DIGITAL SOVEREIGNTY BY DESIGN

Building IT research infrastructure at TECH, Aarhus University

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
AARHUS UNIVERSITY

DEIC CONFERENCE 2026
28. OCTOBER 2026

ESKE CHRISTIANSEN

QUESTIONS TO ASK YOURSELF WHEN BUILDING RESEARCH INFRASTRUCTURE

- Who is it for, and what problem does it solve for them?
- Will it support reproducibility?
- Who maintains it after the grant or project ends?
- What are the security, privacy and legal constraints?
- Should we build it, or adopt something that already exists?
- How easy is it to adopt?
- Does it work with the rest of the ecosystem?
- How will you measure success and give credit?

Will this still be usable, maintained, and trusted in five years? Who is responsible for that?

LEGAL STATUS?

- Clear legal status of data stored on home network on home resources.
- Clear legal status of data stored in public clouds that are owned and placed by European companies.
- Unclear legal status of data stored on European ground but owned by US owned companies.
- Unclear legal status of DeiC operated shared services e.g., DeiC Dataverse!
- That implies that
  - The data stays under Danish and EU law, and the institution is the data controller with direct control. This is the safest default for sensitive data.
  - The data is only subject to EU law, and GDPR applies through a standard data processing agreement.
  - Having the datacentre in the EU isn't enough. US laws such as the CLOUD Act and FISA 702 can require a US parent company to hand over data, wherever it's stored. The EU–US Data Privacy Framework addresses this for now, but it has been challenged in court, and its predecessors (Safe Harbor and Privacy Shield) were both struck down. The legal basis could disappear again.
  - The issue with DeiC shared services is jurisdiction. It's that the roles and responsibilities aren't clearly defined: who is the data controller and who is the processor, which agreements cover the service, what kinds of data it's approved for, and who is liable.

Data sovereignty is defined by control, not location. A shared national service is only as useful as its legal status is clear.

CORE DESIGN PRINCIPLES

- Use AU services as much as possible
  - For example, GitLab and ERDA, rather than building our own or using external alternatives.
- Keep data on the AU network whenever it's technically possible.
- If data must leave AU, use EU-owned providers.
  - Use US-owned providers only after a documented risk assessment and with an exit plan.
- Run simple tasks close to the user
  - On a laptop, workstation or local server
- Classify data before choosing where it lives
  - The classification (public, internal, confidential or sensitive/personal) determines where the data may be stored and processed.
- Make GDPR and security the default
  - Use least-privilege access, AU identity (federated login via WAYF/eduGAIN), encryption, and audit logging for sensitive data.

AU first · Data stays home · EU before US · Compute close to the user, heavy work to HPC · Classify first · FAIR and reproducible · Secure by default · Every service has an owner · Adopt before build ·

Researchers' time matters

DOG FOODING!

- Move from US owned public cloud service that we are using now.
- Choosing a European owned public cloud provider that
  - Is only subject to EU law.
  - Makes GDPR compliance straightforward.
  - Is our preferred option when data can't stay on the AU network.
  - Is transparent about its ownership and subcontractors.
  - Has predictable, transparent pricing.
  - Meets recognised security standards
- We went looking in Europe...

COLLECT AS MANY WE COULD FIND

Name	Country (Owner)	Link	Run Containers	Scaling	HPC	ORCE2024	Prices	Fully EU Cloud!	Comment
Hetzner	DE	https://www.hetzner.com/cloud	Yes	?	No	No		Yes, Have datacenter in US and Asia	
Open Telekom Cloud	DE	https://public.t-cloud.com/en	Yes	Yes	Yes	Yes		Yes	
OVHcloud	FR	https://www.ovhcloud.com/en/	Yes	Yes	Yes	Maybe (no danish rep?)		Yes	
Scaleway	FR	https://www.scaleway.com/en/	Yes	Yes	?	No		?	
UpCloud	FI	https://upcloud.com/	Yes	Yes	?	No		?	
Exoscale	CH	https://www.exoscale.com/	Yes	Yes	?	Yes		No	
IONOS	DE	https://www.ionos.com/	Yes	Yes	?	Maybe (no danish rep?)		?	
ELASTX	SE	https://elastx.se/en	Yes	Yes	?	No		Yes	
gridscale	DE	https://gridscale.io/en/	Yes	Yes	Maybe	No		Yes, have two datacenters in CH.	An OVHCloud Company.
Cyso Cloud	NE	https://cyso.cloud/	Yes	Yes	?	No		Yes	
Seeweb	IT	https://www.seeweb.it/en	Yes	Yes	?	No		?	
Aruba Cloud	IT	https://www.arubacloud.com/	Yes	Yes	?			No	
Merrymake	DK	https://www.merrymake.eu/	?	?	?	No		?	Looks shady
Cloud Temple	FR	https://www.cloud-temple.com/en/	Yes	?	?	No		?	
Clever Cloud	FR	https://www.clever.cloud/	Yes	Yes	?	No		?	Uses other cloud providers. Some US based...
STACKIT	DE	https://stackit.com/en	Yes	Yes	?	No		Yes	
		https://www.scannet.dk/cloud-							

THEN SCORE THEM ALL.

	Corporate Profile (40%)				HCP Ecosystem (15%)			Internal Services (20%)			Pricing & Efficientcy (25%)			Total Overall Score	
Name	Country owner	Jurisdiction (50%)	Infrastructure (20%)	Track Record (30%)	SUM	Hardware Ceillings (50%)	HPC Readiness (50%)	SUM	Uptime SLA (25%)	Backups & Snapshot (75%)	SUM	Standardized benchmark (75%)	Burstable & Idle Savings (25%)	SUM	SUM	
Hetzner	DE	5	5	5	5	3	1	2	3	3	3	5	1	4	3,9	Cheap - but pure IaaS - no HPC tools
Open Telekom	DE	5	5	5	5	5	5	5	5	5	5	1	5	2	4,25	Very good - but expensive
OVHcloud	FR	5	5	3	4,4	5	3	4	3	3	3	5	5	5	4,21	Also good
Oracle Cloud	US	1	5	3	2,4	5	5	5	5	5	5	3	5	3,5	3,585	Current
Scaleway	FR	5	5	5	5	5	3	4	3	1	1,5	5	5	5	4,15	Very good - but lacks backup
ScanNet	DK	5	3	5	4,6	5	1	3	3	3	3	1	1	1	3,14	
Exoscale	CH	5	5	5	5	3	1	2	3	3	3	3	1	2,5	3,525	
UpCloud	FI	5	3	5	4,6	5	1	3	5	3	3,5	3	1	2,5	3,615	
Ionos	DE	5	5	1	3,8	1	1	1	3	5	4,5	3	5	3,5	3,445	
Elastyx	SE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<- Didnt seem fit so didnt finish them
Contabo	DE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<- Didnt seem fit so didnt finish them
Evroc	SE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	<- Didnt seem fit so didnt finish them
Nine	CH	5	3	5	4,6	5	1	3	3	3	3	1	1	1	3,14	
Cleura	SE	5	3	5	4,6	1	1	1	3	3	3	1	5	2	3,09	
Stackit	DE	5	5	5	5	5	1	3	3	3	3	3	5	3,5	3,925	Good - but expensive and no HPC tools
Civo	UK	3	3	5	3,6	5	3	4	3	3	3	5	5	5	3,89	Good - but UK is gray area
Cyso	NL	5	3	5	4,6	3	1	2	3	3	3	1	5	2	3,24	

Category & Weight	Specific Metric	Score: 1 (Poor / High Risk)	Score: 3 (Acceptable)	Score: 5 (Excellent / Low Risk)
1. Corporate Profile (40%)	Jurisdiction (50%)	Headquartered outside the EU (Subject to US CLOUD Act, etc.).	EU datacenters, but foreign parent company.	100% EU headquartered and operated (e.g., Gaia-X compliant).
	Infrastructure (20%)	Reseller/wrapper of another cloud provider.	Mix of owned and leased facilities.	Fully owns and operates their own datacenters and network.
	Track Record & Disasters (30%)	Recent major unmitigated disasters; poor transparency.	Has had incidents, but communicated and recovered adequately.	Exceptional uptime history; transparent post-mortems for any minor issues.

2. HPC Ecosystem (15%)	Hardware Ceillings (50%)	Low RAM caps; only predefined VMs; outdated or unavailable GPUs.	Standard modern hardware; basic GPU availability, able to define VM stats ourselves.	Access to latest-gen GPUs (H100/A100) and massive RAM nodes (1TB+).
	HPC Readiness (50%)	"Bring your own everything"; zero HPC tooling.	OpenOnDemand articles and experience from cloud providor	Managed job schedulers (Slurm), pre-configured ML/HPC environments.
3. Internal Services 20%)	Uptime SLA (25%)	No financial SLA, or sub-99.9% guarantee.	99.9% SLA with standard service credits.	99.99%+ SLA with aggressive financial penalties for downtime.
	Backups & Snapshots (75%)	Manual backups only; requires custom scripting.	Native full backup of disks, but limited scheduling and restore availability.	Fully automated, granular VM-level backups and one-click restores.

4. Pricing & Efficiency (25%)	Standardized Benchmark (75%) *(Calculate exact cost of the 3 VM sizes)*	Top 25% most expensive among evaluated vendors.	Middle 50% average market pricing.	Top 25% most cost-effective among evaluated vendors.
	Burstable / Idle Savings (25%)	You pay 100% for the allocated hardware 24/7.	...?	Excellent burstable instances; deep discounts for stopped/idle VMs.

SELECT TOP 3 TO SCORE IN DETAIL!

- Selected **OVH Cloud**, **Hetzner** and **Open Telekom Cloud** to final round.
- Created 51 requirements to evaluate in detail:
  - Compliance and data protection.
  - Security controls.
  - Compute and storage.
  - Networking.
  - Reliability and continuity.
  - Operability and automation.
  - Vendor and exit terms.

SCORING

- Primary plus OPTIONAL requirements
  - Requirements, weighted 1/3/5, maximum 4 points each.

Provider	*Weighted score*	*Maximum*	*Normalised*
Hetzner	415	608	68.3%
OVH Cloud	460	608	75.7%
Open Telekom Cloud	525	608	86.3%

  - Cost model based on approximation of what we want to run

Provider	*Monthly*	*Year 1*	*Three-year*	*vs. cheapest*
Hetzner	€1,192.14	€14,305.68	€42,917.04	+€17,547.48 (+69.2%)
OVH Cloud	€704.71	€8,456.52	€25,369.56	— (cheapest)
Open Telekom Cloud	€2,427.95	€29,135.40	€87,406.20	+€62,036.64 (+244.5%)

EASY SELECTION

- We went with **OVH Cloud** because
  - Nearly the same score as Open Telekom Cloud.
  - Hetzner had problem with encryption of data on rest. Some of the reason for the low score. That was a disqualifier pr requirements.
  - Network ingress and egress on OVH Cloud is free. Part of the cheaper price
- Plan
  - Movement of services as we speak.
  - Done before new year.

SECURING THE PLATFORM

- OpenBao
  - Identity-based secrets and encryption management system.
  - Secure passwords, API keys, and credentials by design.
  - Linux Foundation project - https://openbao.org/.
- WireGuard
  - Secure connection between our instances on/off premise.
  - Efficient performance with reduced latency.
  - High security.
- CVE scanner
  - Daily automatic running CVE scanner for our services.
  - Other automatic scanners like goaudit, npm audit etc.

SERVICES

- Available
  - Consultant on HPC and data
  - AU Overleaf
  - AU TECH Whisper
  - AU TECH LogSearch
  - AU TECH TimesFM
- Coming
  - AU TECH Dataverse.
  - AU TECH AI Lab (code name Paddy).
  - AU TECH Applied AI (code name Stampe).

AU TECH OVERLEAF

- Overleaf, huh?
  - Browser-based LaTeX editor for writing scientific documents (and other). Google docs for LaTeX.
- Why an AU version?
  - Data should be stored on AU-controlled storage, under EU law.
  - It follows our core principles: *AU services first and data stays home*.
- What we build?
  - An AU version of Overleaf, based on the open-source Community Edition.
  - Added track changes and comments for real co-authoring.
  - Isolated compile: each compile runs in isolation.
  - Templates, IEEE, ACM, and AU templates for Ph.D, master, project etc.
- Coming next
  - Offline editing, using browser storage as a local cache that syncs when you're back online.
  - WAYF and MyAccessID login, so you can collaborate with researchers at other Danish and European universities.
- Known gap
  - Invitations for external collaborators are currently outside the control of our identity provider.

AI MODELS ON YOUR LAPTOP!

- Why?
  - Safe compute sandbox. Uses the browser's sandbox with access to CPU and GPU
  - Easy install
  - It follows our core principles: *AU services first and data stays home (it never leaves your laptop).*
- How?
  - Utilize WASM
    - Program it in Rust, C++, C, Go or what you would like.
  - Thin UI layer in JavaScript or TypeScript with no npm installs.

WILL WASM WORK FOR ME?

- Near-native speed in the browser.
- Compute close to the user, with less server load.
- Data never leaves the device.
- Reuse C/C++/Rust libraries.
- Sandboxed and secure.
- Portable across browsers, operating systems and servers.
- Offline-capable.
- Wasm computes, JavaScript handles the interface.
- Small, fast and cached.
- Open standard, reproducible, no lock-in.

- No DOM access: JavaScript glue tends to grow.
- Overhead from crossing between wasm and JavaScript.
- Large binaries and slow first load.
- Memory limits (4 GB, can't shrink). Browser sandbox.
- Threads and SIMD are hard to do.
- Timeout Detection and Recovery for Windows/Linux. Not on Mac. 2s limit on GPU commands.
- Debugging is harder.
- More complex build toolchain.
- Speed depends on the user's device.
- Browser storage isn't guaranteed.

WHAT HAVE WE BUILT?

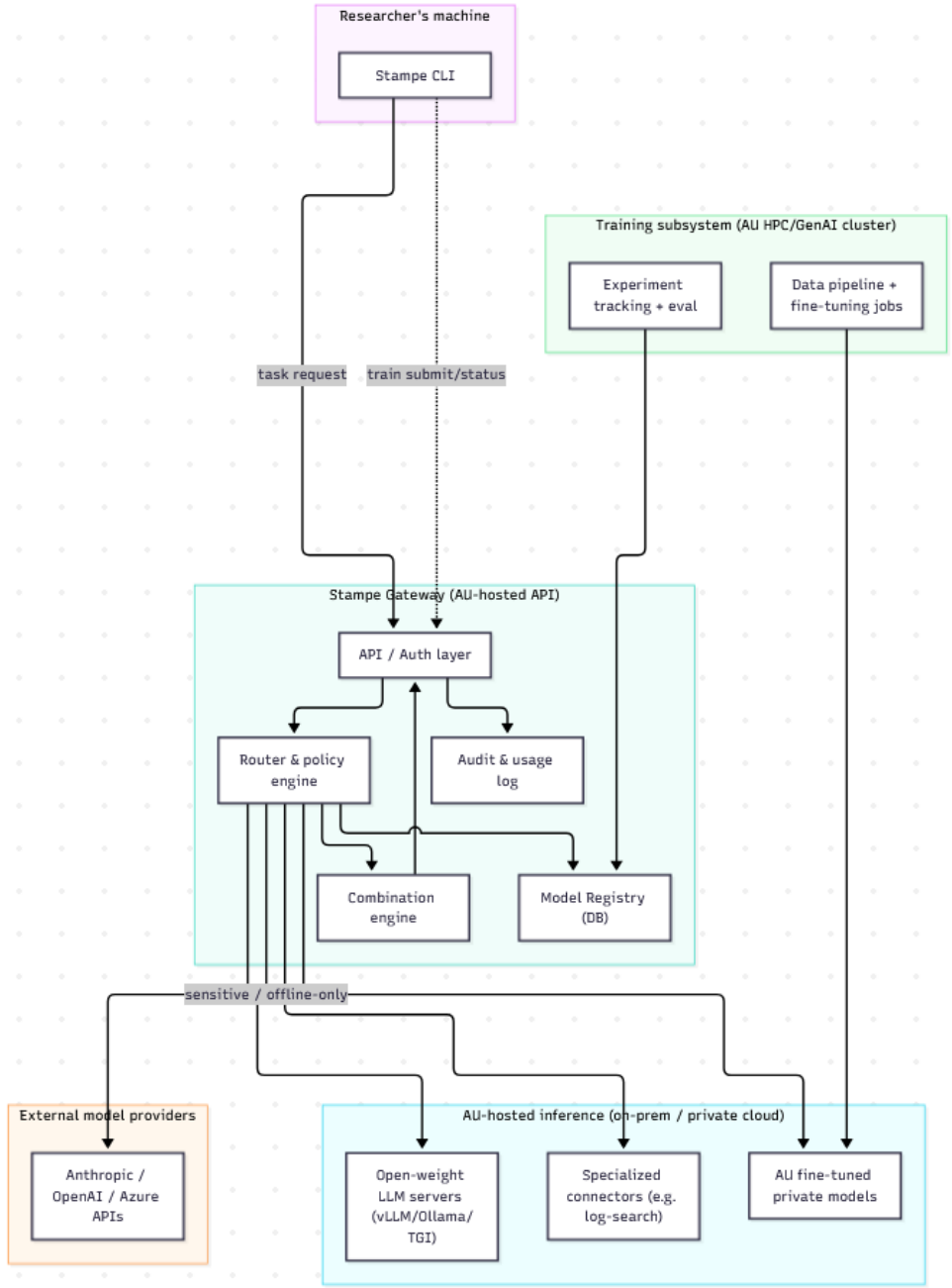
- AU Whisper
  - In-browser speech-to-text using GPU.
  - OpenAI's Whisper model.
  - Upload an audio or video file and it is transcribed locally on your machine.
  - Experimental live transcription in real time from the microphone.
- AU LogSearch (Browser Native Log Search)
  - Utilize AI models to search and find patterns in log files.
  - Search and analyze large log volumes fully in-browser, accelerated by GPU.
  - No log data ever leaves it, so it is safe for sensitive production logs.
- AU TimesFM
  - AU TimesFM (Time Series Foundation Model) is an adaption of Google Research.
  - AI model for time-series forecasting.
  - Covariate Support: It accepts both historical data and known future events (such as holidays, weather patterns, etc)
- DEMO it? Go to https://portal.trit.au.dk/tools/ (QR code later)

WHAT ARE WE RELEASING NEXT?

- AU TECH DATAVERSE
  - Local instance of Dataverse for public data
    - Until DeiC's Dataverse legal status is clear.
  - It is public data with data on AU ERDA system.
  - Rollout late 2026.
- AU TECH AI Lab (code name Paddy).
  - Student/research access to many different AI machines.
    - Nvidia GPUs, AMD GPUs, Intel GPUs, and Mac MetalGPU.
  - Build on OpenOnDemand with SLURM for user apps.
  - Proxmox VE for hypervisor with clean VMs on machines. Fast and safe.
  - FreeIPA + keycloak for IDMs etc -> integrate with MyAccessID or WAYF. Integrate with OpenBao.
  - Shared filesystem with Ceph for native S3 support and persistent user/project folders.
  - Rollout late 2026.

WHAT ARE WE WORKING ON? (ONE OF MANY THINGS)

- Codex/Claude/Cursor/CoPilot/etc alternative for AU TECH researchers
- Model for apply own data to own hosted models at AU
- Run multiple models in parallel. Chain models to refine on work
- Add your own model (specific for your field of research) for your colleague to use
- CLI/Browers interface like in Claude/Codex.
- This system have the code name “Stampe”.



THANK YOU!




SOLIDUM PETIT IN PROFUNDIS
UNIVERSITAS ARHUSIENSIS