A Metaregister as a Tool to Curb the Information Requests from the Government

Sjur Eigil Dahl & Vemund Riiser
Senior advisers
Department of Information Technology Planning and Co-ordination
Statskonsult - Directorate of Public Management
Norway

Abstract
The aim of this paper is to share some of our experiences with an analysis of a metaregister established by the Norwegian Government for the purpose of curbing the explosive growth of information demand from the government to private businesses.
To do this we will present sketches for what we hope will end up as a general framework for evaluating registers and data-collection processes. This framework consists of some criteria to judge the quality of the data collected and some prerequisites concerning the contents of the information requested, regarding both the part involving the collecting of data and the part concerning the supply of data and the institutional/infrastructure environment linking these two parts together.
The paper sums up by pointing out that to be able to present workable solutions it is necessary in the analysis to transcend the usual divisions between technological, organisational and data aspects of system developments.
We would like to make it clear that these are preliminary results and propositions, which we throw into the discussion here to hopefully get some reactions and criticism back, as well as some help for our further work. The points put forward here must be viewed as expressions by the authors and not as an expression of official views by the directorate or the Norwegian government.

Keywords: Government policy, Metadata, Data re-use, Data quality, Information infrastructure, Actor network

1. Introduction

The paper concentrates on findings and questions raised by a work to increase the Norwegian governments ability to reduce the amount of information requested from the business community and to increase the ability to re-use existing data.

The paper is organised as follows:
• First we will take a brief look at the background for the governmental initiative
• Then we will present the metaregister.
• Then we will present our framework and the background for it.
• Then we will present the analysis of the metaregister.

We will conclude this paper by providing some general impressions regarding the benefits, presumed and actual, as well as limitations of such a register.

2. Background and institutional framework

2.1 The country

Being one of the richest economies and also one of the technologically more advanced countries in the world, Norway has a business structure which might seem more like that of a developing country. We have a large SME sector consisting of generally less technologically advanced companies, and a few big advanced industrial conglomerates, mostly oriented toward the
processing of natural goods, like oil, timber, metal etc. Although we do have small technologically innovative companies located around the major cities, this is not the general picture. Especially when you move outside the major cities, this picture of many small companies with few but relatively large enterprises processing raw material is a dominant one and even more extreme. If you add to this a general large governmental involvement in the economy (the largest banks and the largest companies are either wholly or partly owned by the government), this will give you a more correct picture. In the words of the CIA factbook: “Norway is a prosperous bastion of welfare capitalism”.

This combination of an economic structure where the major part (96%) of the companies are either small (0-4 employees) or “medium sized” (5-10 employees) and high labour costs, do gives certain policy challenges. In addition there is a strong governmental commitment to a regional distribution of economic activity in a rather large country (300.000 km², equivalent to the size of Italy or Poland) with relatively few inhabitants (4 mill).

2.2 The problem

By March 99 the central government had a total of more than 640 different forms, a figure which had been growing in 1998 with a speed of almost one per week. Although not all companies get all the forms, we have done an analysis that shows that on average a company has to answer to 50-70 different forms some each year some more seldom. Calculations has been done show that just providing this information adds up to the continuous work of approximately 6000 persons per year, which is about 0.3% of the total number of people (1.9 mill) employed in Norway.

It’s no news that the business community’s obligation with excessive and demanding reporting to the government is a problem. Regrettably neither are the Government’s attempts to solve it. According to an earlier report the Social democratic Government in power at that time established a committee in 1978 to simplify forms from the government to the public [7]. In 1988 an official report addressed the need to co-ordinate the governmental use of data [6]. As a result of this and similar initiatives it was decided to introduce a unique ID number for all businesses operating in Norway. To administer the ID the Government saw the need to establish a register containing basic information regarding the companies. This register called the “Enhetsregisteret” or “The Central Co-ordinating Register for Legal Entities” and the accompanying laws was established in 1996. This register is a register of all legal entities (enterprises) in Norway, and they are all represented with a unique organisation number. It is linked with other registers collecting legal and economic information from Norwegian companies. This register was meant to be accessible to all governmental institutions, and with a right for anyone to request the public domain information.

Neither is this initiative an exclusive Norwegian attempt. In 1977 the US government established the Paperwork Commission with a similar but broader task. The committee engaged 200 full- and part-time workers, produced 36 official reports and made 770 suggestion [4]. The work resulted in the “Paperwork Reduction Act introduced first in 1980 and later renewed.

The SME bias by reducing this number is a result of the simple argument, that if it takes a certain amount of work to complete a form, this represents a relatively larger load for a small company than for a larger one and a small company rarely has a specialised function which is accustomed do deal with such issues. And given the above mentioned characteristics of the Norwegian economy it is obvious that reducing the amount of work attached to this reporting is a policy goal. There are of course other arguments than economic aspects. The business community’s generally suspicious attitude towards providing the government with internal information is also one important aspect, as well as the modernity symbolism behind the suggested solutions. Another aspect of the initiative is to increase the quality of the collected data. A recent study by the tax authorities showed that 75% of the declarations from businesses had major faults. In fact the faults were so great that the forms did not provide the requested information necessary for a correct
treatment, by the tax authorities’ own measures! The situation had probably been like this for years because they had never thoroughly investigated it.

There is a Government initiative underway to make Norway’s public sector business-friendly and modern. And the present central/liberal coalition now in power has made a 25% reduction of the reporting burden their commitment.

2.3 Oppgaveregistret (OR)

In addition to the introduction of the unique ID number and the establishment of the central co-ordination register, another major initiative has been to establish a central governmental body, responsible for collecting information about all the information the government demands from the business community. This meta-database called “Oppgaveregisteret” or “The Register of Reporting Obligations of Enterprises” (OR) has as its task to: Co-ordinate and maintain a complete record of all information the government requests from the business community. This register contains, among other, such information as what kind of information is requested, how many companies has to deliver the requested information, and to what kind of companies (classified by NACE) is obliged to provide information.

The OR shall booth record what information is collected and shall work as a co-ordinator between the different government bodies. In principle information about one company can be exchanged between registers by using the organisation number, and the OR should tell the different governmental bodies where to find information which already has been collected. In principle it should therefor not be necessary to ask for the same information twice from the companies.

3. Analytical framework

As mentioned in the introduction, we are working to establish a framework consisting of criteria to judge the quality of the registers, and some prerequisites to the parts regarding the collective and delivery of data and the institutional arrangement between them. We would like to mention some of these here, but would emphasise that this is not a complete list.

3.1 Criteria concerning quality

‘Quality can be measured as:

1. Formal criteria
   a) how large a part of the total population regularly file
   b) what part of the data in the form is critical for the administrative process

Ad a

Matters of compliance are an issue in every administrative process, and there is difference between a large proportion of willingly supplied data and data that has been filled in as enforcement. Making assessments to meet production terms in one government area can make things complicated in another. For example, income tax assessment for people from whom collections is impossible, makes the collection of pensions interesting if a just figure for income is used.

Ad b

Not all information in a form is used with the same intensity. In last year’s forms for direct taxes for Norwegian enterprises, more than 5000 possible data could be filed. But of these less than five make it possible for the local tax office to stipulate the enterprise’s tax. (We do not propose this is the rule.) Studies show that over time data that is not strictly necessary will not be filed – that is, if the request for data is not followed up actively.
2. Standard: how accurate is the data as a record of the substance?
   The data only reflects what the officials have registered, and this is more or less close to reality. The relative distance seems dependent on several factors:
   The supplier has a real interest (not an ideal, moral or normative one) to give correct information.
   The supplier has a lot to lose if the data supplied is incorrect, and there is a felt possibility of efficient control.
   There is a mechanism with a third person with his/her own interest in scrutinising the data, e.g., an employer giving the tax authorities data on salaries with a copy to his/her employees.
   The point is that standards only apply to data that is influenced by any incentives in place. The other data in the form is a completely different matter. One example is the Norwegian VAT return. Traders have to stipulate the reporting period’s total turnover, and the turnover on which they have collected VAT. This last figure is of great interest to the tax office. The first one is actually rather complicated to understand for many traders, because there is a difference in having a sale exempt of taxes like “zero-rating”, and a sale where tax do not apply. The tax office dully registers the data supplied, but there are quite a number of errors in the data. And this is then quite contrary to the data on VAT-collected sales, which is extraordinarily good.

3. Timeliness in the sense of the quantity of due filing and proportions filed late, after reminders have been issued or data not filed. (It doesn’t really matter if the data is due periodically or triggered by an action. The same need for a major proportion of due filing is required for successful processing.)
   If the timeliness is good, the data is held until the day it is possible for others to use it without knowing the “internal work schedule” of the register. This last point makes it very awkward to set up workable forms of co-operation.

4. Maintenance: the ability to keep the records updated.
   This is the effort made to systematically update data. Some data is treated as being static, which is not or not followed up efficiently. Over time data deteriorates. A typical case in Norway is “bransjekoder” (NACE classification). This data is rather well controlled and checked by several public entities when an enterprise is registered for the first time. Afterwards no one really has an active interest in maintaining the records. It is no real problem that a dentist and a hairdresser’s records remain unchased if they do not file something themselves. For many enterprises changes of business are not uncommon and enterprises often file several areas of interest without exploiting them all. This data is a quagmire and no critical eye is shown the consequence of lack of maintenance.

5. Usefulness in relation to the goals under which the data shall be used. The most important reminder of quality is the relative usefulness of data. What is actually recorded? That seams obvious or even tidies, but from experience it is very critical. What does it really describe and for what can it be used. One example of this is the Norwegian Population Register, which is a record of where people live in Norway described by address and postal code. But it actually is not. It describes where people living in Norway should vote and pay tax. This might be very important. Since we in Norway positively discriminate people living in some parts of Norway as to how much tax revenue is collected as well as the number of voters needed to get representatives to the parliament, the regulations relation to where you live according to tax municipality and voting region becomes essential. And for many reasons then people do not register where they live in a more practical sense, but where they should or can stay in a technical understanding for tax reasons and the right to vote. For example Oslo, the capital, has one tenth of its dwellers not registered in its municipality because they are students and for several reasons (some enforced by the government itself) pay tax and vote at their homestead. If
the register is used to calculate what kind of public transport is needed to serve the capital’s university, they would probably end up with incorrect figures and the wrong mode of transport.

### 3.2 Criteria concerning the information supplier side

1. The data supplied should be recorded or aggregated from recorded data in the enterprises in a structured and formalised manner, as with bookkeeping standards. This means that the routines on the informant side must support the data. This again assumes that there is a certain level of shared understanding about “how the world is” between the two parties. For instance: If the government asks for information about the number of people in different positions in a company, the company must in some way either use the same categories of positions. Or there must be possibilities to easily translate one party’s implicit or explicit categories into the categories that the other party can use. In many cases this is not the case. The Norwegian standard for classification of positions consists of more than 4000 different entities. For most companies, this level of detail is far beyond what they use. A finding from the analysis of the OR gives another example: There are over 150 different registers in OR based on data from enterprises. The information is categorised on the basis of three different references. There are accordingly 5, 40 and 13 different categories in each reference, and in principle this might give us 2 600 possible kinds of data elements. From our point of view this is much too much compared to the kind of information there is in companies (either electronically codified or not).

2. The data should be transparent in some way, so that another party has an interest in scrutinising its correctness, or other sources can make matching possible if necessary.

### 3.3 Criteria concerning an information demander

1) Make charts of responsibility concerning the roles of capture, registration, maintenance and tending of the data, and management of these charts.

2) The data must be used, and this usage relevant to the criteria for quality.

3) It must be the case that same milieu makes use of the data and is principally responsible for in data capture and maintenance, or the feedback between these milieu must be immediate. This means that the collector must take an active part in using the data, and that the evaluation of the collector’s own work in some visual way depend on the quality of the data.

### 3.4 Criteria for an institutional arrangement between the supplier and the demander

By this we mean there must be a way to communicate between the two parties in order to obtain information other than the actual data. The information we are thinking of here is connected to the points above. There must be some communication between the two parties to ensure that the data satisfies the criteria above. And in case they do not there must be communicated which criteria they fail to satisfy, why, and any acceptable solutions to this. As a part of this infrastructure, there must also be regulations or praxis established in both parties to ensure that this information about the data is being taken care of and the appropriate actions to correct errors are being carried out.

### 4. Analysis of “OR”

OR itself, along with the regulations, does provide both the formal power and factual know-how to make alterations to governmental conduct, and there is in policy as well as politics a real interest to make changes. There is an administration active to make propositions for co-ordination, and the people connected to OR have done fieldwork to promote the use of knowledge in collaboration with some institutions responsible for data collection. And OR has made a difference, it has managed to make governmental entities share some information and to make some forms
obsolete. But this effort has not cut through the bureaucratic undergrowth of demands for data and new wishes for more; it has merely just tampered with a small proportion of it.

The quality of OR (judged by the reasons for its creation) is as we have argued earlier connected to. 1) To what extent the register keepers or information gathers share the same view of the world as the party delivering the information. 2) How institutional arrangement between the different parties functions. In the case of the OR, we have three participants. 1) The different bodies producing forms, 2) OR which in this context is supposed to work for the best of 3) The business community.

Our impression so far has been that the links for communication and sharing the same view of the world have not yet been fully established. When it comes to infrastructure arrangements, we have so far not been able to discover any institutional links between the OR and the different form producing bodies which are maintained beyond the initial registration, with the purpose of keeping both parties informed about actions being taken or considerations about such. When it comes to the links between the business community and the other two parties (and vice versa) we must regretfully admit that this does not seem different from the usual relations between the business community and the government.

5. Conclusion and policy recommendations

The conclusion so far has been that the everyday task of data collection and maintaining data in a governmental context is far too complex to be adequately described by a metaregister. By this we do not mean that the metaregister is useless, our experience is rather the opposite: it is useful and necessary. It simply means that the establishment of such a register is far from sufficient to accomplish the goals for which it was established.

The shortcomings are according to our view two: First if we look at our criteria for judging the quality of registers, we find that the information in the OR is far from sufficient to judge the actual quality of the register and therefore to act as a basis for the sharing of information on a large scale. This is not due to bad design or programming with regards to OR but simply a matter of the kind of the information it is possible to codify in such registers. Such information can broadly be categorised as contextual. Correctly categorised and codified information is an important perquisite if one is to establish a large-scale information infrastructure where different users share the same information. This task becomes even more difficult when different actors request the “same” information for different use (in different contexts). It is even difficult to talk about sharing the “same” information. Since it is not a programming or technical problem the solution is not about refining the technology or clever systems design in the strict sense of the word but has to be found elsewhere. A starting point will have to be a thoroughly investigation into the classification of the knowledge represented in the register [1].

The second shortcoming in our view is that establishing a register is far from the same as establishing the connections or a sufficient infrastructural arrangement between businesses and the government. In our view it is important to analyse all actors in this process with that in sight that they act according to the rationality of their ever-changing context. In this we mean that instead of viewing the different player as a part of an infrastructure [3] where every player has a inflicted task, agreed to by all participant, we must view them as a network of actors [2] [5] dedicated both to the common and their “private” tasks. And that this network is characterised not by a static outlay but by a continuos process where institutions, persons, technology, and information is negotiating and re-negotiating their positions.
6. References


*In Norwegian*