

Nomenclature Usage and Stability Study

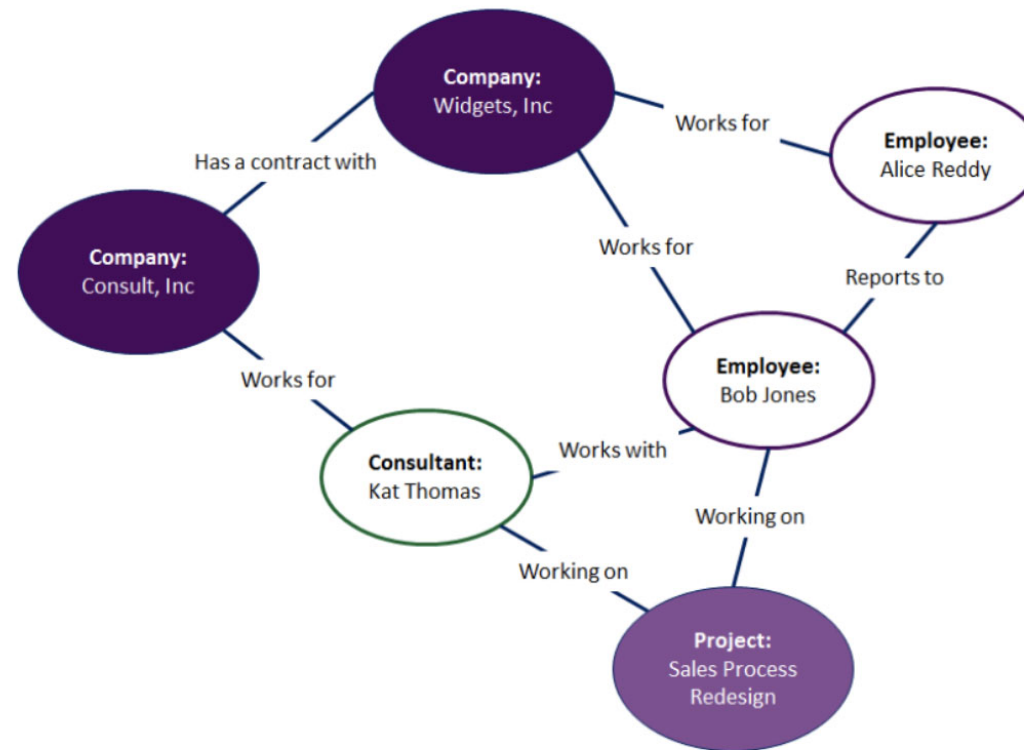
Sandra Young

University of Brighton

Supervisors: Dr Roger
Evans, Dr Gulden Uchyigit

Ontologies

Image from:
<https://towardsdatascience.com/ontology-and-data-science-45e916288cc5> accessed on
13/02/2020



- Yusof, Norlia et al. "Ontology modeling of Malaysian food composition." 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP) (2016): 149-154.

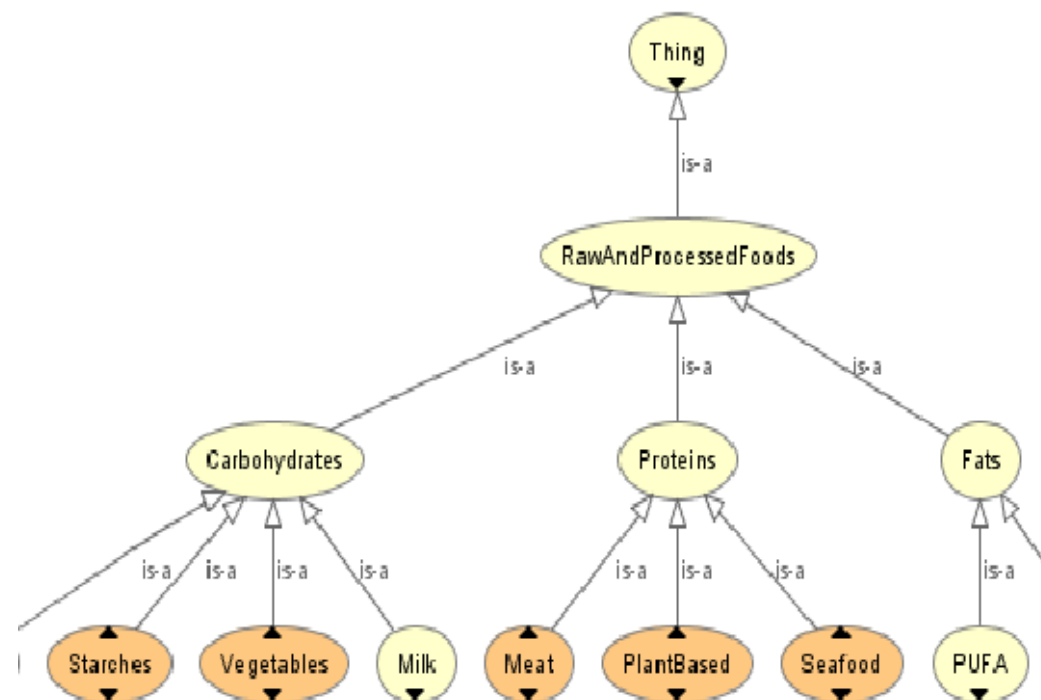


Figure 5. The class hierarchy of raw and processed food.

Problem

Ontologies are really useful

However, question as to accurate integration of data

Ontologies by nature have to exclude at some point (explicit conceptualisation), don't necessarily take into account how presented in the data

Biodiversity and scientific nomenclature and why identified

CONCORDANCE

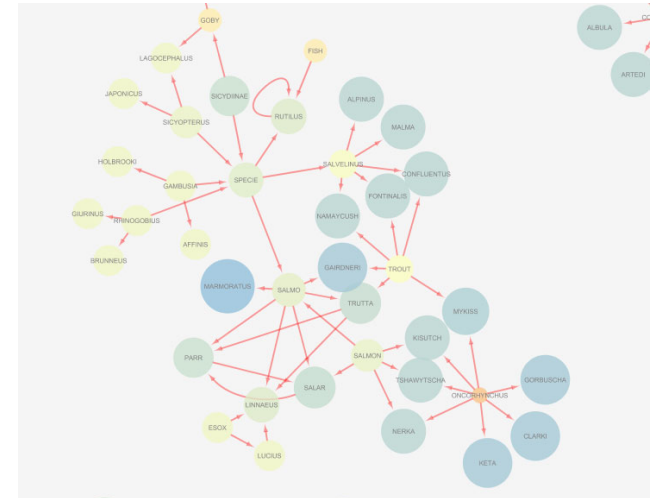
cd mykiss • Oncorhynchus 212 (42.2 per million)

	Details	Left context	KWIC	Right context
1	<input type="checkbox"/>	docP0	nclosure; laboratory; diet Introduction in small cool- Rainbow trout	Oncorhynchus mykiss and longnose dace Rhinichthys catarractae are two species
2	<input type="checkbox"/>	docP0	bas. </s></s> The Southwestern Naturalist 42: 148-154. </s></s> (Oncorhynchus mykiss) Miller, R.R., Williams, J.D. & Williams, J.E. 1989. </s></s> E
3	<input type="checkbox"/>	docP9	ion scale. </s></s> We individually marked and recaptured juvenile	Oncorhynchus mykiss over the course of a year in Murderers Creek, a semi-and r
4	<input type="checkbox"/>	docP9	P97. </s></s> For Atlantic salmon Salmo salar and steelhead trout	Oncorhynchus mykiss , a life history trade-off exists between anadromy and resid
5	<input type="checkbox"/>	docP9	of newly PIT-tagged (New) and previously PIT-tagged (Recaptured)	Oncorhynchus mykiss (- SYM x>75 mm fork length) captured in Murderers Creek
6	<input type="checkbox"/>	docP9	ine September-January Fig. 3. </s></s> Estimated linear density of	Oncorhynchus mykiss at three sampling sites in Murderers Creek during June 20
7	<input type="checkbox"/>	docP9	</s></s> Seasonal pattern in relative growth rates of 100 PIT-tagged	Oncorhynchus mykiss in Murderers Creek from December 2004 through January
8	<input type="checkbox"/>	docP9	Relationship between detection of upstream migrating PIT-tagged	Oncorhynchus mykiss (bars, left vertical axis) and daily stream temperature (line
9	<input type="checkbox"/>	docP9	trut a r e p e t m e r i s Table 2. </s></s> Matrix of PIT-tagged	Oncorhynchus mykiss recaptured in three reaches of Murderers Creek. </s></s> R
10	<input type="checkbox"/>	docP9	within seasonal monitoring of different stream reaches. </s></s>	Oncorhynchus mykiss in Murderers Creek achieved the majority of their annual PI
11	<input type="checkbox"/>	docP9	dance on territory size and population density of steelhead trout (Oncorhynchus mykiss). </s></s> Oecologia 138: 371-378. </s></s> Juanes, F. Li
12	<input type="checkbox"/>	docP9	</s></s> Endanone and resistance in steelhead and rainbow trout f	Oncorhynchus mykiss 1: a review of the non-natives and native </s></s> Parafin

WORD SKETCH

Oncorhynchus as noun 897x

modifiers of "Oncorhynchus"	nouns modified by "Oncorhynchus"	"Oncorhynchus" and/or ...
Salmo ... in Salmoninae : Salmo , Oncorhynchus and Salvelinus	mykiss ... rainbow trout (Oncorhynchus mykiss	Salvelinus ... Salmo , Oncorhynchus and Salvelinus . Reviews in
Oncorhynchus ...	kisutch ... coho salmon (Oncorhynchus kisutch	
	nerka ... sockeye salmon (Oncorhynchus nerka	
	tshawytscha ... Chinook salmon (Oncorhynchus tshawytscha	
	gorbuscha ... pink salmon (Oncorhynchus gorbuscha	
	keta ... chum salmon (Oncorhynchus keta	
	clarki ...	



My approach

Plan for today

Introductions

Nomenclature Profile Studies: an introduction

Hierarchy identification

Knowledge representation: your experiences and my analysis

Nomenclature and vernacular usage: discussion

Corpus data: analyses and discussion

Plan for today

Introductions

Nomenclature Profile Studies: an introduction

Hierarchy identification

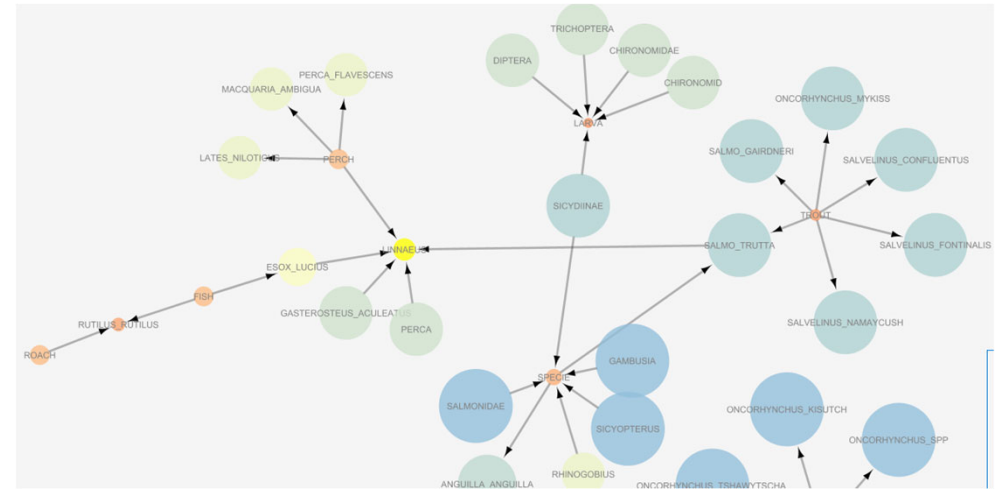
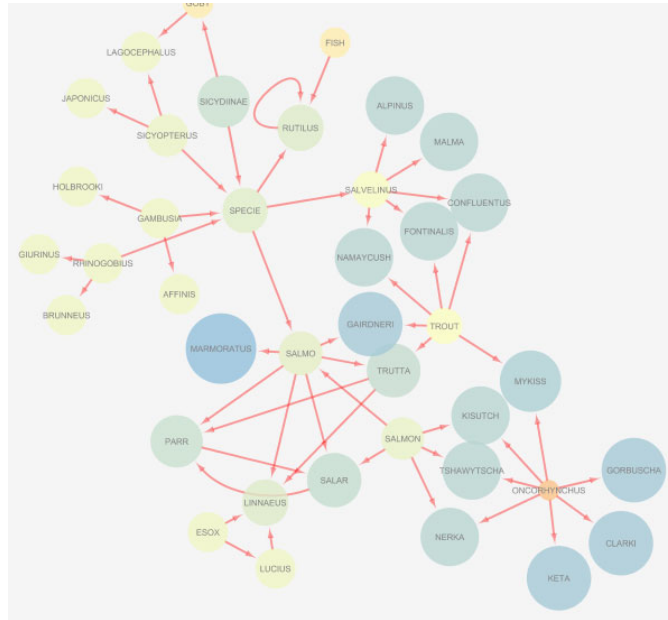
Knowledge representation: your experiences and my analysis

Nomenclature and vernacular usage: discussion

Corpus data: analyses and discussion

- Application of method
- Choose taxon
- Compare knowledge representation resource entries
- Look at representation in two separate corpora

Nomenclature profile
studies



Hierarchy identification

- Left image of original corpus (no extra processing)
- Right image of unified corpus (multi-word names unified)

Plan for today

Introductions

Nomenclature Profile Studies: an introduction

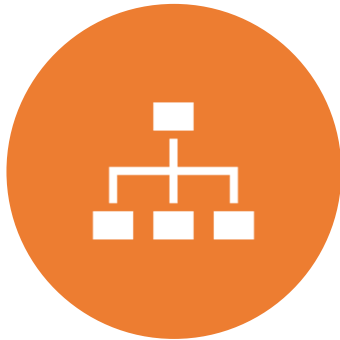
Hierarchy identification

Knowledge representation: your experiences and my analysis

Nomenclature and vernacular usage: discussion

Corpus data: analyses and discussion

Knowledge representation resources



VERTEBRATE TAXONOMY
ONTOLOGY



INTEGRATED TAXONOMIC
INFORMATION SYSTEM



CATALOGUE OF LIFE

Accepted scientific name: *Oncorhynchus mykiss* (Walbaum, 1792) (accepted name)

Synonyms:

- Fario gairdneri* (Richardson, 1836) (synonym)
- Onchorhynchus mykiss* (Walbaum, 1792) (synonym)
- Onchorrhynchus mykiss* (Walbaum, 1792) (synonym)
- Onchorynchus mykiss* (Walbaum, 1792) (synonym)
- Oncorhynchus gairdnerii* (Richardson, 1836) (synonym)
- Oncorhynchus kamloops* Jordan, 1892 (synonym)
- Oncorhynchus mykiss nelsoni* (Evermann, 1908) (synonym)
- Oncorhynchus myskis* (Walbaum, 1792) (synonym)
- Parasalmo mykiss* (Walbaum, 1792) (synonym)
- Parasalmo penshinensis* (Pallas, 1814) (synonym)
- Salmo gairdneri* Richardson, 1836 (synonym)
- Salmo gairdneri gairdneri* (Richardson, 1836) (synonym)
- Salmo gairdneri irideus* Gibbons, 1855 (synonym)
- Salmo gairdneri kamloops* (Jordan, 1892) (synonym)
- Salmo gairdneri shasta* Jordan, 1894 (synonym)
- Salmo gairdnerii* Richardson, 1836 (synonym)
- Salmo gairdnerii gairdnerii* Richardson, 1836 (synonym)
- Salmo gairdnerii irideus* Gibbons, 1855 (synonym)
- Salmo ailberti* Jordan, 1894 (synonym)

Oncorhynchus mykiss (CoL)

***Oncorhynchus mykiss* ([Walbaum, 1792](#))**

Taxonomic Serial No.: 161989

[Download TWB](#)

[Download DwC-A](#)

[\(Download Help\)](#) *Oncorhynchus mykiss* TSN 161989

Taxonomy and Nomenclature

Kingdom:	Animalia
Taxonomic Rank:	Species
Synonym(s):	Salmo mykiss Walbaum, 1792 Onchorhynchus mykiss (Walbaum, 1792) Salmo gibbsii Suckley, 1859 Oncorhynchus mykiss gibbsi (Suckley, 1859)
Common Name(s):	rainbow trout [English] trucha arcoiris [Spanish] steelhead [English] truite arc-en-ciel [French] redband trout [English]

[Taxonomic Status:](#)

Current Standing: valid

[Data Quality Indicators:](#)

Record Credibility Rating: verified - standards met

Oncorhynchus mykiss (ITIS) |

```

[Term]
id: VTO:0058231
name: Oncorhynchus mykiss
namespace: vto-namespace
synonym: "Oncorhynchus kamloops" RELATED [CASSPC:10637]
synonym: "Oncorhynchus mykiss aguabonita" RELATED [NCBITaxon:857568]
synonym: "Oncorhynchus mykiss gairdneri" RELATED [NCBITaxon:857570]
synonym: "Oncorhynchus mykiss irideus" RELATED [NCBITaxon:857569]
synonym: "Parasalmo mykiss" RELATED [NCBITaxon:8022]
synonym: "Rainbow trout" RELATED COMMONNAME [FISHBASE:239]
synonym: "Salmo aquilarum" RELATED [CASSPC:10641]
synonym: "Salmo gairdneri" RELATED []
synonym: "Salmo gairdneri beardsleei" RELATED [CASSPC:33202]
synonym: "Salmo gairdneri gilberti" RELATED [CASSPC:60820]
synonym: "Salmo gairdneri shasta" RELATED [CASSPC:33314]
synonym: "Salmo gairdneri stonei" RELATED [CASSPC:10638]
synonym: "Salmo gairdnerii" RELATED [CASSPC:10628]
synonym: "Salmo gibbsii" RELATED [CASSPC:10632]
synonym: "Salmo iridea" RELATED [CASSPC:10629]
synonym: "Salmo irideus argentatus" RELATED [CASSPC:10646]
synonym: "Salmo kamloops whitehousei" RELATED [CASSPC:10647]
synonym: "Salmo masoni" RELATED [CASSPC:10634]
synonym: "Salmo mykiss" RELATED [CASSPC:10653]
synonym: "Salmo mykiss" RELATED [NCBITaxon:8022]
synonym: "Salmo nelsoni" RELATED [CASSPC:10643]
synonym: "Salmo newberrii" RELATED [CASSPC:10633]
synonym: "Salmo purpuratus" RELATED [CASSPC:10654]
synonym: "Salmo regalis" RELATED [CASSPC:10644]
synonym: "Salmo rivularis" RELATED [CASSPC:10630]
synonym: "Salmo smaragdus" RELATED [CASSPC:10612]
synonym: "Salmo truncatus" RELATED [CASSPC:10635]
synonym: "Salmo whitei" RELATED [CASSPC:33316]
xref: NCBITaxon:8022
xref: TTO:1010653
xref: urn:lsid:globalnames.org:index:84b04c69-51b8-591a-b42a-efcc06ba6de8
is_a: VTO:0058218 ! Oncorhynchus
property value: has rank TAXRANK:0000006

```

Oncorhynchus mykiss (VTO)



Accepted scientific name: *Sander lucioperca* (Linnaeus, 1758) (accepted name)

Synonyms:

- Centropomus sandat* Lacepède, 1802 (synonym)
- Lucioperca linnei* Malm, 1877 (synonym)
- Lucioperca lucioperca* (Linnaeus, 1758) (synonym)
- Lucioperca sandra* Cuvier, 1828 (synonym)
- Perca lucioperca* Linnaeus, 1758 (synonym)
- Stizostedion lucioperca* (Linnaeus, 1758) (synonym)
- Stizostedion lucioperca* (Linnaeus, 1758) (synonym)
- Stizostedium lucioperca* (Linnaeus, 1758) (synonym)

Common names:	Common name	Transliteration	Language	Country
	Luci	-	Albanian	Albania
	Lucioperke	-	Albanian	Albania
	Sharmaku heshtor	-	Albanian	Albania
	Sharmaku i argjendë	-	Albanian	Albania
	Sharmaku i egër	-	Albanian	Albania
	Biala ribba	-	Bulgarian	Bulgaria
	Byala riba	-	Bulgarian	Bulgaria
	Sulka	-	Bulgarian	Bulgaria

Sander lucioperca (CoL)

***Sander lucioperca* ([Linnaeus, 1758](#))**

Taxonomic Serial No.: 650172

[Download TWB](#)

[Download DwC-A](#)

[\(Download Help\)](#) *Sander lucioperca* TSN 650172

Taxonomy and Nomenclature

Kingdom:	Animalia
Taxonomic Rank:	Species
Synonym(s):	Stizostedion lucioperca (Linnaeus, 1758)
Common Name(s):	pikeperch [English] sudak [English] zander [English]

[Taxonomic Status:](#)

Current Standing: valid

[Data Quality Indicators:](#)

Record Credibility Rating: verified - standards met

Taxonomic Hierarchy

Kingdom	Animalia – Animal, animaux, animals
Subkingdom	Bilateria
Infrakingdom	Deuterostomia
Phylum	Chordata – cordés, cordado, chordates
Subphylum	Vertebrata – vertebrado, vertébrés, vertebrates
Infraphylum	Gnathostomata
Superclass	Actinopterygii – ray-finned fishes, spiny rayed fishes, poisson épineux, poissons à nageoires rayonnées

Sander lucioperca (ITIS) |


```
[Term]
id: VTO:0044696
name: Sander lucioperca
namespace: vto-namespace
synonym: "Centropomus sandat" RELATED [CASSPC:51163]
synonym: "Lucioperca linnei" RELATED [CASSPC:36990]
synonym: "Lucioperca lucioperca" RELATED [NCBITaxon:283035]
synonym: "Lucioperca sandra" RELATED [CASSPC:36989]
synonym: "Perca lucioperca" RELATED [CASSPC:19774]
synonym: "Pike-perch" RELATED COMMONNAME [FISHBASE:360]
synonym: "Sander lucioperca (Linnaeus, 1758)" RELATED [NCBITaxon:283035]
synonym: "Stizostedion lucioperca" RELATED [NCBITaxon:283035]
xref: NCBITaxon:283035
xref: TTO:1019774
xref: urn:lsid\:globalnames.org\:index\:19a4e6cb-89f0-5926-a6bb-6e1de9e17d91
is_a: VTO:0044695 ! Sander
property_value: has_rank TAXRANK:0000006
```

Sander lucioperca (VTO) |

Accepted scientific name: *Salmo trutta* Linnaeus, 1758 (accepted name)

Synonyms:

- Fario argenteus* Valenciennes, 1848 (synonym)
- Fario lacustris* (Linnaeus, 1758) (synonym)
- Fario trutta* (Linnaeus, 1758) (synonym)
- Salar ausonii* Valenciennes, 1848 (synonym)
- Salar ausonii parcepunctata* Heckel & Kner, 1858 (synonym)
- Salar ausonii semipunctata* Heckel & Kner, 1858 (synonym)
- Salar bailloni* Valenciennes, 1848 (synonym)
- Salar gaimardi* Valenciennes, 1848 (synonym)
- Salar spectabilis* Valenciennes, 1848 (synonym)
- Salmo albus* Bonnaterre, 1788 (synonym)
- Salmo albus* Walbaum, 1792 (synonym)
- Salmo brachypoma* Günther, 1866 (synonym)
- Salmo caecifer* Parnell, 1838 (synonym)
- Salmo cambricus* Donovan, 1806 (synonym)
- Salmo cornubiensis* Walbaum, 1792 (synonym)
- Salmo cumberland* Lacepède, 1803 (synonym)
- Salmo eriox* Linnaeus, 1758 (synonym)
- Salmo estuarius* Knox, 1855 (synonym)
- Salmo gilberti* Jordan, 1894 (synonym)

Salmo trutta (CoL) |

***Salmo trutta* Linnaeus, 1758**

Taxonomic Serial No.: 161997

[Download TWB](#)

[Download Dwc-A](#)

[\(Download Help\)](#) *Salmo trutta* TSN 161997

Taxonomy and Nomenclature

Kingdom: Animalia
Taxonomic Rank: Species
Synonym(s):
Common Name(s): brown trout [English]
truite brune [French]

Taxonomic Status:

Current Standing: valid

Data Quality Indicators:

Record Credibility Rating: verified - standards met

Taxonomic Hierarchy

Kingdom [Animalia](#) – Animal, animaux, animals
Subkingdom [Bilateria](#)
Infrakingdom [Deuterostomia](#)
Phylum [Chordata](#) – cordés, cordado, chordates
Subphylum [Vertebrata](#) – vertebrado, vertébrés, vertebrates
Infraphylum [Gnathostomata](#)
Superclass [Actinopterygii](#) – ray-finned fishes, spiny rayed fishes, poisson épineux, poissons à nageoires rayonnées
Class [Teleostei](#)

Salmo trutta (ITIS) |

```
[Term]
id: VTO:0058176
name: Salmo trutta
namespace: vto-namespace
synonym: "Fario argenteus" RELATED [CASSPC:27830]
synonym: "Salar ausonii" RELATED [CASSPC:10701]
synonym: "Salar bailloni" RELATED [CASSPC:27831]
synonym: "Salar gaimardi" RELATED [CASSPC:27837]
synonym: "Salar macrostigma" RELATED [CASSPC:10688]
synonym: "Salar spectabilis" RELATED [CASSPC:51499]
synonym: "Salmo albus" RELATED [CASSPC:64487]
synonym: "Salmo brachypoma" RELATED [CASSPC:27810]
synonym: "Salmo caecifer" RELATED [CASSPC:64491]
synonym: "Salmo cambricus" RELATED [CASSPC:27832]
synonym: "Salmo caspius" RELATED [CASSPC:10679]
synonym: "Salmo cornubiensis" RELATED [CASSPC:27833]
synonym: "Salmo cumberland" RELATED [CASSPC:27816]
synonym: "Salmo eriox" RELATED [CASSPC:27812]
synonym: "Salmo estuarius" RELATED [CASSPC:27835]
synonym: "Salmo fario" RELATED [CASSPC:10699]
synonym: "Salmo fario forestensis" RELATED [CASSPC:47390]
synonym: "Salmo fario loensis" RELATED [CASSPC:61299]
synonym: "Salmo gadoides" RELATED [CASSPC:47492]
synonym: "Salmo gallivensis" RELATED [CASSPC:27836]
synonym: "Salmo illanca" RELATED [CASSPC:27842]
synonym: "Salmo islayensis" RELATED [CASSPC:27844]
synonym: "Salmo lacustris" RELATED [CASSPC:10700]
synonym: "Salmo lacustris rhenana" RELATED [CASSPC:64496]
synonym: "Salmo lacustris romanovi" RELATED [CASSPC:10675]
synonym: "Salmo lacustris septentrionalis" RELATED [CASSPC:64497]
synonym: "Salmo lemanus" RELATED [CASSPC:27857]
synonym: "Salmo levenensis" RELATED [CASSPC:27822]
synonym: "Salmo microps" RELATED [CASSPC:27818]
synonym: "Salmo mistops" RELATED [CASSPC:27823]
synonym: "Salmo orcadensis" RELATED [CASSPC:27863]
synonym: "Salmo oxianus" RELATED [CASSPC:10692]
synonym: "Salmo phinoc" RELATED [CASSPC:50415]
```

Salmo trutta (VTO)

Plan for today

Introductions

Nomenclature Profile Studies: an introduction

Hierarchy identification

Knowledge representation: your experiences and my analysis

Nomenclature and vernacular usage: discussion

Corpus data: analyses and discussion

Plan for today

Introductions

Nomenclature Profile Studies: an introduction

Hierarchy identification

Knowledge representation: your experiences and my analysis

Nomenclature and vernacular usage: discussion

Corpus data: analyses and discussion

Variant coverage – summary page

Oncorhynchus mykiss

Total SCI	64
Total COM	24

		JEFF	WEB
VTO SCI variants	26	31%	50%
VTO COM variants	1	100%	100%

		JEFF	WEB
CoL SCI variants	36	0%	0%
CoL COM variants	23	57%	91%

		JEFF	WEB
ITIS SCI variants	4	0%	0%
ITIS COM variants	3	100%	100%

Sander lucioperca

Total SCI	16
Total COM	3

		JEFF	WEB
VTO SCI variants	8	38%	50%
VTO COM variants	1	100%	100%

		JEFF	WEB
CoL SCI variants	9	0%	0%
CoL COM variants	3	100%	100%

		JEFF	WEB
ITIS SCI variants	2	0%	0%
ITIS COM variants	1	100%	100%

Salmo trutta

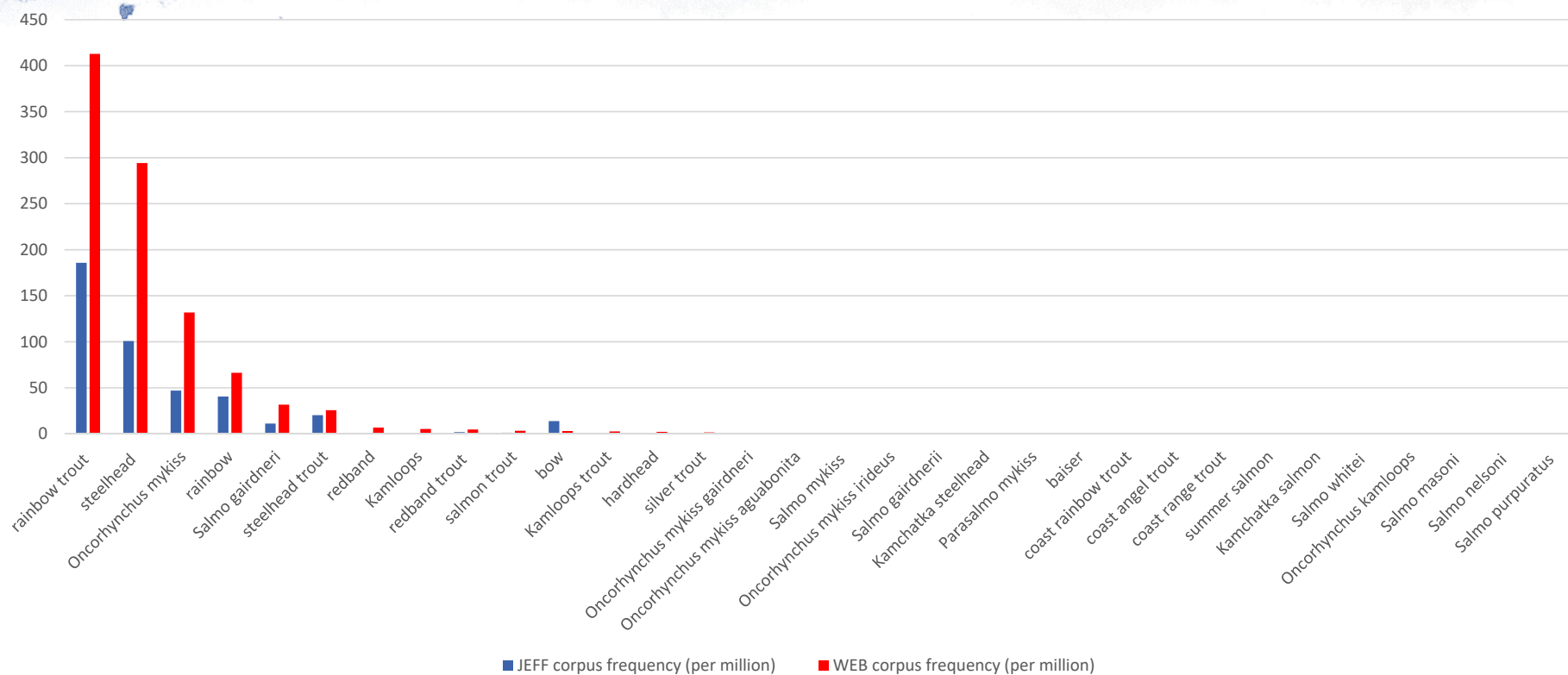
Total SCI	116
Total COM	27

		JEFF	WEB
VTO SCI variants	50	2%	8%
VTO COM variants	1	100%	100%

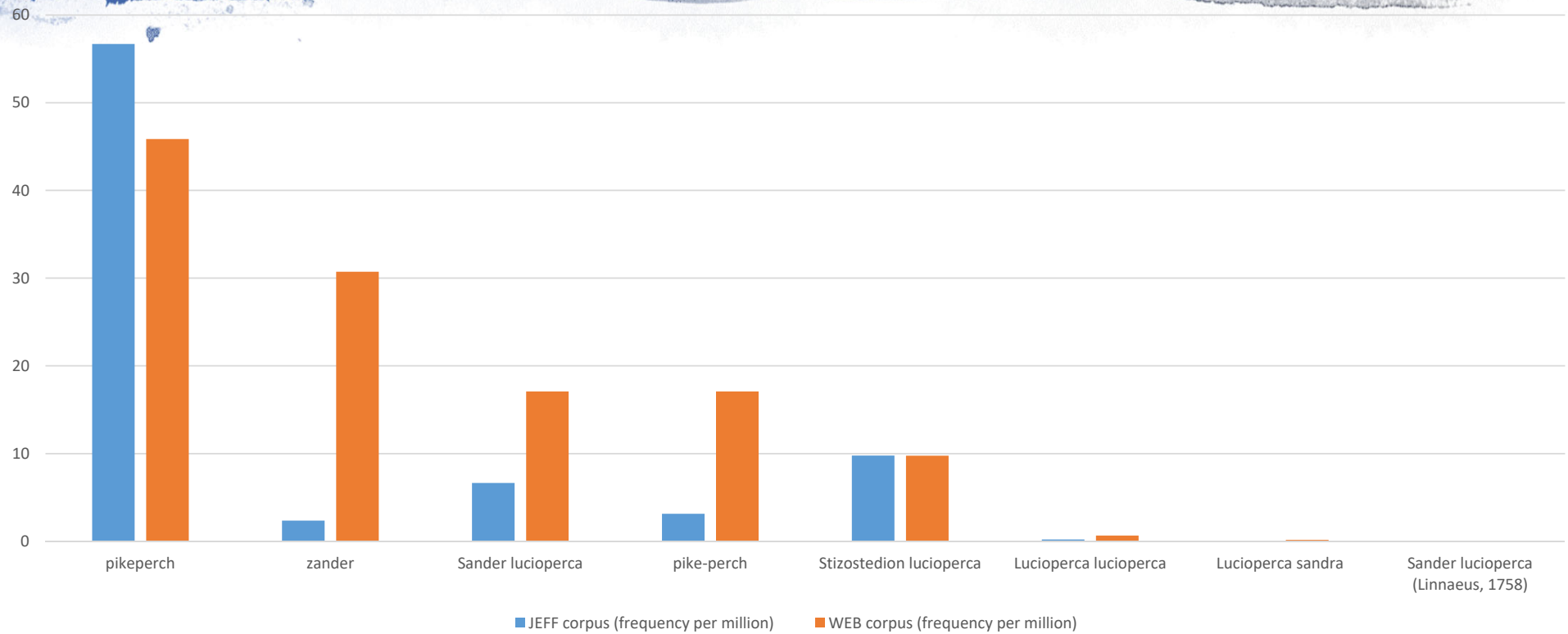
		JEFF	WEB
CoL SCI variants	65	0%	0%
CoL COM variants	27	41%	67%

		JEFF	WEB
ITIS SCI variants	1	0%	0%
ITIS COM variants	1	100%	100%

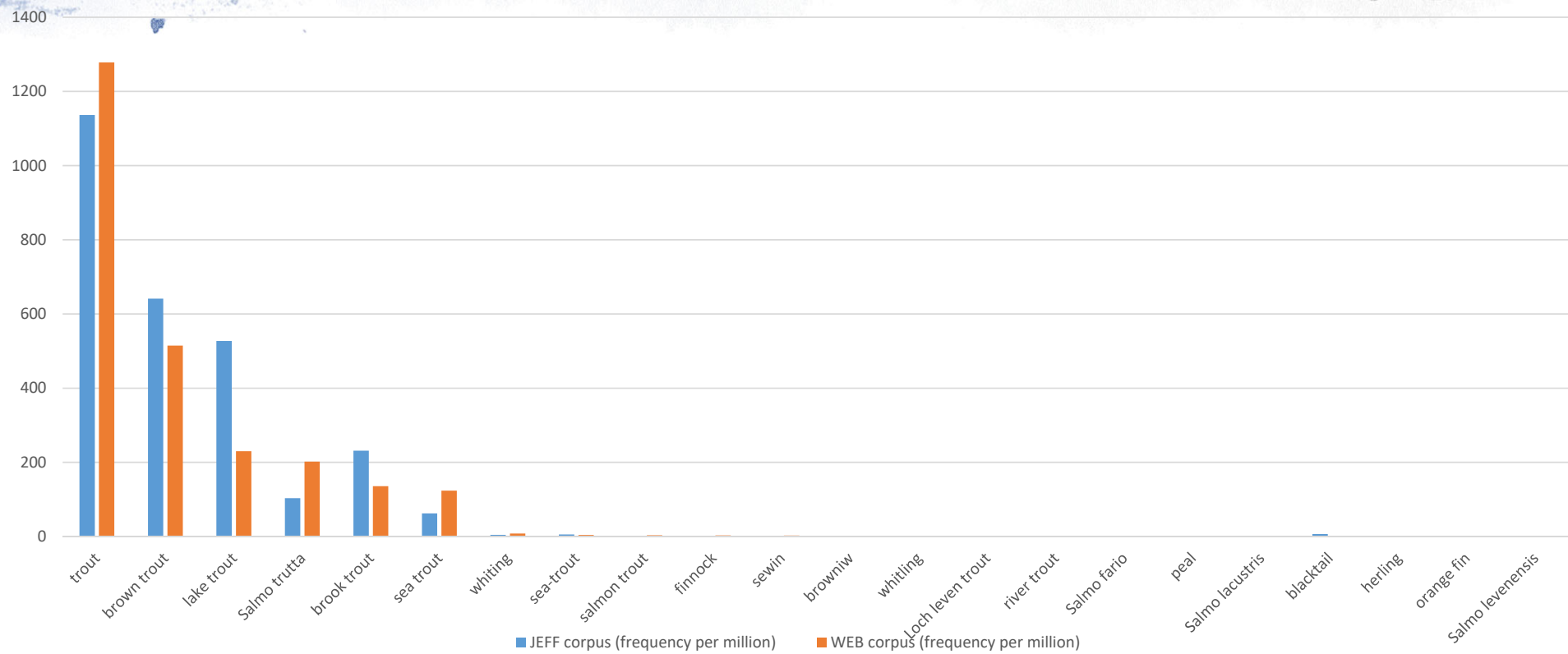
Variant variation – *Oncorhynchus mykiss*



Variant variation – Sander lucioperca



Variant variation – *Salmo trutta*



Ambiguities identified

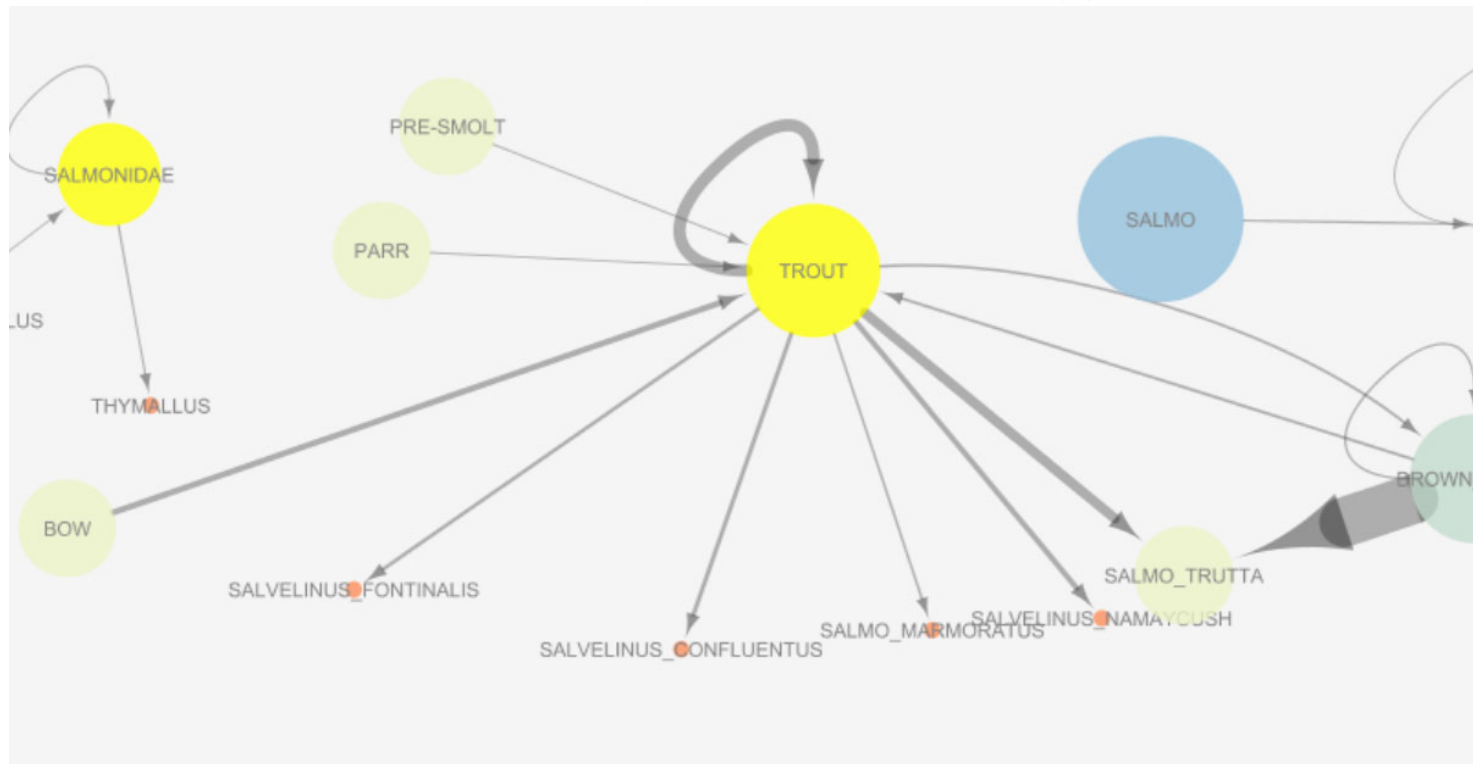
Vernacular

Spelling

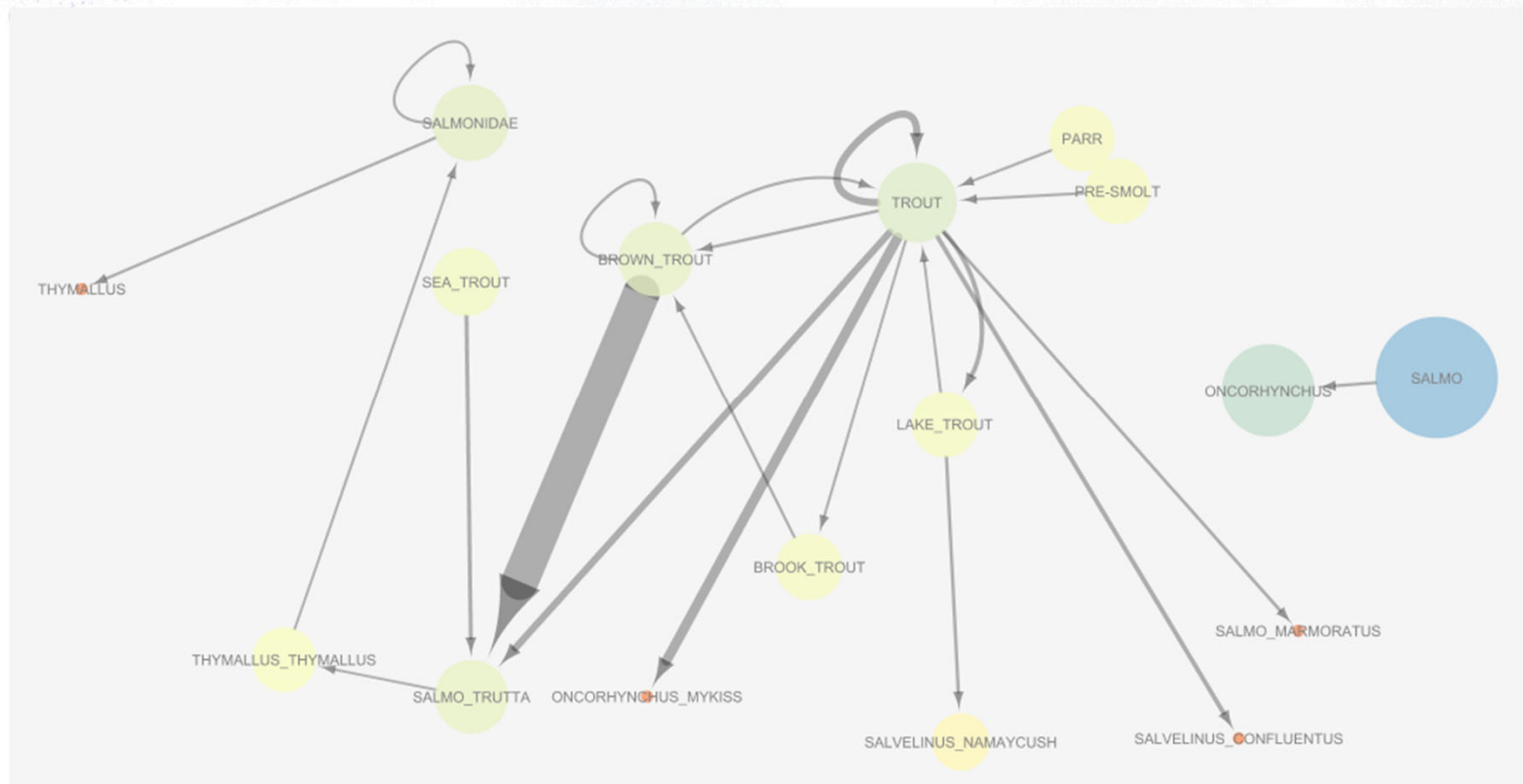
Accepted term
versus most used

Contradictions
between test data
and knowledge
resources

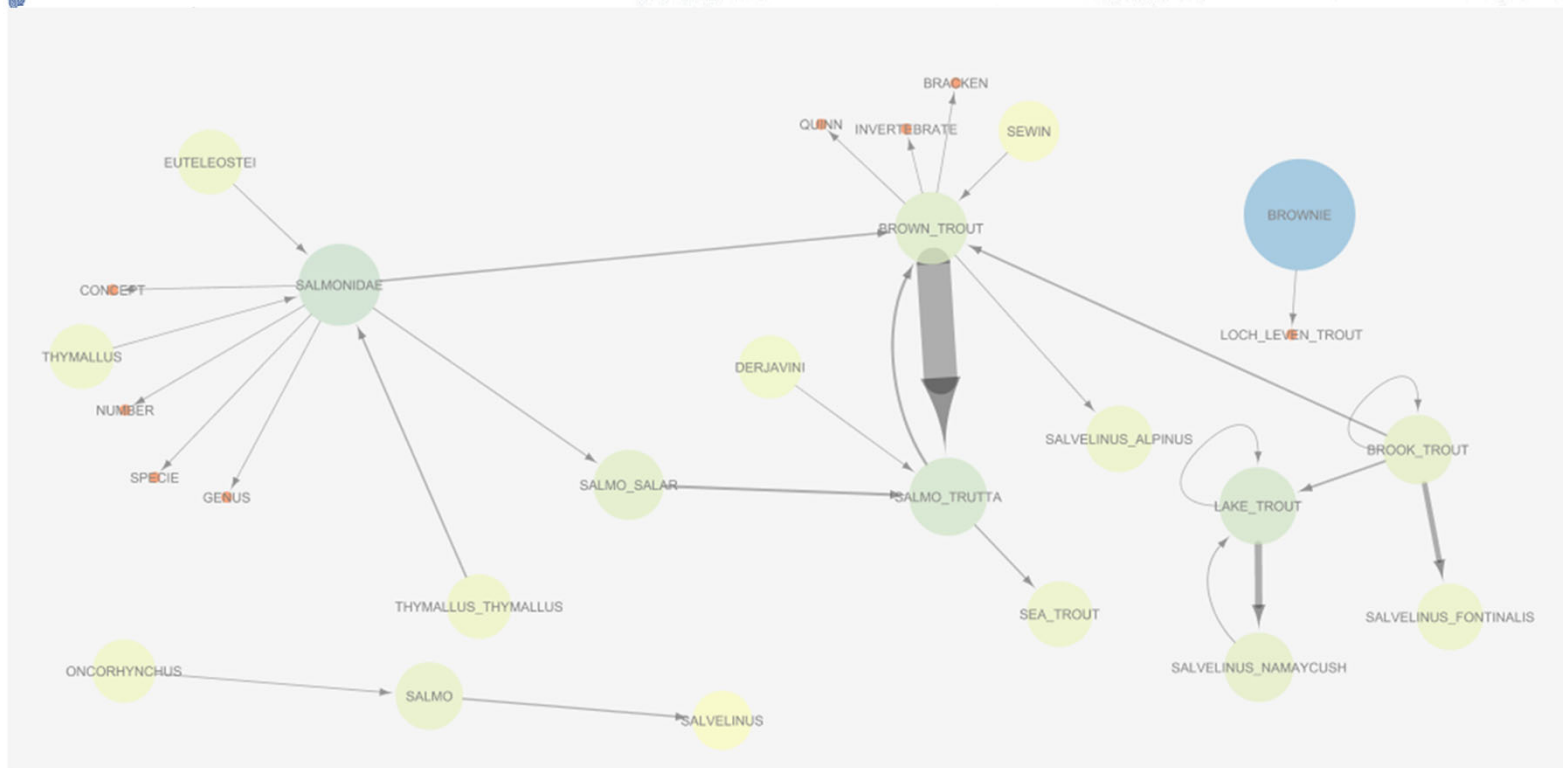
Vernacular ambiguities: broader meaning



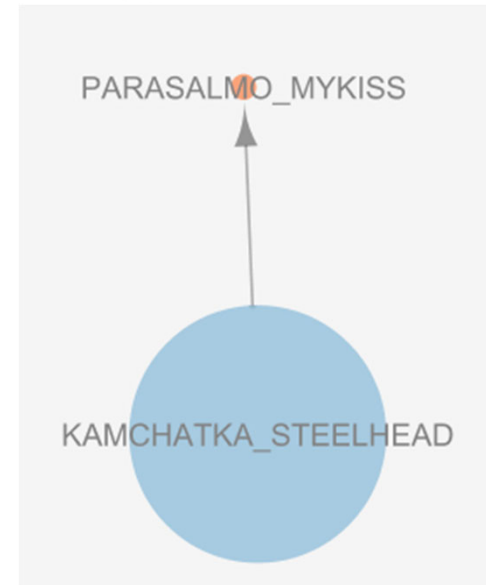
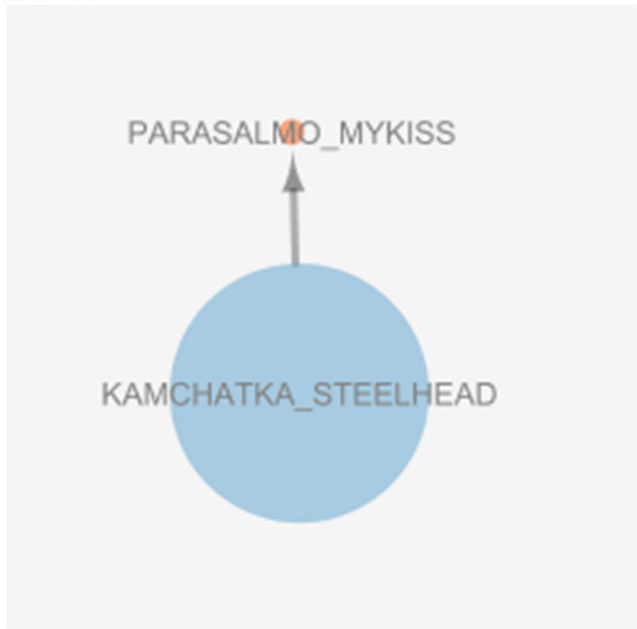
Vernacular ambiguities: broader or narrower meaning



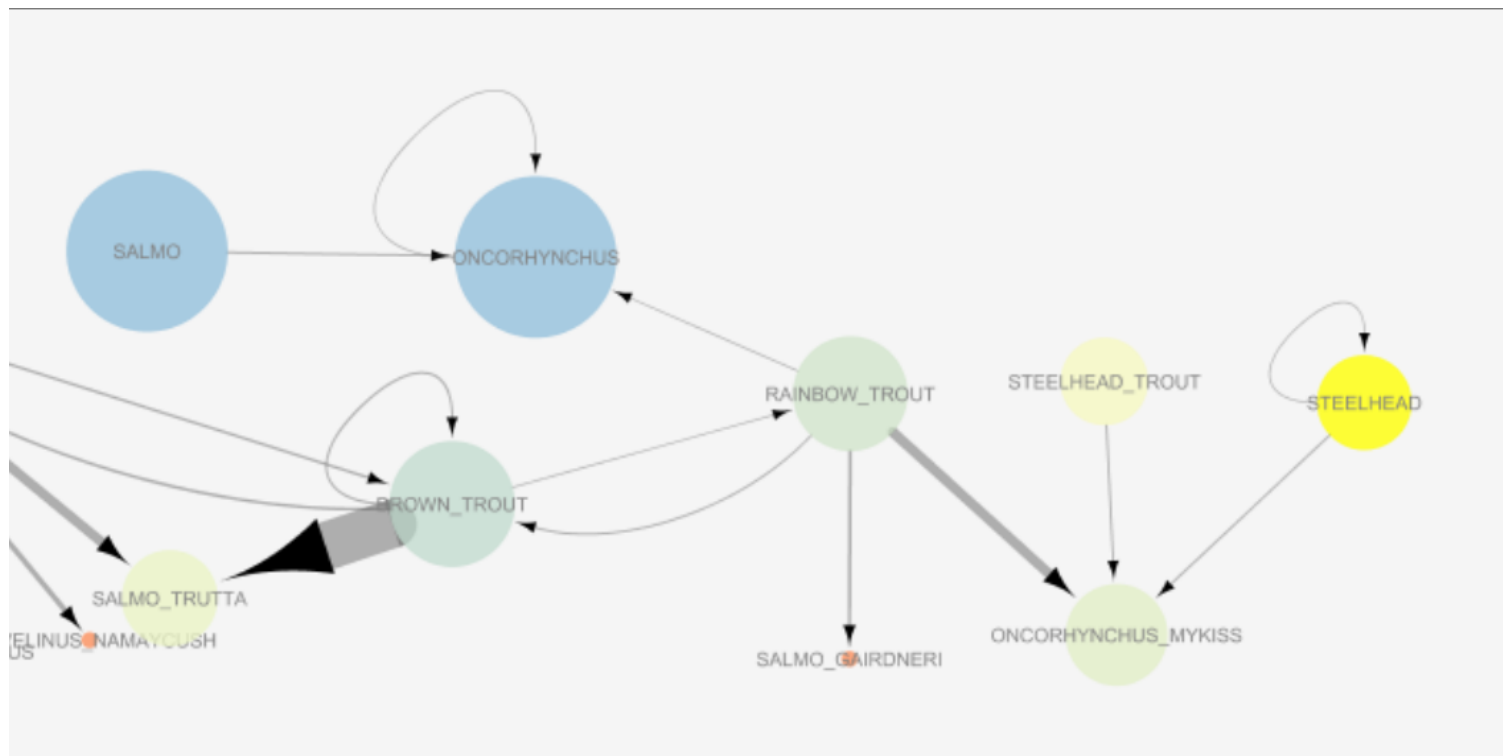
Vernacular ambiguities: broader or narrower meaning



Ambiguities: increased specificity of usage



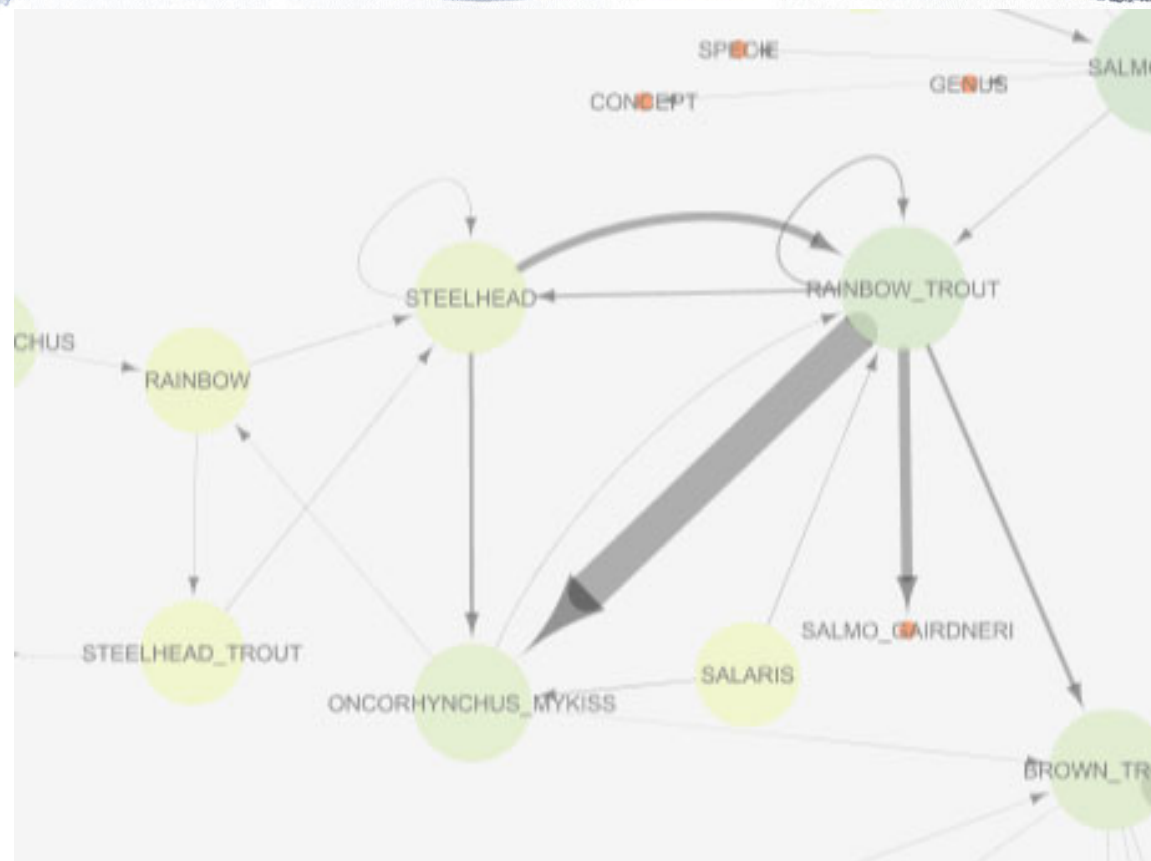
Ambiguities: increased specificity of usage (JEFF)



Ambiguities: increased specificity of usage (JEFF)

Cooccurrence	Co-occur docs	No. of poss. Docs	% co-occurrence docs
All four in same document	4	47	9%
Steelhead trout and rainbow trout	31	47	66%
Steelhead and rainbow	8	48	17%
Steelhead trout and rainbow	7	48	15%
Rainbow trout and steelhead	35	71	49%
Rainbow trout and rainbow	23	48	48%
Steelhead and steelhead trout	24	47	51%

Ambiguities: increased specificity of usage (WEB)

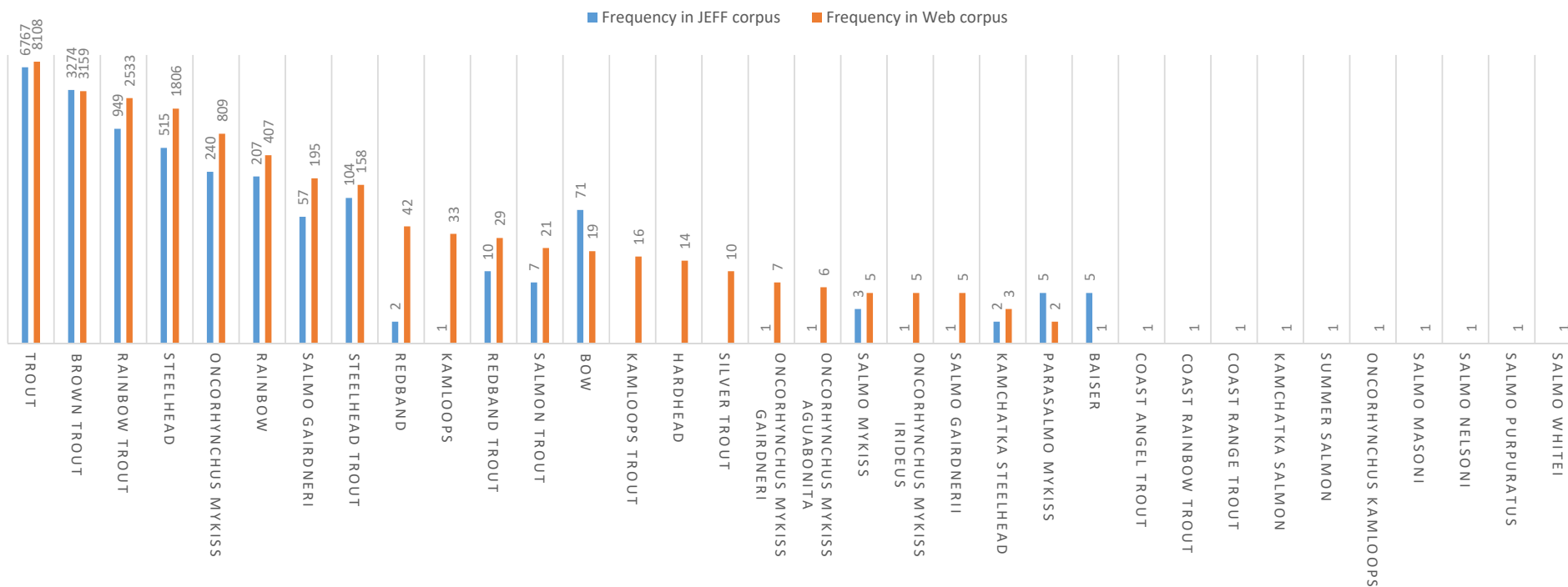


Ambiguities: increased specificity of usage (WEB)

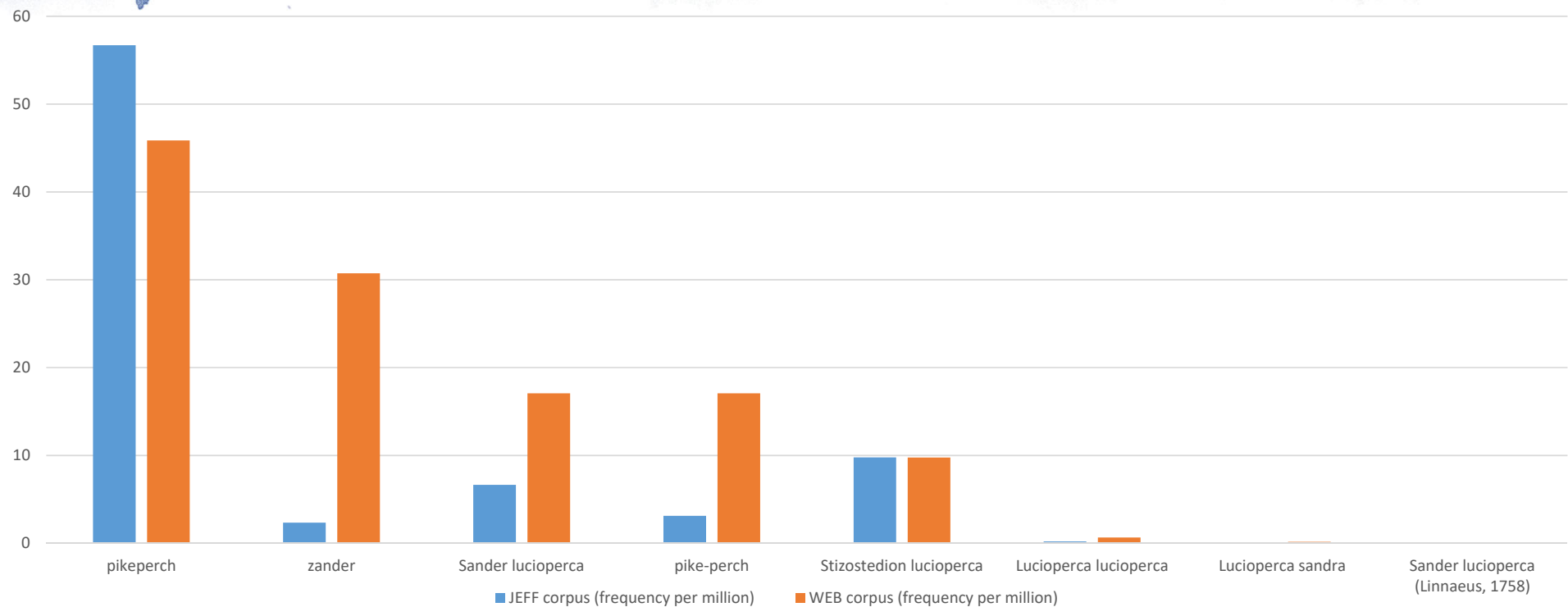
Co-occurrence	Co-occur docs	No. of poss. Docs	% co-occurrence docs
All four in same document	15	72	21%
Steelhead trout and rainbow trout	57	72	79%
Steelhead and rainbow	41	120	34%
Steelhead trout and rainbow	27	120	23%
Rainbow trout and steelhead	101	120	84%
Rainbow trout and rainbow	96	139	69%
Steelhead and steelhead trout	48	72	67%

Ambiguities: nomenclature spelling

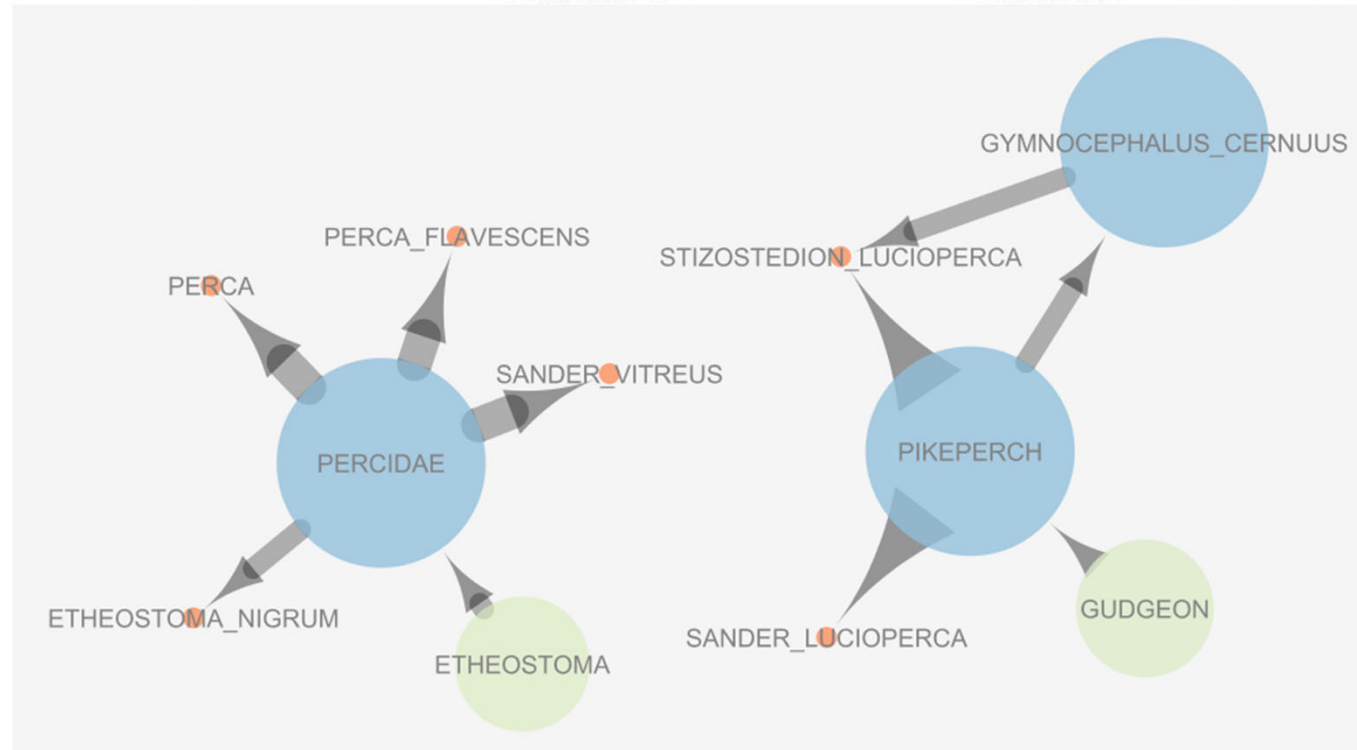
FREQUENCY COMPARISON BETWEEN JEFF AND WEB CORPUS



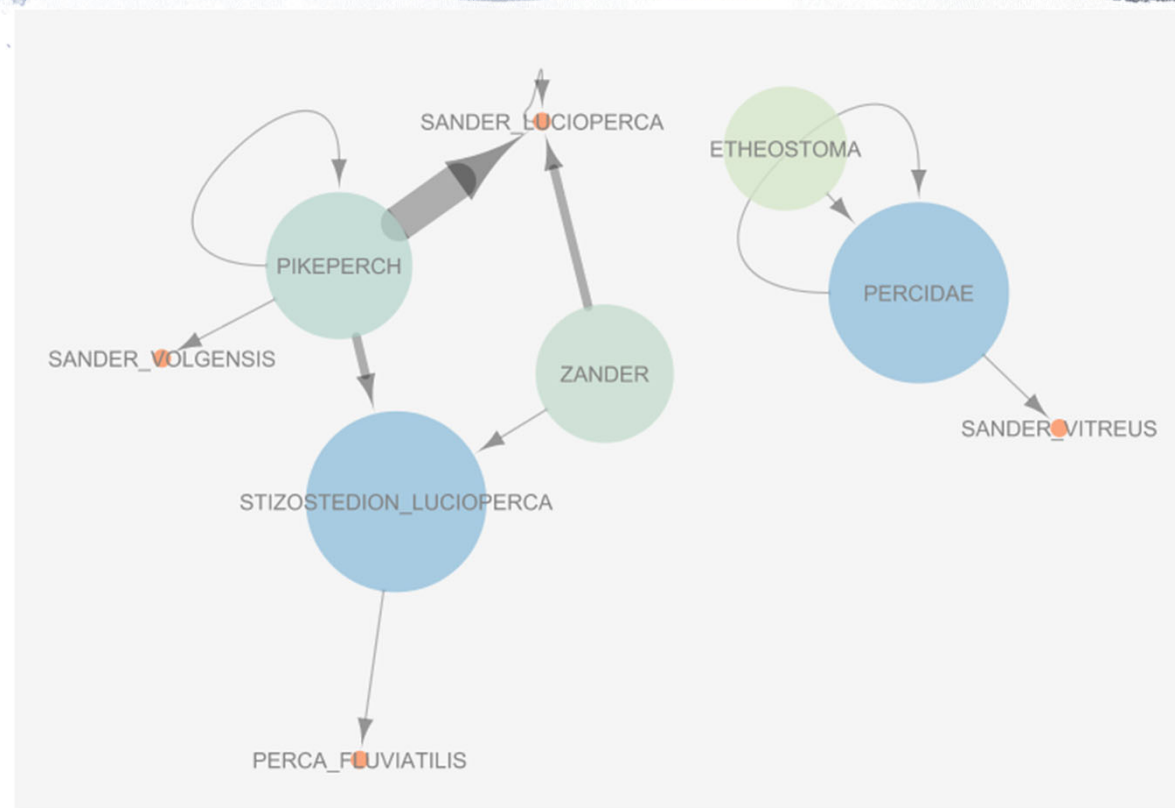
Ambiguities: accepted name versus variant usage



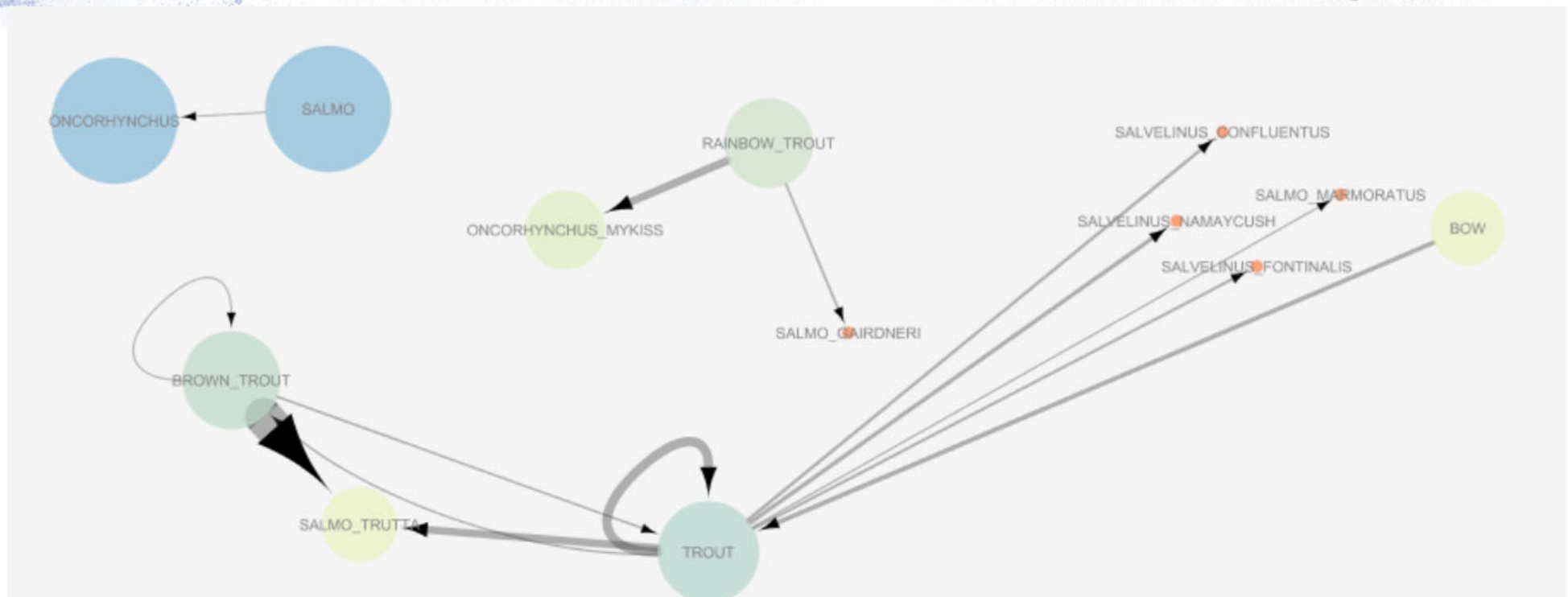
Ambiguities: accepted name versus variant usage



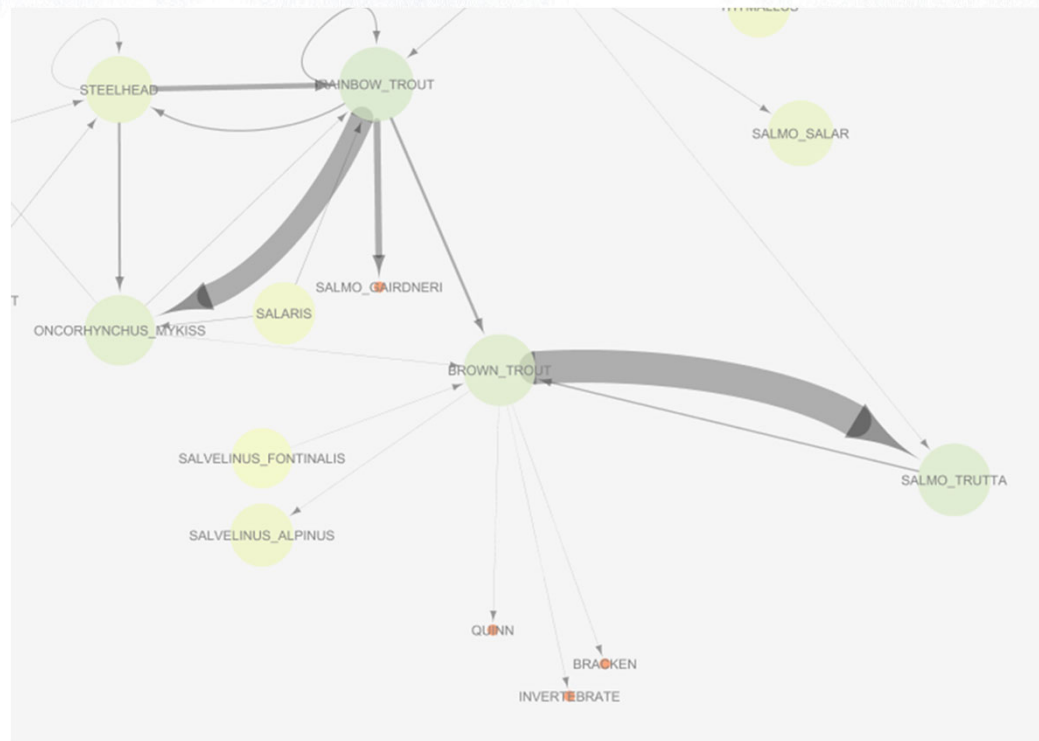
Ambiguities: accepted name versus variant usage



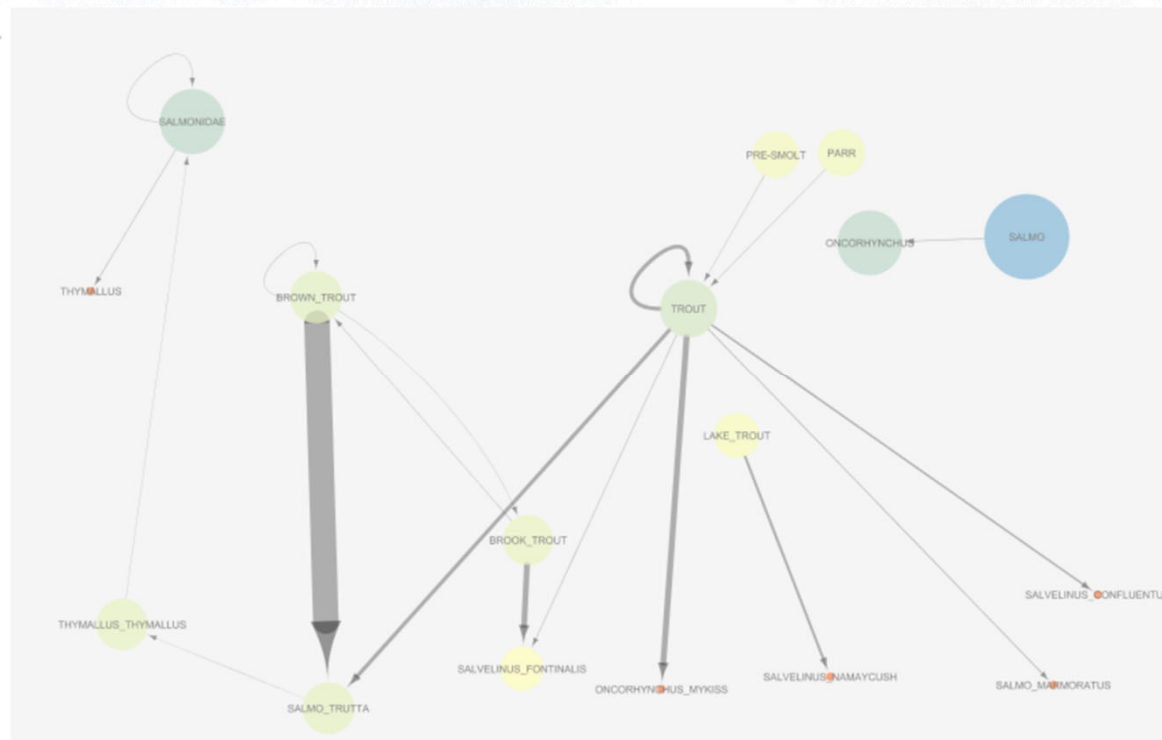
Contradiction with knowledge resource (JEFF)



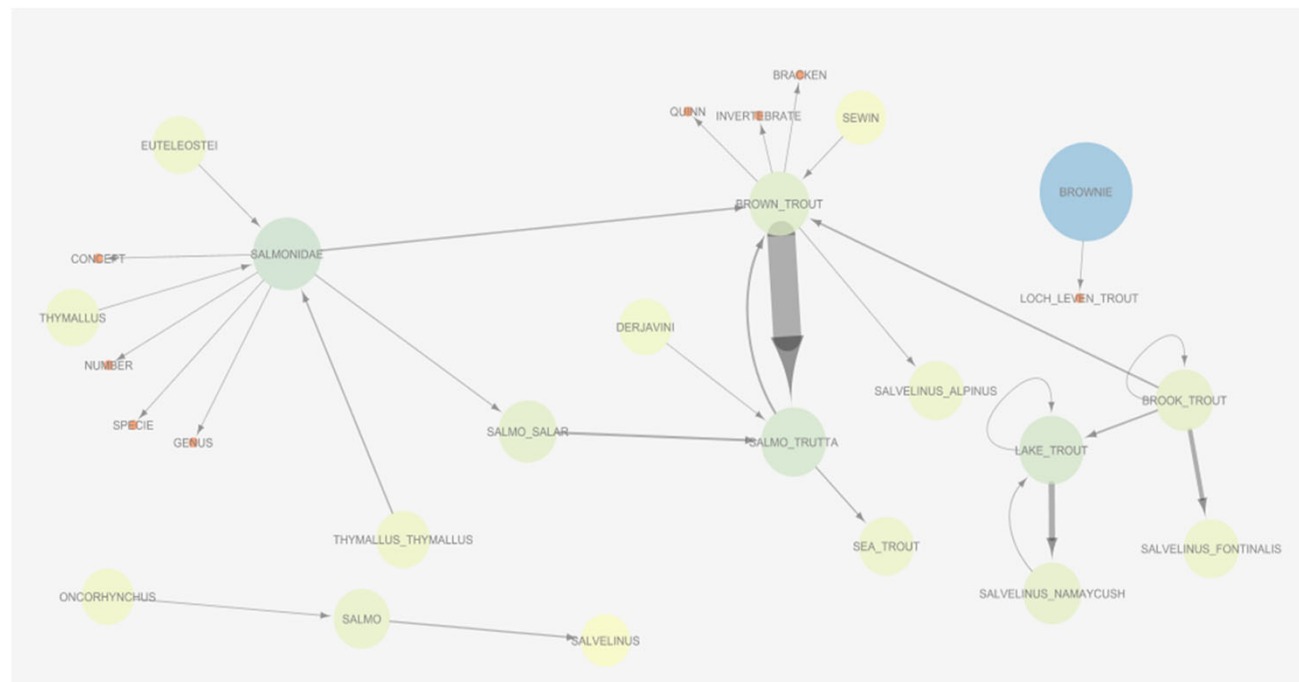
Contradiction with knowledge resource (WEB)

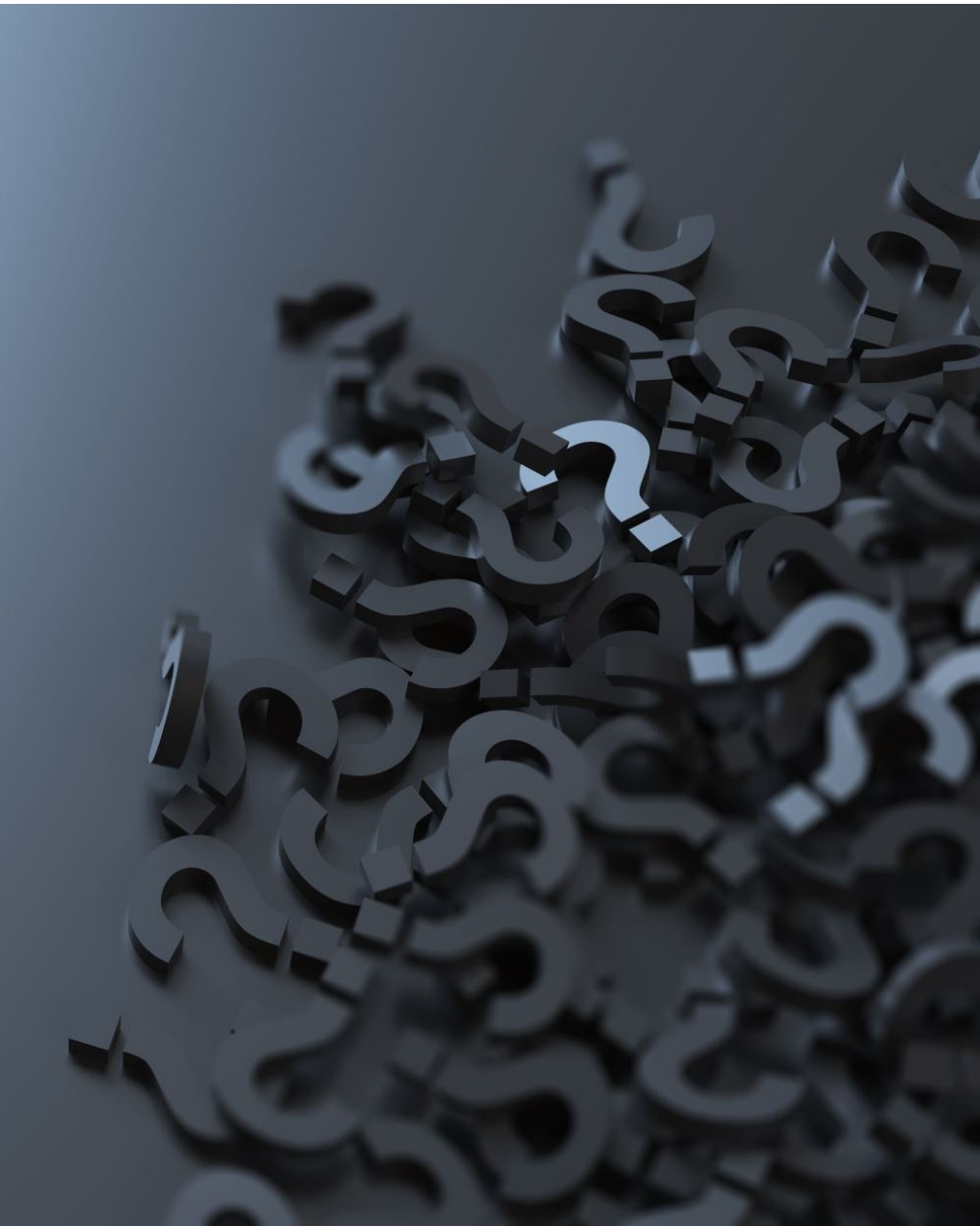


Contradiction with knowledge resource (JEFF)



Contradiction with knowledge resource (WEB)





Questions, comments?

- Network graph observations
- Differences and similarities between the data sets
- Knowledge representation resource quality