Data Summaries: Distilling Best Practices

October 7, 2020

@danphillips_ca
https://orcid.org/0000-0002-0386-0132

DALHOUSIE UNIVERSITY

MEOPAR

MITACS

Definitions

> **What is a “Summary”?**
>
> For the purposes of this talk, “summary” is the text that is intended to supply the primary overview of the contents of a dataset.
>
> Synonyms: description; abstract

Definitions

Repository
A platform to store datasets.

Deposit
A dataset and/or collection of datasets (and their metadata) entered into a repository.

Federated Search
A system in which a user can search multiple repositories at the same time.

Who are you?

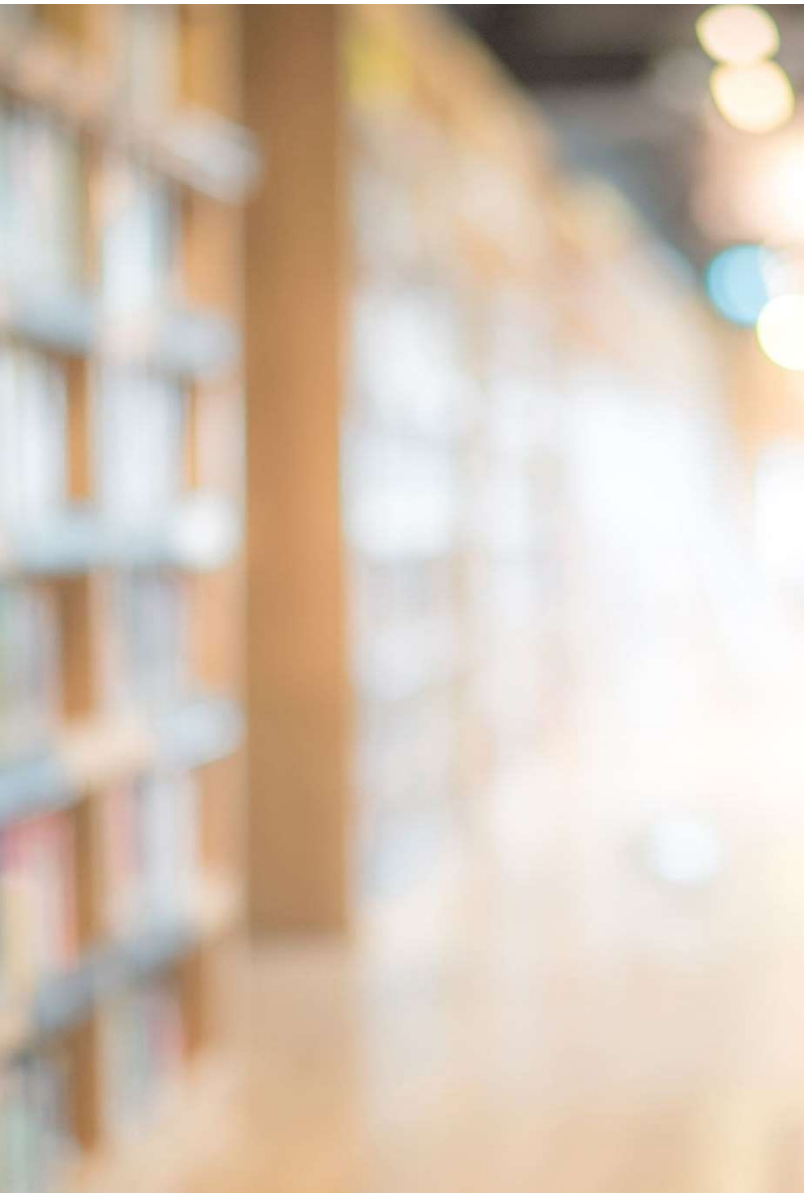
A. Librarian

B. Data Specialist

C. Researcher (New)

D. Researcher (Seasoned)

E. Other



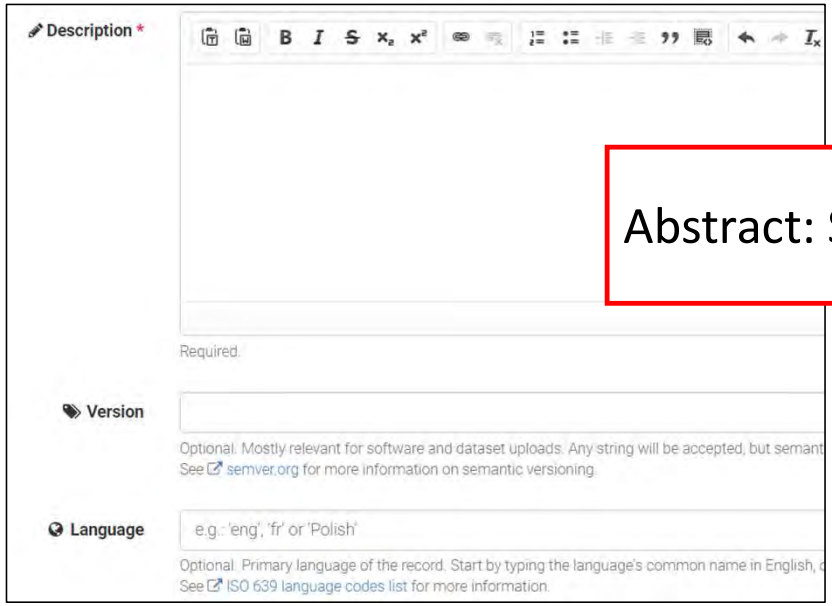
Overview

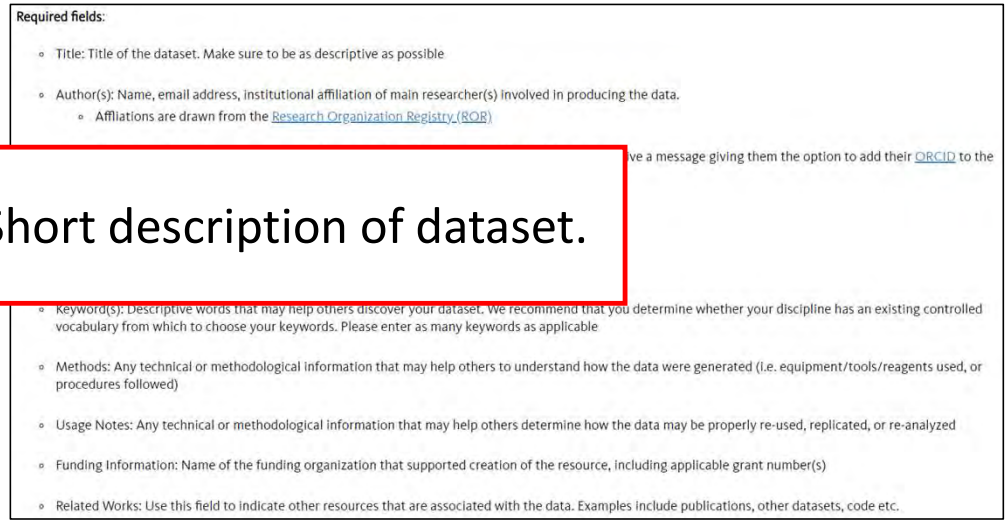
- Observations
- Advice from other fields
  & Proposed Best Practices
- Discussion

Observation 1:

Repositories
don’t offer guidance
on how to write a summary.

Zenodo and Data Dryad

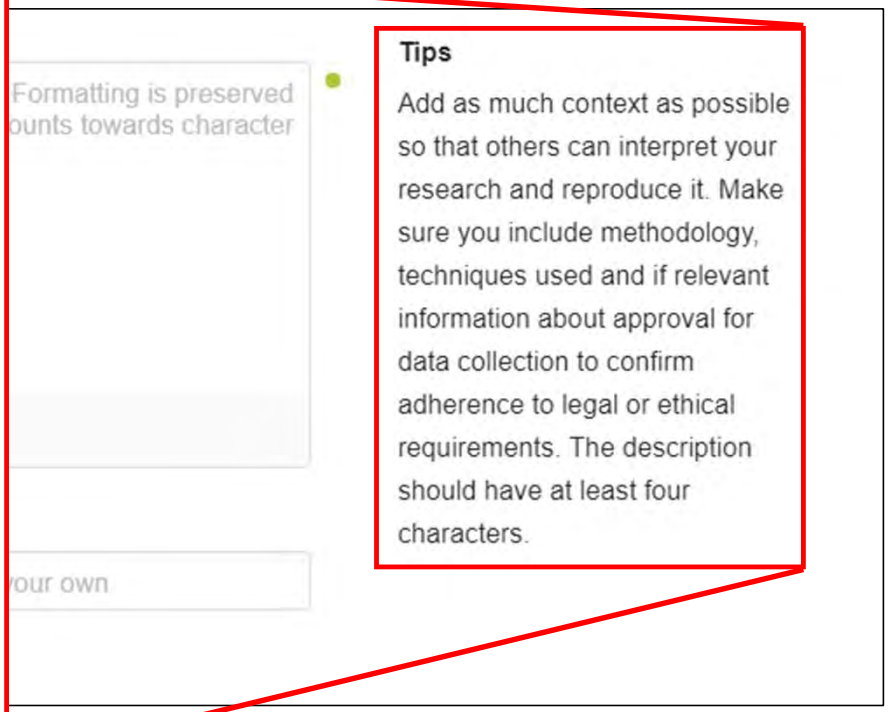




Abstract: Short description of dataset.

Figshare

Tips
Add as much context as possible so that others can interpret your research and reproduce it. Make sure you include methodology, techniques used and if relevant information about approval for data collection to confirm adherence to legal or ethical requirements. The description should have at least four characters.



Dataverse

A summary describing the purpose, nature, and scope of the Dataset.

			g Teens survey was nmunication Development media communication Canada. The survey collected responses from Canadian youth using an online questionnaire that asks about social media use including, platform type, frequency of use, activity type, and location of use. This information is supplemented with the respondent's demographic and household characteristics.
Date	In cases where a Dataset contains more than one description (for example, one might be supplied by the data producer and another prepared by the data repository where the data are deposited), the date attribute is used to distinguish between the two descriptions. The date attribute follows the ISO convention of YYYY-MM-DD.	optional	2018-01-18

(Dataverse North Metadata Best Practices Guide Version 2.0, 2020)

Google Datasets

Required properties	
`description`	A short summary describing a dataset. • The summary must be between 50 and 5000 characters long. • The summary may include Markdown syntax. Embedded images need to use absolute path URLs (instead of relative paths). • When using the JSON-LD format, denote new lines with `\n` (two characters: backslash and lower case letter "n").
`name`	`Text` A descriptive name of a dataset. For example, "Snow depth in the Northern Hemisphere". Guidelines • Use unique names for distinct datasets whenever possible. **Recommended**: `"Snow depth in the Northern Hemisphere"` and `"Snow depth in the Southern Hemisphere"` for two different datasets. **Not recommended**: `"Snow depth"` and `"Snow depth"` for two different datasets.

(Google, 2020)

What about the standards?

Data Description Initiative

Unformatted

Listing of major variables in this study

Describing the purpose, nature, and scope of the data collection, special characteristics of its contents, major subject areas covered, and what questions the PIs attempted to answer.

Mention of using multiple abstracts

bit.ly/DDIabstract

DataCite Metadata Schema

Description is mentioned as the most important of the non-mandatory properties.

Two descriptionType subproperties are recommended for discovery: Abstract and Methods.

Line breaks for new paragraphs. Otherwise, no formatting.

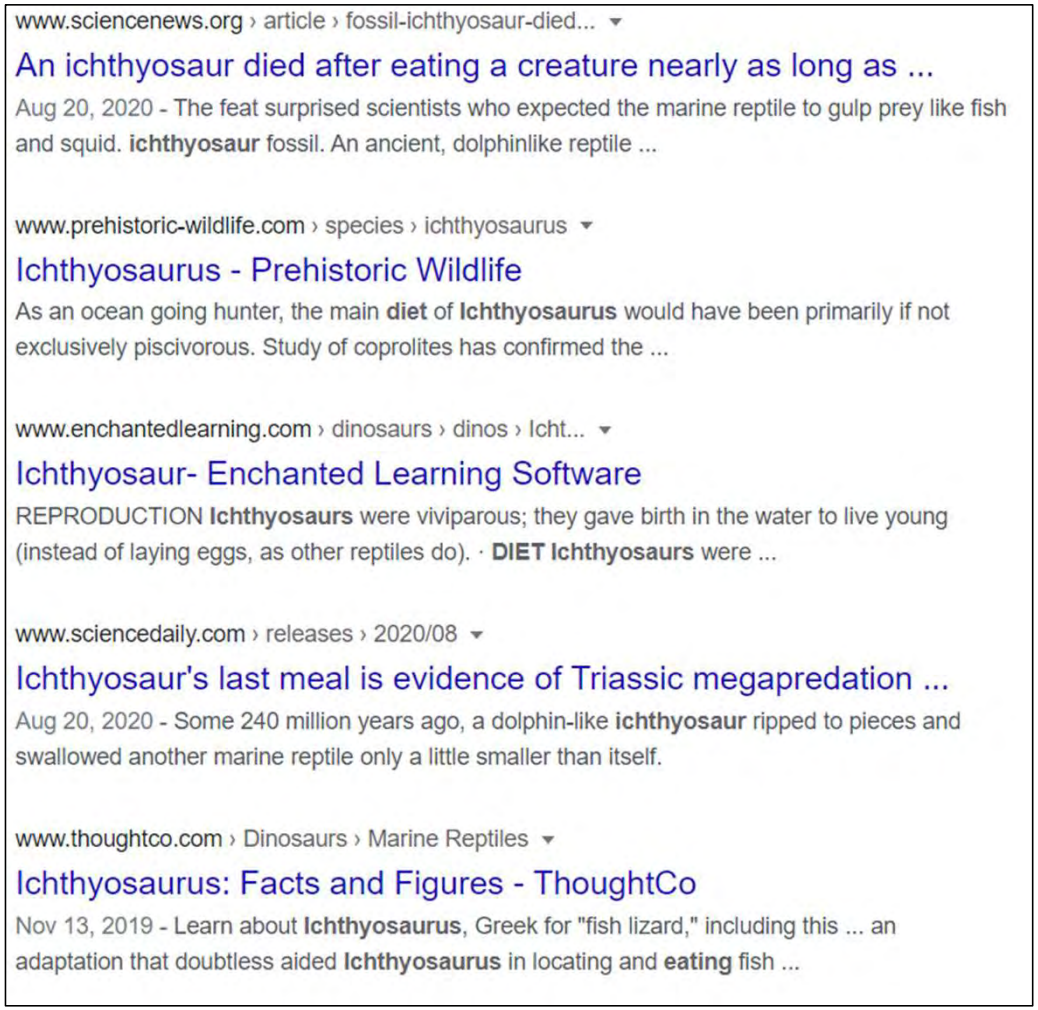
bit.ly/DataCiteAbstract

Platform	Typically Used By/As
Data Dryad	Public
Figshare	Public
Pangaea	Public
Zenodo	Public
ScholarsPortal Dataverse	Academic Platform
FRDR-DRDR	Federated
Google Dataset Search	Federated
CKAN Datasets	Government Open Data

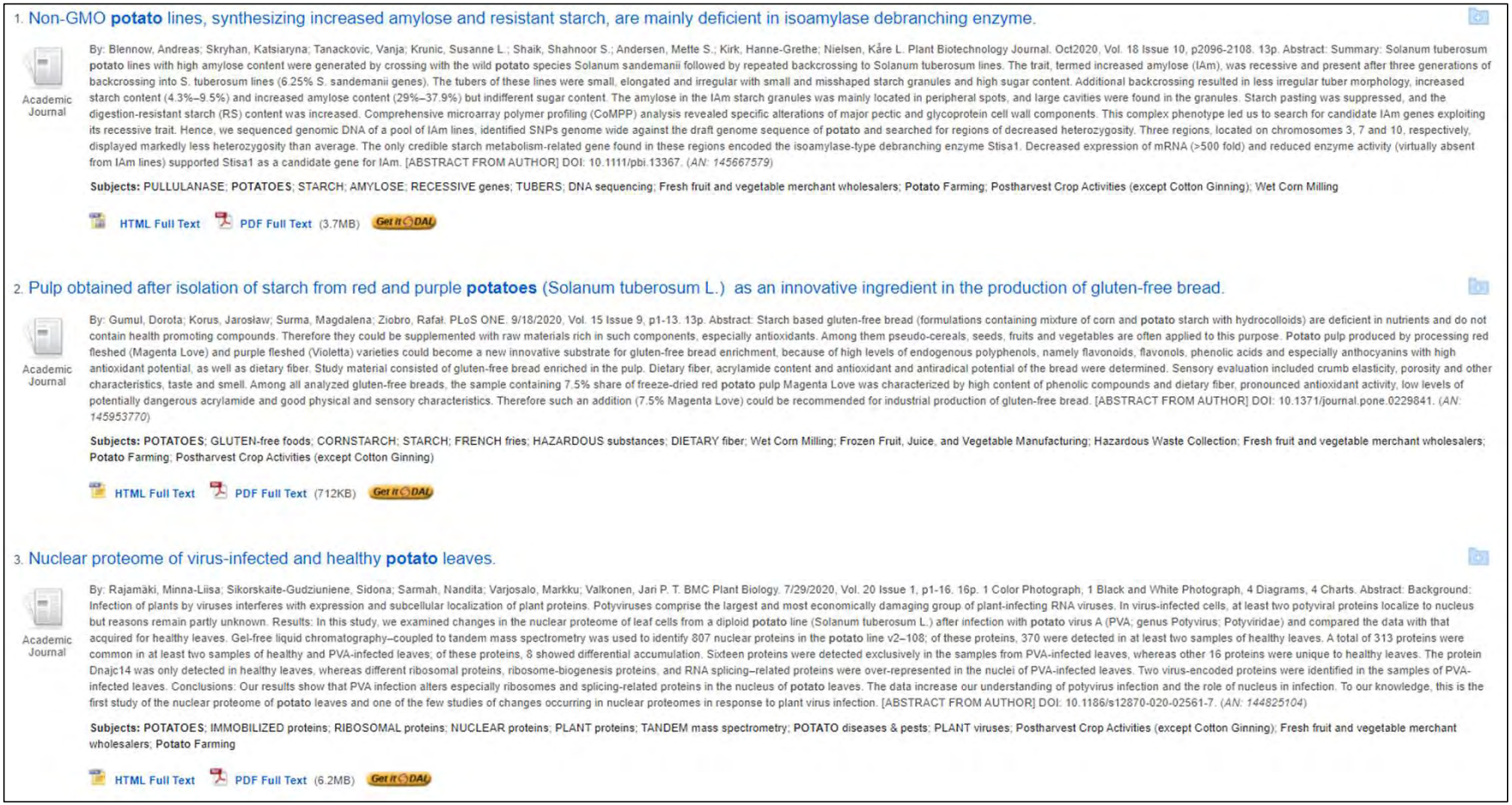
Observation 2:

Searching
for data
feels… different.

Google


www.sciencenews.org › article › fossil-ichthyosaur-died...
An ichthyosaur died after eating a creature nearly as long as ...
Aug 20, 2020 - The feat surprised scientists who expected the marine reptile to gulp prey like fish and squid. ichthyosaur fossil. An ancient, dolphinlike reptile ...
www.prehistoric-wildlife.com › species › ichthyosaurus
Ichthyosaurus - Prehistoric Wildlife
As an ocean going hunter, the main diet of Ichthyosaurus would have been primarily if not exclusively piscivorous. Study of coprolites has confirmed the ...
www.enchantedlearning.com › dinosaurs › dinos › Icht...
Ichthyosaur- Enchanted Learning Software
REPRODUCTION Ichthyosaurs were viviparous; they gave birth in the water to live young (instead of laying eggs, as other reptiles do). · DIET Ichthyosaurs were ...
www.sciencedaily.com › releases › 2020/08
Ichthyosaur's last meal is evidence of Triassic megapredation ...
Aug 20, 2020 - Some 240 million years ago, a dolphin-like ichthyosaur ripped to pieces and swallowed another marine reptile only a little smaller than itself.
www.thoughtco.com › Dinosaurs › Marine Reptiles
Ichthyosaurus: Facts and Figures - ThoughtCo
Nov 13, 2019 - Learn about Ichthyosaurus, Greek for "fish lizard," including this ... an adaptation that doubtless aided Ichthyosaurus in locating and eating fish ...

EBSCO Academic Search Premier



1. Non-GMO **potato** lines, synthesizing increased amylose and resistant starch, are mainly deficient in isoamylase debranching enzyme.

Academic Journal

By: Blennow, Andreas; Skryhan, Katsiaryna; Tanackovic, Vanja; Krunic, Susanne L.; Shaik, Shahnoor S.; Andersen, Mette S.; Kirk, Hanne-Grethe; Nielsen, Kåre L. Plant Biotechnology Journal. Oct2020, Vol. 18 Issue 10, p2096-2108. 13p. Abstract: Summary: Solanum tuberosum **potato** lines with high amylose content were generated by crossing with the wild **potato** species Solanum sandemanii followed by repeated backcrossing to Solanum tuberosum lines. The trait, termed increased amylose (IAm), was recessive and present after three generations of backcrossing into S. tuberosum lines (6.25% S. sandemanii genes). The tubers of these lines were small, elongated and irregular with small and misshaped starch granules and high sugar content. Additional backcrossing resulted in less irregular tuber morphology, increased starch content (4.3%–9.5%) and increased amylose content (29%–37.9%) but indifferent sugar content. The amylose in the IAm starch granules was mainly located in peripheral spots, and large cavities were found in the granules. Starch pasting was suppressed, and the digestion-resistant starch (RS) content was increased. Comprehensive microarray polymer profiling (CoMPP) analysis revealed specific alterations of major pectic and glycoprotein cell wall components. This complex phenotype led us to search for candidate IAm genes exploiting its recessive trait. Hence, we sequenced genomic DNA of a pool of IAm lines, identified SNPs genome wide against the draft genome sequence of **potato** and searched for regions of decreased heterozygosity. Three regions, located on chromosomes 3, 7 and 10, respectively, displayed markedly less heterozygosity than average. The only credible starch metabolism-related gene found in these regions encoded the isoamylase-type debranching enzyme Stisa1. Decreased expression of mRNA (>500 fold) and reduced enzyme activity (virtually absent from IAm lines) supported Stisa1 as a candidate gene for IAm. [ABSTRACT FROM AUTHOR] DOI: 10.1111/pbi.13367. (*AN: 145667579*)

Subjects: PULLULANASE; **POTATOES**; STARCH; AMYLOSE; RECESSIVE genes; TUBERS; DNA sequencing; Fresh fruit and vegetable merchant wholesalers; **Potato** Farming; Postharvest Crop Activities (except Cotton Ginning); Wet Corn Milling

HTML Full Text PDF Full Text (3.7MB) Get it DAL

2. Pulp obtained after isolation of starch from red and purple **potatoes** (Solanum tuberosum L.) as an innovative ingredient in the production of gluten-free bread.

Academic Journal

By: Gumul, Dorota; Korus, Jarosław; Surma, Magdalena; Ziobro, Rafał. PLoS ONE. 9/18/2020, Vol. 15 Issue 9, p1-13. 13p. Abstract: Starch based gluten-free bread (formulations containing mixture of corn and **potato** starch with hydrocolloids) are deficient in nutrients and do not contain health promoting compounds. Therefore they could be supplemented with raw materials rich in such components, especially antioxidants. Among them pseudo-cereals, seeds, fruits and vegetables are often applied to this purpose. **Potato** pulp produced by processing red fleshed (Magenta Love) and purple fleshed (Violetta) varieties could become a new innovative substrate for gluten-free bread enrichment, because of high levels of endogenous polyphenols, namely flavonoids, flavonols, phenolic acids and especially anthocyanins with high antioxidant potential, as well as dietary fiber. Study material consisted of gluten-free bread enriched in the pulp. Dietary fiber, acrylamide content and antioxidant and antiradical potential of the bread were determined. Sensory evaluation included crumb elasticity, porosity and other characteristics, taste and smell. Among all analyzed gluten-free breads, the sample containing 7.5% share of freeze-dried red **potato** pulp Magenta Love was characterized by high content of phenolic compounds and dietary fiber, pronounced antioxidant activity, low levels of potentially dangerous acrylamide and good physical and sensory characteristics. Therefore such an addition (7.5% Magenta Love) could be recommended for industrial production of gluten-free bread. [ABSTRACT FROM AUTHOR] DOI: 10.1371/journal.pone.0229841. (*AN: 145953770*)

Subjects: **POTATOES**; GLUTEN-free foods; CORNSTARCH; STARCH; FRENCH fries; HAZARDOUS substances; DIETARY fiber; Wet Corn Milling; Frozen Fruit, Juice, and Vegetable Manufacturing; Hazardous Waste Collection; Fresh fruit and vegetable merchant wholesalers; **Potato** Farming; Postharvest Crop Activities (except Cotton Ginning)

HTML Full Text PDF Full Text (712KB) Get it DAL

3. Nuclear proteome of virus-infected and healthy **potato** leaves.

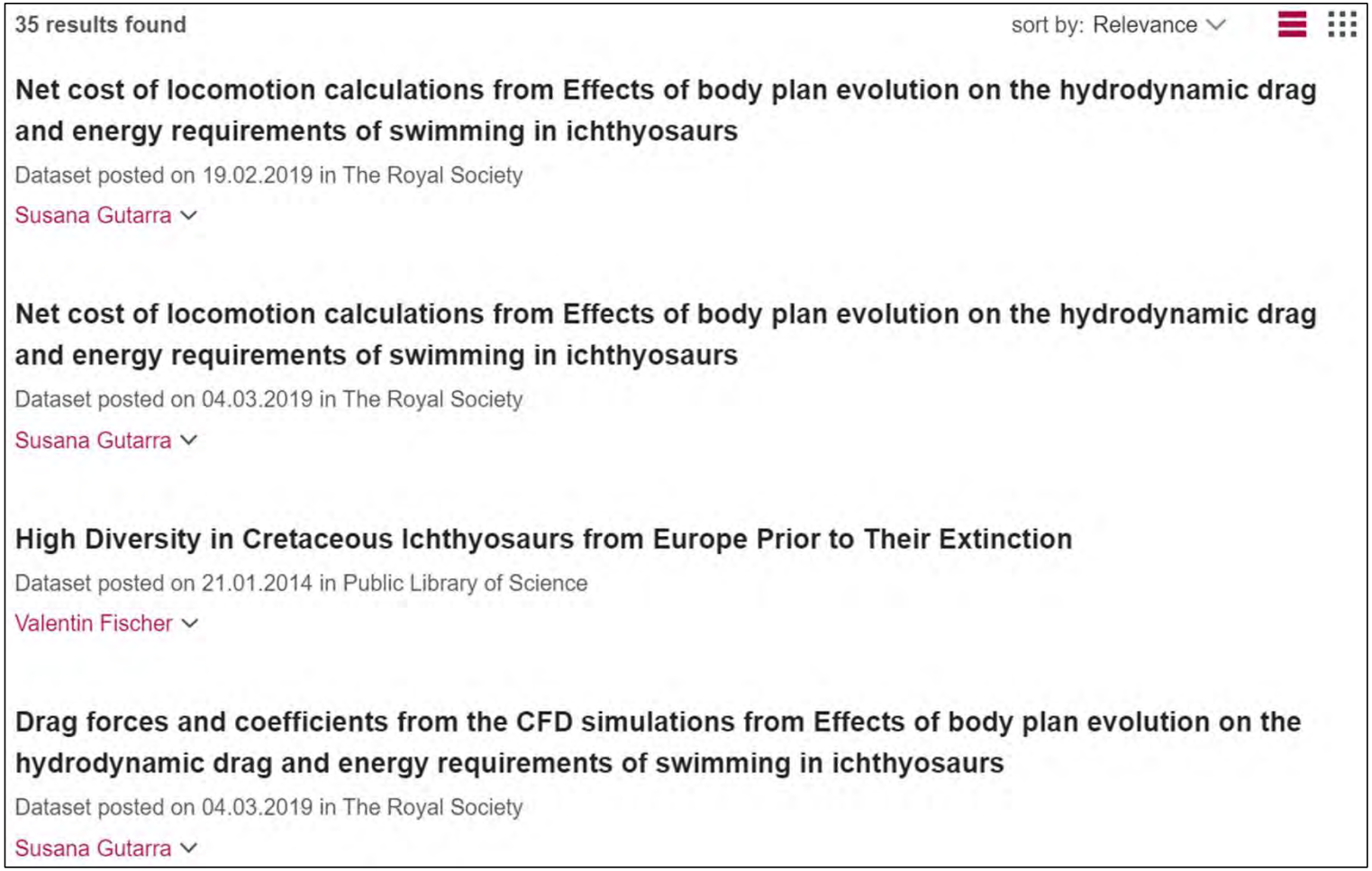
Academic Journal

By: Rajamäki, Minna-Liisa; Sikorskaite-Gudziuniene, Sidona; Sarmah, Nandita; Varjosalo, Markku; Valkonen, Jari P. T. BMC Plant Biology. 7/29/2020, Vol. 20 Issue 1, p1-16. 16p. 1 Color Photograph, 1 Black and White Photograph, 4 Diagrams, 4 Charts. Abstract: Background: Infection of plants by viruses interferes with expression and subcellular localization of plant proteins. Potyviruses comprise the largest and most economically damaging group of plant-infecting RNA viruses. In virus-infected cells, at least two potyviral proteins localize to nucleus but reasons remain partly unknown. Results: In this study, we examined changes in the nuclear proteome of leaf cells from a diploid **potato** line (Solanum tuberosum L.) after infection with **potato** virus A (PVA; genus Potyvirus; Potyviridae) and compared the data with that acquired for healthy leaves. Gel-free liquid chromatography–coupled to tandem mass spectrometry was used to identify 807 nuclear proteins in the **potato** line v2–108; of these proteins, 370 were detected in at least two samples of healthy leaves. A total of 313 proteins were common in at least two samples of healthy and PVA-infected leaves; of these proteins, 8 showed differential accumulation. Sixteen proteins were detected exclusively in the samples from PVA-infected leaves, whereas other 16 proteins were unique to healthy leaves. The protein Dnajc14 was only detected in healthy leaves, whereas different ribosomal proteins, ribosome-biogenesis proteins, and RNA splicing–related proteins were over-represented in the nuclei of PVA-infected leaves. Two virus-encoded proteins were identified in the samples of PVA-infected leaves. Conclusions: Our results show that PVA infection alters especially ribosomes and splicing-related proteins in the nucleus of **potato** leaves. The data increase our understanding of potyvirus infection and the role of nucleus in infection. To our knowledge, this is the first study of the nuclear proteome of **potato** leaves and one of the few studies of changes occurring in nuclear proteomes in response to plant virus infection. [ABSTRACT FROM AUTHOR] DOI: 10.1186/s12870-020-02561-7. (*AN: 144825104*)

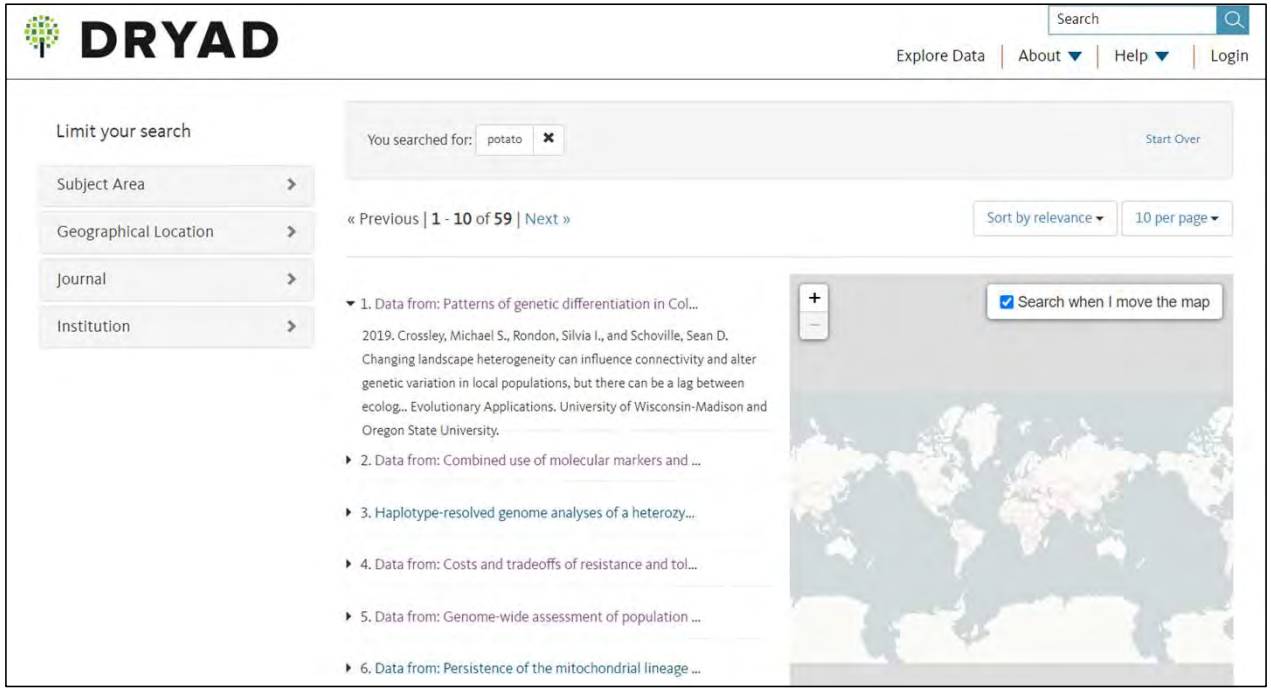
Subjects: **POTATOES**; IMMOBILIZED proteins; RIBOSOMAL proteins; NUCLEAR proteins; PLANT proteins; TANDEM mass spectrometry; **POTATO** diseases & pests; PLANT viruses; Postharvest Crop Activities (except Cotton Ginning); Fresh fruit and vegetable merchant wholesalers; **Potato** Farming

HTML Full Text PDF Full Text (6.2MB) Get it DAL

FigShare


35 results found
sort by: Relevance
Net cost of locomotion calculations from Effects of body plan evolution on the hydrodynamic drag and energy requirements of swimming in ichthyosaurs
Dataset posted on 19.02.2019 in The Royal Society
Susana Gutarra
Net cost of locomotion calculations from Effects of body plan evolution on the hydrodynamic drag and energy requirements of swimming in ichthyosaurs
Dataset posted on 04.03.2019 in The Royal Society
Susana Gutarra
High Diversity in Cretaceous Ichthyosaurs from Europe Prior to Their Extinction
Dataset posted on 21.01.2014 in Public Library of Science
Valentin Fischer
Drag forces and coefficients from the CFD simulations from Effects of body plan evolution on the hydrodynamic drag and energy requirements of swimming in ichthyosaurs
Dataset posted on 04.03.2019 in The Royal Society
Susana Gutarra

Data Dryad


DRYAD
Search
Explore Data
About
Help
Login
Limit your search
Subject Area
Geographical Location
Journal
Institution
You searched for: potato
Start Over
« Previous | 1 - 10 of 59 | Next »
Sort by relevance
10 per page
1. Data from: Patterns of genetic differentiation in Col...
2019. Crossley, Michael S., Rondon, Silvia I., and Schoville, Sean D. Changing landscape heterogeneity can influence connectivity and alter genetic variation in local populations, but there can be a lag between ecolog... Evolutionary Applications. University of Wisconsin-Madison and Oregon State University.
2. Data from: Combined use of molecular markers and ...
3. Haplotype-resolved genome analyses of a heterozy...
4. Data from: Costs and tradeoffs of resistance and tol...
5. Data from: Genome-wide assessment of population ...
6. Data from: Persistence of the mitochondrial lineage ...
Search when I move the map

Pangaea

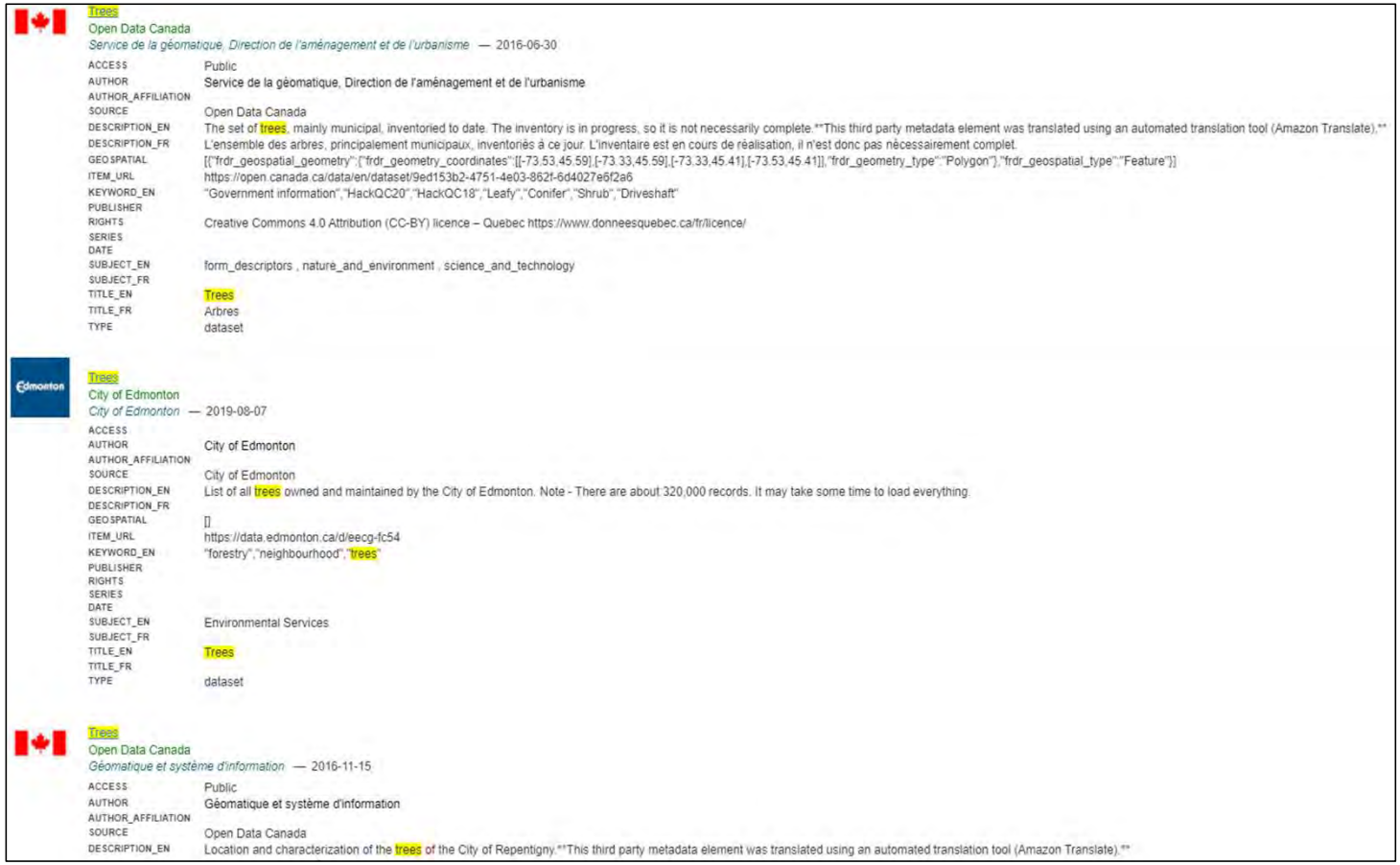
1. **Zweifel, R; Etzold, S; Haeni, M et al. (2020):** Dendrometer, sap flow, meteorology and soil volumetric water content measurements during a long-term irrigation experiment in a Scots pine forest at Pfynwald, Swiss Rhone valley (2011-2017)
   Size: 4 datasets
   https://doi.org/10.1594/PANGAEA.918631 - Score: 35.2

2. **De Beenhouwer, M; Aerts, R; Hundera, K et al. (2015):** Epiphytic orchid diversity found on trees and shrubs in coffee cultivated Afromontane rainforest, SW Ethiopia
   Supplement to: **De Beenhouwer, M; Aerts, R; Hundera, K et al. (2015):** Management intensification in Ethiopian coffee forests is associated with crown habitat contraction and loss of specialized epiphytic orchid species. *Basic and Applied Ecology*
   Size: 4 datasets
   https://doi.org/10.1594/PANGAEA.834569 - Score: 35.08

3. **Pérez-Luque, AJ; Gea-Izquierdo, G; Zamora, R (2020):** Resilience to drought of relict Mediterranean Quercus pyrenaica populations in the southern Iberian (Sierra Nevada, Spain)
   Size: 4 datasets
   https://doi.org/10.1594/PANGAEA.922054 - Score: 34.79

4. **Achterberg, IEM; Eckstein, J; Birkholz, B et al. (2017):** Dendrochronological dates of subfossil pines (Pinus sylvestris L.), relative dates of floating chronology, and peat stratigraphy from site "TOMO_south" at Totes Moor, northwest Germany
   Supplement to: **Achterberg, IEM; Eckstein, J; Birkholz, B et al. (2018):** Dendrochronologically dated pine stumps document phase-wise bog expansion at a northwest German site between ca. 6700 and ca. 3400 BC. *Climate of the Past*
   Size: 3 datasets
   https://doi.org/10.1594/PANGAEA.884249 - Score: 33.79

FRDR



Trees
Open Data Canada
Service de la géomatique, Direction de l'aménagement et de l'urbanisme — 2016-06-30

ACCESS	Public
AUTHOR	Service de la géomatique, Direction de l'aménagement et de l'urbanisme
AUTHOR_AFFILIATION	
SOURCE	Open Data Canada
DESCRIPTION_EN	The set of trees, mainly municipal, inventoried to date. The inventory is in progress, so it is not necessarily complete.**This third party metadata element was translated using an automated translation tool (Amazon Translate).**
DESCRIPTION_FR	L'ensemble des arbres, principalement municipaux, inventoriés à ce jour. L'inventaire est en cours de réalisation, il n'est donc pas nécessairement complet.
GEOSPATIAL	[{"frdr_geospatial_geometry":{"frdr_geometry_coordinates":[[-73.53,45.59],[-73.33,45.59],[-73.33,45.41],[-73.53,45.41]],"frdr_geometry_type":"Polygon"},"frdr_geospatial_type":"Feature"}]
ITEM_URL	https://open.canada.ca/data/en/dataset/9ed153b2-4751-4e03-862f-6d4027e6f2a6
KEYWORD_EN	"Government information","HackQC20","HackQC18","Leafy","Conifer","Shrub","Driveshaft"
PUBLISHER	
RIGHTS	Creative Commons 4.0 Attribution (CC-BY) licence – Quebec https://www.donneesquebec.ca/fr/licence/
SERIES	
DATE	
SUBJECT_EN	form_descriptors , nature_and_environment , science_and_technology
SUBJECT_FR	
TITLE_EN	Trees
TITLE_FR	Arbres
TYPE	dataset

Trees
City of Edmonton
City of Edmonton — 2019-08-07

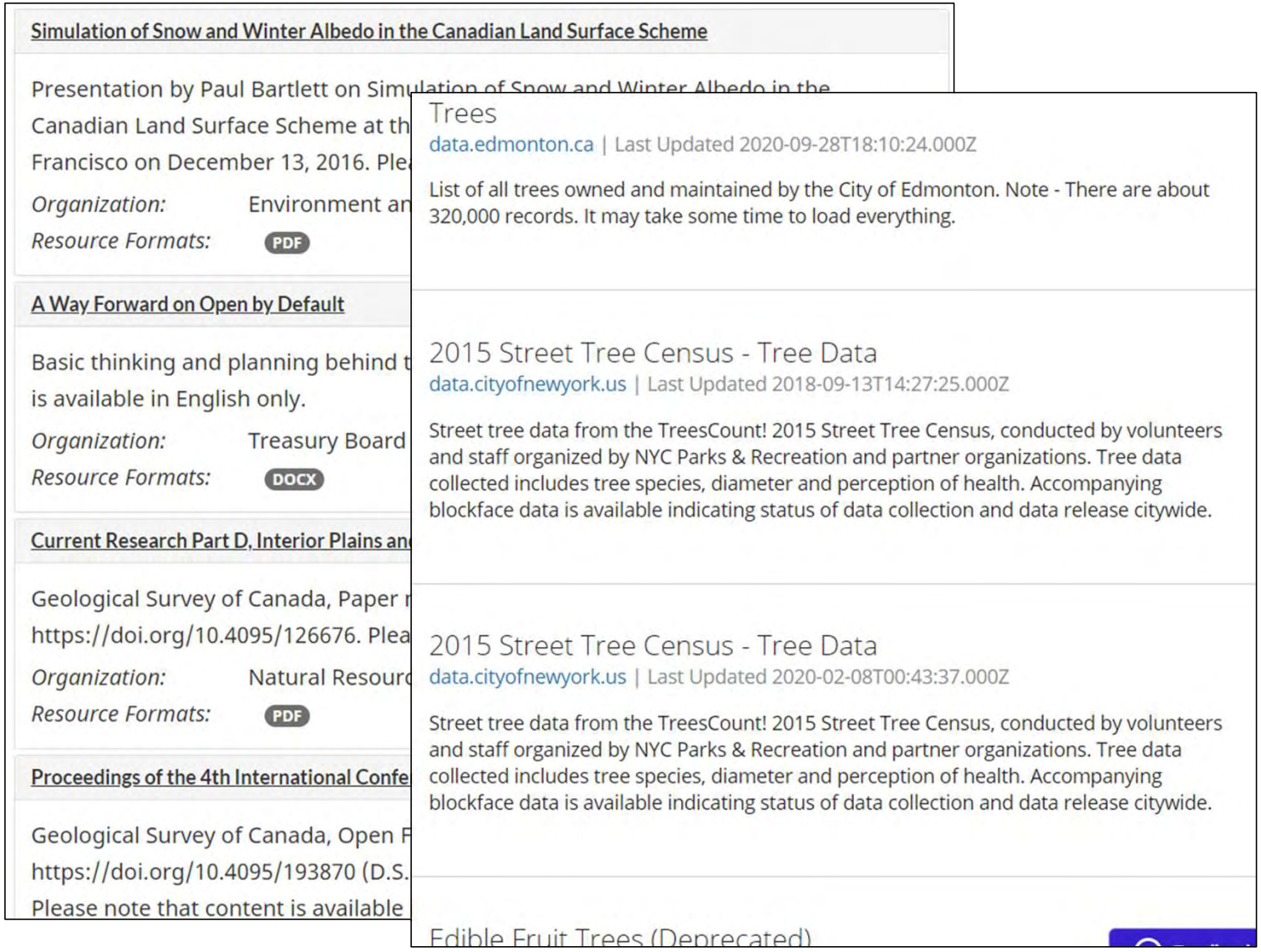
ACCESS	
AUTHOR	City of Edmonton
AUTHOR_AFFILIATION	
SOURCE	City of Edmonton
DESCRIPTION_EN	List of all trees owned and maintained by the City of Edmonton. Note - There are about 320,000 records. It may take some time to load everything.
DESCRIPTION_FR	
GEOSPATIAL	[]
ITEM_URL	https://data.edmonton.ca/d/eecg-fc54
KEYWORD_EN	"forestry","neighbourhood","trees"
PUBLISHER	
RIGHTS	
SERIES	
DATE	
SUBJECT_EN	Environmental Services
SUBJECT_FR	
TITLE_EN	Trees
TITLE_FR	
TYPE	dataset

Trees
Open Data Canada
Géomatique et système d'information — 2016-11-15

ACCESS	Public
AUTHOR	Géomatique et système d'information
AUTHOR_AFFILIATION	
SOURCE	Open Data Canada
DESCRIPTION_EN	Location and characterization of the trees of the City of Repentigny.**This third party metadata element was translated using an automated translation tool (Amazon Translate).**

CKAN

(and other Open Data Platforms)

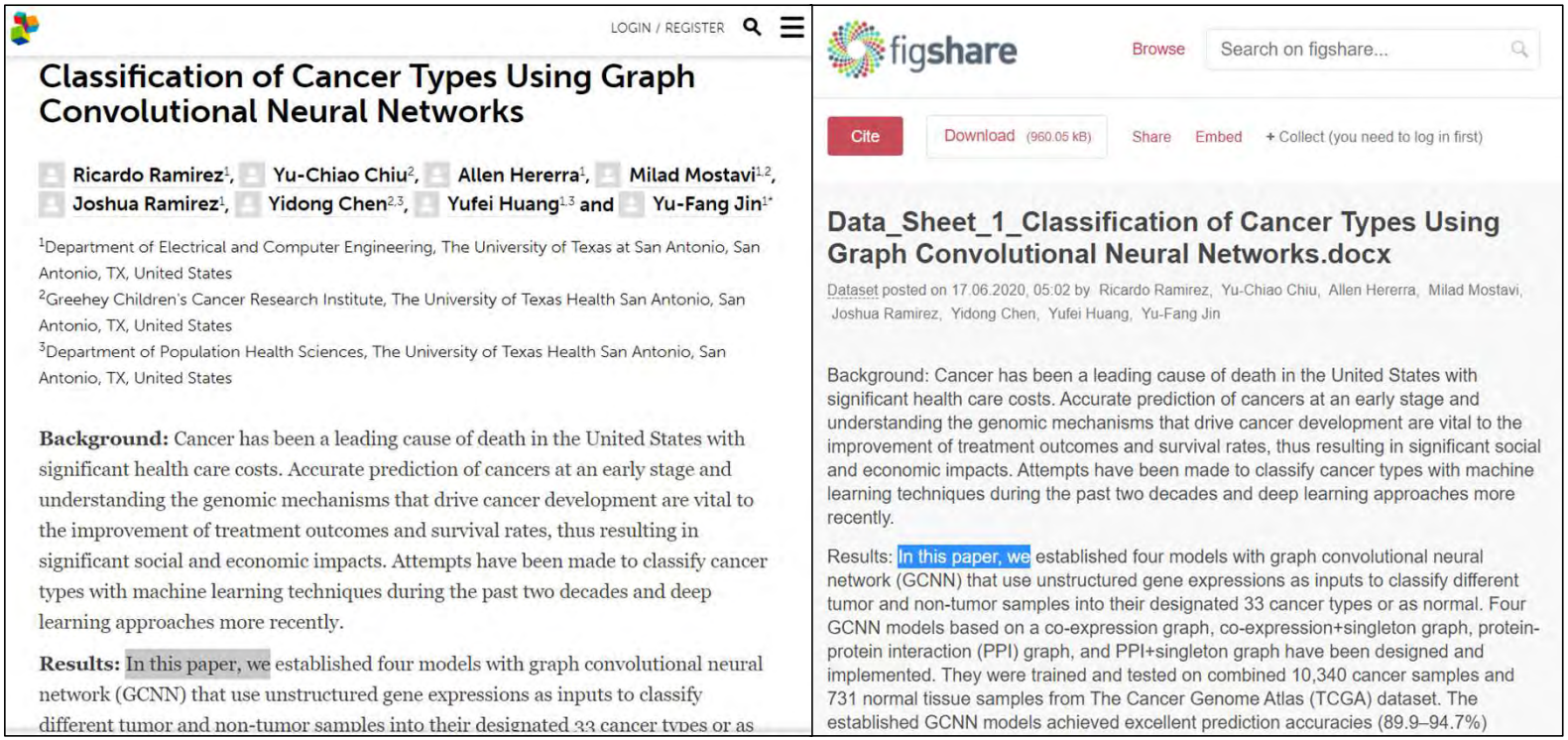


What does this mean in terms of *writing* a summary?

Observation 3:

With what we have,
what do Data Summaries
end up looking like?

Not ideal.



LOGIN / REGISTER

Classification of Cancer Types Using Graph Convolutional Neural Networks

Ricardo Ramirez[1], Yu-Chiao Chiu[2], Allen Hererra[1], Milad Mostavi[1,2], Joshua Ramirez[1], Yidong Chen[2,3], Yufei Huang[1,3] and Yu-Fang Jin[1*]

[1]Department of Electrical and Computer Engineering, The University of Texas at San Antonio, San Antonio, TX, United States
[2]Greehey Children's Cancer Research Institute, The University of Texas Health San Antonio, San Antonio, TX, United States
[3]Department of Population Health Sciences, The University of Texas Health San Antonio, San Antonio, TX, United States

Background: Cancer has been a leading cause of death in the United States with significant health care costs. Accurate prediction of cancers at an early stage and understanding the genomic mechanisms that drive cancer development are vital to the improvement of treatment outcomes and survival rates, thus resulting in significant social and economic impacts. Attempts have been made to classify cancer types with machine learning techniques during the past two decades and deep learning approaches more recently.

Results: In this paper, we established four models with graph convolutional neural network (GCNN) that use unstructured gene expressions as inputs to classify different tumor and non-tumor samples into their designated 33 cancer types or as

figshare

Browse Search on figshare...

Cite Download (960.05 kB) Share Embed + Collect (you need to log in first)

Data_Sheet_1_Classification of Cancer Types Using Graph Convolutional Neural Networks.docx

Dataset posted on 17.06.2020, 05:02 by Ricardo Ramirez, Yu-Chiao Chiu, Allen Hererra, Milad Mostavi, Joshua Ramirez, Yidong Chen, Yufei Huang, Yu-Fang Jin

Background: Cancer has been a leading cause of death in the United States with significant health care costs. Accurate prediction of cancers at an early stage and understanding the genomic mechanisms that drive cancer development are vital to the improvement of treatment outcomes and survival rates, thus resulting in significant social and economic impacts. Attempts have been made to classify cancer types with machine learning techniques during the past two decades and deep learning approaches more recently.

Results: In this paper, we established four models with graph convolutional neural network (GCNN) that use unstructured gene expressions as inputs to classify different tumor and non-tumor samples into their designated 33 cancer types or as normal. Four GCNN models based on a co-expression graph, co-expression+singleton graph, protein-protein interaction (PPI) graph, and PPI+singleton graph have been designed and implemented. They were trained and tested on combined 10,340 cancer samples and 731 normal tissue samples from The Cancer Genome Atlas (TCGA) dataset. The established GCNN models achieved excellent prediction accuracies (89.9–94.7%)

Quaternary Science Reviews
Volume 32, 16 January 2012, Pages 25-47

Marine geological constraints for the grounding-line position of the Antarctic Ice Sheet on the southern Weddell Sea shelf at the Last Glacial Maximum

Claus-Dieter Hillenbrand [a], Martin Melles [b], Gerhard Kuhn [c], Robert D. Larter [a]

Show more

https://doi.org/10.1016/j.quascirev.2011.11.017

Get rights and content

Abstract

The history of grounded ice-sheet extent on the southern Weddell Sea shelf during the Last Glacial Maximum (LGM) and the timing of post-LGM ice-sheet retreat are poorly constrained. Several glaciological models reconstructed widespread grounding and major thickening of the Antarctic Ice Sheet in the Weddell Sea sector at the LGM. In contrast, recently published onshore data and modelling results concluded only very limited LGM-thickening of glaciers and ice streams feeding into the modern Filchner and Ronne ice shelves. These studies concluded that during the LGM ice shelves rather than grounded ice covered the Filchner and Ronne troughs, two deep palaeo-ice stream troughs eroded into the southern Weddell Sea shelf.

PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

Hillenbrand, Claus-Dieter; Melles, Martin; Kuhn, Gerhard; Larter, Robert D (2011): Last glacial maximum sediment record of the southern Weddell Sea shelf. *PANGAEA*, https://doi.org/10.1594/PANGAEA.771774,

Supplement to: Hillenbrand, C-D et al. (2012): Marine geological constraints for the grounding-line position of the Antarctic Ice Sheet on the southern Weddell Sea shelf at the Last Glacial Maximum. *Quaternary Science Reviews*, **32**, 25-47, https://doi.org/10.1016/j.quascirev.2011.11.017

Always quote above citation when using data! You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Copy Citation | Facebook | Twitter | Show Map | Google Earth

Abstract:

Abstract: The history of grounded ice-sheet extent on the southern Weddell Sea shelf during the Last Glacial Maximum (LGM) and the timing of post-LGM ice-sheet retreat are poorly constrained. Several glaciological models reconstructed widespread grounding and major thickening of the Antarctic Ice Sheet in the Weddell Sea sector at the LGM. In contrast, recently published onshore data and modelling results concluded only very limited LGM-thickening of glaciers and ice streams feeding into the modern Filchner and Ronne ice shelves. These studies concluded that during the LGM ice shelves rather than grounded ice covered the Filchner and Ronne troughs, two deep palaeo-ice stream troughs eroded into the southern Weddell Sea shelf.

Summary from associated article abstract

Data Dryad

95.3%

n = 338

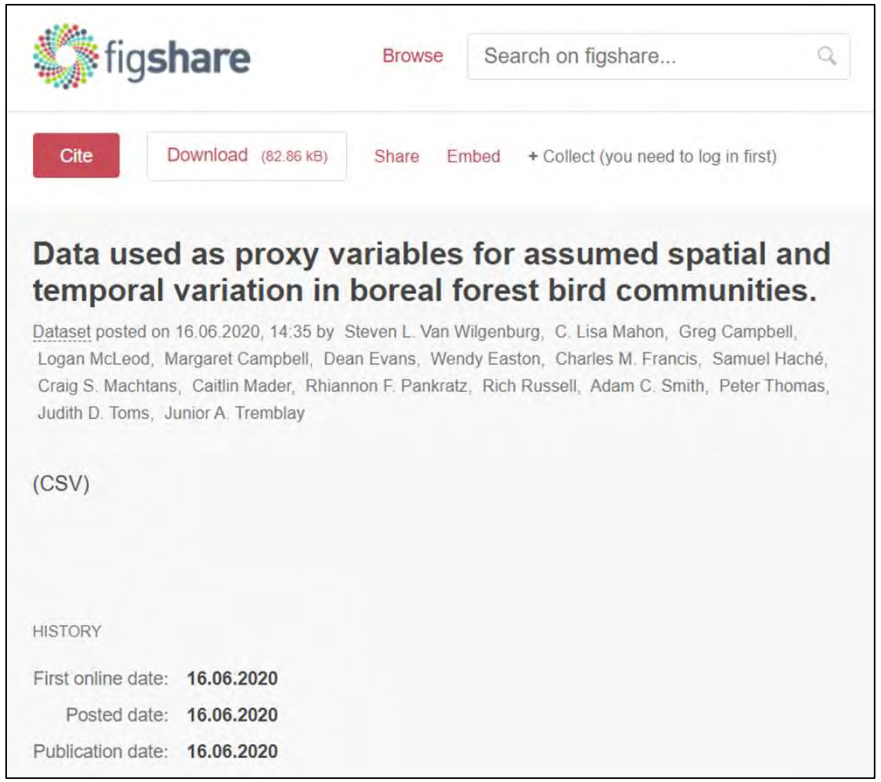
Figshare

29.1%

n = 350

Also not ideal.

Yep.



Distilling some Best Practices

Principles

FAIR Principles	Statement of International Cataloging Principles	Rules for Archival Design	Principles for Digital Data Management (Tri-Agency)
Findable Accessible Inter-operable Reusable	Convenience of user above all.	Description should be integrated and proceed from a common set of rules.	Roles and expectations related to data management.

Studies on Effective Abstracting

- Use of specialized language between disciplines (Montesi & Urdiciain, 2005)
- Use of structured abstracts (Hartley & Sydes, 1995; Ganzi, 2011; Hartley, 2014)

Datasets should be described with attention to the searcher.

The summary should begin as being appropriate for all audiences.

Datasets should be described as independent objects.

The context of data creation should be described.

The dataset should have structure.

Datasets should be described with attention to the searcher.

The summary should begin as appropriate to all potential audiences.

Datasets should be described as independent objects.

Context of data creation must be described.

The dataset should have structure.

Assessment of the reproductive success of lake sturgeon at the Drummondville spawning ground in spring 2015

The evaluation of reproductive success is based on the estimation of the number of eggs deposited on the spawning ground in relation to the number of larvae that have drifted.

First, we defined the area and the spawning period and captured eggs using an egg sensor and a drift net. Then, we carried out a sturgeon capture-mark-recapture campaign to estimate the size of the spawning contingent. Finally, we caught drifting larvae.

CONTEXT Lake sturgeon (fish) have been designated a threatened species in the Upper Great Lakes/Saint Lawrence River area. These data were collected from May through June 2015 as part of a project to determine their reproductive success in the forebay of the power station of Drummondville, Quebec. Reproductive success is based on the estimation of eggs deposited on the spawning ground in relation to the number of larvae.

METHODOLOGY To collect these data, we first defined the area and the spawning period and captured eggs using an egg sensor and a drift net. Then, we carried out a sturgeon capture-mark-recapture campaign to estimate the size of the spawning contingent. Finally, we caught drifting larvae.

CONTENT This collection contains three tables and one report. Tables include counts of lake sturgeon eggs, counts of lake sturgeon (including sex, size, and weight), and counts of drifting larvae. Data also include water temperature at that time nets were lowered and lifted.

NOTES

- Similar studies were conducted in 2014 and 2017
  See also: https://catalogue.cioos.ca/dataset/f97b9be8-c231-4fd3-b26d-8e7da408dcc9
- Dataset is in the French language.

Discussion

How do you use the summary fields when searching for data?

- Is it different than other types of search?
- Do you use it early or late in your search process?

What issues haven't been addressed? (Are they a problem?)

- Length
- Non-interdisciplinary-friendly data
- Relationship between summary and other metadata fields

How would these concepts impact practice?

Datasets should be described with attention to the searcher.

The summary should begin as appropriate to all potential audiences.

Datasets should be described as independent objects.

Context of data creation must be described.

The dataset should have structure.

See Also

Bascik, T., Boisvert, P., Cooper, A., Gagnon, M., Goodwin, M., Huck, J., … Taylor, S. (2020, February 29). Dataverse North Metadata Best Practices Guide : Version 2.0 [R]. doi: http://dx.doi.org/10.14288/1.0388724

Cory Chapman. (n.d.). *Usability Recommendations* [Wiki]. Datadryad.Org. Retrieved June 24, 2020, from http://wiki.datadryad.org/Usability_Recommendations
Dryad. (2019). *Dryad Submission Process*. https://datadryad.org/stash/submission_process

The Dataverse Team. (2020, April). *Dataset + File Management*. Dataverse. http://guides.dataverse.org/en/latest/user/dataset-management.html

Google. (n.d.). *Dataset | Search for Developers*. Google Developers. Retrieved June 24, 2020, from https://developers.google.com/search/docs/data-types/dataset

Hartley, J., & Sydes, M. (1996). Which layout do you prefer? An analysis of readers' preferences for different typographic layouts of structured abstracts. *Excerpt from Their Structured Abstracts in the Social Sciences. British Lib. 1995*, *22*(1), 27–37. Library Literature & Information Science Full Text (H.W. Wilson).

Montesi, M., & Urdiciain, B. G. (2005). Abstracts: Problems classified from the user perspective. *Journal of Information Science*, *31*(6), 515–526. Library Literature & Information Science Full Text (H.W. Wilson).

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Wang, H.-R. (2016). *How to write a good Data in Brief article*. 5.