DALHOUSIE UNIVERSITY

Inspiring Minds

INFO 6540 Data Management

Data Management Plan

- **Jamie Dunbar (B00493505)**
- **Brett McNeill (B00612674)**
- **Xiaoyue Chen(B00736746)**
- **Ankita Zaveri (B00750459)**

This report describes an overall approach to managing all portions of a data life cycle, from data collection, to data storage and distribution for three separate data management plans of researchers at Dalhousie University. This project generates a mix of data types. It collects a blend of raw, experimental and simulations data and each require unique data management plans.

CASE 1: Professor Periwinkle - Proposed Data Management Plan

Types of Data Being Collected and Produced

Ongoing data collection are comprised of raw data which includes data collected from the sensors and monitoring equipment such as tags that are surgically implanted in animals, static sensors, and remotely-operated vehicles(ROMV). The volume of the raw data generated is about 300 MB daily. The above data is then converted to NetCDF format. The data is in uncompressed form and is generated at the rate of 500MB each day. Other ongoing data collection includes citizen reports and field notes are from Dr. Periwinkles website in the form of TSV. Total size of the data would approximately equal 3GB.

The data is typically used for analysis purposes and running complex simulation models. The results of the simulation data are stored in CSV format. The total size of the data is approximately 200 GB. Based on which the field notes, Reports, Research papers, Exercise papers and other inferences are stored in Darwin core format. The data collected is approximately 2GB

The data should be organized and managed. Choosing a logical and consistent methodology for naming and organize files would allow users and others to easily locate and access the files. The structure of the directories/folders should also have a clear labeling system. The Top- Level folders or directories would be categorized based on Data type. And the files can be named in the following format consistently in a form of title, timestamp, and format. For example, “Title+ dd_mm_yyyy+ Format”. It is also suggested that process document that explains and defines each data sets be created, in addition to verbal instruction for how the data is formatted and structured. This document will be handy for end users other than the project team members to understand the data sets and its original source.

For performing advanced analytics on the data on a wider scale, different data types and formats can be cumbersome. To overcome this complexity, converting all documents in a same format such as CSV or json is recommended. NETCDF can be easily converted to CSV with general purpose programming. Having data in CSV would make it easy to have all data

visualized in same visualization tools and can also be used easily for Machine Learning purposes.

What Licenses Apply to the Data

The licensing option allows the end users of your data to know what rights they have regarding the use and sharing of Dr. Periwinkles data. There are several types of licenses that can be applicable in this case such as standard license, open data commons open database license, open non-commercial government license, or multiple licensing. For the data collected and owned by Professor, it would be suggested to use Creative Commons Attribution-ShareAlike License (CC-BY-SA license). This license would allow the end users of the data to share, modify with a proviso that the user gives proper attribution to the original owner of the data and Indicate if any changes are made to the original data sets. Further, the contributions must be shared under the same license. The professor can also apply Creative Commons Attribution (CC-BY license) which is similar to CC-BY-SA excluding the Sharalike part of the license. CC-BY-SA license would ensure that the future adaptations of the data would be released under the same or a similar license.

Data Management Practices

In this case, the data size is large and is continuously renewing every day. According to the case, the first part of data is the ROMV data, which is collected monthly and converted to NetCDF formatted data for storage. Every day, there will be in total 800MB raw and NetCDF data being produced, and the NetCDF data size so far has reached 500GB. The Second part of the data is the field notes, the size of the data collected from field notes is roughly 2GB. The third part is the CSV files of complex simulation models, which accumulated to 200GB now. Our group conducted an estimation of a 3 - year data storage. First for the NetCDF data, yearly storage that required is 800MB per day * 365 days equal to 285GB. In 3 years, NetCDF data was conservatively estimated to reach 285GB* 3 years which is 855GB. Second, for the File notes the estimation is going to be doubled which is a minimum of 4GB in 3 years, due to the increase in funding and research being conducted. There are additional considerations is the data was to increase at the same speed as the NetCDF data, therefore in 1 year it will reach (200GB/500GB) * 285GB=114GB. In three years would reach 114GB *3= 342GB. Therefore, in total the storage that Dr. Periwinkle need will be at least 1201GB over a 3 - year period.

Also due to the increased funding, Dr. Periwinkles lab is receiving, there is substantially more data to collect. The size of the data files are expected to increase therefore the estimated

storage has potential to be larger than what our team predicted above. The storage requirement is high, approximately over 2TB. The current and historical data are stored in hard drives but as the data size keeps increasing the hard drive is not a very good choose. Hard drives are expensive and have the risk to be lost to disaster and will create unnecessary difficulty to organize the data in practical sequence. The data is going to be renewed and accessed very frequently, making the hard drive a very inconvenient option for quick data access and sharing. Therefore, our team suggest Dr. Periwinkle to move all the data to the cloud drive. Cloud allows the data owners to set up access control and also can set up access limitations by IP address and other types of user authorizations. Dr. Periwinkle can set up constraints on who can edit the internal data and can provide view only version for student's general use. For the CSV files, which contain the simulation model, the Version control is necessary. The Cloud server would allow Dr. Periwinkle's research team members to track the changes and keep different versions of the model. Backups are also important; the hard drive is useful, and it should save the all the raw data periodically on a hard drive and on another backup cloud drive, to prevent the loss of data, unauthorized access and incorrect modifications.

Policies for Access, Sharing, and Re-use

Who will own and have access to the data

There are two opportunities to discuss ownership and access to data when considering Dr. Periwinkles research. There is the data collected by Dr. Periwinkle and her team, and there is the data collected from other universities. Regarding Dr. Periwinkle and her team, the professor herself should take full ownership of the data collected from her students and research associates. This will place the burden of control under one person and allow for a centralized decision about the data. In the case it states, "As students graduate they share their data with current students in the lab or take it with them on their personal devices". Dr. Periwinkle should request that her students pass their data into her control, so she can continue to grow her collection of research specific data. This will handle all internal data ownership issues and anyone at Dalhousie can easily reach out to Dr. Periwinkle and gain access to data.

Regarding the second issue, external users, sharing this data across Canada with other universities can really open the door for more robust research. Dr. Periwinkle should request a shared stewardship of research data agreement that makes use of a network such as Portage. Such an agreement would help to bring all major data collectors across Canada together and allow for easier access to research data. The portage principles states that research data are a

public good and additionally allows open access while still respecting privacy and promoting stewardship. Not only can such an agreement help Dr. Periwinkle with advancing her own research but others across the world can make use of the vast data that she herself has collected and push the field of research even further.

Which data will retain value after the life of the project

There is a wide range of data being collected by Dr. Periwinkle and her team. The daily buoy data, citizen reported sightings, field notes, and simulation models that use the buoy data. The buoy data is taken in daily and needs to be manipulated into a usable format. The raw data from the buoys does not retain value after it has been transferred into a usable format. This is especially true if data sharing agreement is made, the end users would prefer to work with formatted, user friendly data, and large amounts of raw data could clutter the sharing network. It is suggested that the raw data, once transformed, and used in models, can be stored feasily on the cloud server. The second type of data are field notes made by citizens, students, and other researchers. This data is not currently being used by Dr. Periwinkle, but it could retain value after down the road as more behavioral data is analyzed by other researchers. These notes should be transcribed into a user-friendly format and made available to individuals should a data sharing agreement be made. The third type of data is that collected by the simulation models run by Dr. Periwinkle's team, this is the most valuable data as it has been already analyzed to predict animal populations and movements. This simulation data will retain its value long after the life of the project and should be curated and maintained with that respect.

What metadata and linked open data strategies will be employed

Dr. Periwinkle has already created a help document that explains how to work and contribute to the data sets. This document can be updated and used as the basis for metadata for the research data as well as a user guide made available on a data sharing platform. Linked open data refers to a set of design principles for sharing data. It makes use of uniform resource identifiers (URIs) that allows users to quickly look up items that have been standardized. Not only does linked data help the user make sense of data quicker, it also provides an extension to that data that makes it readable by computers. Metadata and linked open data can help make Dr. Periwinkle's data easier to use for the general public and help facilitate greater research on the topic.

Long-term Storage and Data Management

Data storage and preservation of access

During the new data collection periods and in the subsequent years, all data will be stored in digital form on a secure university server as well as backed up on a third-party cloud-based server. This will allow for a combination of two separate storage locations that will help prevent against any potential loss due to disaster. Both of these options are institutionally supported and will be free for researchers, students and staff who work with or closely tied to the research that Dr. Periwinkles. The fact the storage methods are supported by the institution allow for increased data sharing internally.

d

There will also be the addition of data sharing protocol that was implemented to address the information sharing protocol for individuals in the research community that would like to access Dr. Periwinkles raw data. With the expansion of the data sharing protocol to deal with external researchers of the university, it can also be concluded that these individuals will have access to the cloud-based storage location via personalized login information. As for individuals who do not work in the university or haven't been granted authorization to access raw data there is still the public facing OceanViewer.org. OceanViewer.org is open online platform for sharing oceans-related findings and where citizen scientists can report information about wildlife.

In addition to how various types of researches will be able to access data, there will need to be preservation plan that is revisited every 5-10 years, this will ensure that any outdated technologies used to collected data will be upgrading to match current trends and mitigate against data being lost due to outdated formats or storage types.

Budget

The costs associated with the various aspects of the proposed data management plan for Dr. Periwinkles research data starts with the new cloud based online backup server and the time required to transfer and maintain the files. Staff time has also been allocated for additional cost to transfer and archive existing and new data. There is also the cost of assessing the data management plan at the end of the 3-year plan. Costs have been calculated and can be seen in Appendix A.

CASE 2: Professor Green - Proposed Data Management Plan

Types of Data Being Collected and Produced

This project collects confidential information and sensitive data studying employee teamwork in hospital environments. The ongoing data is acquired via interviewing key informants. Approximately, 15 audio recordings of the interviews in mp3 format are stored via Dropbox. The volume of the data would range to 600 megabytes – 1 gigabyte. The interviews are transcribed and are stored in word file which are then uploaded in google drive storage.

Professor Green collects data sets regarding healthcare outcome, expenditures, staffing, enrolment by discipline, demographics, pairing this with the hospital environment data and interview data. These data sets are collected from open data sources through healthcare providers and internal network and are downloaded in TSV formats. Data created from textual/Content analysis of 383 individual documents are stored in different formats such as Word, PDF, Plain, text and other documents which stored in Zotero database. Quantitative data for the above files is stored in excel format. The total volume of the data collected amounts to 24 GB.

The data is in different formats. Although the formats justify the types of data, it is recommended that tabular data to be converted to CSV format from TSV and excel formats. The reason for the same is that CSV is accepted by almost all existing software, and applications making it easier to analyze data. It would also be possible to have all data in one place with a proper structure. The data should be structured and manageable. Since some of the data is confidential, it is essential to have control access to information in the form of password protected and encrypted hard drive.

What Licenses Apply to the Data

Most data sets used by Dr. Green are confidential patient data, hence under contract it is not allowed to be shared. However, the data can be stripped from direct identifiers, by anonymizing the information. This anonymized data sets, along with other non-confidential data such as open data sets used for the project can be applied for license. Licensing data allows the end users of the data sets to understand the user's rights to use and share your data. There are several types of licenses that can be applicable if Dr. Green considers sharing his data in the future. Licenses such as standard license, GNU General Public License, or multiple licensing can be applied. The data owned by Dr. Green which will be anonymized for and does not reveal any confidential material, can use CC-BY-SA license which would allow the end users

of the data to share, modify with a proviso that the user gives proper attribution to the original owner of the data. Under this license the user must also Indicate if any changes are made to the original data sets. Further, the contributions, adaptations or future work build on this data sets must be shared under the same license promoting open and free sharing of data. This is all dependent on Dr. Green wanting to share his data and results to other researches and peers in the future for validating his research.

Data Management Practices

The most important information for Dr. Green, the experimental data is not going to increase, so the biggest concern is the confidentiality of the data. According to information provided by Dr. Green and information provided in the discussion, the textual/content analysis file and open data is around 24 GB so far and Dr. Green estimate the size of the data may tripled to 72GB in the future. The second type of data collected is the MP3 files of the interview records. According to AudioMountain.com, one-hour 128kps Mp3 is 57.6MB (Appendix B), therefore, 15 one-hour interviews would need 864 MB storages. The transcriptions would be all word files which is going to take very small amount of storages. Since the size of the experiment data is not going to increase significantly, in total the estimation of the storage that Dr. Green needs is going to be less than 100GB.

The research data needs to be stored and protected over a ten years duration. Multiple backups should be saved to ensure the data safety during a long duration. Also, since the primary concern in this case is the security and confidentiality of the data. Dr. Green refuses to share the data with anyone, therefore there is no external access control needed to be considered. Therefore, our team suggests Dr. Green keep two encrypted hard drives to ensure the data security. A local copy could be stored in the personal computer located on the institutions campus. Additionally, the second hard drive that Dr. Green's is current using to keeps his master copy should continue to be used. Also, transferring his data to a secure intuitional cloud repository could be used for daily access by Dr. Green and where his grad student can upload data.

Currently, Dr. Green stores the MP3 files in Dropbox, transcriptions in Google doc, and textual files in the Zotero database. Since the information in the Dropbox and Google doc is the same, therefore our team suggests the data should be moved together under a single cloud repository that is supported by Dalhousie University, SharePoint. SharePoint is not only a more logical cloud database for Dr. Green to securely centralize his research data but is now stored in Canada. Dropbox and Google have storage and disaster locations in the U.S., which comply with different data privacy policies than storage facilities hosted Canada. Due to the sensitive nature

of the data, having the confidential data in a location that complies with Canadian data privacy laws is imperative.

Policies for Access, Sharing, and Re-use

Who will own and have access to the data

Dr. Green collects very sensitive data that he has promised not to share with anyone outside of his research team. There are two types of data, textual documents, and audio recordings, these could potential be very damaging to the individuals should they be made available to anyone. Due to the nature of the data collected by Dr. Green he must maintain sole ownership of the data, like what he currently does, by restricting access to only those in his research team. Within his team the data is shared using Zotero, for documents, and Dropbox, for audio files. Dr. Green also maintains master copy of all data on a USB that he keeps on him at all times. It is recommended that Dr. Green continue to operate and take ownership of his research data in similar term during this project. Protecting the confidentiality of his participants is critical for the continuation of his research.

Which data will retain value after the life of the project

Once the data has gone through the data management project it can be removed from the cloud storage vendor and only retained on a master copy that Dr. Green keeps on himself if he wishes to remain completely secretive of the data and research. This will help to eliminate the chances of the data falling in to the wrong hands by limiting the potential for data contact with unauthorized individuals. The data involved in the project will retain its value years down the road as the evolution of teamwork in healthcare can continued to be evaluated. The confidentiality of the participants does need to be considered however and so limiting the number of copies for this data, especially the audio files, is critical in ensuring that confidentiality. It is also expected that the majority of the confidential information will be de-identification to ensure confidentially during and after completion of the research.

What metadata and linked open data strategies will be employed

Since the data is suggested to be protected and not shared with anyone outside Dr. Greens team the requirement for metadata and linked open data strategy are very stringent at best. It is not suggested that any metadata be associated with Dr. Greens data as that will help to protect the data should it fall into the wrong hands by making it harder to read by the perpetrator. Since access to the data is restricted to just that of Dr. Green and his team the

requirement for information regarding the data is very limited as Dr. Green can just explain the data to his team or anyone new joining him. This should be in the hopes of securing the data from outside use.

Long-term Storage and Data Management

Data storage and preservation of access

Currently Dr. Greens sentiment towards the data and research he is conducting is that it will be personally used by him to protect the hard work he completed on his own and to protect the sensitive personal information of patients. To comply with Dr. Greens wishes, it is suggested that he retain the only ability to access his research data for conducting work. However, Dr. Green should consider in future if he wishes to have his research validated by peers that the data be de identified and made available to select individuals in his field to replicate his work.

For short term archiving purposes this data will be on an automatic backup schedule and retained for the duration of the research on the dedicated Dalhousie University repository with strict access limited to Dr. Green. However, because of the sensitive nature of the data, we also suggest that Dr. Green backup the data and research daily on his two hard drives. Dr. Green will also be responsible for preparing data for long-term preservation and for updating information. There have been staff hours dedicated for de-identifying the data if he wishes to share the results at the end of 10 years.

Budget

The cost associated with the various aspects of the proposed data management plan for Dr. Greens are relatively small. Staff time has been allocated for the additional cost of transferring the files from their current storage locations (Drop Box, Google Docs) to Sharepoint. There is also the cost of maintenance and support of the system over 10 years and costing out a secure and portable encrypted Hardware. Costs have been calculated and can be seen in Appendix C. In the future if Dr. Green would like to explore alternative cloud repositories outside of what is currently used at Dalhousie University reporistory, table 1 highlights two potential options.

Cloud Provider	Cost and Services
Sync.com	2 TB secure storage in Canada, $300 annually for 2 users, advanced sharing , restriction controls and automatic backup and sync.
EazyBackkup.com	1 TB secure storage (soft cap) in Canada PIPEDA and PHIPA compliant, $700 annually for 2 users, advanced sharing, restriction controls and automatic backup and sync.

Table 1: Cloud Storage Options

CASE 3: Professor Pinkerton - Proposed Data Management Plan

Types of Data Being Collected and Produced

Dr. Pinkerton has been accumulating a large amount of research data over the last 10 years. The data acquired is dynamic in nature and are collected from a variety of sources. Total volume of the data collected and saved is approximately 60 GB. Data creation and collection is unsystematic and often lacks specific purpose. The data is created through several different ways such as, conducting systematic literature review and performing quantitative textual analysis on the data.

The data is then stored in excel format or sometimes CSV format. Further, Professor Pinkerton has also collected job descriptions for entry level positions in her field and stored in excel. Additionally, she has collected 12 years of student performance data from the courses she teaches. She also receives data from other researchers. Some of the data she stores may not be used or accessed regularly, hence it is recommended to archive the data. The data is already stored in the professors' laptop in an organized and structured manner. However, it is essential to migrate the data onto a cloud sever.

What licenses apply to the data

The licensing option allows the end users of your data to know what rights they have regarding the use and sharing of data. There are several types of licenses that can be applicable in this case, such as standard license, open data commons open database license, open non-commercial government license, or multiple licensing. For the data collected and owned by Professor Pinkerton, based on her willingness to share the database, it would be suggested to apply a Copyleft license. Creative common Attribution license would allow the end users of the data to share, modify with a proviso that the user gives proper attribution to the original owner of the data. Professor Pinkerton can also apply for Creative Common Attribution-Share Alike License which would allow the end users of the data to share, modify with a proviso that the user gives proper attribution to the original owner of the data, also the user has to Indicate if any changes are made to the original data sets. Further, the contributions and future work adaptations by the user must be shared under the same license or similar license.

Data Management Practices

The long-term goal of Dr. Pinkerton is to build a database for both internal and external use. According to the information provided in the discussion, the average size of the excel file is 3.5MB, and with total 17384 spreadsheets, the storage requirement is 60GB. Dr. Pinkerton is continuously collecting data from external source, therefore the storage that is need will increase in the future, however since excel files generally do not require large storage, the estimation of the storage would not be exceeds 100GB for our proposed three-year plan.

According to the information from Dr. Pinkerton, she keeps local copies of all the data and currently has a postdoctoral fellow tasked to organize the data on the University SharePoint. The new back-up plan is to successfully migrate Dr. Pinkertons' data on to SharePoint categorized as open and non-open data. This responsibility of the migration will be taken off the postdoctoral student and assigned to a dedicated staff member. The internal research data and data collected from open data source should have different level of access control. External user database servers are open data sources so there will be no access controls, but the internal research would need special requests and authorizations to gain access. Moreover, for the data Dr. Pinkerton does not use very often could be stored in the separate files on the hard drives and located in appropriately titled folders or categories in SharePoint.

Because Dr. Pinkerton is willing to build an external use database serve as an open data resource. Therefore Dr. Pinkerton should adapt the Fair Guiding principle for the data management practice in order to "maximize the added-value gained by contemporary, formal scholarly digital publishing", and the four principles are "To be Findable", "To be Accessible", "To be Interoperable" and "To be reusable" (Wilkinson,2016).

Policies for Access, Sharing, and Re-use

Who will own and have access to the data

Professor Pinkerton has a vast collection of data files that she retains control and ownership of. She also works closely with her postdoctoral fellow, Neil Gaiman, who has complete access of the data collected and was requested to structure a folder in the cloud to manage the data. In terms of ownership of the data this can be done with partnership between professor Pinkerton and Neil Gainman, this partnership will help promote strong stewardship of the data. Most of the data is from external sources, 95%, so references to initial publication will need to be maintained.

Professor Pinkerton will freely share her data with anyone that asks, however of late she has been to busy to respond to requests for data. With the creation of the cloud-based folder structure anyone can be granted access to the data and they can freely browse and download anything from the collection. There should be no privacy concerns in this regard as the data has all been collected from open source or sent to her by fellow colleagues without restriction. Allowing full access to the public will eliminate the need to handle data requests personally.

Which data will retain value after the life of the project

Professor Pinkerton has collected 17, 384 spreadsheets and hasn't made use of them all yet. This is a huge collection of data that she does want to do research on eventually. Ensuring all that information is available to her as well as others after the life of the project means maintaining those data files. Professor Pinkerton also stated that she does not trust the external sources to provide the data into the future meaning that all the data will retain value in her eyes after the project. Luckily spreadsheets do not take up that much space and can be stored easily, this will also allow Professor Pinkerton to relinquish maintaining local copies for herself if they can be stored long term.

What metadata and linked open data strategies will be employed

With the large amounts of data that will stored for the project proper information will be required to document it. This means ensuring there is metadata available for all spreadsheets, either through original sources, or created by Professor Pinkerton and her research associate. The metadata will allow for external users to be able to quickly understand and figure which data sets are of most use to them. This will also allow Professor Pinkerton to revisit old data sets when she wishes to pursue them in the future and be quickly caught up to speed. Employing an open data strategy can be useful as well as it allows for users to access, use, and share this data freely. Making this data linked can help users maneuver and understand the data as uniform resource identifiers (URIs) allow users to quickly look up items that have been standardized.

Long-term Storage and Data Management

Data storage and preservation of access

The research data from Dr. Pinkertons work will be deposited with the Dalhousie University SharePoint repository, there will be dedicated staff member for this who is not Dr. Pinkertons postdoctoral fellow assigned this task. The location on SharePoint will have

dedicated research data storage, ensure that the research community has long-term access to the data. The date will be stored and categorized under specific folders such as open and non-open data. Once the data is successfully migrated, there will be automatically daily scheduled backups and syncs used on the research conducted by Dr. Pinkertons is immediately available on the cloud. It also suggested that the daily research conducted by Dr. Pinkertons is backed up onto an encrypted hard drive at the end of the day.

Having a dedicated and easy accessible storage location for Dr. Pinkertons data on SharePoint will allow other researchers to update their findings in relevant folders, and access research from Dr. Pinkerton without having to deal with access request. It's is suggested that Dr. Pinkerton identify and label sensitive data, so the data management team can place this information in separate folders with access restrictions that require initial authorization from Dr. Pinkerton for access. Other than sensitive data that requires initial authorization from Dr. Pinkerton, all other forms of data and repository are accessible to other researchers in her department and area to access. SharePoint has a version control feature built in, this will allow for the preservation and integrity of her work. Additionally, there will need to be preservation plan that is revisited every 5-10 years, this will ensure that any outdated technologies used to collected data will be upgrading to match current trends and mitigate against data being lost due to outdated formats or storage types.

Budget

The cost associated with the various aspects of the proposed data management plan for Dr. Pinkerton are mainly influenced by staffing costs. Staff time has been allocated for the additional cost of transferring the files from their current storage hard drive location to SharePoint. There is also the cost of maintenance and support of the repository over the three years and costing of new a secure and portable encrypted Hardware. There is also the cost of assessing the data management plan at the end of the 3-year plan. Costs have been calculated and can be seen in Appendix D.

Appendix A - Case 1 Cost Breakdown

Cost Drivers	2018	2019	2020	Total
Cloud Based Backup Solution *	$ 700.00	$ 700.00	$ 700.00	$ 2,100.00
Documentation and Maintenance**	$ 7,800.00	$ 7,800.00	$ 7,800.00	$ 23,400.00
Reassessment cost***	$ -	$ -	$ 2,600.00	$ 2,600.00
Total	$ 8,500.00	$ 8,500.00	$ 11,100.00	$ 28,100.00
*Pricing from EazyBackup. 3 workstations, 2 servers, Hot SQL and MySQL database backups, 24/7 support, unlimited file versions and backup schedules.				
** Based on one $20/hr staff with dedicated 7.5hr weekly towards documentation and maintenance or other support roles as needed.				
*** $650/day -four day consulting fee for reassessing data management plan.				

Appendix B-- Mp3 file size Calculations

MP3 File Size Calculations

Formula:
Kbps = bits per second / 8 = Bytes per second x 60 seconds = Bytes per minute x 60 minutes = Bytes per hour

Bitrate	File size per second	File size per minute	File size per hour
8 Kbps	1 KB	60 KB	3.6 MB
16 Kbps	2 KB	120 KB	7.2 MB
32 Kbps	4 KB	240 KB	14.4 MB
40 Kbps	5 KB	300 KB	18.0 MB
48 Kbps	6 KB	360 KB	21.6 MB
56 Kbps	7 KB	420 KB	25.2 MB
64 Kbps	8 KB	480 KB	28.8 MB
80 Kbps	10 KB	600 KB	36.0 MB
96 Kbps	12 KB	720 KB	43.2 MB
112 Kbps	14 KB	840 KB	50.4 MB
128 Kbps	16 KB	960 KB	57.6 MB
160 Kbps	20 KB	1.20 MB	72.0 MB
192 Kbps	24 KB	1.44 MB	86.4 MB
224 Kbps	28 KB	1.68 MB	100.8 MB
256 Kbps	32 KB	1.92 MB	115.2 MB
320 Kbps	40 KB	2.40 MB	144.0 MB

http://www.audiomountain.com/tech/audio-file-size.html

Appendix C - Case 2 Cost Breakdown

Cost Drivers	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	Total
Staff Cost *	$ 1,650.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$15,150.00
Hard Drive x 2**	$ 700.00	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ 700.00
Total	$ 2,350.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$ 1,500.00	$15,850.00
* Staff cost includes 1 week of dedicated file transferring, plus an additional 10 days per year in staff support and maintenance (Staff 20/hr)											
** Apricorn Aegis Padlock 2 TB USB 3.0 256-Bit AES XTS Hardware Encrypted Portable External Hard Drive.											

Appendix D - Case 3 Cost Break Down

Cost Drivers	2018	2019	2020	Total
Hard Drive*	$ 350.00	$ -	$ -	$ 350.00
Migration and Maintenance**	$ 7,800.00	$ 7,800.00	$ 7,800.00	$ 23,400.00
Reassessment cost***	$ -	$ -	$ 2,600.00	$ 2,600.00
Total	$ 8,150.00	$ 7,800.00	$ 10,400.00	$ 26,350.00
*Apricorn Aegis Padlock 2 TB USB 3.0 256-Bit AES XTS Hardware Encrypted Portable External Hard Drive.				

** Based on one $20/hr staff with dedicated 7.5hr a week per year towards document migration and maintenance or other support roles as needed.
*** $650/day -four day consulting fee for reassessing data management plan.

References

Wilkinson, M., Dumontier, M., Aalbersberg, J., Appleton, G.,, Axton, M., & Baak, A. (2016). Data Archiving Networked Services. The FAIR Guiding Principles for scientific data management and stewardship. *Nature, 3*, Urn:issn:0028-0836.