Contents

1 Foreword

I would like to thank Hannes Leutloff, who has spent countless hours on this project voluntarily and without whose tireless work on the user interface, the build infrastructure, mentoring and code reviews, this project would not have been possible.

2 Introduction

2.1 Terminology

Block access An access to a block of secondary storage. Since accessing secondary storage presents itself as the bottleneck in database latency, time complexity of database operations is often represented by the number of block accesses rather than the number of operations in RAM.

data subject

data client

survey item

Tool provider In the context of LTI, the term *tool provider* is used to describe a system which provides an external tool to an LMS, extending the LMS's capabilities.

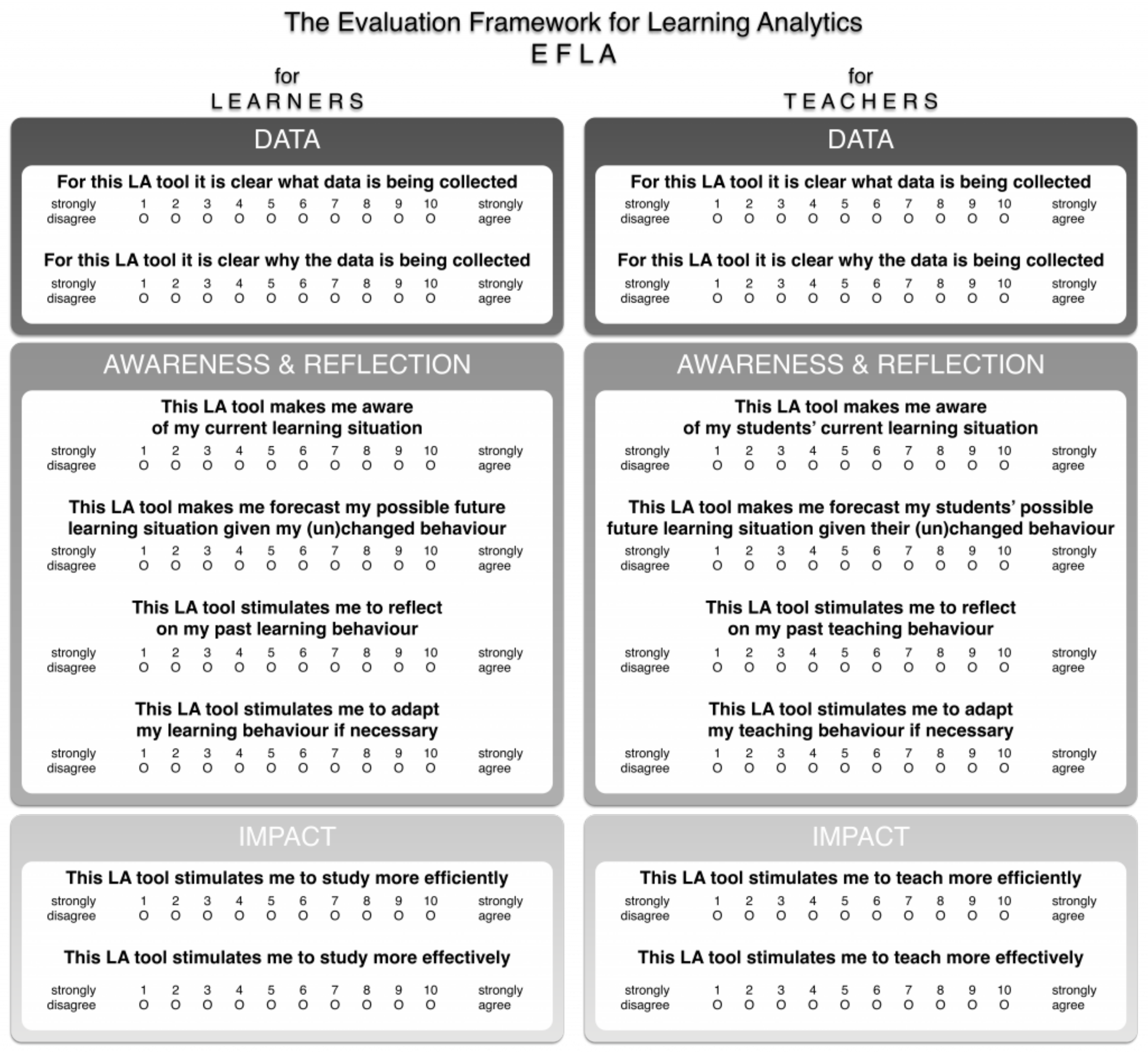
Learning record store A database system, possibly including analysis and visualisation capabilities, for storing data of interest to learning analytics applications. In the context of this thesis, LRS refers to a system for storing and analysing xAPI statements in particular. Examples for LRS systems are the TLA facts engine an HT2Labs's Learning Locker **??**.

Learning management system A content management system, specifically designed for e-learning applications. Examples for LMSs are Moodle and OLAT.

2.2 Previous Work

As part of the computational humanities seminar, which was held by Prof. H. Drachsler in winter 2017/18, Hannes Leutloff and I were tasked to digitize the evaluation framework for learning analytics (EFLA) (Sch17a) and to develop an online platform where the survey could be taken.

The result is an online survey platform, where surveys similar to the EFLA can be created and hosted. While it is possible to create arbitrary survey items, some restrictions specific to the EFLA use-case apply (for a full list of features see table 2). The original version of the survey tool is written in Python 3 and JavaScript for the server and user interface respectively. To be able to re-use code, the choice of language did not change with the new version.



The Evaluation Framework for Learning Analytics
EFLA

for LEARNERS

DATA

For this LA tool it is clear what data is being collected
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

For this LA tool it is clear why the data is being collected
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

AWARENESS & REFLECTION

This LA tool makes me aware of my current learning situation
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool makes me forecast my possible future learning situation given my (un)changed behaviour
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to reflect on my past learning behaviour
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to adapt my learning behaviour if necessary
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

IMPACT

This LA tool stimulates me to study more efficiently
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to study more effectively
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

for TEACHERS

DATA

For this LA tool it is clear what data is being collected
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

For this LA tool it is clear why the data is being collected
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

AWARENESS & REFLECTION

This LA tool makes me aware of my students' current learning situation
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool makes me forecast my students' possible future learning situation given their (un)changed behaviour
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to reflect on my past teaching behaviour
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to adapt my teaching behaviour if necessary
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

IMPACT

This LA tool stimulates me to teach more efficiently
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

This LA tool stimulates me to teach more effectively
strongly disagree 1 2 3 4 5 6 7 8 9 10 strongly agree
O O O O O O O O O O

Figure 1: Template for the EFLA survey (Sch17b)

Feature	Description
Editable surveys	Edit titles, question texts and colors.
Export to CSV	Export all results of a questionnaire to as CSV file
Challenges	Validate data subject's email addresses. White- and Blacklist email addresses. Protect surveys by password.
Internationalisation	Survey items may have multiple translations.
User accounts	Users may sign up and host surveys.
Template surveys	Choose from a fixed set of template surveys.
Visualisation	View survey results as a box plot.

Figure 2: Features of the previous version of the survey tool

2.2.1 Data Model

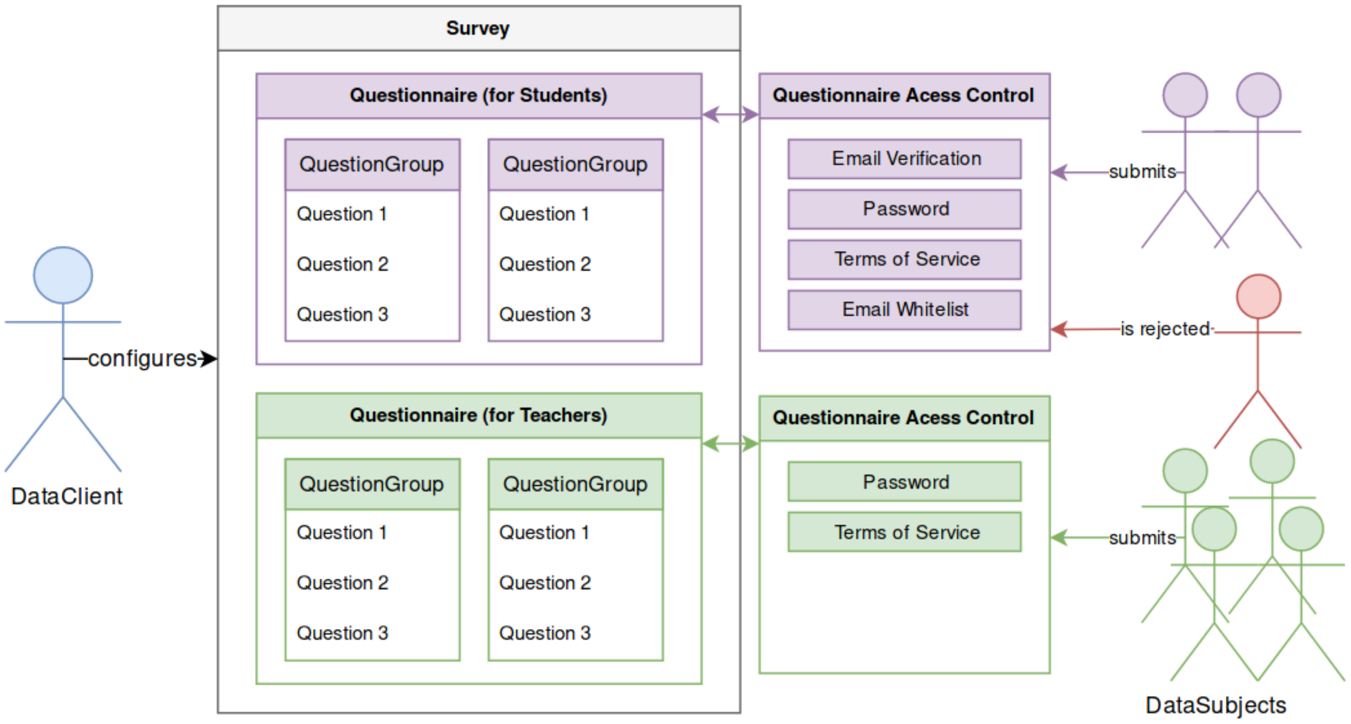


Figure 3: Data model of the previous version

The data model for the previous version is closely coupled to the specific needs of the EFLA survey, where a single survey consists of two questionnaires, targeting learners and teachers respectively. Each questionnaire controlls DataSubjects' rights to submit by it's own set of access control modules, so that audience groups can be discriminated. Each questionnaire contains one or more question groups, which contains of one or more questions.

2.2.2 Architecture

The architecture for the previous version follows a simple client-server paradigm, where a web browser takes the role of the client. The server software consists of an application server, responding to API requests, a database and an in-memory key-value store for storing session data. There are two different user interfaces, one for data clients and one for data subjects. The data subjects' user interface is the page, where the survey is filled out and submitted. This page is rendered on the server using a templating engine and sent to the browser as a static HTML docu-

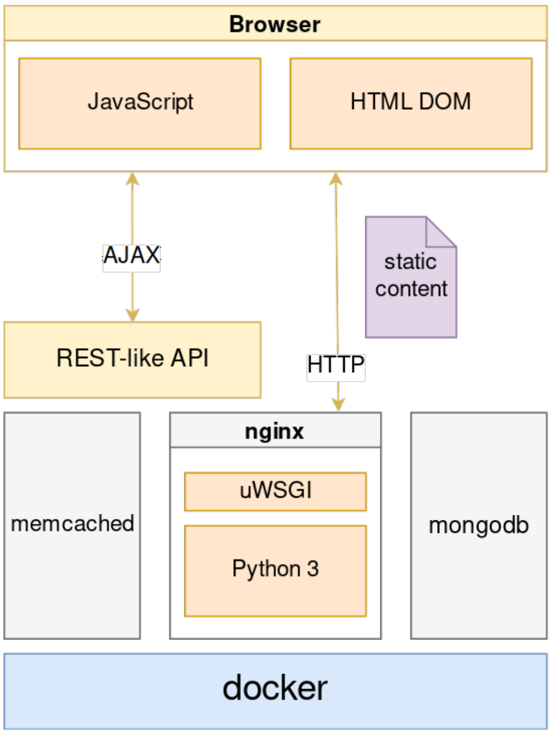


Figure 4: Architecture overview of the previous version

ment. The user interface for data clients presents
the user with an editor, where survey content can be created and modified. This user interface is realised as a one-page javascript application and was developed by Hannes Leutloff. Deployment is handled through Docker, with the help of the docker-compose tool.

2.3 The TLA infrastructure

Prof. H. Drachsler's work groups' current efforts are targeted towards the implementation of a trusted learning analytics (TLA) infrastructure. This infrastructure will provide big data storage and analysis capabilities while complying with the european general data protection regulation (GDPR). A central part of this infrastructure are data providers. These data providers are tools which collect data and publish it to the TLA facts store, using the xAPI statement format as a common data representation. The aim is to not only create the infrastructure to collect and analyse data, but also to build and ecosystem of external tools and data providers which interface with TLAs core components.

2.4 Goal

The goal of this thesis is to extend the functionality of the already existing survey tool and to contribute to the TLA ecosystem by integrating the survey tool as a data provider.

2.5 Motivation

Online survey tools are not a new invention by any stretch of the imagination. An online search yields dozens of already existing survey platforms, both commercial and free to use. Some of the biggest competitors in this business are compared in table 1. For most use cases, one of the existing solutions will be sufficient. The survey tool described here aims at solving a unique challenge not solved by any competitor the author is aware of. This challenge is integration with Moodle and the TLA infrastructure for use at the Goethe University in Frankfurt Main. Features needed for this integration are single sign on, i.e. automated authentication of survey participants between Moodle and the survey tool, support for embedding via the LTI protocol, as well as support for xAPI as a data exchange protocol. These requirements are discussed in greater detail in the Section "Requirements Analysis". While most survey providers, provide single sign on features, most only provide these capabilities to enterprise customers. Also, only one of the competitors looked at in this thesis supports CAS as a mechanism for this. None of the competitors supports data export as xAPI statements out of the box and require either manual programming for every survey or data conversion after export. Out of the listed competitors, only one supports the LTI protocol for embedding the survey into Moodle and the only documentation of this feature is an online video **??**. For these reasons, an alternative survey tool was developed.

Provider / Feature	Google Forms	SurveyMonkey	QuestionPro	qualtrics	survey gizmo
SSO	For managed accounts in G-Suite via SAML and LDAP **??**	Supported, but further information is only provided to potential customers. **??**	Through SAML and HMAC-SHA1 **??**	Through CAS, LDAP or SAML **??**	Through SAML **??** **??**
xAPI	Not supported, but has limited support for programmable event handlers. **??**	Not supported, but supports polling of survey results through an API. **??**	Not supported, but has support for web hooks. **??**	Not supported, but supports polling of survey results through an API. **??**	Not supported, but has limited support for web hooks. **??**
LTI	Unsupported	Unsupported	Unsupported	Alledged experimental support, no official documentation exists.	Unsupported

Table 1: Comparison of survey providers

2.6 Scope

In this thesis, the overall structure of the survey platform and the server-side software stack, which was designed and developed by Noah Hummel is discussed. Implementational details are discussed only if they are central to the workings of the software, or solve a particular conceptual challenge. The user interface was designed and mainly developed by Hannes Leutloff, and is therefore not discussed in this thesis.

3 Requirements Analysis

Requirements analysis was performed on the basis of previously collected use cases. Required features and parts of the existing software that had to be refactored were identified.

3.1 Use Cases

3.1.1 Accessing Students' Self-Regulated Learning Behaviour

In the context of TLA, the survey tool will be used to access students' self-regulated learning behaviour. Several psychometrical instruments exist for this, but for the purpose of this thesis only instruments based on or resembling the motivated strategies for learning questionnaire (MSLQ) were considered. One such instrument described in the 1994 paper "Lernstrategien im Studium : Ergebnisse zur Faktorenstruktur und Reliabilitat eines neuen Fragebogens" (SW94) in particular was used as a guide to define new requirements for the platform.

3.1.2 Embedding in Moodle

Moodle is an LMS used at Goethe University. The platform supports embedding of external tools inside of a course context. Because previous and simultaneous efforts by Prof. Drachsler's work group and collaborators already target Moodle, the new survey tool should also integrate with it. Surveys are still created using the survey tool's own user interface. Once a survey is created, it may be embedded into Moodle's course context via the LTI protocol. Data subjects already using the LMS may then directly participate in the survey without leaving the website.

3.1.3 Stand-Alone Survey Platform

In addition to directly embedding the survey tool into an LMS, it should be possible for a data subject to participate in a survey, if no LMS is used. This use case was carried over from the previous version.

3.1.4 Data Provider as Part of the TLA Ecosystem

As metnioned before, the survey tool will contribute to the TLA ecosystem by providing data to TLA's facts engine, using xAPI to communicate information about responses and events on the platform.

3.2 Refactoring & Restructuring

3.2.1 Database Abstraction

The previous version of the survey tool uses a self-authored database abstraction library, the object-document-mapper (ODM). This library is capable of persisting python objects and relationships between them. It does this by serializing objects to JSON and storing them using MongoDB. Runtime

ORM / Feature	SQLAlchemy	PeeWee	PonyORM	SQLObject
Declarative API	Yes	Yes	Yes	Yes
Eager loading of relationships	Configurable	No	Configurable	No
Lazy loading of relationships	Configurable	Yes	Configurable	Yes
Query caching	Yes	Yes	Yes	Yes
Cascades	update, delete	update, delete	delete	delete
Inheritance mapping	Yes	No	Yes	No
JSON columns	Using PostgreSQL	No	Yes	Using PostgreSQL
First released	2005	2010	2014	2003

Figure 5: Comparison of python ORM libraries

interactions with these objects are intercepted and converted into database queries using the descriptor pattern (https://docs.python.org/3.7/howto/descriptor.html). This approach worked well for the limited use case, but it proved difficult to implement a transaction model that follows ACID properties. To achieve transaction safety without compromising API transparency, an existing object-relational-mapper (ORM) was chosen to replace the ODM. Key factors for the choice of ORM were:

Maturity of the project. Data integrity is critical to the application. There is an increasing chance of bugs being found and resolved with increasing project age.

Support for inheritance mapping. Not being able to effectively use class hierarchies had proven itself to be a major burden on development speed during development of the previous version. Attributes of related classes had to be explicitely duplicated. This was not only confusing to read but also difficult to maintain.

Support for JSON or hash map columns. JSON columns provide a convenient way to store multiple translations of a text column, without producing additional joins between the main table and supplemental translation tables. Storing multiple representations of a string inside a relational database usually involves a 1 to n relationship between the internationalised entity and it's translations. With the use of JSON columns, multiple translations of a string can be stored as a dictionary inside a single column of the entity itself.

In table 5 a comparison between the some of the most popular ORM libraries available for Python 3 is presented. Because of the project age, previous experience with the library and it's reliability in production environments and the large feature set, SQLAlchemy was chosen to replace the ODM.

3.2.2 User Interface

As a simultaneous effort, a major rewrite of the user interface was done by Hannes Leutloff.

3.3 Survey Content

To allow other surveys apart from the EFLA survey to work well with the platform, the following requirements were established:

Response ranges should be adjustable. Responses will still be on a discrete scale with a step size of 1. Start and end of the scale should be adjustable.

Labels for the response ranges should be adjustable. The EFLA survey uses the phrases "Strongly disagree" and "Strongly agree" to describe the lower and upper ends of the response scale respectively. Other surveys may use different descriptions, hence these descriptions should be adjustable.

Question order should be configurable and optionally randomizable.

3.4 Template Management

The set of questionnaires which are of interest is changing over time. Initially, the MSLQ was considered to be the main questionnaire of interest for accessing self regulated learning behaviour. Because of the questionnaire's size, alternatives were researched by Prof. H. Drachsler and associates. As new research is published and questionnaires become better statistically validated, the number of survey items needed to access a certain research question may decrease. For this reason, more current questionnaires may be preferred over older ones. It should be easy to integrate new survey items in the future. At the same time, once a questionnaire is reasonably well verified, multiple users may be interested in using the same questionnaire for different groups of participants. The previous version has rudimentary template support, but templates are supplied as static YAML files on build and can not be updated from the user interface. To overcome this limitation, the following requirements were established:

Templates should be user-contributed. Data clients with special privileges, called contributors, should be able to publish new templates using just the user interface.

Individual survey items may be templates. To create slightly modified versions of a questionnaire, it is useful to re-use existing survey items and not just entire questionnaires.

Templates should be modifiable after creation. Otherwise, editing errors would cause the template to become unusable and the entire template has to be created again.

Changes should be traceable. When the maintainer of a template modifies the template, other users who own copies of it should be able to review the changes.

3.5 Support for xAPI

The previous version only supports limited data analysis capabilities. The new version should interface directly with the TLA infrastructure to allow for further data processing by TLA's analysis engine. TLA uses the xAPI statement format as it's common data representation. To provide data using xAPI to TLA, the following requirements were established:

Survey results should be published to an LRS via xAPI. A suiting xAPI representation of survey results has to be defined and should be transmitted to the LRS vie HTTP as defined by the xAPI specification (Adv16).

The destination LRS should be configurable. This allows the location of TLA to change in the future. It also decouples the survey tool from TLA and allows interoperability with other LRSs.

xAPI statements must not be lost due to failure. When publishing data to an external storage, data integrity is no longer just determined by database integrity. Appropriate measures must be taken to ensure re-transmission in case of network failure. If transmission is not possible due to misconfiguration, an offline fallback method has to be provided.

3.6 Embedding & LTI Launch

To achieve compatibility between Moodle and the survey tool described in 3.1.2, the following requirements were established:

The survey tool has to implement the LTI launch protocol. The LTI launch protocol is used to embed and launch external tools by the Moodle LMS. The survey tool has to recognize LTI launch requests and respond with an embeddable version of the survey.

Data subjects should not have to authenticate. When using the survey tool from within Moodle, data subjects have already authorized with CAS in order to log in to the LMS. Requiring another form of user authentication after this point would break seamless user experience.

LTI uses OAuth 1 to sign requests. This provides a way for the tool provider to verify the identity of the LMS and the authenticity of the LTI request. It was explicitely not required to implement OAuth as part of this work.

3.7 Privacy Considerations

Since May 25th 2018, the general data protection regulation applies to members of the EU (The16). The author is not in any way qualified to provide legal interpretation on the regulation. However, in consultation with Prof. H. Drachsler the following requirements were identified to conform with the regulation:

Data deletion for data subjects. Some personal data has to be stored on the server in order to identify data subjects between requests. Because of the right to be forgotten, there has to be a way to delete personal data stored for a specific data subject. It is sufficient if this is not automated and can be performed by the site's administrator.

No user accounts are available to external data clients. If third parties are allowed to use the service, it can not be guaranteed that they will follow GDPR guidelines. To avoid responsibility for third parties' actions on the site, there will be no way to sign up as a data client that does not involve contacting the administrator.

Other rights covered by the GDPR include the right to view and export personal data. Since all personal data collected is published to the TLA infrastructure, this will be implemented as part of the TLA infrastructure and is not part of the survey tool.

4 Concept

4.1 Architecture

The overall architecture is the same as for the previous version. The survey tool uses a classical client-server approach, where the client is a javascript application running inside a web browser. This approach enables everyone with a modern web browser to use the platform without having to install any additional software locally. Notable exceptions are Internet Explorer, Opera Mini and the Blackberry Browser, as they do not support the CSS `grid` property fully. The server-side software stack is deployed by using docker and exposed through a containerized web server. Communication between different containers on the server takes place on a virtual network which is not exposed to the internet. Communication between client and server is handled by a RESTful API which is provided by the server.

4.2 Survey Content

During refactoring, the top-level organisational unit “survey” was removed and replaced by “questionnaire”. Grouping multiple questionnaires inside a single survey only makes sense in a small set of use cases. For other use cases, this has proven to be confusing to users. Most of the time, a survey contains only a single questionnaire. Use cases, where grouping of multiple questionnaires into a single survey is appropriate can still be achieved by creating multiple questionnaires without any explicit grouping. Apart from that, the overall survey structure did not change significantly.

4.2.1 Questionnaire

A questionnaire is a collection of dimensions, which also contains administrative information. A questionnaire’s life cycle is modeled using three distinct states. The questionnaire starts out as not published, which means that it is only accessible to it’s owner via the API (with the exception of templates, which will be accessible to other authenticated data clients) and will not display when accessed by data subjects via it’s public URL. This state is meant for questionnaires which are incomplete and worked on. Once a questionnaire reaches it’s published state, it will be visible to the public via it’s public URL. It will not accept submissions via the API at this point. The publicly accessible representation of the questionnaire will not display form controls to submit and display an information page before accessing the survey content. This page informs the data subject, that submissions are not accepted at this point. Once data subjects are allowed to submit answers, the questionnaire enters the “accepts submissions” state. The public page will now display form controls to submit and new submissions are accepted by the API. Once the survey has concluded, the questionnaire returns to the “published” state and can be viewed for future reference. If this is not wanted, the data client has the option to retract the survey and return it to it’s “NOT published” state. This life-cycle model is illustrated in figure 6.

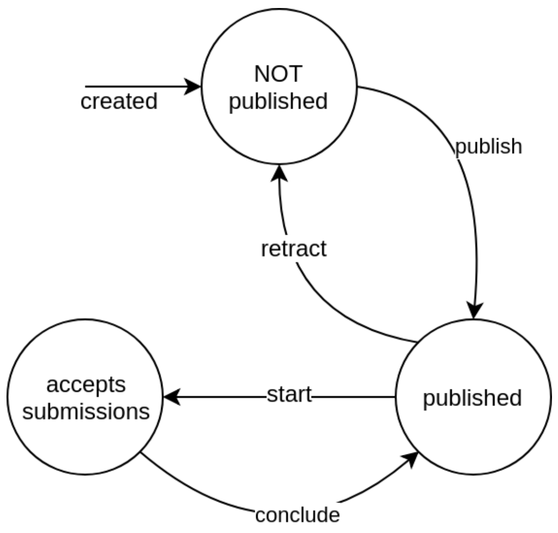


Figure 6: Questionnaire life-cycle

Along with information about it’s life cycle, the questionnaire also stores information about any access control challenges presented to data subjects. As in the previous version, challenges are presented as additional form inputs when submitting. In contrast to the previous version, email addresses are now always validated when submitting by sending a one-time-use token to the entered email address. Data clients may choose from these challenges:

Password: A password has to be entered in order to submit. The password is chosen by the data client.

Email whitelist: Only emails that are present in a list of email addresses are allowed to submit. The email whitelist also supports wildcard expressions to allow all email addresses following a certain schema.

Email blacklist: The same as email whitelist, but instead of allowing certain email addresses, this blocks certain addresses from participating.

4.2.2 Dimension

A dimension is a collection of questions usually pertaining to a specific topic. The only additional piece of information handled by the dimension is whether the order of it's questions should be randomized when the survey is taken.

4.2.3 Question

A question is a single statement to which the data subject may respond on a numerical scale. Lower and upper bounds of this scale may be adjusted by the data client, as well as the descriptions for these bounds. Individual questions may be used as templates, which is why the information on lower and upper bounds and scale labels is stored on a per-question basis. Convenience workflows for updating this information for an entire dimension allow easy editing despite the great granularity of this approach.

4.3 Internationalisation

Every time an API request is made, a request language is determined by the server and the request is handled in this language. For human readable attributes of survey items, multiple translations may exist. When a survey item is created, the language it is created in becomes the item's original language. Translations in different languages may be added later by updating the survey item using the desired language as the request language. When no translations for the requested language exist, survey items are served in their original language instead. To communicate information about existing translations and what translation was served, the API includes the item's current language, original language and a list of available languages in the JSON representation of the item.

4.4 Ownership, Parties, Roles

To control API access to survey items, survey items had a single owner in the previous version. Owner refers to a data client who has access to the resource. In the next version, the ownership model was expanded to allow n-to-n relationships between owning parties and owned resources. Ownership is also no longer restricted to data clients. These changes became necessary for two reasons. Survey item are no longer the only resource with access restrictions, as will become clear in section **??**. While survey items ususally only have a single owner - the author - tracking information has to be accessible to all parties sharing read access to the tracked resource. Because of the template management, data clients may have read access to survey items of which they're not the author (templates). This requires a single resource to have multiple owners. Of course, a party may own more than a single resource, hence the n to n cardinality of this relationship. The reason for expanding ownership to data subjects is that the mechanism makes it easy to identify personal data. Data subjects own all resources which contain personal data. The "right to be forgotten" may then be implemented by simpy retrieving all owned objects for a given data subject and deleting them from the database.

For some priviliged users, access control should not be enforced. Also, the right to publish templates is reserved for a select group of users. To accomplish different levels of privilege, roles were introduced to all parties. A party may have a number of different roles, depending on the actions they may take and the data they may access. There are currently five different roles:

Role	ID	Description
Root	0	All available access control methods grant access to users with this role.
Admin	10	An administrative user, who may view and modify other user's data in order to ensure proper operation of the platform.
Contributor	20	A user who may make survey items available to other users as templates.
User	30	A regular user who may create and modify their own survey items.
Unprivileged	40	A user who may view and participate in published surveys, but not create or modify survey items. This role is used for all data subjects.

Access control may then be enforced by checking if the role required for a certain action is held by the user in question.

4.5 Modification Tracking

To track modifications of survey items, every time a modification is made to an item, a record of this modification is stored in the database. To keep storage space needed for this feature to a minimum, only the most recent modification for each attribute is stored. During development it was discovered that different kinds of modifications require different information to be stored in order to create a meaningful record of the change, which can also be presented to the data client. Such information includes: The modified item, the modifying data client, previous and new values, as well as the point in time when the modification occurred. Five different types of records were identified:

#	Description	Stored information
1	An attribute was updated	Modified item, modifying data client, attribute name, previous value, new value, timestamp
2	A language map was updated	Modified item, modifying data client, attribute name, modified language name, previous value, new value, timestamp
3	A child item was added	Parent item, modifying data client, child item, timestamp
4	A child item was removed	Parent item, modifying data client, child item name, timestamp
5	A questionnaire was removed	Name of the questionnaire, modifying data client, timestamp

The modified item is stored as a reference to the actual record of the item in the database. This makes it possible to quickly find the modified item based on a tracking record. In the user interface, this is used to display the modified item as a clickable link, which will show the modified item instantly. For types of modifications, where an item was deleted, this kind of data model is not applicable, as the referenced record won't exist in the database anymore. In these instances, only the name of the item is stored. A special case exists for the deletion of questionnaires, as they don't have parent item needed to construct the tracking record. In this case, only the questionnaire name is stored. To deliver a personalized stream of modifications to data clients, including only those modifications which are of interest to them, tracking records also use the ownership model. Tracking recors are owned by all parties interested in the tracked item.

4.6 Template Management

A template refers to a survey item, from which other survey items ay be created as copies. A copy of a template should always mirror the template's content, including it's relationships to it's children, as the children are part of the item's content. This is true, since all parent-child relationships of survey items are aggregations; a dimension has no purpose without any questions and a questionnaire has no purpose without any dimensions. Every survey item may optionally also be a template. Survey items may be made available as a template by any data client who has at least the `Contributor` role. Templates are visible to all data clients.

4.7 xAPI Support

4.7.1 Introduction to xAPI

xAPI, formerly known as TinCanAPI, is a data exchange standard closely resembling activity streams (Wor17). It was developed by the Advanced Distributed Learning (ADL) Intiative as a successor to SCORM (Adva, Advb) and allows the exchange of experiential information in the form of statements (Advc). These statements use JSON as their data format and include a minimum of three semantical objects, “actor”, “verb” and “object”. The data is transmitted using HTTP, following REST principles.

```
{
    "actor": {...},
    "context": {
        "contextActivities": {
            "grouping": [
                {...}
            ],
            "parent": [
                {...}
            ]
        },
        "extensions": {
            "http://activitystrea.ms/schema/1.0/place": {...}
        },
        "language": "en",
        "platform": "st3k101 via localhost"
    },
    "id": "896ef7f1-8d5c-4729-b028-e9d72df47fe8",
    "object": {...},
    "result": [
        {...}
    ],
    "timestamp": "2018-09-27T14:32:15.009513",
    "verb": {...}
}
```

Figure 7: Anatomy of an xAPI statement

An xAPI statement may also include a “timestamp” stating the issue date and time of the statement, a “context”, providing additional information about the event and a “result”, detailing the outcome or outcomes of the event. The format is also extensible, additional information can be provided in the “extensions” object inside the “context”.

4.7.2 xAPI Statement Design

There are several actions which will trigger sending of an xAPI statement:

1) A data client logs in into the survey platform.

2) A data subject launches an embedded survey.

3) A data subject answers a single question.

4) A data subject answers a known survey item (whole questionnaire or dimension)

5) A data client updates the xAPI activity ID of a survey item.

Figure 8: List of cases where xAPI statements are emitted

For each of these cases, an xAPI statement had to be designed. All of these statements share at least some of their structure. The `context` object of all xAPI statements emitted by the survey tool includes the `platform`, `language` and `extensions` attributes. The `language` attribute always contains an RFC 5646 (PD09) compliant representation of the language the action was performed in. The `platform` attribute always starts with the string `st3k101`, identifying the origin of the statement. The `platform` attribute may also include the URL of the source LMS in the case LTI was used. In this case, the URL is appended as `st3k101 via SOURCE_LMS_URL`. The `extensions` object of the `context` also includes a geolocation for the client in RFC 7946 compliant GeoJSON format (BDD+16). Below is a concept for what additional data should be included in the statements listed in 8.

#	Actor	Verb	Object	Result	Context
1	data client	logged in	login page	-	-
2	data subject	accessed	questionnaire	-	-
3	data subject	answered	question	response value	parent dimension AND questionnaire
4	data subject	answered	qestionnare OR dimension	response value	parent item, if any
5	data client	updated	questionnaire OR dimension OR question	new acitivity ID	-

Table 2: Concept for information included in emitted xAPI statements

Objects are identified by their `objectType`, `type` and `id` in xAPI, whereas verbs are just identified by their `id`. For the purpose of the survey tool, all objects share the same `objectType`, the `Activity`. It is common practice to use a URL as identifier or type, which will return a human readable description of the item via HTTP GET. In theory, there's no correct identifier to use when designing xAPI statements, as the standard does not prescribe the use of any specific verbs or objects. In practice, several registries with commonly used verbs and objects exist and should be consulted when choosing which identifier or type to use. This avoids re-definitions of already existing items and increases homogeneity among statements by different adopters of the standard. Examples for these registries are `xapi.vocab.pub` and `registry.tincanapi.com`. The verbs and objects used in table 2 had to be translated into already existing verbs and objects. The results are detailed in table 3. For some of the objects, no suitable definitions existed. For those objects, dummy identifiers or types were used, which follow the URL format, but use `http://fantasy.land/` as a prefix.

In order to corellate objects in emitted xAPI statements with survey items in the survey tool, all survey items have a user-modifiable xAPI acitivity ID associated with them. This identifier is used as the object's `id` in xAPI statements. These identifiers are not modifiable in copies of templates, which means that all instances of a copy will use the same xAPI activity ID. This is useful for conducting meta-analyses, where all results for a certain template could be included in the analysis, regardless

Verb	Identifier
logged in	`https://brindlewaye.com/xAPITerms/verbs/loggedin`
accessed	`https://w3id.org/xapi/dod-isd/verbs/accessed`
answered	`http://adlnet.gov/expapi/verbs/answered`
updated	`http://activitystrea.ms/schema/1.0/update`

Object	Type
login page	`http://activitystrea.ms/schema/1.0/page`
questionnaire	`http://id.tincanapi.com/activitytype/survey`
dimension	`http://fantasy.land/dimension`
question	`http://adlnet.gov/expapi/activities/question`

Table 3: Used xAPI verb identifiers and object types

```
"object": {
    "definition": {
        "description": {
            "en-US": "This is a particular scale of a survey, it usually contains multiple questions."
        },
        "name": {
            "de": "5. Anstrengung"
        },
        "type": "http://fantasy.land/dimension"
    },
    "id": "<bla@blubl.net>:lernstrategien_wild_schiefele--5_anstrengung",
    "objectType": "Activity"
}
```

Figure 9: Example of how a dimension is represented as an xAPI acitivity object

of who conducted the survey. During testing, this presented itself as an issue, because results for the same template, which were collected by different data clients would not be distinguishable, as they all used the same identifier. For this reason, the email address of the data client who conducts the survey is added as a prefix to all xAPI activity IDs before sending. In figure 9, an example of how survey items will be represented in xAPI is given.

Actors may be represented in four different ways using xAPI. The identifying feature is either the person's email adress in the case of the `mbox` and `mbox-sha1sum` actor types, an account id in combination with a URL where the account is located in the case of the `account` actor type or OpenID identitiy in the case of the `openid` actor type. Data clients are represented as `mbox` actor types, while data subjects may be represented by as `mbox-sha1sum` actor types or, in the embedded use-case, as an `account` actor type using their LTI user identifier. The latter is necessary, because the email address might not be available though LTI. The LTI user ID is, contrary to prior suppositions, not useful for data analysis, as Moodle will use the user entities database ID. Moodle does however communicate a username via LTI, which for the specific Moodle instance at Goethe University is the same as the data subject's CAS ID. To retain compatibility with other LMSs while taking this finding into account, the username will take precedence over the LTI ID, if it's present in the LTI request. The LTI ID is also not globally unqiue, as there may be separate LMSs which use the same IDs for different users. For this reason, the LTI user ID is prefixed with the identifier of the source LMS, which is always present in LTI requests.

The recipient of the statements is determined by the object which is acted upon. This makes intuitive sense, as the person owning a certain survey item is the one interested in collecting data on it. All survey items are on the highest level organized into some questionnaire and there’s no use-case for different data consumers for individual parts of the questionnaire. For this reason, recipients are configured on a per-questionnaire basis. For objects which don’t have an owner, for example the survey tool’s login page, a default recipient is configured system-wide. To recover from failure in the case that an xAPI statement can not be transmitted over a prolonged period of time, the statement as well as the recipient and timestamp of the failure are logged to file and may then periodically be recovered.

4.8 Support for LTI

The embedded user interface is launched by the LMS using the LTI protocol. To identify the source LMS, a random token, the consumer key, is generated in the back-end for every questionnaire. To embed a questionnaire within the course context, an LTI request with the correct combination of request URL and consumer key has to be made by the LMS. Information about the user is already present in the request body and is used to identify the data subject. Since there’s some time between LTI launch request and survey submission, user information has to be stored on the server for this period of time. To achieve this, the required data is stored in the data subject’s account. If a data subject accesses the service through LTI for the first time, a new account is provisioned for them. When launching the survey tool via LTI, a session is created for the data subject and a session token is embedded into the user interface. Actions by the data client in the embedded user interface will use the embedded session token for authentication with the API. This mechanism allows the API to identify the data subject on every subsequent request. A valid LTI request is treated as sufficient authentication for the data subject, as the source LMS already authenticated the user prior to them accessing the questionnaire.

4.9 Email validation

When data subjects participate in a standalone survey, it is difficult to recognize repeated submission by the same person. Most features used for identification, for example the client’s IP address or cookies present in the browser can easily be modified by the data subject. Even more advanced measures, like browser fingerprinting, are ultimately controlled by the client, as communication with the server is handled by a publicly available API. For this reason, a third party has to be involved in validating the user’s identity. The survey tool achieves this validation through email. When submitting responses, the data subject has to enter their email address. At this point, the responses are stored on the server, but no xAPI statements have been emitted and the responses do not yet count towards the generated statistics. A randomly generated token is embedded into a URL pointing back to the survey tool. This URL is then sent to the submitted email address. Once the data subject follows the URL, the server will associate the token with the user’s responses in order to validate them. After validation, the appropriate xAPI statements are emitted.

4.10 Privacy Considerations

As mentioned in section 4.8, when a data subject participates in a survey, a user account is provisioned for them. Removal of this account and all associated personal data may be performed via the API when authenticated as an admin user. In order for the admin to know which account to communicate to the API, the API allows admin users to query for existing user accounts. Data removal is limited to admin users, as removal of accounts by data subjects themselves would require some sort of authentication mechanism for data subjects. When an email address is present for the data subject, authorization via email is a possible solution for this. In this scenario, the data subject would receive an email with a randomly generated token, which can be used to remove their personal data. If there's no email present for the data subject, for example when the account was provisioned using LTI, or if the data subject hasn't got access to their email account, this mechanism fails. Therefore, email as sole mechanism for data removal is not a viable option at the moment.

5 Implementation

5.1 Architecture

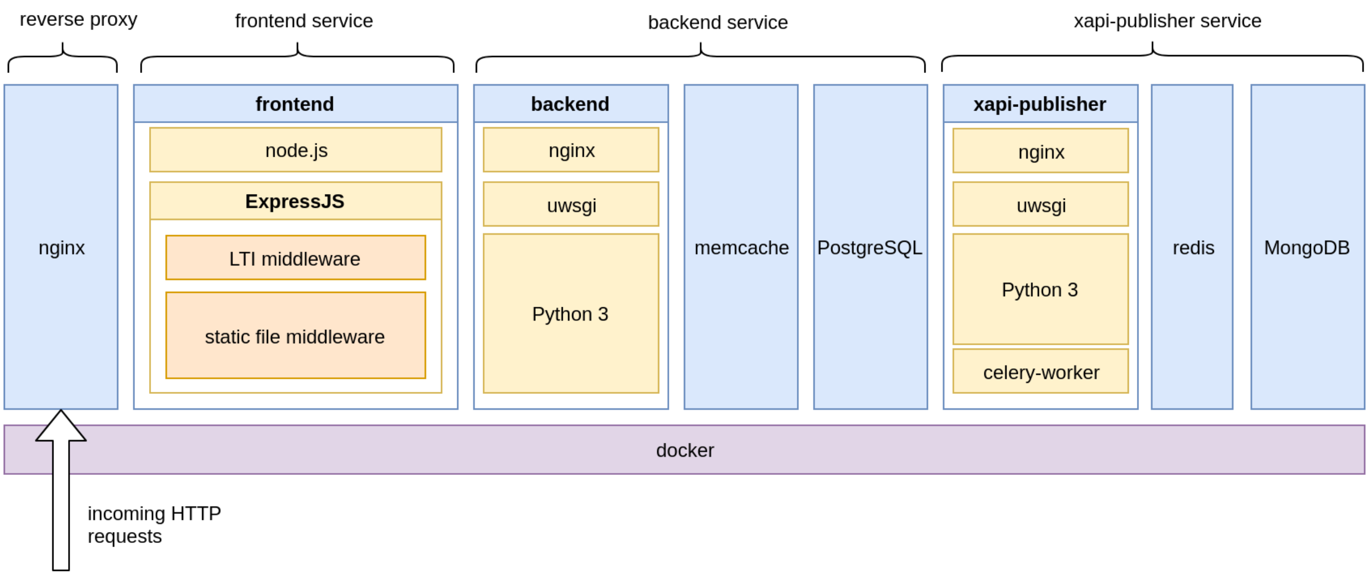


Figure 10: Server-side software stack

The server-side architecture is composed of multiple services. The “frontend” service is responsible for serving the user interface statically and intercepting LTI requests. The “backend” service is responsible for serving the API, user and session management and communication with the database. The “xapi-publisher” service is responsible for transmission of xAPI statements generated by the backend service. Each service uses slightly different technologies, depending on the requirements of the service. The API and “frontend” service are exposed through an nginx web server, which acts as a reverse proxy, forwarding incoming HTTP traffic to the appropriate docker containers.

5.1.1 The backend Service

The backend service is implemented using the Flask library, which allows Python to interface with a web server using the web server gateway interface (Eby03). Similar to PHP or CGI extensions, the nginx web server will accept incoming requests. It will then, using uWSGI, fork a python interpreter which will respond to the request. This means that no data can be persisted in the Python application itself, as the Python context will be discarded for every request. To persist data, a PostgreSQL database and a memcached in-memory key-value store are used. Persistent data, for example survey content, is stored in the database, whereas semi-persstent data, like user sessions, is stored in memcached. This design allows for horizontal scalability, as multiple instances of the `backend` containers can be used with the same database. No data dependencies between the `backend` containers exists. The only bottleneck in this scenario is the shared database. This could be solved long-term by replacing the single PostgreSQL instance by a database cluster.

5.1.2 The `frontend` service

The frontend service consists of a simple node.js application running the ExpressJS web server. For the web server, two middlewares are provided, one for serving static files and another for intercepting LTI launches. This is necessary, since the user interface has to be parameterized for each LTI launch before serving it to the data client.

5.1.3 The `xapi-publisher` Service

The xapi-publisher service mimics the design of the backend service closely. In addition to Flask and uWSGI, it also runs several worker threads, which are responsible for asynchronous sending of xAPI statements. Since xAPI statements are JSON documents, a document based database - MongoDB - is used instead of a relational database to persist xAPI statements between requests. For inter-process communcation between web server and worker threads, a task queue consisting of the celery library and the redis in-memory database is used.

The functionality provided by the xapi-provider service is separate from the backend service, duplicating most of the architecture already present. This might seem sub-optimal at first, as it violates the DRY (don't repeat yourself) principle to the extent that configuration for two separate containers has to be created and maintained. It does however allow for better separation of concerns. The backend container already handles business logic and user management. There is a unique set of challenges that has to be solved for publishing xAPI statements, which would unnecessarily increase the complexity of the backend service. One such challenge is, that sometimes data needed to create an xAPI statement is present in a request before the data is actually valid an should be transmitted. When a data subject answers a survey using the standalone survey interface, their answers are submitted to the server but only become valid, after they have verified their email address. The required data to build the xAPI statement, including the data subject's IP adress, has to be stored on the server until this point. In order to avoid convoluting the backend service's data model and business logic to account for this, a separate service solely responsible for temporarily storing and safely transmitting the xAPI statement is used. This allows the backend service to only use minimal logic for communicating with xapi-publisher service. It also allows maintainers of the codebase to make changes to the publishing behaviour without having to know the internal workings of the backend service.

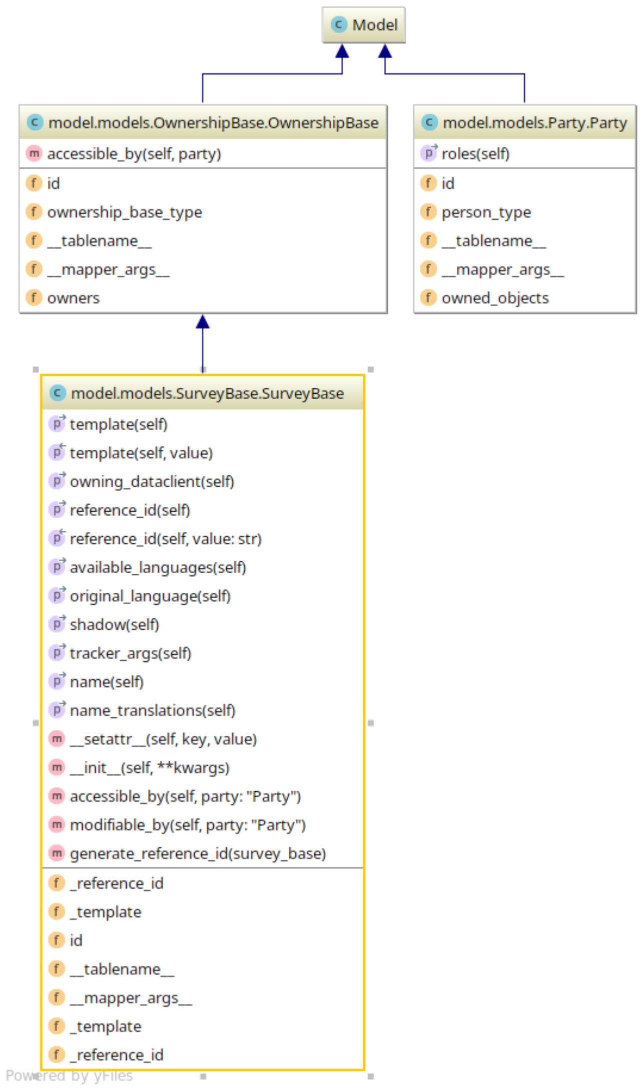


Figure 11: The three main base classes

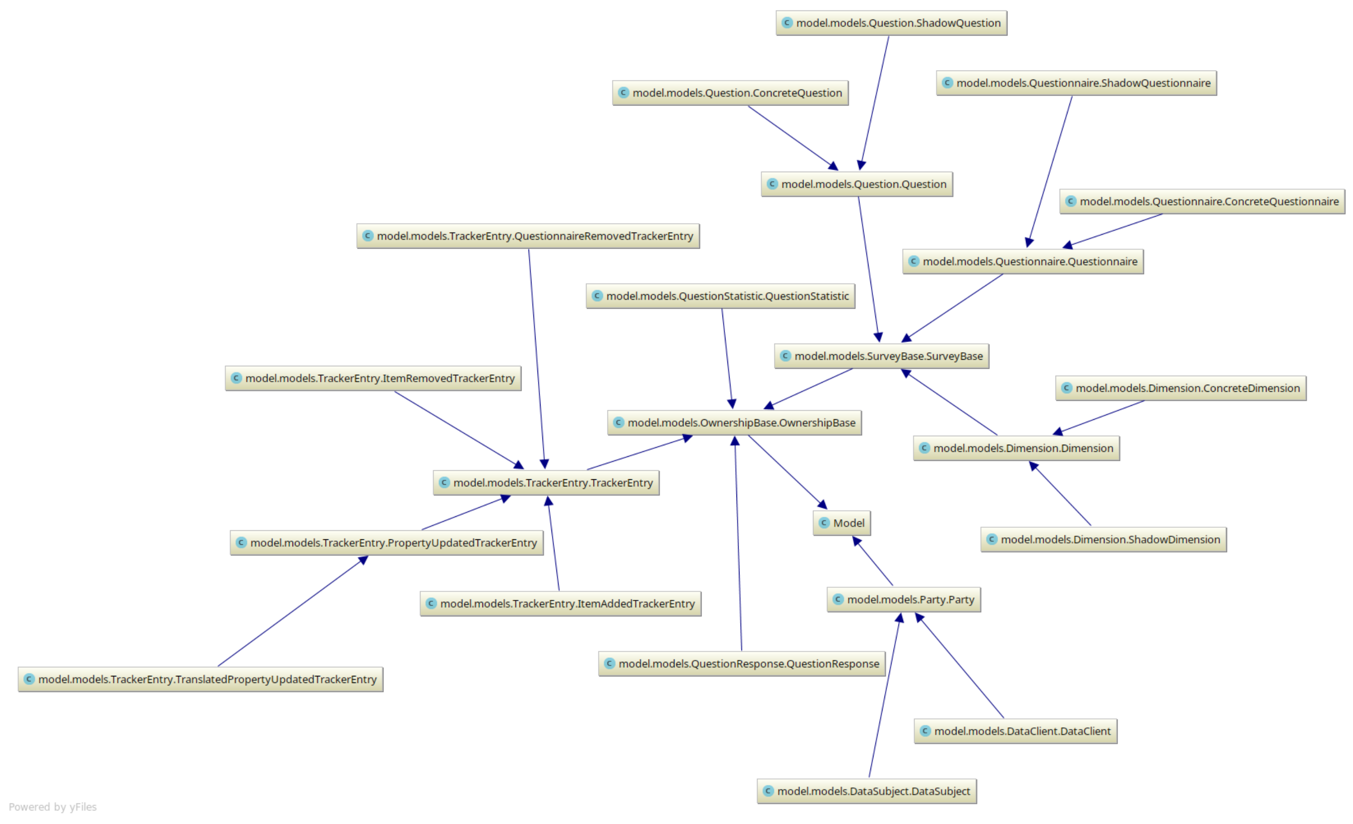

model.models.Question.ShadowQuestion
model.models.Question.ConcreteQuestion
model.models.Questionnaire.ShadowQuestionnaire
model.models.Question.Question
model.models.Questionnaire.ConcreteQuestionnaire
model.models.TrackerEntry.QuestionnaireRemovedTrackerEntry
model.models.Questionnaire.Questionnaire
model.models.QuestionStatistic.QuestionStatistic
model.models.SurveyBase.SurveyBase
model.models.TrackerEntry.ItemRemovedTrackerEntry
model.models.Dimension.ConcreteDimension
model.models.OwnershipBase.OwnershipBase
model.models.Dimension.Dimension
model.models.TrackerEntry.TrackerEntry
Model
model.models.TrackerEntry.PropertyUpdatedTrackerEntry
model.models.Dimension.ShadowDimension
model.models.TrackerEntry.ItemAddedTrackerEntry
model.models.Party.Party
model.models.QuestionResponse.QuestionResponse
model.models.TrackerEntry.TranslatedPropertyUpdatedTrackerEntry
model.models.DataClient.DataClient
model.models.DataSubject.DataSubject
Powered by yFiles

Figure 12: Model class hierarchy

5.2 Data Model of the `backend` service

5.2.1 Class Hierarchy

The class hierarchy is centered around three abstract base classes, `SurveyBase`, `OwnershipBase` and `Party`. `OwnershipBase` is the base class used for all classes which can be owned by a `Party`. `Party` is the base classes used for all user-type classes. `SurveyBase` is the base class for used for all survey items. Each of these classes provide common functionality and interfaces to their subclasses.

5.2.2 Inheritance Mapping via SQLAlchemy

The AQLAlchemy library provides three distrinct mechanisms for mapping class hierarchies to relational database schemas, *joined table inheritance*, *single table inheritance* and *concrete table inheritance*. Joined table inheritance uses a separate table for every class along the hierarchy, with each table only containing data declared by the corresponding class. Attributes inherited from superclasses are associated with subclasses by one-to-one relationships between superclass table and subclass table. When loading an instance of a subclass, a join statement is generated which also loads the appropriate record from the superclass table. Single table inheritance uses a single table for all subclasses of a certain class. Fields not used by a certain subclass are populated with `NULL` values. Concrete table inheritance uses a separate table for every class which contains all data needed to load an instance of the class. This means that inherited attributes will be duplicated in subclass tables. From a perfromance perspective, single or concrete table inheritance outperform joined table inheritance, since no join operation is needed to load an instance. From a storage perspective, joined table inheritance is optimal. Since SQLAlchemy does not support as many operations on concrete table inheritance as for the other types of inheritance mapping (SQL), concrete inheritance mapping was not used. While

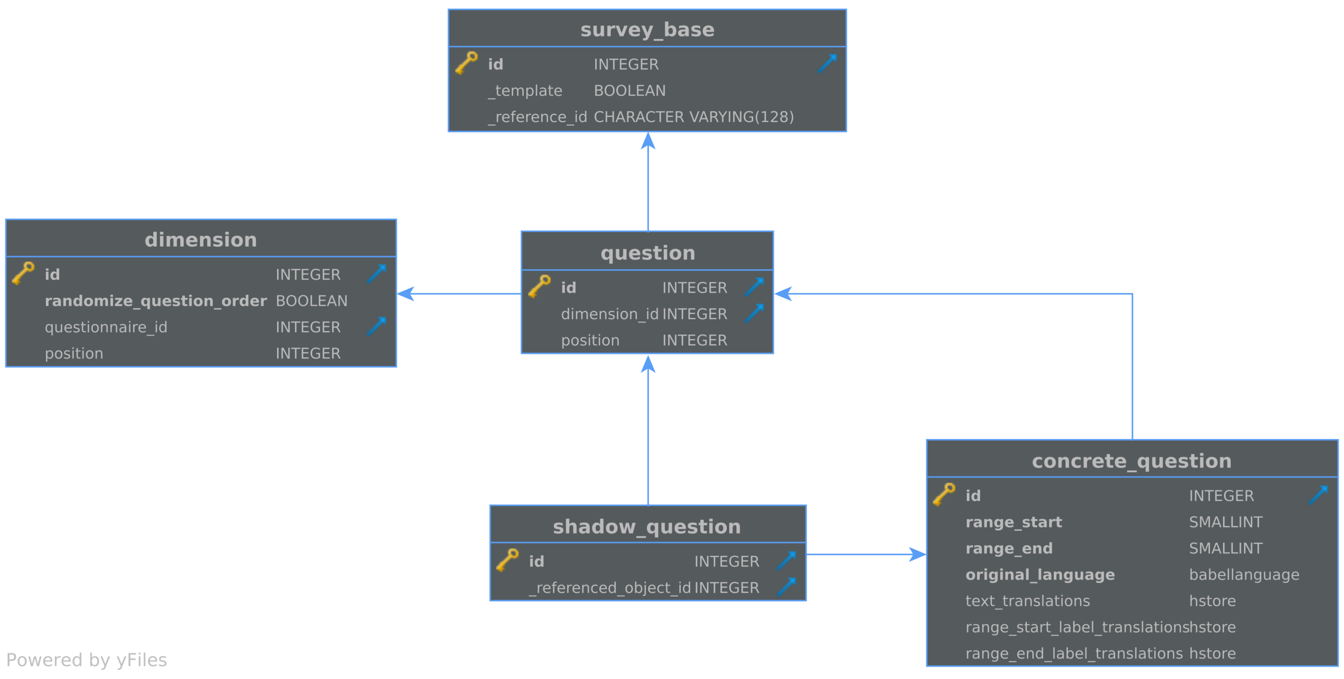


Figure 13: Database schema for a single question including foreign key constraints

single table inheritance works well for shallow class hierarchies, more complex hierarchies with multiple levels of inheritance produce only sparsely populated database records, as more attributes will be unqiue than shared for most classes. For the survey tool, this approach would only create two tables. For storage optimisation, the joined inheritance approach was used.

5.2.3 Template Management - The Template Triad

To allow user-contributed templates which may also be modified after creation, the same data structures are used for templates as for regular survey content. Creating copies of these templates is done by reference instead of by physically copying the template's content. The reasons for this is, that template content may be modified at any time, and these modifications have to be propagated to all existing copies. If copies were created by physically copying the template's content, updating a template would also cause all copies to be updated in the database. The asymptotical number of block accesses of this approach scales linearly with the number of copies present and potentially produces join operations with large result sets. It also stores large amounts of redundant data. Instead, copies are proxy objects, which delegate read operation to the referenced template. Survey items containg actual data are called concrete instances, whereas copies which do not contain the actual data are called shadow instances. For each direct `SurveyBase` subclass, `Questionnaire`, `Dimension` and `Question`, there are two more subclasses for the concrete and shadow representations of the class. In figure 14 and 13, this *template triad* is depicted. For the sake of brevity, the direct descendents of `SurveyBase` will be called the *super-classes*, it's descendents will be addressed by *concrete* and *shadow* respectively. For template management, each of these classes fulfills a certain role. All actual data is stored in the concrete- and super-classes, while shadow-classes only contain a reference to a concrete. As shown in figure 14, some data is stored in the super-class instead of the concrete class. The reason behind this design is, that some attributes of shadow classes may still be modified and should not just reflect the state of the associated concrete. An example of this is the access control configuration of the `Questionnaire` class. Data stored in the concrete class can be categorized as *survey content*, while data stored in the super-class can be categorized as *administrative data*. The super-class also acts as an interface for the

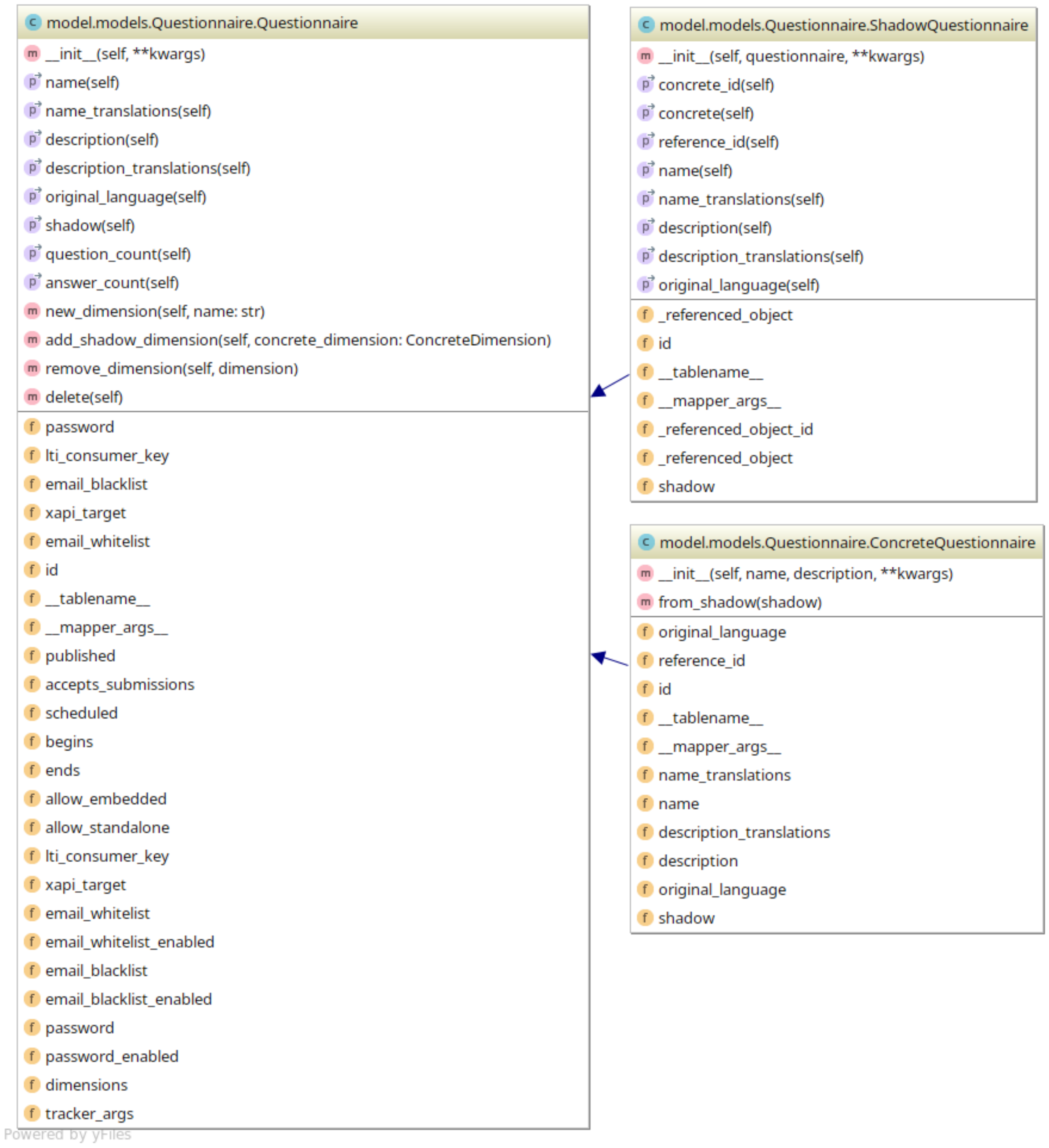


Figure 14: The template triad, exemplarily depicted by the `Questionnaire` data model

concrete and shadow classes, specifying all accessors that should be available for it's subclasses. The shadow classes act as proxies, which use Pythons `@property` decorator to provide transparent read access to the referenced concrete's attributes.

```
@property
def reference_id(self) -> str:
return self._referenced_object.reference_id

@property
def name(self) -> str:
return self._referenced_object.name
```

Figure 16: Example of transparent proxying of concrete attributes in shadow classes

5.3 Template Management

5.3.1 Creation of Shadow Instances

Creating a shadow instance from a concrete instance which does not have any relationships to other items is trivial. When creating a shadow instance from a concrete instance which has children, the resulting shadow instance should also duplicate the relationship structure of the reference concrete instance. In figure 17, an example for this is given. In the example, shadow A was created from concrete A. Parent-child relationships are modeled as a directed graph with edges from parent to child for the purpose of this example. The creation of shadow A will cause the entire subtree starting from concrete A to be duplicated with shadow instances, pointing to their corresponding concrete counterparts. The result is a copy of the concrete instance and it's relationship, which is always in synchronism with it's source, but can not be modified.

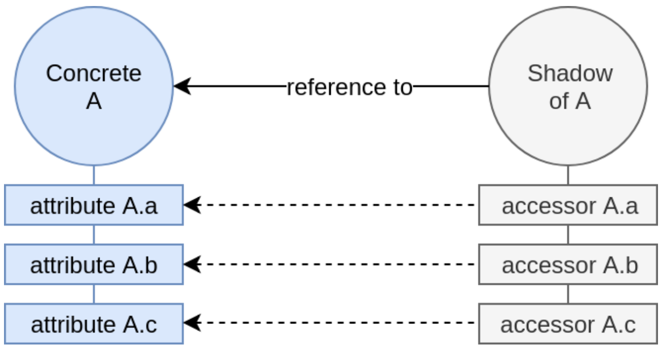


Figure 15: Schematic depiction of the relationship between concret & shadow instances

A special case occurs, when a shadow is created from a concrete, whose subtree also contains shadows. While it would be possible for a shadow insrance to reference another shadow instance, a shadow instance should always reference a concrete instance directly. The reason for this is the asymtpotical number of block accesses needed to read from shadow instance. If a shadow references another shadow, the `reference_to` relationship would have to be traversed multiple times until the concrete instance is found. Access times would then scale

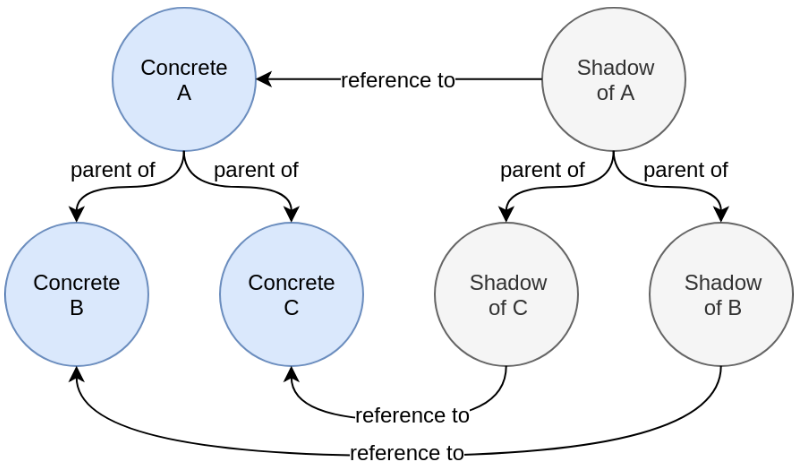


Figure 17: Shadow instances duplicate the concrete items' structure

linearly with the degree of separation between the accessed shadow instance and the referenced concrete. To avoid this, if a shadow instance is encountered while traversing a concrete's subtree on shadow creation, the shadow's `reference_to` relationship is traversed, until a comcrete instance is found. The new shadow instance is then created as a reference to this concrete instance. The example in figure 18 depicts this case. In the example, shadow A was created from concrete A. Concrete A's subtree contains a shadow instance, shadow #1 of B, which should be duplicated in shadow A's subtree. Instead of pointing shadow A's right child to concrete A's right child, shadow #1 of B's `reference_to` relationship is followed. Concrete B is encountered and becomes the referenced template for shadow #2 of B. This approach keeps the degree of separation between shadows and concretes always at one.

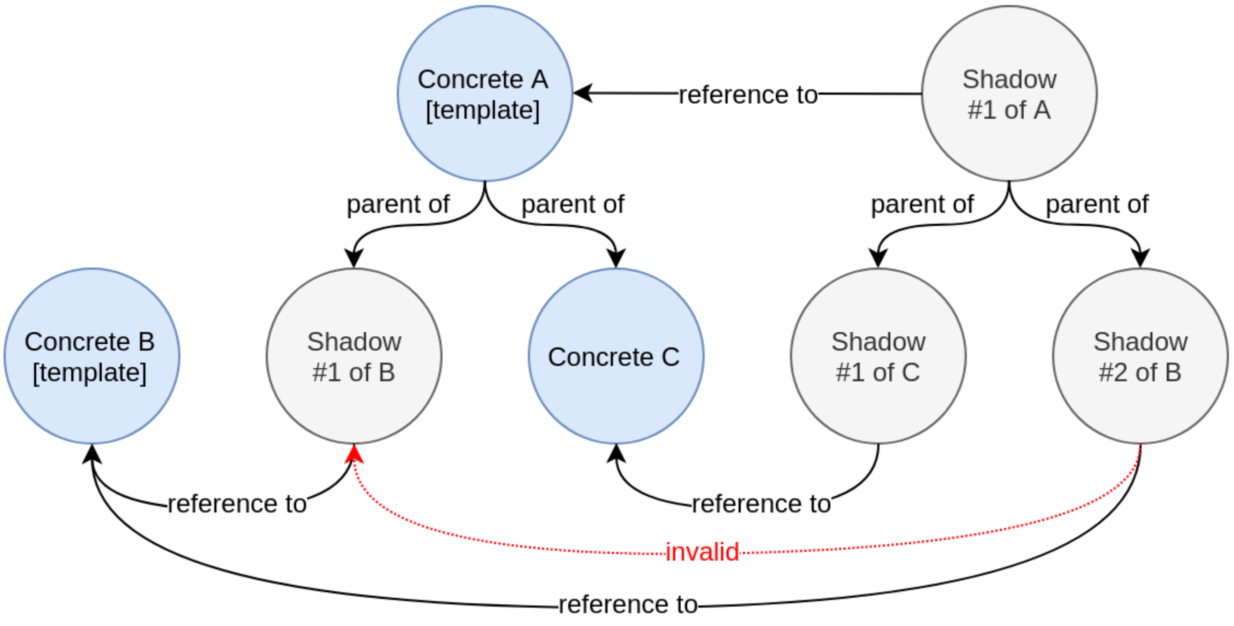


Figure 18: Duplication of a concrete structure which contains shadow instances

5.3.2 Modification of Templates

Modifications of template content is trivial, as shadow instances will always read from the referenced template. When modifying a template's relationships to it's children, the changes have to be duplicated in all copies of the template. To achieve this, the `reference_to` relationship is traversed backwards to find all copies of the template. The copies are then notified of the modification and apply it to their children as well. This operation is depicted in figure 19.

5.3.3 Deletion of Templates

When a template is deleted, copies of it become invalid, as they don't have any instance to point to anymore. Two possible solutions for this are either deleting all associated shadows, or converting all associated shadows into concrete instances. The former is less complex implementation wise, while the latter is more user-friendly, as it doesn't result in unexpected deletion of content. There are also use cases, where deleting all associated shadows might be intended, for example if the survey item is retracted and should not be used anymore. There is no default behaviour which satisfies all use cases. For this reason, a compromise between the two approaches was made. By default, when a template is deleted, all associated shadows are also deleted. Contributors are therefore encouraged to edit existing templates instead of deleting them. Templates will also show a counter showing the number of associated shadows in the user interface, making the contributor aware of possible repercussions. If the contributor chooses to delete a template anyway, the modification tracking feature will provide accountability and transparency to the affected data clients. When deleting a template questionnaire however, the associated shadows are converted into conrete instances, as the questionnaire might already be published and available for participating. In this case, deleting the questionnaire would interfere with another data client's survey and would delete already collected results, which is not acceptable.

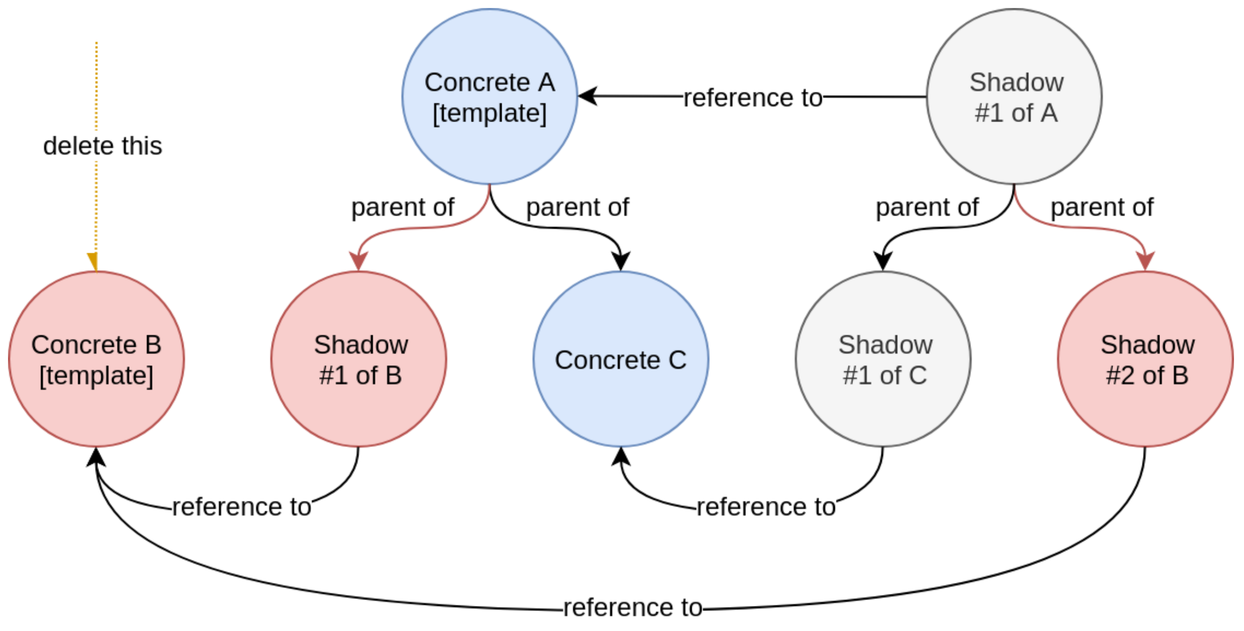


Figure 20: Deletion of a template is propagated to all associated shadow instances. Items marked in red are deleted.

5.3.4 Modification Tracking for Templates

The modification tracker for a given data client should not just show changes to concrete instances which are owned by the data client. Rather, it should also include changes made to templates, of which the data client owns copies of. This provides accountability and transparency to data clients who choose to use templates. As changes are immediately applied to copies, this is an important feature for consistent user experience. To achieve this, every time a modification is made to a template, ownership of the resulting tracking record is not only given to the template owner, but to all owners of associated shadow instances.

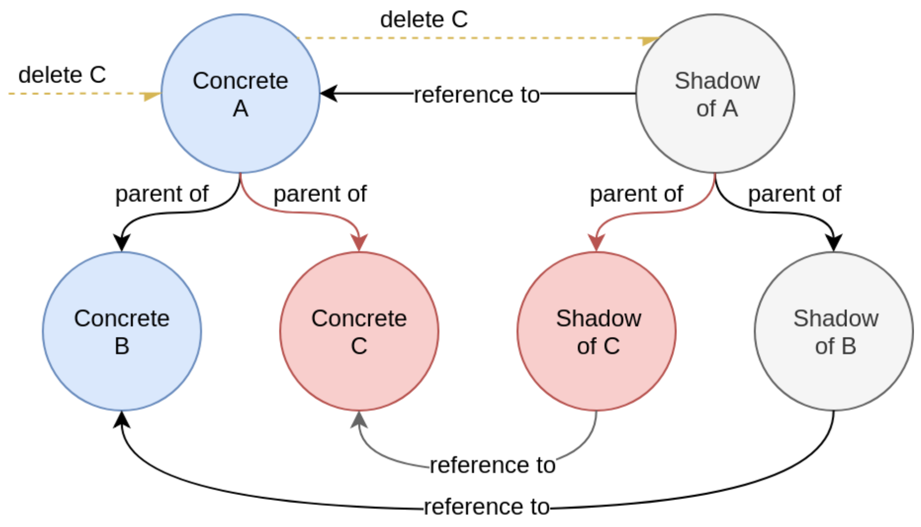


Figure 19: Modifications of templates are relayed to copies. Items marked in red are deleted.

5.4 API

The API is organized using the `Flask-RESTful` library (FLS). For each type of survey item, a canonical endpoint and a set of contextual endpoints exist, which are all mapped to the same handler. Canonical endpoints follow the `http://HOST:PORT/api/TYPE/ID` format, where `TYPE` is the type of survey item. Each survey item is identified by it's unique ID, which may be used for accessing the item via it's endpoints. If a survey item has child elements, the child elements may be accessed by appending `/CHILD_TYPE/` or `/CHILD_TYPE/CHILD_ID` to any of the parent's endpoints. The former returns a list of all children, while the latter accesses a single child. These endpoints are called contextual endpoints.

Each survey item is also associated with at least one JSON schema, which is used for serialisation and deserialisation of the classes instances. The schemas are implemented using the `marshmallow` library. Each survey item's schema also includes an URL pointing to the canonical endpoint of the item, which may be used for fetching the item again in the future.

Before any other request handler is invoked in the backend, information about the request language and user session are parsed. The request locale is determined by three mechanisms, which take precedence over each other in the order they're mentioned here (the latter overrides the former). First, the HTTP headers are inspected for the `Accept-Languages` field and the best match is chosen from the list of available languages. If cookies were sent with the request, they are searched for a `locale` cookie. Lastly, the request parameters are inspected for the `locale` parameter. Session information is communicated using the `Authentication` field of the request headers, follwing the bearer authentication scheme. If a session token is found in the HTTP headers, the token is validated. On success, the associated `Party` object is loaded and injected into the request context.

5.4.1 Access Control

Access to API endpoints and actions is handled by two separate mechanisms. The first mechanism uses Python's function decorators to block or grant access to an entire endpoint by checking the current user's role. The second method involves checking the ownerhip status of the accessed resource. For both mechanisms, different convenience methods exist. These methods are listed in table 4

5.5 Authentication

5.5.1 Data Client Authentication

Authentication for data clients can be performed by providing a valid combination of email and password. These parameters are sent to the backend in JSON format. The request can be encrypted by the client using TLS, when a valid certificate is provided for the load balancer. The load balancer will then decrypt the request and pass it to the backend unencrypted on the virtual network. When creating a new data client, a random salt is generated, using the operating system's random device. The provided password and salt are then hashed using the native Python implementation of the Argon2 password hashing function (ARG). The password hash and salt are stored in the database and can be used to validate login attempts, by re-applying the hashing algorithm to the salt and the provided password, and comparing the resulting hash with the stored hash.

5.5.2 Session Management

Once a `Party` has sucessfully authenticated, a *session record* is created and published to the `memcached` instance. This session record includes information about the `Party`, a timestamp of the last performed action by the `Party` and a randomly generated *session token*. The session token is handed out to the client via the API and may be used to by the client to identify themselves in subsequent requests. Every time the session token is used in a request, the timestamp stored in the session record is updated to

Method	Arguments	Description
@needs_role	[Either Role [Role]]	Grants access to the wrapped endpoint, if the user holds all roles in the given list. List items may be tuples of roles. In this case, only one of the roles contained in the tuple is sufficient. The expression [User, (Contributor, Admin)] would match any user, who has the User role as well as either the Contributor or the Admin role.
@needs_minimum_role	Role	Grants access to the wrapped endpoint, if the role's ID value is lesser or equal to the ID of the given role. Roles can be sorted by their ID, with Root having the smallest and Unprivileged the largest ID. This method is useful for restricting access to an endpoint to all users with a a certain level of privilege while also allowing all users with a higher level of privilege.
SurveyBase.accessible_by()	Party -> bool	Checks whether a SurveyBase may be read by the given party. This methods defaults to True for Admin and above. Read access is also granted, if the SurveyBase is a template. Otherwise, this method only returns True, if the Party owns the SurveyBase.
SurveyBase.modifiable_by()	Party -> bool	Checks whether a SurveyBase may be modified by the given party. This methods defaults to True for Admin and above. Otherwise, this method only returns True, if the Party owns the SurveyBase.

Table 4: Convenience methods for enforcing access control restrictions

the current time. If the difference between the timestamp and the current time exceeds a certain limit in any request, the token is rejected and the session record is removed from `memcached`. This ensures the eventual removal of unused session records from the cache and protects the user against re-use of their session token if they forgot to log out. Another protection against token stealing is IP pinning. When a user succesfully authenticates, their IP address is included in the session record. If any subsequent request with the associated token is using a different IP address than was used for authenticating, the token will be rejected and the session will be removed from the cache.

5.6 Support for xAPI

5.6.1 xAPI in the `backend` Service

xAPI statements are created in the backend service using an object-oriented API. Statements are queued locally using the `XApiPublisher` class, which acts as a transaction manager for xAPI statements. When a request context is created by the WSGI middleware, a new transcation is started. Queued statements are only sent to the `xapi-publisher` service, when the transaction is committed at any point during handling of the request. By default, when the request was handled without raising an error, and a rollback was not explicitly executed during the request, the transaction is committed automatically at the end of the request.

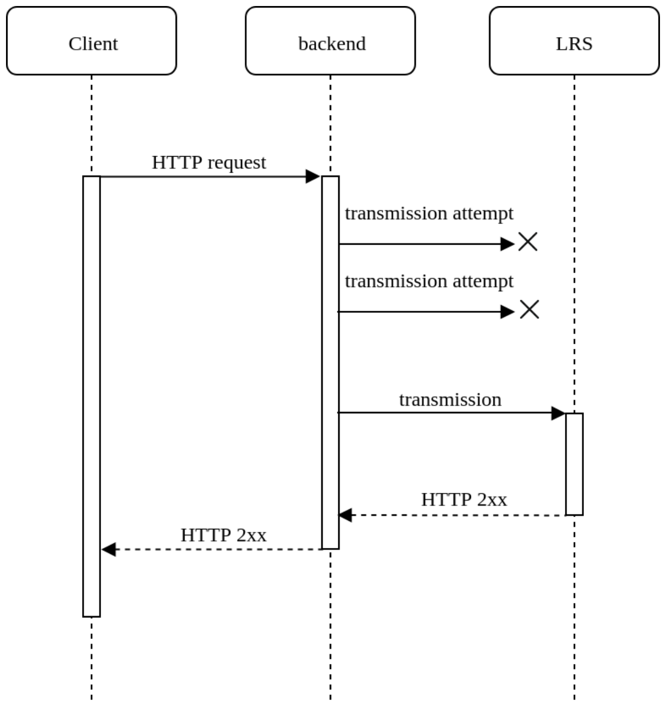


Figure 21: Synchronous (blocking) transmission of xAPI statements

5.6.2 The `xapi-publisher` Service

Sending of xAPI statements is separated from the API, as sending and possible re-transmissions should not block the API from responding to requests. This was discovered during testing, when a misconfigured xAPI recipient was used. HTTP timeouts are usually in order of seconds and unsuccessful connection attempts are retried. This resulted in the site becoming unresponsive when submitting the survey, as the submission of a survey would cause a large number of xAPI statements to be sent. Figures 21 and 22 illustrate this issue.

5.7 LTI Middleware

As mentioned before, recognition of data subjects uses the same token based mechanism that is also used for recognizing data client sessions. Since session management is performed by the backend service, but the frontend bundle is served by the frontend service, performing an LTI launch involves both services. The frontend provides an HTTP endpoint for requesting the LTI launch. The supplied information is then forwarded to the backend service, which validates the supplied combination of consumer key and requested questionnaire. If the LTI launch is valid, the backend service starts a new session for the data subject. A session token is then returned to the frontend service. The frontend service embeds the session token as a global variable into the JavaScript bundle which is the

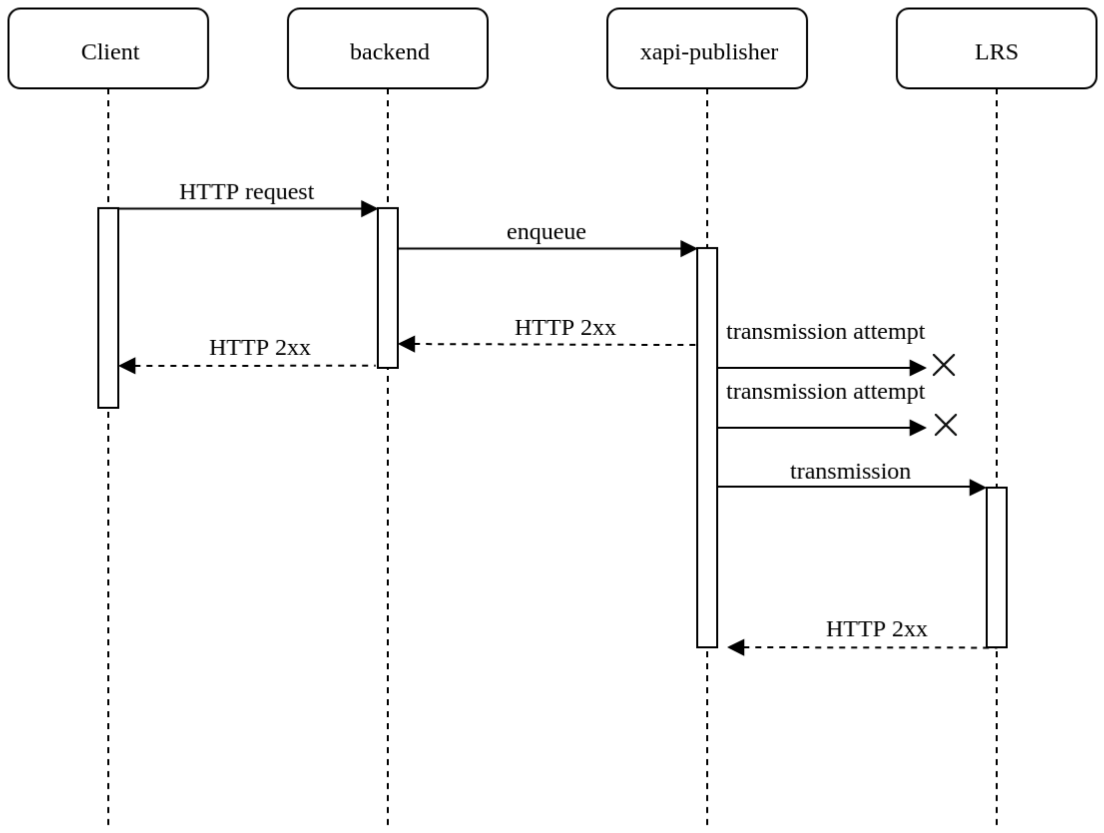


Figure 22: Asynchronous (non-blocking) transmission of xAPI statements

user interface and serves the parameterised bundle to the client. The user interface's bootstrapping process may then check for the presence of the session token and switch to it's embedded version. The embedded version of the user interface uses a different endpoint for submitting responses than the standalone version, as no information about the data subject has to be present in the respose. A sequence diagram for the LTI launch is provided in figure 23.

6 Analysis

6.1 Known Issues & Caveats

Joined table inheritance is inferior to single table inheritance for this application, even if it looks ugly. Why: performance, no foreign key constraints that mess with you

The template feature is too sophisticated, import and export would have sufficed this took time away

No tests exist for the backend, major burden on maintainability

the email adress of the data client who conducts the survey is added as a prefix to all xAPI activity IDs before sending. -> Drawback: not normalized, potential fix: use Instructor <-

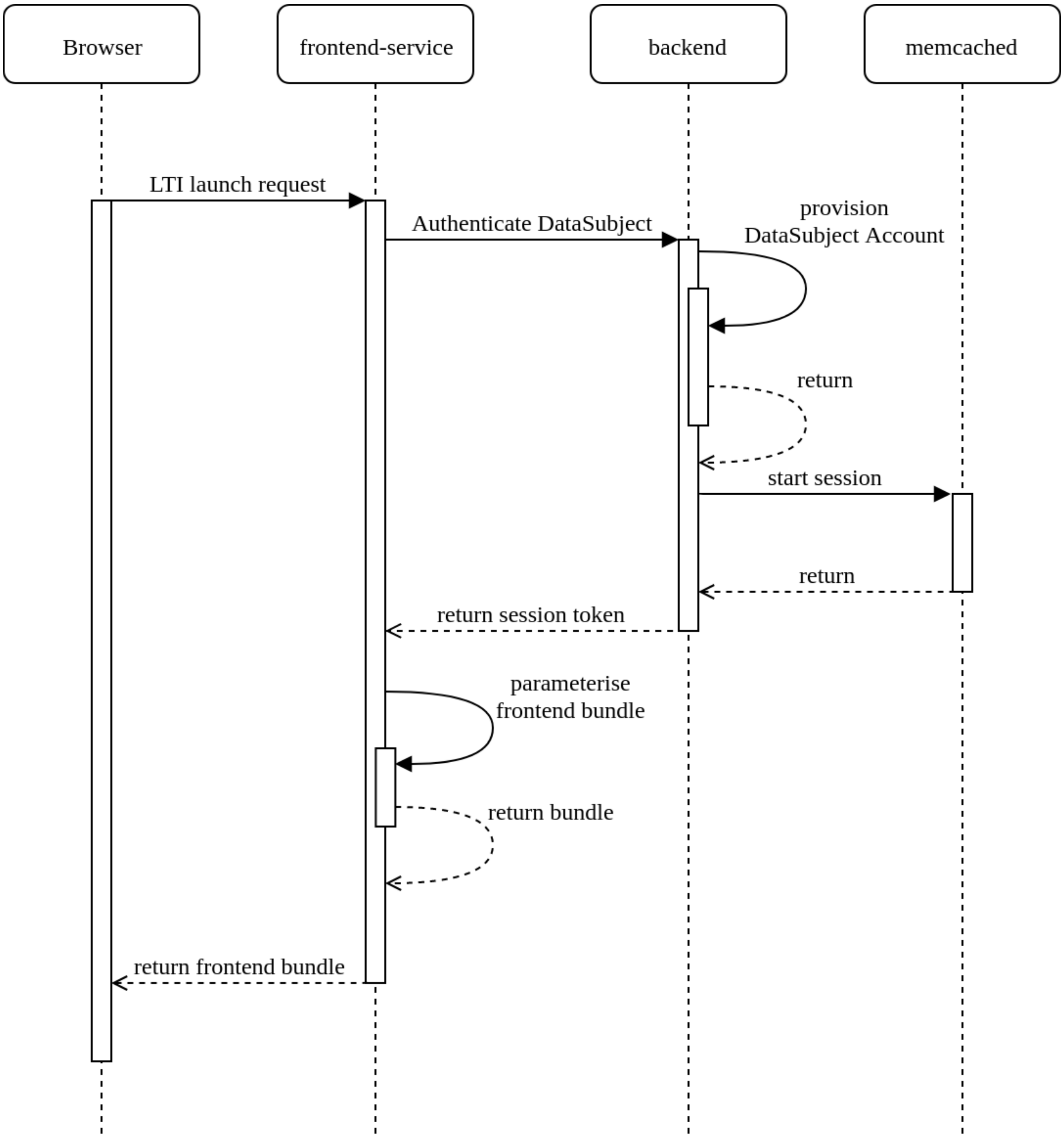


Figure 23: Sequence diagram depicting the LTI launch

7 Appendix

7.1 List of xAPI Statements

```
{
    "actor": {
        "account": {
            "homePage": "http://lmsng.school.edu",
            "name": "292832126.333"
        },
        "name": "292832126.333",
        "objectType": "Agent"
    },
    "context": {
        "contextActivities": {
            "grouping": [
                {
                    "definition": {
                        "description": {
                            "en-US": "This is a survey hosted at st3k101."
                        },
                        "name": {
                            "agq": "Questionnaire name"
                        },
                        "type": "http://id.tincanapi.com/activitytype/survey"
                    },
                    "id": "<bla@blubl.net>:Questionnaire-n-f7f705a0c1",
                    "objectType": "Activity"
                }
            ],
            "parent": [
                {
                    "definition": {
                        "description": {
                            "en-US": "This is a particular scale of a survey, it usually contains multiple questions."
                        },
                        "name": {
                            "de": "4.2 \u00dcberwachung"
                        },
                        "type": "http://fantasy.land/dimension"
                    },
                    "id": "<bla@blubl.net>:lernstrategien_wild_schiefele--4_2_&#252;berwachung",
                    "objectType": "Activity"
                }
            ]
        },
        "language": "agq",
        "platform": "st3k101 via lmsng.school.edu"
    },
    "id": "fcc317f7-d3ce-4597-ad88-245792ecb9c9",
    "object": {
        "definition": {
            },
            "name": {
                "de": "Ich bearbeite zus\u00e4tzliche Aufgaben, um festzustellen, ob ich den S
            },
            "type": "http://adlnet.gov/expapi/activities/question"
        },
        "id": "<bla@blubl.net>:lernstrategien_wild_schiefele—4_2_&#252;berwachung−−7",
        "objectType": "Activity"
    },
    "result": [
        {
            "score": {
                "max": 5,
                "min": 1,
                "raw": 2,
                "scaled": 1
            }
        }
    ],
    "timestamp": "2018−10−12T09:41:15.151759",
    "verb": {
        "display": {
            "en−US": "Indicates the DataSubject answered something."
        },
        "id": "http://adlnet.gov/expapi/verbs/answered"
    }
}
```

7.2 API Reference

7.2.1 backend

7.2.2 xapi-publisher

7.3 Abbreviations

API Application programming interface; a software interface for interoperability between applications.

CSV Comma separated values; a plain-text data exchange format for homogenous data.

ORM Object-relational-mapper; a software library for persisting objects in object-oriented languages in a relational database.

ODM Object-document-mapper; a software library for persisting objects in object-oriented languages in a document based database.

xAPI Experience API, formerly TinCan API; an API specification for exchanging data about learning activities.

REST Representational state transfer; a paradigm for API design.

JSON Javascript object notation; a plain-text data exchange format.

ACID Atomicity, consistency, isolation, durability; a set of properties for database transaction, aiming to guarantee data validity in the event of failure.

SQL Structured query language; a language for interfacing with database systems.

TLA Trusted learning analytics; a big data storage and analysis infrastructure developed at Prof. H. Drachsler's work group.

LRS Learning record store; a big data store for xAPI statements.

LTI Learning technologies interoperability; a protocol for integrating third-party services with an LMS.

LMS Learning management system; a content management system specifically designed for learning environments.

CAS Central authentication service; the user authentication service provided by Goethe University's HRZ.

HRZ Hochschulrechenzentrum; Goethe University's data center and IT service provider.

List of Figures

List of Tables

References

[Adva] Advanced Distributed Learning Initiative: *SCORM Overview*. `https://adlnet.gov/scorm`. – Accessed October 25, 2018

[Advb] Advanced Distributed Learning Initiative: *xAPI Background & History*. `https://adlnet.gov/research/performance-tracking-analysis/experience-api/xapi-background-history/`. – Accessed October 25, 2018

[Advc] Advanced Distributed Learning Initiative: *xAPI Statement Data Model*. `https://drive.google.com/file/d/0BxhK5TH2EsphZFBXeVNnSGozWEE/view`. – Accessed October 25, 2018

[Adv16] Advanced Distributed Learning Initiative: *xAPI specification*. `https://github.com/adlnet/xAPI-Spec/releases/tag/xAPI-1.0.3`. Version: 2016. – Accessed: October 25, 2018

[ARG] *Argon2*. https://github.com/P-H-C/phc-winner-argon2

[BDD$^+$16] Butler, H. ; Daly, M. ; Doyle, A. ; Gillies, S. ; Hagen, S. ; Schaub, T.: *RFC-7946: The GeoJSON Format*. https://tools.ietf.org/html/rfc7946. Version: 2016. – Accessed October 25, 2018

[Eby03] Eby, Phillip J.: *PEP 333 Python Web Server Gateway Interface v1.0*. https://www.python.org/dev/peps/pep-0333/. Version: 2003. – Accessed October 25, 2018

[FLS] *Flask-RESTful*. https://pypi.org/project/Flask-RESTful/

[PD09] Philips, Addison ; Davis, Mark: *RFC-5646: Tags for Identifying Languages*. https://tools.ietf.org/html/rfc5646. Version: 2009. – Accessed October 25, 2018

[Sch17a] Scheffel, Maren: *The Evaluation Framework for Learning Analytics*. http://hdl.handle.net/1820/8259. Version: 2017. – Accessed: October 25, 2018

[Sch17b] Scheffel, Maren: *The Evaluation Framework for Learning Analytics, greyscale template*. http://www.laceproject.eu/evaluation-framework-for-la/. Version: 2017. – Accessed: October 25, 2018

[SQL] *SQLAlchemy 1.2 Documentation: Mapping Class Inheritance Hierarchies*. https://docs.sqlalchemy.org/en/latest/orm/inheritance.html

[SW94] Schiefele, Ulrich ; Wild, Klaus P.: *Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilitat eines neuen Fragebogens*. https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/deliver/index/docId/3182/file/schiefele1994_15.pdf. Version: 1994. – Accessed October 25, 2018

[The16] The European Commission: *Data protection in the EU*. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en. Version: 2016. – Accessed: October 25, 2018

[Wor17] World Wide Web Consortium: *Acitivty Streams*. https://www.w3.org/TR/activitystreams-core/. Version: 2017. – Accessed October 25, 2018