Recommended Practices

for

STEP File Compression

Version 1.3

12 April 2023

Contacts	
Jochen Boy	Phil Rosché
PROSTEP AG	ACCR, LLC.
Dolivostraße 11	125 King Charles Circle
64293 Darmstadt / Germany	Summerville, SC 29485, USA
jochen.boy@prostep.com	phil.rosche@accr-llc.com
	Frederic Darré
	CIMPA
	frederic.darre.external@airbus.com

Table of Contents

List of Figures

Document History

Revision	Date	Change
1.0	2013-02-12	Initial release
1.1	2013-11-25	Change recommended file extension back to ".stpZ" (one dot)
1.2	2016-08-15	Added clarification for compression algorithm and file structure; Update file extensions for Domain Model XML
1.3	2023-04-06	Updates to compression algorithm; Addition of MIME Types; Addition of *.stpA file extension

1   Introduction

With the growing scope covered by STEP due to the latest developments, the average size of STEP files is increasing. The new capabilities included in AP242 "Managed Model Based 3D Engineering", which will be the core data format in the aerospace and automotive, will accelerate this process even further. Hence, there is a need to decrease the size and optimize the performance of file exchange.

This requirement is generic and will be applicable for data exchange based on all STEP application protocols; foremost AP214, AP203 Ed.2, AP209 Ed.2, and all editions of AP242. The main types of information in focus are 3D exact geometry, 3D tessellated geometry, 3D PMI, as well as product and assembly structure with associated process information (configuration, kinematics, etc.)

This document provides a basic approach to compress STEP files, using the same approach as defined in STEP Part 21 Edition 3 (ISO 10303-21:2016), but with some differences concerning the file management within the created ZIP archives.

This Recommended Practices document extends the use of the compression method defined for Part 21 files (ASCII format) to be applied to Domain Model files (XML format) as well. It shall not be used with the Part 26 files (binary format), as it would provide little to no benefit in this case.
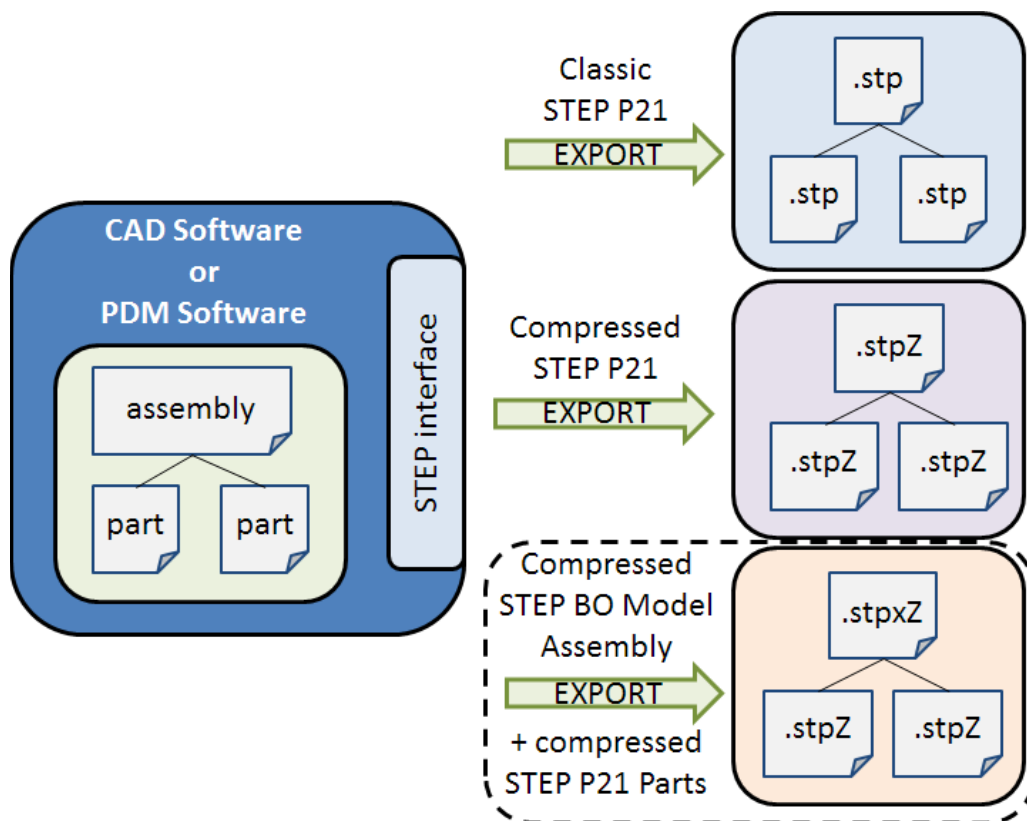

Figure 1: Illustration of STEP interface options

2   Scope

The following are within scope of this document:

- Definition of a compression algorithm to be used for STEP files

- Definition the associated file extensions and MIME types

- Handling specification for import / export interfaces

- Extension for External References

- Example analysis of STEP file compression benefits

The following are outside the scope of this part:

- Description of STEP information related to a specific AP

- Compression of binary (e.g., STEP Part 26) files.

3   Fundamental Concepts

Particular concepts of interest within the scope of this recommendation are described in the following sections.

3.1   Compression Algorithm

According to ISO 10303-21:2016 Annex A.4, all compressed STEP files shall be written using the ZIP compression algorithm, according to the ZIP file format specification version 6.2.0[1] provided by PKWARE (see Annex A [1]). This compression format is well-established and widely used. It is, among others, compatible with:

- pkzip/pkunzip, version 2.04g or later

- Windows Compressed Folders

- winzip

- info-zip

- zlib

- Mac OS X (10.3 and higher)

This algorithm is referenced in STEP Part 21 Edition 3 section 3.10.13. It is also the basis for ISO 29500, the Open Office XML formats, as well as ifcZIP (see Annex A [2]).

The same compression mechanism shall be used for the compression of Part 21 as well as Domain Model XML files.

Many CAx-IF members however use GZip / Zlib for compression. PKZip libraries can read GZip files; the differences are only in the file header. Both shall be supported on import. PKZip allows for additional features such as folder/file structures (see section 3.4 below).

Restrictions:

- Compressed files shall not be encrypted.

- No Unicode support for file names.

---

[1] ISO10303-21:2016 mentions PKZIP v2.04g. This is the earliest versions of the PKZIP software supporting the file format specification 6.2.0, released in 1993. Earlier versions of this document referenced PKZIP v2.05, but there is no difference to the file format.

3.2 File Extensions

In order to distinguish compressed STEP files per this recommendation, the following files extensions shall be used.

For compression of single files:

- *.**stpZ** ("STEP zipped") for single compressed Part 21 (ASCII) files
- ***.stpxZ** ("STEP XML zipped") for single compressed Domain Model XML files

For compression of multiple files in one ZIP package:

- ***.stpA** ("STEP archive") for sets of files, where the main (root) file is a STEP file. These archives can contain a mix of Part 21 files, Domain Model XML files, and other file types.

3.3 MIME Types

A media type, also known as a MIME type, is a two-part identifier for file formats and format contents transmitted on the Internet. They can help determining how a certain file should be handled during data transfer and which applications are suitable for processing it.

The Internet Assigned Numbers Authority (IANA) is the official authority for the standardization and publication of these classifications. The list of registered MIME types is available online at:

https://www.iana.org/assignments/media-types/media-types.xhtml

The following MIME types for STEP files have been added in 2022:

model/step

> *Applications which use this media:* Applications that display, create, edit, import, or export ISO 10303 product data models, which are encoded in the exchange syntax defined in ISO 10303-21, use this media type. While the ISO 10303-21 syntax is also used in application/p21, model/step provides assurance that the model represented is a STEP product data model.

model/step+xml

> *Applications which use this media:* Applications that display, create, edit, import, or export ISO 10303 product data models, which are encoded as XML, use this media type

model/step+zip

> *Applications which use this media:* Applications that display, create, edit, import, or export ISO 10303 STEP models, which are encoded in the exchange syntax defined in ISO 10303-21 and compressed as ZIP, use this media type. While the ISO 10303-21 syntax is also used in application/p21+zip, model/step+zip provides assurance that the model represented is a STEP model.

model/step-xml+zip

> *Applications which use this media:* Applications that display, create, edit, import, or export ISO 10303 product data models, which are encoded as XML and compressed as ZIP, use this media type.

3.4 STEP – ZIP File Structure

There are two approaches for compressed STEP files, "*.stpZ" for compressing single files and "*.stpA" for creating archives with multiple files and folders.

3.4.1 *.stpZ / *.stpxZ:

The primary use case for these file types is long-term archiving.

- There shall be only a single STEP file in each compressed file.

- *.stpZ denotes a compressed Part 21 file.

- *.stpxZ denotes a compressed Domain Model XML file.

- The name of the STEP file, and the name of the STEP compressed file shall be the same, i.e., *sample_123.stp* will be compressed in *sample_123.stpZ*.

Note: When using the basic GZip / Zlib algorithm to manually inflate a *.stpZ or *.stpxZ file, the file extension of the compressed file will get lost. This is a known restriction which is deemed uncritical, since whoever would be working manually with files this way would know which extension to restore.

Note: Part 21 Edition 3 Annex A.4 defines the name of the root file in a ZIP package to be "ISO-10303.p21". This document overrides this recommendation for the following reasons:

- The file name is often automatically generated by the exporting system and contains valuable information in itself, such as part name, part number, part version, etc. This information would get lost when renaming the file to "ISO 10303-21" and would also be problematic in case several such files are handled simultaneously.

- While the extension *.p21 conveys the file is encoded according to ISO 10303-21, using the extension *.stp instead adds the fact that it represents a STEP data model (also see section 3.3 above)

- This recommendation applies the ZIP algorithm also to files other than STEP Part 21 files, namely Domain Model XML files.

3.4.2 *.stpA

The primary use case for these file types is data exchange.

- There can be multiple files in the compressed file. The files can be of different types.

- The compressed file may contain a folder structure.

- The root folder of the archive must contain an INI file indicating the root file(s) for reading the contents, as well as optional additional information.
    - File name: **stpa.ini**
    - See https://en.wikipedia.org/wiki/INI_file for the file format definition

Examples for stpa.ini

1. Usual STEP CAD model with one single root STEP file

```
; root file of the STEP archive
[root files]
Names = "root_assy.stp"
;
```

2. STEP PDM use case with several root files, some of them being in a subdirectory in the archive

```
; root files of the STEP archive
[root files]
Names = "root_assy.stpx" "root_assy2.stpx" "directory/root_assy3.stpx"
;
```

3. Example for future use with a JSON file included:

```
; root file of the STEP archive
[root files]
Names = "root_assy.stp"
; scenegraph file of the STEP archive
[scenegraph]
File = "dirscenegraph/assy_scenegraph.json"
;
```

3.5   Compressed File Handling

The file extensions defined in section 3.2 (and the MIME types listed in section 3.3) can be associated with any STEP consuming software that is capable of automatically decompressing the file upon import.

For STEP processors which are not capable of handling this STEP compressed format, there is the possibility to manually unpack the files using any available unzip tool, and then using the uncompressed STEP files directly with any STEP compatible software (see section 4.3).

4 Compression Process

4.1 Export

On Export, up to three options shall be available to the user, depending on the content to be exported and support by the exporting interface:

- Uncompressed export of one or several files

- Export of a single file as *.stpZ / *.stpxZ

- Export of a larger model consisting of multiple files in one package as *.stpA

Note that it is also possible to export an assembly as a set of *.stpZ / *.stpxZ files.

These options can be preset in the software configuration or might be manually selected by the user.

4.2 Import

Just as for classic import, the root file needs to be selected for import. The STEP interface of the target system needs to decompress each *.stpZ / *.stpxZ file or *.stpA package to a temporary location, and then import and process the contained files as usual, starting with the root file as identified per section 3.4.

The distinction between compressed and uncompressed STEP files will have to be made by the importing processor, based on file extension, MIME type, and file type (ASCII, XML, binary).

4.3 Manual Handling / Compatibility

Compression of exported files, and decompression of files to be imported can also be done manually (or as part of a batch process) in case the involved STEP interface cannot handle the compressed files as per this recommendation directly.

For a single STEP file, compress it with a ZIP tool compatible with the agreed compression algorithm. Make sure the file is included immediately in the ZIP package and is not located in a subfolder. The rename the ZIP package to *name_of_the_part21_file.stpZ* or *name_of_the_domain-model_xml_file.stpxZ* respectively.

For a model consisting of multiple files, use a simple text editor to create a *stpa.ini* file in the root folder of the structure to be packaged, indicating the root file(s) of the package as defined in section 3.4.2. Then pack all constituent files and subfolders into a compressed archive using the agreed algorithm. The name of the resulting package should indicate its contents, i.e., if possible, rename it to *name_of_the_root_file.stpA*.

For manual import, unpack the compressed file with a compatible archiving tool, then handle the extracted files as usual, starting with the root file as identified per section 3.4.

This method can be used as a backup solution, but a direct integration of the compress / uncompress method to the STEP interface is the preferred approach.

5 Extension for External References

This paragraph describes how to use compressed STEP files in the context of External References, which are described in the "Recommended Practices for External References" for Part 21 files, and the "Recommended Practices for STEP AP242 Edition 3 Domain Model XML Product & Assembly Structure" for AP242 XML files (see Annex A). In these cases, it is required to clearly identify when a reference links to a plain STEP file, or a compressed STEP file.

5.1 File Name Reference

It is important to note that, when creating the external references, the name of the referenced external file shall always be the name of the uncompressed file, i.e., the external reference will always point to *sample_123.stp* or *assembly_123_top_node.stpx*, regardless of whether it is compressed or not.

The reason is that the compression procedure shall have no impact on the file contents at all. If the STEP files are uncompressed manually first (as described in 4.3), the external references still must work.

5.2 Reminder – Document Format Properties

For Part 21 files, the document format properties are defined in section 2.3 of the Recommended Practices for External References as follows:

The Document Format is an additional property which completes the structure displayed in Figure 2 below and is attached to the `document_file` entity. It provides information about the data format of this referenced file.

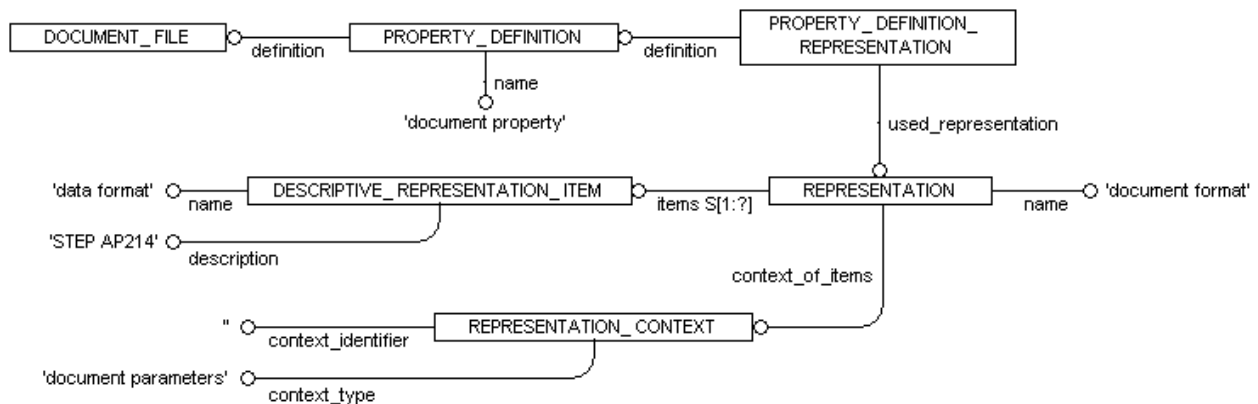The format of the referenced file is stored in the `descriptive_representation_-item.description`:

Figure 2: Instantiation diagram for Document format Properties

For Domain Model XML files, the document format properties are defined in section 10.1 of the Recommended Practices for STEP AP242 XML Product & Assembly Structure as follows:

The `FormatProperty` entity is the specification of characteristics of a File or of a DocumentDefinition that specify the format of the object.

DataFormat: the convention that was used to structure the information in the characterized object. Use ClassString type if one of the values below is used, otherwise use "Class" template.
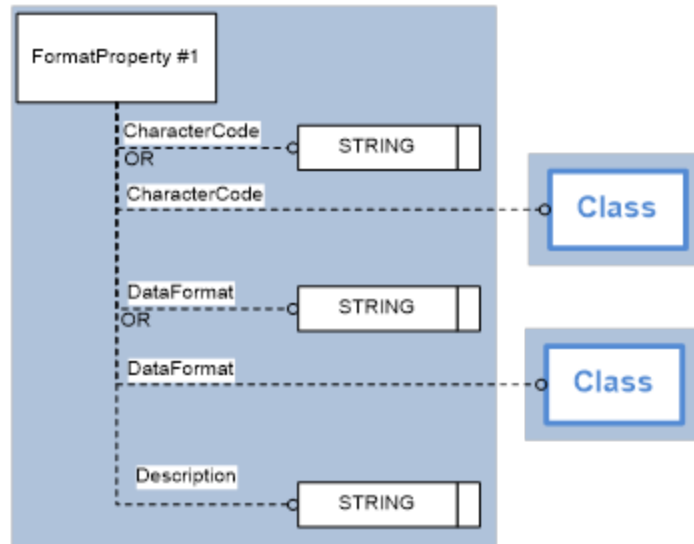
Figure 3: Domain Model XML Template "FormatProperty"

Based on the convention stated in section 5.1 that the external reference shall always point to the name of the uncompressed file, the same applies to the document format properties, i.e., no distinction will be made between compressed and uncompressed STEP files. The reason is that manual handling of the files on either end of the process would lead to inconsistencies.

The importing STEP processor can still unambiguously detect the file format using the file extension or MIME type, together with the fact the file is either an ASCII, XML, or binary file.

For STEP Part 21 files as well as for Domain Model XML files, the same list of recommended values defining the file format applies. For STEP files, these are:

Format property for plain file	Description of referenced file
`'ISO 10303-203',` `'STEP AP203'`	Eventually followed by the release number (E2, E3): The document contains data in ISO 10303-203 Part21 format
`'ISO 10303-214',` `'STEP AP214'`	Eventually followed by the release number (E2,E3): The document contains data in ISO 10303-214 Part21 format
`'STEP AP214 CC06'`	The document contains data in ISO 10303-214 Part21 format according to Conformance Class 06 (product structure only, the file contains no geometry, but refer-ences to external geometry files)
`'ISO 10303-242',` `'STEP AP242'`	The document contains data in ISO 10303-242 Part 21 format
`'ISO 10303-242` `Domain Model XML'`	The document contains data in ISO 10303-242 XML format

Figure 4: Table of recommended Document Format Properties

6   STEP Compressed File Example

This section presents a STEP compressed file using AP242. The AP242 file in Part 21 format is compressed inside the STEP compressed file:
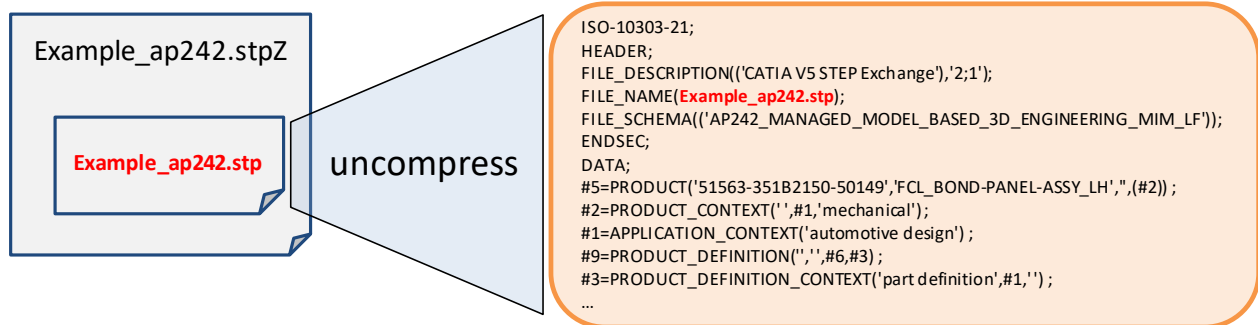

Figure 5: STEP compressed file

The following Part 21 excerpt is taken from the Recommended Practices for External References, and highlights the references to the file illustrated above:

```
#18 = SHAPE_REPRESENTATION('',(#17),#8);
#26 = PRODUCT('10001','L-BRACKET','NOT SPECIFIED',(#25));
#28 = PRODUCT_DEFINITION_FORMATION('','/ANY',#26);
#29 = PRODUCT_DEFINITION('design','',#28,#24);
#30 = PRODUCT_DEFINITION_SHAPE('','SHAPE FOR L-BRACKET',#29);
#31 = SHAPE_DEFINITION_REPRESENTATION(#30,#18);
[...]
#48 = DOCUMENT_TYPE('');
#49 = DOCUMENT_FILE('l-bracket_prt.stp','','',#48,'',$);
#51 = PROPERTY_DEFINITION('external definition','',#49);
#52 = PROPERTY_DEFINITION_REPRESENTATION(#51,#18);
#53 = EXTERNAL_SOURCE(IDENTIFIER(''));
#54 = IDENTIFICATION_ROLE('external document id and location',$);
#55 = APPLIED_EXTERNAL_IDENTIFICATION_ASSIGNMENT
      ('Example_ap214.stp',#54,#53,(#49));
[...]
#428 = DOCUMENT_REPRESENTATION_TYPE('digital',#49);
#433 = PRODUCT_RELATED_PRODUCT_CATEGORY('part',$,(#26,#72,#117,#159,
       #213,#245,#300,#374,#402));
#455 = APPLIED_DOCUMENT_REFERENCE(#49,'',(#29));
#456 = DOCUMENT('','',$,#458);
#458 = DOCUMENT_TYPE('configuration controlled document version');
#459 = DOCUMENT_PRODUCT_EQUIVALENCE('equivalence',$,#456,#440);
#460 = ROLE_ASSOCIATION(#461,#455);
#461 = OBJECT_ROLE('mandatory',$);
[...]
#856 = PROPERTY_DEFINITION('document property','',#49);
#857 = PROPERTY_DEFINITION_REPRESENTATION(#1856,#1855);
#855 = REPRESENTATION('document format',(#1853),#1854);
#853 = DESCRIPTIVE_REPRESENTATION_ITEM('data format','STEP AP214');
#854 = REPRESENTATION_CONTEXT('','document parameters');
```

7   STEP File Compression Benefits

Based on experience and depending on the file contents, a compressed STEP file is between 15% and 40% the size of the uncompressed original STEP file.

The following table represents test cases from CATIA V5 R20 converted to STEP AP242 with Geometric Validation Properties.

Test Case	Native File Size	Plain STEP File Size	Compressed STEP File Size
Low complexity	1.73 MB	2.2 MB	311 kB
Medium complexity	37.9 MB	57.1 MB	8.1 MB
High complexity	60.2 MB	172.3 MB	26.7 MB
Assembly (12 elements)	82 MB	188 MB	37.6 MB

Annex A   References

[1] ZIP Application Note v6.2.0 (File format specification):
www.pkware.com/documents/APPNOTE/APPNOTE_6.2.0.txt

[2] ifcZIP Implementation Agreement:
https://standards.buildingsmart.org/documents/Implementation/IFC_Implementation_Agreements/CV-2x3-154.html

[3] CAx-IF Recommended Practices for External References, v2.1:
https://www.mbx-if.de/documents/rec_prac_ext_ref_v21.pdf

[4] MBx-IF Recommended Practices for STEP AP242 Edition 3 Domain Model XML Product & Assembly Structure:
https://www.mbx-if.de/documents/rec_prac_ap242xml_assy_struct_v3.1.pdf