A Generic Toolkit for the Successful Management of Delphi Studies

Jacqueline Day and Milena Bobeva
Business School, Institute of Business and Law, Bournemouth University, UK
jday@bournemouth.ac.uk
mbobeva@bournemouth.ac.uk

Abstract: This paper presents the case of a non-traditional use of the Delphi method for theory evaluation. On the basis of experience gained through secondary and primary research, a generic decision toolkit for Delphi studies is proposed, comprising of taxonomy of Delphi design choices, a stage model and critical methodological decisions. These research tools will help to increase confidence when adopting the Delphi alternative and allow for a wider and more comprehensive recognition of the method within both scientific and interpretivist studies.

Keywords: Research method, Delphi, Research Design, Research Evaluation

1. Introduction

Delphi is a structured group communication method for soliciting expert opinion about complex problems or novel ideas, through the use of a series of questionnaires and controlled feedback. Delphi has been well explored in a variety of areas, including government, medical, environmental and social studies, as well as business and industrial research (Linstone and Turoff, 2002), but had limited use in Information Systems (IS) research (e.g. Brancheau *et al.* 1996; Galliers *et al.* 1994, Schmidt *et al.* 2001). In all subject domains Delphi has been primarily employed for forecasting, planning, issue identification/prioritisation, or for framework/strategies development (Okoli and Pawlowski, 2004). Thus the method has mainly been used for theory generation, rather than testing and evaluation (Holsapple and Joshi, 2002). Furthermore, particularly in the IS research, the analysis of the Delphi has concentrated upon reporting the results of the study: there has been limited reflection and evaluation upon the use of the research method itself. To address this gap, the present work provides details of the design and application of the Delphi for empirical evaluation of a framework for Information Architecture (IA) for business networks.

This research builds upon previous work that argues for wider and more comprehensive recognition of the Delphi seeing it as a method that supports either scientific or interpretivistic studies (Day and Bobeva, 2004). The paper also seeks to further address the gap in research methodology by consolidating the authors' experience with Delphi into a toolkit that includes generic stage and critical decision factor models (Fig.1). The purpose of this toolkit is to enable the viability and appropriateness of a Delphi inquiry to be established and to identify the practical limits upon its use. Key decisions have been identified to help guide the conduct of an inquiry, by isolating those decisions that have the most influence upon the standard of the final research deliverables. If substantiated as a well-proven and robust set of communication processes, the Delphi enquiry could gain wider use for many, if not most, studies of the managerial and business aspects of information systems and technologies.
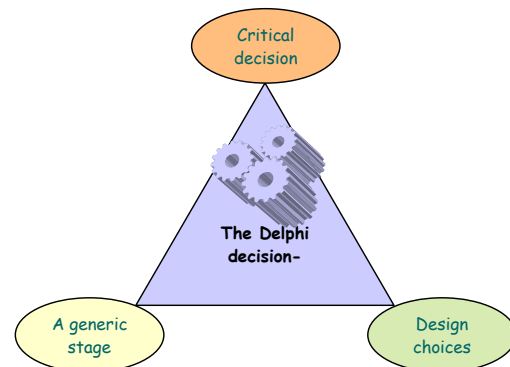

Figure1: The Delphi decision-making toolkit

2. Characteristics of the Delphi method

The Delphi is founded upon the use of techniques that aim to develop from a group of informants an agreed view or shared interpretation of an emerging topic area or subject for which there is contradiction or indeed controversy. Delphi shares that part of the research method continuum with several other consensus-seeking approaches. The family of methods have been mapped within Figure 2, which shows a simple arrangement, reflecting the relative degree to which the researcher and informants are interactively engaged with each other during the course of the research process. Most of the ways for acquiring data or evidence will be quite familiar

to business researchers, except perhaps Nominal Group Technique (NGT). This is closely related to brain storming but follows a highly structured agenda for building the base of knowledge through the use of formal means to involve all participants. An example of NGT is information systems requirements definition by the use of Joint Application Development workshops (Andersen 2000).
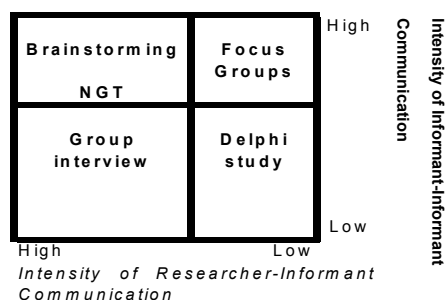
Figure 2: A classification of consensus seeking methods

Although Delphi emerged in the 1950s in a project with strategic importance and was first reported in 1964, it could be argued that in comparison with the usual survey, Delphi is still in its developmental stage. However, the method has gained recognition mostly amongst communities dealing with complex problems, where Linstone (1978) argues that Delphi can be invoked as a method of 'last resort', being particularly suitable for situations where:

i. the gathering of subjective judgements moderated through group consensus is the only approach possible in the absence of precise analytical techniques, or

ii. personal contact is not possible due to time and cost constraints, or is not desirable, due to concerns about the difficulty of ensuring democratic participation.

Whatever the perceived reasons for its choice, the method offers reliability and generalisability of outcomes, ensured through iteration of rounds for data collection and analysis, guided by the principles of democratic participation and anonymity. However, Turoff and Hiltz (1996) warn that because of its emphasis upon communication, Delphi can be in danger of dismissal as merely a form of data collection, when it is much more than this. Its iterative feedback method develops an insight, which in its totality, is more than the sum of the parts.

A problem with Delphi is that it is difficult to draft an explicit all-encompassing definition of the method. Here the present authors concur

with Linstone and Turoff (2002). A preliminary taxonomy of Delphi design variations demonstrates that in addition to subject domain criterion the wide spectrum of Delphi applications may be categorised in terms of the following:

- Purpose of the study: building, exploration, testing, evaluation. The prevalence of studies focusing on developing and exploration has been discussed at the beginning of the paper.

- Number of rounds: varying between two and ten (Lang, 1994; Errfmeyer *et al*, 1986) but most commonly restricted to two or three rounds. Gottschalk (2000), however, in his comparison of methodological choices identifies Delphi studies with only one round.

- Participants: homogeneous or heterogeneous groups. The profile of the participants could be defined by age, nationality, knowledge, expertise, qualifications, occupation or position and thus could be used to further differentiate between two applications of the method. Of particular importance to potential users of Delphi is establishing the expertise of the participant (Gordon, 1994) that affects the quality of the outcomes.

- Mode: face-to-face discussion or remote access. This classification is linked to the anonymity of the participants. Participation through postal or electronic communications allows ensuring full anonymity of the informants.

- Anonymity: full or partial. This was a key element of the original Delphi process. It is a principle that can be sustained even with face-to-face contact, if the study is appropriately designed. Although with these cases it is not possible to ensure full anonymity, as the participants will know each other, their contributions to the study can remain anonymous. Similar considerations apply to technology-based Delphi implementations (see below).

- Media: paper-and-pen based, through telephone/fax, or computerised. Developments in information and communication technologies (ICT) have stimulated and been driven by changes in business and society. The convenience of electronic communication has steered the evolution of the Delphi toward computer-mediated studies. This could foster further developments, including support from multi-media, simulation and modelling tools and altogether boost new research opportunities for the method:

> *"When this technical capability is coupled to the knowledge being gained in the area of Delphi design, all sorts of opportunities seems to present themselves."*
> (Linstone and Turoff, 2002, p. 483)

- Concurrency: classic sequential set of rounds or real-time online conferencing. A consequence of ICT, such as video conferencing, is opportunities to vary concurrency modes, depending upon the nature of the problem and the urgency of its resolution.

This rich menu of Delphi applications could be developed further through designs based upon the different combination of types (Table 1), which forms the tool for Delphi studies:

Table 1: Taxonomy of Delphi Inquiry designs

Criteria	Choice
Purpose of the study	building, exploration, testing, evaluation
Number of rounds	between two and ten
Participants	homogeneous or heterogeneous groups
Mode of operation	face-to-face or remote access
Anonymity of panel	full or partial
Communication media	paper-and-pen based, through telephone/fax facilitated, computerised
Concurrency of rounds	sequential set of rounds or real-time online conferencing

3. The Delphi Stage model

The second component of the toolkit is best understood through a detailed examination of an example. The empirical case chosen to illustrate the Delphi Stage model is work that was done to test the desirability and feasibility of an information Architecture to support e-business alliances. Although the example is, of course, specific to a particular variant of the method, the model itself is grounded in Delphi research undertaken by many authors. Generalisation allows the model to be used for all Delphi inquiries, regardless of its subject field or purpose, including the building, exploration or evaluation of strategies, policies, theories, etc.

3.1 The Delphi test for an Information Architecture framework

Information Architecture could be defined as a blueprint for principles, guidelines, standards and models for information requirements, systems development and management. It has long been seen as major issue for IS managers (Galliers *et al*, 1994, Galliers, 1995; Pavlia and Wang, 1995; Pavlia *et al.*, 2002;

Watson and Branchau, 1992). Although the spectrum of research topics is generally very rich, to date, little attention has been paid to specific architectures for electronically integrated businesses. The work carried under the banner of IA has either been constrained by organisational/sector boundaries or the conflation of web architectures with information architectures (Rosenfeld and Morville, 1998). Furthermore, since e-mediated business networks are at an embryonic stage, research was needed to sense the opinion of experts in this field and challenge them to achieve a consensus about the key components of a theoretical framework for an e-business IA.

Delphi was thought to be suitable because it would enable a clear epistemological stand to be taken in this study, i.e. the post-positivist position that enforces the merits of the scientific inquiry with interpretivist features. The fact that Delphi has not been fully exploited for defining the characteristics of complex entities, such as IS/IT architectures (Mulligan, 2002), neither for theory testing/evaluation, was also a strong incentive for testing the feasibility of the method.

Apart from the constraints imposed by the goal of evaluating the desirability and feasibility of the IA framework, the study were influenced by the specifics of both the research population, which was geographically distributed and a high degree of reliance placed upon electronic communications with information specialists. To maximise the quality of the outcome and address concerns for methodological rigour, the Delphi study was triangulated with a parallel electronic survey and follow-up evaluation interviews. The survey employed a similar data collection instrument and was used as a control group. However, the interviews and their use for evaluating the quality of the IA framework are beyond the scope of this paper. An example showing the use of design choice selection tool is presented as Table 2.

Table 2: The Delphi design choices in th e study of Information Architecture

Criteria	Choice for IA Study
Purpose of the study	testing
Number of rounds	three
Participants	heterogeneous group
Mode of operation	remote
Anonymity of panel	full
Communication media	paper-and-pen based
Concurrency of rounds	sequential

3.2   Key stages in a Delphi study

An analysis of the process followed and the relevant literature yielded a generic Delphi model, comprising three stages: Exploration, Distillation and Utilisation. Details of the activities and results of each stage of the empirical research are shown in Figure 2 and discussed below.

The first stage, called 'Exploration' (Linstone and Turoff, 2002; Ziglio, 1996) is a free-flowing and unstructured investigation of the issues, limitations, challenges and problems that affect or are affected by the elements within the study domain. It includes the following activities:

- establishing criteria for selection of participants
- establishment of a Delphi panel
- design of the data collection and analysis instruments
- eliciting the initial set of issues to be tested through the Delphi rounds, and
- piloting of the toolkit.

Previous Delphi research has recognised that preparatory effort is necessary before the start of the rounds, but does not distinguish this work as a separate stage. Furthermore, the development of the initial base of knowledge has been traditionally considered as a first round of the study. The starting position for the Delphi (the first version of the questions) can be established by either exploratory or confirmatory in nature. The former is best implemented by seeking the views of the informants through initial open-ended question or a set of preliminary interviews (Hasson *et al.* 2000). This approach is particularly apposite for the more vague, ill-defined or contradictory situations often found in social, political and/or organisational worlds. It emphasises the qualitative dimension to the Delphi and hints that it might be successfully integrated with other research methods (see Section 4). The confirmatory form of the Delphi initial stage is traditionally carried out by circulating a predefined list of issues to the panel (Niederman *et al* 1991). This is typical mode of working for follow up studies (Brancheau *et al* 1996, Gottschalk 2000). For the present work the introduction of the stage-organisation allows for the initial round to be distinguished from the rest of the Delphi iterations, based upon the differences in goals: that is, 'generation' vs. 'evaluation and extension'. This refinement allows for further development of Delphi variants, allowing application for theory testing and extension,

where the theory is generated through either secondary or primary research. In this particular application of the method for evaluation of a conceptual IA framework, the list of issues presented to the participants reflected the perspectives and components of the proposed framework, thus negating the need for an initial 'generation' round.

This modification to Delphi has been employed in other studies (Custer et al. 1999; Doke and Swanson, 1995; Mulligan, 2002). Its merits are seen primarily for addressing the weaknesses arising from the unstructured polling of participants' views, by replacing this initial collection of opinions with synthesis of key issues identified in the literature or through preliminary interviews with selected domain experts.

The remainder of the present study was focused on developing a group consensus about the architectural framework components and adhered to the normal form of inquiry for a 'ranking type' Delphi (Schmidt, 1997). This included a series of three rounds, conducted over a period of twelve months. After collecting the participants' scores, a convergence ratio was determined. This represents the extent of participant agreement about the ranking of the architectural elements. Measuring the percentage of votes that fall within a prescribed range is a common approach to assess consensus. However, Scheibe, Skutsch and Schofer (1975), cited in Linstone (1978), argue that this is not the most reliable of measures and accordingly two further tests were implemented:

- Stability: monitoring the permanence of respondents' vote distribution over successive rounds. This is idea is based upon Linstone's (1978) view that opinion stability reflects consensus. He suggests that marginal changes of less than 15% offer a working definition of a threshold for stability, which might be used as a criterion for termination of the Delphi exercise.
- Participation: whether the numbers of participants will drop below a critical level. Ziglio (1996) asserts that useful results can be obtained from small size, homogeneous groups of 10-15 experts; whilst Dalkey, cited in Linstone (1978, p.296), found that seven is a suitable minimum panel size. This was a useful measure in a case such as this study, where the drop in response rates was about 40% after at each round.
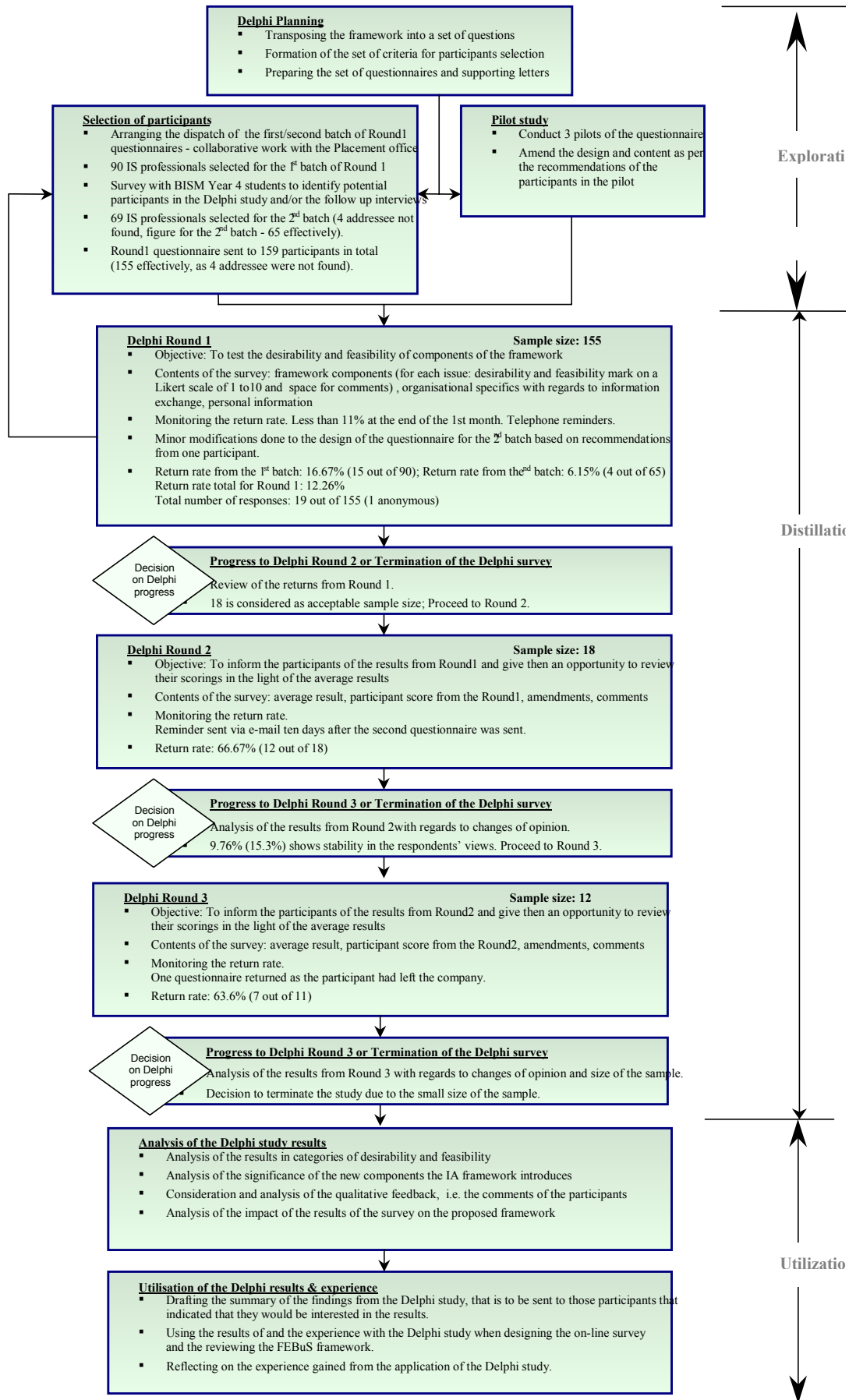
Figure 2: Delphi study implementation process

Repeated attempts at opinion-seeking and subsequent analysis to see whether or not the Delphi had reached a critical point for study termination, forms the Distillation stage in the generic Delphi model. It is a distinctive feature of this study that the three tests were applied at each round. Anecdotal evidence from Delphi studies across various disciplines suggests that few inquiries use more than one of these tests.

To be consistent, the first two stages have been named in accordance with the terminology devised by Ziglio (1996), although with non-similar content. The last stage, Utilization, has not been recognised as distinct stage until now. The present paper proposes that the activities within this stage be classed as 'short' or 'long' term in nature. The former includes the development and dissemination of the final report on the study. Long-term activity is knowledge dissemination about the Delphi exercise.

4. Some critical issues in Delphi inquiries.

The third and last component of the generic Delphi toolkit is a checklist that can be used to ensure that key operational decisions about the management of a Delphi study have been thought through. A link may be made with the stage model discussed in Section 2.2. The researcher's knowledge, ideals, awareness, flexibility and openness to feedback will affect the extent to which these key decisions are able to be fully addressed (Fig.3).
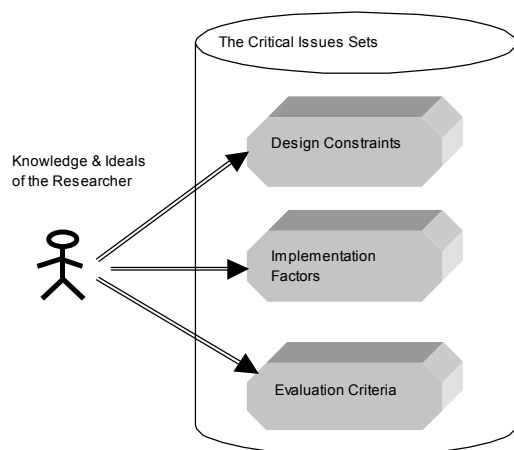

Figure 3: Critical methodological issues in a Delphi inquiry

4.1 Design constraints

A Delphi study, even if simplified as an integrated series of surveys, each

progressively building upon findings, requires more time and effort than many of the methodological alternatives. As such, it is unlikely to be helpful where knowledge can be relatively easily and fully acquired through direct measurement, experiment or simulation. Thus a clear justification for the choice of a Delphi inquiry, are the characteristics of the participants, sample size and design of the data collection instruments. These are the first critical decisions a researcher will face.

4.1.1 Critical issue 1: Choosing the approach

Given that the theoretical foundations of an inquiry, such as the epistemological position and research axioms have already been established, an early decision that confronts the researcher is the choice of an approach that meets the requirements of the research philosophy.

For positivist research, Delphi can be used for defining the characteristics of complex areas (such as IT strategy) and for testing general propositions (nomothetic perspectives), including those concerning the attitudes of social actors. The latter was the case with the IA research that has been used as the exemplar. Since Delphi inquiries are anchored in aggregations of opinion, they are not helpful for investigating psychosocial conditions of an individual, so Delphi is not recommended for research where nuances and experiences of human individual behaviours must be studied in situ. Here evidence must necessarily be gathered through observation of people and the context for their actions. However, since it does not disturb the naturalness of the setting, Delphi can support some types of field studies. The iterative cycle of Delphi is supportive of longitudinal investigations. For example, theory development directed toward building a process model from a factor-based framework.

4.1.2 Critical issue 2: Initial Selection of the Panel

A Delphi panel should consist of individuals with knowledge about the substantive area of research, the motivation to engage with the inquiry process and be able to articulate judgements. The first problem with the selection of informants is when there is no clear definable community that can act as source of expertise, knowledge or opinion. This is often the case of future studies, where the existing knowledge base is of limited value for predicting the future and is best served by constituting the panel from experts drawn from

wide range of subject areas. Similarly, there can be difficulties where informants are identifiable or the group is so homogenous and constrained by conditions that diversity of opinions cannot be represented sufficiently strongly to drive the goal-seeking heuristic of Delphi. Here, the researcher might be in a quandary as to the attitudes of the 'expert' to be recruited to the panel: should they hold strong, positive opinions or is it better for them to be well-balanced and impartial? Goodman (1987) in Hasson *et al* (2000) argues that individuals are more likely to get involved in the Delphi inquiry if they are to be affected directly and/or profoundly by the outcome of the study. Thus for political and medical topics studied neutrality is likely to be less useful than for business management or IS research. It is the nature of Delphi to 'trim' the extreme cases but we argue later that in all situations these extreme or maverick opinions should be carefully conserved because they can act as compare-and-contrast case for establishing the validity and generalisability of consensus opinion.

If there is a large potential population of experts then the researcher must decide the basis on which the sample is to be chosen to form the panel. The choice remains, as for other forms of research, between probability and non-probability (purposive) sampling (with all the possible variations possible within the latter). In practice, the strategy selected is likely to depend upon the nature of the research problem: the narrower the scope, the greater the depth and specificity of expertise needed and the more likely a purposive approach is appropriate. On the other hand, for wide-ranging, social and marketing Delphi studies, an initial 'random' sampling can be used.

The sampling regime that is implemented will be also be influenced by the form of the initial contact made with participants: cold calling or after some groundwork has been done. It is probable that a more personal approach will result in a positive response to an entreaty to enrol as a panellist. However, no matter what initial sampling strategy is adopted for all types of Delphi study, the research outcomes will be derived form data collected from a self-selected sub-set of the original population.

In the empirical IA study, panellists were selected through non-probability sampling. More than half of the respondents were IS project managers with widely differing levels of experience. Other individuals were employed in managerial positions or as consultants in systems analysis, data architecture, outsourcing or sales. Dalkey, Brown and Cochran (1969), cited in Linstone (1978, p.296), argued that self-rating of expertise in the area of research can be used to improve the efficacy of a Delphi inquiry.

Finally, with regard to the constitution of the expert group, it is important to recognise that the Delphi researcher is not immune to implementing good general practice, common courtesies and ethical standards. The principle to be followed here is meeting expectations. For example, clearly informing participants upfront of what is expected from them and when and by the researcher ensuring that promises made to panellists concerning their anonymity or within the subsequent dissemination of the study results are kept.

4.1.3 Critical issue 3: Ongoing Management of the Panel

The drop out rate for a Delphi study could be high after the first few cycles thereby reducing the sample population for subsequent statistical evaluation and indeed this is affect was experienced by this study. The size of the panel that is needed at each round is a decision that must be taken before the execution of each these iterations. Although the largest Delphi undertaken in Japan involved several thousand people (Linstone 1978), most studies use panels of between 15 to 35 people (Gordon, 1994), though Dalkey (2001) suggests that seven as the minimum number. Once the requirements for participant expertise (Issue 2) have been clarified, different sampling strategies can be used to constitute the panel. In the IA project, 155 IS professionals were approached, of which 19 (12 %) agreed to take part in the initial Delphi round.

4.1.4 Critical issue 4: Instrument design

The design of the data collection instrument is critical for both the Exploration and Distillation stages. Creativity is always a good watchword for any data collection mechanism, so different ways of conceptualising the topic and structuring the questions should be tried to match communication mode with the individual preferences of the informants. The design of the instrument will obviously depend a good deal upon the number of questions asked. There are no clear rules for this. Hasson *et al* (2000) suggest that the minimum number of issues should be six, but this does not exclude the possibility of there being more, e.g. ten

(Niederman *et al* 1991) or 34 as is the case for the present study. The number issues explored will reflect the complexity of the problem and to the type of data collected.

Further, the researcher should be aware of structuring the questions in a way that implies an answer or does not properly allow for different views or for a novel re-framing of the problem. The results of the present study also bear out well-proven ideas about survey design. The key is formulating clear, concise and unambiguous questions, together with clear instructions for the participants. Experience supports the common sense view that aesthetically pleasing and easy to fill in questionnaires positively influences an informant's decision whether or not to take part a study.

For the IA study a 10-point Likert scale was used, where 'one' coded the least desirable/feasible constituent and 'ten' for the most desirable/feasible IA component (see Figure 4 and 5 for example of questions and layout). The flow of the questions deliberately did not reflect the logical organisation of the framework, with some elements assigned a higher priority than were others. Rather, questions were grouped by perceived simplicity of the subject matter, starting with questions about the more straightforward topics of IA. Space for comments and/or clarifications was also provided adjacent to each answer block. On completion of the core part of the instrument, further questions were then posed about information exchange, so as to probe the extent to which the participant's organisation was a true member of an e-business network. This was also done to identify appropriate informants for the follow-up interviews. The Delphi survey instrument concluded by asking personal information, such as contact details and preferred method for communication, as well as ascertaining whether the panellist would be interested in the results from the completed study.

Figure 4: Questionnaire layout for Round One (Page 1 of 3 shown only).

	Things you want to know/need to know about the information you use	Desirability (5-level desirable; 10=most desirable)			Feasibility (5-level feasible; 10=most feasible)			Your comments
	Brief description	Average score	Your score	New score	Average score	Your score	New score	Suggested changes, questions
1	Knowing who the source or recipient of the information is;	8.42	«M_1D»		7.09	«M_1F»		
2	Knowing which team/department/organisation is the source/recipient of the information; Is it internal or external for the organisation;	8.74	«M_2D»		8.05	«M_2F»		
3	Knowing what the role of the source/recipient within the project is;	7.56	«M_3D»		7.29	«M_3F»		
4	Knowing what processes use this information;	7.59	«M_4D»		6.12	«M_4F»		
5	Knowing what the importance of the information (e.g. strategic/operational/general; administrative; adding to organisational or personal knowledge) is;	8.16	«M_5D»		7.79	«M_5F»		
6	Knowing what risks exist related to this information; What security measures are needed for this information on the sender's/recipient's side;	8.16	«M_6D»		7.28	«M_6F»		
7	Knowing who the "owner"/originator of this information is;	7.42	«M_7D»		7.61	«M_7F»		
8	Knowing who the controller of the quality/performance of the information is;	7.00	«M_8D»		6.50	«M_8F»		
9	Knowing whether the information is stable or dynamic; How often it is upgraded;	8.09	«M_9D»		7.32	«M_9F»		
10	Knowing what the format of the information carrier is (e.g. text file, diagram, spreadsheet, presentation, document image, image);	7.58	«M_10D»		8.32	«M_10F»		
11	Having access to the information in electronic format, rather than paper or verbally;	8.16	«M_11D»		8.05	«M_11F»		
12	Knowing what the style of the information is (e.g. formal, informal, personal); How structured is it;	6.26	«M_12D»		7.11	«M_12F»		
13	Knowing the level of aggregation, i.e. how detailed or summarised the information is;	7.00	«M_13D»		6.68	«M_13F»		
14	Knowing how current/up-to-date the information is;	8.79	«M_14D»		8.26	«M_14F»		
15	Knowing how the information is described (e.g. what languages, models, tools are used to describe and process the information);	7.06	«M_15D»		6.65	«M_15F»		
16	Knowing whether templates are available for this information and if so, how to obtain them;	6.44	«M_16D»		7.22	«M_16F»		
17	Knowing what events affect this information (e.g. year end);	8.08	«M_17D»		7.59	«M_17F»		

Figure 5: Questionnaire layout for Round Two and Three (Page 1 of 2).

4.2 Implementation factors

In addition to the instrument design and sample size there are a few other decision points throughout the implementation of the Delphi rounds. These are discussed below.

4.2.1 Critical issue 5: Timing of the Delphi survey

This issue was encountered within the IA study and it can have a significant impact upon the stability of the results and the response rate. Greater volatility of the participant scores was observed in the third, compared to the second Delphi round. This could be attributed to the time that elapsed (ten months) between the second and the third iterations, compared to two months between Rounds One and Two. The wider time gap allowed for more change in an individual's circumstances, knowledge and situational context. Although the period chosen for dispatch of the questionnaire was designed for the maximum availability of potential respondents (and thus the highest expected response rate) the first round rate was 12%, and a third of the participants opted out of the study at each of the next two rounds. This limited response may be due to 'survey fatigue': perhaps, nowadays there are too many surveys and some antipathy arising from the failed promises of e-business!

4.2.2 Critical issue 6: Proactive management of the study

This refers to the researcher's ability to maintain a high level of communication during the rounds to enable an adequate level of response to be obtained and to know when to terminate the study. The convergence, stability and participation tests outlined of Section 2.2 provide a good sense of the development of the inquiry after round and could be used as criteria for termination of the study. As mentioned earlier, anonymity is a key feature of Delphi studies, though computer-mediation can enable conversations between informants to take place and may help panel members to more readily consider acceptance of a wider diversity of views. Complementing the traditional version with conferencing features may add significant advantages but presents a challenge to the skilful management of Delphi, because the researcher must be able to facilitate on-line discussions.

4.2.3 Critical issue 7: Documentation of the results

In addition to the obvious task of clearly documenting the results of all rounds, the authors argue that documentation of results includes consistently recording divergent views at a similar level of detail. It is also important that informants are encouraged to provide reasons for the change of views and these also should be captured as part of the

base of data. Restrained use of graphical presentation of the development of the consensus (Malhotra *et al* 1994) is recommended here not only to help the researcher manage the operation of the rounds but to enable informants to better locate their individual views within the consensus. By doing this ownership of the research is engendered and continued participation is maintained.

4.2.4 Critical issue 8: Analysis of the inquiry results

Problems can arise in cases where the data collection instruments employ an interval scale (such as Likert) to measure the importance of an item since a correct choice of statistical analysis methods must be made. The statistics included as part of reported Delphi studies mainly give details about significance in t-test. However these are parametric tests normally used with continuous variables, instead of non-parametric tests that should be used for analysing ordinal data. A standard way to present the Delphi results of a single iteration can be based upon the relative importance of the issues, as determined by the mean values for each element of the instrument. Both descriptive statistics and non-parametric tests were employed in the IA Delphi for exploring relationships and for comparing data groups. These analyses included a Friedman test of the changes to desirability and feasibility scores across the three rounds. Mann-Whitney 'U' tests were also used to examine the difference between the set of scores on feasibility and desirability of Delphi study participants compared to those in e-survey. A closer examination of the change patterns of the responses revealed that there were four extreme cases, one with an exceptionally high rate of changes and three with no change. Other Delphi studies were consulted for guidance about how to treat extreme results, but the literature provided no details about the change patterns but only change rates.

There was also a need to analyse qualitative evidence, so biases of the participants and researchers could be properly acknowledged. A difficulty with this is that at the moment there are few tools available for processing a large number of non-numerical, unstructured, and rich data sets that can be captured with in Delphi studies.

4.3 Evaluation criteria

The final set of decisions is concerned with making sense of the results obtained from a

Delphi inquiry and ascertaining the quality of the final outcome.

4.3.1 Critical issue 9: Reviewing results

Given the claim that Delphi can span the divide between the positivist/quantitative and interpretative/qualitative ideals (Day and Bobeva, 2004), then range of evaluative perspectives partaking of both traditions is needed to review the quality of the Delphi findings. The trustworthiness criteria of confirmability, credibility, transferability and dependability could complement or replace the positivist criteria of objectivity, validity and reliability.

Confidence levels

In the exemplar the researchers acted purely as facilitators and not participants, but were aware of the threat that their subjective interpretation could have when reflecting upon changes in and the motives for the results from each round. Failure to understand the context for the consensus may lead to subsequent failure to capture important contextual information. In the absence of self-assessment tests for individuals forming the panel, systematic biases can arise from the judgements made by the participants reflecting the heuristics employed. A common problem with any non-administered survey is that psychological factors, causing random and systematic errors to impact upon the study and which may be difficult to detect. These 'random' errors can be attributed to factors such as work pressures, the time of the day when the survey was completed, the mood of the informant, as well as characteristics of the instrument itself, such as unclear instructions or ambiguous questions.

To enable confidence to be placed in the primary data, in addition to the criteria for reliability and validity of questionnaires suggested by Mitchell (1996), a detailed rationale should be provided showing how the question set was initially established and employed. This could be accompanied by an audit trail providing a chain of evidence through the data analysis stages.

Rigour

When evaluating the rigour of the research findings the criteria of internal validity could be strengthened by examining plausibility and consistency of the participation results of the panel members. Delphi implements these tests of quality through the continuous

—

feedback and confirmation by the contributors but researchers must exercise care in the execution of the rounds those important contextual changes are detected and properly acknowledged.

To maximise confidence in the example project, the formulation of the architectural components was based on existing models and theories. Evidence was accumulated to confirm the most common components, derived from these existing architectures and how they should be represented in the proposed framework. The elaboration and comparison strategy proposed by Reichardt (2000) was then used to improve development by the reporting of average scores for each construct in the framework. However, a significant issue for the present research was the ability of the questionnaire to properly represent the proposed framework, as in some instances, a single question addressed more than one information category. Usually this reflected a relationship between the data categories.

Credibility

Lee (1999) rightly observes that the use of multiple informants helps to strengthen credibility. For the IA study this meant comparing the results derived from one part of the research population with another. That is, by testing the views of the panellists against the on-line participants and interviewees to support a form of 'triangulation.' Minor variations found between the groups could be well attributed to the different contextual settings of individuals, as well as the capability of each person to understand the scope and objectives of the study. The strategy adopted to reduce this threat to plausibility was the inclusion of reasons for the investigation and by providing the researcher's contact details for further clarification.

According to Lincoln and Guba (1985) replication of outcomes from another context is an acid test for external validity but this non-meaningful for Delphi studies. Gordan (1994 p.1) explains why this is so:

> *"Because the number of respondents is usually small, Delphi's do not (are not intended to) produce statistically meaningful results; in other words, the results by any panel predict the response of a larger population or even a different Delphi panel. They represent the*

> *synthesis of opinion of the particular group, no more, or less."*

For Delphi studies it is more appropriate and useful to take a qualitative perspective by examining the results of the Delphi for their cogency, relevant and plausibility, by identifying the explicit limitations upon transferability of the results to other contexts. This requires that questions must be devised about informant's personal situations for the judgements offered by panel experts. This was not possible for this study since the different situations of each participant precluded testing the general applicability of the framework.

A summary of the critical issues discussed above is given in Table 3. In addition, Table 4 could be used as a reference for the set of issues pertaining to each Delphi stage and/or as a timeline of each issue throughout the Delphi stages.

Table 3: Critical issues in Delphi studies

Design constraints	Implementation factors	Evaluation criteria
(1) Choosing the approach	(5) Timing of the survey	(9) Reviewing results: Complementing qualitative with quantitative quality criteria.
(2) Informant selection	(6) Proactive management	
(3) Sample size	(7) Documentation of results	
(4) Instrument design	(8) Analysis of results	

Table 4: Critical issues cross-referenced to Delphi process stage model

Issues \ Delphi Stage	CI 1	CI 2	CI 3	CI 4	CI 5	CI 6	CI 7	CI 8	CI 9
Exploration	✓	✓	✓	✓	✓	✓			
Distillation			✓	✓	✓	✓	✓	✓	
Utilisation						✓	✓	✓	✓

5. Further development of Delphi and of the toolkit

There are two significant trends to the environment for the conduct of Delphi studies that will require the continued development of the tool kit. The first and perhaps most obvious new element is the advent of personal electronic communication. Incorporating this into the picture means that it is possible to recognise some hybrids derived canonical forms depicted earlier. The importance of this is the recognition that Delphi does not have to be used as a single approach but can be used profitably with other techniques.   Figure 6

shows attested examples in the literature of the mixed forms and also indicates (question mark) where there is an opportunity for future research and development of other multi-faceted research methods that draw upon the strengths of several parent methods and different modes of communication. For example, the embedding of focus groups, brain storming or interviews within the initial stage of Delphi panel formation or the use of computerized conferencing for real-time Delphi studies (Turoff and Hiltz 1996; Linstone and Turoff 2002).

The other major dynamic concerns the evolving social and organisational settings for group consensus seeking approaches in general, and for Delphi, in particular. Traditionally, the PESTE (Political, Economic, Social, Technical and Environmental) factors have thought to apply to business but they also are likely to impact upon the ways and means for conducting business and management research. This could include, for example, 'extended' Delphi studies, where the content of the feedback is enriched to include more anecdotal and qualitative information to enable a more informed decision to be made on the degree of consensus (Bobeva and Day 2005). There is much work to be done in reflecting legislation requirements (such as privacy) in research design and by exploring ways to tailor research processes to properly reflect international settings. For example, in some societies, open debate and consensus seeking is a pervasive cultural norm in others a more formal and authoritarian management style applies. In latter situation, maybe Delphi needs shifted towards NGT model with electronic mediation that enables individuals to contribute ideas without being worried about being seen as 'stepping out of line.'

Brainstorming	**NGT**	**Focus Groups**	**High** **Informant-Informant Communication** Face-to-face
Brainstorming ?	*NGT ? Computerized Delphi*	*Focus group (computer-based conferencing)?*	E-mediated
Group Interview	**Extended Delphi Study** **Survey**	**Traditional Delphi Study**	Limited communication **Low**
High Participatory	Non-intrusive	**Low** Anonymous	

Researcher-Informant Communication

Figure 6: New technologies and consensus seeking methods

6. Conclusions and recommendations

The toolkit discussed in this paper has been proposed as a means of managing a Delphi enquiry. The design of the three tools recognises the critical decisions in the conduct of the inquiry and identifies practical techniques that if implemented, should help to ensure a rigorous and valid study. Much work, however, remains to be done, particularly in the areas of communications, results analysis and in ways of monitoring and controlling research quality. Although the experience of using the Delphi has encouraged the authors to expand their personal research repertoire, generally the power of Delphi as an effective research method has unfortunately remained obscured through a lack of understanding. The contributions of other researchers are therefore vital to break through this conceptual barrier. Their efforts will be well-rewarded since they will acquire a flexible and simple way for exploring and evaluating many challenging topics in the realm of technological, managerial and organisation studies.

References

Bobeva, M. and Day, J. (2005) Managing Delphi Studies: A decision-making toolkit, in Remenyi, D. and Cass, A. (2005) *Proceedings of the 4th European Conference on Research Methods in Business and Management Studies*,

Universite Paris-Dauphine, Paris, France, 21-22 April 2005, pp.55-66.

Brancheau, J., Janz, B. and Wetherbe, J. (1996), Key Issues in Information Systems Management: 1994-95 SIM Delphi Results, MIS Quarterly, Vol.20 No.2, pp.225-242.

Custer, R., Scarcella, J. and Stewart, B. (1999), The Modified Delphi Technique – A Rotational Modification, *Journal of Vocational and Technical Education*, v15 n2, Available at: http://scholar.lib.vt.edu/ejournals/JVTE/v1 5n2/custer.html [Accessed on 10 January 2005]

Dalkey, N.C. (n.d./2004), *The Delphi methodology* [online]. Available from: www.fernuni-hagen.de/ZIFF/v2-ch45a.htm [17 January 2005]

Day, J. and Bobeva, M. (2003) Seeking the truth: The use of Delphi studies for IS research, in Grant, K., Edgar, D.A. and Jordan, M. (eds) (2004) *Reflection on the past, ,making sense of today and predicting the future of information systems*, 9th Annual UKAIS Conference Proceedings , Annual Conference, 5-7 May 2004, Glasgow Caledonian University, Glasgow.

Doke and Swanson (1995) Decision variables for selecting prototyping in information systems development: A Delphi study of MIS managers, *Information & Management*, 29, 173-182.

Duggan, E.W. and Thachenkary, C.S. (2004) Supporting the JAD Facilitator with the Nominal Group Technique, Journal of Organization al and End User Computing, 16:2, 1-19.

Errfmeyer, R.C., Erffmeyer, E., & Lane, I. (1986), The Delphi Technique: An Empirical Evaluation of the Optimal Number of Rounds. *Group & Organization Studies*, v.1:1-2 (March-June), pp.120-128.

Galliers, R.D., Merali, Y. and Spearing, L (1994), Coping With Information Technology? How British Executives Perceive The Key Information Systems Management Issues In The Mid-1990s, Journal Of Information Technology, Vol.9 pp. 223-238.

Galliers, R.D., (1995), A Manifesto for Information Management Research, *British Journal of Management*, No.6 (Special Issue, December), pp.45-52.

Gordon, T.J. (1994) The Delphi Method in *Futures Research Methodology*, *AC/UNU Project*, Available online:

www.futurovenezuela.org/_curso/5-delphi.pdf [6th June 2004].

Gottschalk, P., (2000), Studies of key issues in IS management around the world. *International Journal of Information Management*, 20, 169-180.

Hasson, F., Keeney, S. McKenna, H. (2000) Research guidelines for the Delphi survey technique, Journal of Advanced Nursing, 32:4, 1008-1015.

Holsapple, C. and Joshi K, (2002), Knowledge manipulation activities: results of a Delphi study, *Information & Management*, Vol.39 No.6, pp.477-490.

Lang, T. (1994), An overview of four future methodologies, The Manoa Journal of Fried and Half-fired ideas (about the future), Available at: *http://www.soc.hawaii.edu/future/j7/LANG .html* [24 February 2004]

Linstone, H.L.(1978), The Delphi Technique. In: J. Fowles, Ed. *Handbook of Futures Research*. London: Greenwood Place.

Linstone, H.A. and Turoff, M. (Eds.) (2002), *The Delphi Method: Techniques and Applications,* Addison-Wesley Publishing Company Inc, Reading, M.A. Available at: http://www.is.njit.edu/pubs/delphibook/ [25 June 2004]

Malhotra M.K., Stelle D.C. and Grover V. (1994) Important strategic and tactical manufacturing issues in the 1990s. *Decision Sciences,* 25:2, 189-214.

Mitchell, V., (1996), Assessing the Reliability and Validity of Questionnaires: an empirical example. *Journal of Applied Management Studies*, 5: 2, 199-207.

Mulligan, P. (2002). Specification of a Capability-based IT Classification Framework, *Information & Management*, Vol.39, pp.647-658.

Niederman F., Brancheau, J.C. and Wetherbe, J.C. (1991) Information systems management issues for the 1990s. *MIS Quarterly*, 17:4, 475-500.

Okoli, Chitu and Pawlowski, S. (2004) The Delphi method as a research too: an example, design considerations and applications, Information & Management, Vol.42 pp.15-29.

Pavlia, P.C. and Wang, P., (1995), An expanded global architecture issue model: an addition of newly industrialised countries, *The Journal of Information Technology Management*, V1: 2, 29-39.

Pavlia, P.C., Pavlia, S, Whitworth, J.E. (2002), Global Information Technology: A Meta Analysis of Key Issues. *Information & Management*, Vol.39, pp.403-414.

Rosenfeld, L. and Morville, P. (1998), *Information architecture for the World Wide Web*. Cambridge: O'Reilly.

Schmidt, R. (1997) Managind Delphi Surveys Using Nonparametric Statistical Techniques, *Decision Sciences*, Vol.28 No.3, pp. 763-774.

Schmidt, R., K., Lyytinen, M. Keil, and P. Cule (2001), Identifying Software Project Risks: An International Delphi Study, *Journal of Management Information Systems*, Vol.17 No.4, pp.5-36.

Turoff, M. and Hiltz, S.R (1996), *Computer Based Delphi Processes*. Available from: http://eies.njit.edu/~turoff/Papers/delphi3.html [17th January 2005]

Watson, R.T. and Brancheau, J.C. (1992), Key Issues in Information Systems Management: An International Perspective. In: R.D. Galliers, (Ed.), *Information Systems Research: Issues, Methods and Practical Guidelines*. Oxford: Blackwell, pp.112-119.

Ziglio, E. (1996), The Delphi Method and its Contribution to Decision-Making. in: Alder M. and Ziglio, E. (Eds.), *Gazing into the Oracle, The Delphi Method and Its Application to Social Policy and Public Health*. London: Jessica Kingsley Publishers, 3-33.