CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

FINAL DEMONSTRATION REPORT on Taxation Metadata in Secondary Data Collection - How to connect the metadata of taxation to numeric taxation data and use them at the same time

1. Background

In Finland National Board of Taxes (Taxfin) delivers taxation data of indivual level to the Statistics Finland (Statfin) to produce statistics. In this paper an experiment that is made in Codacmos-project is described. In this experiment metadata that describes taxation is connected to the taxation data. Experiment is made by using personal taxation material and its metadata. National board of Taxes, Statistics Finland and TietoKarhu have been involved in this work.

1.1. The Situation today

The data in Taxfin is saved in relational databases. The metadata that describes the data in relational databases is described in handbooks and partly also in relational databases and in applications that describe the relational databases. The aim of this project is to change the metadata into structural electronic form and to connect it to the data in relational databases by linking.

1.2. Metadata sources

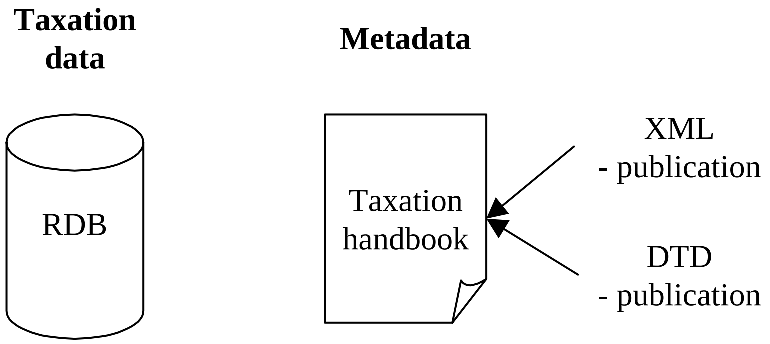
The main source for metadata is the handbook of personal taxation. This handbook is an edition of about 800 pages. Both Taxfin and Statfin can use it. In the book the whole procedure of personal taxation is described very detailed. There are also references to legal cases and to taxation laws. To produce this handbook of personal taxation there is a DTD-definition, which describes mainly the outside structure of the book. It consists also some structure of the contents. (for example legal cases and keywords). The structure does not follow the logic of taxation data. The text is written in XML so that the printed edition can be easily published from it.

There is also some metadata of taxation data saved in relational database or in applications that describe the database. In this project this data is not important.

The situation today can be described like this:

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Fig. 1. The present state

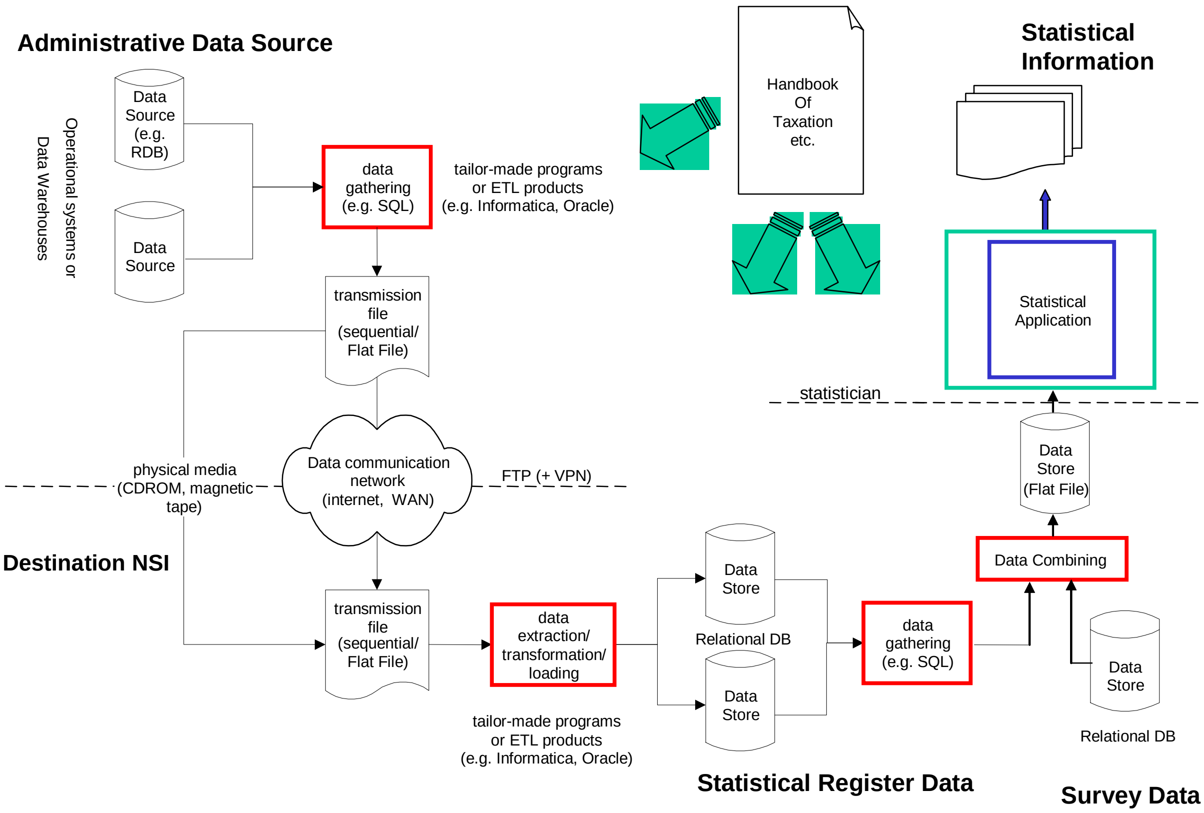


1.3. Moving the taxation data to Statfin

Statfin gets the data directly from the databases of Taxfin (see fig 2.). Statfin defines all the variables they want to get data of. According to these variables Taxfin gathers the material from databases and delivers it to Statfin as a sequential file. This file does not contain any metadata. To interpret the content and meaning of the data Statfin has to use the printed handbook.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Fig. 2. The present state of Secondary Data Collection



2. What is a tax

Tax is a payment of money required by law to finance public expenditure. Taxes can be categorised by the way they are being formed.

Different kinds of taxes in this meaning are:

- Progressive tax, tax applied at a rate which is higher the greater the income or wealth being taxed. Examples of progressive taxes are the tax on earned income, inheritance tax and gift tax.
- Flat rate tax, applied at a fixed rate regardless of the amount of income or other tax. Flat rate taxes include municipal income tax, church tax, corporation tax and tax on capital income.

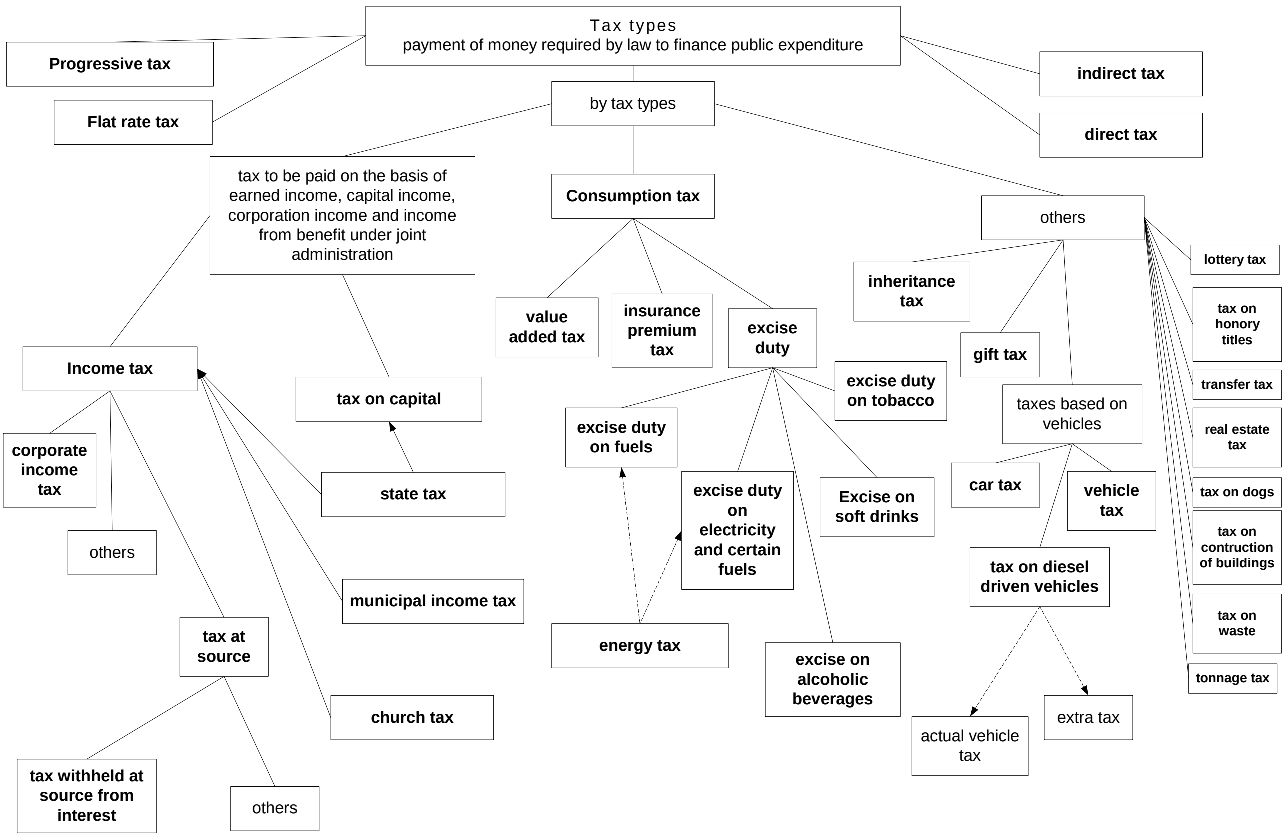
- Indirect tax is a tax that is passed to others through inclusion in the price of goods and services. Indirect taxes include value-added tax and excise duties.
- Direct tax is to be paid directly by the taxpayer. They include income tax and wealth tax.

2.1. Tax types

Taxes can be categorized also by type according to their origin. By source categorized tax types are:

- Income tax, tax to be paid on the basis of earned income, capital income, corporation income and income from benefit under joint administration. Income tax is to be paid to the state, to the municipality and to a parish of either the Finnish Evangelical Lutheran Church or the Finnish Orthodox Church.
- Capital tax (wealth tax) is levied on the basis of net worth. The net worth tax is paid to the state. It applies almost exclusively to natural persons and death estates. Partnerships and most corporations do not pay net worth tax.
- Consumption tax is tax levied on the consumption of goods and services. Consumption taxes are for example value-added tax, insurance premium tax, excise duty on alcoholic beverages, tobacco products or electricity and certain fuels.
- Others taxes are for example inheritance tax, gift tax, vehicle tax, lottery tax, transfer tax and real estate tax.

Fig 3. Tax types



2.2. Types and sources of income

The types of income are earned income and capital income. These are the bases for determining the amount of income taxes to be paid by natural persons and death estates.

2.2.1. Earned income

Earned income is other than capital income and examples of it are wages and pension

2.2.1.1 Salary

Salary = wages is remuneration based on regular employment or the performance of personal labour or services.

In Finland , for purposes of taxation , this concept includes meeting and lecturing fees as well as fringe benefits. In labour law, it denotes compensation for labour or services performed under management and supervision. In general Finnish usage, salary refers to regular payment for work and fee to one-off monetary compensation for occasional work or service. In English, the principal difference between the terms salary and wages lies in their connotations. Salary typically refers to regular compensation for white-collar work expressed as monthly or yearly earnings. Wages more commonly refers to regular compensation for work on a daily, weekly or contractual basis. The terms are, however, interchangeable in many contexts.

2.2.1.2 Fringe benefit

Fringe benefit is received from an employer in other than monetary form as part of normal wages or salary.

2.2.1.3. Pension

A retired person gets pension and not salary.

2.2.2. Capital income (investment income)

In income accruing from wealth. The forms of capital income include capital gains, income from interest, income from rent and stock dividends.

2.2.2.1. Capital gains

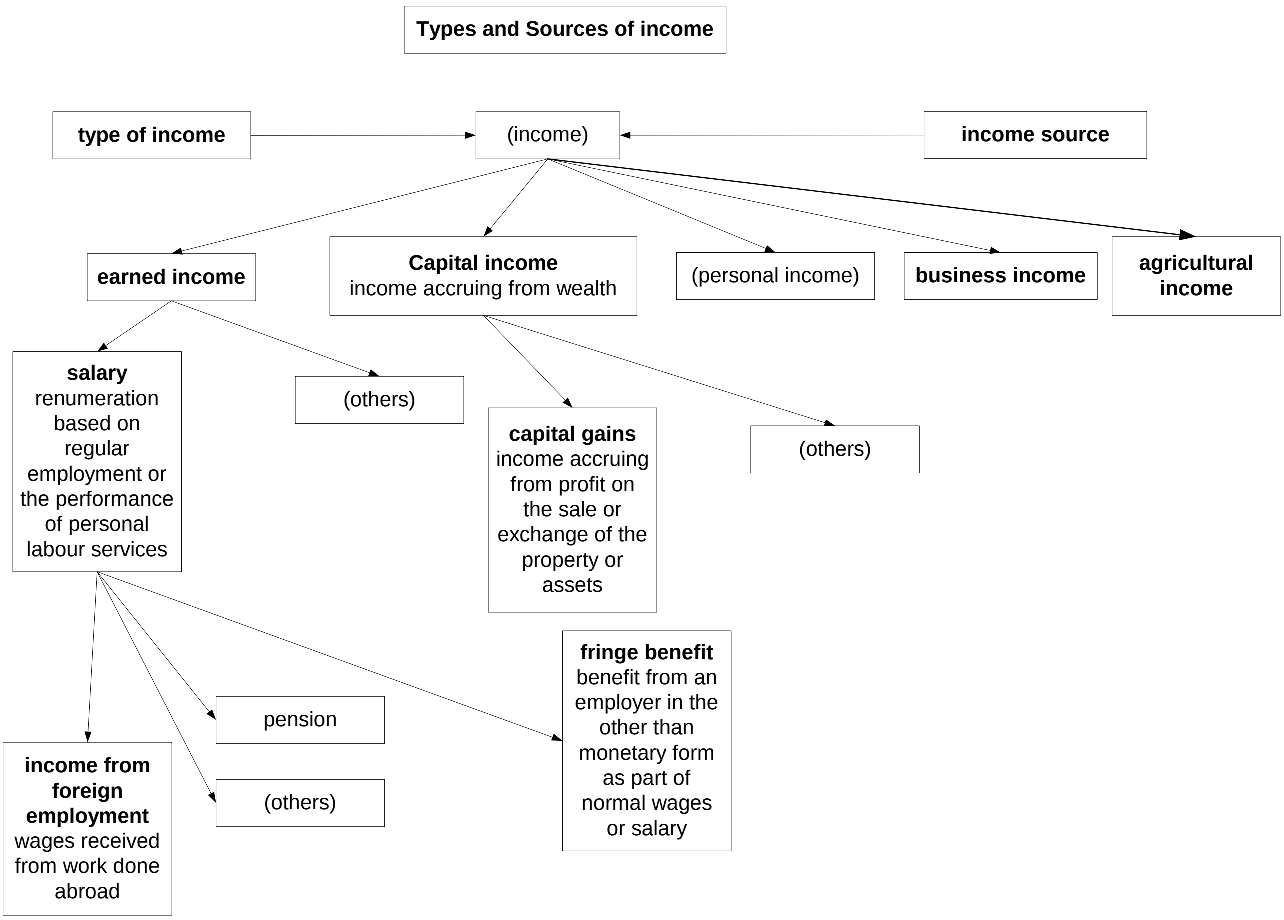
Capital gains is income accruing from profit on the sale or exchange of property or assets.

2.2.3. Income source

Income source indicates the activity in which the income of a taxpayer originates. A taxpayer may have three sources of income: business (business income), agriculture (agricultural income) or other activities (personal income). The source of income determines which law is applied in calculating a taxpayer’s taxable income.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Fig 4. Types and sources of income



2.3. Income tax deductions

A deduction (tax deduction, allowance or credit) is an amount deducted from tax, income subject to tax or net worth before final determination of tax to be paid. In some tax types, the deduction is granted on the tax basis.

2.3.1. Deduction from earned income

Some deductions are from pure income. These kinds of deductions are for example deduction for pension insurance premiums, discretionary allowance for circumstantial incapacity to pay taxes, pension income allowance, earned income allowance and low-income allowance.

2.3.1.1. Standard deduction for work-related expenses

Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

This is a deduction from wages granted ex officio by the tax authorities, replacing up to a certain sum the deduction for expenses incurred in acquiring or maintaining income. The standard deduction for work-related expenses is granted in state and municipal taxation.

2.3.1.2. Deduction for travel expenses

This deduction is allowed from earned income, determined on the basis of the costs of travelling to and from home and work.

2.3.2. Deduction from income tax

These deductions are made from the taxable income and they include for example: child maintenance credit, domestic help credit and credit for capital income deficit.

2.3.2.1. Child maintenance credit

A deduction from income tax granted in state taxation, based on child maintenance payments made by the taxpayer.

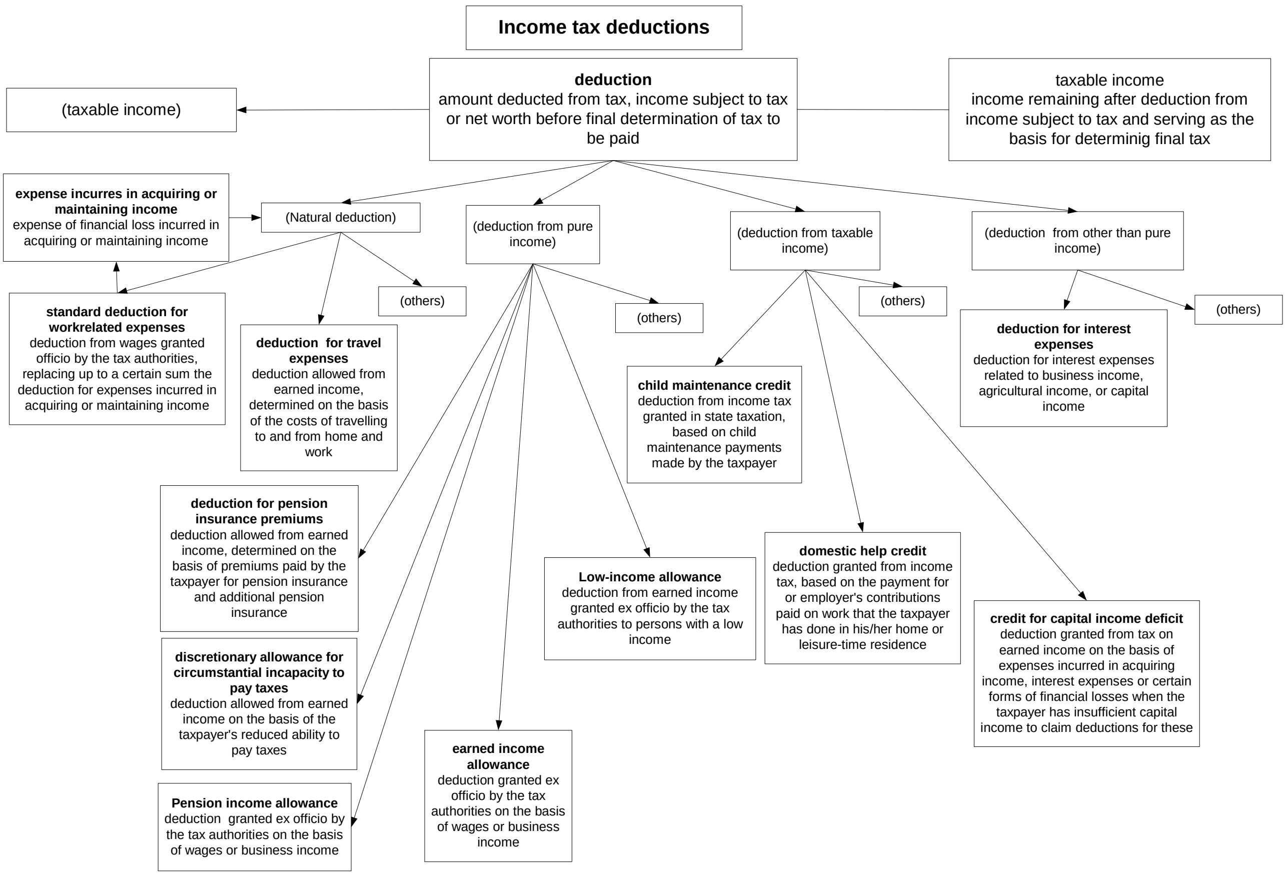
2.3.2.2. Domestic help credit

A deduction granted from income tax, based on the payment for or employers' contributions paid on work that the taxpayer has had done in his/her home or leisure-time residence.

2.3.2.3. Credit for capital income deficit

A deduction granted from tax on earned income on the basis of expenses incurred in acquiring income, interest expenses or certain forms of financial losses when the taxpayer has insufficient capital income to claim deductions for these.

Fig 5. Income tax deductions



2.3.3. Final income tax

Taxable income is calculated from income. Income – deductions from income = taxable income
Income tax is calculated from taxable income.
Final income tax, which is to be paid, is calculated from income tax:
Income tax – deductions from tax = final income tax

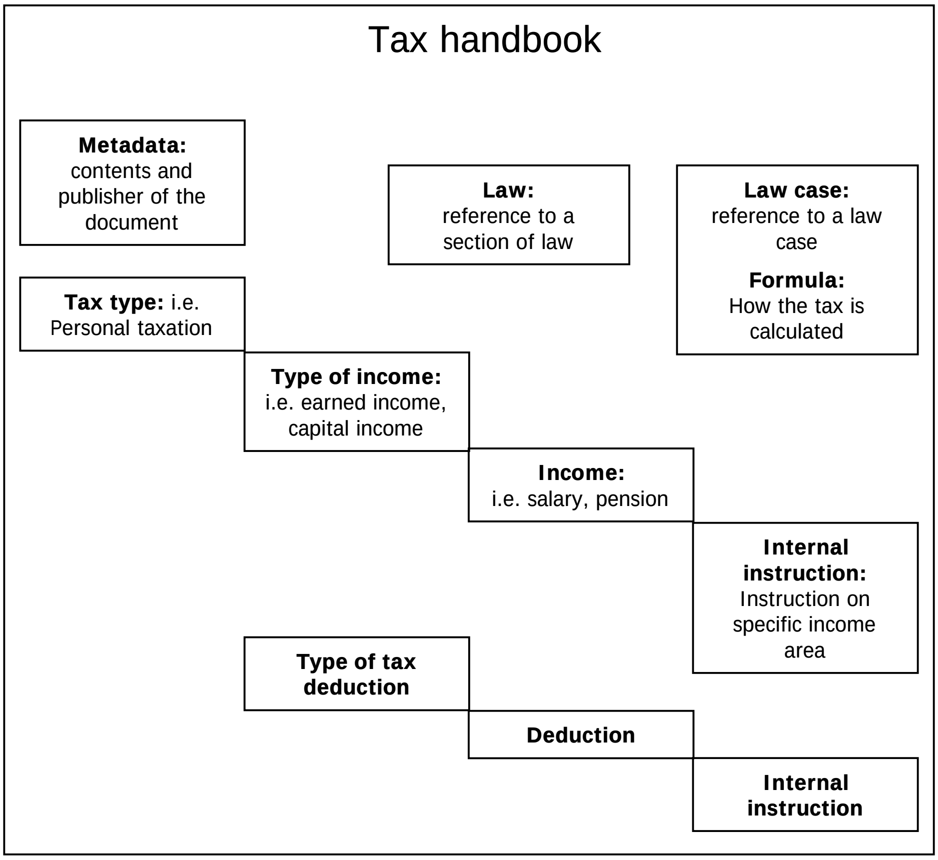
3. Structuring of taxation metadata

3.1. The logical structure of personal taxation metadata

The structure of the metadata connected to the personal taxation is described in the following picture.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Fig. 6. The logical structure of personal taxation metadata



The structure of personal taxation is hierarchical. From a common term (for example earned income) you can go deeper to a more detailed term (earned income -> salary -> fringe benefit) and vice versa. There is the same hierarchy in taxable amounts and in deductions that are made from them. In addition, you can connect descriptions, legal cases and references concerning the case in question. The document concerning taxation as a whole can be described with metadata also.

3.2. Structure definition (DTD) of the metadata of personal taxation

The logical structure of personal taxation is a basis for the implementation of the structure definition. The structure definition was implemented by DTD. It consists of three DTD-modules. The main module is tax dtd, which describes the structure of personal taxation. Two different parts are connected to this main module:

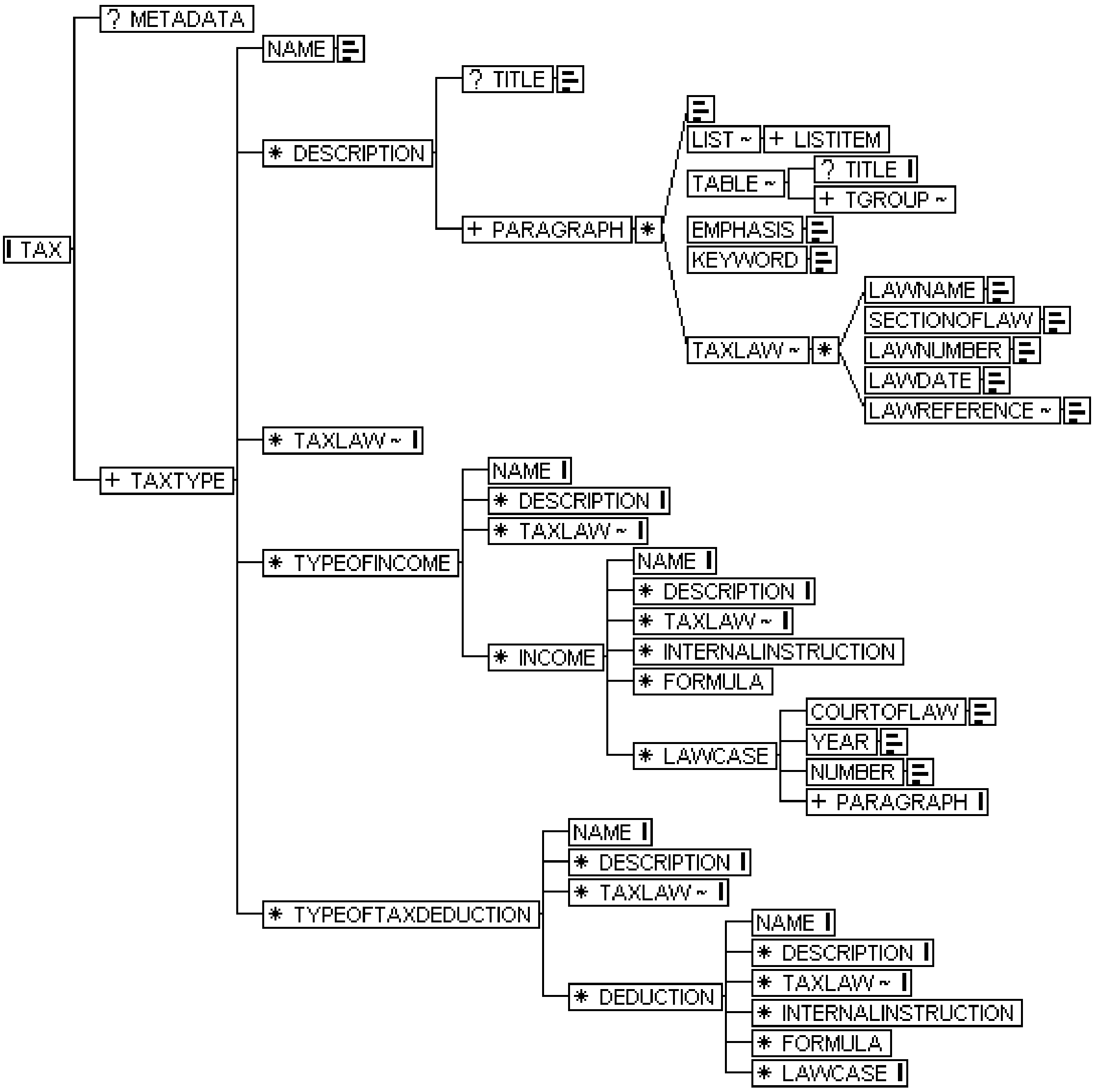
–Metadata module (metadata.dtd) is for the describing metadata of personal taxation handbook. In the metadata part there are described the writer, the subject, the publisher, the taxyear for example. This part is compatible with Dublin Core definitions.

–A CALS-tabledefinition (soextblx.dtd) is inserted for the tables in handbook.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

The first version of the element structure of the personal taxation metadata is described in pictures 7 and 8.

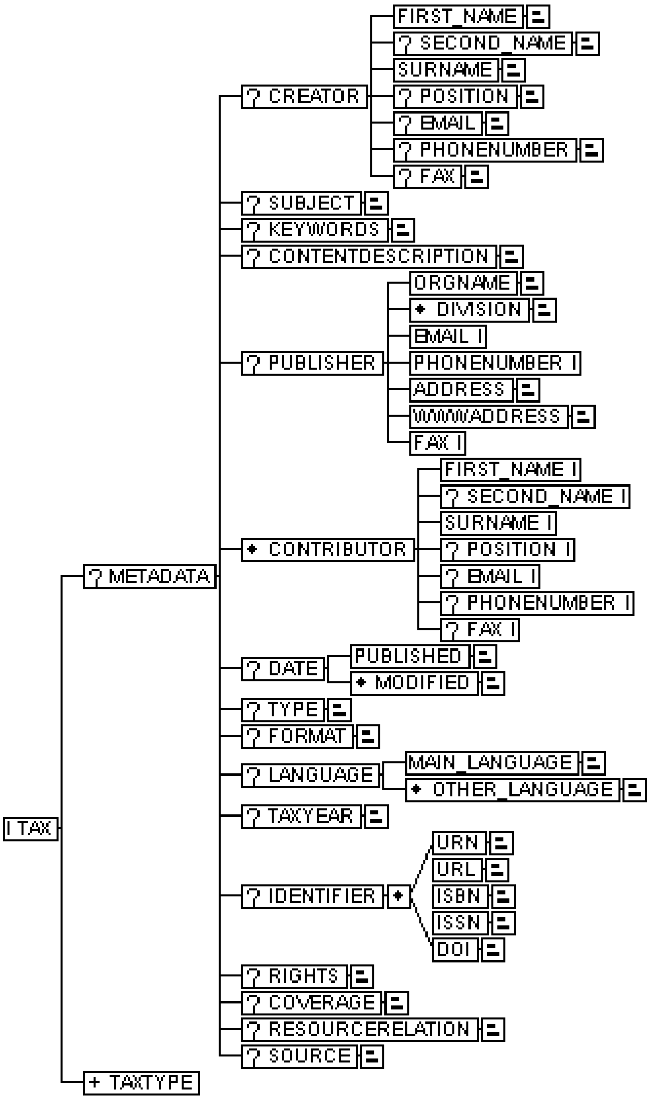
Fig. 7. The structure elements of the personal taxation



CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS

Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Fig. 8. The metadata elements



3.3. Formulating the metadata structure

The objective was to change the current xml-files made for personal taxation handbook to follow the previous structure definition. The xml-structure that was made for printing purposes is changed to correspond to the content based structure. In this way, metadata can be better electronically used. Changing can be partly automated, but the checking of the change and a part of the data has to be handled manually. However, these have to be done only once while implementing the system.

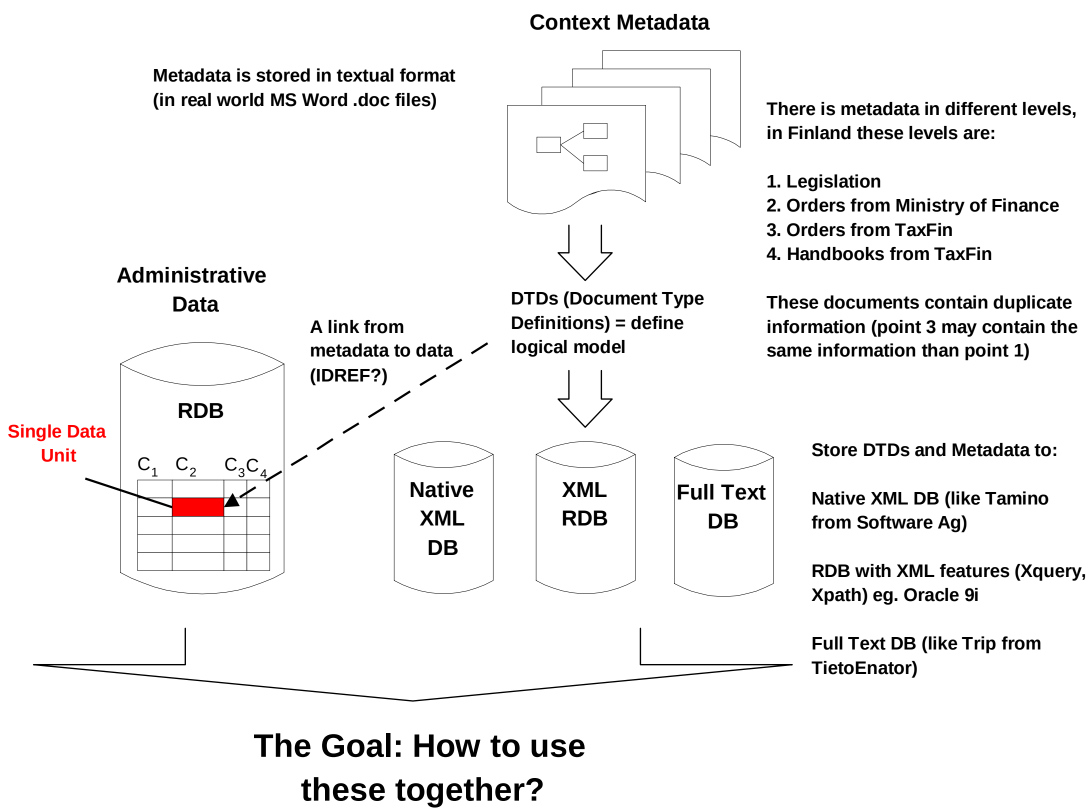
CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

As a result we have a structural XML-version of the personal taxation handbook. The structure is based on the structure of the contents of the book. After this the main question is: how to connect this metadata to the taxation data.

4. A Model to make links between the data in a relational database and metadata in XML-form

The problem of linking the context metadata to the numeric data could be generally described like in figure 9.

Fig. 9. Administrative databases and contextual metadata



In the chosen model, metadata in XML-form is saved in a native XML-database. The structured metadata can be saved as such in an XML-database and the structure remains as it is. With the help

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

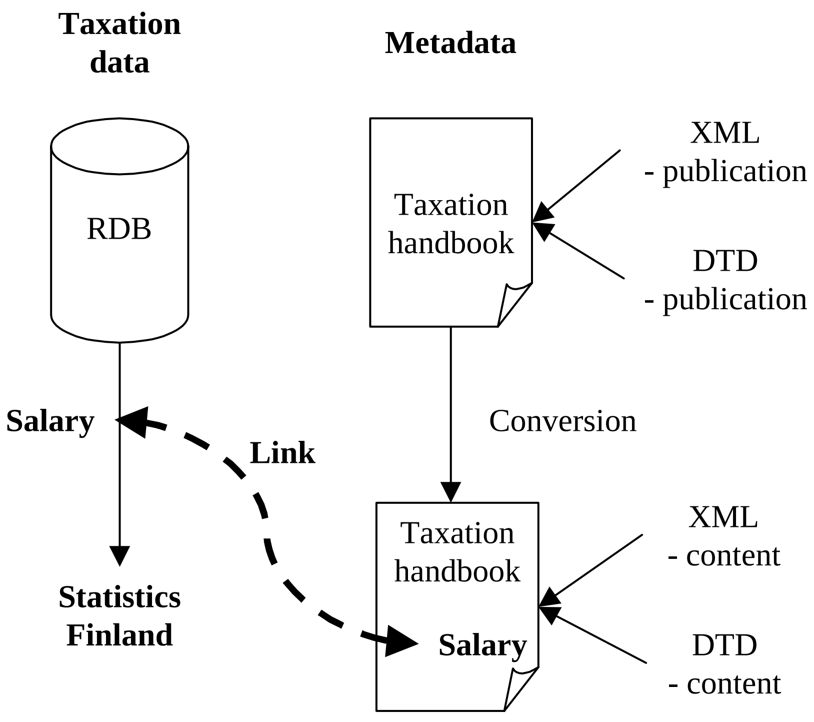
of some features of XML-database the data can be examined very precisely and the data can be gathered and browsed in a way that is needed.

One task of this project was to gather experience of the use and possibilities of XML-databases. XML-database is quite a new technology and there is only little information of its usability and advantages. By saving the metadata of personal taxation in this kind of database we get more experiences and models about the handling of XML-based data.

There is no knowledge (as far as we know) about the linkage between the data in relational database and the metadata in XML-database. This is undoubtedly going to be an essential question in the future. The taxdata and metadata are offering good material to test and work out this problem. The main point in this experiment is to find out how is it possible to connect two different types of databases and to make them work together.

The architecture of the model is described in the figure number 10.

Fig. 10. Joint use of a relational database and the data in an xml-database



CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

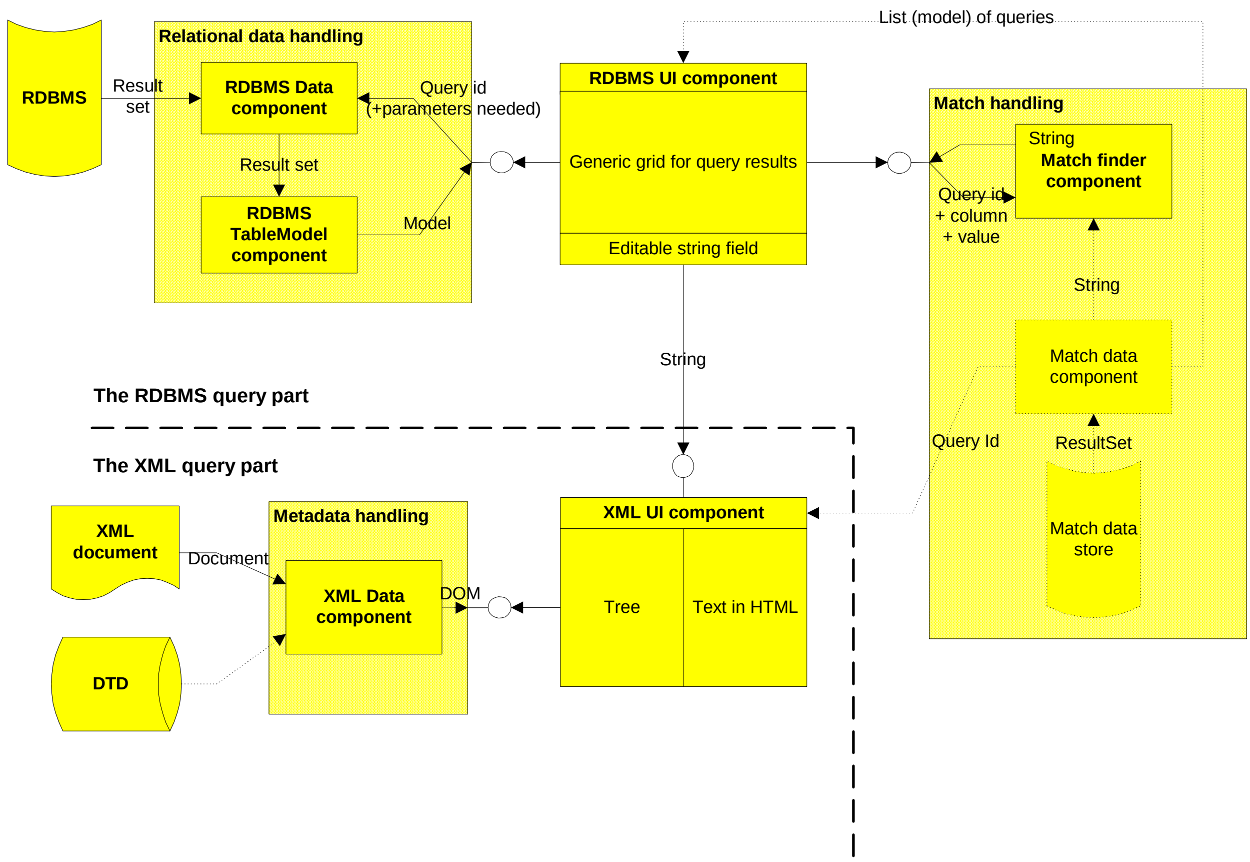
4.1. Technical implementation

The purpose of this demo building work has been to find a logical and general model for combining the operational data stored in a relational database and corresponding metadata stored in XML-documentation. The purpose has not been to optimize run-time performance of the system. The target has also been to implement the solution technically (to work in real life). The solution has been implemented in development environment (IDE) only. If a deployment package is needed, it must be produced separately.

There are models (to combine the operational data stored in a relational database and corresponding metadata stored in XML-documentation) where both or the other data is manipulated to correspond somehow each other. Instead, it is possible to operate with existing data sources or to develop them independently from each other in this model.

The structure of the demo system is described in the diagram below.

Fig. 11. The structure of the system



The implementation is *loosely coupled* and components are *layered* in the way that logical functions (like database handling and user interface) are separated. The connections between layers are implemented using different abstract models instead of direct data transfer. This solution makes also *late-binding* possible because data and its structure shown in each component are populated run-time and there is no need to know original data-sources or data presentation format.

The part of the diagram, which is drawn using dotted lines are not implemented in the current demo version. However, the idea of these components is discussed later.

4.1.1 The RDBMS query part of the system

- **RDBMS**
  The structure of relational data stored in these tables corresponds closely to real world examples of taxation production environment. Data content is not directly from production environment and all data values in customer identification fields are manipulated to meaningless ones.

- **RDBMS UI component**
  The user interface is implemented as generic and generalized as possible. All selection lists (possible queries and customer identification codes) are populated using list models in run-time and query results are shown in generic grid. The connection to the XML query part of the system is opened here. The parameter string, which is used to find corresponding metadata to queried relational data structure or data values can be get straight from data values in the grid, input manually or get from match finder component.

- **RDBMS Data component**
  All database operations are executed by this component. This is the only component, which has a database connection.

- **RDBMS TableModel component**
  This component makes possible to use only one generic grid in the user interface to show results from all possible queries. This is done with the inherited abstract TableModel. There are couple of other model creating components in the demo system to populate select lists in the user interface (these components are not included in the diagram above – to keep the diagram less complex). Although in this demo system, the used data source for this component is RDBMS, the data source for this component doesn't need to be a relational database. In other words, this kind of model (the object) between the data source and the data presentation layer makes the generic solution possible.

- **Match finder component**
  The exact match rules (or links) between relational data and XML-document are collected here. Because all these matching rules are handled in one component, these rules can be developed to more specific and more coverable, bit by bit, without technical influence to the

other parts of the system. This development can lead to a Rules Engine like solution.

- ***Match data component***
  This component is not implemented in the current demo version. The purpose of this component would be an interface to stored match data (see below).

- ***Match data store***
  There is no match data store in the current demo version. Use of a match data store would enable to add functionality to the system and make the system more generic:

  Matching rules stored in a match data store instead of coding them in the match finder component code makes it possible to develop rules without component programming changes.

  Queries stored in a match data store enables to generate queries select list model to be always consistent with available queries.

  Query parameters stored in a match data store enable to generate even more generic user interface.

  Query statements stored in a match data store enable to make the structure of the data component more straightforward and consistent with available queries.

  Query data stored with match data in a match data store enables connections to both directions (from the XML document to the relational data and vice versa).

  Also, this kind of stored queries system (where you can add or remove query statements depending on situation and where query results are shown in generic way without need to make programming changes) could be useful itself (comparing to the situation where each query result needs to program to use separate form or report document or comparing to the situation where you have to grant privileges on the relational database objects and queries are made using some database management tool).

4.1.2. The XML query part of the system

- **XML-document**
  The metadata of the subject is stored in this XML-document. In this demo system the XML-document is part of real world documentation (a fragment of official handbook written in XML).

- **DTD**
  The DTD-file describes agreed structure of the XML-document and so it is the shareable,

common metadata for all parties involved in the subject.

- **XML Data component**
  This component generates the DOM from the XML-document. The DTD-file could be used for validation, but this is not implemented in this demo version. In this demo system the DOM is used to generate a general model (the object containing data and methods) between XML-documentation and its presentation layer. This makes possible to release connection to original data source and same time to preserve the structure and the data of the original document (which is not possible if navigation through the document is done by using e.g. XPath from other layers of the system).

- **XML UI component**
  This user interface component has two panes. There is a navigable tree in the first pane (to show the document hierarchy) and there is a selected part of the document in HTML in the second pane. It is possible to navigate through the document either by moving in the tree or by giving a search string when the system finds corresponding node from the document (this method is used automatically when the XML query part of system is called from the RDBMS part of system). In this demo system, certain default selection level (the node in document hierarchy) is used to find a corresponding document fragment. This selection level value is parameterised, but it could be made selectable from DTD-file. This kind of component can be used to show any tree structure (not necessarily a XML-document).

4.1.3. The connection and dependency between the RDBMS query and the XML query parts of the system

The data in a relational database is stored typically in code value pairs or as plain values (and how it is intended to be). The data in XML-documents is close to written human language. Typically, the XML-document can be read and understood as it is. On the contrary, the data stored in the relational database cannot be reliable decoded without knowing structure and relations (i.e. the correct SQL query statements) related to the particular database.

In the beginning of this demo building work there were some ideas about connecting the relational data and XML-documentation by using e.g. relational database repository descriptions but it was realized that this kind of approach would lead to less generic and less generalized solutions.

As described earlier, the both parts of this system (the RDBMS query part and the XML query part) can be used independently and the connection between them can be arranged very loosely (using strings or kind of search words) and one-way from the relational data to the XML-documentation. To create more precise connection, some kind of matching rules are needed (like earlier mentioned Match finder component and Match data store).

4.1.4. Tools used

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

- **Eclipse IDE** for programming (**Java**)
- **Eclipse VE-plugin** for user interface (**Java swing** components)
- **Oracle 9i** to represent a RDBMS
- **JDBC Thin** protocol for RDBMS connection
- **eXists** to represent a native-XML database

5. Conclusions

5.1. About considered alternatives

During the design phase of building the demonstration application several technical solutions were under consideration. Here is brief reasoning about chooses.

There arose an idea of using database repository and design tool for metadata source. It was ditched because it would have meant binding the solution to one software vendor's product and its named version. It would have also meant several, possible conflicting sources for same metadata. The content of repository was also seen as a technical description of relational database in stead of information in it. A more open solution was still to be found.

The XML document is not described with schema because a DTD was available. It is possible use them when building new metadata sources to be used beside the existing one.

Different linkage tables / columns in relational database or separate linkage files were not used because they would have been bound the relational data source's structure to metadata structure with additional data structure. This would have meant one more component to be maintained always either data source's content is changed, not to mention situations when data structures evolve or new sources are added. Target was to achieve a solution without any changes or needs to existing data sources' contents or structures.

5.2. Analysis of result

The main conclusion is that the model above is working in practice. This is achieved by using loose coupling, standard interfaces and generic user interfaces. Technically this demonstration proves that solution, which has loosely coupling between components by using interfaces based on standard models, gives possibility to process and combine information from different sources. It is essential to cut loose information from its technical source and represent it with standard model based way.

It can also be concluded, that building user interface from generic objects reduces workload when building and maintaining the application. It enables to change the data source without updating the source code of application because the user interface is build during run time.

The expandability of solution can be achieved by building ready to use plug-in interfaces. In this demonstration application there are couple of functional parts (e.g. Match Finder Component) build in a way that they can be plugged out and be replaced by a new, more sophisticated plug-in component.

Finally it was concluded that building on pure, open, standardized models in every aspect means little more research work at the beginning but saves the burden and workload during implementation and especially during maintenance and further development.

5.3. Evaluation

Results of internal and external evaluation are gathered in Appendix 1. Evaluation.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Appendix 1. Evaluation

Evaluation criteria

For all CODACMOS demonstrations, the following general evaluation criteria have been identified:

- effect on data quality (including timeliness)
- effect on data collection costs (for respondent, data collector, statistical institutes)
- value of concept
- range of applicability (sectors, countries, tools)
- contribution to standards
- data security and confidentiality

The appendix presents the findings of the evaluation process.

A1. Internal evaluation

Tero Pernu/Taxfin

The following section presents the findings of the internal evaluation by TaxFin.

A1.1. Effect on data quality (including timeliness)

Demonstration has a drastic influence on data quality. Advantages concern both primary and secondary data collection, not mention further use of data. It is vital to ´speak´ same language and understand the meaning or content of data. Tax metadata is excellent way to unify the logical deduction of content of data.

Timeliness is not that obvious to respondent but data collector and secondary data users.

A1.2. Effect on data collection costs (for respondent, data collector, statistical institutes)

There is minor increase on costs when designing and implementing the metadata system. Naturally there is maintenance cost also. These new costs are significantly less compared to the present situation where numerous of manpower and time is used to interpret the data.

Respondents site the cost is not an issue because metadata is not visible for them.

A1.3. Value of concept

Concept is very strong.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

A1.4. Range of applicability (sectors, countries, tools)

Concept is applicable in all sectors and countries. Concept requires at least co-operation between two actors, but can (should) be one issue when practising e-government. Based on this demonstration, it can be said that model is applicable for sure in every EU countries at least between tax administrations and statistical institutes.

Although the final version may not follow exactly the demonstrated model, the idea and practical implementation is very good. TaxFin will continue to evaluate and participate to development of the model.

A1.5. Contribution to standards

There is no special contribution to the standards. Document metadata follows the Dublin Core standard. Taxmetadata is based on CoSSI model developed by Statistics Finland.

A1.6. Data security and confidentiality

There is no negative or positive influence towards data security or confidentiality.

A2. External evaluation

Following the presentation by three partners, 20 Workshop participants have used the pocket PCs to complete the evaluation questionnaire on this demonstration of primary data collection with handheld computers. Of these 20 respondents, 13 were members of the CODACMOS consortium, 7 were not; 13 worked for national statistical institutes, 7 for other organisations. The answers to the respective evaluation questions are summarised below.

A2.1. Effect on data quality (including timeliness)

Compared to more traditional solutions, do you think the effect of the demonstrated solution on ***data quality*** *would be:* mean score **4.05** (‘positive’)

Compared to more traditional solutions, do you think the effect of the demonstrated solution on ***timeliness of statistical data*** *would be:* **3.85** (‘positive’)

A2.2. Effect on data collection costs (for respondent, data collector, statistical institutes)

Compared to more traditional solutions, do you think the effect of the demonstrated solution on data collection costs (for the data collector) would be: **3.75** (‘positive’)

Compared to more traditional solutions, do you think the effect of the demonstrated solution on data provision costs (for the data provider) would be: **3.55** (‘positive’)

Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

A2.3. Value of concept

Do you think the 'value of concept', the idea behind the demonstrated solution is: **4.25** ('strong')

Of respondents 45 % *thought that the idea behind the demonstrated solution is* strong and 40 % that it is very strong.

A2.4. Range of applicability (sectors, countries, tools)

Do you think the 'range of applicability', the extent to which the demonstrated solution could be applied in other sectors, countries, etc. is: **4.1** ('broad')

Most of respondents (70%) regarded that the demonstrated solution could be applied in their organisation and it could have added value.

A2.5. Contribution to standards

Do you think the contribution to standards', the extent to which the demonstrated solution could contribute to the emergence of valuable widely shared standards is: **3.9** ('strong')

A2.6. Data security and confidentiality

Compared to more traditional solutions, do you think the effect of the demonstrated solution on data security and confidentiality would be: **3.5** ('neutral')

A3. Additional comments

Jozef Olenski (National Bank of Poland)

The following summary is from the Report for the session made by chairperson Jozef Olenski after workshop discussions:

> The demonstration was good example of effective integration of metadata on taxation and numeric taxation data as the method of adjustment of information collected in the tax system for statistical purposes. The project could be used as the pattern of metadata based interoperability of secondary data sources and statistics. It is also proving that integrated common metadata platform for data sources and statistical surveys is the prerequisite of integration and interoperability of statistics and source data.
>
> Methodological gap between data sources (primary and secondary) and statistics could be reduced or eliminated by introducing common metadata standards by the stakeholders of statistical data collection processes: respondents and statisticians. The results achieved by this demonstration should be widely propagated as "best practice" to be followed by other statistical offices and governments managing administrative records useful as administrative data sources.

CODACMOS
CLUSTER OF DATA COLLECTION INTEGRATION AND METADATA SYSTEMS FOR OFFICIAL STATISTICS
Type of document: Working document
File name: Taxation Metadata Demonstration Report.doc
Date of current version: 18 October 2004
Authors: StatFin: Heikki Rouhuvirta, Harri Lehtinen; TaxFin: Seija Karevaara; TietoKarhu: Aki Laavola, Sakari Harlas

Appendix 2. Glossary

DOM
The Document Object Model is a programming interface, which lets a programmer create and modify HTML pages and XML documents as program objects.

DTD
The Document Type Definition is a specific document defining and constraining definition.

RDBMS
The Relational Database Management System is an application that lets you create, update, and administer a relational database.

SQL
The Structured Query Language is a standard interactive and programming language for getting information from and updating a database.

UI
The User Interface is the part of an application, which invites human interaction and responds to it.

XPath
XPath is a language that describes a way to locate and process items in XML (Extensible Markup Language) documents.