# Identification and Control of Business Records and Metadata at information and process modelling

First Draft : 1997-05-16

Author: Ulf Andersson
Information Systems
Information Systems & Technology
Astra AB

# Foreword

Ensuring longevity and evidential value of electronic business records is becoming an important issue in modern enterprises. More and more information is, though there may be a final printout, held in computers during a considerable part of its life cycle. How the record is managed, during the period when it is stored in a computer, is critical regardless the final presentation form.

The concept of the Sesam report[1] tried to prove that it is possible to obtain both longevity and evidential value of electronic business. The theoretical concept was validated and found promising at a joint international workshop[2] held by Astra AB and the Swedish National Archives. A theoretical records management concept is however not enough. It must be developed further to a concept for system design, expressing legal requirements and record management requirements in a way that makes them understandable and meaningful for systems engineers. In order to achieve this, the terminology of the previous refereed documents has been changed slightly to achieve the preciseness and consistency of terminology, necessary for information modelling.

---

[1] Astra AB Sesam, Philosophy and Rules Concerning Electronic Archives and Authenticity. Astra AB "( February 1996)

[2] Workshop on electronic Archiving. An evaluation of the sesam report by Ulf Andersson Astra AB 1996, published 1997.

# Table of content

# Introduction

## Business needs

The amount of and demands on electronic information generated and administered during the life cycle of a product has dramatically increased during the last 10 years, and will continue to do so. Information technologies are generating a new industrial revolution. It is a revolution based on information.

How an enterprise responds, and turn current opportunities into real benefits, depends on how well It can manage information as a resource, independent of the systems that operate on it. Information must be made independent of its creator and of the actual systems handling the information. Formats and structures in our current systems lock information into the creating systems. This makes it very difficult to use information outside the system and across organisations.

During the research, development, manufacturing and maintenance of product a vast amount of information, describing several aspects of the drug is collected, analysed and structured. This information is generated from different activities utilising a great number of different of computer systems. The information needs to be logically maintained by an organisation in such a way as to allow its use at a later date. This requires understanding of what data is available including the context of that data. There is also the issue of technical maintenance of electronic business records. Both the physical media and logical formats utilised have a shorter life cycle than the information it self. In other words the records have to be migrated, viewed by tools that emulates the original environment or by a combination of both. This has to be conducted by other specialists.

There is today no formal methods that secure the usability of information over business process borders and different releases of a product or application.

An enterprise must establish a corporate framework that facilitates managing product information as a corporate resource. This must include a standardised way to achieve uniformity concerning both logical and physical structuring of data and its context. One of the key issues to achieve those goals, especially for complex information, is metadata (i.e., data about the data itself). The recorded metadata shall support both reusability, maintainability and evidential value of enterprise 's business records. Electronic archiving must have the same reliability and the same legal status as paper archives.

# Reality to handle

The way we handle information with computer support is becoming more and more complex. This fact has consequences for the long term preservation. It escalates the need for specially trained professionals to handle the problem of this area. The critical issues can not be handled afterwards, but must be included in the design process of information management. Though to the fact that more and more commercial standard packages are implemented this does not solve the problem of having full control of both context and content of stored information. In worst case the result can be the opposite.

## Information primary structured to be machine interpretable

During the 60:th and 70:th the capacity of the CPUs of the computers were poor. To get systems that had good performance on those machines, the database applications were constructed with detailed models of the data see Figure 1. A result of this method was that any change in operations of the system made great amount of system design and programming work necessary. Most systems were used as isolated islands of information, some of them had of course ways to interchange information. This situation made long term preservation easy. The context in which the system was used was well defined and the data model is giving a good overview of how different information objects relate to each other.

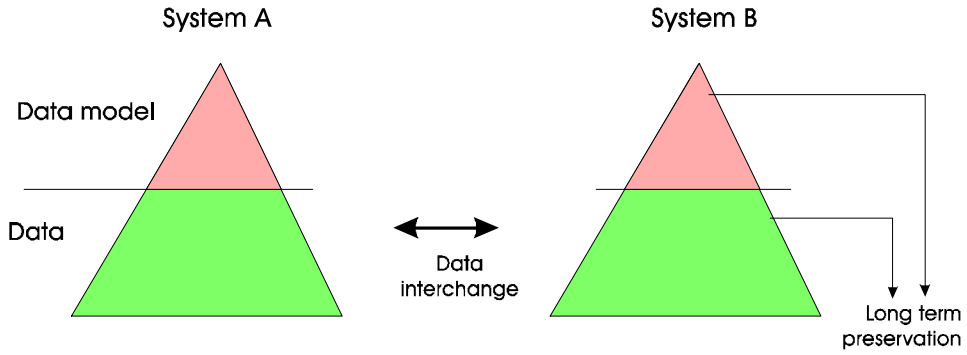



*Figure 1 Relation between data model and data in systems designed during the 60th and 70:th*

During the end of the 80:th and during the 90:th computers become more powerful. Those computers do not need detailed data models in order to get good performance. This possibility was used to decrease the need of involving IT professionals when business changes. The data models of the systems have become more generic. How such a small more generic model shall effect the system, is handled by user entered parameters. How this is done has a major impact on both how the system operates and how to interpret the data, especially when only the data, not the system, is available any more, see Figure 2.

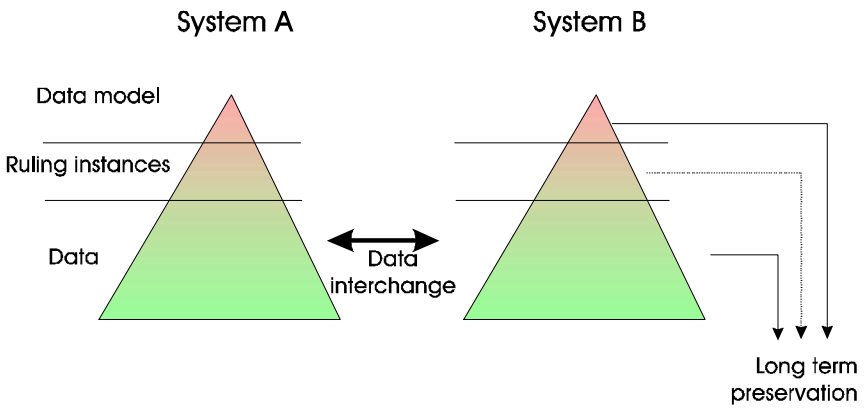



*Figure 2 Relation between data model and data in systems designed during the 80th and 90:th*

The introduction of such elements, controlled by the users, which are as important for the understanding of the stored data as the information model will effect long term preservation. It is absolute necessary to create and maintain “model like” descriptions of this information to ensure long term understandability.

In order to achieve data records which together create business records it is necessary to assure the information model and the derived data model and the ruling instances guarantee:

1. Activity which is documented
2. Rules for the activity
3. Information describing the activity
4. Context of the activity
5. Tractability and change control
6. Responsible organisation
7. Responsible person

In order to optimise efficiency of business support, the amounts of information interchange between systems have increased. It is essential that information interchanged has the same meaning in the delivering and the recipient system.

An other problem for long term preservation is the fact that the data model only poorly reflects the users view of the information. It represents a possible physical model for how to implement the users' view. The information model, reflecting the users view regardless of implementation technology, must be maintained.

## Information primary structured to be interpretable for humans

During the 90:th computers become more powerful enough to handle multimedia documents. This is a kind of business records that are structured to optimise human understanding. There is no database management system that control neither the structure or used syntax compared with a traditional database schema (no blobs), compare Figure 2 and Figure 3.

The complexity, that can be created and has to be preserver for the future, is unlimited. The 7 topics in the list of the previous paragraph must however still be fulfilled.

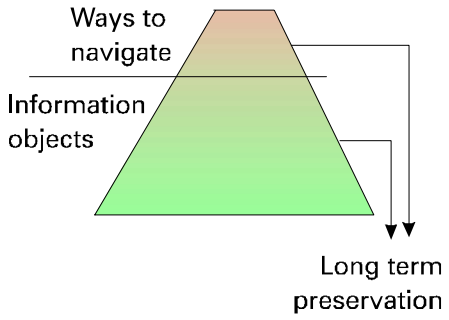



*Figure 3 Multimedia does not need to have a predefined logical structure*

# Strategy

## Vision for electronic long term preservation of electronic business records

The goals of long-term electronic management of information and electronic archives are to fulfil today's and tomorrow's requirements on:

- authenticity
- legal status
- security
- reliability
- possible to use stored information in operational business activities as a "corporate memory".

To be able to achieve this re-use advanced navigation and search functions are necessary, capability that makes an electronic archives superior to a paper-based traditional archives, both in functional and economical terms.

## Mission

It is essential that an enterprise creates IT environments producing, storing, transferring and maintaining electronic business records which can be used as evidence, whenever needed. An electronic business record is distinct from an electronic data record since it can consist of several electronic data records. It contains not only the data which is traditionally stored but also information (metadata) about the context of the data.

The necessary minimum level of quality of information management will be decided by the authorities, not by the private enterprises. A complicating factor deciding the quality is the elevation of the necessary quality demands that will occur during the time span between creation and re-use.

## General

A business record has a life cycle. It is created, used, archived, migrated, reused and deleted.

Three factors are essential for a successful management during the entire life cycle:

1. Define data structures (see page 23, information packages) which allow description of the wider context of the data related to the mission (see page 22, *causa*) to be documented. This concept referees and take advantage of Swedish administrative tradition by defining the “ärende”. The term is not translatable directly into English but in some areas, case mean the same. Those structures (information packages) must contain the necessary amount of metadata to allow full understanding. They shall maintain their evidential value even after EDI to another system.
2. Information in databases must be archived as flat ASCII files together with the information models and process model also documented in flat ASCII files. The syntax must be in accordance with established standards.
3. The base for maintaining text or multimedia is simple pieces, individually maintainable, linked to achieve the full presentation.

An archiving philosophy shall be created and implemented which in principle may be applied to storage of any representation made in accordance to the stated rules and standards. The documentation shall also be made searchable and extractable from the archival application.

Business records must be created with the decided quality as to their logical and physical structure and their formats. The archiving environment must operate according to the established rules and only accept business records with the decided quality. Those factors will guarantee the integrity of the business records.

## IT environment

An infrastructure for archiving that is physically and logically separated from the production environment shall be established.

Information packages, with legal status, shall be transferred to the archives when they have been approved by the responsible person(s).

The organisation of context and content must meet both internal needs and external demands for re-use of information. By using appropriate stable standards when structuring and storing information, we will be able to use the full potential of new technology and still have a stable storage environment.

The archival application must log the access and extraction of business records.

It shall be possible to invoke certification of how event-logging, is performed by an external public body, if and when it becomes available.

## Permanence

Formats and logical and physical structures shall be identified and utilised in such a way that migration of stored business records is a migration only of the formats used for storage, not involving changes of the structure of user context and content.

Migration of business records shall not include a conversion of the logic of the applications used to manage them.

## Evidential value and authenticity

**General**

The enterprise must create an environment that can maintain full evidential capacity of the business records during the whole life cycle of a record. This is necessary since the enterprise must guarantee that information leaving the company in an electronic way is correct regardless of when the record was created. Another aspect is that the employees must be able to rely on the systems so that their work can not be manipulated afterwards.

The lifetimes of media and formats in electronic archives are often shorter than the lifetime of the information of the business record. Nevertheless, the archival process must guarantee the ability to use the business records as evidence in legal proceedings for example patent litigation. Migration of business records must be carried out in such a way that quality is maintained and the accuracy of the converting procedure can be guaranteed. Authorisation and quality control of these transactions must be documented. This documentation must be maintained during the whole life cycle of the business records. Multiple possibilities of verifying the authenticity of a business record must be established.

The documentation produced by the pharmaceutical industry is ruled by various regulations from different countries. There are regulations which determine how long the business records have to be maintained. It is essential for a pharmaceutical industry to be able to maintain large volumes of documentation for several decades. After this period it is unlikely that the formats for storage are still usable in the original state so we must rely on migrated versions. The authenticity must not be lost in the migration process.

**Differences in between countries**

There are major differences between administrative tradition in different countries. It will effect how courts will interpret the evidential value of records.

This makes it necessary to have an overview of the legal systems in important markets concerning business critical issues.

## Security

**General**

The different levels of confidentiality decided for the enterprise must be implemented in a way that is interpretable for computers. Information and process models have to be designed to support all necessary levels of confidentiality.

The need of confidentiality for a certain business record will vary over time. In most cases it will decline.

**Security against encroachment**

Electronic interfaces for navigation in information, such as the World Wide Web on the Internet, give an overview of information of a breadth that does not exist in traditional manual systems. This increases both the danger of an encroachment on electronic stored data and the extent of a possible damages.

“Fire walls” are used to connect internal LAN of an organisation to the Internet. In spite of this there will be encroachments from the outside into the internal LAN. In manual systems there are several barriers for protection. Electronic environments must also have multiple barriers. A professionally managed firewall for Internet connection is crucial.

Encroachment into archival servers must be stopped whether it is done via the archival application or directly into the archival server.

## Economy

Management of electronic business records is expensive. The extra benefits e.g. multiple access and paste capabilities are therefore essential to achieve a balanced economy.

Business records are to be stored once, with full resolution, in stable formats and in safe and cost-efficient archives. It must be possible to present the business records on different media depending on user needs.

## Responsibilities

Management responsibility:

- Establish rules concerning information management which includes the results from a continuous monitoring of regulations from the most important markets.

- The production environment, the event-logging environment and the archiving environment are, from an organisational point of view, to be totally independent of each other. No one is allowed to have full access, outside operational systems, to more than one area.

The user organisation responsibility:

- The producer of business records is responsible for ensuring that these records are adequately structured to meet both operational and archival demands.
- Business records, both content and context, are produced with the necessary quality to make long term maintenance possible with maintained authenticity. Used structures shall ensure that context descriptions are easy to understand.

Archiving organisation responsibility:

- Make information retrieval possible according to user demands.
- Maintain business records in a format and with a syntax that make them usable for the production environment.
- Maintain authenticity of archived business records.
- Select appropriate standards to express context and content for archiving.

# Methodology and definitions

# Methodology

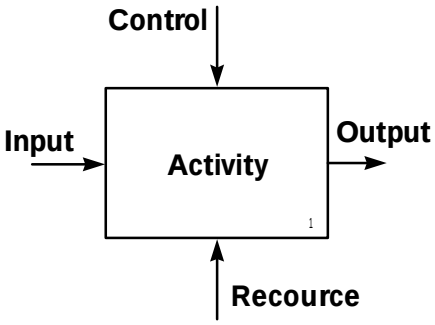
## Process models

### IDF0 syntax

**Reading activity models**

In this methodology we will use the IDEF0 activity / functional modelling language to define our activity models. The basic ideas of this language will be provided in the following pages for the purposes of the domain experts. It is assumed that the information modelling experts already know and understand the IDEF0 language and techniques.

IDEF0 is a very simple diagramming technique which consists of boxes and arrows. The boxes represent activities and the arrows represent interfaces to these activities. The starting point is to identify the single activity which covers whole scope of the model, this is usually termed the top of the decomposition and identified as A-0. This activity is then decomposed into between 3 and 6 other activities which together perform the higher level activity.[3] This decomposition is then carried out for each element of the diagram until the detail becomes too small for the purpose or the viewpoint of the model.




The activity is identified by a name which should be verb based. Each activity should also be identified by a number indicating where the activity occurs in the breakdown structure of the next higher level activity. If a particular activity is

[3] The limit of between 3 and 6 was decided to ensure that enough detail is described in the breakdown, but not too much detail as to confuse the reader.

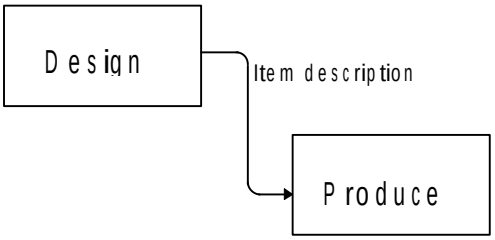
not further decomposed a small line is put in the top left corner of the box to indicate this.

The role of the arrows which are connected to the activity is determined by which edge of the box the arrow is connected to. Each arrow role is defined below:

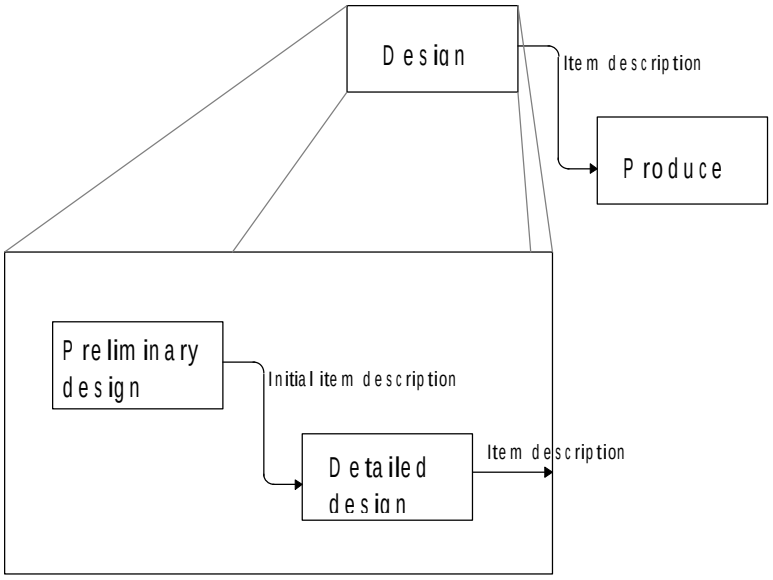
- Inputs to the activity are presented as entering the box from the left edge. An input is defined to be the information required for the activity to be performed. This information will in some way be transformed by the activity.
- Controls on why the activity is performed are presented as entering the box from the top edge. A control is defined to be the constraints or governing information which are used when performing the activity.
- Outputs from the activity are presented as leaving the box from the right edge. An output is defined to be the data created or modified by the activity.
- Recourses on how the activity is performed are presented as entering the box from the bottom edge. A recourse is defined to be the means by which the activity is performed.

Arrows are connected between boxes as follows:

- on the same level of decomposition. The arrows are directly connected, the output from one activity may be an input to another or act as a control




- between boxes on one level and the box which they have been decomposed from on the next higher level. The arrows at the lower level are connected between the boxes and the boundary of the diagram representing the edges of the higher level box.




An arrow which is connected to an activity as a control or an input constrains the activity in that the data supplied by the arrow should be present before the activity can proceed. In the case of feed back loops the constraint is not applied unless feed back is provided.

The arrows may branch, indicating that the same class of data is used by more than one activity. If the data implied in the branched paths are different then they should be labelled differently after the branch, if the branches represent the same things as before the branch new labels are not needed. Arrows may also join, indicating that the same class of data is produced by more than one activity. Similar naming rules apply to the joining arrows.

# Information models

## Express G syntax

**Reading information models**

In this methodology we will use the EXPRESS language (ISO 10303-11:1994) to define our information models. The domain experts will be using a graphical subset of the language called EXPRESS-G which will be described in the following pages.

The EXPRESS-G language was developed to provide a simple interface to the EXPRESS language which ignores many of the more complex constructs. It was intended that the EXPRESS-G language should be easily read by people who were not experts in information modelling.

EXPRESS-G is basically comprised of three different types of symbols;

- definition symbols;

- relationship symbols;
- composition.

Each of these type of symbol will be described.

**Definition symbols**

These are symbols that denote simple data types, named data types, constructed data types, and schema declarations. Definition symbols consist of a rectangular box enclosing the name of the item. The form and line style of the box indicates which kind of definition applies.


An ENTITY data type represents a collection of conceptual or real-world physical objects which have common properties. These common properties are represented in the information model in terms of attributes and constraints of the entity data type.

Entity is the fundamental modelling concept in the EXPRESS language and are used to represent those things of interest in the real world.

A simple type is a basic type within the EXPRESS language and defines a domain of values which can be used in the definition of an attribute.


These types are the types which can usually be stored in a computer with no conversion or mapping.

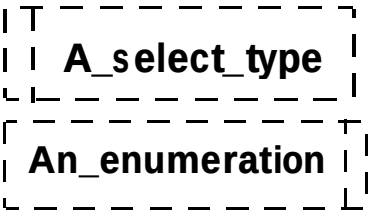
The simple types are:

- NUMBER, either a REAL or an INTEGER;
- REAL, a scientific number which includes a decimal point;
- INTEGER, a whole counting number;
- LOGICAL, either TRUE FALSE or UNKNOWN;
- BOOLEAN, a special type of LOGICAL which is either TRUE or FALSE;
- STRING, a value which is made up of a sequence of characters

BINARY, a value which is made up of a sequence of bits (similar to STRING).

A defined data type is a type which has added semantics identified by the identifier associated with the type.


The domain of values is given by the base type of the defined type indicated by an unnamed relationship to the base type from the defined type.




A select type is a special form of defined data type which groups together named data types. It defines the union of domains of named data types in its select list. This is a generalisation of each of the named data types in its select list.

defines an ordered set of names. The names represent values of the enumeration data type and are referred to as enumeration items, these are not presented in EXPRESS-G.

**Relationship symbols**

The relationship symbols are used to relate the definition symbols with each other and are presented a lines of various styles. The different types of relationship between the definitions will be described with the line style indicated in the following pages.

An attribute represents a trait, quality or property which is a characteristic of all things represented by that entity. It relates the entity with a type used to define the domain of values associated with that characteristic. Attributes are given a name by which they are identified in the entity. There are three kinds of attribute in EXPRESS:

- explicit attribute, these are the key attributes which define the characteristics of the entity;
- derived attribute, these are characteristics which may be derived from other characteristics of the entity;

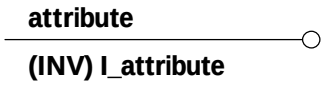
inverse attributes, these are characteristics which may be found by looking at the roles the entity plays in relationships with other entities (inverse of explicit attributes).

Required attribute, this symbol represents an explicit attribute which is required by the entity.

The name of the attribute is “Required_attribute”, and the circle identifies the type used to define the domain of the attribute.

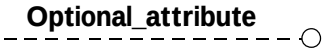
This symbol represents a derived attribute, identified by (DER), with the name “attribute”.


This symbol represents both an explicit attribute “attribute” and the corresponding inverse attribute




“I_attribute” for the same relationship. Since an inverse is the inverse of an explicit attribute the two are always found together.

This symbol represents an optional explicit attribute.




This identifies an attribute for which the value does not always have to be present. An example of this could be a middle initial of a name.

This symbol is used to represent the relationship between a defined data type and its base type, the base type being identified by the circle.

To support the above relationships there are characters which identify aggregations and bounds on these aggregations. The aggregation types allowed by EXPRESS are as follows:

ARRAY (symbol A) This is a fixed size collection. It is indexed by a sequence of integers. You can specify whether or not an instance can have duplicate elements.

BAG (symbol B) This is an unordered collection and duplication is allowed. The number of elements in a bag can vary and can have constraints associated with it.

LIST (symbol L) This is a sequence of elements that can be accessed according to their position. The number of elements in a list can vary and you can constrain it

by the data type definition. You can also specify whether duplicate elements are allowed.

SET (symbol S) This is an unordered collection of elements where duplicates are not allowed. The number of elements can vary and you can constrain it by the data type definition.

The bound specifications, minimum and maximum which can be held in the aggregate, can also be specified using [l:u] where l and u are the lower and upper limits respectively.

This symbol is used to identify inheritance and relates entities together into subtypes and supertypes.

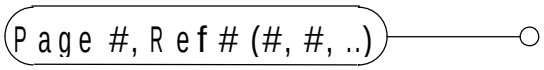
The circle identifies the subtype in the relationship.

A subtype is a specialisation of the supertype either by reducing the domain of the supertype or by adding more specialised attributes.

**Composition symbols**

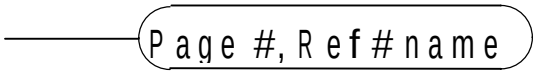
Since information models can be very large and humans can only comprehend a number of objects at any one time the EXPRESS-G specification allows a single information model to be distributed over a number of pages. To allow references between these pages to be followed there is a page referencing notation which has accompanying symbols.

This is a reference onto this page.




The page # is this page number, the ref # is a unique reference number on this page, and the numbers in parenthesis are the pages from which the reference is made.

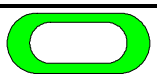
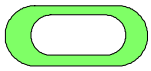
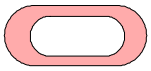
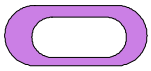
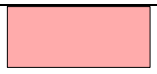
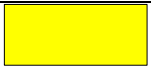
This is a reference onto another page, it indicates the page, reference number and name of the thing being referenced.



# Definitions

Electronic business records have different characteristics compared to traditional paper based business records, basically depending on the lack of intuitive understanding of an electronic business record, for example that you cannot touch and feel it. Due to this difference in characteristics, a set of definitions has to be established for electronic management of information.

## Usage of colours and shapes

| | |
|---|---|
| Archival model |  |
| Causa record |  |
| Configuration |  |
| Activity record |  |
| Activity record metadata (envelop) |  |
| Information model |  |
| Information model with business records and metadata |  |
| Library information |  |
| Process model |  |
| Activities |  |
| Product model |  |
| Reality or action |  |
| Rules |  |
| Stored data or information package |  |
| User interface |  |

## Business process model

A process model is a description of activities within business segment, including their sequence and the information flow between them.

## Business record

A business record is the representation and proof of a business activity or the context in which the business activity occurs. In order to achieve evidential value it must not only contain the information related to **the result of the activity** (causa or business activity): library information, it must also contain information of the **circumstances** of its creation as well as information concerning **organisation and business process**, see Figure 4.

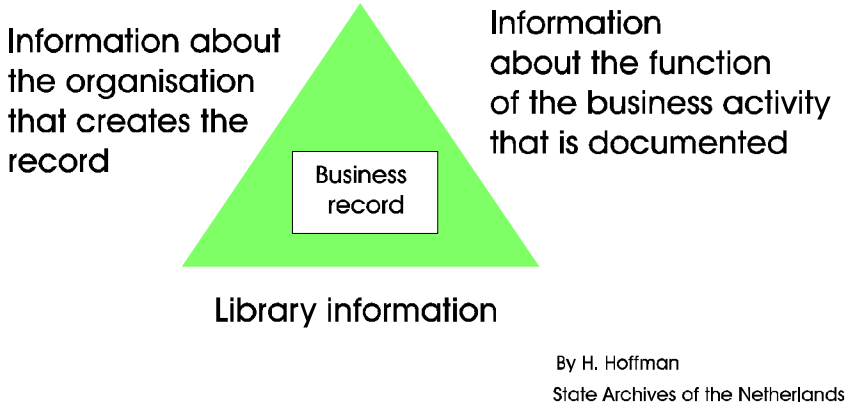



*Figure 4 The content of a business record*

### Causa

A *causa* (Latin) is a defined administrative routine limited in time and extent. It is performed according to administrative rules. It can be thought of as a work programme. A *causa* consists of a sequence of activities that forms a logical unit. A *causa* is represented as an **information package in the electronic environment** and as a case documents in the paper based environment.

To obtain a self-explaining information package, models of both the business processes and information content must be realised. Parts of those models must be included in the business records as metadata. The relationship between reality (the causa and its activities) and its representation is shown in Figure 5.

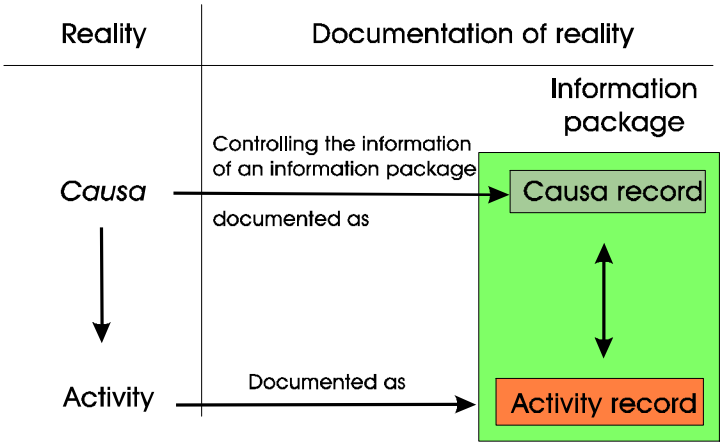



*Figure 5 Relations between how reality is structured and how it is documented*

Business record and its subtypes, used in this document, is shown in Figure 6

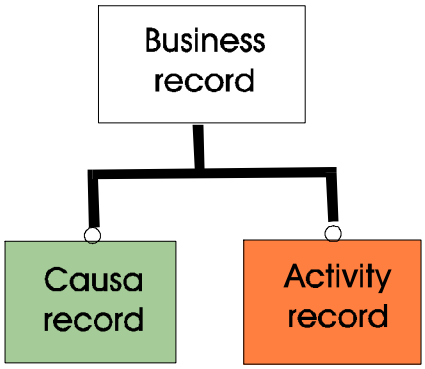



Figure 6 Specialisation of business records

## Information package

An information package is logically the electronic equivalent of a traditional paper-based case file. It embraces one or several activity records that belong to the corresponding *causa*. Each activity record must have sufficient metadata explaining its context. This collection of metadata is called envelope. In most cases is the format and syntax of the library information of activity record **not** capable to carry the metadata. In this case it has to be stored in a file of its own. Causa record is identifying and explaining the information package and must utilise a format and syntax fulfilling those demands on structuring capability, see Figure 7 and Figure 19.

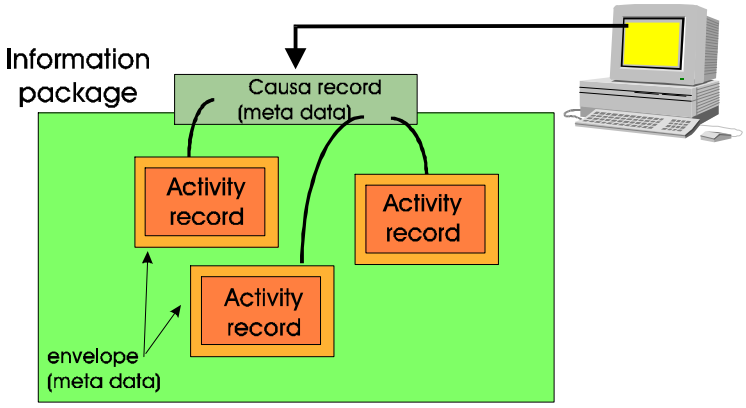



*Figure 7 An information package with* activity records*, envelopes and causa record*

For computers the causa record will provide a number of items where the local information model of the information package can be interfaced with product models relating to company needs. Compare by putting a book on the shelf its table of content and index would automatically be a part of the indexing system of all the gathered books of the library.

If electronic muniments are unnecessary to create and all activity records to all *causae* are located in the same database, the information package and its components are all just parts of the archival database extract file, a STEP part 21 file.

## Causa record

The causa record is a subtype of a business record. A causa record is the identification and physical control of an information package, it holds the metadata for the entire information package. It is therefore the "point of connection" to the archival structures. It is the only allowed point of access into an information package. Activity records are never connected directly to the product model and archival structure, they are always connected through a causa record. See Figure 8

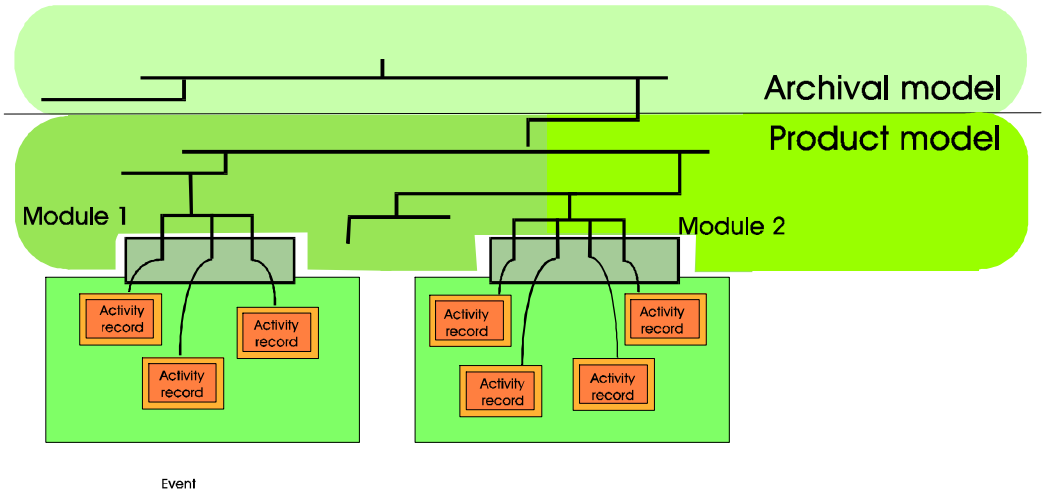



*Figure 8 Information packages connect to the archival structure, consisting of the archival model and the different modules of the product model, through the causa record*

Compared to a traditional case file, the causa record contains the diary or log in the case file, as well as a table of content for the case file, keywords for searches, and references to and into activity records. This would not make too much sense in a paper based case file, but in the electronic information package it is the key to search and retrieval of information. The causa record provides keywords, structures, links (electronic references) and "change history" of the representation of a *causa*. This would provide a functionality to find and read or retrieve information packages and activity records

## Envelope

An envelope is the identification and description of an activity record. It holds the metadata of the activity record. Compared to a paper based record, the envelope contains responsible organisation and person, signatures, and a description and key to the format and syntax of the record (i.e. "it is a paper document written in English").

The envelope provides the relevant context to an activity record, constituting the real content of the information package.

## Information model of an information package

The equivalent of a traditional files in archives are the information packages. An information package will **gather** all information concerning a specific mission, a causa. It will store the **life cycle** history of all included activity records. All gathered activity records will provide each other with **contextual information** to ensure understandability and evidential value of the included activity records.

For humans the causa record will provide a standardised way for **access** and understanding, compare table of content and index of a book. For computers it will provide a number of items where the local information model of the information package can be interfaced with product models relating company needs.

The envelope is just a technical solution to put metadata related to business objects in a computer interpretable file and structures. Very few formats have the capability to allow your own design of metadata structuring. If the activity record is a SGML instance the envelope will physically be a part of the record. The information model of an information package is described in Figure 9.

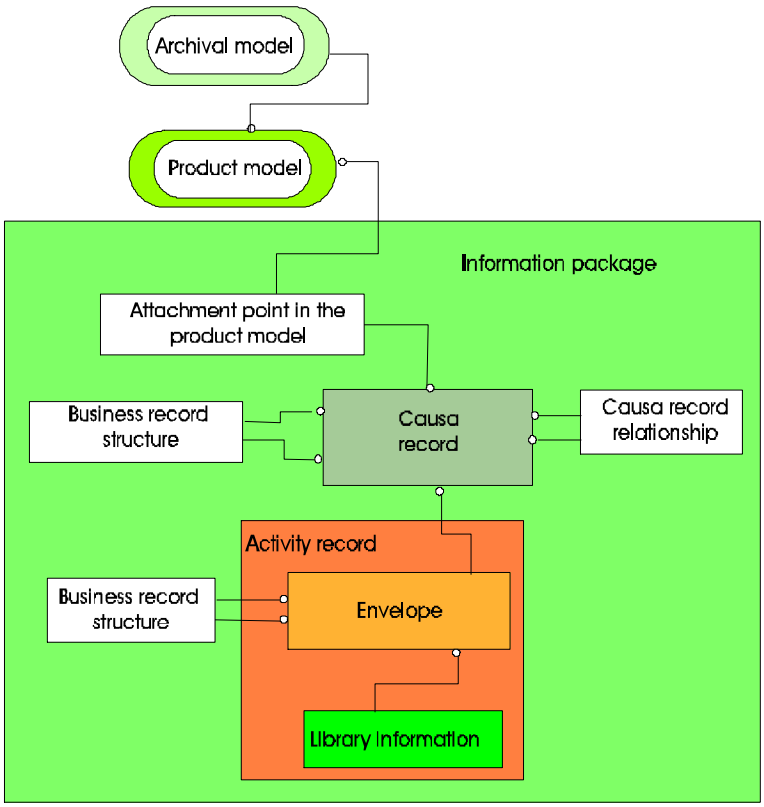



*Figure 9 Generic information model of an information package and how it is mapped to the product and archival models*

From information modelling point of view, the causa record is a specialised business record, see Figure 6.

**How to decompose and classify actions and things**

From process and information modelling point of view we must separate between actions and physical representation of an action. They will both be decomposable into different parts with relations in-between. The parts may have an internal structure, see Figure 10.

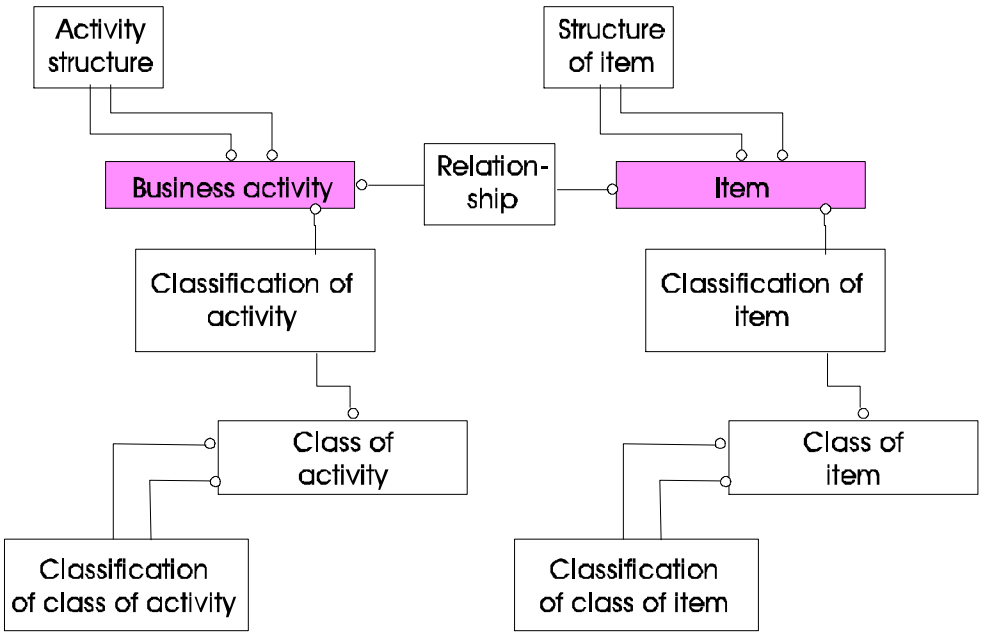



*Figure 10 How to classify and relate business actions or items*

**The context of where to apply metadata and other data terms**

Scope of the information model

All parts of an information package are business records, including the causa record. It is from information modelling point of view a decomposable assembly of different types of business records. The model shall show the relations of involved components from a system design perspective, see Figure 11.

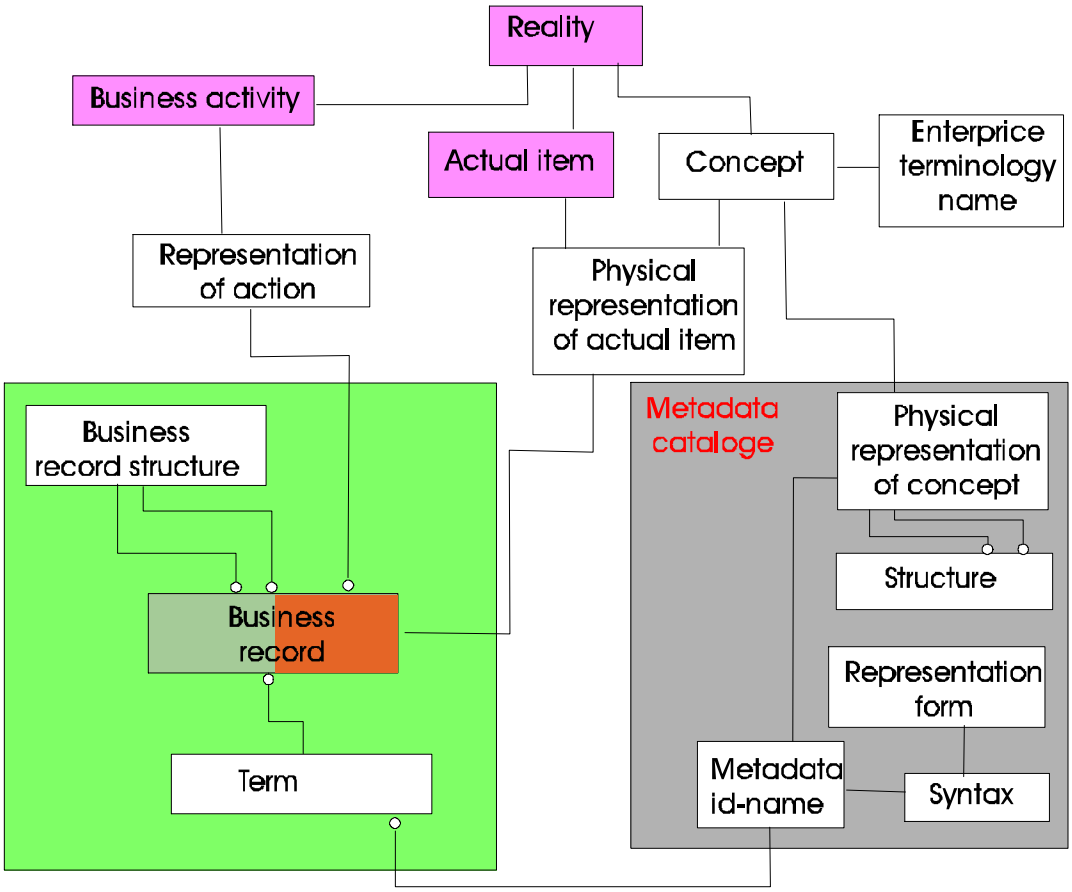



*Figure 11 Generic information model of action classification*

Figure 10 shows the complexity and possibility to decompose activities and business records. The actual complexity of the reality will effect which metadata terms that are to be added to a business record. The necessary metadata for each level of decomposition of a business record is to be found in the catalogue, for definition of the catalogue see “Process model how to derive metadata” on page 50.

**Information packages and business records storage syntax**

If a business record is to be stored as an electronic muniments or by utilising database storage[4] is due to risk assessment.

The main difference between data base storage and the electronic muniment in this aspect is that the modern database has a multidimensional capability to express structures and relations by utilising the syntax of the DBMS. The electronic muniment has a structure for human reading. This will give the muniment a more sequential structure than a database solution. The syntax used

[4] where the format is fully controlled by the DBMS-software, not a BLOB (Binary Large OBject)

for electronic muniment must how ever have the capability to strictly identify the same terms used in a DBMS solution giving the capability to relate to the terms and structures of the information models. Only few standards have this programmable possibility for identification of terms.

# Requirements for records management

# Activities and their representation in an enterprise

## The model of the enterprise

### Objectives of the model

None in an enterprise can operate alone, all employee acts within the network of the enterprise

The enterprise consists in general of 5 to 10 business domains. Each domain consists of several working units which together correspond to the business needs. The enterprise model shall show how a business activity is involving different business areas and producing a documentation that is to be utilised in several business areas.

### Creation of information and metadata

An enterprise is an integrated structure of different business areas. Performed business activities will always involve more than one business area. As a consequence the created information must be possible to find, understand and re-use across the organisational borders. To achieve this capability it is necessary to have a basic level on standardisation on structuring, and naming of information and metadata. Figure 12 gives an overview of this complex situation.

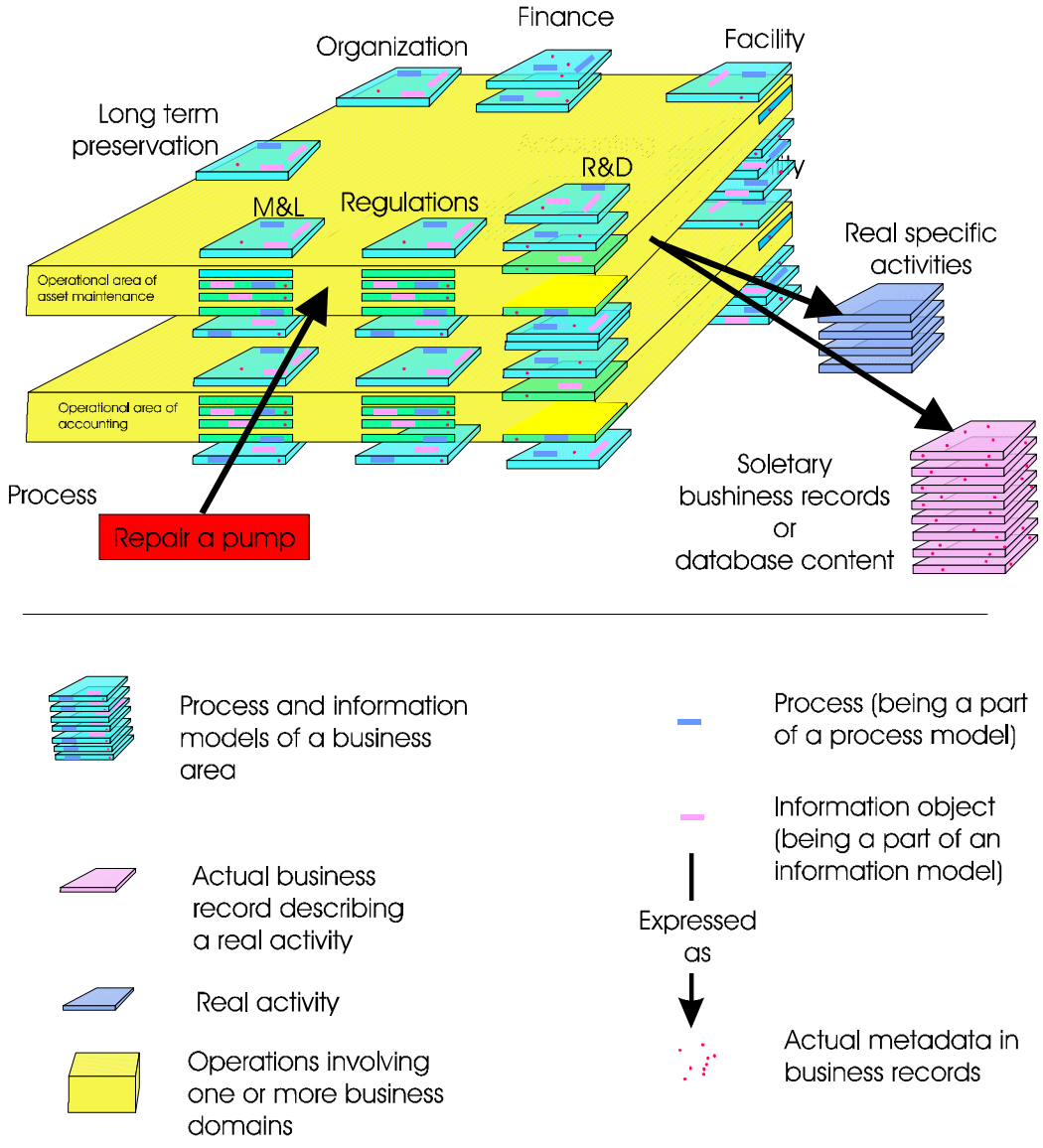



*Figure 12 The enterprise model*

As described earlier most activities done by any operation, involve of several operations from different business domains. An example, Repairing a pump in the production area:

The mission (process) to repair the pump is the *causa* see Figure 5. To do the different activities necessary to conduct the *causa*, several business areas will be involved. The conducted activities must be described in activity records. This documentation must be possible to identify and utilise from all involved business areas.

In the enterprise model it is obvious where entities are regarded as “data” and is modelled as such. This organisation is regarded as concept holder. The same

information is regarded as metadata in an other business domain. When information classified as “data” in one business area is used as metadata in an other, it must be handled in accordance with demand of the concept owner.

In practice this means that information classified as metadata in the activity records produced to document an activity is classified as data in an other business domain. To make information transferable we need to ensure an unambiguous naming if information. List of value is to be decided by the concept holder.

### How to control identification of data and metadata

Creating common structuring within an enterprise is a difficult task. To much structuring produce bureaucracy and to little do not support the user need. The way to achieve a proper level is to allow each level within the enterprise to define the minimum level necessary for their operations to handle information between and within business domains at their level. The deeper down in the organisation the more need of high volume control of details. By letting each level specify their needs top down, redundancy and overkill can be avoided.

An other possibility to minimise the amount of metadata on activity record level is to use the information package structure to putt all common metadata on the causa record level.

## Business processes

In the computer world business activities are described in an abstract business process model. The model describes:

1. The sequences of activities
2. Incoming data
3. outgoing data
4. Rules
5. Resources

The business process model will define the causa, its activities and workflow utilised and produced information.

Not the meaning of “process” in this context. It is different to that of used in reference of executing programs. Here is abstraction of a working procedure.

# Documentation of activities

The is a major difference to just document an activity or create an evidence of the activity and its result. To day most computer systems are designed just to document the result from an activity. This may not be enough, we may want to produce evidences, that can satisfy the enterprise need for product liability in court. In this case systems development and information storage must also fulfil the rules from record management.

## Include record management in system development

Record management is the area of knowledge how to handle the life cycle of a business record. It must be handled in a way that the evidential value is maintained. It is an area that will be more and more regulated from the authorities. The rules must be included in the design work at an early stage.

**Scope of identification of business records, identification of controlling models for metadata and their naming**

The scope of metadata management is:

- Define ways and responsibility to identify metadata / data
- Describe the structure of an information package and of its business record
- Minimise the amount of metadata necessary at the lowest level of activity records
- Allow mapping from the enterprise product model to the local structure of the metadata for navigation purposes
- Support the user to interpret and understand the stored business records
- Ensure that the stored business records contains the necessary information to achieve the necessary evidential value
- Support migration to achieve long term maintainability

The control system shall be a part of the **user support environment** to assure that business records are created with the necessary quality, see Figure 25 and Figure 26.

No specific model how the different metadata elements relate to each other will be established. The description of information objects will be found in the “Core Information model” or in “local information models” or in the “entity frame work”. Metadata is derived from these models and it shall always be possible to determine the context of the data (i.e. the model from which it is derived) by the unique identification rules applied to the data and metadata.

A process model for controlling the identification of metadata shall be established An information model for naming defining group wide rules on naming (including identification of synonyms and homonyms).

**List of issues to be supported by the process and information models**

a) Identify data / metadata for a business record, related to the different demands from the top level of the enterprise down to the producing units. Each level must define the minimum metadata, necessary for their scope.

b) This data / metadata shall allow mapping of ways for navigation from the enterprise product model to the local structure of the *causa.* This will make local structuring possible on the lowest level in order to minimise the need for top down structuring.

c) Minimise the amount of metadata necessary at the information object level and maximise the amount that can be put at the *causa* level.

d) Support the user to interpret and understand the stored business records e.g. context information identification r description.

e) Support information retrieval so it can be as precise as possible.

f) Support long term maintainability

g) Support the stored business records with necessary information to achieve business records that have the sufficient evidential value.

h) The control system shall be a part of the **user support environment** to assure that business records are created with the necessary quality. (Support that mandatory and necessary optional metadata is identified and support users with already known metadata and document embryos.)

i) Usage of names for terms as metadata

j) Define in which business record a specific identification is to be used

k) Define optional and mandatory terms in a specific business record

l) Define a suitable way of storage structure and implementation that gives a long term evidential value in countries with different administrative traditions. It must be decided if a certain information can be stored in a database or has to be stored in an electronic muniment

## Business records in information modelling

Traditional systems design includes often only operational aspects on library information. This will still be the operational part of the combined model, showing the operational aspects from the users point of view, see Figure 13 *e* and *i*. This information in not sufficient to create a business record (see Figure 4) with evidential value. It must also contain information concerning the organisation and used tools as well as a description of the activity that created the business record see *a*. Information concerning organisation, tools and ruling regulations and Standard Operating Procedures (SOP) are rather static. This information can preferably be hold in centralised archive to which other operational business records can make references, see Figure 13 *b*, *c* and *d*. The activity is to define*d* in a process model see Figure 13 *f* and *g* and documente*d* in *j*.

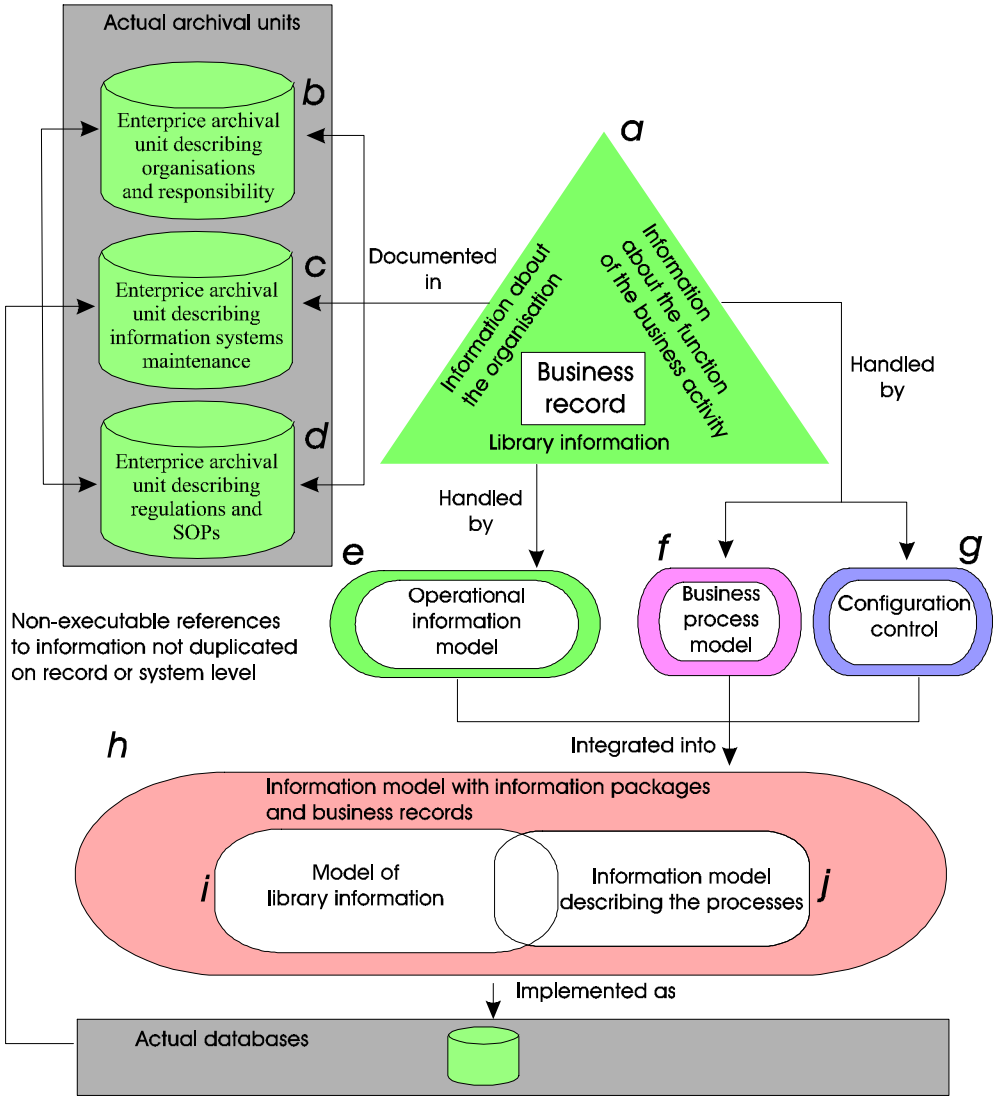



*Figure 13 How to model and document a business record*

Rules, internal and external, are an important part of record management. External rules have to be interpreted and defined, Figure 13: *d*. That will effect not only the system development but also how a system is operated and maintained. The planned activity in partly a part of the rules and can, if they are standardised, be documented in the rules database. The actual database must be documented in the operational database.

**Information model of a business record**

Scope of the information model

The model shall show the relations of involved information components from a system design perspective. The principle is applicable for both database storage and storage of electronic muniments. The business record is in this model is

represented as one physical record, but it may its own complex structure related to the operational need. See Figure 14.

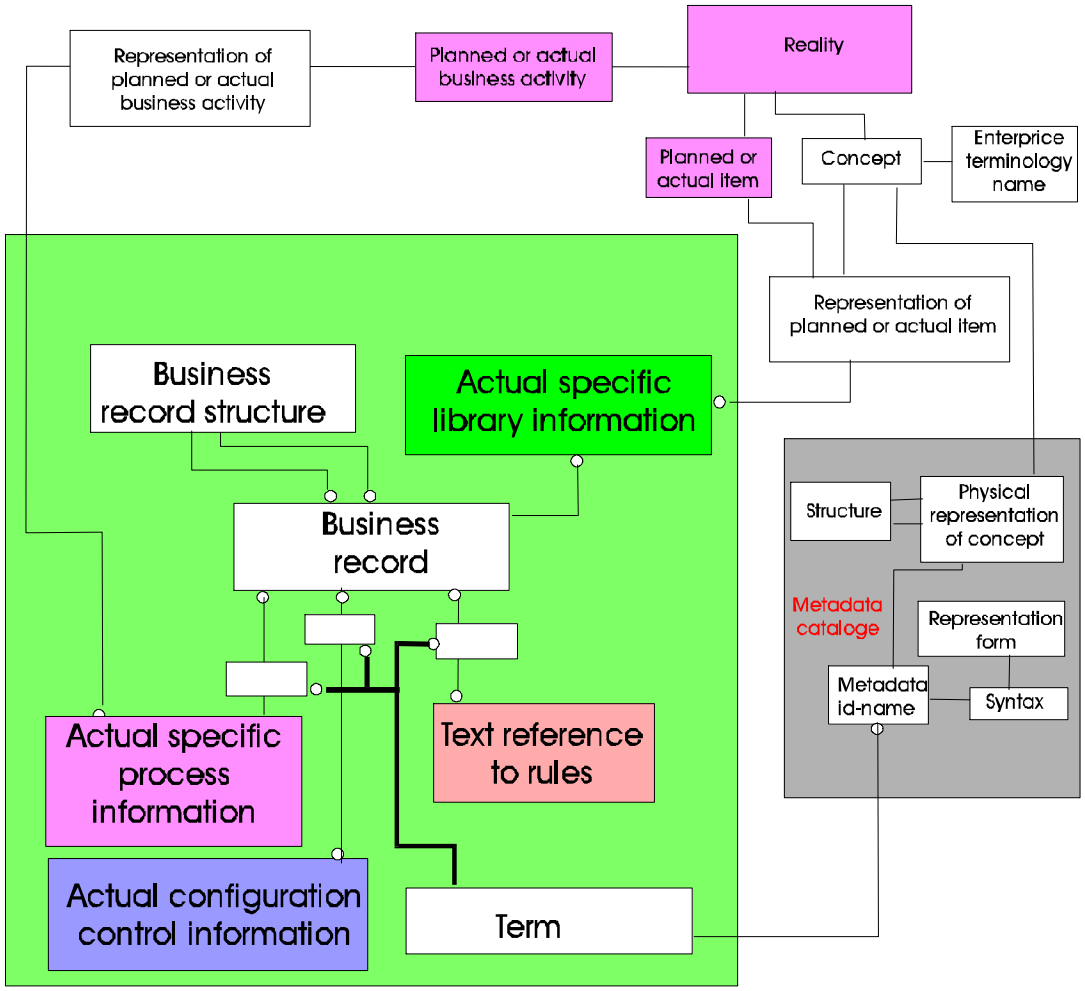



*Figure 14 Generic information model of a business record with its metadata and library information*

**Which entities to model in order to achieve a business record**

The way information is modelled is critical: Can the derived application and its database contain and handle **business records** or not. Two examples of different attempts that fulfil the requirements have been identified: STEP based Data Management for Petrochemical Industries[5] and The Product Model -CALS Technical Goal[6].

[5] Publication for CALS Europe '95, given on Oct 5th at Hamburg, Germany

[6] Published by FMW CALS Office in co-operation with EuroSTEP AB and Försvarsmedia AB

The model of a business record described in Figure 15.

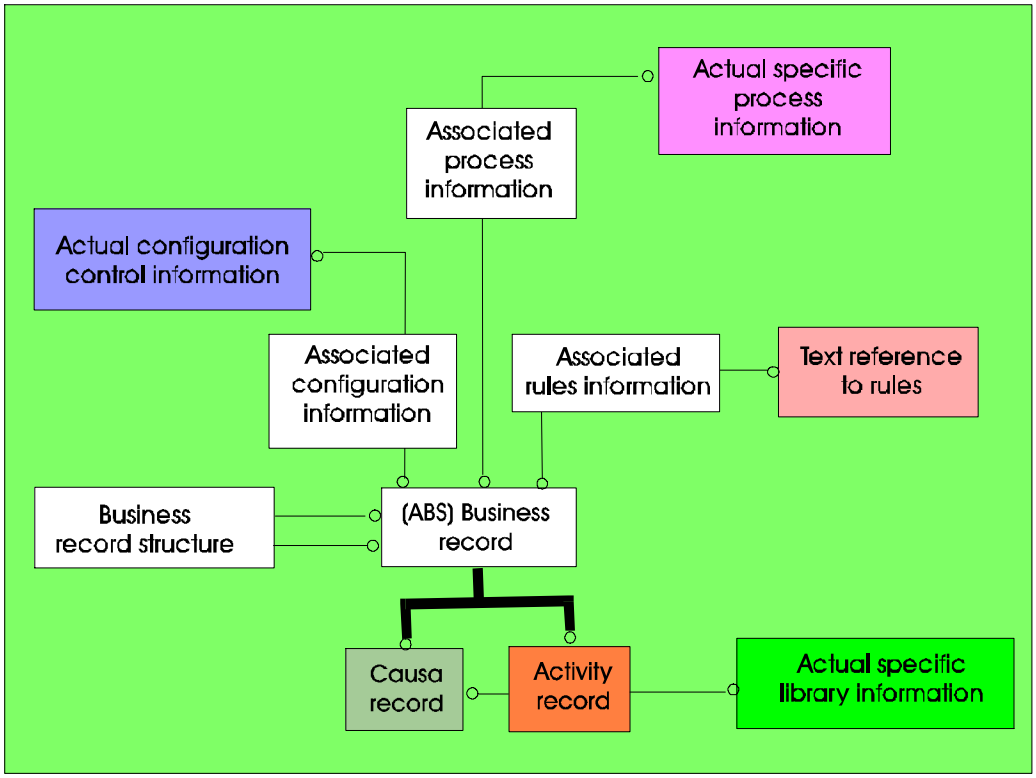



*Figure 15 The entities of a conceptual business record*

This model must be combined with the model of business activity and item at a conceptual level, compare Figure 10. The combination of the model of a business record with a the way of doing information modelling to achieve definition of all the necessary information, is described in Figure 16.

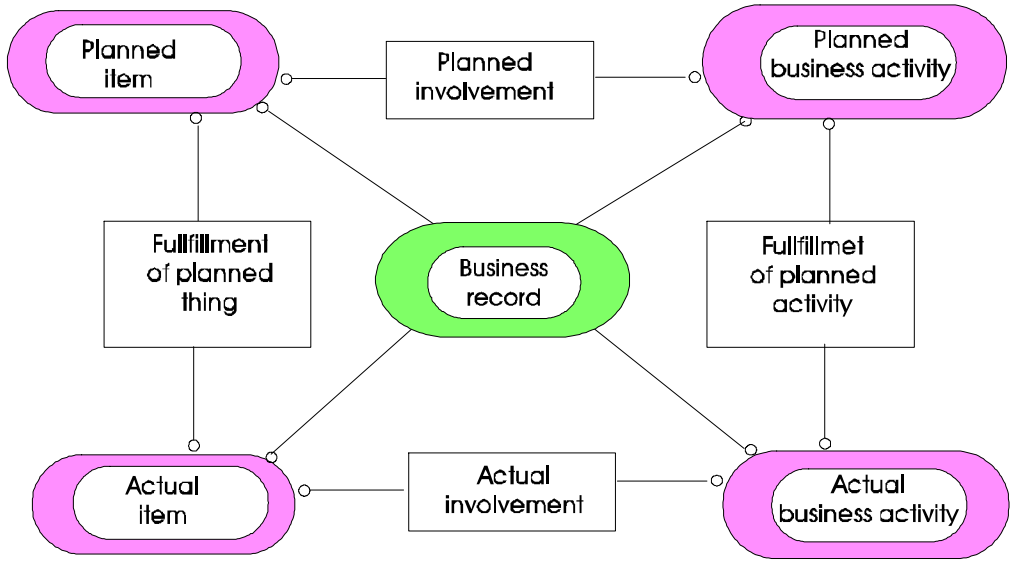



*Figure 16 Aspects of reality captured in a business record and their relationships*

# How to handle rules

The product liability is of great importance in several industry groups. The aerospace and pharmaceutical industry are heavily regulated by specific agencies. Their product liability is extent. A consequence of this is that internal rules, SOPs, must reflect law and legislation from various countries. This kind of operations will be necessary for all types of industries having a multinational product liability

## The process of keeping internal rules up to date with external demands

It is necessary that the information models and process models are enabling and expressing the requirements of all relevant user categories of the enterprise. This must include in a more active participation from the Legal Affairs and Regulatory Affairs and records management.

The process and information models will help the users to understand the requirements of other categories. All requirements must be sufficiently detailed for use in system development. At this point of the analysis, the information and process models must **not** include aspects of the actual technical implementation. Structuring must be related to the user's view of reality.

If laws and regulations are not interpreted correctly, generic models will lead to a false sense of security and to widespread ("generic") defects. It is important to

know which parts of the regulations are not included in the models. Those demands must be part of the system's program code.

To create correct models, it is essential to:

- Define the scope of the archives
- Employ skilled legal expertise from all jurisdictions concerned
- Properly implement external (mandatory) rules related to the scope
- Ensure that legal requirements are superior to (override) internal business rules, if this is necessary
- Build a legal risk assessment into the analysis model (what documentation is likely to be required in case of a judicial proceeding)
- New legal demands must be constantly monitored on an international scale (knowledge-base of legal text {including interpretation}). See Figure 17.

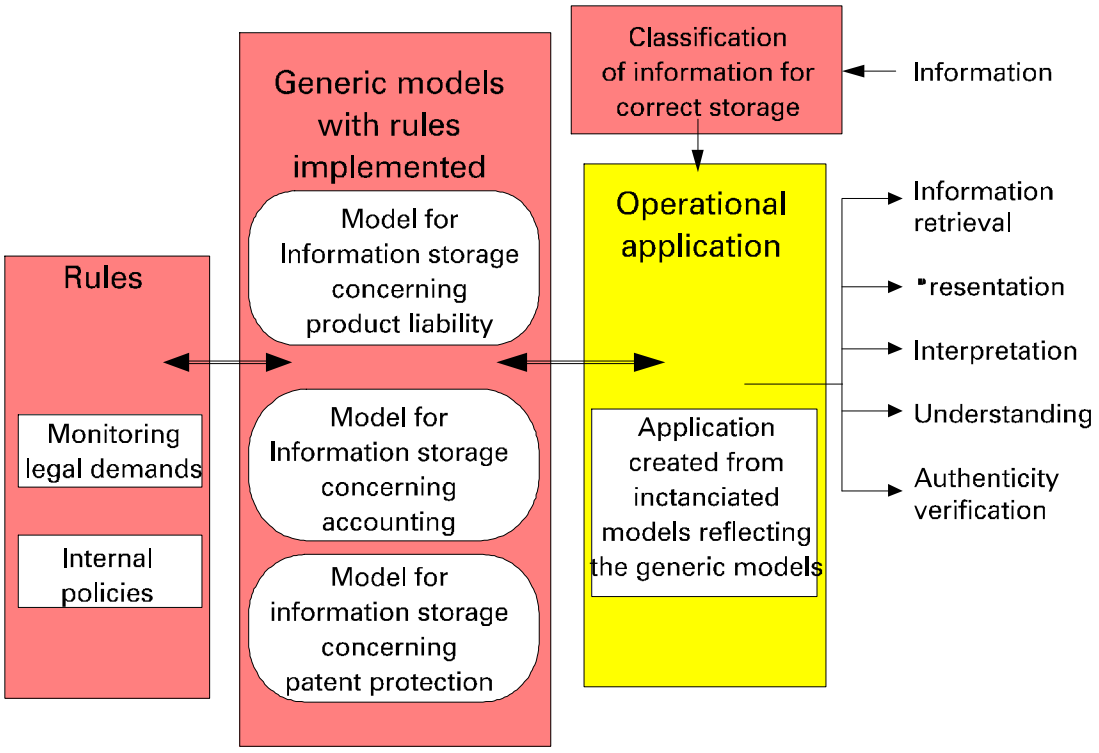



*Figure 17 How to monitor and implement changes in legislation from various counties into operational and archival systems concerning product liability accounting and, patent protection*

In practice the work will be performed in an interactive way. Sketches of policies will lead to modelling which in its turn leads to redefinition and refinement of policies. The formal requirements part will take on different character for different contexts - The product liability context is more difficult to deal with than the accounting context. Compare Figure 13, the rules stored in *d* will effect *c*, *d*, *e*, *f*, *g* and *h*.

## Legal aspects on storage and physical distribution

Both the logical and physical structure is important for maintenance of a record. Both the way it is created and maintained must be understandable for non IT professional. This possibility to understand and thrust a business record will effect the evidential value in court. There will be major differences, between countries from deferent cultural areas in the world, on what they regard as reliable.

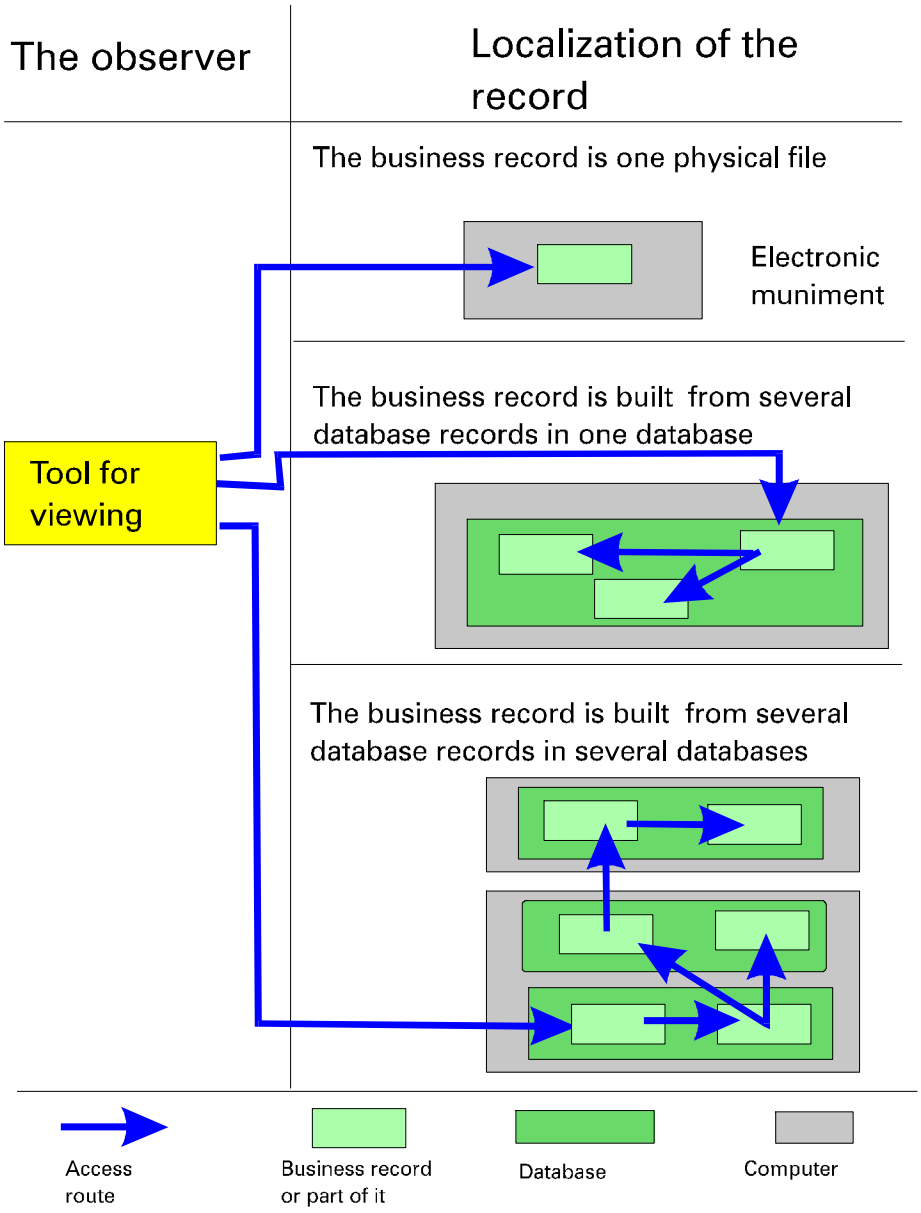



*Figure 18 Physical storage of a business record*

# Concepts required as metadata by every business record

Traditional information modelling concerns operational functionality of the user's information. In order to assure that the stored information is valid business records with evidential value some additional concepts have to be instanciated in the model. See Figure 4 The content of a business record.

The information we require for all business records covers areas such as identification, transactions etc. The amount of metadata necessary for an activity record is mainly a subset of that for a causa record. The causa record gives a possibility to store metadata common for all activity records with the same context. This allows us to reduce the amount of metadata replicated on a number of activity records.

The concepts to handle is shown in Figure 19

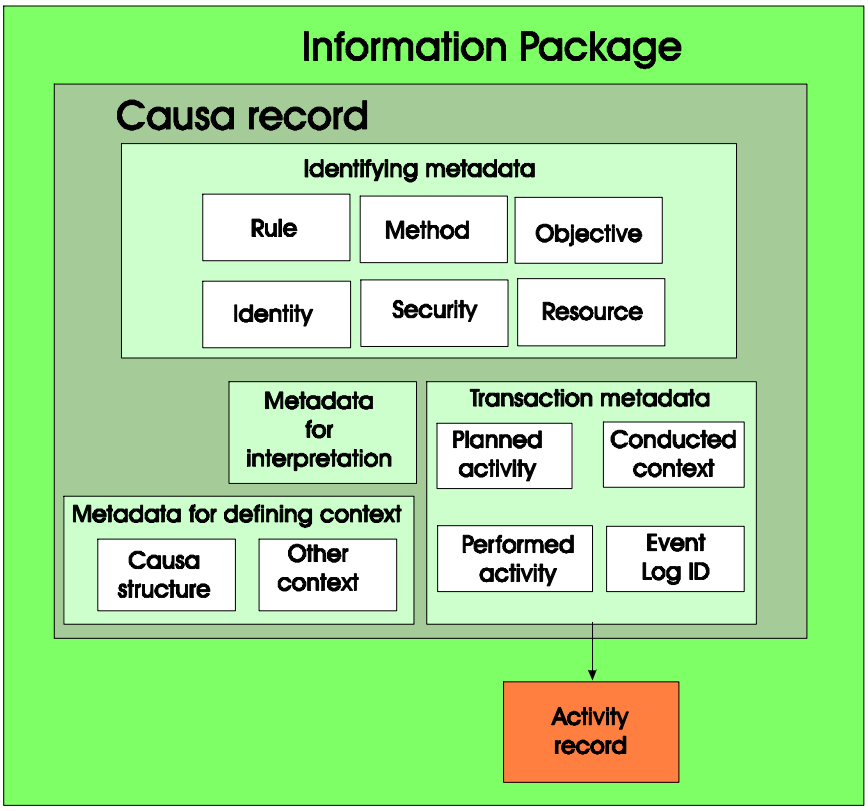



*Figure 19 Demands on metadata for a* causa record

## Identifying metadata

Metadata necessary for identification of business records

| **Objective:** | **The aim of the business action** | **Examples** |
|---|---|---|
| Rule: | External | Laws and regulations |
| | Internal | Policies<br>Guidelines<br>SOP[7] |
| Planned business action | The activities necessary to conduct a specific business action in accordance with the stated rules | . |
| Actual business action | The actual activities of an actual business action including explanations for deviations from rules | |
| Resources | The resources (materials and personnel) necessary to conduct the business action in accordance with the stated rules. | |
| Method[8] | How to conduct the business activity to fulfil the aim of the business activity | |
| Business record structure | Information about owning business records and records owned | |
| | | |
| Security | Metadata necessary for access control | |

[7] An SOP could also be regarded as an planned typical business action (less precise compared with a planned specific business action)

[8] Could be regarded as a SOP (rule)

## Metadata for interpretation

Metadata necessary for interpreting and maintaining the actual specific library information associated with the business record.

| **Objective** | **The aim of the business action** | **Examples** |
|---|---|---|
| Syntax | | e.g. SGML, STEP part 21 file |
| Format | | e.g. ASCII, Unicode |
| Natural language | | e.g. Swedish, English |
| Free text | Descriptive text in the specified natural language | |
| | | |

## Metadata for defining context

Metadata necessary for understanding in which context the actual specific library information was created

| **Objective** | **The aim of the business action** | **Examples** |
|---|---|---|
| Necessary parts of the process model including relevant business rules | Description of how this action relates to other actions | |
| Necessary parts of the information model including relevant business rules | Description of the structure in which the item is used or has meaning | |
| Creating organisation | | |
| Other business rules | | |

## Transaction metadata

Metadata necessary for defining how a business record is created and how it has changed over time.

| **Objective:** | **The aim of the business action** | **Examples** |
|---|---|---|
| Planned actions | The relationship of the planned actions | |
| Actual specific actions: | The relationship between planned and actual actions (control purpose) | |
| Version and history of business records | | |
| Electronic signatures | | |
| Event-log | Logging to sequential write once media (WORM) in order to support authenticity | |

# Organisational responsibility

**Introduction**

In order to access business records created in different operational areas within an enterprise it is necessary to have a consistent definition and naming convention. If no harmonisation happens the transferred information will be miss-interpreted and information retrieval will be inaccurate.

**Mission**

Metadata should be defined by the business area or operational area that has the concept as a part of their information or process models. The definition made, could then be used as a definition on any of the three levels of Figure 20. This concept should be named in accordance with the enterprise naming conventions.

**Restrictions**

The amount of metadata that is added to a business record and not directly related to the operational need should be as little as possible. The local operations will very easily drown in the proposed volume of metadata. It is important not to extend the scope for metadata described on page 42.

The described models and procedures can also be used for identifying and naming of data which is to be understood in more than on business area. In this case the volume of terms implemented is related by the business needs and its cost-benefit.

## Implementation

Information must be possible to be transfer between different environments without degraded evidential value. To achieve this all handling of information, including creation, within an enterprise must be based on rules. The rules must be based on a top down pyramid, where one area is responsible to add rules that are necessary for all common operations below, See Figure 20.

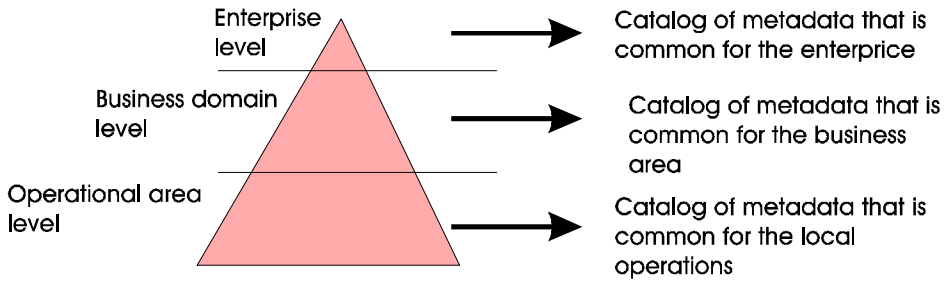



*Figure 20 The three different levels where metadata is to be controlled*

**List of business domain responsibility of adding metadata to an information package**

The following list will define responsibility for defining metadata to the different structures of a *causa.*

**Metadata for business record**

**General for enterprise**

**Not related to a specific business area**

Necessary for maintenance and evidential value

**Term owned by a specific business area**

Maximum 5 to 10 for each business area

**Specific for an business area, owned by the area**

Used only within the area. No limitation in number, No local variation allowed within the area

**Specific for part of an operational area, owned by the area**

Used only within a specific part of an area. No limitation in number, Local variation allowed within the area

## How to derive metadata

Metadata on a business record is always relating to information objects defined in the information models of one business areas of the enterprise.

**Relations between the information models of the business areas and metadata**

The business record is related to the activity that it is describing. The activity is defined in the process model.

The metadata terms are terms deviated from the process and information models of the enterprise. The naming of the term can vary depending on the syntax used to carry the term.

Figure 21 shows how metadata is derived from the information models to the metadata catalogue and then added to the business record.

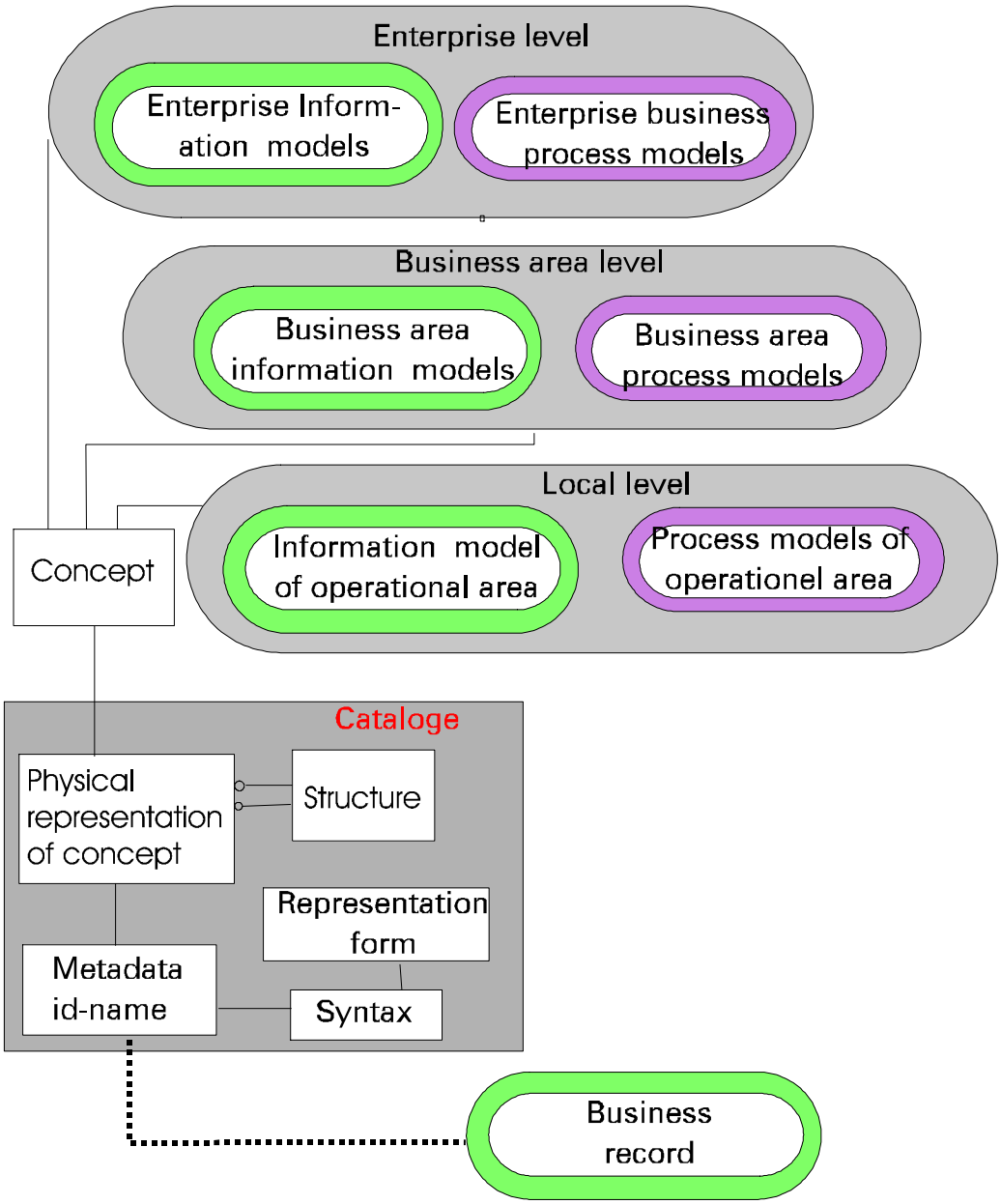



*Figure 21 The relation between metadata and information models and process models*

**Process model how to derive metadata**

Scope of the process how to derive metadata

Figure 22 shows how to derive metadata from process and information models at the enterprise to achieve the 3 categories of metadata shown in Figure 20.

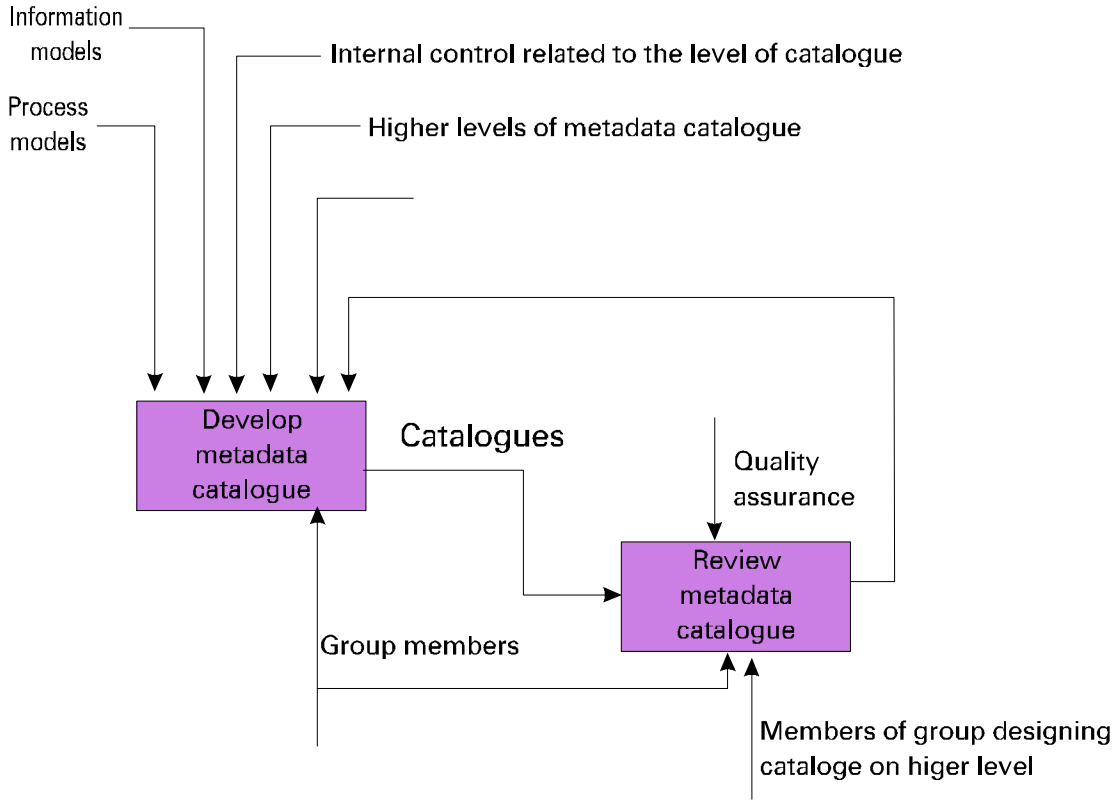



*Figure 22 Process model how to derive metadata*

Activities of the process

Identification of which terms that are used in between business areas or are necessary for information retrieval and understanding in between the different business areas. See Figure 21

Identification of terms necessary for evidential value and long term preservation of a business record.

Identification of external rules and regulations including risk assessments

Identification of external terminology

Unique naming, related to syntax, internal use and external receiver

Input

Output

There catalogues of common metadata For the enterprise level only 5 to 10 items are to be selected for each business area

Resources

The following disciplines must be involved:

1. All business areas
2. Legal affairs and regulatory affairs
3. Records management
4. Information modelling
5. System designers of existing systems

Control

Organisation and functional knowledge of the enterprise, existing information models, data models and process models

*Start criteria*

End criteria

## Process model making a conceptual information model with business records

**Mission**

Process model of how to develop a traditional operational information model into a conceptual information model (see Figure 16) with business records with evidential value (see Figure 15)

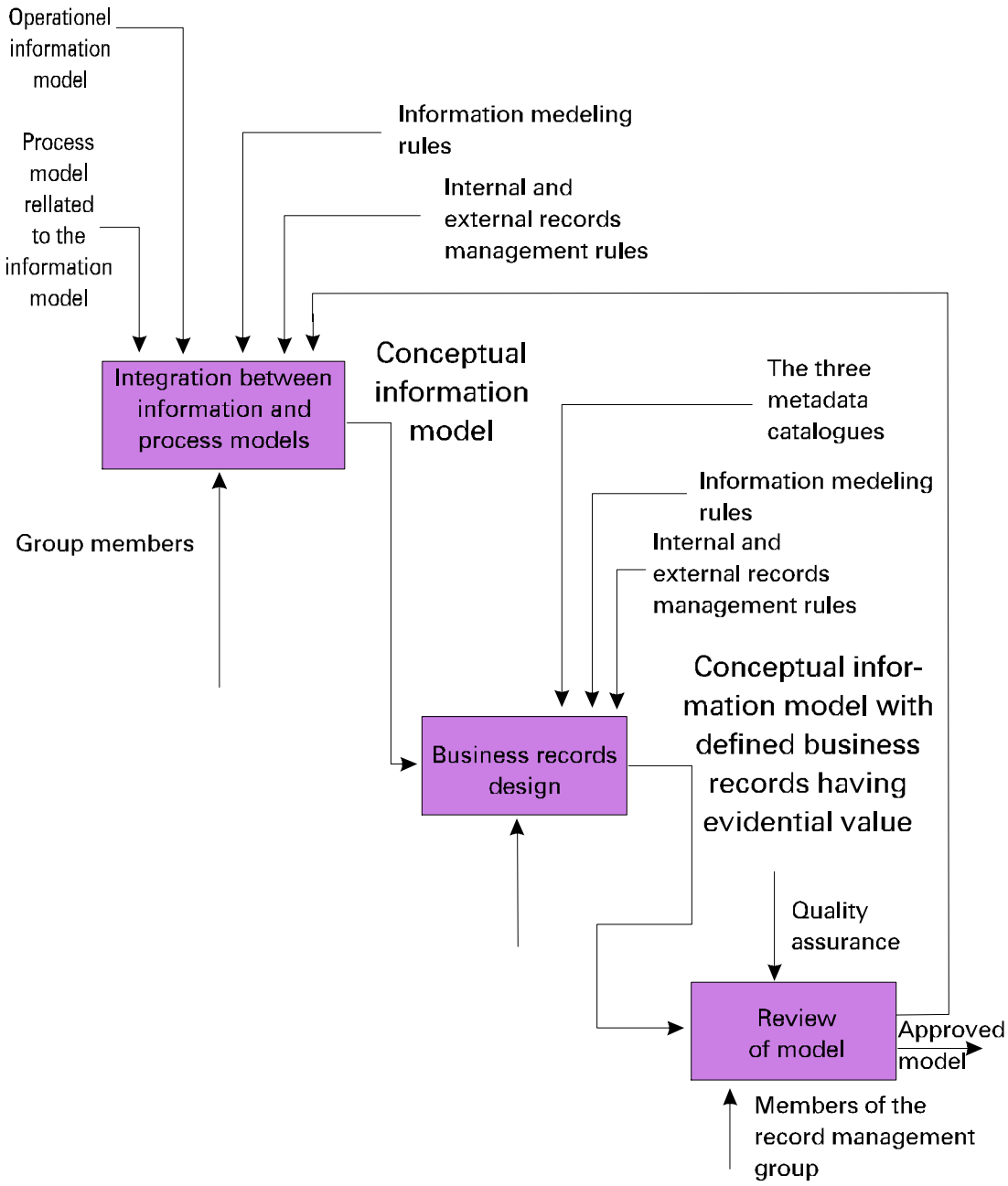



*Figure 23 Development of conceptual models with business records*

**Integration between information and process model**

Scope

Make an information model in which the business records can be identified.

Activities of the process

Identify:

1. Internal rules and regulations including
2. Information interchange, internal and external

Risk assessment

Different record storage will have different evidential value in court, see Figure 18. There will be major differences between different countries. The benefits and costs must be evaluated. Three possibilities are available.

1. A pure DBMS solution will satisfy the need of evidential value of stored records
2. It is necessary to create electronic muniments. No DBMS solution is necessary
3. It is necessary to create both the electronic muniments and a DBMS solution

Input

Output

**Conceptual information models in which business records and information packages can be identified. Two different outputs are possible if electronic muniments are necessary or not:**

**Traditional conceptual information model defined by using the EXPRESS language. This model contains complete business records.**

**If electronic muniments are found necessary, those parts of the model that are defining the causa record and the activity records must also be expressed in a way that includes a presentation sequence, e.g. a DTD for SGML. In this case, the SGML instance is to be created at first by the uses. This action includes creation of the electronic signatures. The information of this SGML instance is then to be loaded into the operational database. The database in this case does not need to contain complete business records, since this storage is not aimed to have evidential value.**

Resources

The following disciplines must be involved:

1. The involved business area
2. Information modelling
3. Legal affairs and regulatory affairs

Control

Record management rules.

Operational information modes and process models.

Start criteria

End criteria

**Business records design**

Scope

This is the process model for final amendment of metadata to an information model. It will show how to apply metadata to the different parts of an information package.

Activities of the process

Input

Information model with defined causae and business records

Output

Information model with satiated terms of metadata added to the defined causae and business records

Resources

Control

Two types of conceptual models are possible

1. Traditional conceptual information model defined by using the EXPRESS language. This model contains complete business records.
2. If electronic muniments are found necessary, those parts of the model that are defining the causa record and the activity records must also be expressed in a way that includes a presentation sequence, see Output on page 54

Start criteria

End criteria

**Review of model**

Scope

The models for control of metadata objectives will be used as rules to control applied metadata in the information models.

**Activities of the process**

Identification of:

1. Relevant rules, internal and external
2. Record life cycle
3. Used syntax

Input

Information model containing defined business records and metadata

Output

An, from records management, approved information model

Resources

The following disciplines must be involved:

1. The involved business area
2. Necessary legal and regulatory knowledge
3. Information modelling

Control

Control schema for metadata see: Concepts required as metadata by every business record on page 43

Start criteria

The functional information model in which business records with metadata is completed.

End criteria

The model is approved by all included disciplines.

# Life Cycle Management

The above described metadata (list of values) and metadata naming needs change management where:

- all historic data is stored, even ‘misspelling’
- all changes are traced
- all changes are kept as proposal of different state until it is approved on a certain date. No draft versions shall be maintained after a version is finally approved
- all product data is viewed by date, normally current date is used and then toady’s products and product data is displayed, or a time interval normally from current date to maximum date (9999-12-31) could be used then also all future changes displayed.

# IT architecture

## System architecture

The major difference from existing system structure is that we clearly have to separate these three different areas:

1. **Information production environment** Information must be produced and stored with the necessary quality to achieve evidential value.
2. **Archival systems** Those system must maintain the business records in a way that the evidential value is maintained. Business records will is to be collected here for future re-use
3. **Re-use of information** Information will be re-used either for simple information retrieval, warehousing used for new and complex analyses or for creation of information assemblies e.g. NDA.

A proposed implementation architecture is shown in Figure 24.

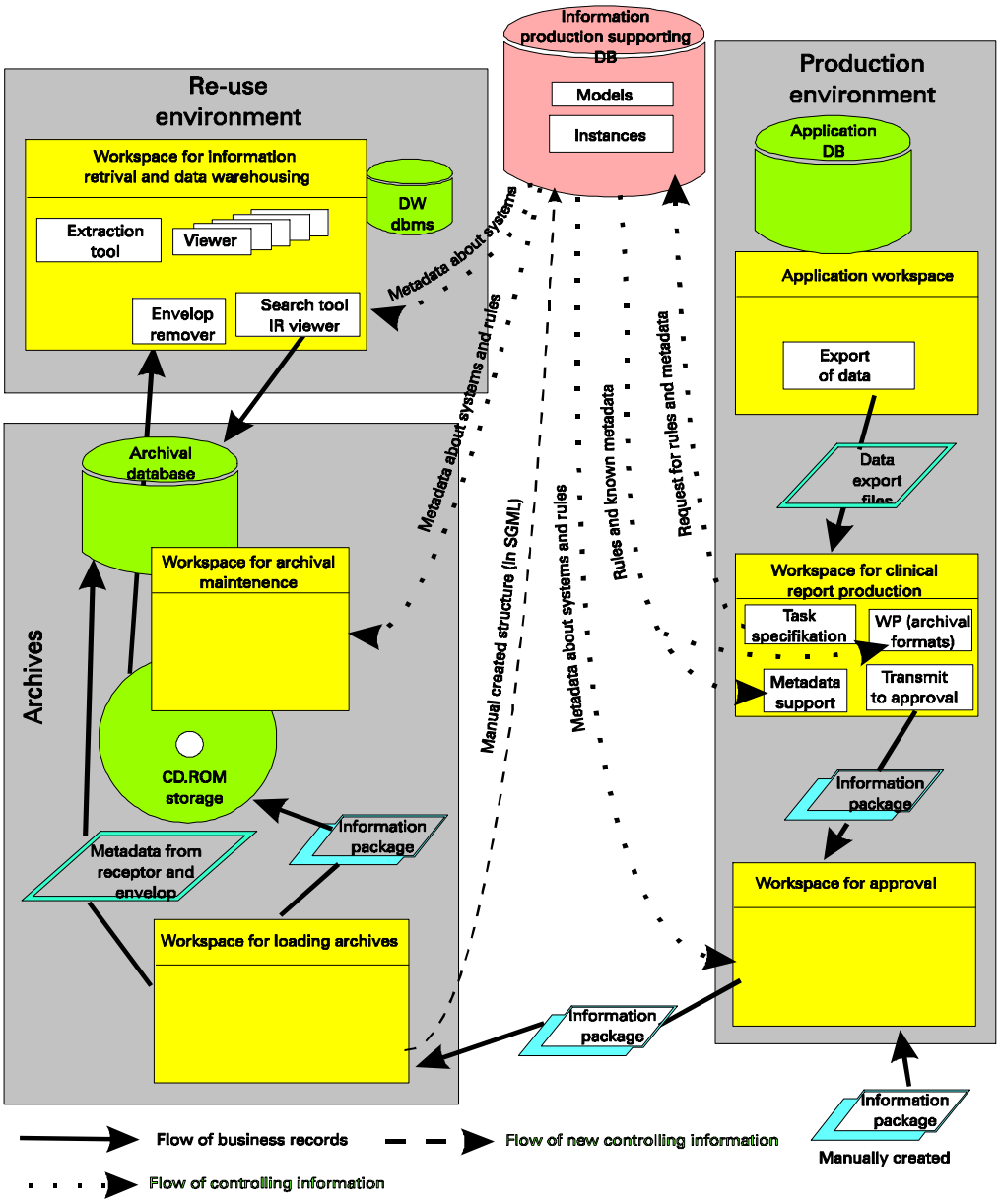



*Figure 24 A proposed implementation architecture supporting the three areas*

## How to produce business records with the correct metadata

The quality demand to achieve the necessary uniformity in structuring and identification will force us to produce systems that both support the users and control the information they produce. If each project or company will solve information management in proprietary ways the difficulties in sharing, re-using and maintaining information within the enterprise will escalate even more.

An example, how the supporting environment that be structured in an R&D environment is described in Figure 25.

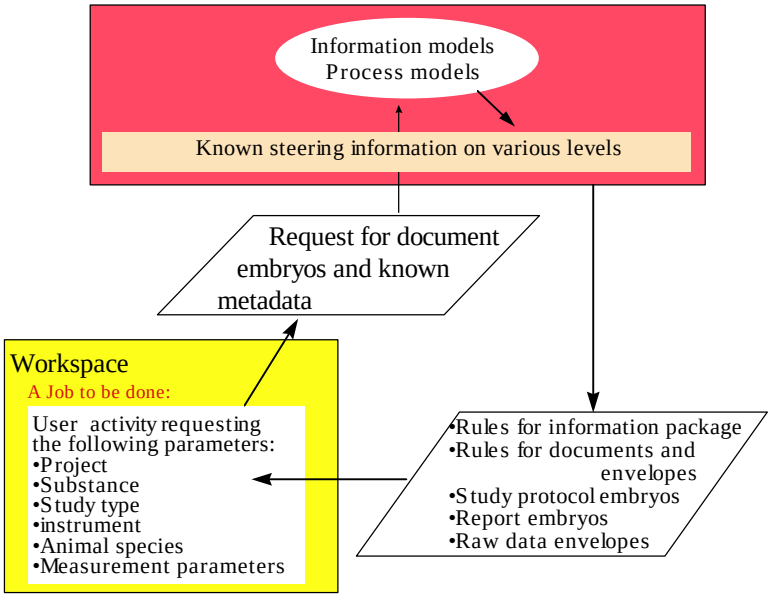



*Figure 25 How to use known rules and known information to support users to create business records with high quality*

# Conclusion for the IT environment

To be able to transfer information between different business areas and IT environments it is necessary the all the environments involved can relate back to common models of information and processes. An example of the implementation of those rules in the IT environment is shown in Figure 26.

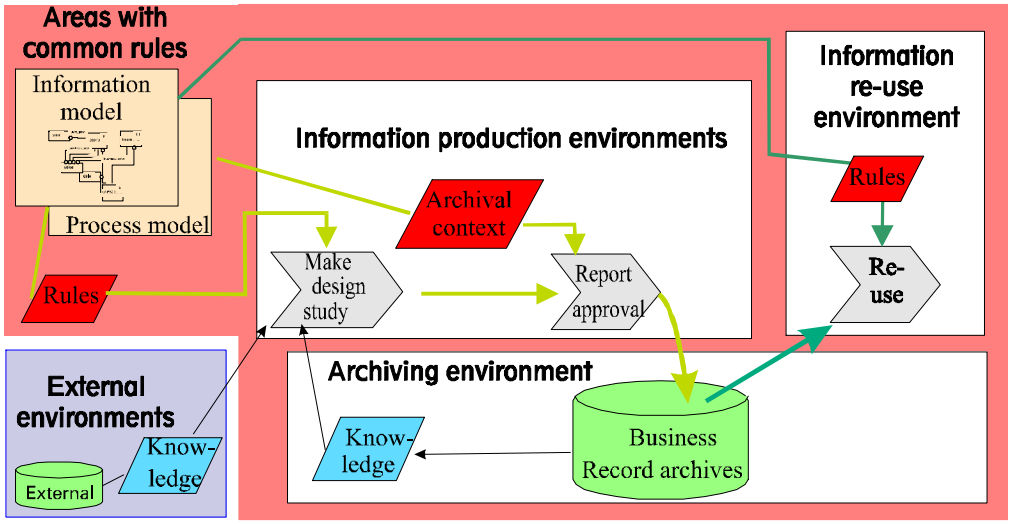



*Figure 26 Utilisation of common rules*

# Table of figures