Controlled Vocabulary Enhancement through Crowdsourcing
Dr. Youngok Choi & Mr. Joseph Koivisto
Department of Library and Information Science, Catholic University of America

Project Goal
The purpose of this study is to test controlled vocabularies to describe resources based on domain-specific scholarly practices for the Project Andvari, a digital humanities imitative to aggregate early medieval northern European artifacts.

Micropasts – a crowdsourcing platform sponsored by the British Museum—a Pybossa-based application platform through which researchers, graduate students, and the public could test the thesaurus by applying terms and concepts to digital objects collected.

Method
Holding a workshop of domain experts
Creating a crowdsourcing platform to test controlled vocabularies
Building a skeleton of vocabularies

Facets
Abstract
Figure
Natural World
Object Type
Subject Matter

Results

• About a half of user-generated terms were related to abstract or built environment attributes. Often users provided terms describing materials or other physical characteristics of an artifact.

User Comments
• “Wish I could zoom into this pic; too hard to see the features.”

Summary
• Users tended to tag an artifact focusing on pictorial representation instead of interpreting meanings of an artifact.
• A thesaurus will need to expand vocabularies.
• Additional facets would be needed to describe materials.
• For a similar test with a crowdsourcing platform, a better resolution or a zoom in feature for an object would be necessary to see details.

More information
Project Andvari: http://www.andvari.org/
Project Workshop Blog: https://projectandvari.wordpress.com/
Micropasts site: http://crowdsourced.micropasts.org/project/andvari/

Sample Objects: 251 collected from the British Museum and the Swedish National Heritage Board

Participants: 57

• More than 60% participants did not assign any terms to describe a figure attribute from a list of available terms. There are two possible reasons: one is that an object may not contain any figural iconography, the other might be a lack of enough vocabularies to choose.
• One third of users did not use any terms to describe each facet.