GreenDIGIT

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Environmental Sustainability in Digital Research Infrastructures - the GreenDIGIT Project

Catalin Condurache *(EGI Foundation)*
Gergely Sipos *(EGI Foundation)*
Yuri Demchenko *(University of Amsterdam)*

Outline

- GreenDIGIT project scope and goals
  - Shared responsibility model for sustainability
- Landscape survey
  - Objectives, targets
  - Responses, findings
  - Opportunities, recommendations
- Environmental sustainability metrics
- ‘Green’ optimised frameworks
- Conclusions

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

GreenDIGIT Project - Founding Digital RIs (ESFRI)

- **EBRAINS** - An open research infrastructure that gathers data, tools and computing facilities for brain-related research
- **EGI** - International federation delivering e-Infrastructure and open solutions for advanced computing and data analytics in research and innovation
- **SLICES** - Scientific Large-scale Infrastructure for Computing and Communication Experimental Studies
- **SoBigData** - Distributed, pan-European, multi-disciplinary research infrastructure aimed at using social mining and Big Data to understand the complexity of our contemporary, globally interconnected society

https://greendigit-project.eu

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

GreenDIGIT Project (2024-2027) - Objectives

GreenDIGIT

- **O1: Assess status and trends** of low impact computing within 4 DIGIT RIs (EGI, SLICES, SoBigData, EBRAINS) and in the broader digital service provider community of ESFRIs, to produce **recommendations and roadmaps** for providers during and beyond the project.
- **O2: Provide reference architecture and design principles**, as well as an actionable model for RIs about environmental impact assessment and monitoring, reflecting on the **whole RI lifecycle** and including the digital infrastructure components and their interaction with the broader environment.
- **O3: Develop and validate new and innovative technologies, methods, and tools** for digital service providers within European Research Infrastructures through which they can reduce their energy consumptions and overall environmental impact.
- **O4: Develop and provide for researcher technical tools** that assist them in the design, execution and sharing of environmental impact aware digital applications with reproducibility, Open Science and FAIR data management considerations.
- **O5: Educate and support digital service providers and researchers** in the RI communities about good practices on environmental impact conscious lifecycle management and operation of infrastructures and services.

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Sustainability Aspects: Energy Efficiency – Decarbonisation – Environmental Impact

- **Energy Efficiency in Digital Infrastructures**:
  - **Definition**: This refers to optimizing digital infrastructures to consume as little energy as possible for a given workload or service. It's about achieving more computational or storage results with less energy input.
- **Decarbonization of Digital Infrastructures**:
  - **Definition**: This specifically targets the reduction of carbon emissions associated with the operation and maintenance of digital infrastructures.
- **Reducing Environmental Impact of Digital Infrastructures**:
  - **Definition**: This is a more comprehensive consideration of the various ways digital infrastructures might affect the environment, going beyond just energy consumption and carbon emissions.

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Sustainability Aspects: Energy Efficiency – Decarbonisation – Environmental Impact

- **Energy Efficiency in Digital Infrastructures**:
  - **Definition**: This refers to optimizing digital infrastructures to consume as little energy as possible for a given workload or service. It's about achieving more computational or storage results with less energy input.
- **Decarbonization of Digital Infrastructures**:
  - **Definition**: This specifically targets the reduction of carbon emissions associated with the operation and maintenance of digital infrastructures.
- **Reducing Environmental Impact of Digital Infrastructures**:
  - **Definition**: This is a more comprehensive consideration of the various ways digital infrastructures might affect the environment, going beyond just energy consumption and carbon emissions.

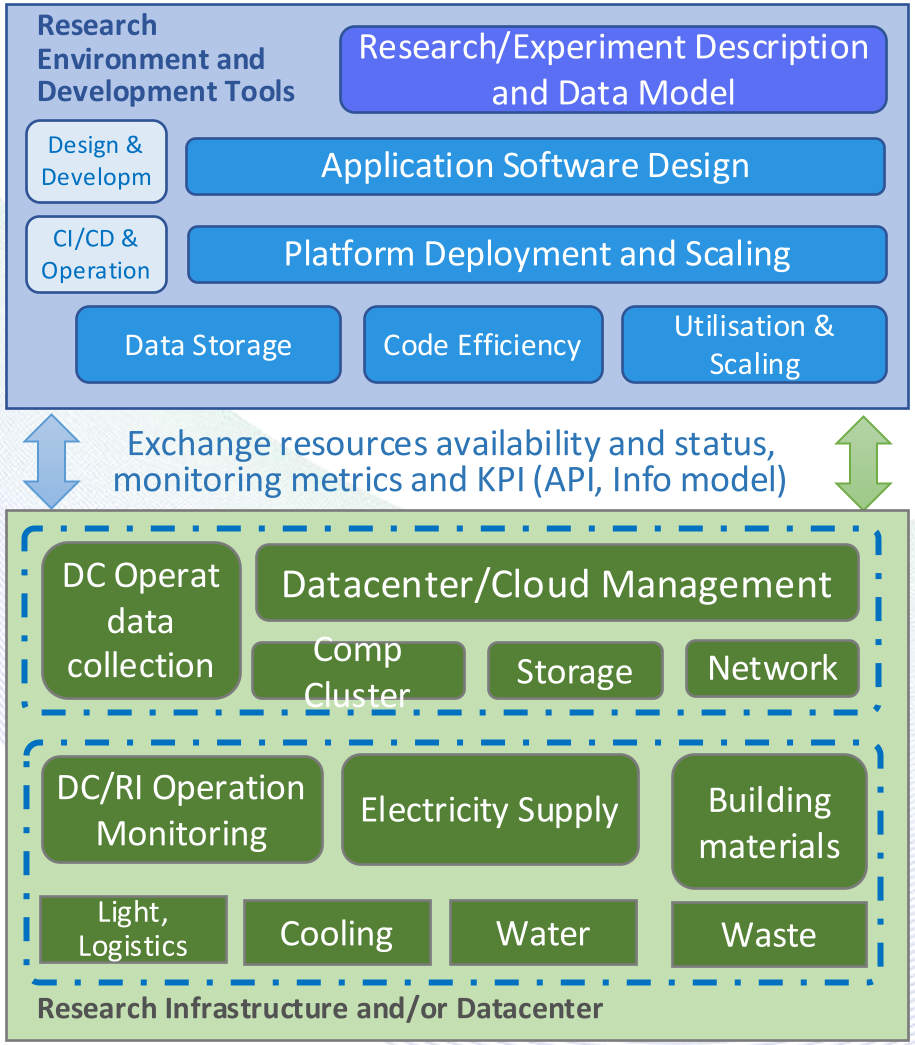
Architecture, Design, Recommendations

Operation, Monitoring, KPIs

Lifecycle, Policy, Training

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Shared Responsibility in Sustainability



Four categories identified

- Manufacturers
- RI operators
- SW developers
- RI users

Each category must assume their share towards the common goal

GreenDIGIT brings various solutions to help last 3 groups and incentivise first group

Standards and regulations
Software Development Quality and Design Patterns

Standards and regulations
Datacenter and RI Building and Operation

Project/Researcher Responsibility on RI
Applications Development, Deployment, Operation, Energy usage and KPI monitoring

Provider/Operator Responsibility of RI
Research Infrastructure or Datacenter, Monitoring Energy and environmental impact metrics and KPI

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

RIs Landscape Survey

GreenDIGIT

- Objective
  - To collect and analyse RIs practices, needs and gaps, and opportunities-barriers-good practices
  - What did the RIs do so far? → Good practices
  - Where are they going? → Initiatives
  - Where could we help? → Opportunities
- Target
  - GreenDIGIT consortium (EGI, SoBigData, EBRAINS, SLICES)
  - Thematic RIs with **significant** digital infrastructures
- Methodology
  - Preparation, survey and interviews, analysis of results and identification of key aspects

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

RIs Landscape Survey - Responses

PREPARATORY PHASE → PRE-PRODUCTION PHASE → PRODUCTION PHASE

slices RI | SOBIGDATA RESEARCH INFRASTRUCTURE | EBRAINS | egi

ICT facilities at 14 sites
(5G, Networks, Clouds)

Data, Computing and Digital

6 Computational

25 Research facilities

3 HPC centers

~200 HTC sites

~30 cloud sites

+

CERN (Particle acc., Intern.)
ELI-Beams (Laser facility, CZ)

Physical Sciences

CLARIN (Language resources)

Social Sciences & Humanities

GÉANT (Networks, Intern.)
Switch Cloud (CH)
University of Freiburg (DE)

Data, Computing and Digital

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

RIs Landscape Survey - Findings

- Included separated responses from EGI Federation sites
- Responses represent over 50 academic providers from 20+ countries:
  - 80% offering "Computation resources" as the primary resource
  - 60% offering "Networking / Communication resources"
  - 40% offering archives and scientific data
  - 30% offering also "Scientific equipment / instruments"
  - Each of the ~50 centers running between 50-10,000 physical core systems

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

RIs Landscape Survey - Findings

RIs served (*https://ri-portfolio.esfri.eu/*)

- **Energy**
  - EUROfusion
- **Environment**
  - LifeWatch
- **Health & Food**
  - ELIXIR, French Bioinformatics, Instruct, METROFOOD
- **Physical Sciences & Engineering**
  - WLCG, CTA, HESS, SKAO, LOFAR, BELLE II, ESA, AGATA, FOCAL
- **Social Sciences & Humanities**
  - CLARIAH, CESSDA, ODISSEI (NL)

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

RIs Landscape Survey - Findings

GreenDIGIT

- **Energy** is primary, **water** is secondary environment impact
  - Tracking total energy consumption is nearly everywhere (paying bills)
  - Power Usage Efficiency (PUE) is most common (bw 1.1 - 1.7).
  - Others: Watts/VM, Watts/GPU, Watts/vCPU, FLOPS/Watt
- Initiatives to **lower consumption/waste** exists, but isolated
  - Physical upgrade - Not trivial because of investment cost (modernise HW, A/C, building)
  - Software based optimisation at few places. Workflows are not designed for energy saving
  - Few practices for waste heat reuse (heating offices/homes - significant investment)
- Green computing **policy** is rare. An example at IRES, France: https://scigne.fr/resources/policies/green-computing-policy.pdf
- Very few centres **encourage/educate** staff on good practices
  - Switching lights or equipment off, reduce travel, cycling to work
  - Both staff operating the IT equipments and the end users are targets

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Survey -> Opportunities -> GreenDIGIT Plans

GreenDIGIT

- **Framework** for managing the full digital RI lifecycle from the environmental impact perspective
  - For status assessment, Design, Implementation, Operation, Decommissioning
- **Software tools** for power usage efficiency
  - Metrics infrastructure, Schedulers, Reproducibility frameworks, Resource managers (VMs, containers, workflows, AI/ML models, Data, etc.)
- **Develop skills and collaborations**
  - Training resources, events, courses, forums

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Planned SW Tools for Power Usage Efficiency

PROVIDERS' SIDE

- **Metrics infrastructure**
  Gather relevant environmental sustainability metrics from providers
  Serve metrics to decision making systems (schedulers, frameworks, VREs)
- **Energy-efficiency aware schedulers**
  Reshuffle jobs based on its type (CPU bound or I/O bound)
  Delay execution based on electricity cost, regulations, carbon intensity

- **Developer tools for energy efficient software design**
  POS, JupyterHub, Greenspector
- **Scientific reproducibility and resource management platform**
  JupyterHub, Binder, DataHub, SoBigData, etc.

USERS' SIDE

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

GreenDIGIT

Initial Recommendations

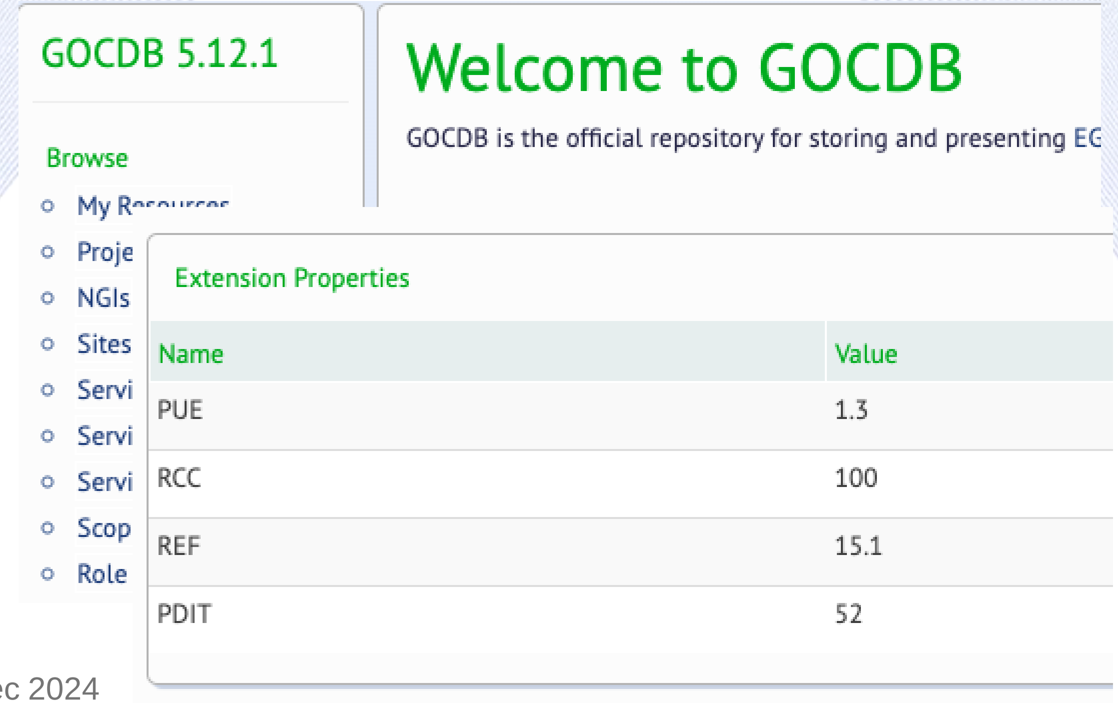
Recommendation	How to achieve	How to measure
Use less (per flop/Byte/bit)	HW: Modern building; Modern hardware (e.g. flash instead of disks); Led lights SW: Optimised code (AI/workflow); Job monitoring (to avoid ill configured jobs on HPC nodes; Switching off unused nodes Physical: Less staff travel; Data centre in cold region	W / YEAR (energy) m^3 (water)
Use for the right purpose	Higher A/C temp; Hot/cold aisles; Turn off lights	PUE
Use from renewable source	Change supplier; Deploy solar panel; Delay workload to when energy is greener	Green %
Share the infrastructure	Invest into shared infrastructure (use) for better cost-value ratio, to eliminate ‘downtime waste’	In-house - Cloud ratio
Reuse your waste	Repair; Waste heat reuse (office heating); Circular economy recycling (paper, plastic, …)	Waste reuse %
Train people, support new ideas	Office training; Information dashboards	# training hours # new initiatives
Measure and improve	Regularly update the metrics	

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Environmental Sustainability Metrics

EGI partners in consortium involved in choosing and proposing a use for them

- A very small set to start with from
  - PUE, WUE, ERF, REF, Carbon intensity, Energy consumption, …
  - EC Delegated Regulation (EU) 2024/1364 of 14 March 2024
  - Likely only 2-3 for the beginning
- EGI Configuration Database (goc.egi.eu) proposed to collect and manage them
  - Extension Properties feature (*Name, Value*) to be used to store these metrics
  - If not available for a site, the country value (from official reports) to be used



'Green' Optimised Frameworks

GreenDIGIT

GreenDIGIT will implement energy aware workflow/orchestrator optimisation algorithms within relevant frameworks

- DIRAC Workload Manager(@CNRS)
  - Perform validation through scientific user cases (WeNMR, Biomed) for HTC and Cloud
  - Possibilities to alternate production runs with/without 'green' metrics to validate algorithms
- AI4EOSC framework (@CSIC)
  - Validation through iMagine ML/AI models

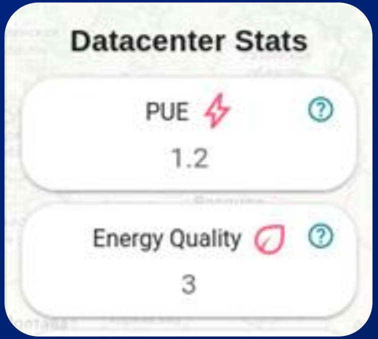
DIRAC THE INTERWARE

AI4 eosc

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

AI4EOSC Power Consumption & Environmental Impact

AI4 | eosc



AI4EOSC aims to make users aware of the impact produced by training and deploying AI models because it can leverage significant computing power.

We need to instrument at different levels our platform in order to:

Produce metrics at datacenter level.

- PUE, energy quality, efficiency, hardware characteristics.
- Currently static, working to make data collection dynamic.

Produce metrics at model level.

- Impact on training phase. (i.e. how much does it *cost* to build a model)
- Impact on inference/prediction. (i.e. how much does it *cost* to use the model)

The approach is to measure the power consumption of the platform through monitoring the power consumption of the processes that runs the VMs where is deployed on. This has to be done from bare metal machines (It requires access to Infrastructure Level).

- Not trivial, as GPUs and other specialized hardware must be taken into account.
- Nowadays, it is not possible to automate the instrumentation of the whole OpenStack cloud stack without cooperation from resource providers.

‘Green’ Optimised Frameworks

GreenDIGIT will implement energy aware workflow/orchestrator optimisation algorithms within relevant frameworks

- Terraform DevOps tool (@SZTAKI)
  - Validation via Environmental science use cases using Terraform containers
- Data Center Energy scheduler (@CESNET)
  - Power-grid aware computing algorithms
  - Control consumption based on
    - Electricity cost, power-grid regulation needs
    - Current carbon intensity of regional electricity
    - Reused heat demand

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

GreenDIGIT, EGI and WLCG

GreenDIGIT

- Sites and/or experiments are welcome to join
  - Either during validation process
  - Or during integration and acceptance within RIs
- EGI Green Computing Task Force
  - Re-activated following EGI-ACE project
  - The entity that links interested EGI stakeholders/sites to GreenDIGIT
  - https://confluence.egi.eu/display/EGIBG/Task+Force%3A+Green+Computing
  - Email list: green-computing-TF@mailman.egi.eu

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Conclusions

- GreenDIGIT works on an overarching framework for environmental impact reduction methodology in digital sciences
- The framework considers the RI lifecycle and shared responsibility models
- Technological solutions to be provided for compute centers and software developers
- Get involved via the EGI GC TF - green-computing-tf@mailman.egi.eu

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

GreenDIGIT

Thank you!

Questions!

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Backup slides

GreenDIGIT

COMMISSION DELEGATED REGULATION (EU) 2024/1364 of 14 March 2024 on the first phase of the establishment of a common Union rating scheme for data centres - https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401364 - result of Directive (EU) 2023/1791

Proposed metrics - KPIs (Annex II)

- **Total energy consumption** (“E_{DC}”, in kWh) of the reporting data centre shall be measured as defined by, and by using the methodology in the CEN/CENELEC EN 50600-4-2 standard or equivalent. Total energy consumption includes the use of electricity, fuels and other energy sources used for cooling.
- **Total energy consumption of information technology equipment** (“E_{IT}”, in kWh)
- **Total water input** (“W_{IN}”, in cubic meters) shall be measured as defined by, and by using the methodology set out in the CEN/CENELEC EN 50600-4-9 standard WUE Category 2
- **Waste heat reused** (“E_{REUSE}”, in kWh) shall be measured as defined by, and by using the methodology set out in, the CEN/CENELEC EN 50600-4-6 standard or equivalent
- **Average waste heat temperature** (“T_{WH}”, in degree Celsius) shall be measured as the temperature of the fluid used to cool the information and communication technology equipment in the data centre computer room, averaged over the year, and across every measurement point.

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207

Backup slides

GreenDIGIT

- **Total renewable energy consumption** (“$\mathbf{E_{RES\text{-}TOT}}$”, in kWh)
  $\mathbf{E_{RES\text{-}TOT}}$ is the sum of $\mathbf{E_{RES\text{-}GOO}}$, $\mathbf{E_{RES\text{-}PPA}}$ and $\mathbf{E_{RES\text{-}OS}}$, as defined below
- **Total renewable energy consumption from Guarantees of Origin** (“$\mathbf{E_{RES\text{-}GOO}}$”, in kWh) shall be determined as the sum of the Guarantees of Origin purchased and retired by the reporting data centre
- **Total renewable energy consumption from Power Purchasing Agreements** (“$\mathbf{E_{RES\text{-}PPA}}$”, in kWh) shall be determined as the amount of energy from Power Purchasing Agreements made by the reporting data centre
- **Total renewable energy consumption from on-site renewables** (“$\mathbf{E_{RES\text{-}OS}}$”, in kWh) shall be measured as the energy generated from on-site renewable energy sources within the data centre boundary.

Proposed Data Centres sust. Indicators (Annex III) - shall be calculated based on the information and key performance indicators communicated to the European database on data centres in accordance with Annexes I and II.

- **Power Usage Effectiveness (PUE) = $\mathbf{E_{DC} / E_{IT}}$**
- **Water Usage Effectiveness (WUE) = $\mathbf{W_{IN} / E_{IT}}$**
  E_{IT} in MWh
- **Energy Reuse Factor (ERF) = $\mathbf{E_{REUSE}/E_{DC}}$**
- **Renewable Energy Factor % (REF) = $\mathbf{E_{RES\text{-}TOT} / E_{DC}}$**
  Default values from EC Country Report

Acknowledgement:
This project has received funding from the European Union's HE research and innovation programme under the grant agreement No. 101131207