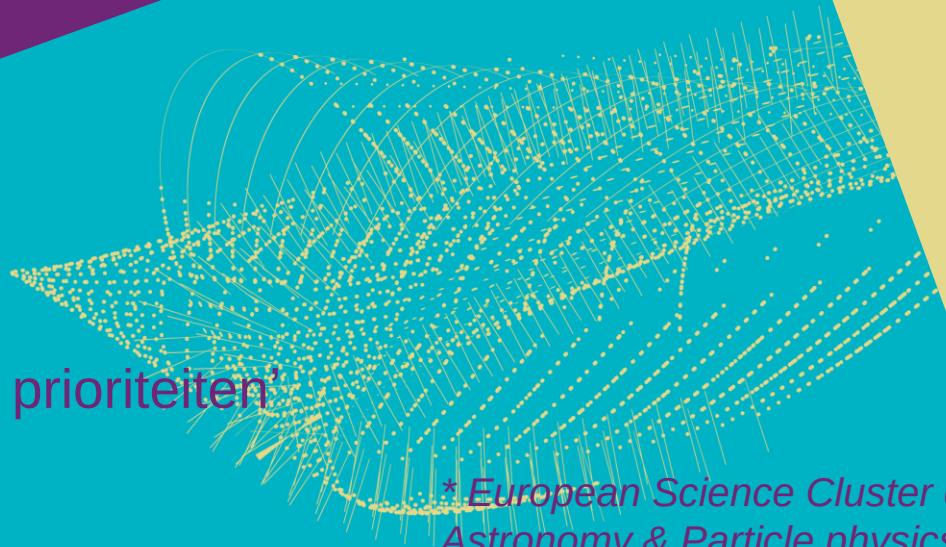
Nikhef



‘FAIR in de fysica: waarde en prioriteiten’
NPOS werkconferentie

Data Management from the ESCAPE* room

** European Science Cluster of Astronomy & Particle physics ESFRIs*

David Groep,
Nikhef PDP programme,
November 2020

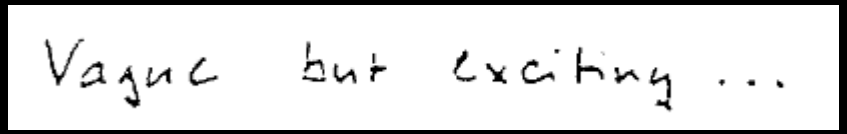
A tradition of open and accessible data

Article II

Buts

1. L'Organisation assure la collaboration entre Etats européens pour les recherches nucléaires de caractère purement scientifique et fondamental, ainsi que pour d'autres recherches en rapport essentiel avec celles-ci. L'Organisation s'abstient de toute activité à fins militaires et les résultats de ses travaux expérimentaux et théoriques sont publiés ou de toute autre façon rendus généralement accessibles.

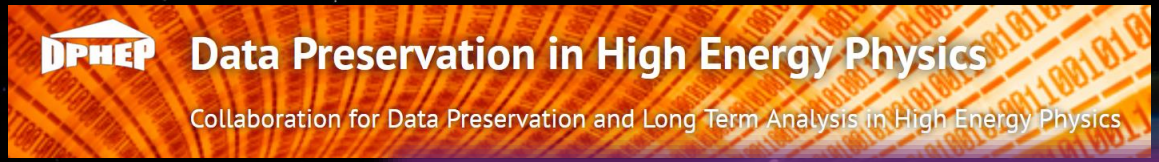
CERN Convention, done in Paris, 1st July 1953

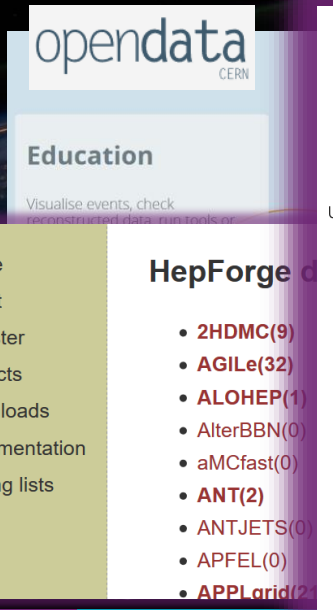


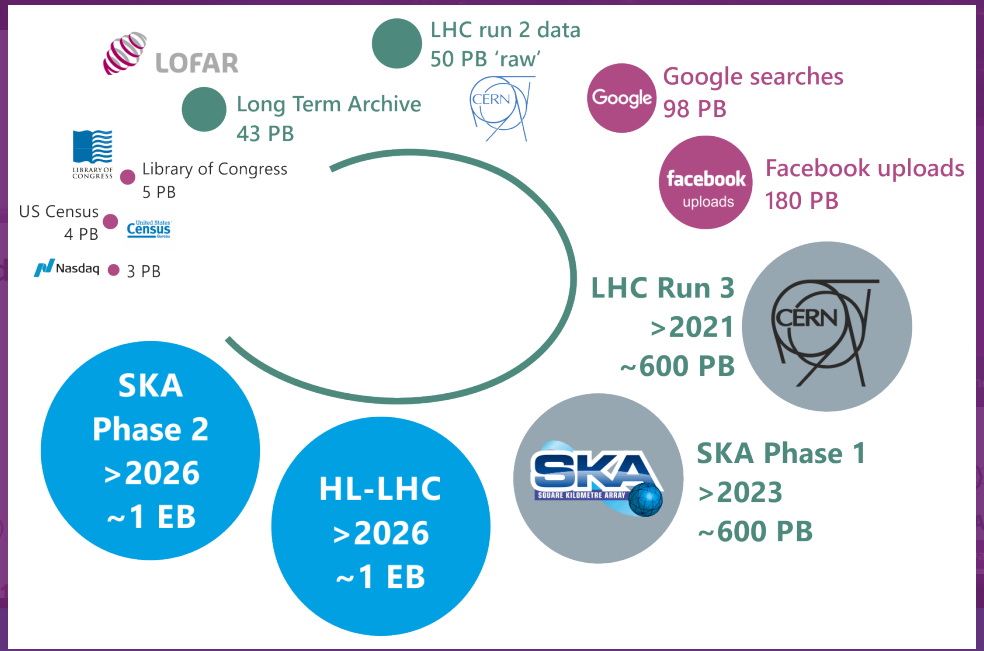
CERN-DD-89-001-OC at cds.cern.ch/record/369245/

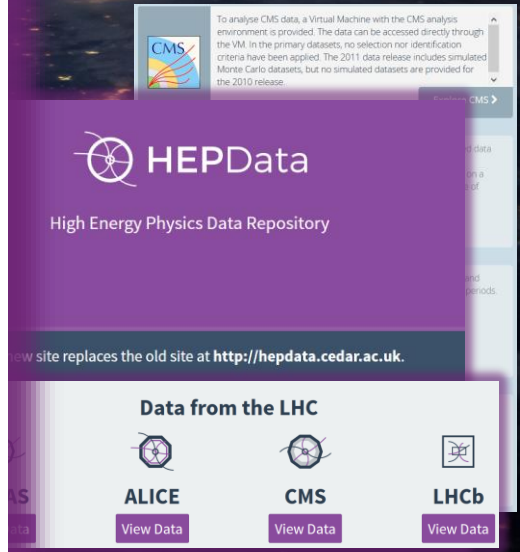
FAIR data needs its domain context

Considering both data as well as its necessary processing environment









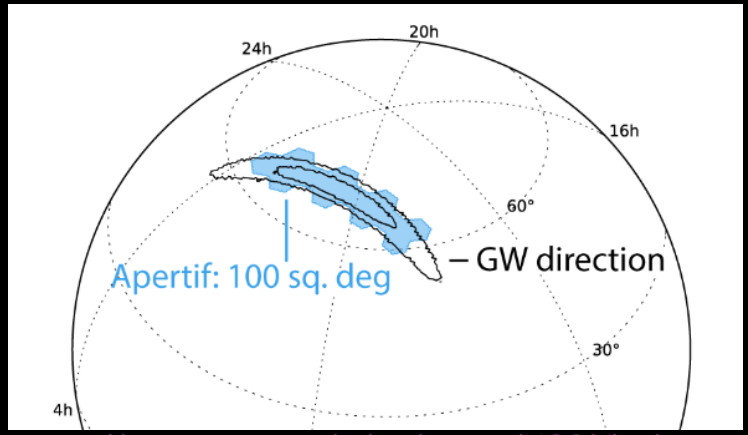
Data from various sources, for public entities: data ca. 2018, indicative, within ~ factor 2
LHC volumes: LCG Resource Scrutiny Group & CERN; 2020
SKA and LOFAR volumes: ASTRON/Michiel van Haarlem, 2020

Nikhef

Value of ‘DPHEP’ and of common data structures

- since much data that’s unique just cannot be reproduced today
- ‘its just essential’
  experiment time: 30+ year data taking
  ‘researcher time’: ~4 years
- meta-data and cataloguing are embedded in analysis frameworks
- for more powerful statistical analyses

multi-messenger
physics and astronomy enabled by common domain formats (Root, IVOA)



multi-messenger: gravitational waves (LIGO/Virgo) and radio-astronomy (APERTIF) combined searches

but it’s only valuable if methodological integrity is maintained! Easier for high-level data products, yet cannot be expressed in ‘generic meta-data’ for much of the ‘raw’ data

image: Jouri van Leeuwen, ARGO: The Apertif Radio-Gravitational wave Observatory, 2018
7th Belgian-Dutch Gravitational Waves Meeting, https://indico.nikhef.nl/event/1099/

Challenges and priorities

Much of the data should *really* be interpreted using domain expertise

- knowledge is much more than can be written in any ‘broad language’
- open-ness does not instil understanding

High-level data products - in publications or as images - are ‘simple’ to re-use and make accessible, but this is quite a restricted view

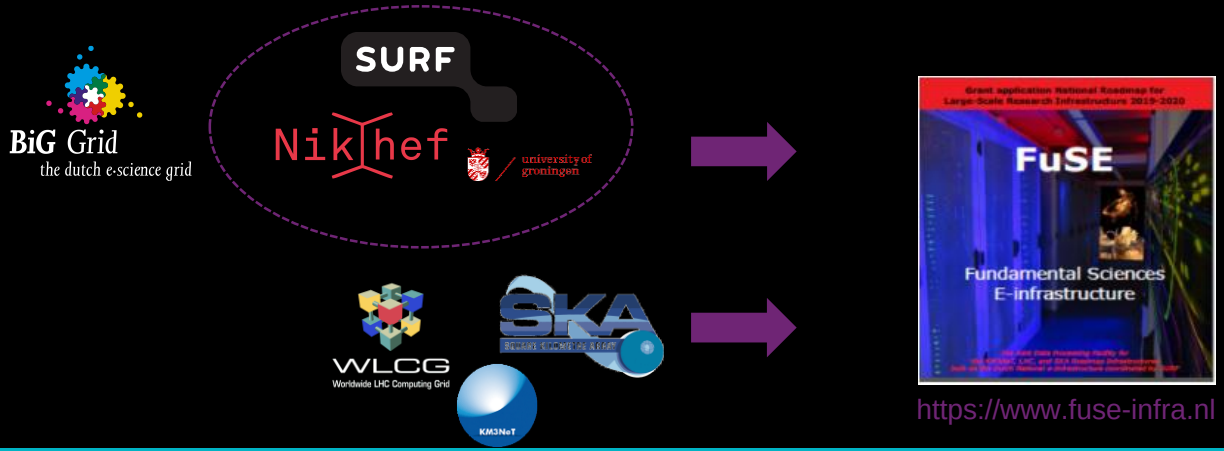
- thematic approaches like FuSE bring actual multi-messenger knowledge – based on domain understanding & distributed participation
- leveraging existing clusters and research infrastructures

FAIR data must come with *means* to re-use data, or it will just be ‘dead data’

Towards Valuable FAIR Data

Considering data management as **regular part of the research workflow**

- data management must be an integral part of the ability to analyse and process
- domain expertise remains essential to ensure scientific integrity of results
- research infrastructure clusters provide solid basis for both *interoperability* and *re-use*
- our existing national e-infrastructure with SURF has this proven ability to integrate



https://www.fuse-infra.nl

Leveraging our existing interactions for RDM …

our thematic
Roadmap & ESFRI Cluster

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

SURF

our local DCC
RDM cluster ‘RDIX’

CWI
Centrum Wiskunde & Informatica

ASTRON
Netherlands Institute for Radio Astronomy

SRON
Netherlands Institute for Space Research

Aanvragende organisatie	CWI uit naam van NWO-I DCC in oprichting
Contactpersoon aanvragende informatie	*< Indiening dient te gebeuren door trekker/kwartiermaker/leider van de lokale DCC (in oprichting) >* Peter Boncz
Email adres contactpersoon	Peter.Boncz@cwi.nl (trekker lokale NWO-I DCC), Carl.Schulz@cwi.nl (contact persoon voor dit project bij CWI)
Telefoonnummer contactpersoon	+31 20 592 4109
Postadres contactpersoon	CWI Science Park 123 1098 XG Amsterdam NETHERLANDS
Naam van het project	**RDIX (Research Data meta IndeX)**

DIFFER

AMOLF

Nikhef

nscr

Royal Netherlands Institute for Sea Research

ARCNL

Data Archiving and Networked Services
DANS

OpenAIRE

EUDAT

egi

GÉANT
Networks • Services • People

Nikhef

David Groep
davidg@nikhef.nl
https://www.nikhef.nl/~davidg/presentations/
https://orcid.org/0000-0003-1026-6606

all logos and trademarks are owned by the respective organisations and projects, and shown here because of joint collaborative participation of these organisations and projects