A Working Paper on Full Disclosure

B. A. Wichmann
Brian.Wichmann@bcs.org.uk

1 Introduction

This document considers the following problem: given an election in which preferential voting is used and the count is conducted by computer, what information should be disclosed? Running an election consists of several stages, but here we are concerned with the counting process only. This process must not only be trustworthy, but needs to be seen as such by the electorate.

With the manual count, the full result is typically declared by a *result sheet* which contains the great majority of the information gathered during the counting process[1]. If a witnessed count is undertaken, which is, of course, the case with public elections, then all the critical information that would be available to the witnesses appears in the result sheet. The same degree of transparency is needed when a computer count is undertaken.

In the USA, under their Freedom of Information Act, full information of the ballot preferences is available for public elections. Of course, although this information is available, the identity of those who voted in a specific way is not available — ballot secrecy is maintained.

In the case of the experimental use of computers in the Irish Dáil elections in 2002, full information was available for the three constituencies polled by voting machines. It appears that the Republic has a similar Freedom of Information Act to the USA.

There are at least three different types of election in which the full disclosure questions arise: public elections; private elections performed by an independent party; and lastly, private elections performed internally. All three types of election occur with the Single Transferable Vote (and computer counting).

2 Data Protection Legislation

Public elections are typically covered by national laws, but private elections would also need to adhere to appropriate national laws. For the EU, this is largely the national laws which enact the European Directive on Data Protection. This gives data subjects the right to information held about them, and for those holding information the need to register and control access to the information.

There are two cases to consider here: those relating to the candidates in an election and those relating to a voter. Assuming that the voter is not specifically identified, then, in effect, no information is held and therefore nothing needs to be disclosed.

For the candidate, it is clear that information is held and therefore the candidate has a right to be told the information held. For a preferential voting system, it has been my opinion (based upon the 1984 Act, which was straightforward to follow), that the candidate should be informed as to how many preferences were recorded against him/her at all the various levels. Of course, the number of first preferences would be available from the result sheet, but the other preferences may not be. Hence, with a computer count, there seems little doubt that more information should be available to candidates than is provided in the result sheet.

The situation is rather more confused when one considers disclosure of more than the above. It is clear that ballot secrecy is paramount and therefore disclosure may be limited by that need. The limitation is surely minimal since ballot secrecy has not been called into question in the USA, where full disclosure takes place.

---

[1] Practices vary in this area. Working calculations should be published but may not be. For some elections, the ballot boxes are opened individually allowing a careful witness some information about the relative strengths of the candidate vote.

We consider secrecy in the next section and hence for the moment, we note current practice.

For the 2002 Irish Dáil elections, full disclosure took place. Some reservations have been expressed about this in a recent Irish report [6]. Also, in the context of public elections, Otten has pointed out a means of making bribery effective by the use of an unlikely sequence of preferences [1]. It seems that this problem has not been raised in the USA.

In the case of an independent balloting organisation undertaking a count, it is not immediately clear who 'owns' the ballot data. If it is the balloting organisation, then disclosure rests with them, otherwise it rests with their client.

Currently, Electoral Reform Services maintain that full disclosure is not possible even when the client requests it. I cannot understand this position and I am not alone in this.

3   Secrecy

Less that 150 years ago it was argued by some that secret voting was not desirable, but nowadays everyone seems to accept that secrecy is paramount. Given that, then the question arises as to whether this imposes some restrictions in applying the principle of full disclosure.

Secrecy has an important limitation. If the entire electoral process is clothed in secrecy, then the validity of the result will be open to question. Hence public elections are open to substantial external scrutiny. In our context, we are concerned with elections in which the count is undertaken by computer. It is far from clear how the process of validating a *count* should be undertaken under such circumstances. Again, we are assuming that the other parts of the electoral system perform the intended function in a manner acceptable to the electorate. The integrity of the count was part of the concern in the report on the Irish system [6].

One means to overcome part of this problem is full disclosure. Then anybody can use the data to repeat the count in order to confirm the result. (Counting software is needed, but that is readily available for almost all counting rules.) This is a *stronger* validation method than the traditional method of a witnessed manual count. When an STV manual count has been checked afterwards by using a computer, some errors are almost always found — sometimes even affecting the result!

Is ballot secrecy compatible with full disclosure? There are two possible problems: firstly, elections with a small number of votes, and secondly, the problem of a long preference list which can act as a signature for the voter.

3.1   An example — census data

It seems to me that there is a good analogy between the problem here and that in handling census data. Complete disclosure occurs after 100 years. People can also request their own data. However, substantial statistical information is made available without restriction — a clear need for Government planning. The apparent conflict is overcome by grouping information into sufficiently large numbers so that an individual return cannot be identified.

It is my understanding that the protocol that the Office of National Statistics uses was agreed with the Royal Statistical Society.

It is my contention that a similar protocol needs to be agreed for preferential election data.

4   Technical measures to ensure secrecy

It seems that there is no concern about the information available from a result sheet. I have been informed of an example in which the result sheet could be regarded as problematic. This was for the 1999 North Tipperary local election in which a candidate got no first preference votes. One could envisage a situation in which such a candidate was then hostile to his/her friends, family, employees, etc.

The preferences themselves can be revealed. Let us say one is voting in an election in which your preferences are A, B, C and finally D. It is not possible to exclude the possibility that the existence of such a voting pattern will be evident from the result sheet. For an actual example in which a long preference list was evident, see [2], which was evident due to full disclosure.

In practice, the percentage of preferences actually used in an election is quite small, so it is usual for long preference lists to consist mainly of unused preferences (see [4]). It is therefore possible to provide a form of disclosure in which some of the preferences on the ballot papers are omitted or changed, but still provide data which confirms the result of the count. In other words, there is plenty of room to provide a form of disclosure which allows for count validation but nevertheless ensures ballot secrecy.

The statistical analogy to the census data problem would perhaps be to disclose a fraction of the ballot papers. This is not a good method, since the data would then not provide a means of validating the count. I have written a program myself to make a number of changes to ballot data so that both the election and the candidates could not be identified. Unfortunately, such changes make it impossible to perform some reasonable forms of analysis, like determining if there is an alphabetic bias in the voting data.

It is certainly true that if ballot data is provided only for some forms of statistical research that a sampling method could be effective. Such a form of disclosure would be of use, but only to a very limited audience.

I am unclear how small any election should be before full disclosure could not reasonably be undertaken. If full disclosure is not provided, then the issue of count validation remains.

Finally, it should be noted that once any public form of disclosure takes place, the use to which it is put is uncontrolled. Here, we are not concerned with making information available under some form of non-disclosure agreement that might restrict its use for research purposes.

5   Conclusions

From the above, I make the following conclusions:

1. In the interests of openness and the validation of computer counting, full disclosure should be the default.
2. Legal advice should be obtained on any caveats to full disclosure as a result of the Data Protection Directive.
3. Technical measures should be agreed on how full disclosure should be implemented, given the paramount importance of ballot secrecy.
4. Purists may well object to anything other than making the ballot data available without change, but disclosure which is sufficient for count validation is surely required.

6   Postscript

Drafts of this paper have been sent to several people who I know are interested in this subject. I have tried to reflect the views of those who commented on the drafts, but this has not always been possible. Those who provided comments include: James Gilmour, Steve Todd, Joe Otten, Colin Rosenstiel, Anthony Tuffin, Jeffery O'Neill and David Hill.

David Hill was strongly of the view that no change should be made to the ballot preferences. I would prefer that, but think that it is better to have effective disclosure (in which small changes are made), rather than no disclosure which is the position with the majority of STV computer counts at the moment.

7   References

[1] Otten J. Fuller disclosure than intended. *Voting matters*, Issue 17, p8. 2003.

[2] Hill I. D. An odd feature in a real STV election. *Voting matters*, Issue 18, p9. 2004.

[3] Hill I. D. Full disclosure of data. *Voting matters*, Issue 18, p10. 2004.

[4] Wichmann, B. A. A note on the use of preferences. *Voting matters*, Issue 18, p9. 2004.

[5] Hill I. D. What would a different method have done? *Voting matters*. Issue 16. p5. 2003.

[6] Interim Report of the Commission on Electronic Voting on the Secrecy, Accuracy and Testing of the chosen Electronic Voting System. `http://www.cev.ie/htm/report/download_report.htm`