

# Outbreak data analysis platform

## What it is

The outbreak data analysis platform exists to provide an accessible, usable data resource by linking and curating data from **clinical records**, **research studies** and **audit data** for UK patients. The platform combines trusted research environment infrastructure with >£100M of exabyte-scale computational capacity on the UK National Supercomputer. The solitary aim of the platform is to accelerate scientific understanding of new outbreaks for the benefit of patients and the protection of the public.

## What it isn't

- It is not a viral sequence analysis platform. The excellent MRC CLIMB resource provides excellent resources and infrastructure for analysis, presentation and sharing of viral sequences.
- It is not a replacement for public health activities. The platform will work closely with public health agencies across the 4 nations and with UK HSA, providing data feeds where useful to augment surveillance capacity, and a tried-and-tested route to engage additional analytic capacity and expertise from the academic sector.
- It is not a replacement for existing TREs. The analysis platform will provide curated data feeds to TREs across the UK to facilitate and supplement data held elsewhere.

## Design of the platform

### Data held

The platform already contains a unique aggregation of UK sovereign data assets, including the complete data resources of the ISARIC4C/CO-CIN, GenOMICC, PHOSP and UK-CIC studies, together with viral sequence data from COG-UK, and linkage to NHS clinical records and structured clinical audit data. This creates a unique opportunity to combine clinical, biological, genomics and virology research in a secure, openly-accessible framework. Manual curation of these linked datasets, in a single platform, is a key step to maximise data quality and usability.

### Compute power

- processing power
- 2.5Pb of storage
- GPU arrays for massively-parallel applications such as genomics and image analysis
- API structure for real-time data sharing with other trusted research environments across the UK
- data pipelines to supply surveillance feeds to public health

### Structure

There are two routes of access to the analysis platform (Figure 1):

1. NHS Trusted Research Environment (Safe Haven) for access to personal clinical data and data collected without explicit consent.

2. Rapid-access flexible compute for access to non-disclosive research data collected with explicit consent.

Within both of these environments there are two levels of access, governed by the data contributors:

1. Publishable “open access” data which any user can use and report as they wish, according to data protection and privacy rules;
2. Embargoed active research data, shared by academic investigators and available for linked analysis but not for publication without agreement from all contributors.

This design is intended to build trust in order to encourage immediate contributions of research data from academic collaborators. It makes data available immediately for discovery, whilst protecting the rights of data creators and contributors.

## Research outputs

By generating, integrating and analysing clinical, biological, genetic and virological data on patients with Covid-19 in UK hospitals, the outbreak analysis platform has facilitated research that:

- provided essential weekly updates to SAGE that guide the public health response [isaric4c.net/reports/](https://isaric4c.net/reports/),
- identified host genetic mechanisms of disease,<sup>1</sup>
- quantified the role of age, comorbid illness and obesity in disease severity,<sup>2</sup>
- created the global standard ISARIC4C score for prognostication [isaric4c.net/risk](https://isaric4c.net/risk/),<sup>3</sup>
- determined the impact of long Covid following hospitalisation<sup>4</sup>
- identified the substantial effect of nosocomial transmission of Covid-19 within hospitals [SPI-M/SAGE report](#),
- provided key evidence underlying the choice of therapeutic agents for clinical trials<sup>1,5</sup>
- provided real world data on vaccine effectiveness and failure (SAGE 87 Egan et al, Egan et al.)
- observed data supporting identification of high risk groups for vaccination (highlighted in No10 briefing)
- described the complications of Covid-19 in hospitalised patients.<sup>6</sup>

## Governance of the platform

### National core studies

The platform is funded by UKRI and the National Core Studies programme (Data and Connectivity).

### Advisory committee

A UK-wide advisory committee, initially convened by HDR UK to solve data linkage problems in Covid research, will provide guidance on all aspects of platform management and delivery. The committee comprises representation from public health agencies in the 4 nations of the UK, representatives of key studies, TREs and data resources, and representatives of government and industry.

### Executive group

An executive committee will be charged with managing resources and delivering on the goals set by the advisory committee.

## Current data content

This platform now serves as a hub for a coordinated UK national research response to COVID-19. Data are included from:

- ISARIC4C tier 0: (unconsented) prospective clinical data from 215,825 cases
- ISARIC4C tiers 1 and 2: serial multiomic assays from research samples of blood, respiratory secretions, urine, and stool from 2,401 cases

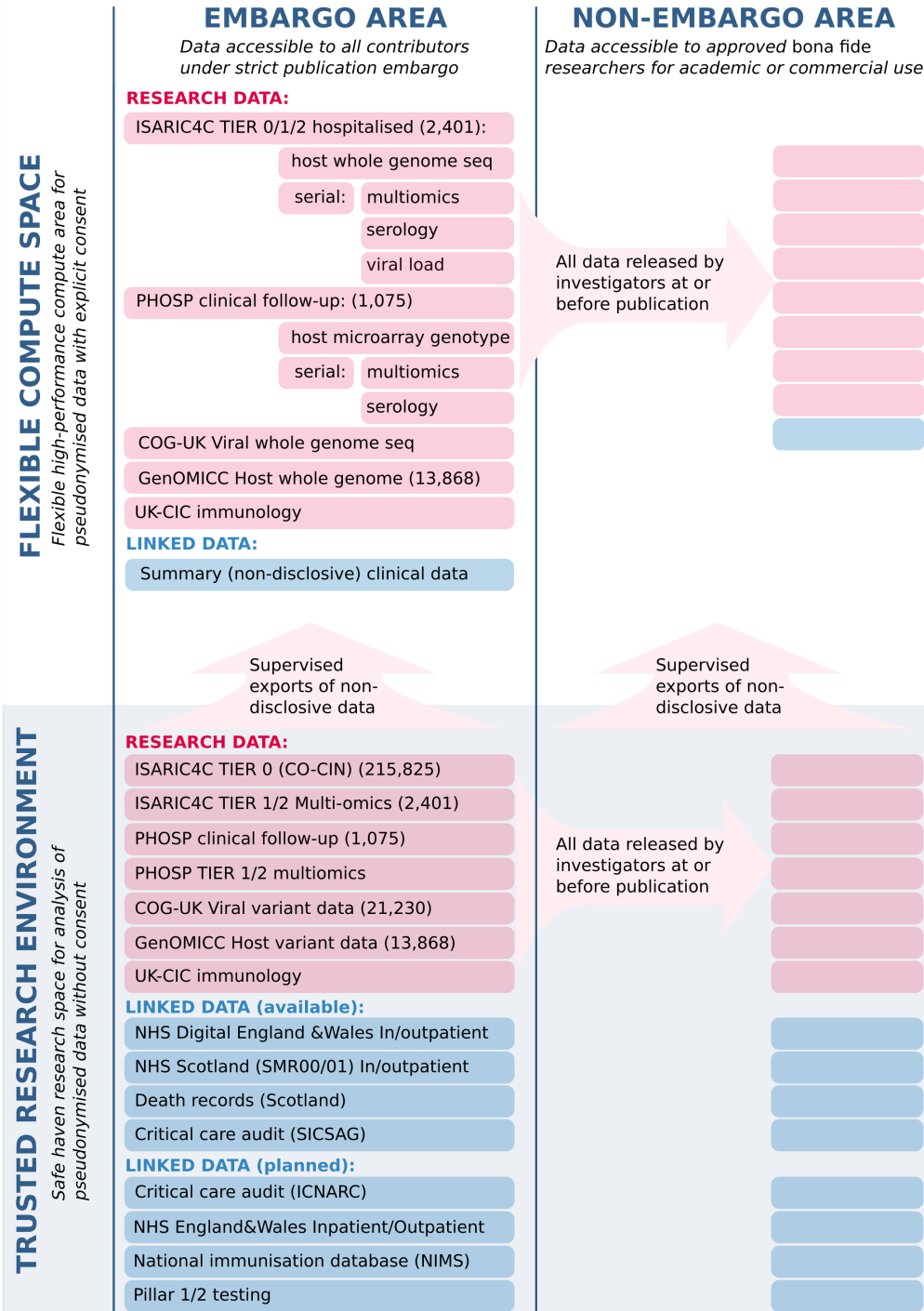


Figure 1: Structure of the Outbreak Analysis Platform

- COG-UK: (unconsented) summary variant data from COG-UK viral sequencing study is already included for matched patients
- GenOMICC study complete data: microarray and whole genome sequence data from 13,868 cases
- PHOSP complete data: follow-up clinical and biological data generated by the Post-Hospitalisation for COVID-19 follow-up study (1,075 cases)
- UK-CIC: deep immunological phenotyping data from across the UK Coronavirus Immunology Consortium, using ISARIC4C samples and local collections.

Research data within the analysis platform is already linked to:

- NHS Scotland primary, secondary care and death records
- NHS Digital health records data

In future, plans are in place to transfer data to link with:

- ICNARC and SICSAG critical care audit databases
- NIMS National Immunisation Dataset
- Pillar 1 testing
- Pillar 2 testing
- ONS

## References

1. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., Furniss, J., Richmond, A., Gountouna, E., Wrobel, N., Harrison, D., Wang, B., Wu, Y., Meynert, A., Griffiths, F., Oosthuyzen, W., Kousathanas, A., Moutsianas, L., Yang, Z., Zhai, R., Zheng, C., Grimes, G., Beale, R., Millar, J., Shih, B., Keating, S., Zechner, M., Haley, C., Porteous, D.J., Hayward, C., Yang, J., Knight, J., Summers, C., Shankar-Hari, M., Klenerman, P., Turtle, L., Ho, A., Moore, S.C., Hinds, C., Horby, P., Nichol, A., Maslove, D., Ling, L., McAuley, D., Montgomery, H., Walsh, T., Pereira, A.C., Renieri, A., Shen, X., Ponting, C.P., Fawkes, A., Tenesa, A., Caulfield, M., Scott, R., Rowan, K., Murphy, L., Openshaw, P.J.M., Semple, M.G., Law, A., Vitart, V., Wilson, J.F. & Baillie, J.K. Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98(2021).
2. Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C., Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K. & Semple, M.G. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ* **369**, (2020).
3. Knight, S.R., Ho, A., Pius, R., Buchan, I., Carson, G., Drake, T.M., Dunning, J., Fairfield, C.J., Gamble, C., Green, C.A., Gupta, R., Halpin, S., Hardwick, H.E., Holden, K.A., Horby, P.W., Jackson, C., Mclean, K.A., Merson, L., Nguyen-Van-Tam, J.S., Norman, L., Noursadeghi, M., Olliaro, P.L., Pritchard, M.G., Russell, C.D., Shaw, C.A., Sheikh, A., Solomon, T., Sudlow, C., Swann, O.V., Turtle, L.C., Openshaw, P.J., Baillie, J.K., Semple, M.G., Docherty, A.B. & Harrison, E.M. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: Development and validation of the 4C mortality score. *BMJ (Clinical research ed.)* **370**, m3339(2020).
4. Group, P.-C.C., Evans, R.A., McAuley, H., Harrison, E.M., Shikotra, A., Singapuri, A., Sereno, M., Elneima, O., Docherty, A.B., Lone, N.I., Leavy, O.C., Daines, L., Baillie, J.K., Brown, J.S., Chalder, T., Soyza, A.D., Bakerly, N.D., Eason, N., Geddes, J.R., Greening, N.J., Hart, N., Heaney, L.G., Heller, S., Howard, L., Jacob, J., Jenkins, R.G., Jolley, C., Kerr, S., Kon, O.M., Lewis, K., Lord, J.M., McCann, G.P., Neubauer, S., Openshaw, P.J., Pfeffer, P., Rowland, M., Semple, M.G., Singh, S.J., Sheikh, A., Thomas, D., Toshner, M., Chalmers, J.D., Ho, L.-P., Horsley, A., Marks, M., Poinasamy, K., Wain, L.V. & Brightling, C.E. Physical, cognitive and mental health impacts of COVID-19 following hospitalisation – a multi-centre prospective cohort study. *medRxiv* 2021.03.22.21254057(2021).doi:[10.1101/2021.03.22.21254057](https://doi.org/10.1101/2021.03.22.21254057)
5. Thwaites, R.S., Uruchurtu, A.S.S., Siggins, M.K., Liew, F., Russell, C.D., Moore, S.C., Fairfield, C., Carter, E., Abrams, S., Short, C.-E., Thaventhiran, T., Bergstrom, E., Gardener, Z., Ascough, S., Chiu, C., Docherty, A.B., Hunt, D., Crow, Y.J., Solomon, T., Taylor, G.P., Turtle, L., Harrison, E.M., Dunning, J., Semple, M.G., Baillie, J.K., Openshaw, P.J. & Investigators\*\*, on behalf of the I. Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. *Science Immunology* **6**, (2021).
6. Drake, T.M., Riad, A.M., Fairfield, C.J., Egan, C., Knight, S.R., Pius, R., Hardwick, H.E., Norman, L., Shaw, C.A., McLean, K.A., Thompson, A.A.R., Ho, A., Swann, O.V., Sullivan, M., Soares, F., Holden, K.A., Merson, L., Plotkin, D., Sigfrid, L., Silva, T.I. de, Girvan, M., Jackson, C., Russell, C.D., Dunning, J., Solomon, T., Carson, G., Olliaro, P., Nguyen-Van-Tam, J.S., Turtle, L., Docherty, A.B., Openshaw, P.J., Baillie, J.K., Harrison, E.M., Semple, M.G., Baillie, J.K., Semple, M.G., Openshaw, P.J., Carson, G., Alex, B., Bach, B., Barclay, W.S., Bogaert, D., Chand, M., Cooke, G.S., Docherty, A.B., Dunning, J., Filipe, A. da S., Fletcher, T., Green, C.A., Harrison, E.M., Hiscox, J.A., Ho, A.Y., Horby, P.W., Ijaz, S., Khoo, S., Klenerman, P., Law, A., Lim, W.S., Mentzer, A.J., Merson, L., Meynert, A.M., Noursadeghi, M., Moore, S.C., Palmairini, M., Paxton, W.A., Pollakis, G., Price, N., Rambaut, A., Robertson, D.L., Russell, C.D., Sancho-Shimizu, V., Scott, J.T., Silva, T. de, Sigfrid, L., Solomon, T., Sriskandan, S., Stuart, D., Summers, C., Tedder, R.S., Thomson, E.C., Thompson, A.R., Thwaites, R.S., Turtle, L.C., Gupta, R.K., Palmieri, C., Swann, O.V., Zambon, M., Dumas, M.-E., Griffin, J., Takats, Z., Chechi, K., Andrikopoulos, P., Osagie, A., Olanipekun, M., Liggi, S., Lewis, M., Correia, G. dos S., Sands, C., Takis, P., Maslen, L., Hardwick, H., Donohue, C., Griffiths, F., Oosthuyzen, W., Norman, L., Pius, R., Drake, T.M., Fairfield, C.J., Knight, S.R., Mclean, K.A., Murphy, D., Shaw, C.A., Dalton, J., Girvan, M., Saviciute, E., Roberts, S., Harrison, J., Marsh, L., Connor, M., Halpin, S., Jackson, C., Gamble, C., Plotkin, D., Lee, J., Leeming, G., Law, A., Wham, M., Clohisey, S., Hendry,

R., Scott-Brown, J., Greenhalf, W., Shaw, V., McDonald, S.E., Keating, S., Ahmed, K.A., Armstrong, J.A., Ashworth, M., Asiimwe, I.G., Bakshi, S., Barlow, S.L., Booth, L., Brennan, B., Bullock, K., Catterall, B.W., Clark, J.J., Clarke, E.A., Cole, S., Cooper, L., Cox, H., Davis, C., Dincarslan, O., Dunn, C., Dyer, P., Elliott, A., Evans, A., Finch, L., Fisher, L.W., Foster, T., Garcia-Dorival, I., Greenhalf, W., Gunning, P., Hartley, C., Jensen, R.L., Jones, C.B., Jones, T.R., Khandaker, S., King, K., Kiy, R.T., Koukorava, C., Lake, A., Lant, S., Latawiec, D., Lavelle-Langham, L., Lefteri, D., Lett, L., Livoti, L.A., Mancini, M., McDonald, S., McEvoy, L., McLauchlan, J., Metelmann, S., Miah, N.S., Middleton, J., Mitchell, J., Moore, S.C., Murphy, E.G., Penrice-Randal, R., Pilgrim, J., Prince, T., Reynolds, W., Ridley, P.M., Sales, D., Shaw, V.E., Shears, R.K., Small, B., Subramaniam, K.S., Szemiel, A., Taggart, A., Tanianis-Hughes, J., Thomas, J., Trochu, E., Tonder, L. van, Wilcock, E., Zhang, J.E., Flaherty, L., Maziere, N., Cass, E., Carracedo, A.D., Carlucci, N., Holmes, A., Massey, H., Murphy, L., Wrobel, N., McCafferty, S., Morrice, K., MacLean, A., Adeniji, K., Agranoff, D., Agwuh, K., Ail, D., Aldera, E.L., Alegria, A., Angus, B., Ashish, A., Atkinson, D., Bari, S., Barlow, G., Barnass, S., Barrett, N., Bassford, C., Basude, S., Baxter, D., Beadsworth, M., Bernatoniene, J., Berridge, J., Best, N., Bothma, P., Chadwick, D., Brittain-Long, R., Bulteel, N., Burden, T., Burtenshaw, A., Caruth, V., Chadwick, D., Chamblor, D., Chee, N., Child, J., Chukkambotla, S., Clark, T., Collini, P., Cosgrove, C., Cupitt, J., Cutino-Moguel, M.-T., Dark, P., Dawson, C., Dervisevic, S., Donnison, P., Douthwaite, S., Drummond, A., DuRand, I., Dushianthan, A., Dyer, T., Evans, C., Eziefula, C., Fegan, C., Finn, A., Fullerton, D., Garg, S., Garg, S., Garg, A., Gkrania-Klotsas, E., Godden, J., Goldsmith, A., Graham, C., Hardy, E., Hartshorn, S., Harvey, D., Havalda, P., Hawcutt, D.B., Hobrok, M., Hodgson, L., Hormis, A., Jacobs, M., Jain, S., Jennings, P., Kaliappan, A., Kasipandian, V., Kegg, S., Kelsey, M., Kendall, J., Kerrison, C., Kerslake, I., Koch, O., Koduri, G., Koshy, G., Laha, S., Laird, S., Larkin, S., Leiner, T., Lillie, P., Limb, J., Linnett, V., Little, J., Lyttle, M., MacMahon, M., MacNaughton, E., Mankregod, R., Masson, H., Matovu, E., McCullough, K., McEwen, R., Meda, M., Mills, G., Minton, J., Mirfenderesky, M., Mohandas, K., Mok, Q., Moon, J., Moore, E., Morgan, P., Morris, C., Mortimore, K., Moses, S., Mpenge, M., Mulla, R., Murphy, M., Nagel, M., Nagarajan, T., Nelson, M., Norris, L., O'Shea, M.K., Otahal, I., Ostermann, M., Pais, M., Palmieri, C., Panchatsharam, S., Papakonstantinou, D., Paraiso, H., Patel, B., Pattison, N., Pepperell, J., Peters, M., Phull, M., Pintus, S., Pooni, J.S., Post, F., Price, D., Prout, R., Rae, N., Reschreiter, H., Reynolds, T., Richardson, N., Roberts, M., Roberts, D., Rose, A., Rousseau, G., Ryan, B., Saluja, T., Shah, A., Shanmuga, P., Sharma, A., Shawcross, A., Sizer, J., Shankar-Hari, M., Smith, R., Snelson, C., Spittle, N., Staines, N., Stambach, T., Stewart, R., Subudhi, P., Szakmany, T., Tatham, K., Thomas, J., Thompson, C., Thompson, R., Tridente, A., Tupper-Carey, D., Twagira, M., Vallotton, N., Vancheeswaran, R., Vincent-Smith, L., Visuvanathan, S., Vuylsteke, A., Waddy, S., Wake, R., Walden, A., Welters, I., Whitehouse, T., Whittaker, P., Whittington, A., Papineni, P., Wijesinghe, M., Williams, M., Wilson, L., Sarah, S., Winchester, S., Wiselka, M., Wolverson, A., Wootton, D.G., Workman, A., Yates, B. & Young, P. Characterisation of in-hospital complications associated with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol UK: A prospective, multicentre cohort study. *The Lancet* **398**, 223–237(2021).