**Editors' comments:**

The authors have improved the manuscript and the figures. There is still a need to edit the language of the manuscript and to improve clarity. The reviewers pointed out specific sentences and paragraphs and made useful suggestion that we encourage the authors to follow.

Thank you, we believe that we have addressed the recommendations presented by the reviewers.

**Reviewer 1 Comments to Author: Open RT Structures: A Scalable Solution for TG-263 Accessibility**

General comments:

Thank you for the opportunity to read this paper. The submission is interesting and can potentially add significant value to the radiation oncology community.

1. Please comment on the importance of this question and the originality of the findings for the readers of Red Journal.

The paper questions how standardized naming conventions can be efficiently and consistently obtained and used in radiation treatment planning. While the extent of this problem and the repercussions of mislabeled structures is not well described, the problem is significant enough to warrant an AAPM Task Group on the problem.

Thank you for the comment. We have elaborated on the importance of properly labeled structures as noted below.

*“The proper labeling of these structures is important in the evaluation of generated treatment plans, when comparing with plans from other individuals/institutions, and when curating large data sets from multiple institutions. In large groups, incorrectly labeled data would be extremely difficult to identify, and so proper and consistent labeling is exceedingly important.”*

The originality of the work stems from a few things. First, the authors provide open-source code to create these structures. It needs to be noted that not every clinic will have the skills necessary to deploy this in their clinic. Second, the structures abide by standardized structure nomenclature and provide DICOM RT Structure Set data. Finally, although less clear to the reader, the datasets generated are theoretically vendor agnostic as they would, at the very least, adopt DICOM (SS) standards.

Thank you for the comment. To the note that every clinic might lack the skills to deploy this, we have released a series of YouTube videos explaining every process: from installation, to changing templates in Eclipse, or running it as a standalone server.

2. Please comment on the appropriateness of the study approach and experimental design.

The submission is not a study, and thus details for experimental design and study approach are unnecessary; however, there are still areas where there is a mix-up between 'results' and 'methods' as this reviewer would prefer. As it stands, the 'story' jumps between methods and results. This needs to be a bit clearer in the text. This reviewer encourages the authors to clearly define what was done in the methods and the findings of that work in the results (how to use the code).

Thank you for the comment, we agree that the storyline does jump around. We have rewritten and clarified the Methods section to better explain the overall purpose and flow of the program.

4. Please comment on the analysis and interpretations of the data. Do you agree with the proposed conclusions?  
  
Again, this revision does a nice job addressing concerns raised by reviewers, particularly by including a high-level figure. However, a number of changes to the document should be made to make it more readable and highlight the shortcomings in addition to the issues raised above.

5. Please comment on weaknesses or limitations of the study. (Examples are: selection biases, sample size limitations, missing data.)  
  
The range of usefulness challenges is now addressed, and it now may be left to readers to judge the quality of the tool created.  
  
6. Please comment on the writing and organization of the paper. Is the paper overly wordy? Is the English language acceptable?  
  
The paper remains well-written with few (if any) grammatical issues, but some of the text sometimes reads a bit loose and informal.  
Thank you for the comment. We have gone throughout the manuscript to try and remove some of the more informal language.

The sentence including *headache* has been rephrased to be

*“This plugin was coded to alleviate tedious manual work in exporting Varian templates to the program.”*

We have removed the word ‘dummy patient’ in favor a more detailed explanation.

*“If the user instead wishes to create an anonymous CT and load RT Structure files, to later save as templates, they can select the ‘Create folder with loadable RTs’. This will create a folder with a previously anonymized four-slice CT and generate the available structure templates as described above.”*

*“While adoption of the standard nomenclature has popular support, in a recent survey2 conducted by the TG-263U1 Task Group, the majority of respondents had not yet adopted the standardized nomenclature. From this survey, the largest difficulties were lack of time/resources to create new templates, and/or difficulty with retraining staff if templates are not available.”*  
  
  
7. Please comment on the necessity and clarity of the figures and tables. Can they stand independently of the text?  
  
Again Figure 2 is not helpful with low-resolution text.

We have recreated Figure 2 with higher resolution text.

8. Please comment on any need for a formal statistical review.  
Na

Specific comments  
  
Title: Again, 'scalability' is loosely defined and, in this reviewer's opinion, inaccurate.  
The authors state : "Thank you. We hope this addition to the introduction will help to clarify: "Our aim in this work was to lower the barrier to adoption of TG-263 nomenclature in English, Spanish, or French by disseminating standardization that may facilitate data sharing. We have developed a tool which runs on any Windows system to easily create TG-263-compliant structure template libraries. Our tool can monitor folders and automatically add patient-specific structure sets, or create loadable RT structure/.xml templates and is a scalable solution focused on compatibility with all Treatment Planning Systems (TPS) utilizing the DICOM standard." If this is not agreeable, we will gladly remove "scalable solution" from the title of the manuscript. "  
  
When one thinks of 'scalability', the first thought is that the software is able to accommodate things like workload or performance handling or resource utilization, handle various volumes of data, accommodating new features or functions without dramatic changes to the major components of the software, or allow the system to expand easily to accommodate other functions. Only one or maybe two aspect of 'scalability' in this work is presented: utilizing and processing different structure templates based on patient treatment protocols and accommodating different sizes of datasets. We do not know, for example, if the system can be easily expanded for other functions or accommodate different datatypes. I do not think scalability is the correct term and could misrepresent the work.

We have removed ‘scalability’ from the title, it now reads “Open RT Structures: A Solution for TG-263 Accessibility”, we have also removed the word ‘scalable’ from the paper.  
  
Keywords: na  
  
Abstract:  
L25: need to clarify better what was being done in relation to brachytherapy (templates for HDR or LDR? various sites?) Details are needed or should be removed in the abstract. What working group(s)?

Thank you for the comment, we rephrased the section in the abstract and added the following clarification

*“Brachytherapy templates were created with the guidance of the TG-263 brachytherapy members who share membership on the AAPM Brachytherapy Subcommittee and Working Group on Brachytherapy Clinical Applications  
L33: does the tool 'create' structures and rename them? Clarity needed.”*

Thank you for the comment. To clarify this in the abstract we have changed the phrasing to be

*“We have developed a tool which runs on any Windows system to easily create* ***and name*** *DICOM-RT structures sets that are TG-263-compliant for all planning systems utilizing the DICOM standard*”

L37: (I don't see a need to include the last sentence since you're already providing a public forum for the software)

Thank you for the comment, we have reduced the last sentence to better flow and remove unnecessary words.

*“The program and source code are publicly available via GitHub, encouraging feedback from community users for improvement and guide further development.”*  
  
Introduction:   
L10: is it more accurate to say DICOM-RT standards specify structure formats? DICOM is the global phrase used but generally refers to images.

Thank you, to avoid confusion we have added the specifier ‘RT’ to DICOM-RT.  
L33: strictly speaking, you are not disseminating standardization, you are providing a means for ensuring structures have consistent nomenclature as per TG-263.

Thank you, we have changed this statement to more accurately reflect what the program is capable of.

*Our aim in this work is to lower the barrier to adoption of TG-263 nomenclature in English, Spanish, or French by providing a means of standardization to facilitate data sharing and consistency*

L37: please consider writing out what xml format is (eXtensible Markup Language) before using its acronym.

Thank you, we have added this to the manuscript

*“update Variant eXtensible Markup Language (.xml) templates.”*  
  
Methods and Results:  
While I commend the authors for presenting the work, the current layout of the Methods and Results remains awkward, inherently because of the nature of the material presented. I recommend having a detailed Methods section that describes exactly what was done and why (e.g., Figure 1) and having a separate Results section to describe the working interface and provide a real-world example (Figure 2). Information on how fast it takes can be included in the results. The authors discussed using brachytherapy templates: it would be valuable to the reader to see an example of this in the results.

Thank you for the comment, we have split up Methods and made extensive changes.  
  
Stylistically, consider rephrasing sentences like "We wanted to make..." into something more descriptive, such as "The software was designed to ensure the model was compatible with different manufacturers and versions of treatment planning software."

Thank you for the comment, we have adjusted the manuscript accordingly.  
  
P9, L54: 'headache' is not a good word. Please clearly identify what the concern/challenge is.

Thank you for the feedback. We have rephrased this sentence to be more formal and express the challenge associated with this action.

*“This plugin was coded to alleviate tedious manual work in exporting Varian templates to the program.”*  
P10, L54: consider replacing 'dummy' with something more descriptive. You are creating a patient dataset to permit users to create multiple structure templates that can be exported for treatment planning templates.  
Thank you for the comment, we have changed the sentence to better express why you might want to import this anonymous CT.

*“If the user instead wishes to create an anonymous CT and load RT Structure files, to later save as templates, they can select the ‘Create folder with loadable RTs’. This will create a folder with a previously anonymized four-slice CT and generate the available structure templates as described above.”*  
  
Discussion:  
P11: L40: rephrase to state what the program does more concisely: "... the program never..." should be replaced with something along the lines of "... the software is coded to ....? check what? etc. )  
  
Thank you for the comment, we agree this is confusing. We have rephrased this section to better clarify the risk of continual recreation of DICOM RT structure files.

*“The largest risk that we could foresee is that the program continually updates its own previously generated RT Structure files. To ensure this does not happen, the program internally tracks which images have been previously viewed (via Series Instance UID), and creates each RT Structure file with that same Series Instance UID. The program is coded to check if each template RT Structure file already exists, and so prevents a continuous recreation of the same set. As an additional safety measure, the software is coded to only create new RT structure files and will not open or edit an already existing RT Structure file, and so presents no risk to existing work flows present by the user.”*

References: na  
  
Figures:  
Figure 1: great!  
Figure 2: I still believe the text is too small here.\

Thank you, we have expanded the size of the figure to better allow the reader to view the text.  
Tables: na

**Reviewer 2 Comments to Author: 1.** Please comment on the importance of this question and originality of the findings for the readers of Red Journal.

The software application is original. Approaches to simplify adoption of standards into clinical practice is very important for creating large scale data sets. The premise of the present manuscript is that implementation of the application they developed requires less effort than manual approaches to implement TG-263 standardizations, and subsequent updates to it, into clinical practice. Information supporting that point is lacking.

Thank you for the comment.  
  
2. Please comment on the appropriateness of the study approach and experimental design. (Examples: retrospective or prospective cohort, case-control, cross-sectional, ecological, case series; clinical trial or secondary analysis of clinical trial; registry-based; critical review; metaanalysis or systematic review; experimental, based on cell cultures, animal models, physical models, or method/technique development.)  
This is a description of a software application.  
  
  
3. Please comment on the appropriateness and reproducibility of the data collection and experimental techniques. (If applicable, does the study comply with the CONSORT, PRISMA and/or REMARK statements? If applicable, was the study IRB-approved or registered on [clinicaltrials.gov](http://clinicaltrials.gov/)?  
No collected data was evident.  
  
  
4. Please comment on the analysis and interpretations of the data. Do you agree with the proposed conclusions?  
No the information presented does not support the assertion that the application will drastically reduce effort. That may be true, but they have omitted data that would help them to support that point. They have also not clarified how this remains directly linked to TG-263 as updates proceed over time, so that it is maintained as the standard.    See response at item #6

Thank you for the comment. We apologize for not stressing this enough, and have added the following to better express that this project is endorsed by members of TG-263, and will be maintained by them as updates progress.

***“maintained by members of the TG-263U1 Task Group which ensures continuous access to up-to-date templates”***

“*An online spreadsheet, Airtable6, was utilized to house all templates in an evergreen fashion, which will be continually updated by members of TG-263 as recommendations continue to develop over time.”*  
  
5. Please comment on weaknesses or limitations of the study. (Examples are: selection biases, sample size limitations, missing data.)  
Lack of detail on implementation, see response at item #6  
  
6. Please comment on the writing and organization of the paper. Is the paper overly wordy? Is the English language acceptable?  
  
The paper continues to have issues with clarity of writing. Reviewer #1 had specifically identified several which remain uncorrected in the second revision.  The article needs to be proof read, avoiding ambiguous or colloquial language and correcting typos.

Thank you for the comment. We agree that the language was too colloquial and have worked hard to correct the manuscript in this regard.  
"An online spreadsheet, Airtable6, was utilized to house all templates in an evergreen fashion." What is an "evergreen", why is it relevant?

Thank you for the comment. We have clarified that ‘evergreen’ in this form means continually maintained by members of the 263 Task Group as ROI recommendations evolve over time.

*“An online spreadsheet, Airtable6, was utilized to house all templates in an evergreen fashion, which will be continually updated by members of TG-263 as recommendations continue to develop over time.”*

"Target color selection was based on preferences at UCSF (Dr. Sue Yom), MDACC (Dr. Anna Lee), and Michigan (Dr. Charles Mayo), reflecting the Clifford Chao's IMRT book." Reformt the sentence, making less colloquial and shifting citations to individuals into the references. e.g. private communication, or adding an acknowledgement section where the individuals are cited for their contribution of information.

Thank you for the comment, we have addressed this into the acknowledgements section and references, removing the names.

*“Target color selection was based on discussions with physicists and physicians at UCSF, MDACC, and Michigan, and reflect the available literature5. Due to the informal nature of color selection and difficulty achieving consensus, colors remain fully customizable within the program.”*  
  
"The program was written (BMA) using C# , ensuring it's computability with windows systems." What is BMA?

Thank you for the comment. (BMA) refers to the initials of the programmer, these have been removed.

"computability" do you mean "compatibility", even so why is the phrase needed?  Further along in the paragraph the author is suddenly discussing Python and DICOM. The paragraph seems to be attempting to condense in a much longer technical discussion into a single paragraph, at the cost of clarity.

Thank you for the comment, we have broken this section up into what we feel are more appropriate sections of the manuscript.

*“The subsequent sections expand on each step of the program’s workflow. A series of videos explaining each step of the program, including installation, is available on YouTube. The link to these videos is found at the bottom of the program GitHub page: anon for review and are continually updated to demonstrate current features and performance of the program.”*

*Creation of DICOM-RT files*

*“DICOM-RT files are created via the publicly available FellowOakDicom package8, and a C# wrapper for the ITK coding package, SimpleITK9. The framework for creating RT Structure files in Python has been previously reported10, and a similar process is used here.”*  
  
The online structure template has FMAID codes, it does not have SNOMED codes. This was a specific request of reviewer #2, which the responses said had been met.

Thank you for the comment. We apologize for the confusion, this request has been met. We believe the reviewer is referring to the column ‘FMAID’ being present. This column will continue to remain for those users who wish to continue using FMAID. However, this is another column called ‘SchemeCode’ and ‘Scheme’, specifying both ‘SCT’ and the associated code number. We apologize for any confusion this has caused.  
  
The source for structure names and mapping to coding systems (SNOMED, FMAID)  needs to be the TG-263 standards group AAPM SC-263. This includes updates, e.g. TG-263-U1. It is not clear how the application pulls directly from the standards created by that group, or how updates are incorporated. Does this not imply that the Airtable effectively becomes the standard source of information?

Thank you for these comments. Everything within the structure name and mapping is AAPM SC-263 and 263-U1. We ensured everything programmed within the table is directly from AAPM SC-263.  
  
Why aren't  the templates more simply provided as a downloadable PDF or excel document? Additionally, why are they not in a format that can be imported directly into the treatment planning systems used. If anyone can make changes to the spread sheet at any time, then how does this work as a reference? How is direct linkage to TG-263 and updates maintained, so that it is not promoted as an alternate standard?

Thank you for the comment. We agree that it would be very useful to have a format which is directly importable into TPS used. Unfortunately, what is required for this varies based on TPS: Eclipse can import an anonymous CT and save the structure templates, but only the latest version of Raystation enables importation of ‘empty’ structures. Similar with Pinnacle. This is why we created the ability to make RT-Structures on incoming patients in a ‘server’ form.

We have added the following statement as clarification that only the members of TG-263 are able to make edits to the reference AirTable.

*“While the Airtable can be downloaded by anyone, only specific members of the working group can make changes to the spreadsheet.”*  
  
The manuscript attempts to provide a technical discussion of the application developed and how to implement it in a clinic. The current writing is difficult to follow. Hand-offs and dependencies between the technologies them are not clear.

We agree that this is not clear. We have worked hard to clarify this in a manner that we hope makes things more clear.

The key piece of evidence for ability to implement it would be discussion of the range of clinics which had implemented the software. This is only given passing mention.  This was requested by Reviewer #2. The present treatment still has not provided detail on number of clinics implementing the application, challenges faced in implementing, gains and reduction of burden in creating and  
maintaining TPS structure templates in using it, etc.  The most that the reader gets is "This software was tested at multiple sites and ensured to be compatible with Pinnacle v16.2.1, Raystation v12.1, and Eclipse v15.6, although output should be compatible with all TPS utilizing the DICOM standard. "

We piloted this informally at five institutions with different TPS to ensure compatibility with various TPS. We have also added the following text:

*“Early rollout required hands on training with video conferences to demonstrate how to use the software. Therefore, we had to create videos for other physicists to watch to learn how to use the software and hope this paper serves as a narrative review of its capabilities.”*  
  
Note also that TPSs utilize the DICOM standard for input of images, that does not mean that functionality extends to import of templates for structures.

Thank you for the comment. We agree, this was the leading factor in our decision to include a patient-specific option, for TPS which do not allow blank templates, and unfortunately makes it difficult to create a ‘format which is imported directly into the TPS’ for all TPS, as requested above.  
  
The opening sentence in the Conclusion, highlights the challenge for this paper with writing that mixes evidence with aspiration.  
"We have created open-source software that may drastically reduce the burden of creating and maintaining TPS structure templates and facilitates the adoption of TG-263 standardized nomenclature."

Thank you for the comment, please see the closing comments below.

Ease of installation and use are key to ability for the application to reduce effort. No data is presented to assess that point or the assertion of "drastically". In the sentence "may" further highlights what we don't know.  We do not know how many people used the software in this testing, who they were (physicist, dosimetrist, physician, …) , or what specific skills were needed to set up the application or how many and what types of templates they successfully implemented using it. Setting up Dicom servers and applications reaching out from behind institutional firewalls to internet based applications, such as Airtable, is not uniformly easy dependent upon the policies of institutional IT groups.  What was the experience in multiple institutions with implementation? Airtable is not a standard software application and is not under the control of the authors. How does their application function if Airtable is not available? Because the application is dependent on at least  
four different technologies used in concert (Airtable, their C# application, their Python scripts, DICOM server) understanding challenges and risks in implementing is relevant.

Thank you for the comment, and we apologize for the confusion. We have added further clarification within the manuscript of the following: 1) Python scripts are not used, it is simply that the author of this program has written a similar one in Python, which the DICOM manipulation is based on. 2) The DICOM server is locally run, not a cloud server. 3) Airtable is like a Google Excel sheet, the user has no obligation to use it if they do not desire. 4) C# is one of the base programming systems for Windows, this simply indicates that the program is designed to run on Windows computers.

We have clarified to say physicists or physicians piloted the importing of the generated structure sets within their TPS. In our experience utilizing this software at institutions there were not issues. But if issues do arise, the entirety of the AirTable can be downloaded to an excel sheet from a non-institutional computer. We can add a video expressly showing how to manually do this, if desired.   
The authors are to be commended for the application that they have developed. It has promise, but the present manuscript leaves important details unclear or unanswered. It is my hope that the authors will revise their manuscript. I strongly encourage them to have the manuscript proof read, by at least two IJORBP readers (e.g. physician and physicist) who are not part of their work.   

We do believe this automated structure set generator reduces burden and facilitates adoption of TG-263 standardized nomenclature.

This program can quickly generate all 53 templates in Table S1 in under a minute, which is much faster than any human could build these ASTRO-endorsed site-specific templates.

We hope that rewriting the paper in a more narrative tone helps to answer some of these very important discussion topics.  
  
7. Please comment on the necessity and clarity of the figures and tables. Can they stand independently of the text?  
Yes figures are acceptable  
  
  
8. Please comment on any need for formal statistical review.  
Statistical review is not needed