On Identity Analytics: Setting the Context

Marco Casassa Mont, Adrian Baldwin, Simon Shiu
HP Laboratories
HPL-2008-84

Keyword(s):
Identity Analytics, Modelling, Simulation, Security Analytics, Economics, Trade-offs, Policies, Identity Management

Abstract:
This paper aims at setting the context for "Identity Analytics" within enterprises and paving the path towards new R&D opportunities. In our vision, Identity Analytics is about explaining and predicting the impact of identity and identity management (along with other related aspects, such as users' behaviours) on key factors of relevance to decision makers (e.g. CIOs, CISOs), in complex enterprise scenarios – based on their initial assumptions and investment decisions. Ultimately the goal is to provide rigorous techniques to help decision makers gain a better understanding of the investment trade-offs within the identity space (e.g. investing in technologies vs. changing processes vs. investing in users' education, etc.). This means providing "decision support" and "what-if analysis" capabilities to decision makers enabling them to explore these investment trade-offs, formulate new policies and/or justify existing ones. Our vision of "Identity Analytics" is introduced and discussed, along with the methodology that we intend to adopt.

There are many research opportunities and challenges in this space: we believe that a scientific approach is required, involving the usage of modelling and simulation techniques, coupled with the understanding of involved technologies and processes, human behaviours and economic aspects. To ground some of the concepts discussed in this paper, we provide an illustration of Identity Analytics focusing on emerging "web 2.0 enterprise collaborative data sharing", where unstructured information is created, stored and shared by people in collaborative contexts, within and across organisations. We demonstrate how trade-offs can be explored using the modelling approach hence allowing decision makers to explore the different impacts of policy choices.

External Posting Date: July 6, 2008 [Fulltext] Approved for External Publication

Internal Posting Date: July 6, 2008 [Fulltext]

On Identity Analytics: Setting the Context

Marco Casassa Mont, Adrian Baldwin, Simon Shiu
Hewlett-Packard Laboratories, Systems Security Lab, Bristol, UK
marco.casassa-mont@hp.com, adrian.baldwin@hp.com, simon.shiu@hp.com

Abstract

This paper aims at setting the context for "Identity Analytics" within enterprises and paving the path towards new R&D opportunities. In our vision, Identity Analytics is about explaining and predicting the impact of identity and identity management (along with other related aspects, such as users' behaviours) on key factors of relevance to decision makers (e.g. CIOs, CISOs), in complex enterprise scenarios – based on their initial assumptions and investment decisions. Ultimately the goal is to provide rigorous techniques to help decision makers gain a better understanding of the investment trade-offs within the identity space (e.g. investing in technologies vs. changing processes vs. investing in users' education, etc.). This means providing "decision support" and "what-if analysis" capabilities to decision makers enabling them to explore these investment trade-offs, formulate new policies and/or justify existing ones. Our vision of "Identity Analytics" is introduced and discussed, along with the methodology that we intend to adopt.

There are many research opportunities and challenges in this space: we believe that a scientific approach is required, involving the usage of modelling and simulation techniques, coupled with the understanding of involved technologies and processes, human behaviours and economic aspects. To ground some of the concepts discussed in this paper, we provide an illustration of Identity Analytics focusing on emerging "web 2.0 enterprise collaborative data sharing", where unstructured information is created, stored and shared by people in collaborative contexts, within and across organisations. We demonstrate how trade-offs can be explored using the modelling approach hence allowing decision makers to explore the different impacts of policy choices.

1. Introduction

The aim of this paper is to set the context for "Identity Analytics" and pave the path towards new research opportunities in this space.

In our vision "Identity Analytics" is about a set of methodologies, approaches and solutions to explain and predict the impact that identity and identity management (along with other aspects, such as users' behaviours, environment and social aspects, etc.) have on factors of relevance to decision makers (e.g. CIOs, CISOs), such as costs, exposure to security risks, compliance, trust and reputation, in well defined scenarios. This area represents a new research opportunity, specifically for enterprise identity management.

Current solutions in the space of (enterprise) identity management primarily focus on "control point" solutions and compliance-driven solutions. The former includes access control, authorization, authentication, single-sign-on (SSO), federation, etc. solutions. The latter includes auditing, data, events and logs processing capabilities, reporting tools and assessment of compliance against established processes and policies.

These solutions are deployed in consolidated enterprise IT infrastructures and environments, and are configured to implement and deal with well defined sets of security policies. However, other important aspects have an impact on the effectiveness and value of these solutions including users' behaviours within organisations; their education and awareness of policies; the introduction of new, emerging technologies; and social trends. Specifically, new emerging enterprise technologies and users' behaviours are right now creating new challenges and issues: how to control the explosion of (potentially confidential or personal) information and data that is generated, stored and exchanged by using web 2.0 collaborative tools; how to cope with multiple identities and personae that employees and users might create and use within and outside organisations; how to control data flows involving interactions both between employees and externally to the organisation, with social networks and other external sites; how to deal with the increase of identity and credential thefts, identity phishing and privacy violations.

New technological trends in this space include: the increasing growth of heterogeneous and unstructured data needing to be managed by organisations because of employees' adoption of enterprise web 2.0 collaborative solutions (including user-centric collaborative tools, such as Twiki, MS Sharepoint, etc.); the rise of social networks and the blurring

boundaries between organizational and personal/private worlds.

On one hand the adoption of new mobile solutions, communication and collaborative tools, enterprise web 2.0 technologies, etc. provides users with unprecedented ways to share information, collaborate and make progress on common projects and tasks.

On the other hand (identity management) "control point" solutions are showing their limits in these new contexts, in terms of their efficacy in controlling and protecting sensitive, personal and confidential data (and related applications and services) and related flows of this information: this is due to the lack of a clear understanding of the (security and business) risks due to the implications of these changes. It is becoming increasingly hard, for decision makers, to fully understand and predict the implications and impacts that these new technological, social and behavioural trends have on enterprise businesses and their ITC infrastructures, given the many factors that need to be kept into account, their interdependencies and the complexity of their combined effects.

In the security and identity management realms, current solutions that provide "compliance assessment" capabilities (for specific laws and policies, against potential exposures to risks) can only partially address these issues. Some solutions in this space make "Identity Analytics" claims. Despite the value and the key features they provide, they are usually reactive and driven by a "bottom-up" approach. Again, their decision support capabilities are limited to reporting compliance violations and highlighting potential risks, based on existing policies, common security criteria and current IT operational environments.

These solutions do not provide strategic, predictive capabilities based on analysis of trade-offs in investments and they do not take into account the current strategic transition from a pure compliance-based approach to a risk-based approach, driven by the CIOs/CISOs' needs to cope with limited budget/resource issues and prioritize their investments.

In an increasingly challenging business and IT environment, decision makers need to understand how to manage their investment portfolio for IT security, and, in this context, aspects related to identity management. They need to have better insights about strategic matters, understand (in advance) the implications of making IT investments, evaluate the impact on factors of relevance, keep into account the various contextual factors (including business goals, risks, ITC technologies, people's behaviours, legislation, etc.) and explore the potential trade-offs on aspects they can act on.

For example, given particular organisational situations and known risks (e.g. people misusing their own user accounts and credentials), making a technological investment in identity management (e.g. on single-sign-on and federated identity management that could further amplify the involved risks) might not be the most appropriate decision, compared to investing on education and training to change people's behaviours and attitudes to security.

We believe that a "top-down" approach, driven by business needs and decision makers' priorities, would help address these issues. In this context, "Identity Analytics" can help decision makers to explore the implications and consequences (what-if analysis) of their potential decisions in the "identity and access management" space and evaluate multiple scenarios/alternatives. Mechanisms, solutions and techniques need to be developed to support decision makers in formulating new policies and/or justifying current ones. In our view, modelling and simulation are key enabling technologies if based on strong scientific and mathematical foundations.

"Identity Analytics" is a "green field" research area, open to innovation and contributions: it has the potential to shape the next generation of Identity Management solutions. Our work on "Identity Analytics" is part of an ongoing research effort at HP Labs, in the wider context of Security Analytics [2], a project lead by the Systems Secure Laboratory (SSL) [1]. Among other things, work in Security Analytics aims at explaining and characterizing the evolving threat environment in organisations, creating meaningful quantitative metrics that capture the security exposure of an enterprise, helping to understand the attack surface and to manage the gap between security policy and operations. In this context, "Identity Analytics" is a specific "sub-area" of concern and interest, that focuses on aspects and issues derived from identity, identity management and related issues (such as user behaviours, data management, privacy, etc.).

The remaining part of this paper is structured as follows: Section 2 provides additional background on current identity management (IdM) approaches and solutions; Section 3 describes in more details our vision on "Identity Analytics"; Section 4 compares this against related work in this space; Section 5 illustrates the suggested methodology to make progress and move towards Identity Analytics; Section 6 walks through a significant example of Identity Analytics, focusing on the exchange of unstructured data by employees within an organisation by using collaborative "enterprise web 2.0" tools; Sections 7 and 8 respectively describe research opportunities and

issues in this space, along with our next steps. Section 9 draws a few conclusions.

2. Background on Identity Management

This section provides additional background information on identity management (IdM) and an overview of current solutions and products in this space. The main purpose is to illustrate the different areas of technological investments that decision makers (such as CIOs and CISOs) might need to consider, in their daily job. However, technology is only one (important) aspect of identity management. Other important elements, having an impact in this area are users' behaviours, social aspects and new emerging trends (along with enterprise business needs, priorities and available budget). These aspects are usually overlooked whilst they are of key importance and relevance for CIOs/CISOs, in their decision making process, to determine investment decisions.

The area of enterprise identity management is reasonably mature and, from a technological perspective, it is going through a consolidation phase. It includes aspects and functionalities such as authentication, single-sign-on (SSO), authorization, auditing, compliance and assurance management, provisioning, data storage, link to legacy systems and data consolidation. Figure 1 shows the main components and functionalities provided by current enterprise identity management products and solutions.

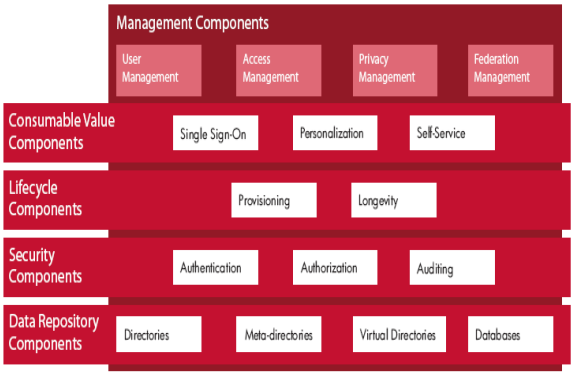

Figure 1. Enterprise Identity Management

Many identity management products and solutions are currently available on the market. They target different types of users and contexts including e-commerce, service providers, enterprises and government institutions, even if most of them primarily focus on the enterprise, corporate, government and healthcare markets. In the last few years new identity management solutions and approaches have emerged in the e-commerce and consumer space, especially in the areas of federated identity management and privacy management.

Most common identity management components and related functionalities include:

- **Directory services, meta-directories, virtual directories and databases** deal with the representation, storage and management of identity and profiling information and provide standard APIs and protocols for their access [3], [4]. In particular, meta-directories address the important problem (especially for large organizations and enterprises) of consolidating, integrating and preserving the consistency of data, disseminated in a variety of heterogeneous systems, geographically spread across organization sites.
- **Authentication, authorization and auditing** are core identity management functionalities. Authentication, in particular, is provided in a variety of ways ranging from local authentication on a system (with login/passwords, certificates, authentication tokens, etc. or combinations of them) to complex distributed authentication [5], [6], including single-sign-on (SSO) within and across organizational boundaries [7], [8]. Recent initiatives, including Liberty Alliance Project [9], [10], aim at the provision of SSO for a federated environment [11], by leveraging identity providers acting as trusted third parties. We are also assisting to the increasing relevance and adoption of other SSO and federation identity management solutions, including Microsoft CardSpace [15] and OpenId [16]. Similarly, authorization functionalities are provided in a variety of forms, usually coupled with auditing capabilities. Authorization can include simple access control management at the Operating System (OS) level, more sophisticated role-based access control - RBAC [12] - up to flexible, distributed, policy-driven authorization, at the application and service levels.
  In the last few years, auditing products and solutions have further evolved towards providing compliance management and assurance capabilities. These solutions, driven by legislation and laws' requirements, such as SOX [17] and GLB [18], have been targeting the increased needs of enterprises to demonstrate good IT governance, by analyzing their processes, comparing them against evidence (e.g. log files, events, etc.) and producing reports, which highlight compliance to

policies and violations, along with required action items.

- **Provisioning and longevity** solutions [13] are used by enterprises, organizations and e-commerce sites to deal with the lifecycle management of identities, including the enrolment, customization, modification and removal of accounts associated with users, employees and customers along with associated identity information (including rights, permissions and access control information). Related functionalities deal with the issuance, certification, management and revocation of digital entitlements and credentials in a secure and trusted way. In particular PKI-based solutions [14] are available for this purpose but their adoption is not so widespread, especially in inter-organisational contexts, because of the intrinsic trust management problems, the complexity of certificate authorities (CA) hierarchies and related costs.
- **Self Service, Personalization and Single-Sign-On** components provide core functionalities to end-users (data subjects) in terms of self-registration and management of their personal information and identities along with mechanisms for single-sign-on across multiple systems and services (within and across organisational boundaries).

These components can be used to provide core identity management functionalities in the following areas:

- **User management:** management of the lifecycle of user accounts associated to data subjects, within organisations;
- **Access control management:** management of access rights and permissions associated to users within organisations;
- **Federated identity management**: management of identity information, access rights and permissions across organisational boundaries;
- **Privacy management:** management of identity information in a way that is compliant to data subjects' requirements, laws and organisational guidelines. This area is increasingly important, because of the growing expectations dictated by laws and legislation and the rise of incidents due to identity thefts and identity phishing activities. Despite various proposals in this space, most of the work is still at a R&D stage, including: (1) "control point" technologies such as privacy-aware access control [19] and privacy-aware information lifecycle management [20]; (2) assurance and compliance-checking technologies [21].

Many papers, whitepapers and solution brochures have been written to describe in great details the various technical capabilities offered by identity management and how they can address security issues, compliance and business needs of organizations.

In this paper we want to focus on the perspective of a decision maker (e.g. CIO/CISO) that must decide on which IT areas to invest and understand the impact of their decision on aspects of interests, such as (security) risk exposure, costs and financial losses, impact on trust and reputation. In this context, it is important to keep into account the current strategic trend in shifting from an approach to security (and identity management) based on "compliance management" to an approach based on "risk management".

Decision makers need to assess and explore different possible types of investments, which might include not only technological options (e.g. identity management solutions) but also other aspects such as process re-engineering, education, etc. A list of possible alternative options might include:

- **Investments in technical solutions**: for example, investments in enterprise directory solutions to consolidate access rights and profile, single-sign-on (SSO) capabilities for employees to reduce duplication of user accounts, automation in provisioning and deprovision solutions to enforce separation of duties, etc;
- **Investments in processes**: for example, investments in processes to handle privacy management and enforcement aspects, auditing and reporting on SOX compliance, etc.;
- **Investment in changing users' behaviours**: for example, investments in user/employee education, detection of misbehaviours and punishment, training, creating awareness about the need to meet legislation, etc.

Many trade-offs of the above types of investments are possible: for example, investments in identity management technologies and solutions need to be justified against the option of investing in other security areas (such as patching, remediation, intrusion detection tools, etc.) or against investments that aim at shaping users' behaviours (such as training and education) or detecting misbehaviours. Trade-offs between security investments and their impact on business functions (i.e. keeping into account business agility and alignment to business needs) are equally important. For example, the introduction of two factor authentication mechanisms on an organisations order processing site may reduce the productivity of the sales

force (because of the involved complexity and impact on usability) and understanding the productivity vs. security trade-offs is critical. Other identity technologies can produce positive benefits for business functions such as deploying automated account provisioning systems, which can help ensure that new high-value staff can be operational as soon as they join a company.

Decisions are influenced by many aspects, including the specific organizational contexts, business needs and priorities, the current IT infrastructure, scarcity of (financial and human) resources and perceived risk exposure. Decision makers are increasingly interested in (and asked to) better understanding the implications and the added value of investing in identity management and justify decisions made in this space against other options.

Being able to explore possible trade-offs, and predict their potential impact and outcomes would be instrumental in making informed decisions. For example, in contexts where users or employees' behaviours are irresponsible, the IT infrastructure and business data and services might be at high risk; the negative impact of frauds or misbehaviours on business assets could be high. Hence, strong investments might be required in basic security and education (coupled with detection and punishment), rather than in enabling a better user experience and simplicity of access to resources, via identity management automation (such as single-sign-on and federation). Being able to predict the impact of such decisions (e.g. in terms of costs or risk mitigation) would be a great bonus.

As anticipated in the introduction of this paper, current identity management solutions can primarily be classified either as "control point" solutions (i.e. enforcing policies, access control management, deployment and enforcement of processes, etc.) or "compliance and assurance" solutions (e.g. auditing, governance solutions, etc.). They do not provide the high-level (top-down) decision support capabilities that can help decision makers to analyse and explore trade-offs and predict the impact on factors of interests (such as costs, trust, reputation, etc.) by means of what-if analysis.

This is the gap that, we believe, can be filled by "Identity Analytics" approaches and solutions, to support, at the right-level of abstraction, the decision-making process. Identity Analytics can help to drive the shift from current "reactive approaches" (provided by current compliance and governance tools) to "proactive approaches", by factoring in the involved complexity and actively supporting the decision making process by means of predictive decision support solutions.

Next section provides more details about our vision of Identity Analytics and the key aspects characterizing approaches and solutions in this space.

3. Identity Analytics: Our Vision

In our view, "Identity Analytics" consists of a set of approaches, techniques and methodologies to explain and predict the impact of identity, identity management and people's behaviours on aspects of relevance to decision makers (e.g. CIOs/CISOs), such as on security exposure/risks, (financial) costs, compliance, trust, reputation, effect on productivity and business (e.g. on business alignment and agility) etc., in well defined context and scenarios - based on initial assumptions and investment decisions.

In this context, "Identity Analytics" aims at providing decision makers with decision support tools and services (based on modelling, simulation and analysis techniques) describing the "levers" (e.g. acting on identity management technologies, automation & centralisation, education, other security investments, policies, etc.) they can act on and the consequences of their decisions (what-if analysis) along with exploring potential trade-offs (e.g. investing on identity automation vs. security patching and intrusion detection).

Figure 2 illustrates the key aspects involved in Identity Analytics.

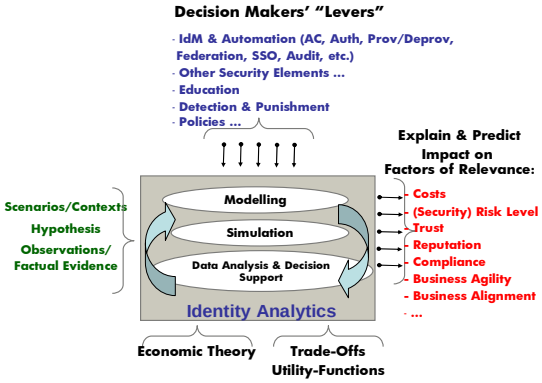

Figure 2. Aspects characterizing Identity Analytics

The focus is at the business level, targeting key decision makers, such as CIOs/CISOs. Identity management is likely to be an area where even the experts have little intuition as to how to invest for the best (security) outcomes. The complexity and tight relationship with business and compliance mean it will remain high priority for CIOs and CISOs. As such it is

likely to be a high profile and rich problem area for Identity Analytics.

Figure 2 shows some of the "levers" that a decision maker can potentially act on in the "identity management" space, to influence "factors of interest". Some of these levers include:

- **Identity Management technologies** and related options, i.e. centralization, automation, etc.;
- **Education and training** of users/employees;
- **Detection of misbehaviours and punishment** (for example via HR);
- **Other security aspects,** such as patching, remediation, vulnerability management;
- **Policies**.

The "factors of interest/relevance" (for decision makers) depends on the context, business needs and priorities. They might include:

- **Operational costs**;
- **Financial losses**;
- **Exposure to security risks**;
- **Trust**;
- **Reputation**;
- **Compliance**;
- **Business agility and alignment**;
- **Robustness and sensitiveness**.

Specifically, business agility and alignment factors are of key relevance. Investments in the IT space are likely to expand in case the business expands (to better support its functions) and the other way around. This applies also to security investments and, more specifically, to investments in the identity management space. The security decision maker may want to use the model to test the decisions under a variety of different business scenarios. This helps them gain an understanding as to how well the policies will react in an agile business situation and ensure that the security policies can easily stay aligned to a changing business environment.

In addition to predicting outcomes, the models should allow decision makers to vary their assumptions about the future business threat environments. This will allow them to assess the robustness and agility of their proposed identity solutions.

Decision makers are usually driven by informal "utility functions", intuitions and experience matured overtime, which keep into account their preferences, priorities and criticalities. In this context, predicted outcomes on "factor of interests" provide additional added-value information to carry out the decision making process. Part of these concepts and aspects can also be factored in the model itself to provide further automation and alignment of the results with decision makers' believes and expectations.

Some of these factors are not independent. For example exposures to security risks have impacts on financial losses and potentially negative implications on the reputation and trust in the organisation.

In the context of well defined scenarios (i.e. with a clear understanding of involved entities, processes, technologies, interactions and human behaviours), it is possible to make hypothesis on the involved entities, interactions and events of relevance; observations can be made on real systems and factual evidence collected to underpin these hypothesis. Experiments can then be carried out to explore and predict the impact of decisions on these factors of relevance, also based on which "levers" (e.g. identity management technologies, education, detection & punishment, etc.) a decision maker has decided to act on. Section 5 provides more details about the methodology we are pursuing in this space, based on the "scientific model".

To achieve this, "Identity Analytics" relies on the following key aspects:

- **Models**: models need to be built to capture aspects of relevance in a scenario, including events, entities, interactions, people behaviours, processes and information flow. Different modelling techniques are potentially available, including deterministic and probabilistic (stochastic) models. Given the nature of the problem we are trying to address, some of the modelled aspects might be subject to probabilistic aspects [51] and randomness (e.g. the likelihood of a person to write down a password, the probability of a system to be hacked, the average time required to provision/deprovision a user account). Models need to be validated and checked for their actual predictive capabilities against real-word expectations or measures. Section 5 describes in more details our approach based on a probabilistic discrete-event modelling approach;
- **Simulation**: simulations, based on the run-time execution of models in a predefined timeframe, help to explore how the "modelled environment" evolves over time and experimentally collect metrics on observed events (e.g. number of compromised user accounts, written down passwords, compromised/stolen identity information, etc.);
- **Data Analysis and Decision Support**: simulation results are analysed to understand how "factors of interest" for decision makers have been impacted (e.g. costs, losses, reputation, etc.). The analysis of

this information, as a first step, can help to explain the current situation, based on known patterns and expected behaviours. In a second stage, they can help to predict outcomes by making initial assumptions and testing them. Different initial hypothesis and alternative settings can be explored, based on the same model, to carry out a what-if analysis.

In this context, decision makers can explore the trade-offs of interests. In our vision the exploration of trade-offs must be driven by economic models, by representing the incentives of the different involved entities (e.g. employees, customers, administrators, etc.) and their impact on aspects of interest (e.g. security, protection of information, disclosure of confidential data ,etc.).

A rigorous theoretical mathematical foundation is required to deal with modelling and simulation aspects. In addition, the representation of trade-offs can be based on mathematical foundation by leveraging analogous work based on the "economic theory", e.g. [22,23,24]. Recent research, e.g. [25,26,27,28,29,30], shows the feasibility of this approach by applying economics and economic theory in the information security realm.

We believe this can help decision makers in understanding the impact of their decisions, in complex, multidimensional/multifactor scenarios and provide added value, at a strategic level. Ultimately this can help decision makers to formulate and/or justify their policies (e.g. IT security policies).

Of course decision support systems based on modelling and simulation techniques are not new and have been successfully used in many disciplines to explain and predict various trends and phenomena.

The novelty consists in coupling these techniques, along with economic and probability theory, in the context of security, identity and identity management and aiming at providing a rigorous mathematical foundation to observations and the decision-making process. As previously discussed, "Identity Analytics" is not just restricted to identity and identity management. A wider context needs to be considered, involving business needs, security requirements, people behaviours and organisational processes. To make progress in this space it is necessary to focus on scenarios of interest (for decision makers), in order to ground concepts and analyse in details various implications. Part of the research challenges in the "Identity Analytics" space consists in identifying common patterns and elements (in terms of identities, identity management implications and human behaviours) that can be reused across multiple contexts and scenarios.

From a research perspective, relevant scenarios that would be worth exploring (because of their complexity and because they take into account new trends), include:

- Collaborative sharing of unstructured data in enterprises;
- Data flow and data lifecycle management within organisations and across boundaries;
- Adoption of new authentication technologies (e.g. authentication tokens), including multi-factor authentication approaches along with different users' reactions and approaches to these technologies, based on their complexity and usability;
- Adoption of social networks and web 2.0 technologies by employees, within and outside organisations;
- Identity thefts and phishing;
- Underground economy underpinned by stolen identities and credentials;
- Privacy management and personal data protection;
- etc.

The goal is to analyse and compare outcomes in these different contexts, try to capture common, reusable patterns and generalize them. Ultimately we want to enable a fast prototyping approach, where we can leverage knowledge from experts in the field and common patterns in building models and running simulations.

Modelling complex scenarios is not trivial: a great deal of effort has to be made to ensure that these models are an appropriate abstraction of the reality, i.e. not too complex or too simplistic. This is one of the important challenges to address. In addition, analysing results and extracting patterns relevant for the decision making progress might be quite time consuming, especially when dealing with complex scenarios, where multiple aspects and interdependent factors need to be considered and factored in. Section 5 provides more details about how we plan to make progress in this space: it illustrates our approach and the methodology we use.

The next section describes related work in this space and how our vision and approach to "Identity Analytics" compare against it.

4. Related Work

Identity Analytics is currently an "overloaded" term, with multiple meanings. It is used to refer to approaches, technologies and solutions that are applied in different fields, noticeably: (1) analysis of personal data and profiles, in order to extract meaningful

"identity patterns", characterizing individuals or classes of individuals; (2) analysis and processing of organizational systems' log files, events and configuration information to assess its compliance to guidelines, policies and legislation, in the space of identity management, privacy and security, and report violations; (3) provide indications to the management team about potential risks and security exposures an organization might incur into, as a further processing step of information gathered by reporting solutions described in the previous point. There is no agreed, common definition of Identity Analytics.

Most current commercial work, solutions and approaches currently making claims in the Identity Analytics area are "bottom-up"-driven solutions, dealing with compliance and governance issues. Their main functionalities are around analysis of log files, events to report on compliance and violations based on current processes, policies and guidelines. Solutions in this space provide indications of risk levels and exposures, based on predefined priorities and processes, e.g. [42,43]. Some other initiatives mentioning Identity Analytics capabilities are pretty much about business intelligence and data mining of identity information, for profiling purposes, e.g. [44,45,46].

This work is complementary to what we are aiming to do. They can provide observational and factual data in specific contexts, by processing and analysing identity and other information collected within the organisation.

Our approach to Identity Analytics is driven by decision makers' needs and aims at exploring and predicting the impact of their decisions along with possible trade-offs in making investment choices. It is a top-down approach, driven by models of scenarios and contexts under examination, based not only on current situations but also hypothetical ones (what-if analysis), along with related simulations and analysis of results.

Existing solutions in the Identity Analytics space focusing on compliance management, decision makers will only be able to assess decisions and policies that have already been made in an organisation. Instead, our work focuses on the current shift from compliance to risk management and can provide upfront support to decision makers, at the decision making time. Decision makers using our approach based on modelling and simulation will be able to understand the implications of their possible decisions (before actually making them), choose the most suitable trade-offs, shape policies and/or justify current ones.

There is related work in Identity Management and Privacy in the space of modelling of simulation, but just in well specific, vertical IdM areas, such as on formulating password policies [47], role of web servers on identity phishing, etc.

This is important work and provides valuable analysis and experimental data that can be leveraged in our work. However, we are not aware of any current research or commercial work that aims at modelling and simulating the overall complexity and different dimensions (various identity management technologies, human behaviour, various interactions between involved entities, enterprise processes, legislation, etc.) that concur in influencing an organization and that need to be taken into account when making strategic investment decisions. We are also not aware of related analysis of trade-offs (by factoring in economics aspects) involving identity management, keeping into account this underlying complexity.

Our work focuses exactly on these aspects and aims at using modelling and simulation techniques to cope with this complexity and provide useful decision support capabilities in this space.

Standards such as ISO 27001 [48], CoBit [49], ITIL [50], etc. describes best practices and methodologies respectively in terms of information security management, IT governance and service management. These standards define valuable common methodologies and guidelines on how to address these management aspects, including aspects of Identity Management. Decision makers still need to understand them, interpret and instantiate them in their specific operational environments. We can use them as drivers and references but our work in the space of Identity Analytics will add the value of grounding the reasoning to specific contexts and related needs and predicting the impacts. Further they represents a one size fits all approach to security and companies wanting to move from a compliance driven to a risk driven mentality need tools to understand the impacts of deviating from best practice.

Our work in Identity Analytics relies on mathematical models and related simulations. The use of mathematical models in engineering has a long and distinguished record of success. From earthworks to suspension bridges, from bicycles to spacecraft, mathematical models are used to predict behaviour and give confidence that necessary properties of the constructions — such as capacity, resilience, and cost — obtain. Such applications of applied mathematics in engineering are useful, and usable, by virtue of the scientifically rigorous modelling methodology, where observations about the external environment and the parameters that the system depends upon are interpreted and a range of properties of the

mathematical model are deduced. In the worlds of traditional engineering, ranging over mechanical, civil, environmental and electrical/electronic engineering, the mathematical methods used are mainly concerned with continuous phenomena and typically use techniques from calculus such as differential equations etc. For modelling security and identity management operations the appropriate mathematical methods are more discrete, being drawn from algebra, logic, theoretical computer science, and probability theory. In order to apply these methods, we require a conceptual analysis of the relevant aspects of the systems of interest.

In our work, we leverage the seminal work done by HP Labs in the Open Analytics project [33,34,35], that we will consider as a reference. We are using Demos2k [36,37] as the reference tool for our modelling and simulation activities. More details about Open Analytics and Demos2K are provides in Section 5.

Finally, an important aspect of Identity Analytics is the studies in the space of Social Science, in terms of understanding, modelling and simulating human behaviours, drivers and motivations and the impact of actions on the surrounding environment. We aim to leverage work done in this space, such as [32], in order to build mathematical models that realistically reflect users' behaviours and the associated impact.

5. Moving Towards Identity Analytics: Methodology

The methodology we are planning to use to make progress in the Identity Analytics area is based on the "Scientific Method" [31], tailored to this specific domain. Given a specific scenario/context, this requires building a "theory" of a specific "phenomenon" we are interested in, making related hypothesis, design "experiments" to prove/disprove these hypothesis (and hence the theory) and, based on the outcomes, potentially refine the initial theory. This involves gathering observational facts, using them to create models via an inductive process, using simulation techniques on top of these models to draw conclusions and validating these conclusions against the real world.

For example, theories might be built, in an enterprise scenario, about the impact and effects that some policies - e.g. password length policies - and some related identity management solutions - e.g. automatic password expiration and renewal solutions - might have on business aspects (e.g. operational costs) and users' behavioural aspects. Hypothesis could be formulated about how users' are likely to react to policies imposing specific lengths and patterns in defining passwords, including the likelihood of forgetting these passwords or actually writing them down. Observational information could be based on metrics gathered from the field, in terms of actual users' requests for password renewals, complaints, operational costs to handle passwords etc. Models and simulation can be built to explain and predict the outcomes (e.g. in terms of actual number of passwords that need to be re-issued or passwords that have been written down) by changing some of the initial assumptions and policies. These results are validated against known observational data, to check the predictive accuracy of the model. This might require refining and/or changing the model, till the point the model is confidently matching observations in the real world and/or intuitive aspects. At this point the model can be used to make predictions about other real world consequences, including non-intuitive aspects.

All these steps might need to be repeated to refine the models, based on reality-check in the field and validation of their actual predictive capabilities.

Figure 3 illustrates the key involved aspects and steps:

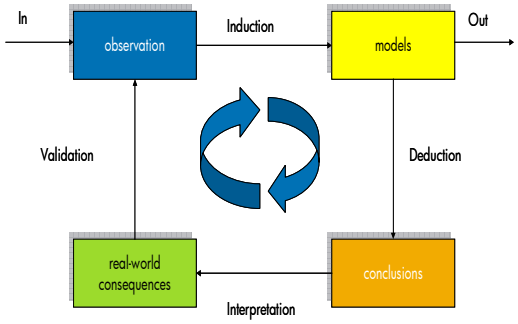

Figure 3. Modelling Methodology

Given a specific scenario and context, empirical studies can be conducted to gather observational facts and evidence. For example, we could consider a scenario focusing on the sharing of unstructured data by employees, within and across organisations, by using collaborative tools; it involves people, their behaviours and interactions, enterprise services, application and data repositories, as well as identity management solutions (e.g. for provisioning/ deprovisioning of user accounts, access control and authorization). Details of this scenario are provided in Section 6.

Observational information and relevant events can be collected from the field (by means of surveys, analysis, etc.) to describe, in probabilistic terms [51], some of the involved interactions (such as the likelihood of users performing specific actions on enterprise resources) and processes where the above entities are involved. For example, in the data sharing scenario described above, this might include extracting by means of surveys (interviews with people) and/or analysis of systems data logs, the probability distributions about how people act in terms of sharing data, handle user accounts and passwords. This might help identifying "users' behavioural profiles" describing common patterns in terms of events generated by these people (e.g. writing down passwords, sharing user accounts, sharing confidential data, etc.) and the likelihood they are going to happen, etc.

Two kinds of events (and related observations) can be analysed:

- **External events**: these are events that just happen, where there is little degree of control, at least in the short term. In the data sharing scenario example, this could include aspects related to users' behaviours, that in the short term are hard to control (but might be influenced in the long term);
- **Internal events**: these are events that can be influenced and/or for which there is a degree of control. In the above example, this might include events related to identity management solutions (e.g. automatically expiring passwords, provisioning/deprovisioning user accounts, etc.) aiming at protecting the access to data and information within an organisation.

This information is then going to be used in an inductive step, to produce (semi-) formal modelling components (such as probability distributions and statistics) and iteratively build one or more mathematical models of the involved entities, interactions, processes and systems.

For example, in the data sharing scenario, this would mean modelling users, their behaviours, the relevant enterprise data sharing systems (e.g. Twiki, MS Sharepoint, shared file systems, etc.) and applications, involved interactions and processes (e.g. the process of accessing confidential data or creating and granting access rights to a user), along with the core capabilities and functionalities provided by the involved identity management solutions.

Simulations, based on these models, are used to generate experimental results. A subsequent analysis of these results drives the deductive process towards formulating conclusions on factors of interests for decision makers (e.g. involved costs, impact on reputation, impact on trust and compliance, etc.) and presenting the impact of different trade-offs (e.g. investing in a technological approach to identity management vs. investing in an educational approach).

For example, in the data sharing scenario, experimental results generated via simulation could be in terms of number of confidential documents that are likely to be exposed or compromised, by keeping into account the probabilities that people write down passwords or share accounts with others. This information can provide indications of potential financial losses and involved costs.

These conclusions are interpreted and checked against the real world (and their consequences), for validation purposes. Experts in the fields, including decision makers, are likely to be actively involved in this process. Multiple iterations of the entire process might be required, before a suitable model is built, that matches expectations and can provide meaningful predictions to decision makers about non-intuitive situations and aspects.

To summarise, the aim is to build models that covers the current and potential alternatives and/or future situations. The model is verified against the 'known situations' with the hope that it generalises appropriately to the unknown ones. In this sense the goal is to both look at the outcome of the model results on the known situations as well as having a structural review to ensure that the assumptions and abstractions underlying the model are 'reasonable' to the expert. The simulation then plays out the interactions between the variety of assumptions that have been made.

At HP Labs, this methodology has already been successfully explored and applied both internally and in the service consulting context by the Open Analytics project [33,34,35].

It is also consistent with the methodology that is used in the context of the UK TSB Trusted Economics Project [32], driven by security and economics aspects and aiming at involving empirical studies, gathering and analyzing observational facts, deducing semi-formal models, modelling and simulation, trade-offs analysis, reiterations and refinements with customers, and stakeholders. HP Labs are actively involved in this project. Specifically, this project will influence the direction of our Identity Analytics project, by helping us to factor in the "economic theory" aspect and using analogies to drive the formulation of trade-offs in the identity management space. This is a green field, open to innovation and contributions.

In the specific context of Identity Analytics, we are well aware that the complexity of the explored scenarios (including users' behaviours, external

influential factors, emerging technologies, etc.) might pose key challenges, in particular in terms of effectively gathering valuable observational data and factual information from the field. This data might not be available; it might be too expensive to retrieve or might be business confidential.

In this case, we envisage the possibility to tackle this kind of problems by means of "thought experiments", still based on the methodological principles illustrated in this section, but where the absence of specific, observational data is replaced by "common sense" (qualitative) observations and the outcomes are "qualitatively" validated against the real-world and "expectations" of experts in the field. We believe that this approach is still valuable to conduct "what-if" experiments, to explore the implications of choices and their "qualitative" impact on aspect of relevance to decision makers. As an example, Section 6 is going to walk through a "thought experiment", focusing on an enterprise data sharing scenario, involving heterogeneous user behaviours, different data sharing solutions and the impact provided by identity management technologies.

One of the objectives of this kind of experiments is to explore the shape of the involved outcome space. An exploration of the space defined by utility function, as specified for a given model (i.e. capturing the decision makers' priorities and expectations), can provide useful insights into where the decisions will be effective - as the input parameters to the model change. Looking for the "cliffs, valleys, plateaus" and any regions of instability (chaos) can help gaining an understanding of when and how security decisions take effect. Without observational data we may not be able to give exact quantitative information but suggest their presence and still provide useful insights into the decision making process.

We also aims at following a rigorous, scientific approach in analysing complex contexts where the outcomes of the modelling and simulations steps might not be trivial and intuitive, given the complexity and non-deterministic aspects of the involved interactions. This is where we see the greatest value of Identity Analytics, i.e. in providing insights and analysis in complex contexts where intuitions and expertise can only help till at one point.

This is particularly true when trying to provide indications to decision makers about the outcomes of trade-offs. As discussed a few times in this document, in the specific field of Identity Analytics, trade-off analysis might involve comparing the impact of different, heterogeneous aspects, such as identity management technologies vs. behavioural and educational aspects vs. legislative aspects vs. punishment and HR-driven approaches to ensure compliance to enterprise policies. The combined effect of these aspects might indeed be not intuitive, hence the help provided by Identity Analytics.

The remaining part of this section describes some possible modelling techniques that can be used in the space of Identity Analytics.

5.1 Modelling Techniques and Tools

Analytical and predictive mathematical modelling approaches are potentially suitable to carry out modelling and simulation activities in the Identity Analytics area. Our current preliminary work and exploration of this space has been based on a "simulation-based predictive modelling" approach. Based on our initial investigation, the predictive modelling approach provides advantages over the analytical approach as it allows to explore (in a more natural way), via experimental results, the dependencies among different involved entities, processes and decisions. This is particularly of relevance for those scenarios involving the modelling of business process aspects, interdependencies with identity management solutions and probabilistic users' actions and behaviours.

Specifically, we have used a specialised simulation oriented language Demos2k [36,37], which implements a modelling framework based on the mathematical foundations of a synchronous calculus of resources and processes, together with an associated modal logic [38]. Demos2k supports the development of discrete-event stochastic (predictive) models. Because of its strong mathematical foundations and sound semantics, we have assurance that simulations based on the models developed in Demos2k language are robust and reliable – thus, meaningful observations can be taken. The code is executed via repeated experimental simulations in the specially developed experimental environment [39], where statistically significant information is gathered.

The mathematical framework behind the Demos2K programming language revolves around four key concepts:

- **resources,** capturing the essentially static components of the system;
- **processes,** capturing the dynamic components of the system;
- **location,** capturing the spatial distribution and connectivity of the system;
- **environment** within which a system functions.

A full description of this mathematical framework can be found in [38].

In the domain of Identity Analytics, "resources" could be any valuable asset or element we might want to model. For example, this could include confidential and personal information, user accounts and related passwords/credentials, identity and authentication tokens, etc.

Modelled "processes" could include, among other things, identity management processes and systems, data lifecycle processes and data flows, enterprise business processes and human activities and behaviours. Of particular importance are those processes that have interdependencies: in this context, the Demos2k language provides great support in explicitly representing dependencies and synchronisation aspects.

"Location" modelling aspects (supported by Demons2k) are also of particular importance in Identity Analytics: they are required to represent spatial distributions aspects of identity management systems, data repositories (depending where they are they could be exposed to security risks and compliance violations) and people's locations and interdependencies.

Finally, the "environment" aspect is used to model additional characteristics of the scenario under observation that are of relevance for the simulation steps.

Section 6 illustrates some of these modelling aspects, in an example, related to a scenario involving unstructured data sharing within organisations.

In general, the model of a specific system in a scenario usually consists of multiple processes which either consume resources, or take certain time to finish. As a result of one process finishing, another might be triggered. Alternatively, they might start concurrently depending on the structure and complexity of the system to be modelled. Some of the processes could be triggered by events from the environment.

Because of the strong semantic properties of these concepts in the Demos2k programming language, we have strong guarantees that during the execution of the model, the processes are executed as it was intended by the modeller. Demos2k efficiently handles concurrency, queuing, and prioritization among processes.

Based on our current investigation, Demos2k seems to be suitable to address most of the modelling and simulation needs for Identity Analytics. A critical aspect for Identity Analytics is the capability of modelling and simulation the behaviour of large populations. This is where we are still testing how to better use Demos2k: Section 6 illustrates a possible way to achieve this.

We are also interested in exploring the suitability of other modelling and simulation approaches to Identity Analytics: this requires further analysis and investigation. This is an area open to contributions and innovation.

Hence, we are planning to investigate other potential approaches to (mathematical) modelling and simulation, compare them and understand their pros and cons, driven by our needs and objectives in this space. Alternative/complementary approaches we are planning to investigate include:

- **Probabilistic rule-based modelling:** this includes the usage of rule-based models coupled with probabilistic mechanisms, Bayesian Networks [ref Pearl]and Probabilistic Boolean Networks;
- **Modelling based on Swarm Theory and Swarm Intelligence [40]**: this area looks promising in terms of exploring complex human behaviours in large population, an aspect that is of relevance in Identity Analytics, given the key impact of human behaviours;
- **Modelling based on Chaos Theory [41]**, in particular exploring this theory from an organisational perspective;
- **Probabilistic Agent-based modelling and simulations:** this approach might be an alternative way to explore complex interactions of users, and the outcomes of their behaviours**;**
- **Game theory, game modelling and simulations.**

At the current stage, we do not exclude the fact that hybrid approaches, using two or more different modelling and simulation approaches might be required in Identity Analytics.

6. Identity Analytics: Walking Through an Example

This section aims at grounding some of the concepts described in this paper in terms of Identity Analytics by walking through a scenario and providing additional information on how we deal with modelling, simulation aspects and data analysis.

It is beyond the scope of this paper to enter in the details (papers will follow with this kind of information): the main goal is to illustrate the feasibility of our approach, along with showing the complexity of the involved factors and the value of providing predictive capabilities to decision makers in the identity management space and supporting related what-if analysis. We will follow the methodology illustrated in Section 5: at the current stage, the results

illustrated in this section are the outcomes of a "thought experiment" that is evolving over time.

Within the overall methodology this sits in the early experimentation stage of trying to model based on experts' intuitions and descriptions of a scenario. This then forms the basis for a more rigorous experimental design including user studies to better understand the details of how people act in a given scenario.

The scenario we considered in this example is about "sharing of unstructured data" by people within an organisation. This scenario focuses on an emerging enterprise trend, consisting in the adoption and usage by employees of collaborative, customisable data sharing tools, such as Wiki, Twiki and Microsoft Sharepoint tools. These collaborative tools provide an unprecedented level of flexibility and simplicity of usage, in terms of creation of related data sharing sites, creation, posting and retrieval of unstructured data and information (compared to more traditional structured data stored in databases and LDAP directories), collaborative generation of content and wide options for sharing unstructured material and information.

On one hand, people within organisations might be encouraged to share data and information, to improve communication among parties involved in projects and being more effective. On the other hand this data sharing presents the security risk that data may be shared inappropriately. Data sharing sites could be created and installed within an organisation IT environment, without fulfilling the basic security and compliance constraints dictated by enterprise guidelines and policies. A potential relevant risk is that confidential information might be stored and shared on insecure and unapproved sharing sites, hence creating harm to an organisation, in terms of: financial losses due to data leakages; exploitation of information by malicious internal personnel; negative impact on organisational trust and reputation.

The dilemma that decision makers (such as CIOs/CISOs) have to face is about how to act in these situations. For example, they might need to understand what the implications are in terms of allowing/disallowing the usage of these tools, in a complex environment. By completely forbidding the usage of data sharing tools, they could undermine collaboration, creativity and innovation (or staff may use other mechanisms, such as Facebook sites). Alternatively they could decide to directly supply these data sharing services by means of centralised enterprise IT services, hence meeting the basic security requirements and being compliant to security policy but undermining the level of flexibility and customisation of these tools.

In the context of this scenario there is a wide range of assumptions we could make. We considered two categories of data sharing tools:

- **Central IT (CIT) Data Sharing Sites**: these sharing sites are hosted and run on enterprise approved IT infrastructures;
- **Shadow/Self-IT (SIT) Data Sharing Sites**: these sharing sites are run on "shadow/self IT", i.e. not officially approved IT infrastructures, such as using personal servers or individual's PCs to run them.

Identity management (along with traditional security) plays a key role in this scenario, in terms of providing basic mechanisms for authentication, authorization and protection of data and information along with supporting audit capabilities. User provisioning and deprovisioning solutions can be used to automate the management of the lifecycle of user accounts and their credentials. Federated identity management and Single-Sign-On (SSO) solutions can simplify the access to various systems and sites by reducing the number of required user accounts and credentials.

In our scenario, we considered different situations and options for providing these identity management capabilities:

- **Central IT Identity Management Solutions**: these solutions are provided by the organisation's central IT services, hence they are compliant with the required security requirements and meet organisational policy. Specifically, in this scenario we considered central IT identity management solutions supporting: automatic provisioning/ deprovisioning of user accounts; automatic expiration of user passwords and requests for renewal; single-sign-on (SSO) functionality (i.e. a unique user credential that allows access to multiple sites) for centrally managed IT systems, including central IT data sharing sites;
- **Ad-hoc Identity Management solutions**: these identity management solutions (or degrees of them) are provided on ad-hoc basis, by Shadow/Self-IT sites, for example in terms of ad-hoc management and setting of user account and passwords. These identity management solutions might or might not be compliant with requirements imposed by the organisation, in terms of security.

These are a few assumptions (and simplifications) that we made based on our experience in enterprise organisations. We could change and/or extend them to include other aspects and phenomena, such as the increasing reliance of social networks (such as

LinkedIn, Facebook, etc.), their identity management solutions and their approaches to protect data.

Employees (users) and their behaviours play a key role in this scenario. Ultimately, they are the entities that influence and drive the overall data sharing process. Based on their behaviours there are different levels of (security) risks. We considered the case where a few categories of users could be identified, based on "common behavioural" profiles:

- **Compliant Users:** these are users that are policy aware and act at the best of their knowledge in terms of complying with prescribed policies and guidelines. In this scenario, these users are likely to use data sharing site approved by the central IT organisation and obey to organisational policies;
- **Loose Users:** these are users that do not quite know how to act/react in specific situations, as they are not fully aware of policies or guidelines. Some of them might act driven by common sense; others might just take actions without fully understanding the implications. In the scenario under examination, these users might or might not use central IT data sharing sites, depending on circumstances, other people they interact with/influence them, etc.;
- **Non-Compliant Users:** these are users that might be well aware of policies and guidelines but deliberately act against them, not necessarily because of bad intentions but because they perceive that some of the existing policies can undermine their work and business objectives, hence they adopt their own approaches and ad-hoc solutions. In this scenario, these users might deliberately decide to adopt and use new, unapproved technologies, if this makes sense to achieve their goals. This category of people could be very relevant to organisations, as they bring innovations and diversity. On the other hand, they could expose organisations to unnecessary risks;
- **Traitors:** these are users that deliberately act against the interests of an organisation and create potential losses, for whatever reason that motivates them. In the context of this scenario, traitors might try to get hold with sensitive information by engaging in teams that share sensitive and important data and leveraging security weaknesses (such as getting hold of shared credentials to access sites, written down passwords, manipulating other people via social engineering, etc.).

In this scenario we focused on specific users' actions and events (as an example) that could create risks for an organisation, based on their behaviours:

- Writing down passwords (associated to data sharing accounts or single sign-on accounts);
- Sharing accounts with other people;
- Leaving data sharing sites without unsubscribing, hence leaving hanging accounts that could be misused;
- Accidentally leaking data by means of other communication tools, such as emails.

People might bypass "safe" and "policy compliant" data sharing sites mandated by organisations and adopt alternative solutions (e.g. creating their own Shadow-IT data sharing services) that can expose the organisation to unnecessary security and financial risks. Organisations and decision makers have complex dilemmas to deal with:

- Collaboration and exchange of data, under certain circumstances, should be encouraged as it empowers people and could boost opportunities for better business outcomes;
- The adoption of "Shadow IT & Self-IT" could be a "plague": it might be discouraged or people might be encouraged to be "more compliant" when creating or suing them;
- Automation (in identity management, security etc.) can partially address some of the security issues but it can still be circumvented by people and their behaviours (such as by creating Shadow-IT & Self-IT collaborative sites)
- Education, Detection & Punishments could be additional ways to address the problem, beyond a pure technological approach.

From a decision maker's perspective (CIOs, CISOs) the dilemma is concerned with making the "right" decisions, given the complexity of the situation, existing constraints and the many different, interdependent involved aspects (including business priorities, available budget, etc.). Some legitimate questions might arise:

- What is the best way to deal with this situation?
- What are the potential trade-offs?
- What could be the consequences of making certain decisions?

Figure 4 summarises the key aspects that are involved in this scenario along with the involved aspects that need to be modelled, i.e.:

- Different types of existing sharing sites (CIT and SIT sharing sites) that have been deployed within an organisations;
- Identity Management solutions provided by the central organisation's IT services;
- Different categories of users with their behaviours;

- Actions that these categories of people can perform, in terms of: joining, leaving, creating, closing a sharing site; adding, retrieving and sharing data; asking for new access credentials, renewing credentials;
- Additional actions that can expose the organisation to risks: writing down passwords, leaving sharing site by leaving hanging accounts; sharing their user accounts or credentials with other people; leaking data; misusing or attacking sharing sites.

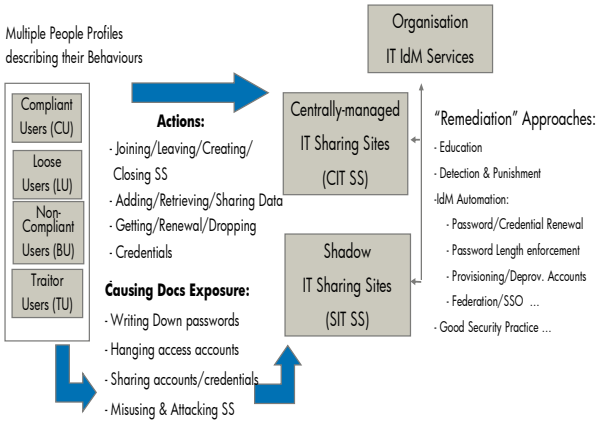

Figure 4. Data Sharing Scenario and Involved Entities and Processes

Figure 5 illustrates, with additional details, some of the "complex interactions" happening between the modelled entities, the effects of users' actions and some of the "levers" that could be adopted by decision makers to deal with this situation.

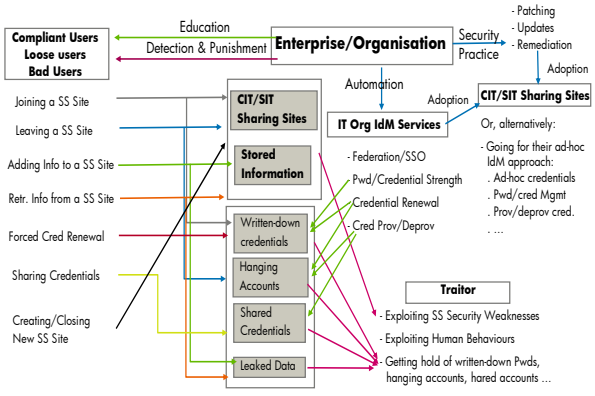

Figure 5. Data Sharing Scenario – Modelled Aspects

As already stressed a few times, modelling users' behaviours and users' actions is as important as modelling the involved systems and processes. This helps to better understand the dynamics in specific contexts, identify potential critical aspects and make predictions on the implications. For example, in this model we represent the consequences (in terms of exposure and leakage of information/data) of users' behaviours such as installing and using non-compliant sharing sites, writing down passwords, sharing credentials, and deliberately taking advantage of this for personal gains.

By representing, in our model, a variety of aspects, ranging from human behaviours, security aspects and technological components, we can start tackling the analysis of their aggregated effects and consequences and explore trade-offs. For sake of simplicity, in this experiment we considered three main levers that decision makers can act on to explore the implications of their investment decisions:

- Investments in automation and security, including further centralization of identity management solutions and further security of various sharing sites;
- Investments in user education, via training courses, awareness campaigns, etc.;
- Investments in detecting misbehaviours and punishment via HR.

We have built a (mathematical) model representing the involved systems (i.e. different types sharing sites, identity management solutions, etc.) the involved categories/classes of users, their interactions and behaviours when dealing with sites and stored information.

In this model the *"external events"*, with some degree of approximation, include the actions and behaviours of users. Of course these behaviours can be shaped on the long term, for example with education campaigns or simply detection and punishment, but in the short term are externalities. Modelled *"internal events"* include the way various systems and processes works, such as events related to identity management, including forcing the expiration of passwords/accounts, dealing with their provisioning/deprovisioning, etc.

Related observational facts could have been collected on the "field", for example with empirical studies and surveys, to describe the behaviour of people, to identify and characterise different kinds of user profiles (e.g. good, loose, bad, traitor). This could have given us indications about probability distributions [51] describing (for example):

- Probability distribution/likelihood to perform some of the described actions on data sharing sites, by differentiating between Central IT and Shadow IT sites;

- Probability distribution/likelihood to write down password information, based on imposed constraints (length, patterns, etc.);
- Probability distribution/likelihood to share credentials with other people;
- Probability distribution/likelihood to leak information, via other tools, such as social networking tools, emails, etc.

As this model is in the initial exploratory stage within the methodology, we make some educated guesses for the probability distributions based on our and other experts' intuitions about the situations. This allows the area to be explored, with the next refinement being to carry out experiments that better ground these distributions.

Similarly, Central IT sharing sites and Shadow IT Shadow IT sharing sites have been characterised by means of:

- Probability distributions/likelihood to support and adopt central IT sharing sites, such as single sign on and their user account management solutions;
- Probability distributions/likelihood, in case ad-hoc Identity Management solutions are adopted, to enforce password renewal, automate the provisioning/deprovisioning of user accounts, etc;
- Probability distribution/likelihood to be compliant to security policies.

In terms of simulation, some additional assumptions have been made on the initial number of involved sites, amount of information stored in these sites and amount of information that can be generated and accessed by various users, etc.

In addition, a few "factors of interest" to decision makers have been identified as potential outcomes that we want to measure and analyse. In this example we focused on the following "indicators" both for Central-IT (CIT) and Shadow/Self-IT (SIT) data sharing sites:

- **Number of "Exposable Documents"**: this measure indicated the number of shared documents that can potentially be at risks, e.g. due to written down password, shared accounts, security weaknesses on sharing sites, etc. In other words, this give a very raw indication of the "exposure surface";
- **Number of "Exposed Documents"**: this is the actual number of documents that have been compromised, due to the activity of traitors or other misbehaviours;
- **Number of Written Down credentials/ passwords**: this is the number of passwords that users might be writing down, as they cannot remember them or are too lazy;
- **Number of Hanging Accounts**: these are accounts that are not used anymore by people but that, nevertheless, are still active and can be misused;
- **Number of Shared Accounts**: these are accounts and credentials that are shared by a set of people.

All these "coarse-grain" indicators are calculated during the simulation process by keeping into account the combined effects of all the involved entities and processes. Nothing prevents, in case of needs, to have fine-grained indicators, showing the "contributions" of each category of users.

In our initial experiments we considered (as an example) a population of 1000 users, distributed as follows: 20% compliant users, 60% loose users, 19.5% non-compliant users and 0.5% traitors. The model has been designed to allow us to run a simulation of the (probabilistic) actions that can be carried out by each individual of this population on a daily basis. In Snapshots of all the involved indicators have been captured on a monthly basis with an overall simulation period of a year. These settings can be changed as we refine the experiment and understand the details of a specific scenario.

We have built a full working prototype of our model by using Demos2k [36], i.e. a stochastic, discrete-event modelling tool and framework.

Based on our initial settings of various probability distributions, etc. simulations have been used to generate experimental values for the "indicators/factors of interests" described above. We have analysed and processed them to provide information for the CIO/CISO's decision making process.

For example, Figure 6 shows the amount of "exposed documents/information" over the simulated period of time (12 months, with monthly observations).

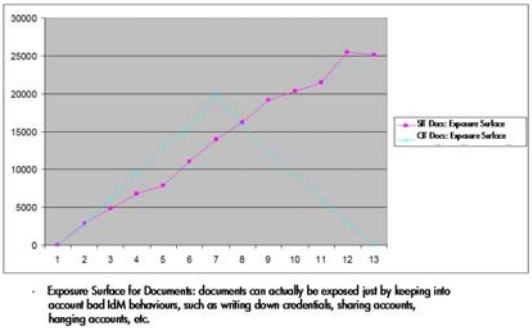

Figure 6. Experimental Result – Exposure Surface in terms of Shared Documents

This figure illustrates that the number of "Exposable Documents" in Shadow/Self-IT (SIT) data sharing sites sensibly grows over time. This is due to (1) non-compliant users' behaviours (e.g. writing down passwords and sharing accounts) and (2) bad security and identity management practices adopted by the administrators of these data sharing sites.

On the contrary, this figure shows (given the initial assumptions we made in this example) that the situation for Central IT (CIT) Sharing Sites is definitely better. The impact of users' misbehaviours is still initially high but then identity management solutions (that we assumed are properly deployed and fully working in CIT sharing sites) catch up, by automatically expiring passwords/credentials after a predefined period of time (in this example we assumed that the renewal period of passwords is normally distributed [51] with a mean of 6 months and a variance of 1 month). The effect of this is clearly shown in Figure 6. The usage of these identity management solutions automatically reduces the exposure surface, as expired credentials are of no value and cannot be misused.

Similarly, Figure 7, shows the impact that automation in identity management (specifically password expiration) has on the number of single-sign-on (SSO) credentials that could be compromised (because, for example, they have been written down), that have been actually compromised or that have been automatically expired. In this scenario we assumed that each user has a SSO credential provided by the organisation's central IT identity management services, to access common enterprise services and systems.

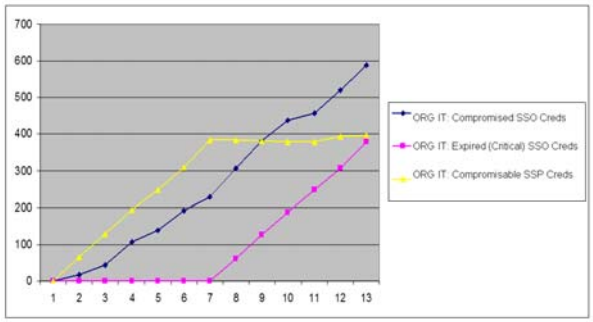

Figure 7. Impact of Identity Management automation on centralised Single-Sign-On credential management

Once again, the impact of users' misbehaviour is determinant in exposing the organisation to risks; under the assumptions made in our model, Figure 7 shows that identity management automation can help to reduce the impact of these misbehaviours, by stabilising the number of "compromisable" SSO credentials (i.e. credentials that are, for example, written down by users and can be misused by "traitors").

These kinds of graphs (e.g. Figures 6 and 7) can confirm some of the intuitions that decision makers might have, based on their expertise and understanding of the involved technologies, processes and users' behaviours. These are examples of "simple" experimental results and outcomes that can be provided by our modelling and simulation approach. Of course, in a real modelling and simulation activity, these conclusions must be checked and validated against the real world. In a real modelling exercise, many iterations will be required to refine the model and initial assumptions, till the outcomes are match the observable facts in the real world.

The power and strength of this approach, driven by modelling and simulation, is that it can then help decision makers to explore trade-offs and carry out what-if analysis on aspects that might not necessary be so intuitive. For example, given a model that reflects the current situation, a decision maker could play out the potential effects of a decision to explore the value and impact of the decision. Variation in the outcomes of the model, given different possible effects of the decision, may help design metrics that allow the decision maker to ensure that the new policies are working as necessary.

In our experiment we assumed that Central IT (CIT) data sharing sites and Shadow/Self (SIT) IT data sharing sites have different degrees of compliance to basic security practices, such as patching vulnerability, running antivirus scanners and other remediation solutions (e.g. firewalls, etc.). Specifically we assumed that 99% of CIT data sharing sites is compliant whilst only 50% of SIT data sharing sites are compliant.

Lack of security compliance increases the exposure to risks, as "traitors" (and/or external agents) can take advantage of these vulnerabilities, compromise data sharing sites and get hold of confidential data and information.

A decision maker could be interested in exploring the sequent questions: "What if I invest resources in ensuring that also Shadow/Self-IT (SIT) data sharing sites are compliant to security practices?". In this context the decision maker might recognise the value of these sites but be worried about the risks they introduce.

What would be the impact on "exposable" and "exposed" documents if, let's say, actions are taken to ensure that 70% or 90% or 100% of the SIT data sharing sites being compliant?

This is an example of "what-if" analysis that can be easily supported by our modelling and simulation approach. We simulated different cases where we supposed that different percentages (70%. 90%, 100%) of SIT sharing sites are security compliant, by maintaining all the other assumptions constant. In this context, based on our assumptions, CIT sites are already compliant, so no point in taking further actions with them.

Figure 8a shows that by acting on this "lever" it is possible to sensibly reduce the number of exposed documents, but till to the point where the number of exposed document is still high and relevant and no further improvements can be achieved (i.e. case of 100% of SIT data sharing sites being fully compliant).

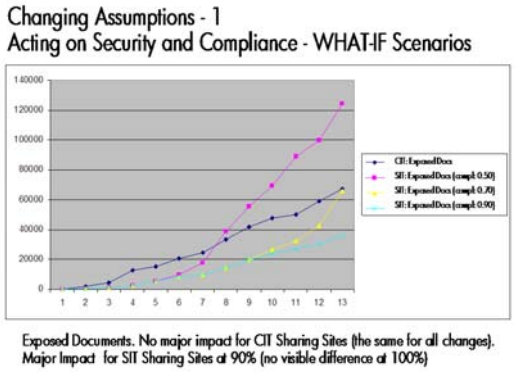

Figure 8a. What-IF analysis – Acting on Security and Compliance – Impact on Exposed Documents

On the other hand, Figure 8b, shows that by acting on this "lever" it is possible to dramatically reduces the number of compromised SIT sites (over the observed period of time), as expected.

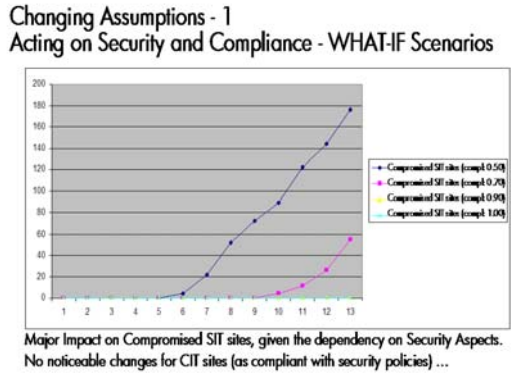

Figure 8b. What-IF analysis – Acting on Security and Compliance – Impact on Compromised Sharing Sites

By comparing the outcomes shown in Figure 8a and Figure 8b, it is possible to infer that users' behaviours still have a big impact on the final number of "exposable" and "exposed" documents.

This can be explained by the fact that, no matter how much the CIT/SIT data sharing site are secure, if a user writes down passwords or shares account information with other people, the risks of data leakages and exposures of documents are still there.

In a more complete and realistic approach, our modelling and simulation activities should have included (for example) the analysis of the involved costs and financial losses and compared them against the predicted outcomes in terms of "exposable" and "exposed" documents. This would have given decision makers more compelling information about the implications and impact of their decisions also from a financial perspective.

Of course this is just one of the possible "what-if" analysis and predictions that a decision maker might be interested in exploring. The decision maker might have been interested in exploring the impact of investing in education and/or detection and punishment instead of investing in technology or security. What are the implications of acting on this "lever", requiring changing users' behaviours? This can be potentially predicted by using modelling and simulation-based approaches.

As part of our experiment, we considered the case where actions taken by decision makers force the distribution of the user population to be shifted towards more and more compliant users (starting from the initial situation described at the beginning of this sections, that is based on a population of 1000 users, where 20% were compliant users, 60% were loose users, 19.5% were "bad" users and 0.5% were traitors).

Figure 9 illustrates a few possible alternative distributions of this population, where the number of compliant users increases whilst the number of loose and non-compliant users decreases.

Changing Assumptions - 2
Acting on Users Behaviours - WHAT-IF Scenarios

- What if the Organisations Manages to change Users's Behaviours (e.g. via education, detection and punishments, etc.) by shifting the number of people that belong to each identified profile (apart from Traitors). NOTE: the profiles are the same, people move around ...

	Compliant	Loose	non-compliant	Traitors
Ref. Model	0.20	0.60	0.195	0.005
Alternative #1	0.40	0.45	0.145	0.005
Alternative #2	0.60	0.30	0.095	0.005
Alternative #3	0.80	0.15	0.045	0.005

Note: we still consider the case of 1000 users. Above are related percentage

Figure 9. What-IF analysis – Changing User Behaviours by Shifting Population's Distributions

Of course the aim is not to suggest which specific action and process should be carried out by the decision maker to achieve this, but provide a way to predict what the implications would be, should these changes be made (all other aspects being the same).

Figure 10a and 10b shows the potential impact that these changes could have on "Exposed Documents" and "Document Exposure" respectively.

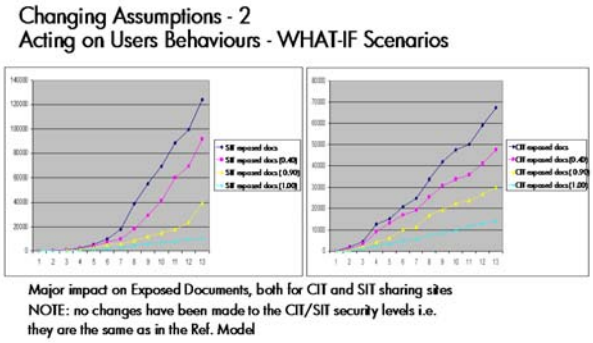

Figure 10a. What-IF analysis – Acting on User Behaviours – Impact on Exposed Documents

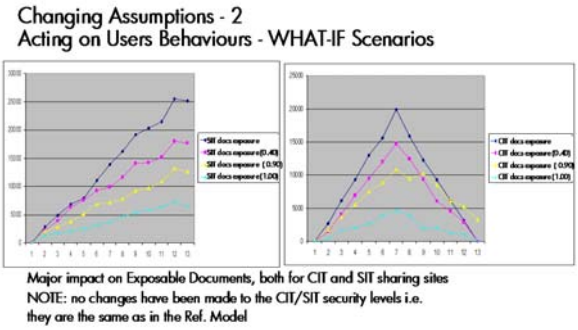

Figure 10b. What-IF analysis – Acting on User Behaviours – Impact on Exposable Documents

These figures show that the amount of exposed and exposable documents can substantially be reduced, by having more and more compliant users i.e. by ensuring that their behaviours can be "improved". This trend is confirmed both for CIT and SIT data sharing sites, both in terms of "exposable" and "exposed" documents.

Of course other what-if and trade-off situations could be analysed, for example considering hybrid cases where a decision makers acts both on the security and educational "levers".

Once again, all the results shown in this section are based on an initial model where assumptions are based on our intuitions of the situation. The next stage in the modelling process is to validate and refine these assumptions with experts and carry out experiments to validate or refine the basic probability distributions used within the simulation. Even this initial model can nevertheless demonstrate the power of this approach, in providing "qualitative" predictions in complex situations, once the "qualitative" predictive capabilities of the models have been confirmed by experts in simple an basic aspects.

To conclude, this section aimed at describing an example in the Identity Analytic space, by means of an experiment, to illustrate some of the kind of predictions and analysis that can be provided to decision makers, by mean of an approach driven by predictive, discrete-event (mathematical) modelling and related simulations.

7. Research Opportunities and Challenges

Identity Analytics is a green area, plenty of opportunities open to innovation and contributions. Specifically, there is the unique opportunity to explore the space of identity and identity management from different perspectives i.e. not just technology-driven but also by keeping into account social, behavioural and economical aspects.

Whilst the area of identity management is quickly maturing and commoditising from a technological and solution perspective, little has been done so far to understand implications of these solutions in complex enterprise contexts by factoring in human behaviours, policies, social aspects and legislation.

Hence, the main research opportunity we see in this space is in "turning the table around" and focusing on the decision makers' perspective (rather than on the usual IT perspective), by providing decision support tools and solutions that allow them to explore and predict the impact and consequences of their decisions, by keeping into account all the above aspects – on factors that are of relevance to them, such as costs, security risks and exposures, financial losses and impacts on trust and reputation.

This is a very promising area considering the current enterprise trends in the strategic/executive decision-making area from a compliance-driven approach to a risk-driven approach.

Research in this space can potentially be very challenging, as the predictive capabilities of models and simulations depend on the availability of observational data, expertise in this space, access to CIOs/CISOs and their perspectives and validation of the outcomes on the field. The collection of observational data could be particularly challenging,

due to the potential lack of this information and related statistics, especially in emerging scenarios.

A related challenge to be addressed is how to make use, at the best, of "thought experiments", providing valuable qualitative predictions to decision makers (in absence of additional data), the other based on modelling and simulations based on real-world data and related observations.

Another important challenge to address is how to properly model and simulate complex human and social behaviours, in particular large populations where different categories of behaviours could apply and where these behaviours could change over time.

Equally, bringing economic ideas of understanding the incentives behind individuals and organisational behaviours can help build more realistic models.

We are planning to carry our explorations in this space and identify suitable modelling and simulation approaches and tools (see Section 5) to deal with the involved complexity.

To deal with these issues we also aim at establishing collaborations with universities having track records in this space and engaging in joint collaborative projects (e.g. [32]) to make progress.

8. Next Steps

We believe that to make steady progress in Identity Analytics it is important to identify and work on relevant scenarios where this methodology can be applied and proved, i.e. where the impact of identity and identity management solutions can be explored, along with involved processes, human interactions and behaviours.

So far we have identified a few scenarios we believe could provide meaningful insights. These scenarios include: exploring the introduction of web 2.0 technologies and social networks within enterprises; implication of password policies in heterogeneous contexts, involving multiple cross-organisational entities (e.g. by involving identity federation); lifecycle of confidential and personal data within and across organisations; identity thefts and phishing coupled with users' behaviours and education.

As anticipated in the previous section, we are planning to engage with decision makers in HP business groups, customers and joint collaborative project (e.g. [32]), to refine these scenarios.

We also plan to incrementally build significant (and validated) models covering various aspects of identities, identity management and related users' behaviours, characterising the factors of relevance (for decision makers) that are influenced by these aspects. The goal is to create a "library" of models and associated knowledge that can be reusable and adapted in new contexts, under new circumstances.

For example, some of these common and reusable models could provide validated predictions about the implications of password policies and password lifecycle management to reduce unauthorised access, could describe the impact of identity provisioning and deprovisioning in terms of automation and risk reduction, could describe the impact of using different and/or multiple authentication mechanisms (e.g. device tokens, etc.) on users' behaviours, etc.

Expertise must be created on how to build these models in such as way that they can be easily reused (with minimal efforts in terms of changes and adaptation) in new scenarios and contexts. This includes defining validation patterns to ensure the suitability of these models to the new circumstances. This is an area where we plan to work.

We believe that similar progress has to be made in identifying "repeatable" set of trade-offs, that can be meaningful in different scenarios and contexts and be able to link back to a well defined set of models underpinning this kind of analysis. This is another area we are planning to work.

Libraries of reusable models and associated trade-offs, along with common patterns and a suitable knowledgebase are the very foundation of "Identity Analytics' Decision Support Tools" (Figure 11), tools that can be quickly adapted and reused in various circumstances and scenarios.

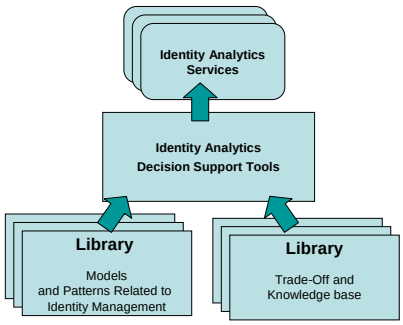

Figure 11. Key Elements of Solutions based on Identity Analytics

Ultimately, in the long-term, we believe there is an opportunity to explore the value proposition of "Identity Analytics as a Service", i.e. providing Identity Analytics as "customisable services" accessible directly by decision makers and/or to be used for strategic consulting purposes [33]. We believe

this will also shape the future funding landscape of identity management, in terms of technologies and solutions.

Our longer terms objectives also involve factoring in, in the context of "Identity Analytics", other important aspects, such as exploring the implication of privacy legislation and data flows driven by privacy constraints, consent and data revocation, data protection guidelines and users' privacy behaviours. These are important areas as they bring in legal, social and technological aspects whose implications and impacts are of importance for decision makers. We are looking for collaboration opportunities in this space with other organisations and joint projects.

From a practical perspective, we aim at making progress in all these directions by also engaging with HP business groups and customers, both in terms of gathering valuable requirements and validating our work in their real-world environments.

As anticipated in Section 5, different modelling and simulation techniques need to be evaluated (and potentially new approaches created) to understand their pros and cons and ensure that the most suitable approaches and tools are used in this space. Activities in this direction are part of our next steps.

9. Conclusions

This paper has discussed the concept of "Identity Analytics" in enterprise and set the context for future research and innovation in this space. In our view it is necessary to "turn the table around", by focusing on aspects of identity and identity management from a decision maker's perspective, in emerging organisational scenarios, where technology is only one of the important factors and where human behaviours, social aspects, economics, emerging trends and legislative factors need to be considered as well.

In this context, we described the role that Identity Analytics can play as a mechanism and tool to explain and predict the impact of identity, identity management and other related aspects (such as user's behaviours and social aspects) on key factors of relevance to decision makers (e.g. CIOs, CISOs), in complex enterprise scenarios – based on their initial assumptions and investment decisions. The goal is to provide decision support and "what-if" analysis to decision makers, to explore possible trade-offs (e.g. using technologies vs. changing processes vs. investing in education of users, to change their behaviours) driven by an economic perspective and formulate new policies or justify existing ones.

We discussed our vision and the methodology we intend to adopt, based on the adaptation of the "scientific method" to this domain. To ground some of the Identity Analytics concepts we illustrated in this paper, we discussed, as an example, focusing on emerging "web 2.0" enterprise data sharing scenarios, where unstructured information is created, stored and shared by people in collaborative contexts, within and across organisations. Some qualitative outcomes and what-if analysis have been provided and discussed.

We believe this area is plenty of research opportunities as well as challenges to overcome, in terms of identifying suitable scenarios, gathering relevant observational data, being able to access the expertise and judgment of CIOs/CISOs and validate work in this space in real-world scenarios, by means of trials. We described our plans to address these challenges along with our next steps.

10. Acknowledgements

We would like to thank our colleagues Yolanta Beres, Jonathan Griffin, Brian Monahan, David Pym and Mike Wonham for their input and valuable feedback.

11. References

[1] HP Labs, Systems Security Laboratory (SSL), HP Labs, http://www.hpl.hp.com/research/systems_security.html, 2008

[2] Security Analytics, HP Labs, SSL, http://www.hpl.hp.com/research/systems_security.html, 2008

[3] Penn, J., IT Trend 2002: Directories and Directory-Enabled Applications, IdeaByte, 2002

[4] Neuenschwander, M., Meta-directory Services and the Emerging Enterprise Data Network, The Burton Group, 2002

[5] Smith, R.E., Authentication: From Passwords to Public keys, Addison-Wesley, 2001

[6] Burton Group, User Authentication, Burton Group, 2002

[7] Volchkov, A., Revisiting Single Sign-on. A Pragmatic Approach in a New Context, pp. 39-45, IT Pro, IEEE, 2001

[8] De Clercq, J., Single Sign-On Architectures, proceedings pp. 40-58, InfraSec 2002, Bristol, UK, 2002

[9] Liberty Alliance Project, Liberty Project web Site, http://www.projectliberty.org/, 2008

[10] Liberty Alliance Project, Liberty Architecture Overview, http://www.projectliberty.org/, 2008

[11] Blum, D., Toward Federated Identity Management, Burton Group, 2002

[12] Ferraiolo, D., Kuhn, R., Role-based Access Control, NIST, 1992

[13] Penn, J., Market overview: user Account Provisioning, GIGA Information Group, 2002
[14] Housley, R. Ford, W., Polk, W., Solo, D., RFC2459: Internet X.509 Public key Infrastructure Certificate and CRL Profile, IETF, 1999
[15] Microsoft CardSpace, Introducing Windows CardSpace, http://msdn.microsoft.com/en-us/library/aa480189.aspx, 2006
[16] OpenId, OpenId Foundation, http://openid.net/, 2008
[17] SOX: "Sarbanes-Oxley Act", http://www.sarbanes-oxley.com/, 2005
[18] GLB: "Gramm-Leach-Bliley Act: Privacy of Consumer Financial Information", http://www.ftc.gov/privacy/glbact/glboutline.htm, 2003
[19] Casassa Mont, M., Thyne, R., Privacy Policy Enforcement in Enterprises with Identity Management Solutions, Journal of Computer Security (JCS), Volume 16, Number 2/2008, 2008
[20] Casassa Mont, M., Privacy-aware Information Lifecycle Management in Enterprises: Setting the Context - Information Security Solution Europe 2006, 10-12 October 2006, Rome, Italy, ISSE 2006, 2006
[21] Baldwin, A., Casassa Mont, M., Beres, Y., Shiu, S., On Identity Assurance in the Presence of Federated Identity Management Systems, ACS CCS 2007 Workshop on Digital Identity Management, DIM 2007, 2 November 2007, George Mason University, Fairfax, VA, US, 2007
[22] Aliprandis, C.D. (Editor), Economic Theory Journal, http://www.springer.com/economics/economic+theory/journal/199, 2008
[23] Nobay, R.A., Peel, D.A., Optimal Monetary Policy in a Model of Asymmetric Bank Preferences. London School of Economics, Mimeo.
[24] Varian, H., A bayesian approach to real estate management. In S.E. Feinberg and A. Zellner, editors, Studies in Bayesian Economics in Honour of L.J. Savage, pages 195–208. North Holland, 1974
[25] Anderson, R., Bohme, R., Clayton, R., Moore, T., Security economics and the internal market, European Network and Information Security Agency, 2007
[26] Anderson, R., Moore, T., The economics of information security. Science, 314:610–613, http://www.cl.cam.ac.uk/˜rja14/Papers/toulouse-summary.pdf, 2006
[27] Anderson, R., Why information security is hard: An economic perspective. In Proc. 17th Annual Computer Security Applications Conference, 2001
[28] Gordon, L.A., Loeb, M.P., The Economics of Information Security on Information and Systems Security, 5(4):438–457, 2002
[29] Yearworth, M., Monahan, B., Pym, D., Predictive modelling for security operations economics (extended abstract). In Proc. I3P Workshop on the Economics of Securing the Information Infrastructure, Proceedings at http://wesii.econinfosec.org/workshop/, 2006
[30] Beautement, A., Coles, R., Griffin, J., Ioannidis, C., Monahan, B., Pym, D., Sasse, A., Wonham, M., Modelling the human and technological costs and benefits of USB memory stick security, to appear in Proceedings of WEIS 2008, 2008
[31] Wilson, E. B., An Introduction to Scientific Research McGraw-Hill, 1952
[32] Trust Economics, UK DTI grant P0007, Trust Economics Project, 2008
[33] Pym, D., Taylor, R., Tofts, C., Yearworth, M., Monahan, B., Gittler, F., Systems and services sciences: a rationale and a research agenda (Open Analytics Project, HP Labs, Bristol, UK), http://www.hpl.hp.com/techreports/2006/HPL-2006-112.html, 2006
[34] Taylor, R., Tofts, C., Model Based Services Discovery and Management, PICMET 2008, 2008
[35] Taylor, R., Tofts, C. Taking a RaSP to Enterprise Stakeholder Dissonance, accepted EDOC 2008, 2008
[36] Demos2k, Demos 2k, http://www.demos2k.org/, 2000
[37] Birtwistle, G., Demos,discrete event modelling on Simula. Macmillian, 1979
[38] Pym, D., Monahan, B., A Structural and Stochastic Modelling Philosophy for Systems Integrity. HP Labs Technical Report Series, HPL-2006-35, Feb 2006
[39] Brian Monahan, DXM - The Demos eXperiments Manager, HP Labs Technical Report, 2008
[40] Beni, G., Wang, J. Swarm Intelligence in Cellular Robotic Systems, Proceed. NATO Advanced Workshop on Robots and Biological Systems, Tuscany, Italy, June 26–30, 1989
[41] Thietart, R. A., Forgues B., Chaos Theory and Organisations, *Organization Science*, Vol. 6, No. 1, Focused Issue: European Perspective on Organization Theory pp. 19-31, Jan. - Feb., 1995
[42] SailPoint, SailPoint – Identity Risk Management, http://www.sailpoint.com/product/reporting.php, 2008
[43] Beres, Y., Baldwin, A., Shiu, S., Model-based Assurance of Security Controls, HPL Technical Report, HPL-2008-7, 2008
[44] Oracle, Reporting and Auditing Solutions Roadmap, ftp://ftp.oracle.com/sales/outgoing/oam/roadmap.pdf, 2008
[45] IBM, Identity Analytics, http://www-935.ibm.com/services/us/gbs/bus/pdf/g510-6527-ibm-identity-risk.pdf, 2008
[46] IdAnalytics, IdAnalytics, http://www.idanalytics.com/, 2008
[47] Shay, R., Bhargav-Spantzel, A., Bertino, B., password policy simulation and analysis, DIM 2007, 2007
[48] ISO, ISO 27001, Information Security Management, http://www.iso.org/iso/catalogue_detail?csnumber=42103, 2005
[49] ISACA, Cobit, IT Governance, http://www.isaca.org/Template.cfm?Section=COBIT6&Template=/TaggedPage/TaggedPageDisplay.cfm&TPLID=55&ContentID=7981, 2008
[50] ITIL, ITIL IT Infrastructure Library for Service Management, http://www.itil-officialsite.com/home/home.asp, 2008
[51] Grimmett, G. and Stirzaker, D. "Probability and Random Processes", 3rd ed., Oxford UP, 2001