

From selfish silo to collaborative culture – embracing data-enabled cancer research

openaccessgovernment.org/article/collaborative-culture-embracing-data-enabled-cancer-research/178254

19 June 2024

Aedin Culhane and Mark Lawler, Co-Leads of the eHealth Hub for Cancer, reflect on their data-enabled cancer research journeys, how their collaborative team science approach has reaped significant dividends in cancer research and policy and how the hub is inducing a paradigm shift in how health data are deployed on the island of Ireland

An article in The Wall Street Journal (not the normal reading material for scientists) in 2011 highlighted a new approach to performing scientific research that was gaining significant credence at the time. Entitled ‘The New Einsteins Will Be Scientists Who Share’ (with the strapline ‘From cancer to cosmology, researchers could race ahead by working together – online and in the open’), the article presaged an unprecedented change in how scientists interact with each other, ushering in a culture of collaboration and cross-disciplinary research. Nowhere was this change more obvious than in the genomics and data science community, where a bottom-up movement led to the creation of the Global Alliance for Genomics and Health ([GA4GH](#)), bringing together researchers from around the world to work together to address some of human health’s greatest challenges through the deployment of data and data tools.

The GA4GH Cancer Task Team (Lawler co-lead) was particularly active, breaking down silos and empowering the community to share data, code, and expertise for common benefit. ⁽¹⁾ Tools such as the Cancer Meta Knowledgebase were created, a data ‘translator’ that harnessed knowledge on potential/actual mutations linked to a particular cancer, aggregated that information in a single knowledgebase and shared it freely online with the entire community, thus ensuring a more collaborative team science approach to cancer research. ⁽²⁾ The principle of a Cancer Knowledge Network was defined with a determination to move from a closed ‘selfish silo’ mentality to a more open-source collaborative culture. ⁽³⁾

Embedding the patient in the centre of a data-enabled cancer research revolution

In the UK, this focus on data gave rise to DATA-CAN, the UK’s Health Data Research Hub for Cancer (ML Scientific Director), part of Health Data Research UK (HDR UK), the UK’s health data science institute. A key DATA-CAN philosophy is the concept of fair value, which means that all stakeholders, whether they be patients, the National Health Service (NHS), researchers, or industry, get fair value from the use of patient data. Patients are the DNA of DATA-CAN. Patients were involved in the original funding bid, sit on all steering, management and project committees and are in the room for all

discussions, including with industry. ⁽⁴⁾ DATA-CAN's Patient and Public Involvement and Engagement (PPIE) activities have led to it being recognised as a PPIE exemplar; ⁽⁴⁾ a philosophy best captured in the words of DATA-CAN's PPIE group member and breast cancer survivor Jacqui Gath: 'Patients want their data to be used to improve care and enhance research. In fact, they're often surprised it's not used already.' These words reflect a common viewpoint from patients and need to be heard more widely in the debate on data privacy and secondary data use for research. Data privacy and trust in the use of data must be balanced with the need to deploy data to help enhance health and wellbeing. ⁽⁵⁾ During the COVID pandemic, out of necessity, data privacy rules were relaxed, and data access for research was simplified in order to access the data needed to both develop and test emerging COVID vaccines. Cancer has killed many more people than COVID and deserves the same considerations for pragmatic data privacy and secondary data use. ⁽⁶⁾

Changing the paradigm: bringing the solution to the data rather than the data to the solution

As genomic and clinical datasets increase in resolution and size, transferring large datasets between institutions/jurisdictions has become costly and cumbersome. Data privacy regulations also impact the handling and sharing of data, necessitating the development of novel solutions to ensure the privacy-preserving sharing of sensitive medical data. Federated data analytics provide an innovative solution. Data remain in place; each dataset is protected within a local or cloud-secure data environment (Trusted Research Environment (TRE)). Compute is performed locally. We have been at the forefront of developing standards for code sharing and data harmonization that enable federated clinical genomics research. The NHS Research Secure Data Environment (SDE) Network recently announced it will adopt this OMOP data standard. The European Medicines Agency's DARWIN EU project, is utilizing the OMOP CDM to conduct network studies in the UK and 12 EU countries to advance treatments for rare cancers.

Federated data networks offer privacy, regulatory compliance, governance, and scalability advantages but require that data are harmonized to a common standard. A common data model empowers researchers to perform cross-institute and cross-border data analysis at each local site without transferring or sharing the data itself. ⁽⁷⁾ We recently helped create clinical genomics data standards for federated sharing between 25 EU countries within the 1+ Million Genomes Genomic Data Infrastructure Program ⁽⁸⁾ and are also developing training and frameworks to support the adoption of OHDSI (Observational Health Data Sciences and Informatics) Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The NHS Research Secure Data Environment (SDE) Network recently announced it will adopt this OMOP data standard.

Cancer (and data) know no borders

Our data journey has reached an important inflection point, where we now both work on the island of Ireland, one island but with two distinct jurisdictions, each with its own data policies and legislation. However, the principle of a collaborative culture is still very

relevant. Recognising this and the need to enhance the deployment of data on the island of Ireland to supercharge cancer research and its translation to improved cancer care, we have established the eHealth Hub for Cancer, an emerging hub of excellence funded through the Higher Education Authority North South Research Programme. This Hub has acted as a magnet to coalesce the cancer community on the island of Ireland, providing a data-empowered milieu to work together to develop innovative solutions to a series of cancer challenges that are best addressed using data. ^(6, 9) Growing applications of clinical next-generation sequencing and the success of precision oncology mean the number of clinically-actionable molecular features is increasing, and to maximally benefit patients, the eHealth Hub for Cancer is addressing the challenges associated with the management and standardisation of these data. The Hub facilitates the employment of common data models such as OMOP that enhance the analysis and sharing of data across the island of Ireland, underpinning data-driven insights to drive future research discoveries and inform cancer care and cancer policy. Cancer (and cancer data) know no borders. Neither should we.

References

1. Lawler M, Siu LL, Rehm HR, Chanock SJ, Alterovitz G, Burn J, Calvo F, Lacombe D, Teh BT, North KN, Sawyers CL. All the World's a Stage. Facilitating discovery science and improved cancer care through the Global Alliance for Genomics and Health Cancer Discovery 2015; 5:1133-6. doi: 10.1158/2159-8290.cd-15-0821. PMID: 26526696
2. Wagner AH, Walsh B, Mayfield G, Tamborero D, Sonkin D, Krysiak K, Deu-Pons J, Duren RP, Gao J, McMurry J, Patterson S, Del Vecchio Fitz C, Pitel BA, Sezerman OU, Ellrott K, Warner JL, Rieke DT, Aittokallio T, Cerami E, Ritter DI, Schriml LM, Freimuth RR, Haendel M, Raca G, Madhavan S, Baudis M, Beckmann JS, Dienstmann R, Chakravarty D, Li XS, Mockus S, Elemento O, Schultz N, Lopez-Bigas N, Lawler M, Goecks J, Griffith M, Griffith OL, Margolin AA; Variant Interpretation for Cancer Consortium. A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. Nat Genet. 2020; 52:448-457. DOI: 10.1038/s41588-020-0603-8 PMID: 32246132
3. Lawler M, Haussler D, Siu LL, Haendel MA, McMurry JA, Knoppers BM, Chanock SJ, Calvo F, The BT, Walia G, Banks I, Yu PP, Staudt LM, Sawyers CL. Clinical Cancer Genome Task Team of the Global Alliance for Genomics and Health(GA4GH) Sharing Clinical and Genomic Data on Cancer – The Need for Global Solutions. N Engl J Med. 2017; 376:2006-2009. DOI: 10.1056/NEJMp1612254 .
4. Wheatstone P, Gath J, Carrigan C, Hall G, Cook Y, DATA-CAN Sujenthiran A, Peach J, Davie C, Lawler M. DATA-CAN: a co-created cancer data knowledge network to deliver better outcomes and higher societal value. BMJ Partnerships in Practice 2021 <https://blogs.bmj.com/bmj/2021/08/11/data-can-a-co-created-cancer-data-knowledge-network-to-deliver-better-outcomes-and-higher-societal-value/>

5. Lawler M, Morris AD, Sullivan R, Birney E, Middleton A, Makaroff L, Knoppers BM, Horgan D, Eggermont A. A roadmap for restoring trust in Big Data. *Lancet Oncol.* 2018; :1014-1015. DOI: 10.1016/S1470-2045(18)30425-X.PMID: 30102210
6. Lawler M, Crul M. Data must underpin our response to the covid-19 pandemic's disastrous impact on cancer. *BMJ.* 2022 ;376:o282. doi: 10.1136/bmj.o282.PMID:35115384
7. Roel, E., Pistillo, A., Recalde, M., Sena, A.G., Fernández- Bertolín, S., Aragón, M., Puente, D., Ahmed, W.-U.-R., Alghoul, H., Alser, O., Alshammari, T.M., Areia, C., Blacketer, C., Carter, W., Casajust, P., Culhane, A.C., Dawoud, D., DeFalco, F., DuVall, S.L., Falconer, T., Golozar, A., Gong, M., Hester, L., Hripcsak, G., Tan, E.H., Jeon, H., Jonnagaddala, J., Lai, L.Y.H., Lynch, K.E., Matheny, M.E., Morales, D.R., Natarajan, K., Nyberg, F., Ostropolets, A., Posada, J.D., Prats-Urbe, A., Reich, C.G., Rivera, D.R., Schilling, L.M., Soerjomataram, I., Shah, K., Shah, N.H., Shen, Y., Spotniz, M., Subbian, V., Suchard, M.A., Trama, A., Zhang, L., Zhang, Y., Ryan, P.B., Prieto-Alhambra, D., Kostka, K., Duarte-Salles, T., 2021. Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* 30, 1884–1894. <https://doi.org/10.1158/1055-9965.EPI-21-0266>
8. Riba, M., Sala, C., Culhane, A.C., Flobak, Å., Patocs, A., Boye, K., Plevova, K., Pospíšilová, Š., Gandolfi, G., Morelli, M.J., Bucci, G., Edsjö, A., Lassen, U., Al-Shahrour, F., Lopez-Bigas, N., Hovland, R., Cuppen, E., Valencia, A., Poirel, H.A., Rosenquist, R., Scollen, S., Arenas Marquez, J., Belien, J., De Nicolo, A., De Maria, R., Torrents, D., Tonon, G., 2024. The 1+Million Genomes Minimal Dataset for Cancer. *Nat. Genet.* 1–4. <https://doi.org/10.1038/s41588-024-01721-x>
9. <https://www.rte.ie/brainstorm/2022/0519/1299916-cancer-treatment-diagnosis-covid-delays-data-ireland-europe/>

More About Stakeholder

The Patrick G Johnston Centre for Cancer Research:

Relieve the human suffering from cancer

Addressing one of the greatest challenges to human health and well-being, the Patrick G Johnston Centre for Cancer Research at Queen's University Belfast (QUB) is recognised internationally as an impactful research centre committed to relieving the human suffering from cancer.

