Definitional drift: Human–AI collaboration with MatSci-YAMZ

(An Interactive Demonstration)

Scott McClellan, Metadata Research Center, Drexel University
Colton Gerber, Toberer Group, Colorado School of Mines
Addy Ireland, Pennsylvania State University
Chris Rauch, MRC, Drexel University
Robert Sammarco, Drexel University
John Kunze, ARK Alliance
Mat Kelly, MRC, MRC, Drexel University
Jane Greenberg, MRC, Drexel University

DREXEL UNIVERSITY
Metadata Research Center
College of Computing & Informatics

NSF-HDR-OAC #2118201 Institute for Data Driven Dynamical Design

MatSci-YAMZ, a Very Brief Overview

- Collaborative vocabulary application where users can define terms of relevance
- Stackoverflow-style voting and commenting to determine the most relevant or “best” definitions
- Allows users to provide definitions which capture and accentuate meanings which may be novel or discipline specific
- Communities can vote on terms which best fit their collective understanding
- Integrates Gemma 3.27b LLM to produce alternate definitions
- LLM support allows for alternative definitions based upon broader linguistic pool of sources
- Why is this important?
    - Findable, Accessible, Interoperable, Reusable (FAIR) Data
    - Improved communication in publications by reducing ambiguity in usage
    - Clarification of terminology which is more fluid

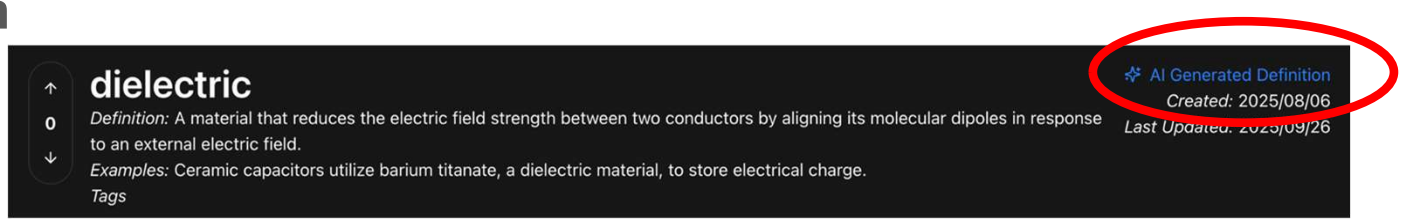
MatSci-YAMZ, an Extremely Brief Overview

- Website where communities of users can define terms
- Stackoverflow-style voting and commenting to find best definitions
- Understand the differences between LLM- and human-generated definitions
- Integrates Gemma 3.27b LLM to produce alternate definitions
- Why is this important?
  - Findable, Accessible, Interoperable, Reusable (FAIR) Data
  - Improved communication by reducing ambiguity in usage
  - Clarification of terminology which is more fluid

Time to participate!

Go to matsci.yamz.net, log in using a Google account

Browse terms, find one you have opinions about, **find the AI Generated Definition**



- If no terms speak to you, take **2-3** minutes to add one, then look at its AI-Generated twin.

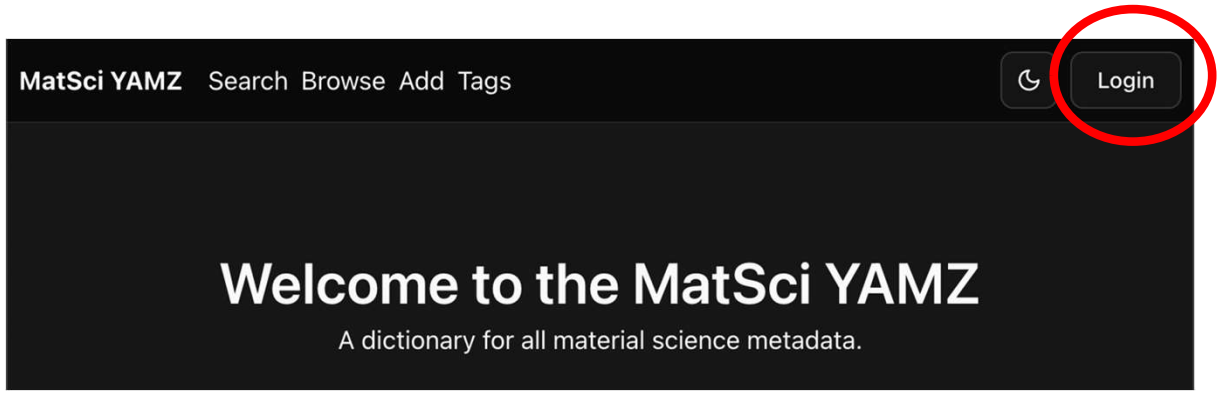
Please **comment on the AI-generated terms.** Comments will cause the LLM to update the definition, so **refresh after commenting!**

Up-vote up or down-vote terms.

https://matsci.yamz.net

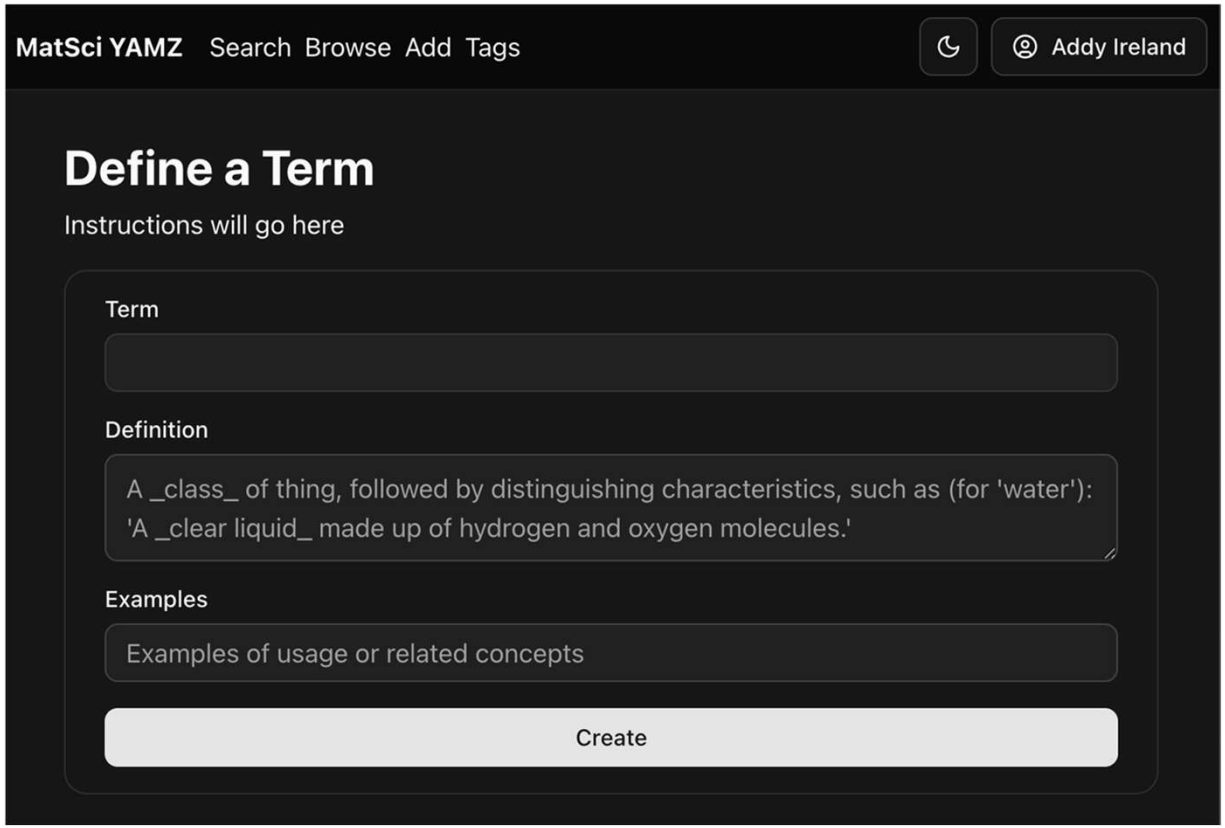
First Steps

- Getting Started
  - Go to https://matsci.yamz.net or use the QR code
  - Start by logging in, updating your profile, and saving your changes
- Tabs Overview
  - **Search**: Search terms and definitions
  - **Browse**: Browse through terms and how many definitions they each have
  - **Add**: Create a definition
  - **Tags**: Browse through tags and definitions associated with them



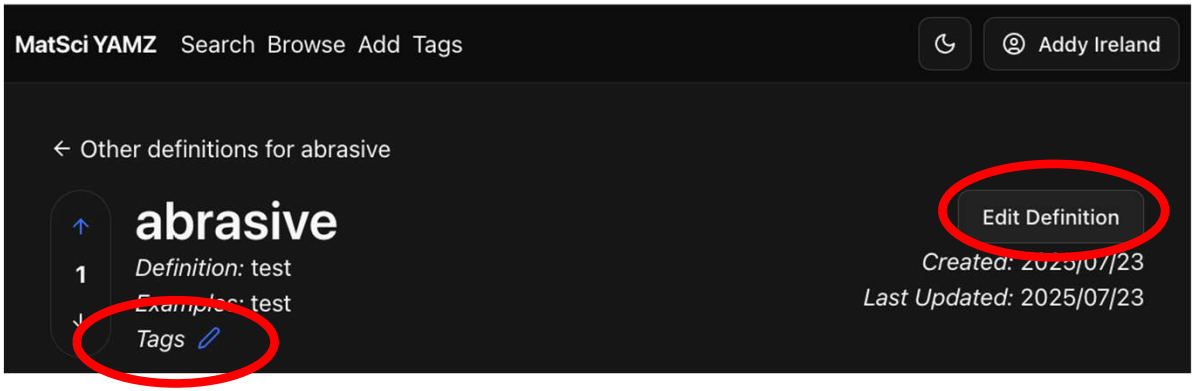
Creating a definition

- Enter your term
  - Either define a new one or select one from the dropdown
- Enter a definition and example
  - Can be as brief or extensive as necessary
  - Example is required
- Finally, create your definition
  - You can also edit your definition after posting it



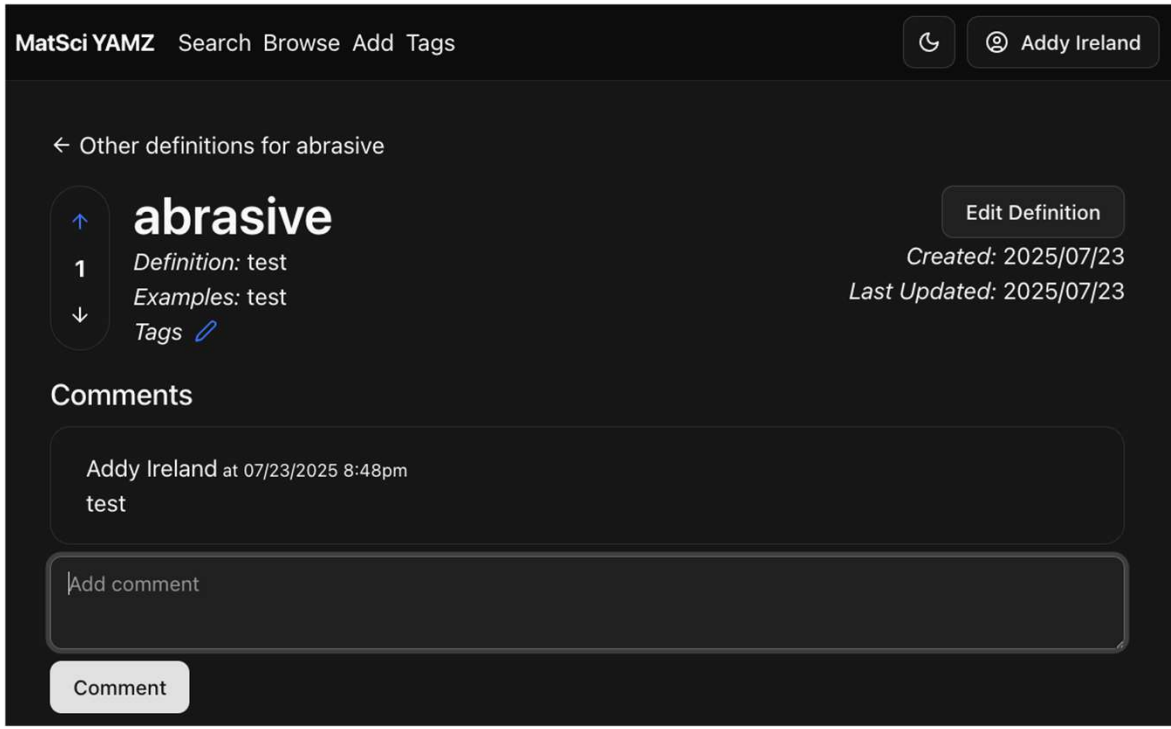
Editing Your Definition

- Once you've created a definition…
- Edit it by clicking the "Edit Definition" button
- Tags, to denote community or category
  - New tags can be created by clicking the tags button and clicking "Add Tag"
  - Add a tag by clicking the blue edit button next to tags
- Browse all your definitions by
  - Clicking your name in top right, then selecting "Definitions" from the dropdown



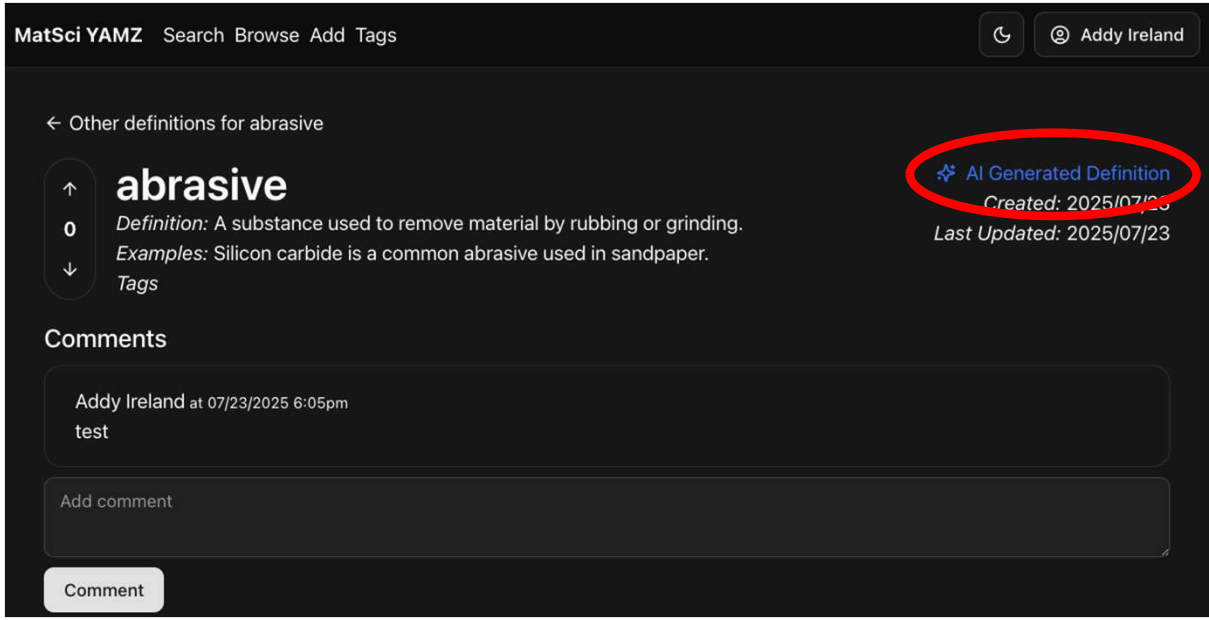
Commenting + Voting

- Find terms by going to the search or browse tab
- Upvote/Downvoting
  - Vote for the best definition using the up/down arrows
- Comments
  - Post a comment by writing in the “Add comment” box and clicking comment



AI Definitions

- Denoted by “AI Generated Definition”
- Automatically generated from user-provided examples by an LLM (Gemma3)
- LLM refines its definition based on users’ comments to it



Feedback

Let us know what works (and what doesn't), what features we might incorporate, or any other thoughts on MatSci-YAMZ.

sm4522@drexel.edu

Thank you!
&
Questions?

Citations

McClellan, S., de Oliveira, I. M., Rauch, C., Adriaenssens, S., & Greenberg, J. (2024). Exploratory analysis of a crowdsourcing metadata tool for building terminological consensus in civil engineering. *Automation in Construction*, *166*, 105627-. https://doi.org/10.1016/j.autcon.2024.105627

Greenberg, J., McClellan, S., Rauch, C., Zhao, X., Kelly, M., An, Y., Kunze, J., Orenstein, R., Porter, C., Meschke, V., & Toberer, E. (2023). Building Community Consensus for Scientific Metadata with YAMZ. *Data Intelligence*, *5*(1), 242–260. https://doi.org/10.1162/dint_a_00211