Emma Arata and Reagan Callahan
LIS 546 Data Curation II
Instructor Bree Norlander
June 6, 2021

Good Games! Repository Curation Protocol

I. *Context*

A. Project Definition and Scope

Through this project, we hope to learn how to standardize data from various sources and then present said data in a user friendly way. We plan to investigate the nuances of data licensing agreements, iterate on existing protocol designs, and by refining our practical understanding of various data curation principles, ultimately construct a protocol that prioritizes the needs of our end users. The protocol we build will ideally empower users to explore the different measures of a video game's popularity.

B. Audience

This repository will primarily serve policymakers. Collection developers may refer to this data when looking to either introduce or expand a library's video game collection and video game developers may refer to this data when conducting market research. While there are numerous websites specifically designed to provide video game consumers with review aggregates, there is also a chance said consumers will derive a similar benefit from this repository. For example, usage data could help a potential consumer determine whether or not an older multiplayer title is still worth playing.

II. *Dataset Assessment and User Stories*

A. Dataset Assessment

Dataset:

Name: Steam Store Games (Clean dataset)
Link: https://www.kaggle.com/nikdavis/steam-store-games
License: Attribution 4.0 International (CC BY 4.0)
Overview:

The dataset contains information about the published video games on the online video game store, Steam. Data was collected in May 2019 using the Steam Store and SteamSpy APIs and is considered open through a CC license. A Kaggle account is needed to download the CSV files.

Assessment:

The dataset is considerably clean and is easy to maneuver. Variables that might be particularly important to developers include the title of the game, release dates, categories, genre, steamspy tags, the number of positive and negative ratings, number of owners, and the price of the game. Users might be able to filter observations by genre or categories, perhaps similar to that which they are developing or have developed already. Being able to compare information surrounding the number of owners, average playtime and price for games with similar genres or categories would be helpful for marketing endeavors. While this dataset includes a wealth of information, not all game developers create games to be used on Windows, Mac, or Linux, so including similar information for console games would be beneficial.

B. User Profiles

Profiles in Brief

Video Game Developer

- Video game developers will reference this data while conducting market research in order to evaluate market trends in terms of popular genres or features. In a similar vein, video game developers can submit relevant data about their own games to the dataset.

Library Collection Developer

- While either introducing or expanding their library's video game collection, collection developers will refer to this data to determine which titles they would like to include.

Video Game Consumer

- Those interested in purchasing video games may also consult this data prior to purchasing a game.

Video Game Developers

Video game developers go to great lengths to create a fun, interactive experience for consumers. Market research is an essential component of the production pipeline and ensures a successful game release. The GG! repository will be particularly helpful as an aid for market research, as it provides a fair deal of data regarding the types of games that are most frequently purchased via online stores, including Steam.

Our hypothetical developer accesses our repository and the Steam Store Games dataset to better understand what the most popular Steam titles have in common. The developer intends to use this new understanding to inform design decisions for their own game, which they hope to release successfully through Steam. After accessing this data and using it to their benefit, our developer will in turn be keen to submit data related to their own game. Doing this would also enable them to more easily compare their game's data to the data of competitors. If the developer has previously released titles through Steam, they may also choose to submit data not included in the Steam Store Games dataset, either because those titles were released before or after May 2019, or because the data pulled from their API was incomplete.

Library Collection Developers

Much the way video game developers go to great lengths for consumers, library collection developers go to great lengths to ensure their library's collections are enriching for library patrons. More and more libraries are expanding their catalogs to include video game titles. When creating or expanding a library's video game collection, the collection developer aims to select titles from a variety of different genres, operating systems, and ratings. Patron demographics may also influence these decisions. For example, if the patrons most interested in the video game collection are very young, the collection developer may prioritize titles with an appropriate ESRB rating. Our repository will make this searching process easier by investing in faceted search options, so that collection developers can quickly identify popular titles with specific traits, such as a particular rating, price point, genre, or operating system.

Video Game Consumers

Video game consumers are demographically diverse. There are already a large number of resources specifically designed to help guide purchasing decisions, such as gaming magazines, review blogs, review aggregating websites, and popular forums. While this repository is unlikely to be as user friendly as these resources, the data our repository offers may appeal to a numbers-oriented subset of consumers. Multiplayer video games that rely on an active player base

may only be worth purchasing if there is sufficient evidence of activity, especially if the title in question is an older release. A video game consumer can easily use our repository to track popularity over time, something that is more difficult to accomplish with other purchasing guides.

Goals	U/ Video Game Developer	Library Collection Developer	Video Game Consumer
Successfully navigate the repository of video game data	"I want to refine my search to locate the most recent data regarding games purchased on Steam."	"I would like faceted search options, such as by operating system, to ensure my search results contain data relevant to the consoles I know are most popular with my library's patrons."	"I want a user-friendly repository that is easy to navigate and search."
Compare and analyze information within the repository	"I want to be able to see how my game fares against other games and quickly identify datasets related to my game."	"I want to evaluate popularity through sales metrics as well as ratings."	"I want to identify qualities, such as genre and price point, that different titles have in common."
Easily contribute data for future use	"I want to submit our game's Steam data, which was released after May 2019."	"I'm interested in sharing anonymous data regarding the video games most popularly checked out at my library."	"I found a dataset on a website that I think fits this repository, but I've never submitted a dataset before and need help making sure I'm doing so correctly."
Download datasets in a convenient format	"I want to be able to download datasets as a simple CSV to ensure the file will be supported by my data upload interface."	"I want to be able to download CSV datasets so I can compare them with the circulation statistics of my library's video game collection."	"I want to be able to download datasets as an XLSX, because I am most familiar with this file format."

III. *Metadata Needs*

The Good Game! (GG!) Video Game repository uses the The Dublin Core Metadata Initiative (DCMI) schema. The elements used by the repository are listed out in the ***Elements*** section. Following that is a table, which depicts each element and its namespace, namespace definition, optionality status, any relevant rules, and an example. Each element follows an encoding scheme that can be used in XML.

Required elements ensure a baseline consistency across all submitted datasets and are derived from the DCMI schema. Optional elements should only be included if they are applicable to a submitted dataset. These elements are also derived from the DCMI schema.

Supplementary elements are inspired by the Video Game Metadata Schema (VGMS) and the Video Game Metadata Schema designed for the Seattle Interactive Media Museum. These elements are also optional and serve to improve the repository's automated authenticity check by providing the bot with keywords to crosscheck with the submitted dataset. In addition to bolstering authenticity checks, these elements will enable users to more easily locate datasets with a specific theme. For example, a user interested in seeing datasets specifically related to the "horror video game genre" or "Mass Effect series" will be able to narrow their search accordingly.

<u>Required Elements</u>

- Dataset Title
- Creator
- Identifier
- Created Date
- Temporal Coverage
- Description
- Source
- License
- Language
- Format
- Type

<u>Optional Elements</u>

- Modified Date
- Contributor
- Spatial Coverage

<u>Supplementary Elements</u>

- Game Title
- Game Series
- Year Game Released
- Genre
- Operating System
- Developer
- Publisher
- Number of Players
- Rating
- Suggested Retail Price

Metadata Elements and Examples						
Element	**Namespace**	**Label**	**Definition**	**Status**	**Rules for Content**	**Example**
DCMI:title	http://purl.org/dc/terms/	Dataset Title	A name given to the resource.	Mandatory	String	Steam Store Games (Clean dataset)
DCMI:creator	http://purl.org/dc/terms/	Creator	An entity responsible for making the resource.	Mandatory	Literal Value (i.e. name of creator)	John Smith
DCMI:identifier	http://purl.org/dc/terms/	Identifier	An unambiguous reference to the resource within a given context.	Mandatory	String	
DCMI:created	http://purl.org/dc/terms/	Created Date	Date of creation of the resource.	Mandatory	This will be automatically generated upon submission. It follows a YYYY/MM/DD format.	2021/05/02
DCMI:temporal	http://purl.org/dc/terms/	Temporal Coverage	Temporal characteristics of the resource.	Mandatory	Preferred format is YYYY/MM/DD, but YYYY/MM, YYYY, and a range of dates are also accepted.	2019/02/14; 1998-2001
DCMI:description	http://purl.org/dc/terms/	Description	An account of the resource.	Mandatory	String	This dataset contains content related to PS4 game sales.
DCMI:source	http://purl.org/dc/terms/	Source	A related resource from which the described resource is derived.	Mandatory	Controlled Vocabulary: Proprietary CV (PS4, Steam Store, etc). Alternatively, users can submit a string attribute if the source is not found in the CV.	Steam Store
DCMI:license	http://purl.org/dc/terms/	License	A legal document giving official permission to do something with the resource.	Mandatory	Literal Value	CC-0

DCMI:language	http://purl.org/dc/terms/	Language	A language of the resource.	Mandatory	Controlled Vocabulary: ISO 639-2 https://www.loc.gov/standards/iso639-2/php/code_list.php	eng
DCMI:format	http://purl.org/dc/terms/	Format	The file format, physical medium, or dimensions of the resource.	Mandatory	Controlled Vocabulary: Proprietary CV consisting of accepted file formats and sizes.	CSV; 2 MB
DCMI:type	http://purl.org/dc/terms/	Type	The nature or genre of the resource.	Mandatory	Controlled Vocabulary: Proprietary CV (ratings, reviews, usage, or sales) used to indicate a dataset's focus.	Sales
DCMI:modified	http://purl.org/dc/terms/	Modified Date	Date on which the resource was changed.	Optional	This will be automatically generated upon modification. It follows a YYYY/MM/DD format.	2021/05/23
DCMI:contributor	http://purl.org/dc/terms/	Contributor	An entity responsible for making contributions to the resource.	Optional	String Value	University of Washington
DCMI:spatial	http://purl.org/dc/terms/	Spatial Coverage	Spatial characteristics of the resource	Optional	Controlled Vocabulary: Getty Thesaurus of Geographic Names (ISO 3L) https://www.getty.edu/research/tools/vocabularies/tgn/	USA
VGMS:gameTitle	http://metadataregistry.org/uri/schema/VGMS/100029	Game Title	Proper names used to refer to a video game, assigned by the creator.	Optional	String Value	The Sims 4

VGMS:series	http://metadataregistry.org/uri/schema/VGMS/100014	Game Series	The Series entity type allows for the expression of ordered and hierarchical groupings of video games belonging to a particular series as established by game companies or by users.	Optional	String Value	Mass Effect Series
VGMS:retailReleaseDate	http://metadataregistry.org/uri/schema/VGMS/100139	Year Game Released	The date of the public/commercial release of the video game.	Optional	Preferred format is YYYY/MM/DD, but YYYY/MM and YYYY are also accepted.	1998/02/22
VGMS:gameplayGenre	http://metadataregistry.org/schemaprop/show/id/22903.html	Genre	The overall nature of a video game's interactivity based on its objectives, types of rules, distinctive characteristics, modes of action, and manners of gameplay.	Optional	Controlled Vocabulary: OLAC Video Game Genre Vocabulary https://www.olacinc.org/olac-video-game-vocabulary	Horror video game
VGMS:platform	http://metadataregistry.org/uri/schema/VGMS/100079	Operating System	The hardware and operating system on which the video game was designed to be played.	Optional	Controlled Vocabulary: UCSC-Stanford Libraries controlled vocabulary for Computer Game Platforms. https://gamemetadata.soe.ucsc.edu/platform The repository will update this CV to include newer generation consoles, such as the PlayStation 5.	Microsoft Windows 10
SIMM:developer	https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/33392/IJDL_GameMetadata.pdf?sequence=1&isAllowed=y	Developer	An organization or group of individuals and/or organizations responsible for creation and/or realization of a game	Optional	String	Electronic Arts

SIMM:publisher	https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/33392/IJDL_GameMetadata.pdf?sequence=1&isAllowed=y	Publisher	An organization or group of individuals and/or organizations responsible for the manufacture, marketing, and distribution of a particular product manifestation.	Optional	String	Tencent Games
VGMS:numberOfPlayers	http://metadataregistry.org/uri/schema/VGMS/100095	Number of Players	The number or range of the number of players the video game can accommodate either separately or concurrently.	Optional	String	1-2
VGMS:rating	http://metadataregistry.org/uri/schema/VGMS/100120	Rating	The classification of the content in the video game used to inform decision making about the video game, provided by organizations such as professional associations, video game distributors, or creators. Note that some older video games do not have this rating information.	Optional	Controlled Vocabulary: ESRB rating and audience age https://www.esrb.org/ratings-guide/	M: Mature 17+
VGMS:suggestedRetailPrice	http://metadataregistry.org/uri/schema/VGMS/100147	Suggested Retail Price	The manufacturer's suggested retail price (MSRP) at the time of initial release in the region where the video game was released.	Optional	Float, with currency, source, and date in YYYY/MM/DD format	19.99 USD, Walmart.com, 2002/08/19

IV. *Policies*

C. Deposit Policy

Determining Relevance

Collected data must relate to the user base, reception, and/or distribution of video games. The data may refer to multiple or a single title. Reception may refer to either critical or popular reception, including both professional reviews as well as user-submitted ratings. Distribution may refer to physical as well as digital sales or download data, which may be more appropriate to track the distribution of titles available free of charge. The data submitted should be curated using APIs directly from the source. Potential sources include streaming platforms, console stores, developer websites, and critic websites.

While the repository primarily gathers data through opportunistic means, it also welcomes original observations or corrections to existing datasets via submission. Developers who wish to submit their own work are encouraged to ensure their data is not already in the repository. To that end, the repository will publish information related to regular data scraping initiatives, including when data scraping occurs and where the data is scraped from. Users are encouraged to submit data that has been scraped, but the repository will also scrape data from sites that have continued value to our users. Download statistics will be monitored to convey in-demand sources. Data is only scraped from sources that welcome such practices. Permission from these resources must be granted prior to each data scrape, either through direct communication with a representative or the resource notes on its platform that data scraping initiatives can occur. Current sources for scraping done by the repository include the Steam Store, Playstation Network (PSN), and ESRB. Scraping efforts from these in-demand sources are scheduled to be done on a quarterly basis. Those submitting data should be aware that repository users have the option to filter out user-submitted datasets from their searches.

Minimum Viable Package and Maximum Collection Size

All submissions must clearly indicate licensing, including any restrictions of reuse. The repository requires all submissions to have open licensing, such as ODC-By, ODbL, PDDL, or CC0, to ensure all data contained therein can be used freely. Files using proprietary software such as Microsoft Excel or Macintosh Numbers will need to be converted into an acceptable file format. Submissions must include a data directory or a README file to clarify what each variable represents. Additional contextual information, such as when and how the data was collected and by whom is also required. Each submission will require the submitter to input particular metadata elements, including a title, a brief

description, and a selection of relevant tags pulled from a controlled subject vocabulary. Each submission should be no more than 4 GB in size.

File and Format

During the data submission process, the user will be made aware of the repository's accepted file formats, which will be used to both store the data and share that data with end users through a download option. Upon submission of proprietary file formats, users will be notified that they must transfer their files into an accepted format. They will also be provided with links to corresponding tutorials and guides to help them transfer their content. As part of an accessible DIP for the end user, submitted files will be copied and transformed from their original formats and deposited to end users as .xls or .xlsx. End users will be able to download the non-proprietary formats of the data as well as the transformed .xls or .xlsx versions.

Formats					
Acceptable		**Unacceptable**			
All submissions must contain the dataset in at least one of the following formats:	These formats are acceptable if there are supplementary graphic representations of a dataset:	Unacceptable text, spreadsheet, and image formats include:	Unacceptable formats include all files formats for vector graphics, audio, video, virtual reality, computer programs, and presentations.		
.txt .csv .xml	.pdf .png	.css .dtd .rtf .html .sgml .odt .docx .xls / .xlsx .sxc .ods .dbf .tiff .jpg .bmp .gif	.svp .cgm .swf .aif / aiff .wav .au .mid / .midi .flac .m4a .aac .mp3 .aifc .snd .ra .rm	.ram .wma .m4p .ogg .mp4 .mj2 .avi .mov .mpg / .mpeg .rv .wmv .x3d .wrl .vrml .c	.cpp .java .js .jsp .php .pl .py .class .ppt / .pptx .sxi .odp

D. Ingest Policy

Data Transfer Policy

Uploading Procedure

Users or curators will upload data to the repository through the user interface. Those who wish to submit data must create a free account with the repository, which will require an email address. Once an account is created, the user will be asked if they wish to submit data. Users will also be asked if they want to submit data upon logging in, and they will be prompted to respond with "submit data now," "not at this time (ask later)," or "never (do not show this again)." If a user wishes to submit data, they will be asked to select and upload a file. During this process, the user will be reminded of acceptable file formats as well as other submissions standards. The user will also provide metadata regarding their submission's title, source, license, collection methodology, and temporal coverage.

While users are encouraged to submit a readme file or a data dictionary as part of their data package, the submission process will also feature a webform that will automatically populate once a dataset is uploaded. Users will have the option to edit this webform for accuracy. This webform will ensure curators do not have to download contextual documentation to access the information. It will also allow users unaccustomed to creating readme files or data dictionaries to submit their datasets. For those who wish to learn more about best practices regarding dataset creation, the repository will provide accessible tutorial material, guides, and templates.

In addition to user-submitted metadata, the submission process will automatically generate metadata regarding the account associated with the submission as well as the date and time of submission. If the dataset is later updated, the updating process will automatically generate the date and time of the update as well as apply a version number to represent how many times a dataset has been altered since its initial submission.

Each data collection will be automatically assigned a persistent identifier, which will serve as a unique identifier within the repository. Specifically, the repository relies on Persistent Uniform Resource Locator (PURL) as the identifier, so that the dataset can be accessed in perpetuity. In addition to the PURL identifiers, users will be able to use a dataset's title, keywords, and/or creator as access points.

Authenticity and Security

The repository will create, store, and maintain checksums to ensure that fixity is intact. These will be conducted on a monthly basis to detect file corruption. In order to ensure submitted files align with their corresponding

metadata webform, the repository will utilize a bot that will check to ensure webform submitted metadata--such as variables--appear in the submitted files as a cursory authenticity check.

Upon ingest, submitted files will be scanned for viruses and malware. Users will also be asked if the datasets include any PII, including but limited to: name, initials, identification numbers, drivers' license numbers, birthdate, home address, phone numbers, email addresses, usernames, and unique device identifiers like IP addresses. If the user indicates that this information is present in the dataset, the repository rejects the submission and suggests how PII can be removed before submitting again. As part of the submission process, users will share a list of variable names. Specific variable names that might contain PII, such as "email address," will flag the dataset for review by repository staff.

The submitting user will be notified that their submission has been flagged for review, and consequently the submission will not appear in the repository until the review process is complete. This notification will briefly describe how the review process prevents datasets including PII from appearing in the repository and why such a thing is important. Once a data curator has reviewed the dataset, the submitting user will be informed whether or not their submission was ultimately accepted. If there appears to be PII present, the dataset will be rejected. Notification of rejection will always include feedback on how to edit the data package so that if the recommended changes are made, the submission will be accepted.

The repository reserves the right to remove datasets that include PII as well as those with metadata found to be inaccurate, including misrepresented licensing agreements or false collection methodology.

E. Transformation Policy

Data Normalization

Data transformations will involve applying consistent variable names as well as labeling and standardizing values. Based on a sample of six open data sources, several common variables were identified and a preferred variable name selected. With the exception of "sales," preferred variable names are singular. Each word is capitalized, and spaces are indicated with an underscore. When submitters create a data dictionary, the repository's preferred terms will be listed; the submitter will be encouraged to make adjustments to their data to reflect these preferred terms. Upon ingestion, the repository will use OpenRefine to adjust any non-preferred variable names to the six preferred variable names that are recurring in each dataset.

Preferred Variable Name	Non-preferred Variable Names
title	Name, Game
year	Release, Release_Date, Year_of_Release
genre	Genres, Category
platform	Platforms, Console, Device
developer	Developers, Creator
publisher	Publishers
criticScore	Meta_Score
criticCount	Count
userScore	Score
userCount	Count
esrbRating	Rating
urlL	Link, Hyperlink
naSales	North_American_Sales
euSales	Europe_Sales
jpSales	Japan_Sales
otherSales	Other
globalSales	Worldwide_Sales

Preferred variable names use capitalization and camel case to create consistency in variable naming (i.e. price, requiredAge, averagePlaytime, etc). Users will be made aware of this upon submission and are encouraged to follow this format, but the repository's staff will make adjustments as needed using data normalization tools such as OpenRefine.

The submitter is responsible for ensuring their dataset is sufficiently tidy. The repository utilizes a bot that detects potentially untidy data, such as cells that include two or more values rather than parsing out the information. In extreme instances of untidy data, the repository reserves the right to remove untidy datasets. For datasets that need minor tidying, a professional at the repository can offer feedback to the submitter. Datasets that were removed can be resubmitted with changes, but a professional at the repository will need to ensure that appropriate adjustments have been made before making the dataset live to the public.

Amongst the various help guides available, the repository will also provide documentation outlining the importance of tidy data for the purpose of computation and reuse. This documentation will include examples featuring a tidy and untidy version of the same dataset. The tidy version will represent changes made based on the repository's own step-by-step guide. Null values will not be left blank. For continuity, cells that are left blank or have the value "NULL", "NA", or "N/A" will be converted to "-". In the data dictionary, it is requested that the submitter indicate what "type" of value is associated with a given variable. Potential "types" include text, number, boolean, and categorical.

Notifying Users of Changes

Changes to a dataset will automatically be recorded in the metadata under the "version" and "updated" variables. Dataset versions will start at "1.0" upon submission, unless the submitter indicates otherwise, and any subsequent changes will increase the version number (i.e. 2.0, 3.0, etc). The "updated" variable relates to the date and time that the most recent update occurred, and this will automatically update within the system.

F. Licensing Policy

Preferred Licensing

Good Games! Video Game repository strives to provide its users with open data free of restrictions surrounding reuse. The following licenses are the least restrictive and are therefore preferred licensure for submissions within the repository:

CC-0 ***Creative Commons Public Domain Dedication***

This license is one of the open Creative Commons licenses and is like a public domain dedication. It allows you, as a dataset owner, to use a license mechanism to surrender your rights in a dataset when you might not otherwise be able to dedicate your dataset to the public domain under applicable law.

PDDL ***Open Data Commons Public Domain Dedication and License***

This license is one of the Open Data Commons licenses and is like a public domain dedication. It allows you, as a dataset owner, to use a license mechanism to surrender your rights in a dataset when you might not otherwise be able to dedicate your dataset to the public domain under applicable law.

Other Accepted Licensing

The repository will also allow licensing with more limitations, such as requiring those reusing or manipulating the dataset to attribute the original dataset owners. The repository will never attempt to override more restrictive licensing. The following licenses will also be automatically accepted by the repository's submission process:

CC-BY ***Creative Commons Attribution 4.0 International***

This license is one of the open Creative Commons licenses and allows users to share and adapt your dataset so long as they give credit to you.

ODC-BY ***Open Data Commons Attribution License***

This license is one of the Open Data Commons licenses and allows users to share and adapt your dataset so long as they give credit to you.

Exceptions

If a dataset is submitted with licensure that differs from those listed above, the submitter will be asked to indicate their wish for an exception on the webform and briefly explain their reasoning in the designated field. Their submission will be flagged for review, at which point the applicability of the requested exception will be evaluated on a case-by-base basis. Exceptions may be granted to developers, publishers, research institutions, or other users with a sufficiently compelling interest in ensuring their data is not used commercially. Exceptions will not be granted to those seeking to publish their dataset without any licensure.

G. Preservation Policy

Preservation Guarantees

The acceptable formats listed above are widely used, openly documented, and supported by a range of software platforms, which makes it easier to ensure that reuse can occur. Our repository is also committed to preservation actions such as creating preservation metadata, including running checksums, and maintaining both onsite and offsite backup copies of submissions.

V. *References*

Carpenter, M., Clarke, R. I., Lee J. H., Tennis, J. T. (2015, June). *Developing a Video Game Metadata Schema for the Seattle Interactive Media Museum*. ResearchWorks Archive. https://digital.lib.washington.edu/researchworks/handle/1773/33392

Data.World Help Center. (2021). Common license types for datasets. Data.World. https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets

Digital Preservation Coalition. (n.d.).*Persistent Identifiers*. Digital Preservation Handbook. https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers

Dublin Core Metadata Initiative. (2020, January 20th). *DCMI Metadata Terms*. Dublin Core. https://dublincore.org/specifications/dublin-core/dcmi-terms/

eCommons. (2021, April 27th). Recommended File Formats. Cornell University Library. https://guides.library.cornell.edu/ecommons/formats

Game Metadata Research (GAMER) Group. (2013, June 3rd). *Video Game Metadata Schema*. Open Metadata Registry. http://metadataregistry.org/schemaprop/list/schema_id/132.html

IDEALS. (2018, March 14th). Digital Preservation Support Policy. Illinois Wiki. https://wiki.illinois.edu/wiki/display/IDEALS/Digital+Preservation+Support+Policy#DigitalPreservationSupportPolicy-Category1

OLAC Catalogers Network. (2021). *Alphabetical List Of Genre Terms In The OLAC Video Game Genre Vocabulary*. OLAC Inc. https://www.olacinc.org/alphabetical-list-genre-terms-olac-video-game-genre-vocabulary

Seattle IT. (2016, July 11th). *PII/Privacy in the Open Dataset Inventory*. City of Seattle. https://www.seattle.gov/Documents/Departments/SeattleIT/OpenDatasetInventory_Privacy_PII.docx

UCSC-Stanford Libraries. (2021). *Computer Game Platform*. Game Metadata and Citation Project Controlled Vocabularies. https://gamemetadata.soe.ucsc.edu/platform

Watson, Alex. (2021, January 4th). *Automate Detecting Sensitive Personally Identifiable Information (PII)*. Gretel. https://gretel.ai/blog/automate-detecting-sensitive-personally-identifiable-information-pii-with-gretel