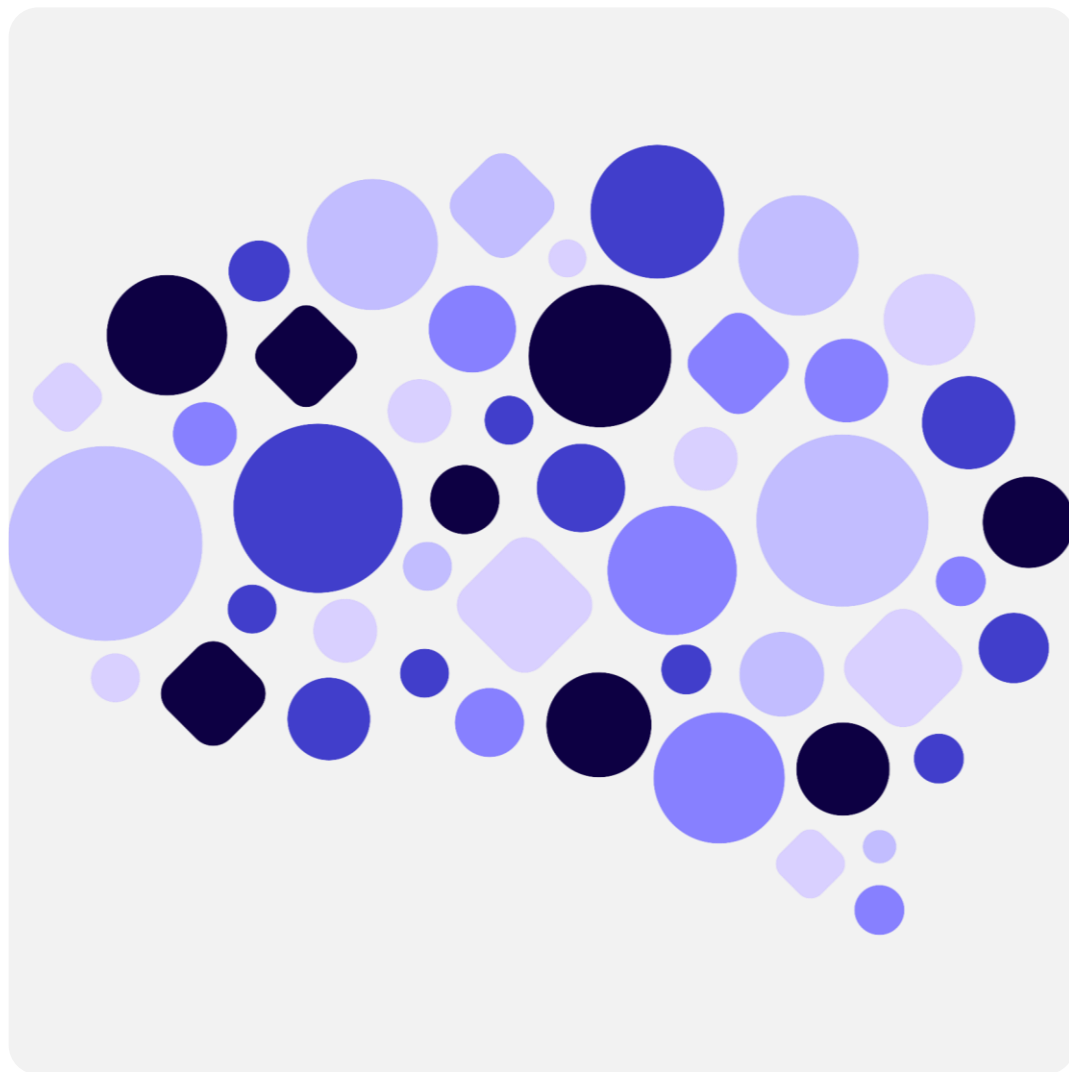


The Norwegian Research Information Repository (NVA)

From Fragmentation to National Infrastructure

Jan Erik Garshol and Anders Eivind Bråten — Sikt



Sikt - Norwegian Agency for Shared Services in Education and Research

- A government agency under the Ministry of Education and Research
- Serves 200+ institutions: universities, university colleges, health trusts, and research institutes
- One-stop shop: network infrastructure, identity & access, student systems, research information, cloud, and library services — all under one roof
- Closest comparators: SURF (Netherlands) and CSC (Finland) — but Sikt's government-agency status gives a stronger universal mandate
- Neutral and non-commercial: Sikt does not compete with member institutions; its sole purpose is to reduce their costs and increase their capability

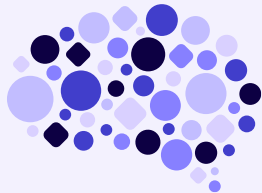


Sikt - Organisation

Educational Services



Research and Knowledge
Resources



Data and Infrastructure

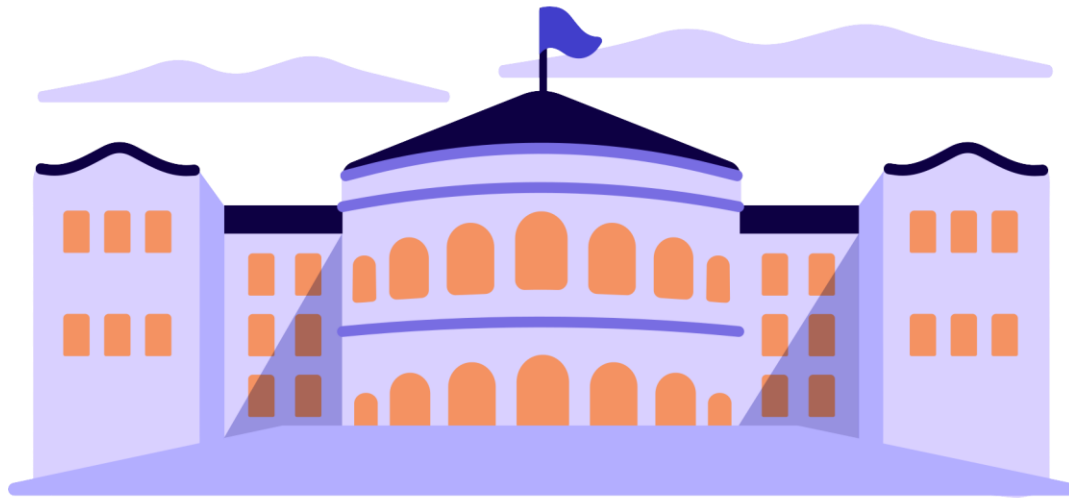


Joint Services



Joint technical services





The Mandate: One Platform, One Data Model

- Ministry of Education & Research:
'Publicly funded research must be publicly accessible'
- 2019 white paper recommended a national master source for research information
- Goal: replace Cristin + 80+ Brage repositories with a single, coherent infrastructure
- Sikt commissioned to design, build, and operate NVA
- Political anchoring at ministerial level
— the single most important sustainability factor

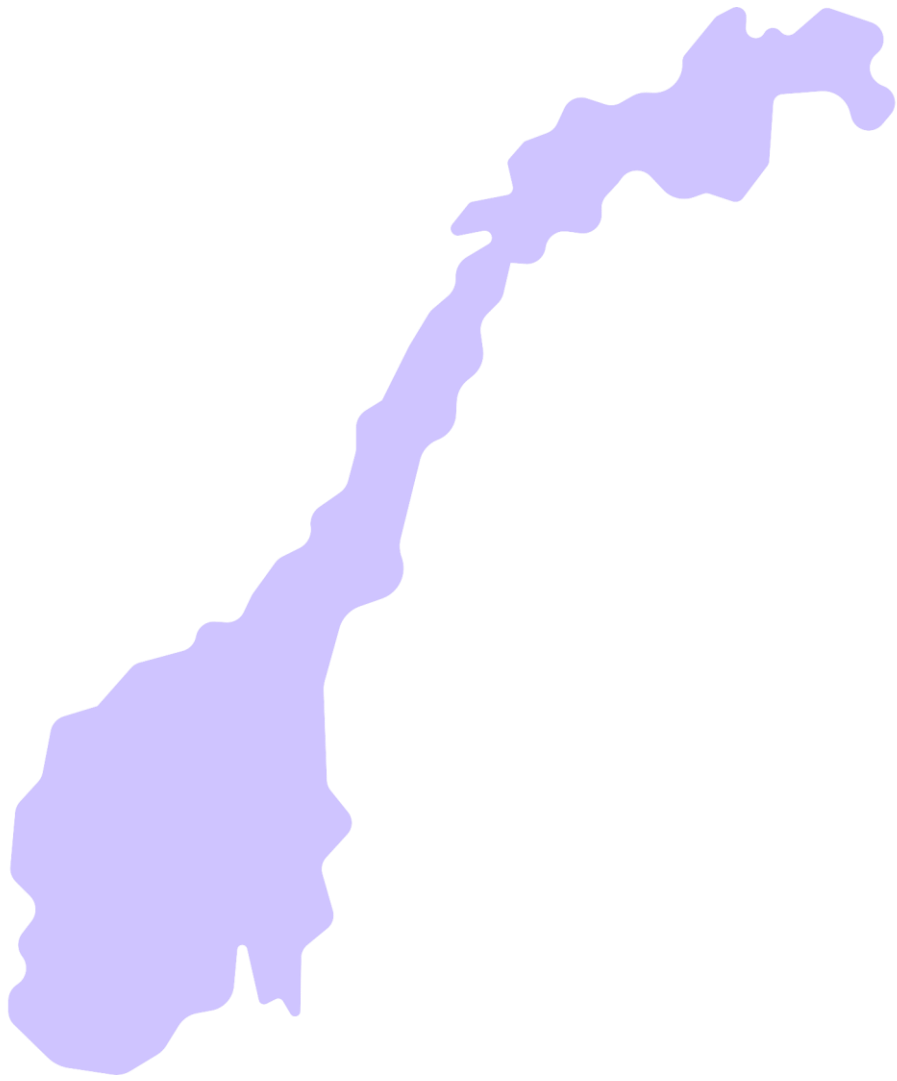




A Fragmented Research Landscape — Before NVA

- CRIS + 80+ separate institutional repositories independently managed
— overlapping scope, duplicated data
- Fragmentation led to variations in metadata standards and practices, inconsistencies in data completeness and quality, parallel rights clearance, and significant administrative overhead
— increased risk of discrepancies, faults and errors when creating reports about academic publishing
- Uneven support for open access across sectors
- Goal: Streamline open access, reduce duplication and cost, improve national statistics



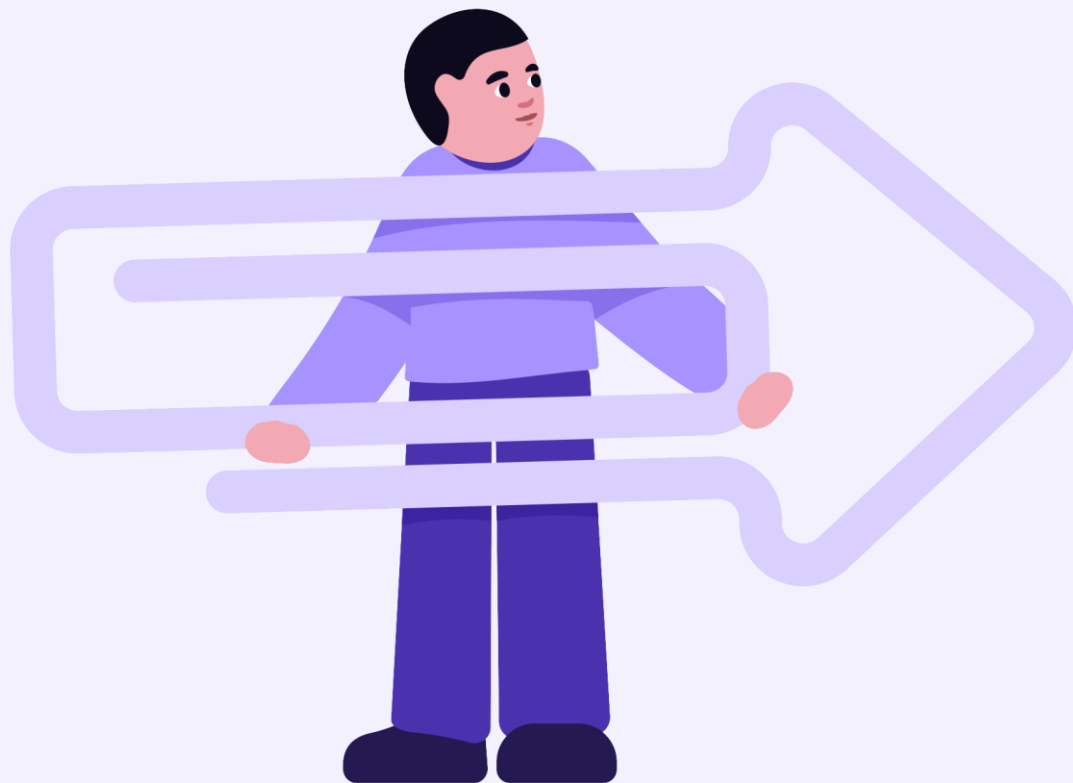


NVA Today: National Scale in Numbers

- 160+ research institutions onboarded
 - 2.5+ million records after migration and deduplication
 - 200,000+ unique accesses per day at peak — 20× original estimates
 - 7 years of iterative development, 40+ contributors
-
- ~30,000 researchers (~70% of all Norwegian academic researchers) are actively using the platform
 - ~3,500 privileged users across curator, editor, and admin roles

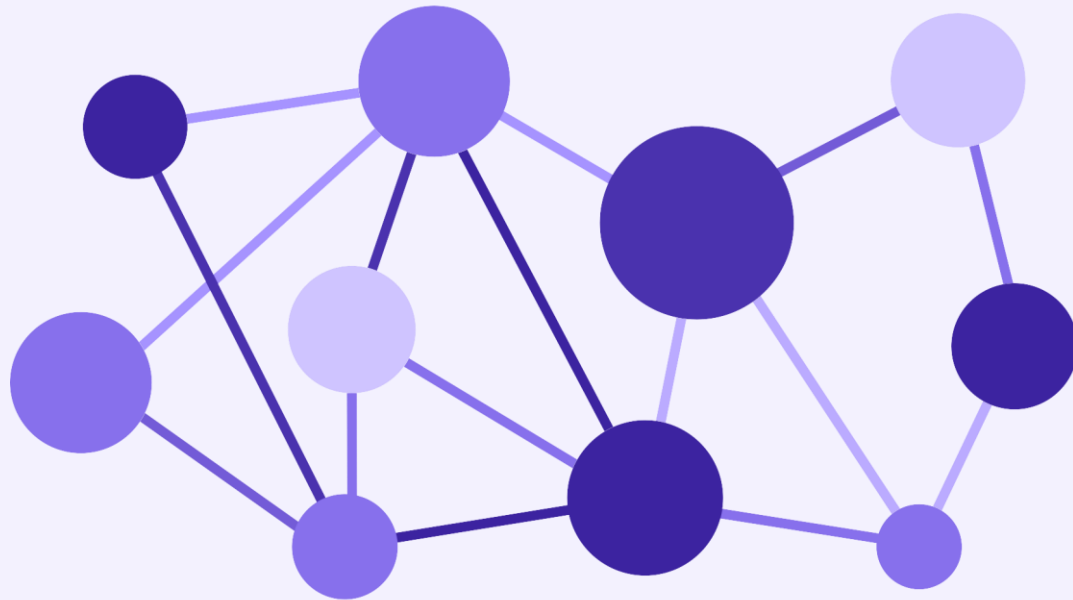


The Platform



A Six-Year Journey: Key Milestones

- 2019: Ministry white paper published
— mandate established
- 2020: First working version of NVA launched
- 2020–2024: Iterative development; domain-expert working groups formed
- Late 2024: Migration from Cristin and 80+ repositories begins
- September 2025: Full-scale production launch across all institutions in multiple sectors
- April 2026: First national reporting to the Norwegian Scientific Index (NVI) done



Technical Architecture: Cloud-Native and Open

- Fully serverless on AWS: Lambda for compute, DynamoDB + S3 for storage
- All infrastructure as code
225 microservices in 16 deployment groups
- Primary AWS region is Ireland
Stockholm is backup with disaster recovery
- Scales to zero cost when dormant; scales to 200K+ daily accesses when needed
- MIT licenced source code on GitHub
- 683 e2e-tests (3×/day) · 13,000+ unit tests (every commit) · 100+ API tests — 10–40 commits/day to the codebase, deployed to production daily
- No downtime from own codebase — except scheduled maintenance outside office hours



FAIR Principles in Practice

- Findable: DOI + Handle for all outputs; public search portal; OAI-PMH → OpenAIRE and other
- Accessible: HTTPS, no login for reading; open-access mandate enforceable via NVA
- Interoperable: DataCite + schema.org; JSON-LD APIs; ROR for institutions
- Reusable:
 - Every metadata record carries an NLOD licence; all attached files carry a CC licence
 - Full provenance metadata; Who, When, Where
 - Semantic graph enables bibliometric queries
 - PIDs enable joins with 3rd part metadata



The Identifier Backbone: PIDs as ***Core Infrastructure***

- DOI — identifies published outputs;
enables duplicate detection and automated import
- ORCID — disambiguates researchers;
bridges HR systems and publication metadata
- ROR — standardises institutional affiliations across all
metadata and APIs
- Handle — assigned to all NVA records;
stable links independent of reorganisation
- Key principle: PIDs designed in from day one
— not retrofitted later



The HR Integration Pipeline

- Scopus batch:
author = name string + affiliation string
— unreliable without PIDs
- Layer 1 — ORCID/ScopusID maps author
→ researcher profile
- Layer 2 — maps affiliation string → institution
- Layer 3 — HR system maps researcher → org unit;
handles career mobility over time
- Layer 4 — manual workflow as needed
(stores ScopusID on researcher profile)
- Critical insight: the researcher connects their own
ORCID to their NVA profile via a double-authenticated
handshake

Results & Lessons



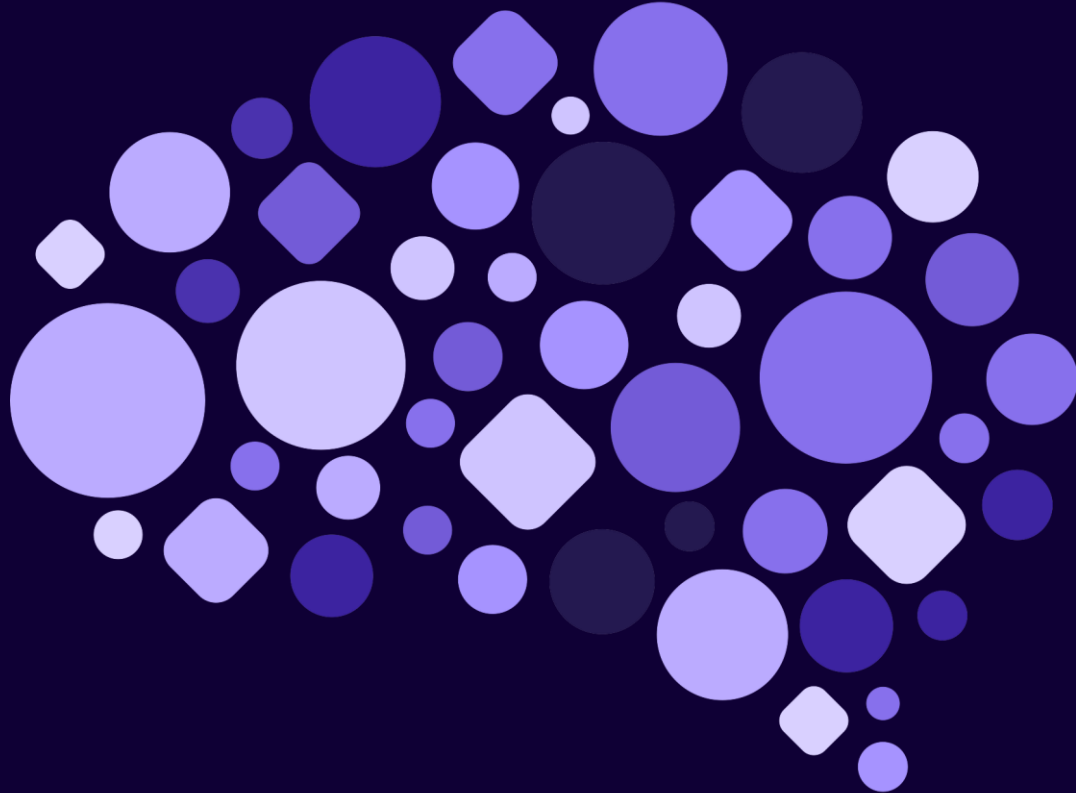
Result: Open Access Mandate Achieved

- Before NVA: Ministry hesitant to mandate open access — no shared infrastructure existed
- After NVA: Mandate in force across 160+ publicly funded research institutions
- Automated file and licence download — no manual intervention for compliant outputs
- NVA did not just document research — it enabled national open-access policy



Result: Automation, Quality, and Shared Capacity

- Metadata from publishers largely replaces manual registration
- One shared metadata record → single best practice emerges across all institutions
- 2.5M records deduplicated after migration from predecessor systems
- Smaller institutions leverage Sikt's advisory role and shared infrastructure
- Artistic research: a previously underserved community now properly documented



Sustaining a 7 Year Project: What Worked

- Ministerial mandate
→ political durability
project protected across budget cycles
- Continuity core (4 people, full lifecycle)
→ institutional memory always preserved
- Working groups
→ distributed ownership
stakeholders invested in outcomes
- Iterative delivery
→ visible progress sustained confidence with funders
and governance
- Normalised team turnover and rigorous
documentation
→ velocity maintained across 40+ contributors



Four Key Lessons for National CRIS Builders

- Collaboration between developers and domain experts is indispensable
- Persistent identifiers are core infrastructure — DOI, ORCID, ROR, and Handle must be designed in from day one
- HR system integration enables scalable metadata quality — without it, assignment is error-prone
- Open source builds trust — institutions can inspect what processes their data; community can contribute and fork



Future Directions and Open Invitation

- Clinical trials reporting: automated import from EU CTIS and Norwegian REK — mandated by Ministry of Health
- Richer external APIs: enabling third-party reuse, ecosystem integrations, new bibliometric services
- Expanding research output types: source code now supported as a first-class output type, with more to follow as the platform matures
- NVA is fully open source — MIT licence, all code on GitHub — contributions welcome