**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.

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.

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).

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.  
  
  
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.  
  
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.  
  
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)?  
L33: does the tool 'create' structures and rename them? Clarity needed.  
L37: (I don't see a need to include the last sentence since you're already providing a public forum for the software)  
  
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.  
L33: strictly speaking, you are not disseminating standardization, you are providing a means for ensuring structures have consistent nomenclature as per TG-263.  
L37: please consider writing out what xml format is (eXtensible Markup Language) before using its acronym.  
  
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.  
  
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."  
  
P9, L54: 'headache' is not a good word. Please clearly identify what the concern/challenge is.  
  
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.  
  
  
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. )  
  
  
  
References: na  
  
Figures:  
Figure 1: great!  
Figure 2: I still believe the text is too small here.\  
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.  
  
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  
  
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.  
  
"An online spreadsheet, Airtable6, was utilized to house all templates in an evergreen fashion." What is an "evergreen", why is it relevant?  
"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." Reform 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.  
  
"The program was written (BMA) using C# , ensuring it's computability with windows systems." What is BMA? "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.  
  
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.  
  
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?  
  
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?  
  
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.  
  
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. "   
  
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.  
  
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."  
  
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.   
  
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.     
  
  
  
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