Long-term archiving and reuse of CAD data: viewpoint of Datakit

Datakit considers that it is essential that the vendors of CAD solutions and industry in general mobilize to look for reliable solutions for optimizing the use of data in the long term. Christian Caillet will be representing Datakit at Micado and in the different forums devoted to STEP, notably CAX-IF with its focus on implementation. CAX-IF is also working with the ISO TC184/SC4 organization on standardization issues and the LOTAR groups for long-term archiving.

Caillet insists on the importance of the quality of archived data and points out that when quality is lacking, flaws will almost certainly be reproduced, causing confusion and misinterpretation when the data is reread. He adds that: "the description of archived data should also fully cover complete models, in other words include all product-related data and the links between this data. Also, this description must be sufficiently open to enable forms of exploration that did not yet exist at the time the data was archived."

With its know-how and track record in data exchange and storage, Datakit discusses three alternative scenarios for data archiving.

- A solution based on the STEP standard
- The use of a more compact existing format
- An approach based on ontologies

Scenario 1: STEP standard

The STEP standard is initially a CAD data exchange standard. It has gained acceptance throughout industry and is supported by various standards bodies and implementation forums. An approach to long-term data storage based on this standard provides good coverage. However, while it does represent genuine progress, offering an advantageous solution for consulting and reuse purposes (as a model defined once and for all), it has limits that are liable to impede its extension going forward.

Scenario 2: Use and extension of an existing more compact format

Formats initially defined for data viewing might be extended for the archiving of CAD models. They have the advantage of being more compact, faster to use (at least for their initial purpose), and possibly easier and quicker to develop. However there are gray areas, and, at least in their current form, they cannot really be considered fully suitable for long-term data storage and reuse.

Scenario 3: Ontological approach

Research aimed at finding an alternative to the above approaches based on existing formats and their inherent limitations is ongoing. Datakit has been focusing on the ontological approach that its development team has been using for some of its work. This approach is based on a method for describing and classifying data. Consequently it uses specific access and query methods.

Description by ontologies enables a body of knowledge to be represented in a form intelligible to a computer. An ontology is a structured set of concepts that gives meaning to information. It is a basis for formalizing knowledge that classifies relations between concepts into categories.

The ontological approach is well suited to capturing, interpreting, reusing and sharing a product's semantic data in a heterogeneous environment. The information becomes intelligible to the machine, enabling the subsequent use of methods of reasoning to find the information in accordance with rules of correspondence or similarities between different resources. This method has many advantages. However it is, as yet, relatively unknown and its supporters are faced with a big challenge in promoting it.

By way of conclusion, it is worth pointing out that the success of long-term data storage and reuse is and will be a key factor in determining the success of businesses. The design and management of industrial products invariably generates ever-greater volumes of data that serve as a reference point throughout the life cycle. This reference, for long embodied in paper drawings, now lies in 3D CAD data, and increasingly in a functional model that has been memorized and can be used for simulations. The data encompassing the "3D CAD" in the broad sense include annotations, symbols, features, kinematics, 2D, and management data, all of which falls under the umbrella concept of PLM: CAD/CAE/CAM data …. It is in the sheer wealth of this data that lie its strengths and weakness for users in industry and software vendors who are seeking for a viable solution.

For a copy of the full paper presented at the conference, contact support@datakit.com