March 16, 2022 - Class lecture notes

Q and A
- MARC21
    - 21 refers to century rather than edition
    - All fields have own archaic rule
    - Anything that start with a 0 = meant for computer reading rather than human
    - Aim of the system when it was made was to make each field as individual as possible, so that there is no cross reference between fields - was made as simple as possible since computers used to clunky and slow
    - Legacy issue when the main point was to make it easy to find
        - Ex card index in libraries
- Assignment on data format
    - Students will likely to do the same thing since only 5 main data formats to analyze
    - When doing the assignment, describe the system and than go into the specifics
        - Ex. why is the X field so low on the list
    - Look at why records were built this way
    - Data formats example
        1. BIBFRAME: what people are trying to get libraries to switch to but little success
        2. XML
        3. JSON
        4. CSV - analyze this if you want to really think deeply about excel sheets
    - Focus on what part is about computational need and what extent is "you" readable
    - Look at the structure, view the actual MARC record (i.e. raw data source) not the end users' version
- Question: thesaurus vs controlled vocabularies
    - Thesaurus is a narrow term of controlled vocabularies (i.e. thesaurus is a type of controlled vocabulary)

Next week
- Start look at things holistically
- Ex. VPL: look at the data format, the CV< content standards, etc.
- Word limit will go up

Bri's presentation
- Talks about linked data
- Linked data = basically made up of 3 parts
    - Subject → predicate → object
    - Ex. Naomi → wrote → Book on Writing

Class activity (more information available in lecture slide)

Class activity discussion
- Machine cataloging
- Students noticed that there were fields with wrong information and some fields should only be written once - a lot of information seems to be computer generated rather than human generated
- Many random incorrect information (ex. UBC not in Vancouver)
- Issue with translation: wikidata has 5 main translation language and if you want to add a language not on the list, it is a process
- Pinyin translation of names: lots of named simplified into very simple characters and some names get translated into the same pinyin translation even if they differ greatly in the original language
- Who posts all of this data?
    - Wikidata is a free resource and non profit
    - Main contributors are librarians
    - Information copied from library websites
- ISSN/ISBN: some books had them and some didn't
- The I icon with a circle around it meant to indicate that something is wrong with what you imputed
- Between computer imputed information and human imputed information, the latter takes priority
- Additional activity: go back a few weeks later and see if the information you imputed today stayed the same or was it changed?