Interactive Data Visualisation: Insights Gained from Developing a New National COVID-19 Seroepidemiology Data Hub in Ireland, 2021-2022

Laurin Grabowsky, Katie M O'Brien, Jennifer Doyle, Thomas Roux, Eoghan McCarthy, Ciara Kelly, Simon Bergin, Liam Murray, Eimear Hayes, Melissa Brady, Derval Igoe
.

BACKGROUND

The Health Protection Surveillance Centre Seroepidemiology Unit was established in 2021 and commenced publishing static reports and data visualisations using data provided by the Irish Blood Transfusion Service and the Laboratory Surveillance Network. A process to design and implement an interactive online data hub was initiated in February 2022 as a means of promptly communicating serosurveillance data and supporting public health planning and decision making.

OBJECTIVES

The aim of this development was:
- to invite audience engagement with interactive output
- reduce the lead time between the receipt and communication of seroprevalence data
- create a streamlined reproducible workflow for data processing, analysis, and display

OUTCOME

- The Seroepidemiology of COVID-19 Data Hub went live in July 2022.
- The hub has received 2,000 unique views with an average screen time of 5 minutes per user.
- Fortnightly updates received repeated media coverage.
- The retrieval, transformation, and communication of serosurveillance data has been streamlined.
- Lead time between data receipt and dissemination has been reduced from approximately 14 days to 2 hours.
- The process has been automated and documented to eliminate of single points of failure and potential delays between data receipt and communication of results.

Fig 1: Historic Seroepidemiology Unit report transformed into interactive data visualisations presented via ArcGIS dashboards
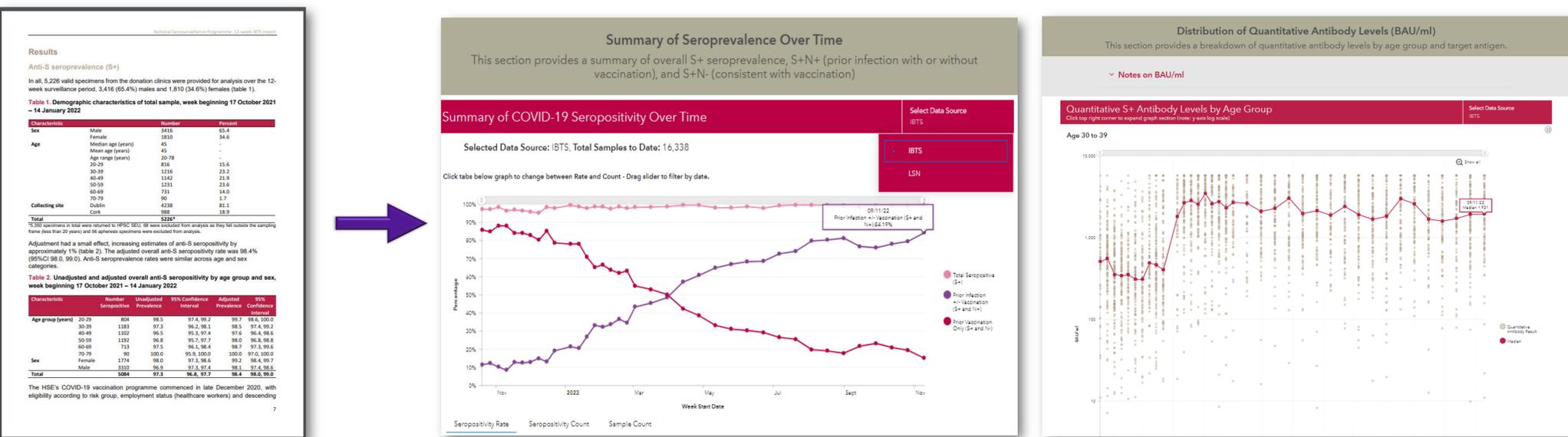
PROCESS

- **Initiation:** Fundamental reporting requirements, data formats, contents, and update cycles were established and agreed with data provision partners.
- **Source Data:** Data cleaning, processing, and validation steps were automated inside a SQL database.
- **Design Process:** Iterative development meetings (c.10) were scheduled with an ESRI ArcGIS technical expert over a period of 5 months. Design and visualisation reference documents were developed to guide the development process.

- **Dashboard Extracts:** Stable dashboard input data extracts were produced using R programming scripts which enabled repeatable automated updating of dashboard content.
- **Update Workflows:** Standard operating procedures were agreed and documented to ensure transparency and repeatability of the data flow.
- **Validation:** A pre-production version of the hub was deployed to provide a testing and validation environment which is updated and refreshed in advance of the live environment.

Fig 2: Seroepidemiology data hub design process flow chart
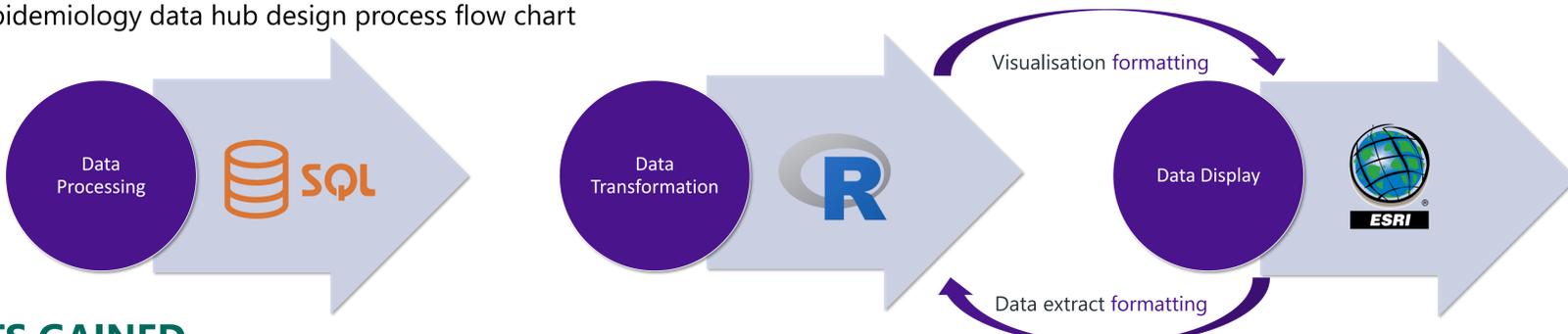
INSIGHTS GAINED

- Consider the structure and content of the raw data, and the implications for sharing these via an online data hub
- Consider the intended audience(s) and whether a dashboard is the most effective and impactful means of communication
- Ensure that the data are refreshed at a frequency that justifies the effort of designing and automating the workflow
- Ensure stability of the raw data structure to allow repeatability of automated data manipulations
- Assess the staffing and expertise required for design and implementation

- Investigate display platforms available (visualisation and design capabilities, and associated cost requirements)
- Scope the infrastructure, software, and technical support required during and after the design process
- Ensure staff receive adequate training and/or technical support required to take ownership of existing display and to develop new visualisations
- Determine and delineate development, update, and maintenance responsibilities and timelines