Haplo

Haplo Research Manager

Technical specifications

November 2017

Haplo Research Manager

Haplo Research Manager is a modern, flexible research management system built on an open source platform. Haplo provides an ideal solution to support modern universities seeking to increase the volume and quality of their research activity. The solution:

- provides the university with **accurate, up-to-date information** about current research activity
- provides efficient access to information about a **wide range of research activity**, including: doctoral researchers, research ethics, research funding, research outputs
- enables **access by all** researchers, doctoral researchers, and administrative support staff
- creates **rich profiles** of research activity per researcher, research group and school
- produces easily accessible **reports and management information** for those with responsibility for research within schools, research groups and across the university
- enables the university to improve the **quality** of its research activity: encouraging good practise; identifying successful researchers and research activity; and identifying areas requiring more support
- ensures **compliance** across the university with contractual, legal, and ethical requirements for research and research data management, including requirements from funding bodies
- applies appropriate **permissions** to enable information to be easily shared and protected as required
- **integrates** with existing university IT systems
- is provided as a **hosted** solution providing fast, reliable access to the system
- stores data in line with **UK data protection legislation** and information security good practise
- is **flexible** and extensible to adapt to changing requirements
- is **cost-effective**
- is **easy to use**

Technical specifications

Architecture principles ***4***
Modular functionality 4
Service architecture 4
Standards based 4
Simplicity 5
Reliability 5
Sustainable 5
Hostname (web address) 6
Branding 6
Data feed 6
Authentication 7
Email notifications 7
Outlook integration 8
Performance 8
Test environments 8
Hosting and security 8
Servers 8
Data integrity 8
Security and availability 9
Monitoring 9
Backup & disaster recovery 9
Archive service 9
Browser support 10
Accessibility 10
Open source 10
Connections to other systems 10
Permissions 10
Audit 10
Documentation 11

The Haplo platform provides a flexible database tailored to storing information about the activities in complex organisations. It enables all types of information to be described and stored within a single integrated system. Sophisticated search enables information to be found, and automatic linking of related items supports serendipitous discovery of useful information.

The web-based user interface provides a familiar experience for users, reducing the need for training and shortening the time to embed the system into everyday use.

The platform includes support for collaboration and workflow, and the storage and processing of files in all major formats. It provides fine-grained permissions and full auditing. The high-levels of customisation are enabled by server-side JavaScript plugins.

Architecture principles

Modular functionality

Haplo Research Manager applications are composed of approximately 100 modular components. This modularity allows highly customised applications to be delivered with the majority of functionality implemented through common modules, with institution customisations layered on top.

Each component provides a well defined interface to other components, through a platform API which enforces isolation and provides mechanisms for declaring and calling internal interfaces. All platform APIs and functionality provide defined extension points for customisations.

This allows us to reuse functionality between systems, while still providing the high level of customisability required to be able to implement all processes and practises of an individual university.

Service architecture

While the development model uses the underlying principles of service orientated architecture, it is deployed within a single server for reliability and ease of deployment.

Because of the high efficiency of the application platform (typical response time is about 0.05 seconds) and the size of the data (typical graduate schools store less than a million objects in the database), there is no need to split the applications across multiple servers, as our largest graduate school uses about 0.1% of the resources of our standard deployment server.

We use virtualisation to manage and distribute resources across our server fleet.

We deploy on servers owned by Haplo Services, and co-locate them in UK datacentres, to guarantee we can meet the data protection requirements of our clients.

Standards based

Haplo Research Manager is implemented as a standards compliant web application which requires no special software to access.

The underlying Haplo platform is open source: https://haplo.org

We only use open source software to develop the platform, with liberal licenses which do not place restrictions on deployment.

Our APIs use simple standards to interoperate, such as REST with XML and JSON payloads.

Simplicity

Our high performance is due to ensuring that all software and architecture is as simple as possible, with complexity only when required to meet business requirements. As well as high performance, this reduces the cost of development as developers only have to engage with the business requirements.

Reliability

Redundancy with seamless failover in all parts of the infrastructure that are most likely to fail: Internet transit networking, firewalls, internal networking, power supply, disk controllers, disk drives. This minimises the chances of a common component failure impacting customers.

Application resilience is achieved by replicating data on a very frequent schedule to a secondary datacentre, and having an automated method of triggering fail over to the secondary data centre after an administrator has confirmed the need.

All these redundant components and data replication use simple widely deployed solutions, so we can have a high level of confidence that they will work. As with any hosting provider, we have had component failures, which have been resolved without impacting customers.

After reducing the likelihood of a customer impacting failure through careful component choice and redundant configurations, we host the application in a very simple manner using monolithic deployment of the application in Solaris Zones on a single server.

We have made this choice as complicated schemes to deliver automatic failover of stateful application data using multi-server distributed systems are prone to failure themselves. Simple architectures with few moving parts and interdependent systems are much easier to run reliably.

We meet our SLA through a comprehensive monitoring system which alerts the on-duty system administrator to investigate and respond, using documented playbooks.

Sustainable

The modular approach and emphasis on well defined platform APIs backed up by an automated test suite means that applications can be enhanced and maintained in the long run.

We have applications running on the latest version of the platform which were initially developed 8 years ago. All our APIs are designed with long term sustainability in mind.

The deployed application has extensive monitoring and telemetry enabling us to monitor application performance, the infrastructure, and identify any issues before they impact customers.

Hostname (web address)

Haplo is hosted under your university domain name, with your choice of hostname, such as [research].youruniversity.ac.uk

The university will need to:

- supply a valid SHA-2 SSL certificate from a public CA, along with any intermediate certificates, ideally with a one-year expiry time. We can provide you with a CSR in any reasonable format.
- create a CNAME DNS record to point the chosen hostname to [university].infomanaged.co.uk

Branding

Your university's logo and branding colours are displayed throughout the system.

Data feed

Haplo Research Manager is a hosted web application which integrates with university infrastructure over simple APIs. Every university stores the information about their researchers and staff in different ways. We work with your IT team to get the data you need from the appropriate system and sent to Haplo. This enables Haplo to understand everyone's roles to route workflow to the right person.

For the feeds of information about staff and students, we recommend using a batch process. A report will be created in each system which outputs all relevant information into one or more files. These are uploaded by a script (examples provided for UNIX and Windows servers) over https to the Haplo servers. Haplo Research Manager works out the changes between the new files and the previous files, and updates data in Haplo Research Manager. Updates are performed online without any service interruption. Logs are provided by email and through an administrative interface.

This approach is quick and easy for the institution to implement, and because all system state is maintained in a set of files, it makes troubleshooting easy.

The feed system is common functionality. Custom business logic is added to interpret the data from the institution, as all institutions configure their systems in different ways, and a mapping interface is provided to maintain the mapping of codes.

The likely data we would need from the university records system is:

* Name (title, first name, surname)
* Username
* Department
* Email
* Role (Researcher, Doctoral Researcher, Administrative staff)

* Job title (Staff only)
* Enrolment status
* Initial registration/enrolment date

* Intended award
* Visa type

While we prefer JSON as an interchange format, we can use TSV (tab separated value), XML or any other structured text-based format.

If the university is able to provide all the required information in an LDAP directory, and allow Haplo to search it, we can use this instead of the data feed.

Authentication

Authentication is a separate mechanism to the user data feed to authenticate users when they log in. While Haplo has built-in authentication, we recommend integrating with the university's identity server for single sign-on.

For authentication, Haplo currently supports:

- Shibboleth
- LDAP (including Microsoft Active Directory)
- OAuth2 for Google Apps

In Q1 2018, Haplo will support:

- Microsoft AD FS

Haplo can add support for extra identity providers, and implement custom login user interface.

If LDAP is chosen, we need the university to:

- Create an LDAP service account: Provide a hostname, username and password for your LDAP server. We only support encrypted connections.
- Create firewall rules for LDAP server: Allow access from our hosted networks (we will provide you with a list of network blocks which may make requests)
- LDAP certificates: Provide details of all SSL certificates which may be presented by the LDAP server (if they do not necessarily have the same hostname as the one we use to connect)
- LDAP search path: Provide details of which parts of the tree we should search for users, and which attribute is used for the username. (We check passwords by searching for the username in the tree to find the user's full DN, then attempt another connection with that DN and the provided password.)

Email notifications

Haplo will send email notifications to users on behalf of the university. There are a number of options for how we provide this, the two main options are 1) the university sets up DNS SPF records, to inform external spam filters that we're a valid source of email, and whitelisting within your email servers, or 2) we

can route email through your own email servers, using authenticated SMTP, which you can then deliver to the recipient.

Outlook integration

Haplo Research Manager integrates with calendar software as an iCalendar feed of alerts, reminders, and all meetings arranged through Haplo. This is a standard format which is compatible with Outlook and all calendaring software on all devices. A direct connection to Microsoft Exchange is a planned feature on our roadmap.

Performance

The Haplo platform 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.

Test environments

Clients are provided with development, test and anonymised test environments, in addition to the production system (with managed failover to a Disaster Recovery site).

Hosting and security

Haplo is offered as a hosted solution, using a fully redundant internet connection, with an 99.9% SLA from our internet provider, at multiple co-location facilities in the UK. There is no single point of failure in the networking: Each server has two physical connections to the network with automatic failover. These physical connections connect to different switches, which connect in a cross-over pair to a redundant firewall cluster, which in turn connects to two network backbone routers, which each have multiple redundant paths to the internet at large. Our network has had 99.995% availability over the last 4 years.

We have the ability to utilise private peering arrangements with other networks, if required and contractually feasible. Our network provider has a fast connection to JANET.

Servers

As a matter of policy, we own all the server hardware so we can offer strict data protection assurances to our customers. Only Haplo employees have access to our servers.

All server components with moving parts are in hot-swappable redundant pairs (power supplies, fans, storage) and monitored by an integrated service processor. Each data centre has two redundant power feeds, with backup generators. The servers can operate with only one power supply connected, and each feed has enough capacity to power the entire data centre.

Our servers have had 100% availability over the last 4 years.

Data integrity

All storage uses mirrored disks for redundancy (equivalent of RAID 1). Future bulk storage may use other RAID levels for space efficiency.

All data is stored on ZFS filesystems with full cryptographic checksumming, for assurance that any hardware errors will be detected and corrected. Daily snapshots are taken, and every week, every bit of data on the disks are checked for proactive detection of hardware faults. In addition to the filesystem level checksumming, files managed by the platform are checksummed at the application level, and those checksums are verified.

Security and availability

Our firewalls only allow http and https traffic from the internet at large. Administrative access is controlled by certificates (not passwords) and requires VPN access. Our deployment system uses cryptographically signed software archives, ensuring that only authorised software is installed.

Each server and Platform instance is independent (no single point of failure within the cluster as a whole), and all traffic is encrypted. The servers are physically located in a locked rack in the datacentre. Access to the rack and datacentre floor is controlled by 24/7 manned security, and all visits are escorted and logged.

Monitoring

We use internal monitoring of server health, and external monitoring of network availability.

Backup & disaster recovery

We maintain equipment in a 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.

Our primary method of recovery is to promote a replica in another datacentre. All live applications are constantly replicated to at least one other datacentre.

A secondary recovery method is to restore from a backup. Our current time to restore from a backup is under 4 hours. In a disaster recovery scenario, we can build a new server, automatically, in under 15 minutes starting from taking a new server out of the box.

To ensure that our backup and disaster recovery mechanisms always work, we use the same processes for our day to day administration. For example, the normal software deployment and upgrade process uses exactly the same mechanism as would be used for disaster recovery.

To verify the backup system is working as designed, test restores are made on a regular basis to test the ability of the system to restore a backup, using Customer data chosen at random.

30 days of complete backups are kept. Daily snapshots are taken and retained for 14 days. One in 5 daily snapshots is retained for 90 days. Backups older than 90 days are deleted.

Archive service

We provide an optional archive service where a copy of a client's data is made available to them to maintain their own independent backup. The archive service is intended as a one-off or less regular service and a small charge is made for each archive operation.

Browser support

To access Haplo, users need to use Web browsers which support the SSL SNI extension. All modern web browsers support SNI.

Haplo can be used on both PCs and Macs, using browsers including Internet Explorer, Microsoft Edge, Firefox, Google Chrome and Safari. It can be used on mobile devices.

We will support any modern standards compliant web browser, and commit to supporting all the browsers supported by Microsoft Office 365 and Google Apps for Education.

Accessibility

The system meets W3C WCAG 2.0 accessibility standards and complies with all relevant legislation regarding accessibility.

Open source

The underlying Haplo platform is open source, released under the Mozilla Public License v2. It's written in Java, JRuby with applications written in JavaScript.

The platform code and documentation are available at https://haplo.org

We are working on open sourcing most of the core functionality of our Higher Education products. Features specific to a single client will not be open sourced for reasons of client confidentiality, but will be made available to the relevant client.

Connections to other systems

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

Permissions

Access permissions will be applied according to the requirements of the institution. Most permissions will be administered in the normal user interface by authorised users, and not require system administration privileges.

Key users can be granted the ability to 'impersonate' any other user, enabling them to see and take action by the other user for providing support.

Audit

One of the key benefits of Haplo Research Manager is the integrated audit trails to demonstrate that the institution has met all obligations to students, the university, and relevant third-party regulators. This is further enhanced by the capability to automate the entire workflow, so every contractual interaction is performed through the system and therefore audited.

User access: A system level audit trail records all successful and unsuccessful authentication events, along with system configuration and core record changes. The audit trail can be searched and used in reporting.

Specific events: All processes include a full audit trail of actions performed, shown to user as a "timeline" of activity on the event's page. This can be used in reporting.

Forms, including forms for events: Forms are fully version controlled, with logs of who made changes and when, and access to the previous version of the form in the user interface.

User profiles: manual changes to user profiles within Haplo Research Manager is recorded in the history of the record, including which user made the change and when. Changes to user profiles from updated information sent in the data feed from university systems, is recorded in the history of the record, including that it is a system-made change and when it occurred.

Documentation

Platform documentation is publicly available at

https://docs.haplo.org/dev/plugin for developing custom integrations.

https://docs.haplo.org/dev/standard-plugin/reporting/rest-api for access to reporting information.

Service levels

Availability target	
Service availability (not including planned maintenance)	99.5%
Planned maintenance	Up to 1 hour per month

Availability (% of calendar month)	Reimbursement rate (% of monthly Charges)	Maximum Downtime (hours per calendar month)
99.5% and above	0	3.6
99% - 99.49%	5	7.2
95% - 98.9%	10	36
90% - 94.9%	15	72
less than 90%	20	More than 72 hours

The Haplo team will notify you by email at least 5 working days in advance of scheduled downtime required for planned maintenance which is expected to exceed 30 seconds or take place within normal UK business hours. Where Planned maintenance is not expected to exceed 30 seconds Downtime, it will occur at the times listed in the Planned maintenance schedule.

Planned maintenance schedule for maintenance not exceeding 30 seconds	
Monday - Friday, excluding Bank Holidays	8am, 7pm (UK)
Saturday - Sunday, and Bank Holidays	All day

The Haplo team will notify you of any report of non-scheduled downtime, and investigate and remedy it using suitably qualified personnel in line with the published response and fix time targets.

Response and fix time targets	Core service hours		Non-core service hours	
Severity	Response	Fix	Response	Fix
1. Critical The problem severely impacts your use of the software in a production environment (such as loss of data or your production systems not functioning.) The situation halts your business operations and no procedural workaround exists.	1 hour	Continuous effort	2 hours	Continuous effort
2. High The software is functioning but your use in a production environment is severely reduced. The situation is causing a high impact to portions of your business operations and no procedural workaround exists.	2 hours	1 business day	2 core service hours	1 business day
3. Medium The problem involves partial, non-critical loss of use of the software in a production environment or development environment. For production environments, there is a medium-to-low impact on your business, but your business continues to function, including by using a procedural workaround. For development environments, where the situation is causing your project to no longer continue or migrate into production.	4 hours	4 business days or as agreed	4 core service hours	4 business days or as agreed
4. Low A general usage question, reporting of a documentation error, or recommendation for a future product enhancement or modification. For production environments, there is low-to-no impact on your business or the performance or functionality of your system. For development environments, there is a medium-to-low impact on your business, but your business continues to function, including by using a procedural workaround.	6 hours	8 days or as agreed subject to product roadmap schedule	6 core service hours	8 days or as agreed subject to product roadmap schedule

Service	Days	Time
Core service hours	Monday - Friday (excluding Bank Holidays)	09.00 - 17.30 (UK)
Non-core service hours	Monday - Friday Saturday - Sunday Bank Holidays	17.31 - 08.59 (UK) All day All day

About Haplo

Haplo Services is an innovative British company, combining Information Management and Information Technology expertise to deliver tailored solutions to complex information management problems.

The Haplo platform has been in commercial use continuously since November 2007. The platform was open sourced early in 2015.

Haplo Services have built an enviable reputation for the quality of its solutions, a relentless focus on information security, and high-levels of expert customer service.

Established in 2007, Haplo is based in London.

hello@haplo-services.com

Unit C 1st floor, Emperor House
Dragonfly Place, London SE4 2FL