ANALYTICS IN GOVERNMENT QUARTERLY

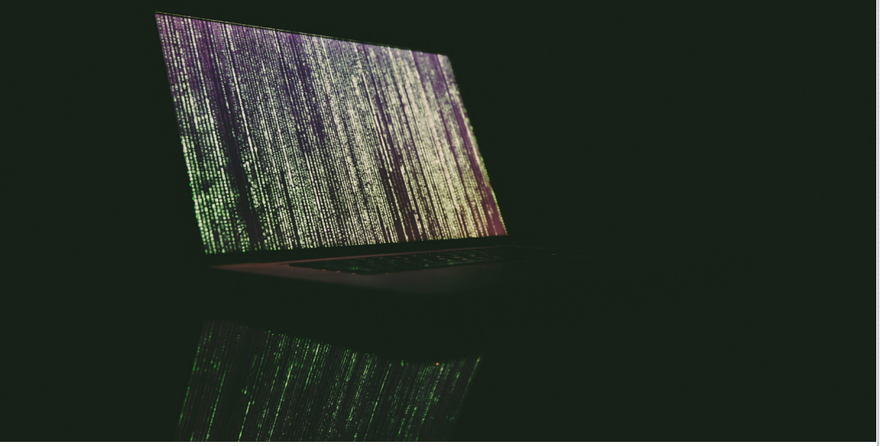
FOR GOVERNMENT DECISION MAKERS

THE FUTURE OF GOVERNMENT SERVICES

INSIDE:

ISSN 2562-9123
JANUARY 2022
ISSUE #6

CONTENTS



ANALYTICS IN GOVERNMENT QUARTERLY

Editorial

Managing Editor |
Gregory Richards, Ph.D.

Editor & Design |
Yu(Yvonne) Leng

Columnists |
Boris Bogatirev
Kevin Kells, Ph.D.
Hubert Laferrière
Rami Abielmona, Ph.D.

Editorial Advisory Board

David Buchanan
Team Lead , Canadian Federal Government, SAS Canada

Stéphane Gagnon, Ph.D.
Associate Professor, Business Technology Management, Université du Québec en Outaouais

Tara Holland
Principal Industry Consultant, Global government Practice, SAS Canada

Alex Ramirez, Ph.D.
Associate Professor in Information Systems, Sprott School of Business, Carleton University

Eduardo Rodriguez, Ph.D., MBA, MSC. Math.
Principal IQAnalytics Inc.
Sentry Insurance Endowed Chair in Business Analytics, University of Wisconsin-Stevens Point
Analytics Adjunct Professor, Telfer School of Management, University of Ottawa

Betty Ann M. Turpin, Ph.D., C.E.
President of Turpin Consultants Inc.

EDITOR'S NOTE

Welcome to the January 2022 issue. We are evolving the articles to include more opinion pieces and to drill down further into how different forms of analytics can enable delivery of government services. We encourage our readers to provide their own opinions of the concepts discussed. We will publish these in the next issue of AGQ.

In "It's not just about the technology", we poll our writers about the future of government services enabled by analytics. The consensus is that analytics can help improve productivity across government services by breaking down silos across departments, building more integrative management systems, and enabling better use of information. In this issue, for example, Kevin Kells and Cameron Hopgood from the Ottawa police discuss co-production of data and the important topic of open data. Thematically, it reinforces the notion of integrative management of complex issues across government organizations.

Hubert Laferrière addresses the role humans play in automated decision making. There is much concern about algorithms gone wild, but Hubert's paper shows that policies and procedures are in place to ensure that humans stay engaged with automated decision making. Nevertheless, some of these policies can fall short since they often narrow the scope of decisions that need to be carefully monitored. Ultimately, the algorithms are supposed to enable the work of humans: we need to be careful not to let them run unfettered. Related to this theme is the fact that bad data can lead to bad decisions. Sunil Meharia and Betty Ann Turpin explore the challenges of integrating Human Resources data from several different source systems.

This issue also introduces a new topic related to forecasting IT project risk through automated semantic analysis. When applied to project management, this approach mines project documentation to identify entities related to project risk. Relationships between the entities are then highlighted to help project managers identify and proactively manage risk. It's a new approach, fresh from research conducted by Franck-Olivier Kwan, Véronique Nabelsi, Stéphane Gagnon, and Wassim El Kass, that demonstrates the power of analytic tools to address complex problems in government service delivery.

We hope you enjoy the issue and as always, we welcome your comments and suggestions.

Gregory Richards, Ph.D.
Managing Editor

Corporate

Analytics in Government Quarterly magazine is published four times per year by the Government Analytics Research Institute, a consortium of the University of Ottawa, Carleton University, the University of Quebec en Outaouais, SAS and the Institute on Governance. The institute conducts research with government organizations who are experimenting with the introduction of analytics of all forms. Professors and students work on proof of concepts, testing of algorithms as well as examining the organizational practices needed to fully integrate analytics into business processes.

All opinions expressed herein are those of the contributors and do not necessarily reflect the views of the publisher or any person or organization associated with the magazine.

General Inquires

Letters, submissions, comments and suggested topics are welcome, and should be sent to agq@governmentanalytics.institute or visit our website http://governmentanalytics.institute

Subscription Information

You can subscribe to the magazine online at http://governmentanalytics.institute/magazine

Reprint Information

Privacy Policy

We do not sell our mailing list or share any confidential information about our subscribers.

ISSN 2562-9123

Is Human Oversight of Automated Decision System Unsolvable?

"Human-in-the-loop", "human control of technology", "human insight" are terms associated with automated decision system (ADS). They involve measures to avoid adverse or harmful effects on people. They have become more important in public sectors such as health, social welfare, taxation, police, and communication, because "harm reduction" pursuit has increased over the years and the means to deal with adverse effects are limited.

ADS is a decision-making process that uses machine and algorithm to automate processing without any human involvement. It may replace human judgement by generating a score (e.g., someone who reached the acceptable score is automatically granted of the benefits of a government program). Or it may assist an agent to perform administrative decisions by bringing in evidence based on factual data and inferred data. Some believe that the continued and increasing use of ADS by states affects governmental approaches to manage public services and functions and that perhaps the replacement of human judgment by an algorithm makes people more vulnerable.[1] The UN High Commissioner Agency is concerned that the continued and increasing use of ADS by states pose negative, even catastrophic risks to human rights, policing and justice.[2] Recent reports in Brazil state numerous errors generated by an algorithm trained on racial biases that has led to people being imprisoned when they should not be.[3]

Addressing Harms: Human-in-Control, Human Insight

In *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI,* the authors found "human control of technology" as one of the eight themes that constitute the "normative core" of a sound AI governance. They distinguished three key principles that embodied the theme: (1) the human review of automated decision principle; (2) the ability to opt out of automated decision; and (3) the opportunity for people to choose how and whether to delegate decisions to AI.

Government authorities have formulated, for some years, policies to regulate the use of ADS in the public sector. They defined the key requirements of the "human-in-the-loop" or "human insight" necessary to ensure that algorithms must do what it must do without adverse effects.

The Canadian government implemented, in early 2020, a *Directive on Automated Decision-Making*. All departments and agencies must apply the Directive and undertake an Algorithmic Impact Assessment. The risk level results determine additional and appropriate requirements such as quality assurance procedures and control measures. Thus, "human-in the loop," which is listed as a mandatory item to be assessed, is being adjusted at the required level to mitigate risks: with minimal impact risk, decisions may be made automatically without direct human involvement. With higher impact risk, a decision cannot be made without having specific human intervention points throughout the decision-making process; the final decision must be made by a human.

The Normative Core Themes

- Privacy
- Accountability
- Safety & Security
- Transparency & Explainability
- Fairness & Non-discrimination
- Human Control of Technology
- Professional Responsibility
- Promotion of Human Value

The *Automated Decision Systems Accountability Act of 2021*, a legislative proposal currently in the process of being adopted by the California Legislature, would impose continuous tests for biases during the development and usage of the ADS.[4] Earlier this year, the European Commission made a proposal for a regulation laying down harmonised rules on artificial intelligence, complementing the General Data Protection Regulation (GDPR). The proposal prescribes prohibiting unacceptable AI practices and banning AI systems that could potentially harm people.[5]

There are already several calls for firmer measures: Michelle Bachelet, the UN human rights chief, called for a moratorium on the sale and use of AI systems, including a ban on AI applications that cannot be operated in compliance with international human rights law.[6] Many cities in the US have shared the same line of thought: they had already banned the use of facial recognition systems used by police and security enforcement agencies because of many mishaps.

The policies are not only calls for rights and ethical principles of compliance and respect. It is an appeal for implementing specific measures to frame the algorithm itself and its operating conditions, this because an algorithmic chain value may hold several defects (up to nine social biases were identified[7]). The policies include specific measures to prevent the negative effects of algorithms (mitigate or ideally end them). Absence or lack of consideration about measures safeguarding a true human insight or control over the algorithmic model is detrimental to the sustainability of ADS.

Flaws in Human Oversight Policies

Conducting a review of forty human oversight policies, Ben Green has found some flaws with the current and proposed control measures on algorithm.[8] He has observed that policies were not based on solid evidence, and they were not sufficient to ensure proper human oversight. As a result, people are unable to perform the sought-after oversight function. Moreover, it engendered a more damaging consequence: the current human oversight policies legitimize government uses of faulty and controversial algorithms.

1 Turkle, S. Technology and Human Vulnerability; Harvard Business Review (2003), 81(9). And: Anderson, J. and Rainie, L. Artificial Intelligence and The Future of Humans (December 2018); Pew Research Center.

2 The Right to privacy in the digital age, Report of the United Nations High Commissioner for Human Rights; (13 September 2021 - Advance Edited Version); A/HRC/48/31.

3 Racisme. Au Brésil, les systèmes de reconnaissance des suspects posent problème; Courrier international (2021-10-10); Courrier international SA, Paris.

4 AB-13 Public contracts: automated decision systems; (2021-2022); version 07/15/21; California Legislative Information

5 Proposal for a Regulation of The European Parliament and Of the Council, Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts (April 2021), European Commission, COM (2021) 206 final.

6 UN News, Sept.15 2021

7 Silva, S and Kenney, M. Algorithms, Platforms, and Ethnic Bias; Communications of the Association for Computing Machinery (ACM), vol.6, number 11, pp.37-39, November 2019.

8 Green, Ben, The Flaws of Policies Requiring Human Oversight of Government Algorithms (September 10, 2021). SSRN: https://ssrn.com/abstract=3921216

For instance, policies aiming "solely" at restricting ADS end up narrowing the scope of cases that could be problematic, leaving the faulty ones in operation. Several policies encourage bypassing some restrictions, like Article 22 of the GDPR (this situation is often referred as rubber stamping). Other policies encourage the work in tandem, algorithm and human judgment, allowing human judgment to overrule automated decision results; this may result in the decision-maker being bound to information generated by the ADS (automation bias/fettered decision). In the case of policies proposing a meaningful human oversight where human is the unique entity who can consider the context (e.g., a "subject" dealing with an ambiguous and conflicting situation), B. Green found the policies did not supply criteria nor standards to clearly define human oversight.

Reinforcing the Human Insight Framework

The findings and conclusions are worrisome as they cast serious doubts on the ability of policies to permit proper supervision and control over algorithms and their operating conditions. If ADS resumes its growth within public sector, the role of human insight becomes a more pressing issue. There is a need to develop more targeted and exact measures to reinforce a human control framework. Green proposes a twofold approach to delineate such measures.

First, policy developers, data scientists, executives, and program management must simply examine whether there is a need for ADS for their targeted business process. To do so, one needs to decide if the process relies only on human judgment. If it is the case, ADS is not appropriate. If there is no or little need to exercise human judgment (often labelled as discretionary judgement) and a thorough due diligence on the algorithm was undertaken, then, ADS use could go forward in tandem with human judgment or can even go forward solely. This is somewhat similar to the Canadian Directive about proportionate risk mitigation measures for the "human-in the loop" requirement.

Green will then insist on filling the information gap about the impact and interaction of the algorithm and human judgement. For instance, experiments on the collaboration between human and algorithm including ongoing quality assurance measures that could shed light on what is the core element of automated decision. This will generate empirical knowledge from the "inside." In return, both information and knowledge could help policy developers to better align oversight policies with tangible and embedded measures that could be easily integrated to the algorithms.

Green recognizes his proposals are based on a technological solutionism approach. He is convinced that evidence-based frameworks "...develop approaches to using algorithms that promote rather than undermine central values of public governance."

We barely scratched the human insight or in-the-loop question. One thing is sure: the need to scrutinize further the effects of algorithms on decision making, including interaction between human judgement and algorithm. This would allow better defined and more precise policies to avoid negative impact on people. Another "thing" is also sure: clearly confirm central values and develop practical criteria for public governance.

About the Author

Hubert Laferrière was the Director of the Advanced Analytics Solution Centre (A2SC) at Immigration, Refugees and Citizenship Canada. He had established the A2SC for the Department of IRCC and led a major transformative project where advanced analytics and machine learning were used to augment and automate decision-making for key business processes.

Ottawa Police Service: Lessons Learned on Co-producing Community Safety Data Analytics

By Kevin Kells, Ph.D. & Cameron Hopgood

With rising calls for police reform and trust at an all-time low, police services must reframe their approach to delivering public safety services. In Ontario, part of this change has been enabled through legislation that aims to address the social determinants of health through Community Safety and Well-Being Plans. From a data and analytics perspective, this means leveraging data across domains to find innovative solutions for harm reduction and improved delivery of service. Open data and analytics are enablers to co-production, the meaningful collaboration with community partners, to address this complex challenge. Expanding access to data also supports the need for greater transparency in policing, while improving community awareness of trends and the police response.

The Ottawa Police Service (OPS) has long supported neighbourhood-based problem-solving through the provision of data, analytical reports, and web-based mapping solutions. Part of this commitment is also reflected in partnerships with Open Ottawa and the Ottawa Neighbourhood Study (ONS) that help increase access to data. Despite incremental improvements, the public appetite and expectation for open data has grown. This includes access to a user-friendly and dynamic environment that helps visualize key indicators and trends with geographic reference. Technological advancements, support from vendors, and awareness of the benefits of open data have led to improved solutions.

With a new approach to change, the following are lessons learned as the OPS moves towards co-producing an open data and analytics platform with the Community:

1. Listen to your community; feedback comes in many forms

The police service has benefited from strong feedback on the need for better data and analytics. This has come through a variety of channels and significant public discourse around demonstrating the value of public investments. Channels that have helped solidify this need and the core requirements include oversight bodies, Councillors, social service partners, academic institutions, residents, and freedom of information requests.

2. Analytics needs to be managed as a product

Public data offerings must be thought of as a product, not a one-off. When you first publish your data or platform, there may be initial joy that now there is something where previously there was nothing. Don't confuse the initial elation with satisfaction. You must be ready to sustain the solution and iteratively enhance the product based on feedback from residents and partners.

3. Maximize your existing technology investments

The OPS journey started with a low cost external solution that bolted on to existing technology investments, but did not integrate with existing systems. This permitted only limited internal control over content and configuration. This balanced basic requirements from the community within a limited budget envelope, and was a low cost alternative to configuring internally available applications to provide this service. With the need for a more robust capability that is scalable and can be regularly enhanced based on community feedback, the Police Service is moving towards an internally hosted service. This will maximize an existing technology investment to enable both the public and police service to leverage a similar platform.

4. Consider total cost of ownership and the benefit of responsiveness

Consider the total costs of ownership (TCO) when comparing solutions and deciding between outsourced vs. in-house configuration. Cheaper is not always the best. An implementation with a lower price tag may deliver a product your organization will not have sufficient control over, nor receive sufficient support, nor meet the service level to fulfill the intended mission. There can be hidden costs as well, such as direct costs when staffing resources must be allocated to make up for a gap in a provider's service or implementation. And indirect costs where poor vendor responsiveness reflects on your organization's reputation of trust and confidence in the community.

Having control over the configuration and support of the product allows the delivery of higher value. The organization will be able to maintain a higher quality offering that better utilizes staff resources, and is more responsive to community input. This reinforces the desire to co-produce this information and this product with the community.

5. A consistent user experience for internal and external users

There is a benefit to adopting the same platform internally as offered externally for data visualization. The internal system shows a higher level of detail for management and operations, while the public-facing component has visibility to the same data without privacy or confidentiality risk. Because both environments are powered by separate instances of the same platform, there is a single system to learn and maintain. This approach helps improve system quality and reduce internal costs

Because the internal and external systems share a design philosophy meant for ease-of-use by non-computer whizzes, both citizens and internal users can benefit from improved data literacy. Consistency goes a long way to improving cohesion in data consumption, helping co-produce insights inside and outside our organization.

Bottom line: Let community value and trust be the drivers

Delivering on the expectation of a user-friendly and accessible platform for community safety data is important to build public trust, demonstrate a duty of care, and support community safety and well-being planning. The development approach for co-producing an open data platform must be fiscally responsible and, more importantly, deliver a valuable service to the community.

About the Authors

Kevin Kells, Ph.D., has worked as an R&D Engineer in software systems in the Financial and Semiconductor industries in Switzerland, Silicon Valley, and Ottawa, and currently works with real-time data and news feed systems at a major market news and data company in New York City. He has extensive experience in non-profit management, both in the area of human systems and IT systems. He received his Ph.D. from the Swiss Federal Institute of Technology (ETH) and holds an MBA from the University of Ottawa.

Cameron Hopgood, Manager Business Performance, Ottawa Police Service. His area supports senior leadership with risk advisory, corporate performance, and business analytic services. Cameron is passionate about developing partnerships and fostering collaborative solutions to address complex community based challenges.

IT IS NOT JUST ABOUT THE TECHNOLOGY

It is not so much the technology that ushers in the positive changes that will accrue from advanced analytics. It is a combination of the technology and the social changes that occur within and between organizations that in some cases, are stimulated by use of the technology.

Artificial Intelligence and other advanced forms of analytics are new tools for many organizations. This inherent novelty means that researchers and practitioners have focused extensively on the technology to try to better understand how to use these tools. But, to fully integrate analytics into organizations, we need to consider that we are dealing with a socio-technical system: technology impacts the way we work even as the way we work influences how we use the technology.

Socio-technical concepts emerged out of research conducted in the late 1940s by the Tavistock Institute. The first observations were noted in coal mining. One interesting fact is that as mechanization of mining at the coal face progressed, individual workers found themselves more isolated. But as more advanced forms of mechanization were introduced building on the initial efforts, teams found that they were able to work collectively on complete tasks thus reducing isolation and improving working conditions. The underlying premise is that the implementation of any technological solution should consider the specifics of the human system it is supposed to enable.

This paper gathers thoughts from AGQ authors related to the introduction of advanced analytics in government organizations. What is the long run expectation of the use of analytics? Will organizations become more productive? Will we see the kind of "mechanization of management" noted in the initial Tavistock studies? Will the wide availability of analytic tools and processes democratize information in such a way that work teams can become simultaneously more collaborative and autonomous?

"SOCIO-TECHNICAL SYSTEM"

How management processes will change as AI becomes more widely accepted? Will the role of "management" change? How does an autonomous work environment square with the policy context of a government organization?

VOICES OF AUTHORS

GREGORY RICHARDS, PH.D.
Vice Dean, Graduate Professional Programs and Director, Executive MBA, Telfer School of Management, University of Ottawa

Conceptually, if done right, the use of advanced analytics helps to digitalize management so to speak. Planning, organizing and operational management can be greatly aided with analytic tools. So far, however, I've seen information overload more than anything else. But I think that's because everything we are doing in this space is so new. In the long run, I believe these tools will enable government organizations to be more effective because we can capture and deliver external data on outcomes realization to managers. And we can account for time lags. The reality is that the things we do today in organizations might not have an impact until five years from now for example. We can also better align attribution: parsing out the organization's impact on outcomes relative to the work of other organizations. These issues are complex and difficult to sort at the moment.

But I also think it will call for a rethinking of the process of management. Most organizations have already been moving in the direction of less "command and control". In the government context, we cannot escape policy directives, but we could at some point, integrate regulatory structures within the analytic framework such that all decisions become "context aware" permitting more autonomy for work teams. Our research on regulatory intelligence, for example, suggests that in a few years we will be able to link program goals to specific regulatory instruments so that managers are continually aware of the policy context as they make decisions enabled by analytic tools.

STÉPHANE GAGNON, PH.D.

Associate Professor, Business Technology Management (BTM), the Université du Québec en Outaouais (UQO)

Indeed, it is not "just" about the technology. But like in many other adoption cycles, it takes time before practitioners awaken to the importance of integrating social-technical systems. In the case of AI and Analytics (AIA), we should be thankful that this phase of "blind tech. euphoria" has been relatively shorter and less damaging than some others, for example, the e-commerce bubble from mid-1990's, or the race for full-scale Enterprise Resource Planning (ERP) implementations of the early 2000's. Also, the rigor that most governments have displayed in emphasizing ethical and explainable AIA, and dutifully implementing policies that follow world-class best practices, are all reassuring trends. The changes in AIA strategy in 2020-2021, linked obviously to the ominous spirit of the pandemic, have nevertheless laid a more solid ground for project success and scalability in coming years.

But some weaknesses remain, and as I wrote in the last issue of AGQ, our AIA strategies must evolve beyond simply focusing on Machine Learning (ML). There is a wide range of AI technologies, including Knowledge Representation, Knowledge Graphs, Decision Rules, Semantic Reasoning, Multi-Agent Systems, Robotic Process Automation, etc. All of these integrate with ML and Analytics, whether descriptive, predictive, or prescriptive. Again, it is not just about these technologies, but public sector AIA shops are bound to develop solid competencies in many of them, depending on the specific functionality required by their policies and jurisdictions.

ALEX RAMIREZ, PH.D.

Associate Professor in Information Systems, Sprott School of Business, Carleton University

In the long run, more government organizations will see the opportunities of using artificial intelligence in their operations. It just makes sense. Productivity is about doing more with less, and one of the resources that is becoming scarcer is human-response time to an ever-increased demand for our attention. In this deluge of data that needs to be analyzed, artificial intelligence algorithms will do most of the grunt work and leave the high-level analysis and decision-making to humans. That will allow them more time to work on the complex cases without being overwhelmed by too many easy-to-deal, simple decisions.

I don't think that incorporating AI solutions will make our jobs more mechanical, on the contrary, by reducing the demand on our limited attention span, we can become more efficient and effective in our decision-making activities. To really benefit from these AI solutions, the data that governments have must be open, that way the AI algorithms will become more robust and accurate. Management will change, because everything else is changing, but the change is one that will make our organizations more agile and ready for the challenges of the future.

HUBERT LAFERRIÈRE

Former Director of the Advanced Analytics Solution Centre (A2SC) at Immigration, Refugees and Citizenship Canada

What will be the future for the public sector? I wish I had a good ML algorithm to predict it.

There is no doubt that using AI, in particular ML, to automate tasks and decisions will contribute to making public administration and service delivery more effective and efficient. Digital technologies are already currently playing a prominent role in shaping up and regulating the behaviors, performances, and standards of our world. The technology has already changed leadership and strategies, employee relations and trust, organizational change, and management science.

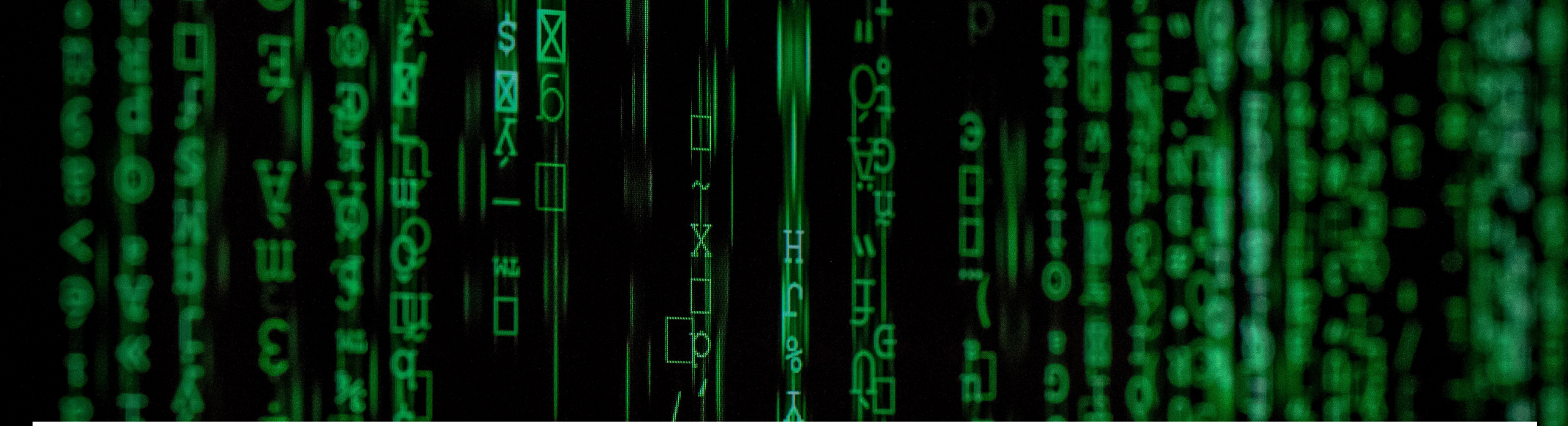
In this socio-technical landscape, the transformation brought about by AI, in fact by information technology in the public sector (Digital Era Governance, Digital Transformation, or e-Government), may attenuate or make disappear the respective limits between the public policy development and service delivery domains. If so, public policy is and will be gradually integrated into a service delivery logic, figured out by computational logic and by a techno-solutionism approach, often driven by practitioners who believe that their own technoscientific expertise is particularly relevant to the identified social problem. Indeed, the terms public policy engineering, computational public policy, political engineering, and computational politics, based on the application of engineering, computer science, mathematics, or natural science to solving problems in public policy. Is that a desirable outcome? The concern is like those aimed at the popular New Public Management approaches at the end of the last century. Some will say that, on the contrary, IT and AI allow for a better focus and precision on public policy formulation away from parochial interests and would contribute to service of public-spirited goals. I do not have an answer, and if I have one, it is incomplete. The questions are imperative.

TARA HOLLAND

Principal, Global Government Practice, SAS

In the context of government mandates, it is important to recognize that the complexity and scope of the business challenges are much greater than for most private sector organizations. To address significant challenges like homelessness, climate change, pandemic response or economic recovery requires a coordination of the legislative, regulatory and policy frameworks across a larger number of stakeholders. The socio-technical complexity for management and decision making is multiplied in this context.

We are seeing those efforts to apply advanced analytics and AI to these large problems are acting as a catalyst to break down the silos within and between government organizations and across jurisdictions. As governments bring together diverse data, technology and domain expertise in governed and open platforms, these ecosystems are driving unique innovation. The opportunity for all stakeholders to become simultaneously more collaborative and autonomous will drive better outcomes for communities.



v

Data Validation in Public Sector Human Resource Systems

By Sunil Meharia & Betty Ann M. Turpin, Ph.D

With the rapid onset of data analytics and artificial intelligence within the Canadian Public Sectors (Federal, Provincial, Territorial), organizations are scrambling to ramp-up their data processes, digital systems, and human resource (HR) complement and competencies. The latter has also given rise to a very competitive market for data analyst/scientist talent. It has also led to a transformation of HR management (HRM) in order to meet the growing demands of competitive hiring and provide fast, accurate data turnaround to support informed decision-making related to any aspect of the employee lifecycle.

HRM spans the continuum of the employee lifecycle (Figure 1) from hiring to departure, and is governed by several legislative and regulatory frameworks at all levels of government. With the ever-increasing need to better understand, predict and respond to workforce changes and requirements, it is paramount to understand how people move through their careers and identify effective methods of learning and the intricate relationships between people-related factors (e.g., employment equity, tenure, gender, sick leave, and multifactorial combinations of factors). This has reinforced the need for HRM to move from being transactional to a strategic decision-making partner, where HRM uses data to make evidence-based decisions and takes a proactive role in advising organizational business sectors on HR matters. One key challenge is the multiple sources of data (internal and external) that HRM needs to access. Most of these data sources are held by external GoC sources (e.g., PSPC), but some are based on the department/agency's Peoplesoft data base (MYGCHR) (Figure 2). Several of these sources provide dashboards and tables related to the employee lifecycle.

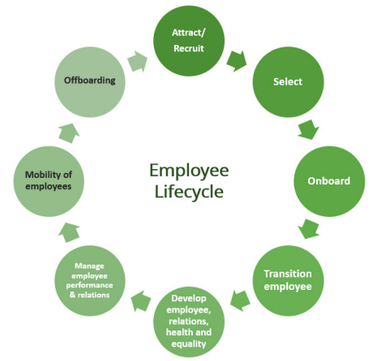


Figure 1: Employee Lifecycle

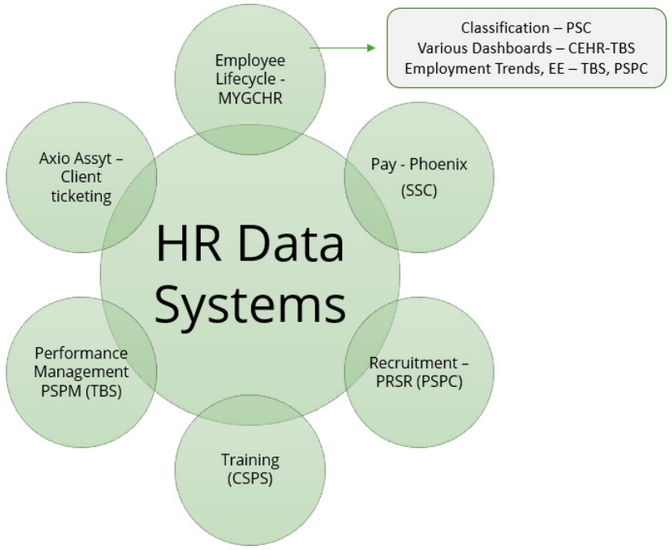


Figure 2: GoC HR Data Sources

Evidence-based decision-making necessarily relies on credible1 data[1]. Data analytics means carefully and logically thinking about the decision(s) that is required; systematically gathering, cleaning, and validating the data; interpreting the data, and reporting on the interpreted information to support informed decisions. The analytics process entails a typical data pipeline and should be the same process, regardless of the program or service area, and

1 Credible data evidence is derived from sound, accurate, and fair assessments of programs and must address: probity, context, reality, quality, integrity – thus it yields the qvuality of being believable or trustworthy.

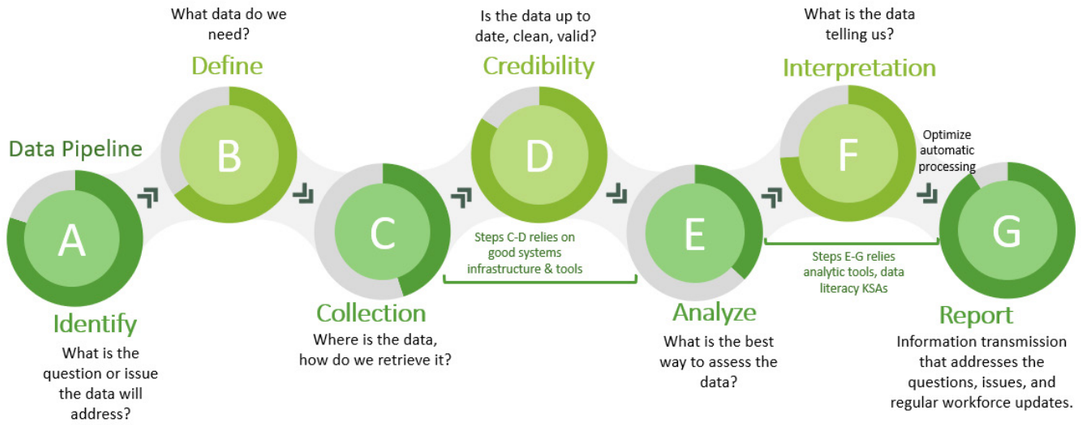


Figure 3v: HR Data Processing Pipeline

entails a typical data pipeline model (Figure 3).

HRM, as the custodian of the organization's HR data, is responsible for its accuracy and usage and is in a unique position to contribute to the organization's vision. To optimize data use, data management and analytical capacities must be enhanced. It will also require enhancing and developing new technologies, tools, policies and procedures, and re-engineering business processes.

In most Public Sector (PS) organizations, while HR units perform analytics, the processes involve a lot of manual workloads. Data entry within a PS organization is done by HR staff, employees, and managers, and as such data errors and/or untidy data[2] occurs. Untidy data requires cleaning and correcting which can be manually laborious and very time-consuming. This is particularly relevant when separate data sets are integrated. For example, currently, within the Canadian Federal PS, there are nine fragmented HR systems that departments/agencies extract data from, and sometimes the data in various systems are not provided in real-time, or methods of analysis differ from what the department/agency requires.

Untidy data can result in issues during the reconciliation of data between different HR systems or within the HRIS. The Audit of Human Resource Data Integrity[3] was conducted as part of Correctional Service Canada's (CSC) Internal Audit Branch (IAB) in 2011. Since some data fields were not accurate, the Salary Management System could not be reconciled with the HRMS. One of the instances where this situation occurred was when an employee was funded by one cost center but was actually working in another functional area. This impacted all downstream HR metrics and often resulted in inaccurate KPIs.

Data validation seeks to uncover data errors or untidy data. It is an automated process that enables the user to check the data for accuracy, completeness, and formatting. This process ensures that the database maintains "clean" data. Data validation should occur at every step along the employee lifecycle.

Due to restrictions or lack of inter-portability of data within various HR Systems, data collected in one system often may not be transmittable to other systems. It is not advisable to manually input the same data across different systems as it increases the potential for data entry errors and redundancy. However, this is often done because the means (e.g., programming, APIs, AI) to automatically integrate the data is lacking.

To overcome this challenge, HRM must have the ability to identify different data quality issues with a validation process that is automated, systematic, and periodic. Every HRM business unit will have its own contextualized rules for how data should be stored and maintained. To be more efficient and effective with data usage, setting basic data validation rules is necessary. The most critical rules used in data validation are rules that ensure data integrity. Once established, data values and structures can be compared against these defined rules to verify if data fields are within the required quality parameters.

To achieve efficient validation tests, with actionable insights and easy to build logic, it is necessary to start at the most granular level of HR Data. Data formatting has been one of the biggest challenges when multiple HR Systems are involved. The main focus of automation of the HR Data validation process is to establish a logic between different HR data attributes (such as Hire date, Seniority Date, Job function, etc.) to clearly understand whether it is meeting the business expectation or not. For example, the termination date of an employee cannot be equal to/or prior to the hiring date. This level of automation can be performed using a scripting language to specify conditions for the validation process. An XML file can be created with source and target database names, table names, and data columns (data attributes) to compare. The validation script will then take this XML as an input and process the results.

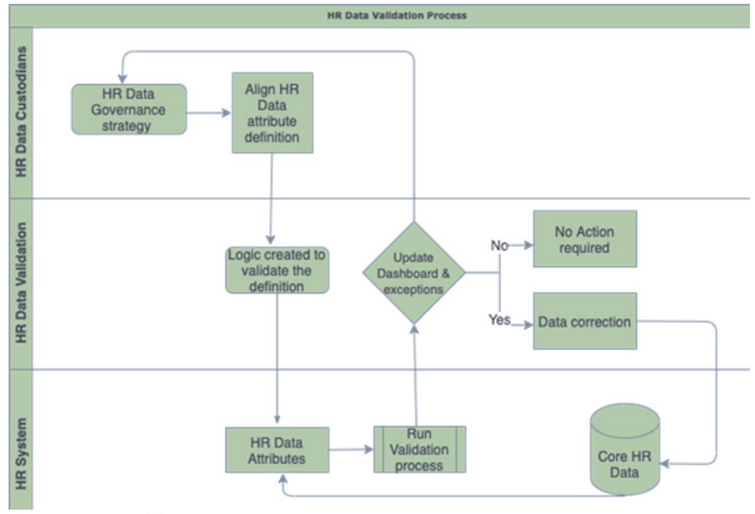


Figure 4: HR Workflow

Based on an HR Technology stack, workflows can be created specifically for data validation or this data validation can be added as a step within other HR data integration workflows (Figure 4). Once developed, these workflows can be reused or could be scheduled. The user gets notified through a validation report or a dashboard if data is invalid. This helps in tracking data integrity issues and helps in data correction, especially when data sets need to be merged/ combined.

In summary, automation of data validation ensures that there are no inconsistencies or errors (correct data), no missing fields where a value is expected (complete data) and aligned with corporate data definition (data compliance). Data validation logics should be continuously upgraded to support new data fields, any enhancement in data definition, and new data integrations.

About the Authors

Betty Ann M. Turpin, Ph.D., C.E., President of Turpin Consultants Inc., is a freelance management consultant, practicing for over 25 years, has also worked in the federal government, in healthcare institutions, and as a university lecturer.

Her career focus is performance measurement, data analytics, evaluation, and research. She is a certified evaluator and coach.

Sunil Meharia is a HR Analytics Manager with a deep interest in how analytics can transform Human Resources. Over 12 years+ of experience across the Public sector, large manufacturing, and research-based conglomerates, his primary focus has been on data analytics in human resources, business strategy, and HR transformation with multiple years of experience in implementing Oracle HCM solutions for Fortune 500 companies.

2 In untidy data, there will exist inconsistencies in the data file such as in variable names, observations are stored in columns when they should be in rows, dates (dd/mm/yy) are not in the same order, etc.

3 Audit of HR Data Integrity: https://www.csc-scc.gc.ca/005/007/005007-2509-eng.shtml#ftn12

Automated Semantic Analysis of IT Project Risks

By Franck-Olivier Kwan, Véronique Nabelsi, Stéphane Gagnon & Wassim El-Kass

LATEST SCIENCE

What are Automated Semantic Analysis (ASA), Ontologies, and Knowledge Graphs (KG)?

Automated Semantic Analysis (ASA) refers to the use of an "ontology" to annotate textual and other unstructured data. An ontology can be thought of as a way of categorizing objects and the relationships among them. When applied to managing IT project risks, this approach can help to automate or semi-automate the many unstructured relationships contained in project management documents. A key benefit for project managers is that the approach enables forecasting of project risks and proactive management of these risks as a project unfolds.

Think of this as going beyond the traditional "keywords indexing" approach, and instead creating a wide network of multiple-flexible relations between concepts in your project management framework. Akin to Object Oriented Modeling (which is in fact related to ontologies through model transformation standards), but without the complex constraints, and without concern for evolving information schema.

Ontologies are often mentioned when dealing with an increasingly popular NoSQL database platform, Knowledge Graphs (KG) Databases. Some well-known trademarks, such as Neo4j and OrientDB among many others, provide primarily "property graphs" that simplify the constraints on inter-concept relationships. A KG is therefore an ontology or a network of entities, normally defined as concepts or classes and their domain-specific instances, with relationships that define a domain-specific logic with a variety of constraints.

Breakthrough in Using ASA to Analyze IT Project Risks

The use of ontologies to describe IT project risks is still a major scientific challenge. IT projects are complex and representing their risks within a temporal context requires extensive practical experience. Risks can only be understood through in-depth knowledge of the variety of issues throughout the lifecycle of IT projects, from architecture to development, from implementation to end-user experience.

This challenge has successfully been overcome, thanks to a recent completed research project that may be of interest to several government agencies seeking innovative solutions for IT project risk analytics.

First, a new IT project risk ontology was proposed that differentiates clearly between risk occurrences and risk mitigation[1]. It is modelled after PMI's PMBOK risk

1 Kwan, F.-O. (2021). Development and implementation of an ontology for the automated semantic analysis of IT project risks. D.B.A. STI, Université du Québec en Outaouais, Gatineau.

management processes and represents a whole breath of academic and practical knowledge found in the literature. As such, this new IT project risk ontology serves as a first step to help extend the knowledge base in the future.

Second, we have demonstrated the feasibility of using an ASA platform entitled Adaptive Rules-Driven Architecture for Knowledge Extraction (ARDAKE)[2], relying on its ontology-driven annotations to serve as a risk management tool in IT projects. We were able to automatically annotate all artifacts of a project, indicating events related to various risks. Annotated information can be found in several types of project documents, such as project status reports, lessons learned document, business case, and meeting agendas. This diversity in the annotation of project documents seems to provide a perspective that is both direct (i.e., project reports, meeting agendas) and overall risk (cost benefit analysis, lessons learned).

Third, we then used a semantic analysis tool to establish potential links between various events and risks, and thus identify the precursors to trace their occurrence. The annotated dates of the project documents make it possible to create a chronology of events and information to support the dated actions of the mitigation strategies of the risk register. Between the annotated dates of the first events and those of the mitigation strategies, it is possible to estimate a lag ranging from a few days to several months. A risk emergence pattern may be found when the dates of the first annotated sentences were earlier than the dates of the mitigation strategy sentence.

Seeking Government Agencies for Further Pilot Studies

Our findings show the potential of ASA tools to help IT project risk analytics. To advance this research, we need partners to help develop a more systematic research process. This research breakthrough used the detailed PM data from an actual, completed IT project. However, we believe we could radically improve IT project risk forecasting by relying on more detailed and diverse datasets.

To broaden the scope of IT and Business (digital) risks we can forecast and monitor, we can reuse the Business Technology Management Body of Knowledge (BTM BOK). Started recently in v.0.1, BTM BOK will become an integrated reference of 60+ open-source professional standards[3]. Once the BTM BOK ontology is created from integrated standards, it will be possible for any ontology-driven tool to rely on its specifications for IT roles, skills, tasks, processes, artefacts, and outcomes to control risks within complex IT projects.

Government agencies are invited to contact Stéphane Gagnon to develop more pilot projects. We can work with ongoing and completed project datasets, and especially compare IT project risk monitoring from both traditional (manual) methods and more automated analytics features in requirements-to-test tools (e.g., JIRA).

Further Readings about Ontologies

As defined by Studer et al.[4]: "An ontology is a formal, explicit specification of a shared conceptualization". Various standards exist to represent ontologies, but in this research, the object of study is only the W3C's Web Ontology Language (OWL). Ontologies can contain a thesaurus, glossary, and taxonomy characteristics. It is possible to create multiple ontologies to reflect specific areas of knowledge. For example, research has created and linked three ontologies (software, bug, and version) in a study of information systems to represent various aspects that generate source code[5]. In addition, ontology or a body of knowledge represents elements of a domain of knowledge that can be used in the labeling process to extract various subjects [6].

Ontologies are valuable tools to help analyze IT project risks. In 2009, a risk ontology was developed for IT projects post-implementation of an integrated management software package. Practitioners can use this detailed ontology in identifying risks in this type of software.[7] More recently, software that uses a risk ontology has allowed practitioners to know the risks and risk management practices according to project phases[8]. These authors also add that ontology engineers must be able to extract information from various knowledge bases to update the concepts of the ontology.

About the Authors

Franck-Olivier Kwan is Manager of the Case Management Competency Centre at Public Services and Procurement Canada (PSPC). He obtained his DBA in Project Management in 2021 at Université du Québec en Outaouais (UQO). His thesis used the Adaptive Rules-Driven Architecture for Knowledge Extraction (ARDAKE) platform to demonstrate how an IT Risk Ontology can be used to detect early warning signs of emerging risks.

Véronique Nabelsi is Full Professor in Healthcare IT and Chair of the DBA in Project Management at Université du Québec en Outaouais (UQO). She is also Scientific Director of the Psychosocial Medicine Research Center and the Primary Healthcare Research Group at the Centre Intégré de santé et de services sociaux de l'Outaouais (CISSSO). Her research focuses on the management of healthcare establishments, including the integration of medical information systems and processes, in an emerging strategic framework of activity-based funding.

Stéphane Gagnon is Associate Professor in Business Technology Management (BTM) at the Université du Québec en Outaouais (UQO) and founding member of the Government Analytics Research Institute (GARI). His research deals with digital transformation and big data analytics, applied to the healthcare, energy, and financial services industries, as well as public administration.

Wassim El-Kass is Assistant Director at the Advanced Analytics Solutions Centre at Immigration, Refugees and Citizenship Canada (IRCC). He obtained his Ph.D. in Information Science and Technology (IST) in 2018 at Université du Québec en Outaouais (UQO). He developed an ontology-driven text annotation platform entitled Adaptive Rules-Driven Architecture for Knowledge Extraction (ARDAKE).

2 El-Kass, W. (2018). Integrating semantic web and unstructured information processing environments: a visual rule-based approach. Ph.D. STI, Université du Québec en Outaouais, Gatineau. https://gagnontech.org/ardake

3 Gagnon, S. (2020). Business Technology Management as Transdisciplinary IS-IT Competency Framework. ICIS 2020 Proceedings. Presented at the International Conference on Information Systems (ICIS), Hyderabad, India: AIS. https://aisel.aisnet.org/icis2020/digital_learning_env/digital_learning_env/8/

4 Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. Data & knowledge engineering, 25(1-2), 161-197.

5 Tappolet, J., Kiefer, C., & Bernstein, A. (2010). Semantic web enabled software analysis. Web Semantics: Science, Services and Agents on the World Wide Web, 8(2), 225-240.

6 Hindle, A., Ernst, N. A., Godfrey, M. W., & Mylopoulos, J. (2011). Automated topic naming to support cross-project analysis of software maintenance activities. MSR '11: Proceedings of the 8th Working Conference on Mining Software Repositories, May 21, 2011. ACM Proceedings Series: 163-172

7 Peng, G. C., & Nunes, M. B. (2009). Surfacing ERP exploitation risks through a risk ontology. Industrial Management & Data Systems, 109 (7), 926-942.

8 Abioye, T. E., Arogundade, O. T., Misra, S., Akinwale, A. T., & Adeniran, O. J. (2020). Toward ontology-based risk management framework for software projects: An empirical study. Journal of Software: Evolution and Process, 32(12), e2269.

ANALYTICS IN GOVERNMENT QUARTERLY