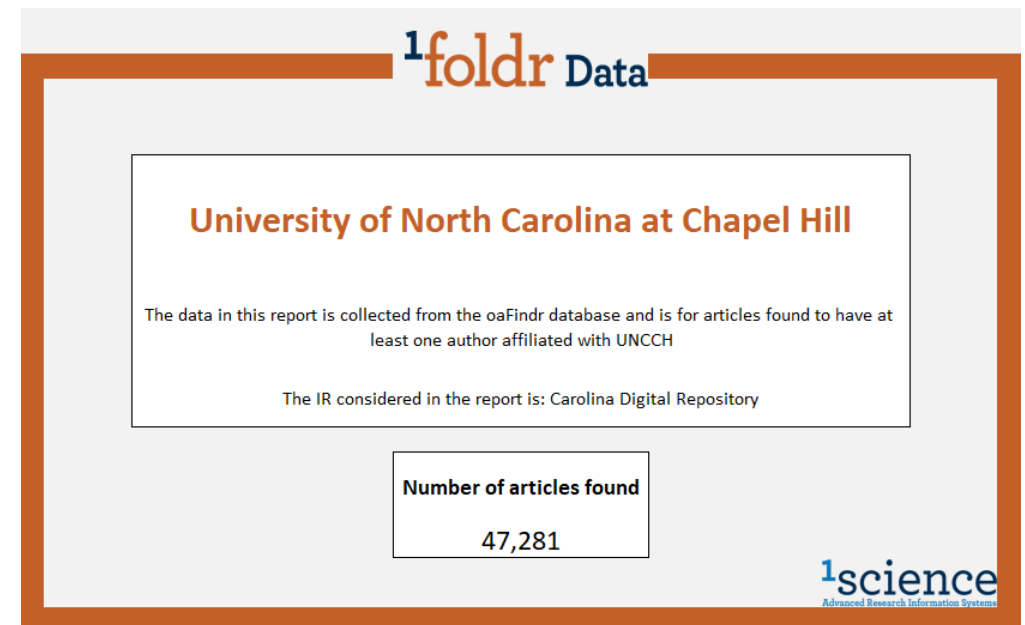# Metadata Mess: Managing Errors in Vendor-Supplied Metadata

Anna Goslen and Rebekah Kati

University of North Carolina at Chapel Hill
Samvera Connect, October 27, 2020

UNC | UNIVERSITY LIBRARIES

The Problem

- Purchased report listing 47K open articles and metadata
- Discovered errors
    - Missing authors
    - Truncated abstracts
    - Download script overwrites
- Missing
    - UNC departmental affiliations
    - Author order
- To do
    - Copyright evaluation
    - Scale

The Investigation

- LDAP for author institutional and departmental affiliations
- Scopus API for article metadata from another source
  - Some articles were missing DOIs
- Lots of test ingests to validate data and that PDFs were matching metadata

The Solution

- Scopus data was best for missing authors, affiliations, and author order
- Mapped departmental affiliations to internal controlled vocabulary
- Corrected download script and merged data

Next Steps

- Load more articles!

- Copyright consultation

- Improvements
  - Scopus
  - Test batches

Screenshot from "Finding Nemo" (2003)

Contact Us

Anna: goslen@email.unc.edu

Rebekah: rkati@email.unc.edu

Carolina Digital Repository: https://cdr.lib.unc.edu
Github: https://github.com/UNC-Libraries/hy-c