3. Please provide detailed requirement specifications for METS, ALTO, and JP2 (preservation and access images).

JPEG 2000 encoding options:

Settings for the preservation image (Lossless compression)
Creversible: Yes
Bit Rate:-,4,2,1,0.7,0.5,0.35,0.25,0.18,0.125,0.088,0.0625,0.04419,0.03125,0.0221,0.015625
Tile Size: 1024*1024
Levels:6
Block Size:64*64
Corder: RPCL
Cmodes: BYPASS
Cuse_sop: yes
Cuse_eph: yes

Kakadu command line options:

```
kdu_compress -rate -,4,2,1,0.7,0.5,0.35,0.25,0.18,0.125,0.088,0.0625,0.04419,0.03125,0.0221,0.015625
Cmodes=BYPASS Cuse_sop=yes Cuse_eph=yes -quiet -double_buffering 8 -num_threads 4 Clevels=6
Cprecincts={256,256},{256,256},{128,128} Corder=RPCL ORGtparts=R Cblk={64,64} ORGgen_plt=yes
Creversible=yes Stiles={1024,1024}
```

Settings for the access image (Lossy compression)
Bit Rate: 1,0.7,0.5,0.35,0.25,0.18,0.125,0.088,0.0625,0.04419,0.03125,0.0221,0.015625
Tile Size: 1024*1024
Levels:6
Block Size:64*64
Corder: RPCL
Cmodes: BYPASS
Cuse_sop: yes
Cuse_eph: yes

Kakadu command line options:

```
kdu_compress -rate 1,0.7,0.5,0.35,0.25,0.18,0.125,0.088,0.0625,0.04419,0.03125,0.0221,0.015625
Cmodes=BYPASS Cuse_sop=yes Cuse_eph=yes -double_buffering 8 -quiet -num_threads 4 Clevels=6
Cprecincts={256,256},{256,256},{128,128} Corder=RPCL ORGtparts=R Cblk={64,64} ORGgen_plt=yes
Stiles={1024,1024}
```

ALTO file specification:

- There MUST be one ALTO-file for each scanned page.
- The ALTO-file MUST contain structural information obtained from the OCR-recognition process conducted on that page
- The ALTO file MUST validate using the ALTO 2.0 specification found at «http://www.loc.gov/standards/alto/v2/alto-2-0.xsd»
- The ALTO file MUST at least contain the following tags:
  - <Description>: Giving information on the processing software used, and the settings for the software.
  - <Styles>: Identifying the Text styles that the OCR-software recognises in the text.
  - <Layout>: Describing the OCR-recognized text in the page with the interpeted layout for the text. Each single recognized word MUST be decribed inside a <String> tag and MUST contain the absolute position and height/width of the word. In addition the <String> tag MUST contain a «CC» attribute estimating the probability of recognition of each letter in the word, and a «WC» tag indicating the estimated probability of the correctness of the whole word.

METS file specification:

- There MUST be one METS file for each scanned newspaper.
- The METS file MUST validate using the METS-specification found at «http://www.loc.gov/standards/mets/mets.xsd»
- The METS file MUST at least contain the following tags:
  - One <fileSec> element containing:
    - One subelement <fileGrp USE="IMAGEpage"> listing all preservation JPEG-2000 files in order from first page to last page. Each file MUST be described using a <file> element, and the element MUST have the following attributes present:
      - ID: MUST be the name of the file without any path.
      - MIMETYPE="image/jp2"
      - SIZE: The file size in bytes
      - CHECKSUMTYPE="MD5"
      - CHECKSUM: The Calculated MD5 checksum for the file.
      - A <Flocat> element: LOCTYPE="URL" xlink:type="simple" xlink:href= «filename_without_path»
    - One subelement <fileGrp USE="ALTOpage"> listing all ALTO OCR files in order from first page to last page. Each file MUST be described using a <file> element, and the element MUST have the following attributes present:
      - ID: MUST be the name of the file without any path.
      - MIMETYPE="image/jp2"
      - SIZE: The file size in bytes
      - CHECKSUMTYPE="MD5"
      - CHECKSUM: The Calculated MD5 checksum for the file.
      - A <Flocat> element: LOCTYPE="URL" xlink:type="simple" xlink:href= «filename_without_path»
  - One <structMap TYPE="physical"> element describing the physical structure of the newspaper. The structure MUST be decribed using ONE <div> element with attributes TYPE="issue" and DMDID set to the name of the folder given to the newspaper. The <div> tag for the entire newspaper MUST contain a list of <div> tags, one for each page, ordered from first to last page. All page <div>s MUST have an ORDER attribute giving the page number in the newspaper and have a TYPE= «page» attribute. All page <div>s MUST have two <fptr> elements refering to the relevant JPEG 2000 and ALTO XML files in the <fileSec> element.
- The METS file MAY contain the following tags:
  - One <DmdSec> element that wraps a <mods> element («http://www.loc.gov/standards/mods/v3/mods-3-3.xsd») which describes content metadata such as:
    - <genre>
    - <language>
    - <originInfo>
    - <relatedItem>
  - Several <DmdSec> elements that describes identified articles in the newspaper using a <mods> element («http://www.loc.gov/standards/mods/v3/mods-3-3.xsd»).
  - One <AmdSec> element that wraps several <techMD> elements, one per file generated from the original pages. Each <techMD> MUST wrap a <premis> element («http://www.loc.gov/standards/premis/v1/PREMIS-v1-1.xsd») describing the file.
  - One <structMap TYPE="logical"> element describing the logical structure of the newspaper.