Transforming Publishing Programme: Checklist for Analytical Teams Preparing to Transform Publications and/or develop Reproducible Analytical Pipeline (RAP) Processes

By Anna Price, John Mowbray and Evelyn Shiel

<table>
<thead>
<tr>
<th>Document Control</th>
<th>Version 0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date Issued</td>
<td>June 2019</td>
</tr>
<tr>
<td>Author</td>
<td>Anna Price</td>
</tr>
<tr>
<td>Comments to</td>
<td><a href="mailto:nss.isdtransformingpublishing@nhs.net">nss.isdtransformingpublishing@nhs.net</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version Control</th>
<th>Version</th>
<th>Changes made</th>
<th>Approved By</th>
<th>New Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>First draft</td>
<td>Anna Price</td>
<td>12/12/2018</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>Second draft</td>
<td>Evelyn Shiel, Esther Morris, Anna Price</td>
<td>22/01/2019</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
<td>Added links to Government Digital Service guidance on open code</td>
<td>Anna Price</td>
<td>06/06/2019</td>
</tr>
</tbody>
</table>
Introduction

The Transforming Publications Programme (TPP) offers two distinct outcomes for improved publications (see diagram below):

1. Creation of a Reproducible Analytical Pipeline (RAP) – this focuses on the back end production process of the publication.
2. The creation of a Transformed Publication (TP) – which may or may not include RAP but delivers publications to the web in formats like the Acute Activity and Psychiatric Inpatients.

This document explains what a Reproducible Analytical Pipeline (RAP) is and its benefits. It explains what a Transformed Publication (TP) entails and the benefits of doing it. It also describes the working practices a team should follow when completing this work. Finally, it outlines what to consider before engaging in either of these processes and what to expect when you do.

Reproducible Analytical Pipeline (RAP)

- When creating a publication, the traditional process takes the data from the data source, manipulates it (often using multiple software programmes) and saves it in a data file format, which will feed an output such as a pdf or word document. RAP automates that process using one software programme. As a result the publisher only needs to make a minimal number of changes (depending on the RAP level chosen). The output of a RAP could be the traditional publication outputs (pdf, word document and/or excel table) or it could be a web-based output.

- Benefits of RAP:
  - Saves both analyst and processing time.
  - Builds a process that is fully transparent, auditable and verifiable.
  - Reduces risks of making errors and improves the quality of the publication.

Transformed Publication (TP)

- Currently most publications are published via pdfs and excel tables. To transform a publication means making the data available on the web, with the use of RShiny, D3, open data etc. A transformed publication is not dependent on RAP and can be completed using your current publication process up until the point when the data files are ready. If data files are currently produced using SPSS they can still be made available on the website via RShiny, D3 etc.

- Benefits of a transformed publication:
  - Allows analysts to easily display many different aspects of the data (data explorer).
Engages users to understand what they need from the publication and tests it with them.
For users who would like to do their own analysis, open data gives them this opportunity.
Saves analytical time.

Each process (RAP or TP) could take up to 6 months. This will depend on the skills within the team, what level of RAP the team decide to do and how much time the team have to commit to this work.

It is important that the whole team are involved in the process even if they aren’t necessarily making the changes to the publication. Below is a list of questions to discuss in your team before contacting TPP:

- When is your publication released?
- What is the available resource in your team for development between now and release?
- What are the R/R Shiny/version control skills within your team?
- Is your team enthusiastic about redeveloping your publication?

Transforming Publishing Programme Working Practices

The Transforming Publishing Programme wants the best for your publication. We want it to be as quick and easy as possible to run, to be used by as broad an audience as possible and to be as useful as possible to that audience. The Reproducible Analytical Pipeline (RAP) and Transformed Publication (TP) are developmental processes. We are here to support you through these processes of improvement but at the end of the day this is your publication. It is really important that your team not only takes ownership of the publication but also the process of improving it. We have identified three key working practices for how best to transform your publication:

Resilience – It’s really important to build resilience within your team so that if there are changes within your team both the publication and improving it can still be completed. It also allows team members to learn from each other more easily. It’s important to ensure continuous communication between members and documentation of processes as we go.

Adaptability – We build in development time for our team to learn new practices/best practice so that we can support you and your publication in the best way possible. We want to incorporate any new learning into anything you have developed/are developing. This may require some change but we hope this will be in the best interest of your publication and the organisation.

Commitment – When all team members are committed to making changes to the publication it makes the work easier and more enjoyable for the team. Everyone on the TPP team believes in the product and it makes a huge difference to the productivity and atmosphere of the team.

Things to be aware of:

- Redevlopment of a publication adds value and may include desired changes that were not possible previously.
- There is a significant time commitment when transforming a publication that may be in addition to the time you would usually take to produce a publication.
Reproducible Analytical Pipelines

Before Starting

All team members:
- Read the RAP paper and discuss the key questions in “Reproducible Analysis” as a team.
- Agree the level of RAP that your team will be aiming for. We recommend that for work to be reproducible, the minimum standard of a RAP project should be level 4 (a or b or both). This decision will also be guided by the TPP team.
- Gather metrics on the current publication process:
  - The overall time from the start of the publication process to release.
  - The time required from each team member (by key process).

Team members carrying out the RAP development (minimum of 2):
- Review the RAP companion.
- Attend the 1-day Jumping Rivers R course arranged by staff development.
- Attend the Transforming Publishing team’s Introduction to Git and GitHub tutorial.
- Review the PHI R Style Guide in order to follow it throughout the development process.

Team members producing the publication in the future:
- Take the 4-hour Introduction to R course on DataCamp.
- Attend the 1-day Jumping Rivers R course arranged by staff development.

During Development Process

Team members carrying out the RAP development:
- Gather metrics on the time taken for development.
- Open up the code on GitHub as soon as possible (see the Government Digital Service security guidance).
- Ensure PHI R Style Guide is followed.

Maintenance/Troubleshoot Stage

All team members:
- If there are issues running your publication then please email the TPP mailbox.

Team members producing the publication:
- Gather metrics on the new publication process:
  - The overall time from the start of the publication process to release.
  - The time required from each team member (by key process).

Team members carrying out the RAP development:
- Ensure that the code has been published on GitHub if appropriate (see Government Digital Service open code guidance).
Transforming Publication Process

Before Starting

All team members:

- Familiarise yourself with the current transformed publications (Acute Activity and Psychiatric Inpatients) and the staff guide to the new platform.
- Read user guidance documents: Initial User Engagement and Usability Testing.
- List the customers that you are aware of (e.g. known users of your publication, customers provided with IRs/FOIs/PQs, etc).
- Review open data requirements.
- Identify roles for the development process (e.g. data wrangler(s), data visualiser(s), etc).
- Gather metrics on the current publication process:
  - The overall time from the start of the publication process to release.
  - The time required from each team member (by key process).
  - User metrics e.g. number of downloads over time, number of views over time, survey results etc.

Team members carrying out the TP development (minimum of 2):

- Attend the 1-day Jumping Rivers R course arranged by staff development.
- Attend the Transforming Publishing team’s Introduction to Git and GitHub tutorial.
- Take the 4-hour R Shiny course on DataCamp.
- Review the PHI R Style Guide in order to follow it throughout the development process.

Team members producing the publication in the future:

- Take the 4-hour Introduction to R course on DataCamp.
- Attend the 1-day Jumping Rivers R course arranged by staff development.

During the Transformation Process

All team members:

- Gather metrics on the time taken for development.
- Follow the Protocol for Release Practices.
- Open up the code on GitHub as soon as possible (see the Government Digital Service security guidance).
- Ensure PHI R Style Guide is followed.

Maintenance/Troubleshoot Stage

All team members:

- Continue to follow the Protocol for Release Practices.
- Continue to engage with your users and maintain the relationships developed at the start of the TP process.
- Update your publication to the platform standards as required.
- If there are issues running your publication then please email the TPP mailbox.
Team members producing the publication:

- Gather metrics on the new publication process:
  - The overall time from the start of the publication process to release.
  - The time required from each team member (by key process).

Team members carrying out the TP development:

- Ensure that the code has been published on GitHub if appropriate (see Government Digital Service open code guidance).