



《大科学数据快报》期刊介绍

Data Express in Big Sciences

第四届量子计算与机器学习研讨会
报告人：刘北江

CONTENTS

目录

1

数据论文的定义

定义及组成部分

2

发表数据论文的意义

国际同类数据论文期刊及其影响力

3

《大科学数据快报》介绍

期刊介绍、主编团队、栏目设计

4

RDTM介绍

期刊介绍

5

交流与互动

期刊联系方式

什么是数据论文？

定义

数据论文（Data Paper）是一种学术论文形式，主要目的在于**详尽描述特定的数据集**。与传统研究论文关注研究结果与讨论不同，数据论文的核心在于提供数据集的**收集、处理、存储和使用方式**的详细信息。

核心目标

促进数据**共享与可重用性**，增强科学研究的**透明度与可复现性**。在数据密集型科学领域，数据论文已逐渐成为重要的出版物类型。

数据论文的典型组成部分

背景介绍

研究方法

数据集描述

技术验证

数据使用指导

参考文献

■ **重要说明：**数据论文**不应**提出新的科学发现或结论，其重点是对数据集本身的描述。

国际同类数据期刊

以下三本期刊是当前国际数据论文生态中影响力显著、覆盖学科领域互补的代表性数据期刊

scientific data

开放获取期刊，发表来自自然科学、医学、工程和社会科学各领域的研究数据集描述以及有关研究数据共享的文章。

Scientific Data
IMPACT FACTOR
• 2024

6.9



Data in Brief

SCOPE: 跨学科短篇数据描述
PUBLISHER: Elsevier
已发表数千篇数据论文



Earth System Science Data

SCOPE: 地球系统观测数据集的描述、发布与质量评估
PUBLISHER:
Copernicus
Publications

典型数据论文案例-1

以下为一篇发表于 *Scientific Data* 的典型数据论文，描述用于无监督新物理探测的 LHC 物理数据集。

SCIENTIFIC DATA

2022 · Vol. 9

LHC physics dataset for unsupervised New Physics detection at 40 MHz

Govorkova, E., Puljak, E., Aarrestad, T. et al.

CITATION

Sci Data **9**, 118 (2022).

DOI

<https://doi.org/10.1038/s41597-022-01187-8>

典型数据论文案例-1

以下为一篇发表于 *Scientific Data* 的典型数据论文，描述用于无监督新物理探测的 LHC 物理数据集。

scientific data

OPEN

DATA DESCRIPTOR

LHC physics dataset for unsupervised New Physics detection at 40 MHz

Esterina Goukova^{1,2}, Ema Puljak^{1,3}, Theo Aarnestedt^{1,4}, Maurizio Pierini^{1,5}, Kinga Anna Wozniak^{1,6} & Jennifer Ngadiuba^{1,2,4}

In the particle detectors at the Large Hadron Collider, hundreds of millions of proton-proton collisions are produced every second. Few could store the whole data stream produced in these collisions, but of terabytes of data would be written to disk every second. The general purpose experiments ATLAS and CMS reduce this overwhelming data volume to a substantial level by deciding in real time whether each collision event should be kept for further analysis or be discarded. We introduce a dataset of proton collision events that emulate a typical data stream collected by such a real-time processing system, pre-filtered by requiring the presence of at least one electron or muon. This dataset could be used to develop novel event selection strategies and assess their sensitivity to new phenomena. In particular, we intend to stimulate economically based efforts towards the design of novel algorithms for performing unsupervised new physics detection, customised to fit the bandwidth, latency and computational resource constraints of the real-time event selection system of a typical particle detector.

Background & Summary

The proton bunches of the CERN Large Hadron Collider (LHC) cross paths 3000 million times per second in each of the experimental halls, possibly generating a collision every 25 ns. In such an event multiple proton pairs may collide, possibly producing thousands of particles to be detected by the detectors located in each experimental hall. LHC detectors record the first of emerging particles in the form of electronic signals for the so-called general purpose detectors (ATLAS¹ and CMS²), which globally account for 300 MB of information. The resulting data throughput of 30 TB/s is too large to be recorded. This is why these two detectors process data in real time to select a small fraction of them (about 1000/s), compatible with downstream computing resources. This strategy is effective in preserving the data needed to discover the Higgs boson^{3,4}.

This filtering system, usually referred to as the trigger, consists of two-stage selection, as illustrated in Fig. 1. With the more upgrade of these data acquisition systems, the two-stage trigger, LHCs and ALICE⁵, have been equipped with a novel data processing capability that avoids the need to select events. Instead, their data acquisition follows a real-time data processing approach^{6,7}: all the events enter a computer farm, where they are processed in real time and stored in a reduced-size data format, consisting of high-level information sufficient for data analysis. For this reason, their use goes beyond the scope of this paper, and it is not considered further.

The first trigger stage, the Level-1 trigger (L1T), runs a set of algorithms deployed as logic circuits on custom electronic boards equipped with fast, programmable gate arrays (FPGAs). This stage rejects more than 90% of the events, reducing the incoming data stream to 100 events/s. Due to the short time interval between two events (25 ns) and limited buffer capabilities, the entire pipeline of L1T algorithms has to be executed within 30 ns (s). The second stage, called the high-level trigger (HLT), consists of a computer farm processing events on commercial CPUs, running hundreds of complex selection algorithms within 3000 ms. The trigger selection algorithms are designed to guarantee a high acceptance rate for physics processes under study.

When designing searches for unsupervised physics phenomena, one typically considers specific theories motivated scenarios. This approach has been proven successful when dealing with strongly theoretically motivated scenarios, e.g. the Higgs boson discovery^{8,9}. However, this approach might become a limiting factor in the absence of strong theoretical guidance. The ATLAS and CMS trigger systems could be interesting alternatives, limiting the possibility of new physics discoveries.

www.nature.com/scientificdata/

Fig. 1 The real-time data processing flow of the ATLAS and CMS experiments. CMS-HM collisions are produced every second and processed by the hardware-based event selection system, consisting of algorithms implemented in high-speed on-chip electronics boards. Of these events, 1000 events/s are accepted and passed to the second selection stage, the HLT, which selects about 1000 events/s for offline physics studies.

Therefore, recent work has investigated unsupervised and semi-supervised approaches to data selection and analysis, focusing on anomaly detection (AD) strategies based on deep learning (DL) algorithms. These studies aim to learn a metric directly from the LHC data, with the capability of ranking the events by typicality. One could select a sample enriched with anomalies, possibly due to new physics processes, by selecting the events in the tail of the distribution of such a metric. Extensive reviews of several proposed methods are given in Refs.¹⁰ and references therein.

This effort, mostly targeting offline data analysis, should be paired with a similar effort to integrate AD algorithms as the trigger system of the LHC experiments, possibly already at the L1T. Then, it would be possible to present an unbiased dataset to the AD algorithm, before discarding any event¹¹. One could then collect new event topologies in a special data stream, either to select new physics or CMS with the existing trigger^{12,13} during the first year of data taking at the LHC. By studying these events, one could simulate new theoretical models of new physics phenomena, that could be tested in future data-taking campaigns.

While the focus so far has been on the HLT, this strategy would be more efficient if deployed in the L1T before any selection has been introduced. Each L1T event has to be processed within a few microseconds and therefore the trigger decision is taken by algorithms hard-coded in the hardware electronics as logic circuits. Algorithm complexity, and possibly accuracy, could be enhanced by adopting DL algorithms that fit L1T logic. To do this, the L1T trigger^{14,15} was introduced as a test to simulate a given L1T model that an electronic circuit. Commercial AI-on-FPGA software libraries are initially designed to exploit the FPGAs as an acceleration device, to support a commercial CPU in its network interface computation. This implies the great need of specific computing for components for many tasks. On one hand, this allows one to execute large networks (as in complex neural applications). On the other hand, the small size of the calculations and the consequent data transfer implies a latency time that exceeds by an order of magnitude what can be tolerated in a L1T system at the LHC. By integrating the DL network in the L1T, performance would be improved, avoiding the data transfer to sequential steps. The consequent latency reduction comes at a cost in terms of resource utilisation. This is why L1T logic, as well as real networks with CPUs, is not easy.

Once this, all the ingredients are there to make it possible for the community to design an optional AD strategy for the L1T, in order to stimulate the development of a community working collaboratively to what was done with the LHC's OpenPDA¹⁶ and the Dark Machine data challenge¹⁷; we present a new dataset designed to the kind of data stream, that one could handle at the last stage of the ATLAS and CMS L1T system.

Methods

Physics content of the dataset. The proton-proton collisions taking place at the LHC can lead to the production and observation of many different phenomena predicted by the Standard Model (SM) of particle physics^{18,19}. A broad summary of the SM particle content can be found in Refs.^{20,21}. The role of each of these processes can be calculated within the SM framework of Feynman diagrams and their relative cross-sections, performed by the experiments (summary of the CMS cross-section measurements is available here). In this paper, we focus on events containing electron (e) and muon (μ) light particles that, together with tau (τ) and heavy neutrino partners, form the three lepton families. In principle, we could have considered a dataset with no filter. While the world would contain a more realistic representation of an unbiased L1T dataset, generating such a dataset requires computing resources beyond our capabilities. Instead, we decided to use the lepton filter and make the dataset simulation realistic.

Within the limited size of a typical LHC detector, electron and muon are stable particles, i.e. they do not decay in the detector as they are directly observed while passing the detector material. On the contrary, τ leptons are much heavier, and hence much more unstable, than electron and muon. They quickly decay into other stable particles. In the absence of these decays, τ and μ are produced at the LHC, and an observable source of high-energy leptons is the production of W and Z bosons²², which are among the heaviest SM particles with a small fraction of cases. These particles are leptons. W and Z bosons are mainly produced directly in proton collisions. A sizable fraction of W bosons originates from the decay of top quarks (t) and anti-quarks ($\bar{t}$). The top quark, being heavy and highly unstable, quickly decays into a W boson and a bottom quark, giving rise to signatures with only collimated jets or hadrons called jets or with one e , μ , τ , or a neutrino and multiple jets.

Sample name	Number of events	Type
$W \rightarrow e \nu$	100000	Signal
$W \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow e e$	100000	Signal
$Z \rightarrow \mu \mu$	100000	Signal
$Z \rightarrow \tau \tau$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	100000	Signal
$Z \rightarrow \mu \nu$	100000	Signal
$Z \rightarrow \tau \nu$	100000	Signal
$Z \rightarrow \nu \bar{\nu}$	100000	Signal
$Z \rightarrow e \mu$	100000	Signal
$Z \rightarrow e \tau$	100000	Signal
$Z \rightarrow \mu \tau$	100000	Signal
$Z \rightarrow e \nu$	1	

典型数据论文案例-2

以下为一篇发表于 *Scientific Data* 的典型数据论文，描述在直线加速器相干光源（LCLS）上对雾化Coliphage病毒颗粒进行单颗粒成像（SPI）实验的数据集

SCIENTIFIC DATA

2020 · Vol. 7

Diffraction data from aerosolized Coliphage PR772 virus particles imaged with the Linac Coherent Light Source

Haoyuan Li et al.

CITATION

Sci Data **7**, 404 (2020).

DOI

<https://doi.org/10.1038/s41597-020-00745-2>

发表数据论文的重要性

数据论文已经成为科研出版生态的重要组成部分

01

专业期刊已成体系

Nature 旗下设有专门出版数据论文的期刊 **Scientific Data**，2014 年创刊。

国内：中科院计算机网络信息中心创办了 **Data Express**。

Scientific Data
IMPACT FACTOR · **6.9**
2024

02

科研范式正在转变

AI 在科学中的应用日益突出，未来的论文不仅为人类阅读，也将为AI解析。

结构化、机器可读的数据论文，是 **AI for Science** 时代科研产出的关键载体。

03

对数据工作者贡献的认可

发布数据论文鼓励数据工作者专注于数据的管理与利用，无需承担额外的科研任务。

让数据本身成为可被引用、可被认证的**独立学术成果**。

Data Express in Big Sciences

《大科学数据快报》

• 主办单位

中国科学院高能物理研究所

• 合办与出版单位

科学出版社

• 预计2026年底上线

• 期刊宗旨

着眼于大科学装置数据资源的构建与共享，规范数据描述论文的发表，重点聚焦具有科学价值、开放共享价值及实际应用潜力的数据集和研究成果。

期刊将探讨大科学装置数据获取与应用、人工智能的融合、技术创新、质量控制、管理模式及共享环境建设，促进学术交流与成果转化，推动开放科学的发展。



DE (Data Express) 期刊集群

“1旗舰 + 19领域” 梯度化布局 · 总计 20 本期刊

DATA express (旗舰)



本刊 **DATA express in Big Sciences** 作为大科学领域数据期刊，深度融入 DE 集群生态，共享投审稿平台、数据标准与品牌影响力。

主编团队介绍

主编团队均来自中国科学院高能物理研究所，在大科学装置数据领域具有深厚积累。

主 编

EDITOR-IN-CHIEF



陈 刚

中国科学院高能物理研究所研究员

国家高能物理科学数据中心 · 主任

ASSOCIATE EDITORS · 副主编



齐法制

中国科学院高能物理研究所 研究员
计算中心 主任



程耀东

中国科学院高能物理研究所 研究员
计算中心 副主任



刘北江

中国科学院高能物理研究所
研究员



郝 权

中国科学院高能物理研究所
研究员

创刊时间线

期刊将于2027年正式出版，诚邀来稿！



2026.09

期刊官网及投审稿系统上线



2026.12

累计上线5篇文章



2027.05

累计上线不少于15篇文章



现在投稿，就是最好的时机

- ✓ 共建新刊，成为前15篇文章的作者

期刊栏目设计

期刊围绕大科学装置数据全生命周期，共设 **5 个核心栏目**，覆盖从描述、采集到应用的完整链条。

01

Scientific Datasets and Databases 科学数据集及数据库

发表大科学领域数据论文（Data Paper/Data Descriptor），涵盖大科学装置产出的数据集、数据库、数据产品及其质量评估与验证。

考虑到数据（集）描述类论文数量可能不足的现实情况，此类论文占比初期设60%以上

02

Data Infrastructure and Computing for Big Science 大科学数据基础设施与计算

数据采集、高速传输、分布式存储、边缘/高性能计算

03

Big Data and AI in Big Science 大科学与人工智能

机器学习、深度学习、科学大模型在大科学数据分析中的应用

04

Data Mining, Management and Standards 数据挖掘、管理与标准

数据治理、元数据标准、FAIR原则、跨装置数据互操作

05

Cybersecurity and Data Governance 数据安全与治理

科学数据安全、隐私保护、数据共享政策、开放科学与伦理

期刊概况

RDTM (Radiation Detection Technology and Methods) 基本信息



主管单位: 中国科学院
主办单位: 高能物理研究所; 中国核学会核电子学与核探测技术分会



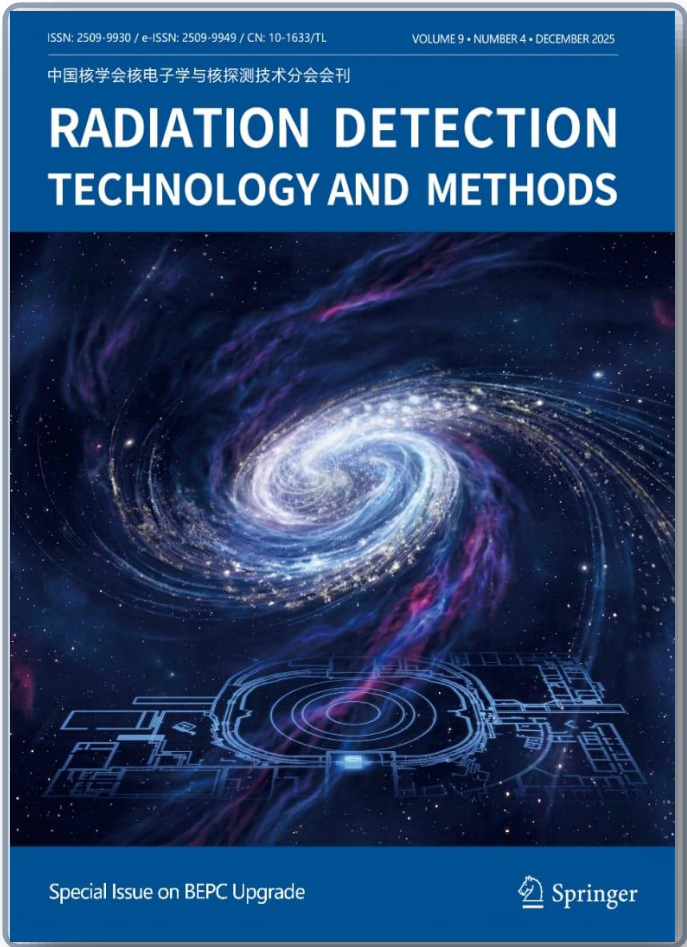
创刊年份: 2017年 | **出版周期:** 双月刊 | **语种:** 英文刊



收录情况: CSCD 和 ESCI 收录



- **完全OA:** 2027年将转为完全OA(Open Access)期刊
- **版面费:** 保持不变, 350元/页
- **编委特权:** 期刊编委会成员作为通讯作者投稿免收版面费
- **系统升级:** 2026年9月起全面启用 ScholarOne 投稿系统



稿件类型:

研究论文

综述

数据论文

交流与互动

欢迎关注DEBS，加入我们的学术社区



DEBS作者群



RDTM作者/通讯群



青年编委/审稿人招募

诚邀符合条件的优秀学生与青年学者加入我们，共同推动学科发展