



## PathLAKE Research Database - Privacy Notice

### **Who are we?**

PathLAKE is a consortium of some of the UK's leading digital and computational experts from the NHS and Universities. PathLAKE will form a computational pathology hub to drive artificial intelligence innovation in pathology for the UK and create the world's largest repository of annotated digital whole slide images. This repository is the PathLAKE Research Database.

University Hospitals of Coventry and Warwickshire NHS Trust is the lead organisation and sponsor of this project. This notice sets out how your data will be used by PathLAKE. We are a controller over the data described in this notice, and if you have any queries regarding our use of your data, please contact us using the details at the bottom of this notice.

### **What is a Research database?**

A research database is a structured collection of individual-level personal information, which is stored for potential research purposes beyond the life of a specific research project with defined endpoints. Research purposes in this context refers to analysis of data to answer research questions in multiple projects.

### **What information do we hold about you?**

The PathLAKE Research Database holds whole slide images from cellular pathology samples taken from patients undergoing diagnostic investigations as part of their routine care at each of the participating hospitals. The data lake will accept de-identified data only. Hospitals providing the data remain responsible for ensuring compliance with GDPR prior to the de-identification, including for the process(es) concerning consent where applicable.

Data categories which may be included in the database with the whole slide images will vary at each participating hospital. The types of data uploaded to the database will be limited to:

- Patient age range at the time of the sample
- Year and/or month the sample was taken

- Specimen type / nature
- Clinical details provided with the sample
- All non-identifiable data included in the pathology report and laboratory information system
- De-identified data held in the pathology information system related to the patient's other samples (e.g. blood tests, cytology, and microbiology results).
- De-identified data held in NHS Trust information databases such as the radiology picture archive, other laboratory databases (e.g. biochemistry results) and communication system may be included.
- The patient's postcode data limited to the first part (the district code) and not the full postcode

Where feasible the database will also be updated with pseudonymised records as the study progresses. The amount of clinical data is expected to increase as resources and experience in collection of this data progress.

All confidential patient information will reside within the NHS Trusts and no identifiable information will be uploaded at any time to the research database. Keys to link the research data to patients will be required for longitudinal updates of clinical outcome; however, this will be strictly controlled, limited only to individual NHS Trusts' data access approved staff.

### **How is your Data collected?**

Clinicians collect your data from electronic hospital records such as the pathology information system and NHS information databases. These are entered onto a secure online data collection tool controlled by PathLAKE and provided by our trusted external data storage providers, who store the data within the European Economic Area. Access to the data entry tool is securely controlled. There is no risk that a patient can be identified from a whole slide image alone. Identifiable data remains at the local sites behind the NHS firewall, subject to the usual data security measures to deter intruders.

### **Research**

PathLAKE aims to build a data lake that comprises annotated whole slide images of patient's histopathology samples, taken for clinical purposes at the collaborating NHS sites, that can be accessible to UK academic groups and SMEs. Access to such high quality data at such a scale is considered essential for the application of Artificial Intelligence in Pathology. PathLAKE will underpin research across a number of areas. Some examples include; colon and prostate cancer screening, immuno-oncology and breast and prostate cancer grading.

We fully anonymise your data before it is used for research purposes. We never use any data which could directly identify you for research purposes, and only the team treating you at your hospital will ever be able to see your identifiable data. We use your data in this way because we have a legitimate interest to provide this resource and it is necessary for reasons of public interest in the area of health research. There is no process in place for withdrawal of data from the database.

## **How long will we use your data for?**

Whole slide images and deidentified data included in the PathLAKE database will be retained by us indefinitely. It is anticipated that PathLAKE will be an ongoing resource as a research database indefinitely and that data collection will continue for the foreseeable future.

## **Feedback**

We have worked hard to produce a short notice that clearly explains how your data is used. Your feedback and suggestions on this notice are very welcome. If you feel that we have overlooked an important perspective or used language which you think we could improve, please let us know by using the contact details below.

September 2020

## **Further information**

The South Central - Oxford C Research Ethics Committee confirmed a favourable ethical opinion of the PathLAKE research database for a period of five years from 16<sup>th</sup> October 2019 (REC reference 19/SC/0363).

PathLAKE is funded from the Data to Early Diagnosis and Precision Medicine strand of the government's Industrial Strategy Challenge Fund, managed and delivered by Innovate UK on behalf of UK Research and Innovation (UKRI; Grant reference 104689).

More information about PathLAKE is available on our website [www.pathlake.org](http://www.pathlake.org)

For more information about University Hospitals Coventry and Warwickshire NHS Trust, please visit: [www.uhcv.nhs.uk](http://www.uhcv.nhs.uk)