

# A Research Design for Measuring Variation in Database Curators' Annotations Through Prospective Randomized Controlled Studies

W. John MacMullen

Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 East Daniel St., MC-493, Champaign IL 61820 USA  
wjohn@uiuc.edu

## Abstract

This project addresses the need for standardized research methodologies for the investigation of variation in workflows and outcomes used and produced by curators of digital repositories when performing standardized tasks, such as indexing, metadata generation, and ontology term assignment.

Research on variation in curators' outcomes (called here 'annotations') is important for several reasons, including: to understand the nature and extent of variation in curators' annotations; to measure internal consistency of curators' work, and to develop related quality metrics; and to learn from best practices to assist in the education and training of new curators.

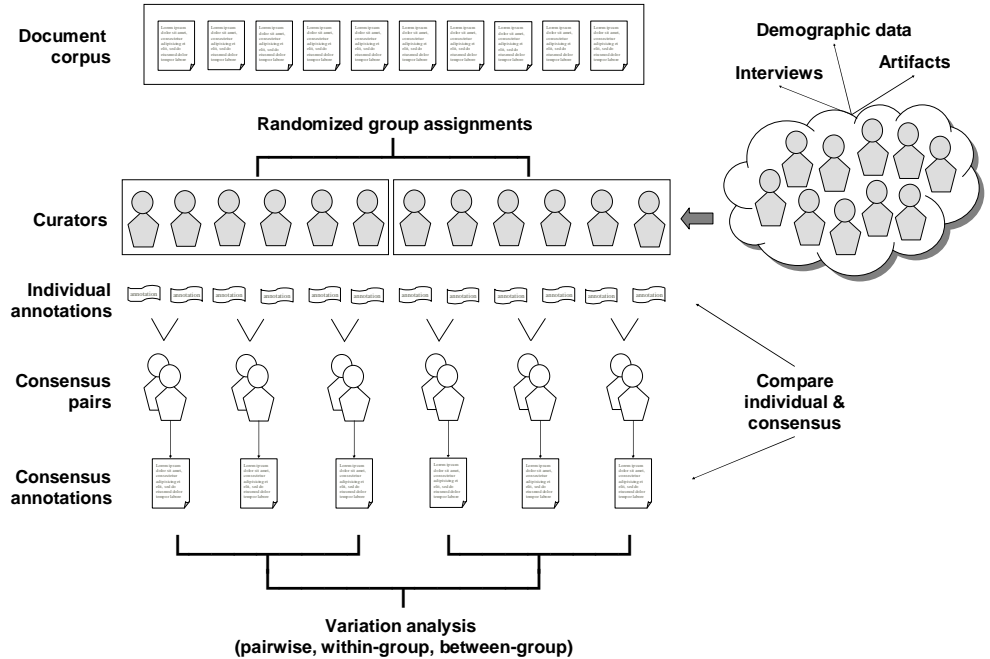
Standardized evaluation methodologies may also allow for the creation of benchmarking metrics, enabling cross-resource comparisons of such quality facets as consistency, reliability, specificity, completeness, and validity [1].

The experimental design was previously used to investigate variation in human curators' Gene Ontology (GO) annotations in model organism databases [2], but is here described such that it is generally applicable to many contexts where multiple curators are performing curation or annotation tasks with documents or other forms of structured data and information.

The research design consists of prospective randomized controlled studies, and includes discussions of document corpus construction, task formulation, document and group assignment, resulting data and analysis, and contextual considerations.

1. MacMullen, W.J.: Facets and measures of Gene Ontology annotation quality in model organism databases. In Proc. of the 69th ASIS&T Annual Meeting, Vol. 43 (2006)
2. MacMullen, W.J.: Contextual Analysis of Variation and Quality in Human-curated Gene Ontology Annotations. Doctoral dissertation, University of North Carolina at Chapel Hill (2007)

## Example: Two-group design with consensus pairing



## Research Design Components

### Participant Selection

- Representative of the population
- Stratified sampling based on attributes
- Sample construction depends on RQs

### Group Assignment

- Random selection
- Different tasks by group for different variables

### Document Assignment

- Random assignment & ordering
- Coverage depends on RQs

### Outcomes

- Data Obtained**
- Controlled vocabulary terms
  - Structured & unstructured data
- Analysis Measures**
- Standard measures of variation
  - Domain-specific measures

### Contextual Considerations

- Experimental Approaches**
- End-user task analysis
- Qualitative Approaches**
- Final annotations may not be sufficient to explain variation

### Document Corpus

#### Selection criteria

- Representative of the objects in the underlying repository
- Content analysis

#### Selection procedure

- New documents (prospective) vs. extant (retrospective)
- Selected by non-participating subject matter experts

### Task Formulation

#### Individual Tasks

- Intellectually and technically similar to normal tasks
- Same interface(s), workflows, outputs

#### Consensus Tasks

- Rationalization of individual outputs
- Annotations used as reference standards

## Acknowledgments

This work was supported in part by an Individual Biomedical Informatics Fellowship (F37 LM009194) from the U.S. National Library of Medicine, National Institutes of Health.