18TH MEETING OF THE VOORBURG GROUP

Tokyo, Japan
6 - 10th October 2003

SESSION: INFORMATION SOCIETY

Statistics based on data from Internet Service Providers - experiences from Statistics Norway

Mona TAKLE
Statistical Adviser
Division for Transport and Tourism Statistics
Statistics Norway
N-2225 Kongsvinger, Norway

Mona.Takle@ssb.no

Abstract

Statistics Norway is conducting a full count survey among the providers of Internet to the Norwegian market. The survey has so far given valuable information on the Internet infrastructure in Norway throughout 2002. Data has also been collected for the two first quarters of 2003. The survey is especially useful as a source to detailed regional information on Internet subscriptions.

The paper presents the survey and deals with problems and challenges that Statistics Norway has encountered in this work. Possibilities to improve the survey are also discussed. The paper summarizes some of the results from the survey.

Introduction

Statistics Norway started the Internet survey in the 1st quarter of 2002, and has at present gained experience from six quarters of data collection. In this paper we will share our experiences from the survey.

Surveys that intend to describe the information society normally collect information from the end-users, i.e. individuals or enterprises using Internet. The Internet survey, however, is based on information from the Internet Service Providers (ISPs). The idea behind the survey was that the ISPs could have better knowledge of the technical aspects of Internet subscriptions than the end-users. We also believed that a survey comprising all the ISPs could improve the possibility to see regional patterns in the development of Internet. A sample survey with the end-users as the entity, will probably imply a higher insecurity both concerning regional divertion and technical matters.

The ISP Population

One of the first and most important steps in the survey is to delimit and update the population. The population in the Internet survey consist of all enterprises that on a commercial basis deliver access to the Internet to end-users. Libraries, Internet cafés etc. are not included in the population. Sometimes municipalities are ISPs on non-commercial basis. These are not included in the statistics.

There are no all-including register for ISPs in Norway. In addition, as telecommunication (NACE 64.2) is not the main activity for all the ISPs, it is not easy to extract the enterprises from the Register of Establishments and Enterprises.

The Norwegian Internet Exchange (NIX) is an institution that organises the exchange of Internet traffic between ISPs. Not all ISPs are members of NIX, but most of the large ones are. Some of the smaller ISPs rent Internet access from one of the NIX members. Prior to the survey all the NIX members was contacted, and the information retrieved enabled us to create a population.

The industry that delivers Internet access has an unstable structure. Permanent monitoring of the branch is necessary in order to adjust the population to changes such as merges and closedowns. The population is also updated quarterly by checking the NIX member list and by information directly from the ISPs (separate question in the questionnaire).

The population consists of approximately 110 ISPs.

The data collection

After updating the population register, the ISPs are contacted quarterly by mail. They are offered the possibility to choose between delivering data in three different formats: Excel, ASCII or XML. The specifications for each format are given at Internet (http://www.ssb.inet/data).

Participation in the survey is obligatory, i.e. the Statistics Act is applied. After the first call for data, the ISPs are given a reminder. The response rate at this point is normally between 57 and 66%. Finally, a warning of fine is sent to those who have not yet delivered data. This is between 14 and 18 % of the population. 97-99% of the ISPs deliver the data within the final dead-line.

In the fist 6 waves of the survey, the collected data has been;
- number of Internet subscriptions at different transmission capacities
- the income of the ISPs
- the data volume downloaded
- the time spent online

The data should be broken down by postal code, and the subscriptions should be devided between business and private.

After the second round of data collection, we saw a need to distinguish between active and passive subscriptions. The main reason for this is the large amount of free subscriptions in the lowest transfer capacities. We can assume that especially many of these are not being used, and how long they are staying in the ISPs' customer register is casual. Thus, to give a correct picture of the reality, the focus should be on the active subscriptions, and perhaps the quarterly changes of the total number of subscriptions. The distinction between active and passive is only used for the subscriptions with transmission capasities under 384 kbit per second. All other subscriptions are considered active.

Challenges in the data collection

Reporting formats

The high degree of details can impose problems for some of the ISPs and it sets certain demands to the reporting format. As mentioned, the respondents can choose among three different formats for reporting. Almost 75 per cent of the ISPs report their data on an excel file. The excel file functions as a questionnaire and has been prepared by Statistics Norway. The ISPs must download it from our web pages (www.ssb.no/inet/data). Around 20 per cent of the ISPs deliver the data by ASCII files and 3 per cent by xml format. There are even some (3 per cent) of the ISPs that deliver their data on paper.

The optimal solution for Statistics Norway would of course be if all the respondents used the same format.While the flexibility can make the reporting less burdensome for the ISPs, it in return makes the handling of data in Statistics Norway more time-consuming. The files must be converted to the same format, before they are fed into the oracle database.

What makes this process even more resource demanding is that some of the respondents do not follow the instructions given. This is most often the case for excel format, where the respondents sometimes make their own versions of the file. Statistics Norway has put some efforts in making our "questionnaire" file more userfriendly, with for instance an example and use of colours. This has to some extent improved the situation.

The largest ISPs has made a program routine that extract the data from their customer registers. If they do not have a program for automatic extraction, the data need to be manually punched. This can be a timeconsuming process, especially if they have subscriptions in many postal areas.

It is difficult to see alternative reporting solutions that could function better on this material. When the ISP has subscription on only a few postal codes, a web solution could be better than excel. Reporting by web could at least facilitate the further handling of the data that is done in Statistics Norway. However, for the ISPs with customers spread on many postal codes, there would still be a lot of manual punching.

Scope of data

At present Statistics Norway has only published data for one of the variables, i.e. number of subscriptions. The other variables have turned out to be more complex and less relevant than assumed at the launch of the survey.

The thought behind the collection of 'income' was that it could help us monitoring the development of profitability, prices etc. in the market. However, this has turned out to be difficult. 'Income' (gross income from the subscriptions) is a quite vague notion and feed-back from the ISPs shows that it has been difficult to operationalize. Some ISPs sell packages containing both Internet subscription,

development of web pages for their customers and other ICT services. In these cases it becomes especially difficult for the ISPs to extract the income on the subscriptions.

Likewise, 'time online' has also been a challenge for the ISPs. Not all the ISPs have the possibility to measure time online of their customers. The variable is only relevant for the lower transmission capacities, as some subscription types are permanently online. By the end of 2002 analogue modem and ISDN, which must be dialled up, still comprises between 80 and 90 per cent of the subscriptions. This means that time online can reveal details about the use of the majority of Internet subscriptions in Norway. As broadband and permanent connection gets more common, we expect time online to decrease, and the variable will by time also become less and less relevant. 20 ISPs are reporting data on this variable, and the quality is not yet investigated thoroughly.

The variable 'data volume downloaded' is reported by around 30 ISPs. Normally the payment of Internet subscriptions do not depend on the volumes of data downloaded, so for the ISPs the measurement of this variable is not necessary. Still, we hope that we will be able to make some estimates on the amount of data downloaded. Such estimates can give us valuable information on the use of Internet and the connection between transfer capasity and the intensity of the Internet use.

The difficulties encountered with these variables can make the reporting quite stressfull and time consuming for the ISPs. In order to simplify the report burden, we concider to exclude some of the variables, e.g. 'income' and 'time on-line'.

Processing of data

Briefly, one could say that the processing of the data in Statistcs Norway include control and convertion of the files, transfer of data to an oracle database and production of tables.

Incoming data are put through logical checks. The excel-files are converted to ASCII-files by an excel-macro. In this process a lot of inconsistancies with the standard format are discovered. This may indicate that the system for reporting data is too complex, and perhaps facilitating the questionnaire would give better quality and less need for corrections in Statistics Norway.

Delivered data is checked to uncover lacking internal consistency. The totals reported from the respondents must correspond with the subscriptions on various transmission capacities. Furthermore, the data is also compared with data from the previous quarter. Great deviatitions are examined thouroughly, and if necessary the respondents are contacted.

Schedule

The first year the ISPs had a total of around 10 weeks to deliver the data. Our aim is to narrow down the deliverance time to aruound 7 weeks. However, in periods of public holidays and summer holidays, the deadline must be extended.

The time needed after the data has reached Statistics Norway has varied, but the aim is to use around 5 weeks. This will of course depend on the quality of the data delivered from the ISPs. Extensive corrections and need for new data from the ISPs may extend this this process.

Published statistics

The results from the survey are published at http://www.ssb.no/inet_en. The results has shown that the number of broadband subscriptions has increased evenly during 2002. From one quarter to another the increase has been approximately 50 per cent. In the survey, broadband is defined as Internet

subscriptions with a transmission capacity (rate of speed in) larger than 384 kbit per second. This delimination is set by Statistics Norway, as there is no generally accepted international or national standard definition of broadband.

Table 1 shows the development of the private Internet subscriptions during 2002. For the first two quarters there was no distinction between active and passive subscriptions. Table 2 shows the corresponding figures for the business market.
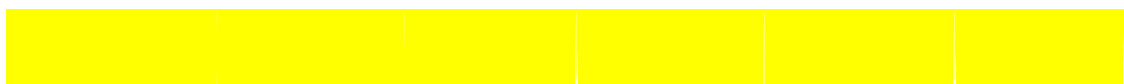
1	**Private Internet subscriptions. Development features. Subscription by transmission capacity and county. 2002**				
	Total	**< 64 kbit/s**	**> = 64 kbit/s = < 128 kbit/s**	**> 128 kbit/s = < 384 kbit/s**	**Broadband**
1st quarter.	**2623839**	**1006251**	**1505775**	**67533**	**44280**
2nd quarter.	**2708077**	**1142549**	**1422673**	**76510**	**66345**
3rd quarter.	**2909357**	**1234590**	**1511489**	**66050**	**97228**
4th quarter.	**3125881**	**1308584**	**1590219**	**68492**	**158586**
4th quarter. Active subscribtions only	**1450028**	**621606**	**607789**	**62047**	**158586**
4th quarter. County. Active subscribtions only					
01 Østfold	75632	32552	33583	3184	6313
02 Akershus	155848	60629	63965	7560	23694
03 Oslo	209464	75288	73059	23021	38096
04 Hedmark	51385	24018	23319	1304	2744
05 Oppland	49668	24206	22609	823	2030
06 Buskerud	69924	30586	31520	2489	5329
07 Vestfold	64241	26038	28686	2219	7298
08 Telemark	44816	19783	20084	1360	3589
09 Aust-Agder	27806	13582	12151	524	1549
10 Vest-Agder	46949	22188	19517	1497	3747
11 Rogaland	118433	48574	50691	6812	12356
12 Hordaland	132971	55487	55244	3254	18986
14 Sogn og Fjordane	31644	16050	14026	500	1068
15 Møre og Romsdal	71538	32799	31953	1889	4897
16 Sør-Trøndelag	83294	33975	32404	2236	14679
17 Nord-Trøndelag	34637	16925	16524	487	701
18 Nordland	71579	34892	29863	1170	5654
19 Troms	46731	22326	18802	919	4684
20 Finnmark	21306	11046	8640	743	877
Unspecified	42162	20662	21149	56	295

2	Business Internet subscriptions. Development features. Subscription by transmission capacity and county. 2002				
	Total	< 64 kbit/s	> = 64 kbit/s = < 128 kbit/s	> 128 kbit/s = < 384 kbit/s	Broadband
1st quarter.	139295	87008	42657	2473	7157
2nd quarter.	158042	109107	36155	2351	10429
3rd quarter.	234569	140009	75299	3507	15754
4th quarter.	246871	149221	74592	4498	18560
4th quarter. Active subscribtions only	176205	126730	26818	4097	18560
4th quarter. County. Active subscribtions only					
01 Østfold	7685	5320	1455	252	658
02 Akershus	17579	12456	2361	473	2289
03 Oslo	56751	41256	6982	994	7519
04 Hedmark	4891	3448	1003	138	302
05 Oppland	4711	3329	1062	54	266
06 Buskerud	7287	5250	1290	169	578
07 Vestfold	6580	4672	920	247	741
08 Telemark	4328	3192	763	114	259
09 Aust-Agder	2143	1720	302	42	79
10 Vest-Agder	4767	3539	844	78	306
11 Rogaland	10483	7517	1713	301	952
12 Hordaland	12918	8718	1994	269	1937
14 Sogn og Fjordane	2576	1888	501	41	146
15 Møre og Romsdal	6199	4328	1138	223	510
16 Sør-Trøndelag	8300	6069	1080	155	996
17 Nord-Trøndelag	5402	4778	480	46	98
18 Nordland	6286	4169	1509	277	331
19 Troms	4037	2937	715	121	264
20 Finnmark	1799	1311	292	69	127
Unspecified	1483	833	414	34	202

The results are also published on municipality level. Due to uncertainty in the data and the fact that the survey is still quite new, we choose to suppress figures where number of subscriptions are 10 or less. We have also chosen to show broadband subscriptions in percentage of households. It is reasonable to believe that there is a one-to-one relationship between private broadband subscription and households.

In table 3 the municipalities are classified by centrality code (Standard Classification of Municipalities, Statistics Norway 1994). Clearly, the most central municipalities have the highest number of broadband subscriptions relative to the number of households.

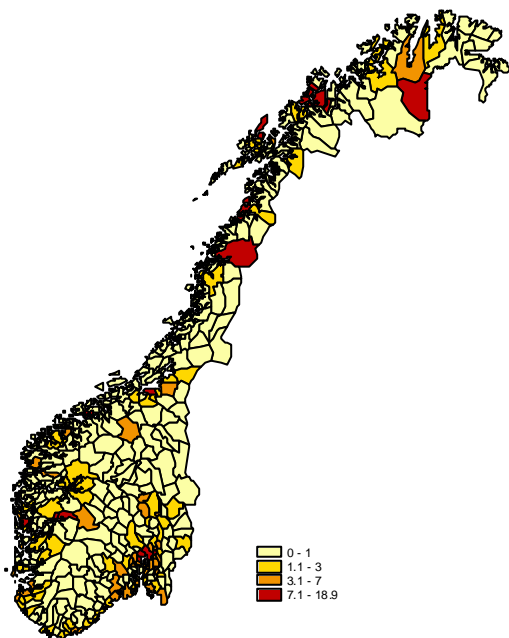| 3	Private broadband 4th quarter 2002. Subscriptions by centrality of the municipalities		
Sentralitet [1]	Population 1st January 2003	Broadband subscriptions	Broadband subscriptions per cent of population
Remote municipalities	625064	3651	0,6
Fairly remote municipalities	336218	4863	1,4
Fairly central municipalities	1143202	33775	3,0
Central municipalities	2447768	134360	5,5

Standard Classification of Municipalities, 1994 (NOS C 192

The distribution of broadband can also be illustrated by maps. By using the same legend each quarter, the increase of broadband can be monitored visually. Figure 1 displays the distribution by household 3rd and 4th quarter of 2002.

Figure 1. Broadband subscriptions in the private market by households.

The private market. Subscriptions with broadband (transimission capacity larger than 384 kbit/s) in per cent of households. 3rd quarter of 2002. Municipalities

0 - 1
1.1 - 3
3.1 - 7
7.1 - 18.9

Map data: Norwegian Mapping Authority

The private market. Subscriptions with broadband (transimission capacity larger than 384 kbit/s) in per cent of households. 4th quarter of 2002. Municipalities

0,0 - 1,0 %
1,1 - 3,0 %
3,1 - 7,0 %
7,1 - 25,3 %
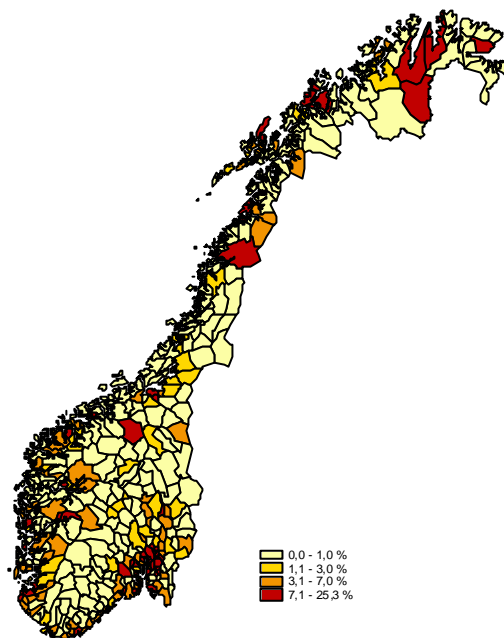
Map data: Norwegian Mapping Authority

Plans for the future

At short term, Statistics Norway will try to consolidate the survey. This include making it more efficient and user friendly for the respondents. We also intend to explore the variable 'data volume downloaded'.

On larger terms, our aim is to follow the development in the marked and adjust the survey when changes are needed. Higher transfer capasities are expected and the capacity categories must be reconsidered from time to time. However, changes in the data structure must be considered thoroughly as this implies changes in the program routines for some of the respondents.