Chapter 3
A Privacy Impact Assessment Tool for Cloud Computing

David Tancock, Siani Pearson, and Andrew Charlesworth

Abstract In this chapter, we consider requirements for Privacy Impact Assessments (PIAs) carried out within a cloud computing environment and explain how a PIA support tool may be constructed. Privacy is an important consideration in cloud computing, as actual or perceived privacy weaknesses will impact legal compliance, data security, and user trust. A PIA is a systematic process for evaluating the possible future effects that a particular activity or proposal may have on an individual's privacy. It focuses on understanding the system, initiative, or scheme; identifying and mitigating adverse privacy impacts; and informing decision-makers who must decide whether the project should proceed and in what form (Stewart B, Privacy impact assessments. PLPR 3(7):61–64, 1996. http://www.austrii.edu/au/journals/PLPR.html. Accessed 30 October 2011).

Keywords Cloud • Privacy • Privacy Impact Assessment • Regulation

3.1 Introduction

A Privacy Impact Assessment (PIA) [1] is a systematic process for identifying and addressing privacy issues in an information system that considers the future consequences for privacy of a proposed action [2]. It is thus, in part, a predictive exercise

S. Pearson
Cloud and Security Lab, HP Labs, Long Down Avenue, Bristol BS34 8QZ, UK
e-mail: Siani.Pearson@hp.com

A. Charlesworth
Centre for IT and Law, University of Bristol, Queens Road, Bristol BS8 1RJ, UK
e-mail: csdjt@bristol.ac.uk; a.j.charlesworth@bris.ac.uk

D. Tancock (✉)
Department of Computer Science, University of Bristol,
Merchant Venturers Building, Woodland Road, Clifton, Bristol BS8 1UB, UK
e-mail: csdjt@bristol.ac.uk

S. Pearson and G. Yee (eds.), *Privacy and Security for Cloud Computing*,
Computer Communications and Networks, DOI 10.1007/978-1-4471-4189-1_3,

designed to prevent or minimise adverse privacy outcomes. Typically, PIAs usually take the form of a series of steps, posing and answering questions and considering options, although they can also be more holistic in nature. In some jurisdictions, an expected deliverable of the PIA process is a document, such as a PIA report [3]. PIAs are primarily a proactive process, whereas other related business processes such as privacy issue analysis, privacy audits, or privacy law compliance checking can be proactive and reactive. For example, a privacy audit can be done in a proactive manner as part of an organisation's attempt to protect private data without it being required by an outside agency or it can be done in a reactive manner by scrutinising existing projects to ensure their continuing conformity with internal rules and external requirements [4]. A PIA permits organisations to design privacy into new systems during the design and development stages, reducing the risk that costly retrofitting of privacy safeguards will be required after implementation.

While a PIA may be perceived primarily as a management tool (i.e. as a threat/risk assessment process), it can be used as a tool for enhancing individual privacy. By surfacing privacy issues at an early stage, and providing system designers with relevant knowledge, as well as the impetus to tackle those issues at the architectural level, PIAs can facilitate the raising of a system's privacy baseline without undue impact on its functionality [5].

Privacy rights are protected and advanced by convincing agencies and businesses to carry out a PIA for the following reasons: to demonstrate legal compliance, to allow organisations to develop better policies, to save money, to develop a culture of privacy protection, to prevent adverse publicity, and to mitigate risks in advance of resource allocation. In the case of cloud computing, the goal of enhancing end user trust by decreasing the risk of exposure of end user's information is particularly important because there is a perceived lack of consumer trust with respect to cloud scenarios specifically where sensitive information is involved.

This chapter considers the possibility of developing a PIA decision support tool for a cloud environment. The structure of this chapter is organised as follows. In Sect. 3.2, we provide some background information on PIAs within major jurisdictions. Section 3.3 considers the problems and issues of privacy and security in the cloud and discusses the challenges of deploying a PIA tool for this environment, which provides the motivation for our approach. In Sect. 3.4, we present details of a PIA tool for cloud environments, outlining what the tool does, how the tool works, and its architecture. In Sect. 3.5, we discuss and present details of the methodology used for our PIA tool including the software development methodology, data collection, analysis, results, and modelling. In Sect. 3.6, we cover related work previously carried out within the context of privacy and security in cloud computing and evaluate whether elements of these approaches are suitable for the proposed tool. Section 3.7 considers the planned next steps for the proposed tool. In Sect. 3.8, we briefly provide conclusions.

3.2 Background

In this section, we discuss PIA processes in different jurisdictions and provide examples of PIAs that have been recently undertaken by government agencies and private organisations.

Our analysis of the various guidance materials indicates that PIAs vary across jurisdictions – sometimes substantially – and that there are many interrelated dimensions. The following subsections describe five major dimensions that we have identified.

3.2.1 The Level of Prescription

The first dimension that affects the type of PIA relates to levels of prescription within different jurisdictions. The requirements for conducting PIAs within different jurisdictions are by "legislation" (e.g. required by law), prescribed by binding "policy" or "recommended" by those with no legal authority (e.g. privacy commissioners), and the landscape can be very complex.

For example, in the Canadian province of Ontario, all three levels of prescription exist for PIAs [6]. Section 6 of the Regulation to the "Personal Health Information Protection Act" (PHIPA) mandates PIAs for Health Information Network Providers (HINP), when two or more Heath Information Custodians (HIC) use electronic means to disclose Personal Health Information (PHI) to one another [7]. In this respect, the legislative and policy drivers for this come from the government. Furthermore, PIAs are required by policy at the detailed design phase or when requesting funding approval for product acquisition or system development work, where those projects involve changes in the management of personal information held by government programmes or otherwise affect client privacy.

The Ontario PIA process is very much seen as part of, or complimentary to, the mandated threat risk assessment process and is designed primarily to aid management decision-making processes. Moreover, it is the responsibility of the information and privacy commissioner to ensure that government and health-care practitioners and organisations abide by the FIPPA and MFIPPA Acts [8]. The commissioner also provides policy advice and training in the areas of freedom of information (FOI) and privacy including PIAs.

In addition, since the "Data Handling Procedures in Government" report published in June 2008 [9], PIAs in the United Kingdom (UK) are mandatory from all government departments that introduce new policy or processes that involve the use of personal data. Thus, all UK government departments will introduce PIAs to ensure that privacy issues are factored into plans from the start and check that they have been carried out as an integral part of the risk management assessment process.

Our analysis also identifies that organisations can conduct PIAs in the absence of any level of prescription (i.e. required by law, prescribed by binding policy, or recommended by those with no legal authority) and instead are based upon self-regulation. The motivations for conducting self-regulation PIAs are based upon the perception of the benefits. For example, private sector organisations typically conduct self-regulated PIAs when they are concerned about reputation.

The next section considers the application of PIAs in private and public sectors, which again affects the type of PIA used.

Table 3.1 Statistical report of Canada's PIAs and PPIAs

Privacy Impact Assessments	Amount
Number of PIAs initiated	172
Number of PIAs completed	89
Number of PIAs forwarded to the Office of the Privacy Commissioner of Canada	78
Preliminary Privacy Impact Assessments	**Amount**
Number of PPIAs initiated	104
Number of PPIAs completed	99

3.2.2 Application of PIAs in Private and Public Sectors

In this section, we discuss the application of PIAs in private and public sectors. In jurisdictions in which PIAs are being currently applied, there is a longer history of regulation within organisations in the public sector than in the private sector. For example, in Canada, New Zealand (NZ), Australia, and the United States (US), public sector privacy legislation has generally predated that for the private sector. Therefore, most PIA requirements apply to public sector organisations such as government ministries or departments and types of public bodies or agencies. However, it is increasingly difficult to determine the limits of the public sector PIAs under current conditions. This is because many public agencies that are outside government ministries now have extensive experience with PIAs. This includes organisations in the health sector, higher education, and statistical agencies. Although there is evidence that PIAs are conducted within the private sector (e.g. self-regulation), we do not know the extent of this in the absence of a mandate. However, private sector organisations have been mentioned by oversight bodies (e.g. privacy commissioners) and central agencies (e.g. Treasury Board of Canada) in relation to conducting PIAs in high-risk situations or initiatives [4]. For example, the Treasury Board Secretariat (TBS) of Canada states in a report of 2010 that 276 PIAs and a short form of PIA called Preliminary Privacy Impact Assessments (PPIAs) were initiated, of which 188 were completed, as illustrated in Table 3.1 [10].

The UK Information Commissioner's Office (ICO) also states in its "Annual Review" of 2010 that "over 300 PIAs have been started across central government and their agencies" [11].

As discussed in Sect. 3.2.4, the PIAs involved in this process include both full-scale PIAs (i.e. those that conduct a more in-depth internal assessment of privacy risks and liabilities) and small-scale PIAs (i.e. those that are less formalised and require less exhaustive information gathering and analysis) [4]. Thus, some examples of PIAs that have been conducted in the UK are outlined in Table 3.2.

However, as illustrated in Table 3.3, some organisations in the UK employ external consultants to carry out a PIA either because they do not possess the necessary skills in-house or because they wish the PIA to be perceived as being as independent as possible from potential influences within the organisation.

Table 3.2 Examples of PIAs conducted in the UK

Organisation	Year of publication	Project/procedure assessed	Type of PIA
Individual electoral registration	2011	Introduction of new policy to help rebuild public confidence in the security of electoral registration	Full scale
UK Anti-Doping	2010	The disclosure of personal data to UK Anti-Doping by the Serious Organised Crime Agency	Small scale
Northern Ireland Statistics and Research Agency (NISRA)	2010	2011 census for Northern Ireland	Full scale
Office for National Statistics (ONS)	2009	2011 census for England and Wales	Full scale
UK Border Agency	2009	Exchange of fingerprint information with immigration authorities in Australia, Canada, United States, and New Zealand	Small scale
National Policing Improvement Agency	2009	Electronic exchange of police intelligence across England and Wales via the Police National database	Full scale

Table 3.3 Examples of PIAs outsourced in the UK

Organisation	Type of privacy impact accessed	Consultancy employed
Aegate (Pharmaceutical authentication services)	Use of RFID technologies to authenticate prescription pharmaceuticals at point of sale	Enterprise Privacy Group
Department for Transport	National time-distance-place road pricing policy. This charges vehicles based on when, where, and how much they drive	Enterprise Privacy Group
Phorm Inc	Behavioural targeted advertising	80/20 Thinking Ltd

Analysis of the outsourced PIAs suggests that in traditional approaches such as an internal distributed network, external consultants (e.g. independent experts, regulators, civil society groups, professional bodies and charities) often bring considerable experience to the PIA process, lending impartially to the process. However, the experiences found in the UK concerning difficulties in organisations conducting PIAs seem to be replicated in most of the jurisdictions studied. These include internal stakeholder resistance such as project managers who often perceived PIAs to be a burden and public relations managers who were wary of engagement with external stakeholders.

In addition, security officers sometimes considered PIAs to be a threat to their expertise, and consequently, employees in that position in the organisation or acting as external stakeholders may often be reluctant to engage with an organisation conducting a PIA. This is through either lack of interest, lack of trust, or lack of resources [2].

Moreover, an exercise such as that conducted by 80/20 Thinking Ltd cannot be accurately described as a PIA, given that the technology and its applications were already fully developed and in use in business operations at the time of the PIA exercise. These exercises might be more accurately characterised as a privacy audit or compliance check [12].

In the next section, we consider the conditions and circumstances for conducting PIAs in the jurisdictions.

3.2.3 Initial Screening

There is variance in the mechanisms for determining the conditions and circumstances for conducting PIAs. Some jurisdictions have developed screening tools to help organisations to determine whether or not to conduct a PIA for any given initiative or to identify privacy issues that may require further analysis.

Commonly, an initial screening exercise is conducted to determine if a PIA should be completed according to the rules or recommendations in the jurisdiction. This can be as simple as determining whether personal information is involved or take the form of a structured instrument that poses a series of questions, as in NZ [13] and the UK [14].

The US screening process is a form called a Privacy Threshold Analysis [15]. Those completing the form provide a variety of information about the system, answering specific questions tailored to their operational context, and the Privacy Office makes an assessment that determines whether or not a PIA is required.

In contrast, within Canada, a Preliminary PIA (PPIA) is similar to a screening tool [8].

In the next section, we discuss the scale of the PIA processes that are conducted in all jurisdictions, as this can vary considerably.

3.2.4 The Scale of the PIA Process

Generally, there are two different types of PIAs conducted in all jurisdictions although the names and the processes vary. For example, names attributed to a short form of PIA are “small-scale” (e.g. UK), “PPIA” (e.g. Canada), and “Privacy Scan” or “Privacy Impact Statement” in other jurisdictions. The short form of PIA is similar to a full-scale PIA but is less formalised and requires less exhaustive information gathering and analysis, usually focusing on specific aspects of a project [4]. A full-scale PIA conducts a more in-depth internal assessment of privacy risks and liabilities. It analyses privacy risks, consults widely with stakeholders on privacy concerns, and brings forward solutions to accept, mitigate, or avoid such concerns. The process guidelines for a full-scale PIA tend to be more comprehensive and suggest the various stages of the process. For example in Australia, the process for conducting a PIA consists of five stages [16]: project description, mapping the information flow, privacy impact analysis, privacy management, and recommendations.

In contrast, the Ontario PIA process consists of three main stages: conceptual analysis, data flow analysis, and follow-up analysis. However, the Ontario PIA process ensures client privacy is considered throughout the business redesign or project development cycle, particularly at the conceptual stage, the final design approval and funding stage, the implementation and communications stage, and at the post-implementation audit or review stage [8].

In the UK, the processes involved in conducting a PIA are again similar to PIAs conducted in other jurisdictions and consist of the following [14]:

- *Initial assessment*: Examines the project at an early stage, identifies stakeholders, assesses privacy risks, and decides whether a PIA is necessary or not and if so, what level of PIA is required.
- *Small-scale PIA*: This is less formalised and requires less exhaustive information gathering and analysis and usually focuses on specific aspects of a project.
- *Full-scale PIA*: This consists of five phases that are usually conducted in sequence and include the following [14]:
    - *Preliminary*: Establishes and ensures a firm basis for the PIA, so that it can be conducted effectively and efficiently
    - *Preparation*: Makes the arrangements needed to enable the following phase (i.e. consultation and analysis) to run smoothly
    - *Consultation and analysis*: Identifies problems early on, discovers effective solutions, and ensures that the design is adapted to include those solutions
    - *Documentation*: Documents the PIA process and the outcomes and delivers a PIA report
    - *Review and audit*: Ensures that the undertakings arising from the consultation and analysis phase are actually within the running system or implemented project
- *Privacy law compliance check*: Examines compliance with statutory powers, duties, and prohibitions in relation to the use and disclosure of personal information.
- *Data protection compliance check*: Examines compliance with the Data Protection Act of 1998. An organisation usually conducts this check when the project is more fully formed.

In the next section, we consider the people involved in conducting PIAs in all jurisdictions.

3.2.5 Who Conducts PIAs

PIAs are usually completed by a senior analyst or a manager with ongoing programme administration responsibilities. The various guidance material suggests a team or committee approach and stipulates what types of expertise should be drawn in to the PIA. This can include, with varying degrees of participation, the following personnel [2]: programme and project managers, privacy policy makers, legal advisors, records management staff, information technology or data security experts, communications staff, and other functional specialists.

3.2.6 Current PIA Tools

As considered above in Sect. 3.2, the processes involved in conducting PIAs across jurisdictions sometimes vary substantially. One important difference is in the PIA tools that each jurisdiction uses. For example, in Canada, the TBS provides an e-learning tool for government employees interested in learning more about privacy and PIAs and how to complete them. The e-learning tool consists of two courses (e.g. Overview and Manage/Monitor) and a PIA assistant to help users complete PPIAs and full PIAs [17].

In contrast, the US Department of Homeland Security (DHS) employs a PIA tool called the Privacy Threshold Analysis that helps users determine whether a PIA is required under the E-Government Act of 2002 and the Homeland Security Act 2002 [18]. In the UK, the PIA Guidelines provide a number of screening questions to help users decide whether a full-scale PIA or a small-scale PIA is warranted. The Guidelines also include a number of questions for a privacy law compliance check and a Data Protection Act (1998) compliance check. Templates are also included within the Guidelines for Data Protection compliance and the Privacy and Electronic Communications Regulations (PECR) [14].

The evaluation processes involved in these PIA tools consist of simple questionnaires, whereby most of the questions require a “yes” or “no” response. Analysis of the PIA tools suggests that they are mainly based upon a simple “decision-tree” approach. This approach is commonly used for simple reasoning, as it is both a knowledge representation scheme and a method of reasoning about that knowledge. In addition, the PIA tools produced by the different jurisdictions are mainly procedure-based (e.g. whereby a number of specified steps are used to reach the desired outcomes), and their granularity is coarse-grained (e.g. consist of fewer larger components). Finally, the PIA tools are Web applications where both data and the applications are at the server-side; therefore, they do not take into account the cloud or any of its characteristics (e.g. on-demand self-service, ubiquitous network access, location-independent resource planning, rapid elasticity, and pay for use).

Furthermore, we contend that deploying a PIA tool (i.e. a tool that is based upon questionnaires in which answers provided by the user addresses the complexity of privacy compliance requirements by highlighting privacy risks and compliance issues) can lead to negative perceptions by organisations and end users including [19]:

- Some organisations find it very difficult to relinquish control or trust third parties to manage their applications and data.

- Some organisations are worried about security and weak data protection in cloud applications.
- Some markets require industry-specific business applications (e.g. military systems) for which solutions such as the software as a service (SaaS) solution are not available.
- Organisations without clear objectives and defined business processes are sometimes no better off with a cloud solution than with an on-premise solution.

3.2.7 Future PIAs

We have seen above that a number of PIAs have been carried out in various jurisdictions and that pressure is mounting from regulators for this approach to be used more widely.

Privacy rights can be protected and advanced by convincing agencies and businesses to carry out a PIA for the following reasons: to demonstrate legal compliance, to allow organisations to develop better policy, to save money, to develop a culture of privacy protection, to prevent adverse publicity, and to mitigate risks in advance or resource allocation.

However, as business becomes more global and moves to the cloud, it will become increasingly difficult to carry out the analysis needed and so more help will be required from a technical standpoint.

Moving PIAs onto the cloud and potentially across and between legal jurisdictions, including processes that are outsourced and data that crosses organisational boundaries, increases risk factors and legal complexity. Therefore, in the next section, we discuss the problems and issues of privacy and security in the cloud and explain further the challenges of deploying a PIA tool for this environment, before considering solutions later in this chapter.

3.3 Issues for Privacy, Security, and PIAs in the Cloud

In this section, we consider the problems and issues of privacy and security in the cloud and discuss the challenges of deploying a PIA tool in this environment, which provides the motivation for our approach.

As discussed in Chap. 1, there are a number of privacy, security, and trust issues associated with the cloud including [20] lack of user control, potential unauthorised secondary usage, data proliferation, transborder data flow and dynamic provisioning, access, availability, backup, multi-tendency, and lack of standardisation. Of these issues, data proliferation, transborder data flow, dynamic provisioning, and virtualisation are very important to PIAs in the cloud. For example, data proliferation is a feature of cloud and this happens in a way that may involve the PIA tool being accessed by multiple customers from different organisations that reside in different jurisdictions, whereby data is not controlled by the data owners. However, Cloud

Service Providers (CSP) ensure availability by replicating data in multiple data centres; therefore, it is difficult to guarantee that a copy of the PIA tool and its data or its backups are not stored or processed in a certain jurisdiction or that all these copies of the PIA tool and its data are deleted if such a request is made. This is because customers of the PIA tool cannot be sure that the PIA tool and its data is in one jurisdiction or that copies that are deleted are really deleted and are not recoverable by a CSP, as currently there are no ways to prove this as it relies on trust.

Furthermore, the movement of data, governance, and accountability of the PIA tool becomes more complex when it moves onto the cloud and potentially across and between legal jurisdictions. This is because processes may be outsourced and knowing the jurisdictions involved can be quite difficult. Moreover, transferring data stored in the cloud to other jurisdictions may violate local laws, because of the difficulty of asserting which specific server or storage device is used, due to the dynamic nature of the cloud [20].

Virtualisation introduces similar concerns due to the separation of the logical entities being assessed from the underlying physical resources. Thus, virtual machines (VMs) are environments that are completely isolated from each other. Although virtualisation makes it safe for users to share the same hardware, the underlying physical resources are the responsibility of the CSP. However, these environments can sometime break down, allowing attackers to escape the boundaries of this environment and have full access to the host. Therefore, organisations should maintain their security based on sound security practices including keeping software up to date with security patches, using secure configuration baselines, and using host-based firewalls, antivirus software, or other appropriate mechanisms to detect and stop attacks.

However, we believe that a cloud-based PIA tool is a novel approach. This is because at the time of writing (e.g. February 2012), no such tool exists. Thus, the PIA tool can provide significant value in increasing trust as a commercial service, in spite of the number of challenges it faces in deployment. This is because we contend that the PIA tool can be accessed in the same way the cloud is delivered: “as a service”. Indeed, the same five characteristics of the cloud (i.e. on-demand self-service, ubiquitous network access, location-independent resource pooling, rapid elasticity, and pay per use) that are used to deploy and access existing applications and tools may be used to deploy and access the PIA tool especially the metering that is already built in for billing and service-level assurance.

In the next section, we discuss our approach for a cloud-based PIA tool.

3.4 Development of a PIA Tool for Cloud Computing

In this section, we present details of a PIA tool for cloud environments, outlining what the tool does, how the tool works, and its architecture.

The PIA tool addresses the complexity of privacy compliance requirements for organisations (both public and private sector), by highlighting privacy risks and compliance issues for individuals within the organisation who are not experts in privacy and security, so they can identify solutions in a given situation. This will allow organisations to identify potential issues at an early stage and hence avoid costs associated with pursuing development paths that are unlawful or pose a higher risk than an organisation can accept or insure against. Where PIAs are mandatory for public sector organisations, the tool can provide evidence that due process has been followed for the purpose of reporting and audit. More specifically, it can help decision-makers within organisations to decide whether a new project (in a broad sense, encompassing scheme, notion or product, etc.) that they wish to develop should go ahead, and if so in what form (i.e. what restrictions there are, what additional checks should be made, etc.). The tool could be run at several stages during the lifetime of a project development process, each time producing different output and advice appropriate to that stage.

The PIA tool also addresses privacy and security risks in the cloud that may be raised, as part of its analysis about the project. This analysis includes those aspects mentioned in the previous section, in relation to the particular context involved: who the cloud service provider is, what their trust rating is, what security and privacy mechanisms they use, as well as other factors that are not specifically cloud-related, for example, to what extent the current project involves sensitive information and for what purposes personal data will be used.

User input for the PIA tool contains project information, such as project name, organisation name, region, brief project description, project lead, and contact details. This is followed by a descriptive analysis of the project such as outlining project documents, identifying stakeholders, and identifying early privacy risks in order to determine if a PIA is required. For example, the user may wish to describe how the organisation collects or obtains personal information or explain if personal information will be transferred outside their jurisdiction including details of the receiving countries. Output for the PIA tool is a report displaying information in several sections: introduction, project and contact details, the summary of findings (which indicates if the PIA tool has found the project to be either compliant or not), risk summary (which indicates the levels of risk associated with each privacy domain), and details of other compliance/non-compliance issues, such as security, transparency, and transborder data flows. Furthermore, the PIA tool provides detailed information about policies in relation to which the project is not compliant or is only partially compliant. In these situations, the tool provides detailed reasons for the partial or non-compliance by highlighting the specific legislation concerned, risks, standards, policies, etc. Finally, recommendations are displayed indicating what the user (organisation) must do to resolve these issues. Throughout the report, clear visual indicators are displayed; these indicate the issues that appear to be compliant with the requirements (i.e. legislation), require further attention, or have failed.

Although our focus for the tool is on privacy and data protection, this approach is also applicable in a broader sense as it can apply to other compliance areas, such as data retention, security, and export regulation.

The following section provides more details of how the tool works.

3.4.1 Architecture and Knowledge Representation

In this section, we discuss the architecture and knowledge representation of the PIA tool. There are a number of traditional programming approaches (e.g. Java, Python, C#, and other object-oriented languages), available for developing a Web-based PIA tool that can address the generic requirements for a PIA system including the collection of data such as project and contact information, the processing of data, and the display of data (e.g. report). Our approach for the PIA tool is a decision support system (DSS) based on a type of expert system [21]. A number of different approaches are available for developing a rule-based system (e.g. expert system) that stores and manipulates knowledge and interprets information in a useful way including Drools [22] and VisiRule [23].

The architecture of the PIA tool that we are currently developing and prototyping is illustrated in Fig. 3.1. This represents one approach (i.e. client-server) of a Web-based PIA tool that can address privacy and cloud environments that is based upon our choice of using the Corvid Runtime environment for a single organisation [24].

The PIA tool has a knowledge base (KB) that is created and updated by privacy experts on an ongoing basis. The experts can be within the organisation (i.e. in-house) or can be outsourced externally (i.e. external consultants). Thus, generic rules for privacy and data protection legislation from a number of jurisdictions (e.g. the UK Data Protection Act 1998, the US Privacy Act 1974) are created and entered into the KB by the experts using a specific user interface (UI). This is important as the tool is to be deployed within a cloud environment, whereby organisations from different jurisdictions may ask to use the application. Initially, the tool will cover jurisdictions that currently conduct PIAs including the UK, the USA, Australia, NZ, and Canada.

There are two types of users: end users (who fill in a questionnaire from which a PIA report is generated) and domain experts (who create and maintain the KB). Typically, users interact with the PIA tool via the Corvid Java runtime that can be Web based (e.g. delivered either as an applet or servlet) or fielded as a standalone Java application [24].

The architecture uses the Corvid servlet runtime [24] that delivers Hyper-Text Markup Language (HTML) pages that contain session-specific data and variables that are sent to the user's browser. Therefore, all processing is done on the server with only HTML pages sent to the client's machine, and it can handle multiple users when questions or results are displayed. Since the servlet engine is already running, starting a new session is very quick as the user does not have to wait for an applet and KB to download. In addition, the full power of HTML and any extensions supported by the browser such as Extensible Markup Language (XML), JavaScript, or Java Server Pages (JSP) can be used to design the user interface screens. This allows for far more complex and sophisticated interfaces to be built than can be done using the Corvid applet approach.

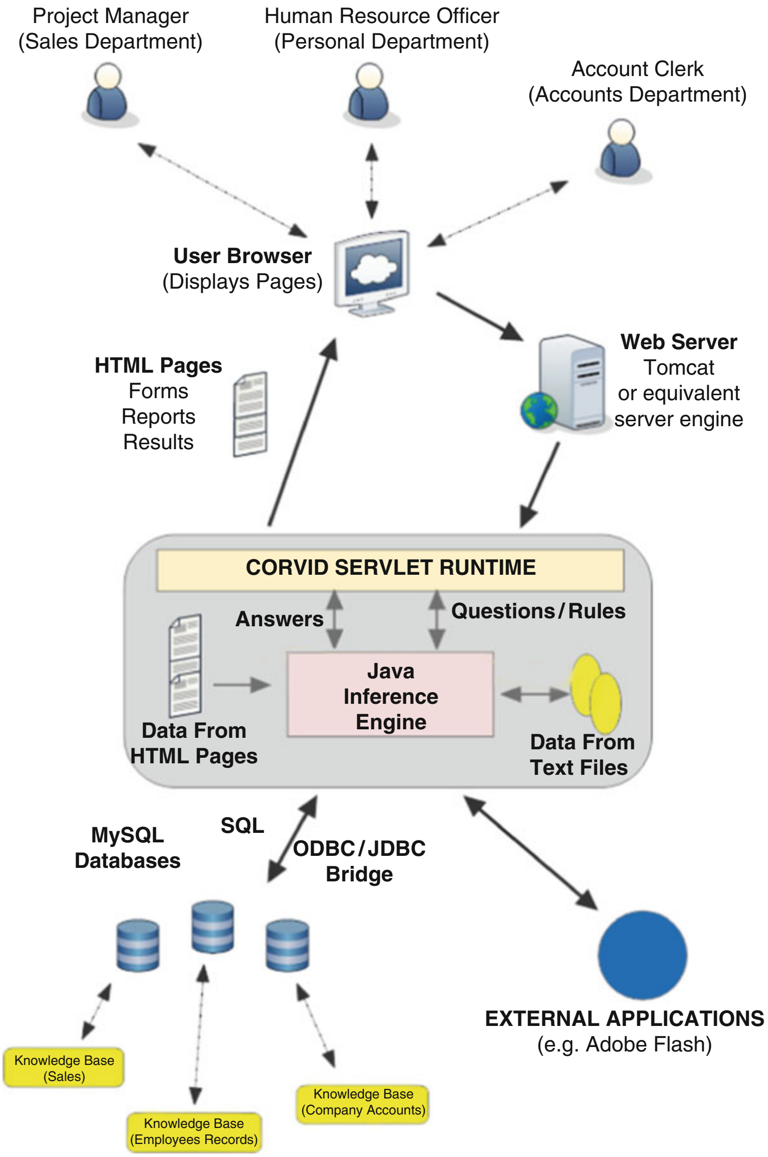


Fig. 3.1 PIA tool architecture for single organisation

The PIA tool may also use multi-tenancy, whereby a single instance of the PIA application may run on a server, serving multiple clients (i.e. tenants) within the organisation. Therefore, it is possible with this architecture to have different KBs for different departments within the organisation that have different privacy and

```
IF
      Your infrastructure is a composition of two or more clouds
AND
      The clouds are bound together by standardised or proprietary
      technology that enables data and application portability
THEN
      The Hybrid cloud model is a good solution
```

Fig. 3.2 Heuristic representation of cloud infrastructure question

organisational policies and acceptable risks. Furthermore, we believe that this architecture is scalable because the system has the ability to accommodate changing load such as the number of users in the organisation that share a single instance of the PIA tool.

Our approach uses a Corvid Exsys rules (i.e. Java) engine [25], which makes inferences by deciding which rules (i.e. those created by the domain expert that are directly associated with questionnaires and questions) are satisfied by facts or objects, prioritises the satisfied rules, and executes the rule with the highest priority. Ontologies can additionally be used for fine-grained reasoning. The engine uses two distinct modes (e.g. backward and forward chaining). In forward chaining (e.g. data-driven), the engine searches the rules until it finds one in which the "IF" condition is known to be true. It concludes the "THEN" condition and adds this information to its data and continues in this way until a goal or conclusion is reached. A meta-level description of the privacy rules for this phase is "IF <trigger conditions> THEN <action>". For example, the National Institute of Standards and Technology (NIST) provides five essential characteristics (e.g. on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service), three service models (e.g. SaaS, platform as a service (PaaS), and infrastructure as a service (IaaS)), and four deployment models (e.g. private, community, public, and hybrid clouds) for their definition of cloud computing [26]. Therefore, a question in the PIA tool may ask the user "Is the Hybrid cloud infrastructure the best option for your organisation?" This question can be converted into a heuristic rule, as illustrated in Fig. 3.2 [27].

In backward chaining (e.g. goal driven), the engine searches for top-level goals, which are the possible answers to the problem or potential recommendations. Therefore, the engine can determine what it needs to meet a particular goal including determining when that goal is met or that a goal cannot be met. However, to meet this determination, the tool requires data on a specific situation being analysed. This data can come from other rules, external sources such as databases and spreadsheets, or asking the user additional questions. For example, an organisation might like to use a cloud provider that uses Representational State Transfer (REST) Web resources and supports multiple accounts with different key management techniques for each customer. The engine checks the rules to find one that would be relevant to making this decision, as illustrated in Fig. 3.3 [27]:

Although in this case the engine has found a potentially useful rule, without more data, it cannot determine if this rule should be used. This is because the engine does not know how many multiple accounts are allowed for each customer by Windows

```
IF
        Representational State Transfer (REST) Web resources are used
        by the organisation
AND
        Multiple user accounts for each customer are required
THEN
        Windows Azure AppFabric Access Control Service
        is a good choice
```

Fig. 3.3 Heuristic representation of REST/multiple accounts question

```
IF
        Multiple accounts for each customer is greater than 6
THEN
        Windows Azure is not a good choice
```

Fig. 3.4 Heuristic representation of new goal

Azure [28]. Therefore, the engine searches for a rule that can tell it something about the maximum number of multiple accounts, as illustrated in Fig. 3.4 [27]:

Although the original rule (i.e. goal) is not forgotten, it is temporarily superseded by the new rule (i.e. new goal). However, to use this rule, the engine needs to know the maximum number of multiple accounts per customer the organisation requires. Thus, this answer may come from a database, a different program, other rules, or by asking the user directly, as the engine determines where and how to get the needed data. This process of having one goal requiring data leading to another goal from the highest level to the lowest level is how the engine in the PIA tool uses backward chaining [27]. In addition, as data becomes available, lower level goals are met and are dropped off the chain and continue until the engine is able to determine which of the conditions for the initial top-level goals are true. Similarly, this approach is used to reason about transborder data flow and other data protection requirements.

Although we use this particular inference engine to run rules, the approach is not reliant on any particular inference engine or specific format beyond the processing of "IF/THEN" rules, and so a variety of mechanisms could be used from production rules systems to "Clips" or "Prolog".

In Sect. 3.5, we provide further details of our approach for our PIA tool including the software development methodology, data collection, analysis, results, and modelling.

The following section will provide more details of our specialised tool, including how it may be used in a cloud environment.

3.4.2 Cloud Deployment of PIA Tool

This section considers the deployment of our PIA tool in a cloud environment and its architecture. The possible deployment of our PIA tool is based upon the advantages and disadvantages between the cloud service models and the major cloud deployment models [26]. As previously discussed in Sect. 3.4.1, there are three cloud service models to consider including:

- *IaaS*: physical hardware such as servers, disks, and networks are abstracted into virtual servers and virtual storage
- *PaaS*: this provides a platform built upon the abstracted hardware that can be used by developers to create cloud applications such as the PIA tool
- *SaaS*: this provides our PIA tool as a service that enables customers to use the cloud without complexities of hardware, the Operating System (OS), or even the PIA tool's installation

For our PIA tool, the SaaS service model appears to be appropriate, whereby the end users (i.e. customers) of the tool do not actually have to own the platforms. However, when deploying our PIA tool as a SaaS service, as with all of the other cloud service models, there is a common set of technological challenges including [29]:

- *User interface flexibility*: UIs must be easy for the end users to use and meet the needs of the customer.
- *Productivity*: The solution for our PIA tool must provide a highly productive environment that focuses on industry best practices.
- *Operational excellence*: Our PIA tool must be always available and scale to the maximum size required.
- *Security and compliance*: The solution for our PIA tool must ensure that the data and application are accessed only by those who are registered to use it.
- *Multi-tenancy*: The solution for our PIA tool must be able to support from one user to many.
- *Integration*: The solution must be able to have the ability to easily integrate with other applications by supporting all relevant standards.
- *Personalisation*: This ensures that our PIA tool must look and work as the tenant and end users want it to. However, each tenant may want different UIs and questions for their particular organisational needs.
- *Costs*: As discussed in this section, initial costs of the deployment of our PIA tool depend upon the adopted solution. However, tenant and end user costs are on a pay-per-use basis.

One of the key elements discussed in the list above is multi-tenancy. In deploying our PIA tool as a SaaS service, multi-tenancy provides several options including [29]:

- *Isolated tenancy*: Whereby our PIA tool, databases, and infrastructure are isolated and are hosted per tenant as separate instances
- *Infrastructure tenancy*: Whereby our PIA tool and databases are isolated, although the infrastructure is shared and hosted in a virtual environment
- *Application tenancy*: Whereby our PIA tool and the infrastructure are shared by all tenants, although the databases are isolated
- *Shared tenancy*: Whereby our PIA tool, database, and infrastructure are all shared by the tenants

Table 3.4 Comparison of multi-tenancy models

Tenancy models				
	Isolated	Infrastructure	Application	Shared
Time to market	Short	Short	Long	Longest
Infrastructure costs	High	High	Low	Low
Economies of scale	Very poor	Poor	High	Highest
Scalability	Poor	Poor	High	Highest
Provisioning	Difficult	Difficult	Easy	Easy
Admin/mgmt costs	Very high	High	Low	Low
Target tenants	Dissimilar	Dissimilar	Similar	Similar
Allows for application changes	No	No	Yes (except DBs)	Yes
Coding difficultly	Easier	Easy	Less difficult	Difficult
Implementation of service-level agreements (SLAs)	Easier	Easy	Less difficult	Difficult
Containment	Easier	Easy	Less difficult	Difficult

In considering which option to choose for our PIA tool as a SaaS service that involves multi-tenancy, a comparison is made of the different models as illustrated in Table 3.4.

The application multi-tenancy model is more suited to our PIA tool because of several reasons including the initial costs are reasonably low, the scalability is high, and the target tenants for our PIA tool are similar in nature (i.e. they will be using our application to check if their projects require PIAs). Therefore, the deployment of our PIA tool as a SaaS application is illustrated in Fig. 3.5.

There are various different options for how this might be provided within a cloud environment. For example, one option may be to deploy our PIA tool in a private cloud infrastructure, whereby the tool is provided as a service for a single organisation. Private clouds rely on virtualisation (e.g. storage and server) and treat hardware as a pool of resources that can be allocated to various functions. Thus, our PIA tool may be managed by the organisation or a third party (e.g. cloud provider) and may exist on premise or off premise [30]. The advantages of using this option relate to control, governance, security, availability, and speed of access. In contrast, the disadvantages of using this option for our PIA tool are minimal elasticity, costs, and scalability since the organisation is responsible for setting up, maintaining, and growing the infrastructure as necessary [31, 32].

Another option is to deploy our PIA tool as a SaaS application in a community cloud. In this cloud, the infrastructure may be shared by multiple organisations and supports a specific community that has shared concerns (e.g. mission, security requirements, policy, jurisdiction, and compliance considerations). Again, this cloud may be managed by the organisation or a third party and may exist (e.g. hosted) on premise or off premise [30]. Some advantages of using this option for our PIA tool are elasticity and a pay-for-use on-demand service. However, because community clouds target a specific industry or concern, some disadvantages for the

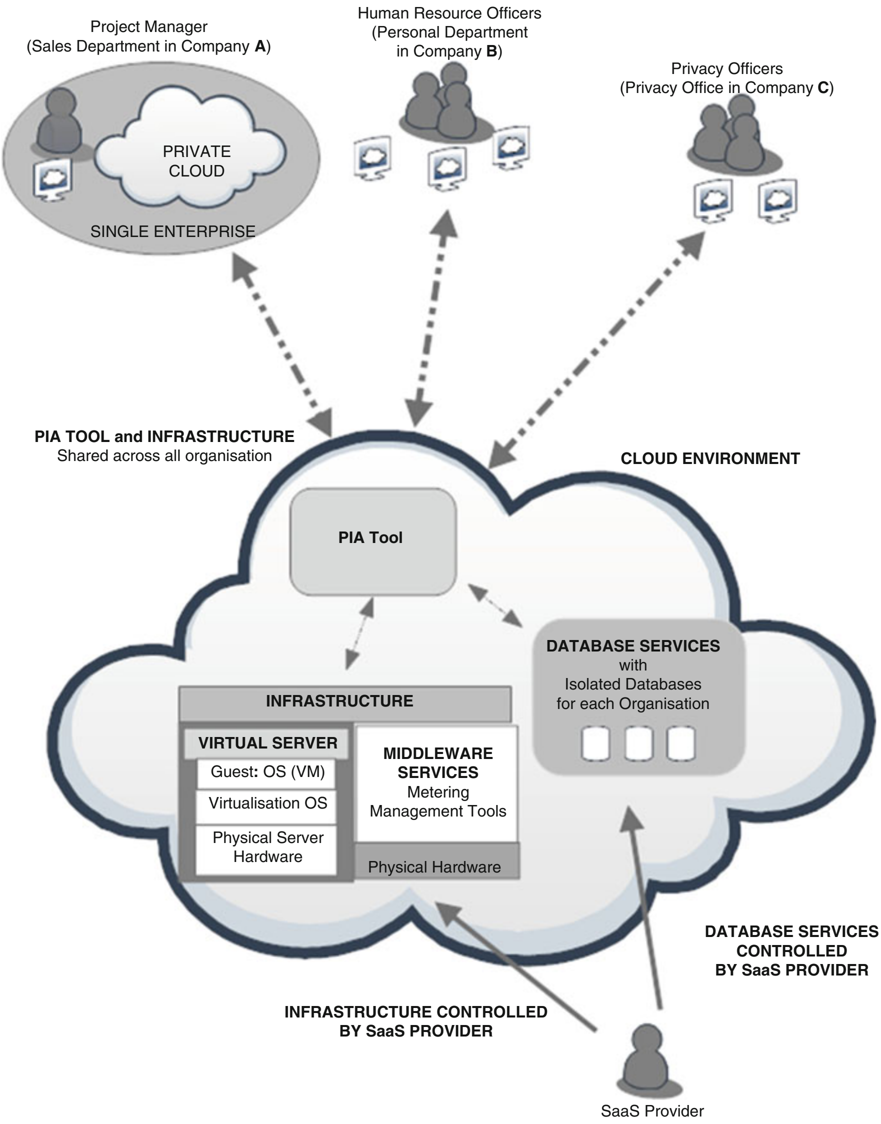

Project Manager
(Sales Department in Company A)
PRIVATE CLOUD
SINGLE ENTERPRISE
Human Resource Officers
(Personal Department in Company B)
Privacy Officers
(Privacy Office in Company C)
PIA TOOL and INFRASTRUCTURE
Shared across all organisation
CLOUD ENVIRONMENT
PIA Tool
DATABASE SERVICES with Isolated Databases for each Organisation
INFRASTRUCTURE
VIRTUAL SERVER
Guest: OS (VM)
Virtualisation OS
Physical Server Hardware
MIDDLEWARE SERVICES
Metering
Management Tools
Physical Hardware
DATABASE SERVICES CONTROLLED BY SaaS PROVIDER
INFRASTRUCTURE CONTROLLED BY SaaS PROVIDER
SaaS Provider

Fig. 3.5 PIA tool in cloud environment

infrastructure exist including low visibility, control, trust, and higher costs due to specialisation in support of specific customer requirements [31, 32].

A possible benefit to PIAs in the community cloud involves the use of a communal KB that may have several PIA questionnaires that represent the needs of the multiple organisations. Thus, the KB can be shared, updated, and maintained by all organisations in the community, and knowledge (such as answers and PIA decisions) is therefore shared between all organisations.

Table 3.5 Advantages/disadvantages of hybrid cloud

Hybrid cloud	
Advantages	Disadvantages
Maximum flexibility	Most of the disadvantages for both private and public clouds (for their respective components)
Dedicated resources on-site (via private cloud)	Additional layer of software is needed to provide governance and brokerage between the cloud services
Pay-per-use resources off-site (via public or community cloud)	Policy must be defined indicating which services and datasets are allowed in which part of the cloud
Off-site resources are pay for what is used. Turn the service off when done	The broker/governance component is an additional software component requiring additional IT skills to operate and manage
Elasticity when needed	
Immediate self-service	

A third option for our PIA tool is deploying it as a SaaS application in a public cloud. This infrastructure is made available to the general public or a large industry group and is owned by an organisation or cloud provider that sells cloud services [30]. Again, in a public cloud, there is no purchase of physical infrastructure, and the organisation (i.e. client) can use the services on a pay-for-use basis (e.g. on-demand self-service) with maximum elasticity. However, using this option for our PIA tool in the public cloud can lead to several disadvantages and issues for the organisation including low visibility, control, and trust. For example, our PIA tool collects information (i.e. contact information such as name, e-mail, and telephone numbers) and produces results and reports that organisations may regard as sensitive data or may regard their exposure as a high risk to the organisation.

This is because the cloud provider takes responsibility for the software and services. Furthermore, governance and policy enforcement is still emerging in public clouds, and from a security perspective, multi-tenancy provides added complexity [31, 32].

A fourth option is to deploy our PIA tool in a hybrid cloud. A hybrid cloud is the composition of two or more clouds (e.g. private, community, or public) that remain unique entities but are bound together by standardised or proprietary technology that enables data and application portability such as the use of cloud bursting for load balancing between the clouds [29]. In this scenario, the sensitive data collected by our PIA tool is stored on its own private servers in a private cloud behind a firewall away from the Internet. Therefore, published PIA results and reports that are then ported to the public cloud for customers would not contain sensitive information. Thus, our PIA tool would use public clouds for less sensitive tasks such as the PIA questionnaires but use a private cloud for vital processing tasks. However, like all cloud deployment models, the hybrid cloud has advantages and disadvantages, as illustrated in Table 3.5 [31, 32].

A public cloud deployment introduces a third party that may lead to several disadvantages and issues including low visibility, control, security, privacy, and trust, as discussed in Chap. 1.

Another consideration is the network bandwidth constraints and cost. For example, if a decision is made to move some of our PIA tool's infrastructure to a public cloud, disruption in the network connectivity between our tool's clients and the cloud service may affect the availability of our cloud hosted PIA tool. Moreover, on a low bandwidth network, there is a possibility that the interaction between our PIA tool and its customers (i.e. users) may also be affected.

There are additional factors to consider before selecting the use between private, community, and public clouds for our PIA tool. One important factor is the amount of storage and time our PIA tool is to be deployed. For example, 10 terabytes (TB) of storage supplied by a cloud provider for 5 years may involve a high pricing structure for our PIA tool in order to recover costs. On the other hand, if our PIA tool uses a temporary storage plan for 1 year, it may be cost-effective to use this private cloud. However, if the plan is to use a community cloud, the costs would be shared between all participating organisations. Thus, it can be seen that one of the factors dictating the use between private, public, or community clouds is the size of storage and how long the storage for our PIA tool is intended to be used.

Of course, cost may not be the only consideration in evaluating which type of deployment cloud model is best for our PIA tool, as some application services such as Salesforce.com (i.e. a popular customer relationship management (CRM) cloud service) offers unique features such as specialised management tools [30]. In addition, other public cloud providers offer services such as capacity planning, procurement, and the management of data centres.

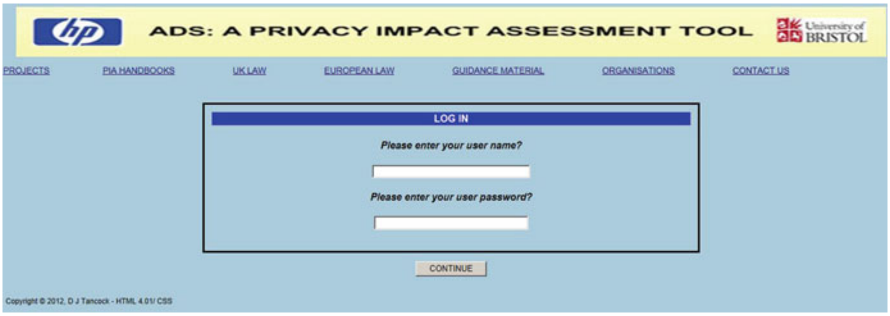
Also there is the context in which our PIA will be deployed. For example, is our tool intended for the UK-based customers only (i.e. those who operate solely in the UK) or for UK customers who operate globally? This difference is critical because sensitive information can mean one thing in the UK under the Data Protection Act 1998, but sensitive information can mean completely something else especially in other countries such as those outside the European Union (EU).

In general, since the deployment models (i.e. private, public, and community) have different characteristics and even different business drivers such as cost, the best solution for our PIA tool may be a hybrid solution that involves all three models.

In the next section, we provide some examples of user interfaces (UIs) that are part of the PIA tool.

3.4.3 Examples of PIA Tool UIs

This section considers the functionality and appearance of some of the PIA tool UIs. However, the examples shown are not the final production UIs, rather those designed as of the time of writing (March 2012).



Please enter your user name?

Please enter your user password?

Copyright © 2012, D J Tancock - HTML 4.01/ CSS

Fig. 3.6 Log-in page

Our PIA tool uses Java servlets that display HTML templates to end users via standard browsers. Typically, the call to start our PIA tool is done with a Uniform Resource Locator (URL) [27]: http://www.myServer.com/CORVID/corvidsr?KBNAME=../../MyApps/MySystem.cvr

The initial screen of our PIA tool consists of a log-in screen that prevents unauthorised users entering our application, as illustrated in Fig. 3.6.

The log-in system asks for specific user names and passwords that allow different user access modes for the tool (i.e. administrator, customer and stakeholder). In a further instantiation, the tool would be integrated with a dedicated authentication mechanism that allow role-based access.

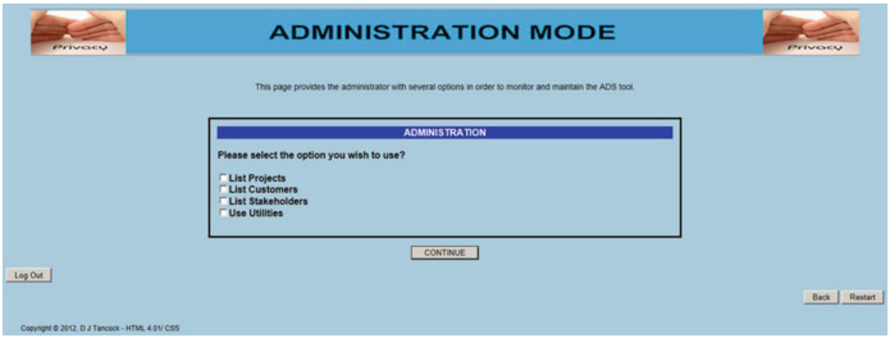
Upon completion of a successful log-in by users, our PIA tool automatically loads a file that contains project contextual information including: contact details, project details, previous PIAs and similar project information, and stakeholder details. This information is the result of previous usage of the tool by those users, although some information may be derived automatically via the login process e.g. via an enterprise directory system.

In the next section, we describe the functionality and appearance of our PIA tool if the end user selects the administrator mode.

3.4.3.1 Administrator Mode

This section discusses the functionality and appearance of our PIA tool when authorization to the administrator mode is successful. The administration main page provides the user with options, as illustrated in Fig. 3.7 . For example, the administrator can view projects, customers and stakeholders in a particular project, or use specific utilities that help maintain our PIA tool. For this particular implementation, restrictions are placed on the layout of some screens by the underlying application used, but future plans include creation of more flexible interfaces for such screens, e.g. via a taskbar.

Projects are listed in a table that displays information including: the overal status of the project, the project name, organisation name, contact name, the date the proj-


ADMINISTRATION MODE

This page provides the administrator with several options in order to monitor and maintain the ADS tool.

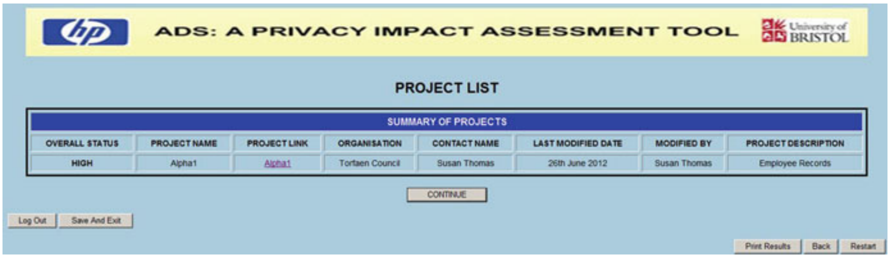
ADMINISTRATION

Please select the option you wish to use?

□ List Projects
□ List Customers
□ List Stakeholders
□ Use Utilities

Copyright © 2012, D J Tancock - HTML 4.01/ CSS

Fig. 3.7 Administration mode


ADS: A PRIVACY IMPACT ASSESSMENT TOOL

PROJECT LIST

SUMMARY OF PROJECTS

OVERALL STATUS	PROJECT NAME	PROJECT LINK	ORGANISATION	CONTACT NAME	LAST MODIFIED DATE	MODIFIED BY	PROJECT DESCRIPTION
HIGH	Alpha1	Alpha1	Torfaen Council	Susan Thomas	26th June 2012	Susan Thomas	Employee Records

Fig. 3.8 Project list

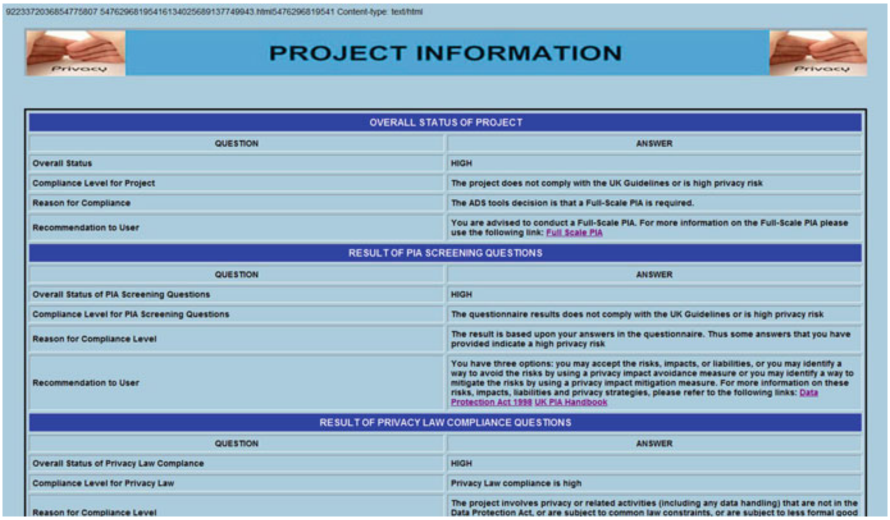
ect was completed or last modified, the person who completed or modified the project and the project description, as illustrated in Fig. 3.8. However, due to the limitations of the servers taht the tool is currently using (i.e. Corvid and Tomcat 7) the returned data is text. It is anticipated that for businesses that use their own servers for the tool the overall status field in the display would include an image that represents the privacy risk.

To access the project the user clicks the project link. A detailsed HTML page is returned that displays the project information, as illustrated in Fig. 3.9. Contact information and stakeholder details can be accessed, in a sanitised form if appropriate for the viewer.

Utilities currently under development such as the web browser, mail log analyser, firewall analyser and multi-router analyser that help the administrator to maintain the tool are accessed via the 'Utilities' checkbox.

3.4.3.2 Stakeholder Mode

This section describes the functionality and appearance of the stakeholder mode. This mode allows stakeholders to view completed reports for particular projects and allows stakeholders to provide feedback without going through the main questionnaires.

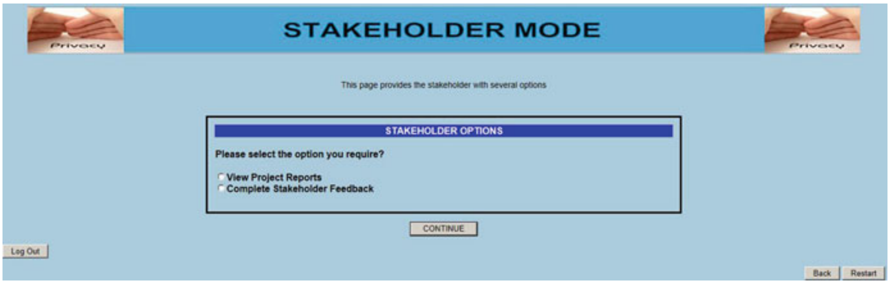


9223372036854775807 54762968195416134025689137749943.html5476296819541 Content-type: text/html

PROJECT INFORMATION

OVERALL STATUS OF PROJECT	
QUESTION	ANSWER
Overall Status	HIGH
Compliance Level for Project	The project does not comply with the UK Guidelines or is high privacy risk
Reason for Compliance	The ADS tools decision is that a Full-Scale PIA is required.
Recommendation to User	You are advised to conduct a Full-Scale PIA. For more information on the Full-Scale PIA please use the following link: Full Scale PIA
RESULT OF PIA SCREENING QUESTIONS	
QUESTION	ANSWER
Overall Status of PIA Screening Questions	HIGH
Compliance Level for PIA Screening Questions	The questionnaire results does not comply with the UK Guidelines or is high privacy risk
Reason for Compliance Level	The result is based upon your answers in the questionnaire. Thus some answers that you have provided indicate a high privacy risk
Recommendation to User	You have three options: you may accept the risks, impacts, or liabilities, or you may identify a way to avoid the risks by using a privacy impact avoidance measure or you may identify a way to mitigate the risks by using a privacy impact mitigation measure. For more information on these risks, impacts, liabilities and privacy strategies, please refer to the following links: Data Protection Act 1998 UK PIA Handbook
RESULT OF PRIVACY LAW COMPLIANCE QUESTIONS	
QUESTION	ANSWER
Overall Status of Privacy Law Complance	HIGH
Compliance Level for Privacy Law	Privacy Law compliance is high
Reason for Compliance Level	The project involves privacy or related activities (including any data handling) that are not in the Data Protection Act, or are subject to common law constraints, or are subject to less formal good

Fig. 3.9 Project information



STAKEHOLDER MODE

This page provides the stakeholder with several options

STAKEHOLDER OPTIONS

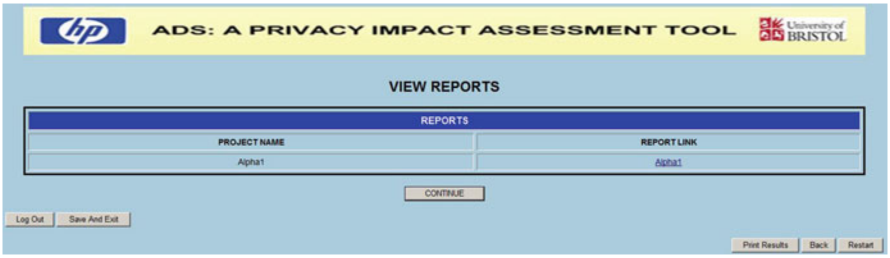
Please select the option you require?

○ View Project Reports
○ Complete Stakeholder Feedback

Fig. 3.10 Stakeholder options

Access and permissions to particular projects that stakeholders are involved with is provided by our tool in several ways. First, a log-in screen is used, whereby users must provide a user name and password that authorised access to a project. Second, permissions are set in the database via the “GRANT” option to restrict stakeholders to particular database tables where the project name equals the stakeholder ID. In addition, organisations can control permissions and access to reports by setting the Internet Protocol (IP) address to individual or group computers because it is often desirable to share a report among others or to have the PIA tool dynamically build a Web page that can be widely accessed. Once authorised, the stakeholder is forwarded to the options page, as illustrated in Fig. 3.10.

One option for a stakeholder is to view the reports for the project. Thus upon clicking “View Project Reports” they are forwarded to the view reports page, as illustrated in Fig. 3.11. Report creation involves the use of a separate HTML template that formats the contents and appearance of the report by using embedded

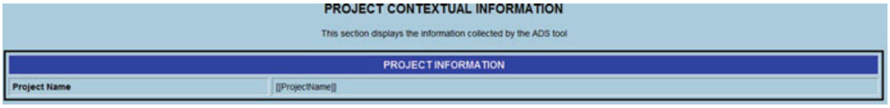


ADS: A PRIVACY IMPACT ASSESSMENT TOOL

VIEW REPORTS

REPORTS	
PROJECT NAME	REPORT LINK
Alpha1	Alpha1

Fig. 3.11 Report list screen



PROJECT CONTEXTUAL INFORMATION

This section displays the information collected by the ADS tool

PROJECT INFORMATION	
Project Name	[[ProjectName]]

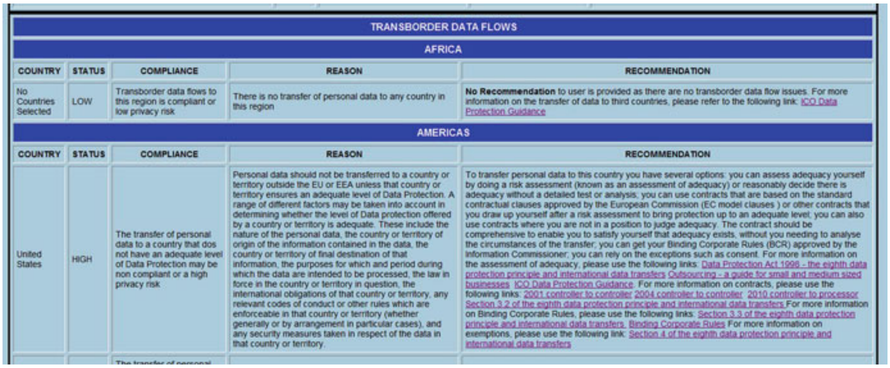
Fig. 3.12 Embedded variables

variables in the form [[…]], before it is added to a collection variable (i.e. a variable that contains a list of strings) and saved, as illustrated in Fig. 3.12 [27]. This is a very convenient way of creating reports because information included in the report can be controlled. For example, one organisation may include information such as personal details, whereas another organisation may want to keep the report confidential.

The report provides in-depth analysis that helps stakeholders and decision-makers determine whether a full-scale PIA assessment is warranted or not and determines whether the characteristics of the UK PIA Guidelines are complaint or non-compliant with the criteria. In addition, the report provides specific reasons for the compliance status and gives advice to the user. The recommendation also includes embedded HTML links to specific information that helps the user understand the advice given by the tool. However, due to the limitations of the servers that the PIA tool currently uses the report displayed is mainly text, as illustrated in Fig. 3.13. However, a display such as a histogram may be developed that indicates the levels of risk associated with each key characteristic.

In addition, the stakeholders can complete a questionnaire about any report that they have read. The questionnaire consists of several questions that encourage communication between the stakeholders involved in the project and the project team, as illustrated in Fig. 3.14, and that allow free text input and ideally attachment of ancillary relevant information. Upon completion of the questionnaire the PIA tool automatically creates a report. Thus interaction between the project team and the stakeholder completed questionnaire is achieved and archived within the customer mode.

In the next section, we consider the customer mode that allows end users to conduct or modify a full-scale PIA initial assessment.



TRANSBORDER DATA FLOWS

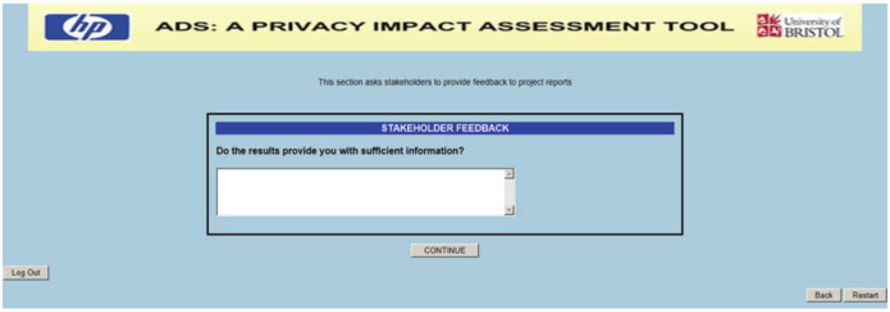
AFRICA

COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
No Countries Selected	LOW	Transborder data flows to this region is compliant or low privacy risk	There is no transfer of personal data to any country in this region	**No Recommendation** to user is provided as there are no transborder data flow issues. For more information on the transfer of data to third countries, please refer to the following link: ICO Data Protection Guidance

AMERICAS

COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
United States	HIGH	The transfer of personal data to a country that dos not have an adequate level of Data Protection may be non compliant or a high privacy risk	Personal data should not be transferred to a country or territory outside the EU or EEA unless that country or territory ensures an adequate level of Data Protection. A range of different factors may be taken into account in determining whether the level of Data protection offered by a country or territory is adequate. These include the nature of the personal data, the country or territory of origin of the information contained in the data, the country or territory of final destination of that information, the purposes for which and period during which the data are intended to be processed, the law in force in the country or territory in question, the international obligations of that country or territory, any relevant codes of conduct or other rules which are enforceable in that country or territory (whether generally or by arrangement in particular cases), and any security measures taken in respect of the data in that country or territory.	To transfer personal data to this country you have several options: you can assess adequacy yourself by doing a risk assessment (known as an assessment of adequacy) or reasonably decide there is adequacy without a detailed test or analysis; you can use contracts that are based on the standard contractual clauses approved by the European Commission (EC model clauses) or other contracts that you draw up yourself after a risk assessment to bring protection up to an adequate level; you can also use contracts where you are not in a position to judge adequacy. The contract should be comprehensive to enable you to satisfy yourself that adequacy exists, without you needing to analyse the circumstances of the transfer; you can get your Binding Corporate Rules (BCR) approved by the Information Commissioner; you can rely on the exceptions such as consent. For more information on the assessment of adequacy, please use the following links: Data Protection Act 1998 - the eighth data protection principle and international data transfers Outsourcing - a guide for small and medium sized businesses ICO Data Protection Guidance. For more information on contracts, please use the following links: 2001 controller to controller 2004 controller to controller 2010 controller to processor Section 3.2 of the eighth data protection principle and international data transfers For more information on Binding Corporate Rules, please use the following links: Section 3.3 of the eighth data protection principle and international data transfers. Binding Corporate Rules For more information on exemptions, please use the following link: Section 4 of the eighth data protection principle and international data transfers
		The transfer of personal		

Fig. 3.13 Sample of completed report



ADS: A PRIVACY IMPACT ASSESSMENT TOOL

This section asks stakeholders to provide feedback to project reports

STAKEHOLDER FEEDBACK

Do the results provide you with sufficient information?

Fig. 3.14 Sample of stakeholder form

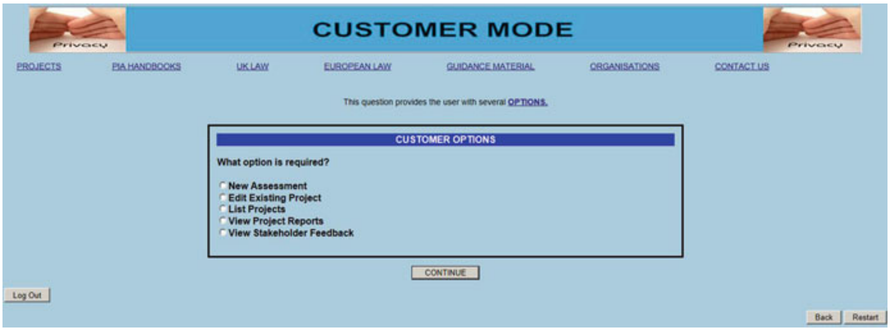
3.4.3.3 Customer Mode

This section describes the functionality and appearance of our PIA tool in the customer mode. The initial customer screen provides several options for users including: the ability to view stakeholder feedback, to conduct a new PIA assessment, the ability to view reports, and the ability to edit an existing PIA assessment, as illustrated in Fig. 3.15.

In both options "conduct a new PIA assessment" and "edit an existing PIA assessment", a specific "user ID" is used to build a unique identifier for saved data for that particular customer, as illustrated in Fig. 3.16. However, in a production version of our PIA tool, there would be a combination of user ID and password to assure that each user's data is protected.

This functionality allows customers to answer some questions and quit mid-session during the PIA assessment, with the ability to return to the same session later as this can be very useful for situations including [28]:

- When there are many questions in the questionnaires that takes the user a long time to answer them.



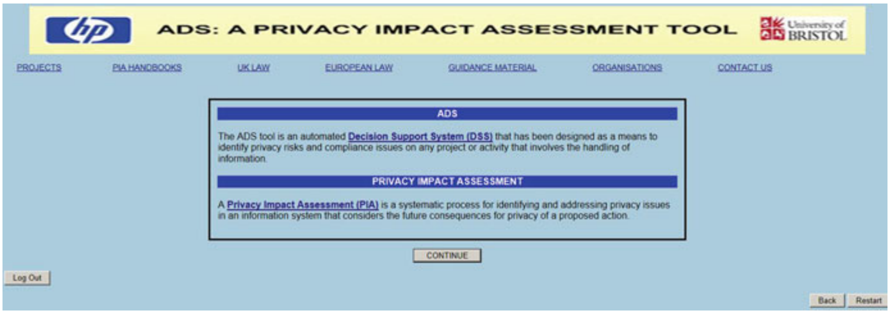
This question provides the user with several OPTIONS.

CUSTOMER OPTIONS

What option is required?

○ New Assessment
○ Edit Existing Project
○ List Projects
○ View Project Reports
○ View Stakeholder Feedback

Fig. 3.15 Customer options



ADS

The ADS tool is an automated Decision Support System (DSS) that has been designed as a means to identify privacy risks and compliance issues on any project or activity that involves the handling of information.

PRIVACY IMPACT ASSESSMENT

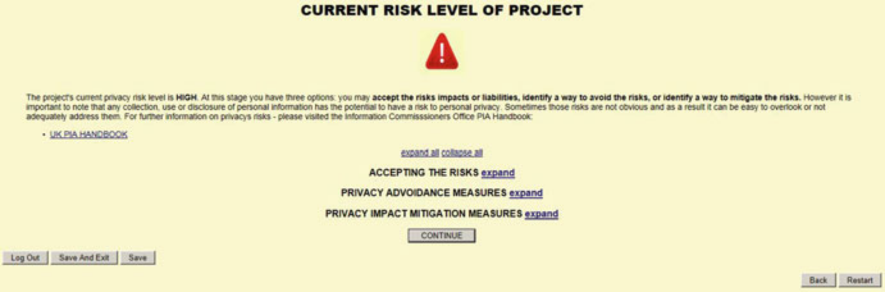
A Privacy Impact Assessment (PIA) is a systematic process for identifying and addressing privacy issues in an information system that considers the future consequences for privacy of a proposed action.

Fig. 3.16 Welcome page

- When there are questions that the end user may not be able to immediately answer.
- When there are questions that require input from several different users, each providing different answers to some questions.

A new assessment typically begins with a welcome page that briefly describes the objective of the tool and gives a brief explanation of what a PIA is. For example, the objective of the tool describes it as being designed as a means to identify privacy risks and compliance issues on any project or activity that involves the handling of information, as illustrated in Fig. 3.17. The main functionality of the page consists of a navigation bar and a submit button (i.e. Continue or OK) that forwards the user to the next page. The navigation bar contains buttons which provide a number of activities and information for the user including:

- *Projects*: Navigates to a different HTML template that lists all previous PIAs and similar projects conducted either by the organisation or by the individual
- *PIA handbooks*: A drop down menu that contains hyperlinks to different PIA handbooks that have been published by major jurisdictions



CURRENT RISK LEVEL OF PROJECT

The project's current privacy risk level is **HIGH**. At this stage you have three options: you may **accept the risks impacts or liabilities, identify a way to avoid the risks, or identify a way to mitigate the risks.** However it is important to note that any collection, use or disclosure of personal information has the potential to have a risk to personal privacy. Sometimes those risks are not obvious and as a result it can be easy to overlook or not adequately address them. For further information on privacys risks - please visited the Information Commisssioners Office PIA Handbook:

- UK PIA HANDBOOK

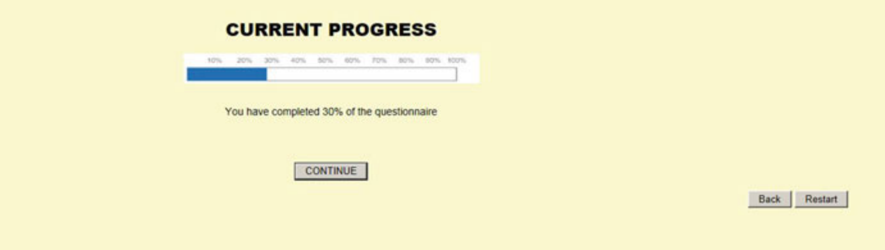
expand all collapse all

ACCEPTING THE RISKS expand

PRIVACY ADVOIDANCE MEASURES expand

PRIVACY IMPACT MITIGATION MEASURES expand

Fig. 3.17 Current risk level



CURRENT PROGRESS

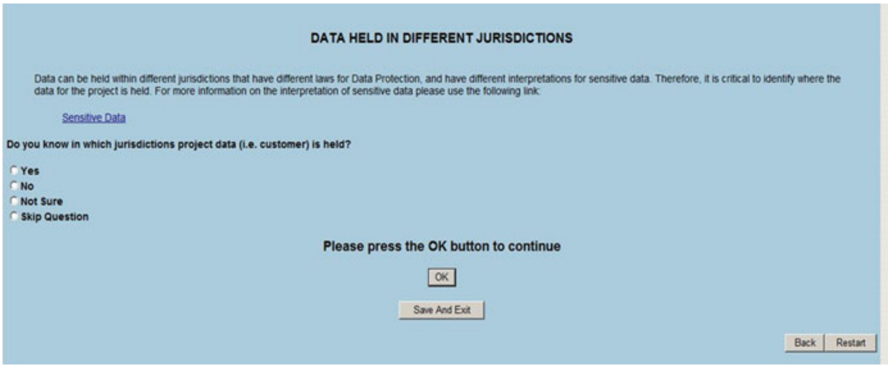
You have completed 30% of the questionnaire

Fig. 3.18 Current progress

- *UK legal topics*: A drop down menu that contains hyperlinks to different UK legal documents including current legislation, regulations, and codes of practice
- *European law*: A drop down menu that contains hyperlinks to different European Directives involving privacy
- *Legal organisations*: A drop down menu that contains hyperlinks to different legal organisations including privacy commissioners websites, privacy advocates, groups, and organisations
- *Contact us*: Navigates to a different HTML template in which the user can e-mail comments and suggestions about the tool or PIAs

The PIA tool can collect a variety of information from databases, files, or by manual user input. For example, contact project, stakeholder, previous PIAs and similar project information, as illustrated in Fig. 3.18. Thus, information is collected via a series of questions that contain free text boxes for user input. HTML hyperlinks are also used in the templates to provide links to instructional help, descriptions, and other websites. Although this information is in the internal results that our PIA tool produces, as previously discussed in Sect. 3.4.3.2, currently in the implementation provided the external reports may not contain this information.

Our tool also provides users with several help pages during a PIA assessment run. For example, at certain stages of the assessment, current risk status and progress HTML pages appear to the users that indicate the risk associated with the



DATA HELD IN DIFFERENT JURISDICTIONS

Data can be held within different jurisdictions that have different laws for Data Protection, and have different interpretations for sensitive data. Therefore, it is critical to identify where the data for the project is held. For more information on the interpretation of sensitive data please use the following link:

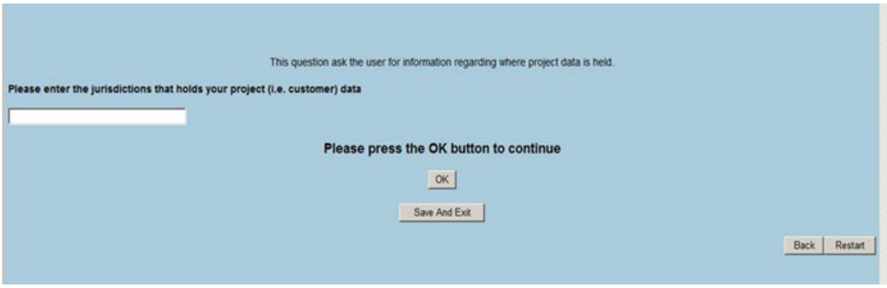
Sensitive Data

Do you know in which jurisdictions project data (i.e. customer) is held?

○ Yes
○ No
○ Not Sure
○ Skip Question

Please press the OK button to continue

Fig. 3.19 Question on customer data held in jurisdictions



This question ask the user for information regarding where project data is held.

Please enter the jurisdictions that holds your project (i.e. customer) data

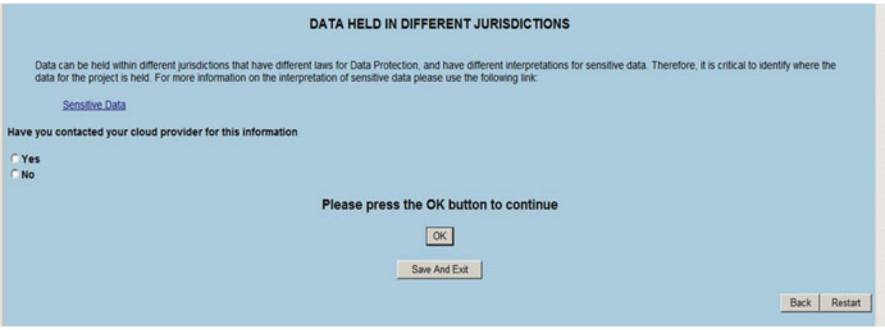
Please press the OK button to continue

Fig. 3.20 Question for manual user input

project at that particular stage and how much of the questionnaire has been completed, as illustrated in Fig. 3.19 and Fig. 3.18.

Typically, the evaluation of questions contained in our tool depends upon the user selecting a single answer from a number of radio buttons. For example, a cloud-related question about customer data held in data centres within different jurisdictions is illustrated in Fig. 3.20 [33]. Therefore, if the user selects any of the possible options, they are given a number of following questions to extract further information. For example, if the user selects "yes" a further question is asked that records the values entered by the user via a free text box (e.g. names of jurisdictions that holds the customer data), as illustrated in Fig. 3.21. These values are then saved in the database in order to produce the results and following report.

However, if the user answers "No" to the question (i.e. Fig. 3.20), a further question asks the user for further information, as illustrated in Fig. 3.22. The possible answers to this follow-up question are "yes" or "no", whereby "yes" triggers a new question that is similar to Fig. 3.20, whereby the user manually enters the values for the jurisdictions supplied by the cloud provider. On the other hand, if the user selects "no", our PIA tool records this answer and provides a recommendation in the results to the user to contact their cloud provider as soon as possible.



DATA HELD IN DIFFERENT JURISDICTIONS

Data can be held within different jurisdictions that have different laws for Data Protection, and have different interpretations for sensitive data. Therefore, it is critical to identify where the data for the project is held. For more information on the interpretation of sensitive data please use the following link:

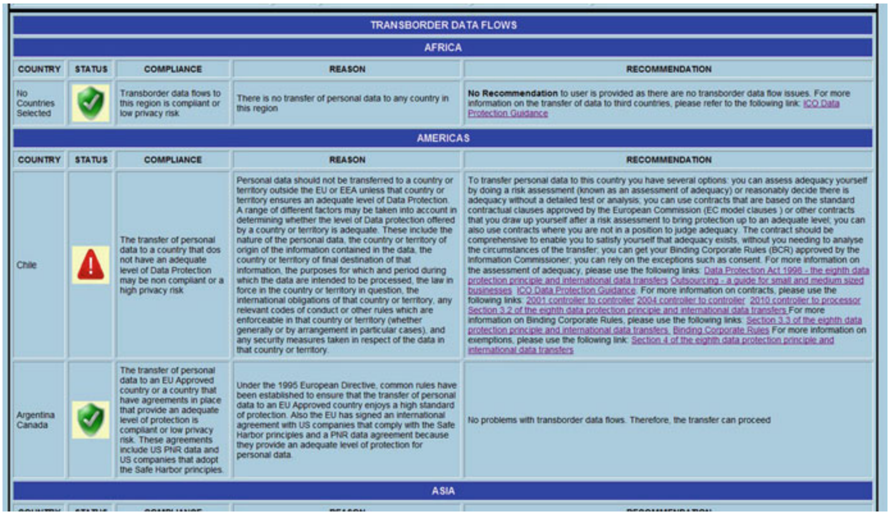
Sensitive Data

Have you contacted your cloud provider for this information

○ Yes
○ No

Please press the OK button to continue

Fig. 3.21 Question on contacting cloud provider



TRANSBORDER DATA FLOWS

AFRICA

COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
No Countries Selected	✓	Transborder data flows to this region is compliant or low privacy risk	There is no transfer of personal data to any country in this region	**No Recommendation** to user is provided as there are no transborder data flow issues. For more information on the transfer of data to third countries, please refer to the following link: ICO Data Protection Guidance

AMERICAS

COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
Chile	⚠	The transfer of personal data to a country that dos not have an adequate level of Data Protection may be non compliant or a high privacy risk	Personal data should not be transferred to a country or territory outside the EU or EEA unless that country or territory ensures an adequate level of Data Protection. A range of different factors may be taken into account in determining whether the level of Data protection offered by a country or territory is adequate. These include the nature of the personal data, the country or territory of origin of the information contained in the data, the country or territory of final destination of that information, the purposes for which and period during which the data are intended to be processed, the law in force in the country or territory in question, the international obligations of that country or territory, any relevant codes of conduct or other rules which are enforceable in that country or territory (whether generally or by arrangement in particular cases), and any security measures taken in respect of the data in that country or territory.	To transfer personal data to this country you have several options: you can assess adequacy yourself by doing a risk assessment (known as an assessment of adequacy) or reasonably decide there is adequacy without a detailed test or analysis; you can use contracts that are based on the standard contractual clauses approved by the European Commission (EC model clauses) or other contracts that you draw up yourself after a risk assessment to bring protection up to an adequate level; you can also use contracts where you are not in a position to judge adequacy. The contract should be comprehensive to enable you to satisfy yourself that adequacy exists, without you needing to analyse the circumstances of the transfer; you can get your Binding Corporate Rules (BCR) approved by the Information Commissioner; you can rely on the exceptions such as consent. For more information on the assessment of adequacy, please use the following links: Data Protection Act 1998 - the eighth data protection principle and international data transfers Outsourcing - a guide for small and medium sized businesses ICO Data Protection Guidance. For more information on contracts, please use the following links: 2001 controller to controller 2004 controller to controller 2010 controller to processor Section 3.2 of the eighth data protection principle and international data transfers For more information on Binding Corporate Rules, please use the following links: Section 3.3 of the eighth data protection principle and international data transfers. Binding Corporate Rules For more information on exemptions, please use the following link: Section 4 of the eighth data protection principle and international data transfers
Argentina Canada	✓	The transfer of personal data to an EU Approved country or a country that have agreements in place that provide an adequate level of protection is compliant or low privacy risk. These agreements include US PNR data and US companies that adopt the Safe Harbor principles.	Under the 1995 European Directive, common rules have been established to ensure that the transfer of personal data to an EU Approved country enjoys a high standard of protection. Also the EU has signed an international agreement with US companies that comply with the Safe Harbor principles and a PNR data agreement because they provide an adequate level of protection for personal data.	No problems with transborder data flows. Therefore, the transfer can proceed

ASIA

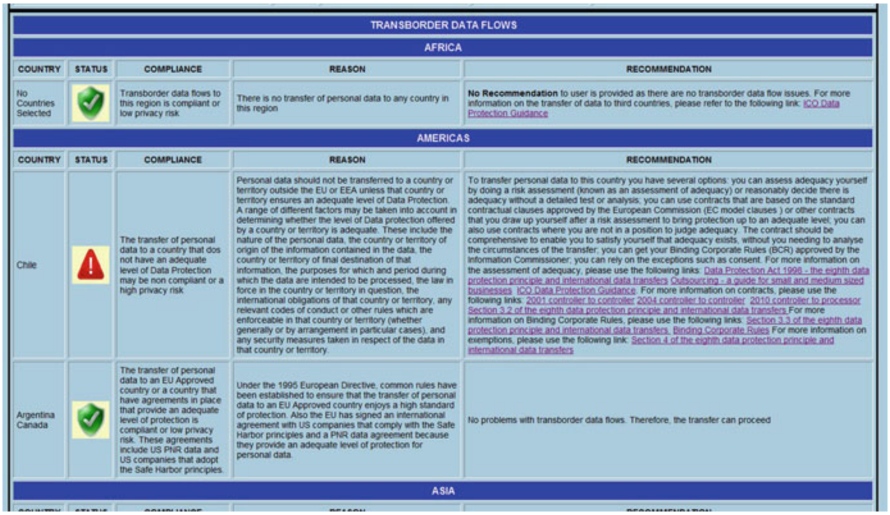
Fig. 3.22 Question for different jurisdictions

Finally, if the user answers “not sure”, a further question is then provided by our PIA tool in the form of a list, as illustrated in Fig. 3.23. In this list, the user can select multiple answers to the question including:

- The selection of multiple answers that are provided in the list including the “other” option, whereby the user can manually enter a value
- The selection of “None” that clears the list and records a value of “no counties have been entered by the user for this particular session” in the database

In addition, the list may be modified to allow the user to select a “not sure” option. In this case, the question is drilled down further to include simple separate questions about cloud providers that provide “yes/no” answers.

Basically, the tool uses rules to generate an output results page and also an audit trail. The output results page provided by our PIA tool is ultimately based on the answers provided by end users, as illustrated in Fig. 3.24.



TRANSBORDER DATA FLOWS

AFRICA

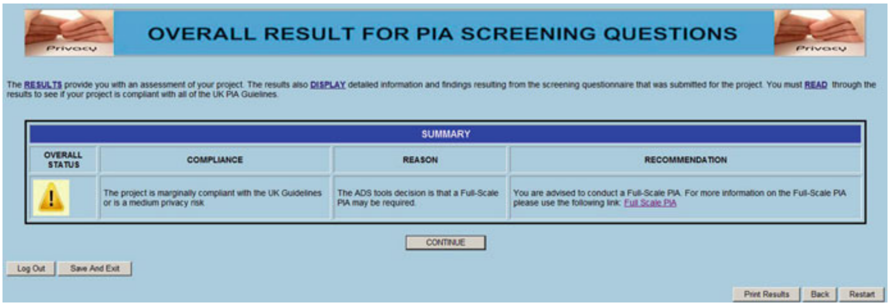
COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
No Countries Selected	✓	Transborder data flows to this region is compliant or low privacy risk	There is no transfer of personal data to any country in this region	**No Recommendation** to user is provided as there are no transborder data flow issues. For more information on the transfer of data to third countries, please refer to the following link: ICO Data Protection Guidance

AMERICAS

COUNTRY	STATUS	COMPLIANCE	REASON	RECOMMENDATION
Chile	⚠	The transfer of personal data to a country that dos not have an adequate level of Data Protection may be non compliant or a high privacy risk	Personal data should not be transferred to a country or territory outside the EU or EEA unless that country or territory ensures an adequate level of Data Protection. A range of different factors may be taken into account in determining whether the level of Data protection offered by a country or territory is adequate. These include the nature of the personal data, the country or territory of origin of the information contained in the data, the country or territory of final destination of that information, the purposes for which and period during which the data are intended to be processed, the law in force in the country or territory in question, the international obligations of that country or territory, any relevant codes of conduct or other rules which are enforceable in that country or territory (whether generally or by arrangement in particular cases), and any security measures taken in respect of the data in that country or territory.	To transfer personal data to this country you have several options: you can assess adequacy yourself by doing a risk assessment (known as an assessment of adequacy) or reasonably decide there is adequacy without a detailed test or analysis; you can use contracts that are based on the standard contractual clauses approved by the European Commission (EC model clauses) or other contracts that you draw up yourself after a risk assessment to bring protection up to an adequate level; you can also use contracts where you are not in a position to judge adequacy. The contract should be comprehensive to enable you to satisfy yourself that adequacy exists, without you needing to analyse the circumstances of the transfer; you can get your Binding Corporate Rules (BCR) approved by the Information Commissioner; you can rely on the exceptions such as consent. For more information on the assessment of adequacy, please use the following links: Data Protection Act 1998 - the eighth data protection principle and international data transfers Outsourcing - a guide for small and medium sized businesses ICO Data Protection Guidance. For more information on contracts, please use the following links: 2001 controller to controller 2004 controller to controller 2010 controller to processor Section 3.2 of the eighth data protection principle and international data transfers For more information on Binding Corporate Rules, please use the following links: Section 3.3 of the eighth data protection principle and international data transfers. Binding Corporate Rules For more information on exemptions, please use the following link: Section 4 of the eighth data protection principle and international data transfers
Argentina Canada	✓	The transfer of personal data to an EU Approved country or a country that have agreements in place that provide an adequate level of protection is compliant or low privacy risk. These agreements include US PNR data and US companies that adopt the Safe Harbor principles.	Under the 1995 European Directive, common rules have been established to ensure that the transfer of personal data to an EU Approved country enjoys a high standard of protection. Also the EU has signed an international agreement with US companies that comply with the Safe Harbor principles and a PNR data agreement because they provide an adequate level of protection for personal data.	No problems with transborder data flows. Therefore, the transfer can proceed

ASIA

Fig. 3.23 Sample of results page



OVERALL RESULT FOR PIA SCREENING QUESTIONS

The RESULTS provide you with an assessment of your project. The results also DISPLAY detailed information and findings resulting from the screening questionnaire that was submitted for the project. You must READ through the results to see if your project is compliant with all of the UK PIA Guielines.

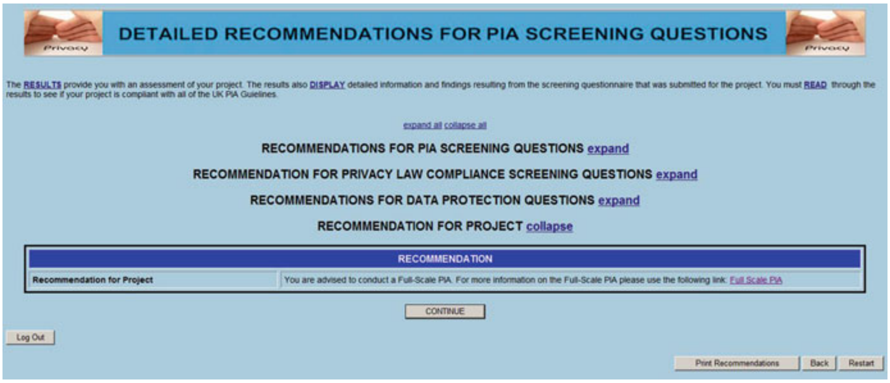
SUMMARY

OVERALL STATUS	COMPLIANCE	REASON	RECOMMENDATION
⚠	The project is marginally compliant with the UK Guidelines or is a medium privacy risk	The ADS tools decision is that a Full-Scale PIA may be required.	You are advised to conduct a Full-Scale PIA. For more information on the Full-Scale PIA please use the following link: Full Scale PIA

Fig. 3.24 Overall results page

The output of the questionnaire (being the answers provided by the user) is matched against the "THEN" condition of the business rules: the corresponding action within the rules contains code that assesses associated risk and groups output into characteristics and categories (transborder data flows, compliance with legislation, jurisdictions, etc.).

The results page provides an in-depth analysis that helps decision-makers determine whether the category is complaint or non-compliant with the UK PIA Guidelines. Although, our tool may use any criteria such as legislation or PIA Guidelines from another particular jurisdiction. In addition, the results provide specific reasons for the compliance status and gives advice to the user. However, at the time of writing (March 2012), the results displayed by the tool are mainly text with a few images. Therefore, in the next iteration of the PIA tool, a display such as



DETAILED RECOMMENDATIONS FOR PIA SCREENING QUESTIONS

The RESULTS provide you with an assessment of your project. The results also DISPLAY detailed information and findings resulting from the screening questionnaire that was submitted for the project. You must READ through the results to see if your project is compliant with all of the UK PIA Guielines.

expand all collapse all

RECOMMENDATIONS FOR PIA SCREENING QUESTIONS expand

RECOMMENDATION FOR PRIVACY LAW COMPLIANCE SCREENING QUESTIONS expand

RECOMMENDATIONS FOR DATA PROTECTION QUESTIONS expand

RECOMMENDATION FOR PROJECT collapse

RECOMMENDATION	
Recommendation for Project	You are advised to conduct a Full-Scale PIA. For more information on the Full-Scale PIA please use the following link: Full Scale PIA

Fig. 3.25 Detailed recommendations page

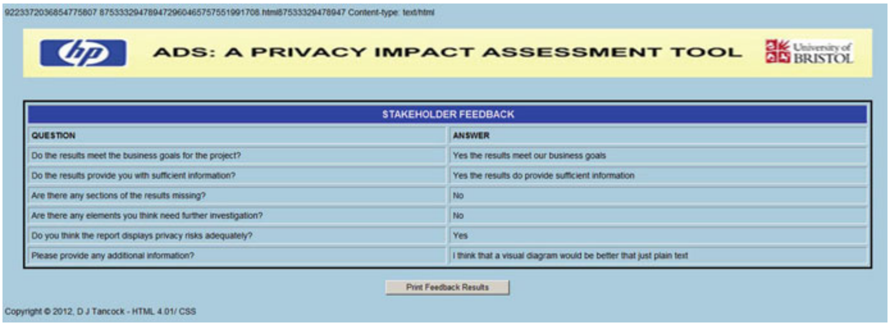
a histogram may be developed that indicates the levels of risk associated with each key characteristic.

Part of the analysis carried out by the tool is to consider legal aspects, such as the UK-US safe harbour process for US companies to comply with the European Directive 95/46/EC on the protection of personal data [34]. The tool has to take into account the rules associated with transborder data flows and cross-border PIAs [35, 36]; moreover, the tool has to consider global organisations and their binding corporate rules. To achieve this, the tool will have a representation of policies related to different legal jurisdictions and will take these policies into account as they apply to a given context.

After the results page, a decision is made by our PIA tool regarding whether the initial full-scale PIA assessment should continue to the privacy law and data protection compliance checks or whether an initial small-scale PIA assessment should be conducted by the organisation, as illustrated in Fig. 3.25. Thus, if our tool recommends that the compliance checks are required, the user is forwarded to the compliance checks upon clicking the “OK” button. However, if the recommendation is that an initial small-scale PIA assessment should be conducted, our tool automatically creates a report.

The privacy law and data protection compliance checks follow the same formats previously described in this section. Thus, three questions are initially asked for the privacy law compliance checks, and a results page is then produced by our PIA tool, which is based upon the users’ answers. In addition, the data protection compliance check consists of one question and a displayed results page. Finally, a results page is produced and displayed by our PIA tool that includes the results from the compliance checks and a report is created.

In summary, our PIA tool helps organisations to ensure privacy concerns are met and supports enterprise accountability, supplying employees with sufficient information and guidance to ensure that they design and conduct their projects in compliance with privacy requirements, such as those outlined in the UK PIA Guidelines of 2009 [14]. In addition, our PIA tool identifies what the user (organisation) must do to resolve these issues.



9223372036854775807 8753332947894729604657575551991708.html87533329478947 Content-type: text/html

ADS: A PRIVACY IMPACT ASSESSMENT TOOL

STAKEHOLDER FEEDBACK	
QUESTION	ANSWER
Do the results meet the business goals for the project?	Yes the results meet our business goals
Do the results provide you with sufficient information?	Yes the results do provide sufficient information
Are there any sections of the results missing?	No
Are there any elements you think need further investigation?	No
Do you think the report displays privacy risks adequately?	Yes
Please provide any additional information?	I think that a visual diagram would be better that just plain text

Copyright © 2012, D J Tancock - HTML 4.01/ CSS

Fig. 3.26 Stakeholder feedback

In the next section, we consider the development suite for our PIA tool, whereby experts can edit and modify questions and rules.

3.4.3.4 Expert Mode and Development Suite

This section discusses the development suite that allows experts to edit and modify questions and rules. The development suite for our PIA tool has been modified into an external file (i.e. a cvd file) that can reside outside the infrastructure on the experts' computer. However, the cvd file must have a link to the cvr file, as illustrated in Fig. 3.26.

However, the cvd file is updated only when the system is saved, whereas the cvr file on the server is updated whenever the system is run by the browser.

The development suite incorporates easy access to our PIA tools internal processes that allow the expert to edit and modify existing questions, rules, and risk levels or create new features without using the application. This is achieved by the expert accessing internal "blocks" including:

- *Logic blocks*: These blocks are made up of rules that can be defined by tree diagrams or stated as individual rules, whereby each block may contain many rules or only a single one. Thus, logic blocks are treated in our PIA tool as objects and are a convenient way to use a group of related rules.
- *Action blocks*: These blocks use a spreadsheet style approach to describe the logic of our PIA tool processes. Thus, action blocks use a procedural approach to solve problems by asking a series of questions.
- *Command blocks*: These blocks control how our PIA tool operates such as what actions to take and what order to perform actions. Fundamentally, these blocks control what variables our PIA tool will try to derive values for and what logic blocks will be used to perform that function. Also, command blocks control the procedural flow of our PIA tool including how the system chains, what blocks to execute, and what results to display.

In the next section, we discuss the confidence variable that is used in our PIA tool to reach a "best fit" for several decisions and conclusions our tool makes.

3.4.3.5 Confidence Variable

This section describes the confidence variable that is used in our PIA tool. A confidence variable is intended to calculate an overall confidence value for the variable. Usually, this is the confidence or likelihood that the result is an appropriate recommendation or solution to the problem that our PIA tool solves. Confidence variables can also be used in other ways, but in all cases, the variable will be given one or more numeric values which will be combined via a formula to produce the overall confidence value.

In our PIA tool, the confidence variable is called “risk level” and is used to measure the probability that the answers in the questionnaire will be selected by the user. The calculation of our confidence variable “risk level” is done by using the sum method, whereby the single value for each question is added together; thus, positive values increase the confidence and negative values decrease the confidence. However, there are other ways of calculation such as average, independent, dependent, multiplication, and the mycin method [27].

Therefore, from and including the technology question in the initial full-scale PIA assessment questionnaire, each possible answer is assigned a value that reflects its confidence. For example, in Table 3.6, risk level values have been assigned to both the technology and identifiers questions.

The use of this feature enables our PIA tool to make multiple simultaneously possible recommendations with differing degrees of confidence to reach a “best fit” for several decisions and conclusions that are then presented to the user. For example, a current status page that may have three possibilities (i.e. the risk level to the project is high, medium, or low) is displayed several times during the assessment run to help the user objectively view the status of the project after a particular set of questions, as illustrated in Fig. 3.27.

Another stage where our PIA tool makes use of the confidence variable “risk level” is in the project summary that is displayed in the results page. Again, this uses a mathematical formula to reach the “best fit” for the project status from three possibilities (i.e. project status is high, low, or medium), as illustrated in Fig. 3.28.

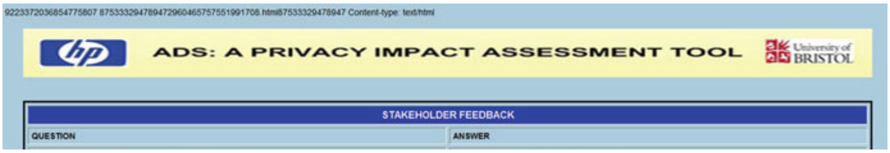
Finally, the confidence variable “risk level” is used in the tool’s full-scale PIA decision. Again, this is a mathematical formula that is similar to Fig. 3.28, whereby different recommendations are displayed to the user, as previously described in Sect. 3.4.3.3.

This feature is used to calculate the answers provided by the user in the questionnaire with the result of the confidence variable “risk level” being an appropriate solution (i.e. displays an compliance indicator and advice to the user for the project status) Thus, in our PIA tool, if the confidence variable “risk level” assigned to each question were modified to meet the needs of an organisation who interprets each question differently, the PIA results and the following report will reflect the change.

This provides an effective solution, whereby KBs can be created that have different values for each question that produce different results and reports. For example, an organisation may have several versions of the KB for different departments.

Table 3.6 Risk levels assigned to questions

Risk levels assigned to questions		
Question	Possible answer	Risk level
Technology	Yes	20
	No	−10
	Not sure	0
	Skip question	10
Technology list	Smart cards	12
	Biometrics	15
	Mobile phone location systems	8
	Global positioning systems	7
	Intelligent transport systems	7
	Visual surveillance	14
	Digital image and video recording	16
	Profiling techniques	10
	Data mining techniques	11
	Logging of electronic traffic	8
	Other	9
	None	−10
Identifiers	Yes	10
	No	−10
	Not sure	0
	Skip question	6
Identifiers list	Digital signature initiative	9
	Multipurpose identifier	7
	Document with identifiable information	10
	Regulation schemes	7
	Biometric identifiers	11
	Other	8
	None	−10



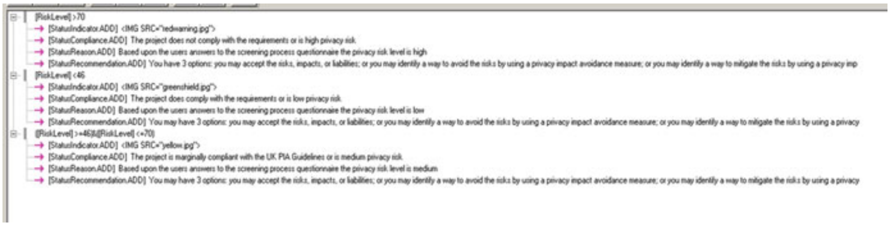
9223372036854775807 875333294789472960465757551991708.html87533329478947 Content-type: text/html

ADS: A PRIVACY IMPACT ASSESSMENT TOOL

STAKEHOLDER FEEDBACK

QUESTION	ANSWER

Fig. 3.27 Link from cvd file to cvr file



```
[RiskLevel] >70
  → [StatusIndicator.ADD] <IMG SRC="redwarning.jpg">
  → [StatusCompliance.ADD] The project does not comply with the requirements or is high privacy risk
  → [StatusReason.ADD] Based upon the users answers to the screening process questionnaire the privacy risk level is high
  → [StatusRecommendation.ADD] You have 3 options: you may accept the risks, impacts, or liabilities; or you may identify a way to avoid the risks by using a privacy impact avoidance measure; or you may identify a way to mitigate the risks by using a privacy imp
[RiskLevel] <46
  → [StatusIndicator.ADD] <IMG SRC="greenshield.jpg">
  → [StatusCompliance.ADD] The project does comply with the requirements or is low privacy risk
  → [StatusReason.ADD] Based upon the users answers to the screening process questionnaire the privacy risk level is low
  → [StatusRecommendation.ADD] You may have 3 options: you may accept the risks, impacts, or liabilities; or you may identify a way to avoid the risks by using a privacy impact avoidance measure; or you may identify a way to mitigate the risks by using a privacy
([RiskLevel] >=46)&([RiskLevel] <=70)
  → [StatusIndicator.ADD] <IMG SRC="yellow.jpg">
  → [StatusCompliance.ADD] The project is marginally compliant with the UK PIA Guidelines or is medium privacy risk
  → [StatusReason.ADD] Based upon the users answers to the screening process questionnaire the privacy risk level is medium
  → [StatusRecommendation.ADD] You may have 3 options: you may accept the risks, impacts, or liabilities; or you may identify a way to avoid the risks by using a privacy impact avoidance measure; or you may identify a way to mitigate the risks by using a privacy
```

Fig. 3.28 Command block showing formulas to display project status

In addition, a global organisation may have different KBs that reflect the PIA Guidelines of several jurisdictions.

In the next section, we discuss the decision-making process of our PIA tool.

3.4.3.6 Decision Making in Our PIA Tool

This section describes the reasoning and decision making of our PIA tool. It appears that simple decision trees that automate business rules are functionally limited by two main factors: the rules are typically black and white with no leeway for special cases and the complexity of logic that can be represented is quite limited such as "yes/no" answers based upon simple logic.

Our PIA tool is different in that it is able to handle very complex problem-solving tasks, involving probabilistic reasoning folding together many factors in reaching a conclusion and recommendation. For example, a typical business rule for stakeholders involved in projects may be "No reports after 10 days", whereby the rule engine would implement this as a simple rule "If days since report created > 10 then refuse access", but what would happen if one of the stakeholders wanted to access the report on day 11? Our PIA tool can be designed to access the stakeholders' history, consider factors that may have delayed the stakeholder in accessing the report, and advise the project manager to make an exception or contact the stakeholder directly, rather than an absolute "NO".

This type of reasoning and decision making in our tool is achieved by the inference engine (IE), allowing complex probabilistic backward chaining (discussed later in this section) logic to be used to solve complex problems in a manner comparable to a human expert. The IE in our tool is used to analyse and combine the individual rules to solve the larger problem and determines [27]:

- What possible answers there are to the problem
- What data is needed to determine if a particular answer is appropriate
- If there is a way to derive or calculate the needed data from other rules
- When enough data is available to eliminate a possible answer, and stop asking unnecessary questions related to it
- How to differentiate between remaining answers
- Which answer is most likely based upon the rules

Backward chaining in our PIA tool is "goal driven", whereby the top-level goals are the possible answers to the problem or potential recommendations. The IE can determine what it needs to meet a particular goal including determining when that goal is met or that a goal cannot be met. Thus, the IE analyses what data is needed to determine if the first possible goal is appropriate for the user. To make this determination, our PIA tool requires data on the specific situation being analysed.

This data can come from other rules, external sources such as databases and spreadsheets, or by asking the user additional questions. Therefore, the IE checks the rules to find one that would be relevant to making this decision. For example, if the HTML report template discussed in Sect. 3.4.3.2 had an embedded variable such as [[ContactAddress]] that did not have a specific rule associated to it. The IE in our tool will ask the user for this value before creating the report because this becomes the new goal of our tool, whereby it supersedes the original goal "create Report". This process of having one goal requiring data, which leads to another goal, can be repeated many times in our tool. Thus, as data becomes available, lower levels goals are met and are dropped off the chain until the IE is able to determine which of the conditions for the initial top-level goal are met, and the recommendation is then presented to the user.

Our PIA tools IE also uses the forward chaining "data-driven" method, whereby the data is already available in the logic of the rules. In this case, the rules are tested sequentially to see what conclusions result. Moreover, in our PIA tool, backward and forward chaining methods are combined, whereby forward chaining is used to run top-level rules and backward chaining is used to derive needed values from other rule modules such as the confidence variable "risk level".

In the next section, we consider the aspects of storing sensitive data in a shared cloud environment and how our PIA tool may minimise the risk.

3.4.4 The PIA Tool and Sensitive Data in the Cloud

This section discusses the storage of sensitive data in the cloud and how our PIA tool may minimise the risk.

Sensitive data encompasses a wide range of information including ethnic or racial origin, political opinion, religious beliefs, memberships, physical or mental health details, criminal or civil offences, as well as PII that relates to customer and contact details [34]. However, as discussed briefly in Sect. 3.4.2, the definition of sensitive data may vary across jurisdictions.

Our PIA tool can record information including contact name, telephone number, project lead name, and stakeholder details. However, answers in the initial full-scale PIA questionnaire may be interpreted by organisations as confidential data, although in some cases organisations may be willing to accept the risk. In addition, the KB itself could be classed as confidential by organisations if the KB was customised to suit their particular needs. For example, if the data gathered and also the customised KB is combined with company policies. To minimise risks, encryption of personal data is feasible, and strongly advisable, if using simple storage. Thus, the PIA tool can make use of a network appliance (or server), called a cloud storage gateway [37].

A cloud storage gateway can provide encryption, authentication, and authorisation, but it is a server that resides at the customer premises and exposes cloud storage services as if they were local storage devices [37]. The gateway is typically packaged as a virtual machine (VM) and translates cloud storage Application Programming Interfaces (APIs), including Representational State Transfer (REST) or Simple Object Access Protocol (SOAP), to block-based storage protocols such as Internet Small Computer System Interface (iSCSI) or Fibre Channel. Additionally, the cloud storage gateway uses local caching to alleviate latency issues and can translate file-based interfaces such as the Network File System (NFS) or Common Internet File System (CIFS) with seamless integration. This is largely due to the fact that cloud storage gateways use standard network protocols and can translate traditional file-based protocols to cached object-oriented storage. An advantage of using this approach is that the administrator (i.e. expert) can modify or update the rules and templates of the PIA tool very easily and quickly without corrupting the application files that are copied by the cloud provider.

Another advantage of using a cloud storage gateway for our PIA tool is the ability to update, at regular intervals, the main files of our PIA tool (i.e. cvr file which is the java runtime servlet file and HTML files that can be accessed in a browser) that are stored in the cloud. For example, Nasuni [37] terms this a “synchronous snapshot”. Thus, after the initial push (where all files are copied to the public cloud and moved into the cache), the snapshot checks each file chunk for changes within the file tree. It then tags new files and altered, corrupted, old, chunks of data as dirty. New files are chunked, and all of the dirty data is then compressed and encrypted. The snapshot then sends each encrypted chunk to the specified cloud and receives the associated keys that allow it to retrieve files in the event of a restore or a cache miss. Once both files and directories have been pushed to the cloud, the snapshot generates a new root directory and tears down the snapshot, ready to start all over again. Therefore, the snapshot uses a number of protection techniques including the duplication, compression, and encryption of each file, before sending them to the cloud. However, the snapshot only forwards changes between the original files and the most recent version and pushes out only what is necessary, thus reducing potential storage costs. Moreover, many cloud storage gateways facilitate the use of encryption techniques and frameworks (e.g. RSA, OpenPGP), whereby the gateway has no access to customer data, as all encryption and decryption happens at the user site.

Also, data at rest which may be used by the PIA tool is generally not encrypted because the problem is that encryption limits data use. In particular, searching and indexing the data becomes problematic. For example, if data is stored in clear text, one can efficiently search for a document by specifying a keyword. This is impossible to do with traditional, randomised encryption schemes. However, there are solutions to this problem including predicate and homomorphic encryption, and private information retrieval (PIR) [38].

Moreover, the data held by the tool cannot be encrypted if processed in the cloud, as it is not yet possible to process encrypted data in an efficient way. Note that techniques for doing this in a non-efficient way are possible including Yao’s protocol for secure two-party computation [39], Gentry’s fully homomorphic encryption scheme [40], and obfuscation (discussed further in Sect. 3.6).

An important factor is that our tool collects information in the form of project and contact details (i.e. names, telephone numbers, and e-mail addresses) that may be considered by organisations and jurisdictional law (i.e. UK Data Protection Act 1998 and EU Directive (95/46/EC)) as sensitive data. Thus, an issue arises when our PIA tool is deployed as a SaaS using an UK cloud provider and accessed by customers that are outside the UK and EU (i.e. transborder data flow restrictions).

In jurisdictions such as the USA, a solution is provided by a framework called "safe harbour". The framework bridges the differences between the US approach on privacy protection with that taken by the EU Directive. Thus, the US organisations who self-certify to the US-EU safe harbour framework ensures the UK cloud provider that they provide adequate privacy protection, as defined by the EU Directive [41].

However, for jurisdictions that do not have any frameworks or agreements with the EU, a possible solution may be the use of redaction software to obscure or remove sensitive information such as names, telephone numbers, and e-mail addresses from our tools results prior to display. For example, RapidRedact [42] is a tool that can be used with our PIA tool to remove the sensitive information and keep it private and confidential. The solution for sensitive data that may be in reports is discussed in Sect. 3.4.3.2 (i.e. Fig. 3.12), whereby manual HTML templates are created using embedded variables. Thus, if an organisation wishes to leave out project, contact, and stakeholder information, all they have to do is omit the variables (i.e. ProjectName, ContactName, etc.).

In the next section, we outline the development methodology adopted for our PIA tool.

3.5 Development Methodology for a PIA Tool in Cloud Computing

In this section, we present details of the methodology used for our PIA tool including the software development methodology, data collection, analysis, results, and modelling.

Stakeholders (i.e. approximately 25) who were generally interested in a PIA tool of some description were initially contacted via e-mail and telephone. These consisted of several backgrounds including software development, security, privacy, records management, networking, and PIAs. Out of the 25, 11 stakeholders were chosen to participate in gathering requirements and providing feedback for our PIA tool and were chosen because of their working experience with PIAs, records management, security, and privacy in organisations in the UK. Typically, feedback from initial conversations and e-mails from the 11 participating stakeholders were mixed but were very encouraging in that several ideas were put forward including the use of open-ended questions for gathering our tools requirements, the use of semi-structured interviews, and the use of MoSCoW rules (discussed in Sect. 3.5.1). The use of MoSCoW rules in gathering requirements for our PIA tool was important, as it helped dictate the style of the interview questionnaire. In addition, after several conversations and e-mails, arrangements were made with the participating stakeholders to hold interviews at their organisations.

The software methodology chosen for the development of the PIA tool is the Dynamic Systems Development Method (DSDM) framework [43]. This is because the framework provides a flexible yet controlled process that can be used to deliver solutions in tight project timescales (i.e. 3–6 months). Furthermore, a fundamental assumption of the framework is that nothing is built perfectly first time but that as a rule of thumb 80 % of the solution can be produced in 20 % of the time that it would take to provide the total solution. This is in contrast to the classical, sequential "waterfall" approach, whereby the next step cannot be started until the current step is completed that results in projects being delivered late, usually over budget, or fail to meet business needs since time is not spent reworking the requirements. Moreover, the DSDM framework incorporates several important techniques that benefit the development of the PIA tool including [43]:

- *Timeboxing*: This is a planning technique that divides the development into time periods (i.e. usually 4–6 weeks long with each part having its own set deadline and a set of deliverables).
- *MoSCoW prioritisation*: This technique reaches a common understanding with stakeholders on the importance they place on the delivery of each requirement of the PIA tool. Thus, *M* equates to "must" have, *S* equates to "should" have, *C* equates to "could" have, and *W* equates to "won't" have.

In the next section, we discuss data collection, data analysis, and present a summary of findings for our PIA tool.

3.5.1 Data Collection, Analysis, and Findings

In this section, we consider data collection, data analysis, and present a summary of findings of the requirements for our PIA tool.

Prior to any data collection, it was agreed with participating stakeholders that the MoSCoW rules were set at the values:

- *Must have* =>4 points
- *Should have* =< 3 points
- *Could have* =>2 points
- *Won't have* =>1 point

These values were set because there was no indication of how many stakeholders would answer questions about the requirements for a PIA tool. Furthermore, an agreement was reached that the development would initially try to deliver all the *M*, *S*, and *C* requirements, but the *S* and *C* requirements will be the first to go if the delivery timescale looks threatened. Moreover, agreements were made that the value of "very high" corresponded to the MoSCoW rule of "must" have, and the values of "low, very low" corresponded to the rule of "won't" have.

The collection of data consists of a questionnaire (e.g. formulated to include both close-ended and open-ended questions) that is used to elicit from target stakeholders their emotional opinions about privacy, PIAs, and the requirements for our PIA tool [44]. To satisfy the research objectives, the study's methodology employed a series of ten semi-structured interviews with a mixture of private and public sector stakeholders in four geographical locations in the UK: the county borough of

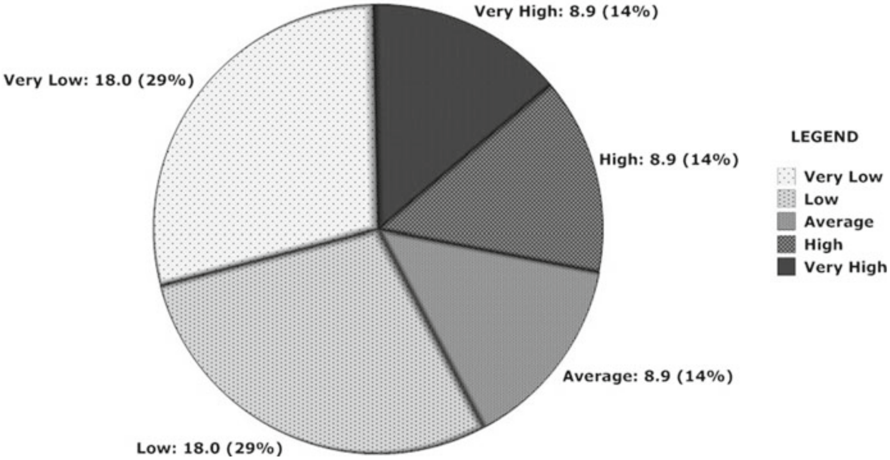


Fig. 3.29 Findings for separate stakeholder analysis page

Torfaen, the metropolitan area of Bristol, the home counties including London, and Essex. Interviews, each lasting approximately 45 min–1 h were conducted between July and September 2011. They were segregated into privacy/security officers, record officers, and information officers to enhance the opportunity for different discussions, opinions, and perspectives.

Analysis of the raw data indicates that opinions and perspectives of the topics discussed differed significantly between the interested parties. For example, regarding the issue of whether privacy was an important factor within their organisation, privacy officers naturally "valued highly" this factor. However, records and information officers suggested that privacy was "not important" or a major concern. Moreover, most stakeholders interviewed (e.g. 80 %) agreed that PIAs are necessary and that they should be adopted for their organisation and that PIAs must start at the "beginning of development". In addition, one of the most notable findings to emerge from the study is that 70 % of the stakeholders interviewed desire an automated PIA tool to help them in this process.

To convert the raw data into requirements (i.e. functional and non-functional), each stakeholder's answer relating to the functionality of the PIA tool is given a value based upon the agreed MoSCoW rules. For example, one question in the questionnaire is about whether the PIA tool should incorporate a stakeholder analysis screen, whereby the findings of this particular question are illustrated in Fig. 3.29.

Each value (e.g. very high, low) is then given a number of points such as high = 4 points. These are then multiplied with the percentage of stakeholders answering that particular value for the question. For example, 14 % of stakeholders answered this question very high, which equates to 14 × 5 = 70. This formula is then applied to all

Table 3.7 MoSCoW rules applied to UI questions

Prioritised list of user interfaces for PIA tool	
Name of user interface	**MoSCoW rule**
Security log-in	Could have
Welcome	Should have
Project information	Must have
Contact information	Must have
Stakeholder analysis	Could have
Communication strategy	Won't have
Environmental scan	Could have
Questionnaire	Must have
Display of results	Must have
Report	Must have

values of the question to give the total number of points for the question (e.g. 255 points). To convert this number into the agreed MoSCoW rule, the total number of points for the question is then divided by 100 (the total percentage) to compute the average value. For example, using this formula, the average value of the question is 255/100 = 2.55, which equates to the MoSCoW rule of "could" have for the question. This method is the applied to all of the other questions about the functionality of the PIA tool, as illustrated in Table 3.7.

Furthermore, correlation techniques such as pattern matching are applied to the raw data to reveal common stakeholder phrases and words such as "I must have that", "I don't like that" or "that is good", and it appears that these phrases and words can be directly interpreted into MoSCoW rules to give the requirements for our PIA tool [44]. For example, a number of functional and non-functional requirements for the project information UI are illustrated in Table 3.8.

In the next section, we discuss modelling the user requirements for our PIA tool.

3.5.2 Modelling of User Requirements for Our PIA Tool

In this section, we discuss modelling the user requirements for our PIA tool. In DSDM, the term modelling refers to Unified Modelling Language (UML) diagrams [43]. To illustrate modelling the user requirements for our PIA tool, we use the project information requirements discussed in the previous section. For example, Table 3.9 describes the use case requirements in detail.

The use case description in Table 3.9 is then converted into a use case diagram. For example, the third iteration for the functionality of the project information screen is illustrated in Fig. 3.30, and the activity diagram for the project name is illustrated in Fig. 3.31.

In the next section, we consider validation of our PIA tool.

Table 3.8 Requirements for project information UI

Requirements for project information		
Functional requirements		
Name	Label	Requirement
Interface	Project name	String data entry only
	Project title	String data entry only
	Project description	String data entry only
	Project lead	String data entry only
	Telephone no.	Numeric data entry only
Business		Data entered by user is stored in database
		Clicking the Back button moves the user request to the Welcome screen
		Clicking the Restart button moves the user request to the Start screen
		Clicking the OK button moves the user request to the next question
Regulatory/compliance		The database will have a functional audit trial
Security		Administrators can edit and delete project information
Non-functional requirements		
Name	Label	Requirement
User		User must be able to access the project information 23 h a day, 7 days a week
		User is not allowed to delete project information
System		System must be unavailable between midnight and 1.00 am for backups

Table 3.9 Use case requirements for project information UI

Use case requirements of project information	
Use case	Description
View project information	User views the project information screen
Enter project name	User enters a string that represents the project name
Enter project title	User enters a string that represents the project title
Enter project lead	User enters a string that represents the name of the project leader
Enter telephone no.	User enters a numeric value that represents the telephone number of the project leader
Enter project description	User enters a string that represents the description of the project
Restart button	If user clicks this button, the system re-starts the process
Back button	If user clicks this button, the system moves back to the last screen or question viewed
OK button	If user clicks this button, the system moves forward to the next screen or question

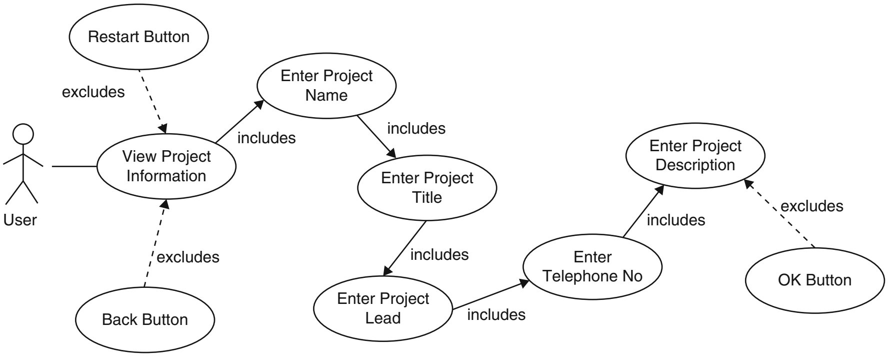


Fig. 3.30 Use case diagram for project information UI

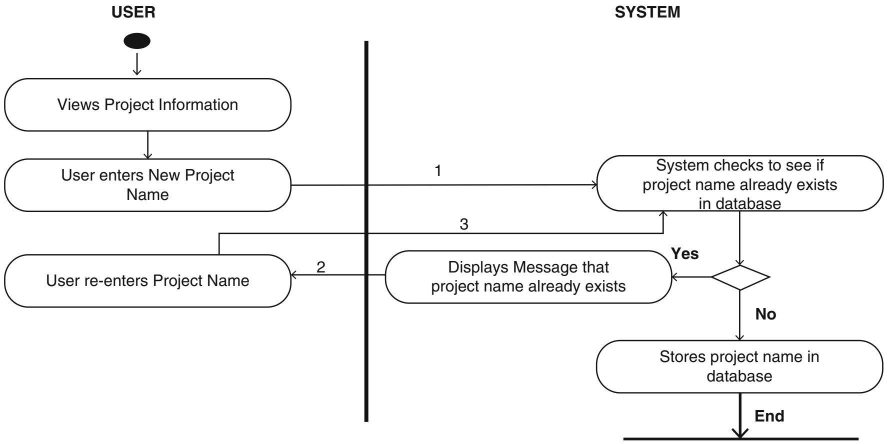

USER
SYSTEM
Views Project Information
User enters New Project Name
1
System checks to see if project name already exists in database
3
Yes
User re-enters Project Name
2
Displays Message that project name already exists
No
Stores project name in database
End

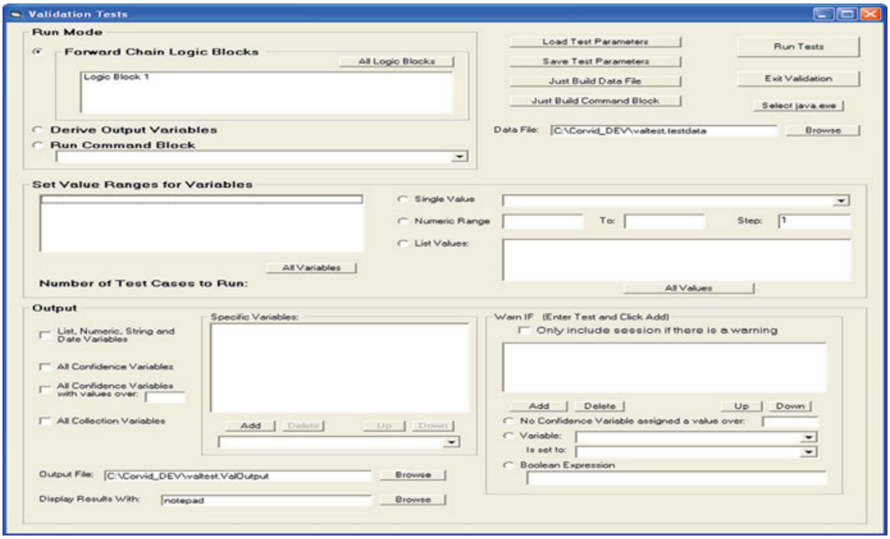
Fig. 3.31 Activity diagram for project name use case

3.5.3 *Validation of Our PIA Tool*

This section discusses validation of our PIA tool. Testing our PIA tool is important as it helps to provide quality assurance, verification and validation, and reliability estimation. Corvid provides a validation function, as illustrated in Fig. 3.32 [27].

This function enables automating very large numbers of tests, along with setting various warning tests to check for specific types of issues in the system.

Validation testing in Corvid allows for a specific logic block or subset of a system (e.g. single or multiple variables) to be tested, allowing thorough testing of even large systems. Once the parameters for the validation test are set, the tests run automatically without additional user input. Thus, Corvid displays the number of tests that will



● Forward Chain Logic Blocks: Logic Block 1

Data File: C:\Corvid_DEV\valtest.testdata

Step: 1

Output File: C:\Corvid_DEV\valtest.ValOutput

Display Results With: notepad

Fig. 3.32 Validation testing in Corvid

have to be run based on the selected parameters and displays the progress as the tests are being executed. For larger tests, they can be allowed to run over night or longer as needed. A file is generated with any errors and problems that are detected and special system warnings that are generated [27]. In addition, the validation test parameters can also be saved to a file, so that the same tests can be run again later to check any modifications to the system. However, it does not understand the actual validity and correctness of the rules in the system. Therefore, it is the responsibility of the developer to make sure that the actual logic and advice given is correct, and only the author and domain expert can assure that a system is giving the correct answers and advice. For example, a specific logic block may be used to set the value for a variable that is based upon user input. Thus, it may be easier to test this block separately from the rest of the system to analyse whether it is setting the correct values. This allows the developer to focus on this detail without the influence of the rest of the system and once that part is validated the logic block can be used in a more extensive test. Moreover, the process of user validation (e.g. tests carried out by the users on the functionality of the PIA tool) has not started, so there is currently no feedback from users on the tool. It is hoped that this process will be completed in the near future.

In the next section, we consider related work in the areas of privacy and security in cloud computing to evaluate whether these approaches are suitable to aid enhancement of our PIA tool.

3.6 Related Work

In this section, we consider related work in the areas of privacy and security in cloud computing to evaluate whether these approaches are suitable to aid enhancement of our PIA tool.

Accountability as a way forward for privacy protection in the cloud is considered by Pearson and Charlesworth [45]. They propose the incorporation of complementary regulatory, procedural, and technical provisions that demonstrate accountability into a flexible operational framework to address privacy issues within a cloud computing scenario. They believe that accountability is a useful basis for enhancing privacy in many cloud computing scenarios, as corporate management can quickly comprehend its links with the recognised concept of, and mechanisms for achieving, corporate responsibility. Accountability in this context is corporate data governance (i.e. the management of the availability, usability, integrity, and security of the data used, stored, or processed within an organisation), and it refers to the process by which a particular goal – the prevention of disproportionate (in the circumstances) harm to the subjects of PII – can be obtained via a combination of public law (legislation, regulation), private law (contract), self-regulation, and the use of privacy technologies (system architectures, access controls, machine readable policies). The approach taken requires a combination of procedural and technical measures to be used and co-designed. In essence, this would use measures to link organisational obligations to machine readable policies and mechanisms to ensure that these policies are adhered to by the parties that use, store, or share that data, irrespective of the jurisdiction in which the information is processed. Companies providing cloud computing services would give a suitable level of contractual assurances, to the organisation that wishes to be accountable, that they can meet the policies (i.e. obligations) that it has set, particularly PII protection requirements. Furthermore, technology can provide a stronger level of evidence of compliance and audit capabilities. However, while the approach appears to be a practical way forward, it has limitations. For example, while contracts provide a solution for an initial service provider to enforce its policies along the chain, risks that cannot be addressed contractually will remain, as data has to be unencrypted at the point of processing, creating a security risk and vulnerability due to the cloud's attractiveness to cybercriminals. Moreover, only large corporate users are likely to have the legal resources to replace generic SLAs with customised contracts.

Obfuscation, as a first line of defence is described by Pearson et al. [46]. This chapter describes a tool called "privacy manager", which they believe reduces the risk to the cloud computing user of their private data being stolen or misused and also assists the cloud computing provider to conform to privacy law. The idea is that instead of being present unencrypted in the cloud, the user's private data is sent to the cloud in an encrypted form, and the processing is done on the encrypted data. The output of the processing is de-obfuscated by the privacy manager to reveal the correct result. The obfuscation method uses a key which is chosen by the user and known by the privacy manager but is not communicated to the service provider. Thus, the service provider is not able to de-obfuscate the user's data, and the un-obfuscated data is never present on the service provider's machines.

Although some obfuscation methods are highly susceptible to known plaintext attacks [46], this does at least protect the data from opportunistic data thieves with access to cloud databases because it ensures that the data is never present in the database in the clear.

Use of DSSs for cloud computing and PIAs is a very new field and there are few systems available. Those that are available for cloud computing are found in the areas of clinical decision applications [47] and life science enterprise solutions [48]. However, very recently, there has been a step change in DSS for PIAs (such as privacy expert systems). Typically, a DSS has a KB that needs to be created and updated periodically by experts on an ongoing basis and a mechanism (e.g. a rules engine, decision tree, or dedicated queries to databases) by which output can be generated, based upon user input via questionnaires. Within this context, we will discuss briefly two DSSs that are at the cutting edge of research.

PRAIS [49] is a research project that has developed a prototype DSS tool for context-sensitive privacy-aware information sharing in children's social care. The DSS is based on the architecture developed for the Identity Governance Framework (IGF) [50], where information sharing is based on a pull model. This means that the recipients are alerted that information is being made available to them, after which it is retrieved from the source. PRAIS uses the IGF architecture as its design choice because it allows the owner of the information to retain liability for the data and to audit each use by using the pull model. Therefore, PRAIS is a DSS tool that enables personnel working with personal information to assess the privacy implications of information sharing actions dynamically and to share information with confidence, whether verbally or electronically. This has been achieved by accommodating the daily routines of social care staff from the outset, whereby it manages users consent and the needs and requests of information from the participants.

However, analysis suggests that the scope of PRAIS is very narrow as it is not intended that the DSS will ever make decisions on behalf of properly trained personnel but instead will assist social care practitioners in making privacy-aware decisions where required. Therefore, it appears that the DSS is designed to assist in the professional's decision-making process and not to replace it. Moreover, one of the main findings is that although PRAIS can be used for sharing information electronically, this may not necessarily be its primary purpose. This is because in social care, information is often shared in multi-agency meetings or over the telephone. Thus, the system can be used by practitioners on an ad hoc basis to explore privacy implications where information may be shared verbally. In summary, PRAIS in its current format is not applicable for the UK PIA tool although some approaches such as the use of an expert system may be considered.

Hewlett Packard's Privacy Advisor (HPPA) is an expert system that captures data about business processes to determine their privacy compliance [51, 52]. The tool helps organisations to ensure privacy concerns are met and supports enterprise accountability, supplying employees with sufficient information and guidance to ensure that they design and conduct their projects in compliance with organisational privacy policies. HPPA uses a rules engine for which rules are defined that are used both to generate questions that are customised to the employee's specific situation and to codify HP's privacy rulebook and other information sources. Based on the employee's response to these questions, it automatically generates an output report that includes analysis of possible privacy risks and a checklist of actions that the employee should take in order to mitigate these risks. This tool has been rolled out to employees within HP.

Analysis of this tool indicates that the methods and techniques used in HPPA are well suited for the UK PIA tool, such as the use of its knowledge representation and inference methods (i.e. rules, dynamic questionnaires, and report generation), and knowledge management (i.e. user modes, interfaces, and its reasoning about global requirements and regulations). However, it will be necessary to modify HPPAs methods and techniques to fit the UK PIA tool because HPPA is based on a customised set of organisational policies, which would need to be different for other organisations; therefore, it is not generic.

A new self-assessment tool, aimed at private sector organisations, particularly small- and medium-sized businesses, was recently launched in Canada (e.g. May 2011). The tool developed jointly by the federal, Alberta and British Columbia privacy commissioners' offices is called "Securing Personal Information: A Self-Assessment Tool for Organizations", where it is hoped that the tool may help businesses better safeguard the personal information of customers and employees and may help prevent breaches of PII [53].

The tool is a detailed online questionnaire that helps organisations gauge how well they are protecting personal information and meeting compliance standards under Canada's private sector privacy law on both federal and provincial levels. The questionnaire is complex and not easy to navigate, as it involves dozens of "yes" or "no" questions divided up into 17 different categories including network security, access control, incident management, and database security. However, it offers some flexibility by allowing users to focus on areas most relevant to their own enterprise. The goal of the tool is for organisations to be able to answer "yes" to each question, and at the end of the process, results for the minimum and higher levels of security are tabulated separately.

The main disadvantage of the tool is that its usage is voluntary, and hence, a comprehensive evaluation of an organisation's internal policies may not be easy to complete. This can also be the case because users who are not experts may have difficulty in understanding the questions. For example, the questions under the assessment "risk management" section indicate that an IT expert is required to provide answers.

Similar tools exist and are freely available from vendors such as Microsoft [54]. For example, the Microsoft "Security Assessment Tool" is also designed to help find weaknesses in an IT security environment and offers a download that takes a snap shot of an organisations current security state. However, the new tool from Canada's privacy commissioners focuses on privacy and protecting personal information rather than the more common security paradigm of protecting intellectual property.

Analysis of these tools indicates that they are composed of simple decision trees that follow a straightforward approach that provides advice based on users answers. For example, the user starts at the first question, and whether they answer “yes” or “no”, they are forwarded to the next question, until they reach the end of the questionnaire when a report is produced based upon their answers. As discussed in Sect. 3.4.3.6, these simple decision trees do not allow for complex reasoning as the rules are typically black and white with no leeway for special cases such as global regulations and transborder data flows, and the complexity of logic that can be represented is quite limited such as “yes/no” answers that is based upon simple logic.

Sander and Pearson [55] outline a DSS for cloud computing that aids selection of appropriate cloud service providers (CSPs). Their approach is a semi-automated DSS tool that gathers context relating to CSPs and inputs to a rule-based system to trigger decisions about whether or not to use that CSP and/or to determine additional stipulations that would need to be made. The tool helps to determine appropriate actions that should be allowed and assesses risk before personal information is passed on through the cloud. For each customer enterprise, an administrator will set up the original questionnaire according to the policies that the customer (i.e. the enterprise) wishes to check or use the default setting offered by the assessment service. When a customer wishes to assess different CSPs offering a service, providers will use the tool via a Web interface in order to provide answers to the questionnaire, and the results will be sent back to the enterprise that wishes to choose between the service providers. These results include reports and automatically generated ratings, which will allow the administrator to distinguish between them. This tool is similar to the HPPA tool, in that it is a form of expert system using a set of intermediate variables (IMs) to encode meaningful information and to drive the questionnaire generation.

Although there are some similarities between this tool, HPPA and our PIA tool, there are significant differences in architecture and deployment, the underlying mechanism for the knowledge representation and for generating questionnaires, and the rules, report structure, and output.

The next section will discuss next steps associated with the development of our PIA tool.

3.7 Next Steps

As the prototype for the tool is only at the first iterative stage, our next planned steps include the following:

1. Conducting another round of stakeholder meetings that includes a presentation of the working tool. This is for validation purposes and to elicit further user requirements.
2. Developing the tool further to include all necessary, and some preferable, requirements.
3. Considering a cloud storage gateway provider for provision of infrastructure that protects the PIA tool’s customer data in the cloud.

3.8 Conclusions

We are currently developing a PIA tool that can be used in a cloud environment to identify potential privacy risks and compliance issues. The tool addresses the inherent complexity and helps both expert and non-expert end users with identifying and addressing privacy requirements for a given context. As part of this approach, we provide mechanisms for privacy experts and other authorised non-technical personnel to modify the KB in our tool in an intuitive way.

If our PIA tool is used as a SaaS application itself, regulatory issues such as transborder data flow can be involved because personal information may need to be accessed from and transferred to different jurisdictions.

References

1. Stewart, B.: Privacy impact assessments. PLPR **3**(7), 61–64 (1996). http://www.austrii.edu/au/journals/PLPR.html. Accessed 30 Oct 2011
2. Warren, A., Bayley, R., Bennett, C., Charlesworth, A., Clarke, R., Oppenheim, C.: Privacy impact assessments: international experience as a basis for UK guidance. Comput. Law Secur. Rep. **24**(3), 233–242 (2008). doi:10.1016/j.clsr.2008.03.003
3. Tancock, D., Pearson, S., Charlesworth, A.: Analysis of privacy impact assessments within major jurisdictions. In: Privacy Security and Trust (PST), 2010 Eighth Annual International Conference, Ottawa, Ontario, Canada, 17–19 Aug 2010, pp. 118–125 (2010). doi: 10.1109/PST.2010.5593260
4. Tancock, D., Pearson, S., Charlesworth, A.: The emergence of privacy impact assessments. http://www.hpl.hp.com/techreports/2010/HPL-2010–63.pdf (2010). Accessed 30 Oct 2011
5. Cavoukian, A.: Privacy by design: the 7 foundational principles. http://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf (2009). Accessed 30 Oct 2011
6. Charlesworth, A.: Jurisdictional report for Canada: privacy impact assessments. International Study of Their Application and Effects (Appendix C). http://www.ico.gov.uk/upload/documents/library/corporate/research_and_reports/lbrouni_piastudy_appc_can_2910071.pdf (2007). Accessed 26 Oct 2011
7. eHealth Ontario: Privacy impact assessment policy version 2. http://www.ehealthontario.on.ca/pdfs/Privacy/PrivacyImpactAssessmentPolicy.pdf (2008). Accessed 27 Oct 2011
8. Cavoukian, A.: Privacy impact assessment guidelines for the Ontario Personal Health Information Protection Act. http://www.ipc.on.ca/images/Resources/up-phipa_pia_e.pdf (2005). Accessed 23 Oct 2011
9. UK Cabinet Office: Data Handling Procedures in Government: Final Report. http://www.cabinetoffice.gov.uk/sites/default/files/resources/final-report.pdf (2008). Accessed 21 Nov 2011
10. Treasury Board Secretariat Canada: Info source bulletin number 33B: statistical reporting. http://www.infosource.gc.ca/bulletin/2010/b/bulletin33b/bulletin33b03-eng.asp (2010). Accessed 15 Nov 2011
11. Information Commissioners Office: Information Commissioners annual report 2009/10. http://www.ico.gov.uk/upload/documents/library/corporate/detailed_specialist_guides/annual_report_2010.pdf (2010). Accessed 14 Nov 2011
12. 80/20 Thinking Ltd: Privacy impact assessments for Phorm Inc. http://www.8020thinking.com/news/9.html?task=view (2008). Accessed 26 Oct 2011
13. The Office of the Privacy Commissioner of New Zealand: Privacy impact assessment handbook. http://privacy.org.nz/assets/Files/Brochures-and-pamphlets-and-pubs/48638065.pdf (2009). Accessed 27 Oct 2011

14. Information Commissioners Office: Privacy impact assessment handbook. http://www.ico.gov.uk/handbook/June.2009 (2009). Accessed 30 Oct 2011
15. Department of Homeland Security: Privacy Threshold Analysis (PTA). http://www.dhs.gov/xlibrary/assets/privacy/DHS_PTA_Template.pdf (2007). Accessed 30 Oct 2011
16. Australian Government: Office of the Privacy Commissioner: Privacy Impact Assessment Guide. http://www.privacy.gov.au/index.php?option=com_icedoc&view=types&element=guidelines&fullsummary=6590&Itemid=1021 (2010). Accessed 29 Oct 2011
17. Treasury Board Secretariat Canada: Welcome to the PIA e-learning tool. http://www.tbs-sct.gc.ca/pgol-pged/piatp-pfefvp/index-c-eng.asp (2003). Accessed 13 Nov 2011
18. United States Department of Homeland Security: Privacy threshold analysis (PTA). http://www.dhs.gov/xlibrary/assets/privacy/DHS_PTA_Template.pdf (2007). Accessed 15 Nov 2011
19. Vallini, M.: Software as a Service (SaaS) Ethical Issues. http://www.marcovallini.com/documentazione/saas_ethical_issues.pdf (2009). Accessed 24 Oct 2011
20. Pearson, S., Benameur, A.: Privacy, security and trust issues arising from cloud computing. In: Cloud Computing Technology and Science (CloudCom), 2010 Second Annual International Conference, 30 Nov –3 Dec, pp. 693–702 (2010). doi: 10.1109/CloudCom.2010.66
21. Giarratano, J.: Expert Systems: Principles and Programming. Thomson Learning, Boston, Massachusetts, Canada (2005)
22. J Boss Community: Drools: Business logic integration/platform. http://www.jboss.org/drools (2011). Accessed 22 Nov 2011
23. Logic Programming Associates Ltd: LPA VisiRule 1.5. http://www.lpa.co.uk/vsr.htm (2011). Accessed 20 Nov 2011
24. Corvid ExSys: Overview of Corvid Knowledge Automation Expert System Software. http://www.exsys.com/pdf/AboutCORVID.pdf (2008). Accessed 18 Oct 2011
25. Corvid ExSys: Java-based Expert System Knowledge Automation Development and DeploymentTechnologiesWhite-Paper.http://www.exsys.com/pdf/ExsysCORVIDWhitePaper.pdf (2009). Accessed 20 Oct 2011
26. Mell, P., Grance, T.: The National Institute of Standards and Technology (NIST) Definition of Cloud Computing Version 15. http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf (2009). Accessed 19 Nov 2011
27. Corvid ExSys: ExSys Corvid Manual Version 5.2.1. http://www.exsys.com/PDF/CorvidManual.pdf (2009). Accessed 16 Oct 2011
28. Microsoft: What is the Windows Azure Platform? http://www.microsoft.com/windowsazure/ (2011). Accessed 21 Nov 2011.
29. Sitaram, D., Manjunath, G.: Moving to the Cloud. Elsevier, Waltham (2012)
30. Sosinsky, B.: Cloud Computing Bible. Wiley, Indianapolis (2011)
31. Rhoton, J.: Cloud Computing Explained, 2nd edn. Recursive Press, London (2010)
32. Trusted Computing Group: TCG Architecture Overview Version 1.4. http://www.trustedcomputinggroup.org/resources/tcg_architecture_overview_version_14 (2010). Accessed 12 Nov 2011
33. European Network and Information Security Agency: Cloud Computing: Benefits, Risks and Recommendations for Information Security. http://www.enisa.europa.cu/act/rm/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport.pdf (2009). Accessed 17 Nov 2011
34. Solove, D.J.: Understanding Privacy. Harvard University Press, Cambridge (2008)
35. Organization for Economic Co-operation and Development (OECD): Guidelines Governing the Protection of Privacy and Transborder Flow of Personal Data. OECD, Geneva (1980)
36. Karol, T.: A guide to cross-border privacy impact assessments. http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/A-Guide-To-Cross-Border-Privacy-Impact-Assessments.aspx (2009). Accessed 30 Oct 2011
37. Nasuni: http://www.nasuni.com/ (2011). Accessed 30 Oct 2011
38. Bethencourt, S., Chan, J., Song, D., Perrig, A.: Multi-dimensional range query over encrypted data. In: IEEE Symposium on Security and Privacy. [City Not Specified] http://www.cs.berkeley.edu/~bethenco/oakland07rangequery.pdf (2007). Accessed 19 Feb 2012
39. Yao, A.C.: How to generate and exchange secrets. In: 27th Symposium on Foundation of Computer Science (FoCS), 27–29 Oct 1986, pp. 162–167 (1986). doi: 10.1109/SFCS.1986.25

40. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: 41st ACM Symposium on Theory of Computing (STOC), [Date not specified] June 2009, pp. 169–178 (2009). doi: 10.1145/1536414.1536440
41. Export.gov: Welcome to the U.S.-E.U. & U.S.-Swiss Safe Harbor Frameworks. 26 October 2011. http://export.gov/safeharbor/ (2011)
42. RapidRedact: Welcome to Redacta. 21 Feb 2012. http://www.redacta.co.uk/ (2012)
43. DSDM Consortium: Handbook Version 2.1. http://www.dsdm.org/atern-handbook/flash.html#/1/ (2011). Accessed 30 Oct 2011
44. Shull, F., Singer, J., Sjoberg, D.: Guide to Advanced Empirical Software Engineering. Springer, London (2010)
45. Pearson, S., Charlesworth, A.: Accountability as a way forward for privacy protection in the cloud. In: Jaatun, M., Zhao, G., Rong, C. (eds.) Cloud Computing, vol. 5931, pp. 131–144. LNCS. Springer, Berlin/Heidelberg (2009). doi:10.1007/978–3–642–10665–1_12
46. Pearson, S., Shen, Y., Mowbray, M.: A privacy manager for cloud computing. In: Jaatun, M., Zhao, G., Rong, C. (eds.) Cloud Computing, vol. 5931, pp. 90–106. LNCS. Springer, Berlin/Heidelberg (2009). doi: 10.1007/978–3–642
47. Preimesberger, C.: IBM, Aetna Join for New Cloud-Based Health Care Support System. http://www.eweek.com/c/a/Health-Care-IT/IBM-Aetna-Join-for-New-CloudBased-Health-Care-Support-System-667092/ (2010). Accessed 30 Oct 2011
48. CambridgeSoft: ChemBioOffice Cloud – An Integrated Decision Support System for CHDI. http://chembionews.cambridgesoft.com/WhitePapers/Default.aspx?whitePaperID=43 (2010). Accessed 30 Oct 2011
49. Harbird, R., Ahmed, M., Finkelstein, A., McKinney, E., Burroughs, A.: Privacy Impact Assessment with PRAIS. http://www.cs.ucl.ac.uk/staff/A.Finkelstein/papers/hotpets.pdf (2007). Accessed 30 Oct 2011
50. Liberty Alliance Project: ID governance – identify privacy and access policy, marketing requirements document. http://www.projectliberty.org/ (2007). Accessed 30 Oct 2011
51. Pearson, S., Sander, T., Sharma, R.: Privacy management for global organizations. In: Garcia-Alfaro, J., Navarro-Arribas, G., Cuppens-Boulahia, N., Roudier, Y. (eds.) Data Privacy Management and Autonomous Spontaneous Security, vol. 5939, pp. 9–17. LNCS. Springer, Berlin/Heidelberg. doi:10.1007/978–3–642–11207–2_2
52. Pearson, S., Rao, P., Sander, T., Parry, A., Paull, A., Patruni, S., Dandamudi-Ratnakar, V., Sharma, P.: Scalable, accountable privacy management for large organizations. INSPEC 2009: 2nd International Workshop on Security and Privacy Distributed Computing, Enterprise Distributed Object Conference Workshops (EDOCW 2009), IEEE, Auckland, New Zealand, 1–4 Sept 2009, pp. 168–175
53. Office of the Privacy Commissioner of Canada: Securing Personal Information: A self Assessment Tool for Organisations. http://www.priv.gc.ca/resource/tool-outil/security-securite/english/AssessRisks.asp?x=1 (2011). Accessed 26 Oct 2011
54. Microsoft: Microsoft Security Assessment Tool Version 4.0. http://www.microsoft.com/download/en/details.aspx?displaylang=en&id=12273 (2009). Accessed 27 Oct 2011
55. Sander, T., Pearson, S.: Decision support for selection of cloud service providers. Int. J. Comput. GTSF. **1**(1) (2010)