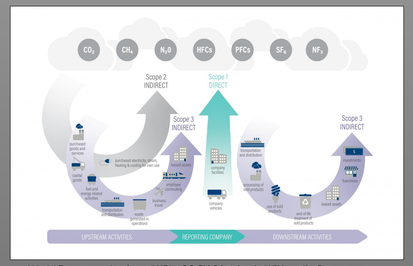
Measuring and Managing the Carbon Footprint of Digital Preservation

Arkivum

matthew.addis@arkivum.com

Emission Sources

Carbon emissions from LTDP can be calculated by following the GHG protocol. This includes the whole supply chain such as use of hosted preservation services and solutions.



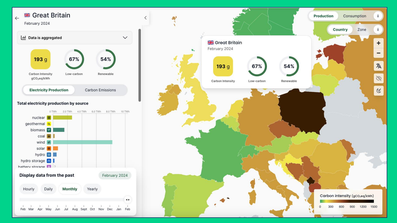
World Resources Institute (WRI), CC BY-SA 4.0 , via Wikimedia Commons

Carbon emissions come from a wide range of sources:

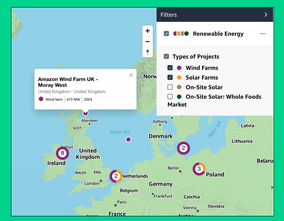
- Energy (power and cooling) for office buildings and ICT equipment such as servers and storage for data.
- Embodied footprint of offices and data centres such as raw materials and building construction.
- Embodied footprint of ICT equipment such as materials, manufacture, transport, maintenance, recycling and disposal.
- Staff, including travel and commuting to work.

Green Energy

Renewable electricity is readily available in many geographies and is used routinely by the major cloud providers, many of which fund / build new capacity for green energy generation. The result is that the carbon emissions from LDTP from energy consumption, especially in the cloud, can be net zero at little or no additional cost, for example by selecting appropriate cloud providers and locations.



https://app.electricitymaps.com/map



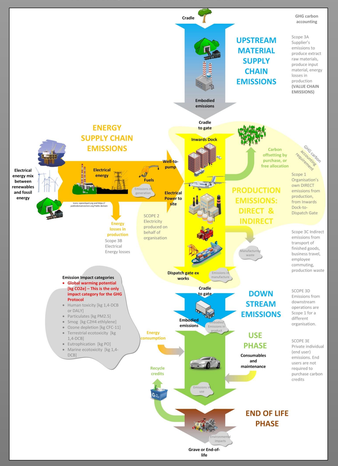
https://sustainability.aboutamazon.com/climate-solutions/carbon-free-energy

Embodied Footprint

Embodied footprint of ICT equipment (servers, storage, networking etc.) used in LTDP is the 'elephant in the room' and typically gets ignored. Life Cycle assessment (LCA) should be used to estimate the contribution to carbon emissions when used on-premise or in the cloud.

Storage	
• Storage Embodied Factor:	kgCO2eq per GB
• HDD lifetime	4-6 years
• Hard Drive	5 kgCO2eq per TB per year
Servers	
• Cloud server lifetime	4-6 years
• Cloud server utilization	50 – 65%
• 1 core-hour ~	0.5 gCO2eq
Data Tape Libraries and media	
• Deep archive LTO	1 kgCO2eq per TB per year

John Pons, CC0, via Wikimedia Commons



Measure Footprint

https://doi.org/10.6084/m9.figshare.20653101

Arkivum has measured / calculated the carbon footprint of our LTDP solution when running in the cloud. We considered both energy use and the embodied footprint of the ICT resources used.

1. Collect resource consumption and carbon emissions from cloud provider reports
   - CPU resource consumption (core-hours)
   - Storage consumption (GB-months)
   - Gross carbon emissions per resource type (kgCO2 eq)
2. Calculate metrics for emissions when using different types of cloud resource
   - kgCO2 eq per core-hour for compute
   - kgCO2 eq per TB-year for storage
3. Measure resource consumption for specific preservation workflows:
   - Large files, small files, inside bagit bags, big ingests, lots of small ingests
   - File format identification, checksum generation, virus scans, metadata extraction, replication etc.
   - Additional processing using Archivematica on-demand, for example file format normalisation
4. Calculate carbon emissions for each of the different use cases
   - For example, kgCO2 eq per TB of data ingested for different scenarios

Preservation Use Cases

https://www.dpconline.org/blog/blog-matthew-addis-enviornmental-23

Large image Datasets

	Gross Emissions from Energy Consumption	Estimated Embodied Footprint
1 PB data stored for 1 year	7800 kgCO2 eq	4000 kgCO2 eq
1 PB ingest of large image files	1600 kgCO2 eq	200 kgCO2 eq

Large collections of office files

	Gross Emissions from Energy Consumption	Estimated Embodied Footprint
1M office files stored for 1 year	5.5 kgCO2 eq	4 kgCO2 eq
Ingest of 1M office files.	140 kgCO2 eq	25 kgCO2 eq

Net emissions from energy use are zero, embodied footprint isn't!

The variation of carbon emissions when running the Arkivum SaaS solution in the cloud for different LDTP use cases (data types, data volumes, workflows) show that it is important to measure real-world scenarios to help understand what are the major contributing factors.

Carbon Reduction Strategies

Steps to help reduce carbon emissions:

- Ensure selection and appraisal only retains what's needed and that unwanted data doesn't build up over time.
- Remove unnecessary carbon intensive steps from preservation workflows, adopt 'Minimal Effort Ingest' / 'Minimum Viable Preservation'.
- Make good use of shared or spare resources, including in the cloud: this helps minimize embodied footprint by reducing the overall use of ICT resources.
- Don't get caught by Jevons Paradox. Green energy doesn't mean use more of it and don't worry: embodied carbon footprint is not zero.

Jevons paradox

https://en.wikipedia.org/wiki/Jevons_paradox

By Unknown author - Popular Science Monthly Volume 11, Public Domain, https://commons.wikimedia.org/w/index.php?curid=11022925

In economics, the Jevons paradox occurs when technological progress increases the efficiency with which a resource is used, but the falling cost of use induces increases in demand enough that resource use is increased, rather than reduced