

A guide to **CKAN** for universities

Transforming research and
improving data management



Contents

<u>Introduction</u>	3
<hr/>	
<u>Data management for universities: what are the challenges?</u>	5
<u>Researchers</u>	6
<u>Data managers</u>	7
<u>IT support staff</u>	8
<u>Institutional requirements</u>	9
<hr/>	
<u>What is CKAN?</u>	10
<hr/>	
<u>CKAN: Meeting the challenges of university data management</u>	13
<u>Improved data discoverability</u>	14
<u>Improved metadata</u>	15
<u>Controlled access</u>	16
<u>Improved data quality</u>	17
<u>Greater collaboration</u>	18
<u>Interoperability</u>	19
<u>A full picture of your data</u>	19
<u>Support for common data formats</u>	20
<u>Easy customisation and scalability</u>	21
<u>Help to measure impact</u>	22
<hr/>	
<u>Case studies</u>	23
<hr/>	
<u>Link Digital: The CKAN expert</u>	29
<hr/>	

Introduction

Data management is one of the most pressing issues facing the higher education sector.

Like governments, businesses and large not-for-profits, universities are generating, collecting and storing more data than ever before.

- Data on their **students and staff**.
- Data on their **teaching**.
- Data generated by their **researchers**.
- Data on their community **engagement**.

University administrative staff and management need to be able to catalogue and store data from a wide variety of sources, and retrieve it fast, to plan and make strategic decisions for the future.



Academic staff not only need a means of securely storing and organising the data they generate. They need to be able to explore, cite and share research data and its outcomes, internally with research collaborators, as well as with the broader public. Indeed, research funders are increasingly requiring their grant recipients to manage, retain and share their research data for transparency, verification and re-use.



The open source data management software known as CKAN – [Comprehensive Knowledge Archive Network](#) – can do everything that other commonly used higher education data software can do.

The following e-book will examine CKAN's functionality and how it can be used to power the data use in the higher education sector. This includes examples of how Link Digital – a globally recognised CKAN expert – has applied the software to help university clients extract the most value from the data in their possession.

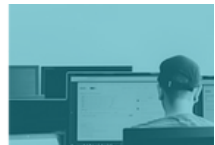
Data management for universities: what are the challenges?

Like many institutions operating within the growing data economy, universities face challenges capturing benefits and actionable insights from their data.

These include:

- Data that is fragmented and siloed
- Inaccurate and outdated data
- Data that is difficult to discover and access
- Poor data governance, which can lead to compliance risks and vulnerabilities
- Data management applications that are difficult or costly to customise or scale up to meet new and/or emerging needs
- Lack of data interoperability

But as large, complex organisations, university operations are also made up of different groups that have different needs in terms of what they might need from a data management system.



Researchers

They require:

- A system that can manage **varied and complex research data** across its entire life cycle, including assistance cataloguing, processing, cleaning and refining datasets.
- **Data sharing in a secure and controlled environment** with research partners and collaborators, then with the wider research community and public.
- **High interoperability.** University research engages not just multiple researchers but a wide variety of different stakeholders, and is often transnational in scope.
- Help to create and maintain **detailed metadata** that includes clear guidance on the provenance, ownership and processing of datasets. For those researchers depositing data directly from scientific instruments into the data management system, the ability to batch input/edit metadata is also important.
- Help **measure research impact** through the provision of standardised citation formats and the automatic creation of digital IDs.

WANT MORE
INSIGHTS?



[linkdigital.com.au](https://linkdigital.com.au/ckan-open-data)
[/ckan-open-data](https://linkdigital.com.au/ckan-open-data)

Data managers

They require:

- Clear data provenance and **metadata that is compatible with other institutional systems**, i.e., library catalogues, to facilitate greater data discoverability.
- **Workflow processes** to manage both formal and ad hoc decision-making. The ability to record and audit all interactions related to the data and metadata they manage.
- A flexible and extensible **permissions system**. The ability to restrict access to datasets for different roles and over different time periods.
- Assistance with the creation of **standardised methodologies** and language to describe, measure and evaluate different types of research data, from raw data collected from the field or in a survey, to scientific data produced in a lab, to qualitative research



IT support staff

They require:

- A **well-designed and modular structure** that is easy to modify, extend and customise.
- Minimum repeated re-architecting of infrastructure, **system implementations**, and **software integrations**.
- Fast, active support for problem solving and troubleshooting, efficient, reliable maintenance and hosting, and comprehensive support documentation.
- **High level security** for the software and the data stored on it.
- Freedom from vendor-imposed constraints on crucial tasks such as **maintenance, support, and upgrades**.
- **Minimal training**/retraining of faculty, staff, and students to use it.



Institutional requirements

They require:

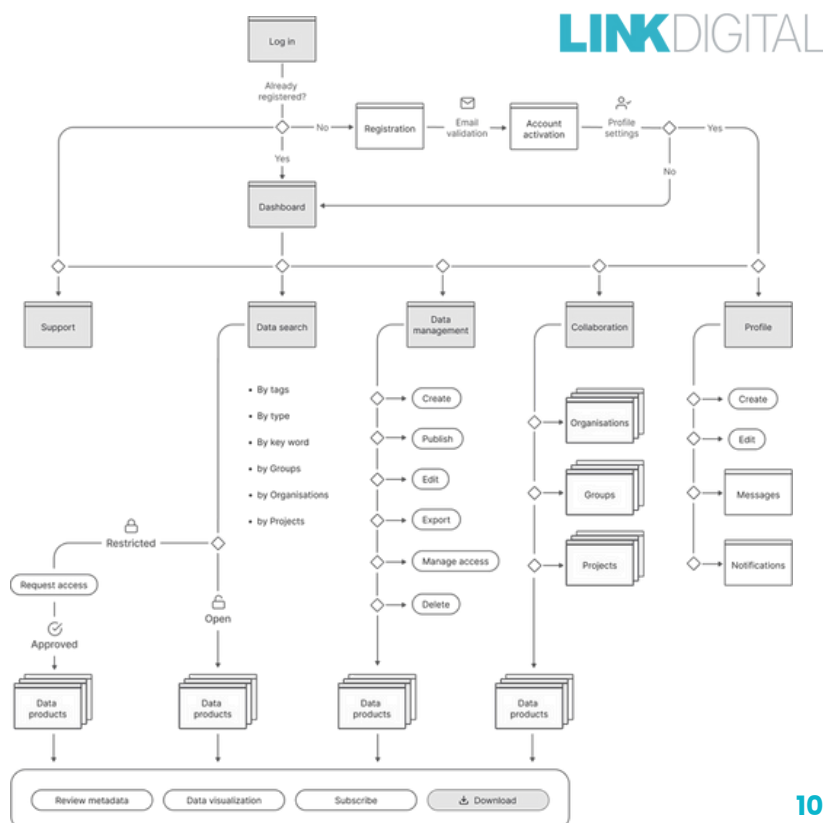
- **Innovative platforms** that meet the needs of staff and students, and support collaboration and promote strategic agility.
- **Control over software lifecycle**, and resilience against vendor driven changes and vendor lock-in.
- Long term **predictability and stability** in terms of budgeting, with minimal or no licensing fees.
- **High level data security** to guard against the potential risks associated with data breaches, unauthorised access and misuse of sensitive information.
- **Interoperability** with existing academic systems, protocols and metadata standards.
- **Data management systems** that can be scaled up easily and seamlessly, without major overhauls and with minimal financial outlay.

Let's look at how CKAN can meet all these needs.

What is CKAN?

CKAN is a sophisticated piece of open source software launched in 2006 that can be configured and set up to function as a public or internal facing data platform or once it is deployed and hosted on a web server.

CKAN's central role in the global data ecosystem is most easily evidenced by the fact that it serves as the key component of some of the largest and most successful open data portals around the world.

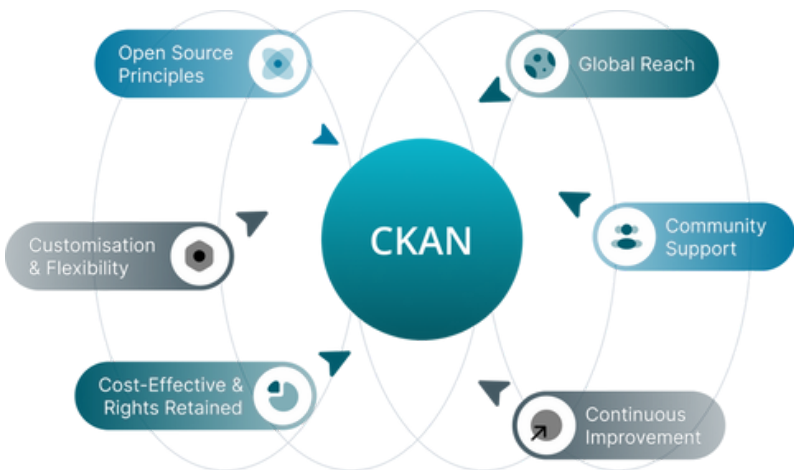


These include the [German Government's open data portal](#), the [United States Open Data Portal](#), [Open Africa](#) – the largest repository of data on the African continent – and [Canada.ca](#), the official website of the Canadian Government. CKAN is also used by the [United Nations](#) and not for profit agencies, businesses and, increasingly, by universities and research institutions.

The non-right restricted nature of CKAN is part of the software's much broader guiding principle – that open data is a public good that should be non-right restricted and easily transferable, and that creates more value the more it is shared. CKAN is non-commercial which allows the project to maintain its open, public-spirited ethos and direction, without being swayed by private financial interests. Indeed, research data management using software such as CKAN can be seen as a continuation of the logic of the Open Access movement.

The fact that CKAN is open source means organisations can avoid license fees and vendor lock-in. It can be customised without restriction to allow data access to be tailored very specifically and appropriately to whatever 'better' access means for an organisation, without any compromise to its core value in terms of machine access. This makes it more accessible and cost-effective. It also means users maintain the rights to their data and metadata.

As well as the support and maintenance provided by third-party vendors, CKAN users can draw on the skills and knowledge of a large global community of users and contributors, including people drawn from private enterprise, academia and the community sector, helps users to keep pace with changes and continuously improve due to the broad range of people contributing to design, development, and defect correction. This is in addition to the efforts of a core development team that contributes to keep the CKAN software in good shape.

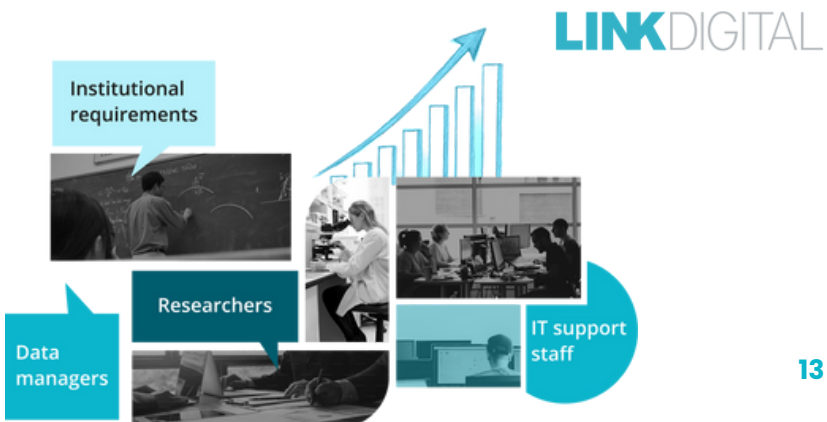


CKAN: Meeting the challenges of university data management

A growing number of universities are utilising CKAN for their data needs, including:

- Open and closed research data catalogues for storing and sharing datasets.
- Catalogues for managing institutional administrative data within universities.
- Portals for student data projects published openly or within courses.
- Inter-departmental portals for collaboration and interdisciplinary data sharing.

Let's look in more detail at how CKAN's functionality can be used in the higher education space.



Improved data discoverability

CKAN can provide a platform for researchers to easily find and access datasets relevant to their work. It makes research datasets more discoverable and reusable, enhancing the visibility of institutional research outputs.

- **Tools to streamline and improve data** publishing, discovery, sharing and re-usage, including powerful search and filtering capabilities.
- **Application Programming Interfaces (APIs)** and faceted search capabilities to allow both humans and machines to access data efficiently. Support for users to search for and discover data assets based on various criteria, such as keywords and tags, across different platforms and domains, reducing the time it can take to locate data and get it ready for use.
- **Assisting researchers** and their institutions to align with the [FAIR Principles](#): Findability, Accessibility, Interoperability, and Reuse of digital data. These principles have gained considerable traction amongst the international research community and adherence to them is now widely seen as a best practice in terms of data management.

Improved metadata

Metadata is vital to organising and extracting insights and value from data. Detailed information about the content, context and ownership of an institution's data also helps to ensure that reuse is carried out properly and adheres to any necessary regulatory and legal frameworks, including data usage rights and licensing.

Comprehensive and consistent metadata is also key to machine actionability – the capacity of computers to locate, access, interoperate, and reuse data with none or minimal human intervention. This includes the ability to connect with various data sources and systems to automatically ingest or update data across different systems and platforms, including legacy systems.



- Consistent, standardised metadata schemas that enable researchers to understand the type of data, how it was collected and what instruments were used to collect it.
- The ability to create a common data dictionary that researchers can use to compare data.
- Customised metadata templates to save time and provide consistent metadata descriptions.
- Support for metadata standards like [Data Catalogue Vocabulary](#) and [ISO 19115](#).

Controlled access

CKAN facilitates cross-organisation and cross-jurisdiction data sharing in a secure manner.

- **Permission settings** to manage who can view, edit or use the metadata and associated data sets, i.e., role-based access control allows different permissions for data publishers, curators, and public users.
- **The ability to make metadata openly accessible** even if the dataset attached to it has access restrictions.
- **Version control and audit trails**, so any changes can be tracked, and data accountability and traceability are enabled.
- **Streamlined approval procedures** for data sharing. These improve transparency and enable data stewards to discern and report on data usage patterns over time and across different levels and functionalities of an institution.

Improved data quality

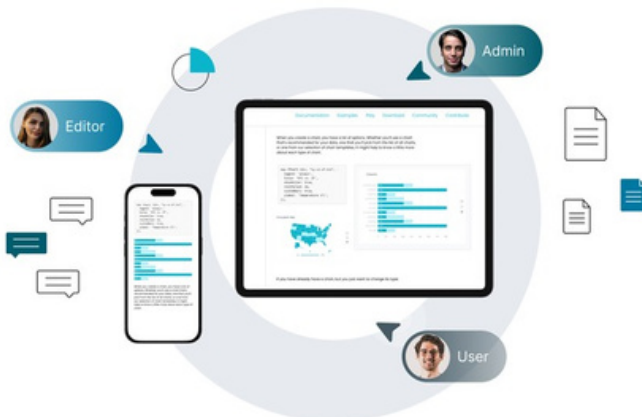
- Assistance to **streamline the collection and dissemination of data**, acting as a single point of truth for researchers.
- Support to **register, update and refine datasets**, improving data quality and ensuring that any research or analysis that uses the data is trustworthy and accurate.
- Help to **track changes to datasets over time**, ensuring that those working on a project are using the most up-to-date data.



Greater collaboration

By acting as a central hub for datasets across disciplines, a CKAN powered portal encourages collaboration and data reuse among researchers and students.

- **User management features** that enable the creation of collaborative spaces, in which researchers and stakeholders from different institutions and across geographic jurisdictions can access and work together on shared datasets.
- **Support for groups** to contribute different data to a shared project, refine datasets and metadata, and explore, independently corroborate, and build upon each other's work.
- **Easy integration** with other research tools and platforms, facilitating seamless data analysis and sharing within research teams, which results in more comprehensive, robust research outcomes.



Interoperability

CKAN powers the interoperability necessary to work in the global research world.

- Integration with a wide range of other tools, learning platforms and institutional research environments.
- Automated data ingestion, harvesting, and synchronization.
- Interaction with all aspects of its open data platform, including extension development, importing data from other services, as well as legacy systems.

A full picture of your data

CKAN can help institutions get a better understanding of their research data beyond acting as a central repository to manage and organise diverse datasets.

- **The ability to link datasets** by project, type, research output, person, etc.
- **Assistance to capture** not just data but research **context, activity and processes**, including rights and license management.
- **Version control of datasets**, enabling researchers and data managers to track changes and access previous versions.

Support for common data formats

- Support for **importing and exporting data in most data formats**, including Excel, CSV, PDF, XML, RDF, etc.

Excel

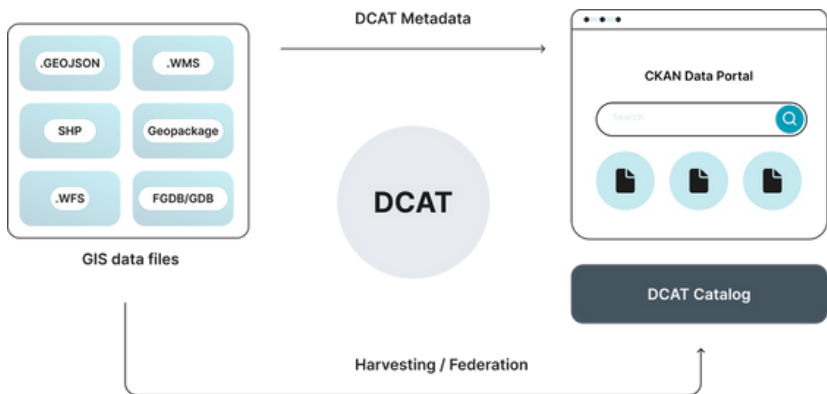
CSV

PDF

XML

RDF

- Support geospatial data hosting**, i.e., data presented in the form of different maps, for example, point or proportional symbol maps, cartograms, as well as standard topographical maps. This includes extensions that allow users to present and explore information from geospatial data or seamlessly integrate with open-source software options that can do this.



Easy customisation and scalability

Structured proprietorial data management solutions not only limit customisation or charge significant fees for feature development. They are inherently unsuited to an environment in which researchers are always adding variables to collect information in a way that makes sense, as is the case in university research environments, or where needs might vary across departments, research programs, and administrative systems.

- **Easily customisation** to meet the specific needs of different institutions, including being scaled up to meet new and emerging needs.
- **Aligned with the funding structure** of higher education, which is mainly grant-based, and thus largely rules out having the budget to fund an expensive, ongoing software licence with a proprietary vendor.
- [A wide variety of extensions](#) can vastly increase its basic functionality and integrate in a seamless manner with a user's existing data and digital operations.

Help to measure impact

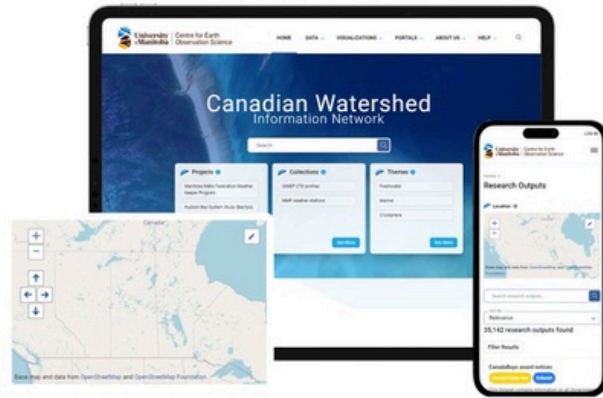
Creating a better understanding of where and how data is used not only helps researchers better comprehend their impact but can contribute to greater transparency and accountability around data use. It also results in researchers receiving credit for their work, further incentivising them to release more data.

- **Assistance to researchers to track citations** and create a citations list for a dataset.
- **Automatic minting of new of DOIs**, which provides a permanent access to knowledge resources, and makes datasets and research findable in other repositories, increasing their discoverability.



Case study

Facilitating greater collaboration around Arctic freshwater data: University of Manitoba



The [Canadian Watershed Information Network](#) (CanWIN) is an open data portal managed by [University of Manitoba's Centre for Earth Observation Science](#) (CEOS), a leading multidisciplinary and collaborative research centre focused on understanding how the earth will respond to climate change. CEOS researchers conduct fieldwork globally, with the Arctic freshwater marine system a particular focus due to the acute impacts of climate change on it.

CanWIN collects and manages a large array of climate and hydrological data, including detailed spatial data, and analyses and shares it with researchers. The original CanWIN data portal was not able to capture and share the complexity of the multi-disciplinary data produced by CEOS's researchers, so Link Digital partnered with them on a new portal design, using CKAN.

The new CanWIN site has improved the ability of researchers and scientists across institutions to collaborate on climate data analysis and share findings for analysis from a hyperlocal to global scale. This has enabled better water management and climate resilience planning.

The CanWIN site also delivers:

- **Detailed metadata catalogue** that enables researchers to undertake far finer grained exploration of datasets and tell far more complex stories about the data used in their work.
- **Greater data storage capability.**
- **Greater collaboration** between researchers,
- **New visualisation tools** to enable researchers to tell more complex stories about the data in their work.
- Added **user permissions** to provide data sovereignty to Indigenous groups to give them granular control over their data.

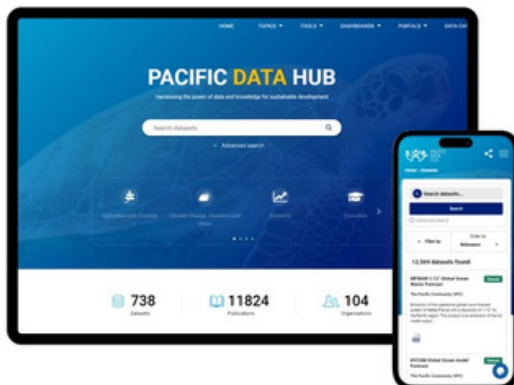
[Read more about Link Digital's work on the Canadian Watershed Information Network.](#)

Case study

Data for more Pacific-centric policy solutions: University of Wollongong

Built by [Link Digital](#) and overseen by the [Pacific Community \(SPC\)](#), the principal scientific and technical organisation for the Pacific, the [Pacific Data Hub \(PDH\)](#) serves as a gateway to the most comprehensive data collection of relevance to the 22 Pacific Island nations and territories. It also pulls in related data from SPC's metropolitan members, Australia, New Zealand, the United States and France, and from international organisations.

Designed to tackle the fact that Pacific related data are hard to locate, fragmented, and held by various stakeholders in multiple geographic locations, the PDH enables the creation of more informed and 'Pacific centric' solutions.



It also acts as a digital platform on which any government, donor, academic, or member of the community can both store and access high quality and reliable data.

One of the many institutions that has worked closely with the PDH is the [University of Wollongong](#). Researchers from the University collaborated with the SPC and regional commodity and country experts to create the PDH's [Food Trade Database](#).

In addition to making quality data on healthy, equitable and sustainable food systems more easily discoverable and shareable, it delivers trade data on food and beverage commodities across the Pacific.

University researchers also worked with the PDH on the first detailed estimates of the [coastal proximity of populations in 22 pacific island countries and territories](#). A collaboration project between SPC, the University and the not for profit research organisation [WorldFish](#), the data is vital data in terms of assessing vulnerability to natural disasters, economic and livelihood issues, and mitigating the impacts of climate change.

[*Read more about Link Digital's work on the Pacific Data Hub here.*](#)

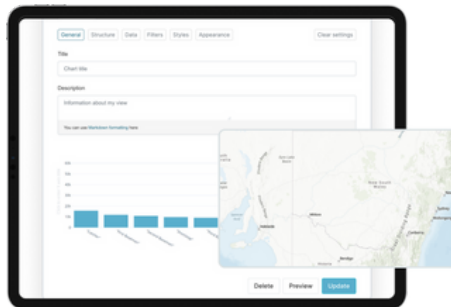
Case study

Helping the design department of a major European University improve internal data discoverability

Link Digital delivered a tailored CKAN data and statistics platform to improve the internal data organisation and search capacities of the design department of a major university in northwestern Europe.

The department delivers a wide range of bachelor's, master's, associate degree, and minor programs in applied sciences, and includes a variety of hands-on educational and research opportunities, many of them taught in English for international students.

Link Digital was engaged to implement a CKAN-based platform for the university's design department.



The CKAN implementation hosts metadata but no datasets and is instead used to aggregate and simplify access to a large array of Power BI dashboards and online resources hosted by the department, making them more discoverable internally within the university. Link Digital was able to create landing pages within the portal, to enable users to search for these dashboards and resources by category, i.e., students, finance, research, personnel, etc.

Clicking on a landing page takes the users to the dashboards and resources specific to that topic, enabling them to navigate back easily to the main portal page, for an improved user experience.

The portal also includes:

- **Integrated single sign-on**, allowing users to log in to the portal once with a single set of credentials and gain access to multiple dashboards and applications.
- **The integration with the department's themed experience/interface.**
- **Multiple CKAN extensions**, built to accommodate different file types, such as PDF, Word, html, CSV and excel, etc.
- **The portal uses the Amazon Web Services Cloud** for hosting, with Link Digital providing support and maintenance.

Link Digital: Your trusted CKAN expert

CKAN can help universities get a better understanding of their data, delivering:

- Improved data **discoverability**
- **Improved metadata**, vital to organising and extracting insights and value from data
- More controlled **data sharing** in a secure environment
- Improved data quality
- Greater **research collaboration** across disciplines and geographic borders
- Greater **data interoperability**
- Easy **customisation** and **scalability**
- Help measuring **research impact**

With nearly a decade of in-depth CKAN experience, Link Digital is one of the few global companies that can provide end-to-end service surrounding open data publishing on CKAN.



Link Digital is what is referred to as a 'co-steward' of the software, serving to support the community of users and contributors that collaborate with the project. Focusing on user centric design, our team of CKAN experts can create a solution that aligns with your institution's unique needs and we can help to maximise CKAN's functionality and versatility by utilising its wide array of extensions and plugins.

Our service suite includes data workflow development, dynamic visualisation systems, and publishing workflows that will allow organisations to control and understand what data is being published to the public, as well as data migration, including complex integrations using the CKAN suite of APIs.

As part of this, we understand that open source data projects are not just about technical or software problems. Attention needs to be paid to the human, cultural, legal and ethical aspects of large scale open data and data sharing projects.



Are you ready to transform your data management?

[Get in touch with us](#) to learn more about how a CKAN-based open data portal or internal data platform can improve data storage and discoverability and facilitate greater research collaboration or more informed decision making at your university.

Get in Touch

