



January 31, 2019

# PHA Review

Prepared For: Department of Environment

## PHA Review

|       |   |    |
|-------|---|----|
| 1     | Executive Summary .....                                       | 4  |
| 2     | Background .....  | 5  |
| 2.1   | Prior to Going Online .....                                   | 6  |
| 2.2   | Moving Online .....   | 6  |
| 3     | What Went Wrong.....  | 6  |
| 4     | Contributing Factors .....                                    | 7  |
| 4.1   | Changes In Procedures .....                                   | 7  |
| 4.2   | Lack of Process Governance .....                              | 8  |
| 4.3   | Poor Communication .....                                      | 8  |
| 4.4   | Staff Attrition .....   | 9  |
| 4.5   | Results Not Reviewed .....                                    | 9  |
| 4.6   | Rushed Implementation and Incomplete Data Conversion .....    | 9  |
| 4.7   | Inadequate Validation for Duplicate Client Profiles .....     | 9  |
| 4.8   | Inadequate and Inappropriate Identity Management .....        | 10 |
| 4.9   | Lack of Development Methodology/Processes/Technology .....    | 10 |
| 4.10  | Running the Lottery Requires Specialised Skills .....         | 10 |
| 4.11  | Fragility of the PHA Process.....                             | 10 |
| 4.12  | Magnitude of the Errors .....                                 | 11 |
| 5     | Recommendations.....  | 11 |
| 5.1   | General Long-Term Recommendations.....                        | 11 |
| 5.1.1 | Project Execution .....                                       | 12 |
| 5.1.2 | Systems Planning .....  | 12 |
| 5.1.3 | Technical Tools .....   | 12 |
| 5.1.4 | ID Management and a Yukon Person ID.....                      | 13 |
| 5.1.5 | Create a PHA Draw Function .....                              | 13 |
| 5.1.6 | Create a Returns and Reassign Function .....                  | 13 |
| 5.2   | Short-Term Recommendations .....                              | 13 |
| 5.2.1 | Appoint a Clear Owner to Govern the Process.....              | 13 |
| 5.2.2 | Implement Proper Document and Data Management.....            | 14 |
| 5.2.3 | Data Cleanup .....  | 14 |
| 5.2.4 | Proactive Registration Invite.....                            | 14 |
| 5.2.5 | Explicitly Invite PHA Applicants to Review Their History..... | 14 |

## PHA Review

|       |   |    |
|-------|---|----|
| 5.2.6 | Review the Results.....   | 14 |
| 5.2.7 | Data Handling .....   | 15 |
| 5.2.8 | Lockdown the Lottery Code.....                                  | 15 |
| 6     | Fixes Already in Progress as of December, 2018.....             | 15 |
| 7     | Summary.....  | 15 |
| 7.1   | The primary causes for the failure of the PHA Lottery were..... | 16 |
| 7.2   | Key Recommendations .....                                       | 16 |
| 7.2.1 | Long Term .....   | 16 |
| 7.2.2 | Short Term .....  | 16 |
| 8     | Appendix A - Weight Algorithm.....                              | 18 |
| 8.1   | Power of 7.....   | 18 |
| 8.2   | Randomness .....  | 18 |
| 9     | Appendix B - As-is System Diagram .....                         | 20 |
| 10    | Appendix C - Future State Diagram .....                         | 21 |
| 11    | Appendix D - Interviews.....                                    | 22 |
| 12    | Appendix E – Acronyms and Definitions.....                      | 22 |

# 1 Executive Summary

The Yukon Government's Department of Environment runs a permitted hunt lottery which results in Permit Hunt Authorizations (PHAs). The process involves hunters applying annually for specific species/hunting areas through a lottery process that favours those that have not been drawn in previous years. It can take many years for a hunter to win a permit so tracking the correct names and draw history is critical.

For many years the application process was based upon paper applications and a small technical team that carefully validated the information. That team used a set of technical tools (i.e. a toolbox) to enter and manage the data and to perform the electronic lottery. It was a semi-automated system, but not a comprehensive and integrated computer system that could run seamlessly without key technical resources being involved.

A new online computer system was built to manage the process which allowed hunters to create user profiles online and then apply for their permits. In 2017, hunters were required to apply using the new system, either directly or in person while front desk staff entered their data. This new online system did not entirely replace the aforementioned semi-manual solution. The previous application and lottery history data remained in the old system so both the new and old systems were required to run the lottery.

To execute the lottery, the profiles and application information were extracted from the online system and combined with the historical application and draw data from the old system - see Appendix B for a diagram. Due to a number of issues, some of the online profiles were duplicated and in a number of cases profiles didn't match the names in the old system.

The PHA lottery system was prone to failure due to a lack of data cleansing and system/data integration between the new Posse data and historical data from the old SAS system. In addition, the departure of key technical employees that understood the systems left the fragile, unintegrated solution exposed.

The failure in the draw was due to manual steps being missed and poor management of data. During the first draw in 2018, an older version of the data that didn't include the previous year's permit returns and reissuances was accidentally used. The second draw problems came from names not matching between the new online system and the original paper and SAS based system.

Due to staff attrition the knowledge required to prepare the data and run the lottery was not available. When the department engaged the Bureau of Statistics to run the lottery the data was not integrated or completely cleaned, and there was no one to liaise with YBS to provide direction, or ownership of the process. Before the Bureau ran the lottery they fixed some obvious profile matching issues (name, date of birth etc.) but did not fix all of them. Due to communication failures between YBS and Environment, the Department of Environment was not aware of the problems and the flawed lottery results were released to the public.

Adopting an integrated e-services based solution that involves citizens creating unique online profiles is not simple. Problems such as mistyped names, name changes, lost login IDs etc. can easily mount. Duplicate profiles, lost login IDs and mismatching names between the two systems are evidence of this.

## PHA Review

Building this type of solution requires higher level technical skills, tools and processes than were required to build the original semi-automated system. In addition, having redundant resources is critical. Environment used in-house resources to build the original and online solutions. Unfortunately, the more advanced technical resources required to maintain and operate the systems are no longer in the department. The Government of Yukon typically outsources software development work rather than building technical solutions of this size and complexity in-house.

It is recommended that the department move toward a single integrated system with corrected hunter profile/application information. In addition, the department should closely collaborate with a government-wide approach on expanding e-services. It is also recommended that the department use more modern tools that are supported by the industry and follow more rigorous systems development processes. Furthermore, an appropriate support model for the system should be put in place.

There were 124 clients with name matching issues between the current applications and the historical data. Yukon Bureau of Stats corrected 29 of them before doing the run, leaving 95 clients with a total of 223 applications that had name matching errors.

We have recommended both long term improvements to processes, technology and contracting methods as well as short term steps to improve the draw for the coming hunting season. These recommendations, combined with initiatives the department has already undertaken to clean the data and improve processes, should ensure that the hunting permit application and PHA lotteries run smoothly in the future.

## 2 Background

The PHA (Permit Hunt Authorization) is a computerised lottery used by the Department of Environment to issue hunting permits for specific zones and species. 1,158 applicants submitted 4,311 applications for the available 247 permits in 2018.

The system is weighted such that for every year a hunter does not receive a permit for a specific animal, the weighting goes up, to increase their odds of winning. A first time applicant has a weighting of 1. A hunter with four unsuccessful attempts at receiving a permit will have a weighting of five, since they are applying for the 5<sup>th</sup> time. A hunter who received a permit the previous year is moved to the bottom of the priority list.

Hunters have the option to return their permits if drawn in a given year. In that case the hunter next in line is given the permit and the hunter who drew the permit is set as if they didn't win, and their weighting is increased by 1 for next year's lottery.

The weighting is then raised to the power of seven, and that becomes the number of 'lottery tickets' the hunter has submitted into the computerised lottery system: i.e. a weighting of 1 gives the hunter 1<sup>7</sup> chances in the lottery, or 1 chance. A weighting of 5 gives the hunter 5<sup>7</sup> chances in the lottery, or 78,125.

For the lottery to be fair, it is critical for the hunter's PHA application history to be correct, because that determines their weighting in the draw.

## 2.1 Prior to Going Online

The original PHA process involved paper applications and a set of SAS Datasets and a few computer “scripts” (written in the SAS language) to manage the lottery. The “system” was operated by two people that updated and operated the system as needed and manually entered the paper applications into the system. They ensured, for example, that names on the applications matched previous applications/permits exactly. A SAS program was used to conduct the lottery as per the rules of lottery. The system involved a lot of manual data verification, correction and handling of data. The process, although not fully automated or robust, appeared to have worked well for almost 30 years due to a small number of people governing the system, as well as much smaller numbers of applicants.

## 2.2 Moving Online

The SAS system, however, was not capable of supporting e-services, so a new online system called Posse was introduced. It started as a system for supporting the online purchase of campground permits and angling licenses in 2015, and then hunting licences, hunting permits and harvest reporting were added. In 2017 it became mandatory to apply for hunting permits online using the new system. Client profiles were loaded from the old SAS system and merged in with the existing client profiles already in the system from angling licenses and small game hunting permits. The profiles were assigned an “Environment ID” - a unique ID that is intended to identify hunters across application years which was also used to log into the e-service. New client profiles were also created during the 2018 application processes, which included their name, date of birth and address information. Historical permit applications were not moved to the new system, and since the old system did not contain an Environment ID, there was no easy way to map the old permit applications to the new client profiles other than by exact name and date of birth matching.

To run the lottery, the profile/application information is extracted from the online system where it is merged with the previous application/permit data from the old SAS system, using name and date of birth to match the data. The lottery is run using a SAS program that is external to both systems.

## 3 What Went Wrong

For the weighting to work correctly, the current applications need to match with the historical results of the lottery. The more years an applicant was unsuccessful, the more chances they are given of winning the lottery. Data needed to be cleaned so that the applicants matched their historical records to ensure a correct weighting.

For the 2018 draw there were 2 draws made. Both had incorrect results.

**The First Draw: Data Management Issues.** The first draw was incorrect because the wrong data was used. The data were stored in SAS datasets and Excel spreadsheets on a file server. The returns and reissues of the hunting permits is a very manual process requiring the manual editing of the spreadsheet with the lottery results. This spreadsheet is then saved with a different name. Naming conventions were not clear, and staff mistakenly chose the older 2017 excel spreadsheet without the returns and reissues to use for the lottery. Because of this, hunters who received a reissued hunting permit in 2017 would still have a high priority in the 2018 draw, and hunters who returned their permits in 2017 would have no chance to win the lottery, even though they should have a very high priority.

## PHA Review

Storing and handling data in this way is risky and prone to errors. This risk was compounded by the retirement of the key person responsible for the data management.

Note that the name mismatch issues of the second lottery were also present in the first lottery, but this was not apparent because the problem with choosing the wrong file, which was more noticeable.

**The Second Draw. Data Integrity and Duplicate Client issues.** In 2018 there were a number of hunters whose application history was not merged correctly with their current application, resulting in their weighting being dropped to a 1 or 2 (since the system showed no history, or at most, one year of history.) The data didn't merge well because the historical data came from the old SAS system, and the current applications were entered into the newer Posse System. The lack of an identifying key between the old and new systems meant that merging data between the two systems was done by name and date of birth, which is not accurate enough.

The system did not have a robust way of determining if a new client being entered was in fact a new client, or a duplicate of a client already in the system. It was easy for the hunters to create a new client profile in the system, but to log in as an existing client required the applicant to use their Environment ID. Many hunters don't know their Environment ID's since they are only used infrequently for their annual hunting, angling and camping permit applications, so many did not know their ID's when the 2018 applications were being entered. So instead of logging in to existing accounts, hunters would create new accounts, resulting in many duplicate clients being entered into the system. The duplicate clients would not have any history attached to them, resulting in a weighting of 1 for the lottery.

For the client to match with their history, they would have to create their profile name exactly the same way it was stored in the old SAS system. In 223 cases, this did not happen. For example, the name could be John A Macdonald in one system and John Macdonald in the other system, and then they would not match. Typically the difference in the names between the new applications and the historical data was simply a missing middle initial or a variation on first name like Mike instead of Michael. There were also a small number of cases where birth dates had transposed values.

## 4 Contributing Factors

There were a number of issues causing the errors in the 2018 PHA lotteries. These issues can be mitigated, significantly reducing the risk of the same problems happening again. In fact, there seemed to be a convergence of a lot of different, unrelated circumstances that caused these errors.

### 4.1 Changes In Procedures

There were two significant changes in procedures for the PHA draw, which contributed to the errors in the draw process.

## PHA Review

**Online applications.** Starting in 2017 the application process switched from paper forms being keyed into a simple SAS system to online applications into a much more complex and comprehensive system. The new, online system stores the applicant data in an Oracle Database. Applicants apply online, or come in to an office and have a clerk enter the data directly. This meant that the system switched from having one or two people who were very familiar with the system and the data, to having hundreds of people (the clients and front desk staff) who were doing data entry. This resulted in less rigorous testing for correct data and duplicate data entry.

**Contracting out the PHA lottery.** Historically, the Department of Environment staff ran the SAS program that conducted the PHA lottery. In 2018 the Yukon Bureau of Statistics (YBS) was used to run the SAS programs because of their expertise in the SAS technology, and to add increased transparency to the process.

Application data was extracted from the new system, and was sent to YBS, along with a spreadsheet containing the historical data to have them run the SAS programs that conducted the lottery. YBS was contracted to run the lottery, but not to verify or clean the data. The resource that would normally act as a liaison with YBS had retired earlier that year, leaving no expert user to aid YBS in understanding the process. YBS, although very competent and rigorous in their methodology, do not know the PHA process, are not familiar with the data, and would not understand the significance of the results or if the wrong name was drawn.

## 4.2 Lack of Process Governance

The PHA lottery process falls under the jurisdiction of the Fish and Wildlife Branch, and was administered by that branch in the past. At some point the position that oversaw the SAS systems that managed the hunting permits, harvest reporting and the PHA Lottery was transferred to the IMT (Information Management and Technology) Branch. When the person occupying the position retired in 2018, no clear owner of the PHA process remained. Time pressure from the upcoming deadlines of the hunting season caused IMT to get the process started, but IMT's role should be one of technical support and data management, while the ownership of the process should belong to the Fish and Wildlife Branch, under whose jurisdiction this process falls. The business owner should be responsible to get the process started, and manage the process from start to finish, while being aware of deadlines, to ensure that there is enough time to go through the steps of the permitting process, ensure resources are in place, and verify the results.

## 4.3 Poor Communication

Environment staff were not aware of the magnitude of the issues with the duplicate client profiles and the problems with the online profiles not matching the historical data. The procedures for conducting the lottery were not well documented, and due to staff attrition with a lack of succession planning, the department was not aware of all of the steps required to conduct the lottery, especially when preparing and cleaning the data.

Yukon Bureau of Statistics (YBS) found approximately 124 clients with errors in their data matching between Posse and the old SAS historical data, and they fixed 29 of them before conducting the run. This left 95 known clients with 223 applications with data matching errors. Environment was not aware of these errors, so the results were released, with Environment staff thinking the run was correct.

### 4.4 Staff Attrition

The key resource responsible for the data management and running the lottery scripts retired earlier in the year with no succession planning. No one else understood the old data, the old system, or the PHA draw process. Given the number of manual steps in the process and the fragility of the sequence of steps, this was a significant risk to the PHA draw.

### 4.5 Results Not Reviewed

Normally the results of the lottery are reviewed by the Conservation Officers (CO's) and Fish and Wildlife staff to ensure they make sense. In 2017 a decision was made due to privacy concerns, to not have the results reviewed. Because the staff are quite familiar with the hunters and their application successes, there was a good chance they would have noticed that something was wrong in the allocation of the permits.

### 4.6 Rushed Implementation and Incomplete Data Conversion

The hunting licence module of the online Posse system was implemented before it was ready. Given the importance of historical data in this system, the historical permit data should have been loaded and tested before the system went live. For the lottery to function, the current permit application needs to be matched to the historical data through a common client profile. The historical permit applications had not been loaded into Posse. This required the lottery process to manually merge the historical data to the current applications using an inadequate identifier (i.e. name and date of birth and not the "Environment ID").

The client profile base was incomplete, with many duplicate records. Client profiles in Posse had been created since 2015, for angling licenses and campground permits, and the client profiles from the SAS system were loaded into the system as a bulk load in 2017, creating many duplicate records. There wasn't adequate time to clean up the duplicates before the 2018 application process.

### 4.7 Inadequate Validation for Duplicate Client Profiles

In 2017 it was mandated that permit applications had to be done through the online (web) interface built in Posse. Some 2018 applicants had difficulty logging into the new system, since they didn't know their ID's, so instead of logging into an existing profile, they created new client profiles when applying for hunting licenses and permits.

The system has inadequate checks in place to warn the client of a potential duplicate client record. It would only flag a potential duplicate if the name and date of birth matched exactly. So if the client typed in Rob instead of Robert or Rob A instead of Rob, the system would not warn of a potential duplicate, even if the birthdate and last name matched exactly. This flaw in the system would have been discovered with adequate application testing.

Upon entering a new client record, the client was required to visit the office in person to validate their residency, so there was the opportunity for the front desk staff to check for duplicate records before validating the application. But the staff did not have adequate training in the system, or clear procedures for determining if the client record was a

## PHA Review

duplicate, and the procedure for validating the application does not alert the staff of a potential for a duplicate record.

This added more duplicate records, which did not link the applications to the historical data, resulting in many “orphaned” history records and incorrect weightings.

### 4.8 Inadequate and Inappropriate Identity Management

As evidenced above, managing profiles and logins is not easy. Not only did people forget, lose or have duplicate “Environment IDs” the problem will grow significantly if new departments create their own online IDs and the citizens will become very frustrated.

Solving the general problem of identity in order to provide e-services to citizens is a corporate problem that Yukon Government needs to solve so that individual departments are not tasked with managing their profiles, user ID, logins and related privacy issues. See Yukon ID below.

### 4.9 Lack of Development Methodology/Processes/Technology

The system was customized in-house, with weak methodology/processes. There was inadequate user input, feedback and testing, inadequate architectural planning, and inadequate data conversion and implementation strategy.

- The historical data should have been loaded into the system prior to go-live
- Critical functions like duplicate client creation and data exports of key reports should have been tested to ensure they worked and met the needs of the users.
- There is no implementation of software version control for the original lottery system SAS code.
- Limited vendor resources were available so internal departmental staff took the technical aspect of the project upon itself.
- There should have been better collaboration with ICT in developing a solution, to help ensure the system fit within the government architecture and was supportable.
- The system was not an integrated solution. The Lottery Program exists outside of the system, requiring data exports, and technical resources to run it.

### 4.10 Running the Lottery Requires Specialised Skills

Running the lottery is not trivial. Two separate datasets need to be created with the correct format, and a SAS programmer needs to make modifications to the SAS code that merges the data and runs the lottery results. There are manual data corrections embedded in the SAS code to correct the 2018 data, because the programmer noticed errors in the data, and chose to fix them by modifying the lottery code.

### 4.11 Fragility of the PHA Process

The PHA lottery required two Excel extracts from two different systems to be handed off to a 3rd party (YBS) to run the programs. YBS loaded the Excel files into SAS to merge them together by name and date of birth, and then forwarded the results back to Environment. There were many manual handoffs of data between two different departments, and several import and export steps in handling the data. Every one of these steps has a risk of failure.

## PHA Review

This entire process has inadequate documentation, apart from comments in the code explaining what each component does, so the programmer running the code has to determine the process by reading the code.

The SAS code is basically a stand alone SAS program that is modified as needed by the current programmer. It may need to be run several times while data is cleaned and the code is tested, before the final run. In 2018 many lines of code were added to clean data as the program ran. There is no 'definitive' run, since it can be run multiple times, so the process is not auditable, and the fairness of the run could be questioned, as different results will be generated each time it is run.

### 4.12 Magnitude of the Errors

The lottery program matches applicants to the historical results by last name, first name, and date of birth. All three need to be an exact match or the historical data is not matched to the current applicant. Even a small difference between the current and historical data on any of the three matching fields would result in no match. Since there are applicants with no history, and historical records with no current applicants, there is no easy way of knowing if there were errors in the matching procedure, since there will always be unmatched records.

The majority of the errors in the data were due to the first names of the applicants being represented differently. Examples are Matt vs. Matthew, James vs. Jim, John A vs. John, or Mike A vs Mike a. When YBS examined the data, they noticed the obvious errors and made adjustments to the matching algorithm in the code to correct 29 of these errors, prior to lottery run. YBS was aware of 95 additional clients with a total of 223 applications that had name matching errors.

Note that there were also matching errors in the historical data. There were at least 28 clients who had different names in the historical data. This means that their history may not be reflected properly in terms of when they had their last success, which would affect their draw weighting.

There were also 3 clients who had errors in their birthdays due to misreading numbers (4's becoming 9's) or transpositions. Of these, one had 2 applications, and the other two only had historical applications.

## 5 Recommendations

The recommendations are broken into two levels: general long-term recommendations and short-term recommendations.

### 5.1 General Long-Term Recommendations

While these recommendations can start to be implemented now, they will take time to accomplish in their entirety.

### 5.1.1 Project Execution

The department should follow industry standard processes with respect to development methodology and documentation including quality assurance and release management and should utilize an experienced development team. During the initial development of the Posse system, the department followed a more ad hoc approach to development, but towards the latter half of 2018 has adopted a more industry accepted approach involving in-house resources for requirements gathering and testing and using an outsourced development team.

It also recommended that the department enter into support agreements with a vendor to ensure its solutions are properly maintained and operated.

### 5.1.2 Systems Planning

It is recommended that the department enter into a planning process to move its legacy systems into an integrated solution. All data and functions should be contained within the system, including the historical data and the lottery function, which are currently external to the system.

### 5.1.3 Technical Tools

Based upon interviews of technical resources we suggest that Posse is likely not appropriate for implementation of Yukon Government's e-services for the following reasons:

- Limited local resources - the department has no developer-level resources today and limited interest from the local resources to provide support.
- The toolset is not readily compatible with the Yukon Government's e-services standards based upon newer web standards (e.g. web-based architecture, Drupal, etc.) and integration with other databases and application services is not easy as demonstrated by trial projects within Yukon Government.
- The database structure is complex and not easily accessible by other applications
- Recently ICT has determined that Posse should not be expanded for use as a corporate solution going forward
- Posse is better designed for simple permits or back-end workflow but not ideal for larger general-purpose applications and specifically the PHA lottery.

SAS has been used by the department for decades, but there are no longer any departmental SAS resources available. It is suggested that any SAS expertise needed in the short term be contracted out to local vendors or the Bureau of Statistics. The department should have the SAS lottery code rewritten by professionals before the next lottery run so that it runs as an encapsulated system, rather than a series of code fragments that are run by a programmer. SAS is an adequate tool to run the lottery until an integrated solution is put into place that contains the lottery function.

Izenda is the reporting tool connected to Posse. It is a cumbersome tool that has several usability limitations. It is not used explicitly for the PHA lottery, but it is used to run reports to validate and examine the data for cleanup purposes. Eventually the department should move towards a data warehouse with better reporting tools.

It is highly recommended that the department work with ICT to move towards modern tools that are supported by the technology industry as a whole. The tools should be compatible with Yukon Government's standards and be web-based.

### 5.1.4 ID Management and a Yukon Person ID

The application matching would have been less of an issue if everyone had a single ID in the system and in the application history. The system assigns IDs to applicant profiles but has inadequate testing for duplicate creation. Having a reliable way to link the data would prevent the problems related to having to match data by name and date of birth.

There needs to be improved checks in the system for duplicate client creation.

The ICT branch has a project in the planning phase to implement a solution called Yukon ID, which is similar to the BC Services Card. They are also looking better user login/identity management solutions. Environment should adopt this solution once it is available and integrate it into their systems. This will make data matching and ID management much simpler and result in fewer duplicates, fewer errors, and better data management. This will become critical as more e-services are rolled out.

### 5.1.5 Create a PHA Draw Function

Once the Department of Environment sorts out which system will be their long-term solution for handling the permitting, they should add a PHA draw function to the system so that it is handled within the system. This eliminates any issues with data extraction, data handling and management, and running of SAS or other code by a 3<sup>rd</sup> party. This could also be handled by implementing a data warehouse and having this as a function in the reporting system that accesses the data warehouse. Either way, it should be a simple request for a report, rather than a lengthy and error prone process.

### 5.1.6 Create a Returns and Reassign Function

The Returns and Reassignment function for permits is completely manual, labour intensive, and carries a high risk of error, and was instrumental in the cause of the first lottery failure. When a permit is returned and reassigned, the draw result spreadsheet is edited manually with the changes. There are many opportunities for errors: editing the wrong record, being interrupted part way through the process, forgetting to save the results, editing the wrong spreadsheet, etc. Assignment and tracking of hunting permits is an important function and should be handled in a secure and robust way.

A function needs to be added to the system that handles the returns and reassignments in a trusted and reliable way.

## 5.2 Short-Term Recommendations

The department must conduct a draw for the 2019 season in early summer leaving it little time to do the planning, technologies upgrades and implement an identity solution in that time-frame. Within the context of its existing tools, it is recommended that the following be completed before the next lottery.

### 5.2.1 Appoint a Clear Owner to Govern the Process

The PHA process needs to have a clear owner, and it needs to reside within the Fish and Wildlife Branch under which the jurisdiction falls. They will rely on the IMT Branch for support, but they will be responsible for the governance of the process from start to finish.

### 5.2.2 Implement Proper Document and Data Management

The initial lottery draw was invalid because the wrong dataset was sent to be processed. Data were stored in SAS datasets and Excel spreadsheets on a file server. There were multiple versions of the datasets with an unclear naming convention. No one was responsible for the data management or version management, and no one knew which file had which data, so sending the wrong dataset to YBS was an easy mistake to make.

There is limited documentation on how to manage the data and the system, so the corporate knowledge resides with the staff. There should be a clearly defined and documented process, including how to handle the data, where to store files and what the naming conventions are.

The data needs to be managed in a relational database. There should be a trusted system with one version of the data that everyone agrees is the most current and accurate. Ideally, the data would be available in a data warehouse that would allow for easy access for reporting and running the permit lottery.

### 5.2.3 Data Cleanup

The department needs to ensure the history data is complete, accurate and properly assigned to the correct clients. Each client needs to have one and only one Environment ID and all current and historical applications with their success status needs to be associated to that ID. Processes need to be developed, documented and implemented to ensure that future applications use the same ID, and that duplicate clients are not created.

The history data needs to be loaded into the system, and linked to the client profile. Regardless of the system being used, having all of the data in one place will make migrating the data to the new system much easier.

### 5.2.4 Proactive Registration Invite

A week prior to the PHA registration opening up for the new year, the department should send out an email to all previously registered PHA applicants, inviting them to register for the upcoming year, and include their login information and permit application history. This will remind them to register, give them the information needed to make it easy, and allow them to review and comment on their history and weighting.

### 5.2.5 Explicitly Invite PHA Applicants to Review Their History

When the applicants register for the PHA, they should be presented with their history, and verify that it is correct, before they can continue with the registration process. This will ensure the applicants take ownership for the completeness and correctness of their data. This will help prevent the applicants from creating new profiles because they will be reminded that their history is important for their success in the draw. It will also help create trust in the process, and give them a sense of ownership of their data.

### 5.2.6 Review the Results

Review the results of the draw. The staff that work with the applicants are fairly familiar with the history of the hunters, and could notice the more obvious issues just by looking at the lists of successful applicants.

## PHA Review

The department should consider engaging in a 3<sup>rd</sup> party to certify the input data for the draw.

### 5.2.7 Data Handling

Ideally, data should not be transported back and forth between departments. The PHA lottery should be run on site accessing data from an integrated system's database.

### 5.2.8 Lockdown the Lottery Code

The lottery code needs to be cleaned up. There are irrelevant pieces of code, like the Bison lottery which should be disabled until needed. There is code that needs to be modified every year to build the profile of the number of permits issued by animal and zone. This should be stored in a database, or at least be done separately from the actual lottery. The lottery code should only contain the lottery code. Everything else should be stripped off, and if it's needed, be run separately. Irrelevant code should be removed.

The lottery run needs to be seen as fair, auditable and reproducible, and safe from tampering. The code should have a clearly defined data input structure, and the code should be locked in an unalterable, stored procedure. It should only be run once, so that the results are fair and agreed upon. If it is run multiple times to allow for changing of input date, and fixing of code, there is no definitive run and the results can come into question.

Once it is run, the random seed, and the input and output datasets should be archived, so the results are reproducible. Currently the lottery run cannot be recreated, because the input datasets are not exactly in the same state as they were during the 'official' lottery run.

## 6 Fixes Already in Progress as of December, 2018

The Department of Environment has already begun the process of improving the PHA process. They have an adequate system with a relational database to manage the data in the short term and have begun work on many of the steps needed to ensure the PHA lottery runs correctly in the future.

The Department has cleaned up the data, and assigned unique Environment IDs to all clients, and removed the duplicates. They have built a clean Excel spreadsheet of all the historical data which they plan to load into Posse by Environment ID.

They are working on implementing a process that requires the applicants to verify their history when they log in to the system, so that the applicants can validate their own history. This will help in preventing duplicate client creation.

## 7 Summary

The following summarises the findings and recommendations.

## 7.1 The primary causes for the failure of the PHA Lottery were

- Mismatched data between the 2018 Applications in the new Posse System, and the historical data from the old SAS system
- Poor communication of the errors in the data by Yukon Bureau of Statistics
- Attrition of a key resource with no succession planning
- Lack of quality assurance of the lottery results by conservation officers and Fish and Wildlife experts familiar with the hunting history of the applicants
- Implementation of a system that did not follow a standard development methodology. It was not adequately tested, and critical historical data was not loaded into the system.
- Inadequate functionality and processes to prevent duplicate client creation, and poor identity management
- A fragile lottery system written in SAS, and set of processes that requires specialized skills to run
- Poor data management practices. Multiple versions of the data stored with similar names on a file server

## 7.2 Key Recommendations

### 7.2.1 Long Term

- Move toward an outsourced technical model with long-term support agreements
- Adopt web-era technical tools that are attractive to the technology industry/developer community
- Work with ICT to sign on to a government-wide identity management initiative once it is available.
- Create a Returns and Reassignment Function so that the process is not the manual editing of a spreadsheet.

### 7.2.2 Short Term

- Assign an owner to the PHA Process
- Develop and document procedures in data management and client management and ensure there is adequate training of front line staff in using the system to effectively follow these procedures.
- Cleanup the data and load the historical data into an integrated solution. This is already underway, and needs to be completed before the next PHA lottery
- Proactively invite all previous applicants to apply for the PHA Lottery. Provide them the login information and their historical data, giving them an opportunity to validate their data
- Require that clients verify their history during the application process before they can complete their application. This is currently under development and will be available for the 2019 season.
- Use knowledgeable staff to review the results of the lottery (as they had done in the past)
- Implement better functionality to help prevent duplicate client creation. Ensure that the staff understand the procedures for checking for duplicate clients before authorising new clients
- Lockdown the lottery code so it is unalterable and archive the input data, output data, and random seed after the final lottery run so it is reproducible. The lottery

## PHA Review

code should be cleaned up to remove irrelevant code. Also, implement error checking reports and audit/archiving features

- Involve the Statistics Bureau in an audit of the new lottery code and involve them in the process of running the lottery to give more credibility to the PHA lottery process - but do so in an integrated technical environment (e.g. avoid passing around data and allowing them to modify the code)

## 8 Appendix A - Weight Algorithm

### 8.1 Power of 7

The random draw for the PHA works as follows. The number of years of unsuccessful applications plus 1, raised to the power of 7 is your weighting. So if you've had 4 unsuccessful draws, your weighting is 5 raised to the power of 7 or 78125. Meanwhile, if you've had 1 unsuccessful application, your weighting is 2 raised to the power of 7 or 128. With this method, you cannot calculate the probability of an applicant winning, but you can calculate the probability of one weighting being successful over another weighting.

As the table below shows, a weighting 1 will beat a weighting 2 less than 1% of the time. However, a weighting 6 will beat a weighting 7 over 25% of the time. Without the power of 7, a 1 would beat a 2 33% of the time, and a 6 would beat a 7 over 46 % of the time. More importantly, with the power of 7, the chance of a 1, 2, 3 beating a 7 is close to 0%. Without the power of 7, a 3 would beat a 7 30% of the time, and a 1 would beat a 7 12.5% of the time. Raising the weight by a power of 7 makes it much more likely for the higher numbers to be successful.

Probability of a lower weighting being successful over a higher weighting

Lower Weighting Result

|                         |   | 1    | 2    | 3     | 4     | 5     | 6     | 7     |
|-------------------------|---|------|------|-------|-------|-------|-------|-------|
| Higher Weighting Result | 1 |      |      |       |       |       |       |       |
|                         | 2 | 0.8% |      |       |       |       |       |       |
|                         | 3 | 0.0% | 5.5% |       |       |       |       |       |
|                         | 4 |      | 0.8% | 11.8% |       |       |       |       |
|                         | 5 |      | 0.2% | 2.7%  | 17.3% |       |       |       |
|                         | 6 |      |      | 0.8%  | 5.5%  | 21.8% |       |       |
|                         | 7 |      |      | 0.3%  | 2.0%  | 8.7%  | 25.4% |       |
|                         | 8 |      |      |       | 0.8%  | 3.6%  | 11.8% | 28.2% |
|                         | 9 |      |      |       | 0.3%  | 1.6%  | 5.5%  | 14.7% |

### 8.2 Randomness

The random algorithm was reviewed both by YBS and Make IT and has been determined that it is adequately random. MakeIT took it one step further and created a simulation to generate test data to run through the system. The results of test runs were as expected and consistent with the table above.

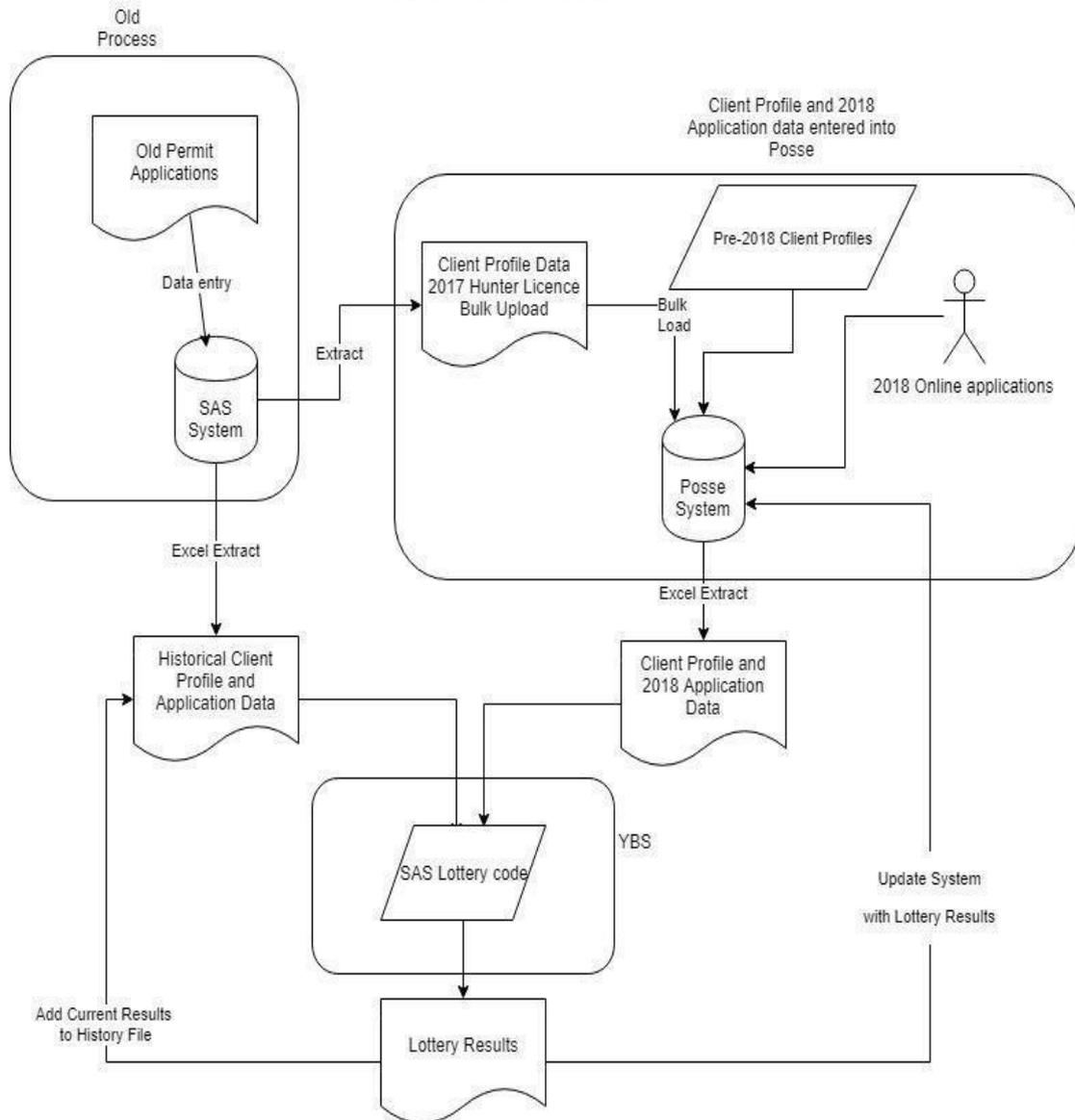
The main issue with it is that the seed used for the random number generator is hard coded in the system meaning that the program will produce the exact same sequence of numbers every time it runs. The advantage to doing this is that the run can be run multiple times with the same results, so if it needs to be rerun because of an error in the data, the actual results will be the same, if the data provided is in the same order.

## PHA Review

The disadvantage potentially is that if the order of the applicants is the same from year to year their sequence of numbers will start out the same way every time. This could be perceived as a bias. It may be better to use a new seed every year so that the randomness seems fairer. If the existing seed was multiplied by the year, that would give a reproducible seed that was different every year.

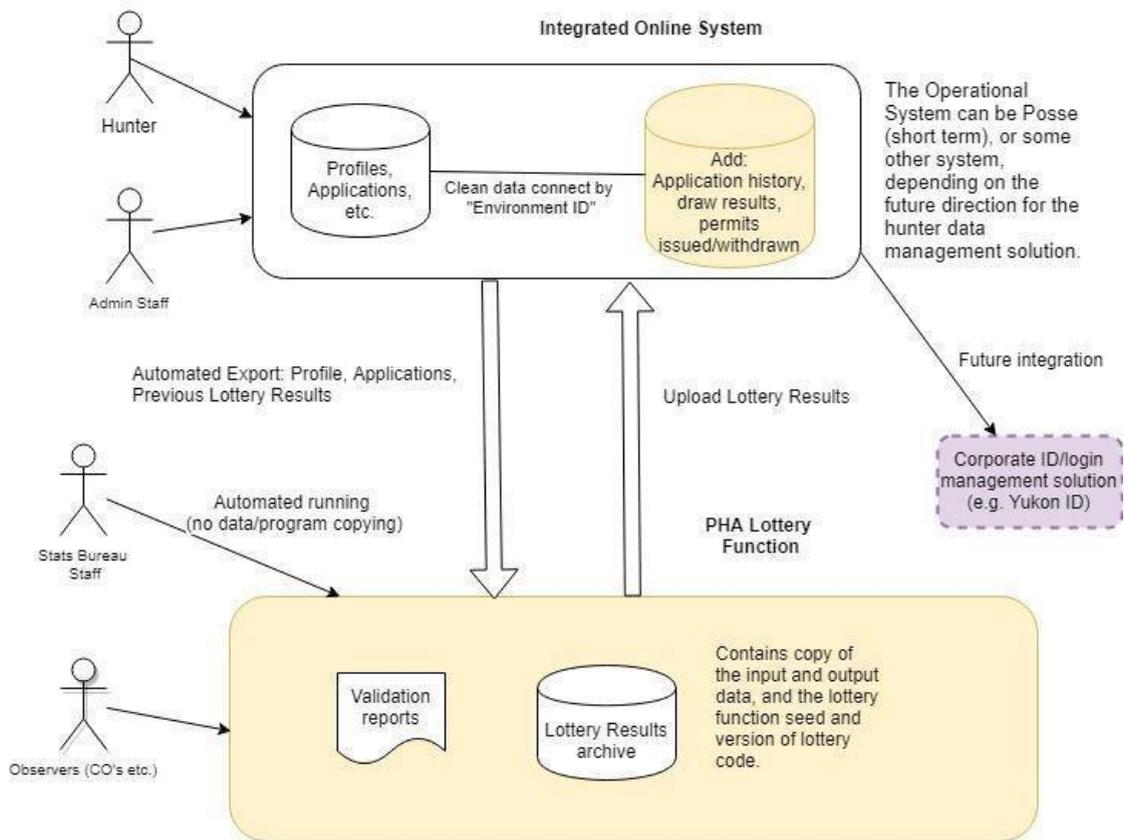
# 9 Appendix B - As-is System Diagram

PHA Lottery Process with Data sources



# 10 Appendix C - Future State Diagram

## Potential Short-term Future State



The lottery function should be a self contained black box that manages the data extract, runs the lottery and produces the results file, and updates the operational system.

The PHA lottery function can be contained within the system or maintained separately. For the short term, it will likely be maintained as a separate, SAS based system.

## 11 Appendix D - Interviews

### Interviewers

Chris Lane, Make IT Solutions

Mike Tribes, Sub-contractor to Make IT Solutions

### Department of Environment, Corporate Service and Climate Change

Director

Functional Analyst

Business Analyst

Operations Manager

Harvest Coordinator

former IT Manager/Developer

### Department of Finance, Yukon Bureau of Statistics

Director, Yukon Bureau of Statistics

Socio Economic Statistician

### Department of Highways and Public Works, Information and Communications Technology

CIO

Enterprise Architect

Director, E-Services for Citizens

## 12 Appendix E – Acronyms and Definitions

ICT – Information, Communication and Technology, the IT branch of the Yukon Government

PHA – Permit Hunt Authorization, a permit that allows a hunter into a specific area, assigned by lottery.

Posse – The system used by Dept of Environment to manage the Wildlife harvest, and hunting, angling and campground permits

SAS – Software used in the old system. Stores the historical data, and runs the lottery draw program.

YBS – Yukon Bureau of Statistics