



| 2021

ENHANCING MEDIA THROUGH THE DEVELOPMENT OF A PUBLIC SERVICE DATA ECOSYSTEM

E. Sharp¹, H. Ricklefs¹, M. Leonard², R. Jones¹, A. Greenham¹, J. Carter¹, T. Broom¹, K. Bird¹, B. Thompson¹

¹BBC Research & Development, UK and ²BBC now Inrupt, UK

ABSTRACT

We are exploring how the BBC can develop services underpinned by trusted management of data in line with the BBC's values. These services are delivered within an emerging model we have named the 'public service data ecosystem' (PSDE), a set of components which work together to provide a secure and effective platform for public service applications, and which are able to integrate personal data with open data, aggregate data, and data from sources such as IoT devices.

In this paper we introduce the PSDE and describe a working end-to-end demonstrator using personal data stores that delivers enhanced media recommendations. We outline and discuss challenges spanning legal, information security, operational privacy, research and interface design. We end by discussing the implications of our work and its relevance for public service and commercial use and outlining next steps for our research.

INTRODUCTION

The successful implementation of the General Data Protection Regulation (GDPR) within the EU, and its positive influence elsewhere, has encouraged governments to explore further regulation to balance privacy and control with public and private sector needs to access and use data to deliver services. At the same time, ideas for new forms of data governance and stewardship are on the rise, suggesting alternatives to big tech monopolies (1). These models, while different in their motivations and manifestations, seek to address power imbalances in the global data economy, including the limited visibility and control people currently have over how personal data is processed.

BBC R&D conducts research and development to "promote technological innovation and maintain a leading role in research and development" (2) and an increasingly important aspect of our work involves looking at the ways data is acquired, stored, and used by public service media organisations. This task requires attending to complex questions of how we balance the increasing need to leverage audience data to deliver our services, while protecting the privacy of audience members and respecting their personal agency. A key tenet of this work is determining how data can be used responsibly and ethically to deliver added value to users and wider society.

As part of a broad research programme into New Forms of Value (3) we have been exploring alternative approaches to data management aimed at promoting privacy and user oversight, control and agency with regards to use of data about identifiable individuals' 'personal data'.



We have developed an early version of a framework that can bring together data from different sources in a controlled way and allow applications to use the results of processing this data without compromising user control. Whilst nascent, we believe this platform and data sharing approach could provide the foundation for a new generation of services.

We characterise such a data ecosystem as a **public service data ecosystem (PSDE)** - a well-defined collection of infrastructure, analytics, and applications used to acquire, store, analyse and operationalise data that delivers public service outcomes.

DESCRIBING A PUBLIC SERVICE DATA ECOSYSTEM

The framework we have developed incorporates a number of components that work together to provide a secure and effective platform for public service applications, mixing personal data with open data, aggregate data, and data from sources such as IoT devices in a secure way. It also provides for interoperability between service providers, in line with the BBC's obligations around fair trading and market impact.

The Public Service Data Ecosystem

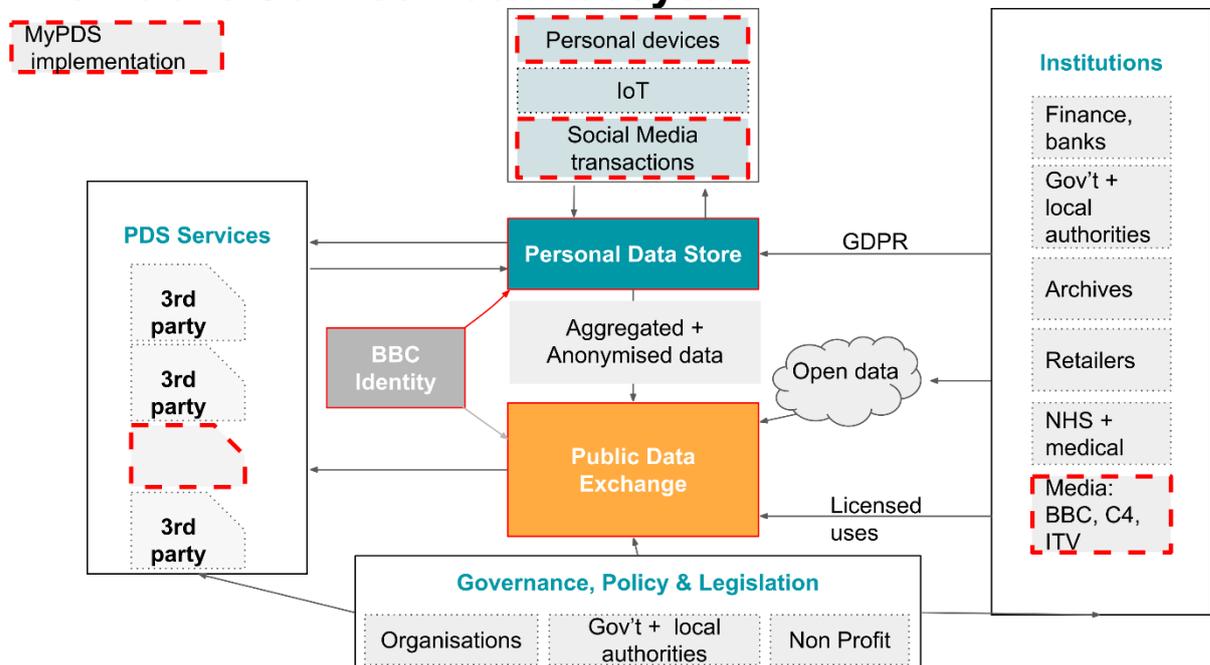


Figure 1: The Public Service Data Ecosystem

The PSDE (Figure 1) incorporates:

1. A place to securely store data: a secure repository for personal data and other data that individuals want to keep control over that afford the user greater oversight, access and control over data about them.
2. A data exchange: a service that offers access to aggregate, open, and licensed non-personal data by enabling data transfer from third party services by exposing a unified interface to data.
3. A set of services for individuals, groups and entities that use the stored data to deliver public value.



Data comes from a variety of sources, under a range of licenses including open data, and may be either stored or referenced for use by services.

A Place to Securely Store Data

Like any online service provider, public service media providers store and use data relating to the ways people engage with their content. The usual way to do this is in a central database or user activity store, often associated with user accounts created for the purpose. The secure storage and effective use of personal data is a challenge for any organisation and most seek to resolve it by storing data securely in repositories that are under their control and implementing policies that can reassure users and regulators.

In order to provide additional protections and enhanced user control, a public service data ecosystem takes a different approach, using a private data store to provide a user-centered and application-decoupled approach to the storage and access of various types of data, including personal data - the repository is not just a store for legally-defined 'personal data', but for any data that the user might want to control.

Each user has their own repository, which may be hosted in a location of their choice or as specified by the provider (including on a device like a mobile phone) and computation is brought to the data. This lets the user monitor, mediate and control the data that flows in and out of the device and the processing that occurs on the data held there. Third parties cannot copy or perform any processing of the data without the user's agreement.

The dominant set of technologies for providing these repositories are called personal data stores or personal data services (PDS). There are many players in the PDS landscape, the business models are still emerging, and they vary in how user data is stored and processed, but they each offer data storage and a range of tools to facilitate the integration of third-party apps. They include Databox, Mydex, HAT, Digime, Solid, Blockstack and others¹.

Data Exchange: access to aggregate, open and licensed data

Data has value when it can be combined with other data which can deliver insights and services that can deliver individual, societal, and economic value (4). The PSDE requires that this is done in a secure, carefully managed and responsible way, so in addition to a secure place to store private data including personal data or data from internet-connected devices like thermostats, doorbells or smart domestic appliances, it needs to provide services with a usable interface to other data sources.

At its simplest the data exchange requires a common interface to data from a range of sources, including the private data stored in a PDS, data from organisations participating in the PSDE and third-party data such as licensed or open data, providing for easy mixing of data from internal and 3rd party APIs and data sources. This allows services to be built using any available technology that can access the published APIs.

Following earlier work developing the Research and Education Space (RES) (5) we believe that there is an additional need for a public data exchange that aggregates this data and makes it available through a single API. The RES platform offered an open-source software stack which finds, indexes and organises rich structured data about a set of archive collections published as linked open data. If the archive assets themselves are available in

¹ Databox <https://www.imperial.ac.uk/systems-algorithms-design-lab/research/databox-project/>; Mydex <https://mydex.org/>; HAT <https://www.hubofallthings.com/>; Digime <https://digi.me/>; Solid <https://solidproject.org/>; Blockstack <https://www.coindesk.com/company/blockstack>



digital form, that data included the information on how to access them, all in a consistent, machine-readable form. While RES established the value of a data aggregator, it did not attempt to make use of personal data. We believe bringing the data exchange together with the personal data store is an important new development.

Digital Public Services

The final key component of the PSDE is a set of services able to access the data exchange and use the data in the data stores. These digital public services provide the main user-facing element of the ecosystem and deliver public value while offering secure and trusted management of the data on which they rely.

DOING DATA BETTER

Within BBC R&D we have been exploring how public service organisations can use personal data for several years. Our first prototype of this kind was built on the Databox platform (6)

Databox was an open-source research project into building secure data stores, funded by the UK Engineering and Physical Sciences Research Council. The core principle behind Databox was that data must not only be stored but also processed in user space. This is to help protect privacy as once data leaves a personal data store and is held by a third party then guarantees of privacy cannot be sustained, since services could copy the data and store it elsewhere.

Databox resolved this by developing a complete operating environment within which user processes and data were co-located and by implementing a strong security policy controlling ingest and export of data. Our BBC Databox-based prototype, the BBC Box (Figure 2), comprised a Databox installation on a Raspberry Pi computer and an associated web app. It implemented processes to ingest user data from BBC iPlayer and Spotify, generate a user profile, and export that profile in a controlled way to selected applications – in our case, a simple application that could use the profile to select from an extended inventory of BBC programmes and local events and play content. It allowed us to do live demonstrations but was a standalone technical demonstrator rather than a fully functioning system (7).



Figure 2: the BBC Box and associated app



Demonstrating the Public Service Data Ecosystem

Building on our experience with Databox, and seeking to explore the characteristics and principles of a PSDE and the potential role the BBC might play in shaping new and emerging data ecosystems, we have developed a fuller implementation of the overall ecosystem consisting of:

- **A web app ('My PDS')** that uses data from Spotify, Netflix and BBC accounts that has been ingested into a personal data store to create a media profile which can be exported and used by other services.
- **An enhanced version of the BBC Sounds audio service application** that combines insights from the media profile with publicly accessible listings data from other sources.
- **Speculative prototypes of a wider PDS proposition** - clickable UX screens to explore how a user could collate, see, and explore their data within My PDS, including health, finance and social profiles, providing a way to organise data and a central dashboard which collates and visualises insights.
- **Speculative PDS-compatible service propositions:** Visual mock-ups of a wider suite of PDS compatible service propositions.
- **Supporting materials** - including explainer videos, privacy policies and service FAQs.

We developed the working PDS web app and media service around the Solid (derived from Social Linked Data) platform (8). Solid is a set of proposed W3C standards that originate from work by web inventor Tim Berners Lee and stores user data in linked data format (RDF) in a secure personal web server called a 'pod'.

Our demonstrator provides each user with a Solid pod and allows a user to import their media consumption data from a variety of sources. Using this data, they are then able to see their combined media viewing history, as well as receive recommendations for BBC content based on their data from other services.

The working system shows how secure data held privately in a PDS and aggregate, non-personal public data can be used to deliver existing and new media services as part of a working data ecosystem.

Capabilities of the Demonstrator

The My PDS service and platform lets a user register, create a Solid pod, authorise the transfer of data from BBC and Spotify and generate a media profile. The user can inspect their media history, view the profile generated from that history, edit it and authorise it for use (or not as the case may be) by third party apps.

The app gives the user greater visibility of their media data, new functionality to directly edit data in their media profile stored on the PDS, and the ability to authorise, uses, offering greater transparency and oversight over how personal data is used to generate algorithmically determined recommendations.

The user can authorise the export of their profile to a web app that delivers some of the functionality of the BBC Sounds audio app and uses the profile to select material from the BBC audio library and the open source Skiddle events listing service. As a result, the user could be recommended content or events based on an analysis both of their Spotify listening and their BBC activity. For example, if a user spent a lot of time listening to Jehnny Beth on Spotify, then the BBC Sounds app would surface interviews with the artist or similar music.



2021

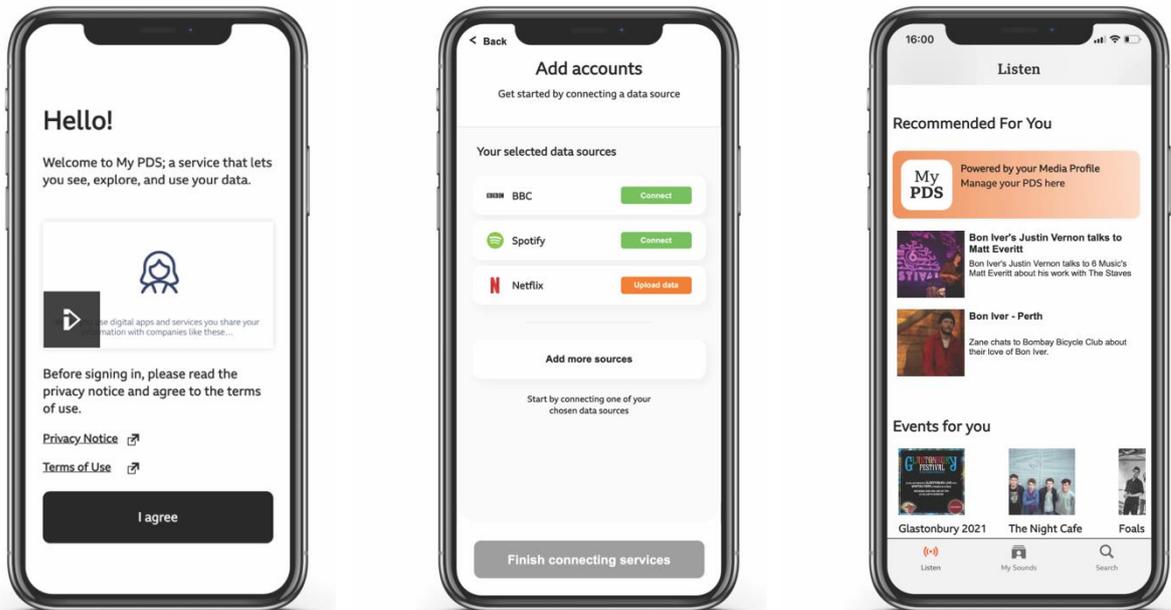


Figure 3: Key screens from the My PDS web app and enhanced version of the BBC Sounds app – combining insights from the media profile with publicly accessible listings data



Figure 4: Clickable UX speculative prototypes to explore what a wider PDS proposition might look like, including health, and social profiles and a central dashboard which collates insights

For the speculative apps we specifically targeted categories/areas that were currently underrepresented within the BBC's product landscape but that initial research had indicated would resonate strongly with our target audience. These were **health & wellbeing**, **everyday money matters**, **the environment**, and **local community and connection**. It should be noted that our research took place whilst many in the UK were coping with huge changes in their lives as a result of the Covid-19 pandemic, juggling isolation, unemployment, childcare



and a global crisis, therefore some of our services touched on activities that worked within the evolving restrictions, for example being out in nature, discovery, and enjoying food.

CHALLENGES

The demonstrator was used to support a closed user trial at the start of 2021 that provided a unique opportunity to explore a range of questions and challenges around the broader notion of a PSDE. Although carried out remotely because of the Covid-19 pandemic and subsequent public health measures in the UK, surveys, evaluation, and an online user community allowed us to explore attitudes to both personal data stores and the provision of PDS-based services by the BBC (results of the research will be published separately).

In order to reach the point where we could undertake this research we had to negotiate many barriers within the BBC in addition to tackling the technical challenge of building a stable research environment and a set of services on top of the Solid/Inrupt server, as detailed in Ricklefs et al (10). These additional challenges required engagement across design, engineering, information security, legal and regulatory teams from the outset, as legal and security requirements shaped technical choices and design decisions as much as the technology posed new legal and security challenges.

While there were specific issues related to the BBC's policies and the legal framework under which we operate, we believe these challenges will be relevant to any organisation seeking to develop or participate in a public service data ecosystem.

Establishing Legal Grounds and Terms of Use for User Data

We needed to establish the lawful basis for data processing for the PDS prototype service. Solid was the underlying PDS storage solution, but the MyPDS prototype was provided by the BBC, and PDS are emerging technologies that are maturing against the backdrop of different and changing regulatory contexts. We therefore had to feel confident both about the BBC's position as a controller/processor according to GDPR and any liability when holding non-BBC data from audiences, even though such data was not 'visible' to us as providers of the Solid pods being used.

Under the UK General Data Protection Regulation (9), tailored by the Data Protection Act 2018 there are different legal grounds for processing data, such as consent - typically for when people give consent for an organisation to process, and contract - where a contract is the primary basis for processing and public task.

Typically, the BBC like other organisations relies on users providing consent for organisations to collect and process personal data but when using a PDS the BBC does not store or collect the data but processes it locally at the user's request.

This led us to the 'public task' provision, where 'processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller' as an alternative to both consent and contract. This was felt to align well with the BBC's remit and the motivations of the project as the BBC is mandated under its Charter to take creative risks with technology and show the most distinctive and high-quality output and services.

Historically this provision has been focused on activities related to TV licensing, the BBC's main source of funding. While it offers an alternative to legitimate interest as a justification for data processing in this case it would not apply to a 'production' service and so as we look to scale provision of PDS services we will be looking for regulatory clarity.



The complexity of the internal debate is a clear indication that the current legal framework is not equipped to deal adequately with a situation within which data subjects are provided with secure data storage facilities by organisations which have no direct access to the personal data being stored but still use that data in providing services. The EU Digital Services Act and Digital Markets Act (11) offer one approach to resolving the issue, and others may emerge.

Informed Consent to Research Participation

As well as establishing a legal basis for processing data, we had to obtain consent for the research trial itself. It was important to make sure trial participants understood what the My PDS service was and how it managed their data differently to existing services. Specifically, we needed to make clear that the BBC could not see, access, or use the data. The language used during the signup process was carefully simplified, with help from a professional copywriter, to be easy to understand and accessible, and included a dedicated permission screen designed to communicate key points about the service to the user in a straightforward way.

UK Data Protection Law requires data protection by design and by default. This means integrating data protection and privacy into processing activities and organisational practices, from the design stage right through the lifecycle. In a decentralised data ecosystem, data is stored and processed in a local account and the user grants access to trusted parties. This creates challenges around provision for informed consent, as designing and communicating permission models around data connections is significantly more complex than allowing a company to store data in a centralised database.

Information about how the service worked was not frontloaded but distributed over different stages of signup, using icons, diagrams, and easily noticeable text pointers, and emphasised at each stage of the user journey. It was also made clear that access could be revoked (12).

Data Standardisation

A key benefit of the data ecosystem is support for interoperability across various data sources and data domains, to allow easier access to digitised collections that can be accessed by services developed by a range of actors, and to do this in a way that users can trust. My PDS and the recommender demonstrate this first part by showcasing an application that uses media data stored on a user's pod to drive a recommender service. For that data can be used across several different applications in needs to be in a format that allows machine-to-machine interoperability. The needs of the ecosystem will therefore drive organisations towards sharing their data formats to support interoperability, bringing us closer towards the model of the Semantic Web, described in Ricklefs et al (10).

Data Requests

Naturally there was a good deal of 'heavy lifting' involved for the user practically setting up their PDS, particularly as there is yet to be a clear standard for people to access their data from the services they are using. We talked a lot about friction and maintaining the need to ensure that the effort of onboarding and set-up didn't outweigh the value of the offering. When using a PDS data is retrieved on a service by service basis so if a user wanted to retrieve their data from 10 different applications, they have to go through the process each time separately for each service.



There are currently three ways for a UK user to retrieve their data: an instant and subsequently continuous connection via an API; an online form to download a snapshot of your existing data; and a manual request for your data (via email) which requires an organisation to respond within 30 days. Although UK and EU data protection law provide for user access to personal data, as it stands organisations are only obliged to fulfil the last of these, and the datasets returned have little commonality.

Ultimately this upfront effort will need to be justified by the value users get from the services, although as more data sources become available, we anticipate that onboarding and data acquisition will become more streamlined and standardised.

Data Management Practices

As currently implemented the PSDE does not imply that participating organisations stop hosting audience data in their own systems, just that this data is also made easily available – as required in the EU by GDPR – to a user-controlled personal data store or pod, where it can be analysed in combination with data from many other sources in ways that, as we have shown, have potential to provide benefits to both organisations and audiences, from privacy preserving access to more complete and potentially accurate audience data stemming from their ability to access, see, explore, edit, interact and authorise access to it in new ways. This could support a wide range of digital services capable of using a variety of data types to deliver public service outcomes.

This has implications for internal data management standards. Currently, media organisations like the BBC log a range of audience interactions, and the resulting data is held in multiple locations. This creates technical complexity but also adds to the cost of meeting subject access requests, data deletion and service design. The PSDE could provide a user centric approach to data storage and processing, where all audience data is held in a PDS. With a PDS at the centre of the system, processing of user data happens just in time, with no need for separate storage.

NEXT STEPS AND FURTHER WORK

Personal data stores are often described as privacy enhancing technologies that can promote greater user control, but standards and best practice in this area have yet to emerge. We see two main findings from our work:

A personal data store offers a basis for constructing a privacy-preserving cross-media recommender that potentially has access to more holistic and representative user insights, which in principle, can enhance recommendations while retaining the user's oversight, agency and control (10)

The demonstrator shows the potential for a more trustworthy *data ecosystem* within which assurances of user control of data can be made more confidently, allowing organisations to make stronger and more credible privacy promises. This may be important in terms of audience confidence and in terms of regulatory compliance, particularly given moves to strengthen the legal controls over data processing such as the EU Digital Services Act (11) and moves in other jurisdictions (13).

An obvious next step would be to store audience data exclusively in the pod. Doing this requires further research/innovation/revision in the way audience data analytics are gathered. Instead of having a user activity store, this data would be stored across a large



number of pods and would have to be processed without breaching the security protocols in place. This would require further development of approaches such as federated learning (14) to train machine models that could model audience behaviour or use of differential privacy techniques (15) to allow data analysis to be done without exposing the data of identifiable individuals.

Creating an ecosystem where the treatment of personal user data is improved in line with new principles and standards, and where audiences can choose to use services knowing the conditions around their data use, requires collaboration between many organisations. This is to ensure that common technical standards are agreed upon to allow data interoperability, and to provide the range of data sources needed to support a full range of services.

To support a PSDE, partnership models will be needed for sharing data between current data holders in the form of agreements that may come in different forms and governance principles around the use of shared data, including the protection of individual's privacy in such use. Such agreements may specify for instance that data is shared only for public services not commercial (potentially a condition to encourage some currently reluctant data holders to share). Or that commercial use needs approval of an independent Data Trust, or equivalent.

CONCLUSION

The work described in this paper was motivated by the observation that data is not being done well insofar as industry standard practices do not reflect the public service values that underpin the BBC's journalism and content commissioning, and that this disconnect was becoming increasingly problematic as the BBC and other PSM sought to make more use of data – specifically audience data – to shape their offers.

We envision a public service data ecosystem as a way of modelling an environment for data services that has the potential to improve access to data to support the delivery of digital public services while building and maintaining trust with our audiences, incorporating new types of functionality and controls to support transparency around data use and active choice. We have successfully implemented an end-to-end demonstration of some elements of an ecosystem, with a clear model for how we could extend this. We have trialled the prototypes with users and shared our work internally in the BBC and with partners, prompting an ongoing debate on the value of these alternative models.

The service and platform we have created are at a very early stage and we faced complex challenges along the way, but we believe that, taken together, this work offers a valuable real-world example that other media organisations can reference as they too look to determine how to make use of personal data to enhance services while addressing changing regulatory requirements and audience expectations.

We believe that interest in PDS and alternative models for data stewardship will grow as more and more organisations begin to develop services around them, and that legislation which puts the user at the centre of their data will be a fundamental driver towards their adoption. The BBC has an opportunity to shape public service data in the manner that it has shaped public service media. The first generation of tools and models have been developed – the BBC and partner organisations could provoke a step change in their deployment, effectiveness, and trustworthiness.



Questions about what healthy, safe and publicly orientated digital services look like for a 21st century digital society are getting more attention. While the BBC will continue to spend most of its income and effort making television and radio programmes and distributing them as effectively as possible to UK audiences, digital online tools and services create additional spaces for engagement that present new opportunities to the BBC to deliver its public purposes and its mission to inform, educate, and entertain. Such services will necessarily involve audience data.

We also believe that a public service data ecosystem is an important element of the digital public space (16,17), an online zone for engagement and participation within which public service organisations can engage with audiences by, for example, providing services, making assets and archive material more freely available, and encouraging creative engagement and participation (18,19). A public service data ecosystem could enable connections between data of many different types and from a variety of sources within a digital public space, support the delivery of public service outcomes, and be of value to individuals, society and the economy.

REFERENCES

1. Open Data Institute, 2020. *Applying new models of data stewardship to health and care data*. http://theodi.org/wp-content/uploads/2020/08/OPEN_ODI_HF_Applying-new-models-of-data-stewardship-to-health-and-care-data-2020-07-16.pdf
2. BBC, 2017. *BBC Royal Charter and Agreement*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/577829/57964_CM_9365_Charter_Accessible.pdf (Accessed 13 May 2021).
3. BBC 2019. *New Forms of Value: A BBC for the Data Economy*. <https://www.bbc.co.uk/rd/projects/new-forms-value-bbc-data-economy> (accessed 13 May 2021)
4. Coyle, D., Diepeveen, S., Wdownin, J., Tennison, J. & Kay, L., 2020. *The Value of Data Summary Report 2020*, Bennett Institute for Public Policy and the Open Data Institute. Available at <https://www.bennettinstitute.cam.ac.uk/publications/value-data-summary-report/> (Accessed May 13 2021)
5. Leeming, R, 2016. *Connecting the past to inspire a future of learning*. BBC Internet Blog <https://www.bbc.co.uk/blogs/aboutthebbc/entries/55ac78ff-d5ec-438d-a0f6-2b6f2268fcdf> (accessed 13 May 2021)
6. Imperial College, London, 2016. *Databox Project*. <https://www.imperial.ac.uk/systems-algorithms-design-lab/research/databox-project/> (accessed 13 May 2021)
7. Thompson, B., Jones, R, 2019. *Introducing the BBC Box* <https://www.bbc.co.uk/rd/blog/2019-06-bbc-box-personal-data-privacy> (accessed 13 May 2021)
8. Solid 2019. *The Solid Project* <https://solidproject.org/> (accessed 13 May 2021)
9. Gov.uk. 2015. *Data protection - GOV.UK*. [online] Available at: <https://www.gov.uk/data-protection/the-data-protection-act> (Accessed 13 May 2021)
10. Ricklefs, H., Leonard, M. Loveridge, J., Carter, J., Mackay, K., Allnutt J., Preece, T., Nooney, T., Bennett, K., Cox, J., Greenham, A., Al-Ali Ahmed, T. & Thompson, B.,



- 2021, Stronger Together: Cross Service Media Recommendations. In IBC 2021 (In press)
11. European Commission, 2020. The Digital Services Act Package. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> Accessed 13 May 2021
 12. Sailaja, N. 2021. *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors Computing Systems* May 2021 Article No.: 271 Pages 1–7 <https://doi.org/10.1145/3411763.3451808>
 13. Zhang, T, 2020. *China's Personal Information Protection Law: Compliance Considerations from an IT Perspective* <https://www.china-briefing.com/news/data-privacy-china-personal-information-protection-law-it-compliance-considerations/> Accessed 13 May 2021
 14. Google AI, 2020 Federated Learning <https://federated.withgoogle.com/> Accessed 13 May 2021
 15. Dwork C. and Roth A., 2014. *The Algorithmic Foundations of Differential Privacy. Foundations and Trends Theoretical Computer Science*, vol. 9, nos. 3–4, pp. 211–407
 16. Hemment, D. and Thompson, B. *We are Digital Public Space - Introducing Digital Public Spaces*. 2018 In *Digital Public Spaces* pp3-5 FutureEverything Available at: <https://futureeverything.org/wp-content/uploads/2018/12/DPS.pdf> (Accessed May 13 2021)
 17. McRoberts, M. 2011 *BBC Digital Public Space Project*. BBC Internet Blog. https://www.bbc.co.uk/blogs/bbcinternet/2011/04/bbc_digital_public_space_proj.e.html Retrieved May 13 201
 18. Thompson, B. 2011. *Partnerships and the Digital Public Space*. BBC Internet Blog https://www.bbc.co.uk/blogs/bbcinternet/2011/10/digital_public_space_partnersh.html. retrieved May 13 2021
 19. Kiss, J. 2015 *A Digital Public Space is Britain's missing national institution*. The Guardian. <https://www.theguardian.com/technology/2015/mar/05/digital-public-space-britain-missing-national-institution> retrieved May 13 2021

ACKNOWLEDGEMENTS

We would like to thank Andy Conroy, controller of BBC Research & Development, whose consistent and visionary support for research into data stewardship and new forms of value made this work possible.