A Brief Description of the Development of the Norwegian Research Information Repository

Jan Erik Garshol[1] and Anders Eivind Braten[1]
[1] Norwegian Agency for Shared Services in Education and Research (Sikt), Norway
jan.erik.garshol@sikt.no, anders.braten@sikt.no

Abstract

The Norwegian Research Information Repository (NVA) is a national, open-source infrastructure for documenting current Norwegian research activities and results and underpins national reporting to the Norwegian Scientific Index (NVI). It is deployed as a service-oriented, serverless platform on AWS, using a graph data model aligned with FAIR principles and persistent identifiers. The platform has enabled the Norwegian government to effectively mandate open access for publicly funded research and provides institutions with shared tools, workflows, and best practices for research documentation. The paper describes the historical context, development process, technical architecture, achieved results, lessons learned, and future work, offering practical guidance when developing national-scale research information infrastructures.

1 Introduction

The Norwegian Research Information Repository (Nasjonalt Vitenarkiv, NVA) is an open-source, unified national platform for documenting contemporary Norwegian research activities and results. It is a curated system for collecting, managing, and sharing metadata and, where permitted, the full content of research outputs. The repository serves as a centralized infrastructure that aggregates information on institutions, projects, researchers, employment, funding, approvals, publications, datasets, and a wide range of other research output types.

NVA is the infrastructure through which the Norwegian knowledge sector, which comprises all higher education institutions, the health sector, the research institute sector, and other research-producing organizations, documents scholarly publishing. It is used to collect, curate, analyse and report the annual Norwegian Scientific Index (NVI) reported to the Ministry of Education and Research, the Ministry of Health and Care Services, and the Research Council of Norway.

In this paper we first describe the background for the project. Next, we explain how we developed NVA and give a description of the platform's technology stack. Then, we list some of the results we have achieved so far. Finally, we discuss lessons learned and provide some practical advice for others who might be planning similar national scale infrastructures.

2 Background

The Norwegian Ministry of Education and Research requested the development of NVA, based on the recommendation put forth in the 2019 white paper "Rapport fra utredning av nasjonalt vitenarkiv" by UNIT [1]. The first working version of NVA was launched in the summer of 2021 and underwent iterative improvements until September 2025, when it was deemed ready for full-scale production across over 160 research institutions. Subsequently, data from the former national Cristin system and approximately 80 institutional repositories were migrated to the platform, resulting in a consolidated dataset of approximately 2.5 million records after duplicate removal.

Until autumn 2025, most universities and research institutions relied on the Brage service (DSpace) delivered by the Norwegian Agency for Shared Services in Education and Research (Sikt) as their institutional repository and used the national Cristin service as their CRIS system. A few universities operated their own DSpace-based institutional repositories. National reporting and data aggregation depended on a combination of centralized and decentralized data submissions across these separate infrastructures. This resulted in extensive duplication of information, inconsistencies in data completeness and quality, and significant administrative overhead. By consolidating previously fragmented information flows into a single infrastructure, NVA aims to improve data quality, reduce duplication effort, increase the quality of research evaluation and strengthen Norway's integration into the global research information ecosystem. This approach is in line with current research from Mabile et al. [2].

3 A Brief Description of the Development Process

NVA was developed by Sikt as a national, service-oriented infrastructure through a collaborative process involving research institutions and national authorities.

The investigation and requirement specification phase, documented in the 2019 white paper from UNIT [1], focused on how to establish 1) a national master source for research information, and 2) a joint information architecture and joint data architecture for the Cristin system and the institutional repositories. NVA was thus designed to rationalize and standardize the collection of research information across institutions, to support national research policy and evaluation, and to promote open access to publicly funded research, as recommended by de Castro [3]

Functional specifications were derived from the national CRIS practices, national reporting needs, and open-science policy requirements. These specifications were then translated into modular services for registration, curation, aggregation, and exposure of research information.

The initial design and development were guided by over 20 years of hands-on domain knowledge and the 2019 white paper, which outlined the goals as requested by the Ministry. As development progressed, several working groups were established 1) to address the need for broader user involvement and 2) focus on specific topics, including artistic development work, research data, administrative tasks, issues related to clinical trials, etc. Each working group consisted of 3–6+ members and met regularly to refine requirements and ensure the platform addressed user needs.

Currently, the core development team consists of 9 people, covering all relevant areas of skills and expertise. Over the 6–7 years of development, more than 40 people have been part of the team, with a core of three members that has been part of the project since the beginning. The development process followed an iterative, requirements-driven approach rather than a single, monolithic implementation.

4 Technical Description of NVA

Since the launch of NVA in September 2025, it has been in a state of continual development. The solution is built on a fully serverless architecture using Java and AWS, leveraging AWS Lambda for compute and DynamoDB and S3 for storage. The data model is a fully semantic, standards-compliant graph, materialized into the required representations as needed. It is influenced by DataCite, CERIF and schema.org and fully aligned with the FAIR principles and compliant with persistent identifiers, especially DOI, ORCID, ROR, and Handle. The platform is fully scalable and can scale down to zero cost when dormant, ensuring both low cost, and high availability and performance when needed.

The platform consists of over 25 million lines of source code hosted on GitHub under a MIT-licence, organized into approximately 200 independent services grouped into around 20 deployment groups. All infrastructure is defined and managed as code and runs from Ireland, with Stockholm serving as a backup region. Disaster recovery, resilience, and data integrity are integral to the architecture, with fully automated scripts ensuring that backups are stored in secure vaults. Large parts of this codebase are dormant outside Norwegian office hours, while the components serving anonymous users scale continuously to support usage around the globe.

Data ingestion and synchronization from institutional systems rely on standardized APIs and batch import mechanisms, while data validation and curation routines are implemented centrally to ensure consistent metadata quality across institutions. Throughout its development, substantial attention was given to scalability, long-term maintainability, and alignment with established open standards, rather than to bespoke, institution-specific customization.

While we anticipated around 2,000–3,000 users, the current usage is 10–20 times higher, with peaks of up to 200,000+ unique accesses per day. There is no zero cost moments.

5 Results and achievements

The Ministry's main goal with NVA was to ensure that research results funded by public money are publicly accessible. Without a national solution, the Ministry was hesitant to enforce this mandate. With NVA accessible to all research institutions, implementing this requirement has become much easier. This objective has now been successfully achieved.

NVA incorporates a central import service through which metadata purchased from publishing companies largely replaces the need for manual metadata registration. Together with the significant reduction of duplicate metadata records across former institutional repositories, this enables institutions to redirect remaining manpower hours toward improving a shared pool of metadata, thereby increasing quality and saving time. The latest iteration enables fully automated downloading of files and licensing information based on metadata records, where the content is preserved and made fully accessible to end users without any manual intervention at all.

A large proportion of institutions are small, with limited staff formally trained to document research and ensure compliance with laws and regulations. Larger institutions serve as a source of knowledge for smaller institutions. Sikt contributes both as an advisor and knowledge broker, aiding where needed, while the shared data platform facilitates tighter integration across institutions.

Most results registered in NVA use Norwegian Register for Scientific Journals, Series and Publishers (Kanalregisteret) to document publishers, journals, and series. This register classifies publication channels into quality levels that are key to the Norwegian Scientific Index (NVI) reporting, helping to steer research publishing toward quality publishers and highlighting areas of concern where quality standards may not be met.

NVA provides all institutions with easy-to-manage access to mint DOIs for scholarly publications, improving the institutions' publication outreach and visibility. If a researcher cannot locate their work in NVA, it can be easily registered using its DOI. Metadata, files, and licensing information are automatically imported into the registration wizard, allowing the researcher to review, make improvements as needed, and approve publication with a few clicks.

Former institutional repositories allowed a wide variety of documentation practices, resulting in diverse metadata quality. NVA enforces a standardized approach to documenting research across all research institutions, enabling a single best practice to emerge, as all metadata records are shared.

NVA offers support for a range of research types, including artistic research, enabling a previously neglected part of the research community to document its work. With a semantic data model at core, NVA can easily accommodate new types of research outputs and support new workflows as needed.

The joint data platform, with its shared best practices, workflows, metadata, content, fundings and research approvals is key to enabling high-quality research on research, as well as other types of reporting and benchmarking across institutions. Norway is known for its high-quality registry data, which can easily be combined or integrated to highlight important areas of interest.

6 Lessons learned and recommendations

Collaboration between developers and domain experts: The sustainable development of a contemporary research archive depends on structured, ongoing collaboration between developers and domain experts [4]. Integrating domain experts into the Product Team as Product Owners was crucial for prioritizing features, validating user stories, and conducting acceptance testing. This approach ensured clearer priorities and enabled better-informed decisions. Early and continuous involvement of domain experts in CRIS systems and open access publishing is essential. Requirements related to embargoes, manuscript versioning, licensing, and funder mandates are nuanced and difficult to interpret correctly. Collaboration helps prevent misalignment and reduce rework.

Importance of PIDs: Systems that rely on local IDs or names might run into ambiguities and inconsistent mappings, especially in cross-institutional contexts [5]. The use of PIDs enables NVA to integrate into the broader scholarly ecosystem. DOI is critical for identification of published research outputs, avoiding duplicate registrations, and to import metadata. ORCID is vital for disambiguating researchers with similar names or changed names; maintaining continuity when researchers move between institutions; and supporting automated import, deduplication, and author claims of publications. ROR identifiers are used to expose and standardize Norwegian institutional affiliations in published metadata, ensuring consistency and interoperability across systems. The same applies when author affiliations are imported. In addition, assigning Handles for all research outputs ensures that: links to publications remain stable over time, independent of reorganization; citations can reference results without link rot; and transparent representation of versions and more. Thus, the use of PIDs is a foundational part of the infrastructure, built into the core architecture, data model, and user workflows from the start. They are not optional add-ons.

Integration with institutional HR systems for metadata import: Reliable central import of metadata from publishing houses (Scopus) requires accurate mapping from publication metadata to institutional structures. However, manual verification of imports does not scale [6]. Without a robust connection to HR data, automatic matching of authors to institutional researchers can become prone to errors. Tight integrations with institutional HR systems enable automated matching of authors to institutional staff by name and affiliation metadata; automatic assignment of publications to correct organizational units; and better handling of edge cases, like multiple affiliations or changes over time. Integration with institutional HR systems is therefore a key enabler of reliable and scalable metadata import. Achieving this goal requires that repository development be closely coordinated with

institutional data governance, ensuring that HR data is exposed, structured, and maintained in ways that support automated metadata workflows.

7 Future work

There are currently three areas within NVA that are under development or are planned to be implemented in a near future: First, we work on a new national report in NVA, requested by the Ministry of Health, covering all patients who have received treatment in approved clinical trials. Data will be automatically imported from the EU Clinical Trials Information System (CTIS) and the Norwegian national system (REK). Second, we are in the process of refining and improving our external APIs and are working on creating several endpoints which will enable us to share data with third-party vendors. Third, we continuously improve existing functionality, enhance user experience, and extend workflows to meet emerging needs.

8 Conclusion

The development of NVA has transformed a fragmented landscape of institutional repositories into a coherent, national research information infrastructure. By consolidating data flows, enforcing shared metadata practices, and embedding persistent identifiers at the core of its architecture, NVA improves data quality, reduced duplication, and strengthened Norway's integration into the international scholarly ecosystem. Crucially, it provides the technical and organizational foundation needed for implementing national open-access mandates and for supporting reliable NVI reporting.

Beyond efficiency gains, NVA enables new forms of collaboration and capacity building. Smaller institutions benefit from shared infrastructure, the expertise of larger institutions, and advisory support from Sikt, while previously underserved communities—such as artistic research—now have an appropriate place to document their outputs. The resulting joint data platform supports robust "research on research", benchmarking, and policy analysis.

The main lessons learned from the development of NVA are sustained collaboration between developers and domain experts is indispensable; standards-based persistent identifiers must be treated as core infrastructure; and integration with institutional HR systems is vital for scalable, accurate metadata ingestion. Future work on expanded reporting, richer APIs, and improved search will further enhance NVA's role as a national hub for high-quality, openly accessible research information.

References

[1] UNIT. White paper. "Rapport fra utredning av nasjonalt vitenarkiv." (2019).
[2] Mabile, L, et al. "Recommendations on Open Science Rewards and Incentives: Guidance for Multiple Stakeholders in Research". *Data Science Journal,* vol. 24 (2025).
[3] De Castro, Pablo. "The role of Current Research Information Systems (CRIS) in supporting Open Science implementation: the case of Strathclyde." *Informačné technológie a knižnice (ITLib)* (2018).
[4] OECD. "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25 (2015).
[5] Downey, M. "Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context." *Journal of Library Metadata*. (2019).
[6] Dagienė, Eleonora. "Mapping scholarly books: library metadata and research assessment." *Scientometrics* (2024).