Haplo

Haplo

Technical specification

NOVEMBER 2019

Technical specification

Architecture principles

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.

Modular functionality

Haplo Research Manager applications are composed of over 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 institution.

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 2% 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 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, and makes it easier to verify correctness and security in code reviews.

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.

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.

All software is developed under a stringent secure development policy, which includes security code review for all code. The platform is "secure by default" to enable safe development, and easy identification of privileged code in a security review.

Integrations

Data feed

Haplo Research Manager is a hosted web application which integrates with institution infrastructure over simple APIs. Every institution 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.

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 that Haplo will provide (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.

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 institution 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.

The likely data we would need from the institution systems is provided at:

https://support.haplo.com/setup/user-sync

This is not a prescriptive list and would likely be amended for each institution.

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 institution's identity service so users can use their institutional credentials and all access is controlled by the institution.

For authentication, Haplo currently supports:

- Shibboleth (SAML2)
- Microsoft AD FS and Azure AD (SAML2)
- LDAP (including Microsoft Active Directory)
- OAuth2 for Google Apps

Our recommended approach is to use Shibboleth or AD FS where possible, and LDAPS otherwise.

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

Instructions on setting up authentication integration are provided here:

https://support.haplo.com/setup/authentication

Access for external users

Access can be granted to users who do not have institution credentials via Haplo's internal access management function. This is commonly used to support access by External supervisors and External examiners (for PhD Manager) and external committee members (for Ethics Monitor.)

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.

APIs are provided to query and modify from other systems in real time, with a several approaches available depending on the requirement.

Outlook integration

Haplo Research Manager integrates with calendars as a feed of alerts, reminders, and all meetings arranged through Haplo.

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.

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 (clients in Australasia may use AWS if preferred.) 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, switches and firewalls so we can offer strict data protection assurances to our customers. Only nominated 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.

Security

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.

The underlying Haplo Platform is a security-focused information management system, which provides secure-by-default APIs to application code, gating of sensitive features behind flags that can be picked up in code review, and tools which prevent security flaws being introduced in the first place.

Monitoring

Haplo monitors all hardware, systems, and services 24 hours per day, 7 days per week, 365 days per year. Our availability over the last four years is at least 99.995%. This is determined by an external monitoring service which checks availability from multiple locations on the internet, and each of those tests performs a health check that all elements of the service are up. This, and an internal monitoring system, send alerts to system administrators who will respond 24/7 to any availability issues.

Back up

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.

The backup processes have no impact on performance and application availability.

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.

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.

Disaster recovery

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.

All recovery is performed by Haplo under the contract. The institution may request a copy of their data at any time, subject to a charge for cost recovery.

Our Disaster recover policy is reviewed annually and disaster recovery testing is undertaken regularly. The last test was successful and without issues.

Penetration testing

Haplo welcomes penetration testing. Several clients have commissioned independent penetration tests of Haplo Research Manager and these did not find any issues to resolve. Should a client wish to perform penetration testing, we would create a separate instance explicitly for the test.

ISO27001

Haplo's products are developed and hosted under an ISO27001 certified information security management system.

GDPR

- Data are stored in the UK, and will never be transferred outside the UK
- Haplo (the hosting provider) is UK-based
- Retention policies can be defined by institutions
- Processes are provided to enable the institution to meet their GDPR compliance, for example, retrieval of all data about a user.

Password management is controlled by the institution's authentication system. No passwords are stored by Haplo.

All data in transit, over the internet and the local network, is encrypted using robust open standards. Traffic is encrypted using TLS/SSL with a configuration that maintains an A grade at Qualsys SSL Labs.

Administrative and support access is only available to employees of Haplo, and controlled by cryptographic authentication and physical one time password tokens. Access to customer data is allowed for only users who need it to support the institution, and all support access is recorded in the audit trail.

Updates

Planned and preventative maintenance

Most updates are in zero downtime. Where downtime is required, it is less than 30 seconds, scheduled in advance, and performed out of hours.

Critical security patches may be applied out of schedule without prior notification, to maintain the security of the infrastructure.

Change control process

Haplo have Change Control Processes in place to meet an institution’s requirements, in particular:

- Following the institution's Change Control Process for approval, and working within the agreed timetable for upgrades.
- Full management of patches, including full testing and validation, applied during scheduled maintenance periods.
- Internal ISO27001 compliant processes for change control and releases.

Accessibility

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

Browser support

Haplo can be used on both PCs and Macs, and all modern standards compliant web browsers are supported, including Internet Explorer/Edge, Firefox, Safari, Chrome and Edge. We maintain ongoing support for the current and previous versions of these browsers (including for administration and authoring interfaces.) We commit to supporting all standards compliant web browsers. Haplo Research Manager can be used on PCs and Macs using Windows 7, Mac OSX and iOS.

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.

Mobile and tablet access

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.

Compliance

Audit trail

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 institution, 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. Only those with the relevant permission levels are able to view this information.

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. Changes are highlighted and amendments are linked.

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 institution systems, is recorded in the history of the record, including that it is a system-made change and when it occurred.

Data retention and disposal

Haplo Research Manager retains and disposes of all data including files in line with institutional records management policies.

Set up

Hostname (web address)

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

The institution 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 will provide you with a CSR in any reasonable format.
- create a DNS CNAME record to point the chosen hostname to [institution].infomanaged.co.uk

Branding

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

Email notifications

Haplo will send email notifications to users on behalf of the institution. There are a number of options for how we provide this, the two main options are 1) the institution 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.

Open source and access to closed source code

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.

- All code used to deploy the application, including both product and institutional customisations, is available at any time.

- A complete copy of the institution's data is available at any time (subject to cost recovery charges).

This provides full protection for our institutions.

Support

Help desk

Haplo Services provides second-line support. The institution IT team normally provide first-line support for issues such as logins and forgotten passwords, and the Research Office for issues regarding institution regulations and processes.

Haplo Services is based in London and our standard service help desk support is available between 09:00 and 17:30 Monday to Friday (excluding UK Bank Holidays).

Issues should be reported by email to our dedicated support email address which will automatically create a ticket in our support ticketing system enabling us to respond promptly and in line with our Service Level Agreement.

Please note that during the implementation and development phase, institutions should directly contact their Facilitator or Developer directly with any issues.

User documentation

Haplo Services provide comprehensive user documentation covering the use and administration of the system.

For end-users, Haplo Services provides illustrated guides covering all main functions and on-line help and guidance throughout the system.

Haplo Services provides an Administrator's manual (for key roles such as Research Office) and an IT Support guide:

https://www.haplo-services.com/support-library

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.

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).

Test systems will be refreshed with data from the live system on request. Our efficient hosting and processes enable us to clone environments with minimal effort. Refreshes are requested through the ticketing system, and require 2 working days notice.

Service Level Agreement

Availability target	
Service availability (not including planned maintenance)	99.5%
Service hours	24 hours a day, 365 days a year
Planned maintenance	Up to 1 hour per month

Service metrics	
Maximum transaction time (read & write, 99.97% of transactions)	2 seconds
Recovery time objective	4 hours
Recovery point objective	5 minutes

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 (UK clients)	
Monday - Friday, excluding Bank Holidays	8am, 7pm (UK local time)
Saturday - Sunday, and Bank Holidays	All day

Planned maintenance schedule for maintenance not exceeding 30 seconds (Australasian clients)	
Monday - Friday, excluding Bank Holidays	12 noon (UK local time)
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. *24/7 support for critical infrastructure issues.*	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

Response and fix time targets	Core service hours		Non-core service hours	
Severity	Response	Fix	Response	Fix
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 hours		
Service	Days	Time
Critical infrastructure issues (with automatic notification from monitoring system)	7 days a week	24 hours a day
Core service hours	Monday - Friday (excluding UK Bank Holidays)	09.00 - 17.30 (UK local time)
Non-core service hours	Monday - Friday Saturday - Sunday UK Bank Holidays	17.31 - 08.59 (UK) All day All day

As most support requests are simple issues which don't require follow-up clarifications from the client, Australasian clients should expect any tickets submitted by the end of their working day to be resolved by the start of their next working day.