FOLIO Data Privacy Roadmap

Written by: FOLIO Privacy SIG, in collaboration with Becky Yoose, LDH Consulting Services

January 2024

About the Roadmap

This roadmap aims to address fundamental questions asked by members of the FOLIO Privacy SIG about how data governance is currently conceived and managed across the FOLIO initiative. Time and availability constraints among members prevented the group from addressing these complicated problems in monthly zoom calls. Adam Chandler, chair of the SIG, contacted Becky Yoose of LDH Consulting Services to assist the SIG in creating the data privacy roadmap. Becky is a noted expert on library privacy, having worked with an array of different library staff and organizations. With funding from Cornell University Library, Becky served as the lead author and consultant to the SIG for the review of data privacy practices in FOLIO development and in creating this roadmap. Members of the SIG provided expert commentary and review during the creation of the roadmap.

Executive Summary

This high-level roadmap serves as a starting point for a broad, multi-level review of FOLIO privacy policy and practices. The roadmap starts with a description of current data privacy practices and responsibilities in FOLIO development, ranging from what has been achieved by the FOLIO Privacy SIG to the complicated question of responsibility and accountability. The roadmap then summarizes the constraints to addressing those problems within the current FOLIO organizational structure. The roadmap concludes with an ambitious vision of how FOLIO could create an approach to data privacy in library open source software that could be a model for other OSS projects. The recommended changes would require adjustments in how FOLIO apps are designed and developed, including the creation of data privacy policies and integration of data privacy frameworks such as Privacy by Design. The conclusion returns to the question of how FOLIO could identify and dedicate resources in order to achieve the vision laid out in the roadmap.

Data Privacy at the Margins

FOLIO's current approach to data privacy centers around the work performed by the Privacy Special Interest Group (Privacy SIG). Formed in 2017 and reconvened in 2021, the Privacy SIG provides subject matter expertise to guide the FOLIO project and its members in securing personally identifiable information (PII) in the FOLIO platform, ideally on both conceptual and practical levels. Privacy SIG members are expected to have extensive knowledge of library data privacy standards, guidelines, and best practices. Members are also expected to know about regulations, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), ensuring that FOLIO users can comply with these regulations.

In practice, the Privacy SIG currently addresses specific data privacy and security issues (e.g., system log file handling and retention) on a case-by-case basis. The Personal Data Disclosure (PDD) form is the primary major project led by the Privacy SIG. The PDD allows FOLIO users to comply with data protection regulations by disclosing what types of PII are stored in each FOLIO module. Integrating the PDD into GitHub modules is a step toward automating the process of recording which modules work with personal information.

Despite the relative success of integrating the PDD form into FOLIO modules, other efforts toward regulatory compliance, such as a FOLIO Privacy Policy and data flows, have not progressed since initial discussions in the SIG and the community. Like many FOLIO groups, the FOLIO Privacy SIG is not resourced adequately for itself to take on a deeply embedded role in the project. Ultimately, while the Privacy SIG provides some data privacy guidance, a successful data privacy approach in a complex ecosystem such as the FOLIO project requires a multi-level approach across the project. This success is currently limited by several factors within the FOLIO Project:

Privacy-by-Compliance (PbC)

Most data privacy work in FOLIO focuses on compliance with GDPR and CCPA, with the PDD as a prime example. Systems should assist users in complying with data protection regulations, but compliance must not be the final goal of a data privacy program. While some data protection regulations incorporate select elements of data privacy frameworks (i.e., Privacy by Design), data protection regulations often do not provide adequate data privacy protections found in general industry standards and best practices, including the library profession.

Another drawback of PbC is that compliance is primarily a reactive approach to data privacy. Data protection bills take years in the legislative process to become law. Any protections that become part of the final regulation will not account for the torrent of changes in technology that took place between the introduction of the bill and signing the bill into law. On the other hand, professional standards and best practices tend to be more flexible in addressing these rapid changes. On the community level, communities of practice in the library and privacy professions are well positioned to provide proactive data privacy guidance beyond regulations' lagging data protection benchmarks.

Tacked-on Privacy

Data privacy, like security and accessibility, are commonly addressed very late in the development process, if at all. FOLIO's development process is no exception. The current process for developing new modules provides limited chances for reviewing possible privacy risks and impacts, with no privacy

requirements during the initial approval through Product Council and the only requirement of having a PDD form during the final checklist process for Technical Council. Currently, there is no approval or oversight process for developing new features in existing modules, which increases the possibilities of data security and privacy related vulnerabilities being introduced into existing modules already being used in production. There are a few user stories that include privacy elements, but there is no systematic integration of data privacy in the Agile process used to develop features and modules.

The lack of integration of privacy into the design and development processes opens the possibility of developing module features that run counter to international standards and ethical codes in the library profession. For example, FOLIO APIs enable developers to interface with various FOLIO modules through another program or system. A developer can use APIs to query and retrieve data in FOLIO, including querying and retrieving user data. There is no shortage of documentation about the "what" and "how" of the APIs, including:

- API documentation (https://dev.folio.org/reference/api/)
- API documentation endpoints (https://dev.folio.org/reference/api/endpoints/)
- How to use APIs (https://dev.folio.org/faqs/how-to-use-apis/)
- Working with FOLIO APIs (https://wiki.folio.org/display/FOLIOtips/Working+with+FOLIO+APIs)

The documentation, however, does not provide guidance around the technical guardrails that should be in place for programs accessing user data in FOLIO via an API. The closest technical guardrail in the documentation is the Okapi Security Model (https://github.com/folio-org/okapi/blob/master/doc/security.md) which discusses how to secure the API gateway software Okapi via authentication and authorization of API users. There is no discussion of the potential privacy risks in working with FOLIO user data in third party programs that use FOLIO APIs, be it an off the shelf third party product or a script that was built in-house for local operational processes.

Another example, the Library Data Platform (LDP) harvests data from FOLIO modules and creates a data warehouse for reporting and analytics. LDP can also include data outside of FOLIO and integrate with third-party applications developed outside the main FOLIO branch. Even with an "anonymization" feature, the LDP documentation includes the following warning:

> "WARNING: LDP does not provide a way to anonymize the database after personal data have been loaded into it. For this reason, anonymization should never be disabled unless you are absolutely sure that you want to store personal data in the LDP database."[1]

Both examples show how, without technical and administrative guardrails in the development process and in production, features and functionalities built into the FOLIO platform could enable use that conflicts with library standards and codes, such as the recommendations in the International Federation of Library Associations and Institutions (IFLA) Statement on Privacy in the Library Environment:

> "Library and information services should reject electronic surveillance and any type of illegitimate monitoring or collection of users' personal data or information behavior that would compromise their privacy and affect their rights to seek, receive and impart information. They

[1] Library Data Platform, "LDP Administrator Guide," July 16, 2022, https://github.com/library-data-platform/ldp/blob/1.8.2/doc/Admin_Guide.md#6-data-privacy .

should take measures to limit collection of personal information about their users and the services that they use."[2]

Technical guardrails put in place throughout the development process can mitigate uses in production that directly conflict with professional standards, codes, and legal regulations. While these guardrails can be put in place during the approval processes in Product and Technical Councils, there must be guardrails throughout the entire development process. However, there are structural and material limitations that must first be addressed for any guardrails to be effective.

Lack of Dedicated Resources for Privacy

Like other open source projects, FOLIO depends heavily on community members in developing and sustaining the project. FOLIO is not without vendor support – EBSCO and Index Data are two major vendors providing dedicated staff and resources for development and product ownership. Nevertheless, most of the community work, particularly in the SIGs, is subsidized by libraries through the work of library workers. Typically, these library workers must balance their work in FOLIO with their other job commitments, with FOLIO community work placed on top of existing work responsibilities. Professional development (e.g., training, workshops, and monitoring news and websites for the latest updates) can be almost impossible for many workers who already struggle finding time and resources to perform their core job duties.

Special Interest Groups, such as the Privacy SIG, try to emulate communities of interest in bringing together like-minded individuals in their interest about a particular topic. We cannot forget, though, that these individuals do not have dedicated time or resources from either FOLIO or their library to train and develop data privacy skills and knowledge. Data privacy is subject to the same rapid pace of change as technology. The same is true with the number of changes that happen in rapid succession, particularly with developments in privacy-preserving and privacy-invasive technologies. Individual library workers whose primary job responsibilities do not center around privacy have the burden of staying current on data privacy matters in addition to their primary job responsibilities. Combined with the ever-present threat of burnout that is commonplace in many open source projects, the community's lack of dedicated resources for privacy work severely limits any attempt at a comprehensive data privacy approach in FOLIO.

The Perilous Balance of Who is Responsible for What

Discussing data privacy within an open source project inevitably brings up the question of responsibility. There have been attempts in creating privacy standards for open source projects (e.g., Open Source Privacy Standards[3]), but there have been no specific discussions dedicated to the responsibilities of the open source community and

[2] International Federation of Library Associations and Institutions, "IFLA Statement on Privacy in the Library Environment," August 14, 2015, https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/ .

[3] Burns, Heather, "Open Source Privacy Standards," last modified June 5, 2018, accessed Nov. 30, 2023, HYPERLINK "https://github.com/webdevlaw/open-source-privacy-standards"https://github.com/webdevlaw/open-source-privacy-standards .

It is this lack of delineation of responsibilities – even at a high level – that has the ability to stall any substantial movement toward a more privacy-preserving FOLIO. For example, the responsibilities around data privacy in the software could be assigned to the FOLIO community, but where does that responsibility shift when it comes to an end user customizing or modifying the FOLIO software? What should happen when a person or organization makes repeated code contributions to the FOLIO project that are clearly out of line with privacy standards? Another example comes in the way of legal matters. End users are ultimately responsible in ensuring that they comply with applicable data protection regulations, but they need a product that is designed to meet those compliance requirements.

The examples about legal and technical responsibilities present some natural delineations of responsibility. The community that contributes to the FOLIO main branch is the primary party responsible in ensuring that the technical components of FOLIO preserve data privacy with the understanding that end users who fork from the main branch then take on that responsibility. End users of FOLIO must take responsibility in data protection regulation compliance matters at their organization, including working with the FOLIO community in the cases that the software prevents said compliance.

Nevertheless, the one obligation that both the community and the operators have is to professional standards and ethics. The community and the organizations that adopt FOLIO have the obligation to develop and use FOLIO in the best interests of the patron, including the privacy of the patron's use of the library. Libraries' reliance on data analytics continues to grow – this is evident with the development and use of the LDP and Panorama (as well as WorldShare Reports, Alma Analytics, and specialized products like LibConnect, Patron Point, and Savannah) – but this reliance does not mean that the profession is no longer obligated to mitigate or eliminate patron surveillance and tracking. The collection, use, retention, and disclosure of patron data in FOLIO must be guided by a shared understanding around data privacy in the library by the FOLIO community and organizations that use FOLIO. This includes a shared understanding of how to mitigate privacy risks in collecting, using, and disclosing personally identifiable patron data, particularly when using a data warehouse for personal data.

What is Holding Us Back – Constraints

The constraints surrounding the FOLIO community's ability to enact a comprehensive data privacy approach in the development process are more extensive than individual groups or processes. Several constraints are systemic and require fundamental changes to the development process and product governance:

- As mentioned earlier, there is no formal process for approving new features in existing modules, increasing the chances for vulnerabilities in the software that can be exploited. The lack of a formal process can also limit the effectiveness of any development governance or policies created or enforced by the community.
- There is an overall misalignment between the Product and Technical Councils regarding FOLIO development approaches, outcomes, and priorities. At the time of writing, the Product Council is creating a formal functional review process. The Technical Council currently has a checklist that all new modules must complete, including a check for the PDD form and checks for data security. Though these processes operate independently of each other, there is no coordination

around functional and nonfunctional requirements, design, and systematic reviews of module or feature impacts on data privacy.

- There seems to be no agreed-upon approach to privacy in the development process among the leads for module development, including Product Owners, Technical Design Owners, and Code Leads. Product Owners are encouraged to attend SIG meetings to understand the SIG's area of expertise better, but outside of discussions of individual issues with data security, there is no uniform approach or collaboration between the module leads around integrating privacy into the development process. However, this is not the only area where integration is lacking. Security and accessibility also struggle to find their way into the development process, with the closest integration found being accessibility guides specifically geared toward developers and product owners.

Other constraints are due to the nature of open source communities and the architecture of FOLIO:

- FOLIO is used worldwide. The international user community has different regulatory requirements, patron privacy risks, and privacy needs for both users and employees. Any comprehensive privacy approach must have the capacity to handle such differences while not settling for the lowest denominator for privacy defaults.
- The modular nature of FOLIO can complicate any initial dependency or data flow mapping. The development of modules or features to existing modules will require additional processes and resources to maintain those processes to minimize the number of privacy-related vulnerabilities introduced into the main FOLIO branch.
- Resource availability heavily depends on institutions' willingness to provide money and staff time to any development process or committee work.

Any comprehensive approach to data privacy in the FOLIO development process must account for these constraints. The constraints of competing user priorities and system architecture will affect what the community can realistically take on and achieve in any data privacy practice or initiative. Nevertheless, a sustainable data privacy approach requires addressing the systemic inconsistencies found in the development process on multiple levels of leadership and governance.

Binding Data Privacy into FOLIO

Libraries are patron data privacy stewards. This stewardship extends to the library service platforms (LSPs) that comprise the core of library operations. FOLIO is one of several LSPs available in the marketplace, but its position as an open source LSP that scales to large academic libraries places it in direct competition with one of the most prominent vendors in the LSP marketplace. FOLIO is in a unique position in that the community developing the LSP has the opportunity to make a solid commitment to align FOLIO's approach to data privacy to the library profession's codes of ethics and standards. Situated within a marketplace dominated by companies that have privacy-invasive data analytics business lines[4], FOLIO could have a potential competitive advantage if the community takes a privacy-first approach to

[4]Lamdan, Sarah. *Data Cartels: The Companies That Control and Monopolize Our Information*. 1st edition. Stanford, California: Stanford University Press, 2022.

their LSP development. This raises the question of how a commitment to privacy would look in FOLIO, particularly in the development cycle.

Fortunately, there are existing privacy frameworks and practices that can guide organizations and projects, like FOLIO, in creating systems that can support sustainable, comprehensive data privacy approaches in their work:

- **Privacy by Design (PbD) and by Default** – one of the most commonly known and used privacy frameworks in general industry, Privacy by Design's seven principles center around embedding privacy in all areas of the development process:
    - Proactive not reactive; preventive not remedial
    - Privacy as the default setting
    - Privacy embedded into design
    - Full functionality – positive-sum, not zero-sum
    - End-to-end security – full lifecycle protection
    - Visibility and transparency – keep it open
    - Respect for user privacy – keep it user-centric[5]

  PbD is included in several data protection regulations, such as Article 25 in the General Data Protection Regulation (GDPR). PbD also became an ISO privacy standard in 2023.[6]
  Privacy by Default, originally one of the seven principles of PbD, is a complementary framework that takes a more privacy-preserving approach by setting system defaults to the highest level of privacy protection for the user.
- **Data Governance** – ensuring data privacy throughout FOLIO requires oversight of data privacy controls, policies, and standards. FOLIO already has steering committees that oversee the product and technical work in FOLIO development. A steering committee consisting of people that represent different areas of the development process and community can shape the direction of data privacy work in FOLIO. A centralized committee with authority to create policies and resolve disputes around disagreements or misalignments in the development process would provide stability needed for long-lasting change to existing development practices. A data governance committee also allows for a single entity to take on work regarding accountability and transparency in FOLIO's data privacy work.

Privacy by Design, Privacy by Default, and data governance require fundamental structural changes in FOLIO. Nevertheless, there are several actions FOLIO can take now or in the near future that work toward fully adopting these frameworks in the development process. These actions fall under three general categories: policy, the development lifecycle, and resources.

Policy and Process

1. **Create a policy stating what library and general data privacy standard(s) FOLIO will use to guide privacy practices in the development lifecycle** – FOLIO operators have specific compliance requirements depending on their jurisdiction – the PDD is an example of the FOLIO project assisting their users in meeting those requirements – and FOLIO will need to continue to

[5] Cavoukian, Ann. *The 7 Foundational Principles. Implementation And Mapping of Fair Information Practices*. Toronto: Information and Privacy Commissioner's Office, Ontario, 2011, https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf .
[6] Consumer Protection: Privacy by Design for Consumer Goods and Services, ISO 31700-1, 2023, https://www.iso.org/standard/84977.html .

look at regulations to ensure that FOLIO users can achieve compliance. However, any guidance policy must go beyond a compliance-first approach and look toward library privacy standards and commitments to bring FOLIO more aligned with them. A starting point that reflects the international nature of FOLIO's community is the IFLA Statement on Privacy in the Library Environment.[7] FOLIO can also use this policy to formalize its commitment to PdD, Privacy by Default, and other general data privacy standards and frameworks. Additionally, the policy can be a starting point to re-align the Product and Technical Councils in their approaches to data privacy when evaluating and assessing new modules and features in existing modules. Finally, this policy can also guide language in the contributor license agreement in which contributors acknowledge that their code will meet the data privacy standards as described in the policy.

2. **Incorporate data privacy and security requirements into the Product Council design evaluation process** – Product Council is working on a new design evaluation process. Incorporating data privacy in the evaluation process creates a consistent approach to new modules and features, ensuring a consistent application of privacy requirements throughout the FOLIO module environment. Several resources listed under the Development Lifecycle section (i.e., incorporating privacy design strategies and PbD into Agile development practices) provide examples of what these nonfunctional requirements can look like in the evaluation process.
3. **Conduct a data audit of the official build** – The PDD is the first step in tracking the flow of personal data through the FOLIO system. Documenting the flow of personal data through FOLIO can serve as a basis for data privacy audits and reviews of existing modules. The audit also provides the opportunity to create data privacy and security nonfunctional requirements for new development.
4. **Create a process for regularly scheduled data privacy reviews of existing modules** – Currently, there is no formal oversight of the development of new features in existing modules. While creating a formal process for said development can help mitigate potential data privacy risks, creating a data privacy review of existing modules can catch potential risks introduced into the code that might not have been caught in the initial development.
5. **Create a policy regarding possible privacy issues in community contributions to FOLIO –** Due to the distributed nature of open source projects on the scale of FOLIO, there will be variation in the skillsets, knowledge, and values surrounding data privacy in the community contributors to the code base. Proactive education (e.g., training, documentation) will help guide community members to make sure that their code contributions meet privacy and security standards and requirements. Nevertheless, FOLIO will need to be prepared to address privacy issues found in community contributions to the code base. The policy will need to consider the context of the issue and if there is a pattern of deliberately submitting contributions that go directly against FOLIO privacy standards and requirements. FOLIO should reserve the right to restrict contributions to the official code base, but only after other venues for remediation have been exhausted (i.e., additional training, code review assistance, documentation review and changes as necessary).

[7] International Federation of Library Associations and Institutions, "IFLA Statement on Privacy in the Library Environment," August 14, 2015, https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/ .

Development Lifecycle

1. **Introduce data-oriented privacy design strategies early in the design process** – In conjunction with the Product Council design evaluation process, Technical Design Owners and, to an extent, Product Owners are in a position in the development lifecycle to incorporate data-oriented privacy design strategies into new modules and features. These four strategies – hide, minimize, abstract, and separate – offer several ways (i.e., tactics) for product teams to implement privacy-preserving functions and designs early in the development lifecycle. The Little Blue Book of Privacy Design Strategies is a starting point for practical guidance on the strategies and tactics for both data- and process-oriented privacy design strategies.[8]
2. **Integrate data privacy design strategies into existing Agile development practices** – FOLIO's Agile development process already includes some data privacy-related elements, primarily including privacy in some user stories. Nevertheless, many opportunities exist for expanding data privacy's inclusion into existing practices. The Catalan Data Protection Authority's Privacy by Design and Privacy by Default: A Guide for Developers and the MITRE Privacy Engineering Framework and Lifecycle Adaptation Guide provide a general overview of how to incorporate PbD and Privacy by Default into a development lifecycle and can be used to create documentation and guidance for product owners and developers.[9] The Catalan Data Protection Authority's guide is a good resource to start the conversation about incorporating privacy-enhancing technologies (PETs) into FOLIO, while the MITRE document provides a guide on incorporating privacy engineering activities into Agile development lifecycles. Agile-specific examples of integrating PbD and Privacy by Default into a development lifecycle include Engin Bozdag's presentation about implementing PbD at Uber, Threat Poker, and Privacy Criteria Methods.[10]
3. **Create privacy design and engineering documentation for developers and product owners** – Documentation describing the application of the standards adopted by the community (see #1 under Policy and Process) will provide guideposts for product owners and developers throughout the development lifecycle. The Accessibility SIG has similar documentation for product owners and developers, and the privacy documentation should use these documents as

---

[8] Jaap-Henk Hoepman, "Privacy Design Strategies (The Little Blue Book)," April 19, 2022, https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf .

[9] Catalan Data Protection Authority, "Privacy by Design and Privacy by Default: A Guide for Developers," February 2023, https://apdcat.gencat.cat/web/.content/03-documentacio/documents/guiaDesenvolupadors/GUIA-PDDD_EN.pdf; MITRE Corporation, "MITRE Privacy and Engineering Framework and Lifecyle Adaptation Guide, Version 2," September 29, 2019, https://www.mitre.org/sites/default/files/2021-11/pr-19-00598-5-privacy-engineering-framework-v2.pdf .

[10] Engin Bozdag, "Privacy at Speed: Privacy by Design for Agile Development at Uber," *USENIX Association*, January 28, 2020, https://www.usenix.org/conference/enigma2020/presentation/bozdag; Rygge, Hanne, and Audun Josang. 2018. "Threat Poker: Solving Security and Privacy Threats in Agile Software Development." In Proceedings of the 23rd Nordic Conference on Secure IT Systems, 2018. Oslo, Norway, https://www.duo.uio.no/bitstream/handle/10852/72016/2/RJ2018-NordSec.pdf; Peixoto, Mariana, Carla Silva, João Araújo, Tony Gorschek, Alexandre Vasconcelos, and Jéssyka Vilela, "Evaluating a privacy requirements specification method by using a mixed-method approach: results and lessons learned," Requirements Engineering 28 (2018): 229-255, https://doi.org/10.1007/s00766-022-00388-2 .

a template when creating the documentation.[11] The documentation can become a part of a larger knowledgebase where community members can find information such as glossaries, templates, standardized functional and non-functional requirements, and data management guidelines.

4. **Determine minimum viable privacy in Agile development practices –** While incorporating data privacy into the development lifecycle, there must be agreement as to the level of privacy that is needed when determining when a feature is considered "done" and ready for [testing/final review/release]. A minimum set of requirements around privacy can help refine the definition of "done" that takes into account data privacy requirements.

Resources

Identify and reserve resources for dedicated data privacy subject matter experts in the project – The limitations of the SIG's ability to fulfill the role as privacy subject matter experts are due, in part, to the reliance on contingency-based resources and staffing through the institutions that participate in the community. A dedicated privacy subject matter expert or a dedicated privacy team directly addresses these limitations. A dedicated privacy person or team creates a more stable environment that can push for a more consistent, sustainable data privacy approach in FOLIO.

Funding for such dedicated positions is not easy for any open source project. Having stable funding come from the wider community behind FOLIO will take time, but reduces the chance of any conflicts of interest through heavy reliance on any one organization or institution to provide the expertise. While the community identifies stable long-term funding for these positions, FOLIO can explore working with vendors, such as EBSCO, that already have dedicated privacy staff to temporarily provide privacy subject matter expertise to the project for the short term.

[11] FOLIO Project, "For Product Owners: Accessibility Testing & Requirements," July 12, 2021, https://wiki.folio.org/pages/viewpage.action?pageId=25724070; FOLIO Project, "For Developers: FOLIO Accessibility Guidelines," July 13, 2021, https://wiki.folio.org/display/A11Y/For+Developers%3A+FOLIO+Accessibility+Guidelines .