SNS-BASED EPARTICIPATION AND CLOUD COMPUTING – A CONSIDERATION OF THE ISSUES RAISED

Richard Beales, Steve Taylor, Paul Walland[1]

Social Networking Systems provide a significant opportunity for governmental policy-makers by allowing them to interact directly with citizens, for example by stimulating new discussions or participating in conversations that are already underway so as to gauge public opinion on a proposal. Because Social Networks are already widely adopted, they provide potential for a much wider citizen base than specialist eParticipation platforms. The WeGov project is building a software toolkit to help policy-makers make effective use of SNSs by stimulating debates, identifying hot and emerging topics and picking out influential individuals or clusters of sentiment. The data storage and processing requirements of these features are significant, and third-party Cloud Computing presented itself as an option to meet them in a way that would be affordable to cash-strapped public-sector organisation. However despite the popularity of Cloud Computing services in the business IT world, concerns about data protection and privacy led us to conclude that political conversations harvested from SNS networks could not legally and ethically be entrusted to such services at the current time. This paper presents our analysis and offers some recommendations to Cloud Providers that we believe must be adopted if the potential of the technology as an economical platform for eParticipation is to be realised.

Introduction

Enterprise IT providers have aggressively promoted Cloud Computing as a transformative technology that lowers the cost of ownership of IT infrastructure while radically improving scalability, availability and agility. This is an attractive claim to anyone seeking to implement a Social Network System-based eParticipation platform, where typically there is intermittent need to harvest and process very large volumes of data. However eParticipation involves collecting politically and socially sensitive personal data; the use of third-party Cloud Computing services demands the release of such data to external organisations whose terms of service may not be adequate for processing sensitive personal data. In this paper we analyse the significant legal and ethical issues that must be addressed if Cloud Computing is to be used successfully for eParticipation systems, reflecting decisions we have made in WeGov, a collaborative project developing an eParticipation toolkit. And we present a series of recommendations that we believe Cloud providers and policy-makers must adopt if the

[1] IT Innovation Centre, Gamma House, Enterprise Road, Southampton, SO16 7NS, United Kingdom, {rmb,sjt,pww)@it-innovation.soton.ac.uk

potential of third-party Cloud Computing as an economical platform for eParticipation is to be realised.

SNS-based eParticipation

For policy-makers seeking to broaden the reach of their eParticipation activities, social networks like Facebook and Twitter offer an exciting opportunity to engage with a wide cross section of citizens on their own turf[1]. These Social Network Systems (SNSs) are used enthusiastically by young and old, male and female, professional and blue-collar alike. However the popularity of SNSs brings its own challenge to any organisation wishing to use them for mass political consultation or dialogue; effectively managing discussions involving hundreds or even thousands of citizens, and then analysing them to understand the opinions and sentiments expressed, requires the support of specialist software tools. Commercial SNS analytics packages like Radian 6, Crimson Hexagon and Mutual Mind are used by the marketing industry to promote brands, track customer opinion, and identify customers who have had negative product experiences or who are early adopters that might trial new products or services. Although at least one of these packages, Crimson Hexagon[2], has its roots in political science, they are all targeted squarely at advertising agencies and corporate marketing and publicity departments. The WeGov project is now developing an equivalent software toolkit for policy-makers, aiming to provide a manageable user interface through which policy-makers can seed, participate in and analyse policy discussion on multiple social networking platforms. The WeGov toolkit will allow policy-makers to initiate and follow SNS discussions, track discussions that relate to a particular policy or policy announcement, identify clusters of opinion, locate individual opinion-formers who might be well-placed to lead a further debate, and track popular sentiment as a discussion or political event unfolds. It will reflect the kinds of services already used for commercial brand management, but SNS management and analysis will be tailored specifically to support the working practices and analytics requirements of policy-makers: for example we are investigating how SNS analysis can be linked to individual parliamentary constituencies.

Handling the Data

The volumes of data involved in SNS analysis can be vast. Depending on the scale of consultation to be undertaken, it may be necessary to harvest and analyse tens of thousands of natural language social network messages. The infrastructure required to store and process this amount of data is significant, with associated costs: even with the economies of scale of commercial brand management, the leading SNS analytics packages charge subscription fees of several hundred dollars per month. If cash-strapped public sector organisations are to adopt the WeGov toolkit it must be as cheap to access as possible. Therefore when the WeGov project was formulated, rather than basing our software architecture around in-house infrastructure, it was our intention to exploit emerging Cloud Computing services. Cloud Computing offers a flexible, scalable model for on-demand provision of data processing and storage resource. Although Cloud Computing can be deployed as a private resource within a single organisation, the term more commonly refers to outsourced computing resource provided on a commercial basis by third-party organisations. As such, it represents an attractive alternative to costly up-front purchase of IT infrastructure that might only rarely be used to its full capacity. However, the success of any eParticipation activity is heavily dependent on user perception. Policy-makers must have confidence that deploying the toolkit

will not provoke a popular backlash or breach privacy or data protection legislation or ethical guidelines. In the following discussion, we describe issues that led us to conclude it was not currently practical to use third-party Cloud Computing as the basis of our own toolkit; we believe consideration of these issues is essential for anyone seeking to deploy eParticipation applications on 'the Cloud'.

Governmental and Media Concerns over Cloud

In December 2009, at the end of the year in which cloud computing really took off in business, Social Computing Journal predicted 2010 would be the year **"**Cloud computing will go big" for eGovernment[3]. Since then, serious concerns have been raised in the IT and popular press regarding the use of commercial Cloud Computing services to handle personal or sensitive data[4][5][6]. Amid calls from European leaders for a global data protection law[7], the European Network and Information Agency (ENISA) has warned government agencies against deploying applications that process sensitive data to external Cloud providers[8]. In the UK, the charitable fundraising sector, which (like eParticipation) involves profiling large numbers of private individuals, has also cautioned against using commercial Cloud systems[9]. Though the UK Government has announced the establishment of a 'G-Cloud' to host a wide range of Government services, it has stated this will be a private, closed facility, hosted within the UK. The Australian government has delayed its timetable for moving services to public Cloud computing, with the movement of private citizens' data taking place last of all, and not for several years[10]. So what is the basis of reported concerns regarding Cloud Computing? The ENISA analysis indicated seven areas of risk: loss of governance, provider lock-in, isolation failure, compliance, management interface compromise, data protection and insecure or incomplete data deletion. Several of these have been echoed by business leaders and IT consultants: the IT Governance Institute reported almost half of the C-level executives they surveyed cited data privacy worries[11], and the Cloud Security Alliance has highlighted the lack of a clear regulatory environment for Cloud Computing providers[12]. These challenges are exacerbated when Cloud providers are located in different jurisdictions to their customers, or move data between jurisdictions without notice.

The Legal and Ethical Issues inherent in eParticipation

Against this background of press and governmental anxiety, the University of Southampton's internet law group, ILAWS, investigated the legal and ethical issues inherent in an SNS-based eParticipation application, including an assessment of their implications for Cloud deployment. Mindful of the need that an eParticipation application is *seen and believed* to comply with ethical and legal best practice, its recommendation was to exercise caution. The analysis focussed on three potential problem areas - data protection and privacy, copyright and intellectual property, and defamation – but as the ENISA report has already indicated, it was in the area of data protection and privacy where issues particularly relevant to Cloud Computing were identified.

At first glance, privacy might not seem to be a major consideration when dealing with SNS postings. The primary purpose of a SNS is, after all, facilitating *sharing* of messages, content or status. Indeed this view was reinforced by a recent ruling of the UK's Press Complaints Commission (PCC), widely reported in the UK and international press. The PCC was asked

by the complainant to consider the right of a newspaper to republish their Twitter posts. The PCC ruled that Twitter posts should be considered public, but this ruling was, logically enough, made in response to a *complaint* – the individual concerned was angered that their Tweets were republished by a mass-circulation newspaper without their permission. Ironically, the complainant was a UK Government civil servant, who had previously blogged in favour of government using SNSs to engage with the public[13], but clearly it would be catastrophic for an eParticipation application to become mired in a similar controversy. This reflects the complex attitudes to privacy that surround what ostensibly is a very public medium. Privacy expectations vary between SNS platforms, and between different user areas of the same platform - consider, for example, the difference between a post by a citizen on a commercial organisation's public Facebook wall, and a post to their own private wall by one of their Facebook friends. In fact ILAWS indicated three reasons why SNS posts are not without privacy concerns:

- Lack of awareness by SNS users of how privacy controls operate, particularly exacerbated when default privacy settings are changed by the SNS after its launch, as happened with Facebook in 2009[14];
- The legal and ethical sensitivity of political opinions and the need for any eParticipation tool to adopt a defensive legal position;
- Social expectation, for example where the use of pseudonym usernames indicates an expectation of, or a desire for anonymity.

The implication of this assessment is that while information might be sourced from publicly-accessible areas of an SNS, the *eParticipation application should handle the information as if it were private and confidential*. Repositories of harvested SNS posts and profiles, the channels through which the content is moved between different software components of the eParticipation application, and the individual software components themselves, must all therefore be operated securely and in accordance with data protection legislation, discussed next.

In the European Economic Area (EEA), the EC Data Protection Directive applies regardless of whether the user has an expectation of privacy; its scope is *personal* data, and this is simply data that "*relates to an identified or identifiable natural person*". The Data Protection Directive prohibits processing of personal data except when specific conditions are met relating to transparency, legitimate purpose and proportionality. The obligations are particularly stringent for *sensitive* personal data, i.e. that revealing an individual's racial or ethnic origin, political opinions, religious or philosophical beliefs, or containing information about their health or sex life. The political data is the very stuff that policy-makers seek to understand, but the policy-maker might find themselves handling sensitive personal data even where it is not explicitly sought; an individual's SNS username, for example, might strongly indicate their ethnicity.

Due to differences in the transposition of EC directives into national law, even within the European Union, the precise legal situation varies from one member state to another. The European Data Protection Supervisor has recommended replacement of the current Data Protection Directive with a self-executing Regulation that would be immediately enforceable across all member states [15], but it is not yet clear if and when this will happen. In the meantime, we have based our analysis on the Directive as currently transposed into UK law,

as the UK Data Protection Act 1998. This contains eight data protection principles, in summary:

- Fair and lawful processing
- Personal data obtained for fair and lawful purposes
- Personal data shall be adequate, relevant and not excessive in relation to the purpose(s) for which they are processed
- Personal data shall be accurate and kept up-to-date
- Personal data processed for any purpose shall not be kept longer than is necessary for that purpose
- Processing in accordance with the data subjects' rights
- Implementation of adequate security measures, including against accidental loss of or damage to personal data
- Transfer of personal data outside the EEA prohibited unless the destination country ensures adequate protection of rights of data subjects in relation to processing of personal data

We can see that consequent on data protection legislation are obligations not only to maintain data security, but also relating to data integrity, retention and movement. It is necessary to ensure that any harvested SNS profiles are up-to-date, and that old versions do not linger in storage. Profiles and messages must only be retained for as long as is necessary for the specific eParticipation exercise for which they were harvested, which means there needs to be an effective mechanism to irreversibly delete this data as soon as it is no longer required. And the physical location of the data must be known at all times, whether it is being actively processed, or sitting in storage. With whom do these obligations lie, and how do they impact on the viability of using third-party Cloud computing?

The Challenges for Cloud

Let us first consider the issue of responsibility. Ethically, it is clear-cut. The policy-maker must take responsibility for the security and proper treatment of the data they harvest. It is essential that policy-makers retain goodwill and confidence of citizens (and in the UK at least, it is front-page news when government is careless with citizens' personal data). The policy maker will also be stimulating the debate in many cases, actively encouraging the posts that they subsequently harvest. In order to understand with whom legal responsibility for meeting data protection requirements lies, it is necessary to examine the distinction between *Data Controller* and *Data Processor,* as different data protection obligations are placed on each role under European Law. According to the EC Data Protection Directive, the Data Controller is defined as "*the natural or legal person, public authority, agency or any other body which alone or jointly with others* ***determines the purposes and means of processing personal data***". Data Processing is defined much more broadly, as "*any operation or set of operations which is performed upon personal data whether or not by automatic means, such as collection, recording, organisation, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction*". SNS providers themselves are data controllers as they determine purposes and means of processing of information published and exchanged by their users. On the other hand, the eParticipation application would itself be

considered a data processor, not a data controller; here the responsibilities of data controller would lie with the policy-maker using the eParticipation application to post to and mine SNSs.

Ethically and in law it would therefore be the responsibility of the policy-maker to ensure that the Cloud provider has suitable measures in place to meet the requirements of privacy and data protection legislation and that these are included in the contract with the provider. Unfortunately many Cloud providers disclaim responsibility, stating that it is up to the customer to meet these requirements. We have already discussed the security concerns surrounding Cloud computing. Hitachi Data Systems' CTO Security and Privacy, Eric Hibbard, has referred to "*data droppings*" in the Cloud, saying "*data retention and media sanitization are unpredictable*" [17]. This is counter to the data protection requirement that data not be kept for longer than necessary. Hibbard also stated that "*the integrity and authenticity of data* [held in the Cloud] *may be questionable*" – this too is counter to one of the tenets of data protection, that data shall be accurate and up-to-date. The mantra of 'in the Cloud' suggests we shouldn't care where our processing or storage takes place, but data protection regulations applicable to SNS content mean the opposite is true – it is vital to know in which territory and jurisdiction data processing and storage are located. Unless there is a contractual guarantee that the cloud provider will not transfer data they receive to another cloud service outside the EEA, the policy-maker may find themselves unwittingly in breach of data protection legislation even when having contracted an EEA service provider. We note that the EC's Safe Harbour programme does permit export of personal data to the US provided the processor to which it is exported has Safe Harbour status, but organisations are permitted to self-certify and self-regulate their compliance – as a result, the German data protection authorities advise that certification cannot be relied on [18]. With Cloud service providers "*sometimes hesitant about disclosing locations or sub-contractors*" [19], again it would be difficult for policy-makers to be sure they were Data Protection compliant.

Why Anonimisation isn't the answer

If we cannot confidently entrust unprocessed sensitive data to Cloud services, perhaps anonymisation of citizens' SNS profiles and messages might offer a solution: if these profiles and messages could be modified such that the individuals to whom they related were no longer "identified" or "identifiable", they would no longer be considered personal data under the terms of the Data Protection Directive. We have considered the viability of anonymisation of SNS data at a private dedicated server prior to uploading it to a third-party Cloud service for further processing and storage. Unfortunately there are several reasons why we don't believe this to be practical:

- Removing information will severely reduce the value of the information – in many cases, rendering it useless.
- It extremely difficult to determine what needs to be deleted to ensure that SNS data is guaranteed to be not personal. For example, the data may contain arbitrary unstructured information that contains personal information. Therefore it is almost impossible to guarantee that the SNS data is not personal.
- Anonymisation cannot be done at the SNS sites themselves as SNSs do not offer anonymised data via their APIs, so the anonymisation service itself would be subject to all the legal and ethical concerns we have outlined in this paper.

Furthermore, we cannot guarantee that aggregation of multiple seemingly innocuous pieces of information will not result in a combined profile from which a single individual is clearly identifiable. Many might hold strong opinions on nuclear power or Middle East policy, and tens of people might care about the siting of a particular mobile phone mast, but it isn't too difficult to imagine that one individual in a community is known for being vocal on all three issues. Indeed networked data of this kind is notoriously difficult to anonymise effectively; there is ongoing research, for example examining the effectiveness of graph structure perturbation [20] (a technique that results in information loss), but Microsoft's Chief Privacy Advisor for EMEA, Caspar Bowden, is among those who have argued that effective anonymisation of 'networked' information simply isn't possible given advances in Computer Science [21]. Although beyond the scope of both this paper and our work in the WeGov project, perhaps a more viable alternative is selective encryption of personal data in such a way that it can still be processed remotely [21].

Recommendations to Cloud Providers

On the basis of our analysis, we make the following recommendations to Cloud service providers, which we believe will allow policy-makers and others dealing with sensitive personal data to use third-party Cloud processing and storage with greater confidence:

- Conduct and make public the results of independent security and data protection audits;
- Provide clear checklists of data protection law compliance;
- Engage with the EC and other regulatory bodies to develop a common regulatory framework for Cloud service provision that will include those operators outside the EEA and working under Safe Harbour status;
- Provide details of their data deletion policies, including maximum time guarantees between the customer requesting deletion of an item of data and all copies of that data being deleted irretrievably and verifiably;
- Clearly specify the physical locations and jurisdictions of all processing and storage within their service level agreements;
- Explicitly detail any sub-contracting arrangements that entail customer data being moved to other service providers;
- Allow Cloud service customers to opt-out of any subcontracting arrangements, even if this means charging a financial premium for a geo-constrained service.

Conclusions

So where does this leave our work with the WeGov toolkit, and what is our advice to others considering developing SNS-based eParticipation applications or looking to use Cloud technology to support other eParticipation activities? We have determined we cannot deploy the WeGov toolkit on third-party Cloud Computing services (as they are currently defined as cheap, commodity, "locationless" services); we are not confident that any such service currently addresses the legal and ethical issues identified in this paper, so in the short term will instead employ our own servers and storage. Having adopted a modular 'software-as-a-service' architecture, it will be relatively straightforward to move CPU- or storage-heavy

processes to a third-party Cloud service if and when a provider emerges that complies with the recommendations made above. What is also clear is that data protection compliance is not cheap, so the commodity pricing of existing Cloud processing services will not apply with a dedicated data protection compliant hosting provider. In the meantime, those policy-makers with sufficient resources might wish to follow the example of the UK Government, and make use of private, in-house Cloud capability to host their eParticipation applications or alternatively seek dedicated relationships with external, private, data protection compliant, hosting companies despite the cost overhead that this involves.

References

[1] Øystein Sæbø, Jeremy Rose, and Judith Molka-Danielsen. 2010. eParticipation: Designing and Managing Political Discussion Forums. Soc. Sci. Comput. Rev. 28, 4 (November 2010), 403-426.

[2] Daniel Hopkins, and Gary King. A Method of Automated Nonparametric Content Analysis for Social Science. In American Journal of Political Science 54 (2010): 229-247.

[3] http://socialcomputingjournal.com/viewcolumn.cfm?colid=868

[4] http://www.fiercecio.com/story/report-execs-privacy-worries-impede-cloud-migration/2011-01-23

[5] http://www.computerweekly.com/Articles/2010/01/12/235782/Top-five-cloud-computing-security-issues.htm

[6] http://news.bbc.co.uk/1/hi/programmes/click_online/8625625.stm

[7] http://www.computerweekly.com/Articles/2010/03/26/240731/Cloud-security-weaknesses-prompt-call-for-global-data-protection.htm

[8] http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment

[9] http://www.fundraising.co.uk/blog/2011/01/18/cloud-computing-sensitive-data-and-data-protection

[10] http://www.computerweekly.com/Articles/2011/01/07/244796/Security-concerns-to-delay-Australian-government-move-to-public.htm

[11] http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/Global-Status-Report-on-the-Governance-of-Enterprise-IT-GEIT-2011.aspx

[12] http://www.computerweekly.com/Articles/2010/03/26/240731/Cloud-security-weaknesses-prompt-call-for-global-data-protection.htm

[13] http://baskersworld.wordpress.com/some-thoughts/

[14] http://www.guardian.co.uk/technology/2009/dec/10/facebook-privacy

[15] http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2011/11-01-14_Personal_Data_Protection_EN.pdf

[16] http://dev.twitter.com/pages/api_terms

[17] http://www.snia.org/events/wintersymp2009/cloud/Cloud-Computing-Legal-Angles.090120.pdf

[18] http://www.bnai.com/GermanyDpas/default.aspx

[19] http://www.readwriteweb.com/cloud/2010/09/tension-are-increasing-in-euro.php

[20] M. Hay, G. Miklau, D. Jensen, D. Towsley, and P. Weis. Resisting Structural Re-Identification in Anonymized Social Networks. In VLDB, 2008.

[21] http://www.silicon.com/technology/security/2011/03/23/anonymity-loophole-puts-uk-data-at-risk-39747185/2/

[22] http://cloudcomputing.sys-con.com/node/1456822