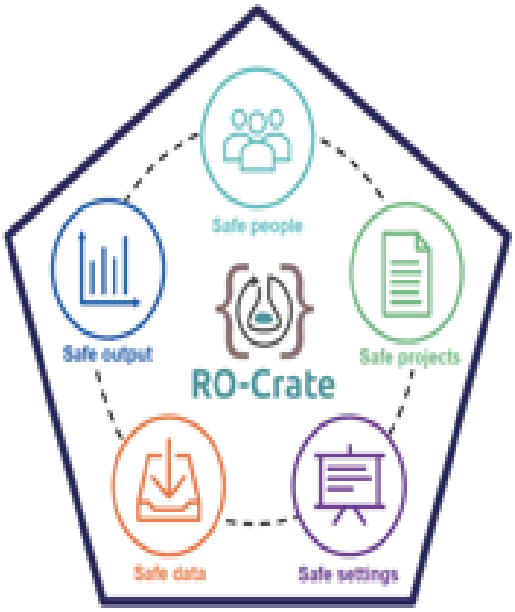
alterline

TREFX Project

Public Consultation Online Focus Groups Report



TRE-FX

DELIVERING A FEDERATED NETWORK OF TRES TO ENABLE SAFE ANALYTICS

July 2023

Compiled by:

Damian Miller,
Research Director
Alterline

DARE UK

Swansea University Prifysgol Abertawe

University of Dundee

MANCHESTER 1824

NHS University Hospitals Birmingham NHS Foundation Trust

University of Liverpool

The University of Nottingham

PIONEER Health Data Research Hub

HDRUK Health Data Research UK

TREFX Online Focus Groups Report

INTRODUCTION

This report represents the full findings from a series of Online focus groups which were conducted to support the TREFX project in engaging members of the public. The research was undertaken by Alterline, a third-party research agency that specialise in Academic and University research. Alterline have worked with the University of Nottingham previously on a TRE4C project as well as student related projects and have built a relationship based upon the quality of the research delivered.

The focus groups were conducted through July 2023 and the findings compiled, analysed and brought together here to highlight the insights that have been surfaced from the research.

METHODOLOGY

To capture the thoughts, feelings and experiences of the members of public, Alterline conducted five initial online focus groups consisting of eight people per group. Participants were citizens of the UK and were a representative mix of ages, genders, backgrounds, education levels and country they live in. An overview of the participants can be found in the appendix at the back of this document.

Each group was taken through the same discussion guide and were moderated by one of Alterline's expert researchers. The discussion guide was created by Alterline and agreed with key stakeholders from the TREFX project group. The guide covered areas to understand the public facing materials to explain the project, thoughts on benefits of the project and concerns arising from their understanding, as well as the role of data privacy and safeguarding in providing confidence in the systems behind the project.

These discussions allowed Alterline to explore how the TREFX concept would be greeted by the public, how it might be perceived and what questions or concerns people had about a project of this potential scale. The discussions allowed Alterline to capture the voice of the citizenry in their own words and allow conversations to develop amongst participants and understand the strength of feeling and to what extent there are any universal or divisive elements.

EXECUTIVE SUMMARY

This report presents the findings from the study that aimed to explore public perspectives related to Federated Analytics. The study examined participants' knowledge of Federated Analytics, their understanding of the public facing material, their thoughts on benefits & concerns of Federated Analytics, the importance of data privacy, control and effective communication.

The participants initially had limited knowledge and awareness of Federated Analytics, but their understanding significantly improved after reviewing the provided documents. The materials effectively explained the concept in simple terms, addressing the participants' initial confusion and highlighting the potential benefits of Federated Analytics.

The findings underscore the potential benefits in advancing research, including research collaboration, integrating diverse data sources, and improving research outcomes. However, concerns were raised about data management issues, differentiating between research and commercial use, and compliance with data privacy regulations.

Participants particularly expressed numerous concerns about data security, potential misuse of personal data, and the lack of control over their information. Trust and accountability were identified as critical factors, with participants emphasising the need for external regulation and accreditation to ensure data security and prevent unauthorised access.

The importance of transparent communication regarding data usage and consent was highlighted, with recommendations for clear and relatable messaging, opt-in and opt-out choices, and the sharing of success stories to engage the public and promote awareness.

The concept of the five safes framework was positively acknowledged, with some recommendations for further clarity and safeguards.

Overall, this report provides valuable insights for the implementation and improvement initiatives to contribute to a comprehensive understanding of Federated Analytics and its role in shaping the future of data-driven healthcare.

HEADLINE FINDINGS SUMMARISED

Limited prior knowledge: Participants had limited knowledge and awareness about Federated Analytics before receiving the documents.

Improved Understanding and Appreciation: The participants' understanding of Federated Analytics significantly improved after reading the provided document. The document effectively explained the concept in simple terms and provided clear explanations with diagrams.

The public facing document is potentially too long for most people to read and understand, and using this version to refine down to the key points to aid understanding and a follow up website or FAQ to deliver more details to those who need it may be a better approach.

Concerns about data security: Participants expressed concerns about data security and potential unauthorised access or breaches.

Trust and reliance: Participants emphasised the importance of trust in the system and reliance on its promised security measures and potential benefits.

Communication and transparency: Effective communication and transparency were highlighted as essential for keeping the public informed about data usage and research progress.

Data privacy and consent: Participants emphasised the need for clear and informed consent, with options for individuals to choose specific research projects to participate in.

Five safes framework: The concept of the five safes framework was discussed, providing reassurances regarding data privacy and security is seen as an important and necessary inclusion, but that more clarity and detail will be needed to reassure participants

Concerns about lack of control: Participants expressed concerns about not having control over their personal data and how companies use it without their knowledge.

SUGGESTED ACTIONS

Based upon the findings from the focus groups there are some things that we would recommend that you consider in your thinking and planning for the future phases of the TREFX project.

Enhance Readability and Accessibility:

Address participant feedback regarding the document's wordiness and length. Consider using bullet points, concise explanations, and more visual elements such as more diagrams and charts to make the content more digestible and engaging. Strive for a balance between providing comprehensive information and ensuring readability for a wider audience.

Provide Clearer Information on Data Usage:

Respond to participant concerns by including more detailed information on how the data will be used, who will have access to it, and how long it will be retained. Clearly explain the measures in place to protect data security and privacy, and explicitly state the commitment to using the data solely for medical and scientific research purposes.

Enhance Transparency and Communication:

Ensure clear and transparent communication about data usage, consent processes, and the purpose, duration, and beneficiaries of data sharing initiatives. Use plain language and accessible formats to effectively communicate with the public.

Strengthen Data Security and Privacy Measures:

Implement robust data security measures to protect against unauthorised access and breaches. Adhere to data protection regulations, such as the General Data Protection Regulation (GDPR) and ensure companies handling data are transparent and fully compliant.

Establish Trust and Accountability:

Build trust by promoting accountability in data handling practices. Implement external regulation and accreditation to ensure data security and prevent misuse. Consider involving independent bodies or an ombudsman to provide oversight and enforce strict penalties for data breaches.

SUGGESTED ACTIONS continued

Empower Individual Control and Consent:

Give individuals greater control over their personal data by providing clear and informed consent options. Implement an opt-in default system with the ability to opt out, allowing individuals to make choices based on their preferences and the specific research projects they want to participate in.

Engage in Public Outreach and Education:

Conduct public outreach campaigns to raise awareness and understanding of federated analytics. Share success stories and case studies to demonstrate the real-world impact of data contributions. Utilise various communication channels, such as text messages, emails, advertising, GP appointments, and targeted outreach efforts to reach a diverse audience.

Develop Clear Governance Frameworks:

Establish clear governance frameworks to ensure fairness, transparency, and accountability in granting access to data. Include experts with relevant knowledge and experience in decision-making processes to ensure ethical considerations are adequately addressed.

Continual Monitoring and Improvement:

Regularly monitor and evaluate the implementation of federated analytics initiatives. Seek feedback from the public and stakeholders to identify areas for improvement. Continuously update and refine processes, security measures, and communication strategies to address emerging challenges and build public trust.

The story behind the findings

RESPONSE TO THE PRE READ MATERIALS

WHAT HAVE YOU LEARNT ABOUT FEDERATED ANALYTICS?

Limited prior knowledge and awareness about Federated Analytics among participants was evident before they received the provided pre read documents. The participants generally lacked familiarity with the topic and did not necessarily perceive it as particularly relevant or interesting before finding out more. 68% said before the session that they had no knowledge of the topic (28% limited understanding understanding). This lack of prior knowledge posed a challenge to the effective dissemination and comprehension of information regarding Federated Analytics, with many saying that they may well not have read it without being prompted for these sessions.

However, the participants' understanding significantly improved after reading the documents and opened them up to what the topic was about and how that was of more interest to them than they had believed. The materials succeeded in explaining the concept of Federated Analytics in relatively simple terms, utilising clear explanations and diagrams to enhance comprehension. This demonstrates the positive impact of well-designed informative documents in bridging the knowledge gap.

> *"I think it was useful. It explained in reasonably simple terms how they intend to set about integrating data together to do research. And rather than pull all the data into one central source, they're going to query the different sources around the country. I think it explained it really well, especially with the little diagrams*

Some participants did experience initial difficulty in understanding the material and they highlighted the need for simpler language and more accessible presentation formats. The complexity of the information presented proved challenging and led to confusion among some participants, who needed to read several times to fully understand the concept.

Suggestions were made to employ more bullet points and concise explanations, allowing for better comprehension and ease of reviewing the details. It is evident that the use of plain language and visually appealing formats could significantly improve the accessibility and understanding of the complex concept of Federated Analytics.

The public facing document did strike some balance in the amount of information presented, effectively addressing both the advantages and disadvantages of Federated Analytics. Participants appreciated the comprehensive coverage of potential concerns, including data ownership and accountability.

Such considerations were surprising to some participants, who had not initially considered some of the concerns until it was discussed in the sessions. This highlights the importance of transparency and openness, thereby fostering a more informed and engaged audience.

The document's readability emerged as an important factor in participants' understanding and engagement. Feedback indicated that the material contained lengthy sentences and paragraphs, which hindered some comprehension and could contribute to information overload. Participants recommended the use of bullet points and visual elements to enhance readability and facilitate absorption of the information.

In conclusion, participants initially had limited to no knowledge about Federated Analytics. However, the provided documents were instrumental in improving their understanding of the topic. The documents effectively explained the concept in simple terms, utilising some clear explanations and diagrams.

Nonetheless, challenges were encountered in terms of comprehension, readability, and information overload. Simplifying language, incorporating more visual aids, and employing concise explanations were recommended strategies for enhancing accessibility and comprehension. Future endeavours in educating individuals about Federated Analytics should emphasise the importance of balancing the amount of information, accessibility, and readability.

WHAT ARE YOUR THOUGHTS ON THE BENEFITS OF FEDERATED ANALYTICS?

The participants in the study acknowledged several benefits associated with Federated Analytics, recognising its potential to enable research collaboration, combine diverse data sources, and enhance the quality of research outcomes. They emphasised the efficiency of analysing data without the need for extensive data transfers, which can save time and resources.

Furthermore, participants highlighted the potential benefits of Federated Analytics in various domains, including healthcare research, pattern identification, focusing on vulnerable groups, and proactive problem-solving. The system was seen as a means to improve communication and access to medical information across departments, facilitating more efficient healthcare services. Additionally, participants recognised the value of readily available data for research purposes, enabling more comprehensive and insightful studies.

The potential benefits of Federated Analytics extended beyond healthcare research and encompassed improved research capabilities, access to a larger pool of information, personalised medicine, and the ability to analyse data from different demographics and regions. Participants acknowledged the usefulness of Federated Analytics, however, some participants expressed the need for more evidence and tangible results before fully embracing the concept, highlighting the importance of empirical validation in gaining trust and confidence.

The participants also recognised the value of a centralised data pool for research purposes. They believed that a comprehensive view of data from various sources would lead to better solutions and advancements in healthcare. Furthermore, accessing and analysing medical data was perceived as beneficial in multiple ways, including facilitating collaboration among researchers, identifying patterns and trends in diseases, and enhancing planning and prevention strategies.

Overall, the participants' acknowledgment of the benefits of Federated Analytics highlighted its potential in driving research, improving healthcare services, and promoting proactive problem-solving. However, these focus groups revealed the importance of further evidence and tangible results to ensure widespread acceptance and utilisation of Federated Analytics, demonstrating the actual impact and effectiveness in order to address any scepticism and promote its broader adoption.

> *"You get a UK wide view, rather than what you've probably got at the moment. But if you could share, then you get a much bigger sample and you can probably understand what's happening across the country a bit better, and then potentially use that to come up with better solutions or better ways of solving the problem*

Federated analytics – why is it needed?

Federated analytics is an innovative approach that enables data analysis and machine learning while respecting privacy and security. They allow organisations and individuals to collectively gain insights and train models without directly sharing their raw data.

The key reasons it is needed are as follows:

- This devolved approach reduces the need for large-scale data transfers, leading to faster and more efficient processing.
- Allows extraction of useful information like trends in a population's health without revealing any individual's specific details.
- Allows analysis on a larger and more diverse dataset, resulting in more accurate and comprehensive insights that represent a broader population or user base.
- Promotes privacy preservation, as data remains under the control of the data owners.
- Enables collaboration across organisations and individuals.
- Supports valuable insights and advancements in healthcare while safeguarding personal privacy.
- Data remains private while contributing to the collective knowledge embedded in the shared machine learning model.
- By enabling collaboration and knowledge sharing, federated analytics supports innovation and discovery.
- Machine learning models can be trained on local devices to capture unique insights.

WHAT ARE YOUR CONCERNS ABOUT FEDERATED ANALYTICS?

Participants in the study raised significant concerns regarding data security in the context of Federated Analytics. It was the main concern that was voiced in every group session. Scepticism stemmed from their belief in previous incidents of data loss or misuse in the healthcare industry, which has eroded trust in data handling practices.

Participants emphasised the need for robust security measures to safeguard against unauthorised access or breaches. Furthermore, concerns were expressed regarding potential misuse of the data for political or revenue-generating purposes, underscoring their view of the importance of restricting data usage exclusively to medical and scientific research.

> *"I'm hoping that the relevant people or organisations that use it have enough. Safety, like firewalls and technology against hacking and things. That wasn't something that necessarily came out in the 5 safes.*

Data management issues within the National Health Service (NHS) were also highlighted by participants. Challenges such as incomplete or inaccessible medical records raised questions about the NHS's ability to effectively manage and utilise data for Federated Analytics. These concerns indicate the need for addressing existing perceptions of data management gaps before implementing a new system.

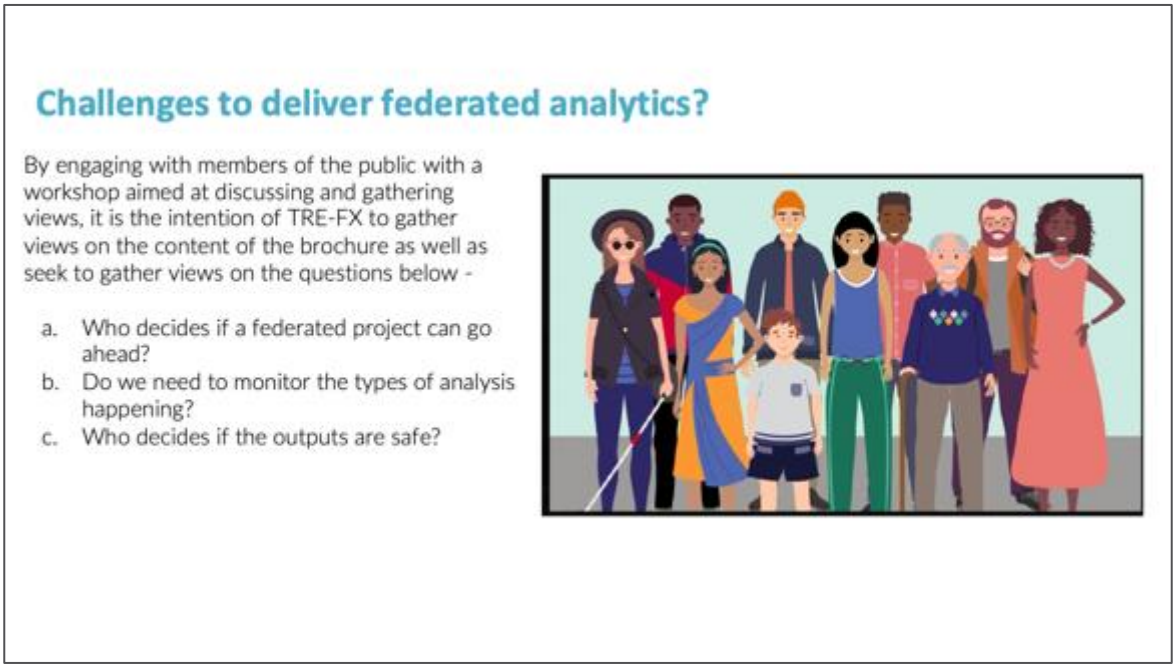
Differentiating between research and commercial use of data emerged as a key consideration. Participants raised concerns about people's willingness to allow their data to be used for commercial purposes, rather than medical research. They stressed the importance of clear guidelines and options for individuals to choose how their data is used, ensuring transparency and respect for individual preferences.

Participants expressed various additional concerns related to data privacy practices and compliance. Consistency in data measurement, differing approaches in different locations, and adherence to privacy regulations were among the concerns raised. Furthermore, questions regarding data ownership, transparency, ethical considerations, and individuals' control over their data were also voiced.

> *"Will an NDA be signed as well, because obviously you don't know the people that are actually getting those access codes, what are they actually going to do with it behind closed doors? Or are they signing something to agree that all the information they see should be kept in that room or that nothing should be taken away?*

Looking toward the future, participants expressed apprehension about the long-term implications of data sharing practices and the potential privatisation of data. They emphasised the need to proactively address data privacy concerns and establish mechanisms to protect individuals' rights and interests.

The security of personal data emerged as a significant concern for participants, given the risks of data breaches and unauthorised access. Participants stressed the importance of implementing strong security measures and stringent controls to prevent misuse of sensitive information.



The complexity and challenges associated with implementing the system were acknowledged by participants. Variations in standards, security protocols, and data quality across different institutions posed significant hurdles. Participants emphasised the need for careful planning, proactive consideration of potential issues, and asking the right questions to ensure the successful implementation of the system and drive meaningful research outcomes.

In conclusion, this study revealed a range of concerns regarding data security, potential misuse, data management, differentiation between research and commercial use, compliance with privacy practices, and the future implications of data sharing practices.

Addressing these concerns through robust security measures, clear guidelines, comprehensive data governance frameworks, and proactive planning could be crucial for successful adoption of Federated Analytics while ensuring privacy and maintaining trust among participants and stakeholders.

WHAT ARE YOUR THOUGHTS ABOUT PRIVATE PERSONAL DATA AND HOW IT IS USED?

Trust and accountability emerged as important considerations in the participants' discussions regarding data handling. Participants stressed the need for external regulation and accreditation to ensure data security and prevent misuse, establishing a system of trust and accountability. The importance of anonymity and trust in the organisations handling the data was emphasised to ensure participants' comfort with their medical data being used.

The participants' perception of data use varied, with some expressing scepticism and concerns about companies profiting from personal data. Trust issues arose, leading to doubts about the motives and potential misuse of data within the system. To address these concerns, participants emphasised the necessity of strong safeguards and ethical considerations to ensure transparency and protect individual rights.

The concept of the five safes framework, a framework for ensuring data privacy and security, was discussed by participants. While some participants found the framework helpful in understanding the security measures in place and appreciated the emphasis on appropriate data use and safeguards, others found it too vague and lacking in some transparency.

Participants suggested improving the clarity and differentiation of certain aspects of the framework to address these concerns.

> *"I like that it appears that each step is a thing, so if that doesn't pass, then you don't get passed on to the next so it's a very good idea. It's sounds very safe and secure.*

> *I feel like it needs to still be looked into, like it should be more than the five safes. The five safes were very woolly and vague. They use lots of soft, cuddly words, like appropriate, trust, minimise, safe.*

In conclusion, trust, accountability, and data privacy were key themes discussed by participants in relation to Federated Analytics. The need for external regulation and accreditation, the importance of anonymity and trust, and concerns about data misuse and profit-making were highlighted.

The participants' perceptions of data use varied, with some expressing scepticism and emphasising the importance of safeguards and ethical considerations. The participants also discussed the five safes framework, appreciating its concept but raising concerns about its clarity and transparency, as well as the need for additional safeguards.

WHAT ARE YOUR THOUGHTS ABOUT CONTROL OR MEASURES FOR DATA PRIVACY

The lack of control over personal data emerged as a significant concern among participants, who expressed apprehension about companies using their data without their knowledge. Trust and reliance on the system were emphasised, with participants drawing comparisons to other online activities and expressing a belief in the greater good and potential benefits of improved treatments.

Acknowledging the likely involvement of external companies or experts in data analysis due to the lack of internal expertise within the NHS, participants emphasised the need to trust both the NHS and the external parties involved.

Data protection regulations, such as the General Data Protection Regulation (GDPR), were discussed, but participants noted that companies often request consent in ways that are not easily understood or are not fully transparent about data usage, raising concerns about privacy. To address these concerns, participants suggested the inclusion of additional reassurances and verification measures to enhance data security and privacy. The idea of allowing individuals to choose what data they share received support, with participants highlighting the importance of consent and the ability to keep certain information confidential.

Trust in the handling and usage of data was a significant concern for participants, who expressed some trust in institutions like the NHS and universities but had reservations about involving commercial entities. They emphasised the need for robust governance, potential involvement of an ombudsman, and strict penalties for data breaches to ensure trustworthiness.

The participants underscored the need for oversight and governance in managing the data, advocating for a governing body or board to ensure fairness, transparency, and accountability in granting access to different parties. They highlighted the importance of involving experts with relevant knowledge and experience in decision-making processes.

Participants recognised the limited control they already have over their personal data in the digital age and questioned the extent of control they would have in a Federated Analytics system. They raised concerns about data collection and usage by various platforms and the potential implications for privacy.

In conclusion, the participants expressed concerns about the lack of control over personal data, emphasised the importance of trust and reliance on the system, and discussed the involvement of external parties, data protection regulations, and the need for additional reassurances.

Trust in institutions and the governance of data were deemed crucial, and participants highlighted the importance of transparency, appropriate messaging, and avoiding actions that undermine trust. The establishment of oversight mechanisms and involving experts in decision-making were seen as essential for ensuring fairness, transparency, and accountability. Concerns about privacy and data control in the digital age were also raised, calling for measures to address these issues in a new system.

> *"Really, we've got no control over what anybody does with our data. This data is leaked all the time. And then, you know, if there's always the same old lines trotted out, which is our mistakes, the mistakes have been made, we'll learn from this and let's move on. And then that's it.*

WHAT ARE YOUR THOUGHTS ABOUT KEEPING THE PUBLIC INFORMED ABOUT DATA USAGE

Transparency and understanding emerged as crucial themes in participants' discussions on data usage and consent. They highlighted the significance of clear and transparent communication, presenting information in an easily understandable manner, including details about the purpose, duration, and beneficiaries of data usage.

To enhance public engagement and awareness, participants emphasised the need for effective communication strategies, keeping the public informed about data usage and research progress. They suggested using success stories and case studies to demonstrate the real-world impact of data contributions.

Participants stressed the importance of clear and concise communication regarding the initiative, providing opt-in and opt-out choices. Informed consent was deemed essential in respecting individual autonomy and ensuring control over personal data. An opt-in default approach, similar to organ donor registration, garnered support from some participants.

The discussion also brought up considerations for vulnerable groups, necessitating special measures such as obtaining parental consent and protecting the rights of vulnerable individuals when engaging in data sharing practices.

> *I think if you did an opt in there probably be a lack of excitement of people doing it to some extent. So, I think in order to get the data, almost having an opt out option would be better. But then obviously, you then do have the complication of how easy is it to opt out and you have a process to manage people opting out, which is going to cost money to do so. There's pros and cons of both sides, but I think you'd get more people involved if it was opt out.*

Recognising the vital role of local General Practitioners (GPs), participants highlighted the value of face-to-face communication and personal interactions in addressing concerns, providing clarity, and building trust. They expressed a preference for NHS endorsement over government involvement, as it could enhance public trust in the initiative, emphasising the importance of independence and credibility.

Clear and personalised communication was stressed by participants, who underscored the need for relatable messaging, including comparisons to the current situation, to effectively convey the benefits and advancements brought about by Federated Analytics.

In conclusion, participants' feedback underscores the importance of transparency and understanding in communicating data usage and consent. Clear and transparent communication, along with the provision of opt-in and opt-out choices, are vital for ensuring individual autonomy and control. Effective communication through success stories and case studies enhances public engagement and awareness. Special considerations for vulnerable groups, the role of GPs, and NHS endorsement further contribute to building trust and fostering a more receptive response to Federated Analytics initiatives.

CONCLUSION

These focus groups examined participants' perspectives on various key themes that emerged from discussions related to Federated Analytics, including limited prior knowledge, understanding of provided documents, concerns, benefits, trust, communication, and data privacy. The findings shed light on important considerations and provide valuable insights for the implementation and improvement of Federated Analytics initiatives.

The insight reveals that participants initially had limited knowledge and awareness about Federated Analytics. However, the provided documents played a significant role in improving their understanding of the topic, explaining the concept in simple terms and utilising clear explanations with diagrams. Despite some initial confusion, participants expressed increased comprehension and appreciation for the potential benefits.

Participants highlighted concerns regarding data security, potential misuse, and the lack of control over personal data. Trust and accountability were deemed crucial, emphasising the need for external regulation and accreditation to ensure data security and prevent unauthorised access. Participants also stressed the importance of transparency and understanding in communication regarding data usage and consent. Clear and relatable messaging, provision of opt-in and opt-out choices, and sharing success stories were recommended to enhance public engagement and awareness.

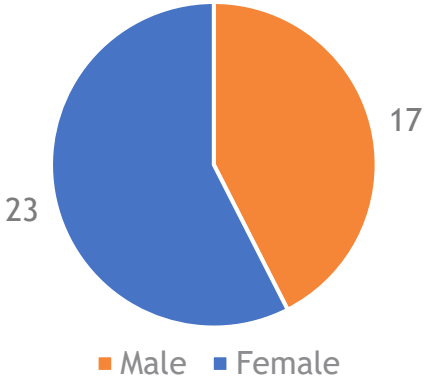
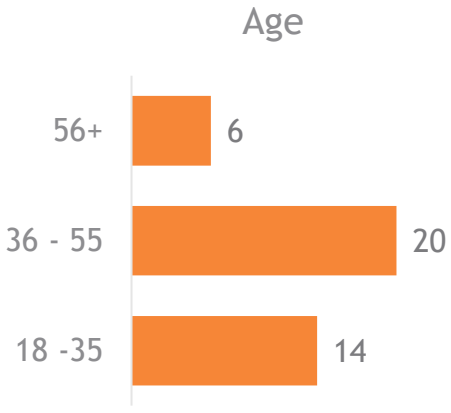
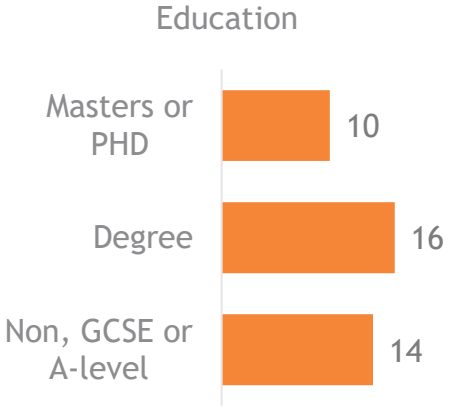
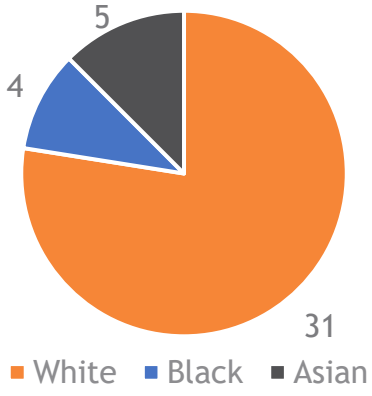
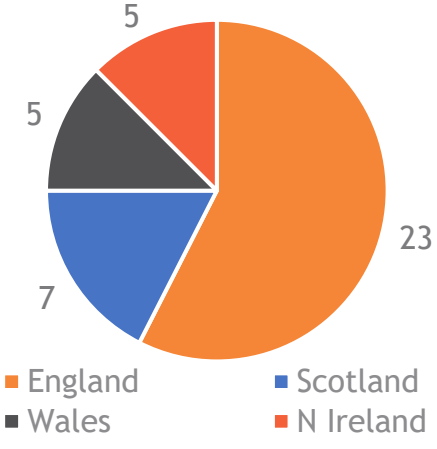
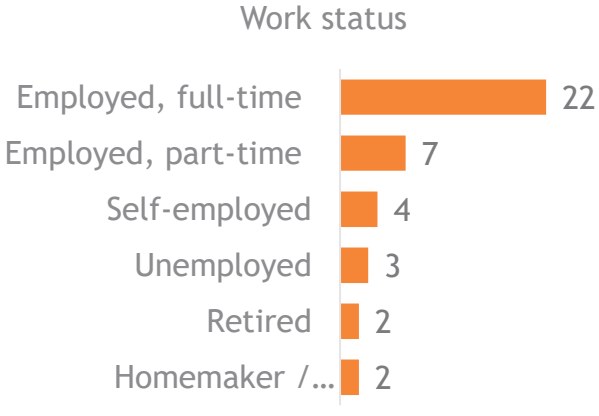
The research highlights that participants understand the potential benefits of Federated Analytics to include, research collaboration, diverse data source integration, and improved research outcomes. Participants recognised the value of a centralised data pool and the potential for enhanced healthcare services. However, concerns were raised about data management issues, differentiating between research and commercial use, and compliance with data privacy regulations.

To address these concerns, participants suggested the inclusion of additional reassurances, verification measures, and stringent governance. They emphasised the importance of involving experts with relevant knowledge, establishing oversight mechanisms, and ensuring fairness, transparency, and accountability. The concept of the five safes framework was acknowledged, with suggestions for further clarity and considerations.

This study highlights the need for clear and transparent communication, trust, accountability, and robust data governance in Federated Analytics initiatives to win over public trust. It underscores the importance of addressing concerns related to data security, privacy, and control, while emphasising the potential benefits of collaboration, improved research outcomes, and enhanced healthcare services.

By implementing the recommendations provided by the participants, Federated Analytics initiatives can build trust, foster public engagement, and ensure responsible data sharing practices. Future research could continue to explore and address the challenges and opportunities associated with Federated Analytics, promoting a comprehensive understanding of its potential in advancing research and improving healthcare outcomes.

Appendix

FOCUS GROUP DEMOGRAPHICS

This report is based upon five focus groups each containing 8 participants. The follow charts are a breakdown of demographic detail about those participants