Loop Users in: The Key to Cross-Platform Data Interoperability

Han Su, Jialing Wu, Lifeng Liu,
Yingxuan Zhu, Jian Li

Futurewei Technologies
Boston, Massachusetts
Email: (hsu, jwu3,
lifeng.liu, yingxuan.zhu,
jian.li)@futurewei.com

Liyang Zhu, Boyan Xu

New York University
New York, USA
Email: (bx376, tomzhu)@nyu.edu

***Abstract*—The lack of data interoperability on today's Internet platforms has led to data ownership problems, which further lead to the deprivation of netizen participation and representation in the data economy. In this paper, we conclude that user-oriented data interoperability is the key to change the siloed app ecosystem toward a more decentralized direction. This paper examines the various attempts made by the industry to increase interoperability at different levels: software level, platform level, and infrastructure level. Web 1.0 granted netizens the right to view online content and Web 2.0 (Apps) has given netizens the right to publish in a participatory manner. In this paper, we envision that the next-generation Internet platforms will enable netizens to access personal storage and computation, which can recuperate netizens' right for data ownership and representation in the data economy.**

Keywords–Data Interoperability; Web 3.0; Data Ownership; App Ecosystem; Data Economy; Privacy.

I. INTRODUCTION

With the rise of Internet companies, Internet platforms have become more centralized in terms of data storage, which has led to myriad problems including data ownership [1]-[3] and privacy issues [4]-[6]. The rise of apps and the decline of the World Wide Web exacerbate the problem [7], [8]. A major drawback of the app ecosystem where Internet companies control most resources is the lack of data interoperability between platforms [9], making the current app ecosystem de facto app silos. Data interoperability [10], [11] not only means different platforms sharing data between each other, but also addresses the ability of users, platforms and other agencies who create, exchange and consume data to have clear, shared expectations for the usage, context, and meaning of the data [12].

App silos (see Figure 1) depict the ill status quo of different apps practicing their own data standards while leaving users out of the loop. The lack of data interoperability has led to two major issues. First, Internet companies have made the data generated on their platforms as private properties and make profits through which, e.g., by enhancing ads with data-driven micro-targeting means, or directly selling user data. Second, netizen online data are shattered on different platforms with different accounts. The popular narrative of big data is that platforms may learn your preferences to provide better-personalized services. However, the app silo status quo thwarts the ideal, since an app can only learn about a limited part of the user. None of them can see the whole picture, nor could the user themselves, which makes it less possible to maximize the potential value from user data.

Moreover, the lack of data interoperability also exacerbates the unvirtuous deeds and competitions in the internet industry, for big companies will always be in an advantageous, if not hegemonic position in data collections and data trades. Since users are left out of the loop of the sales and future usages of their data, these kinds of data sales and analysis deeds in the capitalism market will ultimately commodify personal information and let corporations prey back on users.

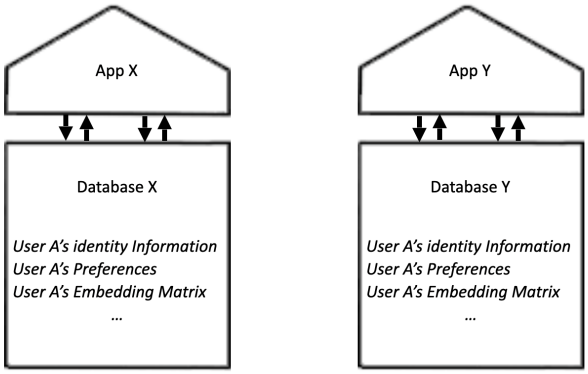

Figure 1. App Silo.

Dissatisfied with the status quo, innovators in the Internet field have been developing new forms of protocols or technologies to increase interoperability between different agents, e.g., the decentralized web movement and linked-data platforms in the Web field [13], [14], and the burgeoning use of Application Program Interface (API) [15] in the Internet industry.

The purpose of this paper is to introduce an ongoing project. Our aim is to protect user privacy, facilitate data management online and enhance data interoperability among web apps. This paper will cover the theories, methods and products of our project.

The paper is structured as follows: Section II describes the three levels of data interoperability with real-world examples. Section III illustrates the architecture and characteristics of our end-user customizable platform design proposed in the Alora project and incorporated in its web based portal app called Alora.

 ISBN: 978-1-61208-796-2

TABLE I. THREE MAJOR ISSUES ON TODAY'S INTERNET

Issue	Description
Online Privacy	Without resorting to third party services, netizens cannot tell if a website is collecting certain cookies or using trackers, nor could netizens disable them.
Data Storage	Online activities only enrich the databases of Internet companies, rather than the power of netizens to control what they want to see. Moreover, the databases of different apps are not interoperable.
Micro-targeting Advertising	Although most of the Internet is free to browse for netizens, yet we are paying with our attention to the free services, which may eventually cost more due to the asymmetric information in the market.

II. Three Levels of Data Interoperability and Existing Examples

The prevailing data ownership issues stem from when Internet companies make it a norm to collect and privatize user data, which has led to the rise of advertising-supported business models. As scrutinized by Wu [16], Internet platforms provide user free diversion in exchange for user attention, which is monetized through advertising, and the efficiency of which is enhanced by micro-targeting and recommendation systems based on the data collected by internet companies. Internet companies play the role of "attention merchants", a term coined by Wu to describe the role played by Internet companies in the advertising model [16]. We have listed three major problems faced by today's Internet users due to the lack of data interoperability in today's Internet platforms, as shown in Table 1.

A. Software Level Interoperability

Software level interoperability refers to the scenario that the owner of the software designs data interoperability protocol and defines how their data will be accessed by outside parties.

The app ecosystem has been constantly making progress towards software level data interoperability. We are seeing an ever-growing API ecosystem. With API, outside developers can access data in the API providers' servers with the provided calls and requests. For example, team collaboration app Slack [17] has integrated more than 2000 frequently used services into their platform, including Office 365®, GitHub, and Google Drive.

Emerging products like Zapier and IFTTT, whose slogan is "get all your apps and devices talking to each other", are de facto a combination of APIs. There are plenty of explorations to improve the interoperability between apps, and API is the best representative of which, yet software level interoperability also has the app itself as the limit–making things beyond the app out of reach.

B. Platform Level Interoperability

Platform level interoperability denotes a parent platform that defines the data interoperability protocol adopted by agencies residing on the platform.

In platform level interoperability, the parent platform promotes platform level interoperability in a top-down manner by enforcing standardized developing languages, data formats, and user interaction components. One recent popular instance is the Mini Program inside a platform. The Chinese social media platform WeChat [18], [19] first published its Mini Program platform in 2017, followed by Baidu and Alipay. The parent platform provides a slew of developing standards and functioning modules, so that all functions are accessible to developers and there is no need to reproduce existing solutions: authentication and authorization, QR code scanning, augmented reality modules, online payment, map and location service, etc. Moreover, data storage is also partially interoperable between Mini Programs and the parent platform. With the help of various interoperable functions, Mini Programs can stay light in size and focus on functions the main platform does not provide. According to the 2019 year-end report, WeChat has more than 3 million Mini Programs on its platform with 330 million daily active users. Moreover, users on average have used 60 different Mini Programs in 2019.

C. Infrastructure Level Interoperability

Infrastructure level interoperability refers to a thorough set of protocols from the front-end interface to back-end computing and storage, which all agents need to have a consensus on the use of the data.

One and probably the only prevailing Internet platform that meets infrastructure level interoperability is the World Wide Web. Backed by W3C standards, the web is an immense platform where netizens read data and write data. However, the interoperability of the web has been declining with the prevalence of platforms with private databases when most of the data collecting actions are underwater.

In this section, we analyze three projects aiming at providing infrastructure level interoperability to internet services: SoLiD, Brave, and Blockstack.

Social Linked Data (SoLiD) [20], [21] is the new project proposed by Sir Tim Berners-Lee, the inventor of the World Wide Web. SoLiD targets at improving data interoperability in the infrastructure level by granting users access to their data storage with the Personal Data Pod, which gives read/write permissions to different apps. Social linked data also guarantees social connections to be a secured property linked with users' WebID [22], [23], which makes authentication and social connection infrastructures on the web. SoLiD aims to rejuvenate the web through solving problems, which we have concluded below: privacy, storage, and advertising.

Brave is an open-sourced browser known for its novel solutions to privacy and advertising issues. Brave aims to build the infrastructure for online content production and consumption business model. As Brave points out on its website, the foremost problem on today's Internet for Brave to solve is: "[a]s a user, access to your web activity and data is sold to the highest bidder. Internet giants grow rich, while publishers go out of business. And the entire system is rife with ad fraud [24]." Brave helps its users to block data-grabbing ads and trackers [25]. Moreover, Brave introduces the Brave Attention Token, a cryptocurrency issued by Brave that allows users to earn frequent flier-like tokens for browsing and support web creators with the tokens. In this way, contents on the Internet will get matched with economic values without resorting to the

 ISBN: 978-1-61208-796-2

advertisement. Brave envisions Brave Attention Token as a new infrastructure for linking valuable content with user attention and money.

Blockstack is a platform with the overarching mission of giving users direct ownership of their internet assets and protecting user privacy [26]. Blockstack is trying to achieve the goal of infrastructure level interoperability for dApps by providing a suite of developer tools and protocols intended to lower the start-up barriers of dApp development [27]. The most important feature is its enforcement of universal login with Blockstack ID. Blockstack keeps a record of user identity on a blockchain database and then asks its user to set up accounts with apps built on Blockstack [28].

III. Architecture and Characteristics of our End-user Customizable Platform–Alora

As discussed, Alora is a web-based platform being developed by our team [29]. Alora addresses the three problems aforementioned, i.e., data privacy, personal storage, and advertising, in a user-friendly way with existing web technologies. Figure 2 is a screenshot of the current version of Alora extension. The code is open-sourced [30].

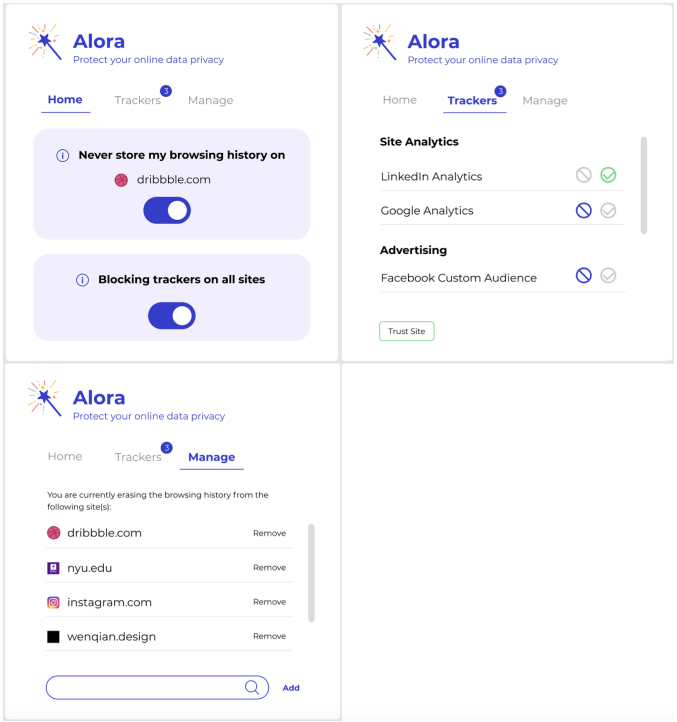

Figure 2. Alora Privacy Extension.

A. Data Privacy

Alora provides a digital footprint management system in the form of a browser extension, as shown in Fig. 3. Based on EFF's privacy badger project [31], the extension can detect third party trackers and automatically block it for the users. Moreover, Alora also automates users' preference on managing privacy-related data on chosen sites including cache, cookies, history, etc. For example, if a user does not want to save history nor cookies on www.amazon.com, the user can switch on the button on Alora, and the extension will delete cookies and history data related to the selected URL automatically. Alora also provides an email for every user to do account management.

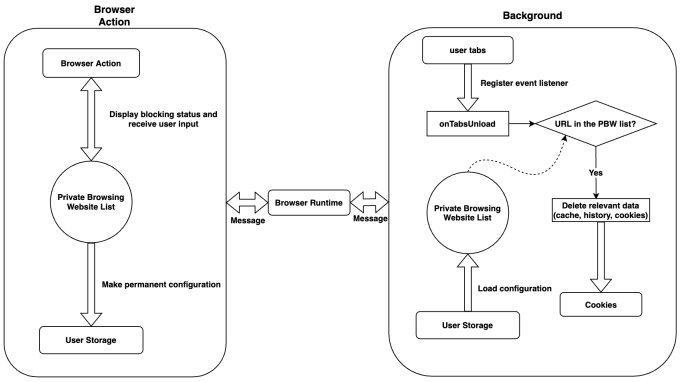

Figure 3. Alora's Customizable Personal Portal User Interface.

B. Personal Data Management.

Alora provides a personal portal page that serves as the homepage of our users, where frequently visited websites will be prioritized into a cluster of lists, as shown in the left column in Figure 4. Furthermore, with the help of APIs of the

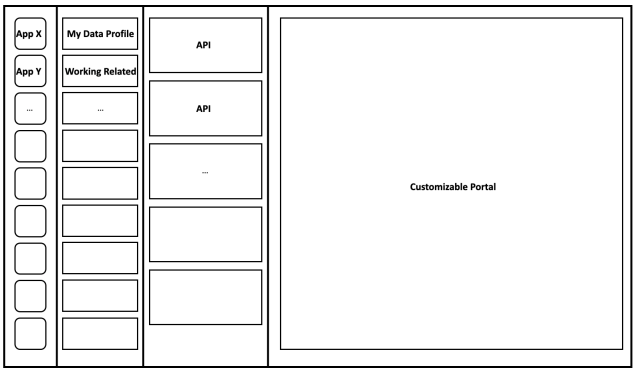

Figure 4. Alora's Automation for User Private Data Management.

frequently used services of the user, users are able to maintain a customizable portal–e.g, email notifications, Twitter feeds, messages, etc. In this way, the features of different services are interoperable with the user's Alora homepage. Moreover, Alora affords users the function to add User Generated Paratext upon resources (see Fig. 4), which means users can save, like, comment, annotate the resources they have browsed online. In this way, user data and metadata are guarded inside the user's personal data zone at the local storage or user's Alora personal cloud, which is analogous to the Personal Data Pod in SoLiD, rather than Internet companies' servers. In future work, Alora also plans to generate user embedding with locally stored user data, or lets users allow third-party services to run federated learning models on certain local user data [32]. The strength of Alora is that the data structure is a thorough one that contains all aspects of the user's online activities, rather than a biased one restricted to a single service. What's more, users have full knowledge and control of the data collecting and management process.

 ISBN: 978-1-61208-796-2

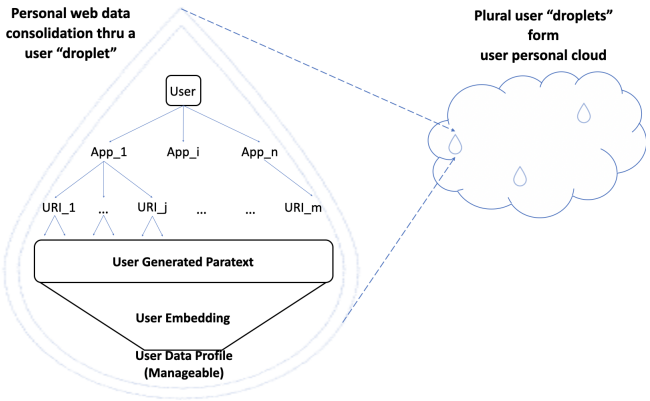

Figure 5. Alora's Customizable Personal Portal User Interface.

As shown in Figure 5, Internet resources generated through Alora include the User Generated Paratext and the User Embedding with a user manageable data profile, which are controlled by the user and accessible by third parties through Alora APIs–which users have the right to decide the read/write accessibility of their resources. Therefore, with the help of Alora, user online data is at the fingertips of users in their personal data storage rather than in custody by cloud service vendors. Such an in-app approach elevates the visibility of privacy control and grants the user convenient control of their personal data management.

C. Data Profile Management (Work in progress)

Based on the user metadata including browsing history, user action data, and the user-generated paratext including likes, comments, and annotations, Alora will be able to train a user embedding, and generate a user-readable and -manageable data profile. The profile will include different tags as in Figure 6. User embedding is a popular method used in today's

TABLE II. USER DATA PROFILE EXAMPLE

Key	Value
Age	18-24 years old
Gender	Male
Interests	Fitness, Adventure Games
Work	Internet Industry

recommendation systems, yet most labels are latent [33]. The goal of Alora's user data profile design above is to make the latent labels transparent to the users, and users will also have the right to manage their data profile and assign accessibilities to their profile and metadata. Furthermore, third party services will no longer need to resort to trackers to collect user data, which is hard to be comprehensive especially for small and medium businesses. We believe the impacts of Alora on the next-generation Internet are huge by looping users back in the game. Data economy and the advertising business model of Internet companies will be fundamentally changed by looping users in the data mining process. The data monopoly by giant Internet companies will get destroyed by granting users the right to maintain a more comprehensive user data profile. Last but not least, the realization of data ownership and willingness to pay for content online will go hand in hand–if users are willing to pay for better data ownership, they will also appreciate the value of information and will be willing to get information as a service.

IV. Conclusion

This paper scrutinized the compelling problems faced by Internet users today: the privatization of user data has led to the ill status quo of today's Internet–lack of data interoperability between platforms and users, myriad data breaches, and various data ownership issues. The quintessential problem is not technical bottlenecks, but a lack of user-oriented infrastructure that protects the data ownership of Internet users, the sources of data that are ironically left out of today's data economy. Alore aims to provide open-sourced, transparent, secure, user-oriented personal cloud computing solutions. By looping users in the data collection and mining process, the latent data collection and data mining process will become transparent to users, resulting in better data interoperability between users and services.

Acknowledgment

This research was partially supported by Futurewei Technologies. We thank our team members from MIT, NYU, CMU who contribute to the Alora project with their expertise, and the global contributors of Alora.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

[2] S. Bertram and C.-P. Georg, "A privacy-preserving system for data ownership using blockchain and distributed databases," arXiv preprint-arXiv:1810.11655, 2018

[3] S. M. Khan and K. W. Hamlen, "Anonymous cloud: A data ownership privacy provider framework in cloud computing," in 2012 IEEE 11thInternational Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2012, pp. 170–176.

[4] S. Zuboff, "Big other: surveillance capitalism and the prospects of an information civilization," Journal of Information Technology, vol. 30,no. 1, 2015, pp. 75–89.

[5] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," inProceedings of the 2011 ACM SIGMOD International Conference onManagement of data, 2011, pp. 193–204.

[6] M. S. Ali, K. Dolui, and F. Antonelli, "Iot data privacy via block chains and ipfs," in Proceedings of the seventh international conference on the internet of things, 2017, pp. 1–7.

[7] Pentina, L. Zhang, H. Bata, and Y. Chen, "Exploring privacy paradox in information-sensitive mobile app adoption: A cross-cultural comparison," Computers in Human Behavior, vol. 65, 2016, pp. 409–419.

[8] V. M. Wottrich, E. A. van Reijmersdal, and E. G. Smit, "The privacy trade-off for mobile app downloads: The roles of app value, intrusive-ness, and privacy concerns," Decision Support Systems, vol. 106, 2018,pp. 44–52.

[9] A. Kadadi, R. Agrawal, C. Nyamful, and R. Atiq, "Challenges of data integration and interoperability in big data," in 2014 IEEE international conference on big data (big data). IEEE, 2014, pp. 38–40.

[10] J. Buck, S. J. Bainbridge, E. F. Burger, A. C. Kraberg, M. Casari,K. S. Casey, L. Darroch, J. D. Rio, K. Metfies, E. Delory et al.,"Ocean data product integration through innovation-the next level of data interoperability," Frontiers in Marine Science, vol. 6, 2019, p. 32.

[11] R. Nawaratne, D. Alahakoon, D. De Silva, P. Chhetri, and N. Chilamkurti, "Self-evolving intelligent algorithms for facilitating data interoperability in iot environments," Future Generation Computer Systems,vol. 86, 2018, pp. 421–432.

[12] "What Is 'Data Interoperability?'." [cited 12 July 2020] The Data Interoperability Standards Consortium , datainteroperability.org/.

 ISBN: 978-1-61208-796-2

[13] N. Vogel, “The great decentralization: How web 3.0 will weaken copyrights,” J. Marshall Rev. Intell. Prop. L., vol. 15, 2015, p. 136.

[14] Zuboff, Shoshana. ”Google as a Fortune Teller: The Secrets of Surveillance Capitalism.” FAZ.NET. [cited 12 July 2020] https://www.faz.net/aktuell/feuilleton/debatten/the-digitaldebate/ shoshana-zuboff-secrets-of-surveillance-capitalism - 14103616.html?printPagedArticle=true.

[15] A. W. Smith, A. J. Moore, D. S. Ebbo, E. B. Christensen, E. B. Olson,F. A. Yeon, J. V. Rajan, K. W. Ballinger, M. Vasandani, M. T. Anderset al., “Application program interface that enables communication for a network software platform,” Oct. 3 2006, US Patent 7,117,504.

[16] T. Wu, The attention merchants: The epic scramble to get inside our heads. Vintage, 2017.

[17] J. Hill, R. LaFollette, R. Grosso, D. Axelson, K. Hart, and E. McDonough, “Using slack to facilitate virtual small groups for individualized interactive instruction,” AEM education and training, vol. 3, no. 1,2019, pp. 92–95.

[18] Lin, J. Qiu, and P. Chen, “Exploration and practice on intelligent teaching patterns based on wechat mini program,” in Proceedings of the 2020 9th International Conference on Educational and InformationTechnology, 2020, pp. 153–157.

[19] L. Hao, F. Wan, N. Ma, and Y. Wang, “Analysis of the development of wechat mini program,” in J. Phys.: Conf. Ser, vol. 1087, 2018, p.062040..

[20] A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem,D. Zagidulin, A. Aboulnaga, and T. Berners-Lee, “Solid: a platform for decentralized social applications based on linked data,” TechnicalReport, MIT CSAIL Qatar Computing Research Institute, Tech. Rep.,2016.

[21] J. Werbrouck, P. Pauwels, J. Beetz, and L. van Berlo, “Towards a decentralised common data environment using linked building data andthe solid ecosystem,” in 36th CIB W78 2019 Conference, 2019, pp.113–123..

[22] G. Huang and K. Mak, “Webid: A web-based framework to support early supplier involvement in new product development,” Robotics andComputer-Integrated Manufacturing, vol. 16, no. 2-3, 2000, pp. 169–179.

[23] P. Mainini and A. Laube-Rosenpflanzer, “Access control in linked data using webid,” arXiv preprint arXiv:1611.03019, 2016.

[24] Secure, Fast Private Web Browser with Adblocker — Brave Browser [Internet]. Brave Browser. 2020 [cited 12 July 2020]. Available from: https://brave.com/?ref=soc369.

[25] Tung, Liam. ”Brave defies Google’s moves to cripple ad-blocking with new 69x faster Rust engine”. ZDNet. Retrieved 1 July 2020.

[26] About [Internet]. Blockstack.org. 2020 [cited 12 July 2020]. Available from: https://blockstack.org/about.

[27] Popper N. Tech Thinks It Has a Fix for the Problems It Created: Blockchain [Internet]. Nytimes.com. 2020 [cited 12 July 2020]. Available from: https://www.nytimes.com/2018/04/01/technology/blockch ain-uses.html

[28] Buchko S. What Is Blockstack (STX)? — The First SEC-Qualified Token Offering [Internet]. CoinCentral. 2020 [cited 12 July 2020]. Available from: https://coincentral.com/blockstack-stx/

[29] The Privoce Project [Internet]. Privoce.com. 2020 [cited 12 July 2020]. Available from: http://privoce.com.

[30] The Privoce Project [Internet]. . 2020 [cited 7 Oct 2020]. Available from: https://github.com/privoce

[31] Privacy Badger [Internet]. Electronic Frontier Foundation. 2020 [cited 12 July 2020]. Available from: https://privacybadger.org

[32] D. Zhang, J. Yin, X. Zhu, and C. Zhang, “User profile preserving social network embedding,” in IJCAI International Joint Conference on Artificial Intelligence, 2017.

[33] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, “Neural collaborative filtering,” in Proceedings of the 26th international conference on world wide web, 2017, pp. 173–182.

 ISBN: 978-1-61208-796-2