Preserving Computer-Aided Design (CAD)

Alex Ball

DPC Technology Watch Report 13-02 April 2013

Series editors on behalf of the DPC
Charles Beagrie Ltd.

Principal Investigator for the Series
Neil Beagrie

DPC Technology Watch Series

Published in association with Charles Beagrie Ltd and Jisc's Digital Curation Centre.

ISSN: 2048-7916
DOI: http://dx.doi.org/10.7207/twr13-02

First published in Great Britain in 2013 by the Digital Preservation Coalition

Foreword

The Digital Preservation Coalition (DPC) is an advocate and catalyst for digital preservation, ensuring our members can deliver resilient long-term access to digital content and services. It is a not-for-profit membership organization whose primary objective is to raise awareness of the importance of the preservation of digital material and the attendant strategic, cultural and technological issues. It supports its members through knowledge exchange, capacity building, assurance, advocacy and partnership. The DPC's vision is to make our digital memory accessible tomorrow.

The *DPC Technology Watch Reports* identify, delineate, monitor and address topics that have a major bearing on ensuring our collected digital memory will be available tomorrow. They provide an advanced introduction in order to support those charged with ensuring a robust digital memory, and they are of general interest to a wide and international audience with interests in computing, information management, collections management and technology. The reports are commissioned after consultation among DPC members about shared priorities and challenges; they are commissioned from experts; and they are thoroughly scrutinized by peers before being released. The authors are asked to provide reports that are informed, current, concise and balanced; that lower the barriers to participation in digital preservation; and that they are of wide utility. The reports are a distinctive and lasting contribution to the dissemination of good practice in digital preservation.

This report was written by Alex Ball, a specialist in digital curation at the UK Digital Curation Centre and employed by UKOLN at the University of Bath. The report is published by the DPC in association with Charles Beagrie Ltd and the Digital Curation Centre. Neil Beagrie, Director of Consultancy at Charles Beagrie Ltd, was commissioned to act as principal investigator for, and Managing Editor of, this Series in 2011. He has been further supported by an Editorial Board drawn from DPC members and peer reviewers who comment on text prior to release: William Kilbride (Chair), Janet Delve (University of Portsmouth), Sarah Higgins (University of Aberystwyth), Tim Keefe (Trinity College Dublin), Andrew McHugh (University of Glasgow) and Dave Thompson (WellcomeLibrary).

Acknowledgements

The author's expertise in this area largely derives from working alongside colleagues in the Innovative Design and Manufacturing Centre at the University of Bath, and its peer centres in institutions across the UK. He would therefore like to express his gratitude to them and their industrial partners for such a thorough grounding.

The author would also like to thank the speakers at the DPC workshop *Designed to Last*– namely Sharon McMeekin (Royal Commission on the Ancient and Historical Monuments of Scotland), Sean Barker (BAE Systems), Kieron Niven (Archaeology Data Service), Chris Puttick (Oxford Archaeology) and Kurt Helfrich (Royal Institute of British Architects) – for a broader understanding of the uses of CAD and the problems faced by organizations with archival responsibilities. Additional thanks must go to the two anonymous reviewers and the DPC editorial team for their many helpful suggestions and insights.

Alex Ball, University of Bath

April 2013

Contents

Abstract 1

Executive Summary 2

1. Historical introduction to CAD 4

1.1. Motivations for the development of CAD 4

1.2. Three-dimensional modelling 4

1.3. Advanced modelling techniques 6

1.4. Integration 7

1.5. Alternative uses of CAD 7

1.6. Importance of preserving CAD 8

2. Issues 10

2.1. Nature of the CAD marketplace 10

2.2. Interoperability with present and future systems 11

2.3. Linkages between files 11

2.4. Relationships with other information 12

2.5. Viewpoints 12

2.6. Security 13

3. Standards 14

3.1. IGES 14

3.2. STEP 14

3.2.1. IFC and NBIMS-US 16

3.2.2. LOTAR 16

3.3. VRML and X3D 17

3.4. U3D, PRC and 3D PDF 17

3.5. JT 18

3.6. AutoDesk formats 18

4. Techniques and Technologies 19

4.1. VDA Recommendation 4958 19

4.2. The KIM Project 20

4.3. The FACADE Project 21

4.4. Sustaining Engineering Informatics 22

4.5. The SHAMAN Project 23

4.6. Archaeological metadata 24

5. Discussion 25
5.1. Preservation of the original CAD model 25
5.2. Rolling format migrations 25
5.3. Normalization to standard formats 26
5.4. Validation 26
5.5. Multi-file CAD models 27
5.6. Links and Annotations 27
5.7. Style 28
5.8. Advocacy 28
6. Conclusions and Recommended Actions 29
7. Glossary 31
8. Further Reading 35
9. References 36
Appendix I: Structure of STEP 42
Appendix II: Structure of LOTAR 43

Abstract

Computer-Aided Design (CAD) systems are used in both industry and academia to create digital models, whether of engineering designs, archaeological dig sites, or virtual worlds. These models can be of long-lasting significance and importance, particularly if they contain irreplaceable data or relate to long-lived products. This report is primarily aimed at those responsible for archives and repositories with CAD content, but may also be useful for creators of CAD content who want to make their models more amenable to preservation. It begins with an introduction to the historical development and basic concepts of CAD systems, then reviews the most pertinent issues associated with preserving CAD models, and indicates the current state of standardization work in the area. The report goes on to present some recent research of relevance to preserving CAD models before drawing conclusions and making recommendations on how archives should handle the CAD models they accept.

Executive Summary

Computer-Aided Design (CAD) is a technology originally developed to assist engineers and architects in producing large and complex designs. At first little more than electronic drawing boards, CAD systems are now capable of producing sophisticated virtual worlds and are used routinely outside their original target community, in fields such as archaeology and the entertainment industry.

The success of CAD means that a great deal of vital, valuable and irreplaceable information is stored in CAD models, from the designs of aircraft carriers and skyscrapers to records of archaeological excavations. It also means that CAD is an area of constant innovation and intense competition between vendors, resulting in CAD systems that are ephemeral and largely incompatible with each other. It is the disconnect between these two – the value of the models and the ephemeral nature of the systems – that makes CAD preservation at once important and challenging.

This report makes the following recommendations to those responsible for archives and repositories with CAD content:

- Determine the likely uses to which the preserved CAD models will be put, and from this determine which aspects of the models should be targeted for preservation.
- Select a set of properties that will be used to determine whether the selected aspects of the models have been preserved (e.g.volumes of solid objects, identifiers for surfaces). These should be recorded for each CAD model ingested into the archive or repository, and used as the basis for validation whenever the CADmodel is migrated to a new format or loaded into a new system. Where models in vendor-neutral standard formats fail validation, re-check them using a different software package if possible.
- Keep native CAD models for as long as they can be read (accurately) by available software. Also, normalize CAD models to at least one, but ideally two or three, vendor-neutral standard formats. The formats defined by STEP (ISO 10303) are ideal.
- With large CAD models consisting of many assembly and part files, archive each file as a separate archival information package, though with all the packages linked and alongside instructions on reconstituting the full model.
- Work closely with depositors to ensure that all the information required to understand the model is archived alongside it. Such documentation might include specification documents, process and rationale models or reports, file naming conventions, layer naming conventions, drawing conventions, materials data sheets, parts catalogues, or supplementary databases. Encourage the use of a documented house style for CAD models wherever possible.
- Express any filesystem links between CAD models and other files in an indirect or relative fashion (rather than using full path names) wherever possible.

The wider preservation community should build a business case that underlines the importance of interoperability and preservation for CAD customers and vendors, and use it to campaign within both groups (and beyond) for better support for standard formats in CAD systems.

1. Historical introduction to CAD

Computer-Aided Design (CAD) is, as the name suggests, the use of computers to assist with the design of manufactured products, the built environment, or fictitious environments. More specifically, it refers to software – and originally computer hardware as well – for creating digital models of physical objects.

CAD systems are typically expensive and complicated pieces of software, and their native file formats are equally complicated, opaque, and in an important sense, incomplete. This makes them hard to preserve, a fact which has driven standardization initiatives for over 40 years and will continue to do so long into the future.

In order to preserve CAD models, it is helpful to understand the context in which CAD systems were first developed, why they evolved as they did, how they differ from one another, and how they are used today.

1.1. Motivations for the development of CAD

The rise of CAD systems in the 1960s was motivated by the sheer impracticality of drawing designs by hand. Not only was the process laborious and error-prone, it could also cause practical headaches. Many designs had be drawn to a scale of 1:1, which was something of a challenge when it came to aircraft wings or ship hulls (Salomon, 2006, viii; Weisberg, 2008, 2.4). A further driver was the development in 1957 of PRONTO, the first commercial computer numerical control (CNC) system, which could be used to automate certain machining processes. Programming such systems from paper plans was, again, laborious and error-prone, and would be considerably easier if the shapes involved were already mathematically defined.

Thus, from the late 1950s to the mid-1970s there was an intensive effort by both industry and academia to find mathematical representations of the paper designs, and to create tools for authoring them. Probably the first recognizable CAD system was SKETCHPAD, developed between 1960 and 1963 at MIT by Ivan Sutherland. User input was via a light-pen, with which the designer drew on the computer screen. Major industry players such as Ford, Rénault and Lockheed developed in-house CAD systems in the 1960s, and the first successful commercial CAD systems appeared in 1969.

1.2. Three-dimensional modelling

The earliest CAD models were two-dimensional, more or less a digital analogue of the blueprint. CAD systems solved many efficiency problems: designers could easily copy and paste repeated design elements, run scripts instead of laying out everything by hand, and avoid or correct mistakes more easily. But what firms really wanted to do was to input CAD models directly into CNC systems, and the CNC systems worked in three dimensions. Three-dimensional shape data was needed.

The first approach used for 3D-modelling involved wire-frames, where shapes were represented solely by their vertices and edges. While computationally simple, the technique could not express complex surface curvatures, however, and intricate designs quickly became unreadable.

The next generation of systems used surface modelling. Several mathematical constructs for representing surfaces were tried, but eventually non-uniform rational B-splines, or NURBS, emerged as the standard. NURBS turned out to unify most of the previous techniques, and are still widely used for representing exact geometry today.

One of the drawbacks of moving from full-scale drawings to computer systems and their small screens was that designers found it harder to detect shape defects by eye (Farin, 2002, 13). In response to this problem, CAD systems started to have a role in analysing the designs they were used to author. There is only so far one can go using surfaces alone, though, and in order to determine if surfaces join up to make realistic objects, the effects of mass and materials, how the objects should properly be rendered and so on, the systems needed the concept of solidity. Various methods of solid modelling were researched and subsequently used in CAD systems, but the two that proved most popular were Constructive Solid Geometry (CSG) and Boundary Representation (B-Rep) (Stroud, 2006, 1–2). Figure 1 illustrates the difference between the two approaches.

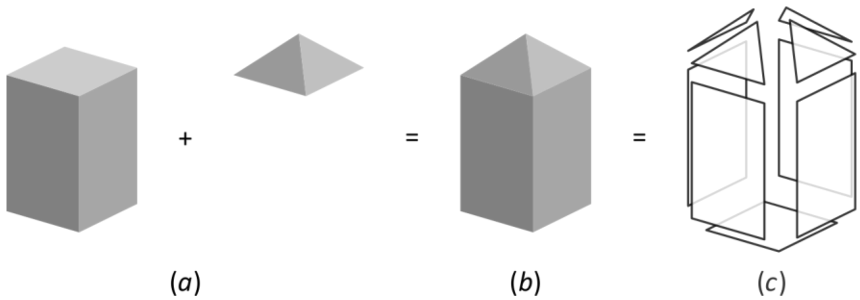

Figure 1: Demonstration of how a solid object (*b*) might be modelled using Constructive Solid Geometry (*a*) and Boundary Representation (*c*).

CSG is generally regarded as the simpler and less powerful of the two methods. It uses a standard set of primitive shapes – spheres, cones, cubes, cylinders, etc.– from which more complex shapes are constructed by deformation, union/addition, difference/subtraction, and intersection. The first solid modeller sold commercially, MAGI's SynthaVision from 1972, used a form of CSG, and went on to provide the computer-generated imagery in the 1982 film *Tron* (Elin, 1976; Weisberg, 2008, 2-13).

B-Rep, meanwhile, takes surface modelling and adds solid geometry intelligence to it. Among other things, it enforces rules for ensuring that a set of surfaces really do join up to form a solid object.

Calculating the geometric implications of shape data is the job of a modelling kernel. Most B-Rep modellers in use today use one of two kernels: either Parasolid or ACIS. (Both were produced by companies co-founded by Ian Braid, who wrote the first B-Rep modeller as part of a doctoral dissertation [Braid, 1974].)

The move to three dimensions was the point at which CAD models stopped being mere conveniences for drawing blueprints and started taking on importance in their own right. With 3D models, it became possible to design shapes that could not be clearly or adequately expressed by three 2D elevations. The ability to analyse designs in 3D meant that more ambitious designs could be attempted, and also that standards for design checks were raised beyond what could be done by eye. In the context of industrial product design, 2D surrogates soon became inadequate records and regarded as dangerously open to misinterpretation.

This fact has, admittedly, taken longer to permeate some areas than others. Regulatory bodies such as the (US) Federal Aviation Administration (FAA) and the European Aviation Safety Agency (EASA) still routinely receive aircraft designs in paper form, but even this is beginning to change. In 2011, Dassault Aviation's Falcon business jet became the first to receive both FAA and EASA approval on the basis of 3D data alone (IBM and Dassault Aviation, 2011; Garrouste, 2012). Dassault Aviation was permitted to submit its designs in this form after demonstrating its archiving processes complied with the emerging LOTAR standard (see Section 3.2.2).

1.3. Advanced modelling techniques

In parallel with these developments, other innovations were being introduced to increase the ease with which designs could be created and reused.

Many CAD vendors have implemented construction history modelling, which takes the idea of an 'undo' function and turns it into a sophisticated tool. The sequence of editing actions leading to the current model is recorded in the file, meaning that a designer can revisit any stage in the history of the model, even from a previous editing session. Once there, the designer can make adjustments, then replay the subsequent editing actions taking the adjustments into account. This function can also be useful from an investigative or reuse perspective, as it allows a second designer to inspect exactly how a model was put together and thereby infer the rationale of the original designer.

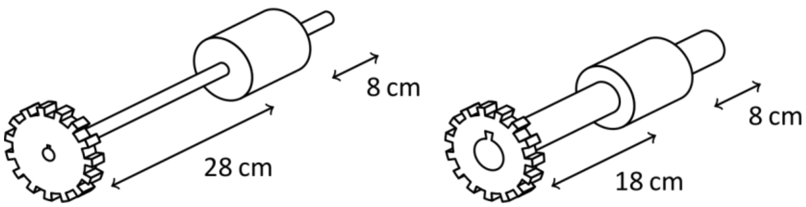

Figure 2: Demonstration of parametric variations on a drive shaft connecting a gear wheel and a pulley.

Parametric modelling is a powerful technique for making designs easier to adjust. Certain aspects of the design are made to depend on one or more variables, and a set of constraints is added so the system knows how to accommodate any changes to those variables. For example, Figure 2 shows a simplified drive shaft connecting a gear wheel and a pulley. In this design, the length and radius of the shaft are parameters; there are constraints that fix the positions of the gear wheel and the pulley relative to the two ends of the shaft, and ensure their inner surfaces remain contiguous with the surface of the shaft. The figure shows how a system might regenerate the design if the shaft were made shorter and thicker.

A logical extension of this is feature-based modelling, which adds a semantic layer to designs. A feature in this sense is a collection of characteristic shapes which are significant for the use or performance of the model. For example, a designer might apply a curved blend feature where two surfaces meet; the semantics associated with the feature will reveal whether the designer is anticipating how the part will be machined (the radius might correspond to an available cutter) or trying to avoid a stress concentration at the corner (or both).This is known as design by features; some systems have a facility for feature recognition, where the system uses pattern matching to apply semantics to a design, and some can perform feature validation to ensure the design is feasible. Feature-based modelling began to appear in commercial systems in the late 1980s, after about a decade of development (Shah and Mäntylä, 1995, 3–20).

Related to this, but somewhat distinct, is the way some commercial CAD systems provide pre-defined parts, often parameterized and with embedded engineering intelligence, from which to assemble a product. Such facilities are by necessity tailored to a particular market, as the components needed by automotive engineers (e.g. sprockets, drive shafts) are not the same as those required by architects (e.g. doors, windows, staircases).

All these modelling techniques embed far more information into a model than would be evident from just the final shape data. This information is highly useful for investigators trying to uncover why a product or building performed unexpectedly; for designers seeking to modify legacy designs in the light of new customer requirements; and for designers wanting to reuse elements in new designs. While some information might be inferred using sufficiently advanced feature recognition, having the original information would clearly be preferably if one were using the designs as evidence in a legal case or academic argument.

1.4. Integration

The current emphasis for CAD development in the engineering sector is not so much on geometric modelling as on integrating data and information from across the product lifecycle. Computer-Aided Engineering (CAE) is an umbrella term bringing together CAD, Computer-Aided Manufacture (CAM), Computer-Aided Process Planning (CAPP), Finite Element Analysis (FEA), and Material Requirements Planning (MRP), and is itself subsumed under the wider concept of Product Lifecycle Management (PLM), which also includes areas such as performance tracking, portfolio management, data archiving and so on. PLM systems are marketed as a way of integrating all these different tasks.

In architecture and construction, the (roughly) equivalent concept is Building Information Modelling (BIM), which seeks to integrate into a single authoritative model all the information needed to plan, design, construct, maintain and eventually demolish a building (Lee, Sacks and Eastman, 2006).

In both cases, multiple contributors with different access permissions need to be able to work collaboratively on the model, possibly from different sites, so version control, access control and synchronization of master models with simplified visualization surrogates are common features of PLM and BIM systems. While this level of integration and control is clearly of benefit, it presents additional preservation challenges. For designs to remain useful as designs, there will need to be ways of reading them not just into future CAD systems, but all the other future CAE systems as well. Archival systems may need to replicate or circumvent the access controls of PLM and BIM systems.

1.5. Alternative uses of CAD

The 3D-modelling capabilities of CAD have been found useful in many contexts outside pure product or architectural design, and this has driven the development of additional features and capabilities.

The use of CAD for animation and real-time simulation prompted research into multiple levels of detail (LOD). This is where a model contains, in addition to the canonical geometric representation that accurately expresses the designer's intentions, one or more simplified versions of an object. The latter are used when the object is distant, small or partially obscured from the viewer's perspective, as a way of speeding up rendering (Klein and Straßer, 1997).

Procedural modelling is a technique where sequences of editing actions are stored as algorithms so they can be applied automatically to 3D models. It was used originally for high quality rendering, but is now also used for generating complicated geometry: applying a layer of peeling paint to a flat surface, for example, or generating an entire city from a few sample buildings (Cutler, 2003; Smelik *et al*., 2009). Since the results tend to be pseudo-random in nature – that is, apparently random though generated deterministically – procedural modelling is used most extensively in the creation of virtual worlds rather than in formal design work.

For about as long as there has been interest in creating digital representations of imagined objects, there has also been an interest in creating digital models of existing objects. Some CAD systems can take data from co-ordinate measuring machines (CMMs) or computed tomography (CT) scanners and generate CAD models. In industry, such models might be used to compare manufactured components with their original designs to analyse wear, or as an initial step towards reverse engineering an artefact (Reed and Allen, 1997; Flisch *et al*., 1999; Kalender, 2006). It is also possible, of course, to import a set of data points into a CAD system and manually 'join the dots' to create a CAD model. This is a typical use case in archaeology, where CAD has found widespread use as a documentary tool for both 2D site plans and 3D models of sites and artefacts.

One of the attractions of CAD software for both archaeologists and architects is its ability to arrange data into layers which can be viewed or hidden as needed. Use of layers is so embedded in the architectural workflow that conventions on what information to include in each one and how to name them are the subject of national and international standards (e.g. *United States National CAD Standard* 2011; ISO 13567). In archaeology, layers are used in a much more varied fashion according to need: to separate out different materials, strata, building phases, functions, etc.; to separate out licensed map data from newly collected data, for example; and to separate out physical features from annotations (Eiteljorg *et al*., 2011). When preserving CAD models, therefore, it is important that the layer conventions are adequately documented so the significance of the data is known. Such conventions should form part of any archaeological data management plan and their documentation passed on to the archive when the CAD model is deposited.

1.6. Importance of preserving CAD

Since archaeology is a destructive science, as each site is dug the records produced by the dig team become the only source of information about that site: the information cannot be recreated later. It is highly important, therefore, that when archaeologists commit such information to a CAD model they can be confident that the model will still be readable and reusable long into the future.

In engineering and architectural contexts, too, there are strong motivating factors for ensuring the designs remain usable in the long term, not least the legal and regulatory requirements for this to be so. These requirements are themselves motivated by the need to know, if things go wrong, why they go wrong and how such incidents may be avoided in future. In such investigations the intent of designers can be just as important as an examination of the product as manufactured or constructed, and for this only the original designs and supporting documentation will do.

The requirements of the customer are likely to change several times over the life of an aircraft or hospital, say, meaning adaptations and modifications will have to be made. This will be considerably easier if the designer or architect has access to the construction history and feature semantics of the original design, and perhaps prohibitively expensive if the design has to be reverse engineered or recreated from scratch.

It is vitally important, then, that effective methods for preserving CAD models are developed and used, but there are many obstacles. The nature of the CAD marketplace, with its many incompatible and short-lived systems, is not conducive to the long-term usability of CAD models. When one considers the range of non-CAD systems that use CAD models, the problem multiplies. Certain characteristics of CAD formats, such as their ability to span many files and dynamically link to data sources, make them hard to transfer between systems, and the commercial sensitivity of industrial CAD data introduces some non-technical challenges. These issues are discussed in greater depth in Section 2.

Fortunately there are standards and techniques that can be employed to meet these challenges. The STEP standard (ISO 10303) in particular has had a positive impact on the portability of CAD data between systems. VDA Recommendation 4958 (VDA, 2005–2007), the emerging LOTAR standard and the FACADE project provide good models for CAD-friendly archives. Standards are discussed in more depth in Section 3, while techniques emerging from various initiatives and projects are described in Section 4. Finally, the themes of the report are drawn together and discussed in Section 5.

This report does not deal with more general aspects of preservation such as bit-level preservation, the collection of generic preservation metadata and software preservation. The interested reader is invited to consult the wealth of literature that already exists on those topics; some pointers may be found in Section 8. The terminology used in this report is intended to be harmonious with that used by the Open Archival Information System (OAIS) Reference Model (ISO 14721:2012), but close familiarity with that model is not required.

2. Issues

CAD models present some particular preservation challenges, only some of which are related to the file formats themselves. Important issues arise from the environments in which CAD systems are developed, licensed and used, from both a technological and a business standpoint.

2.1. Nature of the CAD marketplace

CAD vendors operate in an environment of intense competition. This has a real impact on the nature of CAD software, with important implications for preservation.

A CAD system with no export options would be deeply unattractive to customers, but one with high-quality export filters would make it dangerously easy for customers to migrate to competing products, or to run a mixed-system environment. It is in the interests of CAD vendors, therefore, to supply export filters that are not quite robust enough to satisfy the needs of their customers. It is also against their interests to reveal details of their proprietary file format specifications, as this might allow competitors to produce high-quality import functionality.

Even if the specifications were released, it is doubtful this would provide enough information to exactly replicate a CAD model in a different system. CAD files tend not to be exhaustive descriptions of a model, but rather more like recipes for building the model within the software (Qi and Shapiro, 2006). The modelling kernel and built-in feature logic have a strong influence on how a CAD file is interpreted, meaning that even later versions of the same piece of software, ostensibly using the same file format, might bring up somewhat different models on reading the same CAD file. Interoperability issues between versions 4 and 5 of CATIA have been blamed for delays to the delivery of the Airbus A380 aircraft (Wong, 2006), which is unsurprising given how some version 4 models were rendered by the version 5 software (Barker, 2010, 11).

The issue of poor interoperability between CAD systems and between versions is exacerbated by the rate of software development. In order to maintain a competitive edge, there is constant pressure on CAD vendors to release new versions of their software with increased functionality or fewer limitations. Not only does this create instability regarding file formats and their interpretation, it also means that individual versions of CAD packages can become obsolete rather quickly, especially when compared to the required lifespan of the CAD models they create. To put this in concrete terms, a new version of a typical CAD system might be released every six months, and the system withdrawn entirely after ten years. In contrast, a ship might be in service for 40 years, a public building for a century or more, and archaeological data will be of significance perhaps indefinitely.

With this rapid turnover of software versions, in order to decrease support costs and drive up revenue, it is not uncommon for CAD software to be licensed in a time-limited manner. Where this is the case, it prevents the software being preserved and run on emulated platforms in the future, closing off one avenue for reading legacy CAD models. Even were that not the case, it is doubtful whether future designers would be familiar enough with legacy, emulated systems to be able to use them in a production environment, due to the way design work evolves as CAD systems become more sophisticated.

2.2. Interoperability with present and future systems

A major selling point for CAD systems in most industrial and academic contexts is the ability to reuse the model data in specialist tools, or even better, to integrate with them directly. In the context of industrial engineering, CAD systems need to be able to communicate with FEA packages and CAM systems, integrate with product review systems, and synchronize with bills of materials, supply chain management software and perhaps other downstream performance monitoring systems. In archaeological research, it is often useful for CAD models to be imported into geographic information systems (GIS), for example, and with computer-generated imagery (CGI) the models need to be imported into dedicated rendering software.

Vendors' responses to these drivers are mixed in terms of how they impact on the preservation of the CAD models. Where multiple vendors are involved, interoperability implies a certain degree of openness about the file formats used and a greater likelihood that multiple versions of formats will be supported in the applications. Where the same vendor is responsible for the different components, there is no implication of wider file format support, and a likelihood that the component versions will be locked in step to encourage wholesale upgrading. The latter situation can actually be detrimental from a preservation perspective, since once such interoperability is relied on, it would not be enough to be able to read a CAD model using the original software; in order to reuse the design all the other software components (CAM, FEA, and so on) would have to be able to read the CAD model as well.

2.3. Linkages between files

Many CAD systems have the ability to link several smaller files together to create a single design. In the case of an engineering product model, for example, individual parts may each be saved in a different part file. Information on how to combine the parts into an assembly would be stored in an assembly file, which would reference the part files (or subordinate assembly files) rather than store the part information directly.

There are several advantages to this. It allows different designers to work on different parts of an assembly at the same time. It helps to keep the files small, meaning they can be transferred, opened and edited more smoothly; this is especially important where design teams are geographically distributed. It helps to keep designs modular, so that parts may be reused several times within and across designs, thus preventing duplication of effort and data. If the model is to be exported to another system (e.g. for rendering, finite element analysis or geospatial analysis), sometimes better results can be achieved if the parts are assembled in the external program rather than in the CAD system. It also provides more flexibility for archivists if one part of a design needs to be treated differently to another, perhaps because of differing intellectual property rights.

On the other hand, it is harder to maintain consistency across multiple CAD files than within a single file, and harder for the designer to appreciate the effect that modifications on a single part might have on the wider assembly. From a preservation perspective, the key concern with decomposed models is ensuring that the linkages between files are maintained when the project is transferred to a new location for archiving. This is much easier to achieve if the cross references specify file locations indirectly or in a relative rather than an absolute fashion, and if the interdependencies of the files are separately documented.

2.4. Relationships with other information

Related to both the previous issues is the fact that CAD models are frequently linked to other, non-geometric information resources. An archaeological model of an ancient building might be linked to a database that records the surface texture and tooling marks of each component stone block (Eiteljorg *et al*., 2011). A product model would normally have a corresponding bill of materials (BOM) and systems models (such as pipe runs or wiring diagrams), and may also be linked to a process model explaining how the designers arrived at that design, and a rationale model (or more usually a report) explaining why. While such information is seldom vital to the preservation of the CAD model itself, it provides additional context and may be required for full understanding and reuse of the model in future.

Depending on the capabilities of the software in question, links between CAD models and external resources can take various forms. The link might be a purely historical one, as with a static BOM document that was initially generated by a CAD system. On the other hand it could be more dynamic, as with BOM databases that monitor multiple CAD models, or PLM systems that swap component CAD models in and out of product lines depending on configuration options. The links could be hypertext-style references, where clicking on an icon in a CAD model brings up further information in an annotation file or vice versa. A CAD system might be able to pull information from a linked database and overlay that information on a model, or it might provide an application programming interface API that allows the external database to tweak how a particular model is displayed.

Given this variety, the preservation of such links has to be handled on a case-by-case basis. For some use cases it may be sufficient to preserve the external resource (or a snapshot thereof) alongside the CAD model, without preserving the interactions between them. For others, it may be possible to preserve the interactions by ensuring that files retain their internal bookmarks, anchors or identifiers through any format migrations. In some cases, it might be necessary to preserve the software that facilitates the interactions.

2.5. Viewpoints

As mentioned in Section 1.3, feature-based modelling is a convenient way of incorporating engineering or architectural intelligence into a design. One potential difficulty with it, though, is that it ties the model to a particular engineering viewpoint. For example, from a design viewpoint the model in Figure 3 might represent a base surface with two ribs added for structural support. The parametric information embedded in the model would therefore concern the height and width of the ribs as well as their relative placement. From a manufacturing viewpoint, though, the model might better be thought of as a thick surface with three wide cavities cut into it, in which case parametric information about the width and depth of the cavities would be more appropriate (Lee, McMahon and Lee, 2003).

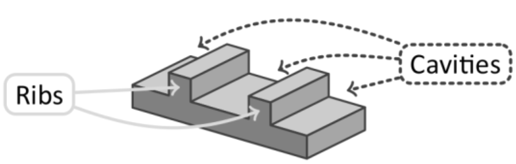

Figure 3: Sample model with an ambiguous feature set.

Issues

A common way to resolve this mismatch is to generate additional versions of the model with the original design semantics swapped for semantics tailored to other viewpoints: not only manufacturing but also structural analysis, process engineering and so on. As these other versions all contribute to the finished product, they also need to be preserved if a full record of the product's provenance is to be maintained; on the other hand, they do share a lot in common, not least the basic shape data. In such cases, preservation techniques and formats that can separate out feature semantics into a separate layer on top of the model geometry would be particularly attractive.

2.6. Security

The CAD models created in commercial contexts often represent significant and valuable intellectual property, and as such access to them is tightly controlled. This helps to protect the company's competitive advantage but in the case of industries such as defence it can also help protect the eventual products from external interference. The most immediate effects of this culture of tight control are felt in matters of contemporaneous data exchange – between partners in collaborative projects, between suppliers and customers – but there are also implications for preservation. For example, it hinders the public exchange of experience and the identification of best and sub-optimal practice regarding preservation tools and techniques. With limited access to real-world data, the vendors of conversion utilities have reduced opportunities for thoroughly testing their products, especially in cases where they have to reverse-engineer the formats. As with other media that support digital rights management, if files have been password-protected, say, this makes it impossible to perform preservation actions on them unless the password is known to the archivists.

In some contexts security is in place not only to prevent unauthorized access to data but also to guarantee its authenticity. The prime example of this is in aerospace, but it can also be true for other industries and in academic contexts. It is not enough merely to preserve a CAD model: one must also be able to trace its provenance and prove that it has not been tampered with or accidentally altered in the meantime. This requires robust security not only for the model, but also for the information used to validate the model.

3. Standards

The problems caused by CAD models being tied to a particular version of software are well known, and from as early as the 1970s there have been attempts to overcome them through standardization. While the emphasis has always been on contemporary exchange rather than long-term preservation, the standards produced by these efforts are certainly worth considering as preservation formats.

3.1. IGES

As the name suggests, the Initial Graphics Exchange Specification (IGES) was one of the first attempts to create a vendor-neutral exchange format for CAD. It arose from a meeting of the Society of Manufacturing Engineers (SME) in late 1979, where industrial CAD customers, frustrated with their inability to transfer CAD data between tools and their internal systems, challenged a group of CAD vendors to develop a common exchange mechanism (Goldstein, Kemmerer and Parks, 1998). Circumstances were such at that meeting that the idea gained traction and a project was funded almost immediately to draw up an exchange specification and file format. The project was led by Roger Nagel of the (US) National Bureau of Standards (later the National Institute of Standards and Technology), with representatives and assistance from major CAD customers and vendors.

Version 1 of IGES was delivered in January 1980 (Nagel, Braithwaite and Kennicott, 1980) and submitted to the American National Standards Institute (ANSI) committee Y14.26 for consideration as a national standard. Version 2 was eventually accepted and published as ANSI standard Y14.26M-1981. It was revised and developed over the years: the last version was 5.3, which was published as ANS US PRO/IPO-100-1996 in 1996 and withdrawn in 2006, when the US Product Data Association (US PRO) closed down.

Even though IGES was successful in terms of widespread adoption and implementation, it had some serious shortcomings. IGES provides several different ways of doing things, and vendors were at liberty to implement only portions of the specification, meaning there was never any guarantee that two tools would support enough in common for data to be safely transferred between them. Furthermore, there was no facility for testing conformance, and so no way of ensuring a consistent implementation of the specification between vendors (Wilson, 1987).

3.2. STEP

STEP, the Standard for the Exchange of Product Model Data, holds the record for being the largest ISO standard (ISO 10303), with over 590 published parts and 18 technical corrigenda and other supplements. Its low-numbered parts are fundamental building blocks that are combined and applied practically by its high-numbered parts; for more information, see Appendix I.

The ISO subcommittee TC 184/SC 4 began work on STEP in 1984, motivated by the shortcomings of existing standards such as IGES and its European rivals SET (*Standard D'Echange et de Transfert*) and VDA-FS (*Verband der Automobilindustrie-Flächen-Schnittstelle*). A quite general problem among these standards was that they did not cover all the types of data that needed to be exchanged, and none could claim the level of international acceptance that was increasingly becoming necessary (Fowler, 1995).

One of the intentions for the standard was that it should be more rigorously defined than its predecessors, with all its data models clearly modelled and separated from concerns about file formats. There was to be an emphasis on testing and conformance, to ensure the robustness of the data exchange process. There was also to be a multi-layered approach to the standard's architecture: an application layer containing data models specific to an application or discipline, a logical layer containing cross-application data models, and a physical layer containing the file format specifications. Initially various modelling languages were used within the application layer, but when this started causing problems at the logical level, a new language called EXPRESS was devised to express all the data models in the standard.

A first draft of the standard was reviewed in 1989, but it was felt to be too unwieldy. It was therefore split into parts that could be maintained separately. An initial set of 12 parts was issued for committee review in 1992, and published in 1995. As the standard grew, it became apparent that the divergent authorship of the application protocols – data models for specific applications such as B-Rep mechanical design – was leading to the same semantics being represented in incompatible ways in different parts of the standard. To remedy this, the approach taken by the standard was modified in two ways. First, an additional layer of application interpreted constructs was introduced between the logical and application layers to handle semantics common to several applications; the first batch of these was published in 1999 and related to B-Rep modelling. Second, modular equivalents of the application protocols were introduced; these sets of modules began to be published in 2001.

The most widely known and widely implemented parts of STEP are AP 203, 'Configuration Controlled 3D Designs of Mechanical Parts and Assemblies', and Part 21, 'Clear Text Encoding of the Exchange Structure', which together define a CAD file format suitable for exchange and archiving known as an AP 203 STEP file (or STEP physical file). This is the format that is probably most widely used for exchanging CAD data in the engineering domain, with AP 214 ('Core Data for Automotive Mechanical Design Processes') files also used extensively in the automotive industry. When using these formats, it is important to note which editions are being used. Edition 1 of AP 203 does not support CSG-based solid modelling, geometric dimensions and tolerances (GD&T), or construction history modelling, for example, while Edition 2 does.

AP 203 and AP 214 have very similar scopes, so work is underway to merge them into a single part: AP 242 ('Managed Model-Based 3D Engineering'). As well as having the capabilities of its precursor parts, AP 242 will also support shape data quality information, semantic 3D product and manufacturing information (PMI), approximate geometry (for visualization) and access rights management. The last of these will of course require special consideration in a preservation context.

An initiative called the CAx Implementer Forum has been set up to assist vendors in testing their CAD conversion tools for compliance with STEP (CAx IF, 2012). The forum is run by the US consortium PDES and the German association ProSTEP iViP, both of which are heavily involved in the development and promotion of STEP as a standard. Currently 27 converters are being tested by 12 vendors, with a particular emphasis on AP 203, AP 209 ('Composite and Metallic Structural Analysis and Related Design') and AP 214.

As well as being hugely influential in the area for which it was intended, that is, contemporaneous exchange of product model data, STEP has also been used as the basis for standards in other areas such as Building Information Modelling and long-term archiving.

3.2.1. IFC and NBIMS-US

The IFC (Industry Foundation Classes) data model was developed by buildingSMART International, a federation of national or multinational regional alliances that are themselves made up of individuals, associations, research institutions, companies and agencies with an interest in the architecture, engineering and construction (AEC) industry.

The purpose of IFC is to express Building Information Models (BIMs) in a vendor-neutral way, so the information may be exchanged between different proprietary software applications. It has been published by ISO as a Publicly Available Specification (ISO/PAS 16739:2005) and is in the process of becoming a full International Standard. It uses the EXPRESS modelling language and STEP's Integrated Generic Resources for expressing geometric data, and is exchanged using either a STEP Part 21 file or using STEP's Standard Data Access Interface. In addition to its development work, buildingSMART International runs a certification scheme that tests compliance with the IFC model.

The North American chapter of buildingSMART International, buildingSMART alliance (a council of the National Institute of Building Sciences), has developed the complementary *National BIM Standard – United States* (2012), which specifies how BIMs, and by extension how standards like IFC, should be used. Other regional alliances, including buildingSMART UK and Ireland, are adapting the standard for use in their home nations.

3.2.2. LOTAR

LOTAR (Long Term Archiving and Retrieval) International is a project involving a consortium of aerospace and defence companies from the US and Europe (LOTAR International, 2012). Its purpose is to develop a multipart standard for archiving 3D CAD models and product data management information, the structure of which is explained in Appendix II.

The draft parts of the LOTAR standard undergo parallel validation at the European Committee for Standardization (CEN) and the (US) Aerospace Industries Association (AIA), for publication within the standards EN 9300 and NAS9300 respectively. The first NAS9300 parts were released in April 2012, while the first EN 9300 part ('Part 003: Fundamentals and concepts') was published in September 2012, followed by a further six in January 2013. The British Standards Institute is republishing the European versions as BS EN 9300.

The LOTAR standard draws heavily from existing standards and best practice such as the Open Archival Information System (OAIS) Reference Model (ISO 14721:2012), *Verband der Automobilindustrie* (VDA) Recommendation 4958 (VDA, 2005–2007), and STEP. Indeed, it may loosely be characterized as a standard for building an OAIS-style archive service using STEP-compliant processes and data, and insights from the development of LOTAR are being fed back into the development of STEP.

The LOTAR approach is pragmatic. Where necessary, several techniques are specified in order to accommodate different levels of available technology. For example, three versions of Part 120 ('Long Term Archiving and Retrieval of CAD 3D Explicit Geometry with Product and Manufacturing Information') are planned. In version 1, the PMI is included in a purely visual fashion within an AP 203 or AP 214 STEP file; cross-highlighting is used to reinforce the associations with the geometry. This should be widely achievable with current CAD systems. In version 2, STEP's semantic PMI constructs will be used, so that the information is both human- and machine-interpretable. Version 3 will support form features and parametric constructs as well as 'static' 3D geometry.

Recognizing the variable quality of format migration tools, the approach places heavy emphasis on checking data against quality and validation criteria. The quality criteria are used to ensure the data are likely to remain useful, while the validation criteria are used to ensure that the key characteristics of the data survive when opened in a new system. Examples of validation criteria include point clouds that trace the location of vertices, edges and surfaces; checksums for PMI data; and hash values generated from the product structure.

Though the LOTAR standard is still under development, it has already earned respect in the aerospace industry. As mentioned in Section 1.2, it was compliance with LOTAR that helped Dassault Aviation gain approval to submit 3D digital designs, rather than paper ones, to aviation authorities.

3.3. VRML and X3D

X3D is an XML-based 3D modelling format tuned to the needs of virtual worlds and in particular virtual reality applications, with an emphasis on usage on the Web. It is developed and maintained by the Web3D Consortium and published as a set of standards through ISO (ISO/IEC 19775; ISO/IEC 19776; ISO/IEC 19777). It is a successor to, and backward compatible with, Virtual Reality Markup Language (VRML), which is also an ISO standard (ISO/IEC 14772-1:1997). Open source libraries and viewers for working with X3D are available.

X3D has various facilities for representing 2D and 3D geometry, including NURBS surfaces, solid geometric figures and swept and extruded surfaces, but does not yet have full B-Rep modelling support. It also has facilities for representing CAD layers and assemblies (including the reuse of component parts), multiple levels of detail, interactivity and animation. It does not, however, support the inclusion of product and manufacturing information or the more advanced modelling techniques such as parametric or feature-based modelling.

3.4. U3D, PRC and 3D PDF

When Adobe Acrobat and Reader version 7.0 were released in 2005, they included capabilities for working with 3D graphics. The specification adopted for this purpose by Adobe was Universal 3D (U3D), a standard developed by the 3D Industry Forum and published and maintained by Ecma International (ECMA-363, 2007). This format only supported surface modelling with polygonal meshes, but did have a facility for incrementally adding and subtracting levels of detail and for reusing parts within an assembly. The fourth edition of the standard added support for NURBS surfaces.

Support for U3D was updated from the first to the third edition for the release of Acrobat and Reader version 8.1 in 2007, but more significantly support for another format was added: Product Representation Compact (PRC, pronounced *précis*). This is a highly compressed 3D modelling format developed by Trade and Technologies France (TTF); after Adobe acquired the company, the format was prepared for submission to ISO and is currently in the process of becoming an International Standard (ISO/DIS 14739-1). As well as supporting both mesh surface modelling and B-Rep solid modelling, one of the key attractions of the format is that it can also express product and manufacturing information. A PRC file can contain several file structures, simulating the separation of model data into multiple files. The format does not support construction history or parametric modelling, though.

Although Adobe still supports viewing embedded 3D models in its applications, it stopped developing conversion tools and 3D PDF generation tools in 2009; the latter functionality is now provided through third-party vendors such as Tech Soft 3D (Yares, 2012). The 3D capabilities of PDF are being considered for the forthcoming PDF/E-2 standard (ISO/CD 24517-2).

3.5. JT

JT was originally developed by Engineering Animation and Hewlett Packard, with further development by subsequent owners UGS and Siemens PLM Software. It was first published by ISO as a Publicly Available Specification, but is now a full International Standard (ISO 14306:2012). Like PRC, it is a compressed 3D modelling format supporting mesh surface modelling and B-Rep solid modelling. The model data can be contained in one file or split across several, and it can contain product and manufacturing information. It does not support construction history or parametric modelling, but unlike PRC, it does support multiple levels of detail and CSG solid modelling.

Being especially suited to the visualization of large assemblies, JT has been adopted fairly widely in the aerospace and automotive industries. ProSTEP iViP has taken an interest in the JT standard and is working to ensure it complements the forthcoming AP 242 of STEP. A cost-free viewer application, JT2Go, is available, and a C++ library for reading and writing JT data is available to members of the JT Open community.

3.6. AutoDesk formats

AutoCAD is one of the most popular CAD packages and hence its native file format, DWG, is considered to be a de facto standard in some quarters. The format is regularly modified, and the specifications are not publicly available. AutoDesk (via third-party vendor Tech Soft 3D) license an official library, RealDWG (AutoDesk, 2012a), for reading and writing the format, but due to the popularity of the DWG format alternative libraries have been reverse engineered. The Open Design Alliance, for example, provides a library called Teigha available only to its members (ODA, 2012), while the GNU Project has started work on an open source library called LibreDWG (GNU Project, 2010).

AutoDesk is also responsible for the DXF exchange format, for which recent specifications are freely available from the AutoDesk Web site (AutoDesk, 2012b). The purpose of the format is to provide the full information in AutoCAD models for exchange with other systems, without revealing (or relying on) the optimizations of the internal DWG format. There is, however, reported to be insufficient detail in some areas of the specifications from version R13 onwards to allow full implementation. Also, in common with IGES there are no defined levels of conformance, so there is a tendency for tools to implement partial support for the format and silently discard unrecognized data (Eiteljorg *et al*., 2011). Nevertheless, it remains a more transparent format than DWG, so in most cases would be more suitable for preservation purposes.

AutoDesk has also produced compressed lightweight formats, DWF and DWFx, for which the specifications are freely available, along with a C++ library and source code. These are intended for visualization (in such contexts as sales, marketing and review) rather than full data exchange. Users of other CAD systems may find their vendor supports a similarly published visualization format; for example, Dassault Systèmes, the company behind the CATIA CAD system, promotes its 3D XML file format in much the same vein as DWF and the simpler standards previously mentioned.

4. Techniques and Technologies

While the majority of standardization work has concentrated on exchange formats for CAD models, standards such as STEP, and more particularly LOTAR, have implications for the processes and techniques used to preserve these models. Outside the world of standards, there have been complementary efforts to find appropriate preservation techniques and to develop new tools that assist in performing them.

4.1. VDA Recommendation 4958

Verband der Automobilindustrie (VDA) is Germany's automotive industry association. Its Recommendation 4958 (VDA, 2005–2007) fleshes out and extends the OAIS Reference Model (ISO 14721:2012), applying it to industrial archives. Much of what it recommends is good records management practice – the certification assessment criteria in particular – but it does contain advice specifically related to CAD models.

It recommends that an archive dealing with industrial designs should define its policies in terms of three layers. The first is a Requirements Model, which enumerates all the information, rules and values which are needed to specify a product, either so that it may be manufactured or to prove the compliance of the manufactured product with the design; in other words, it specifies the significant properties of design documentation. The Requirements Model would typically be written in a formal language such as Unified Modelling Language (UML) or EXPRESS; a file or set of files intended to fulfil these requirements is known as a Core Model.

The second layer is Descriptive Standards. These are standards that are used to interpret the objects specified by the Requirements Model. So, for example, if an archive includes geometric dimensions and tolerances in its Requirements Model, it might specify in the second layer that such information should conform or be mapped to ISO 16792:2006.

The third layer is Implementation Standards and Models. These are the data models and formats used to express the objects and attributes specified by the Requirements Model. These models and formats should be vendor neutral and, ideally, be controlled by a standards body so that they may be understood in the long term. Policies at this level will specify which formats will be used to represent and store the design information, and the conventions used for mapping the semantics of the Requirements Model to the syntax and semantics of these formats.

The recommended workflow for archiving designs is as follows. Someone on the design side prepares a Submission Information Package (SIP) containing the native CAD model alongside relevant metadata as described by the OAIS model. Amid the Fixity Information should be a set of Validation Properties that can be used to check the success of future format migrations. For example, the weight, surface area, volume, or centre of gravity of solids in the design could be calculated, or (as later proposed by LOTAR International) a point cloud generated that indicates where vertices, edges and surfaces occur in the design (Barker, 2010).

The SIP is submitted to the archive, which checks that the package is complete and uses acceptable formats. If the SIP is suitable, it is converted to an Archival Information Package (AIP) as follows. The CAD file is converted into a Core Model as defined by the above policies, possibly supplemented by additional models (e.g. the native CAD file[s], an extended vendor-neutral model containing feature information or design construction history, a lightweight visualization). Any format conversions are checked by regenerating the Validation Properties from the new files and comparing them to those recorded in the SIP. If it is not possible to regenerate the Validation Properties using the new format, additional files should be created by converting the new files back into the old format, and the Validation Properties generated from these 'round-trip' files; otherwise, a bespoke validation technique should be devised. The accompanying metadata is also normalized and expanded. Care should be taken to ensure that any linkages between geometric and non-geometric information remain intact.

For large models it is recommended that each part (or leaf node in the product tree) is archived in its own AIP, and assemblies are archived as AIPs that virtually (rather than directly) include the part-level AIPs. The recommendation calls this 'incremental archiving'.

When a model is retrieved from the archive, a Dissemination Information Package (DIP) is produced. Depending on the need, the DIP may just present the visualization file from the AIP within a viewer application, or it may include a native CAD model generated from the Core Model. Again, any new conversion is verified using the Validation Properties.

4.2. The KIM Project

The KIM (Knowledge and Information Management) Project was a 'Grand Challenge' project which ran from October 2005 to March 2009, jointly funded by the Engineering and Physical Sciences Research Council (EPSRC) and Economic and Social Research Council (ESRC) (KIM Project, 2009). One of the issues tackled by the project was that of creating and curating advanced product information representations, and within that were two strands of research specifically focused on CAD models: LiMMA and RRoRIfE.

RRoRIfE (Registry/Repository of Representation Information for Engineering) was developed as a proof-of-concept preservation planning tool for CAD formats (Ball, 2011). It uses information about the semantic structures within CAD formats, and the capabilities of authoring and conversion software for translating between them, to calculate possible migration pathways. These calculations take account of whether certain types of information should be preserved or removed in the process. RRoRIfE can also be used to predict the data loss that might occur in a given format migration.

In parallel with this, the project developed a complementary set of tools implementing the LiMMA (Lightweight Models with Multilayer Annotations) system (Ding, Ball, Matthews *et al.*, 2009; Ding, Ball, Patel *et al*., 2011). This is an approach for preserving (some of) the information that is lost when a native CAD model is migrated to a lightweight format such as X3D, through the use of additional annotation documents. Each of these documents collects together a set of annotations relevant to a particular audience (defined in terms of domain specialization, access rights, etc.), and links them to the model geometry by referring to named entities (e.g. parts, surfaces), particular co-ordinates or both. Each annotation document can be applied to the model in any of its formats (software permitting), and any particular model file may be supplemented with any number of annotation documents.

From a preservation perspective, the idea is that if the migration of a model from a legacy native CAD format to the current CAD format proves unsatisfactory, the annotations may be layered over the migrated copy and used to reconstruct the original model. The emphasis on lightweight visualization formats is to ensure that at the very least the geometry of the model will survive in a form that may be simply and reliably imported into a future system. An additional benefit of the

LiMMA approach is that, for example, if maintenance crews annotate lightweight visualizations of the design with issues they have encountered in service, these annotations can be layered over the native CAD version should the design ever be revisited, and the issues taken into account.

The prototypical tools developed by the KIM Project to demonstrate LiMMA included a plug-in for the CAD system NX3, JavaScript code for use in 3D PDFs, and a custom Java-based X3D viewer application.

4.3. The FACADE Project

The FACADE (Future-proofing Architectural Computer-Aided DEsign) Project was a collaboration between MIT Libraries and the MIT School of Architecture and Planning (Smith, 2009). It ran from October 2006 for three years with funding from the US Institute of Museum and Library Services. The purpose of the project was to set up a technical infrastructure and procedure for archiving corpora of digital architectural design documentation, especially two- and three-dimensional CAD models.

The corpora used by the project as a test bed included various projects from the MIT Department of Architecture, plus three sets of designs from professional architects external to the university. Each corpus contained not only CAD models but also associated presentations, reports and correspondence.

A workflow was devised by FACADE and tested with MIT Libraries staff. It can be summarized as follows.

1. Upon receiving the files on hard drive, staff make a copy and run automated metadata capture tools over it.
2. A digital preservation specialist reviews the detected file formats to determine, for instance, if any will prove particularly challenging to preserve.
3. An architecture specialist tags groups of files using the facets and terms of FACADE's Project Information Model (PIM).
4. The architecture specialist then selects the most important files (i.e. the ones likely to satisfy 80% of users' needs) for active curation.
5. A CAD specialist creates derivative formats for the selected models: a 3D PDF file for display, an IFC or STEP file for archiving the full model information, and an IGES file for archiving just the simple geometry. The original models are also kept for archiving.
6. The CAD specialist creates detailed metadata for the selected models to place them properly within the context of the other selected files.
7. The architecture specialist performs quality assurance on the metadata.
8. The digital preservation specialist updates the local file format registry.
9. Repository staff load the corpus into the archive, and then set up special user interfaces for accessing the corpus, using metadata copied across from the archive.
10. Finally, the architecture specialist performs quality assurance on the user interfaces.

Various tools were developed or enhanced to support this workflow. One was the Curator's Workbench: a tool for creating metadata for files en masse and individually, designating some as 'selected', and for managing the bulk import of items into a DSpace repository (approximately steps 1 to 9 above). On the DSpace end, FACADE developed a custom ingest tool for importing large numbers (of the order of tens of thousands) of files as DSpace 'items' related together in a 'package', as well as providing some improvements to the import facilities for 2D and 3D PDFs, ingest validation checks, and file format recognition and validation (step 9). An additional tool was developed for exporting FACADE's PIM-related metadata from DSpace for use in the external user interface (steps 9 and 10).

The external interface itself was based on work from the SIMILE Project, which developed tools for visualizing Semantic Web data (MIT, 2009). It has three layers: a simple catalogue of building projects for which documents are held; within each project, a simple catalogue of the 'selected' objects; and within each project, a somewhat more powerful and scalable catalogue of the complete collection of objects. These interfaces are linked so that, for example, full text searches in the extended catalogue interface can be triggered from a search box in the simple catalogue.

As noted in step 5, FACADE recommends keeping copies of CAD models in four formats: the original, a heavyweight standard format, a lightweight standard format and a visualization format. The rationale it gives for this is as follows. The original format provides the fullest amount of information about the design, but is only usable so long as the original software is actually and legally available (in the short term in most cases). The purpose of the standard formats is to preserve the design information in a vendor-neutral, portable way. A heavyweight standard is used to preserve the most information possible, accepting the risk that some information may be poorly converted, leading to an inauthentic expression of the design. A lightweight standard is used to preserve a restricted subset of the information (specifically shape data), in the expectation that this subset will be translated robustly and could therefore be used as a fall-back option should the information encoded using the heavyweight standard prove unreliable. The visualization format was chosen to allow convenient display of the model in-browser, using software that is near-ubiquitous among users of the archive.

4.4. Sustaining Engineering Informatics

Lubell, Rachuri *et al*. (2008) propose a framework for archiving engineering-related digital objects such as product models. One of their insights is that the use cases for such objects typically belong to one of three levels. The least demanding use cases involve *Reference*, that is, visualizing, reproducing and exploring records in a read-only fashion. Beyond that are use cases involving *Reuse*, that is, using and modifying the design using an appropriate system. The most demanding use cases involve understanding the *Rationale* behind the decisions underlying the object. These levels are known as the 3Rs, with each level including the previous one.

The proposal involves analysing how the needs of end users might be met for each of the 3Rs, and working backwards from that to determine which dissemination formats would be most appropriate, how objects should be packaged and formatted in the archive, and what metadata needs to be collected at ingest. For determining appropriate formats, the authors suggest using sustainability factors such as those used by the Library of Congress (Arms, Fleischhauer and Jones, 2011); for example, disclosure (availability of specifications), adoption (level of use by practitioners), and self-documentation (embedded metadata).

4.5. The SHAMAN Project

SHAMAN (Sustaining Heritage Access Through Multivalent Archiving) was a Large Integrated Project co-funded by the European Union as part of its Seventh Framework Programme (SHAMAN 2012). It launched in December 2007 and ran for four years.

The over-arching aim of the project was to develop and test a complete digital preservation framework, capable of supporting a wide range of different content types. It targeted three specific areas: memory institutions (museums, libraries, archives), design and engineering, and e-Science. The work on design and engineering had two foci: one on combatting semantic drift in PLM data, and one on promoting long-term preservation by reducing reliance on a single system.

Within the first research strand for design and engineering, the project envisioned a PLM system handing off data to an archival function at the point at which the design is released for production (Brunsmann and Wilkes, 2009; Maceviciute *et al*., 2011). The archival function would as far as possible normalize the PLM information to standard formats such STEP, and then do either or both of the following:

1. provide a mapping of semantics present in the files (by virtue of their formatting) to a standard external ontology;

2. apply annotations that fill in the engineering knowledge that is either tacit or incompletely expressed in the files (e.g. that a part represents a specific item from a parts catalogue, and what its key characteristics are), using a standard external ontology;

Over time, it is expected that the external ontologies will be modified or replaced. The archival function keeps track of such changes so that mappings may be calculated between the version referenced originally and the current version. These mappings can then be used to translate the format-specific semantic mappings and annotations to use the new ontology, so they may be understood by contemporary users and systems.

This vision places quite some store on the use of standard external ontologies, so that first, any changes are well documented, and second, there are economies of scale in maintaining a robust set of mappings between different versions.

Within the second research strand, the project considered co-design as a driver for creating more preservable models (Jacquin, 2011a). Co-design in this sense is where designers from multiple domains or from different partners collaborate on a design. Examples might be where a mechanical engineer designs an automated teller machine (ATM) with the cabling and electrical harness supplied by an electrical engineer, or where an aircraft manufacturer and a jet engine manufacturer collaborate on an aircraft but do not exchange full design details with one another.

The usual paradigm is for a team of designers all to contribute to a single authoritative master design, which is signed off and then passed to the manufacturing engineers for conversion into CNC programmes and the like. The proposed paradigm takes a more distributed approach: designers and manufacturing engineers work in parallel, using their familiar tools and contribute versions of their designs, simplified according to an agreed set of rules, to a shared visualization space. Despite the fact that the visualization space contains only simplified designs, it is considered the 'master' as it is the place where the various aspects of the total design come together and are harmonized. The advantage of this approach is that it makes cross-domain iterative optimization easier and more transparent to all parties, while allowing more flexibility over how much information is actually shared.

The rules for constructing the visualization space must necessarily be bespoke to each project, but will commonly specify which objects will be shared, what the semantics of the objects are, and what format should be used for the common visualization. From a project management perspective, there also need to be rules for establishing inter-dependencies between objects from different domains, and where responsibilities lie for resolving any conflicts. SHAMAN developed a co-design meta-model, ISP2, for expressing such rules, as well as an Eclipse-based demonstrator for managing a visualization space and a worked example of mechanical and electrical CAD co-design (Jacquin, 2011b).

From a preservation perspective, this way of working has several advantages. First, the meta-model that aligns the models from the different domains provides useful information about what the objects in the models represent. (Incidentally, this is the kind of semantic annotation dealt with in the other strand of SHAMAN research.) Second, the method ensures that as well as the native CAD representations of (parts of) the design, there is always a unified representation of it in a visualization format that, being simplified and of necessity read/writeable from all the native CAD systems in use, is easier to preserve. Third, the method involves contributors creating snapshots of their own design work and updating the visualization space with the changes; this means that the history of the design is recorded. While this record is not as powerful from a reuse perspective as the design history modelling features of CAD tools, it has the advantage of linking design changes to logs of conflict resolution negotiations (reasons for, reasons against, decisions) thereby providing an enhanced record of rationale.

4.6. Archaeological metadata

Eiteljorg *et al*. (2011) suggest a set of metadata that archaeological researchers should provide in order for their CAD models to be properly preserved. The elements of the set may be summarized as follows:

- *Project information*: this is more to do with discovery than preservation, strictly speaking, and applies equally to non-CAD data; see Eiteljorg *et al*. (2011, Section 4.2).
- *File-naming convention*: a description of the convention used; and a list of file extensions and the exact file formats (including version numbers) to which they refer.
- *Data collection documentation*: field data capture metadata can be used to determine whether the data is suitable for use in generating a CAD model; and data sources metadata is used to track the intellectual property rights adhering to a CAD model; see Eiteljorg *et al*. (2011, Sections 4.4–4.5).
- *Layer-naming convention*: the name of the layer-naming convention; and for each layer, its name, content, and drawing conventions (special symbols, fonts and colours, etc.).
- *Other CAD model information*: name and address of agent responsible for creating the CAD model; name of the model; name and version of the CAD software used; files used in the model; locations of files cross-referenced in the model.
- *Linked database information*: name and version of database software used; name of database; data field names and value codes (with definitions); linked CAD models and the field recording the links; file format; creation date.

5. Discussion

The preservation of CAD models is not something that is amenable to a one-size-fits-all solution. Different communities demand different things from their legacy CAD models, and in any case what is technically possible varies enormously depending on circumstance. Nevertheless, the research and standardization work that has been conducted in this area has produced several approaches that might be adopted individually or in combination.

5.1. Preservation of the original CAD model

It is generally the case that the format in which a CAD model was constructed contains more information about it than will ever be extracted by export filters or conversion tools. If forensic-level investigations into a design are required (to uncover, for example, how a particular quirk was introduced), or if reuse relies on the more advanced features of the software, then the original model is invaluable. Given the ephemeral nature of CAD systems and the tight dependency of models on those systems, though, mere bit preservation will almost certainly lead to the information becoming inaccessible.

While one might hope that vendors would release format specifications once they find no competitive advantage in keeping them secret, history to date suggests this is unlikely. There has been some success in reverse engineering certain formats, but given the complexities involved – in particular around interactions with the modelling kernel and embedded feature semantics – it is only for the most popular formats (e.g. AutoDesk's DWG) that archives are at all likely to find converters they can afford with the level of reliability they need.

Maintaining access to the original CAD model therefore implies preserving the original software through emulation methods. It should be borne in mind, though, that the software licence may prevent it being used in perpetuity, so this would need to be checked. Also, the utility of the software is likely to diminish over time, at least for the purposes of reuse and uncovering rationale. Partly this is because proficiency in the software will decline as fewer people active in the community remember it, and partly because the software will be able to communicate information to fewer and fewer currently active systems.

5.2. Rolling format migrations

As a CAD format becomes obsolete, a possible tactic might be to migrate it to a more modern format. For an academic archive, the obvious choice would be the official successor format version, if it exists. For an industrial archive, it would normally be whatever the wider firm has switched to using; indeed, format migration of recent models may be one of the services the incoming CAD vendor is contracted to perform.

The attraction of this approach is that the archive always contains a copy of the model in a format that can be used with current technology. Also, by keeping the migration between highly similar formats, it may be hoped that any data loss or corruption will consequently be small.

On the other hand, each migration has a cost associated with it – the possible acquisition of new conversion software, the time taken to perform and validate the migration, storage of multiple versions – and introduces the possibility of compounding the data loss or corruption of previous migrations. In industry, where there are strict legal and regulatory reasons for demonstrating the faithfulness of any migration, the costs are particularly high, and hence there is a strong incentive to keep any such operations to a minimum. For these reasons, direct migration to a new native CAD format usually only makes sense for models in active use; it is not generally considered as an ongoing strategy for maintaining entire archives.

5.3. Normalization to standard formats

The most widely recommended approach for preserving CAD models is to perform a migration on (or prior to) ingest to one or more vendor-neutral, standard file formats. The advantage of these formats is that they are stable, well documented, and do not depend for their interpretation on a particular modelling kernel or unknown set of feature semantics.

There are a variety of different formats that can be chosen, each with their own set of advantages and disadvantages. In theory, the maximum amount of information may be preserved using the STEP standard and its relations like IFC, and these should certainly be considered first and foremost. They should not be used blindly, though: STEP itself is as comprehensive as standards come, but CAD vendors are slow to introduce all its features in their own import and export filters, and may stick with older parts of the standard (such as AP 203 Edition 1) even when more expressive formats have been released. Moreover, CAD vendors do not always implement STEP formats as fully or as robustly as they might, though initiatives such as the CAx Implementer Forum are addressing this issue.

Beyond STEP there exists a whole family of formats that are more basic, but by the same token are more easily supported by software. These lightweight formats are particularly useful for preserving (just) the shape data, with some tuned to visualization (e.g. X3D, U3D) and others capable of supporting reuse (PRC, JT). IGES is probably the most widely supported lightweight format, though it is more prone to data loss than the others due to that support being incomplete in most cases. Non-standard, vendor-specific lightweight formats should be avoided, as they are unlikely to enjoy widespread support and their long-term prospects are unknown.

Given the variable support for vendor-neutral file formats among CAD systems, it would be sensible to normalize to at least two: one of them should be a STEP file of some sort, and another in a lightweight format. This provides a backup option should one of the files fail to import properly.

5.4. Validation

It is well known that fixity information is a vital component of preservation activity, permitting checks to be made that digital objects have survived long periods of storage. Checksums are useful for detecting corruption in the bitstream, but other techniques are needed to ensure that the significant properties of a CAD model have survived a format migration, or interpretation by a different CAD system. Very few errors of this nature are obvious from inspection, so formal validation properties are needed.

Typical validation properties for solid models include the volume, centre of gravity and calculated weight of each solid in the model. For solid and surface models, surface areas can be used. Another versatile technique is the use of a point cloud: this is where a large set of co-ordinates is calculated such that each co-ordinate lies on a surface in the model. The distribution of these points should not be random: they can be sparse across flat surfaces, but need to be denser where surfaces curve more steeply, and particularly dense along edges and corners.

Such properties should be calculated in the native CAD system and recorded. Then, when the model is loaded by a different system, the properties can be recalculated by the new system and checked against the original set. For point clouds, the procedure is a little different, as the test is whether all the points still lie along the surfaces and edges of the newly rendered model, and that no new surfaces or discontinuities have appeared. It should be noted that the point cloud technique is harder to apply where models have been exported to an approximate visualization format (that is, using triangular meshes instead of exact geometry); doing so involves calculating whether the points in the cloud lie within the tolerance of the approximation algorithm.

The validation checks should be performed immediately after each format conversion so that any errors can be addressed while the relevant software is still available. In the case of vendor-neutral file formats, it is best to check them in at least two different CAD systems or viewers, in case the import filters are at fault rather than the file itself.

5.5. Multi-file CAD models

VDA Recommendation 4958 includes a guideline that large models should be archived at the part level, and these part-level packages included by reference in higher-level packages representing sub-assemblies and assemblies. This aids efficiency where the same part occurs in multiple places within and across designs, and makes each individual package simpler to preserve. On the other hand, if one is normalizing to simple standard formats, not all of them support models represented by several files. In such cases it would involve a significant amount of work to recombine the various parts into a single assembly.

Where a CAD file natively breaks down models into part and assembly files, it makes sense to treat them as per the VDA Recommendation. The assembly files should be checked at ingest to ensure that the cross-references use indirect methods (such as part names) or relative path names, where possible. In the former case, supporting documentation (such as an inventory of part and file names) should be provided, and in the latter case instructions should be recorded for reproducing the required directory and file structure. To avoid having to reconstruct an assembly when using lightweight formats, the lightweight version of the assembly should be generated directly from the native CAD model and associated with the top-level package. If space permits, similar versions could be generated for the intermediate sub-assemblies.

5.6. Links and Annotations

Where links exist between a CAD model and other resources, an archive should be careful to maintain these links where possible. If the links manifest as filesystem references, these should be made to use indirect methods or relative pathnames where possible, with appropriate documentation. Where links refer to internal identifiers or markers within a CAD model, the identifiers should be included in the list of validation properties to check when migrating the model to another format.

As with any digital object, information resources that help to explain the CAD model should archived alongside it as metadata. Typically, such resources might include specification documents, process and rationale models or reports, file naming conventions, layer naming conventions, drawing conventions, data collection documentation, materials data sheets or parts catalogues.

Ideally, information that is known to be lost when converting from a native CAD format to a standard format should be recorded in a separate annotation document that refers to the model through entity identifiers, points/point clouds, or both. Similarly, annotations should be provided that link embedded semantics within the model to semantics of standard vocabularies. In practice, at least for the foreseeable future, this is unlikely to be possible due to a dearth of appropriate tools and vocabularies, though on the latter point there are hopeful signs. CWA 16200:2010, for example, translates the concepts and vocabulary of an ISO standard for materials test data (ISO 6892-1:2009) into linked data, paving the way for similar translations of other standards such as the various STEP parts.

5.7. Style

One way to simplify the task of preserving CAD models is to restrict the way in which they are constructed to a certain 'house style'. If models are known to conform to this style on ingest, then in theory it is easier to correct any deviations from the style that later appear. Furthermore, the style can act as a way of recording and interpreting the semantics of the model that is independent of the (potentially fragile) semantic capabilities of the CAD format that was used.

House styles of various kinds are already common in organizations that deal with CAD models on a regular basis. Large engineering firms have design house rules, for example. On the client side, Harvard University has developed a standard set of conditions it imposes on the architectural CAD models received by its Planning Office (Harvard University, 2009).

The usual motivation for these house styles is to ensure the models are legible, intelligible and navigable. Some aspects of a model are considered so important they are the subject of national or international standards: for example, ISO 16792:2006 for geometric dimensions and tolerances, or ISO 13567 for organizing and naming CAD layers. In the US, the *United States National CAD Standard* (2011) is gaining traction as a house style for the architecture, engineering and construction (AEC) industry and its clients; its explicit purpose is improve communication between client, contractor and supply chain.

The technical side of preservation seldom factors into such house styles, partly because the problems are too intricate to solve with a few simple rules. If an archive routinely deals with only one CAD format, as it might in an engineering or architectural firm, and finds that certain constructions translate better to its preservation formats than others, there may be scope for having this discovery reflected in the house style. Nevertheless, the clarity afforded by a house style is still a benefit for preservation, so if archives are in any position to affect how CAD models are constructed they should encourage the adoption of one. When it comes to archiving a model, documentation of the house style should also be archived and referenced in the model's metadata.

5.8. Advocacy

The issue of preservation is seldom given a great deal of attention by CAD vendors or their customers, so there is little motivation to change the current state of affairs for those in the best position to do so. For any significant improvement in our ability to preserve CAD, the wider preservation community should consider an advocacy programme to raise awareness of the importance of standard formats and high-quality format migration among CAD customers. Any such programme should emphasize the business benefits of a reliable and usable archive of CAD models and the efficiency savings afforded by systems interoperating through common information and data formats. There may also be a role for the community to lobby for legislation that prohibits the use of undocumented CAD formats for exchange on the grounds that they are a trade barrier, in order to promote support for standard formats among CAD vendors.

6. Conclusions and Recommended Actions

- Archives should first establish why they are preserving CAD models, and then which aspects of the CAD models should be targeted for preservation in order to accomplish that aim. If the preserved CAD models will only be held for visualization (Reference in 3Rs terminology), the preservation effort should be concentrated on lightweight, standard formats rather than the native format. If the preserved CAD models might be used to reconstruct how a particular product was designed (Rationale in 3Rs terminology), and software licences allow, the archive should perhaps invest in a secure emulated platform for accessing the native model.
- Archives should determine a set of validation properties they will use to test the integrity of preserved CAD models. These should, between them, cover all the aspects of the model targeted for preservation. Such properties might include internal identifiers, properties of solids and surfaces within the model, or a point cloud tracing the model geometry. For generalist archives offering a low level of preservation, a series of model screenshots – accompanied by a description of viewing angles and zoom factors used, and any elements removed from the view – might be sufficient.
- The validation properties should be recorded for each model ingested into the archive, using the native CAD system in which it was created. If an archive does not have access to the native CAD system, instructions should be provided to depositors so they can record the validation properties prior to submission.
- As native CAD models contain much information that cannot be migrated to another format, they should be kept for as long as there is software available to read them accurately. In some circumstances there may be reasons to keep them longer (e.g. legal compliance).
- CAD models should be normalized to at least one, and ideally two or three, vendor-neutral, standard formats. STEP or IFC formats are ideal. If an archive does not have access to conversion software, it may have to make such conversions a condition of ingest, and provide instructions to depositors on how to export files in the correct format from their CAD system.
- Validation properties should be recorded for each converted model, using whatever software is available, and compared to those recorded from the native CAD system. If discrepancies are found, first these discrepancies should be checked using alternative software, and if confirmed, efforts should be made to resolve them (e.g. using different conversion settings, trying a different converter, trying a different standard format).
- Complex multi-file CAD models should have their part and assembly files archived as a hierarchy of linked packages rather than as one large package. Links embedded in the files should be converted to indirect references or relative paths if possible, and the information needed to implement the links (such as the implied file/directory structure) recorded within the metadata for the package representing the (top-level) assembly.
- Archives should work with depositors to ensure that all the documentation and resources needed to understand or otherwise support the CAD model are archived alongside it. This

might include specification documents, process and rationale models or reports, file naming conventions, layer naming conventions, drawing conventions, materials data sheets, parts catalogues, or supplementary databases. Links between files should be made indirect or relative as outlined above.

- Archives should encourage the use of a documented house style for CAD models wherever possible, especially where the style makes the CAD models easier to preserve on a technical level.
- The wider preservation community should build a business case that underlines the importance of interoperability and preservation for CAD customers and vendors, and use this to campaign for better support for standard formats in CAD systems among customers, vendors and legislative bodies.

7. Glossary

AEC: Architecture, Engineering and Construction The industry sector dealing with the design and construction of the built environment.

B-Rep: Boundary Representation A method of solid modelling where the solids are defined in terms of their boundaries (surfaces).

BOM: Bill of Materials A list of materials needed to construct a product. A BOM is usually highly structured, breaking down the full product into its constituent sub-assemblies, parts and raw materials. It may also include more abstract items that affect budgeting and process planning, such as labour and finishing processes. A typical line in a BOM might contain a part number, description, quantity, cost, wastage (if a raw material), and supplier.

BIM: Building Information Model A digital representation of the physical and functional characteristics of a built environment, used as a through-life co-ordination tool and decision-making resource.

CATIA A CAD system developed by Dassault Systèmes.

CAx Shorthand term encompassing the various industrial activities that are 'Computer-Aided': Computer-Aided Design, Computer-Aided Manufacturing, Computer-Aided Process Planning, etc.

CNC: Computer Numerical Control The method by which computers control robotic tools and machinery in order to automate the manufacturing process.

CSG: Constructive Solid Geometry A method of solid modelling where the solids are constructed from a set of primitive geometric shapes – spheres, cylinders, cones, etc. – by processes such as deformation, union/addition, difference/subtraction and intersection.

Ecma International An industry association that acts as a standards body in the areas of information and communication technology and consumer electronics. Among its high profile standards are ECMAScript (of which JavaScript is a dialect), the Near Field Communication Interface and Protocol (NFCIP) standards, and Office Open XML.

Exact geometry Shape data that exactly replicates the curved lines and surfaces of the intended model. The term is often used to contrast with 'approximate geometry', in which such curves are approximated using a series of straight lines and flat surfaces (usually tessellating triangles).

Exchange format A format that has been designed to be read and written by several different software applications with a minimum of loss. Exchange formats can be vendor neutral (as with IGES and STEP AP 203) or tied to a popular software product, though in the latter case they are typically different from the software's native format. AutoDesk, for example, maintains an exchange format called DXF, which is related to but distinct from DWG, the native file format of its AutoCAD product.

FEA: Finite Element Analysis Analytical technique for predicting the behaviour of a design under operational conditions, for example how a modelled bridge support might deform under the weight of a bridge and its traffic.

Feature A feature in the modelling sense is a generic characteristic or shape with a certain significance, with implications for its relationship with other features and various other parametric constraints. Examples might include a curved blend between two surfaces (which will affect how the boundary behaves under stress) or a keyway (which will need to accommodate a matching key).

GD&T: Geometric Dimensions and Tolerances Also known as geometric dimensioning and tolerancing, this refers to annotations on a drawing or model that explicitly specify the dimensions, locations and orientations of parts, features and spaces, and the tolerances (i.e. limits to acceptable variation) of the same. The form such annotations should take is the subject of various national and international standards.

GIS: Geographic Information System A software environment for creating, editing, visualizing and analysing geospatial data, that is, data points that each relate to a particular location.

IFC: Industry Foundation Classes A data model for representing Building Information Models (which see) in a vendor-neutral way for the purposes of exchange.

IGES: Initial Graphics Exchange Specification A vendor-neutral exchange format for CAD data, notable for being widely but inconsistently supported among CAD systems.

ISO: International Organization for Standardization ISO describes itself as the world's largest developer of voluntary International Standards. The technical work of ISO is done by technical committees and their subcommittees; STEP (ISO 10303), for example, is developed by Subcommittee 4 (Industrial data) of Technical Committee 184 (Automation systems and integration), or TC 184/SC 4 for short. Some technical committees are jointly convened with other bodies; for example, standards marked as ISO/IEC, such as ISO/IEC 19775, are overseen and published jointly with the International Electrotechnical Commission. There are several different routes by which a standard may be drafted and approved:

- A proposal is made to a technical committee or subcommittee (TC/SC), and if accepted a working group prepares a *Working Draft* (WD). Once the working group is satisfied, it is registered as a *Committee Draft* (CD) and considered by the TC/SC; ISO/CD 24517-2 is an example of a standard at this stage. Once the TC/SC has reached consensus on the technical content, it is registered as a *Draft International Standard* (DIS); ISO/DIS 14739-1 is an example of a standard at this stage. The DIS is considered by all ISO member bodies in two rounds of voting, the first to promote it to a *Final Draft International Standard* (FDIS), the second to a full *International Standard*.
- A working group, or a registered standards body or consortium outside ISO, may submit a document for publication by a TC/SC as a *Publicly Available Specification* (PAS). ISO/PAS 16739:2005 is an example of a PAS. The approval process for a PAS is not as strict as for a Draft International Standard, so it has roughly the same weight as a Committee Draft, signifying consensus within the working group. A document can be published as a PAS for up to six years, after which it must either be withdrawn or converted to an International Standard.
- ISO/IEC Joint Technical Committee 1 has a 'Fast Track' procedure whereby an existing or draft standard written outside ISO/IEC may be converted to a DIS (or *Committee Draft for Vote* in IEC terminology) for consideration by all member bodies. X3D became an ISO standard through this route.

Kernel A geometric modelling kernel is the part of a CAD system that handles the geometric calculations, for example whether a point is inside, on, or outside a particular solid, and how a 3D solid should look on a 2D viewport.

LOTAR International An industrial consortium of aerospace and defence companies from the US and Europe. It is developing the LOTAR (LOng Term Archiving and Retrieval) standard for 3D CAD models and product data management information.

Native format The file format that a piece of software uses by default for reading from and writing to disk is *native* to that piece of software. Software can typically read and write several formats; native formats map losslessly to the software's internal data model, while non-native formats typically do not. Contrast with 'vendor-neutral format' (which see).

NURBS: Non-Uniform Rational Basis Spline A mathematical construct for representing arbitrary curved lines and surfaces. Its importance lies in its ability to express exactly both Bézier curves and conic sections. A *basis spline* or *B-spline* is a curve or surface defined jointly by a sequence of control points (which typically lie off the curve or surface) and a sequence of knots (which lie on the curve or surface). *Rational* means that the control points can have different weights, and *non-uniform* means the knots do not have to be equally spaced (Versprille, 1975).

Parametric modelling A modelling technique whereby aspects of design are given a variable value instead of a fixed one, in order to make them easier to adjust and reuse in different contexts. Constraints are used to control how designs should be adjusted in the light of changed variables. For example, if the radius of a gear wheel increases, the CAD system might increase the number of teeth the gear has so their size and spacing remain within pre-defined limits.

Point cloud A set of co-ordinates within a 3D space, so called because when rendered it appears as a cloud of points. Co-ordinate measuring machines output point clouds when scanning real-world objects.

PMI: Product and Manufacturing Information In the widest sense, this refers to the additional information needed to manufacture a part from the shape data present in a 2D drawing or 3D CAD model. At a minimum, it includes geometric dimensions and tolerances (which see) but may include other annotations, and specifications of finishes and materials.

Publicly Available Specification See ISO.

Shape data The points, lines, surfaces and solid objects making up the geometric information in a CAD model, but not the product and manufacturing information (which see), parametric relationships/properties, feature semantics or construction history.

STEP: Standard for the Exchange of Product Model Data STEP is an informal name for ISO 10303, which defines vendor-neutral methods for representing and exchanging a wide range of product data, including but not limited to CAD data. The standard consists of many different parts; the way they are organized is explained in Appendix I.

Vendor-neutral format A vendor-neutral CAD format is one which is not controlled by a single CAD vendor, but rather by a standards body or some other form of consortium or committee. If there is a reference implementation of the format, that is, a system guaranteed to read and write the format correctly, it is either a library or a toy system rather than a serious CAD system in its own right.

X3D: Extensible 3D A lightweight modelling format tailored for creating virtual worlds and animations, especially where users will be interacting with them over the Web. It is published as a series of ISO standards but is developed and maintained by the Web3D Consortium.

8. Further Reading

For an introduction to the generalities of digital preservation, readers are invited to consult the Digital Curation Centre's series of *Briefing Papers* (DCC, NDa), and for more practical advice, its series of *How-to Guides* (DCC, NDb). Subject-based data centres also provide advice to researchers on making their data easier to preserve: see for example the Archaeological Data Service's *Guides to Good Practice* (ADS and Digital Antiquity, 2009) and the UK Data Archive's guidance on creating and managing research data (UKDA, 2013).

For more information about the history of CAD, Bozdoc (2003) provides a chronology of CAD between the 1950s and 2000, noting for example when important research took place, when notable organizations were formed and when particular CAD systems were released. CADAZZ (2004) provides an alternative account that is less comprehensive in terms of dates but provides more narrative commentary. Weisberg (2008) provides a thematic account of the history of CAD, focusing on the personalities and the vendors involved.

For more detailed guidance on preserving CAD models in the context of archaeological research, the (UK) Archaeology Data Service in collaboration with the (US) Center for the Study of Architecture has published a *Guide to Good Practice* on the subject of CAD (Eiteljorg *et al*., 2011).

Although a summary of its key points is provided in this report, if one is interested in preserving CAD designs in an industrial context, it is worth reading VDA Recommendation 4958 in full (VDA, 2005–2007). As well as specific advice on CAD models, it also provides practical guidance on running an industrial archive, and is freely available in both English and the original German. The recommendation is largely harmonious with the LOTAR standard (EN 9300; NAS9300), which should be considered for high quality industrial archives.

For an appreciation of how CAD preservation fits into the wider context of knowledge management issues in engineering, see the *Proceedings of the Atlantic Workshop on Long Term Knowledge Retention 2007* (Ball and Ding, Eds, 2007) and the report on the subsequent workshop *Long Term Sustainment of Digital Information for Science and Engineering* (Lubell, Mani *et al*., 2008).

The scope of this report was heavily influenced by the DPC workshop *Designed to Last* (DPC, 2010), the presentations from which are available from the event's web page.

Other titles in this series of DPC *Technology Watch Reports* may also be of interest, in particular those on the Open Archival Information System Reference Model (Lavoie, 2004), large-scale archival storage (Linden, Martin, Masters and Parker, 2004), geospatial data, for its consideration of geographic information systems (McGarva, Morris and Janée, 2009), and preservation file formats (Todd, 2009). A second edition of the report on preservation metadata (Lavoie and Gartner, 2005) is in preparation.

9. References

ADS and Digital Antiquity 2009, *Guides to Good Practice*, Archaeology Data Service and Digital Antiquity, http://guides.archaeologydataservice.ac.uk/g2gp/Main (last accessed 28/02/13).

ANS US PRO/IPO-100-1996 1996, *Initial Graphics Exchange Specification (IGES)*,version5.3, American National Standard, US Product Data Association, online at: http://www.uspro.org/documents/IGES5-3_forDownload.pdf (last accessed 28/02/13).

Arms, CR, Fleischhauer, C and Jones, J 2011, Sustainability of digital formats: planning for Library of Congress collections, http://www.digitalpreservation.gov/formats/ (last accessed 28/02/13).

AutoDesk 2012a, RealDWG, http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=770257 (last accessed 28/02/13).

AutoDesk 2012b, DXF reference, http://usa.autodesk.com/adsk/servlet/item?linkID=10809853&id=12272454&siteID=123112 (last accessed 28/02/13).

Ball, A and L Ding (Eds) 2007, *Proceedings of the Atlantic Workshop on Long Term Knowledge Retention 2007*, Bath: University of Bath, online at: http://www.ukoln.ac.uk/events/ltkr-2007/proceedings/ (last accessed 28/02/13).

Ball, A 2011, RRoRIfE: Registry/Repository of Representation Information for Engineering, http://rorife.sourceforge.net/ (last accessed 28/02/13).

Barker, S 2010, 'LOTAR: high integrity CAD data for aerospace'. Presentation given at the *Designed to Last* workshop, London, online at: http://www.dpconline.org/component/docman/doc_download/478-2010julycadbarker (last accessed 28/02/13).

Bozdoc, M 2003, The history of CAD, http://mbinfo.mbdesign.net/CAD-History.htm (last accessed 28/02/13).

Braid, IC 1974, Designing with volumes. PhD dissertation, Cambridge: University of Cambridge.

Brunsmann, J and Wilkes, W 2009, 'Enabling product design reuse by long-term preservation of engineering knowledge', *International Journal of Digital Curation 4*(3), 17–28, doi: 10.2218/ijdc.v4i3.114.

CADAZZ 2004, CAD software: history of CAD CAM, http://www.cadazz.com/cad-software-history.htm (last accessed 28/02/13).

CAx IF 2012, CAx Implementor Forum, http://cax-if.org/ (last accessed 28/02/13).

Cutler, BM Sept. 2003, Procedural authoring of solid models. PhD thesis, Cambridge, MA: Massachusetts Institute of Technology, online at: http://people.csail.mit.edu/bmcutler/PROJECTS/PHD/ (last accessed 28/02/13).

CWA 16200:2010, *A Guide to the Development and Use of Standards-Compliant Data Formats for Engineering Materials Test Data*, Brussels: European Committee for Standardization, online at: ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA16200_2010_ELSSI.pdf (last accessed 28/02/13).

DCC NDa, *Briefing Papers*, Digital Curation Centre, http://www.dcc.ac.uk/resources/briefing-papers (last accessed 28/02/13).

DCC NDb, *How-to Guides*, Digital Curation Centre, http://www.dcc.ac.uk/resources/how-guides (last accessed 28/02/13).

Ding, L, Ball, A, Matthews, J, McMahon, C, and Patel, M 2009, 'Annotation of lightweight formats for long-term product representations',*International Journal of Computer Integrated Manufacturing22*(11), 1037–1053, doi: 10.1080/09511920802527616. Ding, L, Ball, A, Patel, M, Mullineux, G and Matthews, J 2011, 'Lightweight product lifecycle information management for small enterprises',*International Journal of Product Lifecycle Management5*(1), 21–36, doi: 10.1504/IJPLM.2011.038100.

DPC 2010, 'Designed to Last: Preserving Computer Aided Design', Digital Preservation Coalition, http://www.dpconline.org/events/previous-events/625-designed-to-last-preserving-computer-aided-design (last accessed 28/02/13).

ECMA-363 2007, *Universal 3D File Format*, 4thedn.Geneva: Ecma International, online at: http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-363%204th%20Edition.pdf (last accessed 28/02/13).

Eiteljorg II, H, Fernie, K, Huggett, J and Robinson, D 2011, *CAD: A Guide to Good Practice*, ed. by S Dobson, R Lancia and K Nive,York and Tempe, AZ: Archaeology Data Service and Digital Antiquity, online at: http://guides.archaeologydataservice.ac.uk/g2gp/Cad_Toc (last accessed 28/02/13).

Elin, L 1976, 'Synthavision: serendipity from the nuclear age'. In R Leavitt (Ed.) *Artist and Computer*, Morristown, NJ: Creative Computing, online at: http://www.atariarchives.org/artist/sec32.php (last accessed 28/02/13).

EN 9300, *Aerospace series – LOTAR – Long Term Archiving and Retrieval of DigitalTechnical Product Documentation such as 3D, CAD and PDM Data*, multipart standard, Brussels: European Committee for Standardization.

Farin, G 2002, 'A history of curves and surfaces in CAGD'. In G Farin, Hoschek, J and Kim, M-S (Eds) *Handbookof Computer Aided Geometric Design*, Amsterdam: North-Holland, 1–22, doi: 10.1016/B978-044451104-1/50002-2.

Flisch, A *et al*. 1999, 'Industrial computed tomography in reverse engineering applications'. In *Proceedings of the International Symposium on Computerized Tomography for Industrial Applications and Image Processing in Radiology*, Berlin: Deutsche GesellschaftfürZerstörungsfreiePrüfung, online at: http://www.ndt.net/article/v04n07/bb67_8/bb67_8.htm (last accessed 28/02/13).

Fowler, J 1995, *STEP for Data Management, Exchange and Sharing*, Twickenham: Technology Appraisals.

Garrouste, G 2012, 'Electronic data exchange'. Presentation given at the 2012 US/Europe International Aviation Safety Conference, 12–14 June, Cleveland, OH, online at: http://www.faa.gov/news/conferences_events/2012_us_eu/agenda/media/Panel3.pdf (last accessed 28/02/13).

GNU Project 2010, LibreDWG, Free Software Foundation, http://www.gnu.org/software/libredwg/ (last accessed 28/02/13).

Goldstein, BLM, Kemmerer, SJ and Parks, CH 1998, *A Brief History of Early Product Data Exchange Standards*, Interagency Report NISTIR 6221, Gaithersburg, MD: National Institute of Standards and Technology, online at: http://www.nist.gov/msidlibrary/doc/proddata.pdf (last accessed 28/02/13).

Harvard University 2009, *CAD and Image Standards for Construction Documentation*, Cambridge, MA: University Planning Office, Harvard University, online at: http://www.upo.harvard.edu/CampusProjects/Stds/CAD_Standards.pdf (last accessed 28/02/13).

IBM and Dassault Aviation 2011, 'IBM Helps Dassault Aviation preserve critical 3D design data for the Falcon aircraft', http://www.dassault-aviation.com/en/aviation/press/press-kits/2011/ibm-helps-dassault-aviation-preserve-critical-3d-design-data-for-the-falcon-aircraft.html (last accessed 28/02/13).

ISO 6892-1:2009, *Metallic Materials – Tensile Testing – Part 1: Method of Test at Room Temperature*, Geneva: International Organization for Standardization.

ISO 10303, *Industrial Automation Systems and Integration – Product Data Representation and Exchange*, multipart standard, Geneva: International Organization for Standardization.

ISO 13567, *Technical Product Documentation – Organization and Naming of Layers for CAD*, multipart standard, Geneva: International Organization for Standardization.

ISO 14306:2012, *Industrial Automation Systems and Integration – JT File Format Specification for 3D Visualization*, Geneva: International Organization for Standardization.

ISO 14721:2012, *Space Data and Information Transfer Systems – Open Archival Information System (OAIS) – Reference Model*, 2nd edn. Geneva: International Organization for Standardization.

ISO/DIS 14739-1, *Document Management – 3D Use of Product Representation Compact(PRC) Format – Part 1: PRC 10001*, Geneva: International Organization for Standardization.

ISO/IEC 14772-1:1997, *Information Technology – Computer Graphics and Image Processing – The Virtual Reality Modeling Language – Part 1: Functional Specification and UTF-8 Encoding*, Geneva: International Organization for Standardization, online at: http://www.web3d.org/x3d/specifications/ (last accessed 28/02/13).

ISO/PAS 16739:2005, *Industry Foundation Classes, Release 2x, Platform Specification(IFC2x Platform)*, Geneva: International Organization for Standardization.

ISO 16792:2006, *Technical Product Documentation – Digital Product Definition Data Practices*, Geneva: International Organization for Standardization.

ISO/IEC 19775, *Information Technology – Computer Graphics and Image Processing –Extensible 3D (X3D)*, multipart standard, Geneva: International Organization for Standardization, online at: http://www.web3d.org/x3d/specifications/ (last accessed 28/02/13).

ISO/IEC 19776, *Information Technology – Computer Graphics and Image Processing– Extensible 3D (X3D) Encodings*, multipart standard, Geneva: International Organization for Standardization, online at: http://www.web3d.org/x3d/specifications/ (last accessed 28/02/13).

ISO/IEC 19777, *Information Technology – Computer Graphics and Image Processing –Extensible 3D (X3D) Language Bindings*, multipart standard, Geneva: International Organization for Standardization, online at: http://www.web3d.org/x3d/specifications/ (last accessed 28/02/13).

ISO/CD 24517-2, *Document Management – Engineering Document Format Using PDF– Part 2: Use of 32000-2 Including Support for Long-Term Preservation (PDF/E-2)*, Geneva: International Organization for Standardization.

Jacquin, T 2011a, *Semantic Annotations and Collaborative Modelling*, Deliverable D 7.3, SHAMAN Project, online at: http://shaman-ip.eu/sites/default/files/SHAMAN-D7.3.pdf (last accessed 28/02/13).

Jacquin, T 2011b, Preservation aware co-design techniques, Xerox, https://services.open.xerox.com/WebApp.svc/codesign/ (last accessed 28/02/13).

Kalender, WA 2006, 'X-ray computed tomography', *Physics in Medicine and Biology 51*, R29, doi: 10.1088/0031-9155/51/13/R03.

KIM Project 2009, Knowledge and Information Management (KIM) grand challenge project, http://www.kimproject.org/ (last accessed 28/02/13).

Klein, R and Straßer, W 1997, 'Generation of multiresolution models from CAD data for real time rendering: the state of the art'. In W Straßer, R Klein and R Rau (Eds) *Geometric Modeling – Theory and Practice*, Berlin: Springer, 324–335.

Lavoie, BF 2004, *The Open Archival Information System Reference Model: Introductory Guide*, *Technology Watch Report* 04-01, Digital Preservation Coalition, online at: http://www.dpconline.org/advice/technology-watch-reports (last accessed 28/02/13).

Lavoie, BF and Gartner, R 2005, *Preservation Metadata*, *Technology Watch Report* 05-01, Digital Preservation Coalition, online at: http://www.dpconline.org/advice/technology-watch-reports (last accessed 28/02/13).

Lee, G, Sacks, R and Eastman, CM 2006, 'Specifying parametric building object behavior (BOB) for a building information modeling system', *Automation in Construction 15*(6), 758–776, doi: 10.1016/j.autcon.2005.09.009.

Lee, K, McMahon, CA and Lee, KH 2003, 'Design of a feature-based multi-viewpoint design automation system', *International Journal of CAD/CAM 3*(1), 67–75, online at: http://www.ijcc.org/ojs/index.php/ijcc/article/view/70 (last accessed 28/02/13).

Linden, J, Martin, S, Masters, R and Parker R 2004, *Large-scale Archival Storage of Digital Objects*, *Technology Watch Report* 04-03, Digital Preservation Coalition, online at: http://www.dpconline.org/advice/technology-watch-reports (last accessed 28/02/13).

LOTAR International 2012, LOng Term Archiving and Retrieval – LOTAR, http://www.lotar-international.org/ (last accessed 28/02/13).

Lubell, J, Mani, M, Subrahmanian, E and Rachuri, S 2008, *Long Term Sustainment Workshop Report*, Interagency Report NISTIR 7496, Gaithersburg, MD: National Institute of Standards and Technology, online at: http://www.mel.nist.gov/msidlibrary/doc/NISTIR_7496.pdf (last accessed 28/02/13).

Lubell, J, Rachuri, S, Mani, M and Subrahmanian, E 2008, 'Sustaining engineering informatics: toward methods and metrics for digital curation',*International Journal of Digital Curation 3*(2), 59–73, doi: 10.2218/ijdc.v3i2.58.

Maceviciute, E, Wilson, T, Forsgren, E and Menzies, K 2011, *Report on demonstration and evaluation activity in the domain of industrial design and engineering*, Deliverable D14.3, SHAMAN Project, online at: http://www.shaman-ip.eu/sites/default/files/SHAMAN-D14.3.pdf (last accessed 28/02/13).

McGarva, G, Morris, S and Janée, G 2009, *Preserving Geospatial Data*, *Technology Watch Report* 09-01, Digital Preservation Coalition, online at: http://www.dpconline.org/advice/technology-watch-reports (last accessed 28/02/13).

MIT 2009, SIMILE Widgets, Massachusetts Institute of Technology, http://simile-widgets.org/ (last accessed 28/02/13).

Nagel, RN, Braithwaite, WW and Kennicott, PR 1980, *Initial Graphics Exchange Specification*, Interagency Report NBS IR 80-1978 (R), Gaithersburg, MD: National Bureau of Standards.

NAS9300, *Long Term Archiving and Retrieval of Digital Technical Product Documentation Such as 3D, CAD and PDM Data*, multipart standard, Arlington, VA: Aerospace Industries Association.

National BIM Standard – United States 2012, version2,Washington, DC: National Institute of Building Sciences, online at: http://www.nationalbimstandard.org/ (last accessed 28/02/13).

ODA 2012, Introducing Teigha, Open Design Alliance, http://www.opendesign.com/the_oda_platform (last accessed 28/02/13).

Qi, J and Shapiro, V 2006, 'Geometric interoperability with epsilon solidity', *Journal of Computing and Information Science in Engineering 6*, 213–220, doi: 10.1115/1.2218367.

Reed, MK and Allen, PK 1997, 'A robotic system for 3-D model acquisition from multiple range images'. In *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, (Albuquerque, NM), vol.3, Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2509–2514, doi: 10.1109/ROBOT.1997.619338.

Salomon, D 2006, *Curves and Surfaces for Computer Graphics*, New York: Springer.

Shah, JJ and Mäntylä, M 1995, *Parametric and Feature-Based CAD/CAM: Concepts, Techniques, and Applications*, New York: Wiley.

SHAMAN 2012, Sustaining Heritage Access through Multivalent ArchiviNg, http://shaman-ip.eu/ (last accessed 28/02/13).

Smelik, RM, de Kraker, KJ, Groenewegen, SA, Tutenel, T and Bidarra, R 2009, 'A survey of procedural methods for terrain modelling'. In *Proceedings of the CASA Workshop on 3D Advanced Media In Gaming And Simulation (3AMIGAS)*, Amsterdam, online at: http://graphics.tudelft.nl/~ruben/RMSmelik3AMIGAS09.pdf (last accessed 28/02/13).

Smith, M 2009, *Final Report for theMIT FACADE Project*, Cambridge, MA: Massachusetts Institute of Technology, online at: http://facade.mit.edu/files/FACADEFinalReport.pdf (last accessed 28/02/13).

Stroud, I 2006, *Boundary Representation Modelling Techniques*, London: Springer, doi: 10.1007/978-1-84628-616-2_1.

Todd, M 2009, *File Formats for Preservation*, *Technology Watch Report* 09-02, Digital Preservation Coalition, online at: http://www.dpconline.org/advice/technology-watch-reports (last accessed 28/02/13).

UKDA 2013, Create and manage data, UK Data Archive, http://www.data-archive.ac.uk/create-manage (last accessed 28/02/13).

United States National CAD Standard 2011, version5. American Institute of Architects, Construction Specifications Institute and National Institute of Building Sciences, online at: http://www.nationalcadstandard.org/ (last accessed 28/02/13).

VDA 2005–2007, *Long-Term Archiving (LTA) of Digital Product Data, Which Are Not Based on TechnicalDrawings*, VDA Recommendation 4958, Verband der Automobilindustrie, online at: http://www.vda.de/en/publikationen/publikationen_downloads/?pubSuchbegriff=4958 (last accessed 28/02/13).

Versprille, KJ 1975, Computer-aided design applications of the rational B-spline approximation form. PhD thesis, Syracuse, NY: Syracuse University.

Weisberg, DE 2008, *The Engineering Design Revolution: The People, Companies and Computer Systems That Changed Forever the Practice of Engineering*, online at: http://www.cadhistory.net/ (last accessed 28/02/13).

Wilson, PR 1987, *A View of PDES*, Internal Report. Gaithersburg, MD: National Bureau of Standards.

Wong, K 2006, 'What grounded the Airbus A380?', *Cadalyst Daily* 6 Dec.online at: http://www.cadalyst.com/management/what-grounded-airbus-a380-5955 (last accessed 28/02/13).

Yares, E 2012, '3D PDF is important again', *Design World*, http://www.designworldonline.com/3d-pdf-is-important-again/ (last accessed 28/02/13).

Appendix I: Structure of STEP

STEP, the Standard for the Exchange of Product Model Data, consists of 593 published parts (at the time of writing), categorized as follows:

- Part 1 (published): Overview and fundamental principles
- Parts 11–19 (3 published): Description methods – these specify the EXPRESS modelling language (and its extensions) used by the remainder of the standard.
- Parts 21–29 (8 published): Implementation methods – these specify file formats for data conforming to an EXPRESS schema, and an API (known as the Standard Data Access Interface, or SDAI) for manipulating such data.
- Parts 31–39 (4 published): Conformance testing methodology and framework – these specify (in abstract terms) how to test conformance to the standard, especially the application protocols and SDAI.
- Parts 41–59 (18 published: Integrated generic resources – these are the most widely applicable building blocks for application protocols, for example how to represent annotations visually, how to represent geometric dimensions and tolerances, how to ensure the quality of shape data.
- Parts 101–199 (9 published): Integrated application resources – these are further building blocks that are a little more specialized, for example parametric modelling, product assembly modelling, representing computational fluid dynamics data.
- Parts 201–299 (24 published): Application protocols – these are data models for specific applications (such as B-Rep mechanical design, sheet metal die planning, or ship arrangement) and represent the 'top-level' of the STEP hierarchy of parts.
- Parts 301–399 (4 published): Abstract test suites – these are suites of data and criteria for assessing conformance to a specific application protocol (e.g. ATS 304 corresponds to AP 204).
- Parts 401–499 (6 published): Application modules – these are modularized versions of the application protocols (e.g.AM 403 is the modular version of AP 203), sometimes known as implementation modules.
- Parts 501–599 (22 published): Application interpreted constructs – these are intended as intermediate steps between application protocols and integrated resources, allowing several application protocols explicitly to share the same semantics.
- Parts 1001–1999 (493 published): Application modules – foundational modules (wrapping concepts from integrated resources and application interpreted constructs) and conformance class modules from which implementation modules may be constructed.
- Parts 5001–5999 (1 published): Guidance that clarifies use of the standard in particular circumstances.

Appendix II: Structure of LOTAR

The structure of the LOTAR standard is similar to that used by STEP, but somewhat simpler due to the standard's narrower focus:

- Parts 001–009 are Basic Parts that provide an overview of the standard and lay out its fundamental concepts, methods and architecture.
- Parts 010–019 are Common Parts that describe in detail the processes involved at each stage of the archival lifecycle (that is, ingest, archival storage, retrieval and so on).
- Parts 020–029 are Common Parts that describe the metadata requirements for information packages accepted by the archive (submission information packages), stored by it (archival information packages) and provided on request (dissemination information packages).
- Parts 030–039 are Support Process Parts that describe in detail ongoing administrative functions such as testing, auditing, and preservation planning.
- Parts 100+ are Data Domain Specific Parts that spell out how particular types of data should be handled. Certain numeric ranges have been assigned to the broad categories of data:
    - Parts 100–199 relate to 3D geometry with product and manufacturing information.
    - Parts 200–299 relate to product data management information.
    - Parts 300–399 relate to 3D composite (e.g.*n*-ply) structure information in CAD.
    - Parts 400–499 relate to 3D electrical harness data.
    - Parts 500–599 relate to systems engineering data.
    - Parts 600–699 relate to analysis data.