Resource-aware Research on Universe and Matter: Call-to-Action in Digital Transformation

Ben Bruers[1], Marilyn Cruces[2], Markus Demleitner[3], Guenter Duckeck[4], Michael Düren[5], Niclas Eich[6], Torsten Enßlin[7], Johannes Erdmann[6], Martin Erdmann[6*], Peter Fackeldey[6], Christian Felder[8], Benjamin Fischer[6], Stefan Fröse[9], Stefan Funk[10], Martin Gasthuber[1], Andrew Grimshaw[11], Daniela Hadasch[9,12], Moritz Hannemann[8], Alexander Kappes[2], Raphael Kleinemühl[13], Oleksiy M. Kozlov[14], Thomas Kuhr[4], Michael Lupberger[15], Simon Neuhaus[13], Pardis Niknejadi[1], Judith Reindl[16], Daniel Schindler[17], Astrid Schneidewind[8], Frank Schreiber[18], Markus Schumacher[19], Kilian Schwarz[1], Achim Streit[20], R. Florian von Cube[20], Rodney Walker[4], Cyrus Walther[9], Sebastian Wozniewski[17], Kai Zhou[21]

[1]Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany.
[2]Max Planck Institute for Radio Astronomy, Bonn, Germany.
[3]Universität Heidelberg, Heidelberg, Germany.
[4]Ludwig-Maximilians-Universität München, München, Germany.
[5]Justus Liebig University Gießen, Gießen, Germany.
[6*]RWTH Aachen University, Aachen, Germany.
[7]Max Planck Institute for Astrophysics, Garching, Germany.
[8]Forschungszentrum Jülich GmbH, Garching, Germany.
[9]TU Dortmund University, Dortmund, Germany.
[10]Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
[11]University of Virginia, Charlottesville, Virginia, USA.
[12]University of Tokyo, Kashiwa, Japan.
[13]Bergische Universität Wuppertal, Wuppertal, Germany.
[14]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany.
[15]University of Bonn, Bonn, Germany.
[16]Universität der Bundeswehr München, Neubiberg, Germany.
[17]Georg-August-Universität Göttingen, Göttingen, Germany.
[18]University of Tübingen, Tübingen, Germany.
[19]Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.
[20]Karlsruhe Institute of Technology, Karlsruhe, Germany.
[21]Frankfurt Institute for Advanced Studies, Frankfurt, Germany.

*Corresponding author(s). E-mail(s): erdmann@physik.rwth-aachen.de;

Abstract

Given the urgency to reduce fossil fuel energy production to make climate tipping points less likely, we call for resource-aware knowledge gain in the research areas on Universe and Matter with emphasis on the digital transformation. A portfolio of measures is described in detail and then summarized according to the timescales required for their implementation. The measures will both contribute to sustainable research and accelerate scientific progress through increased awareness of resource usage. This work is based on a three-days workshop on sustainability in digital transformation held in May 2023.

Keywords: Sustainability, Climate Change, Scientific Computing

1 The challenge

Climate change is real. What numbers tell us since many years becomes more and more tangible for everybody in events such as extreme weather conditions, floods, or wild fires that come closer and happen more frequently. This makes it harder and harder to ignore the consequences that our actions as human species have on the conditions for life on our planet.

The discussion often focuses on an increase of average temperature by 1.5 or 2 degrees. But this number does not convey the severeness the climate change will have. Each human induced change of the climate increases *the probability of reaching a tipping point* that leads to an irreversible and uncontrolled development. A so far underestimated risk due to global-warming-induced weather extremes like droughts, heat waves and torrential rains may be the reduction of global food production and the resulting expansion of regions with hunger, migration, and conflicts [1–3].

We can decide to either accept those risks or mitigate them. If we choose the latter this means we have to *drastically reduce greenhouse gas emissions*. Here it is important to note that the *integrated emission* matters, such that the differential emission per year needs to disappear. Therefore time is a critical factor.

To meet the goal of the Paris agreement [4] and reduce the risk of reaching tipping points, the greenhouse gas emissions must be reduced by 50% within 7 years as illustrated for the worldwide energy production in Fig. 1. Here, the consumption share of data centers is estimated to be about 1% [5]. The curve labelled *Fossil* is very relevant for ErUM-Data [1], the digitization of the research on Universe and Matter in Germany [6, 7], because a large fraction of the CO_2e [2] *footprint of computing in ErUM sciences* comes from the energy consumption for processing and storing data, which is continously growing in many research fields.

If we accept the responsibility arising from our actions and take the Paris agreement seriously, *we as a community must develop a plan* to follow the curve in Fig. 1 for the energy consumption of our research. The general transition from fossil to regenerative energies will help and on a time scale of a decade or two there may be an abundant supply of renewable and cheap energy. However, additional savings will be required on short and medium term which can be reached by technological innovations and changes of attitude and behavior.

With these transformations we can on the one hand *contribute to research for sustainability*. On the other hand we must advance *our own research methods to make them sustainable.*

The German government aims at achieving greenhouse gas neutrality by 2045 [11]. While this is an ambitious goal, it still poses challenges in terms of remaining emissions until that year. Figure 2 shows the time evolution of the energy mix focusing on electricity production in Germany, as well as a possible future development based on the targeted expansion of renewables [11–13] which takes into account both the end of nuclear energy in 2023 and the phase-out of coal by 2038.

The challenge of sustaining both, our own research and the conditions for life on our planet,

[1] ErUM: Research on Universe and Matter

[2] Several greenhouse gases CO_2, CH_4, N_2O, HFC, PFC, SF_6, NF_3 contribute to climate change. Their different effects are normalized to the effect of CO_2, the total effect is reported as CO_2-equivalent 'CO_2e' [8].

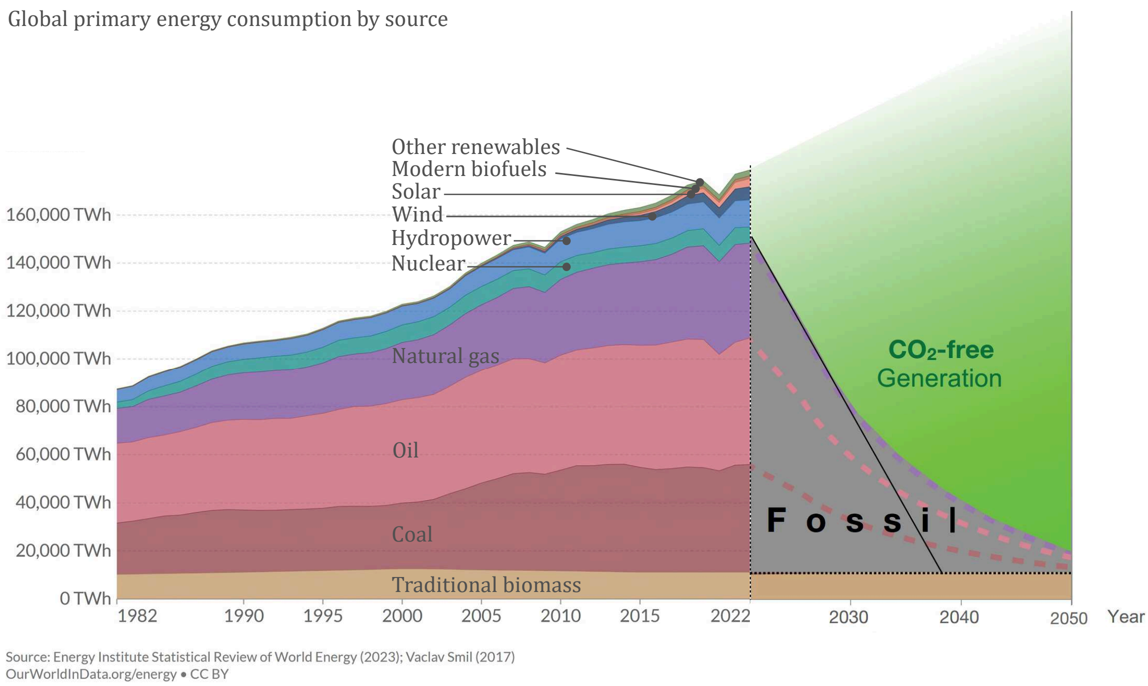

Fig. 1 Graph based on the evolution of the worldwide energy production by sources [9] and projection with required savings according to the Paris agreement by [10]

is huge, and particularly difficult for the ErUM communities. An extensive compilation of generally relevant aspects for ErUM-related sciences can be found in reference [14]. In this work we will focus on aspects of digital transformation. Remarkable advances in compute power, data storage, and algorithm development have catapulted the capabilities of modeling, analysis, and simulations to unprecedented levels. In addition, the 'fourth paradigm' [15] with its data-centric methodology combined with new artificial intelligence (AI) methods enables researchers to tackle and decipher major scientific challenges. On top of that, there is the increasing sensitivity of experimental instruments and a corresponding increase in data rates and volumes. In light of all this, there are still ideas on how to address the seemingly contradictory overall challenge of preserving both nature and research-driven knowledge gain.

In this document *we present practical measures* in the areas of data and software management, algorithms and artificial intelligence, as well as computing infrastructure discussed at a workshop of ErUM-Data scientists [16]. Twelve questions were formulated in advance which gave a structure to the work and discussions (see Appendix A). The measures can be the basis for the development of a realistic plan for the required reduction of greenhouse gas emissions in the next years while having sufficient computing resources for our research. There are three main approaches.

First, *energy can be saved* by avoiding or reducing unnecessary computations, e.g. by reusing existing results. This requires in particular FAIR [3] data [17] and FAIR research software [18] as well as an increased *awareness to properly balancing knowledge gain with resource usage.*

Second, we can *increase the efficiency of the calculations* that are required for our research without tapping into the classical rebound trap of using efficiency gains to do more calculations. The measures range from classical or automated code optimizations to new technologies such as new computing architectures and artificial intelligence.

Third, we can *reduce the* CO_2e *emissions* caused by our computing. This is mainly in the hand of the computing providers and includes measures such as locating sites close to renewable energy suppliers and reusing the produced heat.

[3] FAIR: findable, accessible, interoperable, reusable.

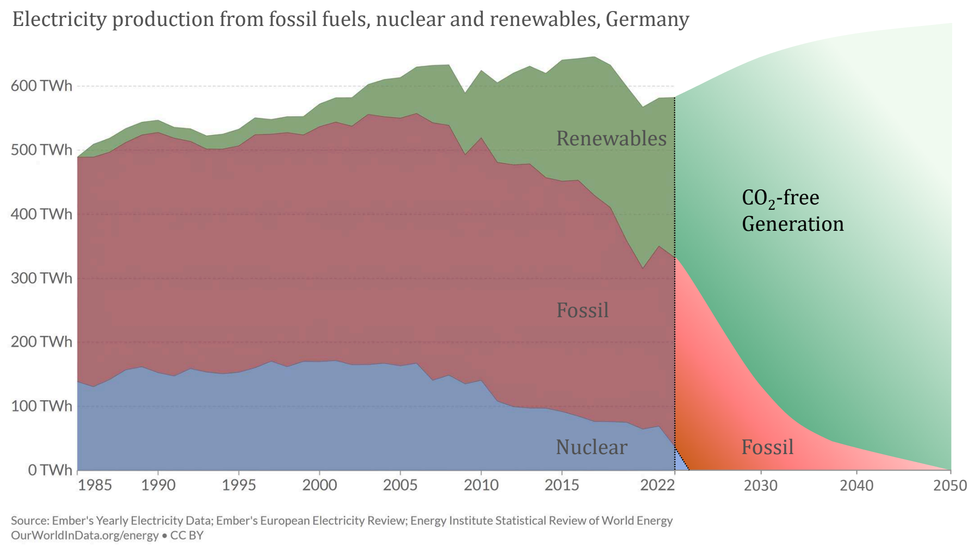

Fig. 2 Shares of the yearly electricity production in Germany [9] together with a possible extension scenario approximately following expectations in [11–13]

To be able to implement the practical measures discussed in this document certain *competences* need to be acquired and the *right support and incentives* need to be provided by funding agencies, research centers, and universities. The acceptance and execution of the technical and non-technical measures in everyday scientific research will demand a difficult, but unavoidable transition period. *Awareness at all levels is a critical prerequisite for a successful transition.*

The portfolio of measures concerns six areas which are detailed in the following sections: smart data transformation, practices in software development and data analysis, algorithms including artificial intelligence, computing with renewable energies, education and training in sustainable research, as well as funding and institutional support.

2 Smart data transformation

Data transformation constitutes an essential aspect of empirical sciences, and advancements in technology are enhancing both its capabilities and complexities, alongside resource utilization. We refer to transformed data designed and susceptible to further exploitation as smart data. As experiments achieve higher resolutions, larger sample sizes or sample quantities, the volume of raw data necessitating processing and analysis increases.

An example are high-energy particle collider experiments, where permanent storage of all raw data is not feasible since several decades. Automated algorithms ('trigger') are designed to identify the raw data with most scientific value and only retain this data for storage. Nevertheless, current technical and economical boundaries are exploited to maximize the scientific revenue. For example in case of experiments at the Large Hadron Collider at CERN in Geneva, this results in the order of 100 PB of data being stored per year.

In the field of astrophysics all data taken by telescopes were archived traditionally and evaluated repeatedly. As computing and algorithms evolved, sensitivity to previously unnoticed features improved such that breakthroughs have appeared from reprocessing of archival data (e.g. [19]). Such research with archived data may be particularly resource-efficient. In the future, storing all data will no longer be possible as newer and more sensitive state-of-the-art facilities such as the Square Kilometre Array Observatory will produce data rates in the range of Terabits per second that can no longer be stored, but must instead be analyzed immediately and deleted after processing.

In photon and neutron sciences, fast automated sample exchange and qualitatively improved detectors cause similar effects. As pioneering examples, 'stream based models' are explored by not storing raw data at all, and performing all processing online [20].

Dealing with such large amounts of data naturally requires significant compute resources with non-negligible footprints in energy consumption (see Section 5). Hence, smart data transformation becomes increasingly important, especially in view of further envisioned detector upgrades and new science facilities.

For achieving our sustainability goals, teams with in-depth knowledge of the relevant experiments are needed to develop comprehensive data management strategies. Their efforts are in direct relation to the goals of the NFDI consortia [21–23]. The teams need to provide a data life cycle with different phases of data use, from instant analyses after data recording to completely new analyses at a much later time (long-term archiving). In the early stages, storage capacities could be substantial to maintain some level of data redundancy for backup and convenient access. In later phases, the volume of data retained could be gradually reduced over specified time intervals.

For the content assessment of the data, far-reaching expertise is needed to define the exact objectives of the data storage. The spectrum ranges from timely thorough data analyses, to reproduction of analyses, later cross-verification or subsequent use in completely new contexts. For the latter, potential future use cases should be anticipated as best as possible.

It is thus mandatory to define which data should be stored in which formats. The criteria under the overarching FAIR data principle [17] for data preservation should also be followed in Universe and Matter research.

For data to be reused, their meaning must be self-explanatory to scientists and convey information about the conditions under which the measurement data were obtained. Such metadata on location, time, device, operating conditions, instrument calibrations, etc. play a critical role in understanding and processing the data in subsequent phases.

Another aspect of discoverability concerns the descriptive language of the data. Within individual research communities, there are often specific terminologies and vocabularies whose linguistic nuances evolve over time and thus can present significant obstacles to the use of archived data. In some research areas, ontologies have been developed to describe preserved data. This approach creates a standardized lexicon and thus simplifies the matching between the description provided with the data and a search query.

Finally, workflows are important for smart data transformation, referring to software and algorithms as well as processes involving humans. In several areas of the ErUM community, it has become common practice to take snapshots of processed data at intermediate stages in workflows throughout the data processing chain and store them temporarily. This approach allows for later extensions and repetition of subsequent steps without having to reinitiate the entire data processing sequence. This efficient strategy saves time and computational resources, albeit at the cost of temporarily increased storage requirements.

With the goal of achieving scientific results as soon as possible, there is an immediate motivation for the efficient choice of such snapshots, and they should improve energy efficiency. Orchestrations of workflows are also being tried in large collaborations where the same data on different aspects are analyzed by individual research teams. However, it requires a strong commitment and dedication to collaborative efforts to prioritize resource efficiency through joint action, especially in highly competitive contexts.

As a summary, we note these actions for the topic of smart data transformation:

- Make data FAIR to promote reuse, which can be particularly resource efficient.
- Reduce and compress data, having the anticipated scientific value of the retained information and the resource requirements in mind.
- Optimize the choice of storing intermediate results against re-calculating them.
- Optimize job orchestration and scheduling in workflows.

3 Software engineering and data analysis

The relevance of human resources invested in software development is increasing in view of the ever-growing volumes of data and the corresponding computing power required. Once written, code is applied to increasingly more data. Therefore, prioritizing software optimization is more than justified.

A common practice in dealing with research data is to code an exploratory analysis that investigates the potential of the data with respect to a scientific question. In case of success, a re-engineering of the used software into a professional structure is mandatory to consolidate and enable possibilities of enhancing the code base. Furthermore, especially in the case of complex data analyses with many intermediate steps, it is essential to introduce a workflow management system to ensure both the reproducibility of the analysis and to avoid unnecessary recalculations [24]. Often, this revising step is skipped and it remains with what is popularly known as 'spaghetti code' together with the researcher who has memorized the order of the calculations for himself, inevitably leading to 'abandonware'. Experience shows, however, that a solid structure of the research software and the use of workflow managers accelerates iterative review processes and thus the publication of the results.

Accordingly, adherence to good software development practices has the potential to save human and energy resources through excellent code quality. The FAIR research software seal [18] now exists in this area as well. Community-maintained, easily accessible open source code is generally more efficient than local stand-alone implementations. Effective coding practices of jointly developed software include tracking systems, version control, rigorous testing, and benchmarking (see, e.g. energy per luminosity in $\mathrm{GWh/fb^{-1}}$ which is particularly suited for particle physics [25]). Furthermore, modularization and reuse of code, possibilities for parallelization of computations (vectorization), comprehensive documentation, and continuous integration are important.

Generally, energy efficiency should play a central role in software development, along with broadly-accepted software-quality metrics. First of all, unnecessary computations should be avoided. This is helped not only by workflow managers, but users can also execute initially small-scale validations and monitor execution for early detection of software or parameter problems. Code developers can implement sanity checks on configuration problems and input data. In this way, wasting energy consumption by large, unsuccessful runs can be avoided.

Great help in conserving resources comes from using established runtime optimization techniques through automated parallelization, adaptation to new CPU architectures (i.e. ARM, RISC-V, etc.) and Co-Processor architectures, vectorization, memory layout optimization, and the use of special compiler flags. All such measures generally lead to more energy-efficient code and are also being further developed in expert groups in the Universe and Matter research area [26].

In order to intensify the necessary awareness for resource consumption, tools for monitoring of resource consumption, trainings on the conscious use of resources (see Sec. 6), practical optimizations of code as well as their computational processes need to be developed and made available.

Ultimately, given the pivotal role of high-quality scientific software in the research landscape, its development and maintenance should receive increased and appropriate valuation. This could encompass recognition through publications and citations, funding allocation, enhanced access to infrastructure resources, and the establishment of dedicated career paths.

Summarizing practices of software engineering, we note these actions:

- Make software FAIR and reliable by following good software development practices and ensuring sustainable support.
- Design software for optimized energy consumption.
- Use workflow management to make processing FAIR.
- Continue research on potential of new technologies for efficient use of resources.

4 Algorithms and artificial intelligence

The energy efficiency of algorithms is crucial for the reduction of computing resources and obviously closely linked to adherence to the above-mentioned good software development practices. Often, there are already optimized algorithms in open-source standard libraries or domain-specific libraries that can be utilized. A recent example is experiments with Big Data in the form of millions of events, in which the formerly sequential event loop has been replaced by parallel processing of large event chunks [27–29]. Beyond this, new challenges may soon arise from dynamic sustainable energy supply of computing centers, to which algorithms may also have to be optimally adjusted.

The rapid progress in the field of artificial intelligence (AI) is opening up completely new possibilities. Machine learning is already accelerating ErUM research in theoretical predictions, in simulations, and in the analysis of experimental data. Moreover, innovative data analyses are becoming possible via these data-driven modeling techniques, in which the physics potential of an experiment is exploited to a much greater extent than originally expected (e.g., [30, 31]).

From many more examples, we mention here also closed-loop experiments in X-ray and neutron sciences with their direct feedback after data taking from quasi-instantaneous data analysis and control of the experimental parameters in film growth [32], and more efficient searches in a dedicated phase space for dispersion relations of phonons or magnons [33]. Finally, we expect that in the near future more and more tasks will be performed by scientists using AI tools.

We argue that the full potential of machine learning algorithms should be prioritized, respecting the goal of reducing energy consumption in ErUM research. The focus should therefore be on developing and deploying models to perform tasks that have the potential of great benefit to scientific progress, or are particularly expensive in terms of computational resources. Given the limited time window for reducing the CO_2e footprint of ErUM computing, developments in areas with high potential for reducing energy consumption must be advanced early.

However, since training machine learning models also leaves a CO_2e footprint, the resources consumed in training should be evaluated and documented for transparency. This should include all training performed, including those to optimize the models. Energy consumption during the inference period is likely to be low in comparison, but should also be documented and compared to the consumption of established alternative solutions.

In light of recent breakthroughs in generative models, this is one area where the deployment of machine-learning solutions has large potential for fields of ErUM that rely on large samples of simulated data. One such example is the detector simulation at the LHC experiments, which at the ATLAS experiment for example consumed 38% of the total CPU resources in 2018 [34] and for which a deep generative model is being deployed that is $\mathcal{O}(500)$ faster than the simulation with Geant4 [35].

Another promising area is the adaptation of large pre-trained models that are only refined for the specific ErUM applications. Thus, refining such pre-trained models would not only require less computational resources in training, but also enable smaller ErUM training datasets. In general, the reusability of previously trained machine learning models has the potential for more efficient training in ErUM research.

Pre-trained models include recent AI developments in terms of large language models, which are available to the general public as new and powerful AI tools. These tools have great potential to increase work efficiency in ErUM research, for example, by helping with documentation tasks. This includes computer code documentation, which is time-consuming and therefore often neglected. Good code documentation has the potential to increase the reusability of code in general and of efficient code in particular. However, large language models can also be used to directly suggest more energy-efficient algorithms. These new capabilities are directly related to the identified needs in Section 3.

It can hardly be overemphasized that scientists have a key role in the choice of algorithms. Their decisions on the use of self-coded algorithms, library algorithms, or artificial intelligence algorithms are crucial to the runtime of the jobs and their resource consumption. Considerable expertise is required especially for successful deployment of new AI algorithms, i.e., deployment that is energy efficient and leads to scientific progress.

Therefore, the strategic use of AI tools in the scientific workflow needs prioritization between the use of human and computer resources and in the expected value of knowledge gain. As a consequence, it needs to be decided where AI should be used sensibly and where not. For example, for some problems, comparable performance could be achieved by human reasoning instead of resource-intensive machine learning training. When AI-generated code or machine learning is used for ErUM research, it always requires extremely careful validation by scientists. Thus, it is clear that human interaction will remain a pillar of scientific progress.

We see the following measures to be central in the area of algorithms and artificial intelligence:

- Continue research on potential of AI or other new technologies.
- Include particularly promising applications of generative and pre-trained models.
- Expand detailed monitoring and documentation of energy consumption and CO_2e footprint in training and inference.
- Use already optimized algorithms in open-source standard libraries or domain-specific libraries.

5 Computing and infrastructures

In this section, we focus on sustainability related to computing hardware and its operation. Discussions of practical measures involve adjusting computing in space and time to the availability of renewable energy, reusing the generated heat, and extending the operating lifetime of hardware. Implementing such measures requires a comprehensive information flow between the stakeholders involved.

Greenhouse gas emissions are widely classified in so-called scopes (e.g. [36]). There are no directly produced emissions from computing for research on the Universe and Matter (scope 1). However, indirect emissions are produced by operating the data centers with electricity that is not produced from renewable sources (scope 2). Finally, indirect emissions arise from the entire value chain, starting with the production of buildings and computer systems and later with their disposal (scope 3).

Renewable energy (scope 2):

First, with regard to scope 2, an essential component of sustainable data processing is a detailed overview of the power consumption of the various systems and services, as well as a detailed accounting of past or planned activities. Data centers generally have detailed measurements and records of the power consumption of their various systems, although this information is generally only available upon individual request and is not yet directly accessible online in a comparable format.

For the annual electricity consumption of large German data centers, we refer to the compilation results in Tab. 1. For comparison, 1 GWh roughly equals the electricity demand of 1,000 single-households per year in Germany.

Computing center	Electricity/GWh	Ref.
MPCDF Garching (2022)	43	[37]
LRZ Garching (2021)	33	[38]
HLRS Stuttgart (2021)	32	[39]
JSC Julich (2012/13)	34	[40]

Table 1 Annual electricity consumption of large computing centers

An example of the approximate breakdown of energy consumption among the major components of the CERN data center is 55% data processing, 21% disk storage, 2% tape storage, 5% network, and 17% services [41]. Details of these numbers can be found in the Appendix B. During the 13-year operation of the Large Hadron Collider, power consumption has been fairly constant, although computing capacities have increased by a factor of 6 from 2012 to 2023. The high service share comes about because of the special role of the CERN computing center in the interconnection of about 170 computing centers in the Worldwide LHC Computing Grid ‘WLCG’ [41], where CERN operates a large part of the central services.

A next big step toward sustainable computing would be to place data centers near sustainable energy sources. Renewable energy supplies generally consist of solar panels and wind farms, but also biogas power plants.

Practical approaches with data centers directly at the producers already exist in Texas in the USA, whose renewable energy capacity will reach approx. 70 GW by the end of 2023 [42]. Transporting electricity is a major problem due to lengthy permitting processes for transmission lines. So the idea of moving electricity consumers to the point of generation was born. While this is difficult for most industries, scientific computing is an ideal candidate for this relocation. On the hardware side, data centers only need power and fiber for the network. The user side of scientists typically computes in batch mode, usually considering temporary interruptions due to insufficient sun or wind and thus moderate delays in the computations acceptable.

For Germany, a comparable data center scenario could be built near the North Sea, where most of the wind is available and most offshore wind farms have either been built or are in the planning stages. Substantial computing infrastructure could be built as green-field sites, as is already explored commercially [43]. It is advisable to keep sufficient storage data capacity along with the computing resources to reduce network latency issues when data throughput is large.

Since in many communities experiments are carried out worldwide and as international collaborations, data of interest to scientists working in Germany could be relocated to such data centers in the north. The cost and effort of laying the required fiber optic cables for the networks likely do not dominate. Transformed data with lower volumes could optionally go to the universities for further processing.

Overall, it makes sense to establish a joint ErUM science cloud initiative timely, starting with moderate equipment and scaling up once significant funding has been successfully acquired. Even in the initial pioneering phase, there are many aspects to develop and explore, as we will discuss below.

When setting up a data center near the power producers, the following criteria should be taken into account. Since the energy comes from renewable sources, the performance of a processor per invested units of electricity (flops per watt) plays a subordinate role. Thus, initially hardware could be operated that does not belong to the latest generation and can therefore be obtained at relatively low cost.

The key challenge for data centers is the dynamics in supply from renewable energies [42, 44, 45]. Data centers must be able to dynamically ramp up or down computing resources or shift workflows as needed. Importantly, centers must be equipped with valid forecasts of weather conditions and expected energy supply (see below paragraph on information flow and middleware).

There are two relatively straightforward ways to reduce power consumption in a data center: First, one can reduce the CPU clock rate. In principle, this can be done immediately; jobs are only slowed down, but continue to run. Studies show that power consumption of compute nodes can be reduced by 50%, with a corresponding slowdown in processing [46].

Second, one can hibernate or power down entire nodes. The order of nodes could follow a prioritization list based on the operating age of the hardware and thus their efficiency. Shutting down CPUs works within seconds, while restarting can take several minutes depending on the memory requirements and the speed of the IO system [42].

However, it must be ensured that running jobs are stopped properly without losing the results achieved up to that point. Ideally, this can be achieved by so-called checkpointing, i.e. the entire program state is stored on disk and can be resumed later. In practice, this feature is challenging for data processing jobs with many open connections to external services. If the processing consists of repetitive, independent steps, as is the case with event processing in particle physics, an alternative is event-level check-pointing, i.e., after each processing step, the output is stored in its entirety. In case of an interruption, processing can be resumed after the last processed event, so that only little CPU time is lost.

However, a minimum energy supply to the data center must be guaranteed at all times. The servers and network switches of the data center should run continuously. A data center with CPU and storage as in use for WLCG requires about 25% of power for hard-disk drive storage (HDD). To avoid damage, these servers should not be shut down frequently, resulting in a continuous power requirement. Therefore, energy storage options should also be planned for the data center from the beginning. Depending on the environment, these could be accomplished for example by batteries, water storage, energy to gas plants, flywheels, and bidirectionally charging electric vehicles.

Efforts to optimally use the supplied energy for computing centers are rated by Power Usage Effectiveness (PUE) which describes the total amount of energy used by a center compared to the energy delivered to the computing equipment. For research infrastructures such as CERN in Geneva, the Prevessin Site reaches PUE=1.1 [46], the National Renewable Energy Laboratory in

Boulder (USA) reported annualized PUE=1.036 [47], and the German HPC center LRZ measured a PUE of 1.06 for the SuperMUC-NG system [48]. Beyond energy-supply discussions, modern computer chips have very high heat output per unit area, exceeding that of a conventional induction stovetop [49]. In terms of sustainability, it is imperative to use the dissipated heat, for example, to provide heating and hot water for the nearby buildings. Also, residential units are provided with data centers where best PUE values were achieved for new residential buildings (PUE=1.024), but also far-reaching improvements were obtained in conversions of e.g. high-rise buildings from PUE ~ 2 to PUE=1.27 (both in [49]).

Hardware lifetime (scope 3):

A subject of its own are the sustainability issues for the above-mentioned scope 3, which takes into account the CO_2e footprints during manufacturing and disposal. It is not easy to get exact data for individual hardware components. There are studies that put the manufacturing CO_2e footprint at $20-30\%$ of the total CO_2e footprint [50, 51]. Obtaining corresponding numbers for e-waste turns out to be challenging.

At this scale, hardware lifetime is a relevant issue. Complementary variables are power consumption and lifetime. Data center equipment runs 24×7 continuously for approximately 5 years, containing the aforementioned approximately 25% of fixed carbon. Extended life in certain mission areas - i.e., larger science centers with dedicated and experienced staff could explore more versatile life extension options that result in 7–10 - or even more - years of operation which is worth when using renewable energies.

However, there is also a finger pointing at typical personal devices such as desktops, laptops, smartphones, etc. Because of the much shorter, integral operating times and sophisticated energy efficiency, the CO_2e footprint of manufacturing here goes in at about 75% bound carbon. Accordingly, life extension for devices for individual use needs to be seriously considered.

Recent announcements indicate that industry is now putting a stronger focus on energy efficiency in data center devices as well [52]. The corresponding component-based replacement and operation of data center devices with renewable energy are clearly going in the right direction to address the challenges of sustainability scopes 2 and 3.

Information flow and middleware:

Successful sustainable computing in the area of research on Universe and Matter requires comprehensive information flows between energy providers, data centers and users. Furthermore, conceptually new middleware tools are needed for dynamic operation of the data centers. These issues are addressed here.

First, monitoring of resource consumption is central to further advancing energy efficiency. What is the consumption per job, per user, per publication, or the entire data center? Details such as the resource profile within a computing job are also required. With such tools, cost-benefit analysis and optimization efforts are made possible in the first place.

Today, only a few researchers, groups and departments receive information on the resource usage of their computations. One example is the National Analysis Facility (NAF) at DESY, which recently set up a monitoring service to determine the energy consumption of a job in the central processing system and record this information in the log file. The service also provides a CO_2e estimate based on the current energy mix at the data center site. Such systems need to be transferred to provide every scientist with comprehensive energy and CO_2e reporting.

The development of easy-to-use tools for explicit measurement and profiling of energy usage and CO_2e footprint for developers and end users is non-trivial. First, the CO_2e footprint depends on the energy resource used, which is known to the energy provider only. The relation between energy consumption and program runtime is not necessarily strictly linear, since modern architectures have dynamic power and frequency scaling. In addition, a poorly optimized GPU implementation may run faster but consume more energy than a CPU version due to differences in thermal design performance.

A bidirectional flow of information between users and the data center is also necessary. Users should be able to estimate the requirements of their job types, at least approximately. Detailed accounting information about the energy consumption and CO_2e of each data processing job, production task or data transfer would be important and useful information. Similarly, users should be able to provide information about what specifications their algorithms tolerate and the dynamics with which their algorithms can respond to power shortage situations. Furthermore, it should be possible to define whether certain delays in batch jobs or even reduced job numbers are compatible with their quality-of-service requirements. An example are High-Throughput Computing applications of WLCG data processing, which are – within boundaries – less time-critical. In other words, for a major campaign that lasts many days, delays of a few hours are not critical.

Conversely, users need information from the data center on what the supply situation is – whether, for example, computing resources will be slowed down or even temporarily shut down – and, finally, what the forecasts are for job completion. In general, one can expect downtime to be acceptable as long as it is predictable on a daily basis. Users would not submit jobs that cannot be completed in the foreseeable future.

Second, we can expect that in the medium term, consumer energy prices will be time-variable, and will depend on the availability of renewable energies and on the overall energy demand. In order for data center operators to perform energy-dependent work planning, it is important to obtain information from energy providers about current pricing and availability of renewable electrical energy, e.g., whether there is an oversupply or undersupply. In addition, forecasts for supply, including weather information and price stability, are needed. Conversely, the energy provider needs the requirements of the data center for its own work planning in a timely manner.

To enable overall asynchronous communication flows, one possible approach to the required information exchange between all stakeholders could be to set up a centralized and scalable or distributed monitoring infrastructure where data is collected in real time and where everyone can query the required information using standard protocols and APIs as well as standardized communication content.

Beyond the aforementioned information flows, data centers require conceptually new middleware tools that enable efficient energy-aware scheduling [53] based on the overall situation of user demand and energy availability and implement dynamic CPU/GPU power modulation and load limitation on existing systems. Good results can be achieved with applications that adapt to changes in resource availability (so-called malleable jobs [54]). Also users need new middleware tools to hand in their job requirements and receive reports.

For longer-term periods of low sun or low wind, energy storage technologies must also be developed and deployed. A scenario must be developed on how to react in such situations. Historical weather data could be used to define storage capacities: how often and for how long do such situations occur.

A possible contingency procedure could be to first use up the energy storage, then freeze jobs or only run high priority jobs for a certain period of time, and finally use non-renewable energy. The guiding principle should be to keep the CO_2e footprint of data center operations as minimal as possible [45].

From the topic of computing with renewable energies we summarize the following demands:

- Monitor and report energy consumption at job level including resource profiling within the job.
- Monitor and report energy consumption at site and project level, provide information of the individual use per scientist/project/publication.
- Extend monitoring of resources beyond CO_2e (water, material etc.).
- Consider carbon footprint for all planned investments and project plans.
- Adjust computing in space and time to the availability of renewable energy, e.g. computing centers close to off-shore wind parks with a job scheduling using only or mainly the surplus of renewable energy available at a given time.
- Develop software and middleware that can respond dynamically to the availability of energy.
- Optimize power usage effectiveness.
- Re-use produced heat.
- Adjust hardware lifetime considering emissions due to procurement and operation.

6 Developing a culture for sustainable science in the ErUM communities

There is no disputing that remarkable advances in computing power, data storage, and algorithm development have greatly accelerated the capabilities of modeling, analysis, and simulation to unprecedented levels. This transformation has opened new frontiers of scientific understanding and progress that were impossible at the turn of the century. Especially data-centric methodology has revolutionized numerous scientific fields, enabling researchers to tackle and decipher grand challenges by studying systems at multiple levels with unprecedented precision (see Section 4).

However, as a result of this development, there has been a significant increase in energy consumption, which needs to be accounted for in relation to scientific progress, so there is an urgent need for comprehensive, strategically designed education that covers all of these areas.

Since the early 2000s, academic discussions and surveys [55] have already underscored the importance of sustainability-focused education and the vital skills needed to equip the next generation of scientists with sustainability awareness and the aptitude necessary for innovative research and development. In addition, UNESCO describe the World Programme of Action on Education for Sustainable Development (2015-2019, 2020-2030) [56] and offer detailed recommendations and strategies on how to effectively integrate AI into education systems in its comprehensive guide [57].

New university courses and workshops designed to empower students and researchers with the knowledge and skills to conduct sustainable digital research can help create a responsible approach to knowledge gain in conjunction with resource usage. Together with the portfolio of measures described in this publication, these educational tracks have the potential to promote informed decision making about the use of computing resources and to support more sustainable practices in scientific research.

Especially when using data-driven methods, a commitment to sustainability is essential. To balance and monitor the seemingly inexhaustible potential of AI, researchers and students must have deep domain-specific knowledge. A solid foundation in mathematics and computation paves the way for the deliberate integration of efficient computation and algorithms, streamlined models, and conscious decision-making throughout the research process. This is where it makes sense to inject one's own intelligence first, rather than reflexively increasing computational power. A further temptation notes that efficiency improvements trigger a rebound effect that increases the demand for computing resources. Thus, regular reviews and controls may be needed to mitigate this phenomenon. This and other incentives under development (e.g. [58]) can certainly be beneficial. Even more robust measures such as CO_2e-based fairshares, CO_2e budget allocations, and specific reduction commitments ('carrot and stick' strategy) are conceivable.

Thus, for early-career researchers, knowledge transfer, mentorship, resource allocation, and networking are central to cultivating their skills, deep understanding and effective use of AI tools. For experienced scientists in leading positions, it is an important task to assess the CO_2e footprint of their current research, to develop, implement, and monitor plans to reduce CO_2e emissions, and to consider CO_2e emissions in future investments and project plans. With these approaches and balance between innovation, awareness, and responsibility, we can continue pushing the boundaries of human understanding while preserving the resources we depend on for a sustainable future.

Overall, there are many technical aspects, guidelines, developments, and scientific measures to improve the current scientific work in our research field with respect to the urgency of climate change. However, the success depends on the implementation in the daily work of individual scientists and therefore requires an urgent change in awareness and responsibility of every scientist and science manager.

For the development of educational concepts and to raise awareness, we summarize here key concepts that should serve to refine priorities for performing research sustainably:

1. *Balance the knowledge gain of any (computational) work against the resources it needs.*
2. *Reliability is sustainability, avoiding unnecessary repetitions.*

3. *Thought-through, well-documented workflows and standardization guarantee reliable results and their re-use.*

A further periodically appearing challenge of working sustainably in ErUM sciences results from project-oriented organization of the funding schemes. The sustainable use of invested resources requires improvements in the long-term organization of knowledge transfer and the exchange of results and methods. For this continuous support, sufficient resources in the areas of software, algorithms, and computation need to be allocated.

For developing a culture for sustainable science in the ErUM communities we summarize the following measures:

- Raise awareness of the climate challenge at all levels.
- Disseminate knowledge of measures to address the challenge.
- Train scientists in good practices.
- Strive to become a role model at all levels and help to establish sustainability in everyday life.
- Enhance awareness of the trade-off between research benefit and climate impact.
- Perform first, rough energy audit and develop an initial CO_2e reduction plan.
- Regularly review and update the CO_2e reduction plan.
- Consider carbon footprint for all investments and project plans.
- Include the resources needed for continuous IT support into project planning.

7 Funding and institutional support

From the perspectives of the federal and state ministries as well as the top management of universities and research centers, there is a great deal of attention being paid to the issue of sustainability. This involves the development of strategic concepts, reporting, measures and funding.

At the level of the German Federal Government, the United Nations *Agenda 2030* [59] has been transferred into a national strategy named *German Sustainable Development Strategy 2021* [60]. A comprehensive general work on the *Digital Strategy Germany* was published subsequently [61, 62] which concerns science and research among many other aspects and is closely linked to the sustainability goals. In more specific contexts, the Federal Ministry of Education and Research (BMBF) developed sustainability measures, e.g. which for science is called *Research for Sustainability (FONA)* [63].

Within the framework programme ErUM [6], the BMBF has strengthened the aspect of sustainability in research on Universe and Matter in their calls for proposals. For example, technological and methodological development work that contributes to climate- and resource-friendly operation of large-scale facilities and experiments can be funded alongside projects (e.g. [64]).

In view of all the initiatives and activities to date, however, major and urgent tasks remain for all stakeholders to be mastered together. The many measures described in this work require appropriate support mechanisms and structural alterations, some of which themselves must be sustainable. For example, computing facilities localized at power generators need infrastructure and personnel. Furthermore, an expert group to advise and support scientists for Digital Transformation measures in ErUM could be implemented, e.g. in a new phase of the ErUM-Data-Hub [65]. In any case, a sustainable impact can definitely be achieved by creating appropriate long-term positions in research data, software, and computing.

Efforts to lower bureaucratic and legal hurdles for the implementation of sustainable dynamic power generation and transfer are most welcome. They are needed along with new communication channels between power companies, data centers and scientists as consumers to enable practical and pragmatic solutions for the sustainable use and operation of computing infrastructures.

8 Conclusions

Based on a three-day workshop on sustainability in the digital transformation, interested colleagues from the ErUM-Data community have developed a portfolio of measures with which our research area will massively reduce the emission of climate gases in the sense of the Paris climate agreement of 2015 [4] and turn to the use of renewable energies. In doing so, we are planning for a transitional period until electricity will once again be available in abundance through the establishment of a sufficient number of renewable energy sources.

Item	Call-to-action
	Immediately or on **short time scale** with little effort these measures can be implemented:
S1	Raise awareness of the climate challenge at all levels.
S2	Disseminate knowledge of measures to address the challenge.
S3	Monitor and report energy consumption at job level.
S4	Consider carbon footprint for all investments and project plans.
S5	Enhance awareness of the trade-off between research benefit and climate impact.
	On a **medium time scale** of a few years the following measures can be realized:
M1	Make data FAIR to promote reuse.
M2	Reduce and compress data having the anticipated scientific value of the retained information and the resource requirements in mind.
M3	Optimize the choice of storing intermediate results against re-calculating them.
M4	Optimize job orchestration and scheduling in workflows.
M5	Use workflow management to make processing FAIR.
M6	Make software FAIR and reliable by following good software development practices and ensuring sustainable support.
M7	Design software for optimized energy consumption and provide tools to measure it.
M8	Continue research on potential of AI or other new technologies for efficient use of resources, but balance gain of research action against resource consumption of these developments.
M9	Monitor and report energy consumption at site and project level, provide information of the individual use per scientist/project/publication.
M10	Extend monitoring of resources beyond CO_2e (water, material etc.).
M11	Train scientists in good practices.
M12	Regularly review and update the CO_2e reduction plan.
M13	Strive to become a role model at all levels and help to establish sustainability in everyday life.
	A **longer term** coordinated planning is required for the following measures:
L1	Adjust computing in space and time to the availability of renewable energy, e.g. computing centers close to off-shore wind parks with a job scheduling using only or mainly the surplus of renewable energy available at a given time.
L2	Develop software and middleware that can respond dynamically to the availability of renewable energy.
L3	Optimize power usage effectiveness.
L4	Re-use of produced heat.
L5	Adjust hardware lifetime considering emissions due to procurement and operation.
L6	Include the resources needed for continuous IT support into project planning.

Table 2 Call-to-action in digital transformation: portfolio of measures to be taken, ordered in terms of effort and time they take.

In light of the dramatic increase in data rates from new or upgraded instruments and the enormous potential for new knowledge gains from developments in data-driven, AI-assisted methodologies, the awareness of each of our scientists is critical during this transition period. Already by applying the portfolio of targeted sustainability measures described in this publication and consciously balancing prospective knowledge gain and resource usage, we expect to achieve efficiency gains that will ultimately accelerate our fields of research.

In the portfolio, we see six areas: 1) considering sustainability as an important factor in the planning of measurement data transformation, 2) good practices in software development and data analysis, 3) use and development of renowned and efficient algorithms including AI, 4) data center locations at sustainable energy providers along with lifetime extension of hardware, 5) education and training in a responsible approach of balancing knowledge gain and resource usage, and 6) targeted efforts of all stakeholders including funding and institutional support.

Our ErUM community can respond now and develop a realistic plan for the required reduction of CO_2e emissions. We have the tools at hand or can develop them. In Table 2, the portfolio of measures has been ordered according to the timescales required for their implementation. Various measures can be launched by us immediately or realized within a medium time period. The third category requires our strategic preparations, which involve coordinated efforts with additional stakeholders.

Overall, the commitment and joint effort of all stakeholders – funding agencies, institutional bodies, computing centers, science managers, faculty, scientists, students – are needed to master the challenges of sustainability in digital transformation in Universe and Matter research.

Supplementary information. Data Availability Statement: No explicit data were used in this study.

Acknowledgments. This work is supported by the Ministry of Innovation, Science, and Research of the State of North Rhine-Westphalia, and by the Federal Ministry of Education and Research (BMBF) in Germany through the ErUM-Data-Hub project 05D21PA1. We acknowledge the useful support of the DeepL translator.

Appendix A Twelve Guiding Questions

The workshop on Sustainability in the Digital Transformation of Basic Research on the Universe and Matter was structured around twelve prepared questions, listed below [16].

Hardware & Research Data

1. Footprint: Constructing a comprehensive picture of the footprint of all ErUM-Data related activities. Where does quantitative knowledge exist, where is it lacking? What resource needs do you see, what opportunities for savings? What innovations are needed to keep sustainable use of resources in balance with demands? To what extent does continuing education play a role? How can feedback reduce a footprint through machine learning methods?
2. (Dynamic) Energy Supply: Where to locate & operate computing systems incl. storage? How could a dynamic energy supply look like, which largely covers the needs of ErUM-Data related activities with renewable energies? What information flows would be required for this? What mechanisms and what dynamics are required on a supra-regional basis to create compensation possibilities for windless/sunless periods?
3. Hardware Lifetime: How could prolonged / optimized usage of hardware resources in view of technology evolution be modeled beyond their usual lifetimes? What short- and medium-term monitoring would be required to signal indispensable replacements on the one hand, and to execute computing jobs matching their algorithmic requirements on prolonged or current hardware on the other?
4. Hardware & Algorithms: Which adaptive measures for hardware and algorithms could have a decisive impact on ErUM-Data? Which types of hardware (including e.g., GPU, TPU, FPGA, neuromorphic computing) could be considered and which automated mechanisms exist for adapting algorithms to non-specific or dedicated hardware?
5. Smart Data: Deciding when and how to discard information without losing scientific value, based on learning from nature and experiment. What mechanisms for transforming data to smart data can be envisioned, and how can evaluation and control of information gain or loss be accomplished? How can archiving and retrieving data be managed?
6. Cultural Change: What could a comprehensive educational area for rethinking, among other things, the use of computer hardware, actually required information (smart data), preparation of data packages (event loops versus event chunks), etc. look like? How can we change to a culture of data reuse? Assessment of ethical implications and risk assessment.

Algorithms & Mindset

7. Autonomization: We witness the transformation from the era of automation to an era of autonomization (e.g., unsupervised learning). Where will ErUM-Data benefit from autonomization, which innovations are necessary and how can the reliability of the autonomously obtained results be ensured?

8. Inquiries & Dynamics: How can input questions be posed to generate the best possible output from the machines? What relevance will dynamic learning algorithms and machines have for the field of ErUM-Data?
9. Algorithmics & Software: Our thinking in algorithms and software has a direct impact on resource requirements. What can sustainable algorithm & software engineering and an associated educational program in algorithm & software development look like to get ErUM-Data to the forefront of developers?
10. Machine Models: Pre-trained and generative models have a high potential for energy savings in both their creation and usage of machine learning. What innovations are needed to achieve a reliable routine operation?
11. Injected Intelligence: How can reasoning by the physicist, mathematician, or any other kind of intelligence speed up the processes of learning or make them more energy efficient? What measures can we apply to avoid constantly reinventing the wheel? What can knowledge discovery of work already performed look like?
12. Workflow & Stakeholders: How can well-defined, reproducible workflows with high user dynamics (data analyses) be captured that remain functional in the long term? How can an overall picture be created with all stakeholders working together on a large-scale project for the benefit of sustainability across their departmental boundaries?

Appendix B Energy consumption of data center components

For data centers as used in the Worldwide LHC Computing Grid WLCG, typical numbers for power consumption of modern computing hardware in 2023 are:

- **CPU** Modern CPUs have power consumption in the region 3–8 W per core, e.g. dual AMD EPYC 7513 node with 480 W for 64 cores [66] and about 1700 HS23 (HS23-table) or 280 W / kHS23 [4]. Extrapolated to the whole available WLCG computing capacity (used by the four LHC experiments) of about **14 MHS23** this would correspond to **3.8 MW** (in reality consumption is presumably larger since average CPU hardware is older).
- **Disk storage** Modern storage servers have a power consumption around 1–2 W per TB, e.g. HP Raid-6 server with 14 × 16 TB HDD and usable capacity of 192 TB consumes 240 W (= 1.2 W/TB). Extrapolated again to full WLCG disk capacity of **870 PB** this would correspond to about **1.1 MW**.
- **Tape storage** Estimates for power consumption for tape storage are typically factor 10 lower than disk storage [41], i.e. around 0.1 W/TB.
- **Networking** Power consumption of network routers and services are rather small, typically at the level of 2–3% for a WLCG data center

There is an additional overhead for cooling ranging from 10% for highly optimized sites with direct warm-water cooling to 40% for traditional air-cooled systems.

References

[1] IPCC: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, (2023). https://doi.org/10.59327/IPCC/AR6-9789291691647 . https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_FullVolume.pdf

[2] IPCC: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, p. 1132. Cambridge University Press, Cambridge, UK and New York, USA (2014). https://www.ipcc.ch/site/assets/uploads/2018/02/WGIIAR5-PartA_FINAL.pdf

[4] HS23 is standard CPU benchmark used in WLCG, see (HS23).

[3] Kornhuber, K., Lesk, C., Schleussner, C.F., *et al.*: Risks of synchronized low yields are underestimated in climate and crop model projections. Nat Commun **14**, 3528 (2023) https://doi.org/10.1038/s41467-023-38906-7

[4] United Nations: Paris Agreement. Treaty Series **3156** (2015). https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVII-7-d&chapter=27&clang=_en, [Accessed: 11-Sep-2023]

[5] International Energy Agency IEA: Tracking Data Centres and Data Transmission Networks. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks. [Accessed: 29-Oct-2023] (2022)

[6] Bundesministerium für Bildung und Forschung (BMBF): Rahmenprogramm ErUM: Erforschung von Universum und Materie. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/7/31339_Erforschung_von_Universum_und_Materie.pdf?__blob=publicationFile&v=4. [Accessed: 17-Jul-2023] (2020)

[7] Bundesministerium für Bildung und Forschung (BMBF): Aktionsplan ErUM-Data. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/7/31640_Aktionsplan_ErUM-Data.pdf. [Accessed: 12-Oct-2023] (2020)

[8] Stiftung Allianz für Entwicklung und Klima: CO2e und GWP. https://allianz-entwicklung-klima.de/toolbox. [Accessed: 29-Oct-2023] (2023)

[9] Ritchie, H., Roser, M., Rosado, P.: Our World in Data: Energy. https://ourworldindata.org/energy. Accessed: 10-Sep-2023 (2022)

[10] Ullrich, J., et al.: Nicht-fossile Energie: Eine Globale Herausforderung für den Klimaschutz. Physikkonkret (66) (2023). https://www.dpg-physik.de/veroeffentlichungen/publikationen/physikkonkret

[11] Bundesministerium für Wirtschaft und Klimaschutz (BMWK): Fakten aus Eröffnungsbilanz Klimaschutz. https://www.bmwk.de/Redaktion/DE/Infografiken/Energie/fakten-aus-eroeffnungsbilanz-klimaschutz-pk-habeck.html. [Accessed: 13-Sep-2023] (2022)

[12] Germer, S., Kleidon, A.: Have wind turbines in Germany generated electricity as would be expected from the prevailing wind conditions in 2000-2014? PLOS ONE **14**, 0211028 (2019) https://doi.org/10.1371/journal.pone.0211028

[13] Wirth, H.: Aktuelle Fakten zur Photovoltaik in Deutschland (Fraunhofer ISE). [Accessed: 13-Sep-2023] (2023). www.pv-fakten.de,

[14] Banerjee, S., et al.: Environmental sustainability in basic research: High Energy Physics, Cosmology, Astroparticle Physics, and Hadron and Nuclear Physics (HECAP+). [Accessed: 31-Oct-2023] (2023). https://sustainable-hecap-plus.github.io

[15] Hey, A., Tansley, D., Tolle, K.: The Fourth Paradigm: Data-driven Scientific Discovery. Microsoft Research, Redmond (2009)

[16] DIG-UM, ErUM-Data-Hub: Workshop: Sustainability in the Digital Transformation of Basic Research on Universe & Matter (2023). https://indico.desy.de/event/37480

[17] Wilkinson, M., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L.O., Bourne, P., Bouwman, J., Brookes, A., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R., Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data **3**(160018) (2016) https://doi.org/10.1038/sdata.2016.18

[18] Barker, M., Chue Hong, N., Katz, D., Lamprecht, A.-L., Martinez Ortiz, C., Psomopoulos, F., Harrow, J., Garcia Castro, L., Gruenpeter, M., Martinez, P., Honeyman, T.: Introducing the FAIR Principles for research software. Scientific Data **9**(622) (2022) https:

//doi.org/10.1038/s41597-022-01710-x

[19] Lorimer, D.R., Bailes, M., McLaughlin, M.A., Narkevic, D.J., Crawford, F.: A bright millisecond radio burst of extragalactic origin. Science **318**, 777 (2007) https://doi.org/10.1126/science.1147532 arXiv:0709.4301 [astro-ph]

[20] White, T., et al.: Real-time crystallographic data processing with streaming data. European Crystallographic Association (ECA), 15th ECA Lunch webinar (2022). https://ecanews.org/education/eca-lunch-webinars_past-events/#15th-webinar

[21] Schörner, T., et al.: PUNCH4NFDI. https://www.punch4nfdi.de. [Accessed: 10-Sep-2023] (2021)

[22] Barty, A., Murphy, B., et al.: DAPHNE4NFDI. https://www.daphne4nfdi.de. [Accessed: 10-Sep-2023] (2021)

[23] Draxl, C., et al.: FAIRmat. \https://www.fairmat-nfdi.eu. [Accessed: 10-Sep-2023] (2021)

[24] Rieger, M.: Design Pattern for Analysis Automation on Distributed Resources using Luigi Analysis Workflows. EPJ Web Conf. **245**, 05025 (2020) https://doi.org/10.1051/epjconf/202024505025

[25] Claudet, S.: Less, better, recover. https://cerncourier.com/a/less-better-recover. [Accessed: 6-Oct-2023] (2022)

[26] Matthes, A., Widera, R., Zenker, E., Worpitz, B., Huebl, A., Bussmann, M.: Tuning and optimization for a variety of many-core architectures without changing a single line of implementation code using the Alpaka library. CoRR **abs/1706.10086** (2017) 1706.10086

[27] Dask Development Team: Dask: Library for Dynamic Task Scheduling. (2016). https://dask.org Accessed 2022-05-26

[28] Gray, L., et al.: CoffeaTeam/coffea: Release v0.7.11. Zenodo (2021). https://doi.org/10.5281/zenodo.5762406

[29] Eich, N., Erdmann, M., Fackeldey, P., Fischer, B., Noll, D., Rath, Y.: Fast Columnar Physics Analyses of Terabyte-Scale LHC Data on a Cache-Aware Dask Cluster. Comput. Softw. Big Sci. **7**(1), 3 (2023) https://doi.org/10.1007/s41781-023-00095-9 arXiv:2207.08598 [physics.data-an]

[30] Abdul Halim, A., *et al.*: Mass Composition from 3 EeV to 100 EeV using the Depth of the Maximum of Air-Shower Profiles Estimated with Deep Learning using Surface Detector Data of the Pierre Auger Observatory. PoS **ICRC2023**, 278 (2023) https://doi.org/10.22323/1.444.0278

[31] Sclafani, S., *et al.*: Observation of High-Energy Neutrinos from the Galactic Plane. PoS **ICRC2023**, 1108 (2023) https://doi.org/10.22323/1.444.1108 arXiv:2307.14842 [astro-ph.HE]

[32] Pithan, Linus and Starostin, Vladimir and Mareček, David and Petersdorf, Lukas and Völter, Constantin and Munteanu, Valentin and Jankowski, Maciej and Konovalov, Oleg and Gerlach, Alexander and Hinderhofer, Alexander and Murphy, Bridget and Kowarik, Stefan and Schreiber, Frank: Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments. Journal of Synchrotron Radiation **30**(6), 1064–1075 (2023) https://doi.org/10.1107/S160057752300749X

[33] Teixeira Parente, M., Brandl, G., Franz, C., Stuhr, U., Ganeva, M., Schneidewind, A.: Active learning-assisted neutron spectroscopy with log-Gaussian processes. Nature Communications **14**, 2246 (2023) https://doi.org/10.1038/s41467-023-37418-8

[34] Calafiura, P., Catmore, J., Costanzo, D., Di Girolamo, A.: ATLAS HL-LHC Computing Conceptual Design Report. Technical report, CERN, Geneva (2020). https://cds.cern.ch/record/2729668

[35] Aad, G., *et al.*: AtlFast3: The Next Generation of Fast Simulation in ATLAS. Comput. Softw. Big Sci. **6**(1), 7 (2022) https://doi.org/10.1007/s41781-021-00079-7 arXiv:2109.02551 [hep-ex]

[36] National Grid: What are scope 1, 2 and 3 carbon emissions? https://www.nationalgrid.com/stories/energy-explained/what-are-scope-1-2-3-carbon-emissions. [Accessed: 11-Sep-2023] (2023)

[37] MPCDF Garching: CO_2-footprint of MPCDF. https://www.mpcdf.mpg.de/about/co2-footprint. [Accessed: 8-Oct-2023] (2021)

[38] LRZ Garching: Jahresbericht 2021. https://www.lrz.de/wir/berichte/JB/JBer2021.pdf#page=101. [Accessed: 8-Oct-2023] (2021)

[39] HLRS Stuttgart: Umwelterklärung 2021. https://www.hlrs.de/fileadmin/about/social_responsibility/Sustainability/HLRS_Umwelterklaerung-2021.pdf#page=24. [Accessed: 8-Oct-2023] (2021)

[40] JSC Julich: Nachhaltigkeitsbericht 2012/2013. https://www.fz-juelich.de/de/aktuelles/news/pressemitteilungen/2014/14-07-01nachhaltigkeitsbericht?expand=translations,fzjsettings,nearest-institut. [Accessed: 8-Oct-2023] (2014)

[41] Campana, S.: A holistic study of the WLCG energy needs for the LHC scientific program. page 9 (2023). https://indico.jlab.org/event/459/contributions/11499/attachments/9236/14205/WLCGEnergyNeedsCHEP2023.pdf

[42] Grimshaw, A.: Concrete Path to Sustainable Computing (2023). https://indico.desy.de/event/37480

[43] Sladek, S.: WindCloud Company. https://windcloud.de. [Accessed: 5-Oct-2023] (2018)

[44] Lin, L., Zavala, V., Chien, A.: Evaluating Coupling Models for Cloud Datacenters and Power Grids. ACM Digital Library, 171–184 (2021) https://doi.org/10.1145/3447555.3464868

[45] DESY Computing Center: Research Facility 2.0: Towards a more energy-efficient and sustainable path. Funded by EU Call HORIZON-INFRA-2023-TECH-01-01 (2023)

[46] Walker, R.: Measures for sustainable computing operation (2023). https://indico.desy.de/event/37480

[47] National Renewable Energy Laboratory (NREL): High-Performance Computing Data Center Power Usage Effectiveness. https://www.nrel.gov. [Accessed: 28-Sep-2023] (2023)

[48] LRZ Garching: Mit Forschung und Supercomputing gegen Krisen. https://www.grid.lrz.de/presse/ereignisse/2022-12-16-Supercomputer-Energieverbrauch. [Accessed: 1-Nov-2023] (2022)

[49] Struckmeier, J.: Building sustainable systems (2023). https://indico.desy.de/event/37480

[50] Life Cycle Assessment of Dell R740. https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/Full_LCA_Dell_R740.pdf

[51] Gupta, U., Kim, Y.G., Lee, S., Tse, J., Lee, H.-H.S., Wei, G.-Y., Brooks, D., Wu, C.-J.: Chasing Carbon: The Elusive Environmental Footprint of Computing (2020) arXiv:2011.02839 [cs.AR]

[52] Standard Performance Evaluation Corporation: SERT 2.0. https://www.spec.org/sert2/index.html. [Accessed: 5-Oct-2023] (2017)

[53] Kocot, B., Czarnul, P., Proficz, J.: Energy-Aware Scheduling for High-Performance Computing Systems: A Survey. Energies **16**, 890 (2023) https://doi.org/10.3390/en16020890

[54] D'Amico, M., Jokanovic, A., Corbalan, J.: Holistic slowdown driven scheduling and

resource management for malleable jobs. In: Proc. 48th Int. Conf. Parallel Processing. ICPP '19. Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3337821.3337909

[55] Wiek, A., Withycombe, L., Redman, C.L.: Key competencies in sustainability: a reference framework for academic program development. Sustainability Science **6**, 203–218 (2011) https://doi.org/10.1007/s11625-011-0132-6

[56] UNESCO: Programm: Bildung für nachhaltige Entwicklung BNE 2030. https://www.unesco.de/bildung/bildung-fuer-nachhaltige-entwicklung/unesco-programm-bne-2030. [Accessed: 17-Jul-2023] (2020)

[57] UNESCO: AI and Education: Guidance for Policy-makers, (2021). https://doi.org/10.54675/PCSP7350 . ISBN 978-92-3-100447-6

[58] Niknejadi, P.: yHEP initiatives in sustainability. https://indico.desy.de/event/41165, V. Lang et al. https://yhep.desy.de/sustainability. [Accessed: 11-Oct-2023] (2023)

[59] United Nations General Assembly: Transforming our world: the 2030 Agenda for Sustainable Development. https://www.refworld.org/docid/57b6e3e44.html. [Accessed: 17-Jul-2023] (2015)

[60] German Federal Government: German Sustainable Development Strategy . https://www.bundesregierung.de/breg-de/themen/nachhaltigkeitspolitik/deutsche-nachhaltigkeitsstrategie-318846. [Accessed: 20-Jul-2023] (2021)

[61] Die Bundesregierung: Die Digitalstrategie Deutschland. https://digitalstrategie-deutschland.de/static/67803f22e4a62d19e9cf193c06999bcf/220830_Digitalstrategie_fin-barrierefrei.pdf. [Accessed: 17-Jul-2023] (2022)

[62] Die Bundesregierung: Aktueller Stand der Digitalstrategie Deutschland. https://digitalstrategie-deutschland.de/static/fcf23bbf9736d543d02b79ccad34b729/Digitalstrategie_Aktualisierung_25.04.2023.pdf. [Accessed: 17-Jul-2023] (2023)

[63] Bundesministerium für Bildung und Forschung (BMBF): Strategie: Forschung für Nachhaltigkeit (FONA). https://www.fona.de/de/themen/nachhaltigkeit-in-der-wissenschaft-sisi.php, https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/7/31638_Forschung_fuer_Nachhaltigkeit.pdf?__blob=publicationFile&v=6. [Accessed: 20-Jul-2023] (2020)

[64] Bundesministerium für Bildung und Forschung (BMBF): Richtlinie zur Förderung von Projekten zum Themengebiet Teilchen innerhalb des Rahmenprogramms Erforschung von Universum und Materie – ErUM. https://www.bundesanzeiger.de/pub/publication/8RKcvxMbxghMjChoMBF/content/8RKcvxMbxghMjChoMBF/BAnz%20AT%2028.04.2023%20B5.pdf. [Accessed: 17-Jul-2023] (2023)

[65] Erdmann, M., Fackeldey, P., Fischer, B., Fröse, S., Steinfeld, J., Warkentin, A.: ErUM-Data-Hub. www.erumdatahub.de. [Accessed: 29-Oct-2023] (2023)

[66] Britton, D.: Reducing the Carbon Footprint of Computing, Int. Symp. Grids & Clouds (ISGC) (2023). https://indico4.twgrid.org/event/25/contributions/1211/attachments/746/901/230324_ISGC-v1.pptx