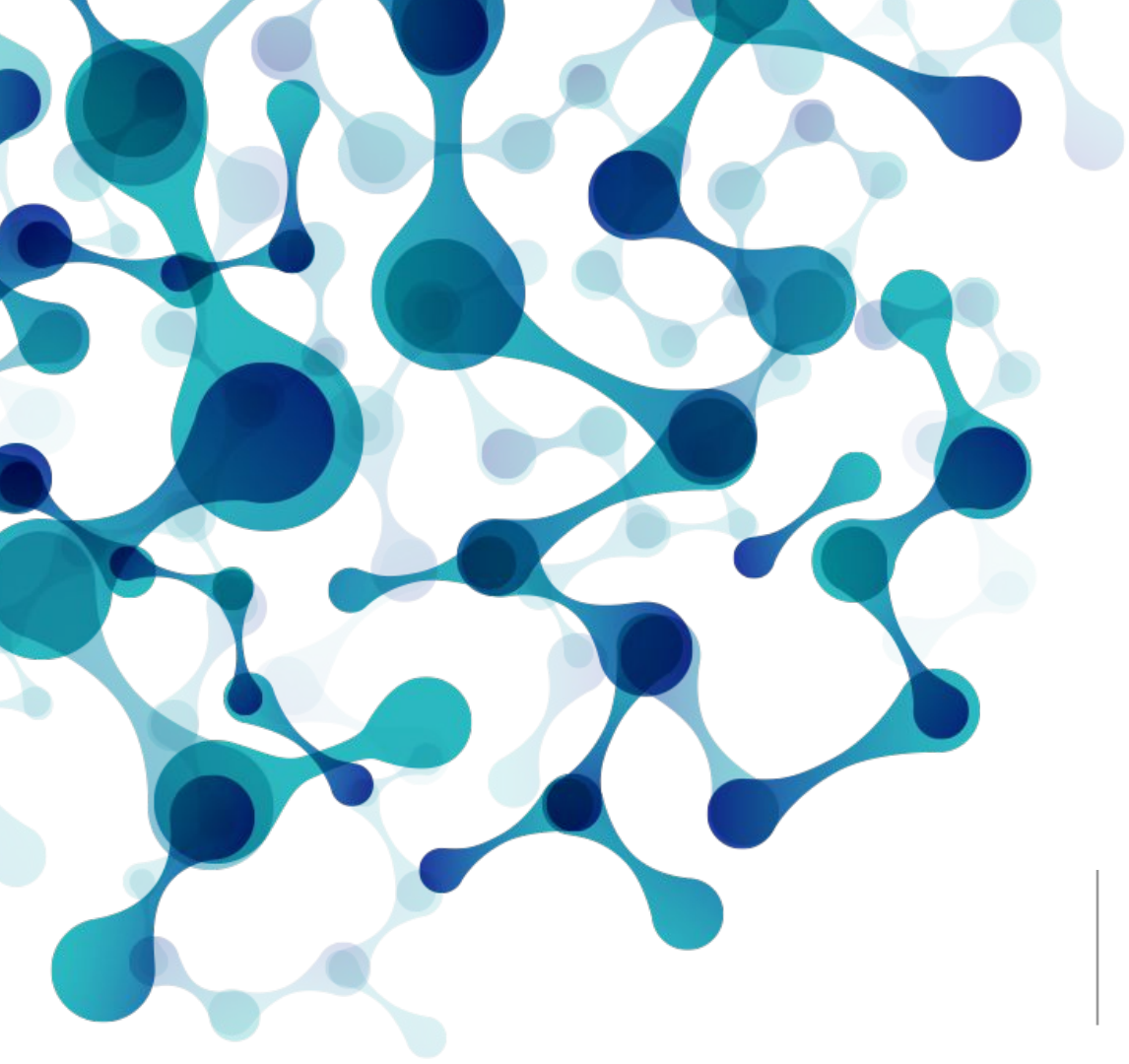


THOR
HTTP://PROJECT-THOR.EU

Where are we with
data citation

Make data ~~great again~~ count

Xiaoli Chen xiaoli.chen@cern.ch
THOR Bootcamp Budapest
2017. 09. 29

Agenda

- Why cite data?
- How data citation works
- Principles of data citation
- Dynamic data citation
- Conclusion

Why cite data?

Why cite data?

playing both its infrared $\Delta v = 1$ vibrational transition (see also fig. 7) and its first allowed electronic transition X $\rightarrow$ A in the ultraviolet. Molecular data on this electronic transition can be found in refs. [65–70]. Finally, we depict in red the sensitivity to absorption of hidden photons heavier than 11 eV onto the first three

https://arxiv.org/pdf/1709.05354.pdf

Why cite data?

playing both its infrared $\Delta v = 1$ vibrational transition (see also fig. 7) and its first allowed electronic transi- tion X → A in the ı electronic transition nally, we depict in r hidden photons heav

[65] D. M. Cooper and S. R. Langhoff, The Journal of Chemical Physics **74**, 1200 (1981).

[66] C. Chackerian Jr, The Journal of Chemical Physics **65**, 4228 (1976).

[67] M. Halmann and I. Laulicht, The Astrophysical Journal Supplement Series **12**, 307 (1966).

[68] P. H. Krupenie, *The band spectrum of carbon monoxide*, Tech. Rep. (NATIONAL STANDARD REFERENCE DATA SYSTEM, 1966).

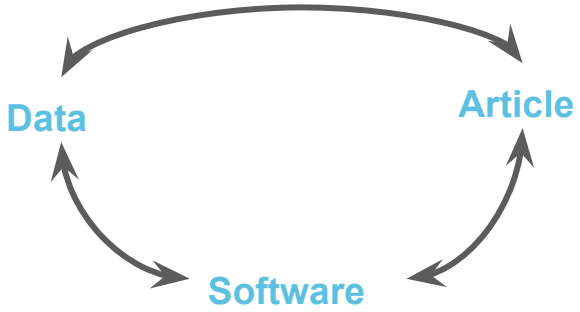
[69] .

[70] S. Tilford and J. Simmons, Journal of Physical and Chemical Reference Data **1**, 147 (1972).

? ? ?

https://arxiv.org/pdf/1709.05354.pdf

Why cite data?



- Support proper **attribution and credit**
- Support **collaboration** and **reuse** of data
- Enable **reproducibility** of findings
- Foster **faster** and more **efficient** research **progress**, and
- Provide the **means to share** data with future researchers

How data citation works

Building a Culture of Data Citation

ands.org.au

ands
NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

Principles of data citation (FORCE11)

Data citation principles (1/3)

1. Importance

Data is the new ~~oil~~ bacon

2. Credit and attribution

3. Evidence

https://www.force11.org/group/joint-declaration-data-citatio

Cranmer, Kyle S.

PERSONAL INFORMATION

Personal Details (HepNames)

Name	Kyle S. Cranmer
Current Institution	New York U.
E-mail	cranmer@cern.ch
Links	http://theoryandpractice.org/ https://www.linkedin.com/in/ky... http://twitter.com/KyleCranmer... https://github.com/cranmer
Fields	HEP-EX HEP-PH PHYSICS
Experiments	FNAL-E-0830 CERN-LHC-ATLAS CERN-LEP-ALEPH
Identifiers	BAI: K.S.Cranmer.1 INSPIRE: INSPIRE-00074922 ORCID: 0000-0002-5769-7094 ARXIV: cranmer_k_1

Period	Rank	Institution
2007	SENIOR	New York U.
2005 – 2007	PD	Brookhaven
1999 – 2005	PHD	Wisconsin U., Madison
1995 – 1999	UG	Rice U.

Update Details

PUBLICATIONS AND OUTPUT

Publications | Datasets | External

1. Data from Table 1 from: Measurement of Z-pair production in e+ e- collisions and constraints on anomalous neutral gauge couplings
2. Data from Table 1 from: Measurement of the Cross Section for open b-Quark Production in Two-Photon Interactions at LEP
3. Data from Table 1 from: Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
4. Data from Table 2 from: Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
5. Data from Table 3 from: Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector
6. Data from Table 0: Final Variable Distribution from: Search for pair production of heavy vector-like quarks decaying to high-$p_{mathrm{T}}$ W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector

Co-Authors

B.Mellado.1 (13)
W.Quayle.1 (11)
C.T.Potter.1 (8)
I.Aracena.1 (8)
M.Wielers.1 (8)
S.L.Wu.1 (8)
A.T.Watson.1 (7)
B.Vachon.1 (7)
C.Santamarina.Rios.1 (7)
G.Louppe.1 (7)
more

Papers

	All papers	Single authored
All papers	851	12
Book	0	0
ConferencePaper	37	10
Introductory	0	0
Lectures	0	0
Published	729	4
Review	5	0
Thesis	1	1
Proceedings	0	0

Subject Categories

Experiment-HEP (779)
Instrumentation (50)
Phenomenology-HEP (27)
Experiment-Nucl (26)

Frequent Keywords

ATLAS (687)
CERN LHC Coll (665)
experimental results (653)
p p: scattering (521)

STATS

Citations Summary

851 papers found, 838 of them citeable (published or arXiv)

	Citeable papers	Published only
Number of papers analyzed:	838	729
Number of citations:	85747	82241
Citations per paper (average):	102.3	112.8
h_{HEP} index [?]	132	132

Breakdown of papers by citations:

	Citeable papers	Published only
Renowned papers (500+)	19	18
Famous papers (250-499)	33	31
Very well-known papers (100-249)	147	145
Well-known papers (50-99)	182	181
Known papers (10-49)	305	293
Less known papers (1-9)	118	59
Unknown papers (0)	34	2

	Citeable datasets
Number of datasets	8900

Click here to view statistics without self-citations or RPP

Warning: The citations count should be interpreted with great care. Read the fine print

"As I watched more and more links turned red, I frantically combed the internet for archived versions of our country's most important polar policies."

https://www.theguardian.com/commentisfree/2017/mar/28/arctic-researcher-donald-trump-deleting-my-citations

Data citation principles (2/3)

Proxy Prefix Suffix

https://doi.org/10.5281/ZENODO.31780

4. Unique identification

5. Access

6. Persistence

https://doi.org/10.5438/55e5-t5c0

Data citation principles (3/3)

7. Specificity and Verifiability

8. Interoperability and Flexibility

Hide Publication Information

Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

The ATLAS collaboration

Aaboud, Morad , Aad, Georges , Abbott, Brad , Abdinov, Ovsat , Abeloos, Baptiste , Abidi, Syed Haider , AbouZeid, Ossama , Abraham, Nicola , Abramowicz, Halina , Abreu, Henso

No Journal Information, 2017

http://dx.doi.org/10.17182/hepdata.78402

INSPIRE Record HepData Resources

Abstract (data abstract)
CERN-LHC. A search for heavy neutral Higgs bosons and Z' bosons is performed using a data sample corresponding to an integrated luminosity of 36.1 fb^{-1} from proton-proton collisions at $\sqrt{s}$ = 13 TeV recorded by the ATLAS detector at the LHC during 2015 and 2016.

The 1l1tau_h channel fiducial region is defined as:
- 1e+>=1tau_h or 1mu+>=1tau_h
- ptLepton > 30 GeV
- |etaMuon| < 2.4, |etaEle| < 2.47 (excluding 1.37 < |etaEle| < 1.52)

Download All

Filter 29 data tables

Table 1
Data from Figure 5A
10.17182/hepdata.78402.v1/t2
Observed and predicted mTtot distribution in the b-veto category of the 1l1tau_h channel. Despite listing this as an exclusive final...

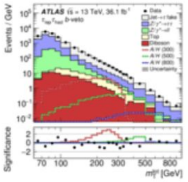
Table 2
Data from Figure 5B
10.17182/hepdata.78402.v1/t3
Observed and predicted mTtot distribution in the b-tag category of the 1l1tau_h channel. Despite listing this as an exclusive final...

Table 3
Data from Figure 5C
10.17182/hepdata.78402.v1/t4
Observed and predicted mTtot distribution in the b-veto category of the 2tau_h channel. Despite listing this as an exclusive final...

Table 4
Data from Figure 5D
10.17182/hepdata.78402.v1/t5
Observed and predicted mTtot distribution in the b-tag category of the 2tau_h channel. Despite listing this as an exclusive final

Table 1 10.17182/hepdata.78402.v1/t2

Resources http://www.hepdata.n JSON



Observed and predicted mTtot distribution in the b-veto category of the 1l1tau_h channel. Despite listing this as an exclusive final state (as there must be no b-jets), there is no explicit selection on the presence of additional light-flavour jets. Please note that the bin content is divided by the bin width in the paper figure, but not in the HepData table. In the paper, the first bin is cut off at 60 GeV for aesthetics but contains underflows down to 50 GeV as in the HepData table. The last bin includes overflows. The combined prediction for A and H bosons with masses of 300, 500 and 800 GeV and tan β = 10 in the hMSSM scenario are also provided.

cmenergies
13000.0

observables
N

phrases
Exclusive
Proton-Proton Scattering
Higgs

reactions
P P --> TAU+ TAU-

SQRT(S)	13000 GeV			
LUMINOSITY	36.1 fb^{-1}			
Channel	1l1tau_h			
Category	b-veto			
Process	Data	SM	$A/H(300)$	$A/H(500$
mTtot [GeV]	NUM EVENTS			
50 - 70	29628	29456.007812 ±356.4	1.056363	0.023294

Visualize

Principle 2: Credit and Attribution (e.g. authors, repositories or other distributors and contributors)

Principle 4: Unique Identifier (e.g. DOI, Handle.). **Principle 5, 6 Access, Persistence:** A persistent identifier that provides access and metadata

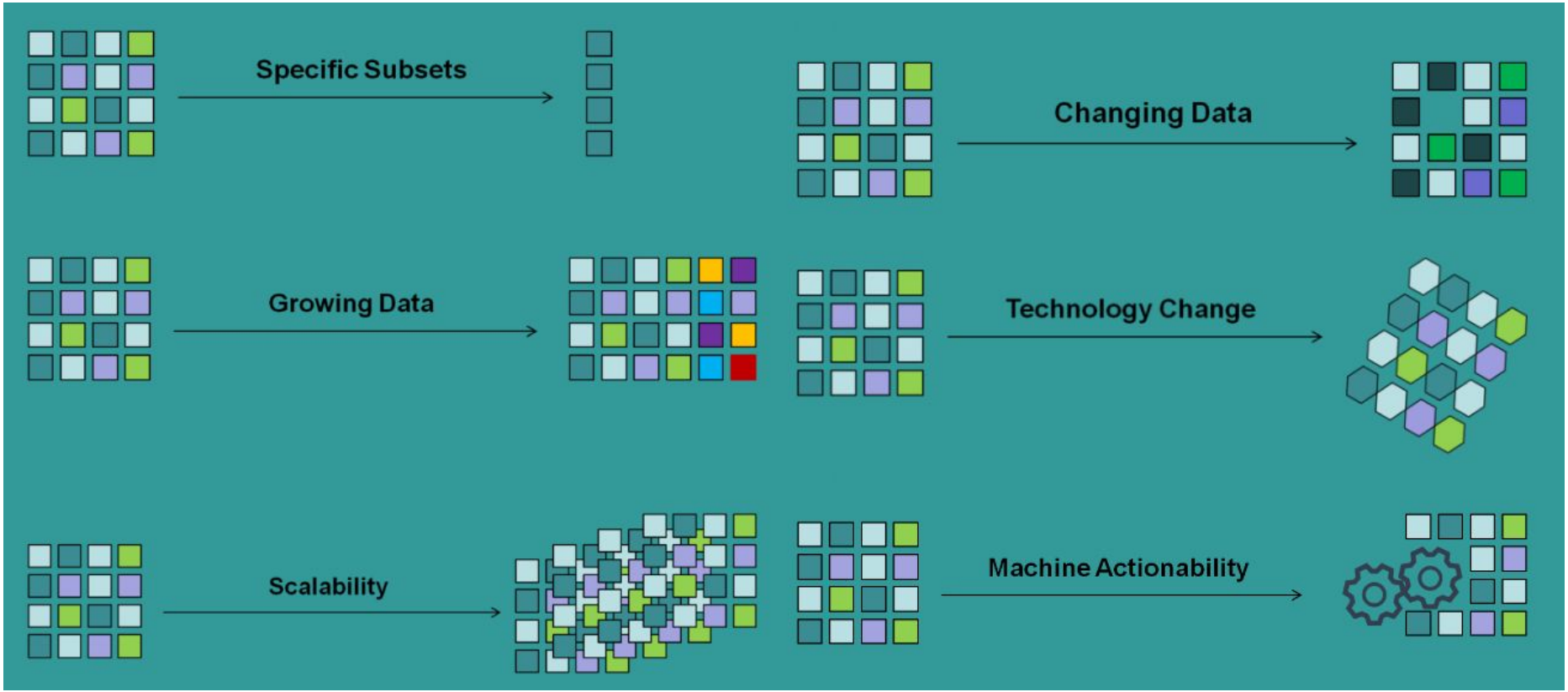
Author(s), Year, Dataset Title, Data Repository or Archive, Version, Global Persistent Identifier

Principle 7: Specificity and verification
(e.g. the specific version used).
Versioning or timeslice information should be supplied with any updated or dynamic dataset.

http://best-practices.dataverse.org/data-citation/

Dynamic data citation

Dynamic data citation



The WG recommends solving this challenge by:

Ensuring that data is stored in a versioned and timestamped manner.
Identifying data sets by storing and assigning persistent identifiers (PIDs) to timestamped queries that can be re-executed against the timestamped data store.

RDA dynamic data citation recommendations

DATA VERSIONING

Keep track of meaningful versions.

TIMESTAMPING

Annotate versions with timestamp.

QUERY STORE

Storage for meaningful queries.

RDA dynamic data citation recommendations

QUERY UNIQUENESS

Only unique queries are kept on record

STABLE SORTING

Defined data sorting property

RESULT SET VERIFICATION

The queries e.g. data retrieval process need to be verified

RDA dynamic data citation recommendations

QUERY TIMESTAMPING

Annotate query with timestamp

QUERY PID

Assign PID to query. PID resolution process = query execution process

STORE QUERY

Query store keeps query metadata

RDA dynamic data citation recommendations

CITATION TEXT

Automatically generate citation text from query store metadata

LANDING PAGE

Human readable landing page with retrieved data and contextual information

MACHINE ACTIONABILITY

Provide some sort of an API

RDA dynamic data citation recommendations

TECHNOLOGY MIGRATION

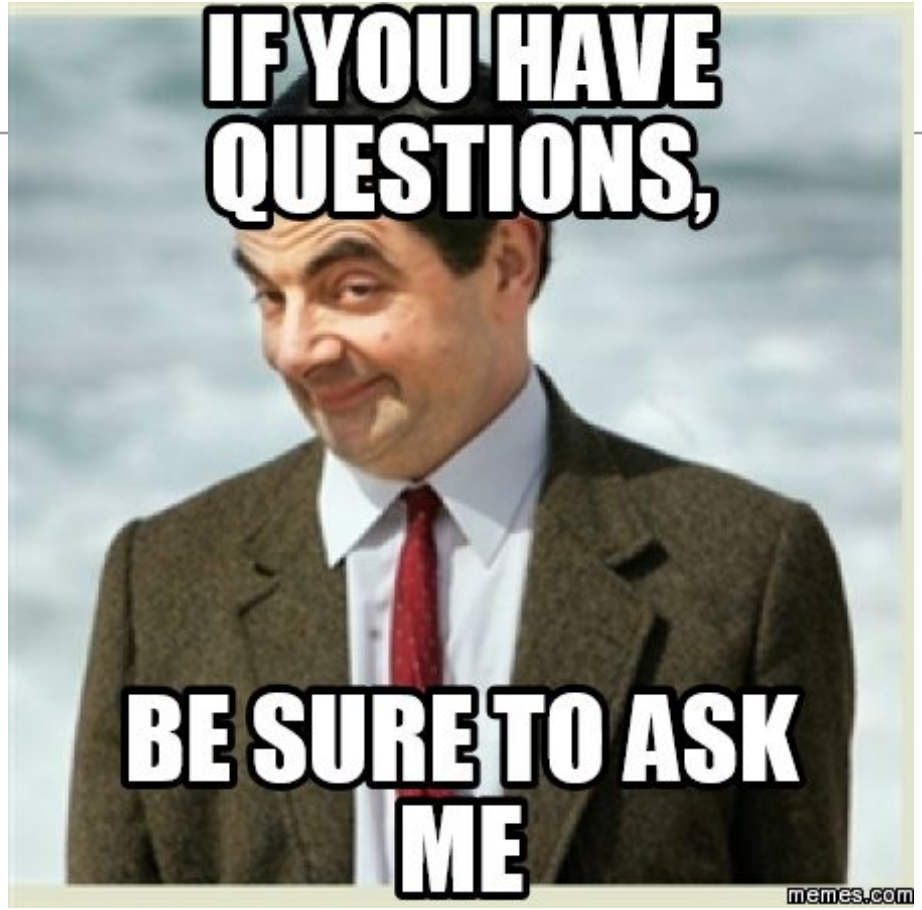
Migratable content

MIGRATION VERIFICATION

… self-explanatory.

Takeaway message

- Build data management requirements based on community needs
- Take advantage of the existing recommendations
- Join the on-going efforts on dynamic data citation, voice your concerns
- Let's make data first-class citizen


IF YOU HAVE QUESTIONS,
BE SURE TO ASK ME
memes.com