APPENDIX B

Organisation of Data

The spreadsheet generated a Unique Number for each person by using the Questionnaire number, a full stop and then the person number.

The technique was as follows: -

Row 4 contains the Questionnaire Number i.e. the three digit number allocated by the in-putter. This is in fact the Household Number. Only the person 1 columns contain a Household Number.

Row 5 contains the person number for each person in a Household from 1 to 6. All columns contain a Person Number (all being in the range 1 to 6)

So in row 3, I put in the equation:-

=IF(I4=0,IF(H4=0,IF(G4=0,IF(F4=0,IF(E4=0,D4+I5/10,E4+I5/10),F4+I5/10),G4+I5/10),H4+I5/10),II4+I5/10)

This looks at the Household Number in I4. If it is zero, column I is not the person 1 column. It then checks H4 to see if this is a person 1 column. If it isn't it checks, in turn G4, F4, and E4. As there are only six spaces for all the members of the household, one of these five columns must be a person 1 column, or it must be D4. When it has found the person 1 column it adds that household number to the person number in I5 with a decimal point in front (i.e. I5/10). This would make the unique number for column I. If column D was the first household column and there were six people in the household, I6 would contain the number 6. The Unique Number for this column and hence this person's data would be 1.6. If, on the other hand H4 contained the household Number 3 and was the closest person 1 column, I5 would contain 2 indicating it was a person 2 column. The Unique Number would then be 3.2 This means that this column holds the data for person 2 in the 3rd household. This equation was then copied across the 404 data columns. Finally Row 3 was "cut" and "pasted special" into Row 2 transforming the formula into a number so it would remain constant however the data was sorted. This was the Unique Number and was generated for each person.

Thus 317.4 would be person numbered 4 on the 17th Questionnaire that the data in-putter number 3 added to the data base. The Questionnaire would have 317 written on the top right of the cover.

Data Entry

The data from the Questionnaire was added by: -

- Replacing a tick by the number 1.
- In Yes/No questions, "Yes" was allocated 1 and "No" became 3.
- In questions that rated the answers from A to E, "A" was allocated 5, "B" became 4, and so on to "E" which became 1.
- Questions where the answer was a number, such as the number of people in a household, then it kept that number. Thus if the answer was 5 then 5 was typed into the data.

Thus each persons opinions were in one column against each of the choices of each question. Each question choice occupied one row on the spreadsheet. So by summing along one row we could get the total number of people that agreed with that choice or statement.

Obvious this would not work with Yes/No questions. This was overcome by adding two rows in the database so that one could contain all the “Yes” answers, the other all the “No”s. This was achieved by adding an IF statement which queried the the original Yes/No row (which contained 1 or 3). Thus =IF(D62=1,1,0) in cell D63 asks column D, row 62 (the one that contains either a 1or a 3) if it contains a 1, if it does it returns a 1 in cell D63, if it doesn't it returns a 0. So if it contained a 3, it would put 0 (which is correct because 3 means No). Also if the question had been left blank (not answered), this would count as a 0 and would correctly return a 0. Similarly D64 the new “No” row would have =IF(D62 =3,1,0). So if D62 contained a 3 then D64 (the new “No” choice row) would put a 1. If it contained a 1 it would put 0 (which is correct because 1 means Yes). Also if the question had been left blank (not answered), this would count as a 0 and would correctly return a 0. These two IF statements were copied along the rows for each of the columns that contained the data.

We now have two new rows, one containing a 1 everywhere there was a “Yes”, the second containing a 1 everywhere there was a “No”. Again by adding up along these two rows we get the correct number of people responding “Yes” in one row and the total of those that responded “No” in the second.

Similarly with questions which rated the answers from A to E. Five new rows were added. The first would contain all the “A” answers, the second all the “B” answers and so on.

Thus if the data was in row 78, the following five rows would each have IF statements as follows:-
Row 79 contains =IF(D78=5,1,0) which returns a 1 if D78 was 5 (“A”) , if not it would return 0
Row 80 contains =IF(D78=4,1,0) which returns a 1 if D78 was 4 (“B”), if not then 0
Row 81 contains =IF(D78=3,1,0) which returns a 1 if D78 was 3 (“C”), if not then 0
Row 82 contains =IF(D78=2,1,0) which returns a 1 if D78 was 2 (“D”), if not then 0
Row 83 contains =IF(D78=1,1,0) which returns a 1 if D78 was 1 (“E”), if not then 0

Thus adding Row 79 would give the total number of people that rated the question with an “A”.
Row 80 the total of “B”s
Row 81 the total of “C”s
Row 82 the total of “D”s
and Row 83 the total of “E”s.

Thus all the simple straightforward analysis was done and the various members of the Steering Group looked at them and tried to summarise the results. Not surprisingly further analysis was required and this is detailed on a section by section basis.

Further Analysis

The first thing that could be done to enhance the results would be to work out how many people answered each question. In the majority of cases it was not simply a question of adding up the totals of each the possible answers, as most questions allowed the respondent to answer more than once. Even in those that did not, did not deter some people who still answered more than once.

The technique was as follows: -

If the question had three possible answers, then sorting the raw data base by these three would isolate those who didn't answer at all to the right of the newly sorted data base. By counting the number of columns up to the this point would give the number of people who answered the question. Three was easy to do because the spreadsheet allows three sorting criteria to b e done at the same time. Also the counting was easy as there is the line of numbers from 1 to 404 in row 2 which could be used. If the question had more choices, then the first three could be sorted as indicated. The next three could be sorted on the remaining data base staring at the column after the last of the positive responses to the last choice, using the same technique. This could be done as many times as was necessary to include all the possible choices and the the number of people who answered the question read off from row 2 again. All the questions which have a total in Appendix A were found in this way apart from simple questions which had, for example, only a yes or no response

Section A:

This section is only answered by the person 1 in each household, and contains questions relevant to the whole Household only. It is the only section devoted to Households. The number of questionnaires which were completed and returned is the same as the number of Households which responded. The number of Individuals that responded corresponds to the number of columns of data that were on the Data Base. The easiest way to count these was to take the vacant row 2 and 'fill' it with a series starting with 1 and increasing by 1, from column D until column OQ. This turned out to be 404. This was also useful as it could be left out of any sorting that was done on the columns and would provide any easy way of “counting” the results.

There is a discrepancy between the total of people in the Households of 448 and the total number in each age group of 457. This is due to the fact that not everybody answered the questions. More people put down the ages than mentioned how many there were in the Household. However there were 38 children who were too young to be included in the rest of the survey and somewhere around 6 to 15 who didn't take part. This section used the standard summation along the rows for most of the answers.

The technique used to determine the number of households that don't have cars was to use the following equation in row 31, a blank row: -

=IF(E4=1,IF(E25=0,1,0),0)

Where row 4 contains the person number and row 25 the number of cars in the Household.

This looks to see if E4 is 1, i.e. is person 1, if it is it then looks at E25 to see if there is a number there. This contains the number of cars, so if this is zero it returns a 1, while if there is a number greater than zero it returns a 0. This means a 1 is placed in row 31 if there are no cars in the Household. Finally the last zero in the equation returns a zero if the person number is not 1. This avoids possible duplication when someone other than a person 1 answered this question. This equation is then copied along the 404 data columns. Then by summing along row 31 gives the total number of Households without cars.

Anybody who didn't answer this question were included in those that had no cars so the number of Households without cars is likely to be slightly fewer.

Section B:

The next question that requires further analysis is B10, What age group do you belong to?

Here it was decided that the population should be split by gender, giving the Male and Female population by age separately. Additional rows were added to hold the extra data. The following equation in row 49 was used to split the data by gender: -

=IF(D$38=1,IF(D40=1,1,0),0)

Where row 38 contains a 1 if the person is male. (note: the $ is used to stop the row number 38 being copied relatively i.e. whatever row this is copied to it will always look in row 38. Without the $ the rows copy relatively, so if you copy this equation to the next row down it would then look in row 39, which in this case would be wrong). Row 40 contains 1 if the age of that person is 11 to 15.

It looks specifically at row 38 column D to see if there is a 1 meaning this person is male. If it is it then looks at row 40 column D to see if there is a 1, meaning this person is in the age range 11 to 15. If it is it returns a 1 and if not it returns a zero. Finally the last zero means it returns a zero if this person is not male, i.e. D38 is blank. Thus this row, 49, is finding all the male people in the age range 11 to 15, by copying the equation along all 404 data columns. However this equation will do more than this. By copying it down the next eight rows, 50 to 57, and along the 404 data columns from column D to column OQ, it will find all those males who are in the age range 16 to 17 in row 50, 18 to 24 in row 51 down to those males who are 85+ in row 57.

A similar equation can be used to get the female population by changing the 38 to 39 as row 39 contains a 1 if the person is female.

Summing along all these new rows will give the population by gender and age range.

However these age ranges are not equal, 16 to 17 contain only two years while 45 to 59 contain twenty years. In order to provide an even age range, it was decided to divide them up into ranges of five years. The 25 to 44 range was easily broken into four ranges 25 to 29, 30 to 34 etc. by allocating each band a quarter of what was in the 25 to 44 band. Similarly the 45 to 59 band could be split into three bands with 1/3 in each, and the 65 to 74 and 75 to 84 bands into two each with ½ of each of the original bands. The new 20 to 24 band will require five out of the seven years contained in the original 18 to 24 band. The new 15 to 19 band requires all those in the original 16 to 17 band plus the remaining two out of the seven in the original 18 to 24 band and one year out of the five in the 11 to 15 band. The new 11 to 14 band will require the remaining four years of the original band. Thus this band will only contain four years while the 85+ band will contain considerably more. It was felt that further manipulation of the data at these ends would not be worth exploring. This population was then displayed as a bar chart so that the information could be readily assimilated.

Section C:

This section is about Employment. After a considerable amount of thought it was considered that this should really only include details of those who permanently reside in the parish. I don't think we should be concerned with the employment details of those who, for example, reside in London and consequently are employed there also.

In the normal way I have included the results for All those who responded for comparison's sake, in the first column.

To get the Resident Data I sorted the Raw Data by B11 - "Are you permanently resident in Loders Parish?" And kept only those that responded "Yes". This reduced the number of people from 404 to 345 i.e. by 59. This is more than the total of people living in second homes. This difference is mainly due to the 21 people who did not answer this question. The difference, 6 people, say they are non resident although the head of the household is, which is understandable.

This has provided a subsidiary data base which I have called the Resident Data and it includes details for all the questions for those who are resident in Loders parish.

These Resident results are shown in the Summary alongside those of the whole Data Base.

This Resident Data was again sorted into those of Working Age, which we defined as above 15 and below 65 years of age The Working Age group was retained and the results shown below the others. The numbers of people below working age and those above can be determined from the relevant answers in column 2. The number of residents of working age was 213.

Finally, the last part shows the results for only those residents who are working. This was obtained by sorting the Resident Data on the rows that contained those that were working, on the row where they were either full or part time, and then on the two rows that contained those that are self employed. The remainder were rejected as they were not working.

Sections D and E:

The summary of straightforward summation is shown in the normal way.

Further detail has been provide for two of the questions.

D22 – Do you have any difficulty in getting to the following?

Hospital

Doctor

Other Medical Services

This produced results of 14, 16, and 15 people respectively. By sorting the Raw Data by these three questions it was possible to show that 9 people had difficulty with all three, and 21 people had difficulty with any of the three. In a similar manner, 12 people had difficulty with getting to the Hospital and Doctor, while 18 people had difficulty with one or the other.

For E24 it was felt that a more accurate result for some of the services would be as a Household rather than an Individual. This means we need to change the data base to reflect the combined Household returns. The method is the same as explained in Section H.

The results are shown to the right of the page at question E24. Although I have shown them for all of E24, they are probably not applicable to the Library or Fast Food,. The Online Shopping aspects are also included in Section H communications as well.

Section F:

Again the summary of straightforward summation is shown in the normal way.

In a similar way to Section C, I wondered if we should we concern ourselves with those that live here as opposed to non residents. If the non residents are second home owners then any transport answers may be about their main residence and not relevant to here in Loders. Also their time spent in Loders may be relatively short and so not representative of the major time in Loders. I have imported the Resident Data from Section C, and included the results in Appendix A: Results. It has not made a large difference to the figures.

Section G:

The summation was done in the normal way and the 'Total answering question' was found by using the sorting technique again. As most of the questions had many alternatives the sorting had to be repeated up to four times to get the final figure for those that answered the question. Again for those that had a choice of 'no opinion' the percentages are included against those that did have an opinion. For G42 the figures were calculated for both working and staying as it is, using the standard sorting technique. For G44 additional figures were produced for those that wanted at least 10 more houses but less than 20, and for those that wanted at least 10 more, again by sorting. Again for G45 for those that wanted smaller homes. Then for G46 figures were obtained for those wanting Housing Association Renting or Shared Ownership to compare to the figures of Affordable Housing for locals, as they should have both been about the same. For the next three questions additional figures were produced as required using the standard sorting technique.

Section H:

The individual aspects of communications are shown in the first column of Appendix A.

Q 54 To the right for question 54 I have expanded these results to include: -

- the totals of people who use the internet
- the totals of who do not or would never consider using the internet
- the totals of those who answered the question H54

I believe that this would give a truer picture of the percentage of people who use the internet. It may, however, be argued that they did not answer the question as they didn't use the internet and it was consequently of no interest to them. It was felt that as the internet is normally provided to each house rather than to each person, it would be more realistic to find out the results per Household. This would entail changing the data base to reflect the combined Household views rather than Individual views.

This involved setting up equations which would: -

- add the result of each person in the Household to that of person 1
- reducing this number to 1 if it was greater than 0
- reducing the value to zero for all persons other than person 1

Then by summing along the row we would have the total of Households that agree with the question.

The following equation sums all of the Household results from row 6 into the column of person 1: -

=IF(E$2-D$2<0.2,D6+IF(F$2-E$2<0.2,E6+IF(G$2-F$2<0.2,F6+IF(H$2-G$2<0.2,G6+IF(I$2-H$2<0.2,H6+I6,H6),G6),F6),E6),D6)

Where row 2 contains the Unique Number.

If we look at just one IF statement, the last one in brackets which is, in fact, in the centre of the nest:-

=IF(I$2-H$2<0.2,H6+I6,H6)

In words this means: - If the unique number in column I minus that of column H is less than 0.2, return the value in H6 plus that in I6, and if not (i.e. larger than 0.2) return only the value in H6.

Thus if the Unique number in H2 was 123.1 and in I2 123.2 then they are in the same household and the difference is less than 0.2 so the value returned would be the sum of both H6 + I6.

Row 6 contains the answers to the question “Do you use the internet for business?”

Example (a)

If H6 did not while I6 did, H6 would be zero while I6 would be 1. In this case the result would be 1, and this would be returned in the column of person 1, i.e. column H.

Example (b)

If both H6 and I6 were both 1 then the result would have been 2. Thus the total in this Household would be given in the person 1 column. It would return the value 2.

If the Unique number in H2 was 123.1 and in I2 124.1 then they are not in the same Household and the difference is more than 0.2 so the value returned would be just that in H6.

In the above examples with the two different households it would return 0 initially, and for the second example it would return 1. Again the result of the Household H (this time of only one) was shown in the person 1 column.

Nesting the IF statements extends the number of columns queried to six as up to six people per

Household are catered for in the Questionnaire. Thus each of the person 1 columns can contain values between 0 and 6 depending on how big the Household is and how many of them use the internet for business.

Equation: -   =IF(D$2=INT(D$2)+0.1,IF(D14>0,1,0),0)

Does two things:-

- It checks to see if this is person 1 by seeing if the Unique Number equals the Integer (without any decimals) of the Unique Number plus 0.1
- If it is,it is person 1 and it reduces the value obtained in the first equation (anything between 0 and 6) to 1 if it is greater than 0,or 0 if the value was 0.
- If it is not person 1, it reduces the value to zero.

These equations are copied 404 times so that all the individual responses are effected. We can then get the number of Households where at least one member uses the internet for business by totalling along the row.

I have further expanded the results to get the total number of Households in which at least one member uses the internet for any purpose, and those that do not. Here, again I used the sorting technique.

This give a total of 191 households that replied which is almost all the households in the survey. Percentages based on these figures are likely to be realistic.

For H55 which relate to problems with the internet, it was again expanded to include only those who have an opinion by removing those who said it was not applicable and basing the percentages on this new total.

The answers were also expanded to find those that had difficulty with reception or reliability using the sorting technique already mentioned. This was done again for those having any difficulty.

Section I:

The majority of these questions could be resolved by using straightforward summation in the normal way. Only I58 could be expanded.

I58 can be looked at closer because the answers cover all possible options, and are mutually exclusive (i.e. Only one response from each person). Although as can be seen this was not quite true. The total of people answering the question was worked out using the sorting technique and gave 374 out of the possible 404 and totalling all the responses gives 375. However this is typical of what is found in practise as somebody must have answered twice. This will have virtually no effect on the results. The percentages can be based on the 369 people who expressed an opinion to give a truer view of the thoughts of the parish as a whole.