



DataExchange: Privacy by Design for Data Sharing in Education

1st International Conference on the
Frontiers and Advances in Data Science

fads.org.uk

23–25 October 2017

Xian, China

Special Session in Data Science for Politics, Policy, and International Development

DOI: 10.1109/FADS.2017.8253202



ZiNET Data Solutions Limited. 25 Russell Street, Hastings, East Sussex, TN34 1QU, UK
Registered in England and Wales, UK (08765226)

http://zinethq.com	@ZiNETHQ
http://dataexchange.education	@DataExchangeEdu
http://opendata.education	@OpenDataEdu

Copyright

Paper is copyright © 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Presentation slides are copyright © 2017 ZiNET Data Solutions Limited. Personal use of this material is permitted. Permission from ZiNET Data Solutions Limited must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Disclaimer

This document and the information contained herein is provided on an 'as is' basis and ZiNET Data Solutions Limited disclaims all warranties, express or implied, including but not limited to any warranty that the use of the information herein will not infringe any rights or any implied warranties of merchantability or fitness for a purpose.

Dr Jon Nicholson, PhD FHEA mIEEE

Chief Technology Officer, ZiNET Data Solutions Limited

Jon has over 20 years experience in software development. He obtained a distinction MSc in Distributed Information Management Systems in 2006, and a Doctorate in Computer Science in 2011, from the University of Essex. Jon's academic interests are in software engineering, visualization and data analysis, with a particular focus to application. Jon has over 10 publications in internationally recognised conferences and journals. Previous to joining ZiNET Jon was a senior lecturer in Computer Science within the School of Computing Engineering and Mathematics at the University of Brighton, including course leadership. Jon also holds a postgraduate certificate in Teaching and Learning in Higher Education, is a Fellow of the Higher Education Academy, and is a member of the Institute of Electrical and Electronics Engineers.



Ian Tasker

Chief Executive Officer and Founder, ZiNET Data Solutions Limited

Ian is one of the UK's leading thinkers on data integration in the education sector. He is the former Co-Lead of the UK Technical Board for A4L Community (Assessment for Learning) and UK Representative to the International Technical Board of the SIF Association (School Interoperability Framework). Ian is perhaps the UK's most knowledgeable expert in the school data arena. Ian has been instrumental in the development of the ground breaking SIF 3.0 Specification and now is at the cutting edge of its successful implementation in the marketplace. His leadership in the UK and Global SIF Community's technical work has put him as a "go to" resource in solutions development!



DataExchange: Privacy by Design for Data Sharing in Education

Jon Nicholson

CTO, ZiNET Data Solutions, UK
jon.nicholson@zinethq.com

Ian Tasker

CEO and Founder, ZiNET Data Solutions, UK
ian.tasker@zinethq.com

Abstract—The UK education data integration and sharing market has long been based on scope or object-level data sharing. However, this approach leaves openings for data leaks and may not be compatible with the forthcoming General Data Protection Regulation. We present DataExchange, a data integration and sharing platform designed around the concept of privacy by design. DataExchange makes use of internationally reviewed education data and communication open standards. DataExchange is based on attribute-level privacy controls which improves visibility of third party data requirements, ensures that third parties can only access the data they have explicit authorization for, and provides transparency as to what data is shared.

I. INTRODUCTION

The United Kingdom has strict data protection laws via the Data Protection Act (DPA) 1998 [1]. The DPA defines a framework for collecting, processing and storing personal information about living individuals (data subjects). It defines the legal obligations of those who manage personal data (data controllers) as well as those who process that data on their behalf (data processors). While the DPA has provided a well understood grounds for sharing personal data, its focus is on data security as opposed to data privacy.

The General Data Protection Regulation (GDPR) [2], which will apply in the UK from 25 May 2018, builds upon the existing measures with a stronger focus on privacy. The GDPR has already had wide ranging impact across all sectors with numerous guides and services produced to support institutions at all levels in being ‘GDPR-ready’ [3], [4]. The GDPR includes new rights for data subjects such as allowing them to request restricted data processing or to data erasure. It also imposes rules over the chain of custody between the data controller and each consequent data processor. The GDPR, therefore, requires far more transparent and privacy focussed data processing than would be acceptable under the DPA.

46% of UK businesses have experienced a security breach in the last 12 months [5]. On average a data breach will cost a UK business up to £138,700 and require over 9 months to recover from the incident [6]. The current maximum fine the Information Commissioner’s Office (ICO), who enforces the DPA, can levy is £500,000 for a serious data breach. Once the GDPR is enforced this maximum will increase significantly to €20,000,000 (£18,053,700¹), or 4% of the company’s annual global turnover, whichever is the greater. Given that only 22%

of education, health or social care firms require suppliers to adhere to any cyber-security standards [5] it is not surprising that compliance violations/legal infractions and insecure third parties are among the top causes of data breaches [6].

In the past year we have observed an increasing level of discussion and queries regarding data protection [7]. Institutions are beginning to (rightly) demand clear and transparent written transcripts detailing what data is shared and who it is shared with. Some existing integration solutions in the UK (see section II) provide limited data filtering, primarily controlled by what data the integrator could obtain from the institution. This is an approach which, by its very nature, is prone to data leaks. A room booking service, for example, may reasonably require a student’s name, registration group, year group, timetable, and so on to function. It is also conceivable such a system would also require data on each student’s special needs in order to help staff select appropriate spaces for their students, such as wheelchair accessibility. It would not generally be appropriate for such a system to have access to much of the student’s sensitive data, such as their date of birth, gender, pregnancy status or if they are in care. Yet if the integrator has obtained this information then the data may be passed on or made accessible to the third party service as part of that student’s record. This could be considered a data leak, or unintentional data disclosure, as the third party vendor has access to more personal data than what is reasonably required to perform their function (a violation of the third principle of the DPA). This is not an uncommon scenario in the current UK education sector: institutions are forced to accept a lack of transparency in what a third party requires and a lack of visibility in what data is collected and shared. To ensure strict DPA and GDPR compliance, data integration in UK education must experience a paradigm shift from facilitating full/partial record exchange to record exchange based on fine grained, attribute-level controls.

Moving toward attribute-level and privacy focussed approaches to data integration presents a significant problem in managing each application’s data requirements as well as each institution’s data permissions. As a data integration provider we, ZiNET Data Solutions Limited, have been working over the last year to develop a solution for this. Our efforts produced DataExchange (<https://dataexchange.education>), replacing our legacy ZiNET Connect product and launched August 2017, a new GDPR-ready data integration and data sharing platform

¹Based on an exchange rate of £1 to €0.9 as of 10am 13th Sept 2017.

that is built using the latest open standards such as OpenAPI [8] and SIF [9], [10]. In section II we discuss existing data integration platforms for education data available in the UK, and then show how DataExchange presents a unique offering (section III). In sections IV and V we discuss DataExchange with respects to securing data integration and data sharing respectively, concluding in section VI by evaluating DataExchange against the seven principles of privacy by design [11].

While DataExchange currently targets the education sector, with appropriate data model standards the underlying technologies and processes can be applied to improve transparency and privacy in other sectors, such as health and social care.

II. EXISTING SOLUTIONS

There are three main data integration products/services in the UK education technology sector²: Groupcall Xporter, Assembly, and Wonde. Table I summarises these products.

Groupcall Xporter [12], and the related family of products, has the biggest market share in UK education data integration. Xporter works by extracting data into files which are then pushed to a file system or out to a vendor over a secure HTTPS connection. Groupcall also offer Xporter on Demand (XoD) which uses Xporter to send the data to their XVault (data warehouse) service on top of which they provide vendors with a web-based API. Groupcall also provide a single sign-on product, IDaaS (identity as-a-service), which is required if an institution is to use XoD. Groupcall Xporter supports both read and write-back requests.

Assembly, part of Ark UK Programmes, is a non-profit joint-venture between Ark and the NEON Foundation. They “exist to help schools use data to improve outcomes for students and to help them succeed” [13]. Assembly primarily provides data analytics and benchmarking tools, but also provides data integration for vendors of school improvement applications. Data is collected into a secure data warehouse and made available in JSON format through an OAuth secured web-based API over HTTPS. Scope-based filtering is applied to limit a third party’s access, dependant on data sharing agreements. Assembly predominantly supports read, as opposed to write-back, requests.

Wonde [14] is a more recent product on the market, which has quickly been gaining attention. Wonde appears to operate in much the same way as Groupcall XoD and Assembly, but little about their processes and controls is currently available. According to institutions using (or looking to use) Wonde [15], their “model is to pull out everything and then as a school, you decide which third party has access”. The lack of accurate published information about the product highlights the need for transparency and visibility for institutions.

All three products work on a very similar scope-based data collection mechanism. Scopes are either derived from system access permissions (what access is granted) or defined by the data integrator (providing some data restrictions). Each scope specifies access permissions for an object or set of

Table I
COMPARISON OF UK EDUCATION DATA INTEGRATION PRODUCTS

Product	Groupcall Xporter	Assembly	Wonde
Data Model	Proprietary ³	Proprietary	Proprietary
Format	XML	JSON	JSON
Push/API	Push ⁴	API	API
System support	14	10	5
Locale support	UK	UK	UK
SDKs	2	2	3

attributes within an object. The problem with scope based data integration and sharing is that scopes typically contain more data than a vendor requires. For example, a demographics scope may include gender, date of birth and ethnicity, so if only gender is required then there is unintentional data disclosure.

III. OVERVIEW

DataExchange is a leading data integration and sharing platform for education. Data is secured in transit and at rest, and only collected, processed and shared when there is declared requirement *and authorization* to do so (Fig. 1). All data requirements and permissions are described at the attribute-level, affording fine grained access controls and explicit data sharing agreements. The unique selling points of DataExchange are:

- **Minimal data collection.** A current 2-party data sharing agreement must exist between the institution (as the data provider) and DataExchange (as the data integrator). This agreement permits DataExchange to extract and store the data as required through the union of the data permissions defined in the institution’s 3-party data sharing agreements. That is, DataExchange will only take what data it needs to provide the services asked for.
- **Minimal data sharing.** A current 3-party data sharing agreement must exist between the institution (as the data provider), DataExchange (as the data integrator) and the vendor (as the data consumer) before any data is shared. The vendor is given access to no more data than is defined in the data sharing agreement.
- **Visibility.** All data requirements of an application are explicitly detailed by the vendor, allowing institutions (data controllers) to make informed choices about what data they share, and with which applications. It also makes visible what data is supported or can be extracted from each data source.
- **Transparency.** All data access permissions are detailed within data sharing agreements and viewable online. The digital footprint of data shared helps both institutions and vendors to perform data audits, produce privacy impact assessments, and inform privacy notices.

³Xporter can support the SIF UK 2.0 data model [10] with some additional tools, but with reduced system support.

⁴Provided through the Xporter on Demand, XVault and IDaaS services.

²Excluding ZiNET Connect, which is being replaced by DataExchange.

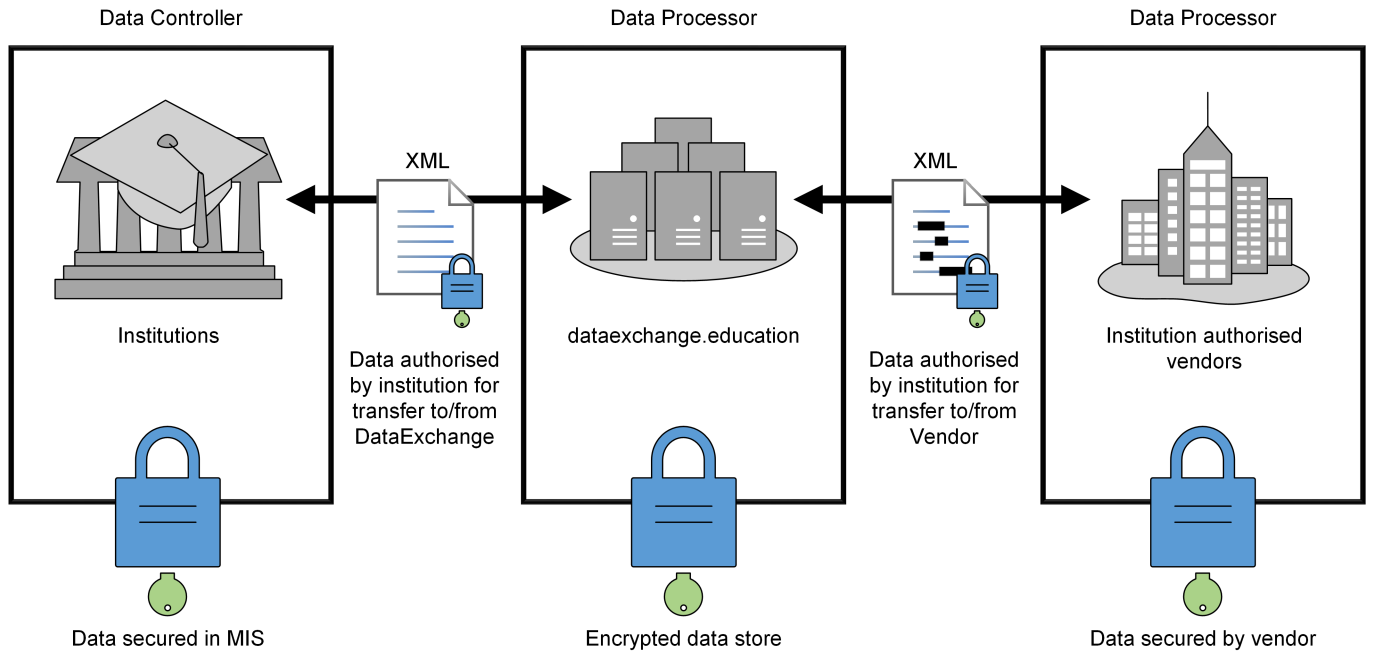


Figure 1. Overview of DataExchange.

IV. DATA INTEGRATION

Educational institutions store their data in one or more database systems called Management Information System (MIS), or Student Information System (SIS) in some countries. The MIS/SIS market is fragmented with large variation in data schema and method of access⁵. This fragmentation is compounded where MIS systems are specialised within a certain range of educational levels. For example, an MIS in the primary education (5–11 years old) sector would not be used in the higher education sector due to the differing learning outcomes and statutory reporting requirements [16, Section 537][17, Section 64/65].

Managing interoperability between heterogeneous systems has received much academic attention (examples [18], [19], [20], [21]). In the domain of data integration in education the typical methodology is “Extract, Transform, Load” (ETL), as used by Groupcall Xporter/XVault and Wonde (section II) which transform into proprietary formats. In DataExchange we use the open Systems Interoperability Framework (SIF) standards published by the Access 4 Learning (A4L) Community.

The A4L Community⁶ is a “non-profit collaboration composed of schools, districts, local authorities, states, US and International Ministries of Education, software vendors and consultants who collectively address all aspects of learning information management and access to support learning” [22] and have locale specific communities in North America [23], Australia [24] and the UK [10]. Each locale community

⁵Typically direct SQL queries (e.g Facility, SchoolBase), local APIs (e.g. Capita SIMS) or web based APIs (e.g. Bromcom and Arbor).

⁶ZiNET Data Solutions is an active member of the A4L Community, holding elected positions at the international level.

maintains its own data model/ontology tailored to their local systems and methods of education. Using standards maintained by an international community ensures their robustness, integrity and consistency. Importantly, it also ensures that no single entity has complete control, specifications are developed collaboratively with input from multiple stakeholders who all share the same goal of improving education.

Alongside developing educational data models the A4L Community defines an infrastructure specification for secure communication between parties. The SIF Infrastructure [9] defines a RESTful interface for either light-weight point-to-point (direct, Fig. 2a) or enterprise (brokered, Fig. 2b) usage. In the case of the direct architecture, the application that provides data (provider) manages what applications (consumers) can access data. In the brokered architecture the task of managing client access rights is moved to an enterprise service bus. This standardises how consumers interact since architectural detail is hidden; a consumer’s view of the world is limited to a set of services and access rights defined in their *environment* presented in XML (Fig. 3).

SIF defines four types of services: Infrastructure, Utility, Object and Functional [9]. Infrastructure services provide consumers with necessary gateways to interact with other services (requestsConnector, which acts like a secure proxy), events, queues and subscriptions. Utility services provide a secondary level of infrastructure functionality, such as exception reporting (Alerts). Object services are authoritative sources of data of a specific type, such as LearnerPersonal (UK) or Student (North America/AU) objects. They can be accessed through standard CRUD (create, read, update and delete) operations, but may also publish events on data changes. Note that in Fig.

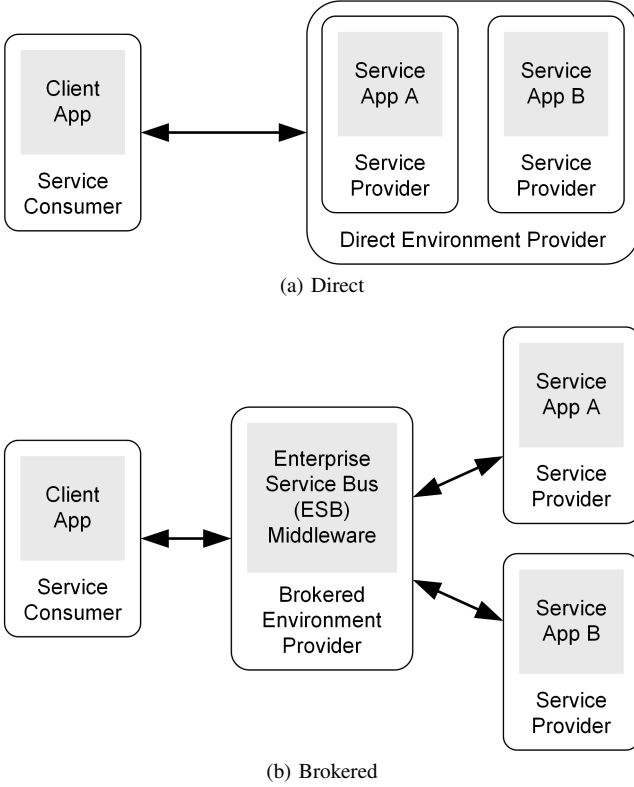


Figure 2. Direct (a) and Brokered (b) SIF architectures.

3 both the LearnerPersonals object services are limited to read only access. Finally, functional services provide a mechanism for encapsulating asynchronous behaviour, such as multi step processes or processes that involve manual intervention.

We will focus our discussion on object services. In SIF object services must be authoritative, they provide a single source of truth for the type of data they provide, where services can be aggregated into zones. Zones are a pre-organised set of services that provide a scope for a set of services. For example, a consuming application for a Multi Academy Trust (MAT) or Local Authority (LA) may have access to many zones — one for each institution it has an interest in. In this way each service provider is authoritative within its scope (zone). Additionally scoping can be achieved with the use of contexts. Contexts provide a local scope to a specific service within a zone so can be thought of as a singleton set containing a single object service. An example use of a context might be used to scope LearnerPersonal object services to current and pre-enrolment students. Given this we can reformulate Fig. 3 as a Concept Diagram in Fig. 4.

V. DATA SHARING

DataExchange works using attribute-level privacy controls obtained by processing the SIF locale data models (XML Schema) on which we provide some additional metadata, for example to identify which attributes should be considered sensitive. This facilitates production of data model oriented client software development kits (SDKs) through the produc-

```

1 <environment
2   xmlns="http://sifassociation.org/infrastructure/3.2.1"
3   type="BROKERED"
4   id="2b1413b0-e898-4bb0-982f-f723d09d8349">
5   <sessionToken>...</sessionToken>
6   <solutionId>United Kingdom</solutionId>
7   <authenticationMethod>Bearer</authenticationMethod>
8   <userToken>...</userToken>
9   <consumerName>Example Consumer</consumerName>
10  <infrastructureServices>
11    <infrastructureService name="environment">
12      https://.../environments/current
13    </infrastructureService>
14    <infrastructureService name="requestsConnector">
15      https://.../requests
16    </infrastructureService>
17  </infrastructureServices>
18  <provisionedZones>
19    <provisionedZone id="School_A">
20      <services>
21        <service name="LearnerPersonals"
22          contextId="DEFAULT" type="OBJECT">
23          <rights>
24            <right type="QUERY">APPROVED</right>
25          </rights>
26        </service>
27      </services>
28    </provisionedZone>
29    <provisionedZone id="School_B">
30      <services>
31        <service name="LearnerPersonals"
32          contextId="DEFAULT" type="OBJECT">
33          <rights>
34            <right type="QUERY">APPROVED</right>
35          </rights>
36        </service>
37      </services>
38    </provisionedZone>
39  </provisionedZones>
40 </environment>

```

Figure 3. An example environment for a DataExchange vendor app in XML.

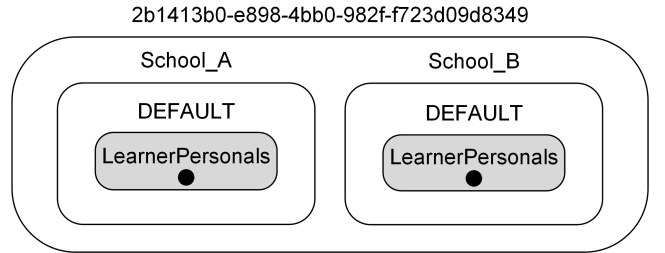


Figure 4. An example environment for a DataExchange vendor app as presented in Fig. 3 reformulated as a Concept Diagram [25].

tion of OpenAPI [8] specifications. Vendors are then required to provide the details of their applications, presented as a series of toggles. Vendors mark data as:

- **Required.** Data necessary for minimum functionality. For example, an individual's name in a lesson planner.
- **Optional.** Data that the institution may opt to deny access to. If present there may be additional functionality. For example, an individual's picture in a lesson planner.
- **None.** Data that is not required and will not be given access to. This is the default.

No other data integrator in the UK education sector requires this level of detail from application vendors, but in collecting it we can generate compatibility tables highlighting an

```

2 <xsl:stylesheet
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns="http://www.sifassociation.org/datamodel/uk/2.0">
4 <xsl:output method="xml" omit-xml-declaration="yes"/>
  <xsl:template match="/LearnerPersonal">
6    <xsl:copy>
      <xsl:apply-templates select="node()|@*" />
    </xsl:copy>
  </xsl:template>
10 <xsl:template match="/LearnerPersonal/@RefId">
    <xsl:copy>
      <xsl:apply-templates select="node()" />
    </xsl:copy>
  </xsl:template>
14 <xsl:template match="/LearnerPersonal/.../FamilyName">
    <xsl:copy>
      <xsl:apply-templates select="node()|@*" />
    </xsl:copy>
  </xsl:template>
20 <xsl:template match="/LearnerPersonal/.../GivenName">
    <xsl:copy>
      <xsl:apply-templates select="node()|@*" />
    </xsl:copy>
  </xsl:template>
24 <xsl:template match="text()">
    <xsl:value-of select="normalize-space(.)" />
  </xsl:template>
28 <xsl:template match="*" />
</xsl:stylesheet>

```

Figure 5. LearnerPersonal XML record filter using XSLT, stripping all but the learner's RefId and name. XPath's shortened for brevity.

application's data footprint. This increases visibility of an applications data requirement *before* an institution engages the vendor or installs the application. Institutions are then given the ability to review and modify their data permissions for that application. Optional data requirements are denied by default, institutions must actively permit data sharing. Permissions are confirmed through digital signature of generated data sharing agreements. These data sharing agreements make what data is shared between each party explicit (to the attribute-level) and transparent, and can be used to produce data auditing documentation. Once all legal authorization is obtained through data sharing agreements DataExchange automatically provisions the underlying SIF infrastructure by generating appropriate environments, zones, credentials etc. It also configures the data store to ensure sensitive attributes are encrypted, making them encrypted attributes within encrypted objects. Part of this process is the generation of data filters in the form of XSLT files. Data filters are used to ensure only authorised data is processed by DataExchange, and only the authorized data for a particular vendor is shared with them. An example filter for a LearnerPersonal record is given in Fig. 5. All data is removed from an object (line 28) by default, data has to be explicitly allowed through the filter. For example the rules on lines 15 and 20 explicitly allow family and given name. Fig. 6 shows this process and the artefacts that are produced by DataExchange.

VI. CONCLUSION

DataExchange is a standards driven, privacy-first data integration and sharing platform. Our evaluation of DataExchange is against the principles of privacy by design [11]:

- **Proactive not Reactive; Preventative not Remedial.** DataExchange has been designed to prevent data breaches before they occur. Minimal data collection and data sharing through attribute-level explicit data sharing agreements ensures no additional data is exchanged than what has been authorized by all parties (Section V).
- **Privacy as the Default Setting.** In DataExchange access must be actively permitted rather than denied. This is applied consistently throughout the platform (Section V).
- **Privacy Embedded into Design.** DataExchange is developed around the core principles of minimal and explicit data sharing through attribute-level data controls. Privacy is integral to DataExchange, and is one of its unique selling points (Section III).
- **Full Functionality — Positive-Sum, not Zero-Sum.** DataExchange is fully-featured and can support all operations defined by the underlying SIF infrastructure, where the common use case is highly streamlined (Section IV), facilitating both privacy and security.
- **End-to-End Security — Full Lifecycle Protection.** Security is embedded from the point that data is collected, through transit over HTTPS, and encryption at rest (including attribute-level encryption for sensitive data). Each vendor has a unique set of access credentials. No more data is collected than has been authorized, which minimises the potential impact of data breaches. Once a data sharing agreement has been revoked, all associated services and data are automatically and securely deleted from the data warehouse (Section IV and V).
- **Visibility and Transparency — Keep it Open.** DataExchange's attribute-level approach ensures visibility of vendor data requirements as well as transparency in data sharing (Section V). Use of SIF specifications ensures transparency of infrastructure, where compliance (certification) also facilitates independent verification of the underlying technologies. The approach used in DataExchange has also been fed back into the A4L's Data Privacy Task Force for review and to share best practice.
- **Respect for User Privacy — Keep it User-Centric.** The data subject is at the heart of DataExchange, but data subjects do not typically use DataExchange directly. Instead DataExchange helps vendors and institutions comply with their legal and ethical responsibilities (Section V). For example, data sharing agreements can form a core part of data auditing and detailed privacy notices. If an institution knows exactly what data is collected and who it is shared with, then that knowledge can be shared with their students and/or their legal guardians.

While our discussion of DataExchange has focussed on commercial applications DataExchange can also be a research vehicle. Research projects requiring education data can be defined in the same way as commercial applications. Future work includes integration solutions for higher education, which will put DataExchange in a unique position to facilitate research into the entire student journey. We are looking to

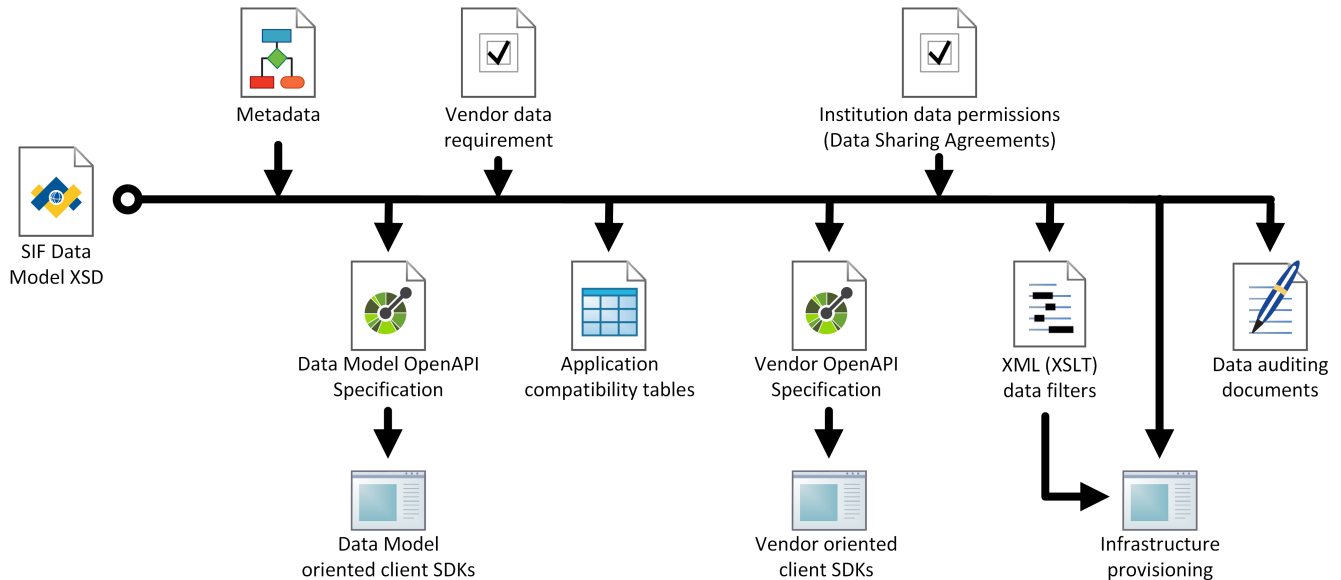


Figure 6. With a data model as input DataExchange produces a series of artefacts including filtering rules and client SDK libraries.

expand DataExchange into new territories/locales and help them establish new SIF communities, enabling privacy-driven access to education data at the global scale.

Finally, while DataExchange currently targets the education sector, with appropriate data model standards the underlying technologies and processes can be applied to improve transparency and privacy in other sectors, such as health and social care.

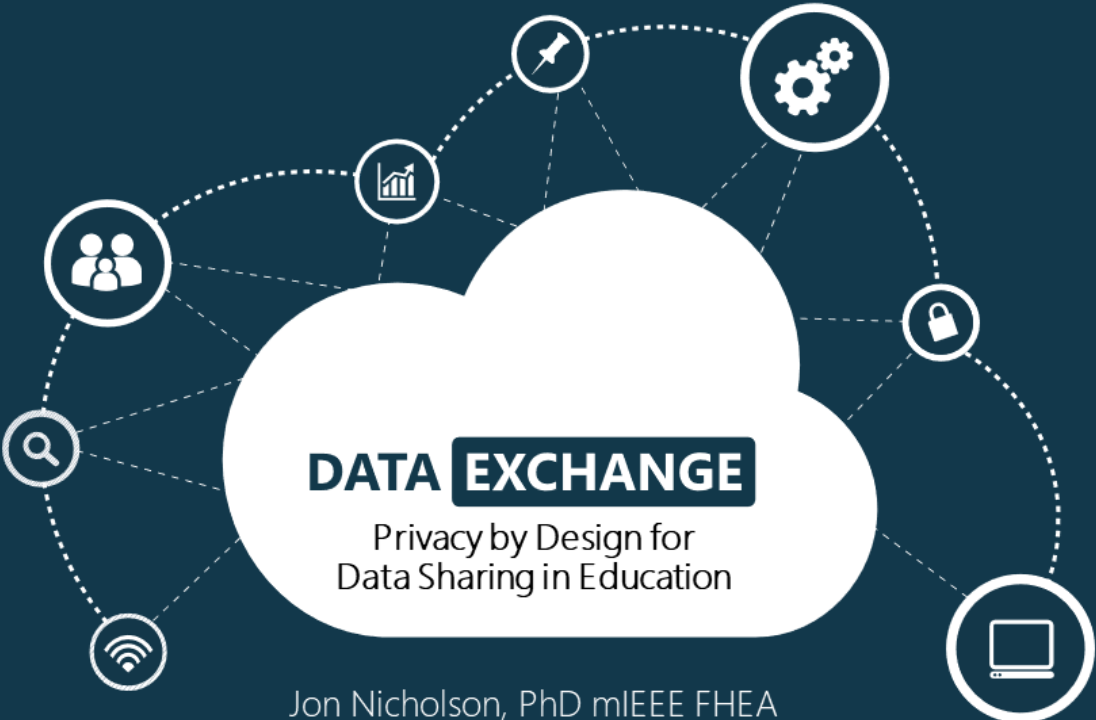
ACKNOWLEDGEMENT

We thank the Access 4 Learning Community, CPSI, The Hastings Academies Trust, and Laura Marquick for their support through the development of DataExchange.

REFERENCES

- [1] HMSO, "Data protection act 1998. (c.29)." [Online]. Available: <http://www.legislation.gov.uk/ukpga/1998/29>
- [2] European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119, pp. 1–88, May 2016.
- [3] GDPRiS, "GDPR in schools," 2017. [Online]. Available: <https://www.gdpr.school>
- [4] J. Kelly, "Data protection," Aug. 2017. [Online]. Available: <https://www.jisc.ac.uk/guides/data-protection>
- [5] R. Klahr, J. N. Shah, P. Sheriffs, T. Rossington, G. Pestell, M. Button, and V. Wang, "Cyber security breaches survey 2017," Apr. 2017. [Online]. Available: <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2017>
- [6] R. Khanna, "Data breaches: the enemy within," *Computer Fraud & Security*, vol. 2013, no. 8, pp. 8–11, Aug. 2013.
- [7] EduGeek, "Data protection and information handling," 2017. [Online]. Available: <http://www.edugeek.net/forums/data-protection-information-handling>
- [8] Open API Initiative, "OpenAPI-Specification," 2017. [Online]. Available: <https://github.com/OAI/OpenAPI-Specification>
- [9] Global Infrastructure Group, "SIF Infrastructure (Global) 3.2.1," Mar. 2016. [Online]. Available: <http://specification.sifassociation.org/Implementation/Infrastructure/3.2.1>
- [10] UK Technical Board, "SIF Data Model Implementation Specification (UK) 2.0," Nov. 2014. [Online]. Available: <http://specification.sifassociation.org/Implementation/UK/2.0>
- [11] A. Cavoukian, "Privacy by design: The 7 foundational principles," Jan. 2011. [Online]. Available: <https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf>
- [12] Groupcall Ltd, "Xporter," 2017. [Online]. Available: <https://www.groupcall.com/product/xporter>
- [13] Ark UK Programmes, "Assembly," 2017. [Online]. Available: <http://assembly.education>
- [14] Wonde Ltd, "Wonde," 2017. [Online]. Available: <https://wonde.com>
- [15] EduGeek, "Wonde and GCSEpod," Mar. 2017. [Online]. Available: <http://www.edugeek.net/forums/mis-systems/180268-wonde-gcsepod.html>
- [16] HMSO, "Education act 1996. (c.56)," [Online]. Available: <http://www.legislation.gov.uk/ukpga/1996/56>
- [17] —, "Higher education act 2017. (c.29)." [Online]. Available: <http://www.legislation.gov.uk/ukpga/2017/29>
- [18] M. Madhavaram, D. L. Ali, and M. Zhou, "Integrating heterogeneous distributed database system," *Computers & Industrial Engineering*, vol. 31, no. 1, pp. 315–318, Oct. 1996.
- [19] P. A. Hepner and W. Zhou, "Integrating heterogeneous databases: a distributed model," in *Proceedings of 3rd International Conference on Algorithms and Architectures for Parallel Processing*, Dec. 1997, pp. 695–702.
- [20] S. McClean, B. Scotney, and K. Greer, "A scalable approach to integrating heterogeneous aggregate views of distributed databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 232–236, Jan. 2003.
- [21] M.-C. Rousset and C. Reynaud, "Knowledge representation for information integration," *Information Systems*, vol. 29, no. 1, pp. 3–22, Mar. 2004.
- [22] A4L Community, "About the A4L Community," 2017. [Online]. Available: <http://www.a4l.org/page/AboutA4L>
- [23] NA Technical Board, "SIF Data Model Implementation Specification (NA) 3.4," Sep. 2016. [Online]. Available: <http://specification.sifassociation.org/Implementation/NA/3.4>
- [24] AU Technical Board, "SIF Data Model Implementation Specification (AU) 3.4.1," Apr. 2017. [Online]. Available: <http://specification.sifassociation.org/Implementation/AU/3.4.1>
- [25] J. Howse, G. Stapleton, K. Taylor, and P. Chapman, "Visualizing Ontologies: A Case Study," in *The Semantic Web ISWC 2011*, ser. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, Oct. 2011, pp. 257–272.

Presentation Slides





The graphic features a central white cloud on a dark blue background. Inside the cloud, the text "DATA EXCHANGE" is written in bold white letters, with "EXCHANGE" in a dark blue box. Below it, the subtitle "Privacy by Design for Data Sharing in Education" is written in white. Surrounding the cloud are eight circular icons connected by a dotted line: a group of people, a bar chart, a pushpin, gears, a padlock, a laptop, a Wi-Fi symbol, and a magnifying glass.

DATA EXCHANGE
Privacy by Design for
Data Sharing in Education

Jon Nicholson, PhD mIEEE FHEA
Chief Technology Officer, ZINET Data Solutions Limited

1st International Conference on the Frontiers and Advances in Data Science
23—25 October 2017. XiAn, China

 @DrJonNicholson @ZINETHQ @DataExchangeEdu 



+44 (0)330 223 1070
www.zinethq.com

Session outcomes

By the end of this session you should:

1. Understand the 7 foundational principles of privacy by design
2. Appreciate how the UK has been impacted by privacy issues
3. Identify current privacy issues in UK Education market
4. Recognise how DataExchange uses a novel attribute-level approach to implement privacy by design
5. Engage with us about our paper and how DataExchange can support your data demands

Privacy by design

The 7 foundational principles of privacy by design are:

1. Proactive not Reactive
2. Privacy as the Default Setting
3. Privacy Embedded into Design
4. Full Functionality
5. End-to-End Security
6. Visibility and Transparency
7. Respect for User Privacy

A. Cavoukian, "Privacy by design: The 7 foundational principles," Jan. 2011.

Impact of privacy issues in the UK

- 46% of UK businesses have experienced a security breach in the last 12 months ^[1]
- On average, data breaches cost UK businesses up to £138,700 and require over 9 months to recover ^[2]
- Only 22% of education, health or social care firms require suppliers adhere to cyber-security standards ^[1]
- Compliance/legal infractions and insecure third parties are among the top causes of data breaches ^[2]

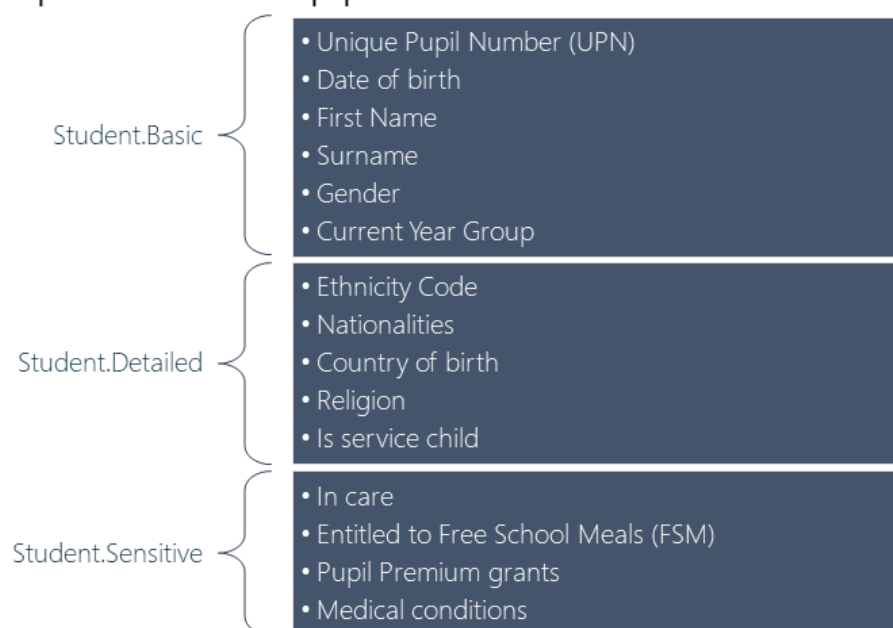
[1] R. Klahr, et al., "Cyber security breaches survey 2017," Apr. 2017. www.gov.uk/government/statistics/cyber-security-breaches-survey-2017

[2] R. Khanna, "Data breaches: the enemy within," Computer Fraud & Security, vol. 2013, no. 8, pp. 8–11, Aug. 2013.

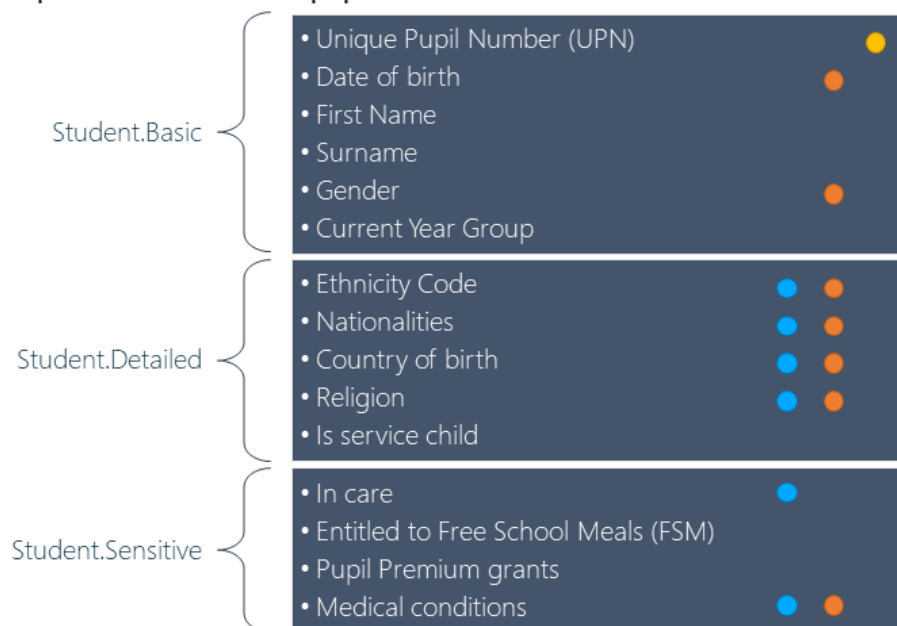


Where is the data going and why?

Scope-based approach

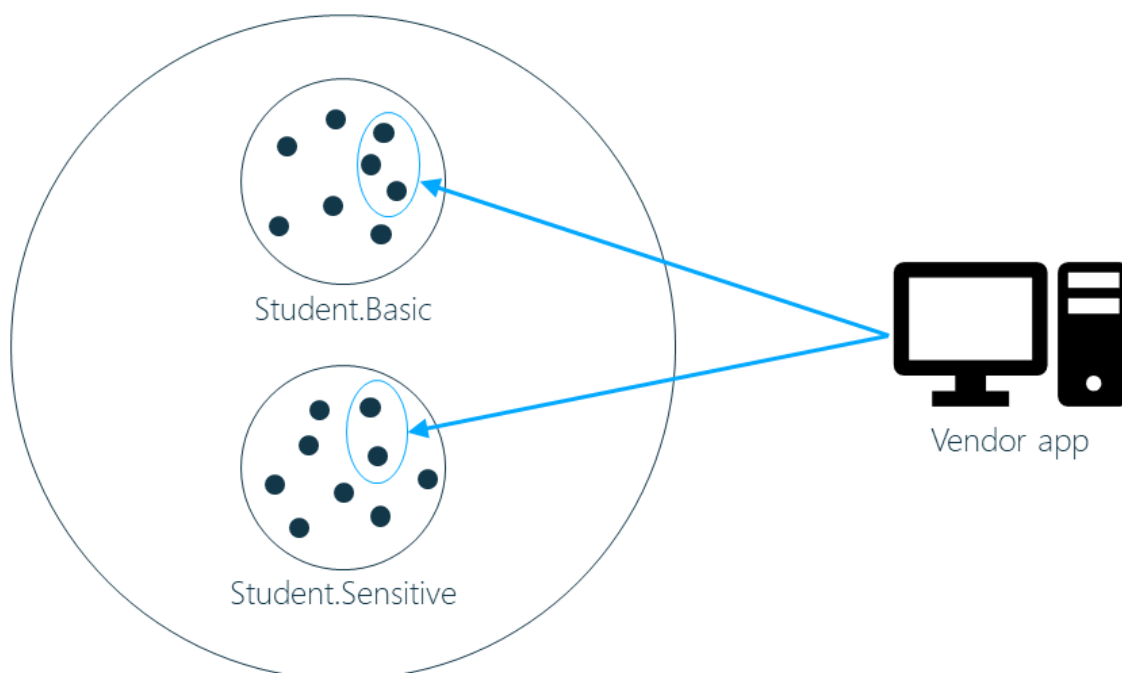


Scope-based approach

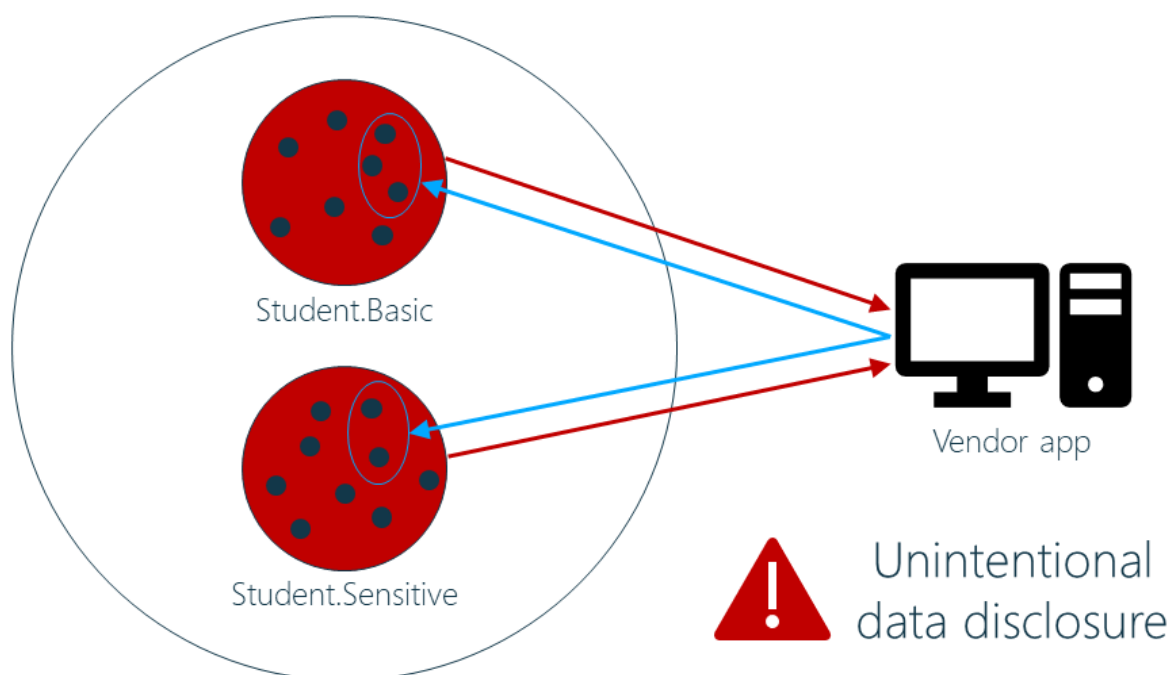


- Sensitive Data (Data Protection Act 1998, <https://www.legislation.gov.uk/ukpga/1998/29/section/2>)
- Protected Characteristics (Equality Act 2010, <https://www.legislation.gov.uk/ukpga/2010/15/section/4>)
- Unique Pupil Number (DfE guidance, <https://www.gov.uk/government/publications/unique-pupil-numbers>)

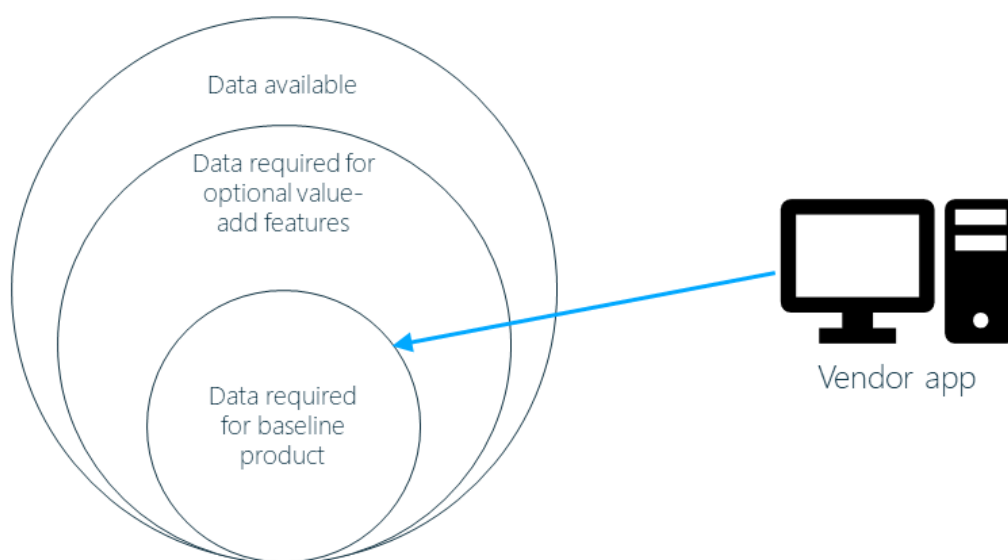
Data available at institution



Data available at institution

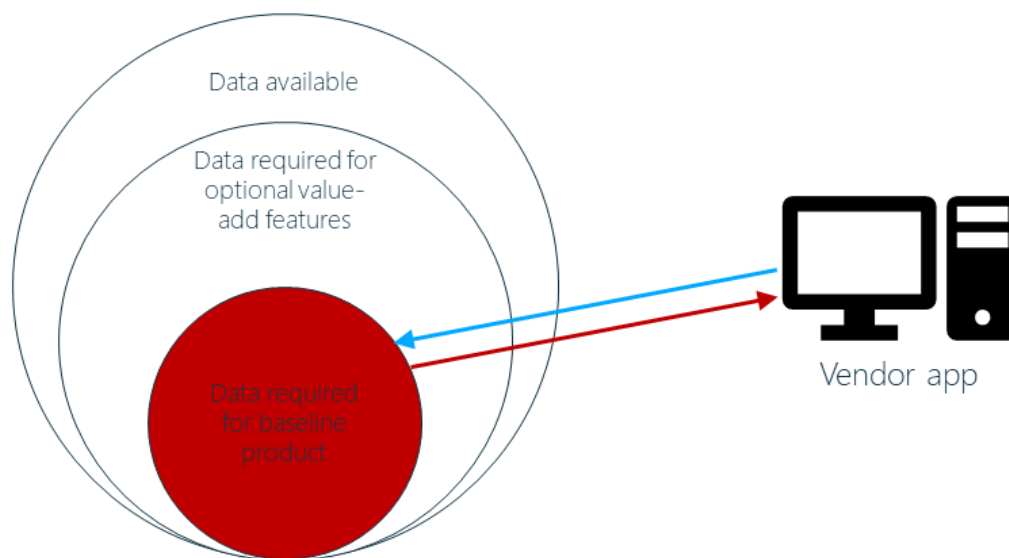


Attribute-based approach



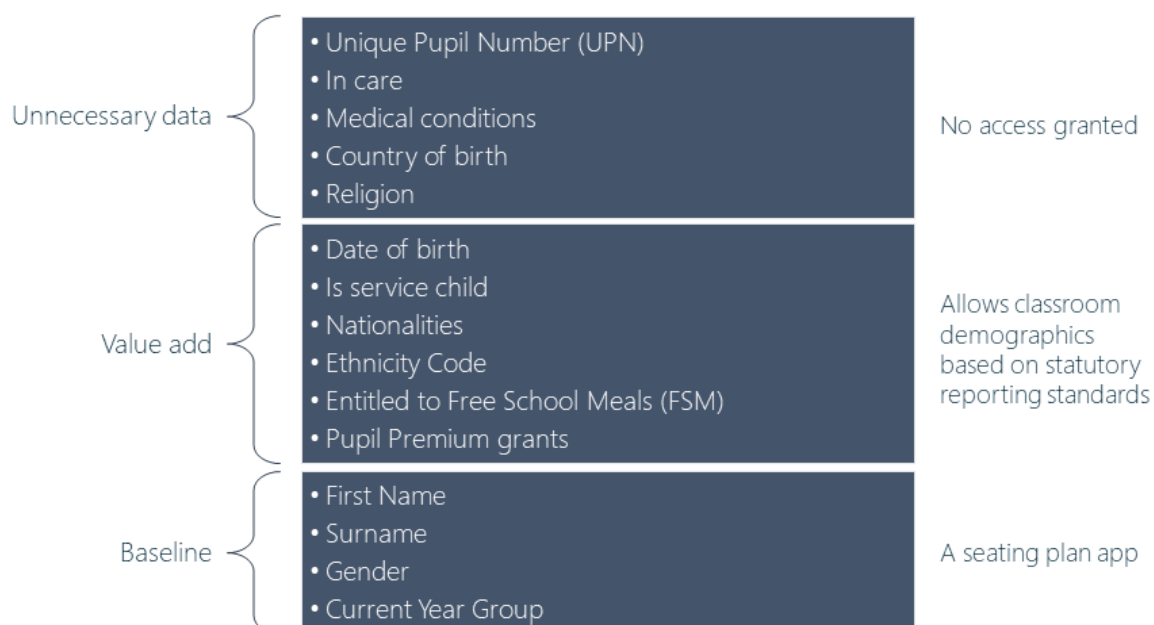
Attribute-level controls afford fine grained privacy management

Attribute-based approach

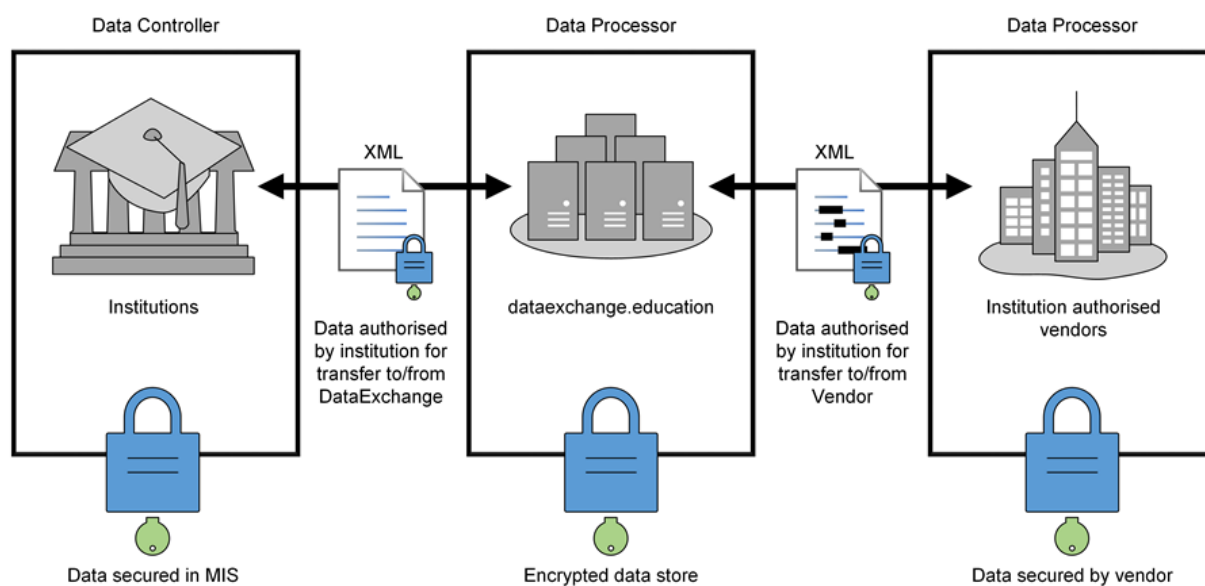


Attribute-level controls afford fine grained privacy management

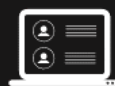
Attribute-based approach



DataExchange: At a glance



Data integration



Data integration

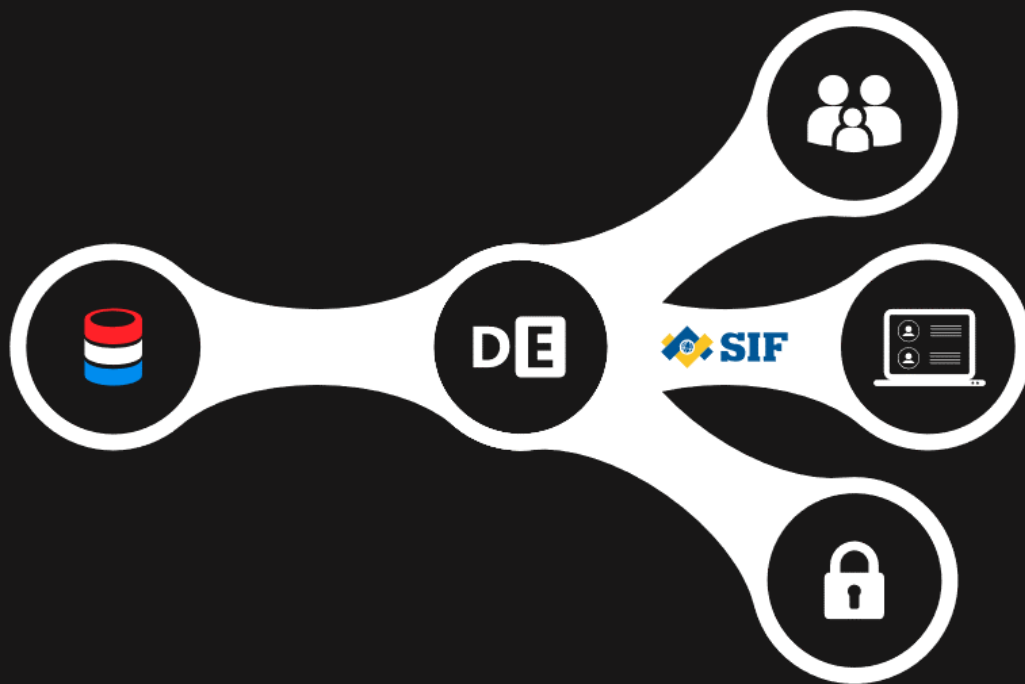


Simple Secure Scalable Standard

Data sharing

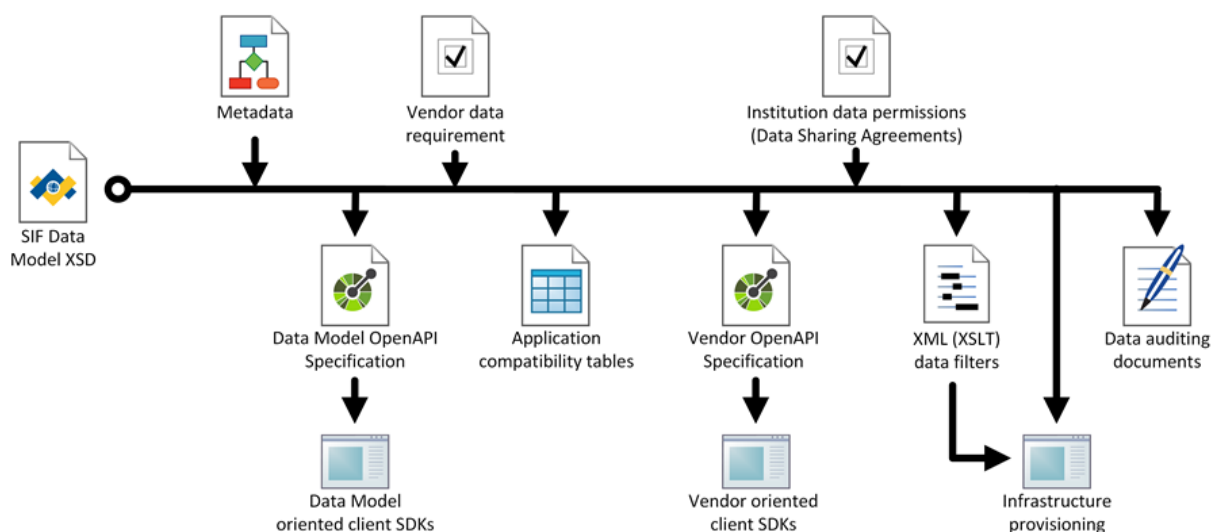


Data sharing



SIF answers the how, not the why

Artefact generation



DataExchange: Unique selling points

Minimal data collection

- Attribute-level data sharing agreement (2-party)
- Extract and store only required data to support sum data sharing

Minimal data sharing

- Attribute-level data sharing agreement (3-party)
- Only specified data is shared to identified vendor services

Visibility

- Data requirements of vendor apps are explicit
- Allows institutions (data controllers) to make informed choices

Transparency

- Data sharing agreements allow calculation of digital footprint
- Data audits, privacy impact assessments, privacy notices, etc.

DataExchange: Privacy by design

DataExchange complies with the 7 foundational principles of privacy by design:

- 1. Proactive not Reactive**
Attribute-level explicit data sharing agreements
- 2. Privacy as the Default Setting**
Access must be actively permitted rather than denied
- 3. Privacy Embedded into Design**
It is one of our USPs!
- 4. Full Functionality**
All operations defined by the underlying SIF infrastructure are possible
- 5. End-to-End Security**
Encryption in transit (HTTPS) and at rest with secure access credentials
- 6. Visibility and Transparency**
Attribute-level data sharing improves visibility and transparency
Use of SIF and other open standards ensures transparency of infrastructure
- 7. Respect for User Privacy**
Ensures students and/or legal guardians can be better informed of their digital footprint

Where next...?

Vertical education projects

Researching the student journey, from Nursery to HE, to improve performance, recruitment and retention at all levels

Expanding into new locales

Australia, America, China, ...

Expand into new markets

Health care, social care, ...

Where can we help you?

Implementation, consultancy, research, ...



Thank you