Sorting out industry classifications in Wikidata

Frédéric Julien
Director of Research and Development
CAPACOA

WIKI DATA CON
2021 - a sustainable
future for Wikidata

Why should we even care?

"Statistics Canada identified that **over half (55%)** of the 400,000 businesses registered by the CRA in 2008 did not have a sufficient business description to allow for the assignment of a NAICS code. Statistics Canada further identified that of the 5 million businesses registered with the CRA, **13.2% had an incorrect NAICS code**."

Canada, Business Registration Evaluation Report, 2012.

NAICS stands for North American Industry Classification System (Q3509282)

WIKI DATA CON
2021 - a sustainable future for Wikidata

Quick fix: state industries in Wikidata

CAPACOA (Q85541788)

national arts service organization for the performing arts domain in Canada ✏ edit
Canadian Arts Presenting Association | Canadian Arts Presenters Association | Association des organismes artistiques | Canadian Association for the Performing Arts

industry	performing arts ⓘ	✏ edit
	▾ 0 references	
		+ add reference
		+ add value

But performing arts (Q184485) is not an industry !

WIKI DATA CON
2021 - a sustainable
future for Wikidata

Strike 2: State the industry code itself

CAPACOA (Q85541788)

national arts service organization for the performing arts domain in Canada — edit

Canadian Arts Presenting Association | Canadian Arts Presenters Association | Association des organismes artistiques | Canadian Association for the Performing Arts

NAICS

NAICS code
classification in the North American Industry Classification System

✓ publish ✗ cancel ❓

+ add qualifier

▾ 0 references

+ add reference

But NAICS code (P3224) is an external identifier !!!

WIKI DATA CON

2021 - a sustainable
future for Wikidata

Strike 3: Try ISIC instead of NAICS

CAPACOA (Q85541788)

national arts service organization for the performing arts domain in Canada ✏ edit
Canadian Arts Presenting Association | Canadian Arts Presenters Association | Association des organismes artistiques | Canadian Association for the Performing Arts

industry	Creative, arts and entertainment activities	✏ edit	⊘
	▾ 0 references		
		+ add reference	
		+ add value	

ISIC 9000
Q29584868

But which is the proper way to use industry classifications?

ISIC stands for International Standard Industrial Classification (Q1666934)

WIKI DATA CON
2021 - a sustainable future for Wikidata

What's an industry anyway?

Industry classifications

Standard industry classifications groups organizations/businesses according to their primary economic activity.

Tell me what you produce,

I'll tell you your industry classification.

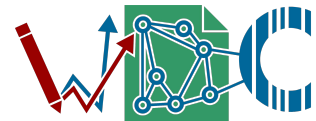# How should industry classifications be implemented in Wikidata?

Current situation

WIKI DATA CON
2021 - a sustainable future for Wikidata

ISIC, NACE and NAICS

	ISIC rev. 4 (P1796)	NACE code (P4496)	NAICS code (P3224)
Data type	String	String	External identifier
Instance of	• Wikidata property for authority control for industries and economic activities • Wikidata property for a classification scheme • Wikidata property for an identifier	• Wikidata property for authority control for industries and economic activities • Wikidata property for a classification scheme	• Wikidata property for authority control for industries and economic activities • Wikidata property for a classification scheme
Type constraint	• industry	• economic activity	No type constraint

WIKI DATA CON
2021 - a sustainable
future for Wikidata

ISIC, NACE and NAICS

	ISIC rev. 4 (P1796)	NACE code (P4496)	NAICS code (P3224)
Intended use in property proposal	"Prospectively, Property:P452 could be restricted to values, that have such property. Probably, this property could be directly assigned to domain of organization (Q43229) (primarily for company (Q783794)) and used as standardized alternative of industry (P452)."	"Planned use: Add the economic activity codes to various institutes that are used as affiliations of people." Further comments suggest the property was meant to be used on individual organization items	"Planned use: associating journal articles with industrial sectors where such data exists" "I would like to create items around those codes and then use main subject (P921) to say that those articles are about those industry sectors. In theory, items about individual businesses could use this property, but I don't think it should work like that—rather, **the code should be applied to the item about the sector, and then the item about the sector used to describe the item about the business**."

ISIC, NACE and NAICS

	ISIC rev. 4 (P1796)	NACE code (P4496)	NAICS code (P3224)
Actual use	Some classes such as automotive industry (Q190117) are implemented, attributed an ISIC code are stated as industry (P452) in organization items.	Same as ISIC	Classes are not implemented, and the property is often used on individual organization items. This is highly problematic since this is an external ID property.

WIKI DATA CON
2021 - a sustainable future for Wikidata

How should industry classifications be implemented in Wikidata?

The question we will probably not answer in 25 minutes...

WIKI DATA CON
2021 - a sustainable
future for Wikidata

The longest road (and fewest redundancies)
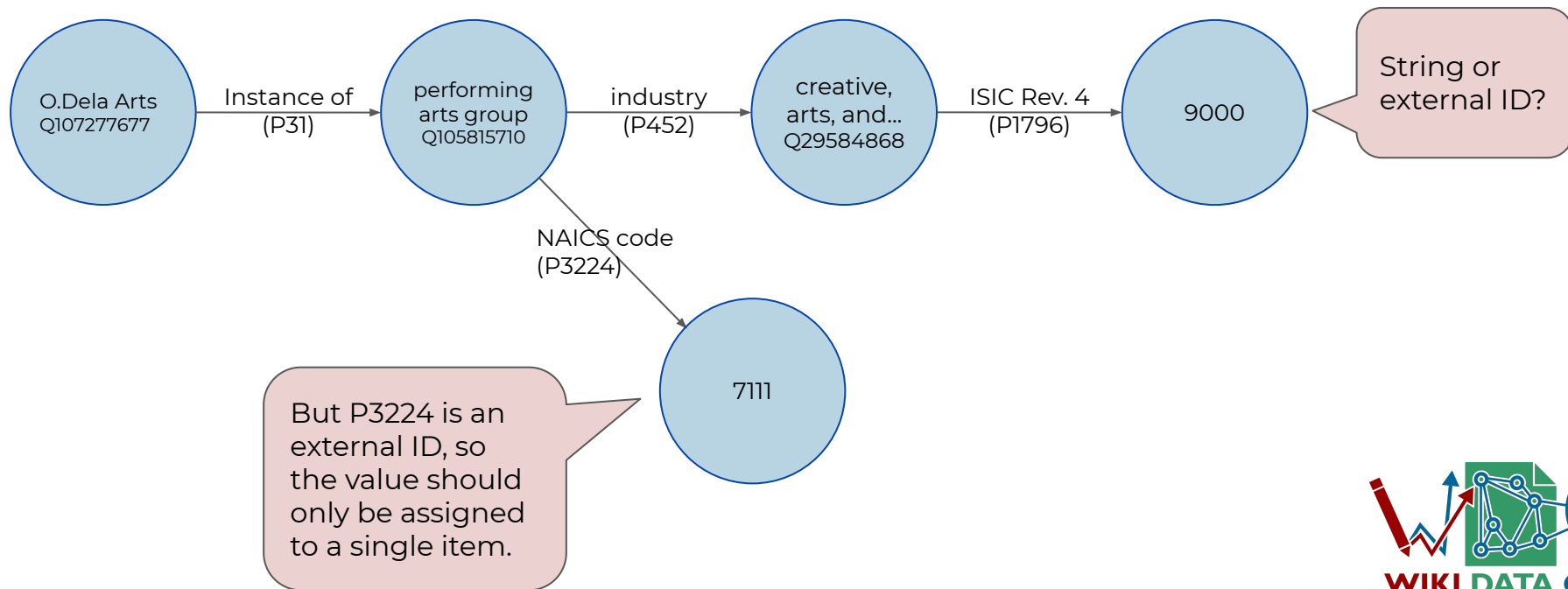
Thanks for your attention!

Is anyone actually still listening?

Frédéric Julien
Fjjulien
frederic.julien@capacoa.ca

WIKI DATA CON
2021 - a sustainable
future for Wikidata