EPSRC
Pioneering research and skills

TOWARDS AN INTELLIGENT INFORMATION INFRASTRUCTURE

3 – 4 December 2012

EPSRC Workshop Report

Matthew Lodge

ICT Portfolio Manager

Engineering and Physical Sciences Research Council

Introduction

This report provides a summary of key points extracted from the discussions at the Towards in Intelligent Information Infrastructure (TI3) workshop held at Jury's Inn Hotel, Broad Street, Birmingham between the 3rd and 4th of December 2012. The workshop attendees comprised 60 academics from a broad range of career stages and with expertise profiles across fields of research related to this cross ICT priority and a number of industrialists from relevant sectors.

The workshop was designed to generate thinking around research ideas associated with the TI3 priority; identify existing research activity, create a vision of a future where TI3 is a success (identify desired outcomes), and identify what stands in our path (identify what we need to do).

This report will explore the outcomes of the final session of the workshop which dealt with the questions, "what needs to happen?" and or "what we need to do?".

Group Sorting Exercise

The majority of the workshop was undertaken in groups to which the attendees declared a preference to join. The groups were each given a set of research keywords to base their discussion around; these keywords were grouped in advance by a pre-workshop sorting exercise. This exercise was undertaken by the academic and industrial participants prior to the start of the workshop with the intention of giving each participant a fair say in the organisation of the groups. To enable better understanding of this report the topic groupings are broadly given below.

Informatics	HCI	Digital Economy	Communications	Networks
Ontologies	Interaction Design	Provenance	Smart Devices	Middleware
Image Interpretation	Accessibility	Trust	Software Radio	Quality of Service
Big Data	Localisation	Integrity	Broadband Antennas	Intelligent Networks
Information Management	User Experience	Security	Terahertz	Content Aware Data Transfer
Data Mining	Visualisation of Data		Modelling of Devices	Cloud Computing
Semantic Web	Multimodal Technologies		Quantum Communications and Computing	Internet of Things
Context Based Information Retrieval	User Interface Design		Cognitive Radio	
Information Ret revel			Optical Infrastructure	
			Tuneable filters	
			Green Communications & Computing	
			Photonic Devices	

In addition to the above there was also a group covering topics associated with the theory of computing. This group did not attain enough member interest to be viable. There were too many people for one informatics group so two separate groups were formed to look at these topics; their outputs are however have been merged.

Executive Summary

Generally speaking the workshop identified two cross cutting challenges, both of which are complex in their own right and in the ways with which they interrelate. These challenges could be broadly identified as ***"The Communications Bottle Neck"*** and ***"Extracting Understanding from Data"***. The workshop has identified important elements within these large challenges, which are themselves significant. In addition issues that cut across the two large challenges and some needs for changes in behaviour have been identified.

There is a clear link between the two challenges identified by the workshop. Working together to address challenges that need to be met to deliver achieve the aims of this priority was possibly the most overarching of the outcomes of the workshop. This included the inclusion involvement of social scientists and collaboration between more disparate areas but there was a strong indication that much of the collaborative working was from within sub-disciplines inside the ICT space, particularly within the Computing, Informatics, Networks and Communications communities. Although the need for working together between those interested in "The Communications Bottle Neck" **and** "Extracting Understanding from Data" was not identified explicitly there is sufficient overlap between the identified challenges to suggest that this is necessary.

Within both of these broad technical challenges the need for some sort of 'facility' or 'test-bed' was identified. To those interested in the Communications Bottleneck this represents an investment in centralised equipment or the enable of sharing existing equipment. For those interested in extracting understanding from data a centralised source of safe usable 'big' data to knowledge sources is indicated as being important.

The workshop attendees recognised the importance of considering the user and the broader society and ethical issues relating to the challenges including the need for 'clean and green' energy efficient technology.

As a result of consideration of the outputs from the workshop and comments from the ICT SAT a decision has been made to announce a call "***A step towards an intelligent information infrastructure***" in early May 2013 in which applications will be asked to address two or more sub challenges presented (further details can be found in the conclusion section of this document).

Please note that workshop attendance will **not** be a prerequisite to apply to the call.

Workshop Outputs

The purpose of the session whose outputs are presented here was to identify what prevents us from achieving the goals and to propose possible solutions. Some groups just presented barriers or obstacles and others also presented solutions. Solutions when proposed are placed after the obstacle to which they relate and are presented in *italics*.

The output of this session was prioritised by vote by the attendees of the workshop. Vote scores are shown inside **square brackets in bold** next to the problem or solution which the attendees identified.

Please note that because of the open nature of the facilitated discussions undertaken during the workshop the different groups were operating at different strategic levels and thus their outputs vary.

Informatics

1. Increasing diversity and complexity of data sources, data links and the volume of data **[22]**. *We need scalable incremental approaches and to bring communities together.*

2. Need to change how we view notion of a system itself (don't know how to characterise socio-technical systems) **[13]**. *Possible solutions include: Specify interactions amongst social actors, design of the system and or platform.*

3. How to ensure access to big data and users for study (taking ethical issues into account). **[12]**

4. Privacy needs, including: government & industry involved, different levels of privacy, how expose to users we need models **[6]**.

5. Making peer review aware of the context of theTI3 priority including identification of appropriate advisers and partners **[3]**

6. Users are not sure what they want, conversely developers don't understand many users (cultural differences across generations) **[2]**. *Potential solutions include: More flexible models of interaction, more generic 'info' architecture, more generic 'info' architecture, need better presentation/representation/interaction to users.*

7. How to manage Big Data now and into the future? **[2]** Make use of University Libraries? *Create a TI3 community dataset (NHS, Web and linked data Social perhaps).*

8. Lack of standards, especially standard ontologies, how to encourage people to adopt them (these are driven by industry and are international). **[1]**

9. Fundamental changes in algorithms design and selection are required (algorithm development incremental and or learning algorithm).

10. Increasing pace of change of technology (keeping up and managing, data exchange/migration across tech)

11. How to manage multidisciplinary nature of the problems presented by the TI3 priority (links with communications infrastructure research for example) and manage links to other projects.

12. How to demonstrate a step change in level of inference to be drawn from MM and MS data, applicants and or impact.

13. Availability of data and knowledge sources to interpret these data.

14. How do we ensure we have sufficient skill base (graduate and postdoctoral) to tackle problems associated with the TI3 priority?

Human Computer Interaction

1. Knowing how to turn data into understandable and actionable information. Our efforts are stuck in various silos (because researchers focussed in progressing own efforts and ideas) and we lack the following: No empirical view, No ground truth, No mechanism which cuts across entire space **[13]**. *Transfer best practice intra-collaboration and international evaluation. Interdisciplinary collaboration – lead to methodologies.*

2. Know where to put computational power to get low overall energy usage while maintaining effectiveness. We don't know dimensions which will allow us to evaluate what is effective, problems are very dynamic and requirements very specific **[7]**. *Involve economists, study use cases, and develop methodologies for analysing the empirical.*

3. Understanding how physical and knowledge infrastructure should be coupled. We don't have a shared understanding of what knowledge infrastructure is; there are language difficulties, multiple interpretations and vocabulary difficulties and there is separation of those in physical and those in knowledge **[4]**. *Possible solutions include: Definitions common language, Identification of the key elements that make a knowledge infrastructure and coupling these thru common endeavour, and identifying bottlenecks where Physical and knowledge infrastructures constrain one another and the boundaries between them.*

4. Human in the loop. Sometimes there is an incorrect assumption that the human isn't needed in loop (also the opposite is true) **[1]**. *Focus on collaborative intelligence – humans and infrastructure – augment human intelligence – performance metrics and methodologies needed.*

Digital Economy

1. Test Beds are required but there are many barriers including; scale of current network investment, coordination and management, who would own it, disparate views. *Possible solutions include: Multi-disciplinary coordination (inclusion of industry)* **[4]***, shared ownership of test bed and data maybe using UKERNA (JANET), access to and curation off public data sets (existing test beds…)* **[15]***, and development of leadership and skills.*

2. A Framework for value and risk assessment of data is required but there are barriers including; commercial commodification and business interests, a lack of shared data and or content, the perception of risk to the public (lack of transparency), the technical problems associated with adequate anonymisation of data. *Possible solutions include usable models of ownership and risk* **[6]** *and being transparent.*

Communications

1. Converged networks (we don't know how to design them). *Possible solutions include: developing Design Tools* **[18]***, and experimental networks (bring the community together)* **[5]**.

2. Spectrum availability and increased demand (compound interest) **[9]**. *Solution must be multi-dimensional making use of, white space, unallocated spectrum* **[2]***. New devices (both RF & photonic) will be required to deal with these problems, the current work is incremental step changes and radical new approaches are required* **[5]**.

3. Changing internet protocols and dealing with entrenched interests (academic and Industrial) in existing internet protocols (analysis of internet protocols, what can they do?) **[9]**. *Engagement across the community to bridge gap between physical layer and abstraction.*

4. Economic cost – energy trade off (bearing in mind embodied energy); New networks that cost users less, Design tools for modelling, how do we design to be more simple? **[5]**

5. We need to draw in different and diverse communities, how do we get them to work together? *Maybe WINES*[1] group must be broader than the people at this workshop. **[2]**

6. To cope with network complexity we need develop an understanding of the fundamental principles of network complexity.

Networks

1. Researchers interested in different layers of the system do not work together **[24]**. *Possible solutions could include: Multi-layer and or "large systems" systems projects, grand challenge, Focus provided by considering 'futurologist' techniques and or potential future scenarios (Has any work been done that we can build upon?) (Any current modelling/scenarios from government etc?), Radical combinations of research fields (outside of usual domains), and Scenario planning with researchers.*

2. Research is not always well-integrated with users/stakeholders **[8]**. *Possible solutions include: Shared application vision, grand challenge, small competitions in conjunction with stakeholder sponsors, and or create a new 'application vision' that is compelling (with respect to what we are trying to achieve) open data or foresight challenges for example (Do we have good visions of what we don't want?, no privacy for example).*

3. There is no current community that spans the TI3 initiative **[2]**. *Possible solutions include: Initiative longevity, carrots for collaboration, spreading out funding opportunities starting with small grants moving to larger grants, consider issues of peer review (including PM), requiring joint research agendas on grants (novel in all disciplines), annual workshops, and activities to force the challenging of traditional assumptions (Tension, detrimental to science against building of community).*

[1] The Wired and Wireless intelligent Networked Systems (WINES) initiative was a series of 17 research projects of total value £22M funded between 2003 and 2007. Outputs from the workshops can be found here www.epsrc.ac.uk/SiteCollectionDocuments/Publications/reports/WINESWrkshpRpt07.pdf (2007) and here www.epsrc.ac.uk/SiteCollectionDocuments/Publications/reports/WINESWorkshopReportvFinal.pdf (2011).

Conclusions

To assist in the dissemination of the outputs of the TI3 workshop the ICT Strategic Advisory Team (SAT) were asked to comment on the outcomes presented in the previous section. The following part of this report represents our conclusions of the outcomes based on the advice of the ICT SAT, some of who were present at the workshop. It finally gives the scope for the forthcoming call and explains why this was reached.

Firstly, the key challenges that need to be addressed for the aims of the TI3 priority to be achieved.

Extracting Understanding from Data

Gaining value and understanding from the vast quantity of data collected in society today is a complex problem that spans many areas of research in the ICT portfolio. This calls for research which cuts across the various domains of informatics research to generate new methodologies for the processing and representation of data and information to yield greater understanding. The use of ontologies and other modes of linking and classifying data are identified as important.

One of the issues identified by the workshop in this area was that of the increasing diversity and complexity of data sources. The workshop also highlighted the importance of being able to convert the massive amounts of data produced into understandable, actionable information. There was a common thread across most of the challenges relating to this broad problem suggesting that solutions would require a deal of cross community working and engagement with both users, in both the sense of the general public and society as a whole (the users of data) and specifically the users of the research output.

Although not as explicit as one would have perhaps expected there was an indication of the link between '*Extracting Understating from Data'* and the '*Communications Bottleneck'*. This manifested itself in recognition that an ever increasing amount of data would require storage and transmission. Given this; decreasing the amount of traffic by turning data into information and information to understanding, and contextualising information prior to transmission both work towards "*Extracting Understanding from Data*" and are key to helping deal with the "*Communications Bottleneck*".

In addition to identifying this important and considerable challenge the workshop identified a need to increase awareness of the ethical and broader societal issues associated with research. This included the identification of a need for some form of repository of *safe* usable 'big' data to knowledge sources.

Communications Bottleneck

The increase in complexity of networks and the growth in the number of networked devices as we move towards the concept of the 'internet of things' makes a strong multidisciplinary approach necessary.

The workshop delegates identified a problem relating the design of converged networks (networks that consist of many different network types/architectures/services/transmitted data types working together) as being important. They also stated that in order to realise a convergence experimental networks and design tools would be required.

Another important point raised by the workshop discusses the understanding of the fundamentals of network complexity as being key to understanding future networks. Although no solution to this was proposed by the workshop delegates a possible solution may be collaboration with those from a more theoretical and mathematical backgrounds.

The issue of availability of spectrum (both radio and optical) and the increased demand for it was identified as bring another key research challenge. In its own right this challenge requires different disciplines to work together. The workshop highlights the need for game changing advances in this area, rather than the incremental and highlights the need for new hardware developments as well as new approaches and theories.

Another important observation of the workshop was the need for test-beds for networks and communications (both optical and RF). The workshop participants recognised that, again, multidisciplinary coordination was key and that equipment sharing and the involvement of organisations such as JANET was important. Linked to this was the importance of the availability of network traffic data from live networks.

Changes in Behaviour

As has become obvious from the previous two sections a quite prevalent feeling generated by the workshop was that a behavioural change in the research communities working toward an intelligent information infrastructure is required.

The bulk of the comments regarding change in behaviour are focused in a lack of collaborative activity between parts of the community that are more closely linked, hardware and software communities addressing different parts of the same applied technology for example. The need to encourage and enable this change in behaviour is recognised.

Other Messages

In addition to the key messages above the workshop highlighted a number of crosscutting societal, economic and energy based problems which are presented below.

> **Energy.** The importance of energy was raised in two ways by the workshop delegates; firstly in terms of economic cost (this includes embodied energy as well as running cost), and secondly the importance of low energy usage and where to place computational power to maximise energy efficiency. This problem applies to both of the technical challenges posed above***,*** "The Communications Bottle Neck" ***and*** "Extracting Understanding from Data"***.***
>
> **The Human Factor.** This centred on the general importance of understanding the needs of the user of the technology, whatever it may be. The workshop delegates also raised an interesting point regarding automation and 'the human in the loop'; when do we want automation and when do we not, and is this always considered. Broader ethical issues and privacy were also raised in relation to users. The issues raised in this area apply to both of the technical challenges posed above, "The Communications Bottle Neck" ***and*** "Extracting Understanding from Data"***.***
>
> **Skill Base.** A reference was made by the workshop to the importance generating and continuing to maintain a skills base; both in terms of post graduate and post-doctoral.

Scope of the Call

The workshop raised the importance of encouraging interdisciplinary activity and although the broader issue of fostering interaction between different research communities could be addressed by our existing networks and community collaboration there is a need for a targeted call in TI3 focusing on highly adventurous, cross disciplinary work that would not naturally take place through standard mode. As a result, and taking account of the technical issues identified, the following 'sub' challenges have been identified for the call:

Protocols for a 21st Century Internet. The internet has changed and continues to do so. The future will demand new and potentially intelligent protocols to deal with the increased complexity and volume of data and information that needs to be communicated and stored.

Energy Efficient Communication and Data Systems. The growth in IT infrastructure has led to it becoming one of the largest consumers of energy in the world. Information and Communication systems need to become energy efficient to remain sustainable.

Building Context and Content Aware networks. Future network hardware and software will need to meet the demands of an ever changing economy. We have moved from voice to text to streamed video, what will the future bring? A joined up approach will be required to build the communications networks of the future.

Extracting Understanding from Data. The need to communicate and store ever increasing amounts of data is putting a massive and ever increasing load on our Information and Communication systems infrastructure. The better we can understand and extract understanding from the data we have the more efficiently we can handle and communicate it.

Seamless mode adaptive communications. Communications systems are becoming ever more complicated and with the beginnings of the emergence of the cloud and the internet of things the future holds the promise of ever more complicated networks. In order for these complex systems of communication to operate effectively the links between its many parts will need to be seamless, adaptive.

The call will be focused on research relating to the sub challenges identified above. Although many of the sub-challenge areas are cross disciplinary in their own right the call will require applicants to address more than one of them in order to encourage the cross disciplinary working seen as necessary to make progress on the TI3 priority. In addition, the call will provide applicants an opportunity to address particular aspects of the sub-challenges which are high risk, 'far forward looking' and potentially game changing.

The total funding amount available for the call will be up to £5M and we expect to fund a number of projects from this total. The call for outlines will be announced in early May 2013 and will close mid June 2013.

The reader is encourage to remember that the call presented above represents a part of the route towards realising an Intelligent Information Infrastructure and should not expected to be the sole funding directed towards this cross ICT priority. There is still a strong need for other applications other mechanisms and schemes to address different aspects raised by the workshop but not addressed by the call.

If you have any questions or comments relating to the content of the report and the conclusions please contact.

Dr Matthew Lodge
Portfolio Manager, ICT Theme

EPSRC, Polaris House, North Star Avenue, Swindon, SN2 1ET

T: 01793 444 552
E: matthew.lodge@epsrc.ac.uk