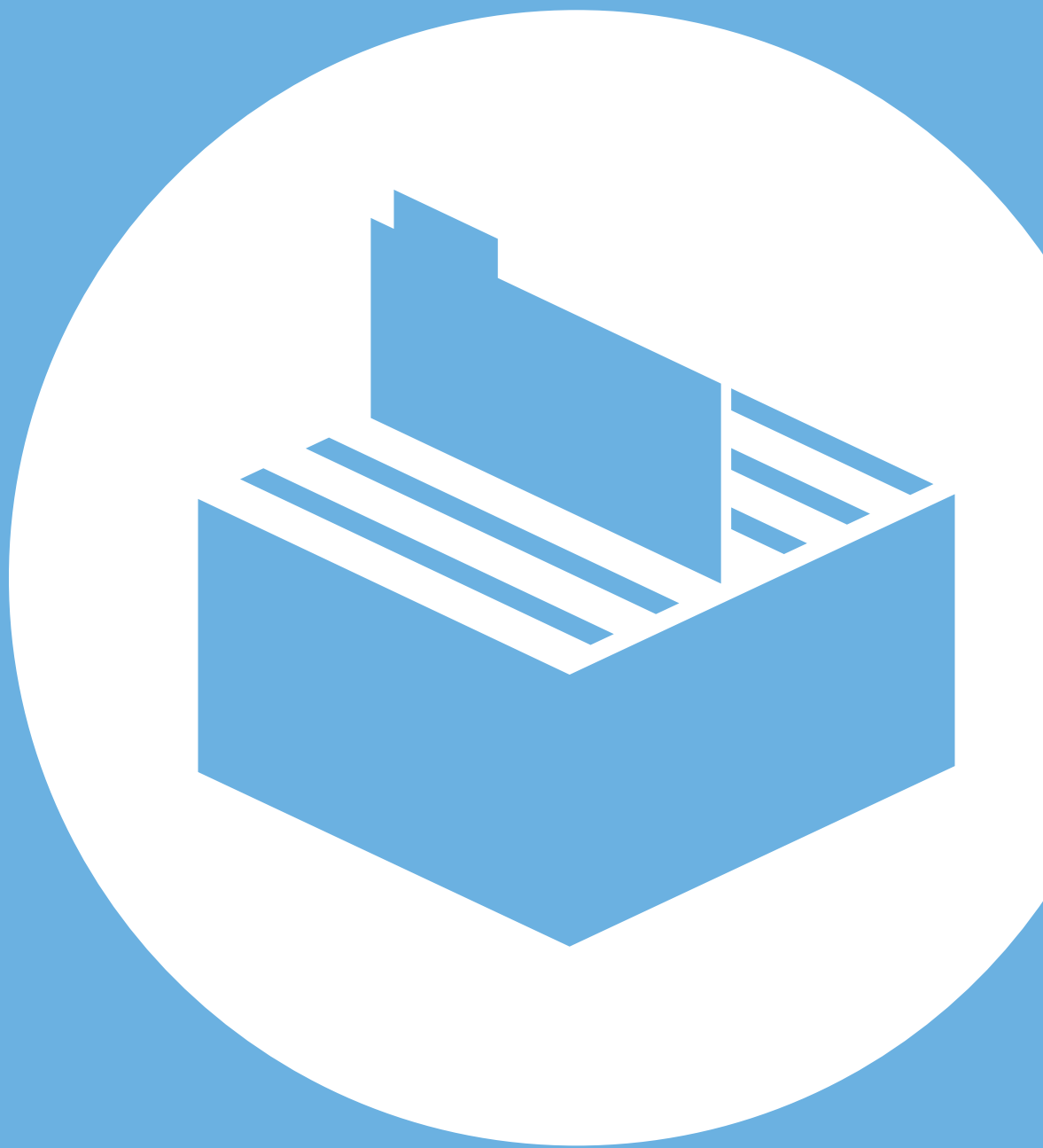


Haplo Repository



Haplo Repository

- 01 About Haplo Repository
- 02 Public Portal
- 03 Researcher Profiles
- 04 Submitting Research Outputs
- 06 Metadata
- 08 Ingest Workflow
- 10 Portfolios
- 11 Finding and Reusing Outputs
- 12 Managing Restricted Files
- 14 Integrations
- 16 Impact and Esteem
- 17 Reporting
- 18 Audit and Version Control
- 19 Data Security
- 20 Questions & Answers



Haplo Repository is a modern open source repository with a flexible architecture that enables your repository to perfectly fit your institution and your research.

Taking an innovative approach to established information management principles, Haplo mitigates common pain points and limitations, enabling institutions to do more with their repositories than ever before.

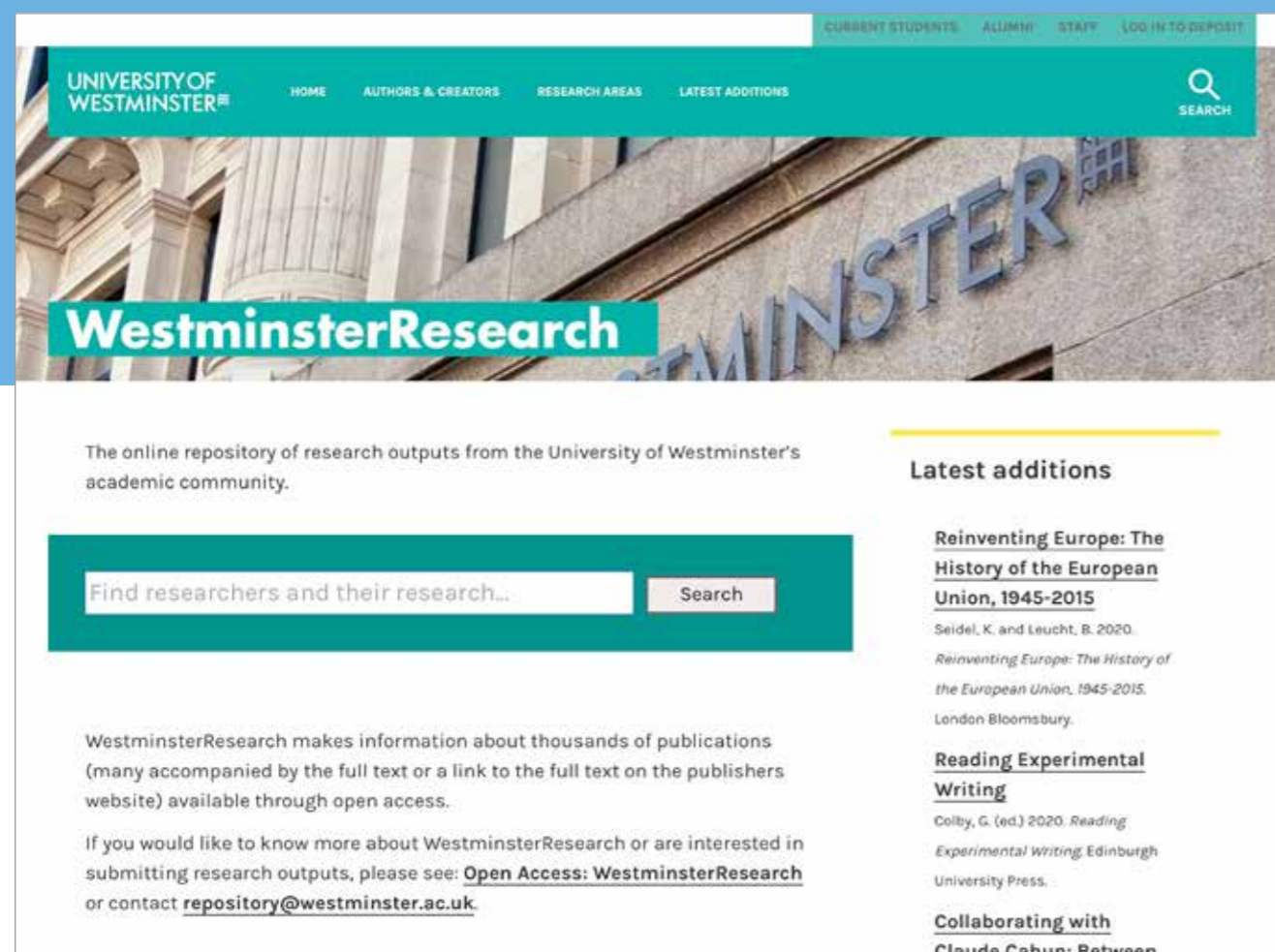
A single repository can manage all research outputs, including traditional scholarly publications, practise based research, and research data. Each output type can have different metadata, workflows and policies.

With integrated workflows and an intuitive user interface, Haplo streamlines processes and offers high rates of engagement with your researchers.

Haplo Repository is offered under the MPLv2 open source license with ISO27001 certified commercial development services, support and hosting available, providing the best of both worlds: a fully-supported modern and professional repository, ww with no lock-in and freedom to develop, manage and self-host the repository in the future.

Public Portal

A public repository displaying all of an institution's research showcases its capability and expertise



Example public repository interface from University of Westminster

Every aspect of Haplo's public portal is customisable, from the kinds of information it displays, to the way that information is structured and presented.

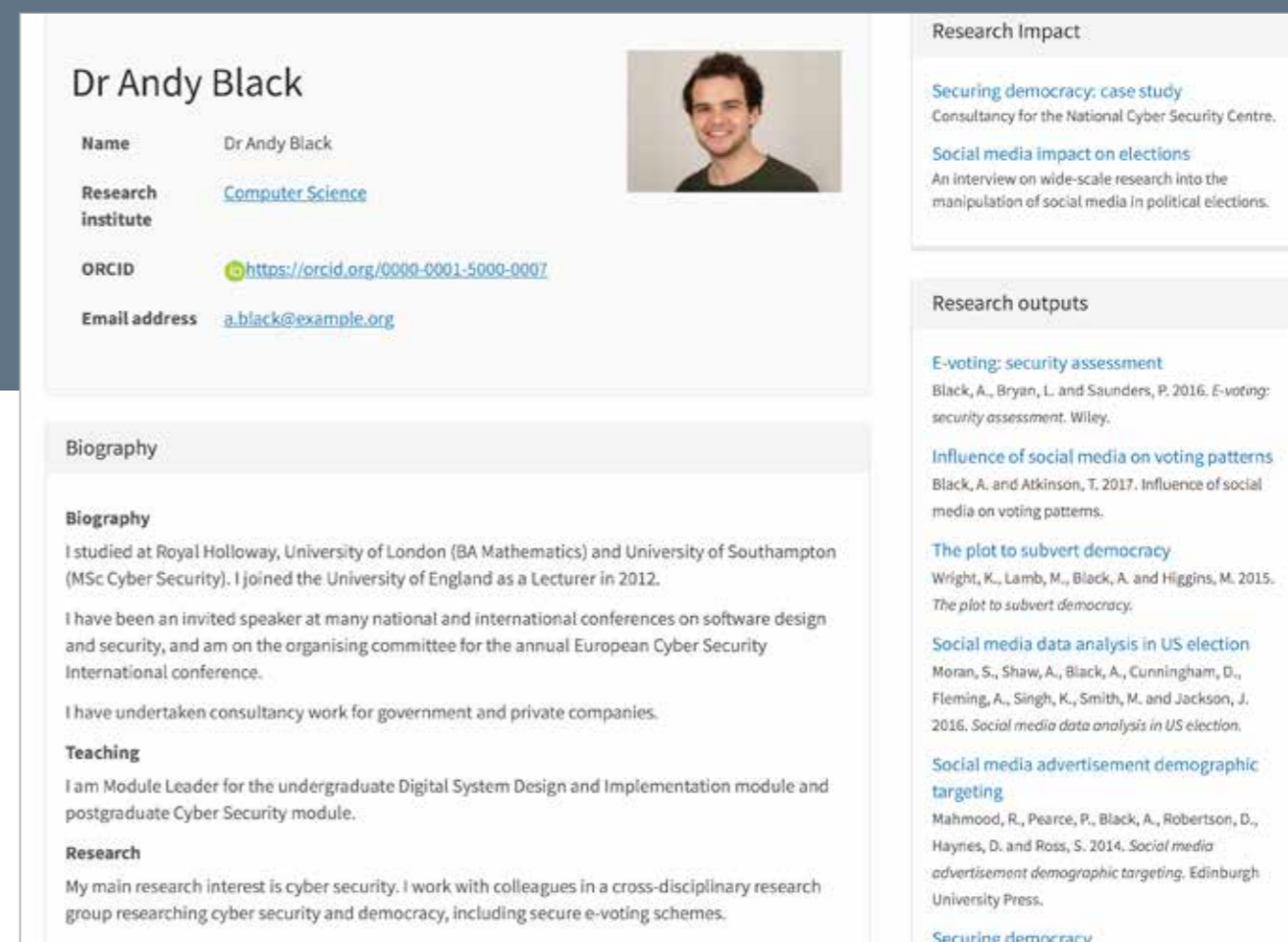
Using open tools, Haplo's flexible toolkit provides institution web teams with full control over all HTML generated, and can modify the HTML, CSS and templates of the public portal to match the look and layout of the institution's visual identity.

Haplo's public research portal can include:

- ✓ A fully browsable and searchable repository of research outputs and research data
- ✓ A searchable directory of researchers
- ✓ Listing of faculties and research groups, and the people and research outputs associated with them
- ✓ Forms and workflows to request access to embargoed or sensitive data

Researcher Profiles

Researcher profiles promote each researcher's interests and achievements



Haplo collates comprehensive, up-to-date profile records for each researcher.

Researcher achievements, interests, publications, and contact details are easily kept up-to-date in a fully searchable public directory.

Researchers can upload their own photos, add social media links and esteem entries. Data held in other institutional systems automatically feeds into Haplo, avoiding manual entry of existing information.

Changes to public profiles can either appear instantly or after approval by a web editor.

Researchers can generate and download an academic CV, with or without institutional branding, incorporating the latest information displayed on their profile.

Researcher profiles key advantages:

- ✓ Researchers can update their own profiles
- ✓ Profiles are automatically populated with data from institutional systems
- ✓ Real-time dashboards show web editors which profiles are up-to-date and complete

Submitting Research Outputs

Researchers self-deposit the outputs from their research to the institutional repository

The screenshot shows a web form for submitting research outputs. The form is organized into sections, each with a plus sign and a minus sign. The sections are: Title (text input), Authors (two text inputs, one with a question mark icon), Abstract (text area), Keywords (text input), Year (text input with a year mask 'yyyy'), Journal (text input), Journal citation (three text inputs: Volume, Number, Page range), Publisher (text input), ISSN (text input with a question mark icon), Digital Object Identifier (DOI) (text input with a 'doi:' prefix and an example '10.1000/182'), and Web address (URL) (text input).

Outputs can be added either via an online form or by looking up and importing existing records by entering their DOI.

“ I’m very impressed with the simplicity of the ingest workflow in Haplo, from both the administrator and end user perspective. ”

Submitting research outputs

Haplo enables researchers to self-deposit their research outputs. Researchers are presented with different metadata records for different types of output so the fields on the form are always relevant to the research being submitted. Embedded guidance notes help improve the quality of information collected.

Data harvesting

Haplo minimises manual data entry by integrating with industry standard systems. This enables researchers to harvest items directly into the repository.

When harvested items are matched to an existing researcher profile, the system prompts the user to confirm or disclaim them.

To speed up metadata quality checks, repository editors are given clear notices in the user interface, indicating that the record was harvested.

Haplo draws publication metadata and citation information from external databases to which the institution subscribes or to which a free licence is available including: Web of Science by Thomson Reuters; Scopus by Elsevier; PubMed; Crossref; ORCID; EuropePMC; Figshare; Altmetric; Mendeley; and JISC Router.

Datasets and non-traditional outputs

Researchers can submit all their research outputs through Haplo. Datasets, practice based research outputs, and traditional research outputs can be stored together, and all outputs related to the same research project are linked together.

Author citations

Haplo creates unique author profiles per person, not per citation.

The user interface enables the author to choose their citation for an output independently of the link to their author profile. This means different outputs can use different citations but be linked to the same researcher profile, and different authors can have the same citation with no ambiguity over which outputs belong to which author.

File types

Any type of file can be stored and managed within the repository, to enable management of written outputs, non-traditional outputs such as images and videos, and the underlying data files.

Text is extracted for full text indexing from common file formats, including PDF, MS Office, OpenOffice, text and HTML. These can be viewed in an embedded viewer directly on the page, along with image file formats including PNG, GIF, JPEG and TIFF.

Audio media players are included for WAV and MP3.

Licence

Researchers can select an appropriate licence from a drop-down list of licences. The repository manager can confirm the licence or select an alternative.

A link to the licence is displayed on the metadata record in the public repository. The licence will affect the behaviour of other parts of the system, e.g. if the licence doesn't allow the file to be downloaded, a "request a copy" button appears which starts a workflow to provide access.

Haplo supports all CC-BY licences and can support any institutional specific licences.

Metadata

A flexible metadata schema enables the repository to accurately reflect the research in an institution making it easier to find and reuse

Haplo can represent any type of information in a single repository. It is ideal for managing datasets, traditional and non-traditional output types together, and has the flexibility to support new templates for new output types at any time.

Advantages of a flexible metadata model

It makes far more sense for researchers to have one system for submitting their research outputs, and for readers to have one place to search. All research outputs can be searched simultaneously and all outputs from a researcher, research project or department can be displayed together.

Repository managers benefit from the convenience of storing all research outputs within a single system rather than supporting separate repositories for different types of output. And institutions build a rich picture of all research undertaken by their researchers.

Supporting highly tailored metadata

Haplo's shared schema support enables a repository to incorporate both industry standard metadata schemas and custom metadata for each type of output. Developed in a highly modular manner, institutions can tailor the metadata, and every aspect of the repository, to accurately describe their research. Haplo's flexibility makes it ideal for even the most specialist of research institution.

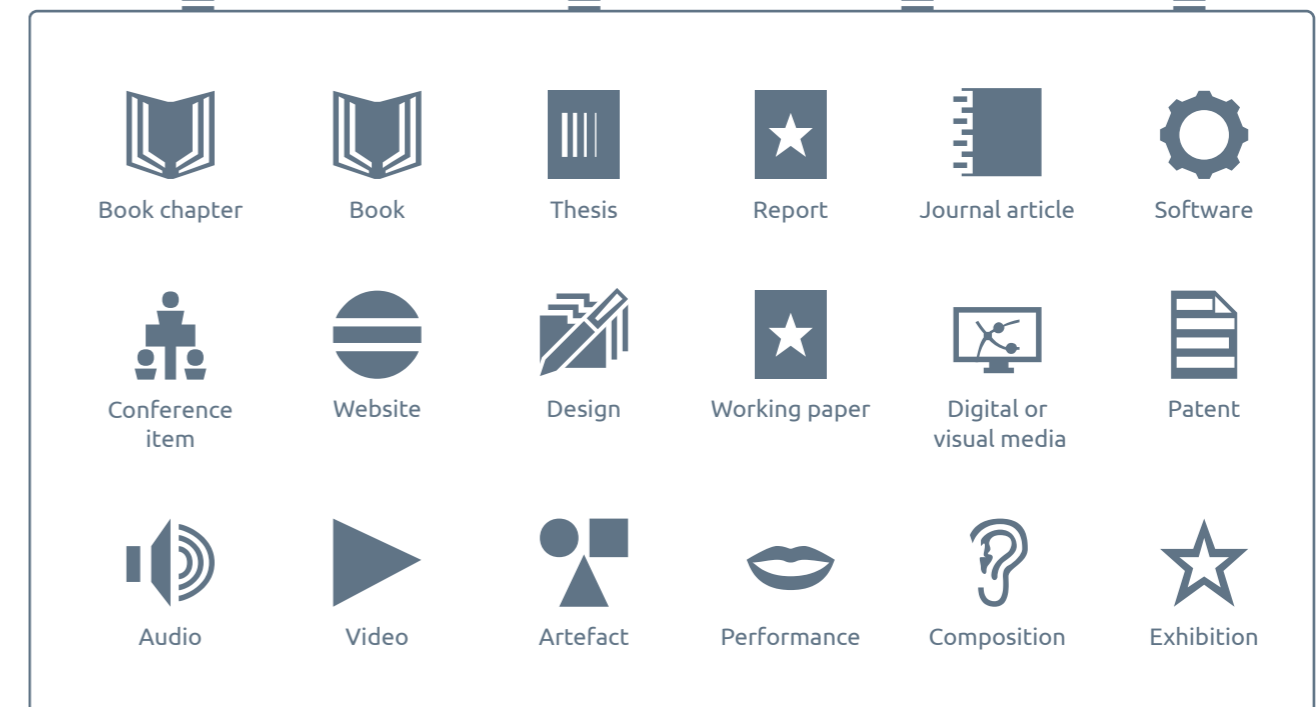
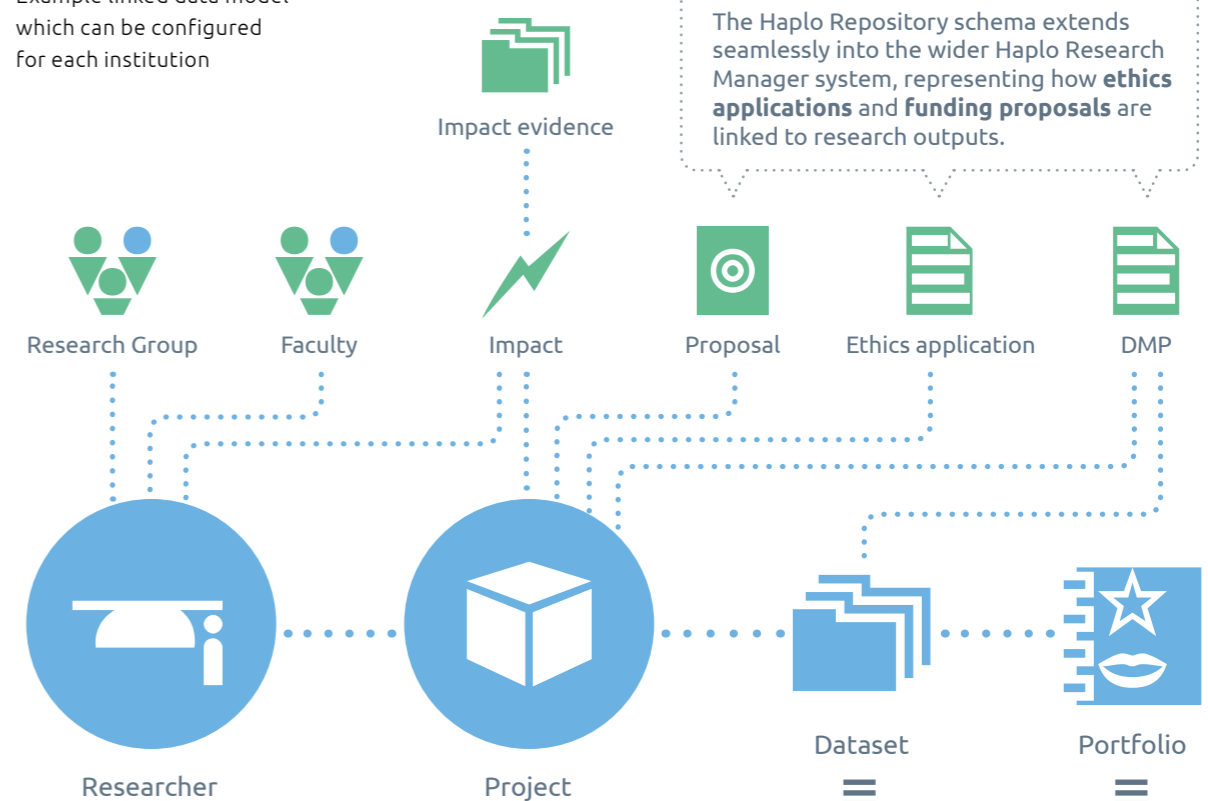
About Haplo's technical architecture

Haplo's data model is a linked data (*semantic web*) inspired object store. It's an innovative mix of a graph database, object store, and search engine. Every field is multi-value, and hierarchical relationships are supported everywhere they're needed.

The technical architecture underpinning the Haplo repository is built through layers of independent and self-contained modules and multiple plugins. By breaking the elements which make up a repository into many small elements, Haplo can offer extensive flexibility in which elements are used and in how a Haplo system is built for each institution.

The ordering of the layers is important and differs from how repository systems are traditionally built. The elements institutions most often wish to customise (*such as the interface and schema*) are in the top layers to enable easy configuration.

Example linked data model which can be configured for each institution



Haplo supports multiple research output types

Different attributes for different outputs. Haplo Repository manages all kinds of research outputs (Datasets, Portfolios, Practice-based research, and traditional research outputs types) within the same system with different attributes used by different outputs.

Ingest Workflow

Repository managers efficiently process submissions ready for publication in the public repository

“ The most impressive part of the ingest process is the review. When an item is reviewed by staff, there is a clear audit trail for both the reviewer and the depositor. All the commenting and discussion is kept within the system. This is really helpful, especially when it comes to time critical deposits for compliance. ”

Deposit process

Repository managers are alerted to new submissions by researchers. Haplo's comprehensive and flexible workflow system streamlines the approval of submitted outputs.

The deposit process includes review stages by authorised users, such as repository managers, and automatic publication of the approved record to the public repository.

Records can be returned to submitters for changes. Haplo maintains an audit trail of all actions together with a record of the conversation between the depositor and repository manager.

Haplo's default workflow is suitable for most institutions, with the reassurance of high levels of configuration and modularity to enable the workflow to be tailored precisely.

Permissions

Fine-grained permissions provide flexible yet highly secure access to the repository system, enabling a greater level of control over access for different users.

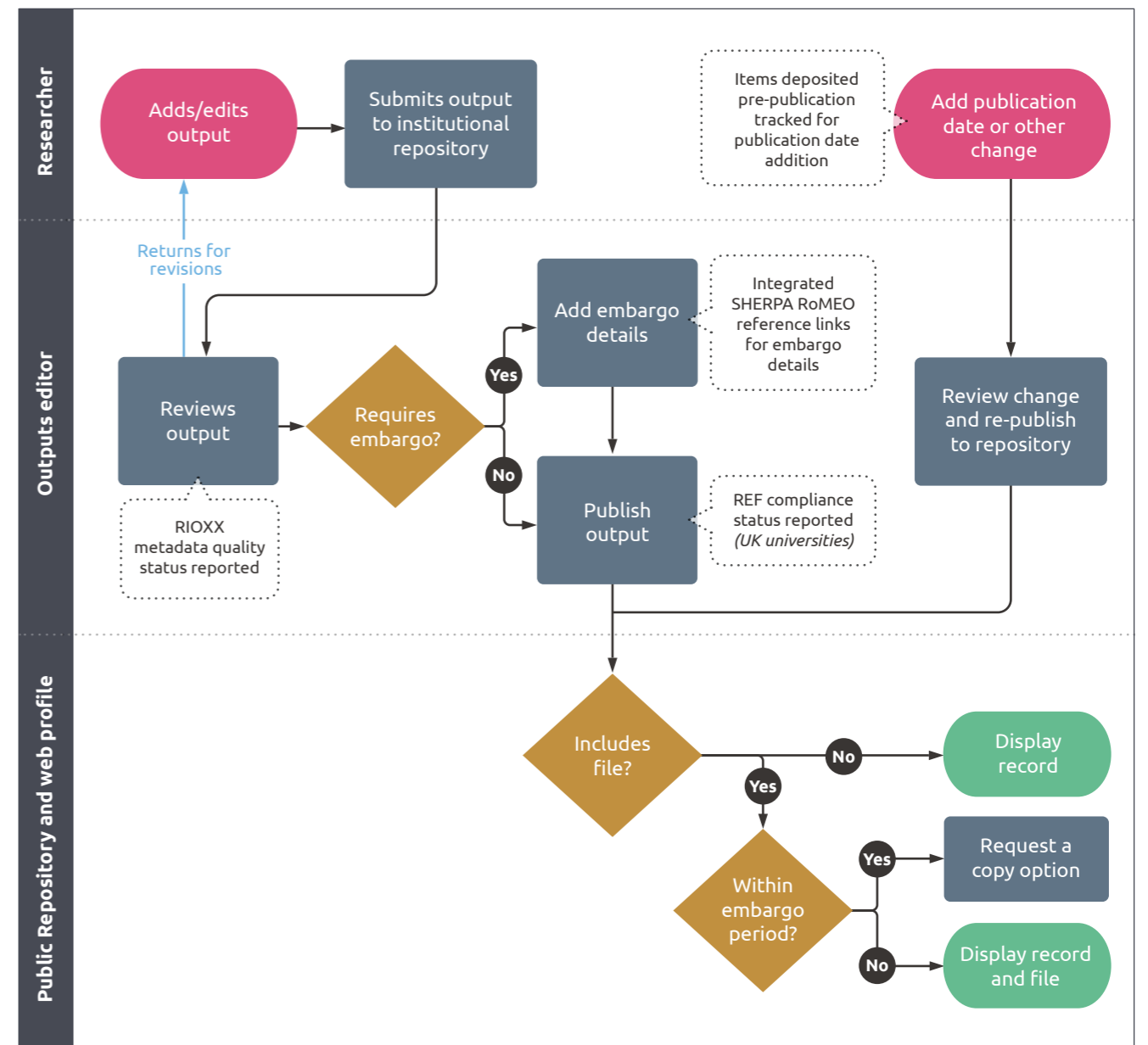
Access rules can be applied down to the individual metadata field level, not just for the whole record. This enables administrative fields to be hidden from everyone except privileged users, or embargoed files to be available only to output authors (*but not other researchers*).

DOI minting

Haplo supports plug-ins to generate and assign DOIs, using DataCite, with institution-specific prefixes on ingest where a DOI is not already recorded for an item. Where a DOI is minted for the item, the DOI is automatically added to the metadata record and displayed publicly.

DOIs can be created for one collection of content only and not for others, or created following any other rule as described by the institution. DOIs may be assigned at a specified point in the ingest process, for example after approval of a record for publication.

Metadata and the link to a public landing page is stored in DataCite, and is updated when the item is revised. Institution specific policy controls which items and what metadata is stored, and how versions of items are managed.



Example workflow which can be configured for each institution

SHERPA

Haplo integrates with SHERPA tools to provide reference and guidance information on funder and journal requirements to users at appropriate stages in the workflow.

This guides users towards compliance with funder and journal regulations, minimising the administrative effort for repository managers and ensuring the compliance of outputs uploaded into the repository.

Multiple embargoes

Haplo supports multiple embargo periods on a

single record, managing different embargo periods for each file type associated with a single output record.

Publisher embargo periods for research outputs can be looked up using the SHERPA RoMEO service.

Embargo date reached

When an embargo date is reached, the embargoed file becomes available for download. This can either happen automatically (*with or without notifying the submitter and repository manager*) or can require confirmation by the repository manager before publication.

Portfolios

Haplo supports the rich variety of research outputs from practice based research

Example portfolio in University of Westminster repository

With highly flexible metadata, configurable visual display, and elegant portfolio functionality, Haplo can support a wide range of outputs from practice based research.

Files and images are collated from each of the portfolio items and presented as a collection. The portfolio record is automatically populated with data from each item which forms the portfolio.

Haplo's extensible schema allows for every contribution to items within a portfolio to be credited. The sophisticated workflow system allows special business logic to be implemented for deposit of portfolios.

Haplo's public interface supports different displays for portfolios, highlighting the curated and collaborative nature of the output.

Finding and Reusing Outputs

All of an institution's repository can be searched to find and reuse research

Example search results display using university branded interface

Users can search Haplo Repository to find the items they want and discover relevant content. Haplo supports the indexing and search of the full text of uploaded files in all major machine-readable formats. Search result prioritisation ensures that search terms matching key fields such as title and abstract are displayed at the top of the search results list.

Haplo supports intuitive 'simple' searching and

advanced search. Searching navigates the links between items within Haplo to support deep interrogation of the data, such as finding all outputs from researchers who have received funding from a specific body.

Haplo is optimised for search engines by presenting clean and semantic mark-up in publication pages and embedding meta tags for machine-readable data. All URLs are maintained and migrated using industry best practices when moving to Haplo Repository, ensuring the addresses of your existing landing pages will work indefinitely, and your search engine rankings are not affected.

Managing Restricted Files

Supporting access requests enables restricted files to be managed confidently in an open repository

Prepare files for release

Select files for release from those that have been previously prepared, and/or upload newly prepared files appropriate for this application's audience.

Select	File	Audience	Action	Comment
<input type="checkbox"/>		External	Anonymised	Fully anonymised dataset.

Newly prepared files

Drag files here or [choose file...](#)

Suitable for *

-- select --

Preparation action taken *

-- select --

Comment

Restricted files access requests

Haplo's access request support enables institutions to manage requests from internal or external readers to view embargoed and restricted files.

Access requests are submitted to the repository manager and author for approval. Files can be routed for processing prior to release, for example to anonymise data sets.

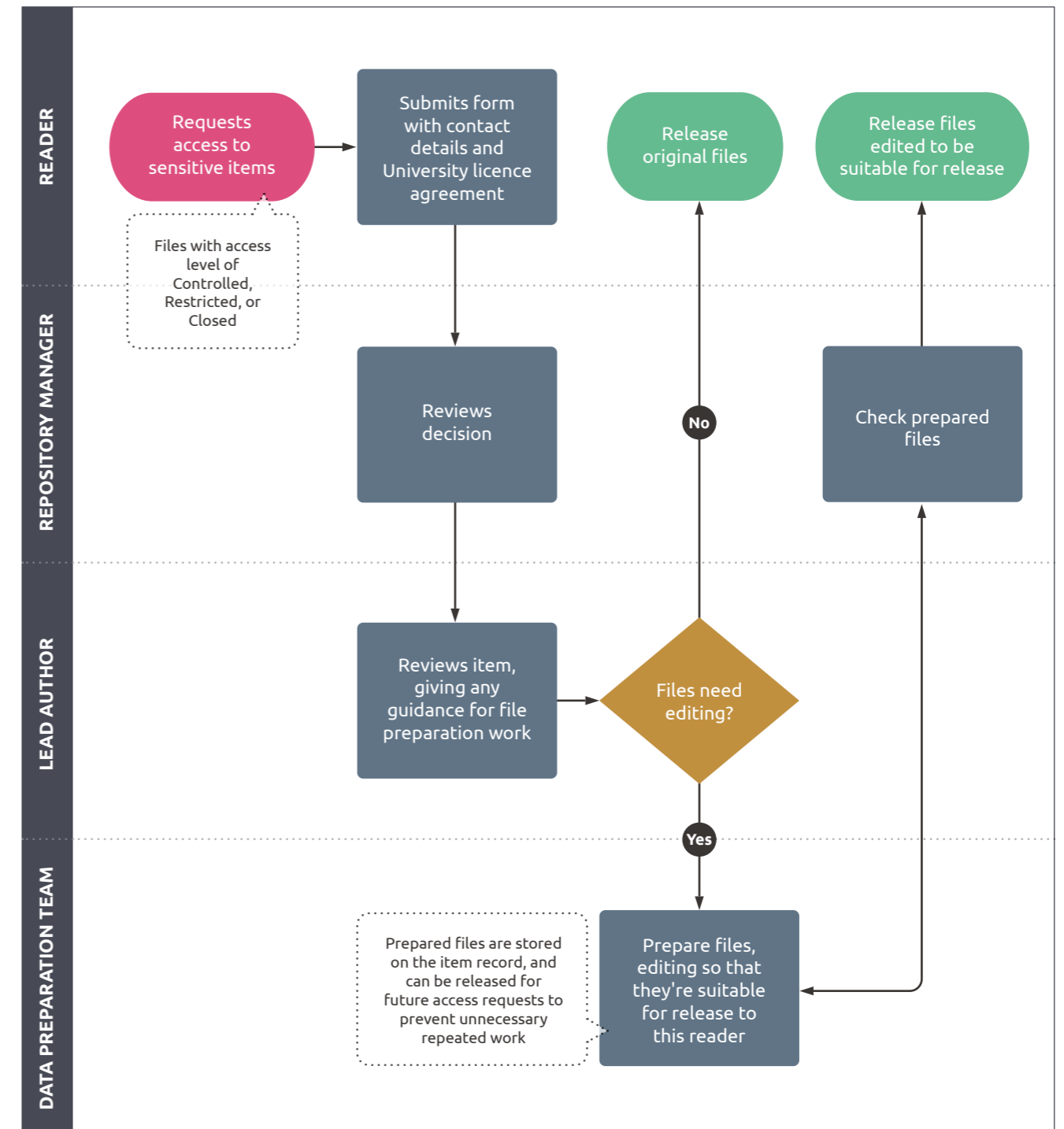
On approval, the reader is granted access to the file for a limited period and the download is logged for

auditing. Users can either be sent a link to the file in an email or be required to create an account.

Re-use licences

Users can be required to agree to or acknowledge a re-use licence before viewing the item. Agreement and download is then logged on the audit trail.

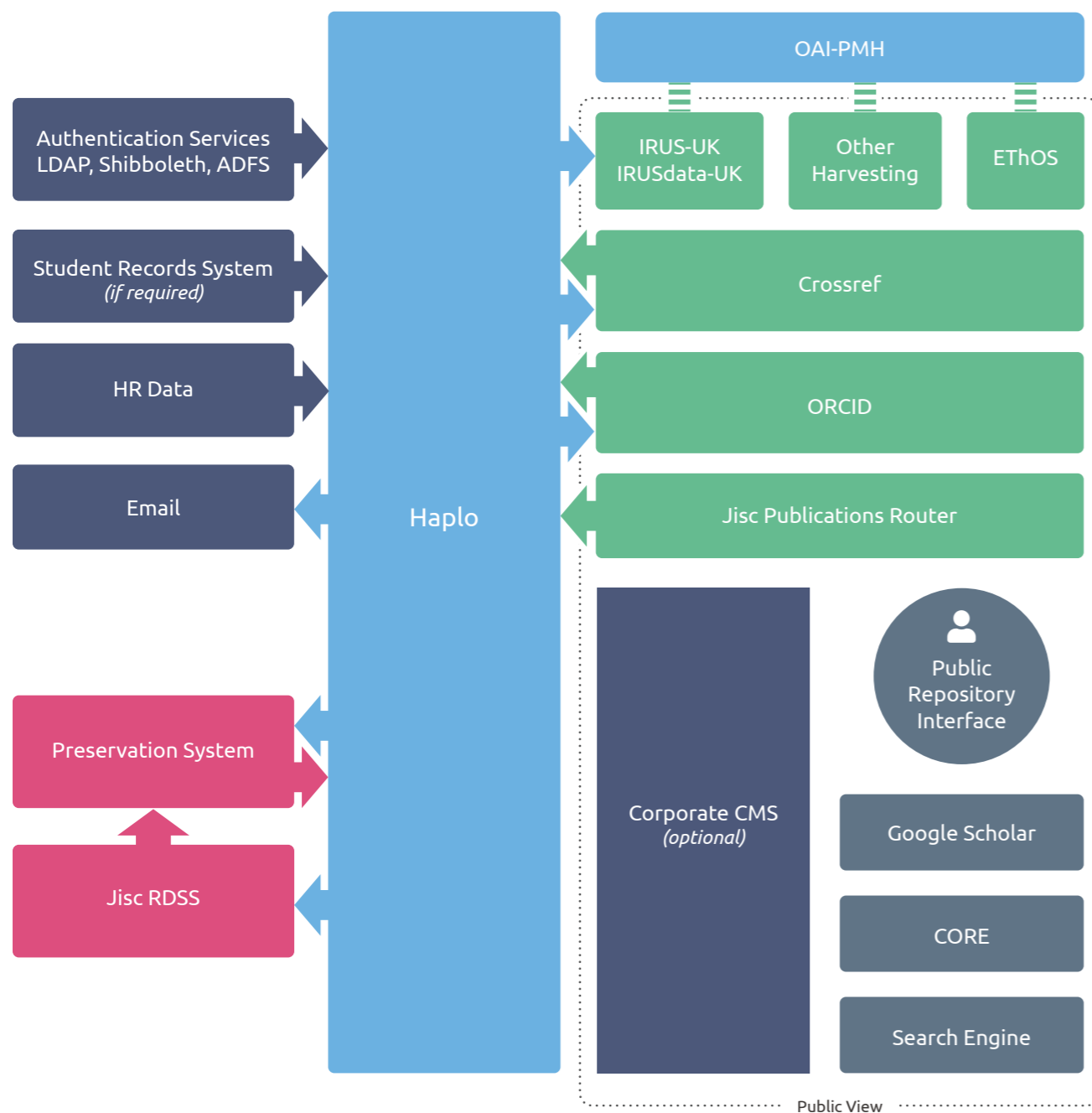
Haplo provides a dashboard for reporting which items use (*and re-use*) licences, and the number and names of readers who have agreed to each re-use licence.



Example restricted file access request workflow which can be configured for each institution

Integrations

Integration with other systems enables data to be entered once and used wherever it's needed



Haplo supports all industry standard integrations for pushing data to and from your repository, including: OAI-PMH, Datacite DOI minting, ORCID authentication and data push.

Haplo can integrate with institutional systems. If you're not using the full Haplo Research Manager suite, Haplo Repository can integrate with your choice of Research Information Management System.

Haplo provides an extensive API to enable data to be fed to and from other systems, with custom APIs developed as required.

Haplo can publish and receive machine-readable, structured data (e.g. JSON, XML), in an open format. This is achieved via a publicly accessible and openly documented API (e.g. using the REST architectural style over HTTP, or SOAP web services.) This can be fulfilled through batch publishing of CSV files, or more structured data formats.

In addition, Haplo has the ability to publish data changes from and to the system in real-time.

Researchfish

Haplo presents data in a format that can be uploaded to Researchfish, as a precursor to an automatic feed when a Researchfish API is made available.

CERIF

Haplo sits within the CERIF framework for CRIS systems, using similar and compatible concepts to describe institutional information.

VIVO

Haplo can be configured using the VIVO ontology for VIVO compliance.

Preservation platforms

Haplo can integrate with preservation systems providing a single streamlined process for ensuring the long-term preservation of your research outputs.

Impact and Esteem

Evidence of the impact of research and esteem factors for each researcher can be recorded alongside researcher profiles and research outputs

Recording impact and evidence of impact

Haplo supports the recording of research impact and evidence of impact.

Evidence of impact can be uploaded in a variety of formats and linked to the impact record, including letters, emails, testimonials and reports. Records can link to external supporting material such as radio or television broadcasts, or other online sources.

Impact records can be updated to capture new evidence and changes over time. Haplo supports version control on all records, enabling all previous versions of impact records to be stored and accessed.

Linking impact to research projects

Records of impact link to the underlying research project, researchers, associated funding, and research outputs. Impact can be linked to REF Units of Assessment (*for UK universities*).

Several projects can be linked to a single impact record and one project can be linked to multiple records.

Haplo maintains links between all related records, enabling users to easily find all impact statements related to researchers within a department.

Workflow review

Records of impact submitted by a researcher can be sent to a reviewer to approve before being confirmed and published.

Reports

Dashboards show impact reports for all researchers. They can be filtered to show records of impact for researchers within a specific department, or for all research associated with a specific REF Unit of Assessment (*for UK universities*).

Public portal

Impact statements and case studies can be published on the researcher's public profile or displayed internally only.

Recording esteem

Researchers can select from a pre-defined list of esteem measures and then add more details about the individual measure of esteem being recorded. Dates and organisations involved are recorded in a structured way to enable effective reporting. The categories of esteem measures can be defined by the institution, and changed as required in response to changes in institutional or external requirements (*such as the UK REF*).

As throughout Haplo, any record can be linked to any other record, so records of esteem factors can be linked to other records such as the relevant research output, postgraduate researcher being supervised, grant or record of impact.

Reporting

Reports provide a comprehensive real-time picture of every aspect of the repository

Haplo reports provide wide-ranging monitoring of the repository, including:

- ✓ Items added each year
- ✓ Status of newly submitted items
- ✓ Items with embargoed or otherwise restricted files
- ✓ Number of file downloads and access requests
- ✓ Last access date of files
- ✓ Administrative metadata including date and time of record creation, name of record creator, date, time and names of users making subsequent changes
- ✓ A comprehensive audit trail including user, date, time and searches performed
- ✓ Funder policy compliance reports on a per-item or per-funder basis
- ✓ Compliance rates recorded on a per-faculty and per-department level
- ✓ Individual item compliance recorded and highlighted to researchers at the point of deposit
- ✓ Reporting information gathered on a per-article basis, allowing article level metrics reporting
- ✓ Configuration of each item record to show number of downloads
- ✓ Optional integration of Google Analytics and other external altmetrics services into the public repository page
- ✓ Custom reports for each institution

Universities in the UK streamline their REF submission process

Haplo provides integrated support for managing REF exercises. With fully pluggable and configurable processes to fit your institution's processes, Haplo is the ideal choice for smoothly managing your data to prepare REF submissions.

Clear and simple reports give Unit of Assessment Leads and REF and repository managers real-time visibility of the current state of REF readiness.

Sensitive data and processes can be managed within Haplo with confidence, due to Haplo's ISO27001 certification, and fine-grained permissions ensure only administrative users access confidential information.

Audit and Version Control

A detailed audit trail and support for version control provides reassurance of the origin and integrity of all changes within the repository

The publications and information stored within a repository are a critical part of a research institution's reputation, so it's important to know who submitted each output and when, and how it's been changed and used since deposit.

Version control

All records within Haplo Repository are version controlled, enabling administrators to view all previous versions of information, and who made the changes - and revert to previous versions if mistakes have been made.

Haplo supports version control throughout the system not just on output records. Profiles of your researchers, information collected in ingest forms, records of licences, publishers, and research projects all benefit from version control.

Audit trail

The audit trail records all activity in the repository. As well as providing a traditional security audit for administrators, it's also used to display useful information to all users about the history of any record.

Working alongside version control, the audit trail records all changes to information on every record, providing repository managers with assurance that they can prove when and how records have been created and edited.

It records key system events such as users logging in, and when and by whom sensitive information has been accessed.

As with everything in Haplo, the audit trail can be used by plugins to record additional audit events, and query the audit trail for auditing or custom activity alerts.

Retention

To maintain a comprehensive and reliable audit trail, by default Haplo never deletes information. Instead, it marks it as no longer required, with different levels of 'hidden' for different purposes.

Deleted records are only accessible to users who have specific permission to view them (*and undelete them*). **Archived** records are generally visible to everyone, but hidden from search results by default. **Deprecated** records are visible, but aren't allowed to be used in new records.

Individual records can be deleted for legal reasons, and retention policies can erase data that's no longer required.

Data Security

Data is kept safe in a repository where security is an integral part of the design

Security

Haplo's focus on security makes it the ideal repository for storing sensitive data and managing embargoed or restricted files.

Safely managing sensitive information is critical for a modern repository. However, traditional repository systems were designed to make all data and outputs available without restriction.

It's challenging to retro-fit security onto a system that hasn't been designed with a suitable threat model in mind, which leaves institutions to shoulder the additional administrative burden of storing and securing sensitive information in alternative systems.

Security is integral to the design of Haplo Repository and Haplo has been managing sensitive personal information to an exceptional standard for over a decade.

Because of the focus on security throughout Haplo Repository, institutions can safely manage all research data within the system, with confidence in the security of their repository, and with the tools to manage access controls within the institution and when working with external researchers.

Haplo is provided as an ISO27001 certified service.

Hosting

Haplo is offered as a hosted solution, using a fully redundant internet connection, with a 99.9% SLA from our internet provider, at multiple co-location facilities in the UK. As a matter of policy, we own all the server hardware, switches and firewalls so we can offer strict data protection assurances to our customers.

Backup

Haplo maintain equipment in multiple datacentres. The live and backup servers mutually authenticate using strong cryptography. Data is replicated and transferred over encrypted connections between data centres using the public internet.

The hosted service continuously replicates data between two datacentres, allowing recovery from loss of a datacenter. In the event of user error, the system can be restored by rollback of individual records, or restoration from backup.

Data is backed-up every night, stored in another datacentre, and retained for 90 days. After 14 days, one in 5 daily backups is retained for 90 days. Backups older than 90 days are deleted. Snapshots can be taken and maintained for longer periods before significant events, for example, import of new data into an existing system.

Questions

Where is our data hosted?

Haplo is offered as a hosted solution, using a fully redundant internet connection, with a 99.9% SLA from our internet provider, at multiple co-location facilities in the UK.

Can Haplo integrate with our research management system?

Haplo provides a suite of integrated research management tools which together form a comprehensive research information management system. In addition, Haplo can integrate with your choice of research management system and any other systems with which you need to share data.

Can we customise the metadata?

Yes, institutions can tailor the metadata, and every aspect of the repository, to accurately describe the research at an institution.

Can we customise the public portal?

Haplo provides an attractive public portal interface by default. For institutions wishing to customise, every aspect of Haplo's public portal is customisable, from the kinds of information it displays, to the way that information is structured and presented. Using open tools, the institution's web team has full control over all HTML generated, and can modify the HTML, CSS and templates of the public portal to match the look and layout of the institution's visual identity.

Which re-use licences do you support?

Haplo supports all commonly used re-use licences, including all CC-BY licences and can support institution-specific licences. If the licence doesn't allow the file to be downloaded, a "request a copy" button appears on the public portal instead of the file which enables the visitor to request access.

What support do you provide?

Haplo provides expert advice and support. Our standard service help desk support is based in London, UK and is available between 09:00 and 17:30 Monday to Friday.

Can Haplo support multi-tenant repositories?

Haplo is the ideal choice for multi-tenant repositories enabling a single institution to host all research outputs in one repository; and a group of institutions to share a repository.

What support and documentation do you provide for developers working on your open source repository?

The Haplo platform is open source, released under the Mozilla Public License v2. The platform code and full documentation are available at <https://haplo.org>. Support contracts are available.

Are there any limitations on what browsers are supported?

Haplo can be used on both PCs and Macs, and all modern standards compliant web browsers are supported.

Can Haplo be used on mobiles and tablets?

Haplo can be used on mobile devices through mobile websites. The system is designed to provide a full experience on tablets, and an experience tailored to reading information on smaller devices. Public interfaces are fully responsive.

Can Haplo perform well when large volumes of data are stored?

Haplo continues to perform well with peak usage loads. Our system performance target is for all user screen interactions to return within 0.5 seconds. Performance is continuously logged and monitored. Current actual performance is between 0.002 seconds to 0.050 seconds.

Can we retrieve all of our data at any time?

We strongly believe you should have full access to all of your data at any time. There is no contractual or technical lock in when using Haplo. Haplo can provide an API for you to export your data at any time. All code used to deploy the application, including both product and institutional customisations, is available at any time under our liberal open source licence.



Haplo
Repository

