Privacy Enhancing Technology Project Parties

NIH NATIONAL CANCER INSTITUTE

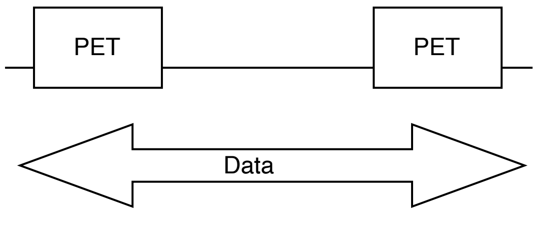
U.S. National Science Foundation

NAACCR North American Association of Central Cancer Registries

OAK RIDGE National Laboratory

NIST National Institute of Standards and Technology U.S. Department of Commerce

eRA



NHS England

NDRS NATIONAL DISEASE REGISTRATION SERVICE

NPIC

Office for National Statistics

Genomics England

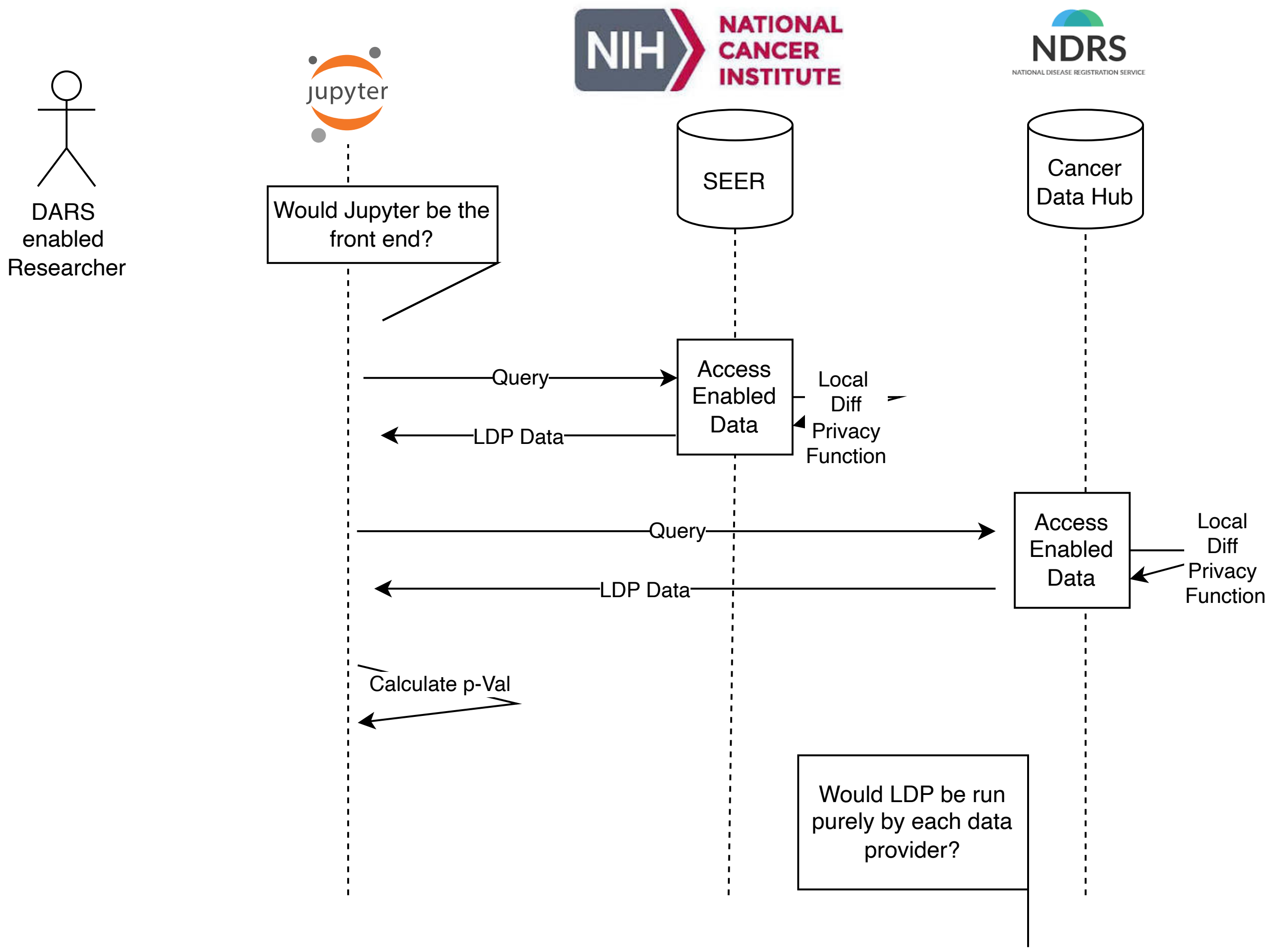
HDRUK Health Data Research UK

	DP Flow	FL Flow	FL-HE Flow	SMPC Flow
Medical Use Cases		**Medical Research**: Institutions can collaborate to train predictive models for disease diagnosis, treatment planning, and drug discovery using patient data without violating privacy regulations. **Personalized Medicine**: Federated Learning allows healthcare providers to personalize treatment plans and interventions based on individual patient data while preserving patient privacy. **Clinical Trials**: Pharmaceutical companies can use Federated Learning to analyze data from multiple clinical trial sites while maintaining data privacy and confidentiality.	Same as FL Flow	
			Highly Sensitive Data: When the data involved is extremely sensitive, such as healthcare records, financial transactions, or personal communications, FL-HE provides an extra layer of privacy protection by ensuring that the data remains encrypted throughout the training process. **Legal and Regulatory Compliance**: In industries or jurisdictions with strict data privacy regulations, FL-HE can help organizations comply with legal requirements while still benefiting from collaborative model training. It allows multiple parties to contribute data without exposing sensitive information to each other or to a central server. **Parties with Differing Trust Levels**: In scenarios where participating parties	

Features			have varying levels of trust or may be competitors, FL-HE mitigates concerns about sharing raw data by allowing each party to encrypt their data before sharing it with others for model training. **Cross-Domain Collaboration**: When collaborating across different domains or organizations with diverse data sources, FL-HE enables joint model training without the need to centralize or exchange raw data, thereby preserving data sovereignty and confidentiality. **Privacy-Preserving Analytics**: In situations where organizations want to perform analytics or derive insights from sensitive data without exposing the data itself, FL-HE enables computations to be performed directly on encrypted data, ensuring privacy while still gaining valuable insights. **Resource-Constrained Environments**: While Homomorphic Encryption can be computationally intensive, FL-HE may still be suitable for resource-constrained environments where the alternative of transferring raw data for centralized processing is not feasible due to bandwidth or storage limitations.	

Differential Privacy Use Case Flow

A permitted Researcher with access to NIH SEER database and a Jupyter Notebook wishes to calculate the p-value between a US rare cancer group and a UK rare cancer group.



Assumptions:

1. Users will have passed Data Access Request permission flows (on a per user basis) with each data provider

2. Users will conform to a 5 Safes style framework

3. ε-differential privacy levels will be set by each data provider

4. SEER will be the data provider from the US

5. NDRS / NHS England will be the data provider from the UK

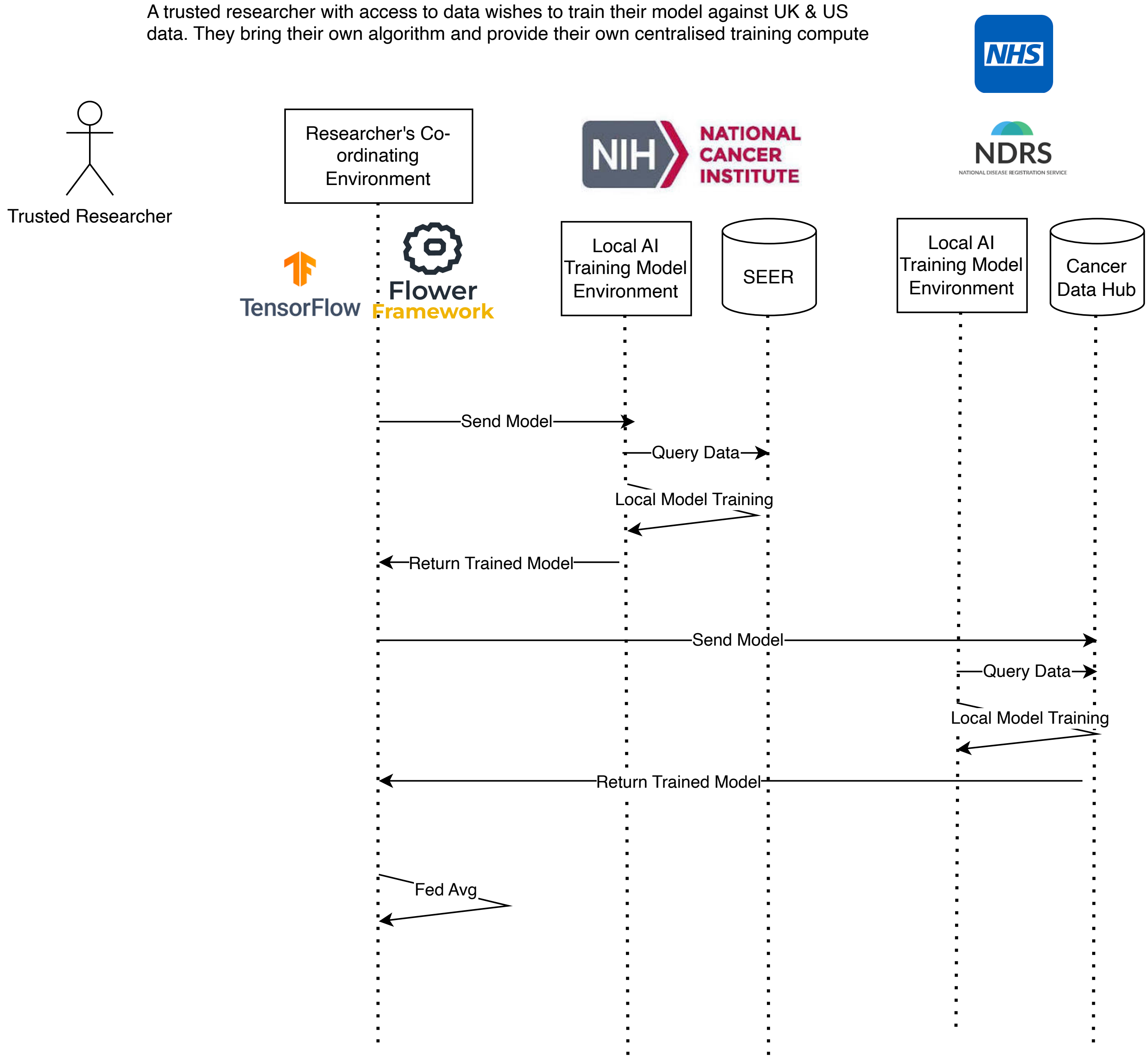
6. Each data provider will implement noise on their data

7. The researcher will use a Python notebook style tool for the front end, which will make remote queries via API?

8. There will be a controlled bastion server on each data provider that the client accesses?

Federated Learning Use Case Flow (Bring your own model + orchestrator)

A trusted researcher with access to data wishes to train their model against UK & US data. They bring their own algorithm and provide their own centralised training compute



Assumptions:

1. Tensor Flow Federated or Flower framework could both be used as orchestrators of a federated learning use case

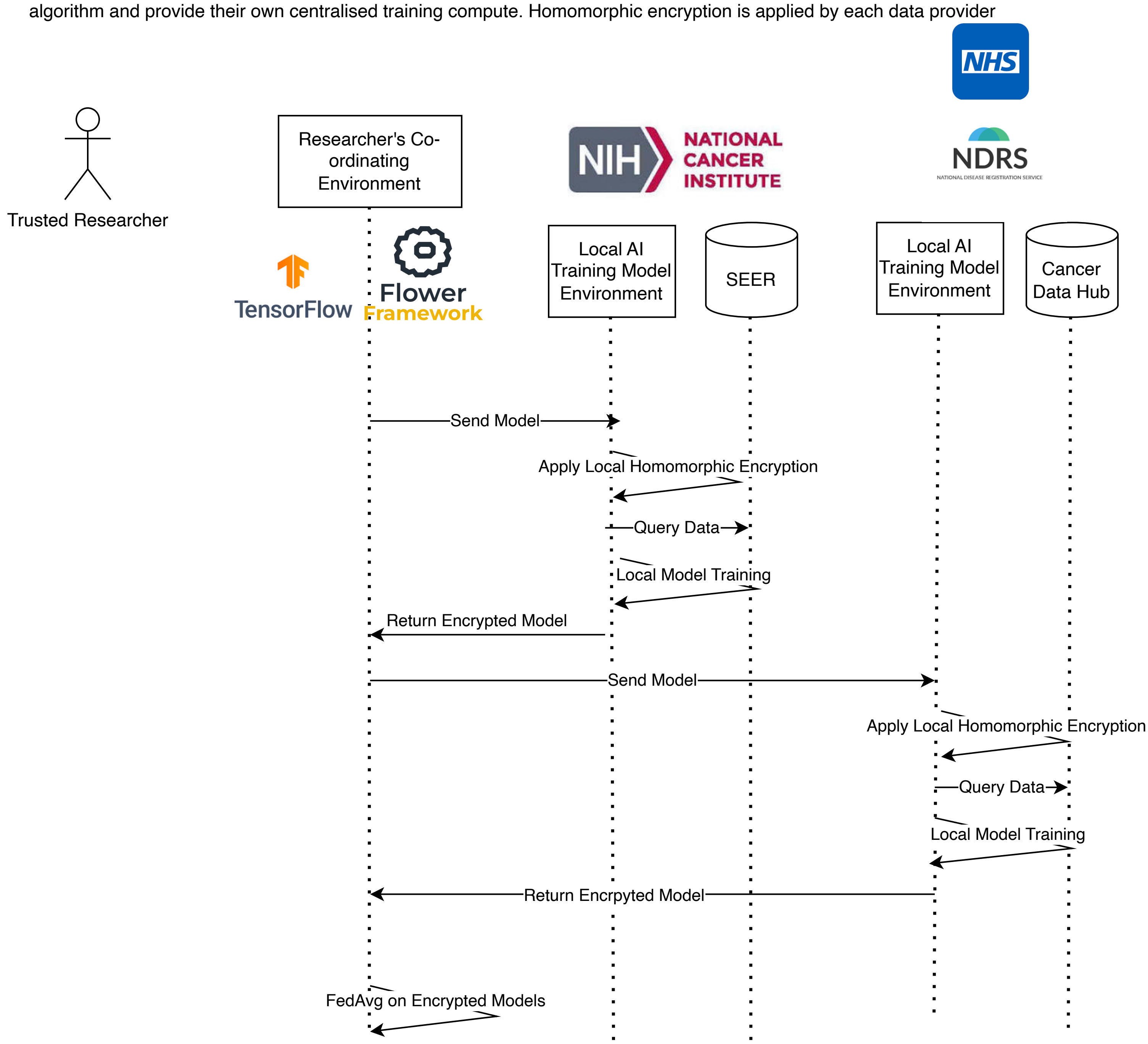
2. Each data provider would be able to provide a local AI training model environment where code would be deployed

3. Models would be trained to work against local data so would need to be deployed to local training environments

4. Orchestration would feed back to the trusted researcher where FedAvg or another algorithm would be applied across the models

Federated Learning + Homomorphic Encryption Use Case Flow

A trusted researcher with access to data wishes to train their model against UK & US data. They bring their own algorithm and provide their own centralised training compute. Homomorphic encryption is applied by each data provider



Assumptions:

1. Tensor Flow Federated or Flower framework could both be used as orchestrators of a federated learning use case

2. Each data provider would be able to provide a local AI training model environment where code would be deployed

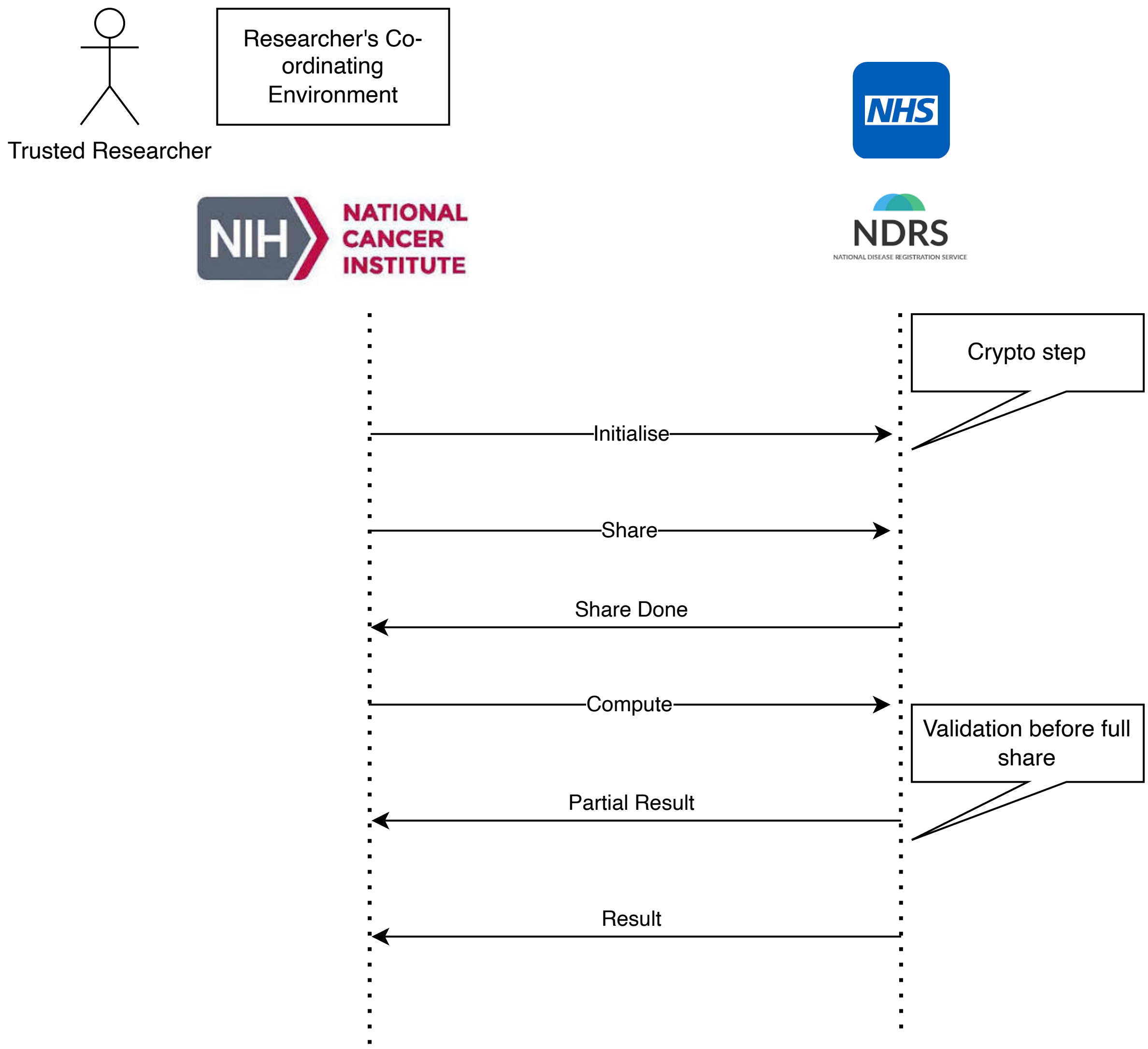
3. Models would be trained to work against local data so would need to be deployed to local training environments

4. Each data provider enforces their version of homomorphic encryption

5. Orchestration would feed back to the trusted researcher where FedAvg or other algorithm is applied across the models in encrypted form

Secure Multi Party Computation Use Case Flow

In a Multi Party Computation, flow the user accesses via one of the parties in the SMPC pair-wise relationships. In this example the access starts via the US.



Assumptions:

2. Participant A and Participant B share their private inputs with each other, typically in a secret-shared form.

3. Once both participants have shared their inputs, they acknowledge to each other that the sharing process is complete. This ensures that both parties are ready to proceed with the computation.

4. Participant A and Participant B independently perform their local computations on the shared inputs. Each participant computes their partial results based on the agreed-upon computation function.

5. Participant A and Participant B exchange their partial results with each other. This allows each participant to verify the correctness of their computation and combine the partial results to obtain the final result.

6. Participant A and Participant B collectively reconstruct the final result of the computation based on the exchanged partial results. The final result is obtained without revealing the individual inputs to each other, ensuring privacy and confidentiality.

Privacy Enhancing Technologies and Test Driven Security

Agree testing framework in advance of the technology implementation

	Functional Testing	Security Testing	Data Minimisation Tests	Privacy Impact Assessments	User Acceptance Testing	Compliance Auditing	T
What	- Test each PET individually - Test data anonymisation - Check data access controls & consent applied properly	- Test against security threats - Pen testing - Verify encryption algorithms	- Confirm only minimum necessary data is collected - Check for unneccesary data fields	- Conduct PIA for implemented PETs & small datasets - Mitigate identified risks	- Involve end-users to test functional use cases are achieved - Collected feedback and review	- Verify that implemented PETs comply with relevant privacy regulations & standards - Ensure documentation is complete	- E exp imp - E priv
How							
Who							

hird Party
Auditing

ngage external
berts to validate
plementation
ngage patient
vacy groups

Anonymity
Tests

- Assess whether anonymisation is effective in avoiding re-identification
- Test for all countries with added external social data, attempt reconstruction with allowed personage

Performance
Tests

- Assess performance impacts especially around encryption
- Measure processing time and resource utilisation