

ON TO A Better Path to Choose Your Best Ontologies

Asiyah Lin, Axle Research & Technology, Rockville, Maryland, US (0000-0003-2620-0345)
John Graybeal, Graybeal.SKI Consulting, Mountain View, California, U.S. (0000-0001-6875-5360)
Anna Maria Masci, University of Texas MD Anderson Cancer Center, Houston, Texas, U.S. (0000-0003-1940-6740)
Juliane Schneider, Pacific Northwest National Laboratory, Richland, Washington, U.S. (0000-0002-7664-3331)
Ruth Duerr, Ronin Institute for Independent Scholarship, Montclair, New Jersey, U.S. (0000-0003-4808-4736)
Eric G Stephan, Pacific Northwest National Laboratory, Richland, Washington, U.S. (0000-0002-8155-6806)
Hande K       McGinty, Kansas State University, Manhattan, Kansas, U.S. (0000-0002-9025-5538)

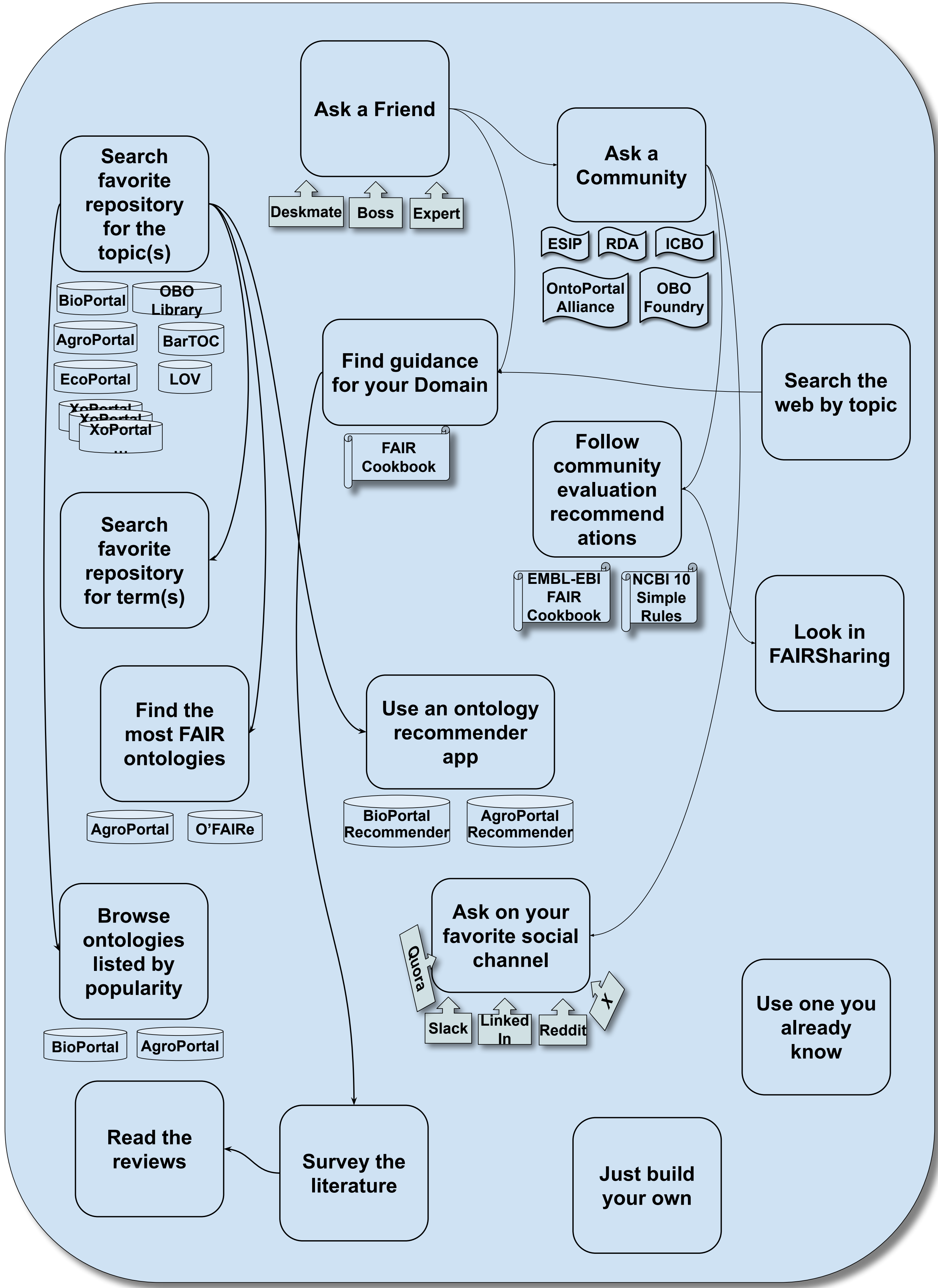


This project is coordinated through the Research Data Alliance Vocabulary and Semantic Services Interest Group. Contact John Graybeal (jgraybeal@sonic.net) to join or learn more, or visit our in-progress draft at the bit.ly link (left) or QR code (above).

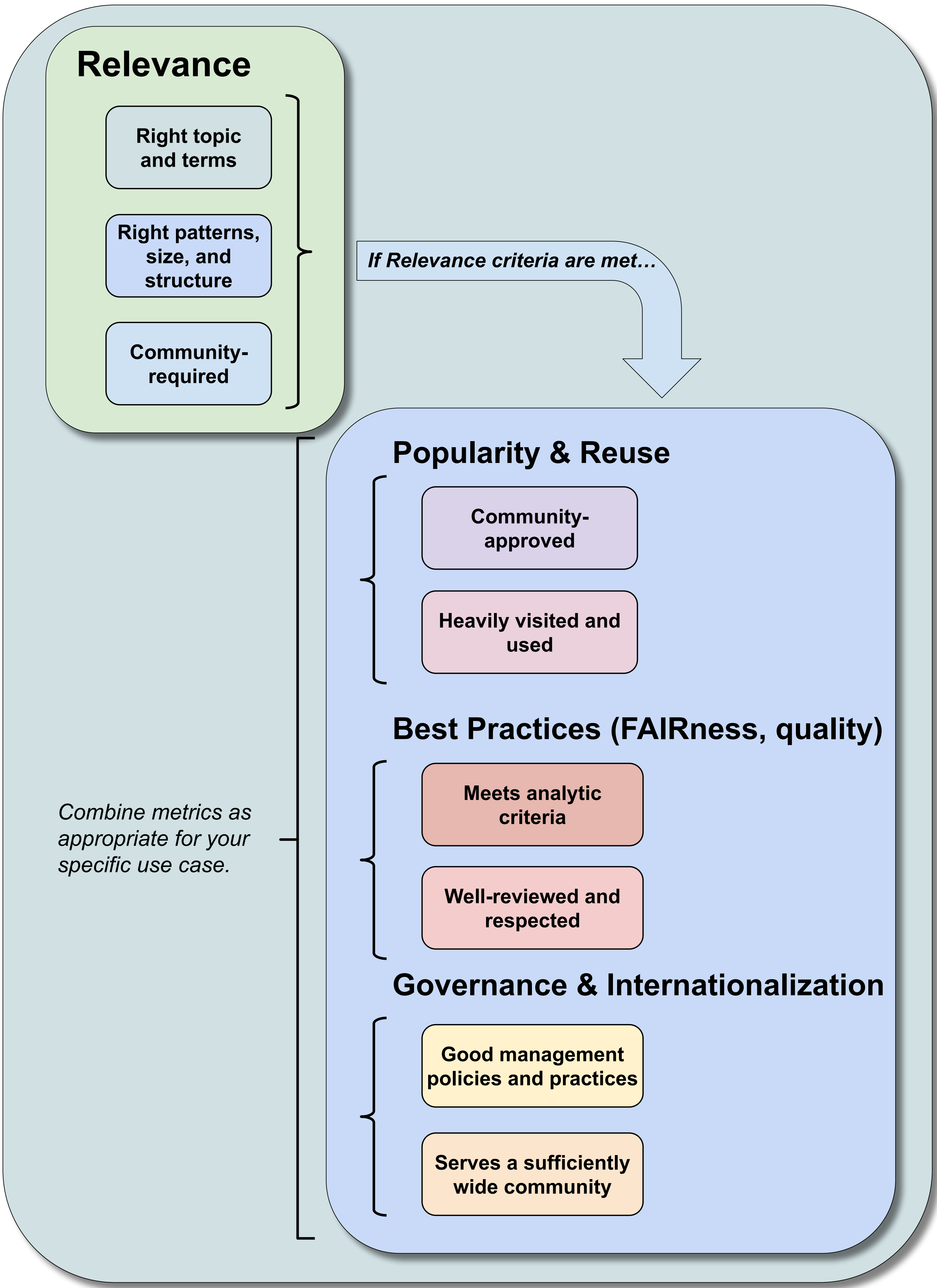
<https://bit.ly/evaluations-choosing-terms-ontologies>

July 17 2024
Version 1.1

Current ‘Strategy’



A Structured Approach



Existing and Future Methods to Evaluate Ontologies

Category	Right topics and terms		Right patterns, size, structure		Community-required		Community-approved		Heavily visited and used		Meets analytic criteria		Well-reviewed and respected		Good mgt policies and practices		Serves a sufficiently wide community	
Metric	Is topic correct, per the title, subject keyword, and description?	Does it cover the concepts of interest? (Exactly all of them?)	Scope and size?	Fundamental patterns (web? restrictions?) and shape (pole, tree, shrub, ivy, mat)	Community edict or constraints declared?	Community expectations or constraints implicit?	Community endorsement or adoption	Community engagement: participation in ontology development and maintenance	Attention: # visits, Google SEO rank, tool popularity, incoming links	# Re-use: citations, web hits for ontology or terms, created views, #includes, term matches, tool lookup matches (e.g. from CEDAR or RedCAP)	Simple metrics: # concepts (classes, properties, individuals), # annotations, # metadata attributes	Complex metrics: % of terms with labels and/or definitions, maximum depth of isA/partOf/ narrower hierarchy, maximum number of children	Approval: # votes, ranking, reviews, papers, citations	Respect: Meets criteria evaluated by communities (e.g., OBO Foundry), contains many metadata attributes (n.b. MOD), positive FAIR evaluation	Defined policies for open, collaborative, community-based governance	Implemented practices that encourage open and collaborative decision-making	Cross-commu nity and international outreach, engagement, and governance	International-ization of content: multiple languages supported for most terms, multi-domain support
Computable Methods *	Ontology search strategies. [?] Title search; keyword search	BioPortal Recommender							SEO services; BioPortal visits and popularity;	GitHub repo followers; GitHub #includes and ont and term matches	Tools often calculate these; ROBOT tool is powerful	OntoPortal has some metrics, ROBOT has many		AgroPortal to check metadata; O'FAIRe to evaluate FAIRness	Machine-discoverable ontology management and governance policies	Issues and changes tracked on open continuous-integration platform with automatic change validation		# languages supported (# two-letter language codes, average # labels + data annotations per language)
Manual Methods			# concepts (vs needed)	# properties, max depth, max children, # leaves	Read community documentation, ask other members, used in tools, read ontology documentation	Concepts specific to, or culturally unacceptable to, community of interest	OBO Foundry validation; OntoPortal groups and projects; website statements	Review contributors or (on GitHub) or author list; ask authors and users			Roll your own code	Spot-check frequency of labels/definitions	Assess citations of primary publication re ontology		Credibility of policies: open, collaborative, timely, consistent, supported by practices	Adopts best practices: versioning, frequent updates, responsive to requests, rigorous change process	Geographic and cultural breadth of leadership; adoption in multiple countries, communities, or organizations	
Future Methods **	More powerful and inclusive search tools	More sophisticated matching	Matches within branch						Tools publishing popularity stats	Tools summarizing reuse stats		All-encompassing analysis tool	Integrative assessment tool		TRUST certification			

* Specific tools cited where known

** New tooling could aggregate all metrics across all ontologies (eventually, across all repositories)