Academic Networking Fund
Application Form

1. Title of Project

Beyond Text: Speech, Image and Social Media Data for the Social Sciences
2. Names, Schools/ Centres and Contact Details of Applicants

Please identify the lead applicant. This person will be responsible for submitting the final report.

**School of Informatics:**
- **Lead:** Johanna Moore, Chair of Artificial Intelligence and Head of School
- Catherine Lai, Research Fellow, Centre for Speech Technology Research
- Peter Bell, Research Fellow, Centre for Speech Technology Research
- Steve Renals, Chair of Speech Technology, Director of Institute for Data Science and Engineering
- Jon Oberlander, Director of Institute for Language, Cognition and Computation, Director of Data Technology Institute
- Bea Alex, Research Fellow, Institute for Language, Cognition and Computation & Alan Turing Institute
- Robin Hill, Lecturer in Cognitive Science/Co-director of the Centre for Design Informatics
- Maria Wolters, Reader in Design Informatics

**School of Social and Political Science**
- Laura Cram, Professor of European Politics and Director of NRLabs Neuropolitics Research
- Juliet Kaarbo, Professor of Foreign Policy; Co-Director of the Centre for Security Research
- Andrew Hom, Lecturer in International Relations
- JP Singh, Chair and Professor of Culture and Political Economy; Director, Centre for Cultural Relations
- Chad Damro, Senior Lecturer, Jean Monnet Chair and Director of the Jean Monnet Centre of Excellence, and Co-Director of the Europa Institute
- Clare Llewellyn, Research Fellow NRLabs Neuropolitics Research

**Schools of Philosophy, Psychology and Language Sciences:**
- Joe Fruehwald, Lecturer in Sociolinguistics in Linguistics and English Language
- Lauren Hall Lew, Lecturer in Sociolinguistics in Linguistics and English Language
- James Kirby, Reader in Linguistics and English Language
2. Finance details

<table>
<thead>
<tr>
<th>Cost centre:</th>
<th>265000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job code:</td>
<td>G00001</td>
</tr>
<tr>
<td>Account code:</td>
<td>1558</td>
</tr>
<tr>
<td>Project name:</td>
<td>Speech as Data Network</td>
</tr>
<tr>
<td>Named budget holder:</td>
<td>Prof. Johanna Moore</td>
</tr>
<tr>
<td>Finance administrator within your School:</td>
<td>Mrs Avril Heron</td>
</tr>
<tr>
<td>Total funding requested:</td>
<td>£3000</td>
</tr>
</tbody>
</table>

3. Project summary (150 - 250 words)

*Please outline the context, aims, objectives and potential impacts of this project. This summary may be used on the IAD website.*

While there has been considerable work on using text as data in social sciences, much less work has been done on audio or video data. However, traditional print media providers are increasingly producing audio content such as podcasts. Similarly, we have seen a proliferation of new content in the form of video blogs, which are primarily distributed through social media. Spoken documents have also become increasingly pervasive in government, as elected bodies keep direct audio and video records of, for example, parliamentary proceedings, alongside traditional edited transcripts. Being able to analyse this sort of data en masse has important implications for understanding the relationship between government, traditional media and social media.

The aim of this network is to accelerate research in digital social science using speech data on a large scale. To do this, we propose to bring together researchers from automatic speech and language processing, political science and linguistics, together with other potential user groups (e.g. BBC, Scottish Parliament). We plan to organize a series of workshops to focus on development of interdisciplinary project proposals, identifying salient problems in social and political sciences, alongside sharing of data and tools. To help make this more concrete we intend to centre the workshop series around the goal of setting up a specific data-driven challenge for the wider speech processing community.
4. **Project activities and timelines (150 - 250 words)**

*Please give details of the planned activities, who will be involved, size and composition of target audience, how will you organise the development and implementation of the project.*

To build our network and organize research collaborations, we plan to hold four one day workshops through 2016/2017. Workshops will involve network members (listed above) and other interested researchers and data providers (40-50 participants). They will be based around answering questions that will guide long term research plans:

- **Introductory meeting:** Introduce researchers from informatics, politics, social sciences, and linguistics. Participants will describe existing related work, tools and data. The workshop will include short talks to describe current research and breakout groups for researchers to make connections.
  - What are the most important research questions that need to be addressed in each field?

- **Data exploration collider workshop:** Three participants will present specific problems/opportunities based on previous discussions. Participants will work in interdisciplinary groups to refine data requirements, hypotheses and challenges based on these issues.
  - What sort of data is available and how can we use it effectively?
  - We sort of data do we need?

- **Prototyping workshop/hackathon:** Based on the outcomes of previous workshops, participants will split into groups to develop prototypes, data reports, and visualizations.
  - What technologies are needed for research in digital social science?
  - What sort of tools would be useful for the wider community?

- **Proposal writing workshop:** Participants will consolidate results of previous workshops. We will draft outlines and work schedules for an overview paper of the field, as well as future grant proposals, and develop guidelines for a potential community challenge/evaluation campaign.
  - Where do we go from here?
  - What papers/proposals can we write in the next year?
5. **Project outcomes and benefits (50 – 100 words)**

*What are the anticipated learning outcomes from your project and who will benefit, including the project team members?*

- Development of a challenge/competition for the speech and language processing community will accelerate work in the field.
- Input from social and political science researchers will ensure that time is spent on problems that are important to digital social sciences.
- Demonstration and sharing of what technologies are currently available to researchers, as well as development of new tools.
- We expect the project team members will benefit in terms of knowledge exchange, but more particularly in developing concrete plans for grant applications and publications.

6. **Feedback and evaluation (50 words)**

*Describe how your project will be evaluated*

In the short term, evaluation will be based on the persistence of collaborations. For example, ongoing reading groups and projects taken forward after the last workshop. In the longer term, we would consider the number of publications and successful grant proposals that come out of the group.

7. **Future plans (50 words)**

*Describe how the network established or supported by this project could continue in the future*

- Publication of an overview paper on speech as data in the social sciences.
- Running of a social science oriented challenge in the speech community.
- An ongoing interdisciplinary seminar series (at least through 2017-2018).
- Completion of interdisciplinary publications and grant proposals developed in the last workshop.
8. **Financial breakdown**

*State total sum requested and a breakdown of expenditure.*

4 x 1 day workshops in the Informatics forum:

**Catering of 4 workshops for 50 people**

<table>
<thead>
<tr>
<th>Item</th>
<th>Unit cost</th>
<th>No. of people</th>
<th>No. of workshops</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lunch</td>
<td>9.50</td>
<td>50</td>
<td>4</td>
<td>1900</td>
</tr>
<tr>
<td>Coffee &amp; pastry</td>
<td>3.00</td>
<td>50</td>
<td>4</td>
<td>600</td>
</tr>
<tr>
<td>Coffee &amp; traybake</td>
<td>2.50</td>
<td>50</td>
<td>4</td>
<td>500</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15.00</strong></td>
<td><strong>150.00</strong></td>
<td><strong>12.00</strong></td>
<td><strong>3000</strong></td>
</tr>
</tbody>
</table>

**Total Cost Request**

£3,000

9. **Support from IAD**

*We may be able to provide additional support for your project, particularly around event design and facilitation. If this is of interest, please contact Lara Isbel (517 143) before applying to discuss your requirements.*
Sign to confirm that the information contained within this application is correct to the best of your knowledge.

Lead Applicant Name: Prof. Johanna D. Moore

Lead Applicant Signature: [Signature]

Date: 10 Oct 2016

Guidelines for Submission

- Please save your completed application as either: ANF_S2016_Lead-applicant-surname.doc OR ANF_S2016_Lead-applicant-surname.pdf

- Email your application to Lara Isbel at lara.isbel@ed.ac.uk

- **Deadline: 5pm, Monday 10th October 2016**