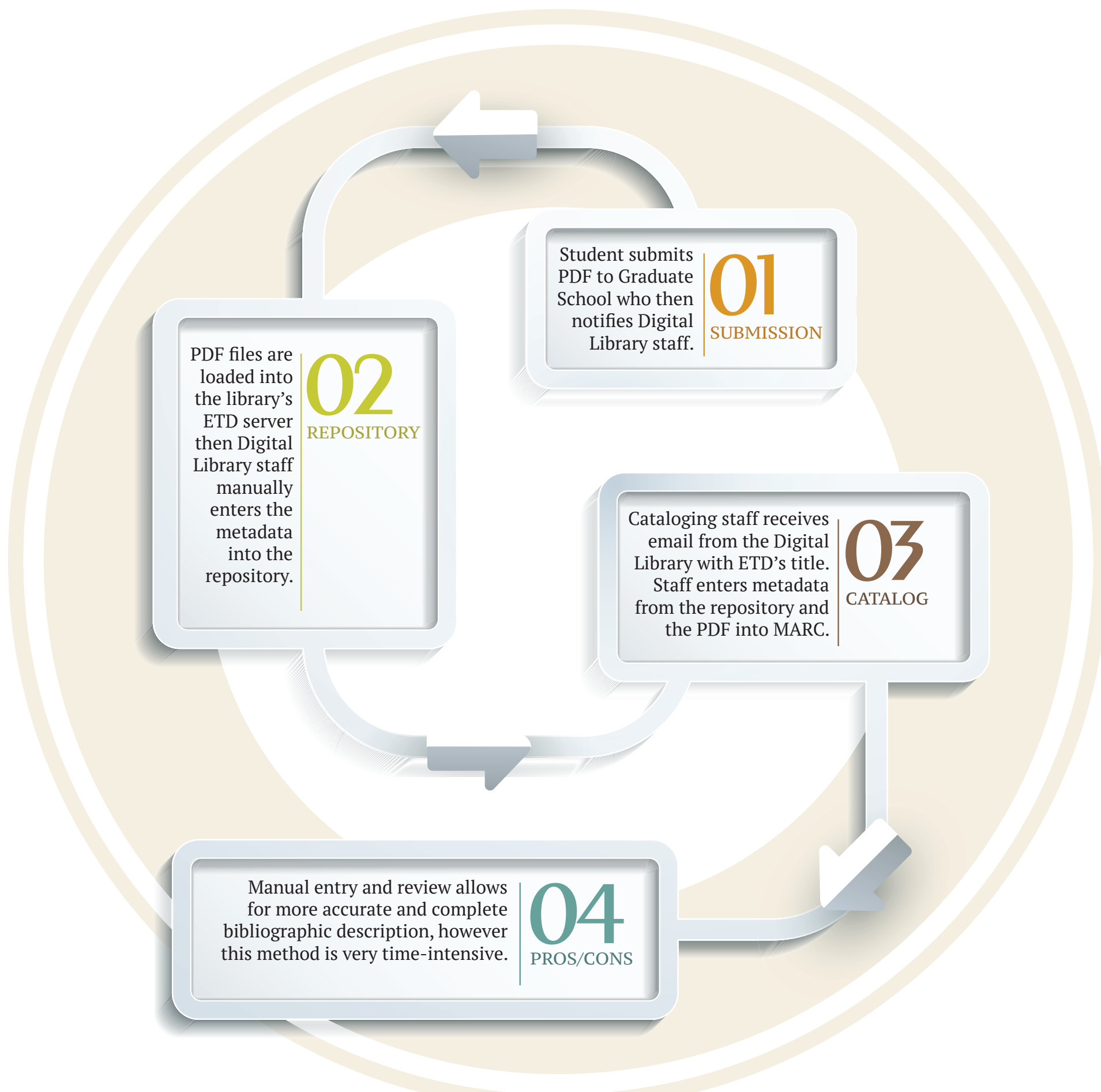


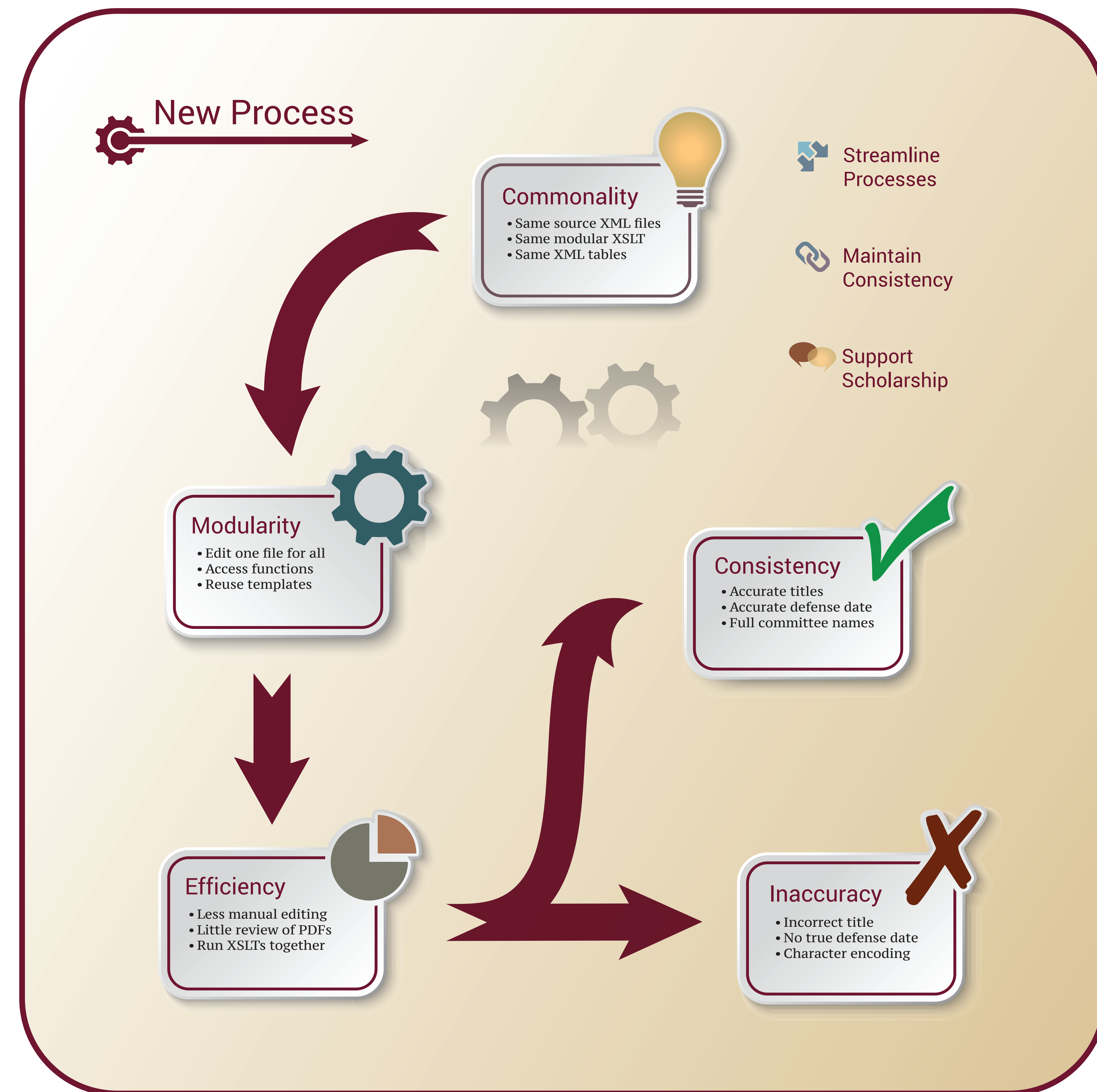
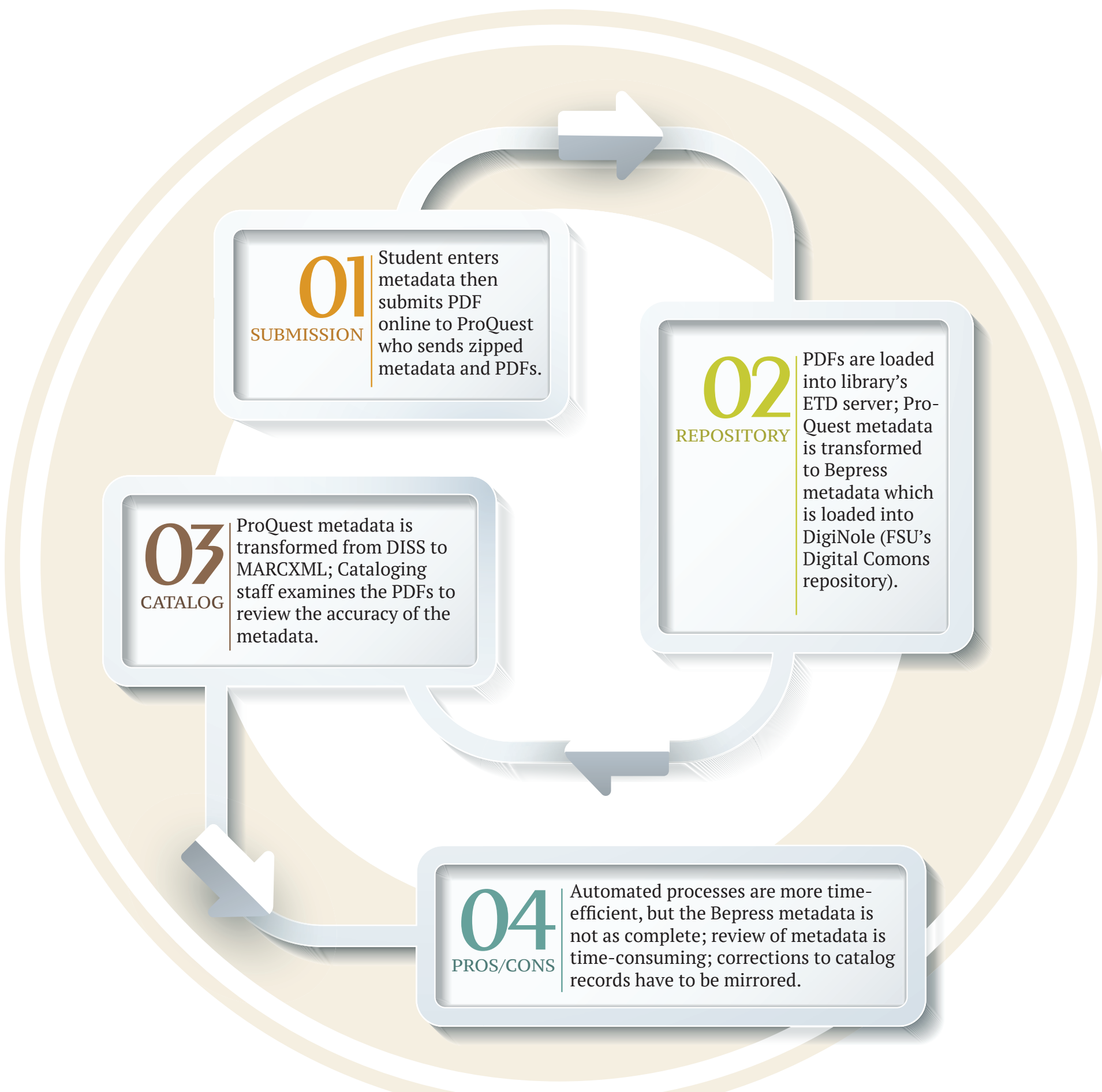
Efficiencies for Quality Control of Repurposed Metadata



First Process



Second Process



Procedure

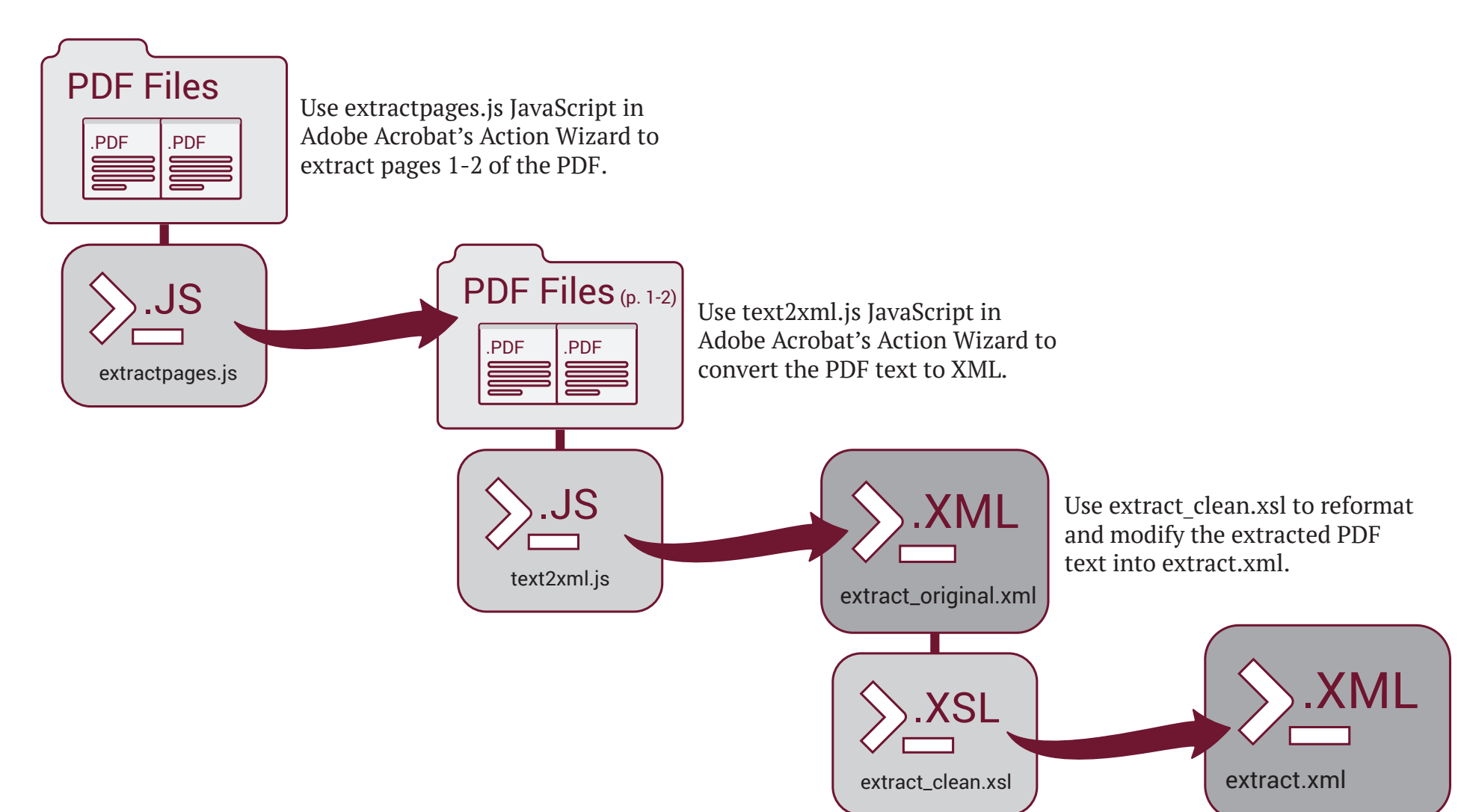
1. Unzip PDF metadata folders received from ProQuest.
2. Move PDFs and metadata into separate folders.
3. Run combine_files.xsl to combine all metadata files into one file.
4. Run Adobe Action Wizard on folder of PDFs to extract pages 1-2.
5. Run Adobe Action Wizard PDF to XML on the extracted pages.
6. Run reformat_adobe.xsl then cleanup_reformat_adobe.xsl on XML.
7. Manually review data.
8. Run trans2bepress.xsl and trans2marc.xsl.
9. Upload Bepress metadata to Digital Commons.
10. Use MarcEdit to convert MARCXML metadata to MARC.
11. Import to OCLC Connexion.
12. Validate the records then correct any character encoding errors.
13. Update the records then export to a file
14. Load that file into the catalog.

Code Snippets

```
JavaScript for Adobe Acrobat Action Wizard Extract Pages
var filename = this.documentFileName.split(".").
shift()+"_bibinfo.pdf";
this.extractPages
({
  nStart: 0,
  nEnd: 1,
  cPath : filename
});

Matching substring of analyze string in caps.xsl
<xsl:analyze-string select="$Element" regex="\w+"
flags="i">
  <xsl:matching-substring>
    <xsl:value-of select="
      if (.= $article and not(position()=1))
      then (lower-case()) else
      if ($list/mixed/pair[term1 = current()]
      then ($list/mixed/pair[term1 =
        current()]/*[(self::term2)]) else."/>
  </xsl:matching-substring>
  ...
```

Extracting PDF Text



Modular Processes

