadata; these tools generally present metadata in a form closely parallel to its original encoding. Radtools[6] specifically addresses the complexity of image metadata, improving upon metadata extraction methods in existing packages. The package implements a layer of processing to convert image metadata to familiar R data structures, eliminating the need for specialized knowledge and custom code to dig into metadata.

Radtools[6] supports the two most common medical image formats, DICOM (Digital Imaging and Communication in Medicine)[11] and NIfTI-1 (Neuroimaging Informatics Technology Initiative)[12]. The industry standard DICOM format combines a header and two-dimensional pixel data into one file, so that an image acquisition typically produces multiple DICOM files. (Some valid DICOM objects do not contain pixel data; these are still supported by radtools.) DICOM header fields consist of a "tag" that identifies the attribute, followed by the attribute value. There is no fixed size for a DICOM header; any number of thousands of possible attributes may be included. Each DICOM file for an acquisition contains its own header; many attributes will be constant across image slices. NIfTI-1 format was developed primarily for multidimensional imaging data as an improvement over the previous ANALYZE format[13]. NIfTI-1 combines header information and the entire multidimensional image acquisition into either a single file or two files (one header file and one image file). Unlike DICOM, NIfTI-1 specifies a particular set of required header attributes, and the header conforms to a fixed size with an option to add extended header information. Radtools[6] provides simple functions to explore and return image properties and header data from both image formats in familiar R data structures. For convenience, radtools[6] also provides wrappers around existing methods for extraction of pixel data and viewing of image slices.

Methods
Implementation

Radtools[6] is provided as a package (extension to the language) for the programming language R. The package is hosted on the Comprehensive R Archive Network (CRAN), and can be installed into the user's local R environment with the command 'install.packages("radtools")'. The package is loaded into an R session or script with the command 'library(radtools)'. Radtools consists of a collection of functions that can be called within R scripts or interactively from the R console. Package usage is documented in a vignette that can be viewed on the GitHub page (https://github.com/pamelarussell/radtools), the CRAN page (https://cran.r-project.org/package=radtools), or from the R console with the command 'browseVignettes("radtools")'. The package reference manual provides documentation of each individual function and is available on the CRAN page.

Radtools implements novel functionality for extraction and processing of image metadata. For implementations of the DICOM and NIfTI-1 standards themselves, radtools uses the existing state-of-the-art R packages oro.dicom and oro.nifti[2]. Radtools builds upon the metadata extraction methods available in those packages, calling their functions under the hood and providing a convenient layer of metadata exploration and processing. In deferring to the implementations in oro.dicom and oro.nifti, radtools is able to process the same file objects supported by those well-developed packages; for files not supported, radtools captures and reports any error messages raised within calls to their functions.

The correctness of our metadata implementations was tested with a diverse collection of 167 DICOM datasets and 23 NIfTI-1 datasets available publically; the tests can be examined and run in the "tests" directory of the package source. Each individual test dataset is documented in code comments.

Operation
The only system requirement is a working installation of R version ≥3.4.0. The radtools workflow consists of calling radtools functions from the R console or within R scripts.

Use cases
Radtools[6] can extract image properties and header data from any valid DICOM or NIfTI-1 file. Image datasets are loaded with the `read_dicom` and `read_nifti1` functions. Several generic functions extract attributes from either data type, including `img_dimensions`, `num_slices`, `header_fields`, which reports the set of header fields present, and `header_value`, which returns the value(s) of a particular attribute. Additionally, functions are provided to specifically address one format or the other. All header data present in a DICOM acquisition can be extracted into a matrix, where rows are attributes and columns are slices, with the `dicom_header_as_matrix` function. As most DICOM headers contain numerous attributes and many of these are constant across all slices, the `dicom_constant_header_values` function produces a named list of common attributes across slices. NIfTI-specific functions include `nifti1_num_dim`, which returns the number of dimensions, and `nifti1_header_values`, which returns a named list of all metadata attributes for the image.

The image itself can be extracted as a multidimensional matrix of intensities for either file format with `img_data_to_mat`. Image slices can be visualized with `view_slice`.

Finally, functions are provided to explore aspects of the DICOM standard itself. The functions `dicom_all_valid_header_tags`, `dicom_all_valid_header_names`, and `dicom_all_valid_header_keywords` return complete lists of valid DICOM header attributes. The functions `dicom_search_header_names` and `dicom_search_header_keywords` return attributes matching a search term.

For a demonstration of package usage including examples with publically available data, see the package vignette available at https://cran.r-project.org/web/packages/radtools/vignettes/radtools_usage.html.

In an additional vignette available at https://cran.r-project.org/web/packages/radtools/vignettes/oro_compare.html, we demonstrate the value of radtools compared to implementing metadata exploration with oro.dicom and oro.nifti. In some cases, similar results can be achieved by developing an understanding of the data representations in those packages and writing slightly more custom code. In other cases, radtools provides useful methods that are not available in those packages. Functions provided by radtools only include: (1) getting the names of metadata attributes present in a DICOM or NIfTI dataset, (2) getting all NIfTI metadata in a single data structure, (3) getting a data structure containing the overall properties of a DICOM acquisition (attributes that are constant across slices), (4) viewing a DICOM image, and (5) exploring and searching the complex DICOM standard itself.

Conclusions
Radtools[6] fills a specific need in the existing ecosystem of R packages for image processing and analysis: namely, the need for convenient extraction of image metadata. The package will accelerate workflow development and provide researchers with easy access to attributes that they may not have otherwise considered using. The inclusion of the package on CRAN, along with clear documentation, make it trivially simple for R users to obtain and begin using radtools.

Data availability
No data are associated with this article.

Software availability
Radtools can be installed with the R command "install.packages("radtools").

Radtools is available from CRAN: https://cran.r-project.org/package=radtools.

Source code available from: https://github.com/pamelarussell/radtools.

Archived source code at time of publication: https://doi.org/10.5281/zenodo.2593175[14].

License: MIT License.

References

1.   van Griethuysen JJM, Fedorov A, Parmar C, *et al.*: **Computational Radiomics System to Decode the Radiographic Phenotype.** *Cancer Res.* 2017; **77**(21): e104–7.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2.   Whitcher B, Schmid V, Thorton A: **Working with the DICOM and NIfTI Data Standards in R.** *J Stat Softw.* 2011; **44**(6): 1–29.
     Publisher Full Text

3.   Fedorov A, Beichel R, Kalpathy-Cramer J, *et al.*: **3D Slicer as an image computing platform for the Quantitative Imaging Network.** *Magn Reson Imaging.* 2012; **30**(9): 1323–41.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4.   Russell P, Fountain K, Wolverton D, *et al.*: **TCIApathfinder: An R Client for the Cancer Imaging Archive REST API.** *Cancer Res.* 2018; **78**(15): 4424–6.
     PubMed Abstract | **Publisher Full Text**

5.   Clark K, Vendt B, Smith K, *et al.*: **The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.** *J Digit Imaging.* 2013; **26**(6): 1045–57.
     PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6.   Russell P: **pamelarussell/radtools: 1.0.1 (Version v1.0.1).** *Zenodo.* 2018.
     Publisher Full Text

7.   **Get Images Out of DICOM Format Quickly.** [R package divest version 0.7.1]. [cited 2018 Nov 6].
     Reference Source

8.   Clayden JD, Maniega SM, Storkey AJ, *et al.*: **TractoR: Magnetic Resonance Imaging and Tractography with R.** *J Stat Softw.* 2011; **44**(8): 1–18.
     Publisher Full Text

9.   **Fast R and C++ Access to NIfTI Images [R package RNifti version 0.10.0].** [cited 2018 Nov 6].
     Reference Source

10. Bordier C, Dojat M, Micheaux P: **Temporal and Spatial Independent Component Analysis for fMRI Data Sets Embedded in the AnalyzeFMRI R Package.** *J Stat Softw.* 2011; **44**(9): 1–24.
     Publisher Full Text

11. **DICOM Standard.** [cited 2018 Nov 5].
     Reference Source

12. Jenkinson M: **NIfTI-1 Data Format — Neuroimaging Informatics Technology Initiative.** 2005; [cited 2018 Nov 5].
     Reference Source

13. **FormatAnalyze - MRC CBU Imaging Wiki.** [cited 2018 Nov 6].
     Reference Source

14. Russell P: **pamelarussell/radtools: 1.0.4 (Version v1.0.4).** *Zenodo.* 2019.
     http://www.doi.org/10.5281/zenodo.2593175

Open Peer Review

Current Referee Status: ✓ ✓

Version 3

Referee Report 14 May 2019

✓ **Volker Schmid** [iD]

Department of Statistics, Ludwig Maximilian University of Munich, Munich, Germany

My concerns have been addressed appropriately. The vignettes now nicely show the value of the Radtools package. Thanks for all the effort you put into the package and the paper.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical analysis of medical images.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 26 March 2019

✓ **Andrey Fedorov** [iD]

Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

The revised manuscript now includes a demonstration and specific summary of how the newly introduced package is different from the already existing functionality. Thank you for this revision and your persistence! I believe this improved the quality of the article sufficiently to make it suitable for indexing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: medical image computing, imaging informatics, applications of DICOM for implementing FAIR principles in medical image computing

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Referee Report 15 February 2019

https://doi.org/10.5256/f1000research.19528.r43610

Volker Schmid   iD

Department of Statistics, Ludwig Maximilian University of Munich, Munich, Germany

Unfortunately the authors have revised their manuscript only slightly. I still see a general interest in the aim and the idea of the provided package. However, the value of the package is not explained in the manuscript, and that may be because the ability of the package is (still) limited. The manuscript needs example code, Figures, Tables etc. to show the value of the package.

The manuscript needs to answer the following question: Why is it convenient to use your package compared to writing code specifically tailored for my data set?

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: statistical analysis of medical images

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Referee Report 12 February 2019

https://doi.org/10.5256/f1000research.19528.r43611

Andrey Fedorov   iD

Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Unfortunately, the manuscript is still lacking details about the added value it provides as compared to oro.* packages. The clarification that it provides "convenient extraction and exploration of image metadata", which, according to authors, "is not provided by existing tools", is not helpful. I do not understand what that statement actually means. Please include specific details how the functionality provided by Radtools is different from oro.*, and justify why that new functionality is important.

The authors refer to the "tests" directory in the package source, but the datasets referenced are from a local Dropbox folder. Considering TCIA provides API for retrieving images, it would make more sense to include code that retrieves all of the tests. Also, it is not clear why the specific datasets were selected, how they are different, and what aspects of the implementation they are testing.

About the updated title: "smooth" is equivalent to "convenient". Both are subjective qualifiers. I understand it was the authors intent to make extraction and navigation "more smooth" than available alternatives in their judgement, but whether this was successful or not will be up to the users of the package. I would drop the subjective qualifier, unless the manuscript includes specific objective criteria that would demonstrate it is "more smooth" or "more convenient" than the alternatives. Not sure how that would be demonstrated though - perhaps blinded user surveys?

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: medical image computing, imaging informatics, applications of DICOM for implementing FAIR principles in medical image computing

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

- -

Version 1

Referee Report 07 January 2019

Volker Schmid iD

Department of Statistics, Ludwig Maximilian University of Munich, Munich, Germany

This manuscript describes the R package radtools. The aim of the R package is "smooth navigation of medical image data".

The idea to provide functions which appear "smooth" for the end user is of great importance. However, the functions provides in the package do not seem to be of much (additional) benefit to the end user. Functions for reading NIfTI and DICOM images are wrappers around function in the packages oro.nifti and oro.dicom; visualisations of medical images are realised in oro.nifti. Only the functions for exploring DICOM headers are genuinely original. This is of course an important part of working with (DICOM) images.

From my understanding, F1000Research requires software tools articles to contain examples of the use of the tools. This manuscript does not contain any examples. An example (or two examples) would not only strengthen the manuscript, but could/would also show the benefit of the R package itself.

Is the rationale for developing the new software tool clearly explained?
No

Is the rationale for developing the new software tool clearly explained?
No

Is the rationale for developing the new software tool clearly explained?
No

Is the description of the software tool technically sound?
Partly

Is the description of the software tool technically sound?
Partly

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
No

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
No

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Competing Interests: I am the co-author of the R packages oro.dicom and oro.nifti, which have been used and cited by the authors (however, I am currently not actively developing either package)

Reviewer Expertise: statistical analysis of medical images

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Referee Report 07 January 2019

Andrey Fedorov (iD)

Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

The authors present a new R package developed to support the use of DICOM and NIfTI files from the R environment. The authors rightfully discuss the popularity of R and the need to support image processing tasks in this environment. The argument for development of the proposed package, radtools, is that "[...] no existing package makes image metadata conveniently accessible. Extracting image metadata, combining across slices, and converting to useful formats can be prohibitively cumbersome, especially for DICOM". The resulting package is available from CRAN, and this reviewer confirmed its installation and basic functions.

The major issues that need to be addressed to make the article sound are the following:
1. Justification of the development of a new package for working with DICOM, or with NIfTI, is not sufficient.
2. No details are provided about how the functionality was tested, and about the capabilities and limitations of the package in terms of supporting specific DICOM objects.
3. Related to 2), no details are provided about how the DICOM files are handled "under the hood", i.e., whether all IO functionality was implemented from scratch, or the package is using some other DICOM libraries.

Through the text, the authors reference other R packages for similar tasks, and most notably oro.dicom and oro.nifti [1]. Those packages have been around for quite a long time, are broadly used, based on citations of the corresponding articles, and arguably provide the functionality of the proposed new package (loading data in the aforementioned formats, examination of the attributes, visualization of the images), plus more (e.g., writing of the NIfTI data).

DICOM is a complex standard, with a lot of ways information can be stored. For example, there are different methods to encode the content (transfer syntax), different character sets that can be used, private attributes. Therefore, often the quality of a DICOM implementation is defined to a large degree by the data that was used to test the implementation. The quality is also usually improved over time with the usage of the implementation. The proposed package is not accompanied by any details about what types of DICOM objects are supported, what was tested and how. Given it is a new package with a short development and usage history, one has to make a very strong argument for introducing such new tools in presence of existing alternatives.

Other suggestions:
- The discussion of the DICOM objects is an oversimplification, which is reflected in the implementation of the functionality. The standard defines various types of objects that can be serialized as files, but those objects are not limited to images. As an example, DICOM defines Structured Reporting object, which will not have PixelData. The proposed package fails to read such object. The authors can find sample SR objects in the familiar to them TCIA (e.g., see QIN-HEADNECK collection).
- "smooth" is a subjective qualifier that is redundant in the title and the text.

I will be happy to reconsider this article after the authors address the above concerns. But my current opinion is that oro.dicom and oro.nifti set a rather high bar for any new implementation of similar functionality.

References

1. Whitcher B, Schmid V, Thornton A: Working with the DICOM and NIfTI Data Standards inR. *Journal of Statistical Software*. 2011; **44** (6). Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
No

Is the rationale for developing the new software tool clearly explained?
No

Is the rationale for developing the new software tool clearly explained?
No

Is the description of the software tool technically sound?
No

Is the description of the software tool technically sound?
No

Is the description of the software tool technically sound?
No

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: medical image computing, imaging informatics, applications of DICOM for implementing FAIR principles in medical image computing

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.