Improving NaNDA Discoverability:

Removing Barriers from a Research Gem

Team Members: Sarah Dababneh (dababneh@umich.edu), Danae Jarrett (danaej@umich.edu), & Lucas Rosendall (lukerose@umich.edu)

Client: The National Neighborhood Data Archive (NaNDA), University of Michigan Institute for Social Research (ISR).

Primary Contact: Lindsay Gypin - Project Senior Manager, SRC-Social Environment & Health (gypin@umich.edu)

Secondary Contact: Philippa Clarke - Research Professor, Survey Research Center (pjclarke@umich.edu)

Course: SI 699 - LAKES Mastery Course

Problem Statement: How should neighborhood-level data be described and represented for easier discovery and reuse?

Word Count: 4472

PART I: REPORT **3**
Executive Summary 3
Section 1 - Introduction: Metadata Systems for Social Science Research Communities 3
1.1 Overview 3
1.2 Motivation & Context 4
1. 3 Problem Statement 5
Section 2 – Background and Research Context 6
2.1 The NaNDA Repository & Its Publishing Environment 6
2.2 The Vocabulary Mismatch Problem 6
Section 3 – Research Methodology 8
3.1 Timeline 8
Section 4 – Findings 11
4.1 User Metric Statistics 11
4.1.2 Thematic Domains & NaNDA Coverage 12
4.1.3 Email Subject Line 13
4.1.4 Publications & Research Topics 14
4.2 User Interviews 16
4.3 Information-Seeking Behavior Survey 16
Section 5 – Knowledge Representation Analysis 17
5.1 Content/Application Distinction 17
5.2 Lack of Rich Relations Between Concepts: : The Scope Relation Problem 18
PART II: METADATA APPLICATION PROFILE (MAP) **19**
Section 6 – Metadata Application Profile: Field Specifications 19
6.1 Field Specifications Table 19
Section 7 – Controlled Vocabulary Crosswalk: NaNDA Domain Coverage 22
7.1 MAP Design Rationale 23
Section 8 – Hierarchical Vocabulary Trees with Node Relations 23
PART III: RECOMMENDATIONS & FINAL REMARKS **24**
Section 9 — Recommendations 24
9.1 Next Steps 24
Section 10 – Conclusion 25
Appendix 26
Appendix A — Link to MAP Document 26
Appendix B — NaNDA Search Term Crosswalk 26
Appendix C — Exploratory Data Analysis: User Metric Data 34
Appendix D – Example Tree Structures 36

PART I: REPORT

Executive Summary

Through a lens of findability and accessibility, this project laid the groundwork for transforming the National Neighborhood Data Archive from underutilized repository to a discoverable research resource. Metadata schema enhancements were made to connect neighborhood-level data with researchers, policymakers, and users.

Section 1 - Introduction: Metadata Systems for Social Science Research Communities

1.1 Overview

Computational Social Scientists have unique information needs that require specialized resources to conduct research and scholarship. These users conduct a range of research-related tasks characterized by the same basic language–a model *M* of how the world works, an inquiry I, a data strategy D, and an answer strategy A. The four together make MIDA, where the *model M* and *Inquiry I* provide a theoretical setting in which a research question can become answerable, and the *Data D* and *Answer A* compromise the empirical strategies for the collecting and summarizing of information (Humphreys et al., 2023). Understanding the information-seeking behaviors – the questions and the research strategies – in which an individual social scientist conducts their study is becoming evermore important as disciplinary insight is increasingly unlocked by data-intensive science. The National Neighborhood Data Archive (NaNDA) is one such initiative that requires further investigation into the system and user-design needs of specialized user communities, namely social scientist researchers interested in the built environment - physical, environmental, economic, and demographic data at multiple levels of spatial scale. This project seeks to address the obstacles and challenges that come with designing

metadata systems that support specialized scholarly-work in the social sciences. We want to specifically know about the impact of global and local metadata description on disciplinary fields that possess idiosyncratic needs: all in an effort to help produce findable, accessible, interoperable, reusable research for the NaNDA scholarly community. To interrogate these questions, we conducted some initial exploratory analysis of the NaNDA user metric data, surveyed the information-seeking behavior of the user population, and conducted semi-structured interviews with researchers. With this methodology in hand, we are at the most fundamental level hoping for information that supports the development of a metadata application profile (MAP) for the NaNDA repository.

1.2 Motivation & Context

The M in MIDA describes a speculation about a possible world, not the world itself. What causes what and how? A model in this sense is a proposition about how the world works (i.e., structural ableism and the built environment); an inquiry is any research question(s) stated in terms of the mode (i.e., what aspects of the built environment are barriers or facilitators to participation?). Because research-design choices have a particular relationship to methodological study, the model and inquiry stage are reflected in the data and the answer strategy (Humphreys et al., 2023). The data collection strategy would be the set of procedures used to gather information from the world – the *units* selected, *conditions* assigned, and *outcomes* measured – and the answer strategy would be the "full set of procedures that begins with cleaning the dataset and ends with answers in words, tables, and graphs " (Humphreys et al., 2023). With MIDA in mind, how should we think about neighborhood-level data being described and represented for easier discovery and reuse? Metadata underpins the findability, accessibility, interoperability, and reusability (FAIR) of research data management overall. Underutilized datasets can lead to lost

research potential, the squandering of public and grant funding, and barriers to equitable access to data intended for broad research use. Since its founding in 2019, NaNDA has grown in both scope and visibility, leading to questions concerning the necessary descriptive infrastructure needed to ensure that datasets are easily discoverable, understandable, and usable. One of the biggest successes in the push for greater research transparency has been the changing norms surrounding data sharing and analysis code after studies have been published; In fact, it is public repositories that have a much better chance of preserving study information for the future, especially if the repository is actively maintained (Humphreys et al., 2023).

The NaNDA repository, which has its own platform, lives inside the Inter-university Consortium for Political and Social Research (ICPSR) universe. Data is ultimately prepared, shared, managed, and archived through ICPSR's Data Management Design and repository architecture (and is part of the Data Document Initiative (DDI) alliance for describing social sciences study-level information). In the structural morasse, the question we are tasked with answering is how we should accommodate subject terms synonyms across ICPSR, JEL, and MeSH vocabularies so that user-supplied subject terms match up with the vocabulary in these classificatory thesauri. Therefore, through user research design methodologies, our project seeks to operationalize or optimize NaNDA metadata at the level of description for future access use.

1. 3 Problem Statement

NaNDA is home to important local data across the entire US. These include topics such as public health and socioeconomic datasets that inform social research ranging from infectious diseases to food deserts to internet reliability, throughout many, if not every, zip code in the US. At the moment, Metadata creation for a new dataset is largely done by a single data librarian. Descriptive metadata inputs for each

dataset are oftentimes limited by time constraints. Moreover, In the last year or so, NaNDA has grown in both scope and visibility. This growth has outpaced the necessary infrastructure needed to ensure that datasets are easily discoverable, understandable, and usable by the repository's stakeholders. There are instances of disconnect between the ISR researchers supplying datasets to NaNDA and the team members that work to curate accurate metadata for user findability. Consequently, underutilized datasets that are not characteristically described may hinder research potential, which may additionally affect public and grant funding, while also enabling barriers to equitable access for broad public use. Altogether, we want to know how neighborhood-level data should be described and represented for easier discovery and reuse.

Section 2 – Background and Research Context

2.1 The NaNDA Repository & Its Publishing Environment

The National Neighborhood Data Archive (NaNDA) is a publicly available data archive containing measures of the physical, economic, demographic, and social environment at multiple levels of spatial scale. It is an initiative out of the University of Michigan’s Institute for Social Research (ISR) – a leader in developing and applying new social science methodologies with an overall commitment to the education of the next generation of Social Scientists. NaNDA publishes its datasets through openICPSR, ICPSR’s self-depository repository.

2.2 The Vocabulary Mismatch Problem

How does metadata affect the ways in which users access and experience the NaNDA repository? Metadata plays a critical role in organizing and contextualizing open research

datasets online. Though it is oftentimes out of site, metadata drives a user's ability to navigate, discover and access research. While the ICPSR's metadata services provide the frameworks or infrastructure for NaNDA's dataset collection, its subject heading metadata is oriented around parameters most useful for categorizing the social sciences. Such an approach is effective for NaNDA, but it can engender inadequate, or at least incomprehensive descriptors for NaNDA data. In our understanding vocabulary mismatch occurs when the terms used by curators to describe the data may not totally align with the search terms researchers use to discover content. There are three dimensions to what we are calling the vocabulary mismatch problem:

- Confusion between content and application: Metadata terms encode what the dataset contains (Pollution Sites, intersection counts, performing arts centers) but data content oftentimes has no particular relationship to methodological study which in turn makes it difficult to discover research applications.
- Divergences in Disciplinary Vocabulary: NaNDA users come from health sciences, sociology, economics, urban planning, and public policy. Each discipline may use different terminology for overlapping constructs, or at least to some extent, there exists variability in word usage across domains.
- Interoperability gaps between schemas: JEL, and ICPSR Subject Headings classify subject matter into hierarchical relationships from general to specific (MeSH more so than JEL and ICPSR). These tree structures are not currently reinforced in NaNDA's metadata. And currently, no crosswalk exists connecting the more equivalent concepts across the three domains.

Section 3 – Research Methodology

This project was sectioned into four different phases. The first phase, discovery and planning allowed us to learn about NaNDA and the different terminologies they use for their datasets. During this phase we often spoke with our client and were able to understand the workflow for importing a dataset into ICPSR and how the NaNDA librarians created metadata.

Phase two, user research is where we began looking at the users of NaNDA to see where there might be a disconnect between search terms and metadata. To gather information on user experiences and thoughts, we conducted four interviews with various NaNDA users and sent out a survey through our client. These interviews allowed us to gather information on users search habits and their level of understanding when it comes to metadata schemas within ISPSR. Also, phase two included exploratory data analysis (EDA) on ICPSR and NaNDA user metric data in order to identify patterns in search behavior.

Phase three, MAP development and crosswalk began with our team taking the statistics and information from our user research and turning it into useful deliverables. In phase four, we presented and handed off our deliverables and findings to our client.

3.1 Timeline

1. **Phase One: Discovery and Planning:** *What do depositors currently understand vs. misunderstand? How do researchers describe datasets, what are the keywords they think of, and what role might that be playing in the creation of metadata?*
    a. Learn more about decision points and data workflows: how researchers currently deposit information and the processes in place to create metadata. **(2/11/26)**
    b. Audit current metadata on NaNDA–explore NaNDA datasets in ICPSR to examine metadata and number of downloads/citations. **(2/16/26)**

c. Preliminary research on mapping LCSH, MeSH, JEL classification systems onto ICPSR subject headings in order to meet the discoverability and accessibility needs of NaNDA; A thinking through of how we might map depositors keywords onto standardized subject headings. **(2/16/26)**

d. Finalize client version of project plan. **(2/16/26)**

2. **Phase Two: User Research:** *What is affecting the way users find, access, and understand the NaNDA repository?*

    a. User behavior analysis part one: Discover more about researchers from depositors (Phillippa) and current data curators and librarians (Will and Lindsay). **(2/11/26)**

    b. User behavior analysis part 2: Email spreadsheet coding–understanding user language and search habits through the 141 email threads containing user questions. **(2/27/26)**

    c. User behavior analysis part 3: A draft survey or interviews with current user communities to discover how they are using the database; clarifying which default metadata fields are potentially confusing for users; constituting what the research data is being used for and the underlying mechanisms that drive users to NaNDA **(2/27/26)**

    d. Comparative analysis to identify areas of overlap, divergence, and gaps in coverage between depositors, librarians, and the NaNDA user communities. **(3/10/26)**

    e. Mid-term report **(3/10/2026)**

3. **Phase Three: MAP and Crosswalk Development:** *What would it mean to construct metadata improvements to better connect NaNDA's user communities?*

a. Introduction & Context: NaNDA's mission and the openICPSR to ICPSR workflow; why good metadata matters for discoverability **(2/10/26)**

b. Shareable Metadata Principles: **context**: quality, shareable metadata must make sense even in an aggregated environment like ICPSR; **completeness**: records must be discoverable in a shared system; **consistency**: standardized vocabulary application; **controlled vocabularies**: Bridge between researcher language and discoverable terms. **(3/17/26)**

c. NaNDA Metadata Cheatsheet and field-by-field guidelines key: updated language and description crosswalking standardized controlled vocabularies like LCSH, MeSH, JEL, and ICPSR according to NaNDA's core subject classes. **(3/24/26)**

d. Appendix: Crosswalk summaries; before and after of improved metadata; glossary of metadata terms; standards and controlled vocabularies quick reference. **(4/7/26)**

e. Client review of deliverable(s) **(4/16/26)**

4. **Phase Four: Finalization, Next Steps, & Handoff**

a. Finalize documentation package: Prepare MAP, share interview data, and set up next steps for future project(s); integrate any client feedback on all deliverables. **(4/16/2026)**

b. EXPO Presentation **(4/20/26)**

c. In-class final presentation; Complete client communication log **(4/21/26)**

d. Final client meeting and handoff of deliverables **(4/22/26)**

e. Individual project reflections **(4/23/26)**

Section 4 – Findings

4.1 User Metric Statistics

User Metric Statistics performs a multi-method exploratory data analysis (EDA) on NaNDA-related Excel files (**Figure 1**) to answer a central question: How do users search for neighborhood data, and where might gaps exist between their search behavior and the current NaNDA metadata system? The goal of such analysis treats a spreadsheet like evidence and follows the same five step pattern for each dataset – load the data and inspect the data shape and types, identify what needs cleaning if and when necessary, filter, group, join, or reshape each table to match the research question, identify the best way to interpret the results (table, plot, statistic, or other visualization), and question the answer strategy or data assumptions that may be present in your model. To begin, a full analytical stack of imports are loaded into our local environment, including a built-in regex engine (import re), numeric & tabular modules for data science (numpy & pandas), a visualization stack (matplotlib & seaborn), and additional modules for natural language processing (NLP) & statistics (scipy for stats & nltk for NLP). From there the excel files are loaded into the local environment. Extra corresponding data related to our exploratory findings is available in **Appendix C** below.

This analysis is primarily concerned with what can be initially visualized and therefore better understood, but it also reveals user vocabulary gaps and metadata recommendations that support our research findings overall. Many users have questions that are geographical and technical in nature. Additionally, there are specific dataset and methodological inquiries related to social determinants of health. Publication data clusters around healthcare research, or coherent themes like urban health disparities, environmental exposure, education outcomes, transportation access, and food environments. MeSH terms related to socioeconomic, demographic,

environmental inequality, land cover, and other social determinants of health could greatly enrich the scope of NaNDA metadata. In what follows our EDA questions are more thoroughly broken down into their constitutive components.

File	Key Contents
nanda_dataset_catalog_2026-02-18	42 NaNDA dataset titles, DOIs, study IDs
NaNDA_ICPSRstats_Fall2025	ICPSR download counts by institution type, user role, email domain
NaNDAPublications	216 publications that cited NaNDA data – titles, journals, types
NaNDAUsageStats	Longitudinal download and citation counts across 11 scrape dates
NANDAUserEmailSubjects	Emails from users who contacted NaNDA regarding research questions

Figure 1: Five Datasets Available for Analysis

4.1.2 Thematic Domains & NaNDA Coverage

We first chose to work with NaNDA catalog data. From the titlesets, we wanted to extract information about thematic domains, geographic units, and time coverage. To do so, we used regex and simple NLP to tag themes, define geographic units, and to extract temporal coverage. In order to identify thematic domains, for example, we chose to map raw title descriptions to roughly sketched thematic domains using keyword dictionaries (**Figure 2**).

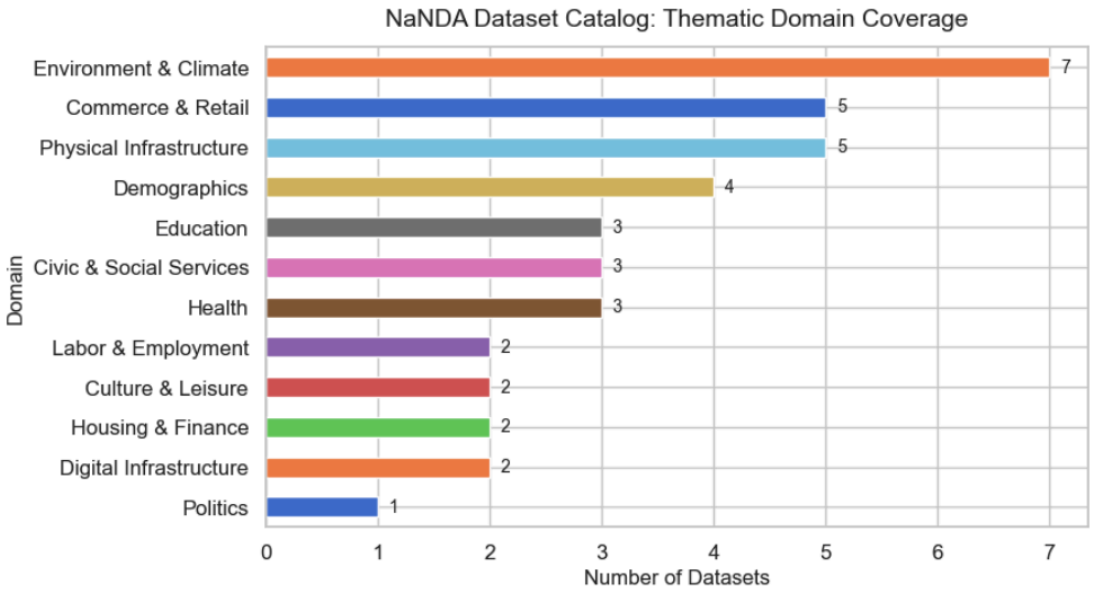


Figure 2: Thematic Domain Tagging with Regex Keyword Matching

4.1.3 Email Subject Line

To understand what users are asking about, we parse 'NANDAUserEmailSubjects.xlsx' to better understand user intent. Since this is unstructured text data, we tokenize raw text for frequency analysis and regex-based query type detection. To begin, we parse and clean email subject strings and refashion email reply threads into subject rows. There are roughly 477 unique email subject lines after cleaning. From there some assumptions are made about stopwords that carry no discriminating information (words like 'nanda', 'icpsr', 'neighborhood', 'hello', ect). Some users had specific questions about a summer workshop, which we chose to leave despite not being directly relevant to our research. Still, some terms are more revealing than others.

Questions regarding "documentation", "variable", and geographic units do arise in the Top 20 Terms list below (**Figure 3**). Of course, some search terms will be inflated by longer email threads, since those threads have been split into individual rows. It is certainly not

foolproof analysis but provides some instances of the vocabulary users use in their email subject headings. Many of the email subject-headings were NaNDA office hours requests – this was something that was also discretionarily left out of the research process in order to make room for other vocabulary terms to surface.

Top 20 Terms

term	count
workshop	96
summer	96
variable	25
net	20
access	18
census	17
social	16
study	14
zcta	14
broken	13
link	13
confirmation	13
tract	13
documentation	12
external	11
published	11
index	11
datasets	11
socioeconomic	11
status	11

Figure 3: User-Email Subject-Heading Vocabulary List

4.1.4 Publications & Research Topics

The `NaNDAPublications.xlsx` file contains 216 scholarly works (journal articles, theses, reports) that cite NaNDA datasets. Understanding what research questions the NaNDA data supports can be tracked back to the disciplines that engage with neighborhood-level research data. But first, the `NaNDAPublications.xlsx` file allows us to track publication type; more importantly, we can also measure publication trends over time (if we treat year as a proxy time index and count publications per year). In the bar chart and line chart below, we see both annual publication count and cumulative growth respectively (**Figure 4**). Further, in an attempt to

understand which academic disciplines draw on NaNDA data most heavily, We classify each of the 216 publications using a journal-first, title-keyword-fallback regex rule set. Publications are first mapped to discipline and then into broader research domains, including Healthcare Research, Social Science, Environmental Science, Information, Social Work, and Interdisciplinary.

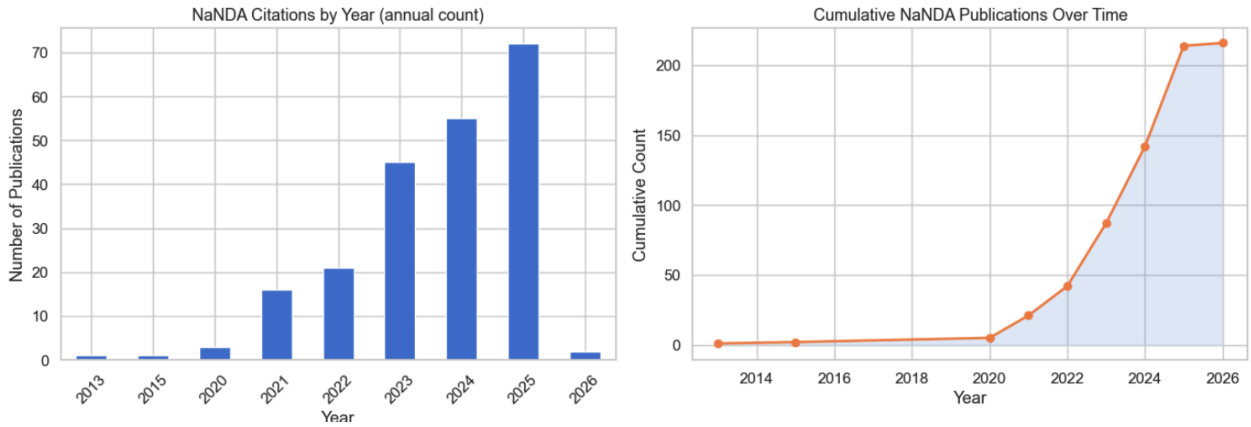


Figure 4: Citation by Year & Publications Overtime

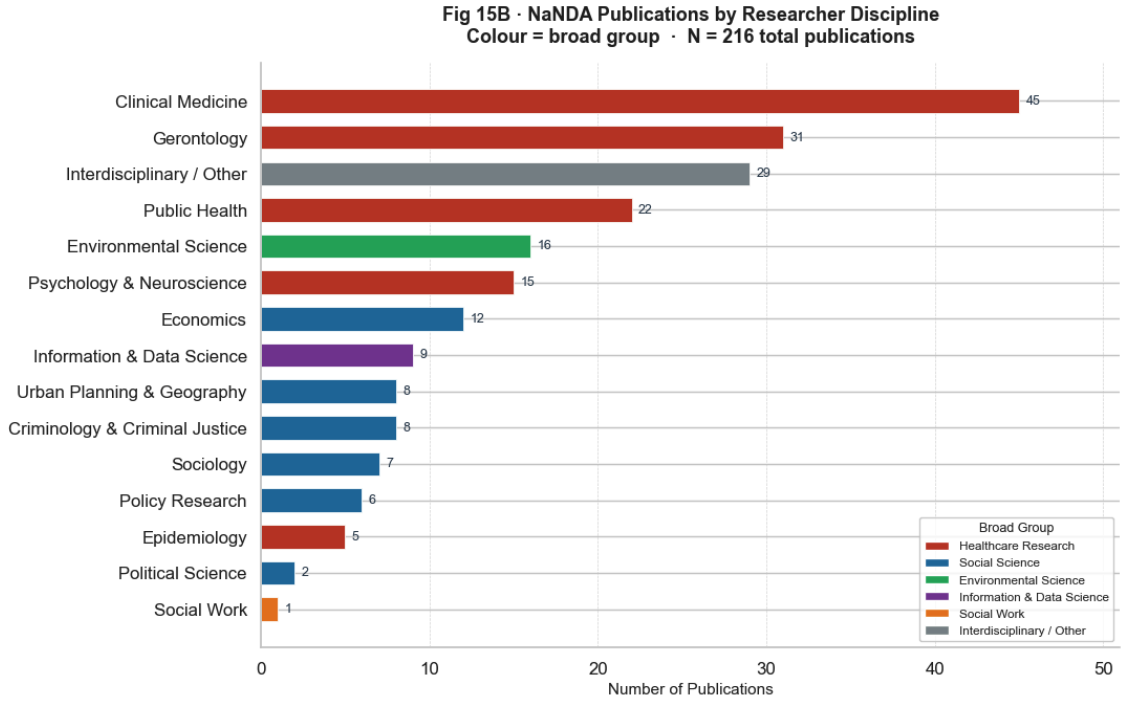


Figure 5: Publication by Research Discipline

4.2 User Interviews

Another method used to gather information about users were semi-structured interviews. We received contact information from our client for four unique NaNDA users. The interviewees included a data manager, a research assistant professor, a doctoral student, and a postdoctoral research student. We conducted these user interviews to understand key issues around metadata within the NaNDA repository and how they shape data discovery. Across interviews, we asked participants about their roles and research needs, how they first encountered NaNDA/ICPSR, how often they use the repository, and what challenges prompted them to seek help from NaNDA staff. We then focused on search behavior and discovery. We asked what terms users try (such as variables, subject terms, and geography), where searches break down, and their experience with documentation and download/versioning workflows. These conversations highlighted recurring difficulties and confusion on documentation, identifying the "right" geography (for example tracts vs. ZIP and shifting boundaries), uncertainty about dataset versions/filenames, and a need for clearer, more filterable metadata that better matches researchers' vocabulary and workflows.

4.3 Information-Seeking Behavior Survey

An anonymous Qualtrics survey was created and distributed amongst NaNDA users in order to gather a more robust understanding of how people come across and interact with the different datasets. The questionnaire sought to find out users' primary research interests, how they first encountered NaNDA, and their personal opinion on the ease of searchability within the datasets and their familiarity with the different search terms that are currently in use. Through a total of 20 respondents, we found that NaNDA's datasets were most commonly found by word-of-mouth recommendation, meaning that users were not likely to find the website

organically through their own search. Research interests ranged from health, economics, public policy, and social sciences. Many users had their specific research needs and needed to make meaning from NaNDA's neighborhood-level datasets. Most users felt comfortable with the interface found on ICPSR and open ICPSR, but still stated that improvements could be made to improve findability of information. Users were most familiar with the MeSH and ICPSR controlled vocabularies, but a majority were unfamiliar with any controlled vocabularies at all. Users suggestions for improvements included creating a dashboard visualization to represent the datasets, narrow dataset downloads to only contain certain variables, and standardization of variable and dataset documentation. These anonymous user insights were helpful in informing the work of creating a smoother process for dataset depositors and researchers to meet community needs.

Section 5 – Knowledge Representation Analysis

5.1 Content/Application Distinction

Overall our project formalizes a content/application contradiction inherent to NaNDA's metadata design. At the moment, data content lacks a particular relationship to methodological study. Of course, this is inherently a difficult thing to assuage – if a dataset is about pollution sites it is about pollution sites and not air pollutants, for example. But our user research data reveals discoverability and accessibility problems: people want to know how a particular dataset applies to their domain of methodological study. At the moment most of the datasets tell researchers what the dataset is without telling them how a particular dataset relates to their workflow. If how the knowledge is represented differs from the information-seeking behaviors of researchers, then what can be done? One could posit that there needs to be richer relations between concepts, one that seeks to improve conceptual and scopus infrastructure in relations to NaNDA's main user community, namely social science researchers interested in the social determinants of health. The fields specification table tries to address this content/application problem

through two differing but complementary structural fields: (1) a dedicated subject- content field requiring vocabulary terms about what the dataset measures, and (2) a separate subject-application field that optionally encodes anticipated research application. Of course, this is only an idea, and if it is not applicable then the problem becomes more about addressing relations between subject term headings.

5.2 Lack of Rich Relations Between Concepts: : The Scope Relation Problem

Computational social scientists routinely need to understand how datasets relate to one another: Which datasets can be merged at the same geographic unit, which datasets cover overlapping time periods, and which datasets operationalize the same theoretical construct using different source data. But descriptive metadata, in our case, can also help with richer relational structure. If descriptive metadata and interface design coordinate according to controlled knowledge nodes and relations, and these relations between concepts move from more general hierarchies to the more specific, then there is an opportunity for more in-depth knowledge representation (**Appendix D**). Moreover, properties of provenance encoded into metadata improve metadata preservation overall. Data provenance tells how descriptive metadata properties come into being. A combination of metadata provenance and in-depth hierarchical classification moving from the more general to the more specific may prove to be a vital aspect of metadata longevity, but more importantly seeks to mitigate a lack of relations between concepts in the NaNDA repository. A Metadata Application Profile is the first answer to this scope relation problem, as it tries to build synonyms in and amongst subject heading namespaces, namely JEL, MESH, and ICPSR.

PART II: METADATA APPLICATION PROFILE (MAP)

Section 6 – Metadata Application Profile: Field Specifications

6.1 Field Specifications Table

This table includes Dublin Core modifiers for already existing field specifications within the NaNDA repository. It describes the necessity of certain elements and how they might apply to datasets and help their description and discoverability. For the most part, these fields already exist across the repository.Nevertheless there are some updates that we hope our client will find useful or at least inspiring. Upon further notice, it may possibly be beneficial to see if any more ICPSR Metadata Elements could be of use to the NaNDA repository, if they are not already included here.

Field Name	DC Element	Status	Qualified Mapping	Definition
Title	dc:title	Mandatory	dcterms:title	Descriptive title naming dataset content, geographic unit, and the time period covered.
Identifier	dc:identifier	Mandatory	dcterms:identifier	The registered persistent digital object identifier (DOI) associated with the data collection.
Creator	dc:creator	Mandatory	dcterms:creator	Principal investigator(s) and institutional affiliation. Linked to ORCID where available.

Publisher	dc:publisher	Mandatory	dcterms:publisher	Distinguishes publication tier (curated vs. self-deposit).
Date	dc:date	Mandatory	dcterms:temporal	Date range of data coverage. Separate from deposit date. ISO 8601 format.
Description	dc:description	Mandatory	dcterms:abstract	Structured abstract: (1) unit of analysis, (2) variable domains, (3) source datasets, (4) spatial scale.
Subject — Content	dc:subject	Mandatory	dcterms:subject	Controlled vocabulary terms describing data CONTENT (MeSH, JEL, ICPSR SH). Must reflect what data measures, not how it is used.
Subject — Application	nanda:applicationDomain	Recommended	schema:applicationCategory	Anticipated research application domains. Encoded separately from content terms.
Spatial Coverage	dcterms:spatial	Mandatory	dcterms:spatial	Geographic unit of observation: Census Tract, ZCTA, County, or State. GeoNames URI preferred.
Temporal Coverage	dcterms:temporal	Mandatory	dcterms:temporal	Start and end year of data series. Encode as ISO 8601 interval.

Spatial Resolution	nanda:spatialResolution	Mandatory	schema:spatialCoverage	Geographic resolution: Tract 2010, Tract 2020, ZCTA, County-FIPS. Link to Census Bureau URIs.
Unit of Analysis	nanda:unitOfAnalysis	Mandatory	schema:variableMeasured	Statistical unit: Census Tract, ZIP Code Tabulation Area, County, Block Group.
Source Data	dcterms:source	Mandatory	dcterms:source	Primary data sources (e.g., NETS, American Community Survey, EPA TRI, USPS). Include version and year.
Variable Names	nanda:variableNames	Recommended	schema:variableMeasured	Key variable names with labels and value definitions. Reference full codebook for complete list.
Geographic Identifiers	nanda:geoIdentifier	Mandatory	schema:geoContains	GEOID format: 11-digit (tract), 5-digit (county), ZCTA5 code. Document crosswalk vintage.
Missing Value Convention	nanda:missingValue	Mandatory	schema:additionalProperty	Explicit encoding of missing value codes (e.g., -1 = CVAP equals zero; -9 = data unavailable).

Relation	dc:relation	Recommended	dcterms:isPartOf	Links to related NaNDA studies (e.g., broader series, predecessor openICPSR version).
Rights	dc:rights	Optional	dcterms:license	Access conditions
Format	dc:format	Recommended	dcterms:format	File formats available: Stata (.dta), R (.rda), CSV, SAS. State software version requirements.
Audience	dcterms:audience	Recommended	schema:audience	Intended user community. Distinguish researchers requiring crosswalk expertise from general users.
Provenance	dcterms:provenance	Recommended	dcterms:provenance	Derivation chain from raw source to NaNDA measure. Include methodology for variable construction.

Section 7 – Controlled Vocabulary Crosswalk: NaNDA Domain Coverage

7.1 MAP Design Rationale

Our MAP is based on our findings from user interviews and survey results. These findings make it evident that researchers search for NaNDA data using specific language that doesn't always correlate with the metadata schemas used in the ICPSR repository. To help reduce the mismatching and improve discoverability, we collected search terms reported by NaNDA users and organized them into research domains such as criminology and public policy. We then crosswalked each term to each of the metadata schemas: ICPSR Subject Terms, JEL, and MeSH (**see Appendix B**). For each line, we added short notes to explain user intent, context, and links across schemas. This design standardizes indexing while preserving the language researchers actually use. It is an example and reference document that can help lead to more consistent metadata application and more successful searching across disciplines.

We also realized that the metadata issue begins with the dataset creators. If they provided a more in depth description of their data with metadata suggestions, NaNDA librarians would have an easier time creating accurate metadata that leads to discovery. To help this, we created a simple guide and explanation for writing better metadata descriptions. This includes clearly stating what the dataset measures, key variables, unit of analysis, and the time and geographic coverage to remove barriers and reduce the workload of data librarians.

Section 8 – Hierarchical Vocabulary Trees with Node Relations

The tables in **Appendix D** provide examples of the hierarchical tree structures within each of the three schemas used within ICPSR datasets. These hierarchical vocabulary trees relate to key NaNDA relevant-domains listed in our Metadata Application Profile. The example tree structures in appendix D below relates to the Race, Equity, and Demographics category – labelled green. From there we dissect an example search term that may fall within the scope of

this category and show how it connects to other terms related to it. This is done with the hope of furthering scopus and conceptual relations between subject term headings MeSH, JEL, and ICPSR. If future work is done in this area, a NaNDA specific controlled vocabulary may begin to emerge. For the most part, each term entry specifies: (1) the terms authoritative heading, (2) its tree number or classification code, (3) its entry terms (synonyms), and (4) its position in the hierarchy from broadest to narrowest. Lastly, (5) we also try to include narrower terms applicable to NaNDA datasets. The tree structure here tries to encode the scope relations required for hierarchical search expansion, and to the broad array of terms users may use to describe their research. For now, it remains incomplete, and our recommendations and final remarks are organized around the implementation of our main deliverable – the MAP.

PART III: RECOMMENDATIONS & FINAL REMARKS

Section 9 — Recommendations

The purpose of the MAP is to serve as a guiding document for data depositors, researchers, and librarians. It can be put into use immediately and be a public-facing document, where it outlines the importance of best metadata practices and standardizes how neighborhood-level data is described in order to improve the findability of the information it contains. The crosswalk across the different controlled vocabularies allows for better interoperability and provides concrete examples for data depositors to best describe their datasets and for researchers to use the best search terms to get the most optimal results.

9.1 Next Steps

The MAP serves as a starting point for improving discoverability of NaNDA's datasets. Standardizing how the datasets and its contents are described will be important future work. This can involve NaNDA creating its own controlled vocabulary based on user search terms and dataset needs. Social science research covers the intersection of many fields of scholarship, which is why it is useful to have a unique vocabulary for researchers and users to refer to.

Section 10 – Conclusion

NaNDA's commitment to providing neighborhood-level data that is foundational to social science research is essential to improving the US's neighborhoods and communities. Improving the discoverability of these datasets ensures that access to this information is not unintentionally lost. Through the implementation of the MAP as a guide and the user experience research, we hope to bridge the gap across stakeholders and reinforce the idea of NaNDA's research and datasets being for the benefit of society as whole.

Appendix

Appendix A — Link to MAP Document

MAP

Appendix B — NaNDA Search Term Crosswalk

Neighborhood Data Repository — Expanded Search Term Crosswalk (ICPSR / JEL / MeSH)

User Search Term	Research Domain	ICPSR Subject Terms	JEL Codes	MeSH Terms	Notes / Neighborhood Context
Crime & Public Safety	Criminology	crime rates crime mapping crime prevention arrests homicide assault burglary drug offenses gang activity police criminal justice system	K42 – Illegal Behavior and the Enforcement of Law K14 – Criminal Law H56 – National Security and War J15 – Economics of Minorities, Immigrants	Crime Violence Criminal Law Police Public Safety Homicide Substance-Related Disorders	UCR/NIBRS incident data, 911 call volumes, hot-spot mapping, arrest rates disaggregated by census tract
Incarceration & Reentry	Criminology / Sociology	incarceration imprisonment prison population recidivism ex-offenders ex-offender employment probation parole	K42 – Illegal Behavior and Enforcement K14 – Criminal Law J15 – Economics of Minorities, Immigrants I32 – Measurement and Analysis of Poverty	Prisoners Recidivism Prisons Criminal Justice Parole Probation Reintegration, Community	Neighborhood-level incarceration rates, reentry services, effects on family stability and local labor supply
Police & Law Enforcement	Criminology / Public Policy	police police-community relations police misconduct assaults on police arrest rates criminal justice policy	K42 – Illegal Behavior and Enforcement D73 – Bureaucracy, Administrative Processes H56 – National Security	Police Law Enforcement Criminal Justice Community-Institutional Relations Police-Community Relations	Officer deployment, use-of-force incidents, stop-and-frisk data, community trust surveys

Racial & Ethnic Composition	Demography / Sociology	race ethnicity racial segregation racial integration racial discrimination ethnic groups Hispanic or Latino Americans Black Americans* Asian Americans White Americans	J15 – Economics of Minorities, Races, Indigenous Peoples, Immigrants J71 – Discrimination R23 – Regional Migration, Neighborhood Characteristics D63 – Equity, Justice, Inequality	Racial Groups Ethnicity Segregation Minority Groups Health Disparities, Minority Racial Discrimination	Census race/ethnicity breakdowns, dissimilarity index, exposure index by neighborhood
Income Inequality & Poverty	Economics / Sociology	poverty income distribution wealth wealth inequality poverty programs low income households affluence	D31 – Personal Income, Wealth, and Their Distributions I32 – Measurement and Analysis of Poverty I38 – Government Policy, Welfare Programs D63 – Equity, Justice, Inequality	Poverty Income Social Deprivation Socioeconomic Factors Wealth Income Inequality	Gini coefficients, poverty rates, median household income, EITC take-up at neighborhood level
Immigration & Migration	Demography / Economics	immigrants immigration status immigration policy refugees foreign born migration* acculturation ethnicity	F22 – International Migration J15 – Economics of Minorities, Immigrants J61 – Geographic Labor Mobility, Immigrant Workers R23 – Regional Migration	Emigrants and Immigrants Transients and Migrants Refugees Acculturation Immigration Policy Foreign-Born Population	Foreign-born share, country of origin, naturalization rates, immigrant enclave formation
Housing & Affordability	Urban Economics / Housing Policy	housing housing costs housing conditions affordable housing rental housing public housing eviction mortgages foreclosure homelessness	R21 – Housing Demand R31 – Housing Supply and Markets R38 – Government Policy (Housing) G21 – Banks, Mortgages I38 – Government Policy, Welfare Programs	Housing Public Housing Affordable Housing Foreclosure Eviction Mortgages Rent	Rent burden, vacancy rates, foreclosure filings, eviction filings, HUD-assisted unit locations

Homelessness & Housing Instability	Public Policy / Social Work	homelessness homeless persons homeless shelters eviction housing instability transient housing	I38 – Government Policy, Welfare Programs R21 – Housing Demand I32 – Measurement and Analysis of Poverty H53 – Government Expenditures and Welfare Programs	Ill-Housed Persons Homeless Persons Housing Instability Eviction Emergency Shelter Transient Housing	PIT count data, shelter utilization, eviction court filings, rapid rehousing program outcomes
Gentrification & Displacement	Urban Planning / Sociology	gentrification neighborhood change housing costs displacement neighborhood conditions residential mobility	R11 – Regional Economic Activity, Growth, Development R31 – Housing Supply and Markets J61 – Geographic Labor Mobility D31 – Personal Income, Wealth	Gentrification Residence Characteristics Urban Renewal Residential Mobility Neighborhood Change	Rent appreciation, demographic shifts, business turnover, long-term resident displacement indicators
Education & School Performance	Education Policy / Economics	education school districts school attendance school dropouts graduation literacy rates test scores educational achievement school choice	I21 – Analysis of Education I22 – Educational Finance, Financial Aid I24 – Education and Inequality I28 – Government Policy (Education) H52 – Government Expenditures and Education	Education Schools Academic Performance Educational Status Literacy School Dropout Achievement Gap	Graduation rates, test score proficiency, chronic absenteeism, school quality ratings by neighborhood
Early Childhood & Child Welfare	Public Health / Social Work	child care child health child welfare child abuse child neglect foster care early childhood education infant mortality	I13 – Health Insurance, Public and Private I21 – Analysis of Education J13 – Fertility, Family Planning, Child Care H53 – Government Expenditures and Welfare Programs	Child Welfare Child Care Child Health Infant Mortality Foster Home Care Child Abuse Early Childhood Education	Head Start access, child maltreatment reports, foster care placement rates, childcare desert mapping

Health Outcomes & Disparities	Public Health / Epidemiology	health disparities health status mortality rates morbidity chronic illness diabetes obesity asthma life expectancy	I14 – Health and Inequality I11 – Analysis of Health Care Markets I12 – Health Behavior I15 – Health and Economic Development	Health Status Disparities Mortality Chronic Disease Life Expectancy Health Inequities Morbidity Diabetes Mellitus Obesity Asthma	Age-adjusted mortality rates, chronic disease prevalence, premature death rates by census tract
Mental Health & Substance Use	Public Health / Behavioral Health	mental health mental health services substance abuse addiction alcohol abuse drug abuse suicide treatment programs	I12 – Health Behavior I13 – Health Insurance, Public and Private I18 – Government Policy, Public Health H53 – Government Expenditures and Welfare Programs	Mental Health Substance-Related Disorders Mental Health Services Suicide Alcohol-Related Disorders Drug Addiction Behavioral Health	Provider access, overdose death rates, treatment facility locations, uninsured behavioral health rates
Caregiving, Dementia & Aging	Gerontology / Public Health / Behavioral Health	caregiving family caregivers informal care dementia Alzheimer's disease cognitive decline aging in place respite care long-term care memory care	J14 – Economics of the Elderly, Economics of Disability I12 – Health Behavior I13 – Health Insurance, Public and Private I18 – Government Policy, Public Health H55 – Social Security and Public Pensions	Caregivers Dementia Alzheimer Disease Cognitive Dysfunction Long-Term Care Aged Respite Care Mental Health	Caregiver burden estimates, dementia prevalence by neighborhood, memory care facility access, Medicaid LTSS enrollment, caregiver support program availability
Health Care Access & Insurance	Health Economics / Policy	health care access health insurance Medicaid Medicare uninsured health care facilities hospitals health care costs	I11 – Analysis of Health Care Markets I13 – Health Insurance, Public and Private I18 – Government Policy, Public Health H51 – Government Expenditures and Health	Health Services Accessibility Insurance, Health Medicaid Medicare Uncompensated Care Health Care Costs Primary Health Care	Uninsured rates, primary care provider density, hospital proximity, Medicaid enrollment by zip code

COVID-19	Public Health / Epidemiology	COVID-19 coronavirus* pandemic infectious diseases* public health mortality rates hospitalization	I18 – Government Policy, Public Health I12 – Health Behavior H12 – Crisis Management E65 – Studies of Particular Policy Episodes J64 – Unemployment Models	COVID-19 SARS-CoV-2 Pandemics Communicable Diseases Public Health Mortality Hospitalization	Neighborhood case/death rates, vaccination coverage, economic disruption, essential worker exposure
Environmental Hazards & Pollution	Environmental Science / Public Health	environmental hazards environmental degradation air pollution environmental monitoring environmental impact hazardous waste water quality	Q53 – Air Pollution, Water Pollution, Hazardous Waste Q51 – Valuation of Environmental Effects Q58 – Government Policy (Environment) K32 – Energy, Environmental, Health and Safety Law	Environmental Pollution Air Pollution Hazardous Waste Water Pollutants Environmental Exposure Toxic Substances Environmental Health	EPA toxic release inventory, lead exposure risk, superfund proximity, air quality index by neighborhood
Climate Change & Resilience	Environmental Economics / Urban Planning	climate change disasters disaster relief flooding environmental planning environmental protection	Q54 – Climate, Natural Disasters, Global Warming Q56 – Environment and Development, Sustainability R11 – Regional Economic Activity H84 – Disaster Aid	Climate Change Natural Disasters Floods Disaster Planning Environmental Health Climate Resilience Urban Heat Island	Flood zone maps, heat island intensity, storm damage records, green infrastructure investment by area
Business Activity & Economic Development	Industrial Organization / Urban Economics	businesses small businesses entrepreneurs retail industry commercial districts business ownership local economy	L26 – Entrepreneurship L11 – Production, Pricing, and Market Structure O11 – Macroeconomic Analyses of Economic Development R11 – Regional Economic Activity	Commerce Small Business Entrepreneurship Economic Development Retail Trade Business Formation	Business license counts, establishment births/deaths, retail square footage, opportunity zone activity
Socioeconomic Status	Economics / Sociology	socioeconomic status social stratification class mobility socioeconomic indicators social class educational attainment occupational prestige wealth accumulation	D31 – Personal Income, Wealth, and Their Distributions I32 – Measurement and Analysis of Poverty J24 – Human Capital, Skills, Occupational Choice D63 – Equity, Justice, Inequality	Socioeconomic Factors Social Class Social Mobility Occupational Status Income Educational Status	Composite SES indices by census tract, intergenerational mobility estimates, occupational prestige scores, educational attainment rates by neighborhood

			I24 – Education and Inequality		
Food Access & Food Security	Public Health / Agricultural Economics	food security food programs food costs hunger nutrition SNAP food supply	Q13 – Agricultural Markets and Marketing Q18 – Agricultural Policy, Food Policy I12 – Health Behavior I38 – Government Policy, Welfare Programs	Food Supply Food Security Hunger Nutrition Food Assistance Programs Food Deserts SNAP	SNAP participation rates, food retailer density, food desert/swamp classification, WIC access
Labor Dynamics	Labor Economics	employment job loss job change unemployment wages essential workers displaced workers occupational mobility workforce job satisfaction	J21 – Labor Force and Employment J31 – Wage Level and Structure J62 – Job and Occupational Mobility J63 – Turnover, Vacancies, Layoffs J64 – Unemployment Models	Employment Unemployment Wages Occupational Mobility Workforce Labor Market	Industry mix, job density, wage levels, commute patterns, unemployment rates at neighborhood scale
Art, Entertainment & Leisure	Cultural Economics / Sociology	arts arts attendance arts audiences arts participation recreation outdoor recreation leisure sports sports participation parks access to arts	L82 – Entertainment, Media L83 – Sports, Gambling, Restaurants, Recreation, Tourism Z11 – Economics of Arts and Literature Z21 – Sports Economics: Industry Studies J22 – Time Allocation and Labor Supply	Recreation Leisure Activities Arts Sports Cultural Participation Community Facilities	Facility counts, participation rates, cultural amenities by neighborhood
Transportation & Commuting	Transportation Economics / Urban Planning	transportation transport infrastructure public transportation commuting air transportation automobile ownership	R41 – Transportation: Demand, Supply, Congestion R42 – Government and Private Investment, Transportation Planning R48 – Government Pricing and Policy L91 – Transportation: General	Transportation Public Transportation Commuting Urban Transport Automobiles Transit Access Active Transportation	Transit access scores, commute times, car ownership rates, bikeability/walkability indices

Recreation	Cultural Economics / Urban Planning / Public Health	parks recreation centers playgrounds open space trails physical activity outdoor recreation community centers leisure access recreation programs	L83 – Sports, Gambling, Restaurants, Recreation, Tourism R14 – Land Use Patterns I12 – Health Behavior H53 – Government Expenditures and Welfare Programs Z21 – Sports Economics: Industry Studies	Recreation Parks, Recreational Leisure Activities Physical Fitness Community Facilities Sports Sedentary Behavior	Park acreage per capita by census tract, recreation center proximity, trail access scores, physical activity rates, park quality ratings, program enrollment by neighborhood
Digital Access & Broadband	Information Economics / Policy	Internet digital communications telecommunications telecommuting technology	L86 – Information and Internet Services L96 – Telecommunications O33 – Technological Change, Diffusion Processes I24 – Education and Inequality	Internet Telecommunications Digital Divide Telemedicine Information Technology Broadband Remote Work	Broadband subscription rates, internet speed tiers, digital redlining, remote work capacity by neighborhood
Built Environment	Urban Planning / Geography	neighborhoods neighborhood characteristics neighborhood conditions housing housing conditions transportation transport infrastructure parks land use	R14 – Land Use Patterns R11 – Regional Economic Activity R41 – Transportation Demand/Supply O18 – Urban, Rural, Regional Analysis	Built Environment Residence Characteristics Urban Health Transportation Parks, Recreational Walkability	Physical/spatial attributes: housing stock, transit access, greenspace, walkability, street design
Aging Population & Senior Services	Gerontology / Public Policy	aging aging population elderly* senior citizens* activities of daily living assisted living adult care services	J14 – Economics of the Elderly, Economics of Disability J26 – Retirement, Retirement Policies I11 – Analysis of Health Care Markets H55 – Social Security and Public Pensions	Aged Aging Geriatric Assessment Medicare Assisted Living Facilities Senior Centers Age-in-Place	65+ population share, senior center access, Medicare enrollment, age-in-place indicators
Disability & Accessibility	Disability Studies / Public Policy	disabilities disabled persons accessibility (for disabled) disability income disability discrimination	J14 – Economics of the Elderly, Economics of Disability I14 – Health and Inequality H53 – Government Expenditures and Welfare Programs	Disabled Persons Accessibility Disability Evaluation Social Security Disability Disability Discrimination	ADA compliance, SSI/SSDI rates, accessible transit stops, disability prevalence by neighborhood

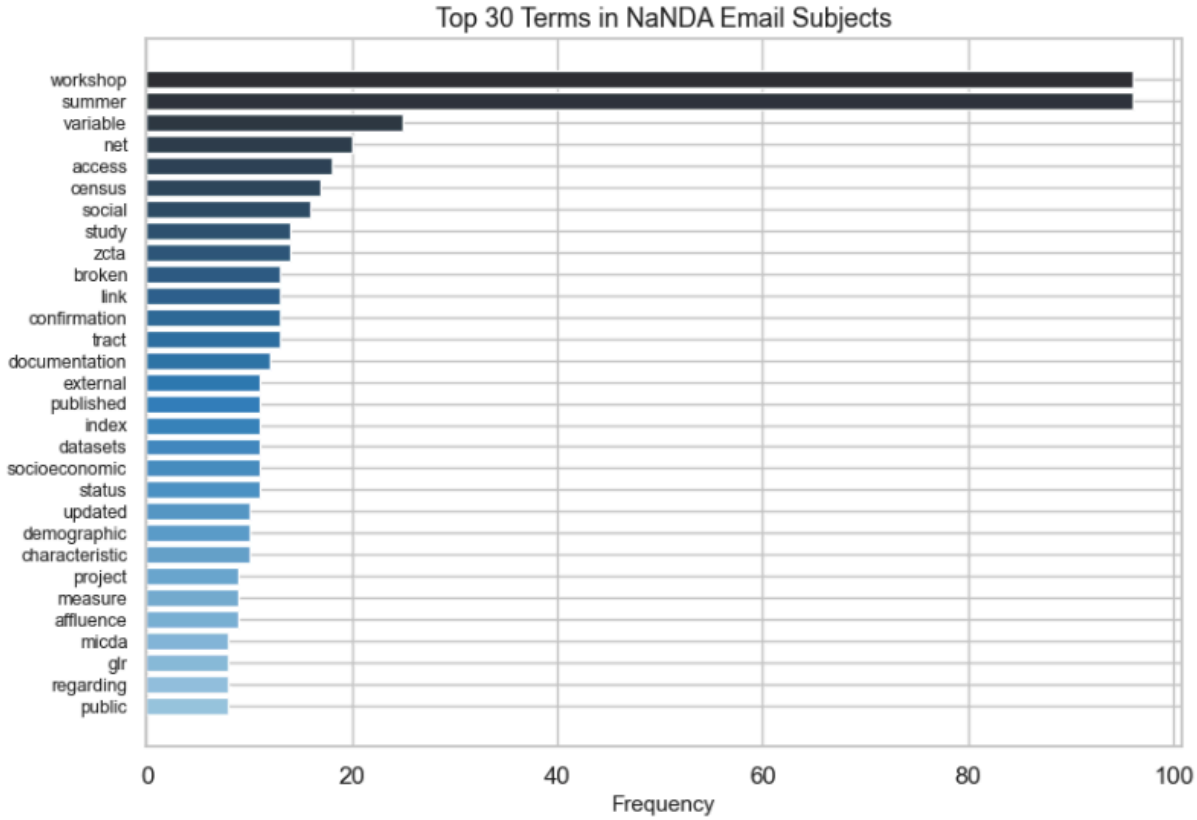
			K38 – Human Rights Law	ADA Compliance	
Gender & Women's Issues	Gender Studies / Economics	gender gender discrimination women domestic violence sexual harassment fertility family planning	J16 – Economics of Gender, Non-labor Discrimination J71 – Discrimination J12 – Marriage, Family Structure, Domestic Abuse I14 – Health and Inequality	Sex Factors Women's Health Domestic Violence Sexual Harassment Gender Equity Maternal Health Family Planning	Gender pay gap by geography, domestic violence shelter access, maternal health outcomes, childcare availability
Family Structure & Households	Demography / Sociology	family background family structure marriage divorce single parent families household composition domestic violence fertility	J12 – Marriage, Marital Dissolution, Family Structure J13 – Fertility, Family Planning, Child Care D13 – Household Production and Intrahousehold Allocation R23 – Regional Migration, Neighborhood Characteristics	Family Structure Marriage Single-Parent Family Household Domestic Violence Fertility Divorce	Single-parent household rates, marriage/divorce rates, household size, multigenerational living by area
Social Services & Safety Net	Social Policy / Public Administration	poverty programs welfare food programs SNAP social services public assistance Medicaid housing programs	I38 – Government Policy, Provision and Effects of Welfare Programs H53 – Government Expenditures and Welfare Programs I32 – Measurement and Analysis of Poverty H55 – Social Security and Public Pensions	Social Welfare Public Assistance Social Services Medicaid Poverty Programs TANF Safety Net Programs	Program take-up rates, benefit cliff effects, social service provider density, TANF caseloads by area
Health Care Services	Social Policy / Health Economics	social services health care services community services human services service utilization integrated care social determinants of health community health workers	I11 – Analysis of Health Care Markets I18 – Government Policy, Public Health I38 – Government Policy, Provision and Effects of Welfare Programs H53 – Government Expenditures and Welfare Programs	Social Welfare Health Services Community Health Services Public Assistance Social Work Patient Navigation	Service provider density maps, co-located service hubs, health-social service referral rates, unmet need estimates by zip code

Voter Registration, Turnout & Partisanship	Political Science	voter attitudes registered voters* elections absentee voters local elections minority voters political partisanship election returns roll call voting records	D72 – Political Processes: Elections, Voting Behavior K16 – Election Law H71 – State and Local Taxation and Revenue	Politics Voting Behavior Elections Civic Participation Political Participation	Precinct-level returns, registration rates by demographic group, mail-in vs. in-person voting

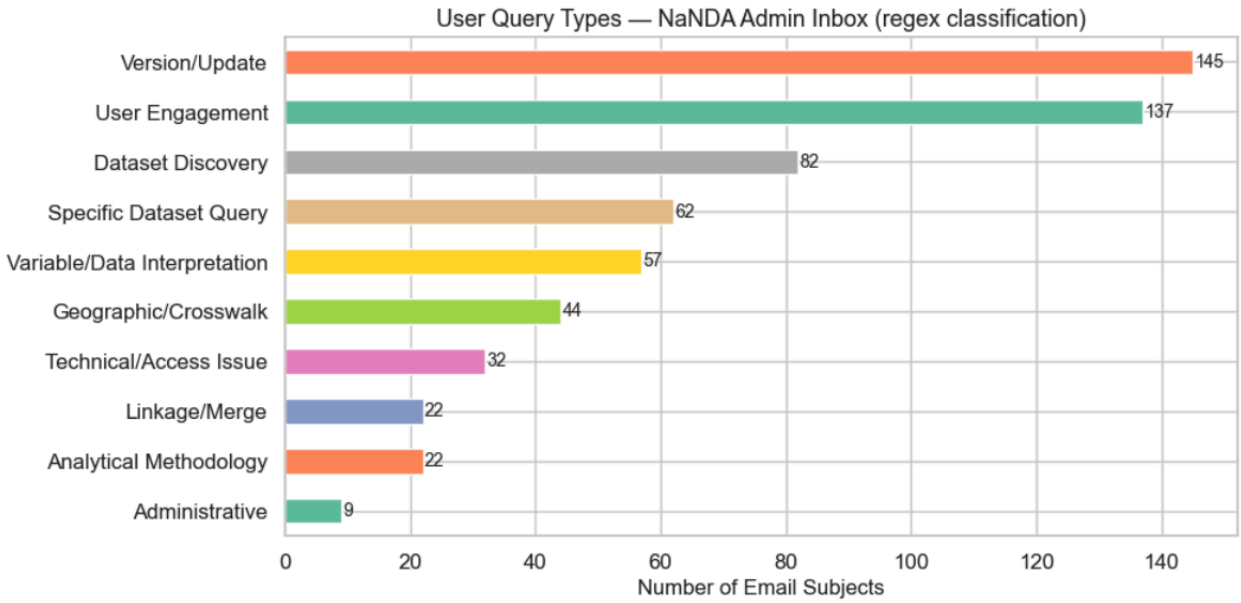
Appendix C — Exploratory Data Analysis: User Metric Data

Geo_unit	Amount
Census Tract + Zcta	14
Census Tract + Zip Code Tabulation Area	13
County	3
Census Tracts	3
Unknown	3
Census Tracts + Zip Code Tabulation Area	2
Census Tract	2
Census Tract + School District + Zip Code Tabulation Area	1
Block Group	1

Geographic unit breakdown from dataset title



Word frequency bar chart for top 30 terms in NaNDA Email Subjects



Regex-based query type classification – how are email subject types distributed?

Appendix D – Example Tree Structures

[MeSH] Residence Characteristics [N01.224.935.700 | D012307]	
Entry Terms	Community; Domicile; Living Arrangements; Neighborhood; Place of Birth; Residential Selection
Hierarchy	N → Health Care N01 → Population Characteristics N01.224 → Demography N01.224.791 → RESIDENCE CHARACTERISTICS ◄
Narrower Terms	Catchment Area, Health [N01.224.791.200] Home Environment [N01.224.791.300] Housing [N01.224.791.400] Independent Living [N01.224.791.550] Neighborhood Characteristics [N01.224.791.775]
NaNDA Scope	*for datasets measuring characteristics of the physical, social, or economic environment of residential neighborhoods; perhaps including NaNDA series on food stores, transit, services, or physical infrastructure more generally. Useful to map to tract-level residential populations.*

Medical Subject Headings (MeSH)

[JEL] Welfare, Well-Being, & Poverty [I3 — Welfare, Well-Being, and Poverty]	
Entry Terms	Anti Poverty, Poverty, Welfare, Well Being;

Hierarchy	I → Health, Education, and Welfare I3 → WELFARE, WELL-BEING, AND POVERTY
Narrower Terms	I30: General I31: General Welfare; Well-Being I32: Measurement and Analysis of Poverty I38: Government Policy; Provision and Effects of Welfare Programs I39: Other
NaNDA Scope	*Covers studies about issues related to general welfare, including basic needs, living standards, quality of life and happiness. for datasets measuring characteristics of the physical, social, or economic environment of residential neighborhoods; perhaps including NaNDA series on Socioeconomic Status and Demographic Characteristics, or physical infrastructure more generally. Useful to map to tract-level residential populations.*

JEL Classification Codes

[ICPSR SH] Neighborhoods [ICPSR Controlled Vocabulary]

Entry Terms	Human settlements, communities, everyday life, neighborhoods, neighbors, residents, influence
Hierarchy	BT (Broader Term): environment BT: Social Environment **home environment ◄**

Narrower Terms	RT: neighborhood characteristics RT: neighborhood conditions RT: urban areas (cities, metropolitan areas, rural areas) NT: urban planning/development NT: Rural Communities (rural development) RT: administrative divisions RT: Socioeconomic Status RT: social influences RT: Housing
NaNDA Scope	*This is arguably a top-level content term for all NaNDA datasets. All NaNDA datasets measure, in the context of the NaNDA repository, characteristics of neighborhoods at the census tract or ZCTA level. This term could possibly appear on every NaNDA ICPSR record as a mandatory subject heading, obviously supplementing with more specific narrower terms as appropriate to the dataset domain.*

ICPSR Subject Headings

References

Allen, R. B. (2020). *Metadata for Social Science Datasets* (SSRN Scholarly Paper No. 5095320). Social Science Research Network. https://doi.org/10.2139/ssrn.5095320

Blair, G., Coppock, A., & Humphreys, M. (2023). *Research Design in the Social Sciences: Declaration, Diagnosis, and Redesign* (SCH-School edition). Princeton University Press. https://doi.org/10.2307/j.ctv34xx5vd

Chenoweth, M., Shallcross, M., & Lyle, J. (2025). *ICPSR Metadata Migration*.

CESSDA Training Team (2017 - 2022). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. https://dmeg.cessda.eu/

"Describing Your Data." *Managing Data | Think and Do at the Library | NC State University Libraries*, NC State University Libraries,www.lib.ncsu.edu/do/data-management/metadata.

Park, J. (2006). *Semantic Interoperability and Metadata Quality: An Analysis of Metadata Item Records of Digital Image Collections~:·*.

Tompkins, V. T., Honick, B. J., Polley, K. L., & Qin, J. (2021). MetaFAIR: A Metadata Application Profile for Managing Research Data. *Proceedings of the Association for Information Science and Technology*, *58*(1), 337–345. https://doi.org/10.1002/pra2.461

Salauddin, N. (2026). *Evaluating the role of metadata standards in enhancing data discoverability and interoperability in academic digital repositories*.