Globalization meets GDPR: implementing privacy safe transactions in modern APIs using personal data

Markus Gylling and Gill Ferrell
[1] IMS Global Learning Consortium, Sweden
[2] IMS Global Learning Consortium, France
mgylling@imsglobal.org, gferrell@imsglobal.org

Abstract

Digital transformation of higher education is bringing enormous benefit. Advances in technology and the development of open standards mean that a seamless 'plug and play' learning ecosystem is now a reality.

Seamless interoperability, for all its advantages, has so far come with certain risks. How can we be absolutely sure that all of our transactions are fully compliant with strict privacy laws? To date, the only way to achieve this has been to make those transactions less efficient than they could otherwise be.

In this paper we show how the work of the European 1EdTech community is ensuring that European public values influence global innovation. We discuss a set of data privacy principles that the community has developed and see how these are being implemented in practice in the Edu-API standard.

By June 2022 we will be able to update this paper with further information on the outcomes of the review of the Edu-API 'privacy ready' approach by the higher education sector.

1. Introduction

Digital transformation of higher education is bringing enormous benefit allowing us to enhance the learning experience and make it available to many more learners by doing new things, doing things better and managing our organizations more efficiently and effectively.

Underpinning this transformation is the digital infrastructure that allows us to support anytime, anywhere learning. Learners have the tools and resources they need at their fingertips and teachers have the information they need to monitor student progress and support each learner in an increasingly personalized way even as class sizes increase.

A seamless, ‘plug and play’ learning ecosystem is no longer a distant aspiration. It is now a reality made possible by advances in technology and the development of open standards to allow platforms and tools to work together (interoperability).

For over 20 years, the standards that provide the life force of this ecosystem have been developed by the community led by IMS Global. From early 2022 the community will begin to operate under the new name 1EdTech to better reflect its nature as a community and a trusted partner in education innovation.

Our European community is very focused on ensuring that European public values influence global innovation and this paper illustrates one of the ways in which we are achieving this.

2. Globalization and GDPR

Whilst the benefits of digital transformation are very evident, the risks are equally clear. The world’s major technology companies have grown in countries where the law and/or culture relating to data privacy is very different to our own. On the face of it, PII (personal identifiable information) in the US and PD (personal data) in Europe are similar concepts but the definitions differ considerably.

Europe has some of the world’s strictest privacy laws and universities, which in Europe are mainly public bodies, are very conscious of their obligations to staff and students in this regard.

The legal instruments governing the European approach are beyond the scope of this paper but at the heart of it lies the General Data Protection Regulation or GDPR (EU2016). In July 2020 the Court of Justice of the European Union (CJEU) issued a ruling, known as the Schrems II Judgement, invalidating the EU/US Privacy Shield arrangement on account of invasive surveillance practices by the US government (European Parliament 2020). Data controllers or processes transferring personal data outside the EU, on the basis of standard contractual clauses (SCCs), now have to ensure a level of protection essentially equivalent to GDPR.

The Schrems II Judgement has caused much consternation across many European public bodies and at a 2021 GÉANT workshop on the topic, there were reports from some NRENs of procurement processes for cloud services being halted so that the feasibility of creating on premise solutions could be considered. EUNIS members have reported some public bodies banning the use of all APIs (application programming interfaces) in commercial products. Creating ‘Fortress Europe ‘in this manner would come at considerable cost and loss of opportunity.

The 1EdTech community is viewing this as a positive stimulus to do things better and ensure that we can support trustworthy data exchange. The challenge of a global standard is that it has to be able to cross the boundaries of policy and legislation on a worldwide scale. In short, it has to be able to meet the most exacting standard.

3. A principled approach

We have come up with a way to approach data privacy which is summed up in these five principles:

1. **Data Minimization**:
n*ever ask for more than you need, never offer more than you were asked for.*

2. **Anonymization and/or Pseudonymization**:
any application that needs personal information is constructed such that it can operate only with the anonymized or pseudonymized form of that information.

3. **Behavioral Documentation and Interrogation**:
verifiable documentation that describes the application's use and handling of personal data, including which personal data it needs to operate correctly, and which personal data it exposes to users and/or other applications.

4**. Supportive data interchange protocols**:
specifications that define the protocols used for data interchange between systems are required and enable practical adherence to the Data Minimization and Anonymization/Pseudonymization principles.

5. **Support for privacy-related administrative workflows**:
enables the institution to act quickly and efficiently on user requests, and also allows the institution to monitor that e.g. data retention policies are followed

4. Applying the principles

The principles are intended to guide the design of specifications such that IMS 1EdTech specifications should support privacy-enabled data exchange.

What that means in practice varies according to context. It could mean complying with national law or it could be down to the detail of what is included in the DPA (data privacy agreement) for a particular peer-to-peer exchange.

Any solution used in IMS 1EdTech specifications must therefore be applicable internationally, regardless of variations pertaining to:

- National and/or state-level legislation
- Institutional-level requirements and/or best practices, if existing
- Local peer-to-peer (data suppliers and consumers) agreements/contracts, if existing (e.g.DPA)

Because of the range of contexts that need to be supported, it is not possible to mandate a single method to ensure privacy, so IMS 1EdTech specifications come with a range of acceptable approaches to choose from:

1. Ensure that fields containing personal data are optional in the specification so that you can miss them out and still have a valid payload.
2. Replace the personal data with something else that is either anonymized (non-reversible) or pseudonymised (reversible).
3. Use OAuth and put the personal data in a special object to which you need an extra key.

Each of these approaches can have advantages and disadvantages depending on how it is applied. For example, making all fields optional can ensure privacy down to a very granular level but sending empty fields could be misused by a lazy implementer and cause problems with data quality for other purposes.

Option 3, on the other hand, can make the API itself less efficient because you need to pass additional data and filter it and because it means the consuming system has to understand a lot about the data model.

5. What solution works best for Europe?

Again, there may be no single answer or, indeed, the answer may be a combination of the existing three options. This is something that the European 1EdTech group is actively working on.

We need to be able to deal with a wide variety of situations i.e. to transmit valid data files that are able to traverse multiple levels of agreement by adapting the payload to what has been agreed. Universities will have very different agreements with different types of supplier. With their core student record system there will be a comprehensive agreement covering a lot of data transfer. At the other end of the spectrum, small scale 'fun 'apps for learning will need to be used in a much more restrictive way.

The concept of direct and indirect personal data fields comes into play here and is very important in relation to European legislation and requirements. For example, neither gender nor ethnicity is personal information specific to an individual but, in a small cohort, they could make an individual identifiable.

New requirements on IMS specifications take this into account:

1. All PII/PD encumbered fields must be flagged:
    1. We have an established format/style convention for these flags to use consistently in all specs
    2. Flags are set for both "direct" and "indirect" PII/PD fields

2. The *default state*[1] of each payload body is one where all PII/PD flagged fields are withheld[2]
    a. Either by way of maxed-out data minimization (with no PII fields being required in spec)
    b. or by way of anonymization/pseudonymization
    c. or by way of OAuth scopes (PII/PD fields in a scope with a dedicated off-by-default grant)
    d. or a, b and c in some combination

3. Mechanisms for anonymization/pseudonymization are clearly defined, if existing:
    a. The privacy framework spec would encourage the use of sourcedId as universal solution for persistent key identifiers; specs would be expected to adopt that.
    b. The privacy framework spec would define a mechanism for spec payloads to *signal that a value has been anonymized/pseudonymized*; specs would be expected to adopt that.

6. Designing the right API endpoints

APIs have to be designed not only to meet legal requirements as regards data privacy, but also to enact efficient transactions.

A weakness of APIs that were established as standards some time ago, is that they can be too primitive i.e. they don't have endpoints to cover all of the common transactions. The outcome of this situation is that the consumer has to draw down more data than is strictly necessary and then filter it for the information that is actually required. Undertaking a complicated query with a REST API can thus result in a lot of calls and processing.

---

[1] The default state applies if no other explicit agreement (e.g. DPA) has been made.

[2] This applies to REST APIs only: a different approach may be needed for asynchronous (pub/sub) versions.

Let's say you regularly need to retrieve a list of the students enrolled on a particular course. An inefficient design may mean that you need to draw down all enrolments at the institution every time and filter the data. If this was already set up as an API endpoint you could get the filtered data back in a single transaction. The goal is to maximize the efficiency of REST APIs whilst retaining privacy.

Transaction oriented endpoints are needed to avoid a consuming site needing a lot of contextual information to make intelligent use of the API. The consumer should be able to receive only the snapshot it needs with nothing further required for it to be able to understand the snapshot.

7. Edu-API as a model for the approach

Edu-API is an open specification currently being developed by the IMS 1EdTech community. Edu-API (IMS 2020) lies at the heart of making core enterprise data from student information systems available to the rest of the learning ecosystem. Data in student record systems is frequently 'locked in ' in such a way that migration to new systems, or cloud services, is time consuming and costly and data exchange with other systems involves complex point-to-point integrations.

Edu-API addresses the problem of getting systems to work together so that student data can be input once and used for different administrative purposes and in different parts of the learning ecosystem. Having a single, coherent industry standard avoids the need for costly custom integrations that need to be maintained and updated over time.

The context of this development is that core administrative data is being put to an increasing range of uses (learning analytics being a good example) whilst at the same time, increased use of cloud services means developers have less direct access to data. In this scenario the existence of robust, flexible APIs takes on increasing importance if we are to develop robust, secure and scalable services.

EUNIS EA SIG (along with a number of EUNIS member organizations) is one of the groups that has contributed to this development by reviewing the approaches and trying to answer the question 'what is the minimum set of endpoints we need to define? 'Complexities arose around the different roles a single person could have within the University and the impact that different academic sessions may have. For example, on a three-year degree course, students in year one may be undertaking a different diet of study units and assessment to those already in the final year. An original set of 33 endpoints was increased to 46 as a result of review by the higher education sector.

The Edu-API MVP (minimum viable product) is due for release during 2022 (this is the REST version of the API. Piloting of an asynchronous 'pub/sub 'version is already underway across Sweden). By June 2022, we will be able to report on the full outcomes of the consultation exercise and the final shape of the MVP.

8. Summary

The data privacy principles have been scrutinized by privacy specialists representing European universities, NRENs and suppliers and received their endorsement.

Edu-API will be the first of the IMS 1EdTech specifications to offer a privacy-enabled API 'out of the box'. The experiences of early adopters will be carefully reviewed as this will provide a model for all future developments by the 1EdTech community.

9. References

European Parliament (2016) General Data Protection Regulation. EU2016/679. Retrieved February 09, 2022, from: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

European Parliament (2020) The CJEU judgement in the Schrems II case. Retrieved February 09, 2022, from: https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf

IMS Global Learning Consortium (2020) Edu-API: a global community initiative driving the next generation of higher education student system interoperability. Retrieved February 09, 2022, from: https://www.imsglobal.org/edu-api

IMS Global Learning Consortium (2021) IMS security framework: trusted exchange of student data. Retrieved January 27, 2022, from: https://www.imsglobal.org/ims-security-framework

IMS Global Learning Consortium (forthcoming) IMS data privacy interoperability primer.

10. Author biographies

Markus Gylling is Solutions architect, IMS Europe at IMS Global Learning Consortium. With 20 years of experience in standards and software development, Markus' primary engagement areas in IMS are accessibility, digital content and data model development. Prior to his work with IMS Global, Markus worked as CTO of the International Digital Publishing Forum (IDPF) and as CTO of the DAISY Consortium, as well as head of the unit for process and architecture at the Swedish National Agency for Education.

https://www.linkedin.com/in/mgylling/

Gill Ferrell has led the EUNIS Learning and Teaching SIG since 2009. She also has an interest in data and information management and led Jisc support for UK HEIs changing practice when GDPR first came into force. In 2021 she joined IMS Global Learning Consortium as IMS Europe Program Director.

https://www.linkedin.com/in/gillferrell/