

SXF をより使いやすく … 「SXF の ZIP 圧縮」の提案

2015/8/20

一般社団法人オープン CAD フォーマット評議会 (OCF)

OCF の技術部会では、SXF の ZIP 圧縮について検討しています。

以下は、OCF 最新動向ブログ(<http://blog.goo.ne.jp/ocf-blog>)の「SXF テクニカル通信」にて、計 8 回シリーズで発信した情報です。

第1回	2015年02月27日	SXF の ZIP 圧縮(1)～ はじめに
第2回	2015年03月22日	SXF の ZIP 圧縮(2)～ ZIP 圧縮するメリットは？
第3回	2015年04月21日	SXF の ZIP 圧縮(3)～ ZIP 圧縮後の拡張子は？
第4回	2015年05月09日	SXF の ZIP 圧縮(4)～ ZIP じゃなくて LZH もあるじゃないか！？
第5回	2015年05月24日	SXF の ZIP 圧縮(5)～ ZIP 圧縮・解凍の簡単なプログラム実装方法は？
第6回	2015年07月13日	SXF の ZIP 圧縮(6)～ ZIP 圧縮・解凍プログラム実装後の課題 1 と対策
第7回	2015年07月26日	SXF の ZIP 圧縮(7)～ ZIP 圧縮・解凍プログラム実装後の課題 2 と対策
第8回	2015年08月19日	SXF の ZIP 圧縮(8)～ まとめ

第 1 回

SXF の ZIP 圧縮(1)～ はじめに

OCF の活動に以下のようなものがあります。

オープン CAD フォーマット協議会設立趣旨(抜粋) <http://www.ocf.or.jp/top/outline.shtml>

- ・ CAD のデータフォーマットに関する情報収集、分析等調査研究
- ・ CALS で利用される、CAD データの交換用ファイル形式の策定支援と普及活動
- ・ CAD データの互換性検証、実証実験、並びに標準化規格の策定及び普及

そこで、現在行っている活動の一部を、ブログという媒体を使って、会員の皆さまだけでなく、広く世の中に発信することとしました。

まずは、現在、技術部会で検討を行っている「SXF の ZIP 圧縮」を、以降 7 回で発信します。

皆様ご存知のように SXF には sfc と p21 があり、いずれもテキスト形式であるため、ファイルサイズが大きくなりがちだったり、画像や属性を付加することでファイルが分割されたりします。

SXF によるデータ交換では、ファイルサイズとファイル分割は弊害となります。

こんな弊害を解決する手法として「SXF の ZIP 圧縮」するのはどうでしょうか？ という提案になります。

SXF の規格は、レベル 2 Ver3.1 が最新です。

その運用が始まってから結構な年数を経て、枯れた(安定した)技術となっています。このあたりで新たな規格を定義し、SXF を盛り上げ、CAD のデータ交換ファイルとして SXF がより広く使われるようになって欲しい、というわけなのです。

さて、いかがでしたでしょうか。

次回は「(2) ZIP 圧縮するメリットは？」をお送りする予定です。

第 2 回

SXF の ZIP 圧縮(2)～ ZIP 圧縮するメリットは？

前回、現状の SXF データ交換の弊害 2 点(ファイルサイズ大、ファイル分割)について軽く触れました。その弊害がどんなもので、どのように解決するのかを、メリットという観点で以下に記します。

1. ファイルサイズが小さくなる

SXF ファイルはテキストファイルなので、圧縮によるファイルサイズの縮小が著しく表れます。約 10% のサイズとなるので、1/10 です。5MB のファイルが 500KB になるわけで、CD への書き込み、メール送信など大いにやりやすくなることでしょう。

SXF に画像が配置されていると、画像ファイルの圧縮率がテキストファイルに対して低くなるため、図面全体での圧縮率は低下しますけれども。

2. 複数ファイルがまとまって 1 ファイルになる

画像や属性を付加すると、SXF は複数ファイルになってしまいます。属性であれば同じファイル名なので探しやすいですが、画像は自由な名前でも複数付加できるので、どれがどこで使われているかを把握するのが困難です。

SXF はデータ交換が目的なわけで、その際ファイルが足りないためにトラブルになったということが少なくない数あったとすれば、その回避策になります。

なお、SXF には朱書きという仕組みがあり別ファイルで管理しますが、これは図面本体とは別に扱う特性が強いので、圧縮の対象外とする案が有力です。

さて、いかがでしたでしょうか。

次回は「(3) ZIP 圧縮後の拡張子は？」をお送りする予定です。

第 3 回

SXF の ZIP 圧縮 (3) ~ ZIP 圧縮後の拡張子は？

ZIP 圧縮後の拡張子の名前をどうするかというお話です。

既存の sfc/p21 をそのまま使ってしまったら、現行 SXF と互換性が維持できなくなるので論外ですし、sfz/z21 みたいに 3 文字拡張子のままとするのは、ロングファイル名が使えるようになって久しいこの時代にイケてない感じがします。

MS Office のように「x」を付けるのがクールであるところ、これは Office ドキュメントが Open XML 化したときの「x」であり、saf 以外で xml を使っていない SXF で採用するのには無理があります。

既存の拡張子に zip の「z」を付けた、sfcz/p21z としようと考えています。

※拡張子の大文字／小文字はどちらがいいのか、どちらでもいいのかという議論もありますが、この記事では小文字で統一して表現します。

ともあれ、拡張子が変わることで現行 SXF との扱いをユーザーおよびアプリが判別できることが重要なことなのです。

さて、いかがでしたでしょうか。

今回は「(4) ZIP じゃなくて LZH もあるじゃないか!？」をお送りする予定です。

第 4 回

SXF の ZIP 圧縮 (4) ~ ZIP じゃなくて LZH もあるじゃないか! ?

圧縮形式として日本で長年親しまれてきた LZH を採用するのもありでしょうか。

答えはノーです。近頃、LZH ファイルを見かけることはないですよ。

UNLHA32.DLL やその他派生アプリの開発は停止していて、LZH の使用中止を呼びかけているという情報が出回ったのが 2010 年 6 月でした。情報の出元は、LZH 開発者の吉崎氏、Micco 氏です。

なぜ使用中止を呼びかけているかという点、

- ・ 多くの圧縮ファイルにおいて、アンチウィルスのチェックを逃れる脆弱性がみつかると
- ・ アンチウィルスや関連団体は、ZIP/CAB/7z 等に対して脆弱性を認め対策を行った
- ・ 一方、LZH については非受理となり、アンチウィルスによる対策が行われなかった
- ・ 非受理になった原因は、LZH が国際的にはマイナーな形式であるためだと思われる

関係する記事は以下の通りです。

[Claybird の日記 Unlha32.dll 等開発停止、LHA 書庫の使用中止呼びかけ](#)

[Gigazine 国産の圧縮形式「LZH」の UNLHA32.DLL の開発中止へ、LZH 形式使用中止を呼びかけ](#)

[CAD 日記 消えゆく LZH](#)

要するに、LHA 書庫に対してウイルス対策ソフトが正しく扱うことができなくなってしまい、バッファオーバーフローやウイルス感染という脆弱性が強まったということなのです。

よって、LZH を採用することはできません。

Windows 標準機能でも zip の圧縮と解凍ができるほど、zip は広まっています。

SXF を圧縮するなら zip 形式、ということで盤石な方針でしょう。

さて、いかがでしたでしょうか。

今回は「(5) ZIP 圧縮・解凍のプログラム実装って難しいの？」をお送りする予定です。

zlib というフリーのライブラリーがあります。もちろん商用利用可能です。 <http://www.zlib.net/> だいぶん枯れてきた安定したライブラリであり、かつ直近では 2013/4 に更新されており、きちんとメンテナンスがされているものです。

Windows の 32bit、64bit で動作可能ですし、CAD のプログラム言語としては主流である C++ での実装が容易に可能です。

zlib のライセンス条項には「できれば zlib の出典を書いてください」とあります。

zlib のライセンス条項の該当部分を引用します。

Permission is granted to anyone to use this software for any purpose, including commercial applications, and to alter it and redistribute it freely, subject to the following restrictions:

1. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

...

翻訳すると以下の通りです。

以下の注釈を守れば、商用であっても自由に改変、再配布してよいです。

1. 製品のオリジナルソフトウェアとして謳うことは駄目です。
できれば製品に出典を書いてください。

...

具体的にどうすればよいのかってことになりますので、C++ での実装概要を以下に記します。

1. zlib のトップページにアクセス

<http://www.zlib.net/>

2. zlib を使ったツールページにアクセス

1 の中をスクロールさせて、真ん中より少し下の「Related External Links」の中の「zlib for Windows 9x/NT/2000/XP/2003 (DLL version, plus related utilities)」をクリックする

3. ファイルをダウンロード

2 のページ上のほうに Current version(1.2.5)のすぐ下、「zlib125.zip」と「zlib125dll.zip」をダウンロードする

4. 使うファイルを探し出す

3 の解凍結果から以下ファイルを使う。

【DLL】	dll32¥zlibwapi.dll
	dllx64¥zlibwapi.dll ※CAD が 64bit ネイティブアプリの場合のみ
【ヘッダー】	zlib-1.2.5¥zlib.h
	zlib-1.2.5¥zconf.h
	zlib-1.2.5¥contrib¥minizip¥crypt.h
	zlib-1.2.5¥contrib¥minizip¥zip.h
	zlib-1.2.5¥contrib¥minizip¥unzip.h
	zlib-1.2.5¥contrib¥minizip¥iowin32.h
	zlib-1.2.5¥contrib¥minizip¥ioapi.h

5. 実装する

zlib のトップページから日本語情報へのリンクが貼ってあります。

下の方の「zlib information in Japanese」

<http://oku.edu.mie-u.ac.jp/~okumura/compression/zlib.html>

zlib の基本を押さえるには、これが参考になります。

すでに zlib を実装されている CAD ベンダーさんは多いかと思えます。

ただ、現在実装されていない CAD ベンダーさんが初めて実装するとなれば、それなりのコストと負荷がかかるかと思えますので、できるだけ容易に実装してもらおうべく、サンプルプログラムを公開することを検討しています。

さて、いかがでしたでしょうか。

次回は「(6) ZIP 解凍プログラム実装時の課題と対策」をお送りする予定です。

第 6 回

SXF の ZIP 圧縮 (6) ～ ZIP 圧縮・解凍プログラム実装後の課題 1 と対策

ZIP 圧縮・解凍プログラム実装後の課題は、大きく 2 つあります。
今回はそのうちの一つ「画像ファイルの扱い」について、お送りします。

CAD の仕様によりますが、ここでは図面に画像を配置した場合に画像ファイルそのものを図面ファイル内に保持せずに、画像ファイルパス情報だけを保持する CAD の場合の課題になります。

CAD で SXF を読み込むときは、圧縮 SXF を作業領域に解凍します。
これにより画像ファイルも作業領域に配置されて、CAD は作業領域の画像ファイルパスを保持します。
作業領域に画像ファイルがあるうちは問題ないですが、圧縮 SXF 読み込みの度に解凍後のファイルが増えていくと都合が悪いので、しかるべきタイミングで画像ファイルを削除しなければなりません。

SXF 読み込み直後に削除するということが考えられますが、その結果画像ファイルの表示や編集ができなくなる可能性があります。

作業領域中の画像ファイルを、どのタイミングで削除すれば不都合がなくなるのかという課題であり、そもそも解凍後の画像ファイルをどのように管理するのかということにまで発展します。

対策は「CAD 側の実装を工夫する」ということになります。

さて、いかがでしたでしょうか。

次回は、今回の続きで「(7) ZIP 圧縮・解凍プログラム実装後の課題 2 と対策」をお送りする予定です。

第 7 回

SXF の ZIP 圧縮 (7) ～ ZIP 圧縮・解凍プログラム実装後の課題 2 と対策

ZIP 圧縮・解凍プログラム実装後の課題 2 つのもう一つ、「ファイル名の不整合」について、お送りします。

ユーザーがエクスプローラを使って圧縮後 SXF をリネームすると、圧縮後 SXF と圧縮前 SXF のファイル名が不一致となります。CAD の機能でリネームするなら両者を一致させることも可能ですが、エクスプローラによるリネームであれば、両者の不一致が必ずおきます。
不一致の場合、どちらを SXF のファイル名とするか？

圧縮後 SXF ファイル名を常に正とすることになるでしょう。

CAD が圧縮 SXF 読み込んだ時は、圧縮前 SXF ファイル名を無視して、圧縮後の SXF ファイル名の図面として扱うことで、両者の整合性が保たれます。

つまり、SXF の ZIP 圧縮におけるルール(実装規約)で、「圧縮後 SXF のファイル名を常に正とする」ということを取り決めればよいということになります。

なお、現状の SXF 実装規約では「SXF ファイルと属性付加機構の SAF ファイルの名前は一致していなければならない」と定義されています。

これまではファイル名を変更する際、両者それぞれをリネームする必要がありましたが、SXF の ZIP 圧縮が実現すれば、圧縮後 SXF をリネームすることで、SXF と SAF を同時に変更したことになるというメリットが出てきます。

さて、いかがでしたでしょうか。

次回は、最終回「(8)まとめ」をお送りする予定です。

第 8 回

SXF の ZIP 圧縮(8)～ まとめ

SXF の ZIP 圧縮というテーマで書いてきたシリーズも、今回は最終回です。
これまでの記事の要点を挙げてみます。

1. 長らく変化していない SXF 規格なわけですが、SXF がより広く使われるようになるために、規格を追加しましょう。
2. ZIP 圧縮のメリットは、ファイルサイズ縮小と複数ファイルが 1 つにまとまることです。
3. 圧縮後ファイルの拡張子は、sfz と p21z が有力です。(第 3 回の配信当時)
※その後の検討で、3 文字拡張子の sfz と p2z に変更しました。電子納品の要領・基準類に、なるべく早く採用されることを意識したためです。
4. ファイル圧縮には LZH 形式もありますが、近頃使われていないし脆弱性の問題もあって、ZIP 形式での圧縮が磐石な方針です。
5. フリーで商用利用可能なライブラリー zlib は、CAD への実装が比較的容易です。
6. CAD への実装課題その 1「画像ファイルの管理」は CAD の実装で工夫が必要です。
7. CAD への実装課題その 2「ファイル名の整合性」は圧縮後ファイル名を正とします。

建設 CALS/EC の一環として、SXF は誕生しました。

国が発注する公共工事で提出する CAD 図面の電子ファイルフォーマットとして創り出されました。

近い将来 SXF が ZIP 圧縮されることで、コンパクトかつ扱いやすいファイルになります。

これからも SXF は進化を続けて、継続的な調達とライフサイクルの支援を促進させるように、OCF は活動してまいります。

-以上