



Project no. 211810

PIREDEU

Providing an Infrastructure for Research on Electoral Democracy in the European Union

Collaborative Project
Capacities Programme

Final Report

Start date of project: 1 February 2008

Duration: 36 months

Organisation name of lead contractor for this deliverable:
European University Institute, Mark Franklin;
University of Exeter, Susan Banducci

| Project co-funded by the European Commission within the 7th Framework Programme (2007-2013) | | |
|---|---|---|
| Dissemination Level | | |
| PU | Public | |
| PP | Restricted to other programme participants (including the Commission Services) | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | X |
| CO | Confidential, only for members of the consortium (including the Commission Services) | |

Contents

| | |
|--|-----------|
| I. EXECUTIVE SUMMARY..... | 3 |
| II. PROJECT DESCRIPTION AND OBJECTIVES..... | 4 |
| II.1 CONCEPT: PROPOSED INFRASTRUCTURE: | 4 |
| II.2 RATIONALE: WHY IS THIS INFRASTRUCTURE NEEDED? | 4 |
| II.3 FEASIBILITY STUDY: ENSURING THE VIABILITY OF THE PROPOSED INFRASTRUCTURE | 5 |
| II.4 RESEARCH USES: SCIENTIFIC IMPACT | 6 |
| <i>II.4.a Research Uses (1): Conducting academic research on elections and voters</i> | <i>6</i> |
| <i>II.4.b Research Uses (2): Assessing the quality of democratic processes in the EU.....</i> | <i>6</i> |
| II.5 BEYOND THE DESIGN STUDY: CREATING A PERMANENT INFRASTRUCTURE..... | 7 |
| II.6 ADDED VALUE OF THE INFRASTRUCTURE | 7 |
| <i>II.6.a Added Value (1): Components of the infrastructure</i> | <i>7</i> |
| <i>II.6.b Added Value (2): Complementary databases.....</i> | <i>7</i> |
| <i>II.6.c Added Value (3): Integration of data regarding National Elections into the Infrastructure.....</i> | <i>8</i> |
| <i>II.6.d Added Value (4): Transforming the infrastructure design into an actual infrastructure</i> | <i>8</i> |
| III. MAIN SCIENTIFIC AND TECHNICAL RESULTS: MAIN FINDINGS REGARDING THE FEASIBILITY OF AN INFRASTRUCTURE..... | 9 |
| III.1 SCIENTIFIC FEASIBILITY OF INFRASTRUCTURE | 9 |
| <i>III.1.a Design Study and Conceptual Map.....</i> | <i>9</i> |
| <i>III.1.b Scientific Validation and Added Value</i> | <i>10</i> |
| <i>III.1.c Contribution of Country Specialists to Scientific Objectives.....</i> | <i>11</i> |
| <i>III.1.d Summary and Evaluation of Scientific Feasibility.....</i> | <i>12</i> |
| III.2 TECHNICAL FEASIBILITY OF INFRASTRUCTURE | 12 |
| <i>III.2.a Data Linking and Technical Design</i> | <i>12</i> |
| <i>III.2.b Cubed data structure.....</i> | <i>14</i> |
| <i>III.2.c Technical Validation</i> | <i>15</i> |
| <i>III.2.d Contribution of Country Specialists to Technical Objectives.....</i> | <i>16</i> |
| <i>III.2.e Ethical Issues: Anonymisation and Protection of Data</i> | <i>17</i> |
| <i>III.2.f Summary and Evaluation of Technical Feasibility.....</i> | <i>17</i> |
| <i>III.2.g Scientific and Technical Results for Stages of the Design Study.....</i> | <i>18</i> |
| III.3 FINANCIAL FEASIBILITY | 34 |
| IV. SCIENTIFIC, TECHNICAL AND SOCIAL IMPACT | 36 |
| IV.1 SCIENTIFIC IMPACT | 36 |
| IV.2 TECHNICAL IMPACT | 38 |
| IV.3 FUTURE POTENTIAL IMPACT: ENSURING THE VIABILITY OF A PERMANENT INFRASTRUCTURE | 41 |
| V. CONTACT DETAILS..... | 42 |

Table of Figures and Graphs

| | |
|--|----|
| TABLE 1: SCIENTIFIC QUALITY AND CORE CONCEPTUAL MAP: DESIGN STUDY | 10 |
| TABLE 2: CANDIDATE STUDY NUMBER OF INTERVIEWS AND DISSIMILARITY COEFFICIENTS FOR TWO RESPONSE SETS* | 28 |
| TABLE 3: VOTER STUDY: MODE, NUMBER OF INTERVIEWS AND RESPONSE RATES BY COUNTRY* | 30 |
| FIGURE 1: THE PIREDEU RELATIONAL DATA BASE..... | 39 |

I. Executive Summary

Elections are one of the primary instruments of democracy and one of the biggest exercises in democracy is the elections to the European Parliament where over 375 million citizens are eligible to participate. These elections offer an unprecedented opportunity to study the functioning of electoral democracy in general and the functioning of European democracy in particular. Scientific evaluations of electoral processes at the EU level have been hampered until now by the lack of co-ordination in the collection of empirical information on which such evaluations are based. Under the auspices of the EU funded project “Providing an Infrastructure for Research on Electoral Democracy in the European Union” (PIREDEU), it has now been found possible to conduct a feasibility study that has produced data of the type that is required.

The main objective of PIREDEU was to test the scientific and technical feasibility of an infrastructure for collecting integrated and linked quality data that allow researchers to address fundamental questions about the representative, accountability and legitimacy functions of electoral processes. As such the feasibility study delivered a pilot study of the administration and coordination required for a centrally conducted research project that covered 27 member-countries at the time of the 2009 European Parliament Elections. In investigating the feasibility of an infrastructure, data were gathered on the attitudes and behaviour of some 27,000 EU citizens, campaign strategies and issues agendas for 1,350 European Parliamentary candidates, issue priorities and positions in 200 party manifestos, campaign news coverage in 140 media outlets and contextual indicators about the political and economic systems of all 27 member countries.

The major findings of the feasibility study (including the pilot data collection) can be grouped according to technical, scientific and financial aspects. Additionally, we comment on findings regarding the need for a permanent infrastructure:

a. Scientific Feasibility – The study has demonstrated its scientific added value by succeeding in bringing together scholars, practitioners and other professionals from across the world to exchange ideas and work with the integrated datasets to promote their research objectives, simultaneously sharing their concerns and visions about the future of this project. At the final user conference in Brussels (November 18-19, 2010), more than one in two (58%) of the authors explicitly linked two or more datasets in the pilot study. Moreover, about one in four attempted a link between three, four or five datasets, whereas a further one out of four authors experimented with linking PIREDEU to external or private data collections.

b. Technical Feasibility – The scientific quality was supported through technical design aspects that allowed the technical and conceptual linking of concepts such as issue saliency and issue preferences. In order to achieve this, the main objectives of the infrastructure were to establish common coding categories for the separate study components in order to facilitate the creation of a cubed data structure that in the future will link these components together, through the use of several different levels of linking variables,

c. Permanent Infrastructure – There is an urgent need for the establishment of an infrastructure for European electoral research endowed with stable funding and capable of compiling, linking, disseminating, and presenting data in a co-ordinated and professional fashion. It is necessary that the permanent infrastructure include both national and European elections. Efforts are already underway with the Consortium for European Electoral Studies (CERES) established at the 2010 PIREDEU final conference.

II. Project Description and Objectives

II.1 Concept: Proposed Infrastructure:

This study aimed at designing an infrastructure for research into citizenship, political participation, and electoral democracy in the European Union (EU). The proposed infrastructure consists of a comprehensive empirical database which will endow our user community with the most essential information required to conduct a regular ‘audit’ that would monitor/ scrutinise all relevant aspects of the electoral process in the European Union. The infrastructure also consists of an organizational network that is able to co-ordinate different data collection activities, so that an integrated database can be created. This database would be so designed as to be accessible not only to academic researchers but also to politicians, political parties, journalists, commercial interests, and even members of civil society.

This project brought together a large network of scholars who had in the past collected data at the time of European Parliament elections on the attitudes, preferences, cognitions and behaviours of the main actors involved in processes of electoral participation: voters, parties, politicians and the media. However, in previous studies no fully integrated data resulted from the separate research projects involved, mainly because of the lack of any formal organisational network to coordinate the activities of the different research projects. Moreover, no facilities exist to provide these data to the research community in an integrated way.

The PIREDEU design has improved on the model provided by the American National Election Study (NES), a permanent infrastructure designed for the study of US elections which has, since 1948, collected and, since the middle 1960s, disseminated to the social science community survey data regarding voter opinions and choices made at the times of all Presidential and most mid-term Congressional elections.

The PIREDEU project has designed an integrated database encompassing not only voter surveys relating to European Parliament elections, but also candidate surveys, media studies, and collections of public record data (including party manifestos) pertaining to the conduct and outcome of the European Parliament elections which are the primary objects of our interest and concern. For members of the academic community, the resulting prototype database has created unprecedented opportunities for cross-national research on electoral representation and behaviour, the role of the media, the emergence and transformation of party systems, and democratisation. It has enhanced the attractiveness of Europe as an object of study and as an environment for comparative social science research. It also holds the potential, for other stakeholders, of opening a window onto processes of electoral democracy that have hitherto remained academic and obscure.

II.2 Rationale: Why is this infrastructure needed?

Elections are crucial instruments of popular control, elite accountability, and political representation; therefore, the quality of democratic governance depends to a large extent on electoral processes. At EU level, democratic rules and procedures are not yet so well established as in most of its member states, and the institutions of multi-level governance undergo frequent reforms. ‘Auditing’ the quality of the electoral process at the EU-level is therefore essential. The conduct of such ‘audits’ at national level already constitute an established practice in non-European countries such as the US, Canada and Australia, and also in some European countries such as the Netherlands, the UK, Italy, Ireland, Sweden and

Norway. Their aim is to assess empirically the processes of electoral democracy, detecting challenges to the quality of these processes.

In order to ‘audit’ the democratic process at European level, the relationships between the behaviour of the three main actors involved need to be investigated: the parties and their candidates for electoral office, the mass media, and the electorate. Due to the ephemeral character of human memory, relevant survey data need to be collected at the time of elections to the European Parliament. The content of news media outlets also needs to be monitored while an election is in progress or the information will be lost. Furthermore, data that is ostensibly part of the public record, relating to the programmatic promises of political parties and to the numerical outcomes of European Parliament elections at the national and regional levels, have in the past proved hard to amass once the election was over.

The data collected as part of an infrastructure of this kind needs to be compiled in such a way as to permit not only linkages between all the elements of the infrastructure but also with other data relating to electoral democracy – data on elite and mass behaviour such as the public record of European Parliament debates, European legislative outputs, and the data collected by other mass and elite surveys at the national and European levels. Procedures we developed for linking the data held in the infrastructure itself do, as far as possible, permit these additional linkages to be made and, above all, permit the infrastructure to be extended to incorporate data collected at the time of future European Parliament elections. As far as possible these procedures also permit the integration of whatever data may exist that was collected at the time of past EP elections.

In addition to procedures for collecting and linking the data, procedures are also needed for making the resulting infrastructure available to the widest possible number of scholars and other users. Indeed, an important purpose of the study was to develop data viewing software that will make access to the data available for those who are not trained social research professionals: especially politicians, journalists, commentators, and even members of civil society; and to those in peripheral regions without easy access to central research installations.

II.3 Feasibility Study: ensuring the viability of the proposed infrastructure

The design of an infrastructure for studying electoral democracy in the EU was accomplished in the context of a pilot study conducted at the time of the European Parliament Elections of 2009. Only by subjecting our proposed procedures to the real-world experience of an actual EP election could we be sure that the procedures were adequate for the task. In particular, the ability to obtain agreement of research teams from 27 countries was a critical test of feasibility which could only be conducted in a context that these teams would take seriously – the context of a real election study.

In order to test the feasibility, we set out to achieve three over-arching aims in the pilot study. First, we aimed to build up an organisational infrastructure. Second, we tested in the context of the European Parliament Elections of 2009 whether this infrastructure was able to organise an integrated data collection of voters, candidates, parties and media. Third, we designed and evaluated a tool to enable easy access by stakeholders to these data. Though the data collected as a test of feasibility for the design related only to 2009, if the infrastructure becomes permanent we hope eventually to be able to incorporate the data already collected at previous EP elections.

The design and feasibility studies were thus conducted in tandem, by means of a number of Work Packages (WPs) through the following stages:

1. consulting the social science research community about the research areas to be addressed;
2. developing scientific and technical guidelines for an integrated data collection effort;
3. conducting in all 27 countries a voter survey, a candidate survey, a content analysis of the news, a content analysis of party manifestos, and a collection of the most important country statistics;
4. validating the quality of these data;
5. designing a cubed data structure that will enable the full integration of these various data sources, and which will allow end users to easily access these data.
6. The feasibility study also provided critical evidence required for a full scientific and technical evaluation of the proposed infrastructure to audit European democracy.

Each stage gave rise to a number of reports, referred to as “deliverables” (referred to as D_n where n is a number), though this report will only refer to some of these (and not always by number).

The feasibility study had two main “products”, additional to the design itself. First, it produced four data sets deriving from: a candidate survey, a voter survey, a content analysis of party manifestos, and a content analysis of the news. All of these data sets contained contextual data (D9.4). These data have been made available to the social science community who have been able to conduct a preliminary audit of the functioning of democracy in the EU. Secondly, a cubed data structure will be designed in which to embed the different data sets, making them accessible to end users (D9.3).

II.4 Research Uses: Scientific Impact

II.4.a Research Uses (1): Conducting academic research on elections and voters

Though pre-existing data have made possible many advances, the proposed infrastructure would go much further, enabling the academic research community to explore interrelationships between the behaviour of parties, politicians, voters and the mass media at national and European levels in an integrated fashion, increasing the attractiveness of Europe as a topic for research and a venue for research activity. Within the political science sub-field of electoral studies and comparative politics, many lines of research could be investigated on the basis of the proposed infrastructure.

II.4.b Research Uses (2): Assessing the quality of democratic processes in the EU

The proposed infrastructure, beyond its innumerable applications in terms of pure science, will have one overriding policy-oriented application in terms of the future of European integration. It will provide the means (in terms of data linkage and data viewing procedures) for evaluating more thoroughly than was previously possible different diagnoses and proposed remedies for European democratic ills. Indeed, provision of ways to represent the contents of the infrastructure in an easily-digestible fashion will make the relevant data accessible beyond the scholarly community to the politicians and commentators most immediately concerned with the political problems of European governance and even to members of the broader European public. Proponents of different approaches to understanding the European malaise would be able to explore observable implications of their diagnoses, thus unearthing empirical and logical connections and anomalies that may lead to more connections between different ‘schools of thought’, rendering each of them less apodictical.

II.5 Beyond The Design Study: Creating a permanent infrastructure

There is an urgent need for the establishment of an infrastructure for European electoral research endowed with stable funding and capable of compiling, linking, disseminating, and presenting data in a co-ordinated and professional fashion. Our project has designed a new infrastructure encompassing the different types of data necessary for investigating and describing the state of electoral democracy in the European Union. The prototype for this infrastructure already contains data on the attitudes of voters, the behaviour of political parties and their candidates, the outcomes of elections, and the contents of mass media reports. These diverse data sets need to be presented to the public in a way that is accessible not only to the academic community, but also to other stakeholders, such as journalists, policy-makers, and members of civil society.

Even while this design study was in progress, efforts were made by way of national representatives on the European Strategy Forum on Research Infrastructures to have PIREDEU listed on the ‘roadmap’ and in due course declared a mature infrastructure. Such a permanent infrastructure will constitute a data repository for social scientists who will employ it to monitor national and European parliamentary elections. It will be continuously updated with data collected at the time of future European Parliament (and perhaps national) elections. This repository will fulfill long-term strategic needs of stakeholders, permit continuing research into the nature and evolution of electoral democracy in Europe and provide regular ‘audits’ of the adequacy of representation processes for ensuring accountability of European policymakers. This will legitimate public policies, and enhance public understanding of European political processes. Once such an infrastructure exists the necessary resources for updating and extending its component data collections should be straightforward to obtain from national, private, and/or EU sources.

II.6 Added Value of the Infrastructure

II.6.a Added Value (1): Components of the infrastructure

The added value of the proposed infrastructure is that it is specifically designed to permit access to data in EP elections in an integrated fashion, thus complementing a number of existing cross-country comparative data collections, none of which is well-suited to audit the behaviour of the main actors involved in European elections. It is innovative in that it will provide an integrated database where data on attitudes and behaviour of voters are linked to data about parties and their candidates, the media reporting that actors are exposed to, and the political and economic context in which all these actors operate. Linking and integrating data in this way is highly complex: no cross-country comparative study has yet created an integrated dataset of this magnitude even for a single election – much less for a series of elections. The design study has investigated the scientific and technical feasibility of building an integrated cross-national database on European electorates, elites and the relevant public records, and has tested this design by means of a pilot study conducted in the context of the 2009 European Parliament elections.

II.6.b Added Value (2): Complementary databases

Our infrastructure will complement three sorts of related databases. These include firstly a number of comparative cross-national surveys focusing on the attitudes and behaviours of citizens: the European Social Survey (ESS), the European Values Studies (EVS), the World Values Studies (WVS), and the International Social Survey Project (ISSP). While these surveys, and the data collections that have resulted, are of great value, none of them are either

conducted in the context of elections, nor are they designed for the study of elections. The database we have designed is specifically tailored for electoral research, and the timing of the data collection activities need to be scheduled accordingly.

II.6.c Added Value (3): Integration of data regarding National Elections into the Infrastructure

The case for studying European democracy in simultaneously national and cross-national terms is unassailable. European Parliament elections occur in the context of democratic elections to national parliaments throughout Europe, and European Governance is a multi-level enterprise involving institutions at national as well as supra-national levels. We have investigated the scientific need for such an integrated infrastructure and report on it in the following section regarding our main findings about the feasibility of the infrastructure. We have also, as individual scholars and not as representatives of the PIREDEU project, taken steps to establish a permanent association of those who study elections in Europe (both national elections and elections to the European Parliament), namely the “Consortium for European Research with Election Studies” (CERES), to serve as a point of contact for scholars interested in seeing the PIREDEU design implemented. This consortium (a Consortium for European Research with Election Studies – CERES) will probably never become an infrastructure and is not intended to serve that role. Rather it is intended to be an association of election studies specialists many of whom would almost by definition have to become involved in any infrastructure that was established. In the meantime, and even after the establishment of such an infrastructure, the consortium will rather play the role of a professional association representing the interests of scholars in this field. And one of the roles of such an association will be to lobby for and attempt to create the preconditions for the establishment of an appropriate infrastructure. The consortium (which is even now in existence) has additional objectives as well, but those go beyond the remit of this report.

II.6.d Added Value (4): Transforming the infrastructure design into an actual infrastructure

The EU is in the process of adopting new procedures for the establishment of EU-wide infrastructures – procedures that no longer provide an obvious route by which even a successful design study can be transformed into a permanent infrastructure. The remainder of this Report assumes that some means can be found for bringing such an infrastructure into being.

III. Main Scientific and Technical Results: Main Findings Regarding the Feasibility of an Infrastructure

III.1 Scientific Feasibility of Infrastructure

Given that the project is a design and feasibility study, our objective has been to assess the scientific benefits of an infrastructure for studying electoral democracy in Europe. Scientific feasibility is assessed in this instance by examining the implementation and added scientific value of linking core scientific concepts in the study of electoral democracy across the 5 study components.

III.1.a Design Study and Conceptual Map

In order to create an integrated infrastructure, it was necessary to develop and then implement a conceptual map that allowed a comparison of key concepts of the study of electoral democracy to be measured across the different data components (party manifestoes, media content, voter survey, candidate survey and contextual data). This core conceptual map was based on a review of components of past election studies conducted since the first European Election Study in 1979 and permitted the establishment of measurement instruments in each component that would be as close as possible to those in each of the other components. The conceptual map also provided the basis for assessing the scientific quality of the design study. The concepts it included are key to the study of electoral democracy and so the scientific quality of the measured concepts across the study components had to be assessed on that basis.

Table 1 contains the core concepts that were included in the 2009 PIREDEU Design Study, (as agreed by the PIREDEU Steering Committee, meeting on 19-21 June, 2008, in Amsterdam). This list only contains the concepts that were to be included across as many as possible of the five data collection instruments. It is not a complete list of all items to be included in each of the questionnaires/coding schemes.

Table 1: Scientific Quality and Core Conceptual Map: Design Study

| Concept | Voter | Media | Candidate | Manifesto | Context |
|---|-----------|-----------|-----------|-----------|-----------|
| 1. Voting: Party choice & turnout | Available | Available | Available | Available | Available |
| 2. Party ID | Available | Available | Available | Available | Available |
| 3. Engagement and mobilization | Available | Available | Available | NA | Available |
| 4. Media usage | Available | Available | Available | NA | Available |
| 5. Institutions | Available | Available | Available | Available | Available |
| 6. EU integration | Available | Available | Available | Available | NA |
| 7. Value orientations | Available | Available | Available | Available | Available |
| 8. Domestic and European issues | Available | Available | Available | Available | NA |
| 9. Representation | Available | Available | Available | Available | Available |
| 10. Identity | Available | Available | Available | Available | NA |
| 11. Demographics | Available | Available | Available | Available | Available |
| 12. Knowledge and experience | Available | Available | Available | NA | Available |
| 13. Recruitment and Nomination | NA | Available | NA | NA | Available |
| 14. Attribution of responsibility and evaluation of performance | Available | Available | Available | Available | Available |

Note: This table is based on “Table 1: Linkages and Core Concepts in Data Components” in the D8.3 “Report on Data Suitability”.

The questionnaires and codebooks used in the data collection efforts were rendered as comparable as possible by being based on this single conceptual map (Deliverable 8.1) of the topics to be investigated along with a single set of codes (D9.1) for recording the same information collected for different data components. These procedures were subjected to both scientific (Work Package 8) and technical (Work Package 9) evaluations. To the extent that they were successful, the data collected as part of the infrastructure will have been essentially “pre-linked” across the different components, though the coding was also intended to facilitate post-linking with other data relating to electoral democracy. Moreover, the infrastructure has been designed so as to permit integration, wherever possible, with data from previous and future EP elections.

III.1.b Scientific Validation and Added Value

The scientific added value of the proposed infrastructure is twofold: it concerns, first, the facilitation of academic research on elections and voters; and second, most importantly, it is intended to be an effective tool for the assessment of the quality of the democratic processes in the European Union. Scientific feasibility is also demonstrated through the added scientific value of the infrastructure. The added value was assessed in D9.5. The purpose of that report

was to provide an overall scientific evaluation of the database for the social scientific communities and the non-academic policy user communities as identified early in the PIREDEU project and summarized above. To that extent, this study has demonstrated its scientific added value by succeeding in bringing together scholars, practitioners and other professionals from across the world to exchange ideas and work with the integrated datasets to promote their research objectives, simultaneously sharing their concerns and visions about the future of this project.

The added value that was visualized at the commencement of this project was to overcome limitations in the availability, quality and utility of earlier data and at the same time provide access to new data on EP elections in a novel integrated fashion. The innovations employed in this study at all stages, from the open consultation with our user communities to the thoroughness and clarity of the conceptual map and from the integration of different data collections to the facilities for adding additional databases, lie at the core of the scientific added value that this study offers not only to its users but to advancement of our understanding of electoral processes and of the quality of democracy in Europe. But even more importantly, added value is inherent in the potential new avenues for research that this study opened.

Since of our main scientific objectives were to provide researchers with the ability to link different datasets with each other, we were interested in finding out how these were linked. We reviewed the papers presented at the final user conference in Brussels (November 18-19, 2010) and found that more than one in two (58%) of the authors explicitly linked two or more datasets. Moreover, about one in four attempted a link between three, four or five datasets, whereas a further one out of four authors experimented with linking PIREDEU to external or private data collections.

This last statistic is particularly important in pointing to the need to see an infrastructure for research on electoral democracy in the European Union as calling for more than just data about European Parliament elections. If a quarter of all papers found it necessary to go beyond the data provided by the PIREDEU feasibility study, this points to a readily understandable limitation in any infrastructure that focuses only on elections to the European Parliament. Democracy in Europe is not only or even mainly about European Parliament elections. It is about all elections at whatever level. The original PIREDEU proposal did contain components that would have constituted pilot projects for linking EP data with data from subsequent national elections in certain member countries, but these components were eliminated for lack of funding. A permanent infrastructure needs to be seen as involving national as well as European Parliament elections, though any such idea involves funding problems that will be addressed in the final section of this report.

III.1.c Contribution of Country Specialists to Scientific Objectives

Though the PIREDEU consortium consisted of fifteen institutions located in nine countries of the EU, and some of these institutions (particularly the European University Institute) themselves housed researchers from all EU member states, still it was deemed important to involve electoral research specialists who resided in each of the countries to be surveyed. These were recruited in the early months of the project, based on lists of researchers who had collaborated on the design and fielding of previous European Election Studies. Country specialists were involved at all stages of the design and feasibility study, checking the conceptual map for country-specificities, checking the translation of the questionnaires, acquiring lists of candidates and their addresses, and overseeing the actual fielding of

questionnaires. These specialists gave their time free of charge for the sake of the public good and were central to the quality control procedures employed in PIREDEU.

The data collection teams asked the country experts to provide a vast range of information and to fulfill highly relevant tasks. The list of tasks was revised and supplemented continuously in reaction to our experience during the fieldwork. In relation to the scientific quality of the data, country experts were asked to assist with the following:

Identifying relevant parties and candidates

Country teams were asked to identify relevant parties and candidates in accordance with the rules established across the data teams. Their decisions were to be based on former elections, recent polls and expertise on the national political context.

Checking the questionnaires

The country teams were provided with the language versions of questionnaires (for both candidate and voter surveys) that had been prepared for fielding in their countries. They were asked to check these documents carefully and to return them with corrections if necessary. Because of missing or poor translations of certain questions, some country teams agreed to do supplementary translation work as well.

Translation of cover letters and reminders

The Candidate Survey Team provided a cover letter and a reminder in English; the Voter Survey Team provided an introductory letter in English. Country teams were asked to translate those letters.

Media Outlets

Country experts were asked to validate the selection of media outlets for the media study, keeping in mind backward compatibility but also taking account of any major changes in media systems.

III.1.d Summary and Evaluation of Scientific Feasibility

The question whether it would prove possible to involve different teams of scholars and different country experts across all EU countries in the process of designing and implementing a full-fledged study of an election to the European Parliament was a primary research question for PIREDEU. The question was answered in the affirmative, providing one critical proof of feasibility.

III.2 Technical Feasibility of Infrastructure

III.2.a Data Linking and Technical Design

Together with the scientific guidelines (see core conceptual map, cf. joint deliverable D8.1 and D9.1) all work packages received technical guidelines for data collection which were agreed upon during the PIREDEU steering committee meetings. The guidelines concerned issues of sampling (especially for the voter and candidate surveys), tendering, questionnaire wording and instruments (e.g. measuring variables like education) and translation. On this basis work package teams submitted their data collection instruments: questionnaires for the voter and candidate studies, codebooks for the media and manifesto studies and a report on data collection from the contextual study.

It is worth noting here some issues that arose regarding the translation of the candidate and voter survey questionnaire. Decisions regarding translation procedures were based on

employing best practices, maintaining compatibility with previous translations, efficiency and cost. Where questions employed in previous surveys had already been translated, these translations were employed. Existing translations were either from previous EES Voter surveys, or from Eurobarometer or similar cross-national surveys (such as the ESS, or the World Values Surveys). Since there was no dedicated budget for translation, this was done ‘in house’, coordinated by the EUI. Again, best practices were followed with translations being discussed by two translators and checked by our country specialists. Other instruments (Manifesto Study, Media Study and Candidate Survey) were not translated as data collection took place in a common language. The codebooks for these studies were developed in English and coders worked on the material in the target language (party manifestoes or news media content) while using the English codebook. All the above are summarized in deliverables D9.1 and D9.2 along with individual Work Packages’ early deliverables.

As mentioned in previous deliverables, specific instructions were given by Work Package 9 regarding post-processing of the collected data in order to insure that the first advance release would contain data that would be as close to the final release as possible (in terms of data cleaning, additional variables, etc.).

However one of the main objectives of the infrastructure was to establish common coding categories for the separate study components in order to facilitate the creation of a cubed data structure that in the future will link together, through the use of several different levels of linking variables, the candidate, voter, manifesto, media and context data collection efforts. Where appropriate we employed standardised, universal codes that are recognised across sectors and disciplines. These variables were¹:

Location - Country

The main location variable employed across studies is country for which the ISO code is used. This is a three digit code that also formed part of the codes of other shared variables.

Media Outlets

Media outlets are found in the candidate, voter and media study. The studies are linked by questions about media use in the voter survey and by campaign media use in the candidate survey. Common codes were provided for use in each of these studies by adding a counter to the standard country identifier.

Political Parties

Across all studies, political parties form categories for several variables. A unique seven-digit code was used to identify political parties in the various study components. This code replaced the raw data entries, and facilitated the integration and linking of the various components. It is comprised of the country identifier, the party ideological family and a two-digit counter.

Candidate identifiers

Linkage with media study would be enhanced if candidate identifiers in the media study could be matched with those in the candidate survey. This involves data protection issues that are currently under review as the data is readied for archiving and release by GESIS.

¹ Please see deliverable D9.4 for a detailed description of the common coding across the different components.

Political issues

Components of the study regarding issues are linked according to the most important issue questions. These involved open-ended questions in the voter and candidate surveys which were coded using the codes from the Media Study topic list (which in turn had been matched to the themes covered by the manifesto study).

III.2.b Cubed data structure

The variables listed in III.2.a represent the ‘key variables’ for which linking across datasets can be expedited using the interface that has been designed for the cubed data structure. As mentioned in previous deliverable (see deliverable D9.3) “the process of linking data will be facilitated by a user-interface that guides the analyst through a structured series of questions that define the particular kind of linking desired. After all questions have been answered, the interface will generate the required syntax for relational database management software and execute this syntax in order to produce the desired dataset. This will take the form of a rectangular data matrix, suitable for analysis by statistical software commonly used in social science research”. Since the advance release of the data, this user interface has been under development. As stated above, in order to achieve this objective, special care was given during post-processing efforts to ensure that the variables from the different components have been coded according to identical protocols and guidelines in each component. The process of linking that is now available to users has been developed with the aid of a number of validation (training) sessions. We refer here to the sessions during the Essex Summer School in Social Science Data Analysis and those that took place during the PIREDEU final conference.

In the first round of validation sessions (July 2010) the linkage was achieved by a lengthy process of recoding and integrating individual datasets (the example that was used was ‘dropping’ media study data into the voter survey). This, however, is not intended to represent the final state of the infrastructure, since this method of data integration does not possess the added value that PIREDEU aspires to. Instead, the intention is for a more automated ‘user driven’ process like the one described in passing above. Such a procedure was available during the second round of validation sessions. Here we have a procedure that uses macros in order to retrieve, reshape and integrate the data. This process is described in van der Eijk and Sapir (2010).² As the authors mention, “The linking application, which is named P-DLc consists of pre-programmed syntaxes (including embedded macros) for data management that can be adapted to provide the end-user full capacity to specify his/her needs. The syntax is defined for a software environment that needs to be available on the computer of the end-user. The application can thus be specified for any statpack that provides syntax and macro functionality, such as R, Statistica, Stata, SPSS, SAS, Minitab, JMP, Systat, S-Plus, R-Plus, BMDP, and Revolution R. Because most of these packages are able to import and export each other’s file formats, there is no compelling need to produce separate versions of the linking application for each individual software environment”. During these sessions users of the data had the opportunity to be introduced to the rationale behind these macros but they were also given the opportunity for hands on experience with using them through lab computer sessions. The new procedure was positively received by the participants.

² The paper “Linking Electoral Data about Citizens, Political Parties, Mass Media and Countries - A General Approach with a User Application for the 2009 European Election Studies” can be downloaded from http://www.piredeu.eu/Database/Conf_Papers/I3_1-Linking_Electoral_Data-PIREDEU-17NOV2010F.pdf

III.2.c Technical Validation

Validation sessions on the linkability of the PIREDEU datasets were organized at the final user conference by Cees van der Eijk and Eliyahu Sapir of the University of Nottingham. All sessions were attended by Theofanis Exadaktylos (Exeter) for the purpose of reporting any substantive comments relating to issues that needed to be addressed before a final release. There were two types of training sessions on offer at the conference, one addressed to non-academic users and a set of repeat sessions addressing academic users.

Training Session for non-academic users

The training session was cancelled due to the server in Nottingham being down. The stability of the web location of the application should be of great priority. Cees van der Eijk suggested the alternative use of Webinars for this training session, which is generally a good idea for future training too. It was also suggested to organize smaller training sessions that would be led by the respective university teams in their home countries for local scholars and non-academic users. Despite the technical difficulties that led to the cancellation of this session, the data are being accessed by non-academic users.

Training Sessions for academic users

The purpose of the training sessions for academic users was to demonstrate how to capitalize on the potential added value of the PIREDEU datasets. Linked data has the potential to overcome problems of omitted variable bias that arise when information from each single dataset is used on its own. Moreover, by acquiring more explanatory variables from different datasets we can enhance a model's strength. This is what the PIREDEU datasets allow us to do as academic users.

The critical idea here is data-linking. This comes with the prerequisite of users having an understanding of which variables and data can be connected with what – something for which a certain amount of training is required. Training is also required, as already mentioned, in order to employ the data-linking tools provided.

The PIREDEU feasibility study has produced datasets incorporating a considerable amount of pre-linking performed during the data collection and data coding phases. However, this has not exhausted the linkage possibilities. In fact, the database is intended as a tool that can be tailored to user-specific needs, limited only by the availability of the original datasets and the users' creativity. The PIREDEU data-linking code works by adapting an SPSS syntax file with embedded macros. It brings datasets together in a two-way fashion, with the user being able to determine the relationship between donor and recipient datasets (which can also be thought of as foreign key and primary key datasets). Information can be added in a mutual way to each data set, and there are upwards and downwards linking. All of this material was covered in the training sessions, which were followed by a Question and Answer period that consisted primarily of technical questions, for example differences between using STATA and SPSS, use of spreadsheets for linking, etc. Macros and SPSS code are included in deliverable 9.3.

The software demonstrated at these sessions will be available as a suite of applications for download as freeware from the PIREDEU website and the University of Nottingham web pages, but functionality has to be finalized before completing the final user interface. The data will be available free of charge from both web locations and also from the GESIS archive, and efforts are currently under way to finalize a web-based interface that removes the need to edit the SPSS syntax files. The objective is to make the web version as simple to use as possible.

On the question of flexibility in combining different data sources, the organizers suggested that there will be some level of pre-structuring of the possibilities, but they stressed that the point of the application would be to provide universal linking of all PIREDEU datasets. The most important feature that they foresee is that the user will be able to eventually link their own datasets and add new variables according to their needs. There will be some restriction of user functionality on the web version, but total functionality will be available on a stand-alone version. This dual solution addresses the trade-off between simplicity (with restricted functionality) and universality (bringing the need to run the application locally). Documentation is will be available to accompany the software.

III.2.d Contribution of Country Specialists to Technical Objectives

As detailed above in the section on scientific feasibility, country collaborators (specialists) played a role in ensuring the technical feasibility of the infrastructure design during the pilot study. In terms of the technical quality of the data, country specialist were tasked with the following:

Brief report about the central features of the election campaign

A brief report, which covers the major topics/issues in the campaign, differences in the campaign strategies of parties/candidates, and other major events/characteristics relevant to the European Parliament Elections should help to put the voter, media and candidate studies from each country into context. It also provides a means of validating the picture of the campaign obtained from the quantitative studies.

Evaluation of mail service

To evaluate the speed of the mail service used, country experts were asked to send back a letter to the Candidate Survey Team with a prepaid envelope similar to the one provided to the candidates. This was provided as a means of testing the feasibility of the postal method for the candidate survey.

Collection of candidate information

Country specialists collected names of candidates for the candidate survey. Due to severe problems in several countries, country teams were asked to write a brief report about the collection of candidate information. This report was to include sources and any problems they were confronted with.

Increasing response rates

Country teams were informed about the respective return rates for the candidate survey on a country and party base. If necessary, they were asked to suggest strategies to increase the returns and to get into contact with certain political parties whose candidates showed a low willingness to participate.

Capturing media content

When not available through other means, country specialists were asked to capture media content for the media study thereby reducing missing data.

Checking question wording and translations for the voters' study

Country specialists played an important role in helping the voters' study with such matters as determining salient issues for their country, coding the educational system, and checking translations of the voters study questionnaire into their language(s).

III.2.e Ethical Issues: Anonymisation and Protection of Data

The pilot study for this project does, and the infrastructure if funded will, contain political data. Members of the mass public have been and will be asked for their opinions on various political issues and about their actual voting behaviour at European Parliament and national elections. However, the identities of the individuals concerned will not be stored in the infrastructure database but only on paper in a locked filing cabinet or on a password-protected computer unconnected to the Internet. That file, in turn, contains no political data, just the identification numbers of the individual respondents, as recorded in the infrastructure database. The identities of these individuals are only needed for quality control and data linkage purposes. They will never be released. And once the data have been validated and successfully linked, information about the identities concerned will be destroyed.

Members of the public will of course be asked whether they are willing to give up some of their time to be interviewed and will not be pressed if they are unwilling. Any respondent is also free to break off the interview at any time, with or without making an appointment for a follow-up session.

The pilot study does, and the infrastructure if funded will, contains data about political opinions of candidates for elected office. However, since we will only interview publicly declared candidates about the political opinions that they would report to a constituent if asked, no question of privacy arises. Candidates for public office lose certain expectations to privacy when they declare their candidacies. Our respondents were told that they should behave as though they were being interviewed by a constituent and that they should not tell us anything they would not want a constituent to know. Nevertheless, EU law requires certain types of protection even for such data, as explained below.

The names of candidates interviewed in connection with the pilot study (and with other components of the infrastructure) do need to be kept in the database since their names provide the only means for linking data collected about them with data regarding their activities (votes, speeches) collected then or later for inclusion in the infrastructure. Such data might be collected years, even decades, after the interviews themselves.

As part of Workpackage 9 on Technical Feasibility, a protocol has been designed to secure anonymity of respondents of the candidate survey and to safeguard their confidentiality. Data archives have standard protocols dealing with these matters and our design study has made use of their expertise. Researchers employing the candidate study data lodged in the GESIS archive are required to agree to maintain the anonymity of candidates and to abide by various precautions in this regard as required by EU and German law.

Other components of the infrastructure contain data that was always intended to be public (party manifesto data, data on the outcomes of elections, and other information that is publicly available, at least at the time of the election concerned). The Scientific Advisory Board had a special role in ensuring that the project kept to these ethical standards.

III.2.f Summary and Evaluation of Technical Feasibility

To promote the technical feasibility of the infrastructure, the infrastructure established: (a) the procedures to ensure quality of data collection and ultimate linkage, (b) the development of the cubed data structure, and (c) the procedures for linking PIREDEU datasets to past and future European Election Studies (EES) data as well as other established social science data collection efforts, outside the EES (e.g. the European Social Survey). The PIREDEU researchers with the assistance of country specialists and the user communities established the

practices that would ultimately lead to a successful data integration and/or took advantage of these data to conduct important research on the quality of electoral democracy in Europe.

III.2.g Scientific and Technical Results for Stages of the Design Study

The feasibility study delivered three products: (i) the infrastructure design, (ii) various datasets, and (iii) a ‘proof of concept’ for the cubed data structure. These were produced in six stages that are detailed in this Section of the Report. In summary, these were concerned with (a) creating the framework, standard operating procedures, and mechanisms necessary for a permanent infrastructure on the study of electoral democracy in the European Union; (b) planning and (c) implementing the collection of four datasets by four Work Packages: a candidate survey, a voter survey, a content analysis of party manifestos, and a content analysis of the news. All of these data sets contained contextual data which was itself collected as a separate Work Package. These data, in preliminary versions, were then (d) extensively validated in house and made available to the social science community for external validation through their use in conducting a preliminary audit of the functioning of democracy in the European Union. Then (e) a cubed data structure was designed in which to embed the different data sets, making them accessible to end users. Finally (f) the outcome of all of these procedures were evaluated in the context of a Final User Conference which provided external validation needed for writing the present Report.

III.2.g.1 Stage 1: Consulting the social science research community

The first stage involved consulting the social science research community about the research areas to be addressed. In this first stage of the design study a kick-off conference was organized and a project website was established as means for consulting the social science research community regarding scientific guidelines as these were developed.

The kick-off conference

The European University Institute in Florence hosted the kick-off meeting and the First User Community Conference of the PIREDEU Design Study from 20 to 23 February 2008. Almost 50 participants, either PIREDEU Consortium partners or representatives from the PIREDEU User Community, took part in the event. This first conference with the user community was used as a unique vehicle to seek a list of projects that would employ the infrastructure both in its initial and final release and to gather feedback on user expectations regarding technical aspects of the design study. The Kick-off conference was designed to serve two purposes: First, the Consortium partners met for the first time in order to discuss concrete details of the design study, and to assign responsibilities for the work in each of the different study teams and for producing the study output. Second, the PIREDEU user community was invited to propose areas of research to address with the data that would be collected in the course of the feasibility study. The kick-off conference was also instrumental in recruiting country collaborators who would then serve as country experts answering country-specific questions (for example, about political parties and media outlets), checking sampling frames and fieldwork plans, and verifying questionnaire translations.

The user community is indeed a large one. It comprises an academic component and a wider set of users outside academia. On the academic strand it includes political scientists, communication scientists, political sociologists, social psychologists and political economists. It has the potential of including at least one thousand academics from around Europe and the United States. Outside academia, the infrastructure can be relevant to all those institutions and individuals who have a professional interest in elections and electoral processes. A first group relates to political parties: elected office holders, party officials, campaigners, research

institutes and think tanks. A second group concerns social groups and organized interest-groups with stakes in the elections, such as labor unions, employer organizations, churches, formal lobbies and others. A third group includes media organizations and journalists providing information on the elections. Finally, a fourth group includes private enterprises that serve all the above groups, such as market research companies, media and campaign specialists, and consultancy firms amongst others.

‘Auditing’ the quality of the electoral process at the EU level is essential to all the users listed above, as they assess the nature of the electoral process and detect challenges and threats to the quality of these processes involving parties and candidates, mass media and voters. An additional feature of the infrastructure is the collection of data which, if not collected at the time of an election, may be lost or recorded in a way that does not abide by the standards set for the collection of other relevant data, constraining the possibilities for future research.

Given the diversity of our user communities it was important to maintain an open forum throughout the conduct of the design study to promote improvement in the original research design. In the kick-off conference that was held in Florence in February 2008 (cf. deliverables D2.1 and D8.2), we consulted these research communities, absorbing ideas about their specific needs and preferences to better design the data collection instruments that would be used in the feasibility study. The First User Conference of the PIREDEU Design Study took place on the second day of the meeting. About 30 users were invited to comment on the presentations of the Work Package leaders and to provide initial recommendations on the design of the study and, more concretely, on the different data collection components.

The consultations begun at the kick-off conference continued through the use of an “open forum” on the PIREDEU project web site (see below for details) which permitted users to suggest additions and changes to the PIREDEU design as this was developed over the course of the following 15 months leading up to the European Parliament elections at which the feasibility of the design was put to the test.

A similar motivation served as a rationale for consulting the user communities in the final months of the project regarding plans for building and extending the infrastructure itself. This included consultation at the final conference in Brussels (November, 2010). Consultation continues on PIREDEU through the open forum. This procedure is still available on the PIREDEU web-site (www.piredeu.eu/public/Open_Forum.asp) and comments that, collected into over one hundred threads, are the living proof of the successful open involvement of our user communities in the realization of this project. PIREDEU has become an essential database for all those interested in electoral democracy in Europe, going far beyond the social scientists who are engaged in comparative and evaluative research on the European electoral process. As a proto-infrastructure it is already extremely useful to a broad community encompassing academics from a range of disciplines, as well as political parties, the media, and civil society.

For members of the academic community the data provide unprecedented opportunities for transnational research on electoral representation and behavior, the role of the media, the emergence and transformation of party systems, and democratization. they increase the attractiveness of Europe as an object of study and as an environment for comparative political science research. For other stakeholders and non-academic users this database opens a window into processes of electoral democracy that have remained academically esoteric and obscure.

While many country collaborators were long established members of the European Election Study, having been involved in at least one previous study, the kick-off conference served a

second purpose in fostering a willingness among these scholars (as well as among collaborators from new member states) to serve as country specialists in the roles described earlier in this report. It fostered a willingness among those with sometimes long-established research agendas to make compromises for the good of the project as a whole – compromises that sometimes could have the effect of changing or slowing these individual agendas. Above all it was necessary to establish a willingness among all collaborators to give up certain cherished questions and question wordings in the interests of producing pre-linked components for later inclusion in a common data structure. “Selling” this idea should be easier in future work because the payoff has now been demonstrated and its value appreciated. Moreover, a widespread interest in backward compatibility as a means of ensuring that subsequent data collection enterprises can be analyzed in conjunction with PIREDEU data, should further facilitate this objective. Still, it should never be taken for granted that an interest in the common good will trump individual research interests, meaning that any future infrastructure needs to give priority to maintaining the view that PIREDEU promoted during its two conferences, its open forum, and elsewhere, of the enterprise being one that was centered on the common good with payoffs that would make compromise worthwhile.

The website

The domain name “www.piredeu.eu” was obtained to provide a name for the project website which became operational at the end of March 2008. Even before that, from February 2008 a presentation of the project was available at the website of the Robert Schuman Centre for Advanced Studies at the European University Institute. The PIREDEU website offered topical information concerning the project. In particular, it presented reports and articles by members of the PIREDEU Consortium. In addition, links to past European Election Studies and other information sources relating to the project's research fields were provided, as well as regularly updated information on events organised in the framework of PIREDEU.

The website contains a public and an intranet section. The Intranet Section contained information and material relevant to the fulfillment of the contractual obligations or relate to the management of the PIREDEU Consortium. In addition, in the intranet section partners could access non-public deliverables, i.e. those that are marked as internal or restricted.

As already mentioned, a key element of the PIREDEU pilot study was consultation with the wider research community, through an open procedure, that allowed us to take the research needs and preferences of the user community into account when designing the data collection instruments.

The PIREDEU Open Forum was the cornerstone of this consultation procedure. The Open Forum welcomed anyone who wanted to make constructive contributions to the design of any aspect of the PIREDEU data collection instruments. After registering, members could write proposals advocating the inclusion of new questions in the questionnaires or new coding categories for those and other data collection instruments. Members were also eligible to post comments about each proposal, and proposal authors could update or revise their proposals in response to advice received. The goal of the Open Forum was to improve the quality and scientific value of each of our data collections, to encourage the submission of new ideas, and to make such experiences more beneficial to individual scholars. We welcomed proposals and comments on each of the five data collection components of the PIREDEU study. The Open Forum also included a General Forum where registered users could make comment about general aspects of the PIREDEU infrastructure and make suggestions that related to more than one of the data components.

The Open Forum allowed PIREDEU decision-making to be open, transparent, and constructive. This Open Forum aimed to constitute a forum for scholars to offer each other constructive feedback, a vehicle for improving and amending proposals, and an opportunity to assess the extent to which a particular proposal is supported by previous empirical work, current theory, and/or a sizeable scholarly community. We warmly invited everyone who is interested in the PIREDEU study, and the individual data collection projects, to take part in this process.

The Open Forum Consultation Phase generated 68 posts in 55 threads across the 5 main forums. It generated 35 proposals, distributed as follows:

1. Voter Survey (VS): 28
2. Candidate Survey (CS): 5
3. EuroManifesto Study (EM): 0
4. Media Study (MS): 1
5. Contextual Data (CD): 3

The Open Forum provided essential feedback that helped in finalizing the Core Conceptual Map mentioned earlier in this report. The Steering Committee met in June 2008 to evaluate each of the proposals posted on the Open Forum for inclusion in and/or amendment of the pilot study. The Steering Committee agreed to first deal with the list of recommendations and the respective concepts, before turning to the individual proposals submitted to the Open Forum. Taking each proposal in turn, the Steering Committee considered it in the light of the Core Conceptual Map, determining whether the proposal would ensure over time and across data instrument comparability. Moreover, given that the voter and candidate surveys could only contain a limited number of question items, priority was given to proposals with succinct question formats. After redesigning the data collection instruments in the light of user proposals there was a second period of consultation where draft instruments (surveys and coding schemes) were made available to the wider user communities and specific comments sought.

III.2.g.2 Stage 2: developing scientific and technical guidelines for an integrated data collection

As the consultation phase required development of instruments and comments on these instruments, the development of the scientific and technical guidelines happened in tandem with the consultation process. In one report the scientific steering committee provided the different teams with clear guidelines on the research topics that had to be addressed in the various data collection activities. In another report, the data committee provided the different teams with clear technical guidelines for organising an integrated data collection. This milestone formed the basis for the further (also technical) activities of the teams, such as the design of questionnaires and coding schemes for content analyses, using the same lists of parties and media outlets in the different studies, and giving each them unique identifiers.

III.2.g.3 Stage 3: Fieldwork preparation and data collection

In the third stage the objective was to conduct in all 27 countries a voter survey, a candidate survey, a content analysis of the news, a content analysis of party manifestos, and a collection of the most important electoral statistics.

Candidate Study

The 2009 European Election Candidate Study was coordinated at the WZB Berlin, in cooperation with all four other Work Packages but especially with the European Election

Voter Study team at the University of Amsterdam Department of Political Science and the 2009 European Election Media Study team at the University of Amsterdam and the University of Exeter. The candidate team provided, as material for the creation of a candidate survey, a document that assembled all elite study questionnaires of relevance, and made particular measurements and concepts easily accessible by hyperlinks within the document. This document was completed in March 2008, and made publicly accessible on the PIREDEU web site (<http://www.piredeu.eu/Database/FORUMDOCS/CoQ-Elites1.pdf>).

Much emphasis was placed on comparability between the candidate and voter surveys for two major reasons. First, the PIREDEU study is an attempt to study electoral democracy also in terms of policy representation. To accomplish this, a number of attitudes and evaluations had to be worded in exactly the same way in both surveys. Secondly, PIREDEU also studies the working and mechanisms of representation, among them the most important one: mobilization. In order to be able to investigate the mechanism of mobilization, the candidate survey asks about activities and resources employed in the campaign. The voter survey mirrors many of these questions in order to evaluate whether voters perceived the candidates' and parties' activities.

In order to be able not only to compare candidates and voters in a cross-sectional perspective but also allowing for comparisons, which help to understand change, most questions for both surveys were taken from earlier studies. The voter survey and the candidate survey share 20 questions that had been asked in earlier comparative mass surveys, among them the European Election Studies and the European Social Survey. Furthermore, some of these questions had not only been asked in earlier mass surveys but also in elite surveys – the European Parliament Survey of 1996, or the Comparative Candidate Survey, for example. Altogether, these amounted to 6 items. In addition, three items, which had not previously been asked in voter, but only in elite surveys, were asked by us of both candidates and voters. Altogether, 29 of the items used in the candidate survey had been used in earlier European elite surveys, mainly among members of the European Parliament or candidates but also among members of national parliaments and candidates.

The final candidate questionnaire consisted of 66 questions, of which 22 were identical with those in the voter's survey, and four additional ones were "mirror" questions, meaning that for example, reported activities by candidates could be compared to voter's perceptions of those same campaigning activities.

The translation of the questionnaires followed an identical strategy for both surveys and involved the same team of translators for both the Candidate and Voter Survey. This was made possible by the close collaboration of the two teams and ensured maximum comparability of both instruments and of the resulting data. In line with the original aim of the PIREDEU project, at all stages decisions were made in close cooperation between the two Survey Teams. Equal importance was assigned to cross-country and cross-time comparability. All decisions were finally approved by the Steering Committee.

These translation procedures would have been impossible without the collaboration of an institution such as the EUI that has doctoral students in political science from most member countries of the EU. However, the recruitment, training and supervision of such students required a faculty member at the EUI who is a central collaborator in the research project. In the absence of such a member of the project there would be little alternative but to employ the translation services of the survey company. It would, however, be essential, in the call for tender that located such a company, to specify that two questionnaires were to be translated even if only one of them was going to be fielded by the survey company. Moreover, close

liaison with the two research teams and their country collaborators would be needed – something that would probably increase the costs of translation beyond what would normally be estimated.

The original plan at the time of the PIREDEU application was to run the Candidate Survey as a web-based survey only. However, discussions at the Kick-off meeting in Florence (Piredeu Conference – 21-23 February 2008) made clear the likelihood of low returns for such an instrument. For this reason, a two-mode design was developed that offered candidates the possibility to participate either in a mail survey or in a web-based survey. The master questionnaire was designed to fit the needs of both modes. The over-all layout and design followed common standards.

Contextual Data Study

Context is almost everything; hence there are many contextual variables that can influence electoral processes, also at the European level. The team of WP7 focused on variables suggested by members of PIREDEU following suggestions articulated in and after each of the meetings of the Steering Committee. In addition to various email exchanges of ideas, the teams also referred to proposals made in the PIREDEU Open Forum.

As the project's main objective is to provide an infrastructure for research, the team evaluated both the research payoff and the feasibility of collecting each domain of contextual variables according to an Implementation Plan established in the first months of the PIREDEU project. The main objective in the plan was to collect contextual data relevant for the European Parliament Elections 2009. The resulting data were sent to the PIREDEU Steering Committee and uploaded onto the PIREDEU webpage (February 2010).

Data collected by the team were divided into seven sections, which differed so far as their importance and feasibility of collecting were concerned:

1. Official election results.
2. Electoral laws.
3. Data on engagement and mobilization.
4. Nomination and recruitment data.
5. Some constituency specific data (mainly demographics).
6. Institutional characteristics of each country (party system, constitutional arrangements, political-cultural arrangements).
7. Survey-based data on political context.

According to the Steering Committee guidelines, three different levels of priority of data to be collected were distinguished (must do – high priority, crucial issues for WP7; possible to do – medium priority; very difficult to do – low priority). The team's main task was to collect contextual data necessary for other WPs (in other words data that was 'functional' for them). This logic implied that data produced by this Work Package be highly compatible with data collected by other WPs (which were collected with other data collection instruments). The research designs of the other WPs largely determined the format for data collected by WP7, ensuring that these data were compatible with other PIREDEU data.

As regards material collected, the WP7 team had to rely on very different sources of information, due to the diverse character of the data. Some of the contextual data were found in official sources. Thus primary sources of information included official statistics, official authorities and official institution archives, databases, websites etc. It is worth noting here that the team's major concern was to collect high quality data; and thus made the assumption that the highest quality data can be obtained from official sources. Electoral commissions (and

similar institutions) were approached in each country, since in theory they should have most of the required data.

However, for some of the data the team had to rely on existing political databases (contextual data collected in previous projects – i.e. CIVICACTIVE, INTUNE etc.). Therefore, reviewing existing political databases (desk research) was an important part of the team's work. These databases (but also secondary sources such as, books, articles, papers and other published documents, which often contain needed data) constituted the second major source of information.

A third important source of relevant contextual information was the project's country-collaborators. This small academic group provided the bulk of those attending the PIREDEU Kick-off Conference in February 2008, when they were officially invited to take part in the project as country-collaborators. WP7 asked them for help when it was impossible to find the necessary contextual data in any other way. Nonetheless, not all of them reacted positively to our queries and in such cases we had to find 'substitutes' based on personal contacts. Still, for some countries it was anyway very difficult to find an appropriate source.

Section 1 data were collected by contacting electoral commissions in each country, and also by using already existing datasets and databases. Sections 2 and 3 data were gathered by contacting electoral commissions in each country and by surveying country-collaborators. The team did not succeed in collecting all of Section 4 data due to collection problems, but the aim was to do it by contacting electoral commissions in each country. Sections 5, 6 and 7 data were collected by using already existing publications and databases. Finally, section 8 data were collected by including additional questions in the voters' survey, and also by using existing datasets (e.g. EES, ESS, CSES, Eurobarometer etc.).

Voter Study

The 2009 European Election Voter Study was coordinated at the University of Amsterdam Department of Political Science, in cooperation with the 2009 European Election Candidate Study team at the WZB in Berlin, and the 2009 European Election Media Study team at the University of Amsterdam and the University of Exeter. Additionally, an important role was played by the country experts, who offered advice and insight on country specific questions, checked translations and the final fieldwork version of the questionnaire. It had been decided to reduce costs by employing telephone interviewing. However, because of country differences in availability of sampling frames, this was not possible in all countries. In countries where the sample was drawn in terms of names and addresses for purposes of face-to-face interviewing, telephone numbers were obtained for a subsample (by looking up names in a telephone directory) and this subsample was interviewed by telephone, so as to permit mode effects to be identified that might otherwise have been attributed to country differences.

The largest problems encountered in preparing for this component of the pilot study arose from the Call for Tender and from translation issues.

Steering Committee Guidelines suggested that a Call for Tender for the Voters Study should be issued in January 2009. However, at that time the EUI was in the process of revising its own rules for Calls for Tender and the PIREDEU call had to wait for that revision to be completed. The new guidelines outlined a laborious process of review which, by the time it became possible to start the process, already meant that the Call would be issued too late for the survey to go into the field on time after election day in each country. So the only way a Call could be issued was by following an accelerated procedure. In the preparation of the Call we also had difficulty meeting requirements of the EUI committees in various ways. All of these difficulties boiled down to a single problem: those who designed the procedures and

who supervised their execution had never envisaged a situation in which it was an absolute requirement that the Call be answered and yield at least one viable tender. The subtext repeatedly was “well, you can always issue another Call”. This is not the case when an election is to be studied. Elections do not wait until procedures can be complied with, and an election study that is late because the Call was late is an election study with vastly reduced value due to the ephemeral nature of human memory. Any future EES and, above all, any future electoral studies infrastructure, needs to realize that misunderstandings on this score can have dire consequences.

Translation issues arose from the need to coordinate the voter study, candidate study and media study in terms of questions asked and coding categories employed. It was not fully realized when the Steering Committee issued its WP instructions that any coordination achieved in the design of English language data collection instruments would automatically be jeopardized at the translation stage if translations of the English language instruments were not also strictly coordinated. This was a primary reason for adopting an in-house translation procedure, organized at the EUI and described in the Candidate Study paragraphs, above. By the time the problem had become evident we could think of no other way of achieving the necessary coordination. But the project benefitted from the fact that a member of the Voter Study workpackage was an EUI professor and could take on the task of recruiting, training and supervising the necessary translators from members of the EUI research community. Without this coincidence the coordination problems inherent in translation would have probably been insurmountable in the available time. In this respect, as in the case of the Call for Tender detailed above, it will be important for future projects to take note of these inherent difficulties and plan early for ways to overcome them. Specifically, integration of translation into the Call for Tender should be investigated.

Fieldwork commenced on the first working day after the 2009 European Parliament elections, with final interviews being held on 9 July 2009. To prevent memory effects, an emphasis was placed on interviewing quickly after the election. Hence, 60 percent of interviews were achieved within one week of the elections, while 80 percent of interviews were achieved within two weeks of the elections. Target sample size in each EU member state was 1000 achieved interviews. Because the sample was a random one, the target was not precisely reached in all countries (see Table 3, below). Sample design and execution was handled by the fieldwork organization (Gallup Europe), to the specifications set out in the tender document.

Media Study

According to the data collection design, media content were captured for the 3 weeks prior (21 days) to the election date in each country (taking into account that the election occurred during a 4 day period across Europe with election dates varying by individual member states). The objective was to capture news media broadcasts across at least 2 stations in each country and at least 3 newspapers in each country. The outlets and sources are listed in appendixes of deliverable D5.2 as well as in the data documentation. In order to minimise risks to losing data or missing recording periods our objective was to capture as much material as possible centrally in either Amsterdam or Exeter. A second objective was to capture as much material as possible in digital form in order to archive the material. These two objectives meant that our goal was to capture as much material from TV news and newspapers via the Internet given changes in the broadcast and newspaper markets as well as technological advances in online media players and digital newsreaders. We met these objectives and our means of capturing content via web-based sources ensures the long term feasibility of the media content study.

We recruited coders whose first language was one of the 22 languages in the project. Coders were recruited through university job bulletins at both sites. Coders were interviewed and then 2 coders were hired for each country. Training was completed in Exeter first and then at UvA. The total training time was 50 hours, including reliability tests and training for the online data entry tool.

Manifesto Study

In order to collect and code all European Parliament election programs of all parties represented in that body we applied the approach of the ‘Comparative Manifestos Project’ (CMP) on elections to the European Parliament. The manual coding of the election programs (“Euromanifestos”) is one of the most important components of the project. Therefore, expert coders, from each country of the European Union were recruited and trained in coding conferences. The expert coders

- collected “relevant” Euromanifestos (relevant parties in our project are those that have been represented in the European Parliament at least once), and
- applied the Euromanifestos Coding Scheme for the coding of each argument of a document.
- entered coding directly into an online coding tool.

Expert coders were able to collect and code most of the political party Euromanifestos that were issued for EP elections in 2009.

III.2.g.4 Stage 4: validating the quality of these data

One of the strengths of the PIREDEU project was the close coordination between data collection workpackages to ensure validity and comparability of measured items. Careful attention was also paid to backward compatibility with items measured in previous European Election Studies. Sources of questions for the candidate and voter survey and indicators for the media and manifesto studies were past studies, instruments from other data components (e.g. questions from the voter survey replicated on the candidate survey) and other cross-national surveys (e.g. European Social Survey and Euro-barometer). Conceptual validity across the studies could thus be more easily achieved. Validation of the data collection instruments was reported in a series of deliverables. The close collaboration also allowed a number of specific procedures to be adopted that it is hoped will have ensured against the sort of data problems that have plagued past studies of European Parliament Elections. This involved separate quality and evaluation reports from each of the data collection WPs associated with the completion of Milestones 3, 4 and 5.

An important role was played in the design study (and hence in the pilot study of the 2009 EP elections) by a Data Committee which was responsible for the overall quality and integrity of the data collected in the pilot study and for evaluating the procedures involved in ensuring data quality in an eventual infrastructure created on the basis of our design. That Data Committee focused on the measurement quality of survey instruments, the quality of the linkage mechanisms embedded in the data, the comparability of the questions asked in different components of the pilot study, and (in two subcommittees) the quality of the sampling frames and of the questionnaire translations employed for the pilot study in the 27 EU member countries. We were strongly aware of problems of sampling and translation that needed to be overcome in any study of this sort, which is why we provided special-purpose subcommittees to focus on these problems. A Deputy Chair for Data Integrity monitored the work of four of the workpackages to ensure comparable coding standards.

The development of the questionnaire, as well as its translations, was carefully designed to increase the quality of the survey, e.g. in terms of wording or validity. A multi-stage translation process started by selecting already tested and widely accepted items whenever possible and decentering original questions by translating them into a test language (German) and then evaluating the result for changes in meaning. These and other procedures resulted in high quality questionnaires and there were no major problems in regard to the questionnaires during fieldwork. Specific issues related to technical validation of the data are detailed below.

Candidate Study

The dual-mode structure of the 2009 EECS (postal and internet) lead to different strategies to ensure data quality in regard to data input and data recording. The data recording of the postal returns was done at the WZB and by an external contractor, FAU GmbH. At the WZB, all non-numeric open answers were typed into Excel sheets. To decrease the probability of errors, an input mask was programmed at the WZB. The design of the Excel sheets was jointly agreed upon with WP5 (media study) which coded the open-ended answers.

The quality of data input into the online survey was directly assured, as far as possible, while programming the survey. For example, respondents were asked to type in their average time per week spent on campaigning. Candidates were only allowed to type in numbers between 0 and 168 hours.³ Needless to say, the data was provided by IVOX in electronic format (including open-ended answers which were added to the typed open-ended answers of the postal survey).

The candidate study team compiled the data from FAU GmbH, IVOX and the original candidate information to produce data sets per country as well as a combined data set. To accomplish this task, a common coding scheme had to be developed and labels had to be defined and assigned. If necessary, this was done in agreement with the Voters Study team (WP 3) and the Data Integration Team (WP 9).

The 2009 candidate study was conducted as a non-random survey and it is confronted with significant differences in regard to response rates (rates so low in some countries as to give rise to Ns in the single digits – See Table 2). Response rates were computed for two different response sets. Set 1 (postal survey) includes all respondents who gave at least one valid response to one of the questionnaire items. Set 2 (internet survey) includes all respondents who gave more valid than invalid answers. For set 2, the response rates range from 4.4% in Bulgaria or 5.6% in Poland to 34.4% in Malta or even 42.9% in Sweden. There seems to be a divide between many of the newer member states and older member states as well as the rather classical differences between southern and northern countries (See Table 2). The cross-country mean is 22.0% and the response rate for all countries combined is slightly above 20% which is a bit lower than originally expected. Therefore, sample weights have been included in the dataset to enable more or less representative analyses.

As a starting point, Duncan indices of dissimilarity have been calculated for three characteristics of candidates. It has to be noted that a comparison with the ‘population’, in the case of the 2009 EECS the over-all number of relevant candidates and their characteristics, is highly limited. The country teams were only able to collect a small number of indicators for this purpose, e.g. gender. In addition to gender, the dissimilarity measures were calculated for party affiliation and proportion of MEPs per country. Dissimilarity measures provide basic information on the proportion of respondents with certain characteristics in comparison of the

³ Obviously, 24-7 campaigning is not plausible but it was decided to define the range of possible answers in accordance to the logical minima and maxima.

respective proportion in the population of the 2009 EECS. The deviations are calculated as the sums of absolute differences: hence, the lower the difference between the proportions, the higher the representativeness.⁴ The dissimilarity measures as well as the response rates are presented in Table 2. Numbers are presented for response sets 1 and 2 (see above).

Table 2: Candidate Study Number of interviews and Dissimilarity Coefficients for two response sets*

| Country | Response Rate | | Gender (Dissimilarity) | | Parties (Dissimilarity) | | MEPs (Dissimilarity) | | N |
|---------------------------|---------------|--------------|---------------------------|------------|----------------------------|-------------|-------------------------|-------------|-------------|
| | Set 1 | Set 2 | Set 1 | Set 2 | Set 1 | Set 2 | Set 1 | Set 2 | |
| Austria | 34.0% | 25.0% | 1.9 | 7.5 | 48.3 | 49.2 | 1.99 | 6.12 | 39 |
| Belgium | 45.3% | 33.1% | 4.1 | 5.6 | 32.1 | 29.9 | 2.65 | 3.52 | 57 |
| Bulgaria | 5.1% | 4.4% | 19.8 | 17.4 | 34.5 | 48.7 | 36.96 | 53.63 | 6 |
| Cyprus | 26.7% | 26.7% | 2.6 | 2.6 | 26.3 | 26.3 | 19.32 | 19.32 | 8 |
| Czech Republic | 20.9% | 15.7% | 4.9 | 1.3 | 32.1 | 36.0 | 0.15 | 5.34 | 21 |
| Denmark | 30.4% | 23.5% | 14.8 | 17.9 | 30.7 | 26.7 | 1.34 | 0.96 | 24 |
| Estonia | 33.3% | 21.9% | 2.2 | 5.6 | 41.1 | 52.7 | 0.15 | 3.14 | 23 |
| Finland | 39.3% | 29.3% | 4.0 | 6.5 | 16.4 | 21.7 | 2.14 | 0.04 | 41 |
| France | 16.6% | 15.5% | 4.9 | 3.6 | 20.4 | 17.4 | 4.51 | 4.02 | 117 |
| Germany | 34.7% | 30.5% | 5.9 | 4.9 | 28.5 | 29.7 | 3.47 | 3.61 | 143 |
| Greece | 21.4% | 12.3% | 19 | 27.7 | 39.3 | 30.7 | 5.37 | 3.66 | 19 |
| Hungary | 22.8% | 19.1% | 5.6 | 6.3 | 56.4 | 56.5 | 3.82 | 8.25 | 26 |
| Ireland | 17.8% | 17.8% | 13 | 13 | 33.6 | 33.6 | 7.71 | 7.71 | 8 |
| Italy | 13.8% | 11.0% | 5.7 | 8.3 | 42.0 | 47.7 | 1.84 | 1.16 | 58 |
| Latvia | 51.3% | 33.9% | 5.7 | 0.8 | 24.4 | 30.1 | 2.53 | 4.46 | 39 |
| Lithuania | 25.5% | 20.1% | 3.6 | 0.6 | 44.5 | 42.0 | 0.72 | 2.37 | 30 |
| Luxembourg | 41.7% | 33.3% | 0.0 | 3.7 | 32.6 | 36.3 | 9.27 | 11.54 | 16 |
| Malta | 43.8% | 34.4% | 0.5 | 3.7 | 39.6 | 42.3 | 8.49 | 7.3 | 11 |
| Netherlands | 26.5% | 24.8% | 0.3 | 0.7 | 35.8 | 18.5 | 3.39 | 4.34 | 73 |
| Poland | 6.2% | 5.6% | 6.6 | 1.3 | 45.0 | 47.2 | 0.19 | 2.13 | 36 |
| Portugal | 14.2% | 14.2% | 18.5 | 18.5 | 13.8 | 13.8 | 1.1 | 1.1 | 17 |
| Romania | 11.3% | 9.7% | 5.3 | 5.9 | 41.4 | 42.8 | 12.16 | 6.32 | 24 |
| Slovakia | 27.3% | 22.7% | 5.6 | 2.3 | 40.7 | 43.1 | 5.9 | 0.43 | 29 |
| Slovenia | 28.4% | 22.2% | 9.7 | 4.4 | 45.3 | 50.3 | 0.37 | 8.33 | 18 |
| Spain | 19.1% | 16.0% | 1.4 | 4.3 | 78.8 | 78.0 | 6.56 | 5.04 | 57 |
| Sweden | 49.0% | 42.9% | 1.9 | 2.4 | 31.3 | 31.1 | 1.15 | 0.49 | 162 |
| United Kingdom | 29.7% | 27.8% | 3.9 | 3.5 | 44.7 | 45.9 | 0.88 | 1.78 | 244 |
| Cross Country mean | 27.2% | 22% | 6.3 | 6.7 | 37.0 | 38.1 | 5.34 | 6.52 | |
| All Countries | 24.4% | 20.6% | 1.3 | 1.4 | --- | --- | 0.6 | 1.3 | 1346 |

* N = 1,346. Response set 1: mail questionnaire; Response set 2: internet survey.

⁴ It has to be emphasized that the representativeness is in addition related to a country's response rate.

In regard to gender, the deviation between the population proportions and the sample proportions are only small or moderate. Only Bulgaria, Denmark, Greece and Portugal show values above 10. The dissimilarity for all countries is especially small. It can be assumed that the representativeness in regard to gender is acceptable.

The dissimilarity indices for parties are presented as averages for all parties in a country. The deviation proportion is calculated as differences between the vote share in 2009 European Parliament Election and the proportion of candidates of a party in percent of all answers in the respective country. In comparison to the other two characteristics (see above and below), these dissimilarities are significantly higher. Primarily, these higher values are based on the fact that candidates of smaller parties, in this case meaning that they received a smaller vote share, were equally or even more inclined to participate in the study. For example, the Austrian Greens received only 9.9% of the vote but 13 candidates (sample 2) participated in the study which corresponds to a share of Austrian respondents of more than 33%. At the same time, some of the major parties, for example in Spain, are underrepresented in terms of response rates.

The last columns relate to the proportions of elected MEPs and show similar dissimilarity figures to those of gender. The results prove to be acceptable, except in Bulgaria, Cyprus and perhaps Luxembourg.

Contextual Data collection

Despite the fact that the task of collecting contextual data seems a relatively easy endeavor, because the aim was collecting high quality data the venture was quite difficult in practice. The main problem is not a gathering of data itself, but rather its validation and cross-checking. Therefore, tools used in order to validate and check the data collected were of crucial importance for the project and require a thorough summary. Country-collaborators' assistance was the major mode of validation of the contextual data collected. They were asked to review the data collected for their country, correct them (if needed), find official sources that might serve as a reference etc. As a result of their work we avoided several mistakes, eliminated a number of incorrect data elements and adjusted much of the information.

Yet, in some countries such a validation procedure was difficult to put into practice and this caused many difficulties and inconveniences in our work. Thus, future research of similar nature should not expect to be able to necessarily find reliable collaborators who are willing to give their time without pay. In this sense, a project should invest some time in finding a proper person, making the relevant checks before selecting, nominating, and hiring. Hiring the relevant individuals would create a contractual obligation to deliver that was missing in this pilot study due to the inability of PIREDEU to pay any of its collaborators. Such a practice can potentially save time and generate more reliable data. Moreover, further selection criteria should include ability and capacity to engage with the project, but also on personal acquaintance and previous record of trust. In some cases, the best idea is to hire younger researchers (PhD candidates, young postdoctoral researchers) since they are by nature most likely to get involved, they have more time, they are happy to work with well-established, older colleagues and help them and thus help the project (and at the same time help their careers). In a nutshell, there was a negative correlation observed between age and involvement (albeit with some remarkable exceptions).

Voter Study

Data for the voters study was captured in the course of interviews conducted by the study's fieldwork agency by means of Computer Assisted Telephone Interviewing (CATI) and computer assisted face-to-face interviewing. Both methods ensured accurate capture of responses

without need for separate data entry. Response rates averaged 27.7% (see Table 3), which was a surprisingly good result for a survey largely conducted by telephone. However, this average masks considerable variation by country, with lows of 11.0 and 11.3 in the Netherlands and Austria, and highs of 48.7 and 51.0 in Malta and Hungary.

After the end of interviewing, the data were delivered by the fieldwork agency in an early-release version. These data were checked by Voter Study team members and subsequently released in a limited distribution to members of the PIREDEU Steering Committee as well as to the country experts. Both were asked explicitly to look for inconsistencies and anomalies to be addressed by the fieldwork agency, which proved very responsive to requests for information, clarification and (where necessary) correction. Overall the number of corrections required was small.

Table 3: Voter Study: mode, number of Interviews and response rates by country*

| Country | Telephone | Face-to-face | N | Response rate (%) |
|----------------|---------------|--------------|---------------|-------------------|
| Austria | 1,000 | 0 | 1,000 | 11.3 |
| Belgium | 1,002 | 0 | 1,002 | 20.3 |
| Bulgaria | 300 | 700 | 999 | 47.8 |
| Cyprus | 1,000 | 0 | 1,000 | 15.4 |
| Czechia | 300 | 720 | 1,024 | 36.4 |
| Germany | 1,000 | 0 | 1,004 | 16.9 |
| Denmark | 300 | 707 | 1,000 | 18.7 |
| Estonia | 1,000 | 0 | 1,007 | 32.3 |
| Greece | 1,000 | 0 | 1,000 | 14.5 |
| Spain | 1,004 | 0 | 1,000 | 20.9 |
| Finland | 1,000 | 0 | 1,000 | 28.1 |
| France | 300 | 705 | 1,000 | 12.6 |
| Hungary | 1,001 | 0 | 1,004 | 51.0 |
| Ireland | 1,000 | 0 | 1,001 | 37.7 |
| Italy | 300 | 701 | 1,000 | 18.8 |
| Lithuania | 300 | 700 | 985 | 40.6 |
| Luxembourg | 1,001 | 0 | 1,001 | 14.2 |
| Latvia | 1,000 | 0 | 1,050 | 46.3 |
| Malta | 1,005 | 0 | 1,000 | 48.7 |
| Netherlands | 302 | 700 | 1,005 | 11.0 |
| Poland | 1,000 | 0 | 964 | 23.8 |
| Portugal | 303 | 700 | 1,000 | 48.0 |
| Romania | 301 | 715 | 999 | 38.2 |
| Sweden | 1,000 | 0 | 1,002 | 17.9 |
| Slovenia | 1,000 | 0 | 1,000 | 18.7 |
| Slovakia | 1,002 | 0 | 1022 | 38.5 |
| UK | 1,000 | 0 | 1,000 | 18.0 |
| Total | 20,721 | 6,348 | 27,069 | |
| Average | | | | 27.7 |

*N = 27,069

Following the initial pre-release described above, the Voter Survey Team in Amsterdam and the Data Integration Team in Nottingham collaborated in producing an integrated and consistently labeled dataset, based on the final data release as delivered by the fieldwork company. This process involved further cleaning and some extensive corrections where, apart from country anomalies, attention was especially paid to between-country inconsistencies. To ensure data comparability, the Candidate Survey team in Berlin also collaborated to ensure identical decisions were taken when problems involved questions that the two instruments had in common. Also in this final stage two political weights were added in this release, to calibrate the data distribution based on the EP election results.

In addition, at this stage the Media Study team in Exeter and Amsterdam contributed by coding the open answers for a number of Voter Study questions, most notably those on what respondents considered the most important problem at the time of the election. Whereas in previous years these answers were coded to a bespoke coding scheme, in 2009 the data were coded to the same coding scheme as used by the Media Study and the Manifesto Study, further aiding comparability and integration of these various datasets. The integrated dataset was made available to the wider research community in mid-April 2010. This release was aimed at the informed user, and the usual ‘user beware’ reservations were attached.

In this way, a middle ground between data availability and integrity assurance was sought. Due to an unfortunate change in identification variables as between the pre-release and final release data as delivered by the fieldwork company, the coded open answers could not be included in this initial public release. They were added in January 2011 and, after extensive checking by PIREDEU Steering Committee members and country collaborators, leading to further minor changes, the first public release was made available to the user community in June 2011. This is also the release that has been deposited in the GESIS data archive, where minor corrections to the data resulted in a second public release.

Overall it can be concluded that the outcome of the voters’ study was satisfactory, but our experience leads us to stress the need for any future studies to build in additional response-enhancement procedures for countries that yielded low response rates in this study, particularly Austria, France and the Netherlands, all of which returned response rates under 14 percent.

Media Study

In order to validate the media content analysis, we followed both best practices for human coding of media content as well as past practice from earlier European Election Studies. The validation of the data proceeded through the data collection, coder recruitment and training procedures as well as reliability tests. Central and online collection of material improved reliability and efficiency and reduced costs.

In the process of data validation for the content analysis project, coders received training in the coding of material on the basis of the final version of the codebook and appendixes (see its appendix IV.1). Throughout the training, and based on feedback from the coders, instructions in the codebook were revised.

Reliability tests were conducted at the Universities of Exeter and Amsterdam based on the coding of 40 stories (30 newspaper and 10 television), as listed in the codebook’s annex IV.4. Coders were given three days to complete this task (taking them about 20-30 hours). Upon completion, the data were emailed to the coordinators who merged the data sets, calculated reliability scores (and where possible Krippendorff’s Alpha), and prepared an overview of ‘problematic’ variables. The variables in need of extra attention were discussed during an additional training session and new material was distributed to conduct test coding on these

variables and to formally re-assess the reliability of the coding. Upon completion of the second test, when satisfactory scores had been reached, material was distributed to coders to start the actual coding. This procedure was replicated at both institutions and reliability scores were remarkably similar in both places. In general, these tests reached acceptable levels of reliability.

Manifestos Study

The preparation of the 2009 manifesto study included explicit efforts to validate the revised and restructured Euromanifestos Coding Scheme (EMCS) and the handbook, which is used as a guide by coders. These validation efforts included checks on both reliability of coders and reliability of the EMCS. First of all, we conducted some experiments in order to test reliability of the revised coding scheme and the procedure for deriving units of text for coding. In a second step, we assessed reliability of the new design by conducting analyses with the experimental data. Using the results as the validation of the revised structure, we began the coding procedure.

Content analysis can be reduced to two basic data generating steps. First, texts are chopped into smaller units relevant to the research question, such as words, sentences, or quasi-sentences. A second step involves coding each unit by assigning a category from the coding scheme to each text unit. Both steps need to be evaluated in order to be able to assess the reliability of the resultant data. The involvement of non-deterministic instruments—such as human coders—in content analysis raises the issue of reliability. Subjective judgments are subject to stochastic variation, and possibly also to bias. Repeated application of the content analysis procedure, at both the unitizing and the coding stage, will yield different results each time because some text units will be misclassified in either or both coding processes. Part of the problem arises from the fundamentally indeterminate nature of human judgment, but this general problem is compounded by the uncertainties of unitizing decisions and the degree of ambivalence that is inherent in any coding scheme. The degree of human misclassification is best assessed through an analysis of inter-coder reliability using measures of coder agreement.

Another way of arriving at reliability scores is to compare the coding decisions with some “true” or “correct” classification (which, in the CMP literature, is known as the “gold standard”). The problems associated with unitizing can be analyzed in an experimental setting that focuses on the overall validity and reliability of the resultant data.

The validation exercise reached several conclusions. First, applying the new hierarchical approach using an online coding tool turned out to be an improvement in the coding procedure. The EMCS 2009 results are better than both the test coder group and the control group from EMCS 2002/04 coders, though these results were statistically not distinguishable. Second, untrained coders drawn from a sample of undergraduate students using an online coding tool statistically speaking performed as well as trained coders drawn from a sample of postgraduate students and junior faculty. Third, the jury is still out whether natural sentence unitizing is better than quasi-sentence unitizing. In terms of reliability our study shows that the two approaches are statistically indistinguishable.

In summary

All in all, the validation procedures employed in PIREDEU yielded considerable information regarding strategic decisions to be made before the launch of a research infrastructure for studying European Parliament elections, in order to avoid difficult and uncomfortable situations such as those experienced by the voter study and contextual data teams and described above, regarding especially the translation of survey questionnaires and the acquisition and validation of contextual data.

III.2.g.5 Stage 5: designing procedures that allow end users to easily access these data

The purpose of this stage was to deliver a “proof of concept” of a cubed data structure that would permit the full integration of the five data sets, with the ultimate goal of providing easy access to the data for academic non-academic users. The data should be post-processed (adding instruments of analytic value such as scale scores), before being integrated into this database, but the PIREDEU project was not funded for either full post-processing of this kind nor full integration, though the procedures for both are fully described in D9.3: Report on cubed data structure design.

The workpackage/data collection teams did implement certain post-processing procedures, in terms of unifying, amalgamating, harmonizing and distributing the data, and guaranteeing the anonymity of respondents of the voter and candidates surveys. Decisions were made on issues of harmonization of the different study components to facilitate use of the data in a cubed structure, with software allowing researchers and other stakeholders an easy and transparent access to the data, as well as means for dissemination of the data and metadata, and training analysts in using it. The various teams embarked on post-processing in late 2009. Data cleaning included correcting/removing invalid answer categories from the data, coding of open-ended questions to allow cross-country comparability and strengthen the respondents’ confidentiality, recoding and standardizing of variables to increase user-friendliness of data, and to ensure a dataset architecture suited to linkage procedures.

All study components used a four-digit country identifier (1 followed by 3-digit ISO code). This Identifier preceded all non-standardized data (i.e., codes that are incomparable cross-country (such as education categories, news outlets etc.) A unique seven-digit code was used to identify political parties referred to in the various study components. This code replaced the raw data entries, and facilitates the integration and linking of the various components. It is comprised of the country identifier, the party ideological family and a two-digit counter.

In addition, education data has been standardized, to allow cross-country comparisons. The data have been recoded into ISCED-97 education categories. Both original (country specific) and derived (ISCED-97) education data are available to end-users. The data was released in SPSS and Stata file formats, and is available for download from the GESIS data archive. The full metadata was released with the data, meeting the DDI 1.0 standard. The documentation includes full details on sampling procedures, field-work and post-processing procedures employed on the data.

III.2.g.6 Stage 6: Scientific and technical evaluation of the proposed infrastructure

As outlined in Sections III.2.g.2 and III.2.g.3, the scientific and technical validation of the proposed infrastructure was undertaken partially by the individual data workpackages. Final evaluations by these WPs were reported in D8.5 Scientific Evaluation of the Design Study and D9.5 Technical Evaluation of the Design Study. The external validation of the proposed infrastructure was the objective of a final user conference.

The final user conference aimed at bringing together scholars who had led the teams comprising the PIREDEU project and users who had been employing PIREDEU data (in pre-release format) in their research. At the same time, it was an opportunity to showcase the advances of our infrastructure platform and demonstrate (a) how we had constructively employed our user feedback to improve the quality of our designs and our data and (b) the potentials offered to researchers from different scientific user communities as well as to those interested in the audit of electoral democracy in Europe. The success of the conference lay in (a) attracting conference participants from the broader user community (cf. deliverable D8.4),

some of whom had already participated in the kick-off conference in Florence three years earlier, and (b) showcasing some 40 research papers that had employed the pre-release PIREDEU data.

The scientific added value to our user communities can be evaluated in many ways. Certainly, the total of 855 downloads that our final user conference papers counted in one month (5 Nov – 9 Dec 2010) and the 907 downloads of our different datasets, questionnaires and other accompanying documentation in the same period attests to the high scientific value of this research project.

To further demonstrate the scientific added value and the extent to which we managed to respond to users' expectations, we conducted a meta-analysis of the conference papers in order to better understand the use of our data. The findings provide us with quite interesting information on the added value of the database. An impressive 75% of the papers attempted a linkage among PIREDEU datasets and 52.8% used additional datasets alongside the above, demonstrating the complementary properties of our research and its ability to link to previous research and other types of databases. These two numbers are by definition two excellent indicators of success in terms of the added value that our design study could offer.

One additional finding of this meta-analysis that reflects well the scientific added value regards the broader objective sought in the course of this project and that was described extensively in Section B.2 of this report, that is to say, the evaluation of its potential contribution to the auditing of electoral democracy in Europe. An important percentage (65.2%) of conference papers were concerned with the main factors that have been identified in past research as lying at the roots of the problems of European electoral democracy. These papers dealt with issues of electoral engagement, information and representation.

As a final indication of the scientific added value of our study, the conference participants agreed unanimously to the creation of a consortium for electoral research in Europe (CERES). The ultimate goal of this Consortium is the establishment of a permanent infrastructure for electoral research in Europe. The feasibility study has established clearly that an infrastructure to study electoral democracy in Europe cannot focus on European Parliament elections alone, but needs to address the electoral process in Europe at all its different levels and facets. CERES will attempt to ensure that high quality data are available for the benefit of researchers and practitioners of our user communities and will support the creation of an infrastructure for gathering, depositing, processing and adding value to election and public opinion data that is efficient, reliable and adheres to international standards.

III.3 Financial Feasibility

The PIREDEU design and feasibility study was conducted on a shoe-string. A great many corners had to be cut in order to reduce the original 4.5million euro estimated cost to the 2.4 million that was ultimately provided by the EU's DG Research. In particular, the study had to depend on institutions and individuals carrying a load that would not be viable in anything other than a feasibility study or for any lesser prospective reward than the implied promise that if feasibility could be assured proper funding would follow. In particular, consortium institutions forewent usual and expected overhead contributions, their employees collaborating on the study worked far longer hours than were provided for by the financial arrangements, and country experts worked without remuneration in roles that were vital to the conduct of the study. Despite the external funding (by the British ESRC) of additional voter study questions, the questionnaire was still short by election study standards, so that a variety of questions that would normally have been asked in a national election study were omitted

from this one and additional questions that would have been helpful in assessing the functioning of democracy at the EU level were not asked.

Above all, PIREDEU suffered the loss of four proposed follow-up studies of elections to national parliaments held in the two years following the European Parliament elections. This prevented us from providing data that would have permitted an evaluation of the scientific benefits obtainable from linked studies of EP and national elections – linkages that we believe to be essential to an understanding of the quality of democracy in the European Union. To some extent such linkages can be provided to externally generated data (and we have seen earlier in this report that such linkages were actually made in a quarter of the papers delivered at the PIREDEU final conference). We return to this topic in the final section of this report.

A second aspect of financial feasibility is equally applicable to other Work Packages and regards resources for the tasks of the work package (a) with regard to tasks to be performed centrally, and (b) with regard to tasks to be performed by country experts.

a) Resources for tasks performed centrally

The success of the voter and candidate surveys and the other components of the PIREDEU pilot study shows that such linked studies are feasible. However, more efforts than anticipated were required to bring that study to fruition. In terms of costs covered by the collaborating institutions, two budgetary implications have to be considered:

- Because of the limited resources for personnel costs, the president of the WZB provided the project with an additional 8,000 Euro to cover costs for student research assistance. A similar contribution was made by the Director of the EUI's Robert Schumann Center in order to cover a shortfall in the funding of additional questions for the Voter Survey that should have been paid for by the British ESRC – a shortfall that was due to currency fluctuations. At the University of Amsterdam, considerable funds were expended from departmental and central resources to hire part-time assistance needed to coordinate the consultations between translators and country specialists. Similar (if lesser) contributions were made by other participating institutions.
- Just as importantly, Work Package leaders had to invest many more working hours than budgeted for in the grants awarded to the EUI, Amsterdam, Nottingham, Oxford and Exeter. This was partially the result of the reduction of the amount asked for in the application, partially a result (in the case of the Candidate Study) of adding a mail survey component, and partially (at all institutions) of underestimating the costs in time and money of coordination needs.

For example, instead of 30 person months requested, the WZB person months were reduced to 16 after the approval of the project, of which 14 were financed by the grant, and two by the WZB. The real investment of person months in fact amounted to the originally planned 30 person months or more, all of which came at the expense of the WZB. Much the same comparisons could be made for the other institutions mentioned above.

b) Resources for tasks performed by country experts

Although resources to cover expenses of country experts in conducting the Candidate Survey were made available after re-organizing the budget, these resources did not cover the real costs even of the candidate survey, as country experts have reported. The call for candidate study cost estimates to country experts already placed some limitations on their demands, to which limits they adapted due to collegiality and scientific curiosity. Without the constraints highlighted in the call for cost estimates, the cost demands would certainly have been higher.

Costs for collaboration by country experts in other PIREDEU components (particularly the voters and media components) were not reimbursed even in part.

For a future study of elections to the European Parliament, all these points have to be taken into consideration, because making up resource deficits by taking advantage of the generosity of collaborating institutions or the motivation of the persons involved cannot be expected on a permanent basis (as noted earlier in regard to problems found in collecting data for the contextual data work package).

The following points are crucial in this regard:

- Provision to the leading institutions of sufficient resources for all central tasks and coordination efforts (e.g. costs for the Candidate Study mail survey and coordination costs).
- The person months for each of the five projects should be extended to 36 in order to be able to start preparations earlier. Two half-time student assistant positions to be filled by advanced doctoral students should be available over the whole project period at the institutions of each Work Package leader.
- More resources for country experts are needed in order to cover real costs and to shorten the time needed for gathering the necessary information (especially candidate and contextual data information). In some countries, final selection of candidates is very close to the date of the election, a cause for low response rates seen in Table 2 for certain countries. In particular for countries in such a situation with a large number of candidates, sufficient resources have to be made available as to permit a timely start to data collection. In general an earlier start can reduce costs, but not in situations where the data are not available before a certain date.

IV. Scientific, Technical and Social Impact

The above section discussed the use of the data produced in the pilot study. In this section, we discuss the impact of the pilot study on scientific research as an indicator of the potential importance of an infrastructure for collecting and disseminating data of this type.

IV.1 Scientific Impact

The pilot study produced 5 linked data sets. Dissemination of the data provided new possibilities for scientific research. Prior to 2004 there had already been a long string of studies focusing on relation between voters of the EU member states and elections for Members of the European Parliament. Some of the classic foci of this research were the type of issues in the campaigns, the relation of European level elections to national issues, the profiles of the candidates, and a wider interest on participation and mobilization of the citizens of Europe. Nonetheless, key deficiencies in this research involved gaps in understanding the broader interplay among voters on the one hand, and, on the other hand, candidates, party platforms and manifestos, the media as information sources on electoral issues and the contextual background in each member state.

More importantly, though, the completion of the Eastern Enlargement of the EU in two waves (2004 and 2007) had recently added twelve new member states of which ten were emerging from a decade of dramatic political and economic transition. Hence, prior to the 2004 EP

election there were only very few, if any, longitudinal studies of elections in those post-communist countries.

In addition, even though elections to the European Parliament have been conducted since 1979, their functionality as mechanisms of democratic control and accountability remains considerably less than those at the level of the member states. It is mainly for this reason that European Parliament elections have progressively been seen as having a secondary character. This phenomenon has led scholars to describe them as second-order elections. Alongside the rejection in popular referenda of the Constitutional Treaty in two founding member states, these events have raised questions about the functioning of electoral process at the EU level, as well as about the democratic legitimacy of EU institutions, and accentuated concerns regarding a democratic deficit in the European Union.

Although previous research of a similar nature was absolutely essential in providing foundations for this project and made an undoubted contribution to our understanding of the relation of voters and elections in the European Union, we have managed to advance this previous experience to a next, more sophisticated level of research through the creation of the EES 2009 databases. Building this new infrastructure with the incorporation of a voter study, a candidate survey, a manifesto study, a media study and a contextual dataset, we have enabled the academic research community to explore in a more fine-grained and integrated fashion the interrelationships between the behavior of parties, politicians, voters and the mass media at national and European levels, and opened up a new universe of potential foci for electoral research in Europe.

More specifically, having seen applications of this infrastructure in the context of the final user conference in Brussels (Deliverable D8.4) and in other academic conference and publication venues, our research project has assisted in research within electoral studies and comparative politics. New light has been shed in assessing the nature, context and quality of the link between EU citizens and their representatives in the European Parliament and at the national level. We have enabled comparative research among the 27 European electorates in terms of their political attitudes and electoral behavior, and in shedding light on the evolution of a European party system. For example, the user community is able to investigate political representation, the role of parties and the responsible party model which concerns the extent to which parties offer voters clear policy choices (see section on assessing scientific quality – section III.2.g.6 above).

Our data offer insight into the existence of a common European public sphere, but also on the impact that national and European developments have on party choices, political communication at the media level, the behavior and attitude of national and European political elites and candidates. Finally, the refined and restructured survey questionnaires have given us much information on national versus ideological motivations of political behavior, on mass awareness and politicization of EU processes, and on aspects of democratization in the EU decision-making processes.

Examples of the types of questions listed above are to be found in the research output prepared for a special symposium in the scientific journal *Electoral Studies*.⁵ The articles in this special symposium examine in a comprehensive and rigorous fashion the state of electoral democracy in the European Union today, focusing on an in-depth analysis of the most recent 2009 elections to the European Parliament. Theoretically, the papers in the

⁵ *Special Symposium: Electoral Democracy in the European Union* Edited by Sara B. Hobolt and Mark N. Franklin. *Electoral Studies* Volume 30, Issue 1, Pages 1-246 (March 2011).

special issue critically review and contribute to existing theories of voting behaviour, second-order elections and responsible party government in multi-level systems of government. By appearing in this way in a high-profile peer-reviewed scientific journal only a year after the first preliminary data release, they demonstrate the importance to our scientific community of the data produced in the infrastructure pilot study and hence of an infrastructure if established. The major findings are outlined in the next paragraph.

In general authors find broad support for the second order model of elections, demonstrating that parties in national government are consistently punished at such elections. But this is not to the exclusion of European issues. This suggests that a pan-European ‘public opinion’ may be emerging in EP elections. Yet these common trends are not only concerned with preferences about more or less European integration, but also other salient issues on the European policy-agenda. There is evidence to support an argument that European issues gain importance when such issues are prominent in the information environment. As the informational environment is crucial to patterns of voting behaviour, articles in the Special Symposium demonstrate that party contestation on the issue of Europe contributes to its visibility in the news. Another finding is that, based on the candidate study, both individual-level, party-level and country-level factors can explain variation in the intensity and nature of individual EP candidates’ campaigning. Finally, yet another article demonstrates that the low-salience, second-order nature of European Parliament elections fails to mobilize new voters who have not yet acquired the habit of voting, with long-term implications for the evolution of electoral participation at both European and national elections.

Further evidence of the importance of scientific findings can be found in the sample of papers that were delivered at the final user conference in Brussels in November 2010. The box below presents some of the original research areas above and just few illustrative examples of the ways in which our users went about answering their respective research questions.

Box 1: Samples from the available research output relating to areas of PIREDEU research

Impact of national and European developments on political communication at the national level:

Schuck, A.R.T, Xezonakis, G., Elenbass, M., Banducci, S. and de Vreese, C. (2011) “Party Contestation and Europe on the News Agenda: The 2009 European Parliament Elections” *Electoral Studies* 30(1) special issue..

Nature, context and quality of the link between EU citizens and their representatives in the European Parliament and at the national level:

Giebler, H. And Wüst, A. (2011) “Campaigning on an Upper Level? Individual campaigning in the 2009 European Parliament elections” *Electoral Studies* 30(1) special issue.

The existence of a European common public sphere:

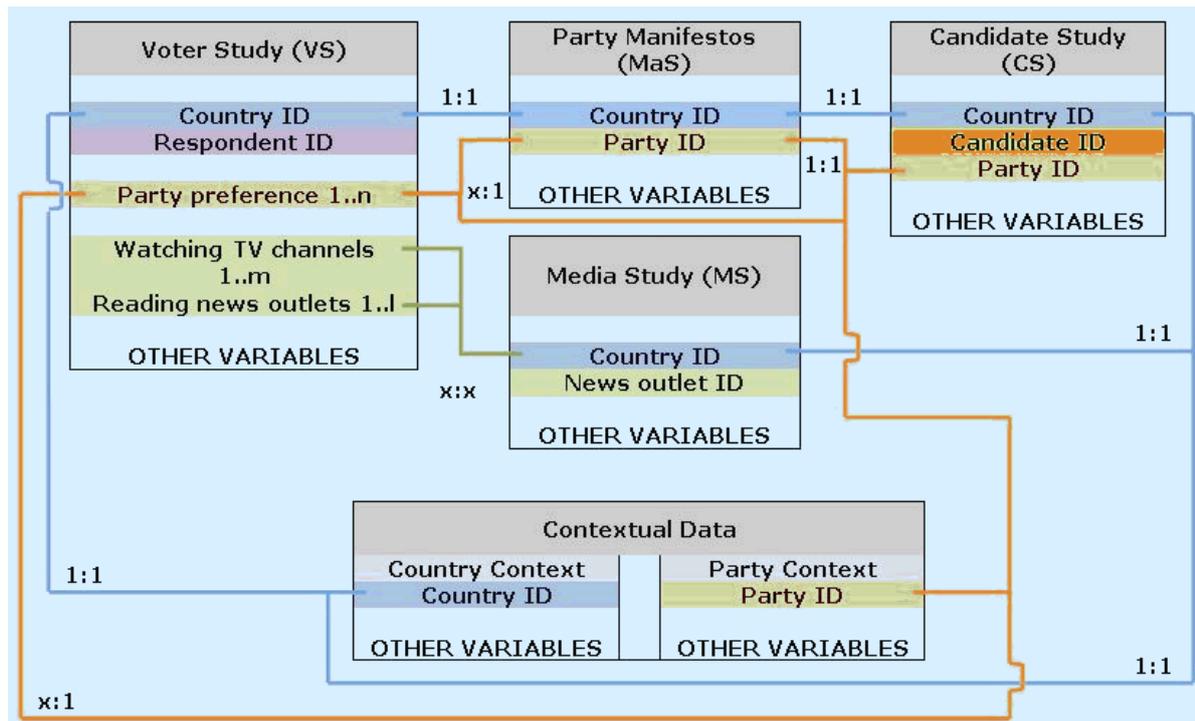
Gschwend, T., Lo, J. and Proksch, S.-O. (2010) “Europe’s common ideological space,” paper presented at the PIREDEU final user conference, 18-19 November 2010, Brussels, http://www.piredeu.eu/Database/Conf_Papers/V2_1-ESSpaper1.pdf

IV.2 Technical Impact

Figure 1 below summarizes the ways in which the component datasets produced by the PIREDEU pilot study are linked. The linkages illustrated there are subtle ones, giving rise to quite nuanced data, whose subtleties need to be retained in the linkage process. In the past,

linkage between datasets has generally only been possible in terms of rather coarse categories – the lowest common denominator of different coding schemes employed in different studies. In the PIREDEU infrastructure design study considerable efforts were made to collect data whose categories can be retained even after linking. This required careful design of concepts and categories, as described in III.2.a, above, so that they are appropriate for all five of the data components illustrated in Figure 1, eliminating the need to collapse categories to their lowest common denominators.

Figure 1: The PIREDEU Relational Data Base



Two different usage modes are envisaged. One consists of user-driven interfaces for linking information from different datasets, the other of cubed data structures. The difference relates to different kinds of data users with different needs, as follows:

- user-driven interfaces are intended for researchers who have access to dedicated statistical software (e.g., SPSS, STATA, R, etc.), and who will generally download data to their own computers for analysis and want to ‘tailor’ the data to their specific research questions. We expect this to be of particular relevance for academic researchers who are familiar with the use of dedicated statistical software packages.
- cubed data structures provide the basis for fast and easy on-line analyses of the available data. This is particularly relevant for data users who have no direct access to statistical software, and whose needs can be addressed adequately by on-line tools for descriptive analysis (frequency distributions, summary breakdowns; graphical displays). We expect this kind of use by parliamentarians and policy makers (and their support staff), journalists, librarians and archivists, analysts working with interest groups and NGOs, and politically interested members of the general public.

In line with the PIREDEU perspective on electoral democracy, the various datasets are seen as providing complementary information about a complex and multifaceted reality. And although each of the datasets provides invaluable information by itself, the project sees it as one of its main tasks to provide possibilities for linking or integrating information from different components in user-friendly and user-driven ways, and thereby to promote a fuller utilization of the joint potential of the data that were collected, so as to improve research on electoral democracy in Europe.

As already explained, special care was given in the post-processing efforts to assuring that the variables from the different components were coded using identical protocols and guidelines for each component. In particular this applied to the keys needed for linking the different data components. There were two types of keys used in PIREDEU data components and cross their boundaries, in a relational fashion, as illustrated in Figure 1.

Each of the five studies has a unique, primary key, accounting for the study's main observable units of observation (e.g. voter ID for the voter study, party ID for the manifesto study, etc.) All studies document data that is related to all other study components. Thus, for example, the interviewees in the voter study were asked to report on their party preferences, media consumption etc. These data were re-coded to allow matching them to other data sources (namely, manifesto data and media content data, respectively).

The main objective of the production of cubed data structures in WP9 was to provide adequate access and usability of data on electoral democracy to a category of end-users that is overlooked in the ubiquitous form of data dissemination in the social sciences, which consists of making micro data sets available for downloading via websites and data archives. This traditional form of dissemination serves well the community of academic data users, who are trained in data analysis and who have access to and familiarity with specialised statistical software. It does not serve others very well, however, (including policy makers and their support staff, journalists and others) whose needs are not defined primarily by multivariate statistical analyses but rather by descriptive analysis and comparisons of sub-populations. The objective of the cubed data structures is thus to cater to this category of end users and thereby to expand the user base of data, the kind of uses made of the data, and the direct policy relevance of the data.

Cubed data structures are generally used to allow very fast and simple on-line querying and descriptive analysis of data. This kind of usage extends to many national statistical offices, the World Bank, regional observatories in the UK, and sundry other institutions that service a clientele of users whose needs for data-analysis are mainly of a descriptive kind (univariate distributions and their summaries in terms of means and other such statistics, breakdowns of such distributions across geographic and political units, graphical displays such as scatter plots and trend lines) without strong needs of statistical inference.

A cubed data structure is in principle nothing else than a high-dimensional contingency table. The user determines which of the dimensions involved are to be activated and in what order. All cells of the non-activated dimensions are aggregated, hence they do not clutter the display desired by the user. The order that is specified by the user determines which dimensions are the primary ones, and which serve as factors to be used in breakdowns. Software for on-line usage of cubed data structures allows the user options such as simple recodings (collapsing categories of the dimensions), filtering (i.e., deleting from the display sets of cases of no interest), rearranging the order of the dimensions to be displayed, and choice of various summarizing descriptive statistics.

The fundamental architecture for the user interface for linking has been defined, presented at the international IASSIST conference, and has received very positive feedback and evaluations. It will form the basis for implementation using dedicated relational data base management software. Still to be finalised are the choice of engine to run the user interface itself, IO architecture and metadata generation for the linked data. A first substantive application of linking data has been reported by van der Eijk, Schmitt and Sapir (2010) in a forthcoming chapter of a *Liber Amicorum* for professor Jacques Thomassen. This application links manifesto study data with voter study data; it served as a pilot for identifying many of the generic choices to be made in the process of data linking. The fundamental architecture of cubed data has been defined, and a software platform for building and analysing a cubed data structure has been chosen. Further implementation in terms of choice of primary variables and breakdown variables will involve consultation with data collection teams, after which technical implementation will be unproblematic. However, the PIREDEU project was not provided funding for this next step in the development of the proposed infrastructure.

IV.3 Future potential impact: ensuring the viability of a permanent infrastructure

We have indicated at several points in this report the damage to the study that was occasioned by inadequate funding. The fact that we were able to conduct a successful pilot study that does establish without question the feasibility of the proposed infrastructure should not be taken to also establish the budgetary needs for such a study.

A specific loss to the project due to budgetary constraints was the collection of data for national elections conducted following the European Parliament elections of 2009. Such data would have permitted us to document the utility of an infrastructure incorporating both national and European Parliament elections. Still, even without such documentation in our Final Report, the case for studying European democracy in simultaneously national and cross-national terms is unassailable. European Parliament elections occur in the context of democratic elections to national parliaments throughout Europe, and European Governance is a multi-level enterprise involving institutions at national as well as supra-national levels. So the case for an integrated infrastructure can be made on logical grounds buttressed by evidence from our Final Conference based on the number of papers (already referred to) that found it intellectually profitable to link PIREDEU data with data collected externally to the PIREDEU project.

Both old and new infrastructure rules require most of the costs of data collection for any infrastructure to be born by national funding agencies. National funding agencies already fund national election studies, and bringing such studies under the umbrella of Europe-wide infrastructure would be no more difficult in principle than bringing together research projects in any other field of study. The particular problem of studying European Democracy arises from the critical position of the European Parliament in that democracy. It is true that national funding agencies (especially the British ESRC, the German DFG and the Dutch NWO) have in the past provided major support for studies of EP elections, and the British ESRC indirectly contributed to the funding of PIREDEU, by funding a battery of the questions included in the voters study. But elections to the European Parliament do not clearly fall within the remit of any particular national funding agency and there is no obvious reason why such funding agencies would be willing to commit resources over an extended period to successive studies such as are envisioned by the proposed infrastructure.

V. Contact Details

Providing an Infrastructure for Research on Electoral Democracy in the European Union
(PIREDEU)

Address:

Robert Schuman Centre for Advanced Studies (RSCAS)
European University Institute (EUI)
Via delle Fontanelle, 19
I – 50014 San Domenico di Fiesole
Italy

Phone: +39-055-4685-1

Fax: +39-055-4685-770

Email: PIREDEU@eui.eu

Co-ordinator: Prof. Stefano Bartolini, RSCAS, EUI (stefano.bartolini@eui.eu)

Chair of the Steering Committee: Prof. Mark Franklin, SPS Department, EUI
(mark.franklin@eui.eu)

Project Manager: Ingo Linsenmann, RSCAS, EUI (ingo.linsenmann@eui.eu)