



Communio

**Use of Data Analytics for Responsible Gambling
Regulatory Directives**

Final Report

Ontario Problem Gambling Research Centre (OPGRC)

Prepared by: David Aro, Laura Jakob and Dr Sally Gainsbury

January 2015
Version: 1.1

Table of Contents

Table of Contents	1
Introduction	2
Project Methodology	5
Types of data	8
Gambling Data	12
Data issues and barriers	17
Research Questions	20
What ‘types’ of data are collected by other stakeholders and how are these being monitored and analysed by regulators and private operators for compliance and responsible gambling purposes? ..	20
What evidence exists to ensure that these processes (data collection, analysis and monitoring) are having their desired effect?.....	28
How has/is data being utilised across industries outside and including gambling to ensure that regulatory standards are being met?	29
How is behavioural data being used to inform and evolve regulation?	36
What constitutes best practice related to the use of operator and/or regulator data to ensure compliance with regulatory standards?.....	41
Conclusions	45

Introduction

Background

Over the last twenty years the quantity and quality of data collected from the general public, commercial operators, and government agencies has grown exponentially. Data collection about personal behaviours is now ubiquitous in our society with large amounts of information being accumulated and aggregated at multiple points. As these sources of data have grown, companies, governments, and individuals have come to realize the role that the information derived from this data may have. Significant advances in the availability, use, and storage of different types of data coupled with technical advances in the ability to analyse diverse datasets have also helped to drive the increased interest in how data can be effectively used. Despite these advances, the integration of big data methodology and approaches into various sectors remains, often times, in its early stages.

Project Scope

The issue of gambling specific regulation is broad. It's development and enforcement ranges extensively across jurisdictions internationally (Graf, 2013). In order to assist in determining an effective scope for this project, the project advisory group, within the context of this project, defined 'regulatory standards' as those relating specifically to those impacting Responsible Gambling (RG) and consumer protection issues.

The scope of the project was also designed to consider data usage beyond the borders of Ontario, focusing on the research questions in a broad manner, unconstrained by specific regulation and/or legislative requirements. Where possible, issues that impact and/or influence the specific use of data to inform regulation within the Canadian environment have been identified. They have not however, limited the scope of the proposed and discussed issues/ideas.

Responsible Gambling and/or Consumer Protection Regulation

Overall, the purpose of RG regulation is to facilitate the sustainability of gambling, by maximizing strengths and minimizing harms (Griffiths & Wood, 2008). Minimizing these harms includes addressing issues such as the protection of vulnerable and under age players, the prevention or reduction of gambling being linked to crime, the conduct of fair and transparent gambling and the provision of tools/standards and resources to help prevent the development of gambling related problematic behaviours (Lycka, 2011; Lycka, 2014).

The method for implementing RG regulation in practice may range from mandatory implementation, supported by legislation to jurisdictions where gambling operators are asked to comply with voluntary codes of conduct (Delfabbro et al., 2012). The approaches to the implementation and enforcement of RG standards within these settings therefore, are by necessity, different.

Further, these variations in approaches to RG regulation are often supported by diverse underlying philosophical approaches regarding the act of gambling itself. This philosophy serves to influence the type of regulation enforced within a jurisdiction, the lengths to which these regulations may impact on a player's gambling experience, privacy and the infrastructure that exists to support gambling within the sector. For instance, many jurisdictions have implemented systems which aim to identify indicators of potentially problematic gambling behaviour. The manner in which systems and/or employees are expected to respond to this behaviour however, differs internationally. Delfabbro et al.

(2012) note that Hafeli & Schneider (2005)¹ reported that people displaying the potential indicators of problem gambling in Swiss casinos would have reports opened and have their behaviour documented. In another jurisdiction, gamblers who are displaying 'high-risk' behaviour may have their Internet gambling account closed and be prevented from using a particular operators system.² Whereas in some jurisdictions no such programs are currently being implemented, thus there is no requirement for staff to actively identify or take action in response to high-risk gambling behaviour.

The Regulatory Environment

This diversity in RG practices speaks to the complexity of the regulatory environment in gambling. Regulators are being faced with new and emerging challenges as they are forced to consider and address issues such as regulation in the Internet gambling sector, responsibilities in relation to data collection and privacy, and accountabilities in relation to duty of care for those who may be identified as gamblers exhibiting high-risk behaviours.

The Internet gambling environment has developed quickly, providing a challenging landscape from which to plan and implement regulation. However, with these challenges come new opportunities with comprehensive player behaviour (account based) becoming readily available. This type of data will be able to provide insights into gambling behaviour that were challenging to study with land based gambling. Evidence suggests that we are already learning about this type of gambling. For instance, it was previously postulated that Internet gambling resulted in greater incidence of problem gambling behaviours, however research has emerged to challenge this assumption (Levi et al., 2009; Gainsbury et al., 2014; Philander & MacKay, 2014). Player behaviour data has the power to provide significant insight into Internet gambling and also to understand the impact of the games upon gamblers in a more meaningful manner.

Previously there had been significant concerns regarding the potential risk for emergence of 'risky' and/or 'criminal' behaviours to proliferate in online environments (e.g., money laundering). As technology has progressed however, there has been some consensus that the ability to scrutinize online behaviour and to track each transaction acts to limit the associated risks of this occurring in the online environment to levels equal to that seen in other online industries (Lycka, 2011).

Internet gambling presents regulators with new opportunities for collaboration given the borderless reach of some of the gambling suppliers. Depending upon the regulations governing Internet gambling within a jurisdiction, the possibilities of sharing common platforms as well as regulatory standards and enforcement can be considered. An example of regulators collaborating on a shared platform is the PlayNow online site developed by the British Columbia Lottery Corporation (BCLC) and regulated by the Gaming Policy and Enforcement Branch of the Ministry of Finance in British Columbia. The Manitoba regulators, The Liquor and Gaming Authority of Manitoba reviewed the regulatory framework for BCLC's internet gambling platform in terms of integrity and RG requirements. The decision was ultimately made to partner and release the site based on BCLC's online platform whilst ensuring compliance with Manitoba's regulatory standards (Manitoba Lotteries, 2012).

Conversely, jurisdictions such as the United Kingdom (UK) are currently enacting changes to the regulatory requirements to require all gambling operators conducting business within the UK to obtain a UK gambling license. This is as opposed to a requirement for only those who are based in the UK with permission granted for sites regulated in 'White Label' jurisdictions to provide services to individuals within the UK. This change will allow the UK to update or change their requirements,

¹ Article referenced was not written in English and thus could not be included for analysis.

² Consultation

including responsible gambling regulations, as required without having to consult with or convince regulators in other jurisdictions to agree (Gambling Commission, 2014).

Terms and Definitions

This report examined the use and availability of data to inform RG/consumer protection regulations internationally. Terminology and game availability across different jurisdictions varies markedly. In order to facilitate interpretation of the findings the following terms and definitions will be used throughout the report:

Term	Definition
EGM	Electronic Gaming Machine
PGSI	Problem Gambling Severity Index
RG	Responsible Gambling

Project Methodology

In order to complete this project Communio employed a three stage project methodology. This included:

Phase	Key Activities
I Rapid Evidence Assessment	<ul style="list-style-type: none"> Complete detailed Rapid Evidence Assessment using multiple sources of references and evidence
II National and International Consultation	<ul style="list-style-type: none"> Reach out to national and international experts across multiple fields to gain insight into current data practices.
III Data analysis and synthesis	<ul style="list-style-type: none"> Data analysis and report synthesis

Research Questions

Several specific research questions were identified to help guide the work of the project. These include:

Primary Research Question

What constitutes best practice related to the use of operator and/or regulator data to ensure compliance with regulatory standards?

Secondary Research Questions

- What 'types' of data are collected by other stakeholders and how are these being monitored and analyzed by regulators and private operators for compliance and responsible gambling purposes?
- What evidence exists to ensure that these processes (data collection, analysis and monitoring) are having their desired effect?
- How has/is data being utilized across industries outside and including gambling to ensure that regulatory standards are being met?
- How is behavioural data being used to inform and evolve regulation?

Phase I - Rapid Evidence Assessment

A REA methodology was employed to review the available evidence related to each of the research questions. The methodology employed mirrored that suggested by the UK toolkit.

Several sources of information were searched in order to identify relevant material. These included academic databases:

- ABI/INFORM Global
- PsycINFO [via Ovid]
- Project Cork
- Academic Search Complete
- MEDLINE [via PubMed]
- Science Direct
- Web of Science [via ISI Web of Knowledge].

Gambling specific databases:

- Canadian Partnership for Responsible Gambling
- Alberta Gambling Research Institute
- Australasian Gaming Council
- Responsible Gambling Council
- Gambling Research Australia.

Grey literature included the websites/blogs of private companies, government departments, regulators and EGM/VLT operators globally as well as YouTube videos and presentations from relevant conference proceedings.

Trade publications were also searched in order to identify relevant information:

- Canadian Gaming Business
- Canadian Gaming News
- International Gaming & Wagering Business
- World Lottery Association Magazine

The following search terms were used throughout the project:

- data analysis
- data mining
- big data
- regulatory compliance
- monitoring
- gambling
- communications
- business intelligence
- regulatory standards
- regulation
- enforcement
- best practice
- gaming
- financial
- health
- marketing

The following inclusion/exclusion criteria were applied to all of the information identified through the searches:

Inclusion criteria include:

- articles published between 2004 – 2014 (i.e., internet era);
- published in English;
- have available abstracts;
- articles relevant to the subject matter in a range of sectors (healthcare, finance, communications etc.).

Exclusion criteria include:

- lack relevance to the research question.

The REA demonstrated that while there was a significant amount of literature that examined data analytics, there was little information identified in the academic literature that pertained directly to the gambling sector and/or the enforcement of regulation. In order to address this gap, the REA protocol included searching for material from sectors outside the gambling field. The majority of the literature identified that pertained to our subject matter was not evaluative in nature. Rather, it was comprised of editorials, articles to inform policy and case reports. Often reference lists pointed to blogs, conference proceedings etc. This supports the opinion that this subject matter is on the leading edge of research and development.

The most valuable source of evidence was extracted from large industry or industry representative reports, social policy articles, government reports, and conference proceedings, which outlined work in specific jurisdictions. As a result, it was not possible for the research team to effectively evaluate the quality of the majority of material sourced during the REA as it was not peer-reviewed.

Phase II - National and International Consultation

In order to build upon the evidence uncovered during the REA, Communio identified a comprehensive list of international key stakeholders with a great deal of expertise from a wide variety of perspectives including:

- Gambling operators (international and national)
- Gambling regulators
- Liquor regulators
- Experts in the RG field
- Representatives from the private gambling industry
- Private industry providing data analytics tool and services
- Pharmaceutical industry experts
- Health industry experts

In order to engage with stakeholders and to facilitate the sharing of valuable information, interviewed participants were assured that they and their organisations would not be specifically identified throughout the report. In order to present their findings comprehensively, the results of the consultation are presented in an integrated fashion with the outcomes of the REA.

Phase III – Data Analysis and Reporting

The results of the REA and consultation were integrated and key themes and patterns identified and analyzed. Dr. Sally Gainsbury, an international expert in RG, provided expert guidance and direct input into the analysis, interpretation of findings, and drafting of the report. Strategies that were employed to facilitate the analysis included multiple research group teleconferences, review of themed content and multiple draft analyses.

Types of data

There are a variety of different types of data that can be collected and used to inform decision-making in a myriad of ways. Currently there is a significant emphasis regarding the use of 'big data' in the popular discourse. Big data differs significantly from the traditional use of data in a number of ways.

An overview of different types of data is included below. Specifically, a comparison between traditional data stores and big data frameworks and analysis is outlined.

Traditional Data Stores

Across sectors, traditional sources of data have provided stakeholders in numerous areas with valuable information to inform decision-making in all areas of business, healthcare, and government sectors. There has been increased need for the use of evidence-based decision making, supported by data collected from a number of sources and by a variety of stakeholders. Within the gambling sector, data has been available and collected for a number of years however, the purpose of this collection, its frequency, reliability, and ultimate usefulness of the data varies considerably.

Large amounts of data have traditionally been saved and stored in a data warehouse. This warehouse environment is a static environment which is understood through the use of a data schema (Roski et al., 2014). The data warehouse environment requires work to extract different forms of data and save them in a format which allows analysis in a similar format to facilitate meaningful analysis (Roski et al., 2014). Although this environment is capable of handling significant amounts of complex data, it may be not be able to cope with the extreme volume and associated storage requirements of big datasets (Chen & Zhang, 2014; Roski et al., 2014).

The nature of this type of traditional data analysis is (by necessity) responsive. Hypotheses are formed and large amounts of data are analyzed to test these statements. To facilitate this type of analysis, data in large warehouses is transformed and subsequently queried to provide insight and direction (Agrawal et al., 2011). This method of analysis is useful and has become integrated across all sectors.

Big data

There are a number of characteristics that differentiate big data from traditional sources of data. Although there is no agreed, formalized definition of big data within the literature, the definition by Gartner (2012) in Figure One is an example gaining popular acceptance

“...high volume, high velocity and/or high variety information assets that demand cost effective innovative forms of information processing for enhanced insight, decision making and process optimization”

Figure One: Big Data Definition

Source: Gartner, The Importance of 'Big Data: A Definition. <http://www.gartner.com/id=2057415>

It is generally agreed in the literature that it is the characteristics included in the Gartner definition that describe big data: volume, variety, and velocity (Australian Government, 2014; Association for Data-Driven Marketing & Advertising [ADMA], 2013; Bates et al., 2014; Gartner, 2012; Manyika et al., 2011; Roski et al., 2014). This means that extremely large pools of data, which have not previously been used, are obtained from multiple sources and in a variety of formats (Chen & Zhang, 2014). Further, it is understood that big data is not static, rather it is constantly evolving, changing, and being gathered (ADMA, 2013).

As a result, some traditional data warehouses and analytic processes are unable to cope with the quantity and complexity of data that is generated in multiple formats (Chen & Zhang, 2014; Manyika et al., 2011; Roski et al., 2014). The most significant barriers to the effective use of big data have been identified as: the heterogeneity of the data, privacy issues, scale, timeliness, and human collaboration (Agrawal et al., 2011).

Big data can be accommodated in new data warehouse environments designed to house the vast amounts of data or in what is referred to as a 'data lake' (Chen & Zhang, 2014; Roski et al., 2014). Data lakes enable an environment where data elements can be used regardless of their varied format. These elements can be tagged and linked together using a series of schemas which can be altered to reflect the questions posed (Roski et al., 2014). The power of this flexible arrangement is not only in the quantity of the data analyzed but that questions can be posed that require input from several sources of information (e.g., text data, electronic formats, pdf etc.) (Roski et al. 2014).

Broadly, what constitutes big data varies depending up on the context and the purpose of the collection and the intended usage of the outcomes. For instance, it may include:

- An individual's web behaviour (usage patterns)
- Social media content developed by individuals (e.g., Facebook posts, tweets etc.)
- Geo data (information about locations)
- Environmental data
- Radio Frequency Identification Data (that tracks or tags objects, people or things, e.g., may be used for manufacturing to track the movements of products or to tag cars for tolling purposes etc.)
- Research data (census data, information regarding health and social welfare)
- Operational data (data produced by business in executing the course of their work, for instance EGMs) (ADMA, 2013).

The power from using these different types of data sources together in the big data context is that it facilitates a multi-layered approach, using multiple sources to detect patterns, answer questions and provide insights. The key difference between traditional data analysis and big data in this regard is therefore, the ability and desire to use the information to answer different questions. Principally, to *"...understand something before it happens. To intuit the future behaviour of individuals."* (Stottardt, 2013. https://www.priv.gc.ca/media/sp-d/2013/sp-d_20131017_e.asp). However, to realize this in practice means to rely heavily on the use of new and emerging data analytical techniques. The difference between these and traditional analytic processes is explored below.

Analytic processes

A variety of data analysis processes can be applied within the big data environment. This varies from basic analytics where traditional forms of data are searched, reviewed and analyzed to make and inform present decisions to big data analytics where complex and real-time analysis is possible, facilitating new forms of gaining insights from data (Booz Allen Hamilton, 2012). The differences between the different types of analysis in terms of analytic complexity and data size are outlined in Figure Two.

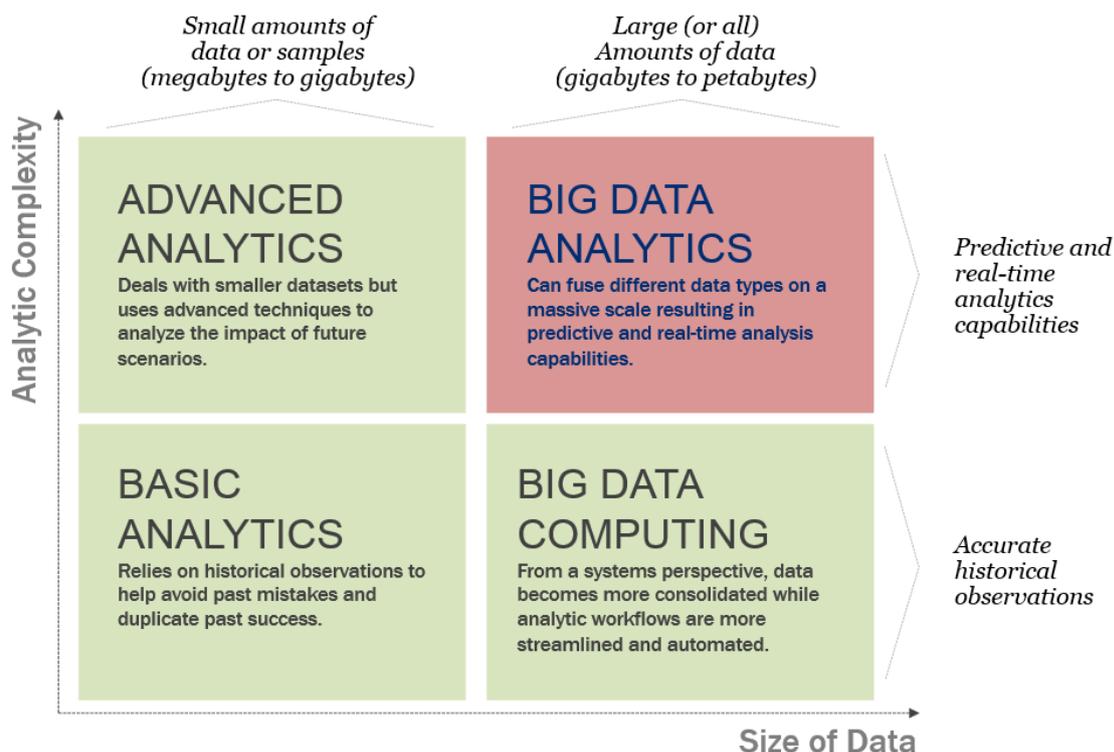


Figure Two: The relationship between analytic complexity and size of data stores

Source: Booz Allen Hamilton (2012) Cloud analytics playbook. Retrieved from: <http://www.boozallen.com/media/file/cloud-analytics-playbook.pdf> (pg. 7)

The key difference in terms of analytic complexity is the ability to inform historical versus present or future (predictive analytics) problems. High level examples of various analytical techniques are included below for contextual purposes.

Data mining

Data mining is a specific analytic technique that is commonly referred to in the context of big data to explore large datasets without being guided directly by hypotheses statements (Bates et al., 2014). It involves the discovery of patterns or relationships across different forms of data (Bierig et al., 2013).

Text analysis

This refers to the identification or source of information from large text based data sets. For instance, open-ended survey questions, customer complaints, website or social media comments etc. The type of information that can be derived from such information stores varies widely but may include explicit, implicit and inferred information, sentiment, metaphors etc. (National Research Council Canada, 2014).

Predictive analytics

Predictive analytics can be defined as the “*use of electronic algorithms that forecast future events in realtime.*” (Cohen, Amarasingham, Shah, Xie & Lo, 2014). This type of analysis relies on the use of a variety of data sources which are integrated to help make specific predictions. The technology has specific applications in healthcare, but also more broadly in terms of consumer behaviour, risk and threats.

Social networking analysis

Uses network theory to organize and view social relationships. It includes consideration of social system design, human behaviour modeling, social network visualization and social networks evolution analysis (Chen & Zhang, 2014).

Gambling Data

Introduction

Within the gambling context there is a vast amount of data collected from a variety of sources. Some gambling stakeholders (regulators and operators) have indicated that effective analysis and use of this data may hold promise for the development and effective enforcement of RG regulation; other stakeholders have expressed reservations about its use due to the perceived risks this may entail.

Current capabilities

The data collected by the gambling sector serves a number of different purposes including the desire to use the data for commercial insights and improvements (e.g., marketing and product updates) and to ensure that gambling operators meet taxation, and legal requirements. The results of this report have indicated that data is currently mostly used for the former, and that there is limited use of data in the regulatory context at this time, both within the gambling sector and more broadly. Factors that contribute to the ability to use data to inform RG regulation and the extent to which it can be used include: the philosophical framework that guides how gambling is perceived and offered, the nature/structure of the regulatory requirements within each jurisdiction, the governing laws (e.g., privacy), the technological infrastructure which exists to support the industry, and the availability of relevant datasets.

There is some demonstrable interest in gathering data and understanding how it can inform RG policy. The UK has taken some significant steps towards this goal. In 2010 the Responsible Gambling Strategy Board in the UK prioritized work examining the relationship and between gambling machines and behaviour. This was supported by a call to understand the nature of the data available from EGMs with the highest stakes and prizes within Great Britain and how it could be used. This project was undertaken and resulted in an EGM data research report by Wardle, Seabury, Ahmed and Coshall, researchers at NatCen Social Research (2013). There are plans to progress this research with a consortium consisting of researchers and consultants, commissioned by the Responsible Gambling Trust (RGT) to investigate whether data from electronic gaming machines more broadly can be linked to problem gambling (Blaszczynski, Parke, Parke & Rigbye, 2014; NatCen, 2014).

Land based gambling

The NatCen EGM data research report (Wardle et al, 2013) highlighted the vast amount of data collected by EGMs across Great Britain and the strengths and challenges associated with using this data in a meaningful way (Wardle et al., 2013). Although the report focused on Great Britain, the similarities between gambling environments and machines, as well as the social and political climates elsewhere mean that many of the findings may apply across jurisdictions. The classifications of data outlined in the report are provided as a framework below to guide this discussion.

Transactional data

Transactional data refers to the accounting data collected by EGMs to provide an auditable trace of financial accounts and records of the money earned for tax, accounting and audit purposes (Wardle et al., 2013). This information may or may not include two basic levels of data, accounting and behaviour data and importantly, is collected and presented by EGM machine level as opposed to the player level (Wardle et al., 2013).

Further, this information can be collected by EGMs in aggregate or at the atomic level. These are described below (Wardle et al., 2013):

Atomic data	'Atomic' level data is game information collected for every game cycle. This means that specific, detailed information is available and it facilitates analysis of the amount of money staked, outcome, how often the game is played, and type of money used (e.g., coins, notes, credit card etc.).
Aggregated data	'Aggregate' data refers to the atomic data that has been aggregated by a variable (e.g., month, day, hour, game etc.) and/or may refer to the fact that the EGM only collects and records aggregated data. This is often dependent on the regulatory requirements for the extraction of data (e.g., frequency and timing). EGM data extracted solely in aggregated form means that analysis at the atomic level is not possible (Wardle et al., 2013).

Table One: Different types of transactional data

Retrieved from: Wardle et al., 2013

The researchers found two important challenges regarding the use of transactional data for research and policy purposes. Firstly, they identified that two separate systems exist within each EGM; the transactional system and the game system. The former is responsible for the transactional accounting data (e.g., cash-in, cash-out). It interacts with the game system which is responsible for the specific game features (e.g., multiple stake options) however, the game system does not produce standardized output regarding game activity and rather the focus of the data is the transactional data collected. The only information that is documented routinely is the amount staked and returned to player and the manner in which this is reported may vary by machine (e.g., frequency of reporting).

Secondly, the researchers found challenges regarding the game cycle data. A game cycle refers to the *"process of placing a bet on one game and one outcome"* (Wardle et al., 2013, pg. 15). The information regarding a game cycle is not always recorded in a standardized manner by different EGMs and thus may not be accessible for analytical purposes. As a result, certain information about game play is not consistently retrievable by some EGMs in Great Britain. This information includes (Wardle et al., 2013, pg. 18):

- Use of specific features such as the autoplay button
- Use of credit transfer and banking features
- Staking patterns
- Browsing activity
- Any game feature activity (e.g., progress towards bonuses).

Player tracking data

Player tracking data refers to game play which is linked to a loyalty card program (or a mandatory player card system). It has previously been defined as:

“a centralized or linked system for recording and storing consumer transactions/customer interactions (e.g. visa, banking, insurance, vehicle maintenance, other customer databases).”

Figure Two: Definition of player tracking data

Retrieved from: Schellinck & Schrans (2008). Player tracking analysis: inside the black box. Using player tracking data to manage risk. European Association for the Study of Gambling Slovenia

Player tracking data allows the entire record of one period of play (which is indicated by the insertion of a card into an EGM) to be recorded including quantity and time spent playing as well as on how many machines (Griffiths & Wood, 2008). This may be recorded in either atomic or aggregated format. An important consideration for the use of player tracking data in jurisdictions without a mandatory player card is that participation in the schemes is optional, as is use of the player card. At any point in the gambling experience, the player may remove their card from an EGM and continue to gamble without traceable records. Thus the impact of this non-representative data (gamblers self-select whether they would like to participate) coupled with the potential for missing data must be considered in any analysis (Wardle et al., 2013).

The NatCen EGM report identified that, despite there being a significant amount of data collected by EGM operators in Great Britain, the ability of the data to inform research (and consequently policy) was seriously impeded by the data quality. The current situation was characterized as *“data rich, information light”* (Wardle et al., 2013, pg. 58). Further, the researchers identified that the attempts by operators to link player data to machine sessional data remained early in development despite a significant level of interest.

Player tracking data in jurisdictions with mandatory player cards

In some jurisdictions (Norway, most recently Sweden and previously in Nova Scotia) mandatory player cards have been introduced to facilitate account based play. This type of play allows for individuals to track their play, facilitates enforcement of regulation in certain environments and allows for data to be collected and potentially analyzed by operators and regulators.

Player tracking data is a significant and important element of data both to understand how gambling impacts upon behaviour and because of the limitations of other methods of data collection within the gambling environment. Recent research by Braverman, Tom, and Schaffer (2014) compared self-report versus actual gambling data as recorded by online player tracking data. The results illustrated that gamblers were not accurate in their recall. Specifically, they found that gamblers did not accurately report their gambling behaviour particularly over the longer term (i.e., 12 months). Further, those with gambling-related problems were more likely to report figures with incongruities (Braverman et al., 2014). Conversely, the opposite is also true. Gambling data is only as accurate as the system that it is measuring. Consequently, it may exclude important information such as the impact of different ‘types’ of gambling (e.g., lottery or table games) and/or gambling that has occurred with another operator (Gainsbury, 2011).

Proxy session data

Proxy session data refers to the use of atomic level data that is broken down by specific parameters to reflect actual gambling sessions. Proxy session data may estimate that a gambling session begins based on several criteria including: the length of time the game was dormant, the starting balance (being equal to zero or a minimum stake) and the act of cash being deposited into the machine (Wardle et al., 2013). All proxy session data/calculations are based on estimations and pre-determined, programmed algorithms. As a result it is difficult to determine their accuracy. This inhibits but does not exclude the use of this type of data for analytical purposes.

Internet gambling

The situation for data collection in the online environment is substantially different to land-based venues. The nature and format of gambling in this context allows for much greater, detailed information about each gambler to be recorded and subsequently analyzed (Griffiths & Wood, 2008).

For instance, a profile is required to be registered before participating in online gambling. In many regulatory environments online operators are required to confirm an individual's identity in order to satisfy age restrictions and 'Know Your Customer' regulations that aim to prevent illegal behaviour (e.g., money laundering) and access to games by underage gamblers (UK Gambling Commission, 2014; Levi, 2009; Lycka, 2014). Registration also ensures that gamblers from locations where internet gambling is prohibited are not able to access the games, provided that operators respect and abide by these regulations (Wiebe & Lipton, 2009). The process of developing a profile may include identifying the individual's name, address, date of birth, email address, phone number, and payment details and validating that the information is correct (Lycka, 2014). In order to effectively verify that the information provided by a gambler is correct, Internet gambling operators must take steps to ensure that the person is who they claim to be. These steps include verifying the customer's data with a third party provider, confirming the validity of hard copy documents supplied by the gambler and/or seeking confirmation from other legal authorities (Lycka, 2014). Although, these steps help to confirm that only one account is established by a customer at one Internet gambling site, it does not prevent a gambler from setting up several accounts at different sites, including offshore providers not licensed within the relevant jurisdiction. This type of account based play does not only facilitate appropriate regulatory compliance activities, it also allows account based play where gamblers activity, playing habits, expenditure (total wins/losses) and behaviour is easily traceable.

In regards to regulators level of access to Internet gambling data, some jurisdictions have direct access whereas others access it through the operator. Previously, there was indication that regulators preferred that operators pass their gambling data through the regulators servers to facilitate access and compliance. However, in many international jurisdictions this has proven to be considered overly burdensome and not feasible, particularly where gambling is offered across multiple jurisdictions (Lycka, 2014). An alternative solution adopted by some regulators is for all gambling data to be consistently accessible to regulators in a live or almost real-time environment. This allows regulators to monitor data to ensure compliance with regulatory standards, including clear identification of all players, the details of all financial transactions and sources of income (Lycka, 2011). This ability to monitor and ensure compliance was identified as good practice in a recently published article titled 'Regulating Online Gambling Markets: Regulatory Guide' (Lycka, 2014). Further, the author states that regulators should additionally be able to seek reports from operators on an ad-hoc or regular basis.

Monitoring data in the online environment

Operators frequently conduct automated reviews of online gambling data to identify risky or suspicious behaviour (Levi, 2009). Levi indicated that although there was no specific standard agreed at the time of his report, at least one operator confirmed that they had a team of analysts examine data that included up to 50 different checks examining behaviour at key points such as deposits and withdrawals in the Internet gambling process. The author states that these automated reviews are based on criteria identified through previous experience (Levi, 2009). These include (but are not limited to the following as some features may have been excluded for confidentiality reasons):

- The sum of the initial deposit
- Deposits not immediately used to gamble
- Money deposited and withdrawn without the placement of bets.

This automated monitoring is easier, faster, and more time sensitive than monitoring for similar behaviours in a land based environment.

Data issues and barriers

As the use of big data continues to grow across a number of industries, so too does the number of identified barriers to its effective use. Issues such as privacy regulations, consent requirements, laws and regulations restricting the effective sharing of information, storing, transmitting and curating of the data sets are all barriers readily identified in the literature and consultation phase of the project.

These challenges represent significant barriers to the effective use of data. They are also reflective of the values and perspectives inherent in each society. For instance, in certain jurisdictions, such as the United States, the government is taking an active and engaged approach to breaking down some of these barriers through the introduction of legislation, programs and initiatives to address issues related to data collection, storage, sharing, analysis etc. (Office of Science and Technology Policy, 2012). These are important steps in terms of designing an appropriate framework from which data sharing across government departments can take place. In other jurisdictions, these barriers are still more significant.

Privacy and security concerns with big data

As the interest and use of data in a variety of sectors continues to grow, so do the questions regarding the safeguarding of information in relation to privacy and consent. Although the impact of these issues may result in different consequences across sites and topics (e.g., sharing health related information as opposed to consumer behaviour choices) there are common themes and challenges faced across sectors.

Consumers often 'give away' their personal information but remain generally unaware of the vast amounts of data that are collected about them throughout the course of their everyday activities and/or how that information is being used (Dusseault, 2013). Records maintained by government departments (e.g., welfare, tax office), 'cookies' left on web browsers, loyalty programs (e.g., from the grocery store, casino etc.), information on social networking sites as well as data retained by healthcare professionals in Electronic Healthcare Records (EHRs) provide a vast volume of data that can be used for both business and profit motives and health/social care needs/interventions.

The public have concerns regarding the misuse of their personal information (Stoddart, 2013). Of particular concern is the use of data that is linked to multiple sources (Agrawal et al., 2011). Traditionally, records that contained potentially sensitive information were de-identified in an effort to preserve a person's privacy and reduce concern regarding the use of this data (Tene & Polonetsky, 2012). Recent reports however, have illustrated that the technology available today means that it is not possible to ensure that records are truly de-identified. With the appropriate mix of knowledge, technology, skill and time, often datasets can be re-identified (Cohen et al., 2014; Tene & Polonetsky, 2012). This raises significant questions regarding expectations of privacy, expectations of consent for personal information to be used and who is authorized to use it.

Importantly, this is an area where the laws governing each jurisdiction may heavily influence the role that data analytics and big data may play. For instance, in the Canadian environment, the use of data must comply with various privacy legislative requirements. These include the Personal Information Protection and Electronic Documents Act (PIPEDA Act) – or similar legislation in provinces where an exemption order has been granted (Government of Canada, 2000). The PIPEDA Act outlines the core principles for the collection and use of data as summarized in Table One.

Notice	Users must be informed when and what information is being collected, how it will be used, how long it will be retained and with whom it will be shared.
Choice	Users must be given a choice whether or not their personal data is collected.
Access	Users must be able to access, edit and confirm the validity of their personal information upon request.
Security	Users' data must be safeguarded from unauthorized access.
Scope	Only the required user information should be collected.
Purpose	Data may only be used for the purpose it was collected.
Limitations	User data should not be held indefinitely.
Accountability	Organizations must ensure privacy policies are followed and are liable if breached.

Table Two: Core Principles of data collection as outlined in the PIPEDA Act

Adapted from: Bradshaw, Harris & Zeifman (2013) Big data, big responsibilities: Recommendations to the Privacy Commissioner. http://www.cigionline.org/sites/default/files/no8_0.pdf

Within the Canadian gambling environment, some regulators have indicated that their use of potentially identifiable data (e.g., player account data) to inform RG regulation is deliberately absent at this point due to concerns regarding the maintenance and compliance with privacy legislation. Further, their concerns extend to issues of liability as well. For instance, if a regulator is able to identify, from the use of loyalty data, that an individual may be displaying signs that are suggestive of disordered gambling of some sort, what is the appropriate duty of care within this context? Some regulators prefer to not have this level of insight available to them about gambler's behaviour until questions of this nature are understood to a greater level. Others acknowledge that if this type of information is collected, it must be held in an anonymous format to avoid such concerns.

In addition, some regulators speak to the fact that if information (such as loyalty information) is to be used by regulators for RG/consumer protection purposes than this needs to be readily disclosed to gamblers in order to comply with regulation. For example, the Nova Scotia Play Now system experienced a challenge in this area when designing their (now discontinued) system. The consent pages which gamblers had to agree to in order to obtain a player card were extensive, with multiple pages of text presented to the gambler before they could provide their consent. Feedback suggested that this was a barrier to system uptake and acceptance. This barrier was made more severe when gamblers expressed their concerns about how the information would be used. Many cited concerns about possible collusion between government departments to identify gamblers who had outstanding tax bills, child support payments etc. This is despite multiple attempts to educate the gambling population that the system developed an anonymous identification number for each registered player and that consequently it was not possible for the system data to be used in this manner. This feedback and reaction clearly demonstrates the distrust for this type of information gathering.

Data access

As explored previously, there is a copious amount of transactional data recorded on a daily basis across EGMs globally however, the usefulness of this data collected in terms of RG is not yet apparent. Furthermore, depending upon the regulatory environment and infrastructure design, it may not be possible to effectively access this data in a reliable, meaningful fashion.

Even in systems which are designed to provide more in-depth information concerning gambling behaviour, access may be a challenge. For example, one jurisdiction reported that they were unable to access their player tracking data freely despite having an arrangement with a specified contractor for this service. These barriers were due to system barriers, restrictions and requirements.

In a broader sense, creating a comprehensive and current legislative and regulatory environment to facilitate data access and sharing in order to make big data and analytics a valuable pursuit is of paramount importance. This is relevant both in the broader context (at a national level) and at a micro level (between the operator and regulator). The consultation phase of the project highlighted the importance of relationships between operators and regulators. The development of a trusting relationship between both stakeholders was emphasized in both the literature and in consultation as being critical to successful information sharing and collaboration (KPMG, 2009).

Resources

The access to appropriate resources and talent available to support the use of big data and analytics is also a significant barrier in many circumstances (WLA, 2013). Not only do people who are actually working with the big data need to have specialized skills and understanding in this complex area, but those using the resulting output need to have insight about the meaning of the outcomes and how best to make use of the information (Agrawal et al., 2011; WLA, 2013). Owners of data need to be able to explain how the conclusions derived through big data analysis came to be. This 'data provenance' information should be able to convey to interested stakeholders, the origin of all data and should be used to help substantiate how the findings or conclusions drawn through the conduct of the analysis have been made (Agrawal et al., 2011). It is essential that users of the output from big data analysis understand and critically examine the data and data analysis processes in order to ensure the proper use of all findings (Agrawal et al., 2011).

Research Questions

What ‘types’ of data are collected by other stakeholders and how are these being monitored and analyzed by regulators and private operators for compliance and responsible gambling purposes?

Key Findings:

- Data collection, analysis, and usage is often dependent upon the nature of the infrastructure available to support the gambling sector (e.g., access to data may be facilitated or inhibited by the presence of a central monitoring system vs. decentralized model).
- Differences in philosophical and regulatory approaches to gambling (e.g., RG regulation mandated by legislation versus a voluntary code of conduct) across jurisdictions impacts upon the desire, perceived need, quality, availability, and usefulness of collecting data to inform compliance with regulation.
- The use of data from a variety of sources should be embraced and utilized in conjunction with gambling data to inform policy and regulatory decisions.
- There are some examples where regulators are embracing data to inform policy and regulatory development; however these are not yet the norm and appear to be heavily dependent on resource availability.
- There are examples where operators are taking the lead in the provision of responsible gambling initiatives that are not required by their regulator. This includes the collection and analysis of data to drive the development of future RG policy and approaches.

Data collection and big data is frequently discussed within the gambling context however, the results from the REA and consultation indicated that gambling operators may be using data/ data analytic tools and software to make advancements in marketing and commercial decisions as opposed to regulators. Consequently, the data collected is often used for marketing and business development as opposed to consumer protection and responsible gambling purposes.

As identified in the methodology, content to inform this section of the report was extrapolated from a variety of information sources including research papers, conference proceedings, PowerPoint presentations and consultation with experts. This consultation phase of the project proved to be essential for developing a detailed understanding of the variety of approaches, stages of readiness, and the desire on behalf of gambling operators and regulators to embrace data. Stakeholders consulted as part of the project assisted the authors to identify the current areas of interest for both operators and regulators and the format in which this data is currently housed. In order to protect the privacy of the participants, the specific jurisdictions have not been identified. They have been grouped geographically and broken down into National (within Canada) and international jurisdictions to account for any significant differences in approaches. Any relevant information sources have however, been cited in the reference list.

Types of data collected

Building on the findings from the NatCen EGM data research report (Wardle et al., 2013), the jurisdictions consulted as part of the project reported the collection of a wide variety of data elements. However, despite the sometimes large volumes of data being collected, few jurisdictions could articulate how this information was used to inform RG or consumer protection regulation development or compliance.

Importantly, stakeholders identified that in addition to the quantitative data collected regarding EGM activity etc. they also collected a wide variety of other 'types' of data. This more holistic dataset was viewed as having relevance to their broader RG platforms. Table 3 provides a summary of the different types of data collected by stakeholders consulted for this report.

<p>Quantitative data</p>	<ul style="list-style-type: none"> • Transactional data <ul style="list-style-type: none"> – turnover (T/O), – gross profit (G/P), – tax, – ticket purchases, – client account: <ul style="list-style-type: none"> ○ debits ○ credits ○ withdrawals ○ deposits ○ balance – date/time – product usage: <ul style="list-style-type: none"> ○ % of time unused ○ time between sessions ○ session lengths – bets: <ul style="list-style-type: none"> ○ bet type ○ bet amount ○ outcome ○ dividend (return) • Cash outs • Data from skill testing • Inventory management (what games are selling well and are popular) • Compliance testing (e.g., EGM) • Field data (e.g., field audits to ascertain compliance with RG regulations – such as the location of EGMs etc.) • Regulatory assessment (e.g., number and size of advertisements on Internet gambling websites)
<p>Qualitative data</p>	<ul style="list-style-type: none"> • Customer feedback data (e.g., letters of complaint, customer service contacts (online and offline), information shared on social media) • Customer interaction with staff • Promotions (special offers, discounts) • Advertising • Health and welfare data (Socioeconomic data, household expenditure data)

Table Three: Data collected by stakeholders consulted by an international sample of gambling operators and/or regulators

Review of national and international jurisdictions

Variances across jurisdictions in terms of the data collected and utilized by regulators can be explained by a number of factors. This may include differences in cultural norms regarding the role of government in regulation (e.g., how much regulation is acceptable), how gambling activity is perceived/framed by the jurisdiction (e.g., is it considered a legal form of entertainment or a risky behaviour?) as these factors contribute to how gambling policy is framed and established and thus, how regulatory frameworks are developed (Korn, Gibbins & Azmier, 2003). Further, the infrastructure implemented to support gambling activity also exerts a significant impact upon the type of data collected and how it can ultimately be used.

Canadian jurisdictions

The Canadian gambling environment is regulated at a Provincial level. The role of the Federal Government in gambling activity is limited to the legal conditions outlined in the Criminal Code of Canada. Historically gambling is illegal under the Code, however amendments made in 1969 sanctioned some specific gambling activities. This was followed by further amendments in 1985 that permitted further gambling activities and transferred their regulation and administration to the provinces and territories (Pruden, 2002).

Each provincial government has subsequently established a body to regulate the gambling activity. This regulatory framework has resulted in each jurisdiction developing a unique system with significant differences existing between them. These systems permit and support different games, lottery schemes and administrative structures.

The majority of Canadian jurisdictions that were engaged as part of this project did not report the meaningful use of data in terms of informing RG or consumer protection regulation at this time. The exception appeared to be that jurisdictions reported access to data regarding self-exclusion which was used to contribute to and inform the development of RG interventions and regulation.

The key barriers identified to the active use of data included concerns regarding privacy and perceived liability issues, lack of access to data, unavailability of meaningful data, lack of a plan for how data could be used, and lack of specialist resources that would facilitate its meaningful use. Examples were found of how structural and policy changes had resulted in situations where access to meaningful data which may have been used to inform regulation is no longer possible due to policy decisions.

Conversely, operators that participated in the consultation expressed both an interest in, and some initial progress regarding, the use of data to inform responsible gambling activity. This pattern was also reflected in some of the international consultations undertaken.

Both operators and regulators acknowledged the limited use of straight transactional data from EGMs for reasons outlined in the NatCen EGM gambling report as without context, this information offers little insight (Wardle et al., 2013). It was reported that greater insight would be available from systems that could report individual level play data (player behaviour) although this was not available in any systematic form except through the use of loyalty data. This specific type of player tracking data was understood to be owned by operators but accessible by regulators if they chose to request it.

In one jurisdiction, player level data had been available through a system which collected player behaviour data on the gambling floor in the casino environment. Specifically, a proprietary software package was developed by a private consulting company and implemented on the gambling floor. This computerized system was connected to the EGM transactional system and analyzed player behaviour to formulate a risk profile for individual gamblers (low, medium and high risk) based on their gambling activity (Davies, 2007; Delfabbro et al., 2012). The software package considered multiple variables to analyze behaviour and included integration of multiple data sources such as customer profile data and other broader datasets (e.g., postcode). The theory behind the system was aligned with research that demonstrates the fluidity of problematic gambling behaviour. That is, many gamblers move between different levels of risk depending upon multiple factors that may be exerting influence on them at a particular point in time (Currie, Hodgins, Wang, el-Guebaly, Wynne et al., 2006; Hodgins, Wynne, Makarchuk, 1999). The purpose of the system was therefore to assist in the identification of potentially harmful gambling behaviours at an earlier point, before they escalated into a gambling problem. Further, the authors of the system indicated that data from such a system could potentially help inform policy in a more meaningful way than traditional data sources as reliance on research which focuses on those with an identified gambling problem may skew the data and hence not provide a true picture of the nature of gambling problems, interventions and outcomes.

Of vital importance to the system were the practices that followed by flagging of potentially harmful behaviour; interaction with the gambler and documentation of that interaction. This human interaction was essential due to the multiple limitations of the system in terms of the quality of the data, the exclusion of gambling behaviour data from activity that occurred in other venues and the potential lack of understanding of the particular characteristics of each gambler (e.g., level of expendable income) and thus the context of play.

The strength of this system, which is no longer in operation, is the provision of real time data to inform RG strategies. Despite a lack of published data evaluating the accuracy of the system, it provides an example of how player behaviour analysis can be applied in the land based gambling environment, as opposed to being limited to the online environment. Further, a system such as this, if used reliably may also be able to identify suspicious gambling behaviour such as money laundering or possible fraudulent activity given the tracking capability.

Similarly, another jurisdiction recently discontinued a system which could have provided meaningful player behaviour data resulting from the mandated use of a player card. The system was implemented in a manner which allowed gambling to occur without registration of a single account. This resulted in significant limitations regarding the potential usefulness of the data collected by the system. Had the system been implemented as originally intended, significant amounts of meaningful data would have been available for analysis at both an individual and aggregate level. In the development phase it had been hoped that this information could help identify patterns of play that could lead to problem gambling. An example of this in practice is assessment of the impact of various game features. For instance, if large payouts of \$500 had resulted in increased, continued play vs. decisions by the gambler to cash-out then these findings could have been used to make policy decisions in accordance with these findings. Further, if specific features of games were resulting in potentially harmful behaviour (e.g., chasing losses), this data would have enabled the identification of these patterns and enabled strategies to be put in place to address these in future game design and regulation.

International jurisdictions

Consultations from international jurisdictions resulted in a wide variety of approaches and infrastructure being explored.

European jurisdictions

Across Europe, fundamental differences exist in terms of gambling regulation and the public support for government control of information and regulation. This has resulted in significantly different systems being developed and the varying use of data. Interestingly, it would appear that even in the jurisdictions where extremely restrictive regulations regarding gambling exist (primarily in the context of EGMs) there are few examples where real-time data is being used to inform regulation.

The nature of regulation

One jurisdiction is in the midst of consultation to review the social responsibility provisions in the licence conditions and codes of practice. Within the proposed changes are several issues relating to the type of the data collected. For instance, it has proposed that in addition to the data supplied by operators, each operator should supply an assessment of the amount of money that they receive from problem gamblers. This would be submitted with an action plan outlining how these figures will be addressed by the operator. It is argued that this type of information would help operators focus their attention and resources on specific areas that warrant further RG intervention.

Further, the approach to regulation may be changing from 'blanket controls' to those that are more targeted. This would be accomplished through an increased understanding of player protection, and strategies to reduce risky or illegal behaviour by gaining new and deepening insights into how to reduce the possible harms that result from gambling. This will likely be accomplished by research, data and analytics to support understanding.

Player cards

In one European jurisdiction, mandatory player cards have been introduced across all business lines. The player card aims to offer gamblers access to technology that may assist them to control their gambling. A variety of pre-commitment tools are integrated within the card and some are mandatory for gamblers to use. Pre-commitment refers to the ability of a gambler to set pre-established boundaries regarding their gambling behaviour in advance of play. For instance, in this jurisdiction gamblers are able to set limits on their time/money, conduct self-assessments to determine risk for problem gambling behaviours, set weekly budgets as well as informative tools to explain how the games work.

In addition the operator has adopted a tool which uses predictive analytics to assess and identify risky gambling behaviours based on a gambler's actions. It is available to customers free of charge and provides gamblers with feedback regarding their behaviour.

This combination of tools facilitates the collection of a wide variety of customer information to be retained, studied, and analyzed in the live gambling environment. This arrangement, where player cards are mandatory for the operators of gambling products, provides the most complete and comprehensive data available in both land based and Internet gambling. Interestingly, in this jurisdiction this initiative was not a regulatory requirement, rather it was the initiative of the monopoly operator who saw several benefits to this action including, separating itself from grey or black market competitors by providing gambler's with a more comprehensive RG experience and providing socially responsible games.

In another example, one jurisdiction, following concerns regarding problem gambling behaviour in relation to EGMs, removed all games from circulation. Modified EGMs were subsequently introduced

in 2009 with strict regulatory requirements governing their use and the requirement for the mandated use of a player card. The monopoly operator now requires that all play be registered, meaning that an individual is required to register a single player card which is required for accessing the games. This, in combination with the developed infrastructure, allows the focus to be on the player's behaviour and consumption patterns rather than on the games per se (Stromslid, 2013).

The nature of the system provides large amounts of data which could be used to gain insight into player behaviour. Examples of the data collected include: demographic data (e.g., age, gender, address, phone number), transaction data which can be broken down by date (day/month/year), medium and game (e.g., games played, how many tickets, cost, time, location), and financial transactions (e.g., amount won/lost, how much, when). It is acknowledged by the operator that the purpose of this data collection is to inform RG activities and interventions however, there is no data currently available to describe or evaluate how this information is being applied within the gambling context. Further, no regulatory requirements stipulate that specific information be shared from the operator with the regulator; however this occurs in a voluntary capacity at regular, agreed intervals.

In these jurisdictions, the operator's position as a monopoly across various gambling products makes this system more comprehensive than would be possible to achieve in a regulated environment with multiple competing operators across and within gambling products. Furthermore, consumer understanding and attitudes towards the collection and use of personal data plays an important role on uptake and usage of such a system. The experiences of these European jurisdictions can be contrasted with the player tracking system introduced in a North American jurisdiction, which faced considerable player opposition and lack of support resulting in an eventual withdrawal.

Asia Pacific

Similar to other geographic regions, regulation and approaches to data collection vary significantly within the Asia Pacific region. For instance, one jurisdiction has specific requirements in terms of reporting data to its regulator. Specifically, every casino is required to have Host Responsibility Program which outlines the responsible gambling approach that will be taken by the organization. An example reporting guide extracted from one of these programs is included as Appendix A.

The data collected and submitted by the casino Host Responsibility Programs extends beyond the typical transaction based data to incorporate a number of problem gambling indicators and is based on proactive policies where staff attempt to identify and intervene with people who may be displaying problem gambling behaviours. This includes making use of their player loyalty data to ascertain gambler's patterns of play to identify the possible development of a gambling problem (similar to efforts cited in the Canadian context).

The data that is required to be provided within this jurisdiction is extensive and structured as outlined below.

Gambling Related Measures
Number of customers about whom there have been observations
Number of indicators reported to Host Responsibility
Number of approaches to casino by third parties (<i>reporting gambling problems</i>)
Number of problem gamblers identified (in the first instance) by requests for exclusion or forthright disclosure, compared to number of problem gamblers identified by the casino.
Number of customers on incident spreadsheet: Number of GOI files by:
– Ethnicity
– Gender
– Age
– Preferred mode of gambling (tables/EGMs).

Number of approaches to customers to offer information about self-exclusion
Number of exclusions by: <ul style="list-style-type: none"> • Ethnicity • Gender • Age • Preferred mode of gambling • Prompted by third party disclosures • Exclusion type • Following re-entry.
Number of customers participating in Multi-site Exclusions.
Number of customers participating in Agreed Limitation programme
Number of Excluded customers agreeing to be contacted by help services on exclusion form.
Number of breaches of exclusion by: <ul style="list-style-type: none"> • Ethnicity • Gender • Age
Number of successful and unsuccessful applications to re-enter following exclusion.
The extent to which customers have been assisted (drawing, inter alia, on feedback from customers and staff).
Number of persons trespassed or required to leave for making loans for financial gain.
Measures related to staff training
<ul style="list-style-type: none"> • HR1 courses • HR2 courses • HR3 courses • Refresher training • Number of staff who need to be trained in each category, and proportion of those staff that have completed the appropriate level training.
Staff recall of knowledge and behaviours related to Host Responsibility and associated policies and procedures.
Staff perceptions of the effectiveness of training.
Other Programme activity and compliance-related measures
Number of internal and external underage incidents.
Number of unattended children.
Number of Requests to Leave the Premises

In the Australian context, gambling regulation and enforcement is the responsibility of each state. As a consequence, several different systems exist to monitor and regulate the environment. This includes significantly different infrastructure to support them.

A central monitoring system in one jurisdiction consulted as part of the project facilitates the collection of data from each of the 40,000 EGMs which are used within the community setting (e.g., outside of casinos). In this particular setting there are multiple providers of EGMs however, the centralized monitoring system facilitates the establishment and formalization of data requirements. This particular jurisdiction collects and utilizes data from the following sources to inform RG compliance with regulation:

- EGM information in terms of turnover and game popularity
- financial data of gambling operators including EGM, keno, lotteries, and wagering (this includes turnover and where the spend may be coming from)
- compliance audits which are conducted in order to demonstrate that the EGM is functioning as it should, in line with any regulatory mandates.

The data are then used:

- to monitor trends around RG compliance to identify the need for further education and support to encourage compliance with a voluntary RG code
- to monitor specific gambling (transactions) to see if it is resulting in problematic gambling behaviours (e.g., to see if winning a jackpot results in riskier gambling behaviour).
- to monitor and track player trends over time.

This jurisdiction provides an excellent example of the use of various data sources to inform gambling policy and regulation. For example, they reported that data sources used by the regulator (in addition to gambling transactional data) included socio-economic survey data, household income data, population growth figures, as well as data resulting from problem gambling prevalence surveys. This type of data is then overlaid on gambling specific data to provide insight into decision making. An example of this in practice is decision making in relation to the appropriate distribution of EGMs within the community. This comprehensive approach to analysis allows the identification of specific community characteristics, benefits, and potential risk factors that contribute to an informed assessment of the overall impact that additional sources of gambling may have within the local community.

Importantly, this jurisdiction acknowledged the need for resources to support these interventions. It has a dedicated team to assist with analysis of this type.

Other jurisdictions report significant data collection processes yet little comprehensive integration of the data into the formulation or enforcement of RG regulation.

Internet gambling

Jurisdictions utilize different approaches to data collection in terms of regulatory requirements in the online environment. Some jurisdictions have strict regulations outlining what sites are appropriate for their population to access whereas others allow and acknowledge that their gamblers are accessing multiple sites.

In terms of the data collected and provided to regulators, consultation revealed that some jurisdictions have access to a copy of online gambling activity data. This allows the regulator to complete their own assessments and reviews of the data. In other jurisdictions, online operators do not provide their regulators this type of access however, operators do provide regular responses to on-demand requests for information in a timely manner. These requests can be for a variety of data which meet the regulators needs.

Where online gambling operators discussed the use of player behaviour analytics, no stakeholders indicated that they provide detailed (potentially identifiable) information to the regulator. Rather, regulators indicated that often with this type of technology (whether mandated by regulation or not) the data resulting from its use is applied in a limited form for RG purposes (e.g., to identify gamblers who are demonstrating potentially risky behaviour) and most often for marketing and promotion information. Regulators in these environments saw their job as one of encouraging the operators to use the available tools and systems to encourage operators to embrace the full possibility of this type of data, including the implementation and use of tools to identify gamblers displaying 'risky gambling behaviours'.

What evidence exists to ensure that these processes (data collection, analysis and monitoring) are having their desired effect?

Key Findings

- There is little evidence available that indicates the data collection, analysis, and monitoring is achieving the desired effect in terms of informing regulatory compliance.
- In order for evidence of effectiveness of any processes to emerge, operators and regulators must work collaboratively to identify the optimal data to be collected.

The evidence obtained in both the REA and stakeholder consultation indicated that there were few examples of data being collected and regularly analyzed specifically for RG/consumer protection purposes. However, the majority of regulators were collecting various data elements to ensure compliance with regulatory standards more broadly. There was also evidence to suggest that work was being done regarding the quality of the data being collected.

Assessing the quality of data collection

In a keynote speech Philip Graf (2013), Chairman of the UK Gambling Commission, it was indicated that the UK had moved away from routine inspections, to focus on individual reviews including a risk based assessment of the data they collect from industry. This model is supported by the need for operators to alert the regulator to possible risks and issues in the RG field in the same manner as they would for other compliance issues.

He noted that in order for best practices to develop, operators and regulators needed to work in partnership in order to collect data and information to progress this agenda. Further, Graf acknowledged that the industry was better positioned to develop new innovations to assist with this using their data and technology to support interventions. This was further elaborated on in a keynote speech given by Graf in 2014 at the ICE World Regulatory Briefing in which he highlighted operators' extensive use of data, tools, and analytics to grow their business. Graf and stakeholders consulted through the project highlighted the opportunities that exist for these measures, tools, and knowledge to be applied to the RG and consumer protection environment to further strategies to address the risks of developing a gambling problem and/or suspicious behaviours such as money laundering. Graf acknowledged that gains were being made in this sector with operators using data and collaborating with researchers however, he acknowledged that more action is required in order to further this agenda and increase transparency in the gambling sector.

How has/is data being utilized across industries outside and including gambling to ensure that regulatory standards are being met?

Key Findings

- Multiple sectors have identified the potential for big data to revolutionize their operations and are using big data to support a range of interventions.
- Few sectors can demonstrate comprehensive integration of this information and approach into their regular business and operations.
- There is evidence to suggest that internationally some data is being used to inform regulatory standards in the finance industry.
- Challenges relating to privacy, adequate infrastructure (analytic capabilities), and data sharing are inhibiting complete use of data in some sectors.

Marketing and Communications

The marketing and communications sectors are leaders in the use of big data to inform business operations. Currently, businesses collect and hold vast amounts of data and this amount of information is expected to double every 1.2 years (Manyika et al., 2012).

Customer segmentation

Data within the marketing sector can be used to gain an in-depth understanding of the customer. This information can be used to tailor communications, advertisements, and products to meet the needs and habits of a particular player base (WLA, 2013).

In order to develop this, organizations have a wide variety of data sources that they can access where consumers, often unaware, have provided significant amounts of data to industry. By using a variety of information sources a 'customer profile' based on an individual's behaviour can be developed. In the Internet environment, this is facilitated through the use of 'cookies' that are left during Internet browsing (Office of the Privacy Commissioner of Canada, 2011). Cookies can be applied and used in different ways, including creating a record of the individual's activity including their searches, browses, purchase history etc. (ADMA, 2013). This information can then be utilized to build a customer profile which can inform targeted advertising and promotion strategies (Office of the Privacy Commissioner of Canada, 2011). When an individual with a profile next arrives on a site, their profile can be identified and an advertisement that is specifically targeted to suit their perceived preferences is automatically identified. This reduces waste in advertising (e.g., marketers purchasing advertising copy and targeting a broad customer base) (ADMA, 2013; Office of the Privacy Commissioner of Canada, 2011).

The process of understanding different 'types' of customers is referred to as customer segmentation. This understanding of a customer and their habits, preferences and behaviour allows a company to communicate with them in highly tailored manners. This information is used and available within the gambling context. Although one would expect that the majority of this work is proprietary in nature, there is some academic evidence that has been published. A recent example of this evidence base was conducted by Chen, Shoemaker & Zemke (2013) and examined the segmentation of EGM players into four clusters including:

- (1) utilitarian gambling seekers;

- (2) excitement gambling seekers;
- (3) multipurpose gambling seekers; and
- (4) relaxation gambling seekers (Chen, Shoemaker & Zemke, 2013, pg. 44).

The information can be used by marketers to specifically design and target new EGM products and games to attract and entice gamblers to their products.

Loyalty programs

Similar strategies are used within the gambling context to target and tailor advertising messages to meet the needs of the gambling operators. Customer profiles can be generated in-house through the use of extensive loyalty programs and supplemented by information from other sources, including aggregated customer data.

Some loyalty programs have evolved to incorporate information obtained through big data sources. This would involve the use of multiple datasets with customer information to further understand and predict the customer's behaviour, preferences and habits. Customer data is gathered and analyzed and used to inform business development initiatives. They have also been able to advance the use of the systems to more specifically target customer's preferences. For instance, cross-promotion of products based on a customer's shopping behavior, targeted promotion of specific products as well as integration with smartphones to allow awareness of specials/adverts etc. (ADMA, 2013).

This particular feature has significant implications for the gambling sector where loyalty programs are commonly available. Gambling operators can use the information gleaned through the loyalty system, for instance the types of games a customer likes to play and frequency of play to strategically target individuals with adverts, promotional offers and targeted messages (WLA, 2013). Gambling operators can also use loyalty programs across multiple business lines, including food and beverage sales and offer venue-based rewards and offers, such as complimentary parking, based on status.

Advanced analytics – data mining and predictive analytics

In addition to the extensive use of data, some gambling operators are able to take their understanding of the market and their customers to new levels. For instance, a large horse racing organization in Asia was able to harness their significant investment in big data analytic technology to mine their customer data and identify that a significant proportion of high-value customers were not participating in races held mid-week. Analyses identified that this was most likely due to work related commitments overseas. As a consequence the company was able to build systems that supported the engagement of this segment of clientele with the races despite their geographical distance. This clearly demonstrates the opportunities that big data has not only for existing clients, but for business development initiatives.

The use of predictive models in this context is useful when identifying customers who may be dis-engaging with a product or service. If a deviation from expected behaviour, as determined by predictive analytics, is noted (e.g., if a gambler does not purchase their lottery tickets at the expected pattern and regularity) then personalized strategies can be deployed to increase engagement with the customer. This targeted marketing essentially provides the right consumer with the right information at the right time (WLA, 2013).

Further, there are examples of how predictive analytics have been used to predict what is happening in the lives of customers based on analysis of purchase habits and the use of appropriate algorithms. For instance, Target in the United States developed an algorithm which facilitated the company predicting whether particular customers were pregnant based upon the products that they purchased and the timing of these purchases. The company was then able to specifically promote certain items to expectant mothers in line with their expected progress through the pregnancy and delivery date (Stoddart, 2013).

Implications for the gambling sector

- There is clear interest within the gambling sector to harness and embrace the use of data analytics to develop an increased understanding of the customer base.
- There is extensive use of loyalty cards and programs in the gambling sector reported internationally. The data collected via these systems is a possible source of primary data.
- There is the potential to harness some of the customer segmentation data to understand the motivations of gamblers in greater detail and thus execute RG initiatives in a more targeted manner.
- Caution is needed when targeting gamblers for example, with promotional offers or rewarding ongoing play, to avoid encouraging excessive gambling. This is particularly important for customers who may be at-risk for experiencing gambling problems or for operators that do not actively identify risky gambling.

Financial regulation

There is significant use of data to inform the enforcement of regulatory standards in financial markets. Although the Canadian markets performed comparatively well during the Global Financial Crisis (GFC), significant international reform has taken place globally since that time (Arjani & Graydon, 2013). Although, an extensive overview of the financial markets is beyond the scope of this project, the use of data to inform compliance with consumer protection regulation will be addressed.

A significant focus of new regulation in the financial industry is devoted to the identification of and management of risk and fraudulent or illegal activities (Davies, 2013; Levi, 2009). The importance of improved regulation and use of data became apparent following the GFC which illustrated that the data available at the time was not sufficient to adequately identify and mitigate risk. It has therefore become a significant issue on the international stage, particularly since the introduction of key pieces of regulation including the Dodd-Frank Act (United States) and the European Market Infrastructure Regulation (EMIR) and Solvency II Regulations (Davies, 2013) designed to address this.

Mandated regulation such as Know Your Customer (KYC), require banks and insurance organizations to understand who their customers are. KYC is defined by the Financial Transactions and Reports Analysis Centre of Canada (FINTRAC) as *“Due diligence measures to identify and verify customers, as well as ongoing monitoring of customer transactions (for example, checking that your customer’s activities are in line with the identity he/she has claimed).”* (FINTRAC, 2014). These requirements have been extended to include requirements to disclose business entity data which outlines the legal structure of a company in the United States thus requiring financial entities to compile a great deal more information regarding their customer base (Davies, 2013).

The role of data analytics in the financial sector also incorporates an important element of fraud detection and prevention. Complex algorithms are utilized to help identify fraudulent behaviour across a variety of technologies (online, smartphones etc.). Tools are also employed to identify real-time risky behaviours and to identify possible fraudulent activities (Bierig et al., 2013).

FINTRAC is an independent government agency that was developed to help detect, prevent and deter illegal financial activity such as money laundering and/or the financing of terrorist activities (FINTRAC, 2013). Its activities demonstrate the use of data analytics in the regulatory environment. The organization receives reports from 'reporting entities' (including casinos) that assist in the execution of its duties. It uses analytical technologies utilizing information from multiple sources to conduct detailed investigations into the allegations (FINTRAC, 2013). These sources include both domestic and international data sources from a variety of fields. Further, data is used to detect trends and patterns in suspicious activities.

Of significant interest in the financial sector is the financial impact that compliance with regulation is having within the sector. Multiple sources of information reported the demands and costs associated with compliance with increasing regulatory requirements was a key challenge for operators.

Despite the available data to inform regulation, there remain significant gaps in data availability and use of analysis to inform regulation in some areas. A report published in August 2014 by the Canadian Foundation for the Advancement of Investor Rights (FAIR Canada) indicated significant gaps in the collection and centralization of data concerning investment fraud. This dearth of data is an example of how regulators are at times unable to effectively use information to inform regulation despite the large amounts of data being compiled.

Transparency and data

Financial regulation of the securities markets in the United States is the responsibility of the US Securities and Exchange Commission. This includes consumer protection, and maintenance of a 'fair, orderly and efficient markets'. This large organization has an extensive remit and uses the services of the Division of Economic and Risk Analysis to execute its duties. This Division is specifically responsible for completing qualitative and quantitative analysis to provide data analytics to support policy and regulatory development and also to support the organisation to identify risks across multiple domains including litigation, examinations, reviews and provides data to support enforcement issues.

Following the financial crisis, the Obama administration responded by signing the Dodd-Frank Wall Street Reform and Consumer Protection Act into law on 21 July 2010. The legislation enforced significant changes to the regulatory environment including the area of consumer protection. The Consumer Financial Protection Bureau (CFPB) became operational in July 2011. The mandate of the organization is to provide a single point for accountability in terms of consumer financial protection. This is executed through a 'data driven approach' which seeks to use analysis and innovative technology to execute its mandate (CFPB, 2013).

This organization's use of data is extensive and with a focus on the education of the public, significant amounts of information are shared and publically available. The organisation highlights its use of complaint data as a source of information for the public making this freely available (<https://data.consumerfinance.gov/dataset/Consumer-Complaints/x94z-ydhh>) (<http://www.consumerfinance.gov/hmda/>) and receives access to extremely large repositories of data (aggregated and atomic) from a wide range of financial sources. For instance, a recent study examining Overdraft Programs included data from a sample of banks regarding their overdraft programs presented in aggregated form which was paired with detailed, transactional data from random samples of de-identified consumer accounts at these banks.

Key learnings applicable to the gambling context

- Significant regulatory requirements have compliance costs associated with them for the financial sector. Thus, new regulations in the gambling sector would benefit from a cost/benefit analysis to determine how the regulation will provide meaningful benefits prior to mandating its implementation. This will help to ensure that all regulatory action in relation to data collection and analysis is aligned with a desired outcome.
- Transparency of the data collected in an important element of regulation. Data collected should be disclosed to customers and/or be made available to researchers in order to maximize its value.

Healthcare Industry

The use of data within the healthcare system has specific and unique opportunities. The costs of healthcare have been continually increasing and there are significant challenges regarding the ongoing provision of service delivery therefore stakeholders are seeking opportunities for solutions that will allow good care to be accessible for less cost. It is a complex sector, where extensive amounts of data are collected by a variety of stakeholders and organizations. However, some commentators argue that the sector has failed to fully transition its viewpoint from one which views this data collection as a consequence of its work to one which views this extensive data as a strength and resource (Murdoch & Detsky, 2013). These large, but disparate data sources have the potential to achieve great understanding if the knowledge that they contain can be harnessed.

There is significant evidence that the healthcare sector has been working to use data and data reporting structures to demonstrate compliance with standards, identify areas for improvements and evaluate outcomes. This is evident in quality healthcare system initiatives such as the Ontario Wait Times Strategy which utilizes regular reporting data through the Wait Time Information System (WTIS) to provide information to the public about wait times for a variety of procedures (e.g., cancer surgery, hip and knee replacement, Emergency Room (ER) wait times). The data used to provide these statistics is gathered and submitted monthly and subsequently published on the Ontario Ministry of Health and Long-Term Care website. Similarly, within the cancer context, a number of quality indicators are examined and reported on provincially by the Cancer Quality Council of Ontario. This data helps to ensure compliance with key directives, goals and regulation. These data are however, reporting on figures from past years. This highlights a key challenge across this sector – although data collection and use have been integrated into decision making systems, there remains a time lag in sourcing and using the data efficiently.

The literature identifies several ways in which big data could contribute to improvements in the healthcare sector. These include improvements in patient outcomes, enhancements to the patient experience, and the reduction of high healthcare costs (Amarasingham et al., 2014) as well as the transfer or dissemination of healthcare best practices (Murdoch & Detsky, 2013). Further, there are also opportunities for data to assist in the reduction and prevention of healthcare fraud (ACL Services Ltd, 2010; Roski et al., 2014).

Data sources in healthcare

Electronic Health Records (EHRs) represent a key opportunity for the effective use of information and data in the clinical context. EHRs facilitate the comprehensive collection of information regarding a patient in a single, electronic record. This information may include data from primary care physicians, specialists, allied health care team members as well as diagnostic results (e.g., x-rays, blood work etc.). The key defining factor about an intra-operable EHR (where the information can be read across multiple systems) is the possibility for this information to be accessed across the healthcare sector thus facilitating comprehensive care without the risks that important clinical information is being

housed in silos; preventing it from being used to make the best healthcare decisions possible. These records are becoming more ubiquitous, particularly in the United States and facilitate the collection of a high volume of clinical data in multiple forms (e.g., open text fields, data, imaging) (Bates, Saria, Ohno-Machado, Shah & Escobar, 2014). Murdoch & Detsky (2013) argue that the EHR data may constitute an “*observational evidence base*” which can be analyzed to answer complex clinical questions and because of the number of different elements of data and numbers of patients that can possibly be collected within an EHR data collection, may help address issues such as generalizability which often limit the interpretation and application of findings from other forms of scientific research (Murdoch & Detsky, 2013, pg. 1351).

An EHR can be thought of as a far more complex and detailed loyalty player card in the gambling context. Every time a patient presents to the healthcare system, the EHR captures all of the relevant information about the visit (e.g., presenting issue, diagnostic tests, results, treatment, outcomes). This is similar to how all transactional information is captured in the gambling sector. If this information is available for broad analysis, greater understanding of complex medical and social questions is possible.

Further, and similar to the gambling context, the health sector needs to seek additional sources of information in order to most effectively use the data housed in the EHR. Although specific analysis would provide a level of insight, data from other disciplines such as sociology, welfare, psychology, and finance has the power to really change the way in which information is understood (Roski et al. 2014). For instance, cell phone and data sourced from social media can provide additional information in the form of text based analysis, providing far greater understanding of healthcare outcomes, challenges and illness (ADMA, 2013; Bates et al., 2014). By analyzing data in a combined fashion, the level of insight is much greater than when each dataset is understood individually. It is this comprehensive picture of a problem which will result in the most significant gains.

Fraud detection in healthcare

The breadth and complexity of the healthcare system provides opportunity for individuals to engage in fraudulent activities. This can occur at both the provider and patient level. Data has and can be employed to identify possible fraudulent behaviours quickly and efficiently by setting appropriate parameters and investigating the data effectively (ACL Services Ltd, 2010; Roski et al., 2014). Specifically, data can be used to identify the following behaviours:

- Identify errors in billing or payment for staff overtime or on-call wages (abuses by professionals).
- Find kickbacks paid in exchange for referring business.
- Highlight “up-coding” of procedures: Statistically outlying numbers.
- Match vendor names/addresses/tax IDs to payroll records for employees (to identify duplicates).
- Summarize large invoices without purchase orders, by amount, vendor, etc.
- Compare list of valid signed-up employees to list of people actually receiving health benefits from an insurance company.
- Highlight billing for medically unnecessary tests.
- Identify false/invalid/duplicate Social Security numbers.
- Highlight excessive use of high risk DRGs (“Diagnosis-Related Groups”).
- Identify excessive billing by a single physician.
- Report entries against authorization records for new or terminated employees.
- Identify multiple payroll deposits to the same bank account.

Source: ACL Services Ltd, (2010). Discussion Paper: Fraud Detection Using Data Analytics in the Healthcare Industry. (pg. 6). Retrieved from:

http://www.acl.com/wpcontent/uploads/DP_Fraud_detection_HEALTHCARE.pdf

Predictive analytics in healthcare

The use of predictive analytics in the healthcare setting may inform the potential use of the technology in the gambling space. Although not used widely at this point, there have been case studies which have demonstrated its value within the healthcare setting. The predictive systems themselves are used in a manner that examines risks and the likelihood of them occurring to develop models that predict outcomes (Bates et al., 2014). In healthcare this can help to identify people at high risk for adverse outcomes and implement strategies or interventions to help minimize these and/or on a broader picture can help to identify potentially large risks at a population level (e.g., communicable disease outbreaks) (ADMA, 2013; Bates et al., 2014).

The ability to identify high risk patients or those at risk for greater complications requires careful consideration. Mirroring the challenges faced when identifying possible high risk gambling behaviours in a subset of gamblers, Bates et al. (2014) argue that it is important to know what interventions to provide the person in order to assist them in the most effective manner prior to their identification. There have been demonstration projects which successfully identify high-risk individuals however; this did not translate into improvements following the interventions. This clearly demonstrates the need for effective and evidence based interventions. Further, challenges arise in the development of effective algorithms which can be used to identify those at risk. Often these are developed by using data from low-risk groups – risks then emerge relating to the quality of the algorithm (Roski et al, 2014). There is a clear need for further evaluative research in order to determine the best practices in this area. In the gambling sector, this use of algorithms is employed when technology is used to examine and interpret the behaviour of gamblers using player tracking data.

The use of predictive analytics in healthcare is an example of how data can be used to help predict outbreaks of disease. This has been accomplished for illnesses like Dengue fever by examining sources of data like search history trends and social media data (e.g., tweets) (ADMA, 2013). It also has implications for resource utilization and staff planning. For example, a Patient Admissions Prediction Tool (PAPT) has been implemented across Queensland Australia and uses multiple sources of data to review and predict when resources will be required in the Emergency Room. This allows administrators to plan resources, scheduling of elective surgeries and in the introduction of innovative solutions (e.g., partnerships) to be created to help manage periods of time where increased activity is expected thus potentially saving significant resources (Australian Government, 2014; CSIRO, 2012).

Key learnings applicable to the gambling context

- The use of predictive analytics holds promise in the gambling sector in terms of the possible identification of gamblers exhibiting risky behaviour and/or to identify possible problematic patterns of play in a broader context using products and tools that use algorithms to analyze behaviour. The key step in terms of regulating this in a gambling context, is to have a clear plan of action to address the varying degrees of risk should they emerge.
- Similar to the gambling environment, the data collected within the healthcare sector provides rich insights into behaviour however, the use of this data in conjunction with other sources of information will provide greater understanding, richness, and problem solving capability.
- The use of data to examine for potentially risky transactions can be built into a compliance and monitoring program which captures risks and information about both customers and staff behaviour.
- The sharing of data between health care providers and with various government agencies and departments is a useful example of how multiple stakeholders and operators can collaborate through the use of a centralized or harmonized system to achieve shared aims and objectives.

How is behavioural data being used to inform and evolve regulation?

Key Findings

- There is little evidence that behavioural data is being used to inform and evolve regulation.
- There is significant progress being made to identify 'risky gambling' behaviours, particularly in Internet gambling.
- There is a lack of available evidence to support the application of predictive analytics to inform gambling regulation.
- Behavioural data excludes other important variables that are essential to effectively understand the context in which regulation would be developed.
- Analysis and effective use of the data is limited in many jurisdictions by the access to effective and appropriate analytical resources.

Behavioural information/data about gambling is obtained through the use of player tracking data. Little to no evidence was found to suggest that behavioural data is currently being used to inform and evolve regulation within the gambling field, or in similar areas. There is little evidence to suggest that the mandated limit was set based on empirical evidence, or behavioural data. However, it is also relatively difficult to accurately ascertain the factors behind any specific legislation due to the complexities of social, industry, and political influences on regulation. Consequentially, this report focused on how behavioural data could potentially inform and evolve regulation.

Identifying behavioural markers of risky gambling

One of the emerging ways in which behavioural data is being used to enhance the provision of RG and minimize harm is to use identifiable markers of potentially risky gambling. Interventions that effectively target gamblers before they reach a point of crisis may reduce harmful consequences of gambling. Regulation of early interventions (that aim to assist people before they seek formal assistance) is consistent with a public health approach that emphasizes the role of industry to provide a safer, less harmful gambling product and environment (Delfabbro, King, & Griffiths, 2012).

OLG currently has a system in place to identify risky and problematic gambling within venues and is working with providers (GTECH and BetBuddy) to implement a behaviour tracking system for online gambling. These systems utilize the available evidence with regards to identifiable markers of risky behaviour to develop algorithms which are then applied to the vast amounts of data collected in the Internet gambling setting. Hypothesized behavioural indicators of risky gambling may include total time spent gambling and the frequency of gambling sessions, monetary losses, ways and frequency of making deposits, escalation in bet size and frequency, the number of different gambling activities used, declined deposits and reversed withdrawals (Delfabbro, King, & Griffiths, 2012; Gainsbury, 2011; Ladouceur, Blaszczynski, & Lalande, 2012). Additionally, customer contacts with operator staff can provide indicators of gambling problems either directly through admission of a problem or impacts of gambling, or indirectly, for example through tone of voice and types of verbal discussions (Haefeli, Lischer, & Schwarz, 2011). Usually classification of gamblers is done through a combination of factors, including consideration of identified behavioural factors by a trained professional (Quilty, 2013). A summary of some of the indicators used to identify potentially problematic gambling behaviour has been included in Table Four.

Indicator type	Indicators	Type of gambling
Physiological	<ul style="list-style-type: none"> Sweaty palms/body Headaches Dry eyes Sick to stomach and nauseated 	EGMs
Frequency, duration and intensity of gambling	<ul style="list-style-type: none"> Gambles for > 120/180 mins per session Multiple episodes per week/hours spent gambling Plays 2 EGMs at one time Gambling almost uninterruptedly Requests that a specific machine be reserved Patron waiting for venue doors to open Difficult to stop patron at closing time Spends ¾ of time at gambling venue Gambles through normal meal times Tries to win on a particular machine Fall asleep at machine/table Chose bets with riskier odds Gambled very quickly or rushing between machines Gambling through meals Losing awareness of surroundings 	Casino gambling EGMs Online gambling
Betting patterns	<ul style="list-style-type: none"> Bets increasing over time/consistent amounts Frequency of gambling increasing over time Decreasing frequency of bets and less risky odds Bet more frequently online Chasing losses Wins are quickly recycled Bets placed late in roulette Repeated use of note acceptors 	Casino gambling EGMs Online gambling
Social behaviours	<ul style="list-style-type: none"> Avoidance of contact with others Asking staff not to let other people know about their presence Angry if others took favourite machine/spot Frequent complaints to customer service in online environment 	Casino gambling EGMs
Fund raising	<ul style="list-style-type: none"> Borrows or tries to borrow money to gamble Gets out more cash to use for gambling Reloads money during online gambling session Tries to cash cheque at venue Uses credit to gamble 	Casino gambling EGMs Online gambling
Emotional reactions	<ul style="list-style-type: none"> Feels/looks sad and depressed Feels/looks angry Cries after losing a lot of money Strikes machines or plays roughly Strikes gambling table with fists Talking to machine Blames machine or venue for losses 	Casino gambling EGMs
Appearance	<ul style="list-style-type: none"> Lack of concern regarding appearance Changes in appearance over time 	Casino gambling EGMs
Concurrent activities	<ul style="list-style-type: none"> Person smokes or drinks a lot while gambling 	Casino gambling EGMs

Table Four – Summary of indicators to identify potential problem gambling behaviours

Retrieved from: Delfabbro, King, & Griffiths. (2012). Behavioural profiling of problem gamblers: a summary and review. *International Gambling Studies*, 12(3), 349-366. doi: 10.1080/14459795.2012.678274

The use of proactive measures to identify potential problem gamblers through staff training and analysis of behavioural or player tracking data raises the challenging issue of how to respond to this information. Once an individual is identified as being at a particular risk level, a procedure needs to be established as to how to proceed. Operators that currently use behavioural tracking have indicated that they interact with customers indirectly, such as sending targeted messages. A range of possible actions can be taken depending on the context and risk level detected. For instance, individuals identified as low risk gamblers may be sent messages encouraging them to keep gambling at affordable levels, while those identified as being at moderate risk of harm may be encouraged to set limits, take a self-assessment, and consider taking a break. Individuals identified as being potential problem gamblers may be contacted personally for a further assessment and encouraged to self-exclude from some or all types of gambling as appropriate.

Analysis of player behaviour and correspondence is in practical use as a method of providing RG by several European online gambling operators. Some European casinos also use behavioural markers to trigger interventions, such as Holland Casino meeting with frequent visitors. Many of the variables used to identify risky gambling require individual players to be tracked over time. This is only possible in gambling systems that require some sort of player identification for gambling through a single account or entry into the gambling venue. Currently, the use of behavioural data for player tracking is limited in most land-based venues to voluntary use of player cards, and potentially voluntarily enabling tracking features within player accounts. Capacity exists for staff to create records and data for individual customers, but this is limited to the capacity for staff to accurately record data about specific patrons. There are some systems currently in development which are attempting to develop algorithms to analyze the risks associated with shorter gambling sessions using proxy or transactional data.

Despite the potential advantages of using behavioural data, there are still major restrictions that limit the extent to which behavioural tracking can be used to inform RG regulations. Firstly, there is still relatively little empirical, published evidence to support the validity of predictive variables. The majority of research has been conducted based on data provided by a single operator, mostly within Europe. Commercially operated player tracking systems mostly do not share the details of their algorithms or how these were derived. Secondly, only data that is collected can be analyzed for the identification of risky gambling behaviours. In most instances the lack of contextual variables (e.g., age, gender, occupation, income, other gambling) limits the implications that can be drawn from behavioural data alone. Thirdly, relying on algorithms may result in an imperfect system, but including high levels of human processing is time consuming and costly.

Concerns have also been raised regarding the potential for false positives, that is, identifying individuals as problem gamblers incorrectly. For all customers, unsolicited scrutiny of their behavioural records may be considered a violation of privacy. It is possible customers may find interventions intrusive and respond angrily and/or commence gambling elsewhere, potentially outside a legal system that incorporates customer protection measures. It is possible for processes to be introduced to increase the likelihood that customers will view contacts regarding their gambling problems positively. This requires significant work to understand customer attitudes and preferences and alert customers to the possibility of contacts, which should be framed in a non-judgemental manner as a service for customers, rather than as an attempt to impose restrictions.

Conversely, the potential for problem gamblers to go undetected is a considerable possibility, leaving open concerns regarding duty of care for operators to have successfully detected individuals with a disorder that is characterized by secrecy. Several legal challenges have been brought against gambling operators within Canada with regards to failing to enforce self-exclusion orders. It is possible that operators that fail to identify problem gamblers may have similar challenges raised.

Importantly, what actions are taken based on the identification of a potentially risky gambler need to be carefully considered. Although directly contacting gamblers who are suspected as having problems may be appropriate in some European jurisdictions, privacy concerns and beliefs regarding autonomy vary considerably between cultures. The most appropriate way to contact gamblers in Ontario should be investigated. It is important to assess gamblers' attitudes and perceptions of operators and regulators using behavioural data to address these in the implementation of responsible gambling policies.

Aggregated data

Collecting individual player tracking data is difficult in land based venues as most transactions are made with cash and privacy concerns exist regarding what identified data can be collected from player loyalty cards. Consequentially, aggregated data may have to be used for regulatory purposes. Some gambling regulators require operators to submit reports containing aggregated data periodically. This is largely to check compliance with integrity standards, such as ensuring that all games are paying at the regulated rate of return. Regulators could also ask for reports, or collect raw data and analyze this, to check compliance with responsible gambling practices.

Consultation with data management professionals in the pharmaceutical and health industries was informative in understanding how aggregated data can be used to understand consumer behaviour. Tools and dashboards can be created to enable large volumes of behavioural data to be accessible to individuals not trained in statistical and analytical procedures. Analytics can be run to determine whether consumers are using products in a way that suggests responsible, or risky behaviours. This can be conducted across all regulated games, as well as enable investigation of the use of specific games, such as new products in the market, or games that may be considered high risk. Analyses can also compare venues, or games within venues, to determine if some venues appear to have practices that encourage responsible gambling or if specific venues appear to have higher levels of risky gambling. This would enable resources to be directed to the most appropriate products and venues. Identifying venues or products with higher levels of risky gambling may trigger further investigation to check whether there is compliance with all regulations. It would also enable recognition of where responsible gambling practices appear to be effective, allowing investigation to understand how these practices could be implemented elsewhere. Comparing use of various gambling products will enable identification of products that are used more intensely, and may facilitate more risky gambling.

Types of analyses that may be run for specific gambling products within a specific venue, or across venues, include:

- Volume of total play (higher levels of betting is an indicator of more intense gambling)
- Proportion of time game is inactive vs. in active use (low periods of inactivity suggest there is a high demand for the product)
- Average length of time between active play (short breaks between activity indicate that there is a high demand for the product)
- Escalation of bets within a single session (data patterns that show a rise in betting intensity in a single session suggest the product encourages intense betting)
- Identification of time periods actively used (products used mostly in the evenings and on weekends are likely being played by gamblers in venues for entertainment, products used heavily during week days and mornings are likely being used by more regular bettors)
- Average increase in spend per session and across time cohorts (morning, afternoon, evening)
- Average speed of bets per session
- Average variation in bets per session
- Proportion of wins that are immediately cashed out

- Average size of credits cashed out
- Proportion of sessions that are played to zero credits

In context where there is not identifying customer information gathered segmentation can understand customer cohorts by looking for patterns in play. For example, customer behaviour can be compared across time periods and between game types.

Examination of behavioural data at an aggregated level can be used to inform regulatory policies regarding RG. For instance, the introduction of any new gambling product requires consideration regarding the extent of harm that this may introduce. Products that are assessed as appropriate may be introduced in a trial manner in specific venues. Behavioural data can then be used to assess the impact of the new product on risky gambling. This may be indicated by very high volume and intensity of play, short breaks between sessions, a low proportion of idle time for the game. The impact of the game as a popular form of entertainment as compared to a risky product can be examined by assessing the times when the game is most used (i.e., in the evenings and weekends compared to during week days) and increases in average bet size over time and within gambling sessions. This type of monitoring and use of data to inform regulation was reported by one international jurisdiction that actively used player behaviour data to assess the impact of EGM games which introduce new features, particularly when the evidence base examining the impact of the feature is lacking.

The decision to introduce a new gambling venue or additional gambling products within an existing venue can be aided by examining behavioural data. If there are already high levels of gambling in the area and short breaks between sessions of play, this may suggest that additional gambling venues or products are appropriate to cater for the high demand for gambling opportunities. However, if the venue or nearby venues appear to have high levels of risky gambling, as indicated by low proportion of players cashing out of sessions, it may suggest that further gambling opportunities may increase gambling related harm. In order to effectively inform these types of decisions some regulators reported using data outside of the gambling realm to assist in this type of decision making process. For instance, use of socio-economic and/or household economic data provides additional context to the player behaviour data. This may assist regulators considering large, impactful policy decisions.

Individual venues or venues grouped by common factors (e.g., manager, location), can be assessed with regards to the extent that they are complying with regulation to facilitate responsible gambling if they are shown to have markers of high risk gambling as compared to similar venues.

What constitutes best practice related to the use of operator and/or regulator data to ensure compliance with regulatory standards?

At the current time it is not possible to provide a detailed overview of 'best' practice in this area due to the relative scarcity of information available. The information provided below illustrates good practices and is informed from a variety of sources and perspectives and recommendations moving forward.

Good practices

- Data analysis to inform policy and regulation should include data from a variety of sources. For instance, socio-economic data, household expenditure data, health and social welfare data and reports. The information gathered from the analysis of these data sources in combination will bring more value than analysis completed in isolation.
- Significant advances in the RG and consumer protection field are being made by gambling operators. It is therefore imperative for regulators to work closely and collaboratively with operators wherever possible to design systems to address complex RG issues in an effective and comprehensive manner.
- Regulators and operators should consult with independent RG experts to ensure that the systems being created are based on the most recent and relevant research and understanding of the field.
- Relationships between regulators and operators are facilitated by the development of trust. To foster this, open communication, understanding differences, and identification of goals and intentions are useful strategies.
- Gambling data, and particularly anonymous gambling data, should be made available for the purposes of RG research to advance the state of the knowledge in this area.
- Regulators may benefit from access to a copy of data gathered via Internet gambling initiatives. Regulators may prefer to be able to complete their own analysis to facilitate transparency around risks, processes and procedures and/or establish requirements for information to be presented upon request.
- Specific, dedicated and knowledgeable staff are required in order to significantly advance the use of big data and complex analytic processes.
- Education regarding the meaning of the data and its potential uses is required in order to ensure that the output from any data analytics is effectively understood by all relevant stakeholders.
- Infrastructure and regulation that supports data sharing and collaboration is required prior to the meaningful use of data across government departments.
- Lack of access to meaningful data from multiple sources will restrict the value of the analytic capabilities that are possible in the big data context.
- Privacy information should be updated to inform gamblers of how their information is to be used (by whom, when and why) where relevant and appropriate. This will help to satisfy privacy regulations and also demonstrates openness and transparency to the public.
- The questions/processes in place to review/analyze the data will be dependent upon:
 - the infrastructure to support data analysis (capacity, storage, tools, resources)
 - the availability of information

- the talent and knowledge of the team.
- Aggregated data analytics reduces the risk of privacy violations.
- A clear understanding of transactional data elements that are available for analysis may assist in the appropriate analysis and use of resultant information depending on the nature of the analytical tools used.
- Technologies and techniques such as text based analytics may be possible within the gambling context by using the multiple information sources available including social media accounts, customer feedback, and information records from RG interventions within gambling establishments. Analysis does not need to be restricted to quantitative figures.
- Player behaviour data provides the greatest level of insight into the gambling context.

Specific examples of how data can be used to inform regulation

In the absence of significant examples of how data can be used to inform the development or enforcement of regulation, the project sought to identify how data can be specifically used in future strategies. Examples of these strategies are outlined below for consideration

<p>Multiple data sources (gambling and related fields)</p>	<p>Multiple data sources including gambling specific data and data from alternate sources such as:</p> <ul style="list-style-type: none"> • Socio-economic • Household expenditure • Healthcare utilization data (e.g., call to gambling help services, addiction services, domestic violence rates) • Financial services data (e.g., number of bankruptcies, loan defaults, overdraft rates) <p>Can be overlaid and used to provide additional insights from which regulatory decisions can be made about the suitability of gambling activities and licenses in a specific geographical location.</p> <p>For example, the data could be utilized to decide upon the distribution of new gambling licenses, the appropriateness of opening new gambling venues and decisions regarding the availability of alcohol within these establishments.</p>
<p>Transactional data</p>	<p>In the future, transactional data may be used to specifically analyze the impact that different features and games have upon behaviour.</p> <p>If discrete gambling sessions can be identified within the transactional data, data analysis should facilitate the understanding of how the game and/or its features are impacting the gambler (e.g., leading to chasing losses, erratic behaviour)</p>

Conclusions

This project sought to identify what constitutes best practice related to the use of operator and/or regulator data to ensure compliance with regulatory standards. This subject matter is currently evolving as evidenced by the relative scarcity of reliable published evidence and feedback from key stakeholders interviewed as part of the project. It is evident however, that the subject is extremely relevant.

Conference proceedings, meetings, consultation and discussions are all currently underway aimed at deciphering the most appropriate manner in which to use the vast volumes of data that are being made available within our current environment. The challenge arises in having the skill, insight and capability and capacity to deal effectively with these datasets within the current frameworks of privacy and liability.

Increased collaboration between stakeholders (operators, third party providers, and regulators) will strengthen the current relationships and enable strong innovations moving forward. Within the Canadian environment there is strong evidence that operators across the country are motivated at advancing RG and there are innovative examples of how progress is being made in this context.

The development of technologies to support analysis and increase the understanding of gambler behaviour and data will be a vital element of how regulation is influenced in the future. The application of this type of knowledge broadly within the gambling community will provide rich insights particularly if that data can be captured and analyzed in conjunction with other, meaningful datasets.

The use of data analytics can be a complex undertaking or a straight forward task, depending on the data set under analysis and the desire to extend the breadth of data informing the answer. Analysis of individual, transactional level data, although challenging for more traditional forms of analysis, may provide a wealth of insights in the big data context. Assessing the desired outcome and the resources available to achieve it will help set and guide the direction taken.

References

- ACL Services Ltd. (2010). Discussion Paper: Fraud Detection Using Data Analytics in the Healthcare Industry. Retrieved from http://www.acl.com/wp-content/uploads/DP_Fraud_detection_HEALTHCARE.pdf
- Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., . . . Widom, J. (2011). Challenges and Opportunities with Big Data 2011-1 Purdue University (Ed.) *Cyber Center Technical Reports. Paper 1*. Retrieved from <http://docs.lib.purdue.edu/cctech/1/>
- Amarasingham, Ruben, Patzer, Rachel E, Huesch, Marco, Nguyen, Nam Q, & Xie, Bin. (2014). Implementing Electronic Health Care Predictive Analytics: Considerations And Challenges. *Health Affairs*, 33(7), 1148-1154.
- Arjani, N, & Graydon, P. (2013). Lessons from the financial crisis: Bank performance and regulatory reform. Retrieved from <http://www.banqueducanada.ca/wp-content/uploads/2013/12/dp2013-04.pdf>
- Association for Data-driven Marketing & Advertising. (2013). Best Practice Guideline: Big Data. *A guide to maximising customer engagement opportunities through the development of responsible Big Data strategies*. Retrieved from <http://www.adma.com.au/assets/Uploads/Downloads/Big-Data-Best-Practice-Guidelines2.pdf>
- Association, World Lottery. (2013). Lotteries and Big Data. *WLA magazine*, 39.
- Bates, David W., Saria, Suchi, Ohno-Machado, Lucila, Shah, Anand, & Escobar, Gabriel. (2014). Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Affairs*, 33(7), 1123-1131. doi: 10.1377/hlthaff.2014.0041
- Bierig, R., Piroi, F., Lupu, M., Hanbury, A., Berger, H., Dittenbach, M., & Haas, M. (2013). Conquering Data: The state of play in intelligent data analytics.
- Blaszczynski, A., Parke, A.J., Parke, J., & Rigbye, J. L. (2014). Operator-based approaches to harm minimisation: Summary, review and future directions. London: The Responsible Gambling Trust.
- Booz Allen Hamilton. (2012). Cloud analytics playbook Retrieved from <http://www.boozallen.com/media/file/cloud-analytics-playbook.pdf>
- Bradshaw, S., Harris, K., & Zeifman, H. (2013). Big data, big responsibilities: Recommendations to the Privacy Commissioner *The Centre for International Governance Innovation - Junior Fellows Policy Brief Series*, (8). Retrieved from http://www.cigionline.org/sites/default/files/no8_0.pdf
- Braverman, Julia, Tom, Matthew A., & Shaffer, Howard J. (2014). Accuracy of self-reported versus actual online gambling wins and losses. *Psychological Assessment*. doi: 10.1037/a0036428
- Canadian Foundation for the Advancement of Investor Rights (FAIR). (2014). A report on a Canadian strategy to combat investment fraud. Retrieved from <http://faircanada.ca/wp-content/uploads/2014/08/FINAL-A-Canadian-Strategy-to-Combat-Investment-Fraud-August-2014-0810.pdf>
- Chen, C.L.P, & Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347.

- Chen, Sandy C., Shoemaker, Stowe, & Zemke, Dina Marie V. (2013). Segmenting slot machine players: a factor-cluster analysis. *International Journal of Contemporary Hospitality Management*, 25(1), 23-48. doi: 10.1108/09596111311290200
- Cohen, I. Glenn, Amarasingham, Ruben, Shah, Anand, Xie, Bin, & Lo, Bernard. (2014). The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. *Health Affairs*, 33(7), 1139-1147. doi: 10.1377/hlthaff.2014.0048
- Consumer Financial Protection Bureau (CFPB). (2014). Strategic plan, budget, and performance plan and report. Retrieved from <http://www.consumerfinance.gov/strategic-plan/>
- Currie, S., Hodgins D., Wang, J., el-Guebaly, N., Wynne, H., & et al. (2006). Risk of harm among gamblers in the general population as a function of level of participation in gambling activities. *Addiction*, 101, 570-580.
- Davies, Bill. (2007). iCare: Integrating Responsible Gaming Into Casino Operation. *International Journal of Mental Health and Addiction*, 5(4), 307-310. doi: 10.1007/s11469-007-9078-4
- Davies, M. (2013, July). New regulations require cleaner data. Retrieved from <http://blogs.reuters.com/financial-regulatory-forum/2013/04/18/new-regulations-require-cleaner-data/>
- Delfabbro, Paul, King, Daniel Luke, & Griffiths, Mark. (2012). Behavioural profiling of problem gamblers: a summary and review. *International Gambling Studies*, 12(3), 349-366. doi: 10.1080/14459795.2012.678274
- Delfabbro, Paul, King, Daniel Luke, & Griffiths, Mark. (2012). Behavioural profiling of problem gamblers: a summary and review. *International Gambling Studies*, 12(3), 349-366. doi: 10.1080/14459795.2012.678274
- Dusseault, P.L. (2013). Privacy and social media in the age of big data *Report of the Standing Committee on Access to Information, Privacy and Ethics* (Vol. 41st parliament, first session).
- Financial Transactions and Reports Analysis Centre of Canada (FINTRAC). (2014). FINTRAC Glossary Retrieved 1 September 2014, from <http://www.fintrac.gc.ca/multimedia/education/g-eng.asp>
- Gainsbury, Sally. (2011). Player account-based gambling: potentials for behaviour-based research methodologies. *International Gambling Studies*, 11(2), 153-171. doi: 10.1080/14459795.2011.571217
- Gartner. (2012). The Importance of 'Big Data': A Definition. Retrieved from <http://www.gartner.com/id=2057415>
- Personal Information Protection and Electronic Documents Act, (2000).
- Graf, P. (2013). *Public confidence – the key to innovation and growth in the gambling industry*. Paper presented at the BISL Annual Conference 2013.
- Graf, P. (2014, Tuesday 4 February 2014). *Innovation as a force for good - exploiting technology to secure public confidence*. Paper presented at the ICE World Regulatory Briefing.
- Griffiths, M.D., & Wood, R.T.A. (2008). Responsible gaming and best practice: How can academics help? . *Casino and Gaming International*, 4(1), 107-112.
- Haefeli, J., Lischer, S., & Schwarz, J. (2011). Early detection items and responsible gambling features for online gambling. *International Gambling Studies*, 11(3), 273-288.

- Hensel, S. (2013). Regulator and industry collaboration: Changing the dynamic. *Canadian Gaming Lawyer* 6, 8-9.
- Hodgins, D.C., Wynne, H., & Makarchuk, K. (1999). Pathways to recovery from gambling problems: Follow-up from a general population. . *Journal of Gambling Studies*, 15, 93-104.
- Korn, David, Gibbins, Roger, & Azmier, Jason. (2003). Framing public policy towards a public health paradigm for gambling. *Journal of Gambling Studies*, 19(2), 235-256.
- Ladouceur, R., Blaszczynski, A., & Lalande, D. R. (2012). Pre-commitment in gambling: a review of the empirical evidence. *International Gambling Studies*, 12(2), 215-230.
- Levi, M. (2009). Money Laundering Risks and egaming: A European overview and assessment Cardiff, Wales: Cardiff University.
- Lycka, Martin. (2011). Online Gambling: Towards a Transnational Regulation. *Gaming Law Review and Economics*, 15(4), 179-195. doi: 10.1089/gltre.2011.15404
- Lycka, Martin. (2014). Regulating Online Gambling Markets: Regulatory Guide. *Gaming Law Review and Economics*, 18(4), 383-390. doi: 10.1089/gltre.2014.1846
- Manitoba Lotteries. (2012, 5 April 2012). Manitoba partners with British Columbia to offer online gambling Retrieved August 2014, from http://www.manitobalotteries.com/uploads/news/news_pdf_207.pdf?t=1333654448
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition and productivity McKinsey Global Institute (Ed.)
- Murdoch, Travis B, & Detsky, Allan S. (2013). The inevitable application of big data to health care. *JAMA*, 309(13), 1351-1352.
- NatCen Social Research. (2014). Research consortium to examine potential for gaming machine data to identify problem gambling. Retrieved from NatCen Social Research that works for society website: <http://www.natcen.ac.uk/news-media/press-releases/2014/february/research-consortium-to-examine-potential-for-gaming-machine-data-to-identify-problem-gambling/>
- National Research Council Canada. (2014). Text analytics Retrieved October 2014, from http://www.nrc-cnrc.gc.ca/eng/solutions/advisory/text_analytics.html
- Neillson, T. (2009). *Using Technology to Assist Player Using Technology to Assist Player Control Control -- An Example from Sweden* Paper presented at the Discovery, Niagara Falls, Ontario.
- Nevada Council on Problem Gambling. (2014). Nevada Gaming Regulation, from <http://www.nevadacouncil.org/responsible-gaming/nevada-gaming-regulation/>
- Office of Science and Technology Policy. (2012, March 29). Obama administration unveils “big data” initiative: Announces \$200 million in new R&D investments Retrieved July, 2014, from http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf
- Office of the Privacy Commissioner of Canada. (2011, May 2011). Cookies – Following the crumbs, 2014, from https://www.priv.gc.ca/resource/fs-fi/02_05_d_49_e.asp
- Pruden, H. (2002). An overview of the gambling provisions in canadian criminal law and first nations gambling. *The journal of aboriginal economic development*, 2(2), 37-40.

- Quilty, L. (2013). Identifying Problem Gamblers in Gambling Venues in Ontario, Canada. Report submitted to the Ontario Problem Gambling Research Centre.
- Roski, Joachim, Bo-Linn, George W., & Andrews, Timothy A. (2014). Creating Value In Health Care Through Big Data: Opportunities And Policy Implications. *Health Affairs*, 33(7), 1115-1122. doi: 10.1377/hlthaff.2014.0147
- Schellinck, T, & Schrans, T. (2008). *Player tracking analysis: Inside the black box. Using player tracking data to manage risk*. Paper presented at the European Association for the Study of Gambling Slovenia. http://www.easg.org/media/file/conferences/novagorica2008/thursday/1610-ses1/schrans_tracy.pdf
- Stoddart, J. (2013). *The age of big data: A new, challenging frontier for privacy and our society*. Paper presented at the Conference Board of Canada's Chief Privacy Officer Council, Montreal, Que.
- Strømshlid, Jan Peder. (2013). *Player Card with ID: A new way to operate and regulate?* Paper presented at the IAGR Conference, Oslo, Norway. <http://iagr.org/wp-content/uploads/IAGR2013-session3a-Jan-Stromslid.pdf>
- Tene, O., & Polonetsky, J. (2012). Privacy in the age of big data: A time for big decisions. *Stanford Law Review*, 63. Retrieved from
- U.K Gambling Commission. (2014). Proposed amendments to the social responsibility provisions in the licence conditions and codes of practice for all operators (the LCCP): Consultation document. Retrieved from http://www.gamblingcommission.gov.uk/publications__consultations/consultations/open_consultations/consultation_social_respons.aspx
- Wardle, H., Ahmed, H., & Coshall, C. (2013). Scoping the use of industry data on Category B gaming machines. Retrieved from <http://www.natcen.ac.uk/media/205543/industry-data-scoping-final.pdf>
- Wiebe, J., & Lipton, M. (2009). Internet Gambling Regulations. *Canadian Gaming Lawyer Magazine* 2.