Term Project -Final Project First Draft:

Animal Crossing New Horizons Catalog

Alex G. Williford

Information School, University of Washington

LIS 545 B Wi 24: Data Curation I: Fundamentals

Dr. Matthew Mayernik

February 23rd, 2024

Term Project - Data and Metadata Profile:

Animal Crossing New Horizons Catalog

Alex G. Williford

Information School, University of Washington

LIS 545 B Wi 24: Data Curation I: Fundamentals

Dr. Matthew Mayernik

January 26th, 2024

Animal Crossing New Horizons Catalog

Introduction

In 2020, Nintendo released Animal Crossing New Horizons, a life simulation game that shook the world (Khan, 2020). The game focuses on the player decorating their dream island while making friends with the visiting anthropomorphic animals known as villagers (Li, 2021; *Animal Crossing*[TM], n.d.). Players are encouraged to curate the world around them, from decorating their houses to designing their island and inviting certain villagers to stay on their island. With a furniture item count of 1,997 (*Furniture*, 2023), 260 more items than the previous edition (*Furniture*, 2023), dedicated fans have created the Data Spreadsheet for Animal Crossing New Horizons, which is comprehensive data on the entire game's catalog. Jessica Li's dataset titled Animal Crossing New Horizons Catalog was created from the original spreadsheet's 2020 version, and there are no expected updates (Li, 2021).

Stakeholders regarding the data include Japanese video game company Nintendo and the staff that helped create it, game distributors, journalists, the game's established fanbase, first-time players, contributors on the original spreadsheet, and Jessica Li, creator of the data set. Researchers and data analysts are also stakeholders in this data since it could potentially be used in academic research.

Dataset Breakdown

The data in question is the Animal Crossing New Horizons Catalog created by Jessica Li. No specific software is needed to access this information; just a basic spreadsheet application or text editor will do. Even though the data set has CC0: Public Domain or no copyright, the creator of the dataset recommends citing the contributors on the spreadsheet. The data files consist of 30 csvs of items, villagers, clothing, fish, bugs, and art, just to name a few. It does not contain images from the spreadsheet.

accessories.csv	achievements.csv	art.csv	bags.csv	bottoms.csv
construction.csv	dress-up.csv	fencing.csv	fish.csv	floors.csv
fossils.csv	headwear.csv	headwear.csv	insects.csv	miscellaneous.csv
music.csv	other.csv	photos.csv	posters.csv	reactions.csv
recipes.csv	rugs.csv	shoes.csv	socks.csv	tools.csv
tops.csv	umbrellas.csv	villagers.csv	wall-mounted.csv	wallpaper.csv

Table 1. Displays an alphabetical list of the 30 csvs included in the Animal Crossing New Horizons Catalog dataset.

Metadata Breakdown

While the metadata embedded in the files is incredibly comprehensive, the metadata for the dataset leaves much to be desired. On the metadata section on Kaggle, the only things listed

under the sections were the collaborator's name (Jessica Li), the CC0: Public Domain rights, and the expected updated frequency, which is never. Sections such as DOI citation, provenance, coverage, and authors were left empty. The authors and the provenance fields are possibly left empty since Li explained that she got the data from the context section. However, further information should be expounded, especially in the provenance field. Knowing why she decided just to carry over just the main parts of the spreadsheet and not the key to understanding the metadata or the images would help the user understand the dataset more.

Metadata embedded in the 30 csv files comprises 630 columns (Li, 2021). Each file usually contains metadata such as the name, the price you can usually buy an item, the prices at which you can sell the item, where to find items, and availability, depending on the season. Other entries might include source notes or an internal number. In saying this, no current structure is applied to the metadata besides the information in csv files. This information should be noted because the original spreadsheet has a small section listing abbreviation/spreadsheet help.

Additional information to help users discover the data set in a data repository is a DOI, subject tags, and a more distinct title. Animal Crossing is a trendy topic, so having appropriate subject tags and a DOI allows for a more satisfactory experience when looking for a dataset like this one. The "Read Me" document could also contain a definition of terms that may be unclear for people unfamiliar with the game's jargon.

Publications

After searching Google and EBSCO Host, as far as I could find, no current publications cite the data set as a source. To search for this data set, I took a couple of approaches. I first searched by the title of the data set and the author, then the author and the data set name separately. In Google Scholar, I looked up Animal Crossing data sets and searched the sources for the data set's author. I repeated these same steps on EBSCOHost. Even though no results appeared for the data set, I got several results referring to the spreadsheet. As noted on the spreadsheet, this information has been used by journalists, video game enthusiasts, apps, and walkthrough websites (*Data Spreadsheet for Animal Crossing New Horizons*, n.d.).

On Kaggle, the discussion boards related to the data set have comments about it being used in exercises from the Google data analytics classes (*Animal Crossing New Horizons Catalog - Discussion*, 2024). When this paper was written, 52 corresponding Dataset Notebooks on Kaggle used Jessica Li's data set to create an entirely new project.

Bibliography

Animal Crossing New Horizons Catalog - Discussion. (2024).

  https://www.kaggle.com/datasets/jessicali9530/animal-crossing-new-horizons-nookplaza-datas

  et/discussion

Animal Crossing^{TM}: New Horizons. (n.d.).

  https://www.nintendo.com/us/store/products/animal-crossing-new-horizons-switch/,

  https://www.nintendo.com/us/store/products/animal-crossing-new-horizons-switch/

Data Spreadsheet for Animal Crossing New Horizons. (n.d.). Google Docs. Retrieved February

  3, 2024, from

  https://docs.google.com/spreadsheets/u/1/d/13d_LAJPlxMa_DubPTuirkIV4DERBMXbrWQs

  mSh8ReK4/edit?usp=sharing&usp=embed_facebook

Furniture. (2023, August 20). Nookipedia. https://nookipedia.com/wiki/Furniture

Khan, I. (2020, April 7). Why Animal Crossing Is the Game for the Coronavirus Moment. *The*

  New York Times.

  https://www.nytimes.com/2020/04/07/arts/animal-crossing-covid-coronavirus-popularity-mille

  nnials.html

Li, J. (2021). *Animal Crossing New Horizons Catalog*. Kaggle.

  https://www.kaggle.com/datasets/jessicali9530/animal-crossing-new-horizons-nookplaza-datas

  et

Term Project - Repository Profile:

Animal Crossing New Horizons Catalog and [Zenodo](#)

Alex G. Williford

Information School, University of Washington

LIS 545 B Wi 24: Data Curation I: Fundamentals

Dr. Matthew Mayernik

February 9th, 2024

For Jessica Li's dataset titled Animal Crossing New Horizons Catalog (ACNHC), the repository Figshare seemed to be a good match. Figshare is an open-access repository with the goal of making research outputs accessible in the ways of citing, sharing, and easy discovery (*What Is Figshare?*, n.d.). As a Committee on Publication Ethics (COPE) member, Figshare has a mission of upholding ethical practices of transparency and academic integrity (*What Is Figshare?*, n.d.). While the repository is open for most submissions, the platform's scope focuses on non-traditional research outputs, defined as "anything that is not a journal article such as figures, supplementary figures, and theses (*Inappropriate Content on Figshare*, n.d.)." The guidelines include but are not limited to copyrighted articles, submissions with sensitive information, or submissions with no academic value and only apply to the uploader's institution. After digging around Figshare's submissions, there was some second-guessing about whether this was the right or good repository for the ACNHC. Unsure about what constituted "academic value,(*Inappropriate Content on Figshare*, n.d.)" this prompted a look at Zendo, another open-access repository. It was developed by researchers who value Open Science (*Zenodo - About Zenodo*, n.d.), a movement that believes research outputs should be available to anyone regardless of background (Science.Gov - Open Science Announcements from Federal Agencies, n.d.). In this repository profile, we will examine Zenodo's open-access repository and how its policies and features appropriately match the ACNHC dataset.

Zenodo has a broader scope than Figshare, stating in their policies that their scope covers "all fields of research" and "all types of research artifacts (*Zenodo - Documentation - General Policies v1.0*, n.d.)." There are no restrictions on who can register and upload content. The repository asks that submitters have appropriate rights for the content they upload, uploads in

English are preferred for textual documents but not required, and they should note that there is a

50GB file size per record (*Zenodo - Documentation - General Policies v1.0*, n.d.). Zenodo also

accepts all data file formats but is working on proper documentation to help foster preservation

practices. Data types include publications, posters, lessons, video/audio, images, presentations,

datasets, and a category titled other.

Regarding uploading the ACNH dataset, Zenodo offers a potential data submitter plenty

of documents and guides to consult. In particular, there's the Quick Start section, which breaks

down publishing into three steps, or for more in-depth instructions, one can look underneath the

Deposit section of the FAQ page for the Creating an Upload subpage. The main steps are

creating an account, uploading the desired files, filling in the required metadata fields, and

publishing. For further information concerning the required metadata, a *Describe Records* page

goes into the eight required metadata fields. The required fields are Digital Object Identifier

(DOI), Resource types, Titles, Publication date, Creators, Description, Licenses and rights, and

Contributors. Zenodo's policies state that metadata is stored in JSON format following a JSON

schema (*Zenodo - Documentation - General Policies v1.0*, n.d.) and can be exported in formats

such as Dublin Core, DataCite Metadata Schema, and MARCXML.

Creating an account is straightforward, but users can find it in the *Get Help* section on

Zenodo's help website if needed. There are three ways to create an account: signing up with

ORCID, GitHub, or creating an account with your email address and password (*Zenodo -*

Documentation - Create New Upload, n.d.). It is to be noted that one does not need an account

to download data or export it as a JSON file directly. When downloading data underneath the

Files tab on the upload page, the user can download single files or a ZIP folder with all the files included. What exactly comes in the dissemination information package (DIP) depends onwhat the author decided to upload. This being said, Zenodo offers the option to download the metadata for the upload, such as Dublin Core XML files to BibTeX files.

While Zenodo's FAQ page is full of guides, its support page is only a form allowing users to submit their issues through a form and upload a file or image. From there, a representative will assist you with your issues. Zenodo also has an open dissemination policy that promotes preservation and access to all (Zenodo - About - Terms of Use v1.2, n.d.). They also state under the "How do I share personal data tab" on the FAQ page that the uploader is responsible for ensuring that the information aligns with open dissemination and doesn't break any laws (*Zenodo - Help - Frequently Asked Questions*, n.d.).

After examining Zenodo, its policies, practices, and upload documentation, I decided it was the appropriate repository for the ACNHC. Zenodo has an open repository, which gives an advantage since anyone curious about the information can easily access, share, and view the dataset. No paywalls align with the dataset's CC0: Public Domain license. Since the ACNHC dataset is not a traditional piece of academic data, having a repository that accepts "all types of research artifacts" (*Zenodo - Documentation - General Policies v1.0*, n.d.) means that it wouldn't be taken down by the repository for not containing enough academic value. The process to upload data is also accessible to academics and non-information professionals with clear, direct instructions to guide the user through the process. With this in mind, Zenodo is the preferred repository for the Animal Crossing New Horizons Catalog dataset.

Bibliography

about figshare. (n.d.). https://knowledge.figshare.com/about

Inappropriate Content on Figshare. (n.d.).

   https://help.figshare.com/article/inappropriate-content-on-figshare

Science.gov - Open Science Announcements from Federal Agencies. (n.d.). Retrieved February

   29, 2024, from https://open.science.gov/

What is Figshare? (n.d.). https://help.figshare.com/article/what-is-figshare

Zenodo - Documentation - Create new upload. (n.d.).

   https://help.zenodo.org/docs/deposit/create-new-upload/

Zenodo - About - Terms of Use v1.2. (n.d.). Retrieved February 29, 2024, from

   https://about.zenodo.org/terms/

Zenodo - About Zenodo. (n.d.). https://about.zenodo.org/

Zenodo - Documentation - General Policies v1.0. (n.d.). https://about.zenodo.org/policies/

Zenodo - Help - Frequently Asked Questions. (n.d.). Retrieved February 29, 2024, from

   https://help.zenodo.org/faq/

Long-Term Preservation Considerations

  The 30 files within the dataset are CSV files, which are not proprietary to any company and are inoperable with many different programs. Common programs users open CSVs with are Microsoft Excel and Google Sheets since the data usually works well in spreadsheets. Still, this data can also be used with a plain notepad application. While there is no immediate danger of this file format going obsolete, it is recommended that the files be routinely checked for bit rot, and the parties responsible for upkeep should be aware of any change of technology that could make CSVs unusable.

Copyright License

It is to be noted that while the video game Animal Crossing New Horizons is owned by Nintendo, the dataset and the information within it are set to be in the Public Domain (CC0: Public Domain.) **Below is an example statement:**

- *Li, Jessica. "Animal Crossing New Horizons Catalog Version 3.0" (2020). Kaggle data sciencecompany.https://www.kaggle.com/datasets/jessicali9530/animal-crossing-new-horizons-nookplaza-dataset/data CC0: Public Domain. The person who associated a work with this deed has dedicated the work to the public domain by waiving all of his or her rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law.https://creativecommons.org/publicdomain/zero/1.0/*

Human Subject Considerations

This dataset has no human subject considerations since no personally identifiable user data was collected.