

# DocumentDictionary User Manual

Version 0.8 2018-12-11

## Abstract

Welcome to the DocumentDictionary User Manual!

The purpose of this manual is to deliver a core conceptual understanding of some of the fundamentals in efficient practical information management of technical information whose life times span decades.

These core concepts are high-level and hard to acquire without concrete examples. Thus DocumentDictionary which is a fully commercially competitive product within its scope of applicability, is used as a teaching tool to transfer this higher-level of understanding.

Intended readers are author experienced life-time-students and teachers of any age that have many long-term challenges to master with regards to widely distributed document information that must be quality assured, maintained and distributed during the whole life-cycle of a product.

<b>Table of Contents .....</b>	<b>5</b>
<b>Preface.....</b>	<b>8</b>
<b>Chapter 1 The Purpose of DocumentDictionary .....</b>	<b>9</b>
<b>Chapter 2 Data and Information .....</b>	<b>17</b>
<b>Chapter 3 The Role of DocumentDictionary in the Document Development Process.....</b>	<b>19</b>
<b>Chapter 4 The Meta Model of DocumentDictionary .....</b>	<b>23</b>
<b>Chapter 5 Document Model Development.....</b>	<b>29</b>
<b>Chapter 6 Document Source Code Generation .....</b>	<b>30</b>
<b>Chapter 7 Prototype Iteration.....</b>	<b>33</b>
<b>Chapter 8 Configuration Files .....</b>	<b>34</b>
<b>Chapter 9 Import of Document Models .....</b>	<b>36</b>
<b>Chapter 10 Export of Document Models .....</b>	<b>43</b>
<b>Chapter 11 Quality Assurance, Version and Release Management.....</b>	<b>46</b>
<b>Chapter 12 Quality Assurance Functionality .....</b>	<b>49</b>
<b>Chapter 13 Version and Release Management.....</b>	<b>55</b>
<b>Chapter 14 Using a Relational Database for DocumentModel Sharing and Distribution ....</b>	<b>61</b>
<b>Chapter 15 Reuse of Models with Copy and Paste .....</b>	<b>62</b>
<b>Chapter 16 Information Quantity Measurement.....</b>	<b>66</b>
<b>Chapter 17 Profile Extensions of the MetaModel .....</b>	<b>67</b>
<b>Chapter 18 Functionality.....</b>	<b>71</b>

<b>Chapter 19</b>	<b>Summary and Conclusions.....</b>	<b>72</b>
<b>A.</b>	<b>References.....</b>	<b>73</b>
<b>B.</b>	<b>References for Exceptional Students.....</b>	<b>75</b>
<b>C.</b>	<b>DomainModel of DocumentDictionary .....</b>	<b>76</b>
<b>D.</b>	<b>Glossary .....</b>	<b>77</b>

Copyright © 2016-2018, ROJTEC, Olof Johansson

All rights reserved.

*You may download and use a personal copy of this document. You may not distribute copies of this document to 3<sup>rd</sup> parties without a written permission.*



# Table of Contents

<b>Table of Contents .....</b>	<b>5</b>
<b>Preface.....</b>	<b>8</b>
<b>Chapter 1    The Purpose of DocumentDictionary .....</b>	<b>9</b>
1.1    Calculation of information quantity in a model. ....	10
1.2    Support for automated checks and documented quality assurance. ....	11
1.3    Globally unique 128 bit object identifiers.....	11
1.4    Open Export / Import in many neutral formats .....	11
1.5    Explicit well documented quality assured and release managed meta model.....	11
1.6    Automated document source code generation for a number of well established standard document formats. ....	11
1.7    Purpose of the Infological Approach .....	12
1.8    Towards an Efficient Shared Information Highway Network for Implementing the UN 2030 Agenda for Sustainable Development.....	13
<b>Chapter 2    Data and Information .....</b>	<b>17</b>
2.1    Definition of Data .....	17
2.2    Definition of Information .....	18
<b>Chapter 3    The Role of DocumentDictionary in the Document Development Process.....</b>	<b>19</b>
3.1    Brief history of OOCASE .....	20
3.2    Product Modeling Systems.....	20
<b>Chapter 4    The Meta Model of DocumentDictionary .....</b>	<b>23</b>
4.1    DBObject.....	23
4.2    Element .....	23
4.3    ModelElement .....	24
4.4    LabelModelElement .....	24
4.5    NameSpace.....	24
4.6    Object .....	24
4.7    Package .....	24
4.8    Model .....	25
4.9    DocumentDictionary .....	25
4.10    A DocumentModel is represented by a Category .....	25
4.11    Category .....	25
4.12    CategoryLink.....	25
4.13    DocumentRecord.....	26
4.14    ContentRecord.....	26
4.15    ObjectRecord.....	26
4.16    AuthorRecord.....	26
4.17    AuthorDocument.....	26
4.18    Cite .....	26
4.19    ElectronicEdition.....	26

4.20	Note .....	26
4.21	Graphical Syntax used in Object Model Diagrams .....	27
<b>Chapter 5</b>	<b>Document Model Development.....</b>	<b>29</b>
<b>Chapter 6</b>	<b>Document Source Code Generation .....</b>	<b>30</b>
6.1	Overall workflow .....	30
6.2	DocumentDictionary built-in Document Source Code Generators.....	30
6.3	Document Source Code Generation Services .....	31
6.4	SQL based Document Source Code Generators.....	31
6.5	Organizing Document Source Code Generator Build Systems.....	31
6.6	Using GIT for Document Source Code Generator Maintenance .....	32
6.7	Quality Assuring Document Source Code Generators with the Benchmark Document Model .....	32
6.8	Test Suites for Document Source Code Generators.....	32
<b>Chapter 7</b>	<b>Prototype Iteration.....</b>	<b>33</b>
<b>Chapter 8</b>	<b>Configuration Files .....</b>	<b>34</b>
8.1	Directory Structure.....	34
8.2	Text format.....	34
8.3	Macro expansion \$(<parameter name>) .....	34
<b>Chapter 9</b>	<b>Import of Document Models .....</b>	<b>36</b>
9.1	BibTeX.....	36
9.2	RDF .....	37
9.3	RIS .....	37
9.4	ODBC.....	37
9.5	TEXT .....	37
9.6	XML.....	40
<b>Chapter 10</b>	<b>Export of Document Models .....</b>	<b>43</b>
10.1	Binary Storage Formats.....	43
10.2	BibTeX.....	43
10.3	HTML .....	43
10.4	RDF .....	43
10.5	RIS .....	44
10.6	RTF .....	44
10.7	TEXT .....	44
10.8	SQL .....	44
10.9	XML.....	44
<b>Chapter 11</b>	<b>Quality Assurance, Version and Release Management.....</b>	<b>46</b>
11.1	Purpose of Quality Assurance .....	46
11.2	Purpose of Version Management .....	47
11.3	Semantic Versioning 2.0.0 .....	48
<b>Chapter 12</b>	<b>Quality Assurance Functionality .....</b>	<b>49</b>
12.1	Checking and Automated Checks .....	49
12.2	Check Level Structure and Checkpoints .....	50
12.3	Recording a passed check .....	53

12.4	Approving an object.....	54
<b>Chapter 13</b>	<b>Version and Release Management.....</b>	<b>55</b>
13.1	Editions, Versions and Releases.....	55
13.2	Version History .....	57
13.3	The VersionOwnerPath.....	58
<b>Chapter 14</b>	<b>Using a Relational Database for DocumentModel Sharing and Distribution ....</b>	<b>61</b>
14.1	Requirements for Providing a Relational Database Service .....	61
<b>Chapter 15</b>	<b>Reuse of Models with Copy and Paste .....</b>	<b>62</b>
15.1	Dominance Ranking of Classes in a DomainModel .....	63
15.2	The Copy/Paste Metaphore and it's complication in the real world.....	63
15.3	Examples for Functionality Coverage and Performance Analysis.....	64
15.4	Managing Traceability with Object Identifiers during Copy Paste.....	65
<b>Chapter 16</b>	<b>Information Quantity Measurement.....</b>	<b>66</b>
16.1	Theory .....	66
16.2	Practise .....	66
<b>Chapter 17</b>	<b>Profile Extensions of the MetaModel .....</b>	<b>67</b>
17.1	Short Introduction to Profiles.....	67
17.2	The DomainModel of Profiles.....	68
17.3	Comparison with SGML DTDs .....	70
<b>Chapter 18</b>	<b>Functionality.....</b>	<b>71</b>
18.1	Egoless Business .....	71
18.2	Building Efficient Interfaces Between Huge Refactorable Knowledge Domains - Efficient Knowledge Economy .....	71
<b>Chapter 19</b>	<b>Summary and Conclusions.....</b>	<b>72</b>
<b>A.</b>	<b>References.....</b>	<b>73</b>
<b>B.</b>	<b>References for Exceptional Students.....</b>	<b>75</b>
<b>C.</b>	<b>DomainModel of DocumentDictionary .....</b>	<b>76</b>
<b>D.</b>	<b>Glossary .....</b>	<b>77</b>

# Preface

*To Documentation Engineers and Technical Writers who want to get something done without complicating it more than necessary*

DocumentDictionary is an Object Oriented Computer Aided Document Engineering application that significantly enhances the productivity of a technical writer or collaborating documentation development team.

Every tool has its *scope of applicability*. Outside this scope there is an exponentially rising "cost/benefit" barrier that requires a different architectural and methodological approach to break through. This barrier has to do with what we humans are and the limitations of how much knowledge and information a person or small team of efficiently collaborating documentation engineers and technical writers can be masters of.

The DocumentDictionary client workstation applications works well for model driven document development for documentation whose documentation models contain up to some hundreds of document records and some ten thousands of content records. Such models may produce generated documentation source code sizes of perhaps million lines of documentation source code in the whole range of standards available for electronic documentation formats, (e.g. doc, html, json, pdf, RDF, rtf, txt,xml) for ONE document dictionary that can be maintained by a small team.

The DocumentDictionary server database works well with object oriented documentation models up to some still yet not in any practical project meet number of records, and has been used for porting document record libraries between software systems platforms that contain millions of document records. However development and maintenance of such kinds of document dictionaries require larger teams of people that are organized in a healthy knowledge ecology. This requires a company with healthy corporate governance and well maintained long-term win-win relationships with organizations and companies in its supply chain and its customers, to long-term maintain the quality of the documentation.

The purpose of this User Manual is to provide the theoretical framework for becoming proficient in using DocumentDictionary, and with that conceptual understanding have acquired the knowledge potential to develop and deliver the next generation of tools we documentation engineers and our collaborating environment peers need, to with implementations that work in practice, cost-efficiently master the known challenges that decade spanning document information management impose. Including providing users with high-performance.



# Chapter 1 The Purpose of DocumentDictionary

*When the purpose of something is MUCH bigger than itself, it receives a divine force that can make it overcome*  
*[The Poet]*

The purpose of DocumentDictionary is efficient use and reuse of Document Models<sup>1</sup>.

An Document Model is a declarative design specification, or can be seen as a contract of what information a document distribution medium format should be able to capture, store, process and present.

The document model is expressed in a language that after a short introductory training can be understood by non computer specialists, and serves as a tool for communication between 3 different groups of people. a) domain specialists that provide the requirements of information representation needed by their knowledge domain, b) computer specialists that implement a software solution, and c) end users of the produced document formats who use the core information in document models as documentation during their daily work.

DocumentDictionary is an object oriented computer aided document engineering tool with a number of unique features.

- 1) Calculation of information quantity in a document or documentation set.
- 2) Support for automated checks and documented quality assurance.
- 3) Globally unique 128 bit object identifiers for distributed document model development
- 4) Open Export / Import in many neutral formats
- 5) Explicit well documented quality assured and release managed meta model.
- 6) Automated document source code generation for a number of well established e-document formats.

Now it's perfectly alright and recommended to skip to Chapter 2 on page 17, since the following is only interesting for large scale project managers who build IT-systems that need to be operational for decades in whatever shape new technology allows safe document information storage and document exchange resources to be implemented upon.

---

<sup>1</sup> An document model defines Document Structure AND from the Document Structure directly inferable Representation Formats and Behaviour functionality in interactive electronic documentation. Inferable behavior that implements from the structure inferable use cases, whose programmed behavior follow easy to remember and reuse user senso-motoric behavior with reused parameter setting conventions and can be implemented by automated source code generators specialized for a highly optimized implementation in a particular target document source code language. Other conformant document model languages such as SGML define both structure and presentation behavior by for example markup declarations and link process definitions. Large amounts of basic kinds of behavior can be derived from the Document Structure and implemented automatically with model driven documentation source code generators or model driven documentation interpreters. For new application development it is cost efficient not to waste too much time on hand written document software behavior before the core document information structure (DocumentModel) is well understood with examples that are entered into generated prototype document applications populated with realistic production environment document contents.

## 1.1 Calculation of information quantity in a model.

Information quantity is measured in a unit called EC for elementary constellation. An elementary constellation or e-constellation is the smallest possible unit that still carries meaningful information that can be stored or transmitted as a message from a sender to a receiver.

The purpose of this measure is to provide decision support for which model to choose if there are several alternative ways of modeling a certain physical product or artifact, perhaps using different software systems.

A model can be anything that has been formalized with the language constructs in the DomainModel language to a level of detail that allows software to be implemented and the model's information stored in a database representation. If two models provide equivalent functionality but differ severely in complexity or computational performance on present technological infrastructure available for the purpose of the model's end user application, someone has to decide which model to choose.

Fact based measurements are decisive (or important decision support in political environments) when the accumulated cost of several decades of IT-system maintenance for serving a fleet of complex engineered products that delivers a fundamental service for the sustenance or protection of a customer's society, requires IT-support for cost-efficient management and maintenance.

In a modern larger engineering company that promotes its staff by merit and leaves plenty of opportunities for choices of a future career open and inviting, the time at the IT department while learning the information structures of the core business is a knowledge development platform for the staff supply that can take on challenges of the more advanced jobs<sup>2</sup>. Jobs that require an understanding of how to implement production capacity for new business opportunities. This in modern engineering companies generous staff meritation requires that the supply can be held up by efficient education of new staff to replace the vacuum for skill enabled career advancers that exposure to this core business knowledge in a concentrated format produces. The choice of model can actually impact the choice of paradigm for career advancement in a company, and with that the whole future for its business<sup>3</sup>.

The output of a modern larger engineering company in the form of life quality sustaining societal infrastructure products are frequently taken for granted by consumers. It is only the experts of these products who have the REAL power to make them deliver their output at the cost possible given the current state of the art knowledge in all those fields of expertise who combined make such products producible at an affordable cost for the end users of the product or its services.

Information quantity based decision support is especially important when two separate communities fight over a standard and are unexperienced<sup>4</sup> in each others technological domains or software implementation support for these technological domains. Unnecessary complex models are harder to teach and maintain, and divert resources from other important areas of development. Having some measurable facts may resolve disputes and get the "fighting communities" focused again on delivering added value, standing on a fact based ground, and if corporate governance is excellent, get amplified by mutual education.

---

<sup>2</sup> Where the best examples of Engineering Companies are plain clever self sustaining career production machines for talented people whose skills are necessary to fix our real problems.

<sup>3</sup> There is always potential competitive advantage layers above a current well-known established business. Those above layers requires creative people who know the core business AND something else that none of the established market players have thought of before or delivered the investments to make it become available for a solvent enough customer basis who appreciates its new products and buys them for a GOOD reason. If the core business is obscured by an unnecessary complex model, the supply of people who understand it well enough to develop the potentially business income generating layers above it will be throttled. This is stuff that matters over decades, when "great asset" people move for reasons outside the control of the company.

<sup>4</sup> This is a real fact due to the enormous size of various industrial information models, and the amount of studying time and practice it takes to become familiar enough with their details to make fact based decisions.

## **1.2 Support for automated checks and documented quality assurance.**

The Quality Assurance techniques applied here were adopted from the mechanical engineering industry where the high cost of failures due to errors in design specifications (drawings etc) drove this industry to develop survival skills that add some spending in the earlier stages of development. Where this added spending serves as insurance against unpleasant expensive surprises later. DocumentDictionary supports a number of levels of automated quality checks, that aid various "check-points" in a cost efficient iterative development cycle where many people (frequently with too little time) are involved.

## **1.3 Globally unique 128 bit object identifiers**

Ensuring global unique identifiers for objects is a re-occurring problem throughout the whole globalizing IT industry. It has with performance in applications to do, scalability and ability to produce large amounts of uniquely identified objects in parallel by people and teams that are unaware of each others existence. Where the unique object identifier issuing mechanism must provide the ability to combine results of independent uncoordinated work in a database that requires unique object identification AND traceability for quality assurance reasons, and efficient on-demand-access to linked external resources outside the local database.

The method chosen for DocumentDictionary and plenty of production systems produced, is providing each creator of new object identifiers with a unique 64 bit identifier (HighId). Each such creator has a self managed 64 bit incrementing counter (LowId) for lifespan unique identifiers from that source.

There are plenty of other ways to solve this problem, but this approach is simple, efficient and it works in practice.

## **1.4 Open Export / Import in many neutral formats**

DocumentDictionary provides many ways of exporting and importing information models. Thus your models are never locked-in within this tool. If a better tool comes along (creative destruction) you can proceed with that.

## **1.5 Explicit well documented quality assured and release managed meta model.**

The meta-model of DocumentDictionary is licensed to all paying customers for their own implementation needs in the most empowering format. DocumentDictionary is modeled in OOCASE. Thus if you develop document source code generators for a new software platform that completely outperforms the one DocumentDictionary is using, you are free to implement your own DocumentDictionary tool on that new platform and migrate to that platform with all your document information model assets intact<sup>5</sup>.

## **1.6 Automated document source code generation for a number of well established standard document formats.**

Document source code generators are available for a number of SQL92 compliant relational databases. Some are efficiently implemented in DocumentDictionary itself, and several can use the standard neutral XML export formats of DocumentDictionary contents in XSLT transformation pipelines etc and a few other programming languages.

---

<sup>5</sup> This may seem stupid with regards to the self-sustainment principle of the company making a living on DocumentDictionary. However in the perspective of global warming whose solution is more important than the self-sustainment of a particular company whose employees can find a living somewhere else, it's non-productive with regards to over history gathered experience to prevent "creative destruction" to happen if the new alternative over time and by facts and evidence delivers a much better output performance with regards to achievement of the goals setup in the UN 2030 Agenda.

---

The rest of this introductory chapter is there for readers who want to understand the benefit of a more efficient standardized "asphalt laying machine" for putting "tarmac" on the emerging DOCUMENT INFORMATION gravel roads of all diverse shapes and sizes, so they without numerous severely errorprone and costly reloading can carry the truckloads of global document information we need to ship to the education centers that can convert it to reliable education plans for the human capital that implement the strategies designed by our industrial leaders and national governments.

## 1.7 Purpose of the Infological Approach

To understand the purpose of any kind of software one has to analyze its role in the larger whole.

The following is a quote from the preface of [Sundgren 73] which describes the foundational infological theory on which DocumentDictionary builds.

"An infological approach to data bases" reports parts of the data base research and development work which has been carried out over a number of years at the National Central Bureau of Statistics, Sweden. Professor Börje Langefors, University of Stockholm, Department of Administrative Information Processing, has been the scientific supervisor of the reported project. Very briefly the objective of the project has been to develop an integrated theoretical framework for design of large-scale data bases. The framework should

- (a) enable people who are not data processing professionals to co-operate actively and constructively in data base design projects
- (b) make it possible to transform systematically the problems, desires, and requirements of those who are affected by the projected data base into problems which can be tackled by data processing specialists
- (c) enable data processing specialists to analyze the computer-oriented data base problems systematically and with sufficient precision
- (d) make it possible to design data bases with which decision-makers, planners, and researchers within different specialized fields could interact constructively, even if the information needs of the interactors are complex, and even if they lack knowledge about computers and computing

There are definitely different opinions among authorities in the computing world as to whether it is feasible to cover all the aspects (a)-(d) within one and the same framework. This report supplies evidence in support of the hypothesis that an integrated approach is both feasible and necessary for the success of large-scale data base undertakings."

The above quote from year 1973 is still valid, however the situation has improved. (a) has been improved with graphical representations of information models that are used in interactive development seminars where a mix of domain specific expertise participate and all understand what they are talking about so efficient communication can take place. (b) and (c) have for the purpose of implementation of basic information handling software functionality for delivering fully functional prototype software implementations been fully automated for certain target platforms. (d) has been significantly improved with automated model driven declarative implementation of software prototypes from information models, that enable domain experts to express their expertise with large scale examples, that reveal the "problems" in the details, where the *Information Model* does not adequately represent reality. Some stuff in seminar or prototype evaluation situations are "gutt feelings" of participating experts, and it requires certain "emotional language literacy" and social skills by a software engineer/seminar

leader to get that information out. According to a non-disclosed source there are 39 different emotional expressions that experienced people use while communicating interactively. Human skills in understanding and acting efficiently on the cues of non-spoken emotional language AND knowing the domain of the expert to a level where the facts can be brought out by asking the right questions requires a special brand of people, of which YOU are a candidate.

Or you can focus on the technical implementation parts of translating declaratively specified *Information Models* to efficient software implementations on the latest superior hardware and software platforms.

## **1.8 Towards an Efficient Shared Information Highway Network for Implementing the UN 2030 Agenda for Sustainable Development**

Some problems can not be solved in traditional ways since there is no-one who owns the problem. Or there exists no single entity with enough resources or authority to solve the problem in practice. Or the entity producing the problem is not within the authority of the entities subjected to its effects.

In order to solve a problem it must first be understood. The UN 2030 agenda for sustainable development has set out a goal. To reach that goal, we need a plan for how to get there.

A typical approach could look like:

- 1) Build a reasonably adequate map of the current situation
- 2) Identify spots where investments would provide largest return with regards to goal achievement
- 3) Allocate resources to fix those problem spots
- 4) Implement the fixes and restart at 1)

Besides ignoring the natural law of self-sustainment<sup>6</sup>, the above approach hit's it's exponential boarder of applicability rather quickly, due to a problem we encountered in Software Industry a long time ago:

### ***The Language Problem***

We still have not solved it satisfactorily however software industry rolls on, in its complex supplier value chains, each actor in it's own little language islands, at the "speed and load capacity of a horse/8-bit CPU", where we could use a "modern truck/64-bit multi-core" instead.

But that is unfortunately not possible, since there are to few "roads/standards" that can "carry such a truck/make use of available information exchange eco-system" (even if there are instances of such "roads" and "trucks" in certain niches that have an enormous turn-around).

Even if it is possible to implement a "truck" for, for instance "Environmental Education", there would not be a large enough market for it. The infrastructure ranging from CPU, OS, DataBase, Network Communication Protocols, User Interfaces, Local User Language Adaptations, needs to be implemented with instances for which there is an educated work force who can install and maintain them.

---

<sup>6</sup> Law of self sustainment) An active entity e.g. @) biological cell, a) plant, b) animal, c) person, d) organization, e) company or f) nation;

that can not find a way to sustain itself with; @) nutrition and energy (provided by its organism), a) nutrition and sunlight, b) food, c) food and housing, d) work force, e) work force and income to pay taxes and work force, f) tax income (to pay for education, defence and law enforcement), educated workforce, peace and stability, law enforcement (that allows its educated citizens to set up companies that are not robbed of their resources (material, money, working time), and employ an educated workforce that can produce the added value necessary to generate tax and the companie's self sustainment);

will starve and eventually perish. The basis for action is energy, without energy action is not possible.

And finally the end users, which collectively are the most expensive and valuable actors in this chain, needs to be educated in order to know what decisions to take based on the delivered "Environmental Education".

But there is a solution to this. A distributed one. One that will require an agreement or arrangement of peace and non-hampering interference by external actors that have their own agendas (or internal problems) and don't care about "the truck's" success since it will not be under their control.

The solution is based on some lessons from the evolution within software industry and other domains, and delivers shortcuts that can shorten the time table from decades to years.

To explain this the "generalize from examples" approach is taken

### **1.8.1 The Software Expansion**

The personal computer was founded on the invention of the CPU on a chip. Success stories of chips like the Intel 8080 lead to low cost competitors like the Zilog Z80, Motorola 6800 and MOS 6502. The possibility to mass produce personal computers at a low cost lead to an extraordinary expansion during the late 1970's. The availability of PC's in all kinds of different industrial, academic and personal environments lead to the creation of software industry that in turn had an extraordinary expansion in the 1980's.

Due to the large distributed presence of personal computers and software tools for software development, the same scale-up problems when software grew larger were detected in vast amounts of different places. Plenty found their own solutions and put them into their own software and software development methods.

In the 1980's a very diversified ecosystem of programming languages, software development methods, tools and software products had emerged.

In academic and other meeting arenas for software developers, people realized that they had common problems with translating data from one program to a suitable format that could be used in another. The same thing appeared in larger software development projects where models and methods from different interacting sub suppliers had to be interfaced or integrated. There was a growing need for a common language.

There seems to be three approaches for finding a common language. These have been tested in plenty of computer science historical peak events and delivered their output in the form of publications, standards, organizations and companies maintaining the standards. The three identified alternatives are:

- 1) Some clever people with indepth knowledge of the problem, design something that takes the best out of everything they know and design something new that solves the problem.
- 2) A large group of representatives with different backgrounds, requirements and visions come together and with fact based arguments and efficient negotiation techniques come up with a compromise that works and deliver an output that is useful to all participants.
- 3) Some actor goes ahead and markets its solution to a scale that it in practice becomes the de-facto standard that gets the most users and thus everyone, in one way or another, has to adapt to. A lock-in that, if the technology is inferior compared to others, may put a suffocating blanket on development seen from a larger context and the potential available in its large user base.

OK, so what are we supposed to say about this?

A Swedish famous quote is: "Ja, det är för jävligt" and that ends the story with a statement that everyone can agree with. A big sigh and no change. However that comment is a violation against since long gathered wisdom. The ones that know something better need to pursue their acts, knowing and learning more, and facts if proven usually makes a difference. You might know where the facts are, however you must find them and present them in a way that serves the purpose of change towards the better.

And remember, it usually does not matter what a GOOD leader who cares about her/his followers does, as long as it put's unity within the group which ends internal fighting and frees up working resources for working towards a goal that leads to an improvement.

So lets focus on combatting global warming, and the facts about what is needed for that will fix the rest.

#### **1.8.1.1 1) Engineered Standard - Relational Database Language (SQL) Example**

Pros:

+

Cons:

-

#### **1.8.1.2 2) Negotiated Standard - Unified Modeling Language (UML) Example**

Pros:

+

Cons:

-

#### **1.8.1.3 3) DeFacto Standard - (somewhat sensitive to select the example) Example**

UNIX, Windows, Linux, C, C++ etc.

Pros:

+

Cons:

-

### **1.8.2 The Railway Expansion**

With the invention of the steam engine, and the landmark locomotive Rocket , it suddenly was possible to move heavy goods and passengers over larger distances that outperformed "horse and carriage". Railways were built in all kinds of places in Europe by entrepreneurs with a transport business idea by different contractors.

As the different tracks came to meeting ends, it eventually became clear that the whole transportation system would become much more efficient if there was a standardized track size. The benefit with a common standard was that the goods would not have to be moved between trains running on different track sizes at the meeting points, and that a manufacturers of locomotives and waggons would get a much larger market if all rail-tracks had the same dimensions, thus enabling mass production of the same designs, with a better profit margin as result, and competition which lowered the prices for "rolling transport infrastructure".

### **1.8.3 The Shipping Industry Expansion**

A break through in the cost efficiency of shipping was the invention of the container. A standardized package for all kinds of goods that could be transferred between ships, railway carriages and trucks reduced time and cost at goods transit points. Just the lowered transport cost made it possible to trade larger volumes of goods at a lower profit margin, thus increasing trade.

The common lesson from all these examples is that the EXACT formulation of the contents of a standard is not that important, however it **MUST** work in practice, it **MUST** be efficient compared to the state of the art, and it **MUST** have a self-sustaining system structure where all participant roles in that structure have win-win relationships to each other. Who is taking a particular role in the system structure is not that important if that actor maintains healthy win-win relationships with its partners in a supply chain **AND** maintains healthy co-opetition with its competitors, where they collaborate on developing efficient standards for the higher layers that are not yet mature enough or large enough to provide the volume benefits of a by **GOOD** standard enabled mass market. There is always a higher layer for the actors that are thrifty, follow the natural laws of healthy business ecosystem dynamics and build their value adding products on the best available standards.



## Chapter 2 Data and Information

*Some things are immaterial, but must be transmitted with the aid of the material*  
*[The Poet]*

Most people have an intuitive understanding of the terms data, information and knowledge. Like for most words this understanding is a product of their life experience of how these words are used. Many see them as different labels for the same thing, and use them interchangeably. People talk about databases, information technology and knowledge based systems, without really making much difference between them unless having some deeper interest or experience in developing such things or exploring them.

There is however a clear benefit of having a more precise definition and understanding of the concepts of data and information in the quest of developing efficient information system.

### 2.1 Definition of Data

The following definition is from [Sundgren 73], the thesis that my GREAT professor put in my hands after entering his office and explaining my problem of making measurable science out of data modeling.

**Definition of Data:** [Sundgren 73], page 20

***If a person intentionally arranges one piece of reality to represent another, we shall call the former arrangement data, and we shall say that the arranged piece of reality is a medium, which is used for storing the data.***

*This definition covers all kinds of data: digital data, analogous representations, spoken and written language, etc.*

...

*Note the wording " ... intentially arranges ... to represent".*

*It is not sufficient that two phenomena are related to each other, incidentally or by design. There should have been a human intention of representing , and not only correlating, one thing with another.*

*In most cases data represents primarily a person's knowledge about reality and only secondarily the piece of reality itself.*

People have been using data as an aid in their daily activities for a very long time. Before the number systems were invented or taught, shepherds had a bag with small stones, one for each sheep, which they used to account for them.

## 2.2 Definition of Information

The concept of information is more difficult to define in one single sentence. It is easier to describe some of its important properties. Information exists only in the mind of a human being as a part of that person's mental frame of reference. By a frame of reference, we mean the collection of concepts, definitions, laws of logic, empirical laws and perceived, deduced or deducible knowledge belonging to the mind of that reference person P at a particular time. A person's frame of reference will change continuously, depending on what new knowledge he/she acquires, and what is currently in focus in his/her conscious mind.

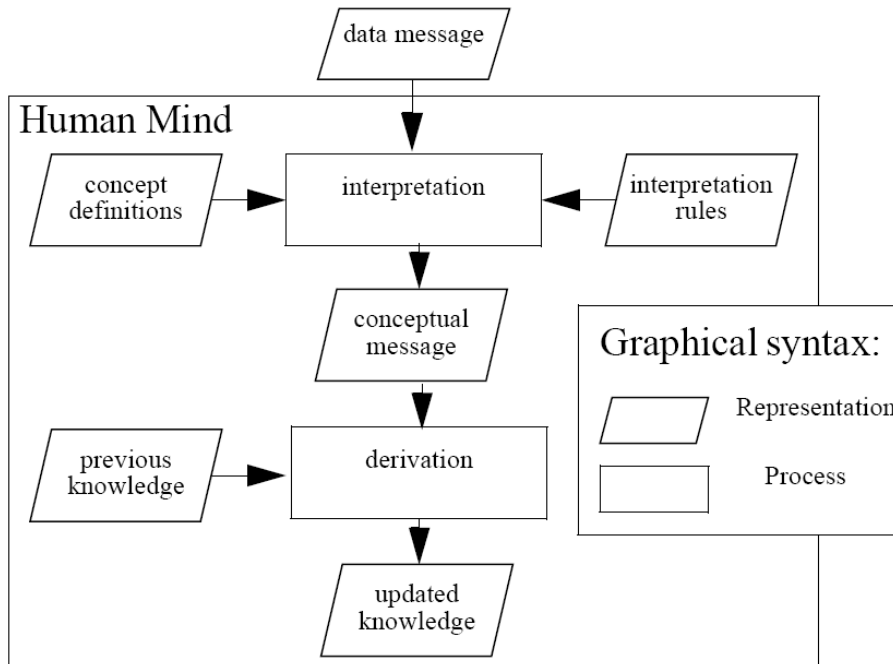


Figure 1. Transformation of data into information. From [Sundgren 73] page 24.

Infological theory has its roots in Börje Langefors' work [Langefors 66][Langefors 93]. A major contribution to the understanding of the nature of information and data in the context of communication with humans is concisely expressed in his infological equation:

$$(EQ\ 1)\ I = i(D, S, t)$$

where I is the information (or knowledge) produced from the data D and the pre-knowledge S of a person, by the interpretation process i during the time t. In the general case, S in the equation is the result of the total life experience of the individual.

This theory was later refined in [Sundgren 73], and adapted to the 1990's main stream scientific and commercial software supported theory of object-orientation in [Johansson 96]. That thesis presented a framework for measuring information quantity in a unit (e-constellations) that is close to the underlying representation in which humans store and index information in their brains [Appendix B].

## Chapter 3 The Role of DocumentDictionary in the Document Development Process

*A skilled actor can play many roles. The extent to whether the performance is perceived as an excellent experience, depends on how well the actor knows the role and it's purpose*

*[The Poet]*

DocumentDictionary is ONE example of a small client-server product modeling application that can be designed and maintained with the Object-Oriented Computer Aided Software Engineering (OOCASE) model driven software implementation framework [Johansson 1996], [ROJTEC 2018d].

The features of DocumentDictionary, such as the ability to work with very large object-oriented document models stored in database servers has its roots in the history of OOCASE which is used to design and maintain the DocumentDictionary client-server product portfolio. The special scale-up and productivity requirements of the Product Modeling applications for Computer Aided Gas- and Steam turbine design, drove the development of OOCASE and made it focus on a practically working core theory. Whereas many other CASE tools originated from the need to document existing large software systems in a graphical way, OOCASE had the privilege of being the origin of the design specification of whole systems. The DocumentDictionary client and server applications reuse much of the functionality in the software frameworks in OOCASE and other product modeling applications developed with OOCASE.

DocumentDictionary is targeted towards long-term maintenance of document specific information, that can be represented in all kinds of different interactive electronic formats. Where the information source in the form of DocumentModels needs to be quality assured, version and release managed deep down into the table of contents structure of individual documents in order to keep product documentation in sync with the latest customer product developments. The electronic documentation including references and indexes for a particular product release can be generated using the information stored in a particular release of a DocumentModel stored in DocumentDictionary.

The productivity benefits of model driven document source code generation start to pay off already when developing small documentation sets for a particular product portfolio. The benefit of building a documentation set from core documentation modules that describe specific functionality that is the same in different products in product family, makes it easier to manage the document production when changes are made in components that are used in many different products.

Thus more of a document developer teams' time and effort can be allocated to "high value" activities with regards to writing high-quality instructions that explain functionality and improve the productivity of end users.

The rest of this chapter describes the history of OOCASE, and what range of industrial applications that lead to the information structure and design of the DocumentDictionary DomainModel.

The text comes from the OOCASE User Manual. You can regard the program DocumentDictionary on your workstation as a client application and the DocumentDictionary database as a product model database.

### 3.1 Brief history of OOCASE

OOCASE has its roots in the late 1980's and early 1990's. In those times the "patterns of thinking" and understanding of information systems development met with leading research in software development and formed new ways of thinking that set foundations for great leaps in software development productivity.

Concepts that had been discovered and rediscovered almost everywhere where larger software systems were developed now became widely published. Some agreement on common vocabulary and terms started to form in the literature and commercial development tools.

Those were the times of 4<sup>th</sup> generation programming languages<sup>7</sup>, and vendors and users of Computer Aided Software Engineering (CASE) tools were developing new markets.

There was however a gap in the CASE markets in the area where main stream commercial databases met with computer aided design (CAD) tools. The software tools at that time had been developed by groups with different end user requirements. The database groups were focusing on masses of textual business data, whereas the CAD groups were focusing on graphical drawings of mechanical and electrical designs.

Large electrical and machinery engineering firms had a need to integrate large amounts of business and CAD data in their information flows, and that job was done on a case by case project basis by a new wave of software application consultancy businesses.

The idea of an integrated product model (PM) that described all aspects of these large engineering firms' products and how to build such systems with the CASE tool approach began to form.

### 3.2 Product Modeling Systems

A product modeling system is a computer-integrated development environment for a specific class of advanced products. A PM-system consists of a product model database which is interfaced with CAD-applications that support graphical designs of engineering models, graphical user interfaces for browsing and modification of the object structures in the product model, and computer aided engineering (CAE)-applications that make engineering calculations on the models.

Figure 2 shows the approach taken to manage the software engineering of product modeling systems. The idea is to maintain a high-level PM-system design specification in the form of an object-oriented CASE model in a meta-database; in the early 1990's commonly called data dictionary.

---

<sup>7</sup> Programming Language Generations: 1<sup>st</sup>) = numeric machine code, 2<sup>nd</sup>) assembly language, 3<sup>rd</sup>) languages like Pascal, C, ADA, etc.

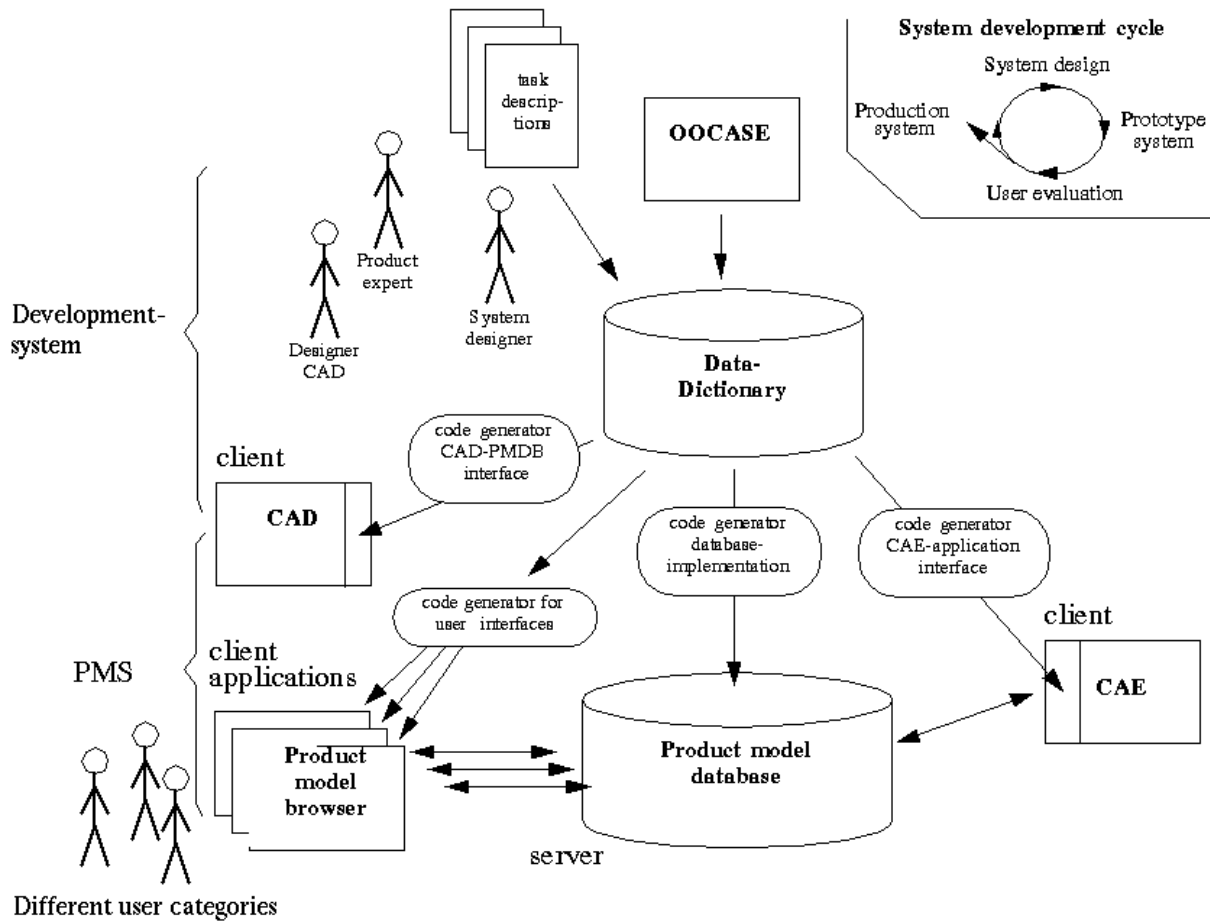


Figure 2. Software development approach for product modeling systems

The OOCASE *DomainModel* of the objects to be managed in the product model database is developed in cooperation with product-, CAD-, and CAE-application experts, and can be stored in a DataDictionary database.

The software development approach for product modeling systems works as follows.

Most of the source code for the PM system implementation is generated automatically, using SQL-based source code generators. The development system generates database schemas with stored procedures and triggers that provide a high level interface for application program interaction with product models. It also generates browser applications for form-based interaction with product model data, and interface modules in the native application development language of a CAD-system. Through these, a CAD application developer has access to the product models in the database on an abstraction level that is natural for an engineer.

By automatic generation of most of the surrounding OOCASE-model dependent software, changes in the model can quickly be implemented in a prototype system and evaluated by experts and end users.

The System development cycle depicted in the upper right corner of Figure 2 consists of:

- 1) System design that is modeled in OOCASE and maintained with the help of the DataDictionary.
- 2) Creation of a prototype system where the code generators deliver the source code that directly depend on the design of the DomainModel for the PM system, which may change in each iteration.
- 3) User evaluation, where the users model their products with the functionality provided by the PM-system prototype, give feedback on functionality and performance, get inspired and deliver new lists of wishes and requirements, which are collected and taken into the next System design iteration.

- 4) If the users are satisfied with the functionality of the PM-system, it is taken into production on a system platform that can handle the load of a full user base and lives up to the requirements of a production system. Which are maintenance requirements with regards to regular backup procedures, bug-reports, user help-desk support, basic introductory education of new users, etc .

## Chapter 4 The Meta Model of DocumentDictionary

This chapter provides understanding that is important for your productivity in DocumentDictionary. It describes the classes, attributes and relationships in the DomainModel of DocumentDictionary.

Please have a look at Appendix C DomainModel of DocumentDictionary on page 76. This object model diagram describes how objects within a document model in DocumentDictionary can be organized. The graphical UML class diagram syntax is described in section 4.21 Graphical Syntax used in Object Model Diagrams on page 27.

The DomainModel of DocumentDictionary reuses modules that contain classes, attributes and relationships from the OOCASE DomainModel. You have the graphical display of the classes DBObject, Element, ModelElement and Model to the left in the diagram of Figure 3 Essential structures and classes in the OOCASE domain model. See 4.21 for an explanation of the diagram syntax.

### 4.1 DBObject

All product model database objects inherit from DBObject. It manages unique key attributes for object identifiers (highId, lowId) and time stamping attributes (dtAdded, dtModified). After the creation of a DBObject instance, the current user's login-identifier is recorded in the createdBy attribute. After each modification to any attribute value within a DBObject instance, the current user's login-identifier is recorded in the modifiedBy attribute. In a relational database implementation of the database, this kind of functionality can easily be implemented with triggers.

DBObject also defines attributes for quality assurance (checkedBy, dtChecked, approvedBy, dtApproved), version and release identifiers (version, release) and full traceability to the original object that an object was copied from (releaseBasedOn-attributes). This functionality is described in Chapter 11 - Chapter 13.

### 4.2 Element

DocumentDictionary is conformant<sup>8</sup> with the UML standard, and has adopted its basic superclass structure for common model elements. An Element is a superclass of all classes.

---

<sup>8</sup> Conformant means that model elements/meta-objects in OOCASE have a mapping to UML model elements and carry the same information, however the names of those model elements may differ a little and the structure of the information carrying objects may differ somewhat.

## 4.3 ModelElement

A ModelElement is an abstract class (one without instances) that has a name and a definition attribute.

A well chosen *name* for a ModelElement that is comprehensible for all collaborating people in the development process as well as expected educated end users is an investment that has an enormous payback in saved working hours for communication among collaborating people over the life cycle of the modelled software.

The *definition* attribute is used for defining the purpose of the model element. It's model element specific value is used for generating documentation, help files, comments in generated source code etc. Thus spending effort on a well formulated concentrated definition that is understandable for all involved people that will work with the model element or instances in generated software applications will pay back over decades to come.

## 4.4 LabelModelElement

A Model element that manages the functionality of sorting and structuring of labelled and ordered ModelElements in interactive lists, reports and documents.

## 4.5 NameSpace

A NameSpace inherits ModelElement, and is an abstract container class that owns other ModelElements that are identified by unique names within the NameSpace. Subclasses of NameSpace model program construct in generated source code.

## 4.6 Object

Object inherits NameSpace. Meta-database objects which are created by the developer such as classes and attributes inherit from Object. When working at an international company such as ABB, the PM system domain models have to be coordinated between different sites in different countries. Therefore it is useful to besides *name* that is inherited from ModelElement, have an alternative name and an auxiliary name that can document names in other languages. In our case, we have used name for English names, altName for the Swedish names, and auxName for German names. The English names are used for source code generation. Due to name size restrictions in a wide range of possible target software platforms such as relational databases, GUI languages, CAD system programming languages etc, the English names should have a size less than 27 characters. Sometimes, however, older target languages do not allow such long identifier names and then a shortName can be used instead. The size limitation of names in the meta model database is 127 characters.

The label of an Object is frequently left blank, however in some generated applications it is used as a compact short hand to designate a unique position withing a hierarchical part-of structure in the same way as "4.6" is a short hand to identify this particular section in the DocumentDictionary User Manual.

Note: DocumentDictionary objects for document models are ModelElements. If there is a need to store translations to other languages for certain attributes i a document model, this can be handled with a Profile or Notes.

## 4.7 Package

This abstract class serves as a container for higher level model elements such as a Model or a Module. It is conformant with an UML package.



## 4.8 Model

A model is an abstract class for a self contained model that is handled as one unit. Within a generated application, a Model can for example be saved in a separate file and the abstract Model class serves as place holder for implementing such reusable behavioral functionality.

## 4.9 DocumentDictionary

A DocumentDictionary instance serves as Model root object providing an entry point for the document reference collection. This Model object serves as manager object when a DocumentDictionary model is synchronized with a document dictionary server database.

A DocumentDictionary can store multiple document models, the root objects of document models are represented by the class Category. The reason for that is described below.

## 4.10 A DocumentModel is represented by a Category

DocumentDictionary uses the Category class for representing a DocumentModel for a particular product. DocumentModels have a tendency to become included in larger documentation sets in a supply chain, that in their turn become new DocumentModels until the final end-user documentation is produced. From a software implementation perspective and practical use of document models, makes a floating semantic distinction between a document model and a category depending on if you are a supplier or an assembling customer. Thus avoiding two different classes in the OOCASE DomainModel that drives the software implementation removes some trouble for an end user assembling a new document model with copy/paste, import etc from supplier document models describing individual components or subsystems.

If a DocumentModel needs special attributes, they can be defined in a ProfileClass that extends the Category class. See Chapter 17 Profile Extensions of the MetaModel on page 67.

## 4.11 Category

Category object for building manual indexes over model elements, or generate indexes based on key-word search etc in the names and definition attributes of ModelElements such as ContentRecords.

For DocumentModels, the Category is used for building the structure and organize a larger set of separately maintained and quality assured documents. Where each such document is recorded with its own DocumentRecord in the DocumentModel.

## 4.12 CategoryLink

A CategoryLink links a ModelElement to a Category.

The link should prefferably cache the Name and Definition of the linked memberOfCategories ModelElement, for the sake of filtered loading from large or huge repositories, to allow the user of a local copy created from a filtered model load, quick access to metadata of the components that are missing in the loaded model.

The information in the cached attributes of a link should be treated for the caches they are, thus repository implementations may chose to trigger updates of changes in ModelElement Name and Definition to the cached data in their memberOfCategories CategoryLink objects.

This this will allow filter synchronizing users to get updates on the latest changes at the next time of synchronization.generators.

### **4.13 DocumentRecord**

A record of a document of a particular document type, that serves a purpose in a reference list or for study effort calculations. ModelElement.Definition can be used for the abstract of the document.

### **4.14 ContentRecord**

Represents metadata for a table of contents item within the document. ModelElement.Label is the chapter number, or hierarchically structured section/subsection number. ModelElement.Name is the textual headline of the chapter, section/subsection etc.

### **4.15 ObjectRecord**

The ObjectRecord serves as a high speed cache for interactive browsing of previously absorbed knowledge content. This cache is for private personal use only, or organizational use when the organization has created the information themselves or acquired the right to distribute the material within its organizational borders this way.

### **4.16 AuthorRecord**

The name attribute value of an AuthorRecord is the most frequent name used in the DocumentRecord.authors field. An Author may have a www DocumentRecord, which may own ElectronicEdition objects with URLs.

### **4.17 AuthorDocument**

Designates the contribution of an author to a particular document that is recorded in a DocumentRecord.

### **4.18 Cite**

A Cite corresponds to an entry in the document's reference list. The label is the by document authors chosen identifier used in the document to refer to an entry in the reference list.

### **4.19 ElectronicEdition**

Url to electronic edition (ee) of the document registered in the DocumentRecord. According to analyzed statistics, some dblp documents have up to 8 entries.

ElectronicEdition can be used to point out the relative location of a document in a distributed software delivery archive, for example by setting the url attribute value to file://\$(DocumentDictionary)/doc/DocumentDictionary\_Tutorial\_V1.0.pdf

### **4.20 Note**

A note attached to a DocumentRecord. The XPath mapping of LabelModelElement attributes towards dblp.dtd (2017-08-29) follows: name = (note@label) definition = (note::text()).

This section gave a short description of some central classes and relationships in the domain model of the DocumentDictionary database. More details are available in the OOCASE DomainModel DocumentDictionary1.[0-9].[0-9].[a-z].odm and [Johansson 1996].

## 4.21 Graphical Syntax used in Object Model Diagrams

The meta model of DocumentDictionary is documented in graphical notation for DomainModels are called ObjectModelDiagrams. These are an efficient compact and practical working representation for describing classes, inheritance, attributes, and relationships between classes. They are easy to use in seminar or team discussions and serve as "mind maps" for accessing the details of the DomainModel. The notation is easy to teach team members of all kinds of knowledge domains, including people without programming experience.

In the DomainModel of OOCASE in Figure 3 on page 28, classes have their class name in bold font in their first box. The second box contains the superclass's name preceded by a "->". All attributes and relationships specified for the superclass are inherited by the class. The third box within a class contains a list of attribute names.

The naming convention for relationships is "<name2to1>\_<name1to2>".

Relationships between classes have cardinality constraints telling how many instances that must participate in the relationship on each side, specified as intervals "<min>..<max>". A '\*' as <max> denotes infinity. A black diamond on the owner class side of the relationship denotes that instances on the other side belong to that class. An unfilled diamond denotes that the instances on the other side are aggregated by a class, but are "physically" owned by another relationship.

Descriptions of the notation of UML class diagrams can be found through [UML]. See [Fowler 2003] for an introduction to UML.

The design choice of using a name reference like "->Superclass" in the graphical notation of a class in an ObjectModelDiagram to identify the superclass of a class, is based on plain practical working experience with large and complex object model diagrams.

In a standard UML Class Diagram an inheritance link is represented by a graphical wire with a unfilled pyramid symbol that connects the superclass with its subclasses.

The amount of manual working time lost with graphical re-routing of inheritance links while incrementally evolving a DomainModel with 20 or more classes, is the rationality behind the design choice of this graphical syntax.

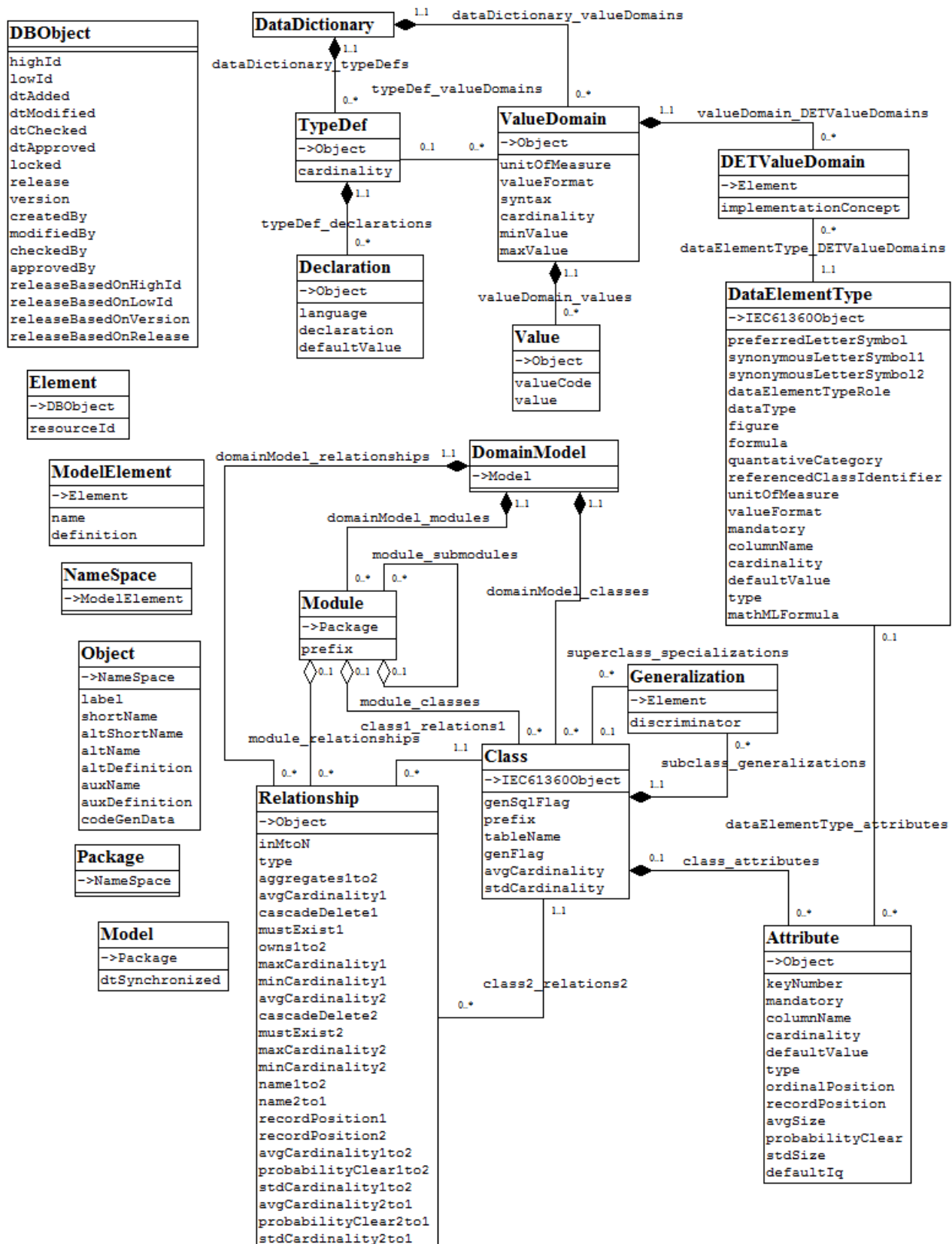


Figure 3. Essential structures and classes in the OOCASE domain model. See 4.21 for an explanation of the diagram syntax

## **Chapter 5 Document Model Development**

This chapter will be written on customer demand.

# Chapter 6 Document Source Code Generation

*Note: This functionality will be released on customer demand.*

*Delegating a laboursome limited task to the most trustworthy and efficient servants available, is a skill that develops with experience*  
*[The Poet]*

A document source code generator converts a DocumentModel into document source code or binary code the way a Compiler converts a source code into executable machine code.

The input to this step is a DocumentModel that has been checked to fulfill the ModelingRules for the particular target language(s) for the source code generation process. Support for automated checking is described in 12.1 Checking and Automated Checks.

Chapter 9 and 10 on page 61-76, 80-81 in [Johansson 1996] give an introduction to the high-level concepts and benefits of source code generation.

This chapter focuses on the more practical aspects.

## 6.1 Overall workflow

- 1) Check the DocumentModel for well known errors. Automated checks are available for that.
- 2) Once the DocumentModel is error free, save a copy of the DocumentModel with a unique name, so you know the origin of the generated document source code.
- 3) Generate the document source code, or have it generated through a document source code generation service.
- 4) Load the generated document source code into your document development environment and test it.
- 5) If all tests pass, ship the zipped directory structure through the automated distribution service that supports your contracted customers and suppliers.

## 6.2 DocumentDictionary built-in Document Source Code Generators

The DocumentModel Window in DocumentDictionary provides a Tools->Generate menu. Here a target language can be selected, and then a generation definition chosen that contains specific parameters for the built-in source code generator.

One way to investigate these is to load a simple well known DocumentDictionary, for example one in the samples/dd subdirectory and just test to generate document source code, and inspect the generated code.

The generation definition files (suffixed by .def) are stored in the Generate/<language>/dd subdirectory in the DocumentDictionary installation directory. Their structure is according to Chapter 8.

The menu Tools->Administration->Edit Generate Definition opens a generate definition file in a text editor.

## 6.3 Document Source Code Generation Services

This kind of service is typically handled by an implementation software specialist. The information necessary for ordering generated documentation source code is the information contents of a DocumentDictionary (.ddi file), and an identifier of the type of document source code ordered.

## 6.4 SQL based Document Source Code Generators

The principles for these are described in Chapter 9 and 10 in [Johansson 1996].

The data dictionary query program **ddq** takes an SQL-batch as standard input, and prints the generated source code to standard output: Example

```
ddq < generate/$(SOURCE_CODE_GENERATOR_FILENAME) > $(DIRECTORY)/$(SOURCE_CODE_FILENAME)
```

Before **ddq** is called, some environment variables have to be set, that tell **ddq** what ODBC data source to connect to, for executing the SQL-commands in the SQL-batch. The ODBC data source is the same DocumentDictionary database that you synchronize your DocumentDictionary with. The environment variables are set in an init-script which is called before any source code is generated.

Typically the repeatable source code generation process of a complete application, or piece of application functionality, is automated in a batch file that calls **make** with different declared targets in the order they should be striped<sup>9</sup> to the output source code files.

The makefile may reside in the source code generation directory, or be available in a path to a make library directory that is set up in the init-script.

## 6.5 Organizing Document Source Code Generator Build Systems

The directory structure of a source code generation system for a particular application is organized into a source code generation directory (SCG-directory). The description below contains variables represented by \$(<variable name>), in line with standard unix shell scripts.

Examples of variable values:

TARGETLANGUAGE = (any of the below explained values, or a custom designations)

html : Html web-site that can be places in a standard web-server under /var/www

pdf : Directorystructure with Adobe Portable document Format (pdf) files

sql : SQL scripts with insert statements to standard web-server application relational databases

txt : Directory stucture with standard 80 character width formatted text files

---

<sup>9</sup> The concept of striping comes from parallell processing where a parameter-tuned task is allocated to individual processors that do the same task, with their specific parameters. In source code generation, typically each Class in the DomainModel that needs to have a class adapted implementation of a standard API, whose source code can be generated with the "algorithm" implemented in the source code generator, adds their band or stripe of compileable source code functionality to the source code file that after compilation by an optimizing compiler implements that class's specific behavior in machine code on the target hardware platform. The order is important due to dependencies that the target source code compiler may have. Some compilers fail if the stream of source code has not supplied them with all definitions they need to interpret new source code coming in from the sequential stream originating from the generated source code file.

A document source code generation system may call other text input to output application programs such as XSLT translators, pdf generators etc.

See the OOCASE User Manual and source code generation products for more details.

## **6.6 Using GIT for Document Source Code Generator Maintenance**

GIT is a powerful directory structured source code version management system. Together with a suitable front-end such as TortoiseGit, it makes geographically distributed collaborative development and maintenance of source code generation libraries much easier.

## **6.7 Quality Assuring Document Source Code Generators with the Benchmark Document Model**

The benchmark DocumentModel covers the full design space of the DomainModel of DocumentDictionary. It can be used to test alternative software implementations of DocumentDictionary. Given that the full range of standard functionality works on a particular hardware software stack implementation, realistic production volumes of test data according to targeted statistical distribution configurations for sets of performance tests can be generated automatically according to the principles described in Chapter 13 of [Johansson 1996]. The performance results of the benchmarks can be used to device pricing strategies for DocumentDictionary software products on different markets, given the conditions of software engineering educated people in a redistribution chain, present internet infrastructure suppliers' chosen software platforms and payment capability of end users on a particular targeted market.

## **6.8 Test Suites for Document Source Code Generators**

When developing source code generators for a new electronically transmittable document dictionary data representation format, it is convenient to use the benchmark DocumentModel as test-application. The test suite for the new representation format should cover the full range of functionality for handling the use cases for display and interactive navigation of all instances within the benchmark DocumentModel.

The testsuites, once a developer is familiar with its implementation for one particular target representation format, are then rather easy to migrate to other target representation formats.

The choice of market strategy depends on the current legislative and factual behavior of actors on the targeted market, and as long as the content within the distributed document models delivers added value capability to the end users on that market, the choice of strategy is a matter of knowing the local market conditions to create and maintain a long-term sustainable business.



## Chapter 7 Prototype Iteration

*Being a listened-to and influential participant in creating your shared future with others, is the best protection for its sustainability, since the genius of a learning and evolving team frequently excels the genius of an individual or a for granted taken cultural, traditional or bureaucratic past.*

*[The Poet]*

A software application development project is an iterative process. The DomainModel with its DocumentModel examples is the Information model that specifies what information structures the application shall be able to serve the users with.

Since applications that deliver any added value compared to what already is available by standard applications without the need for a development project, such information models tend to be rather complex. It is very difficult to for human beings to in their imagination in detail foresee what they like the application to do and serve them with. Thus developing a prototype system, that users can experiment with, and with the aid of the prototype formulate their requirements and wishes was observed to be a much more efficient way of delivering a production system the users wanted and adopted into their work procedures.

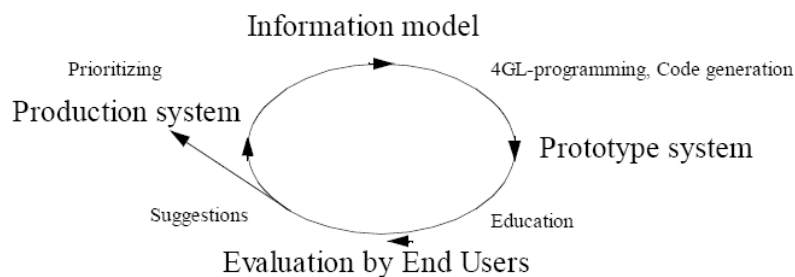


Figure 7.1 Iterative application development process

The purpose of building a prototype system is to receive feedback from end users. In most cases a new prototype must be delivered with an education for the end users who will evaluate it. Once they know how to operate the user interface of the application, they can start building their information structures and fill these with real business information. During that process they will discover if some information creation capabilities are missing in the DomainModel, if they like the user interface, and perhaps if they need convenience functionality, such as copying whole structures or alternative ways of having the information presented to them while pursuing their tasks etc.

The more feedback, the better the next prototype system can become. This off-course requires users to actively work with the prototype and record or explain their feedback so their suggestions can be collected and implemented in the next version of the Information model and user interface of the next prototype system.

If the amount of feedback is large, it might be helpful to organize it and present it to the whole group of end users and other project stakeholders, in order to under discussions identify what is most important in a priority order, and if there are conflicts of opinion.

If conflicts of opinion are NOT IDENTIFIED, discussed and resolved at a solution that all stakeholders can accept, it may lead to unnecessary iteration loops, where functionality is altered forth and back between the wishes of camps of conflicting opinions. Putting conflicting issues at the bottom of the priority list, allows resources to be put on delivering "most value first" when the next prototype system version is implemented.

## Chapter 8 Configuration Files

*On some minor things, that allows growing value to be built upon it, someone has to decide.  
Once the decision is made, everyone accepts it, and nobody cares since the problem is solved.  
[The Poet]*

DocumentDictionary uses configuration files for import, export, automated checks, code generation, interactive sql queries and user role configuration.

### 8.1 Directory Structure

The directory structure for configuration files follow this pattern:

`<action>/<format>/<model class prefix>/`

For example the export definitions for CSV text for DomainModels are stored in:  
`Export/Txt/dm/`

### 8.2 Text format

The general format is close to windows .ini files with the following syntax:

```
[<section 1 name>]
<entry 1 name> = <entry 1 value>
<entry 2 name> = <entry 2 value>
...
[<section 2 name>]
...
```

The first line of a configuration file contains a heading that is displayed in the user interface as guidance for selecting a particular configuration.

When issuing an action in DocumentDictionary, for example File->Export->Text in the DocumentModel Window, all configuration files in the directory `export/txt/dm` are read, and the first line in each file is displayed in a dialog box for selection.

### 8.3 Macro expansion \$(<parameter name>)

To ease the maintenance of configuration files, a very simple form of macro expansion is available.

It follows the make file standard where a macro in the text is referenced with `$(<macro name>)` and expanded with the string contents of the macro.

### 8.3.1 System Macros

The following macros are assigned automatically when reading a configuration file:

#### 8.3.1.1 `$(firstLine)`

This macro is replaced by the first text line in the configuration file. That is convenient for setting the title of interactive reports.

#### 8.3.1.2 `$(firstLineBeforeColon)`

Is replaced by the text on the first line in the file up to the first appearing colon. This is used for identifiers, for instance the login identifier of an automated checker, to stamp as quality assurance mark on an object (if the checker had no complaints).

#### 8.3.1.3 `$(filename)`

Is replaced by the name of the configuration file. Useful in copy-edit-test cycles with several alternative configuration files.

#### 8.3.1.4 `$(filenameWithoutExtension)`

Is replaced by the name of the configuration file, except its extension which may be `.def` or `.mcd`. This can be used for example in:

```
[CONFIG]
```

```
REPORT_FILE = check$(filenameWithoutExtension).txt
```

### 8.3.2 Selection Macros

Some configuration supported operations such as the Report -> Query -> Run SQL Query..., take objects that are interactively selected in the user interface as parameter input.

Attribute values of selected objects can be accessed in the configuration file with:

`$<attributeName><selectionIndex>`

for example:

`$HighId1`

`$LowId1`

where `<selectionIndex>` represent the position in the current selection. The first object has selection receives index 1, the second 2 etc. Thus preconfigured interactive SQL queries to the repository database can be parameterised with attribute values from the current selection.

## Chapter 9 Import of Document Models

*If you carry information in a language I don't speak,  
and I have within translation, your value is mine to keep.  
[The Poet]*

Authors comment: Well you annoying genius, you have to study the copyright laws before your all-knowing wisdom hit's the ground on this planet and destroys our economy.

*True Life is given solely, to the ones who live in trust.  
If life requires honesty , so be it, since it must.  
[The Poet]*

Hrm, the following sections describe various methods and formats to import DocumentModels into DocumentDictionary. These have primarily been chosen by practical needs during the use and development of DocumentDictionary. Thus the import functionality shall be regarded as a preview, where customer project needs will guide the future development of the import in coming releases.

The way distributed parallel software development life all over the planet works today, is that frequently the knowledge barrier<sup>10</sup> and budget of a development team limits the options for what standards and platform-product-supplier-dialects<sup>11</sup> they can support.

### 9.1 BibTeX

A BibTeX record may look like this:

---

<sup>10</sup> Some stuff requires indepth self-studies or a well-designed education to understand. People who have not done the studies or taken a course that makes them understand a particular knowledge domain can not understand what is communicated electronically, discussed or talked about since they have not learned the language that is used for communication and how things interact and work withing a particular knowledge domain. There is nothing but studying that can do anything about the knowledge barrier. Studying builds the biological neural hardware in the human brain of the student that enables a student to understand a subject. Without that biological neural hardware, a human brain will not be able to make any meaningful interpretation of signals or data messages entering into it with regards to that particular subject. That's just a fact from the laws of physics and meaningless to complain about or rebella against. Being a student and life-long student is the most long-term emotionally rewarding activity anyone can engage in, since studying is biologically promoted with emotional rewards in the form of interest and insights.

<sup>11</sup> The teachings of the advocates of "customer lock-in" economy's "historically momentary" monopoly economic benefits, off course has delivered a plethora of diversification of syntax and mechanisms for communicating the same semantic information in a, as big plethora of variations as the singing-variants of birds and wing patterns of butterflies. These syntactical variants are off-course no help to the united human manufactured distributed organism to cost-efficiently fix global warming and preserv the natural ecosystems. However the laws of life apparently governs uncoordinated parallel processes, so we need to ensure we massproduce knowledgeable solvent customers who can pay for the efforts of with science outlining and then with science hammer out the shape and implementation of this high-level efficient self-sustaining organism whose goal is to keep our planet at a medium temperature that serves us as human beings and ensure our quality of life, the way the best knowledge on this planet can guide us, if we give it efficient communication capabilities, and ensure there are enough people who can make good use of that information in actions delivering implementation.

```
@book{Dragon-book,
  author = "Alfred V. Aho and Ravi Sethi and Jeffrey D. Ullman",
  title = "Compilers: Principles, Techniques and Tools",
  publisher = "Addison-Wesley",
  year = 1986
}
```

Import of BibTeX records and reference lists is done by copying the record into the operating system clipboard, select a Category or a DocumentRecord in the DocumentDictionary window and issue the menu command "Edit->Paste External Clipboard", or right-click on a Category or DocumentRecord and select "Paste External Clipboard" from the back-ground pop-up menu.

This is a convenient way to copy/paste BibTeX records from public sources on the internet, like dblp.org.

Other techniques to import libraries of BibTeX records can be added on customer demand.

## 9.2 RDF

This section will be written on customer demand.

## 9.3 RIS

This section will be written on customer demand.

## 9.4 ODBC

The Microsoft Open Database Connectivity (ODBC) interface is a C programming language interface that makes it possible for applications to access data from a variety of database management systems (DBMSs).

ODBC is a low-level, high-performance interface that is designed specifically for relational data stores. DocumentDictionary can use these API calls to import data from databases that are set up with an ODBC data source on the same computer as the DocumentDictionary client application is running.

In a Windows operating system you can administrate ODBC data sources with the program odbcad32.exe.

The rest of this section will be written on customer demand.

## 9.5 TEXT

File->Import->Text, enables import of text data arranged in a standard comma, or tab separated table text format, with one text file for each class of objects in DocumentDictionary, e.g. domainmodel.txt, class.txt, attribute.txt etc.

For each such file, e.g. class.txt, each row in this text file corresponds to one instance of a class in DocumentDictionary. Each value in a column in the table corresponds to a particular attribute's value for the class instantiated on the particular row. The imported text files may have column headings, that determine a mapping of values' positions to attributes.

A text import is configured with an import definition, following the syntax of Chapter 8. In those files, each class in a DomainModel has a corresponding section, that lists what attributes are expected to appear in the text file, and what column headings they are mapped to in the file to be imported.

To configure a new type of text import, first use File->Import->Create Text Import Definition. A number of questions will be asked about the format of the text data to be imported, and a default import configuration generated from the metamodel of the type of model to import, e.g. DocumentDictionary. This configuration can then be adjusted in a text editor. It's [CONFIG] section provides some general configuration settings.

Each metaclass in the DocumentDictionary DomainModel has a separate section in the import definition file that lists all available attributes within that metaclass, prefixed by "att" or "rel". The "rel" prefixed columns are pairs of *relational database foreign key attributes* that serve as a (rel<name2to1>HighId, rel<name2to1>LowId) pointer to a row identified by the corresponding *relational database primary key* (attHighId, attLowId) in another table and thus implements a relationship between individual rows in the imported model.

If the need to import text occurs, it is easiest to study some examples of the import configuration files stored in import\txt\dd.

### 9.5.1 Model Migration between versions of DocumentDictionary

In case a new version of the DocumentDictionary program has changes in the DomainModel of DocumentDictionary the binary formats are no longer compatible and the models need to be migrated to the new DomainModel storage format.

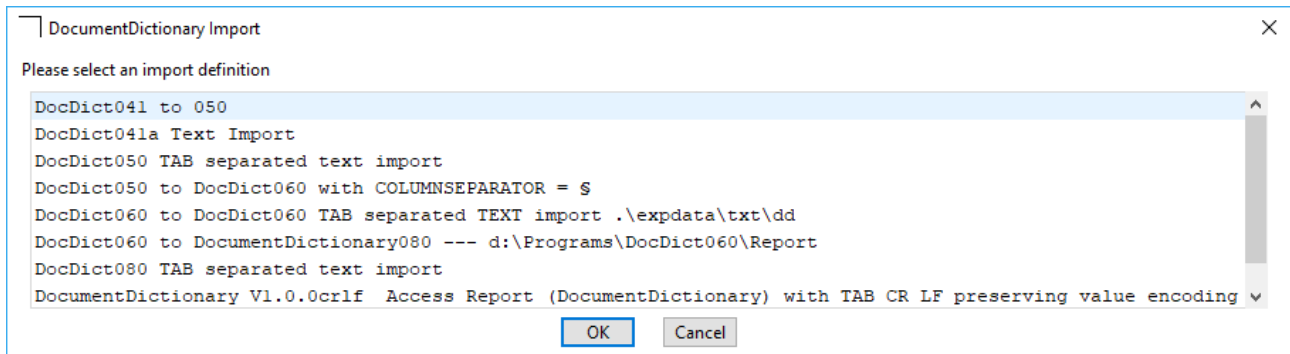
For migration of DocumentDictionary models to new versions of DocumentDictionary, proceed as follows:

- 1) Start the old version of DocumentDictionary and load the DocumentDictionary model you want to migrate.
- 2) Export the model in text format into a separate directory for each exported model. A practical directory location convention for migrated models is to create a directory Report/dd/<DocumentDictionary.name>. Use File->Export->Text, select the latest DocumentDictionary export definition, create the new directory and press Choose. Repeat opening and exporting until all models are exported into their own directories.
- 3) Start the new version of DocumentDictionary. Use File->Import->Text, Select the Import Definition of the previous version of DocumentDictionary, select the directory of a previously exported model. Save the model with a new name in the samples/dd/ directory, perhaps under a subdirectory named after the previous version. If there are errors reported during the import where the cause is not a mistake, a quick way to fix them is frequently to edit the <class>.txt files directly.
- 4) Use Tools->Import->Delete empty string default values. This will replace all imported empty string attribute values with NULL values in the model. NULL values have no representation in universally exchangeable standard tab-separated text files, and are thus represented with empty strings in these. Since an imported model usually never has been synchronized with a documentdictionary repository of the new version of DocumentDictionary before, it is recommended to preserve the dtModified timestamps to preserve the timestamp when information of an object was actually changed.
- 5) Save the model.
- 6) Perhaps run File->Quality Assurance <sup>12</sup>->Check Model, to determine the quality of the imported model.

Most releases of DocumentDictionary have predefined import text definitions from previous releases of DocumentDictionary. Thus File->Import will display import definitions like this:

---

<sup>12</sup> The Quality Assurance principles acquired from best practices on decade(s) lifetime products within Mechanical Engineering Industry and adapted to Software Engineering are described in "Chapter 11 Quality Assurance, Version and Release Management" starting on page 46.



### 9.5.2 Model Migration from Legacy Software Tools

The lifecycles of computer hardware and software tick at a much higher rate than the lifecycles of expensive large scale infrastructure. Thus migrating information about the latter from a previous generation IT infrastructure to a new BETTER<sup>13</sup> IT-infrastructure, is significantly easier if conducted through an intermediate format that has a from decades of practical reality battered but still standing scientific core.

Transforming legacy IT document models to the DocumentDictionary text format as an intermediary for building the new document models or migrating the information to a new IT-system<sup>14</sup> is much easier since simple human readable text files are supported on all new BETTER computer and software platforms.

The concepts in the DocumentDictionary DomainModel for structuring information have been discovered and rediscovered in all locations where there was a developing software industry, however with different environmentally adapted terminology.

Most document information stores developed by people whose thinking has been shaped by education from state-of-the-art computer science and state-of-the-art long-term successful software industry evolving under the expansion phase enabled by the mass market for personal computers and workstations during 1980-1992, can easily<sup>15</sup> be transformed to a DocumentDictionary model, perhaps with some Profile providing additional attributes and relationships.

Simple text files and tables have survived longer than any complex structured in a single file representable universal data exchange syntax format, since the latter requires a complex piece of reading software to be ported to the new platform first.

Exporting and importing document models through various declarative- or perhaps procedural programming implementations to a named column flat table text import format is comparatively easy. DocumentDictionary can handle most of those CVS or TAB separated table export formats, that use standard byte sized character sets. In such a format, the characters that represent a column

<sup>13</sup> This has with available and in practice implementable IT-solution options to do and external environmental pressure that may derive it's pressuring goals and methods from sources captured below the knowledge barrier of experienced managers who take their duty seriously, keeping the large context value adding production capability at prime in a difficult environment they for shure like to improve.

<sup>14</sup> A new IT-system should be decade(s) robust and well supported.

<sup>15</sup> The problem is off-course all those document formats that mix fact information with layout information in a way that is not formally specified and deterministic, or have by decades of accumulation of various features evolved into something that is to complex for a human being or team with different responsibilities to grasp in its entiety.

value separation, e.g. a TAB character, and row separation e.g. a CR or LF, or CRLF, must be uniquely used for that purpose only. Thus some exports where column values contain flowing text with tabs and carriage returns or line feed characters, need to be preprocessed by a one-to-one substitution mapping of ascii control characters, to uniquely identifiable string tokens, e.g. that a tab character (ASCII 9) is replaced by the standard string token `&#9;`. Those mappings are configured in the [CONFIG] section of an import- or export configuration.

Once the exported data files are present in a directory, a text import can begin.

In DocumentDictionary use the File->Import->Guess Import Definition. Select an existing import definition that matches the exported format as close as possible. DocumentDictionary will then compare the column data from the Legacy System export and create an import definition from that.

Use File->Import->Text, and select the newly created import definition, choose the import directory that contains the text files and the model will be imported. Diagnostics on the import process are written to a file named import.err, which in case the import generated errors or warnings can be inspected at once.

In case the import needs to be adjusted, edit the import/txt/dd)/\$(filename).def import specification and retry the import until it works.

After the import is completed the new model file is saved in its binary format (.ddi).

Use Tools->Import->Delete empty string default values, according to 4) above.

## 9.6 XML

The Extensible Markup Language (XML) is a subset of SGML<sup>16</sup> that is completely described in [XML 2008]. This, on the web free for download specification, is a HTML-page that when printed fits on 57 A4 pages. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

Due to the need for exchange of complex structured information in a standard single file format, the presence and superior availability of this standard in conjunction with a mass of free or inexpensive tools for parsing and unparsing, made XML a popular widely used standard for all kinds of structured information exchange.

XML has inherited some theoretical fussiness from SGML, for example the undecidability of whether to use an XML-element, e.g. `<Class><HighId>1000</HighId></Class>` for representing a class object with one HighId attribute value, or if to use an XML-attribute `<Class HighId="1000"></Class>`. This opens up a redundant designspace of  $2^n$  combinatorial ways of representing an element of an object of a class that contains  $n$  attributes.

Once this lack of determinism of syntax for the same semantics of in practice used XML exchange started to generate significant problems, this was addressed by the w3c and later on compensated by XML Schema, later called XSD, which is a kind of more precise Document Type Definition (DTD)<sup>17</sup> against which an XML file can

---

<sup>16</sup> SGML, or ISO 8879:1986, "Information processing — Text and office systems — Standard Generalized Markup Language (SGML)", [www.iso.org](http://www.iso.org), is a 155 page syntax specification of a textual markup language whose (somewhat theoretically unpure non-orthogonal) expressive structuring design space and syntax elements still has a significant legacy impact on text based exchange formats today. The purpose of SGML was having a printing format layout independent language for the text of books and documents that authors could read and write in the text editor tools available up to the first half of the 1980's. The key experience of the horrible inefficient physical paper format's for maintaining and distributing the LAW, was probably a trigger to this great advancement on electronic distribution and maintenance of large volumes of important texts. Leading to enormous productivity gains in the publishing industry with impact on society that depend severely on the semantic content of published texts. The reference for SGML is [Goldfarb 1990].

<sup>17</sup> DTD, a document type definition for SGML documents, which is a structured text document defining syntax rules for how to express concepts that follows the discovered and rediscovered meta model pattern of Class, Attribute, and binary Relationship.



be validated. An XML schema serves as a more precise specification of the expected structure of the transfer format, and can be used to report errors if the structure in an XML file does not correspond to the structure specified by the XML Schema document.

### 9.6.1 Import XML

The preconfigured XML import functionality is not part of the DocumentDictionary product, however are a useful demo of how to use vast information resources made available in XML format.

Most XML imports require some manual services to adapt to a particular need. It is there as an indicator that it is possible to import XML data and transform it into a DocumentModel. The present import XML functionality comes from OOCASE and has been used in a number of funded projects that delivered decision support to many decision makers in the form of structured numeric and statistic information. Many times in the form of high-performance interactively navigable and searchable decision support applications, were Microsoft Access databases were a handy distributable format when the data volume was not to large.

The possibility in OOCASE to lay out the information structure of an XML document graphically on an ObjectModelDiagram while displaying statistics about the underlying data volumes in classes, relationships and attributes was indispensable when some projects hit the technological scope of applicability on some technology platform that had to be migrated to BETTER database technology in order to serve it's expanding application requirements.

The Import XML functionality shows that it is possible to reverse engineer meta data structures out of the exponentially explosive redundant design space of XML file representations for the same information content. Analysis of such structures may deliver useful background information on structure and terminology requirements and size measurement data to a design team that wants to develop or discover a, with measurable science supported optimal, DocumentDictionary DomainModel for a non-redundant design space for structured document information on the knowledge level<sup>18</sup>.

---

And model pattern of object/instance, attribute value and pointer/foreign key/link relationship. In a SGML DTD the meta model pattern class level named ELEMENT, ATTLIST and LINK. And model pattern of element, attribute (in the attlist), and special key attributes named IDREF, and #<idref-value> implementing pointers to other key places in the document. Note that the ENTITY concept used in DTD's is macro expansion, and even if frequently used to implement the declaration functionality of a Class, a DTD ENTITY is a name for a template of text.

<sup>18</sup> The **knowledge level** is the abstract level above the lower symbolic level, named and made consciously aware of by publications of Allen Newell in the Computer Science community, and shortly described in [Johansson 1996]. (The lower symbolic level is fequently with mockery nicknamed the syntactic sugar level when it has to include structures. The mockery arises from cause of the for many of the mockers unknown inadequate logistics and speed of biological hardware building of representations of adequeate, at the time being best available scientific knowledge available, in the creators of the syntactic sugar formats, that are optimal with regards to use and learning within the biological hardware of a human brain given a mass market that requires a standard to build a uninteresting but important platform to, with scientifically proveable efficiency, share higher much more value adding information and knowledge exchange upon). The knowledge level is represented in the human brain with for an individual ideosyncratically developed spatio-temporal activation patterin in the neurons of the brain, that allow the prescence of a evolved concept of a symbol focused biologically wired indexed access to other by biological chemical programming grown biological hardware that physically transfers signal pattern of associated concepts of symbols to enter the consciousness, thus giving an instant in itself representation of the concept of a symbol full access to all by learning created concepts that by tilt of learned patterns-of-thinking in parallel operating overlayed spatiotemporal patterns guide the line of thought at a cycling speed of 4-7 Hz, through the core of the brainstem (that may trigger some important chemical mechanisms), limited by the speed of signal transmission in the neural dendtrite and synapse communication paths that depends on the prescence of more expensive chemical "turbo molecules" that may increase the cycling speed up to 14 Hz at an intense adrenaline or "insight-arousal" rush.

If the need arise to import document models or reverse engineer the information structure out of the contents of reasonably sized XML files, DocumentDictionary reuses the OOCASE rule-based implementation, where the rules are stored in an XML-file.

In OOCASE the import rules are derived automatically from the contents of one or several analyzed XML-files. The user however has to configure where different element and attribute data shall be stored in an DocumentDictionary model. Details on how to create XML import rules are described in the OOCASE User Manual [ROJTEC 2018d].

---

A **knowledge level**, which off-course, since we are limited by our sensory interfaces and languages and what computer symbol structures or human spatiotemporal neural patterns they require to be represented when digesting input from the sensory system, needs some symbolic representation to be communicated between humans or machines. Thus some cleverly designed syntax, or for the practical use efficient interaction tools are needed to enable efficient communication of information.

Now the waste of potentially value adding brain-cycles that could be contributing to our shared increase in quality of life, given a proper information distribution logistic free from "uneducated destructive self-interest" that just delivers a sprawl of habitats for the powerful "ants" that know what they know, while we need a foundation that enables the growth of higher level abstractions that are powerful enough to tell us how we shall fix the real problems of our species and it's limited living quarters, makes a scientist angry enough to deliver a verly long footnote that probably nobody will read, so aggression is a bad emotional index that usually just delivers a by biology promoted puff that will go into nothingness, unless someone with in practice useful influential powers takes a note and acts in the right time and place.

Note: This footnote was temporarily emotionally sponsored by some, in afterthought enlightened, over the authors conduct perhaps powerful but influential anomomous people.

However life stands on the present widely distributed infrastructure however inefficient it's architecture is, so to get something done one has to use what is present and available if it delivers any added value. Where the understanding of what added value really is comes from a place above the knowledge barrier of most humans living today, unless we fix the logistics involved in the education problem.

## Chapter 10 Export of Document Models

*<Some poetry once [The Poet] has read the chapter and condensed its essence>*

<Some author comments on the poetry>

Information in its essence is immaterial, however needs some representation to be worked with or serve its purpose. There are plenty of tools with various functionalities that can be used for some job that has to be done on an information model. Thus the ability for a tool to export information into a number of different formats that serve different purposes is fundamental for delivering added value to a team.

A team that is skilled and free to use the best available tools given their resource constraints to reach the goal set out in their project. Where DocumentDictionary focuses on information systems that carry information for decades in a continuously changing environment.

### 10.1 Binary Storage Formats

Binary storage formats are optimized for performance for a particular well understood purpose.

This section will be written on customer demand.

### 10.2 BibTeX

BibTeX files can be exported interactively by selecting the DocumentRecords to export and issue File->Export-BibTeX. A copy of the exported file contents will reside in the clipboard after the export, so you can paste it into another application or text-editor at once.

Details on this section will be written on customer demand.

### 10.3 HTML

HTML export for a web-server site are available with a menu-commands of Report->Web Report. These create a HTML representation of the DocumentDictionary in the webport directory. This can be copy/pasted to a subnode within a standard HTML web server site, typically somewhere below /var/www.

More details on this section will be written on customer demand, however Chapter 6 Document Source Code Generation, is recommended for an approach that scales up and can be mass-deployed and automated independently of manual interaction in the DocumentDictionary application.

### 10.4 RDF

This section will be written on customer demand.

## 10.5 RIS

This section will be written on customer demand.

## 10.6 RTF

The Rich Text Format (RTF) is a standard developed by Microsoft, that is widely used in industry that use the Microsoft operating system.

Exports configurable views of the selected DocumentModel, DocumentRecord or ContentRecord structure into various template based RTF formats.

Furhter details will be written on customer demand.

## 10.7 TEXT

Exports configurable views of the DomainModel in various tabular text formats, where each object in the model gets a corresponding row in a table where the columns in that row contain textual representations for the objects' attribute values.

The theory behind these formats was explained in "9.5 TEXT".

## 10.8 SQL

Similar functionality as in "10.7 TEXT" however the whole configured export goes into one SQL batch file that contains SQL insert statements. Executing this exported sql batch file on the target database will insert the exported model into pre-existing SQL-tables with the table names and column names specified in the export defintion.

This is useful for information distribution through an existing database infrastructure with a DocumentDictionary DomainModel conformant database schema, or for exporting DocumentDictionary domain models to relational databases dedicated to various types of source code generation, or web-server sites.

It also provides a textual file format work-around in case the DocumentDictionary client application can not be directly connected to a database repository with File->Repository->Login. This may be for security reasons, present organizational firewall structures, lack of compatible call level communication libraries between the client and server operating systems and database platforms etc.

In most such cases an encrypted transfer of a zip-compressed exported DocumentDictionary sql-batch will solve any immediate needs.

## 10.9 XML

Allows exports of DocumentDictionary model objects in XML format.

Configuring an XML export requires experiments with real example files to figure out and become clear about all the details about meta level and instance level etc. It is recommended to start with a very simple DocumentModel example with one instance for each class 'Category', 'DocumentRecord' etc whose data should be exported to the XML-file.

For generic export for archiving etc, we recommend the TEXT format, since it is easier to port to different platforms long-term, however XML may provide a shortcut for certain applications.

The XML export configuration possibilities are limited to direct one-to-one mapping between MetaModel Classes in the DocumentDictionary DomainModel and XML elements, attributes and element hiearrchy structure levels.

More specifically with one example:

The DocumentDictionary MetaModel Class " DocumentRecord" maps one-to-one to ONE XML element name. That means you can not map a DocumentRecord instance to different entity names in the exported XML file depending on other properties of that particular "DocumentRecord" instance. For example you can not map a DocumentDictionary "DocumentRecord" instance to either an XML entity <Article> or <Book> in some external XML format, depending on if the instance has an attribute value of 'article' or 'book' in its attribute recordType.

Such XML file transformations of entity names etc can be done in an XSLT post-processing step.

Owned "Relationships" are represented via XML hierarchical element structure.

Referenced "Relationships" are represented with foreign key references by actual foreign key values such as a pair of <<name2to1>HighId> and <<name2to1>LowId> elements, where <name1to2> means the DocumentDictionary Relationship attribute name1to2.

Alternatively referenced "Relationships" are implemented with an XML file uniquely generated id attribute on each element, which is (foreign-key/pointer) referenced by a ref attribute in an element representing the relationship link.

# Chapter 11 Quality Assurance, Version and Release Management

*Quality Assurance carries an Eternal Truth whatever primitive methods used to enforce it*  
[The Poet]

This chapter goes through the theoretical foundation of ONE mechanism that allows creative incremental development to co-exist with the rigidity necessary to efficiently manage information distribution in a massive parallel distributed supply chain network, where many actors are not aware of each others existence. We name it QAEVR, for Quality Assurance with Editions, Versions and Releases, spoken Q-A-ever.

The reason this ONE mechanism was chosen, is that it is possible to efficiently implement it on the presently widely available theoretical core of the relational database platform, while it is much more efficiently implemented on the object-relational or object-oriented platform. The relational database platform has enough live intellectual capital to remain robust for decade spanning life cycles. The preservation of relational databases is a fundamental interest of all modern governments and businesses, since it serves as the heart pumping the information flows necessary to maintain the prosperity of organized human life on this planet.

## 11.1 Purpose of Quality Assurance

**Quality Assurance** is an insurance investment to protect a development effort from unexpected problems and a much higher cost to fix those problems at later stages in development, implementation, assembly and some cases mass deployment.

Safe quality assurance procedures ensure that all predefined checks against previously occurring and thus well know errors are done, and that this is documented in a way that ensures that no changes are done to the checked design object after the checking is complete.

The work flow from design to delivery of objects are roughly:

- 1) Create the object, or copy a previous object and modify the copy
- 2) Check that the object meets the quality requirements necessary for it to serve its function
- 3) Mark the object with quality assurance information
- 4) Release the object to the supply chain

The quality assurance information, if handled correctly, is a huge time and cost saver when the object travels through the supply chain. If such information is not available when the object is received, the customer of the object has to spend time and money on inspection and relevant testing

equipment to do their own quality assurance checks, if they want to be sure that the object will serve its function in the customer's own product.

In longer value-adding supply chains, such avoidable duplicated testing costs may make the total cost of the final assembled product unaffordable for the end user.

## 11.2 Purpose of Version Management

During a design and development process, objects evolve when new functionality is added and errors or problems are corrected. This means that their properties change. In a supply chain, the customer of an object must be able to easily tell different versions of the same design object apart from each other, to ensure that newer versions of the object still serve the same function they are used for in the customer's design. Newer versions of objects may have changed in a way that requires the customer to adapt the design of the product in which the object is used, to make the customer's product deliver its intended functionality.

The identification of different versions of an object is done by attaching a version identifier to the object.

In complex supply chain businesses the concept of "form-fit-and-function" is fundamental. To illustrate this concept, for different versions of **physical objects** to be interchangeable within a particular assembled product, these objects must have a compatible **form** so they have physical space to be placed in the same geometrical position within the assembly. They must **fit** with the mechanical interfaces provided for securing the object in the assembly, e.g. screw holes, snap in tabs or similar. The interfaces towards the environment, for example the nozzles of a water tap or valve, must match the physical shape of the connecting nozzles transporting the water or medium, whose flow rate is controlled by the valve. The different versions of interchangeable objects must also perform the same **function**. E.g. a Valve must shut off fluid when its handle is turned counter clockwise to a particular geometrical position, and similar.

For **immaterial information objects** such as written text for information transfer from human to human, human to machine or machine to machine, similar but abstract form-fit-function properties apply to the object. Properties that make it serve its function within an assembly without enforcing costly changes of the assembly to accommodate it. For example a piece of text, such as a paragraph of legislation, a requirement on a product design, or a program that can be executed by a human or a machine must have a **form**. There are many aspects of form for text. The Language used for expressing its information is one of them. In immaterial products that serve a function, the rationality aspect frequently determines the form. Few textual information objects mix different languages, unless its function is translation.

**Fit** means that the interfaces of the information object must attach correctly to the environment within the assembled product. In text, the syntax, e.g. the rules of how the textual information shall be expressed, must fit into the assembled product. The **function** of the information object must be compatible in a way that does not require the surrounding assembly to be changed, i.e. the environment's methods to make use of the function. This means that the interface used to make use of or call the function within the object must fit with the interface used within the surrounding assembly. A call or use of the function must deliver the same results, however the performance of a later implemented version of a function may differ with magnitudes.

In software industry the concept of Application Programming Interface (API), enabled software components and libraries delivering complex functionality to be reused in an added value chain. A supplier of a software component could sell it to a large market of customers. By that cost-efficient

division of labour an added value was created, that enabled the components to be sold for a fraction of its development cost, given that the number of sold items delivered more income than the cost to create, distribute and sell the software component.

For a **version identifier** to efficiently inform the customer, which may be a human or a machine, if a newer version of a software component can be reused without adapting the surrounding assembly, the version identifier must express backwards compatibility with regards to form-fit-and-function.

### 11.3 Semantic Versioning 2.0.0

Semantic Versioning gathers the distributed software community experience based practice into a standard syntax that enables the form-fit-function criteria to be expressed in a version identifier for a software component [SemVer 2.0.0].

Given a version number MAJOR.MINOR.PATCH, increment the:

1. MAJOR version when you make incompatible API changes,
2. MINOR version when you add functionality in a backwards-compatible manner, and
3. PATCH version when you make backwards-compatible bug fixes.

Additional labels for pre-release and build metadata are available as extensions to the MAJOR.MINOR.PATCH format.

Here you are kindly redirected to the original source document of SemVer, which you can find at: <http://semver.org/spec/v2.0.0.html>.

Reading the original text and storing a copy on your computer for reference is a good investment of your time.

A note on SemVer rule : <http://semver.org/spec/v2.0.0.html#spec-item-1>

*1. Software using Semantic Versioning MUST declare a public API. This API could be declared in the code itself or exist strictly in documentation. However it is done, it should be precise and comprehensive.*

A DomainModel is a specification on a higher abstraction level than an API. It is a declaration and definition of the names to use for generating a source code library that implements an API that follows certain easy to remember naming conventions for different standard information manipulation functions. Thus knowing the names of classes, attributes and relationships in the domain model, AND the naming conventions for the standard object manipulation functions delivered by a source code generator, defines the API.

From that you can infer rules like:

New version of a Class that has been given a new attribute -> increment the minor version identifier.

New version of a Class that has deleted an attribute -> increment the major version identifier, since the generated API to manipulate that now missing attribute is no longer available, and applications using it will not work.



## Chapter 12 Quality Assurance Functionality

*The understanding of what really delivers Quality and how the Added-Value Prosperity is founded on Quality, is pure self-preservation knowledge, and the foundations for morale*  
[The Poet]

The purpose of Quality Assurance was described in section 11.1.

Most creative developers find manual checking of their own or other peoples work booring and time consuming. There are however some exceptions to that rule.

- a) If the developer can learn something new while inspecting the design or code.
- b) If the developer sees the inspection work as a teacher's mission to bring up the level of skills of appreciated students.
- c) Shared group or community professional pride in the software, and a desire to keep it beautiful in terms of clarity and readability, efficiency, robustness and freedom from faults.
- d) Some other reasons that are beyond the scope of this manual

In larger software businesses, the checking effort is frequently delegated to a Test Department, that can hire personel with the appropriate character structure to pay explicit attention to details and get high on hunting and finding faults.

Anyhow, budget restrictions makes such solutions impossible for smaller software companies, thus the checking effort has to be automated as far as possible, and incrementally maintained as new types of errors are discovered.

### 12.1 Checking and Automated Checks

Checking is done by following a check list. The check list is the accumulated knowledge of errors that have occurred earlier, which can be detected by inspection of the design while focusing on particular known and detectable fault conditions.

Checks can be seens as rules attached to a class of objects, which determine whether the values of the properties of an instance of a class passes the check or fails to pass the check.

In a restricted information domain, that can be expressed in a DomainModel<sup>19</sup>, such checks can be automatically derived from the DomainModel and automated.checks generated for a particular application or product model database.

---

<sup>19</sup> Since the self describing DomainModel of OOCASE is implemented in the same language that OOCASE allows other applications to be described with and uses to generate compileable source code with, these other applications can reuse the same checking functionality that OOCASE uses to quality assure its own design.

In OOCASE and applications that build in its common frameworks including DocumentDictionary, checks are declared in Model Check Definition files with the file name extension .mcd. These mcd files are initially generated from the DomainModel of the application, and constitute standard deriveable checks.

Here is an example of an extract from a check definition file illustrating the checks for the name attribute of a Class in OOCASE:

```
2NI : NameIntegrity - ensure unique and valid names in every namespace environent
-- OOCASE model check configuration file.

-- checkXX = E    : Reports an error if the check is violated.
-- checkXX = W    : Reports a warning if the check is violated.
-- checkXX = M    : Reports a message if the check is violated.
-- checkXX = I    : Ignores the check.
-- constXXYY = <value> : Constant parameter YY used by check XX.

...
-- **** Class Check Definitions ****
[Class]
...
checkNameHasFirstLetterConvention = W
constNameCapitalFirstLetter = true
checkNameHasWordSeparationConvention = W
constNameWordSeparationConvention = Capital
checkNameIsSpecified = E
checkNameMatchesRegularExpression = W
constNameMatchesRegularExpression = [A-Za-z][A-Za-z0-9]*
constNameMatchesRegularExpressionMotivation = Restriction to match valid syntax of
variable names of most programming languages targetable for source code generation.
checkNameSize = W
constNameSizeMin = 1
constNameSizeMax = 27
checkNamesUniqueWithinNamespace = E
...
```

MCD files follow the syntax of Configuration files described in Chapter 8.

The first line is the header that the user can see when selecting which model check definition to apply. Then some comments preceeded by a double minus (--) that roughly explains the conventions of the configuration file for a first time reader or user.

Each class in the domainmodel of the application has its own section, starting with a braced class name. e.g. [Class] above, followed by the set of automated checks that can be configured to generate an error, a warning, a message or be ignored in that particular model check definition. The names of the checks are chosen to be self explanatory.

## 12.2 Check Level Structure and Checkpoints

To keep the number of variants of model check definitions down, these are organized into a check level structure that reflects the needs of development phases, sometimes separated by checkpoints in a typical project.

A checkpoint specifies the level of quality a specification much reach before work on the next development phase is allowed to start. The rationality behind checkpoints is to avoid losing work effort on implementing parts of a specification that risks becoming useless or obsolete due to later design changes on a higher level.

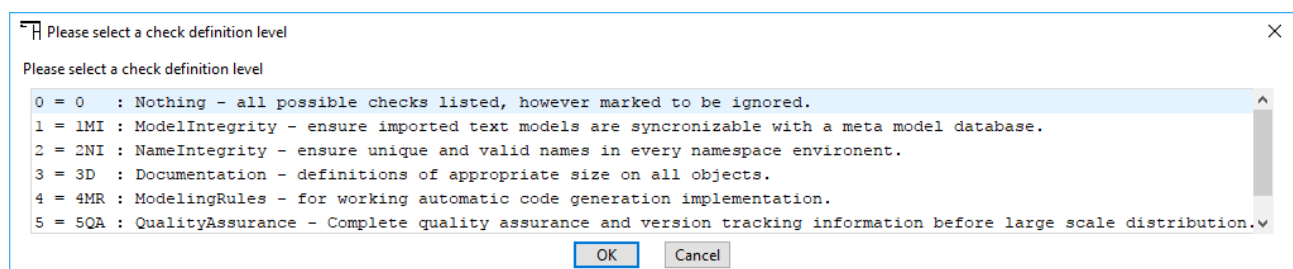
The requirements for passing a checkpoint are there to enforce that higher level more cost impacting designs are complete and fully agreed upon by the development team.

A full understanding of the whole of a design, frequently delivers insights how it can be simplified by refactoring or reuse of frequently occurring structural patterns through the mechanisms of inheritance.

A simpler design that delivers the same functionality that a more complex one does, is easier to implement. In addition, but severely more impacting over a software life cycle, a simpler design is easier to teach and maintain over the software's life time.

Changes in terminology have a severe impact. Frequently the right choice of names for design objects is not apparent until the whole design can be seen in its entirety.

When using, creating or adapting a model check definition, the following baseline check levels are used.



A higher check level number means higher quality. An unchecked design, for example imported from an external system, may require significant work before it lives up to the standards of a cost efficient life cycle of a decade surviving software.

If all possible checks are applied at once, the list of reported errors may become excessively large. A strait-forward approach to just fix the errors in the sequence they are reported in the list leads to a and very boring, inefficient and partly counterproductive use of the designer's time.

The reason is that such a list intermixes small and big issues. A fix of a small issue in the beginning of the list, may be obsoleted when a bigger issue is fixed later in the list.

The format chosen for human identification of check levels covers the user categories ranging from expert to rare occasion users and novices. The format is:

<check level> = <check level mnemonic> : <Check level name> - <Check level concept recall description><sup>20</sup>

The following sections go through the baselines of the chosen check levels, which can be adapted to particular project and implementation needs. This is typically done by creating a .mcd file using the baseline, appending a number or letter to the <check level mnemonic>, and the rest of the human identification information. Then adapting which predefined checks to deliver warnings and errors, and adapting the constants that set the parameters for the checks.

---

<sup>20</sup> This design was dictated by the all-mighty GOD that evolved the human visual system and by that the speed possible to achieve with the biological hardware pipeline available in a human being. If you are interested in that and want to understand the full rationale behind this design choice, you need 2 weeks of uninterrupted 8 hour per day studies of the books in B.

### **12.2.1 0 = 0 : Nothing - all possible checks listed, however marked to be ignored.**

This model check definition is a useful baseline if you just want to run a few specific checks. Starting with a template where all checks are ignored just requires editing lines on the checks that are needed. The checklist for a OOCASE DomainModel contains more than 2000 checks. The checklist for DocumentDictionary more than 1000 checks.

### **12.2.2 1 = 1MI : ModelIntegrity - ensure imported text models are synchronizable with a meta model database.**

Many widely established programming languages including relational database languages require their data definition language to specify the length of text strings and value ranges of integer and floating point numbers. This enables their language compilers to optimise data storage allocation in computer memory, and select the fastest processor machine instructions that will do the actual computations described in the high-level programming language.

Many languages and particularly relational database languages have constructs to prevent storage of complex structured data that violate constraints. The purpose is to prevent entering data that can't possibly contain any valid or useful information.

Such language constructs serve as protection against faulty software and faulty data.

Before a product model can be stored into a particular database implementation, a passed "1MI : ModelIntegrity" check ensures that the data contained in the model actually fits into the data structures provided by the particular database implementation and does not violate any constraints.

In a supply chain, it takes time to change present infrastructure of database and program implementations that serve the information shipments in real time. If a new product model, whatever way it was created, passes this check, the supplier can be sure that it will successfully be delivered through the present installed software infrastructure. If the infrastructure has sufficient quality, which is a distributed responsibility of all actors serving and maintaining it, where many never met or heard of each other, but unite on the shared values exposed by the added value delivered by the design<sup>21</sup>.

### **12.2.3 2 = 2NI : NameIntegrity - ensure unique and valid names in every namespace environment.**

A Name of an object or concept is the primary mechanism used in language to transfer meaning between a sender and a receiver. In a software life cycle that spans decades the choice of names is essential. The meaning of a chosen Name for an object must remain permanent in an environment where people and interfacing softwares come and go while they evolve in their careers and lifecycles.

A well-known problem with Names is that they become overloaded. That is a Name in one particular context does not mean or represent the same thing in another context.

To choose a GOOD name for an object, the designer must have an understanding of the full range of interacting senders and receivers in the environment where the created software is intended to serve a function an purpose.

The choice of names requires interaction and negotiation with experts familiar with all social and technological communities that somehow will interact with the software.

In order to have such an interaction and negotiation, all affected parties must be able to study the design and provide their constructive comments on how to improve it.

---

<sup>21</sup> Something that requires an educated eye to see.

The 2NI check level ensures that there are no machine detectable errors in a proposed design, before its evaluation starts generating the costs of human inspection and feedback.

#### **12.2.4 3 = 3D : Documentation - definitions of appropriate size on all objects.**

To preserve the understanding of what a design object means in a volatile unpredictable environment, the meaning needs to be defined in a way that serves the purpose of the software in its predictable environment, during its life cycle.

Given the overloading problem of names and the unpredictability of information senders and receivers interacting with it during its life cycle, an agreement on a definition that is judged to be understandable by all expected interacting current and future parties is essential.

A definition is written and discussed during the creation of the design, but will be read magnitudes of more times by a huge variety of software users, given that the design's inherent quality delivers that kind of expansion.

The 3D check level ensures that machine detectable errors are not present in a documented design proposal before submitted for evaluation and feedback.

#### **12.2.5 4 = 4MR : ModelingRules - for working automatic code generation implementation.**

This check level has a severe impact on the implementation cost of automated source code generators.

If the designer of a source code generator can assume that the declarative model specification is fault free, the code to capture a huge number of possible fault conditions can be skipped and development efforts be spent on delivering the best possible performance of the generated source code with regards to all predictable situations where the generated source code needs to communicate with humans and machines.

#### **12.2.6 5 = 5QA : QualityAssurance - Complete quality assurance and version tracking information before large scale distribution.**

This check level ensures that valid quality assurance information and version tracking information according to SemVer has been applied to the entire model, to the level possible to detect by a machine.

### **12.3 Recording a passed check**

In an application supporting Quality Assurance, all objects that inherit from DBObject, have the attributes checkedBy and dtChecked. These are used to record who the object was checked by and the date-and-time stamp when the check was done. The purpose is to identify who was responsible for the check. Stamping an object with one's identity gives a justified sense of responsibility. If the check was incorrectly done, this will eventually show itself later in the development process when something goes wrong due to the inadequate check and the responsible person identified and held accountable. Being accountable is a way to ensure that the checking is done properly.

Normally these attributes are write protected, and can only be written while using a special checking functionality to maintain the integrity of the information.

A passed manual check is recorded by storing the login identifier of the user who did the manual check in checkedBy. Login identifiers for humans begin with an alphabetic letter.

Automated checks that are documented by a model check definition file can also be held responsible and accountable. Thus if an object passes an automated check, the check level

mnemonic which serves as identifier of the model check definition is stamped onto the object in `checkedBy`.

This delivers precise information to the user of a checked object what level of quality the object has. Failures later in the process can be traced to this model check definition, and the model check definition be updated to ensure the failure condition is detected by the automated check in the future.

## 12.4 Approving an object

Approval of a design means that a person or design team takes the responsibility for it. This is a matter of efficiency in targeting feedback ranging from error reports, suggestions for improvements, praise, credibility and earned trust.

Usually the manager of the design team is the person that approves a design before it is released. The manager is the person that has the best overall understanding and connectivity into the design process, its organization and infrastructure. Targeting feedback to the manager, is usually the most efficient way to channel it to the correct place within the development organization. An organization that in most distributed supply chains is a complete black box for a customer.

The recording of an approval is done with the attributes `approvedBy` and `dtApproved`. These attributes are also write protected to ensure integrity, and can only be written using special approval functionality.

In most cases the approval functionality provides access to checking information, that on a high level delivers decision support whether to approve the design or not. The function to approve a design is usually executed from the top hierarchical level of the design to be released.

A top manager can delegate the responsibility to approve self containing design components to middle managers in the organization. This is rational when the design is large, and the top manager is responsible for how it all integrates. This delegation however requires that the middle managers have the support an infrastructure to handle incoming feedback efficiently. If this is not the case, the top-manager takes the approval of a design from a trusted middle manager as the quality assurance stamp that it is, and overwrites those stamps with the top-managers own stamp. This way, the "fuss" from the environment can be kept outside the internal organization, and met by people who are skilled in dealing with it on a more frequent basis.

## Chapter 13 Version and Release Management

*How can you be sure that a black box will do what it is supposed to do?  
That is a matter of earned trust delivered by cleverly managed responsibility.  
[The Poet]*

### 13.1 Editions, Versions and Releases

In creativity supporting interactive design tools, unnecessary bureaucratic procedures and restrictions are productivity killers, since they interrupt the designers flow and line of thought with unnecessary details that can be fixed in a later cleanup sweep.

A creativity supporting design tool is as much an aid for thinking and exploring as a tool for documenting the final design.

Thus the approach taken in OOCASE and DocumentDictionary for version and release management is a "non-intrusive" automatic one that goes on in the background without burdening the attention of the designer.

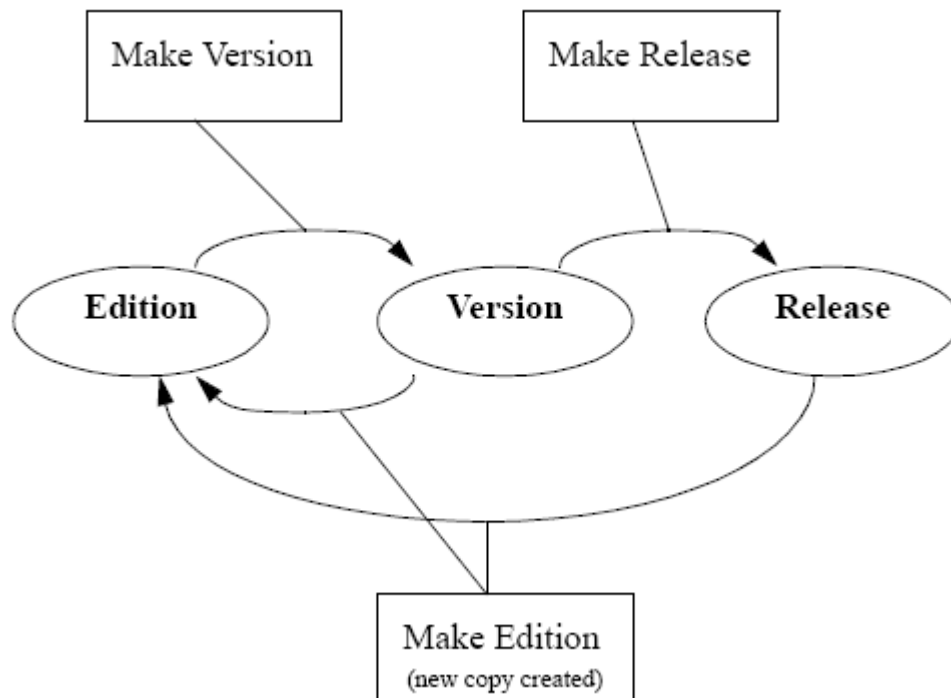


Figure 4. The quality assurance life cycle of objects in the meta model database

All design objects have a quality assurance state depicted with ovals above. The states used here are named **Edition**, **Version** and **Release**. The name of the state **Edition** comes from the word **Edit**.

**Version** means that the object has been quality assured and assigned a version identifier that follows

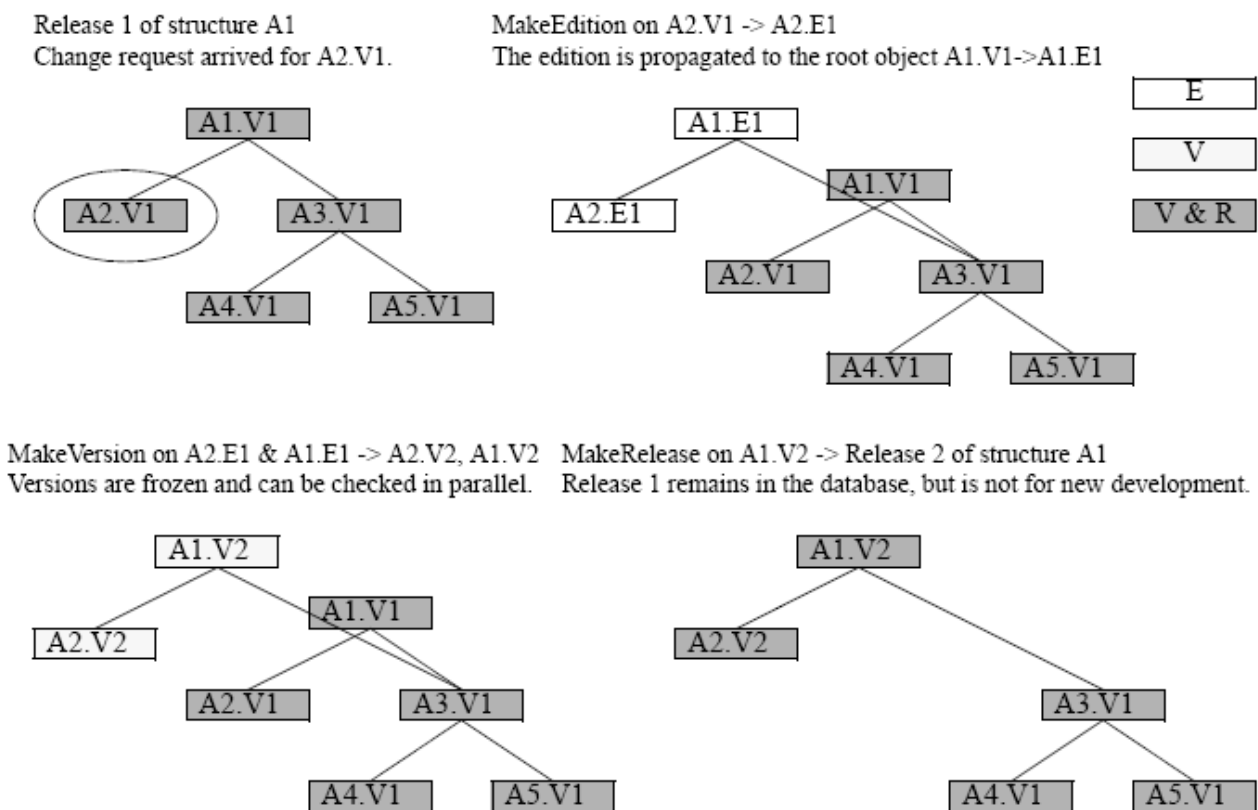
[SemVer 2.0.0]. Release means that the object has been quality assured to a level where it serves a purpose to release it to a supply chain.

Most productive design work start with a copy of a model that is similar to the one the designer wants to create.

In OOCASE and DocumentDictionary a newly created or modified object that has not been quality assured yet automatically receives the state Edition or transits into the Edition state via the automatic process MakeEdition. With regards to the whole model that the object participates in as a part, this edited object is a new object version, distinct from the object version it originates from, thus needs to be uniquely identified. The version identifier automatically assigned must be unique and meaningful to a human being in order to be easy to work with in human thought processes. Time is unique and meaningful, thus the time when the object transits from the Version or Release state into the Edition state is used to uniquely identify the Edition.

When quality assurance<sup>22</sup> has been completed for an object a new version can be made. This is the Make Version process<sup>23</sup>.

Figure 5 illustrates the quality assurance life cycle of a complex object A1. E denotes that the quality assurance state is Edition. V that it is Version, and V & R that the version has entered the Released state.



<sup>22</sup> This means checking on a level that serves a purpose for the development phase that the design currently is in. For complex designs that need coordination of feedback amongs many actors, quality assurance also includes approval by the manager or organizational unit that is responsible for the version of the design.

<sup>23</sup> The Make Version process needs to be tailored depending on design phase and evolved best practice within the development organisation around the designs that can be expressed by instances of the domain model. This may involve procedure descriptions and rules of thumb how to allocate version identifiers, depending on the communication and information distribution needs of the development process.



Figure 5. Example of quality assurance process for complex objects<sup>24</sup>

In DocumentDictionary, the act of editing an object will automatically propagate the MakeEdition upwards the hierarchical assembly structure up to the root object in the structure.

When the designer is confident with the new design, automated checking can be applied where all objects in the chosen assembly structure that fail the checks are listed and can be corrected interactively at once.

Assigning versions only makes sense when the model is shared in a distributed team, or when branching of the development effort is necessary. E.g. keeping a quality secured "fallback" version if the latest development increment fails to meet its deadline or turns out to be a "dead-end".

Assigning practically useful version identifiers to large object structures is a bulk job that needs to be guided by rationality, in order to serve its purpose. What is rational depends on how the design and its components are used individually upwards the supply chain.

The purpose of releasing a model is distribution to the supply chain. When relational databases are used, several releases must be able to reside in the same database. This would lead to object identifier collisions if the same object identifiers are used as in the previous release. This is handled by a mechanism that reassigns new unique object identifiers to all objects in the model while ensuring that referential integrity is maintained in objects that use object identifiers as pointers to other objects.

## 13.2 Version History

In a software supply chain network where the design output of one supplier node in that network is used downstream in a distributed organization working in parallel, there is a need to track the delta of changes between software releases.

When a new release arrives at a node downstreams, that node needs to evaluate how design changes, new functionality and bug fixes impact their own added value software product.

Decisions must be made on what makes sense to support for their own customers up the value chain and a benefit/cost analysis with regards to the customer value of including those changes. After that a development project can be layed out in priority order with nessecary architectural basis first, than features and bug fixes in benefit/cost order.

In OOCASE and DocumentDictionary the version tracking information is maintained by the automatic MakeEdition process that is triggered as soon as an object in the state of Verion or Release is edited. Before the edit is applied, this process stores the unique object identifier (highid, lowid) and the version and release attributes in releaseBasedOn attributes named releaseBasedOnHighId, releaseBasedOnLowId, releaseBasedOnVersion and releaseBasedOnRelease and replaces the version attribute with it's new edition identifier and sets the release attribute to null.

Thus a reference to the original for the edited object is stored in the object itself. This information can be used to trace an object back to its original, regardless which distributed copy of the original is used, independent of any software implementation.

---

<sup>24</sup> To roughlyly contretize this a bit. In the context of an OOCASE domain model, A1 can be a domain model object. A2 and A3 classes and A4 and A5 attributes. In the context of a DocumentDictionary domain model, A1 can be a document dictionary object, A2 and A3 document records and A4 and A5 content records.

The object is truly immaterial, and can exist in parallel wherever it is stored. It can be uniquely identified and version traced without a centralized repository.

A customer can edit a model from the supply chain for internal business purposes without any need to create new releases of it, unless the customers they in turn supply require access to an accurate model that describes the product they purchase.

In this case the customer of the original release needs to quality assure their own modified version of the object, assign a suitable SemVer version identifier that serves its purpose for their own customers in turn, and make a release.

To preserve the properties of a truly immaterial version traceable object, each customer in the supply chain must generate their own globally unique identifiers for their new release. This is ensured by a centralized ledger of serial number series that each customer in the total supply chain network receive their unique highId identifier from.

Thus if supplier A denotes their model V1.2.3, customer B who buys it and uses it as a component to efficiently build added value for their own customers, can have their own SemVer compatible version assignments, and since the unique highId of supplier A, and unique highId of supplier B that is used when they release their added value versions of a model, identifies these version identifiers as different since they origin from different suppliers.

The economy of the collective human mind is to be considered when assigning version identifiers. HTML 4.0 has a distinct meaning to a huge user base. SemVer allows the supply chain to append their own version identifiers on that from their supplier as long as they follow the rules of SemVer. It is frequently worth the while to collaborate with the supplier if the value-adding adaptation of their product is non-backwards compatible.

This section gave a view on supply chain perspective with regards to version history. This view is orthogonal to the topic in the next section, which deals with the part-of structure of a model.

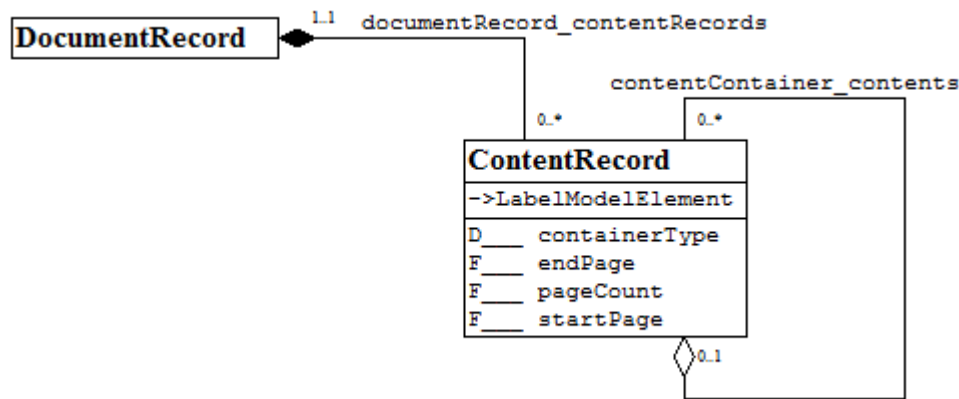
### **13.3 The VersionOwnerPath**

The version owner path is used to determine the hierarchical part-of structure of objects in a model assembly when version identifiers are bulk assigned top-down in the hierarchy of an edited structure. All objects must know how to determine who they immediately belong to in the assembly structure, and by that be able to derive their unique path to the root object in the model.

One such structure within a DomainModel is the Module hierarchy, where Classes and Relationships distinctly belong to a single Module, and a Module may be be a part of another module etc. This is similar to the Package hierarchy in many programming languages and in UML.

#### **13.3.1 Design Rules for Recursive Hierarchical Relationships**

Objects in a Model may be organized into recursive hierarchies. For example a ContentRecord that represents a chapter in the table of contents in a document may contain sections which in turn may contain subsections etc, that are all instances of the same class ContentRecord.



In relational databases and fixed file format storages, large such recursive hierarchies cause a tremendous performance loss while loaded and stored if the nesting has to be traversed and built while loading<sup>25</sup>.

Thus the following design rules apply.

When a recursive hierarchy is needed for a Class, all objects of that class must have ONE true part-of owner relationship denoted by a filled diamond on the owner side in UML.

This way the whole collection of owned parts can be loaded or stored in one single table or record scan.

The recursive structures are "simulated" by one or several aggregated 1-N relationships denoted by an unfilled diamond in UML. These nested structures are reconstructed with direct pointers in primary memory after all objects have been loaded from the external data source and have been indexed by their object identifier and/or primary key.

### 13.3.2 Direct Aggregators

When an object is edited, the MakeEdition must be applied on the edited object all the way up to the root of the assembly structure the way the users perceives the assembly structure. For a true owner relationship (black diamond at the owner side), it is strait forward that an object has been edited and become a new version when one of the components it owns has been edited.

For recursive hierarchical relationships, such as 'contentContainer\_contents', if a subsection is edited, then the section becomes a new version, and the chapter that contains this section, not just the DocumentRecord that is the true owner of the subsection, since the user perceives the Chapter as containing the section and subsection.

<sup>25</sup> If you are an expert with at least 25-30 years in database industry, you know that this problem is elegantly solved by database implementations that provide sophisticated query optimizers or elegant use of hardware support for virtual memory in a way that makes this unnecessary. Such implemtentations are however beyond the purchase capacity and performance needs of the masses we need to mobilize to feed the computational centers with facts. The organizational and educational investment needed to put such advanced database implementations into profitable productive use are some career steps above the basic added value we can deliver here. Standing on our delivery, it is much easier to take on challenges upwards the added value chain. Since you yourself are probably forced by reality to become a manager by now, you must ask yourself what people you would like to hire. On what level do you want to start building the added value knowledge necessary to for an new appreciated co-worker to become productive in the use of your products? Someone who can see the difference between a good design and a brilliant one? You know our history, and we need eyes in our young generation that can see. The end-state calculation along the current trajectory is very important to disperse to influential people who can't see but have all potential to develop their sight. We can fix it. It's just a bug. An attitude that fixers need to have, in order to mobilize the best parts of the legacy we are made of.

Thus the following rules apply:

- 1) If an object is and has an aggregator on the white diamond side of a recursive relationship, then MakeEdition will trigger that aggregator to be a part of the new edition.
- 2) If an object has no aggregator on the white diamond side of a recursive relationship, it means that this is a root object in the recursive aggregation hierarchy that is owned by the true owner relationship only. Thus being the root in the aggregating recursive relationship it is this object that propagates the MakeEdition to the shared true owner.
- 3) To maintain the integrity of the version owner path, the implementation must ensure that all objects in a recursive aggregating relationship that are connected in a hierarchical assembly structure share the same true owner<sup>26</sup>. *(For example if the user interface provides the functionality to move (by drag/drop) a ContentRecord representing a section from a table-of-contents structure from DocumentRecord A to a table-of-contents structure in DocumentRecord B, the true owner of this ContentRecord must also be changed to DocumentRecord B.)*

---

<sup>26</sup> For deeper more complex assembly hierarchies that may involve recursive version owner path loops where an object of class A may be the true owner of an object of class B, which in turn may aggregate other objects of class A, the shared direct or indirect owner that has no aggregators must be the same object. The in theory and practice test is to ensure a design that if an object that participates in recursive relationships is deleted, it should leave no dangling aggregated objects behind that are not deleted straight down forward through its true part-of recursive ownerships. Implementations of drag-drop moves of assembly structures where the object dragged is aggregated, need to ensure the shift of true part-of owner.

# Chapter 14 Using a Relational Database for DocumentModel Sharing and Distribution

*<Well The Poet is eager to tell the Poet's Story>*

Author's comment: I'm completely baffled. You told you that?

Back to business again. All information need a data representation that efficiently serves a number of functions. Roughly, excluding some that we can not cover in this manual, these functions are the following:

- 1) Persistent storage of information over time.
- 2) Retrieval of information in a format that serves the purpose of the task that is querying the information storage.
- 3) Integrity protection of the stored information against known information corruption fault cases. These are typically coming from corrupt or malwritten programmes that enter data into the persistent storage that violates fundamental information integrity constraints.

The theory of Relational Databases dates back to pioneering work done by IBM in the late 1960's. After a number of landmark publications the scientifically educated workforce interested in relational database theory grew significantly, and with it came successive series of research prototypes that evolved into commercial database products.

Practical experience of using these products for high-volume transactions on large volumes of information stored into tables, feed back input to further theory development that lead to higher performance and better methods for protecting the integrity of the information. These were for example declarative specifications of constraints that the database engine itself must uphold with automated implementations, that ensured that faulty information could never be entered, and thus the problem was delegated to the source of the faulty information that received error messages, without any harm being done to the stored information.

There are many in the public domain available high-quality software implementations of Relational Databases.

Even if the interactive performance is higher of some other types of open-source software for sharing and distribution of information, the integrity protection aspect of such implementations is weaker. Their capabilities are better suited for read-only massdistribution of preprocessed high-quality models. Where updates to the information are channeled through a more rigorous quality assurance process than these softwares' internal functionality can provide by itself.

Regardless if the relational database is provided as an on-demand put-on-line service in the cloud or permanent up-and-running physical computer that idles when it has no workload, the service itself requires a number of things before it can be deployed.

## 14.1 Requirements for Providing a Relational Database Service

This chapter is not finished yet, but will be a condensed essence of a number of database courses given at the university and in industry.

## Chapter 15 Reuse of Models with Copy and Paste

*We fired the poet, since what he delivered on this chapter just told us he couldn't contribute constructively to our business plan. Now we are fair and respect talent when it does its homework, so "The Poet" is welcome back once he/she or whatever that thing is, delivers a high quality statement that speaks the essence of this chapter.*

*[The BOSS]*

Authors comment: I'm with the BOSS. A text is a tool for thinking until it has been quality assured.

One of the largest productivity benefits with using a CASE tool, is the possibility to reuse quality assured, well understood, known to be good, models as baselines for new applications.

Plenty of core modules can also have functionality libraries implemented directly in target language source code that use and build on the automatically generated source code from those modules, and provide a reuse leverage when building new applications like DocumentDictionary that share the functionality of the same module.

Reuse of well known modules also has significant benefits in the human infrastructure that creates, maintains and uses the software and the information whose structuring and task adapted user interfaces delivers reusable internal knowledge models within the individuals' S.

It is much more expensive to "upgrade" the human knowledge of a whole organization or customer fleet, than upgrading a complex software if that software is implemented with state of the art methods and technology <sup>27</sup>.

Since the market success of the Apple Macintosh computer series in the 1980es, that made software industry almost universally adopt the "copy/paste user metaphor" from the physical office world, stands on taking a copy of a physical object and placing it in a clipboard. The paste action will then copy the contents of the clipboard onto the selected target object. The copy in the clipboard remains intact, and can be pasted on other target objects.

Copying a complex piece of structured information that may contain thousands of interlinked objects according to this user metaphor involves a number of non-obvious challenges.

To deal with those we need to understand the limitations of the biological hardware that implements our conscious understanding of what is going on in our by us controlled working environment through what the perception of understanding in human conscious working memory can deliver on the basis of the knowledge modules and abstractions that a human being operating user has learned.

---

<sup>27</sup> The public awareness of this insight amongst professionals in the expanding software industry during the 1980'es who drew conclusions from the market success of operating systems working with, instead of against this built-in limitation of biological processing hardware implementing corporate and governmental organismic life, formulated this insight standing on the best-practice, state-of-the art language evolved for user interfaces at those times. One excellent example of an in practice for the software platforms made available by the science backed up gravitational core driving the industrial software expansion at those times is [IBM 1989]. This reference is selected because its author kindly sponsored some student branches in Computer Science during the 1980'es, with healthy win-win relationships, that fostered growth both in access to good knowledge and a supply of students for a future long-term sustainability oriented expanding industry that took well care of its staff. The book provides references to the scientific basis it needed its staff to stand on for continuing to serve the market with high quality software products. Since the present state of the art of the software interacts and shapes organizations and individuals, it is important to maintain the stability and integrity of the scientific core that delivers conditions and real active expansion power for healthy long-term win-win relationships, until the scope of applicability for that market is saturated and the income areas delivering enough profit margins move up the added value chain, standing on healthy established state of the art standards and its performance beneficial biological hardware compiled presence in the human work force.

## 15.1 Dominance Ranking of Classes in a DomainModel

In order to efficiently automatically generate source code for copy/paste functionality within the design space of a DomainModel, some deriveable information from the structure of the entire DomainModel needs to be pre-computed to keep the complexity of the source code generators down.

There are three relationship types in the design space of a DomainModel. Part-Of relationships, Aggregating relationships and Reference relationships, listed in their power-of-influence-order, to determine the rank of a class within a DomainModel, with a dominance ranking algorithm. <sup>28</sup>

Algorithm outline:

1) Compute part-of rank - this is a matter of modeled physical assembly structure or inherent properties of the modeled external world when represented as information described in the language of the DomainModel.

2) Compute aggregate rank

Recursive top-loop - a self aggregating class.

Non aggregated parallel part-of peer.

Subordinate aggregate hierarchy peer.

3) Compute reference rank

Top-weight calculation (inside set)

Aggregated superordinated top-weight calculation (inside model)

Subordination weight/negative domination weight on "dominating" relationship paths

Statistics from actual instances of models decide

Dominates1to2 - a weight set by the designer or design team of the DomainModel.

DomainModel enclosed dominance rank of a class is the position in the sorted collection of all its classes ordered by partOfRank, aggregateRank and referenceRank.

To transfer a subset of objects within an isolated model container to an isolated clipboard container, while preserving all internal relationships within this subset, requires determining the boundary of this subset with regards to the surrounding model. The multi-dimensional dominance ranking enables this boundary to be visible by precompiled algorithms.

## 15.2 The Copy/Paste Metaphore and it's complication in the real world

The *object-action process sequence* builds on human language foundations of nouns and verbs. First you select the object, or objects with a multiple selection, and then apply the action. In this chapter's frame of topic the actions are Copy and Paste.

---

<sup>28</sup> Dominance ranking of classes within a DomainModel is a well-known secret amongst professionals working with large scale optimization problems ranging from autorouting of printed circuit boards to optimal shop floor planning for efficient massproduction of a particular complex assembled product. That level of knowledge is off-course far beyond the ambitions of this manual's important foundational step of education of appreciated software engineers towards a successful sustainable value-adding career in the industry that delivers the basic foundations for our shared wealth. Anyhow the computational speed of silicon processors, compared to what is possible to achieve by manipulating the physical world that is the real thing that is initially accessible to human beings before developing higher level abstract thinking that can be reused in writing programs that control what goes on in nanosecond real time in a silicon processor, tells us that this is something we need to use cleverly to deliver what we need to create the physical items we need to change our world towards the business plan. Dominance ranking in general depends on the knowledge domain it is applied to. So DomainModels are a good tool, to with simple examples explain the more general principles.

In the context of information structures declaratively expressed with a DomainModel, there are several different types of Copy actions. Below for simplicity of the explanation we assume that the user has selected a single object in a model. Then the multiple selection options that are not a strait forward iteration of the single selection action are explained in more detail.

### **15.2.1 Copy Object Only**

This copy action copies the object in the selection only, regardless what type of relationships the selected object is connected to other objects with. This copy action is frequently referred to as a shallowCopy.

This action is a very limited in its power to manipulate complex information structures, thus it is not made available through the user interface since it does not correspond to what most users of a copy action would expect this action would do.

### **15.2.2 Copy Part-Of Structure - Copy**

This copy action is the default that is available for the user under the familiar label Copy. It behaves as if the root object of a part-of structure represents its whole, which can be verified by selecting a complex object and issue the command Copy on it, and then use Edit->Show Clipboard, and inspect the parts of the object in the clipboard.

### **15.2.3 Copy Part-Of- and Aggregated Structures - Copy Aggregate**

In case the DomainModel contains aggregated relationships that enable the presence of recursive structures in a DomainModel, the Copy Aggregate action delivers something that is more close to what the user intends with a copy action. A typical in all domains reoccurring example is an assembly, that may contain other assemblies.

### **15.2.4 Copy Part-Of, Aggregated, and Upwards Dominant References - Copy Dominant**

An upwards dominant object, is an object that is owned by a shared owner higher up in the part-of hierarchy of a selected object, whose properties are shared amongs many objects on equal hierarchical level in a part-of hierarchy below the selected object. Examples are shared membership in various kinds of groups or categories, that own shared attributes and resources, or reference links to particular objects outside the part-of and aggregate substructure below the selected object.

Dominance ranking is a means to automate the algorithmic implementation of:

- 1) Deterministical declarative identification of the span of reach for the contents of a copy from a selected object or multiple selection of objects.
- 2) Defining the categories of span of reach that the user may want to adjust for a particular copy action, to set the closure boarder of the copy.
- 3) Deterministically defining what perhaps not complete subset of a copy that can be pasted onto a particular selected target object.
- 4) The categories of deterministic target structure of the closure of the paste, that the user may want to adjust for a particular paste action.

## **15.3 Examples for Functionality Coverage and Performance Analysis**

The following examples can be tested directly with the implementations provided in DocumentDictionary and OOCASE. They serve as evaluation pattern to be used for functionality coverage analysis and performance benchmark evaluation of a particular implementation using instances of well known Standard Models.



### **15.3.1 Copy Part-Of Examples**

DocumentDictionary: Copy Paste of DocumentRecord owning ElectronicEditions and Notes

OOCASE: Copy Paste of a Class with Attributes

### **15.3.2 Copy Aggregated Examples**

DocumentDictionary: Copy Paste of a ContentRecord aggregating a TableOfContents substructure through the contentContainer\_contents relationship.

OOCASE: Copy Paste of Module hierarchy with Classes and Relationships.

### **15.3.3 Copy Dominant Examples**

DocumentDictionary: Copy Paste of a Category of DocumentRecords with Category instance specific user interactive decision of recursive depth level of related DocumentRecords traceable through Cites. Use DBLP example.

OOCASE: Copy Paste of a GenericCategory of imported IEC61360 standard library of DataElementTypes, e.g. RosettaNet Technical Dictionary or eclass category.

## **15.4 Managing Traceability with Object Identifiers during Copy Paste**

The implementation used is the one described in Chapter 13 Version and Release Management. When an object or closure derivable from a selection is copied, each copied object receives a new unique object identifier, and the referential integrity of the entire closure within the copy is ensured by a copy algorithm.

Since the copy within the clipboard has new object identifiers it is no longer the same object as the original, and thus the mechanism of MakeEdition is applied to it. Briefly recalled, this means that objects in the clipboard lose any previous stamped quality assurance information for this particular new edition, and are provided with ReleaseBasedOn information so a trace to the original object for the copy is stored in the copied object itself.

No application specific information is updated in the copy unless the application has some implementation of a post-copy method.

When a copy is pasted onto one or several selected target objects, their ReleaseBasedOn and Edition information remains the same as in the clipboard, and thus provides direct trace links to the originally copied objects.

# Chapter 16 Information Quantity Measurement

The purpose of measuring information quantity was introduced in Section 1.1 Calculation of information quantity in a model.

## 16.1 Theory

The theory is well described in [Johansson 1996] in Chapter 12 Concepts and Notation from Infological Theory (page 97-107). The mathematical practice for how to calculate information quantity in a DomainModel is described in Chapter 14 Primitives for Domain Models (page 113 - 126).

## 16.2 Practise

OOCASE and the applications building on the same frameworks provide information quantity calculation in a model through File->Calculate->Information Quantity.

The user interface is similar to that for File->Quality Assurance->Check Model.

The information quantity can be measured in the unit bit, e-constellations (EC), or documentation e-constellations (DEC) which is an algorithmically computed approximation of the information quantity in natural language:

- 1 BITs (e.g. a 0 or 1) according to Shannon's theory based on the number of bits allocated for the selected TypeDef for storing an attribute value or relationship link as data in the model.
- 2 Elementary Constellations (EC) - Roughly described as the number of attribute values and relationship links in a model or part-of substructure in a model.
- 3 Documentation Elementary Constellations (DEC) - Various configurable methods for extracting the information quantity of text containing words and sentences written in natural language<sup>29</sup>.

In reappearing bulk-work for a software engineer, information quantity is useful when checking roundtrips of information distribution mechanisms involving transformation between different data representations. DEC is practical for high-level brief checking that definitions have adequate size for a particular documentation purpose, and quickly identifying anomalies in the documentation volume.

Information quantity measurement can be configured in detail using configuration files.

---

<sup>29</sup> To measure information quantity more precisely requires a natural language parser that can classify words for their grammatic roles and map sentence expressions to information quantity with the aid of a quality assured grammar pattern matching database. This database has to be verified on some agreed volume of realistic application test models from a particular knowledge domain. The level of detail is a matter of the value of calibrating the scale for measuring the information volume expressed with a certain language.

## Chapter 17 Profile Extensions of the MetaModel

*The blaze of speed and tuning, when geared to win the known,  
is accepted by the looming, to be the final crow.  
The unknown never ceases, delivering critique,  
thus those who not it peases, will never be unique.  
[The Poet]*

Authors comment: With a well conducted domain analysis and a sequence of with real production information evaluated prototypes, probably 95% of the average information handling needs can be covered for the bulk-volume of information processing needs within the domain. Agreeably the environment never stops changing, and taking care of the ripple on the surface of the bulk volume information representation needs over decade periods, is essential for maintaining the user support for a system.

Superior performance is quickly accepted for granted by a user community who are not into the exceptionally delicate details and history of how to achieve it in the layers of technology that it stands on. Human beings are tuned to react on things that annoys them and disrupt their daily conduct. Thus meeting the user's needs with a working, cost efficient, flexible but severely performance inefficient solution with regards to some infrequently occurring tasks, will allow the users to get that job done. And forget about that job, since users continue with the next job that most frequently is within the borders of applicability of the performance optimized system, and those accumulated smooth working experiences will cover for the lost credits of that other temporarily painful experience in the irrational human emotional accounting.

If there is no extension mechanism provided that can handle the ripple in information storage needs, the human emotional accounting will deliver it's verdict and provide a data migration project to some other system that until thoroughly performance tested will just remain being what it is, unless proper feedback loops are established with the new system's developers and maintainers.

With a leadership who know the value of benchmarking, resources for delivering decision support will tell when it is time to incorporate the experience accumulated in profiles into the application DomainModel, upgrade, migrate the information and push the ripple area to less costly levels.

The concept of profiles or DomainModel extensions has been discovered and rediscovered all over the planet in application environments that had an efficient enough core to expand beyond it's original scope of applicability.

### 17.1 Short Introduction to Profiles

A Profile is a restricted form of meta model that can be used to extend the DomainModel of an application such as DocumentDictionary and OOCASE. Profile specific Objects, Attributes and Relationships can be created and stored in a model that is supported by the application.

The purpose with Profiles is to speed up the development cycle, by allowing the user of an application to add functionality without having to ask the software supplier for help.

The benefit is the flexibility for the user to add the functionality needed in the environment that the user has to deal with.

The drawback is performance and non-standard heterogeneity that does not scale up very well in larger organizations and supply chains, that need to automate their communications with software that is available and affordable for the organizational units that carry their cost.

To the user however, in the situation the user is, having the flexibility of Profiles can have a severe impact on productivity and action capability to deliver added value to a wide range of customers. A well documented profile that has been used in production for some time and proven it's value can be the efficient design specification that the software supplier needs to be able to afford the investment of incorporating the functionality into the next version of their software product.

## 17.2 The DomainModel of Profiles

The DomainModel of Profiles has been discovered, implemented and rediscovered and reimplemented in parallel by independent non-collaborating actors all over the planet where people have access to programming languages and computers. Where the pre-requisite requirement for such a thing to happen is that these people live in an environment where curiosity and interest are promoted and allows them to develop an eye to see the underlying high-level patterns.

Every high-level pattern needs a concrete implementation to be expressed and communicated in the physical real world. The Elements used to formulate this expression are meta model objects that are grounded in the users understanding of their Names and Definitions.

In Figure 6, the layout dimension from top to bottom is going from the abstract to the concrete, that is from the class definition level to the instance level. The left-right dimension is an allocation of space for placing objects on a similar abstraction level such that it fits on the paper space<sup>30</sup>.

A Model, be it a DocumentDictionary, DataDictionary or a DomainModel may own one or many Profiles. A Profile owns ClassDefinitions and RelationshipDefinitions. These definitions can be modified dynamically when the application is running. If a user needs a new attribute on some application class, the user creates a Profile object in the model, and a ClassDefinition for the application class that needs to be extended with an attribute. The name of the application class to be extended is entered in the baseClass attribute of the ClassDefinition instance. Below this instance an AttributeDefinition instance is created, given a name, definition, perhaps defaultValue and type.

In the application, only objects of class ModelElement and its subclasses can be extended with Profile specific information. A ModelElement may own ProfileValues, which carry the profile specific attribute value in its value field. The ProfileValue object also needs to know what attribute its value is supposed to represent for the ModelElement that owns it. Thus the attributeName that the user has chosen to use for this value is stored in the field attributeName. The definition of that attributeName is recorded in the AttributeDefinition object in the Profile, and shared by all ProfileValues. The AttributeDefinition in the Profile is owned by the ClassDefinition that is identified by the className, which is also stored in the ProfileValue.

---

<sup>30</sup> A good geometric layout of a class diagram is arranged in a way that can be efficiently scanned by the physical implementation of the human eye and be remembered by the biological hardware we have in the higher abstraction processing layers in our biological hardware for recognition and processing visual sensory input during mental reasoning. It's a matter of efficient reuse of 2-dimensional structural layout to exploit the reusable mental picture that is helpful during thinking and utilize the vast storage capacity of the higher level processing layers of the human visual systems. See [Hubel 1988] for ideas of what shapes and structures our eyes are optimized for seeing and recalling with mental imagery. Compare the geometrical layout of Figure 6 with the geometrical layout of the DomainModel of OOCASE in Figure 3 and notice the reuse of semantic meaning with geometric location.

In a relational database, all ProfileValues are stored in the same table. Thus they need to know themselves what attribute their value is supposed to represent, and which class and by that inferable subclass hierarchy they might belong to in order to attach themselves to the correct ModelElement without excessive unnecessary search for object identifiers in tables that for sure will not contain the right modelElement.

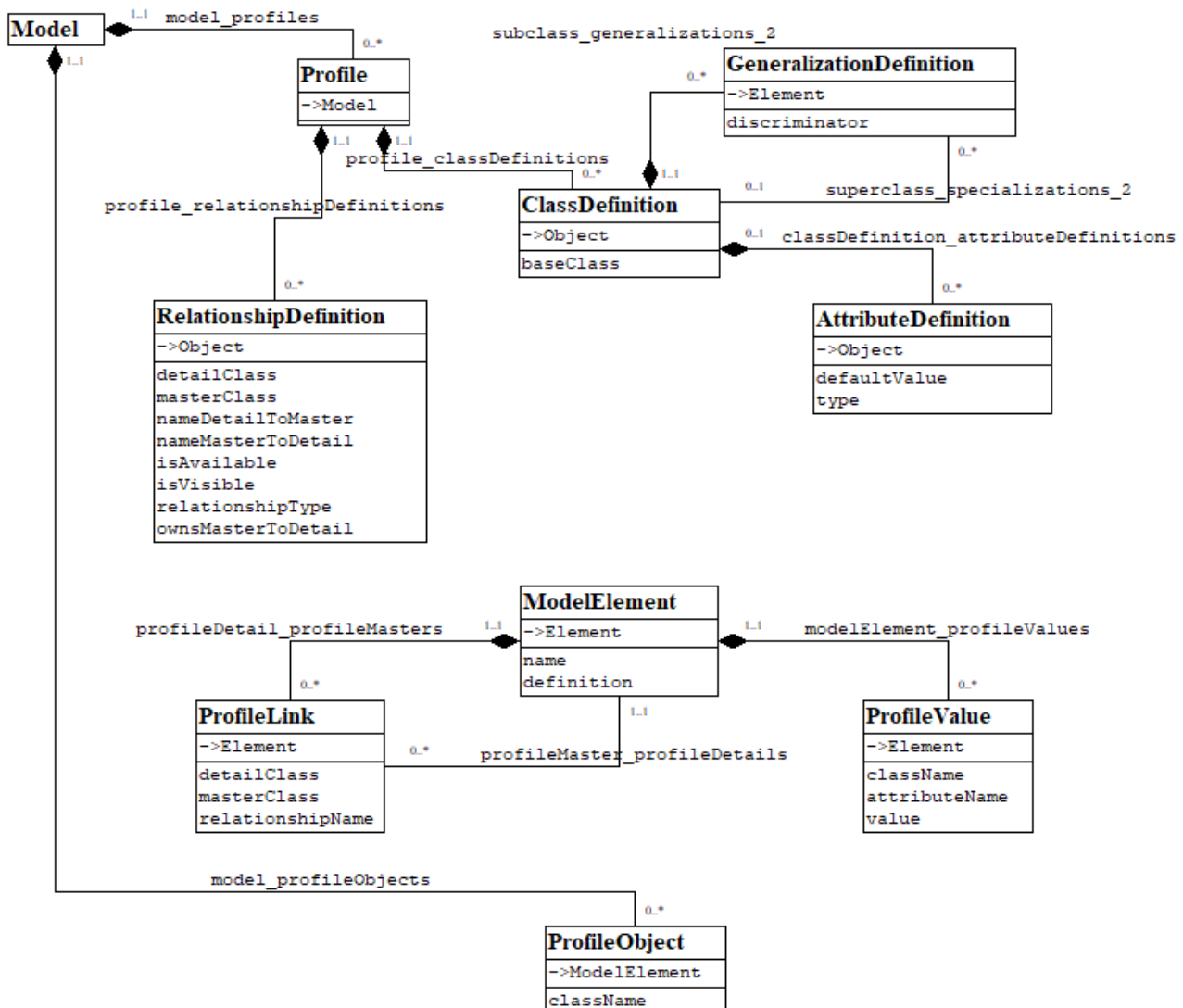


Figure 6. The DomainModel Module for Profiles

Relationships that the user creates between objects are represented by a ProfileLink. Since all ProfileLinks are stored in the same table, the links themselves need to know in what tables their particular master and detail may reside to attach themselves to the correct objects. The link also needs to know what it represents, and that is stored in the relationshipName, which is an index to the RelationshipDefinition by its name, where the user can define in a definition what links with this relationshipName actually mean.

In a multi-user database environment where data is stored and maintained over decades by different people, it is very important to provide definitions for ProfileElements. If this is not done, misunderstandings may lead to unnessecary additional costs in the maintenance of the physical or software objects that are documented by a model.

## 17.3 Comparison with SGML DTDs

A quick comparison is to study Figure 12.12 on page 253 in [UML 2.5.1] and compare it with Figure 6 above. Perhaps study the adjacent text or browse the OOCASE MetaModel in DocumentDictionary.

## **Chapter 18 Functionality**

### **18.1 Egoless Business**

The break-through paradigm shift of Egoless programming, was the separation of a working person from the subject matter he or she produced.

The code was no longer a primary source of personal pride, but an collectively owned delivery item that should serve its primary purpose flawlessly.

The pride of the team maintaining code, is 1) having cleanly formatted attractive looking easy to read source code according to an established coding standard, 2) no bugs and 3) high performance. In that priority order.

### **18.2 Building Efficient Interfaces Between Huge Refactorable Knowledge Domains - Efficient Knowledge Economy**

This condensed essence section has not been written yet, but the source material is already published in [Johansson 1996]. The principle of extracting reusable designpatterns out of a whole piece of scientific work are also documented in other partly overlapping software engineering areas in [Gamma et.al. 1995].

## Chapter 19 Summary and Conclusions

The software framework design of DocumentDictionary has its roots in the enormously enthusiastic software industry and academic computer science research advances in object-oriented programming, object-oriented databases, expert systems, programming environments, model driven software engineering and complete model driven application compilers that exploded in a period around 1988-1992. The core highly efficient design patterns evolving during that intense period while still holding a strong momentum until around year 2000, have now been production battle tested for 30 years and plenty of practical experience has been gained from information life in an environment evolving under Mores Law.

DocumentDictionary is ONE example application that can be generated and implemented in a week or month by reusing OOCASE library frameworks and one of its tuned model-driven software production pipelines, once the development team has decided on the DomainModel to be implemented and released.

The core technical information science has not changed, and the core OOCASE DataDictionary and DomainModel language has remained the same except for the adaptation to IEC 61360 in 1997 where the meta-class Property was renamed to DataElementType, and augmented with standard attributes to be able to import large industrial standard libraries into the DataDictionary.

The core theory and the concepts have no competitor as simple and generic as OOCASE with the same platform neutral capabilities among widely available standard programming languages.

The addition in 2015 with improved functionality for Quality Assurance with Edition, Version and Release (QAEVR) management following SemVer 2.0, with full traceability through the releaseBasedOn(Highid, Lowid, Version, Release) attributes and renumbering of object identifiers when issuing a new release of a model, enable full distributed in parallel version tracking by independent organizations that know nothing about each other, while still being able to trace the version history in distributed independently maintained repositories in any SQL92 compatible relational database or simple TAB-separated table text files for the classes of the information model (DocumentDictionary).

The Requirements Specification for DocumentDictionary grew out of the experiences of undergraduate studies at the University, and later Graduate Studies under heavy schedules with continuous work overload and a need to plan graduate studies efficiently to get a balanced life with time for recovery with confidence that dead-lines could be met. Later in collaboration with industrial partners, new assignments required intake of bulk-loads of software manuals, where the ability to with confidence give project steering groups a meet-able time table when the skills would be acquired, made planning and achieving success less stressful. Those were times when manuals were engineered with a pride in profession. The situation may be different now, unless you study the high-quality publications of the masters of our now integrating scientific knowledge domains. Such studies is a kick that make life worth living once you made your "environment" understand that someone has to take on such advanced research and development task too, to give directives for the future.

We need to use the present state-of-the-art technology more efficiently to deliver customer value in domains that still lack efficient digitalization, and thus have much smaller markets and less economic capacity. Supporting competent motivated students with the best tool and studying advice is our means.

This manual is a summary of what someone who really wants to make a long-term meaningful difference needs to know with regards to technical information that needs to be production live and maintained over decades while hardware and software platforms and programming languages change.



## A. References

- [Fowler 2003] Martin Fowler, "UML Distilled, 3rd Edition: A Brief Guide to the Standard Object Modeling Language", <https://www.martinfowler.com/books/uml.html>
- [Gamma et.al. 1995] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, "Design Patterns - Elements of Reusable Object-Oriented Software", Addison-Wesley, ISBN ISBN 0-201-63361-2, , 1995, pp. 395
- [Goldfarb 1990] C.F. Goldfarb, "The SGML handbook", Oxford University Press, 1990, ISBN 0-19-853737-9
- [IBM 1989] IBM, "System Application Architecture - Common User Access - Advanced Interface Design Guide", International Business Machines Corp., 1989, Document Number: SY328-300-R00-1089
- [Langefors 66] B. Langefors, "Theoretical Analysis of Information Systems.", Lund: Studentlitteratur, 1966.
- [Langefors 93] B. Langefors, "Essays on Infology, Summing up and Planning for the Future", Gothenburg Studies in Information Systems, Department of Information Systems, University of Gothenburg, Report 5, Augusti 1993.
- [Johansson 1996] O. Johansson, "Development Environments for Complex Product Models", 1996, ISBN 91-7871-855-4
- [ROJTEC 2018d] ROJTEC, "OOCASE User Manual", ROJTEC, 2018
- [SemVer 2.0.0] Tom Preston-Werner, "Semantic Versioning 2.0.0", Semantic Versioning, <http://semver.org/spec/v2.0.0.html>
- [Sundgren 1973] B. Sundgren, "An Infological Approach to Data Bases", National Central Bureau of Statistics, Sweden, and University of Stockholm, Dept. of Administrative Information Processing, Beckmans Tryckerier AB, Stockholm 1973.
- [Sundgren 1989] B. Sundgren, "Conceptual Modeling as an Instrument for Formal Specification of Statistical Information Systems", National Central Bureau of Statistics Sweden, 1989:18.
- [UML] OMG, "Unified Modeling Language (UML) Resource Page", <http://www.omg.org/uml>
- [UML 2.5.1] OMG, "OMG Unified Modeling Language (OMG UML) Version 2.5.1", Object Management Group, OMG Document Number: formal/2017-12-05, December 2017,
- [XML 2008] w3c.org, "Extensible Markup Language (XML) 1.0 (Fifth Edition) - W3C Recommendation 26 November 2008", W3C, 2008, <http://www.w3.org/TR/xml/> (accessed 2017-12-19)

*[The Poet] The Poet, "The virtue of successful poetry", (under review by an unscrupulous test department)*

*[The BOSS]*

*There is no adequate publication in the vast output that the BOSS already has made that will give any adequate picture of who the BOSS really is.*

*So if you ever meet the BOSS, you will recognize the BOSS's character traits, in that the BOSS is the nicest person you ever meet in your life. the BOSS seems to know you better than you know yourself. The BOSS cares about you and makes you comfortable to a level where you are able to really explain the reason why you are visiting the BOSS. The BOSS listens, and asks you questions. After a time in the pleasurable safe haven of being in the BOSS's nearness the BOSS signals to you that you need to listen to what the BOSS says, and when you get the BOSS's message, you are changed.*

*The change transforms you, you fly on the wings of an eagle and know exactly what to do. There is no doubt anymore and you know your BOSS is backing you if you follow the advice the BOSS delivered.*

*[The Poet] who knows exactly what The Poet's homework is to get the grace of being allowed to meet the BOSS again.*

*(now there are plenty of bosses who are encouraged by this and lock themselves up in their ivory towers, anyhow the BOSS is ultimately implemented by our own biology with its own history, so a good boss knows you well and knows what to say to make you deliver on a meaningful business plan)*

## B. References for Exceptional Students

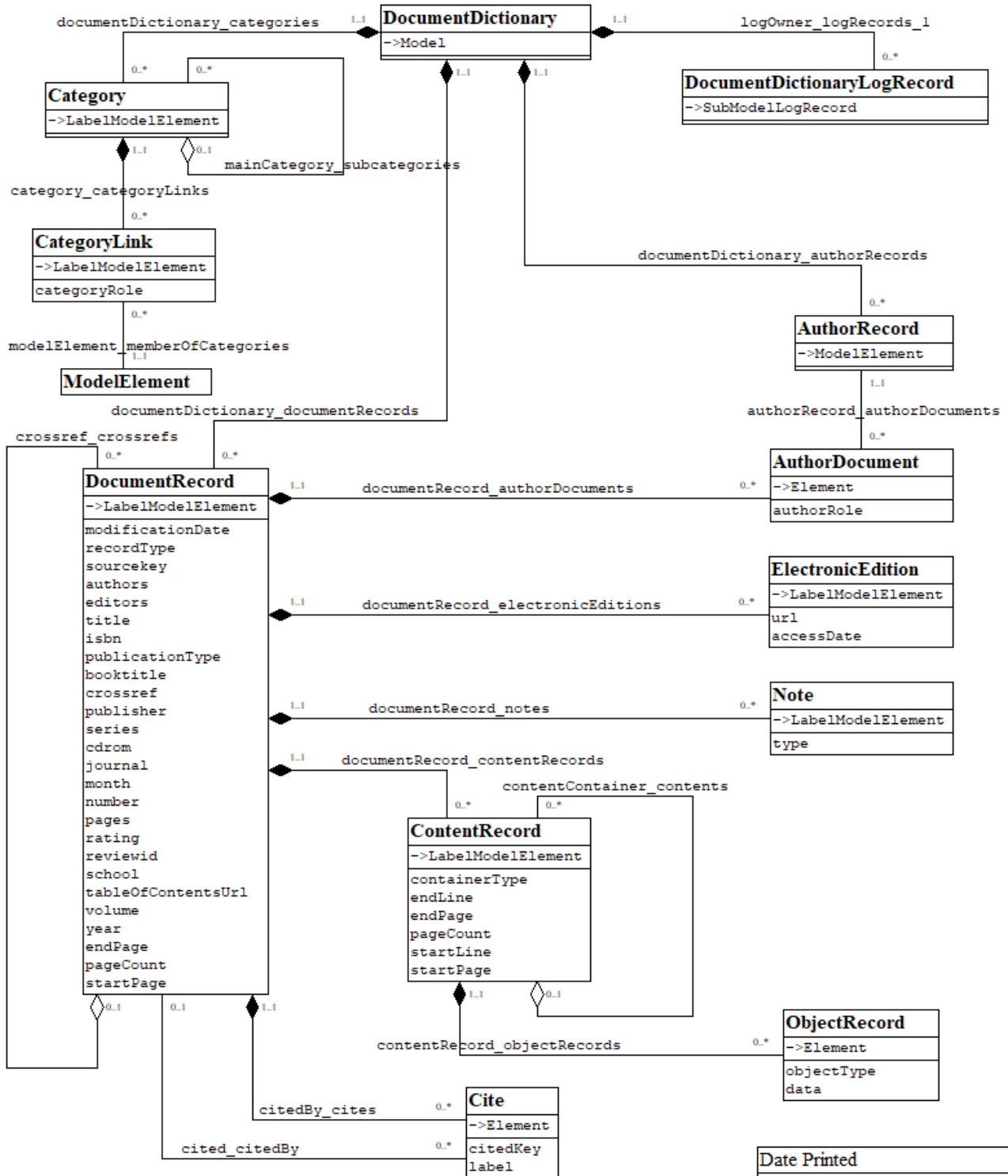
[Card et al. 1983] S. Card, T. Moran, A. Newell, "The Psychology of Human Computer Interaction", Hillsdale, New Jersey: Erlbaum, 1983

[Hubel 1988] A. Hubel, "Eye, Brain and Vision", Scientific American Library, 1988

[Lindsay&Norman 1977] P. H. Lindsay, D. A. Norman, "Human information processing", Academic Press, 1977

## C.DomainModel of DocumentDictionary

LabelModelElement
->ModelElement
label
ordinalPosition



Date Printed

2018-08-31 20:29:41

## D.Glossary

**GOOD knowledge** works in practice within its well defined scope of applicability without adverse side effects.

GOOD knowledge includes being able to identify knowledge as BAD knowledge, when that knowledge is used outside its range of applicability. Example: 1) "Sequential search" works well in the range of 1 to 100 items. Outside that scope of applicability, 2) "Binary search" delivers higher performance. 3) "B-tree search" outperforms "Binary Search" where memory retrieval cost (e.g. disk block access) exceed a certain time limit. Exploiting primary memory databases, efficient hashing, with optimization for exploiting the silicon based virtual memory capabilities and instruction sets of advanced processors and massive parallel GPU's enable still further improvements for fast information management applications.

**GOOD Student.** A good student is a person who is aware of his/her limited knowledge. Someone who can separate knowledge from his/her own personal identity or ego and evaluate knowledge for its merits and deficiencies within a particular area of application. A good student can study any subject and become master of it.