Measuring environmental impact at CLOCKSS

Alicia Wise
Executive Director
awise@clocks.org

May 26, 2024

CLOCKSS

DIMPACT Project Partners - https://dimpact.org/about

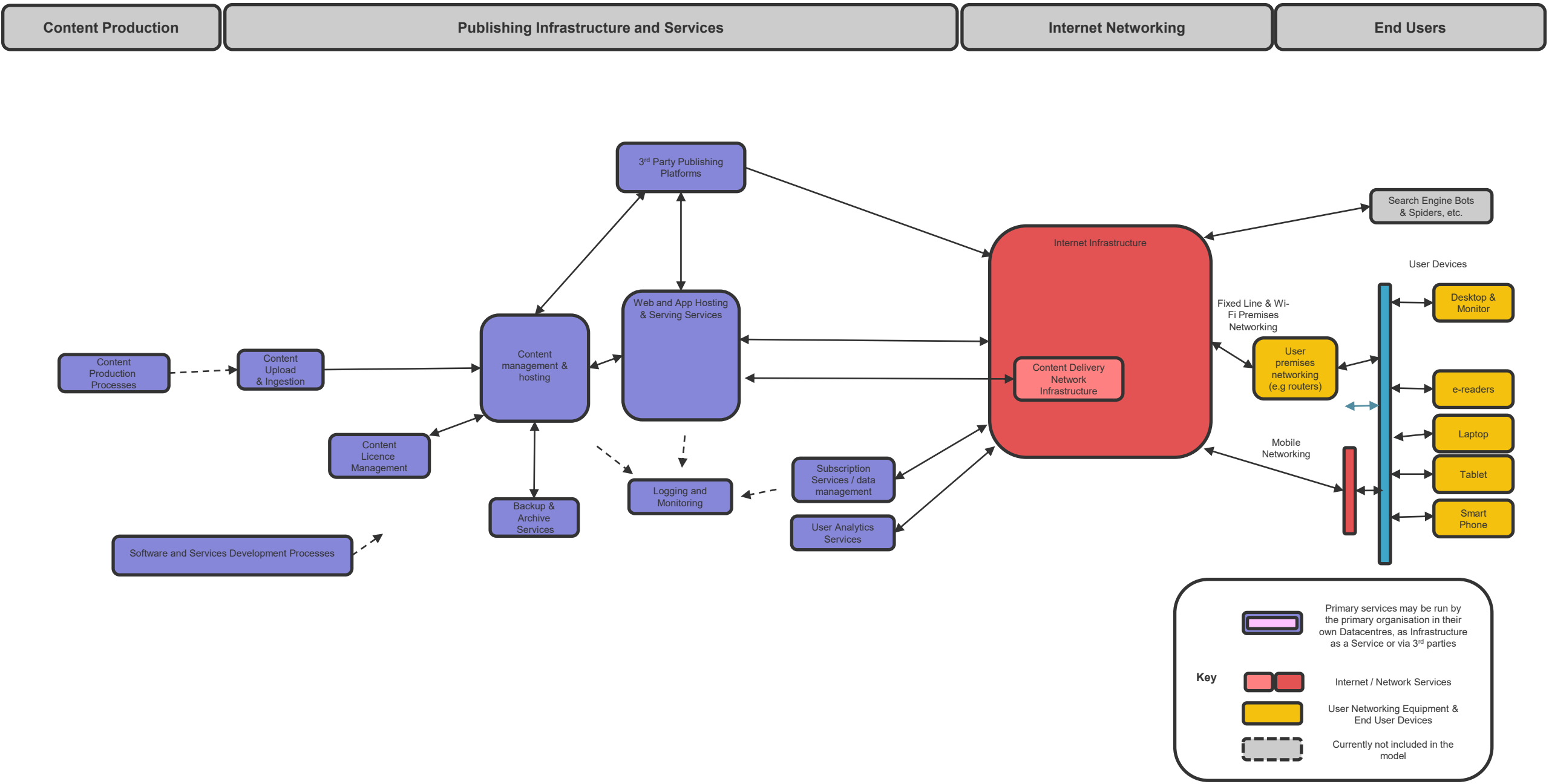
Pioneering tool to manage media industry's digital carbon footprint

Press release issued: 13 January 2020

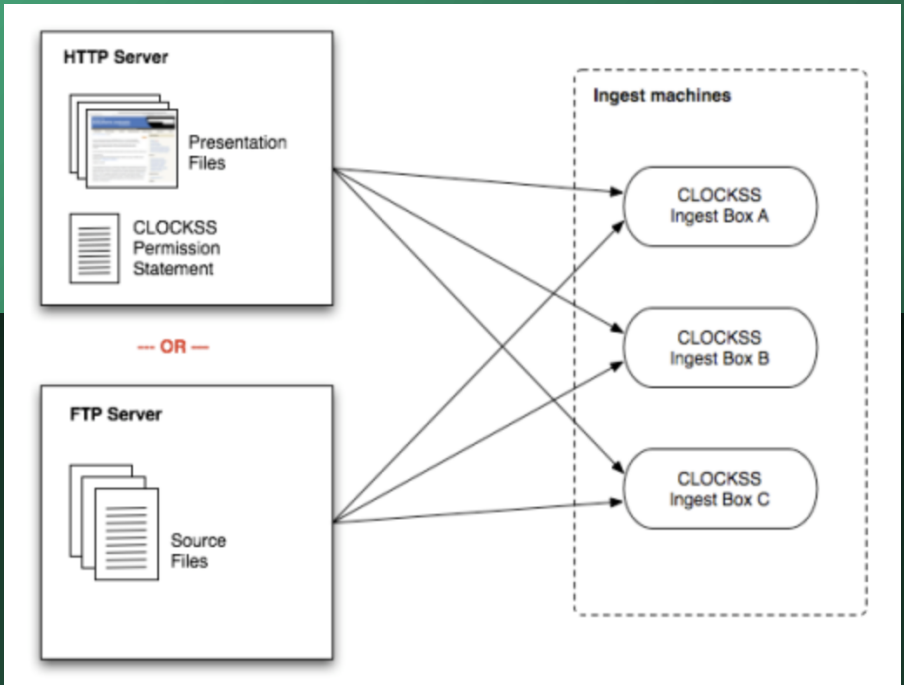
A collaboration between computer scientists at the University of Bristol and nine major media companies, including ITV and BBC, will help the media industry understand and manage the significant carbon impacts of digital content.

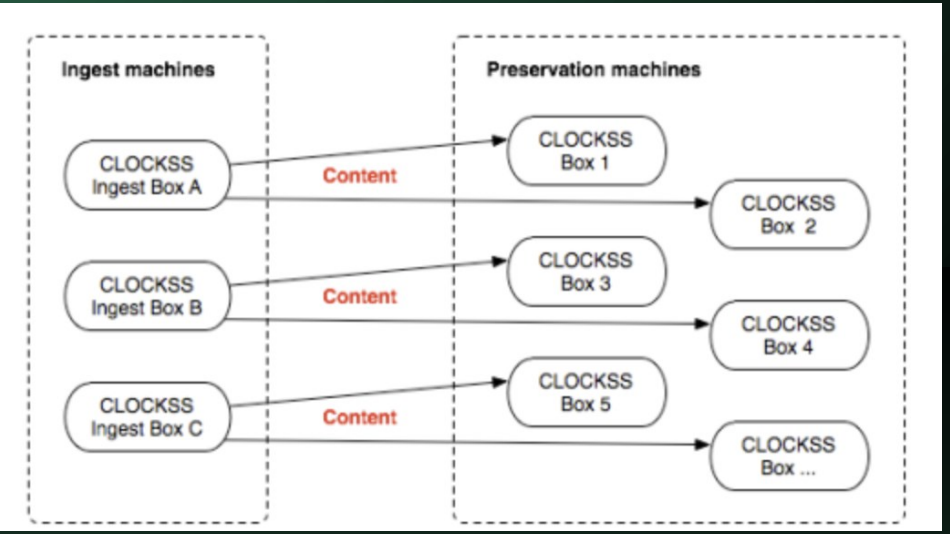
The 12 month collaboration, facilitated by sustainability experts, Carnstone, will see University of Bristol researchers working with sustainability and technology teams at the BBC, Dentsu Aegis Network, Informa, ITV, Pearson, RELX, Schibsted, Sky and TalkTalk, to map the carbon hotspots of digital media content and services.

DIMPACT is the first serious collaborative attempt to create a tool


Content Production
Publishing Infrastructure and Services
Internet Networking
End Users
3rd Party Publishing Platforms
Search Engine Bots & Spiders, etc.
Internet Infrastructure
User Devices
Content Production Processes
Content Upload & Ingestion
Content management & hosting
Web and App Hosting & Serving Services
Content Delivery Network Infrastructure
Fixed Line & Wi-Fi Premises Networking
User premises networking (e.g routers)
Desktop & Monitor
e-readers
Laptop
Tablet
Smart Phone
Mobile Networking
Content Licence Management
Backup & Archive Services
Logging and Monitoring
Subscription Services / data management
User Analytics Services
Software and Services Development Processes
Key
Primary services may be run by the primary organisation in their own Datacentres, as Infrastructure as a Service or via 3rd parties
Internet / Network Services
User Networking Equipment & End User Devices
Currently not included in the model

How CLOCKSS works





1) CLOCKSS is the largest LOCKSS network and uses a network of 12 nodes.
2) These are held at Stanford University (which currently hosts two), Rice University in Texas, Indiana University, the University of Virginia, OCLC in Ohio, the University of Alberta in Canada, Edinburgh University in the UK, Humboldt University in Germany, the Catholic University of Milan in Italy, the Australian National University in Canberra, and the last at the National Informatics Institute in Tokyo.
3) The 12 preservation machines each contain a copy of *all* the content archived with CLOCKSS.

Project Team

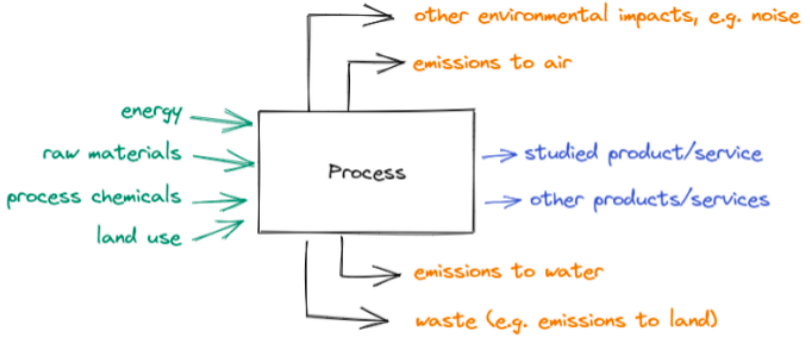
- Thib Guicherd-Callin (Stanford University Library)
- William Pickett (Carnstone)
- Paul Shabajee (University of Bristol, School of Computer Science)
- Danielle Spalenka (Indiana University Library)
- Ianthe Sutherland (University of Edinburgh Library)
- James Wickenden (Carnstone)
- Alicia Wise (CLOCKSS)
- Guanwen "Henry" Zhang (University of Alberta Library)

What we did

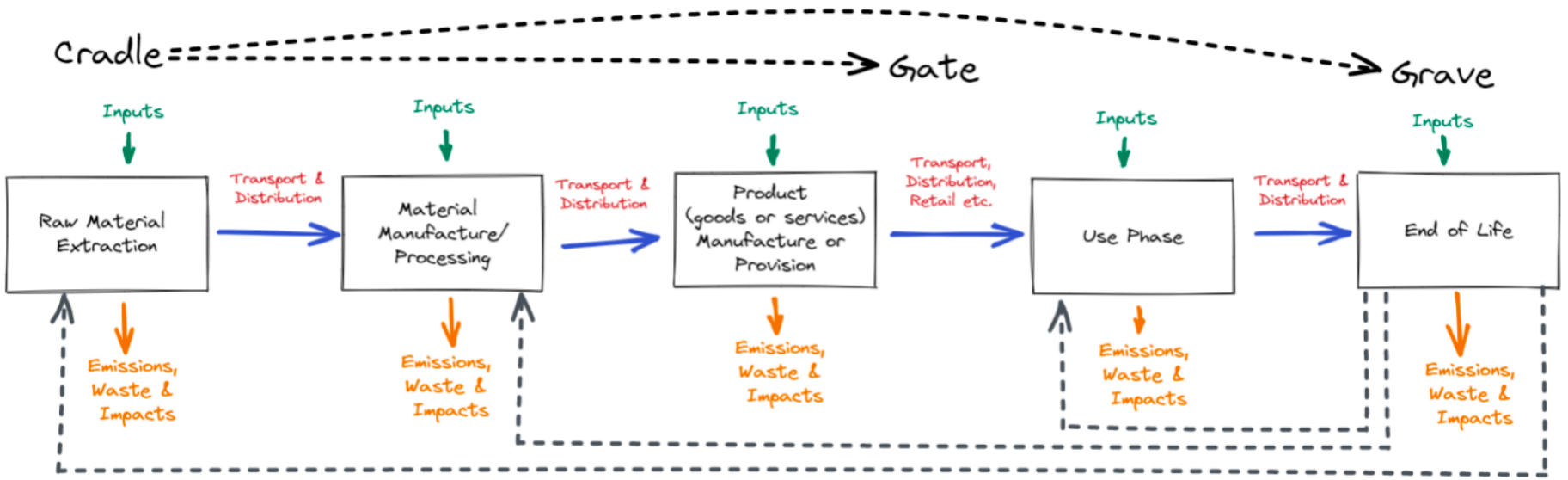
- We organized a 2-hour workshop to talk through:
  - the significant energy-using components
  - how the service is managed and used
- The DIMPACT project team developed a rough map.
- We had a second workshop to refine this and agree on what data we wanted to collect for each machine and host site.
- An unexpected need was for a shared vocabulary of all our machines. This had unintended benefits!

Carbon Footprinting

- Carbon footprinting is a kind of Environmental Life Cycle Assessment (LCA) focused on Greenhouse Gas (GHG) emissions



Examples of inputs, flows, etc. LCA looks at in processes

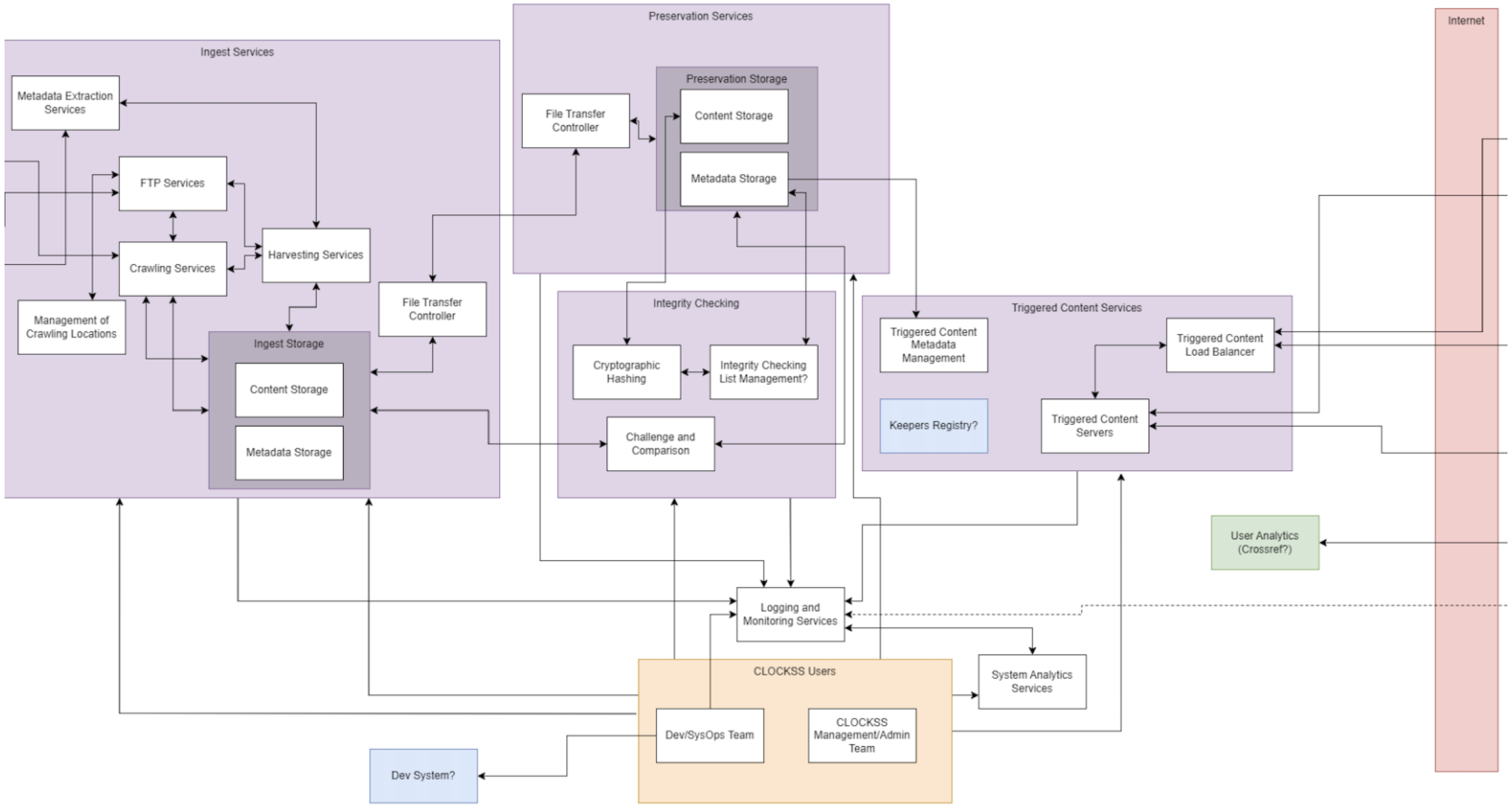


Very Simplified Life Cycle and Life Cycle Stages

CLOCKSS preservation machines

	Stanford	Rice	Indiana	Edinburgh	Others
Physical machines	ingest1 ingest2 ingest5 clockss-stanford clockss-stanford2 clockss-ftp clockss-ftp-backup	ingest3 clockss-rice	ingest4 clockss-indiana	clockss-edinburgh	clockss-virginia clockss-alberta clockss-oclc clockss-humboldt clockss-milano clockss-anu clockss-nii
Virtual machines	triggered-stanford triggered-lb content1 content2 content3 content4			triggered-edinburgh	

CLOCKSS Digital Archive Model (draft)

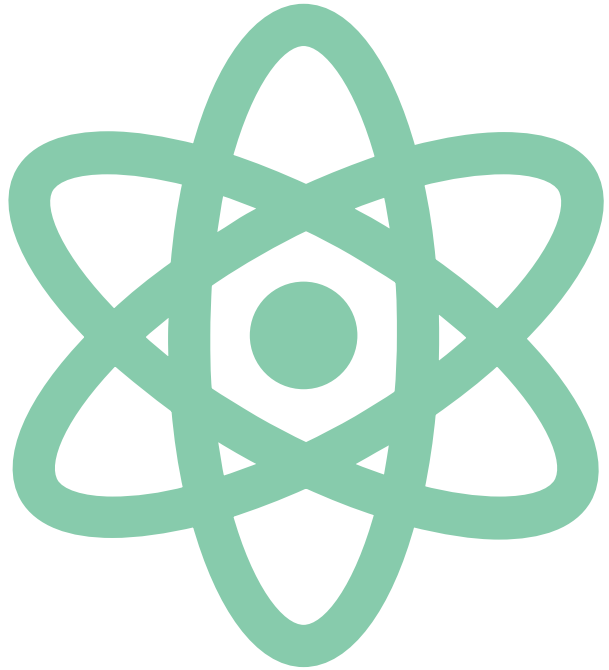


Data Gathering

- What are the specifications for each machine?
- How much energy do the manufacturers claim that these use?
- What is your evidence of how much energy they actually use?
- How much is each idle or active?
- For ingest, what proportion of content is pushed to CLOCKSS vs scraped by CLOCKSS?
- For preservation, what fixity or other checks do you do?
- For egress, how do you migrate and make content available?
- At each site, what is the source of energy and how much of it is from renewable sources?
- What happens to decommissioned machines?

Stanford example

- The majority of CLOCKSS machines are at Stanford.
- Stanford University is using a very high renewable energy profile thanks to the sun in California.
- Stanford owns a solar power plant in southern California which offsets power that they take from the grid in northern California. They also have solar panels on rooftops, and contracts with energy companies that also supply electricity from (mostly!) renewable sources. Exact mix depends on the day.
- For cooling data centres, Stanford uses a chilled water system, whereby giant ice cubes are made in the ground at night and then those ice cubes melt during the day to provide cooling. It is a heat pump and exchange type of process that produces cooling during periods when it's relatively inexpensive to do so.
- We discovered that the university has an Office of Sustainable Energy with lots of experts who were very helpful.

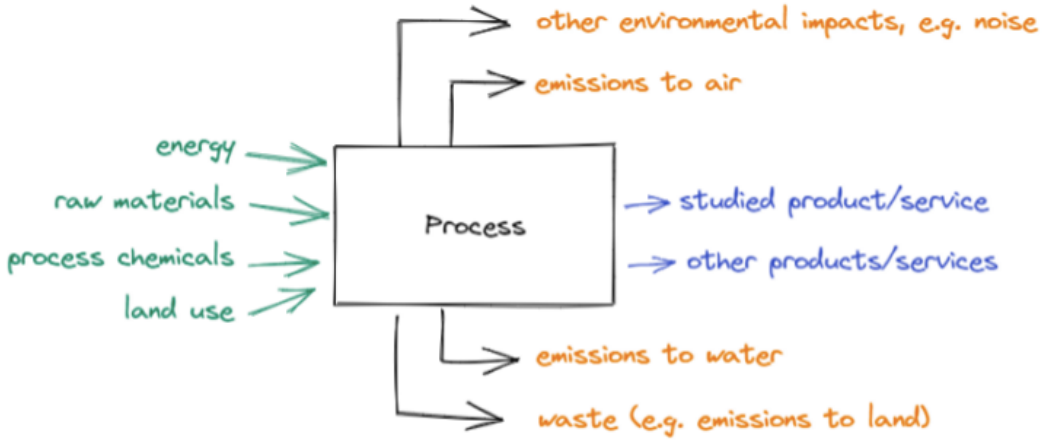


Are we right to focus only on carbon for now?

Yes! This is a best first step.

We could think more about the carbon embodied in the devices we use and how these devices were produced (e.g. raw material extraction, manufacturing, transportation, retail processes).

We may need to consider emissions to air, emissions to water, even perhaps noise which has an environmental impact.



CLOCKSS:

Preserving Scholarship for Future Generations

https://clockss.org/