

Article

Symposium 2008:
Data Collection: Challenges, Achievements and New Directions

Going Web-only in a Complex Enterprise Survey – Experiences and Lessons Learned

by Johan Erikson

2009



Going Web-only in a Complex Enterprise Survey – Experiences and Lessons Learned

Johan Erikson¹

Abstract

Over the last few years, there have been large progress in the web data collection area. Today, many statistical offices offer a web alternative in many different types of surveys. It is widely believed that web data collection may raise data quality while lowering data collection costs. Experience has shown that, offered web as a second alternative to paper questionnaires; enterprises have been slow to embrace the web alternative. On the other hand, experiments have also shown that by promoting web over paper, it is possible to raise the web take up rates. However, there are still few studies on what happens when the contact strategy is changed radically and the web option is the only option given in a complex enterprise survey. In 2008, Statistics Sweden took the step of using more or less a web-only strategy in the survey of industrial production (PRODCOM). The web questionnaire was developed in the generalised tool for web surveys used by Statistics Sweden. The paper presents the web solution and some experiences from the 2008 PRODCOM survey, including process data on response rates and error ratios as well as the results of a cognitive follow-up of the survey. Some important lessons learned are also presented.

Key Words: Web data collection, Collection strategies, Take-up rates, Response rates.

1. Background

Many statistical offices have made significant efforts in introducing web data collection over the last decade or so. There are many reasons why collecting data over the web is attractive. With much political pressure to reduce response burden on first and foremost enterprises, offering electronic alternatives has been seen as an easy way to achieve reductions, both in actual and perceived response burden. Secondly, there is hope that with built-in functionality such as edit checks and help functions, data quality can be higher on the web. And last but not least, the costs for statistical offices of handling data collection may be lowered, since no data entry is needed, costs for postage may be lowered et cetera.

In reality, respondents have been slow to embrace the web alternatives. Therefore, in the last few years many efforts have been made to raise the ratio of respondents choosing the web alternative. In some surveys, the bold step of more or less offering only the web option has been used. With low take-up rates, the supposed gains may turn out to be small or non-existent. On the other hand, concerns have been raised about the effects on overall response rates from promoting the web option and “hiding” the paper alternative. Therefore, it is interesting to study the effects of taking the step to go more or less “web only”.

2. Previous experience

2.1 Collection strategies at Statistics Sweden

At Statistics Sweden, web data collection has been used for more than five years. Before that, electronic questionnaires were offered in some surveys as Excel spreadsheets or such. In the beginning of the 2000s, a few surveys built custom-made tools for collecting data over the internet. In 2004, work was started to create a generalised tool for web data collection, the first version of which was taken in use in April 2005. Since then, this tool has expanded a lot in both functionality and use. Even if the tool was initially built for enterprise surveys, its

¹Johan Erikson, Statistics Sweden, PCA/LEDN, SE-701 89, Sweden, johan.erikson@scb.se

use has been expanded to cover all types of surveys to both enterprises, the public sector and households and individuals. To this date, over 130 different surveys have used the generalised tool.

Previously, most surveys have used what can be described as a “paper offensive” strategy when contacting respondents, who are sent a cover letter and a paper questionnaire. They are given the option to respond by web if they want to, using User IDs and passwords normally printed on the paper questionnaire (since cover letters are generic and not individualised, it is easier to print respondent specific information on the questionnaire). In reminders, new paper questionnaires are sent. Using this strategy, normally between 5 and 20 per cent of respondents choose the web alternative, the take-up rates are higher for enterprises than for individuals.

The last two years, strategies have begun changing, going to more “web offensive” strategies. There are different variations on web offensive strategies, but the most common ones are either not to send paper questionnaires in reminders or, taking it one step further, sending only a cover letter with User ID and password in the initial mail-out. In surveys of the public sector, the web offensive strategy is now actually the most common one used, while in surveys of individuals and households, the traditional paper offensive strategy is still the norm. In enterprise surveys, strategies differ from survey to survey with a tendency to move to more web offensive strategies.

2.2 Steering respondents towards the web alternative

At Statistics Sweden, there have been relatively few studies on the effects of going to more web offensive strategies. Holmberg, Lorenc and Werner (2007) have studied effects in a health survey to individuals. Regarding surveys of enterprises, an experiment was carried out in 2007 (data for 2006) in a longitudinal survey of wages and salaries (“LINDA”). In the experiment, a web offensive strategy was used for one group of enterprises while a traditional strategy was used for the remaining enterprises. The experiment and preliminary results were presented in Erikson (2007) which also gave results from a few other surveys. The final results from the experiment can be found (in Swedish) in Erikson (2008). Figures 2.2-1 and 2.2-2 present the main results of the experiment.

Figure 2.2-1
Response rates over time in the LINDA experiment 2006

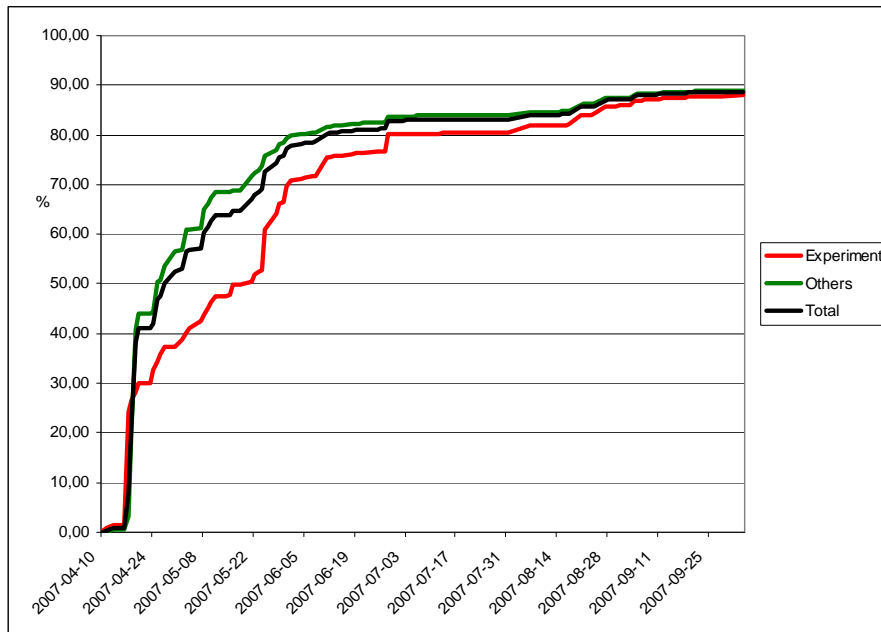


Figure 2.2-2
Web take-up rates in the LINDA experiment 2007

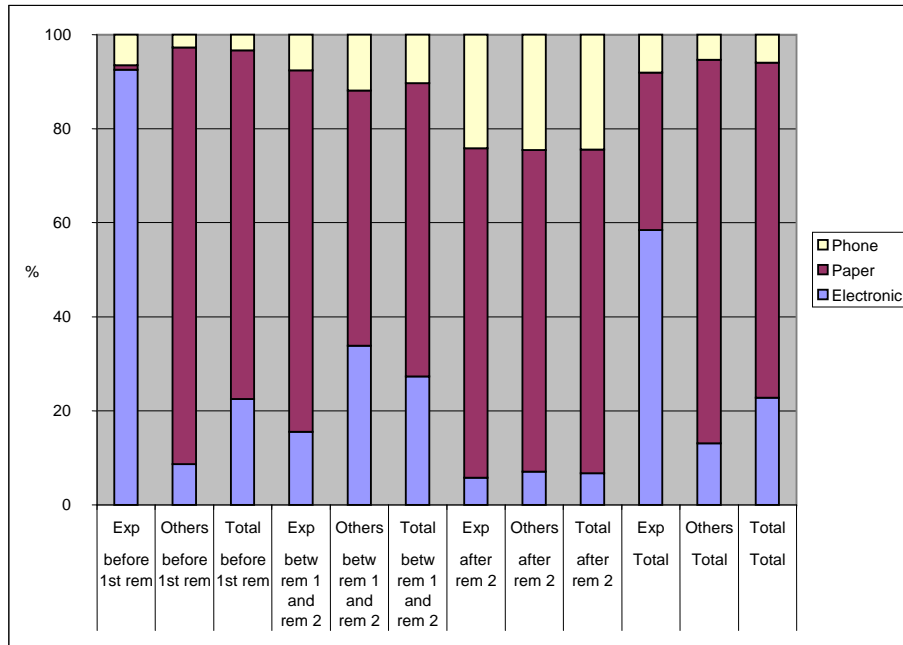


Figure 2.2-1 shows that in the end, the experiment had only a very small effect on the total response rate. What was interesting to observe was the “wait effect” that appeared in the experiment group. It seems that some enterprises waited for the reminder, which contained a paper questionnaire, before answering. This can also be seen in Figure 2.2.2, where the web take-up rate is much higher in the experiment group before the reminder. Figure 2.2.2 also clearly shows that the mode chosen by respondents is affected by the strategy, with web take-up rates much higher for the experiment group, and also in periods where the web alternative is promoted (before the first reminder for the experiment group, and after the first reminder for the others, who did not receive a new paper questionnaire with the reminder).

3. Going “Web-only”

Based on the LINDA experiment, it would be interesting to look at a survey that went more or less web only, using the web offensive strategy, to see if the results from the experiment would be repeated or not. But it was of equal interest to see whether the data collection costs were lowered and data quality higher than before. In 2008, the Survey on Industrial production took that step.

3.1 Survey characteristics

The survey on industrial production is a European survey carried out all across the European Union. Its contents are the production of goods and industrial services distributed by type of commodity. It is regulated by the PRODCOM (production communautaire) regulation, and for the sake of simplicity it will be referred to as the PRODCOM survey hereafter. In Sweden it is an annual survey while it is a short-terms survey in some other European countries. Some of the interesting characteristics of the survey are as follows:

- * It is a mandatory survey, which is the case for many regulated surveys. In theory, respondents can not simply refuse to give data, even if the mode is not their preferred one.
- * The production is to be reported by type of commodity according to a nomenclature of more than 10 000 commodities. This makes reporting data a very detailed and complex task.

* Data to be reported is both values and quantities of goods. For many enterprises and commodities, the concept of quantity is very difficult. Many enterprises don't keep track of quantities in their bookkeeping systems. And for many commodities, there is in reality no relevant quantity to be given. This means that collecting this data is time-consuming, and many re-contacts have to be taken with enterprises to ask questions or collect missing data.

* The survey is a census over a cut-off threshold of 20 employees (10 in some sectors). This means a very low rotation rate, and respondents know the survey.

In 2008 the collection strategy of the PRODCOM survey was as follows: In the original mail-out, only a cover letter with User ID and password was sent. After 5 weeks, a reminder including a paper questionnaire was sent.

3.2 Response rates and take-up rates

Figures 3.2-1 and 3.2-2 present the main results of the PRODCOM survey 2008. The total response rate was not affected at all negatively, it was even a little higher 2008 than 2007. And the web take-up rate was very high, 90 per cent of the enterprises that replied before the reminder chose the web alternative, and as many as 50 per cent even after they had received a paper questionnaire. In total, the web take up rate was around 70 per cent. And, perhaps most interesting, no wait effect was noted whatsoever, the response pattern over the collection period was almost identical to the previous year. Possible explanations for this may be the design of the survey, with low rotation rates and experienced respondents. This can mean that they know they have to answer, and respond at a convenient time, which is about the same time every year. The high web take-up rate shows that it is possible to use web data collection even in large and complex questionnaires, otherwise the take-up rate after the paper questionnaire was sent should be much lower.

Figure 3.2-1
Response rates over time in the PRODCOM survey 2008

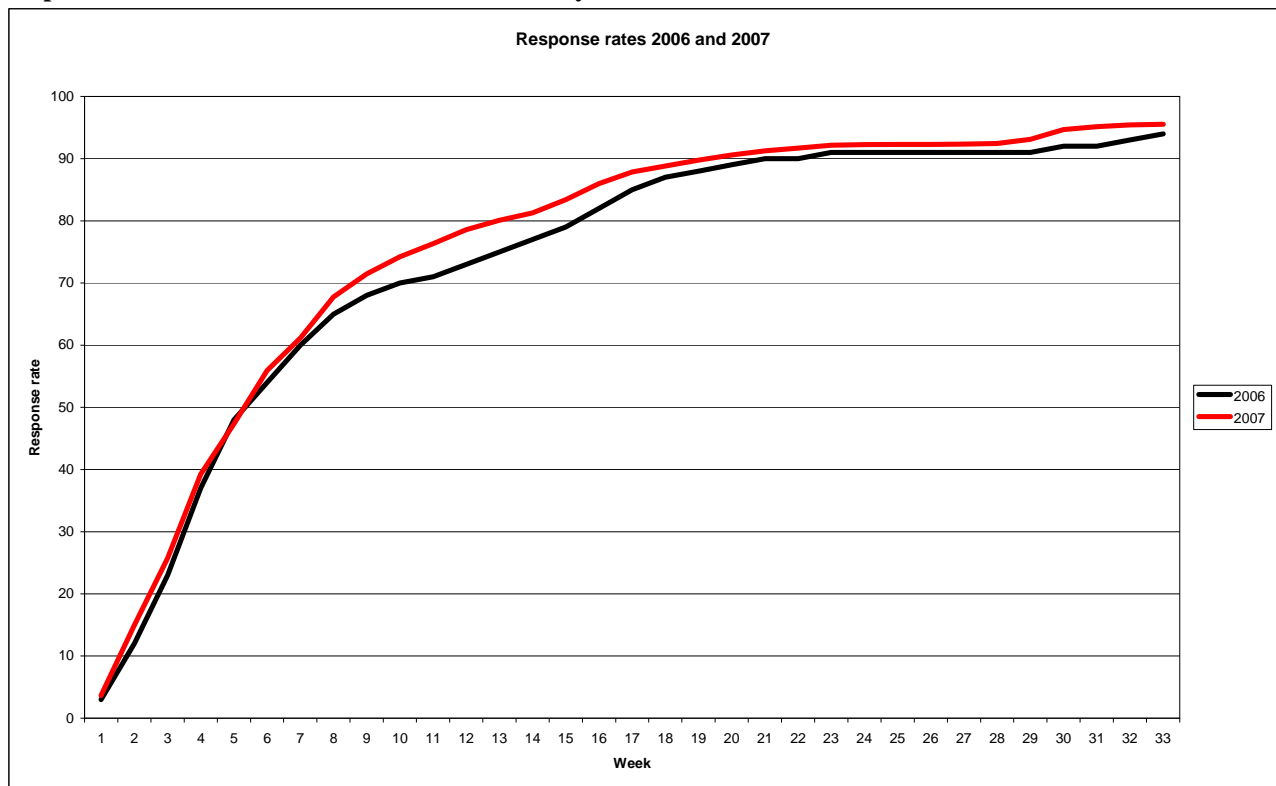
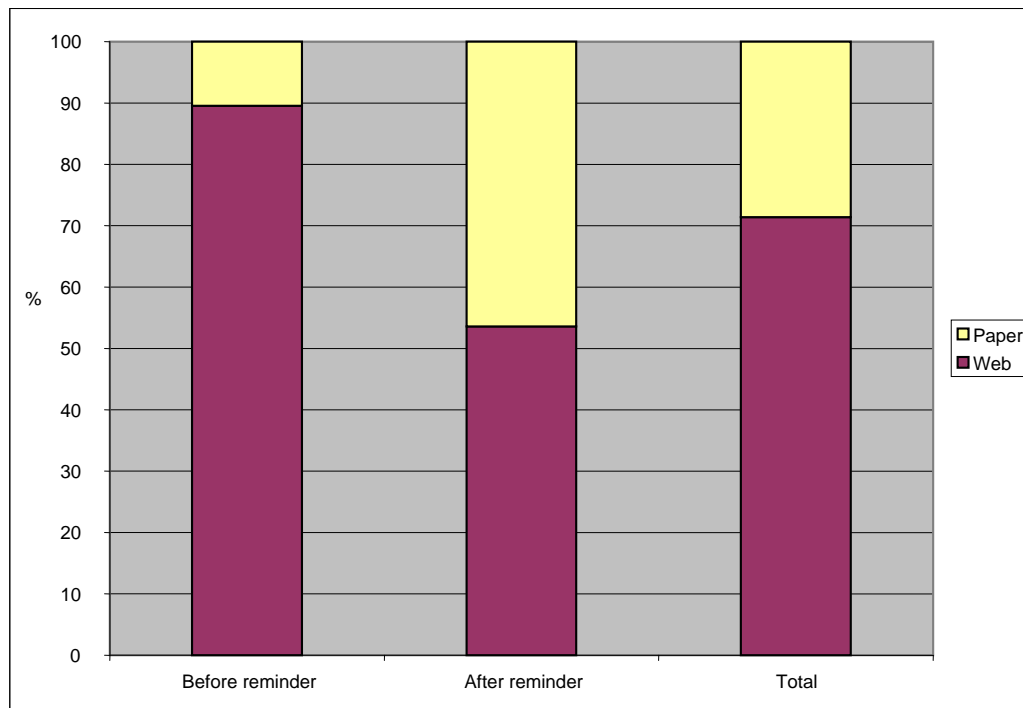


Figure 3.2-2
Web take-up rates in the PRODCOM survey 2008



3.3 Data collection costs and data quality

While the response rates and web take-up rates showed good results, the results regarding cost and quality were OK but not that positive. The costs of postage were reduced by a third due to not sending, but the personnel costs for the data collection and editing processes were more or less unchanged from the year before. Table 3.3.1 shows some process data for the last three years of production.

Table 3.3-1
Process data from the data collection of the PRODCOM survey 2006 to 2008

	2006	2007	2008
	number		
E-mails in	278	352	493
E-mails out	415	408	1022
Phone calls in	482	452	548
Phone calls out:			
No reply	1088	1379	1786
Agreement to call again	342	977	631
Contact made	1243	2126	2026
Sum of phone calls out	2673	4482	4443
Contacted units	1346	2085	2203
Units in survey	3814	4128	4818
Rate of contacted units (%)	35	51	46

The process data shows that incoming calls and e-mails were only a little higher in 2008 than in 2007, which shows that respondents adapted rather easy to the web questionnaire. If there had been high demand to have paper

questionnaires, or large numbers of respondents experiencing technical difficulties, the number of incoming contacts should have been higher. This conclusion is also supported by the response rate and take-up rate data showing no effects of waiting for paper questionnaires. This is a positive sign. On the other hand, if incoming data had been of higher quality than before, the number of outgoing contacts should be lower than before. And even though there is a small tendency of this (46 per cent of the total number of units were contacted compared to 51 per cent the year before), it was not enough to lower the total cost of the collection phase. It was still as time-consuming as before introducing the web questionnaire.

There are several reasons why costs were not lowered and quality improved compared to previous years:

- * The main reason is due to the nature of the survey and its complexities regarding quantities. As was said before, many respondents have problems with quantities, meaning it takes a lot of re-contacts with enterprises to actually get this data at all. The last two years have also seen increased efforts to get quantities from enterprises that have not reported this data previously. This has meant an increase in the number of re-contacts, which is clearly seen from comparing data for 2006 and 2007.

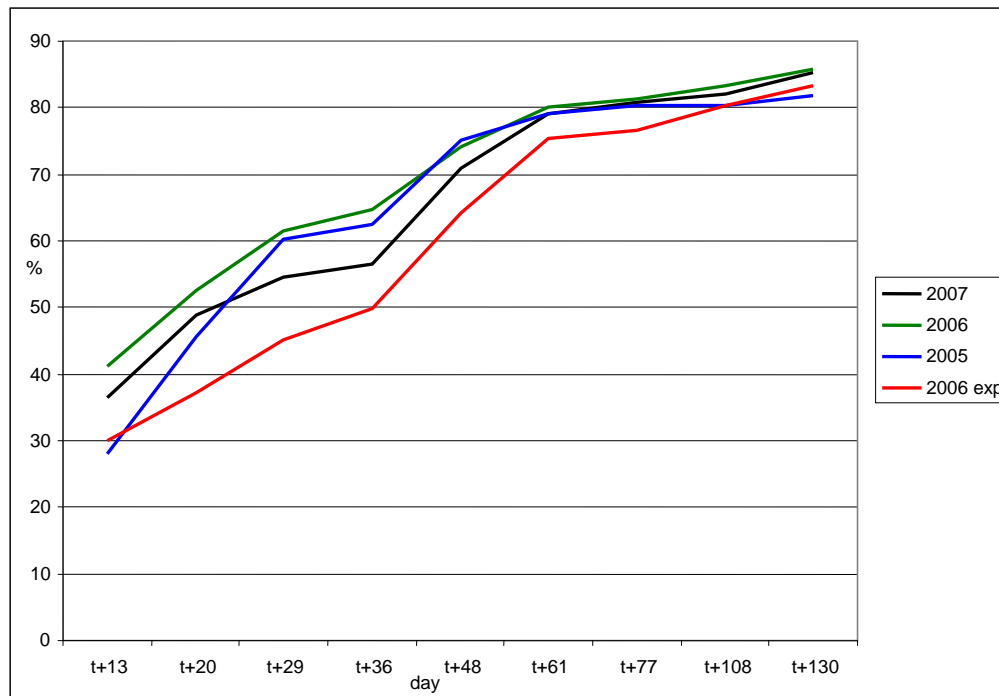
- * When detailed data is asked for using a complicated nomenclature, the data collection process is affected by changes in the nomenclature. Changes are more or less frequent from year to year. In 2008, changes were very frequent. This meant additional efforts in checking data for many respondents, additional re-contacts and more work. It is very possible that nomenclature changes were an important factor why collection costs were not lowered. It will be interesting to see next year, when fewer changes are expected, if more significant effects are seen.

- * The web questionnaire was not good enough. When introducing web questionnaires, it is common that the web questionnaire is made to look as similar as possible to the paper counterpart, even when going to a web offensive strategy. Furthermore, survey managers are often hesitant to add too much built-in functionality such as edit checks, not wanting to risk that respondents choose not to answer at all when faced with failed edit checks. In the PRODCOM questionnaire, a better layout of the questionnaire and more built-in functionality, for example reminding respondents that quantities should be reported, could mean better original data. Not using the possibilities of the web reduces the possibility of lowering costs and raising quality. A cognitive follow-up of the PRODCOM survey (Henningsson and Hartwig 2008) shows that respondents expect intelligence from a web form, and that the PRODCOM questionnaire has flaws, both as a paper questionnaire and a web questionnaire.

4. Further recent experience

Following the successful experiment, the LINDA survey decided to use the web-offensive strategy of not sending paper questionnaires in the original mail-out. Of specific interest in this survey after the PRODCOM experience was checking whether the wait effect observed in the experiment would still be present, or whether the same pattern as in the PRODCOM survey with no wait effect would be achieved. Results are shown in Figure 4-1. The wait effect observed in the experiment is still present, albeit smaller than the year before. One conclusion from this may be that different surveys will show different patterns, and that a pre-analysis of the survey characteristics is important. The LINDA survey is to a much larger extent than the PRODCOM survey directed at small enterprises, and has a somewhat higher rotation rate. This might mean that the respondents are not as prepared to embrace the web alternative as in the PRODCOM survey. The smaller wait effect than last year might in turn mean that respondents get used to surveys being web-only, and therefore turn to using the web. It would be interesting to make more direct comparisons between different types of surveys in the future. The presence of wait effects might also mean that going strictly web-only, never offering a paper alternative at all, might have negative effects on response rates and perceived response burden. This should also be studied more carefully.

Figure 4-1
Response rates over time in the LINDA survey 2005-2007



5. Conclusions

The study carried out for the PRODCOM survey was a case study of one complex enterprise survey. It showed some positive results, with high web take-up rates without any negative effects on the total response rates. This was in line with previous studies and expectations. On the other hand, other results were negative. The total costs of collection were as high as before, and the data quality was not increased. Previous experience stated that lower costs and higher quality could have been expected. Main reasons for the negative results in these areas are the complexity of the survey itself, nomenclature changes and too little intelligence in the web questionnaire. The last two factors should be studied next year, after improvements of the questionnaire and when there will be fewer changes in the nomenclature. Other recent experience shows that results differ somewhat between different types of surveys. A further study should focus on direct comparisons between surveys of different types.

References

- Erikson, J. (2007). Effects of offering web questionnaires as an option in enterprise surveys, *Proceedings of the Third International Conference on Establishment Surveys*, American Statistical Association. 1431-1435.
- Erikson, J. (2008). LINDA-experimentet, unpublished report, Örebro, Sweden: Statistics Sweden.
- Henningson, B. and Hartwig P. (2008). Industrins varuproduktion – Rapport från kognitivt test, unpublished report, Örebro, Sweden: Statistics Sweden.
- Holmberg, A., Lorenc, B. and Werner P. (2007). LOTTA P2 – Optimal kontaktstrategi vid blandad insamling papper och webb, unpublished report, Örebro, Sweden: Statistics Sweden.