Preserving CAD

An Overview

Alex Ball

26 July 2013

It's wonderful to be speaking at the IMechE ... also appropriate ... here twice before ... the KIM Project[1] ... first introduced me to the challenges of preserving Computer-Aided Design (CAD).

Contents

1 What is CAD and why should we care?

CAD is strictly speaking an activity, but we're interested in the tools and outputs of that activity: the CAD systems and the CAD models. They can be used for all sorts of things:¶

- 2D design drawings
- Floor/site plans
- Archaeological site records
- 3D product models
- 3D architectural models
- 3D impressions/reconstructions of buildings
- Virtual worlds
- 3D animations

[1] http://www-edc.eng.cam.ac.uk/kim/

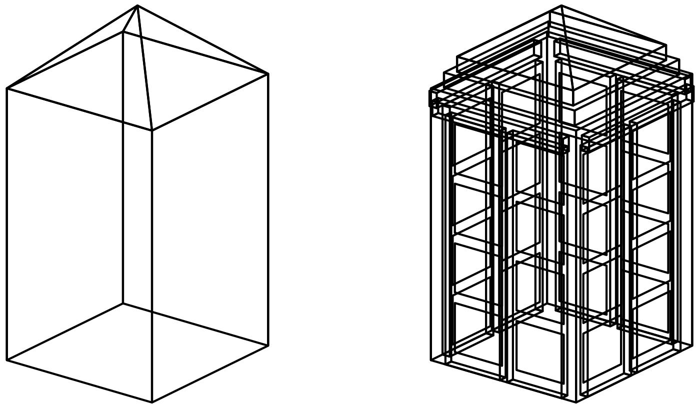

Figure 1: Wire-frame modelling

§ And when you come back to them later, you tend to be interested in doing one of three things: looking up what they say (*Reference*), asking why the model turned out the way it did (*Rationale*), or adapting it for a new purpose (*Reuse*). [2]

¶ Why should we be interested in preserving these things? Archaeology…destructive science…cannot recreate information later, so the records need to be preserved…CAD has proven to be a really useful documentation tool. Engineering and architecture …aircraft carriers and skyscrapers may be in service for a century or more. In that time they will need spare parts, or be refitted several times…things may even go wrong, and the authorities might need to find out why. Paper printouts can't express complex 3D curvature, they can be dangerously ambiguous with respect to handedness and offer few clues as to why a product was designed a particular way. It is also massively wasteful and error-prone to recreate a CAD model from paper plans if a few adaptations need to be made.

2 Why is CAD challenging?

There's such variety in what the systems can do…different systems behave differently depending on the market at which they are aimed. Also the use cases to which a model is put strongly influences which of its properties need to be prioritised for preservation.

The situation is made even more complicated by the different ways in which CAD can be implemented. How do you draw in 3D?

Wire-frame A line drawing but in 3D (¶ see Figure 1). The shape is picked out by its edges. But you just have a frame: there are no real surfaces involved, so you can't define what goes on away from the edges. Plus § things quickly start to look crowded on the screen once the design gets more complex. To solve these problems one needs to introduce the concept of a surface.

Surface modelling Various ways of doing this (¶ see Figure 2):

[2] Lubell et al. (2008), 'Sustaining Engineering Informatics: Toward Methods and Metrics for Digital Curation', *IJDC* 3/2: 59–73. `http://dx.doi.org/10.2218/ijdc.v3i2.58`

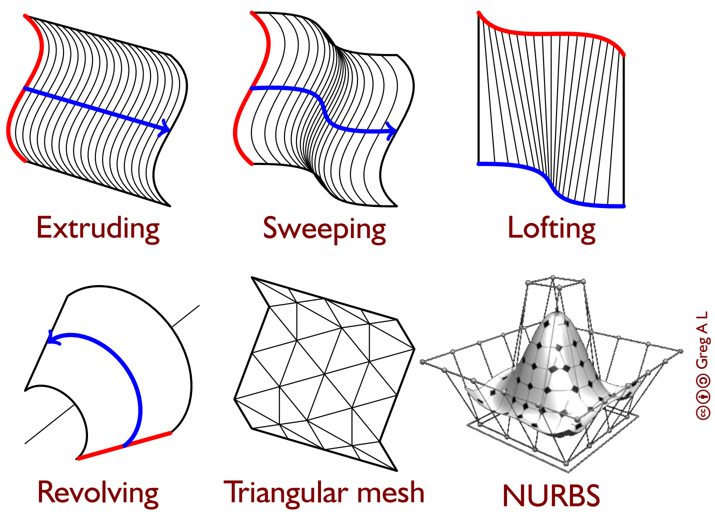

Figure 2: Surface modelling

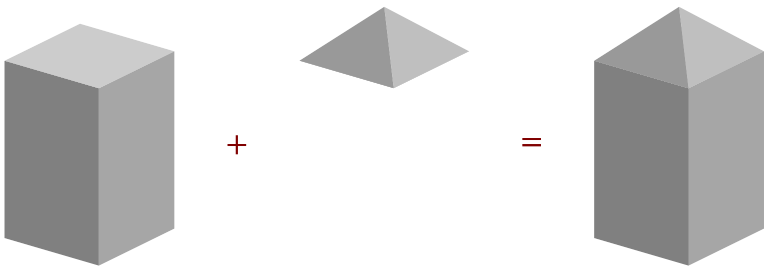

Figure 3: Constructive Solid Geometry

- Extruding – taking a profile curve and moving it along a straight line
- Sweeping – taking a profile curve and moving it along another profile curve
- Lofting – joining two profile curve with lots of straight lines
- Revolving – revolving a profile curve about an axis
- Approximating with a mesh of tessellating triangles
- NURBS (non-uniform rational B-splines)

But it's not easy to ensure all these surfaces join up to form solid objects, nor to calculate where the centre of mass is going to be. To do that you need to introduce the concept of solidity.

Constructive Solid Geometry Building up a solid object by deforming simple shapes and adding, subtracting, and intersecting them (¶ see Figure 3). The CGI for the 1982 film *Tron* was all done in SynthaVision, the first commercial solid modeller.

Boundary representation Building up a solid object by combining surfaces.

So immediately you can see that even in principle, trying to go from any one of these systems to any of the others is going to involve loss of information. And if you think that's bad, there's more to come. Constructing a CAD model is not like producing a bitmap image where the end product can be defined simply and easily. It's more like creating a web page, where you write a set of instructions for the computer to follow. The bit of a CAD system that turns the instructions into geometry is called a modelling

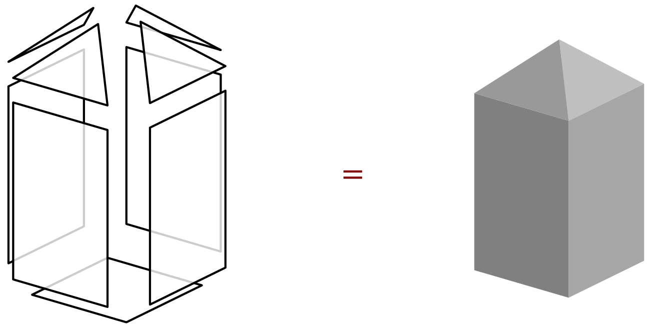

Figure 4: Boundary representation

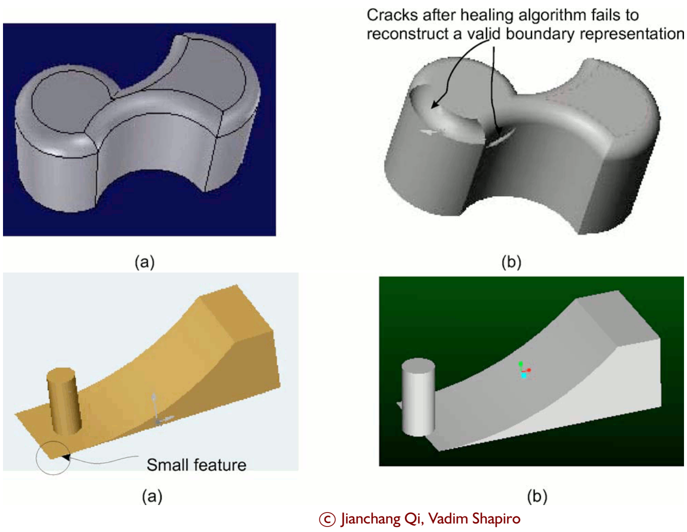

Figure 5: Mistranslation and misinterpretation of CAD data

kernel. And just as web sites can look and behave differently in different browsers, you can end up with quite different geometry by running the same instructions through two different kernels (¶ see Figure 5). And in fact because of the complexity of the maths involved the discrepancies tend to be somewhat greater.

That complexity is just the start. The CAD marketplace is intensely competitive, and vendors are constantly innovating to solve more and more of their customers' problems …systems become obsolete alarmingly quickly.

Construction history modelling (¶ see Figure 6) is a bit like the ultimate undo button. Not only does it allow the designer to step back through previous editing sessions, it also allows the designer to introduce a change in one of the earlier stages of the design and then replay subsequent editing actions.

Procedural modelling (¶ see Figure 7), meanwhile, is like Word or Excel macros writ large…apply algorithms not only to automate repetitive tasks but to give surfaces fine texture, generate realistic snowdrifts or extrapolate whole cityscapes from a single building…not used so much in design work…more for virtual worlds.

Parametric modelling helps to make design easier to adjust (¶ see Figure 8). Aspects of the design are made to depend on variables, and constraints are added so the computer knows how to adjust things should those variables change. Say a design included a

1. Insert cylinder $l = 20$ $r = 1.0$

   ⟵ Change cylinder $l = 40$ $r = 0.5$

2. Insert sprocket $r = 3.0$
3. Fit sprocket to cylinder
4. Group cylinder and sprocket
5. Scale group by $1.75\times$

...

Figure 6: Construction history modelling

Figure 7: Procedural modelling was used to generate the city (left) and to texture and displace the bricks (right)

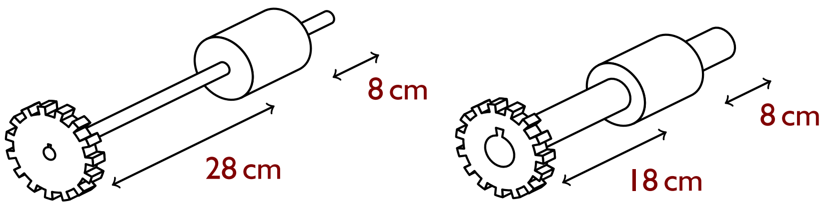

Figure 8: Parametric modelling was used here to adjust the design and position of a gear wheel and pulley to accommodate a shorter, thicker drive shaft

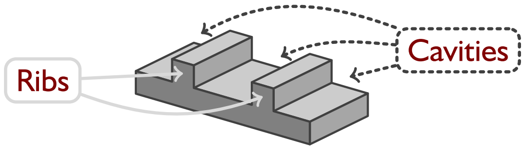

Figure 9: Sample model with an ambiguous feature set

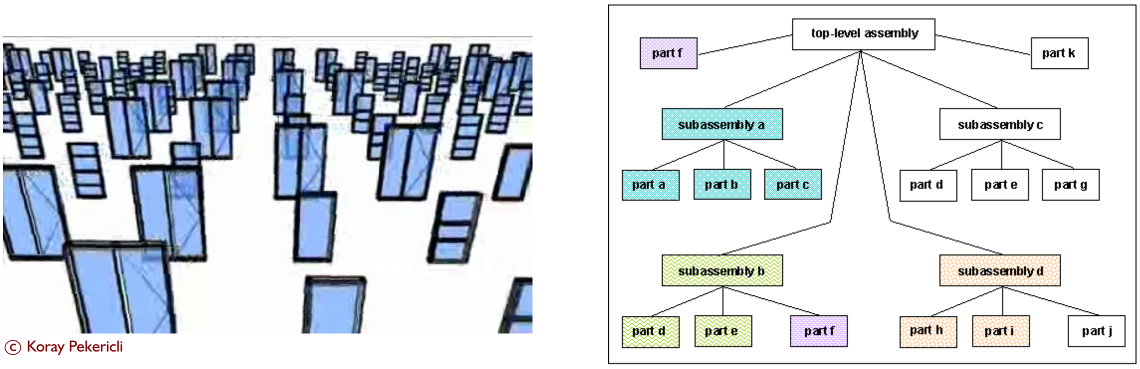

Figure 10: Reusing standard parts

gear wheel... fiddly: the spacing of the teeth is independent of the wheel radius so changing the radius means changing the number of teeth... in a parametric model, the computer can take care of that,

¶ Feature-based modelling is about making explicit the engineering significance of certain shapes in the design. So a curved blend between two surfaces might be there as an artefact of the manufacturing process or as a deliberate design decision to spread the stress; the feature semantics would tell you which. Incidentally, this sort of thing can cause conflicts. Take this shape (Figure 9).

The designer might need this marked up as rib features applied to a base surface in order to calculate the optimum distribution of ribs using finite element analysis, say. The manufacturing engineer, though, might need it marked up as cavities cut out of a thick surface. It's not usually possible to have both in the same model.

Speaking of manufacturing, there's a certain efficiency in using the same part in many different places... crazy to have a bespoke screw for each fitting. Those efficiencies also exist at the design stage, so CAD systems provide mechanisms for including the same part many times in an assembly (¶ see Figure 10). They may have a library of standard parts... preloaded with feature semantics, parametric properties... Architectural CAD package might have doors, windows, staircases... If these parts are hard-coded into the software, it can be very difficult to represent them reliably in a different system.

The other alternative is having the parts in separate CAD files and included in an assembly by reference. Actually, even if the part is only used once there can be advantages to splitting the design into many small pieces. But it does mean having to keep track of many different files and ensuring they point to each other correctly when placed on a different file system... relative path names, IDs managed by a central register... In industrial engineering, this would be done by the Product Lifecycle Management (PLM) system... in architecture and construction, a Building Information Model (BIM) system.

Such systems do more than just assemble CAD files … manage relationships with external databases, finite element analyses (FEA), bills of materials (BoMs), systems models (e.g. wiring diagrams), process and rationale models. Relationships might be static or dynamic … (BoM example). Full understanding and reusability of CAD models may depend on these other resources … some cases might need dynamic relationships intact. PLM and BIM systems also manage the security of the information … preservation systems might need to mimic the access controls of these systems, or circumvent security features built into CAD models themselves.

One last point … when thinking about how CAD will be used … not just CAD systems that will need to read the models … also read by FEA, manufacturing systems, geographic information systems and Computer Generated Imagery renderers and ray tracers. So whatever migration or emulation strategy is chosen might also need to cater for these other types of software.

3 CAD standards

The first commercial CAD systems were sold in 1969 and it soon became apparent that having CAD models tied to a particular software product was a serious problem … need for a standard exchange format.¶

First attempt was IGES (Initial Graphics Exchange Specification). First version 1979. Development led by what is now NIST (National Institute of Standards and Technology) and involving major CAD vendors and customers in the US. Despite being withdrawn as a standard in 2006 it is still widely supported. Problem … it's a very flexible standard. It supports many ways of doing things, and vendors can pick and choose what to support and what to ignore. No way of ensuring consistent implementation across system since no conformance tests.

In the mid-1980s, the French developed SET (*Standard D'Echange et de Transfert*) and the Germans VDA-FS (*Verband der Automobilindustrie-Flächen-Schnittstelle*) … aimed to improve on IGES, but neither won significant support globally. All three standards were narrow in scope compared to the data that needed to be exchanged.

So ISO became involved and started work on an international standard for the exchange of product model data, nicknamed STEP. It took them two attempts and eleven years, but the first version – well, first eleven parts – were published in 1995. The standard has since grown to over 590 parts, of which the most pertinent for us today are AP 203 and AP 214 … define CAD file formats suitable for exchange. We're eagerly awaiting AP 242 which will combine those two and bring in some important new features (shape data quality information, semantic 3D PMI, approximate geometry and access rights management).

STEP has been highly influential, with AP 203 in particular enjoying widespread support. There is an initiative called the CAx Implementer Forum in which CAD vendors test and improve their STEP converters. Several other standards and specifications build on STEP, not least LOTAR which you will hear about in a few moments, and the Industry Foundation Classes which provide an exchange standard for BIMs.

STEP should really be the last word in CAD standards, but there are a few others worth mentioning that were developed to serve different use cases. Sometimes you don't need to exchange a comprehensive set of product information. You might just want to send a design to a client to see if they think it looks right. This is where visualisation standards come in.¶

Typically . . . don't support construction history or parametric modelling . . . do support compression, multiple levels of detail, streaming for quick display over the Internet.

JT was originally vendor-led . . . now an International Standard . . . widely adopted in aerospace and the automotive industry . . . linking up with STEP AP 242.

3D PDF hasn't taken off in quite the same way but there are firms that have found it useful . . . quite nice for interactive design drawings . . . 3D modelling component is supplied either by Universal 3D (approximate geometry; ECMA standard) or PRC (exact geometry; final stages of becoming an International Standard; similar to JT but not quite as powerful).

Also worth a mention is X3D . . . developed for virtual reality applications, so has dedicated support for human animation and interactivity. Also supports layers, which are used extensively in archaeology.

4 CAD non-standards

I haven't talked much about archaeology . . . quite different world to engineering and construction.

Dominated by AutoCAD, to the extent that its DWG format is considered a *de facto* standard. The format changes often and its specification is secret, though you can by the RealDWG library from AutoDesk . . . But being so popular there are efforts to reverse engineer the format: the Teigha library from the Open Design Alliance, and LibreDWG from the GNU Project (open source, pre-alpha).

There is an alternative format for AutoCAD called DXF . . . specifications are available, though recently they have not been as thorough as they might be. And, like IGES, no levels of conformance so no guarantees two systems will support the same portions.

Also has its own visualisation standards, DWF and DWFx.

5 Preservation strategies

As mentioned above, different circumstances have different priorities, so the best strategy to adopt will vary.

Preserve the original CAD model

- Implies preserving software through emulation.

Pros preserves maximum information; easier to guarantee provenance.

Cons need to preserve expertise in the system; need an amenable software licence; hard to maintain integration with current systems.

Good for reference purposes, but not reuse.

Rolling format migrations

- Migration to newer format versions or new CAD systems.

Pros models usable by current designers and software.

Cons cost of validating each migration; incremental data loss/corruption.

Good for models in active development/use, but not for long-term archiving.

Normalisation

- Migration to (*a*) STEP/IFC (*b*) a visualisation format.

Pros only two migrations needed, so limited data loss/corruption; back-up in case a migration goes wrong

Cons cost of validating each migration.

Good for long-term archiving and reuse.

Additional good ideas

Validation Record validation properties (volume, centre of gravity, point cloud) in original CAD system, then test for them after import into another.

Indirect links With linked files, replace any absolute links with relative links or resolvable identifiers.

Split files Archive part files as separate packages, and assemblies as super-packages.

Annotation If information is known to be lost on migration, try preserving it using annotations.

Metadata Supplement the CAD files with specifications, layer naming conventions, file naming conventions, data collection documentation . . .

House style Encourage the use of a (documented) house style, to aid clarity, avoid migration problems and preserve semantics.

6 Final thoughts

- *CAD files can be used for many different purposes, so it is important for an archive to* Establish why a CAD model will be kept, then target the required properties for preservation.
- Create tests that can prove whether these properties have survived.
- Keep native CAD models for as long as they can be read.
- Normalise to STEP/IFC and a visualisation standard (or two).
- Don't forget supporting documentation, especially local conventions and 'house style'.
- Campaign for better support for standard formats in CAD systems!
- Oh, and read the full report: `http://dx.doi.org/10.7207/twr13-02`

Alex Ball. DCC/UKOLN, University of Bath. *`http://alexball.me.uk/`*

The DCC is funded by JISC.

For more information, please visit `http://www.dcc.ac.uk/`