FAIR Linked Data Framework for Glioblastoma Research.

Zoi Katsirea[1], Samira Osterop[2], Valérie Dutoit[3], Tong Li[1], Etienne Doumazane[6], Stephane Pages[2], Omer Ali Bayraktar[1], Josh Moore[4], Andra Waagmeester[5] and Tannia Gracia[1,*]

[1]*Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom*
[2]*Wyss Center for Bio and Neuroengineering, Geneva, Switzerland*
[3]*University Hospitals of Geneva (HUG), Geneva, Switzerland*
[4]*German BioImaging e.V., Constance, Germany*
[5]*Micelio BV, Ekeren, Belgium*
[6]*Paris Brain Institute, Pitié-Salpêtrière University Hospital, Paris, France*

Abstract

Glioblastoma is an aggressive brain tumour with low survival rates and resistance to treatment, requiring innovative approaches to identify prognostic factors and therapeutic targets. Patient-derived imaging, pathological, and molecular data are critical for advancing these efforts. We have developed a dataset integrating sc-RNAseq, spatial transcriptomics, proteomics, neuropathology annotations, and clinical metadata. Efficient access, organization, and sharing of these data are essential for understanding the disease and developing therapies. This linked resource enables querying of cellular states, tissue architecture, and the tumour microenvironment by combining histological features, 3D imaging, and clinical data.

Using FAIR (Findable, Accessible, Interoperable, and Reusable) principles, glioblastoma experimental and clinical metadata from the Wellcome Sanger Institute and Wyss Center for Bio and Neuroengineering were curated and linked. Unified identifiers from ontologies and community standards were assigned, with data formatted into RDF (Resource Description Framework).

We describe efforts to standardize and promote interoperability of multimodal data collected to profile glioblastoma.

Keywords

Glioblastoma, Interoperability, Linked data, FAIR, RDF, Common data space Methods

1. Introduction

Glioblastoma (GB) is an aggressive and highly malignant type of brain tumour. It is the most common and lethal primary brain tumour in adults, with an incidence of approximately 3.2 cases per 100,000 people annually. This type of brain cancer is notoriously difficult to treat due to its rapid growth, invasive nature, and resistance to conventional therapies like surgery, radiation, and chemotherapy [1]. A major challenge in treating glioblastoma lies in its remarkable heterogeneity, both at the molecular and cellular levels, which allows the tumour to adapt and evade therapeutic interventions. This heterogeneity includes a mix of genetic mutations, diverse cell types, and microenvironmental factors within a single tumour, as well as significant variation between patients [2]. Despite advances in research, the complexity of glioblastoma heterogeneity remains poorly understood, hindering the development of effective, targeted treatments.

This multimodal dataset includes cutting edge spatial transcriptomics and single cell technology of deeply profiled glioblastoma tumours from 12 patients at Cambridge University Hospital and 31 patients at the Wyss Center in Geneva, and provides an unprecedented resource for advancing scientific and clinical understanding of this aggressive cancer. By integrating diverse data types, this dataset links

SWAT4HCLS 2025: The 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 24–27, 2025, Barcelona, Spain
*Corresponding author.
0009-0009-3815-4922 (Z. Katsirea); 0000-0002-8233-2660 (S. Osterop); 0000-0002-3469-4237 (V. Dutoit); 0000-0002-8240-4476 (T. Li); 0000-0002-0000-9983 (E. Doumazane); 0000-0003-0618-777X (S. Pages); 0000-0001-6055-277X (O. A. Bayraktar); 0000-0003-4028-811X (J. Moore); 0000-0001-9773-4008 (A. Waagmeester); 0000-0002-8237-1870 (T. Gracia)

the presence and spatial distribution of cellular states, cell types, niches, tumour microenvironment regulators, histopathological features and clinical patient metadata, in a single, comprehensive framework. This unique resource offers a powerful tool for researchers to explore the complex interplay of factors that drive glioblastoma progression and resistance to therapy. Clinical researchers can readily query this dataset to identify biomarkers, regulators of the tumour's microenvironment, and mine for effective therapeutic targets. Ultimately, this integrative approach will accelerate the translation of basic research into actionable clinical insights, fostering the development of more precise and effective treatments for glioblastoma patients.

2. Methods

2.1. Data sharing

Data sharing in this study was enabled through Wellcome Leap's Master Research Funding Agreement (MARFA) [1]. MARFA facilitates efficient and equitable collaboration by pre-negotiating terms such as intellectual property (IP) ownership, confidentiality, and publication rights, requiring only a Statement of Work (SOW) and cost negotiation to initiate projects. A common set of Program-specific data sharing terms and conditions were agreed that define each institution's roles, obligations, and security measures for the physical and virtual data sharing systems. This streamlined approach reduces barriers to sharing, accelerates project timelines, and promotes open access publishing while safeguarding confidentiality and IP through limited review periods. Within this ethical and legal framework, data could be disseminated to international members of the consortium, ensuring timely and collaborative exchange of results.

2.2. Glioblastoma Comprehensive Data set

For the GBM-SPACE project at the Wellcome Sanger Institute, 59 blocks of tumour tissue from 12 patients diagnosed with glioblastoma, clinically known as Glioblastoma IDH-Wild Type (CNS WHO grade 4), were analysed using cutting-edge technologies across multiple modalities of single-cell RNA sequencing (sc-RNAseq) and spatial transcriptomics. These technologies included 10X Chromium Single Cell 3', Parse Evercode-Mega, ResolveOME, 10X Visium CytAssist, 10X Visium HD, and 10X Xenium, enabling an in-depth and high-resolution molecular characterization of the tumour tissues. To validate the transcriptomics analyses, complementary techniques such as Laser Capture Microdissection (LCM), Ion Mobility Spectrometry (IMS), CITE-Seq and iDISCO were applied to selected tissue sections. The resulting data were meticulously linked to curated neuropathology annotations of stained tissue sections and comprehensive patient metadata, providing an integrative framework to explore the spatial organization, cellular heterogeneity, and molecular dynamics of glioblastoma .

The team led by the Wyss Center analysed 52 tumour resections from 31 different patients stored in the biobank of the University Hospitals of Geneva. The cohort included two patient groups: short-term survivors (survival less than 12 months post-diagnosis) and long-term survivors (survival exceeding 24 months post-diagnosis). These samples were processed whole, without sectioning, using the iDISCO clearing and immunolabeling protocol. This work was performed by collaborators at the Paris Brain Institute. This technique enables 3D visualization and downstream analysis of vasculature (CD31) and immune cells (T cells: CD3, macrophages: CD68). Advanced pixel-based classification and segmentation tools were then applied to model the datasets, facilitating the identification of critical survival-associated parameters.

In addition to the 52 samples collected from 31 different patients, 5 samples were shipped by the Sanger Institute to the Wyss Center for further analysis. These samples were initially collected by the Sanger team, where sections were processed for the modalities described above. The remaining tissue was subsequently sent to the Wyss Center for 3D clearing and immunolabelling with the markers GFAP

[1] https://wellcomeleap.org/wellcome-leap-hbnet/

(astrocytes/subset of tumour cells), CD3 (T cells) and CD31 (vasculature). As a result, these 5 samples were processed for the complete set of modalities as outlined in Table 1.

The modalities included in the prototype of Glioblastoma Linked Data were 10X Multiomics, 10X Visium CytAssist, iDISCO, together with the neuropathology annotations and patient metadata (Figure 1a. and 1b.)

Modality	Included in Sanger pipeline (12 patients)	Included in Wyss pipeline (31 patients)	Included in 5 samples exchanged across institutions	Included in Linked Data prototype
Multiomics 10X	x		x	x
Parse	x		x	
ResolveOme	x		x	
Visium CytAssist 10X	x		x	x
Visium HD 10X	x		x	
Xenium	x		x	
LCM (Laser Capture Microdissection)	x		x	
IMS	x		x	
iDISCO (3D)		x	x	x
CITE-seq	x		x	
Manual Neuropathology Annotations	x		x	x

Table 1
Experimental modalities used in this project

2.3. Linked Data Pipeline

Adhering to the standards for FAIR (Findable, Accessible, Interoperable, and Reusable) data[3], we set out a reproducible pipeline to achieve data standardisation and interoperability (Figure 2). We identified broad data structures that describe the experimental and clinical metadata from both research groups. We standardised entries by assigning them unique identifiers from ontologies or community standards and making them machine-readable. The data formats were then represented as RDF (Resource Description Framework), a standard schema for linking health-related data through encompassing data relations in the form of 'subject-relation-object' statements known as 'triples' [4]. In RDF, each triple consists of a subject (the concept being described), a predicate (the relationship or property), and an object (the value or another concept). This structured data was subsequently linked to external sources, enabling more sophisticated querying and integration across datasets, fostering interoperability and discovery.

This linked data pipeline represents a prototype for achieving data interoperability and standardisation. In building the prototype, ontologies and other conceptual spaces were selectively chosen to prioritise feasibility and implementation speed. It is important to note that in an ontology, terms never stand on their own but must be understood within their broader context, including the hierarchical structure in which they are embedded. For this prototype, we primarily relied on the National Cancer Institute Thesaurus (NCIt)[2] as integrated within the OBO Foundry, as it provided a practical starting point with sufficient coverage for the data at hand. However, further study is required to establish a more rigorous ontological foundation for this GBM Linked Data pipeline, ensuring a comprehensive and scalable approach to representing data semantics. This also necessitates the development of robust curation pipelines to continually align new and existing data with the chosen ontologies, maintaining consistency and fostering long-term interoperability across datasets.

2.4. Resources

Resources incorporated in the GBM Linked Data project included over twenty Excel tracker files to register patient metadata, sample description, sample sections, quality control (QC) results (RIN

[2]https://ncit.nci.nih.gov/

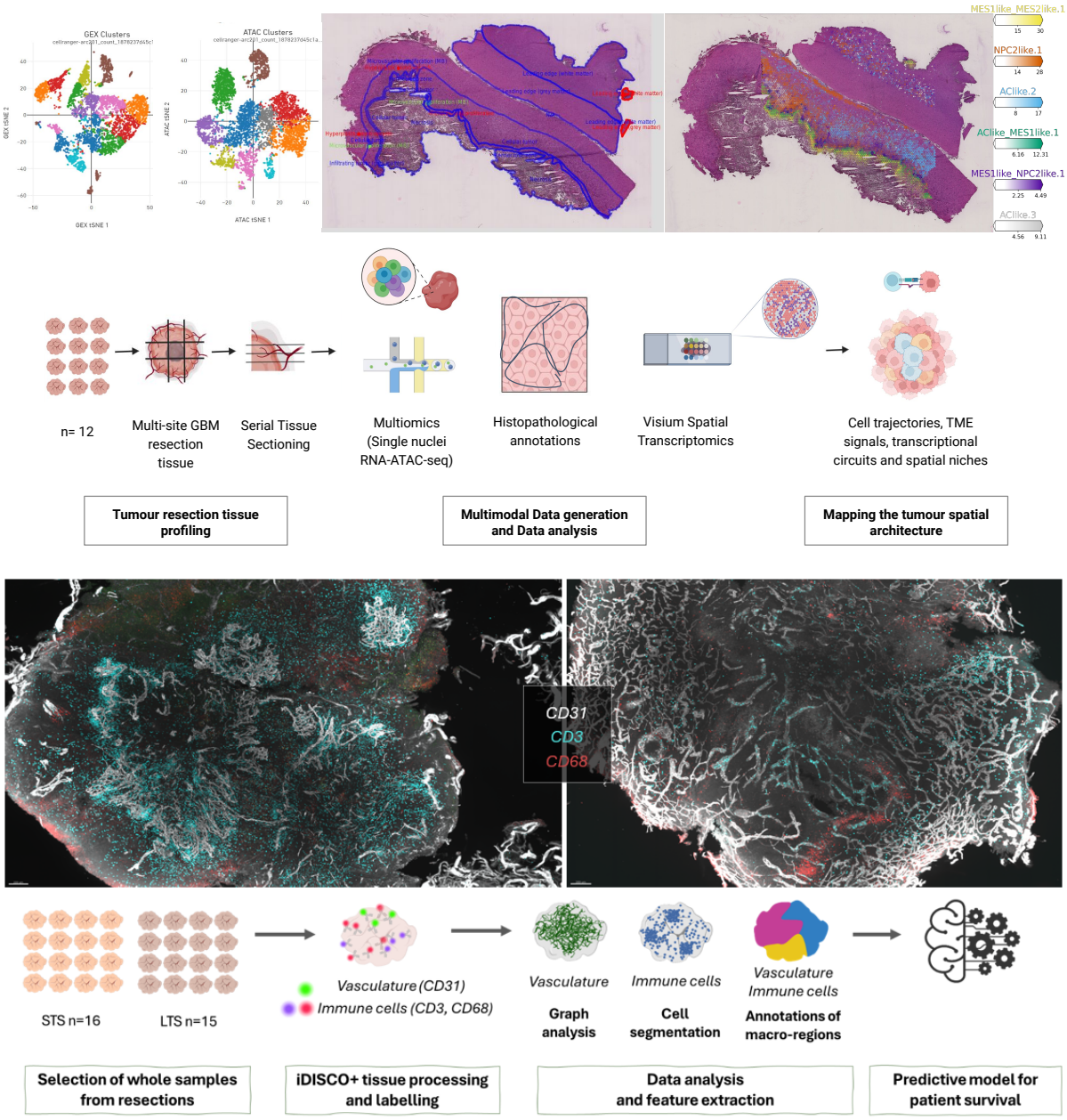

Figure 1: Indicative Experimental Datasets included the Linked Data pipeline with Schematic Representation of the research workflow and aims: Top: Sanger Institute Research Group, Bottom: Wyss Center Research Group

and H&E), sample processing, RNA/DNA libraries sequenced, data mapping, single cell data analysis (gene expression, cellular states, cell type), spatial transcriptomics descriptions (tissue niches) and neuropathology annotations (histological features of tissue sections).

2.4.1. Metatrackers

Metadata from the two research groups were collected in two Excel metadata trackers, referred to as 'Metatrackers'. Each Metatracker includes a curated selection of resources, integrating clinical, experimental, and analysed human glioblastoma (meta)data. Clinical (meta)data and genomic (meta)data were collected in one centralised Excel file and represented in tabular format. The data types supported

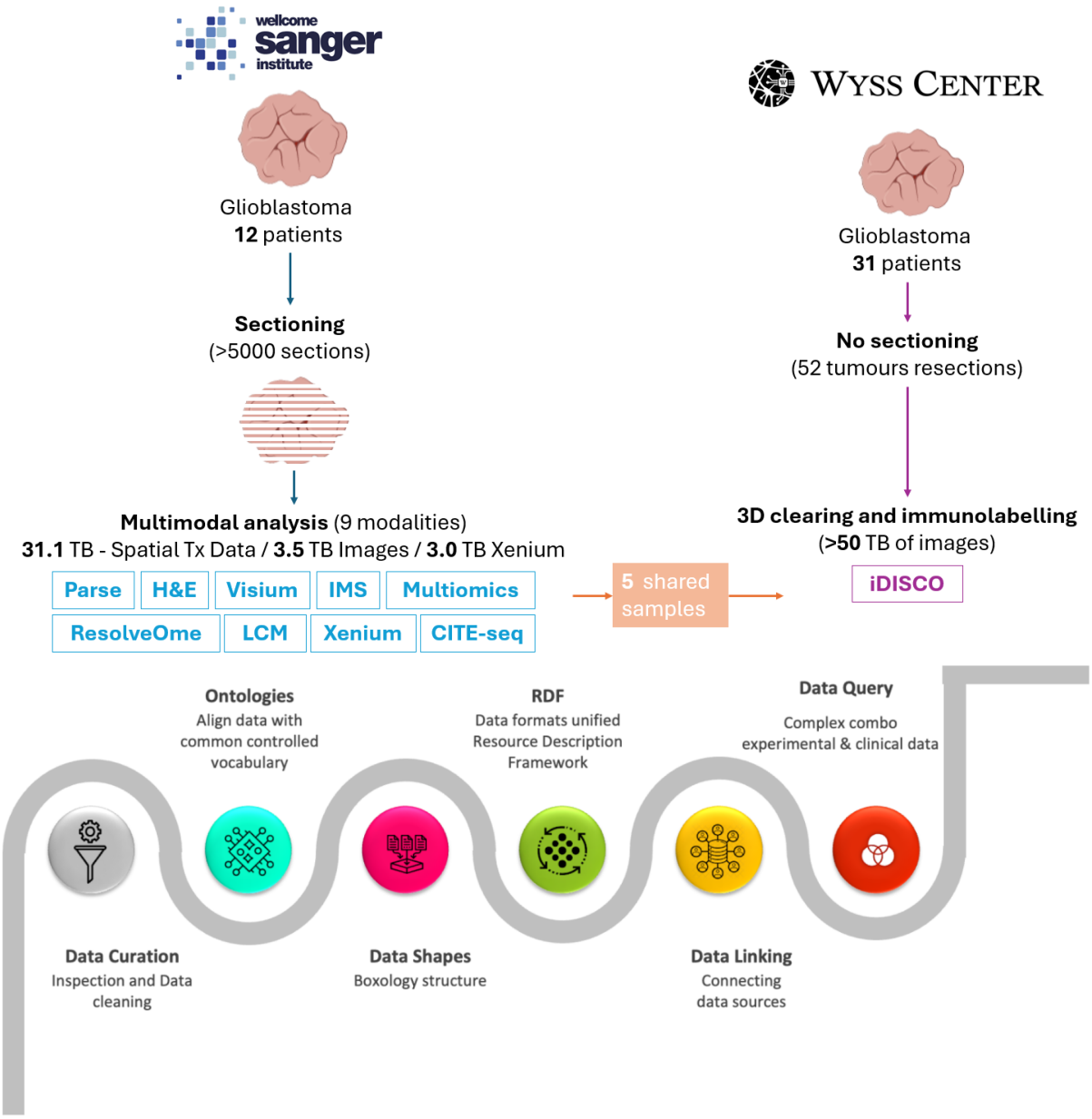

Figure 2: Flow diagram of sample processing (top) and schematic representation of the Glioblastoma Linked Data pipeline (bottom). Together, these figures illustrate the workflow and data integration structure for Glioblastoma research in DT Leap.

by the Metatracker were strings, integers, identifiers, decimals, dates, categorical values, or external links. Clinical (meta)data were stored in the Metatracker, while genomic (meta)data were stored either in the Metatracker or in data repositories such as Globus [5, 6] and OMERO [7]. Both Globus and OMERO adhered to the principles of the MARFA agreement described earlier, ensuring compliance with established standards for intellectual property, data sharing, and confidentiality, which streamlined interoperability and facilitated collaborative research.

In cases where genomic (meta)data were stored externally, we used the Metatracker to achieve interoperability by including shareable links to those data repositories. Each row in the Sanger Metatracker represents one biological tissue section, reflecting the experimental strategy of the project, where serial tissue sectioning was applied for tumour profiling. Each row in the WYSS Metatracker represents

Parameter	patient	patient_id	cohort	age_at_surgery
Format	string	string	categorical	integer
Shape	Subject	Subject	Subject	Subject
Unit				years
Delimiter				

Figure 3: The three first rows in the metatracker capture data type, data structure, and quantitative units of the (meta)data to support the transformation to RDF

one patient, reflecting the experimental strategy used. Each column in the Metatracker represents a different parameter, such as patient identifiers, clinical, or genomic (meta)data. These parameters were described according to the FAIR data principles by assigning specific data structures. The first three rows in both Metatrackers store information on the data type, data structure, and quantitative units of the (meta)data (Figure 3).

2.5. Data Organisation and formats

2.5.1. Data structures

Two biocurators from each research group identified data structures to describe all metadata and their relationships. The identification of these data structures relied on biological semantics, rather than data formats. Structures were represented schematically in the form of a boxology [8]. The boxology follows a three level hierarchical organisation. (Meta)data belonging to one data structure also belong to the parent data structure [9]. The three levels of organisation identified within our metadata were the following: 'Patient' comprising of clinical metadata on the 'subject', 'diagnosis', and 'therapy', a second level including the data structures 'Tumour', 'Sample', and 'Experiment'; these structures mainly describe specimen data, sample identifiers, and experimental processing details, and 'Output' to describe the result of different experimental assays in the format of images, protein markers, datasets and histopathological annotations. The experimental assays incorporated within our linked data were single-nuclei sequencing, spatial transcriptomics and 3D imaging. Deeper data structures for each of these assays, were identified from the parent 'output' data structure to describe respective assay metrics (Figure 4). The boxology is not rigid and additional data structures were incorporated, when required, to reflect the progress of the project. Following the identification of the data structures, they were assigned to the relevant metatracker entries manually.

2.5.2. Data Standardisation

Uniform Resource Identifiers (URIs) for the terms in the Metatrackers were identified and selected through the Ontology Lookup Service (OLS)[3] from the European Bioinformatics Institute. Most of our clinical metadata and protein markers were sufficiently defined through identifiers found in the NCIt, which was embedded within the OBO Foundry and served as the primary ontology for this prototype. In line with the the overall approach of this prototype, term selection prioritised practical coverage and biological semantics rather than a comprehensive ontological foundation. Biocurators assessed definitions independently and collaboratively, selecting those that best aligned with the (meta)data requirements of their respective research group.

While alignment in the selected URIs and data terminology facilitates data integration, our pipeline does not require complete alignment between the two research groups, maintaining flexibility to meet the distinct research needs of each. Selected URIs were also assessed for their contextual relevance within the ontology's class, subclass, and hierarchical taxonomy to ensure accurate representation of the (meta)data. We utilized the abbreviations and definitions outlined in the IVY Glioblastoma Atlas Project [10] and in controlled vocabularies such as SNOMED Clinical Terms (SNOMED CT) [11] to provide standardized definitions for histopathological features. However, transcriptomic assay quality control metrics were not adequately represented in existing medical ontologies. To address this gap, we

[3] http://www.ebi.ac.uk/ols

provided URIs and definitions from 10X Genomics[4], ensuring machine readability and FAIR compliance for these terms. Further work is needed to establish a more rigorous and unified ontological foundation for data interoperability in future iterations of the pipeline.

3. Results & discussion

3.1. Unifying Data Formats and Querying

The RDF schema[5], previously introduced as a standard for structuring and linking data on the Web, was used to unify the diverse data formats included in the two Metatrackers and associated data repositories. Using RDF serialisation and custom code, we ensured that data integration adhered to FAIR principles and facilitated sophisticated data querying. Similar to the data modelling approach employed by Wikidata, which leverages RDF for comprehensive data linking [12, 13], we translated the relationships within our metadata into RDF statements composed of a 'subject', a 'predicate', and an 'object'. In our case, both the 'subject' and 'object' could correspond to any of our defined data structures, such as a patient identifier or a diagnosis (Figure 4). The 'predicate' was used to define the relationship between the two, using descriptors such as 'has' or 'is'.

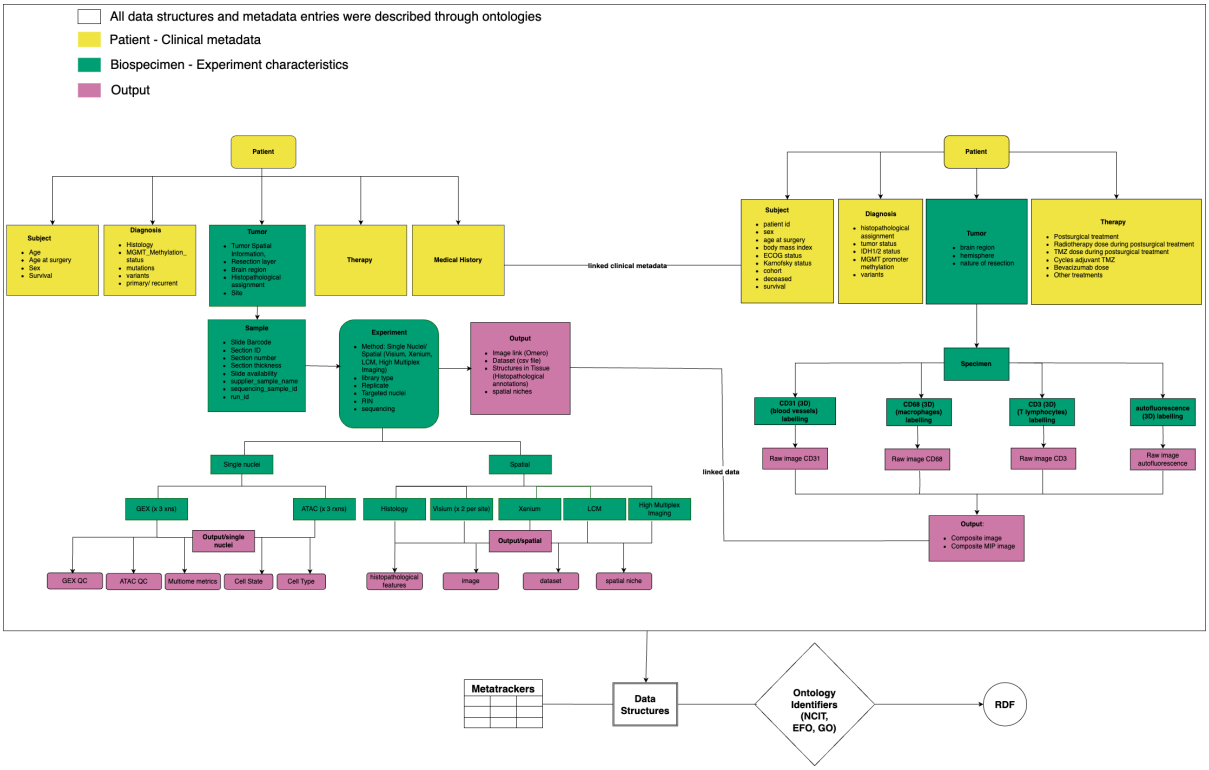

Figure 4: Visualising the Glioblastoma Linked Data pipeline as a boxology. This boxology represents the structures and the relationships between the (meta)data of the two research groups. The colour coding reflects the three level hierarchy of the data organisation. These data structures were used to describe all the metadata and data incorporated in the metatrackers and data repositories used in the project. The standardisation and interoperability were facilitated with assignment of ontology identifiers to all entries and translation into RDF.

This RDF-based representation provided a graph-based structure for linking clinical and genomic (meta)data, effectively creating an interoperable version of our boxology that also preserved relational information. Following RDF serialisation of our datasets, we tested the usability of the linked data by executing proof-of-concept queries using SPARQL, a query language designed specifically for interacting

[4] https://support.10xgenomics.com/
[5] https://www.w3.org/RDF

with RDF data [14]. These queries demonstrated the potential of our RDF-based approach for answering complex questions and enabling deeper data interoperability.

To evaluate the linked-data pipeline, we ran a set of example queries across three levels of querying: local data queries within an institute, collaborative queries across institutes, and exploratory queries using external SPARQL examples. This multilevel capability underscores the flexibility and interoperability of the pipeline (Table 2).

Table 2
Overview of SPARQL query examples applied at three levels of data integration: local (within a single institute), collaborative (across institutes via the DT Leap infrastructure), and global (leveraging external SPARQL endpoints). Each query demonstrates different stages of data availability and reuse, with the final example showcasing how external queries can inform local datasets. The symbols indicate the query results: + (query returned results), ? (query is ongoing or under review), and - (query returned no results).

Query Example	Local	DT Leap	Globally
Retrieve images of pseudopalisade cells around necrosis in patients with methylated MGMT promoter [6]	+	?	?
Find datasets containing CD31 in patients with a tumour in the fronto-temporal lobe aged 64 years	+	+	?
Identify CD68 gene variants predicting positive prognosis across cancer types	-	?	+

```
SELECT DISTINCT ?sangerimageUrl ?omeroUrl WHERE {
  ?diagnosis rdf:type <https://dtdr.sdsc.edu/dev/datashape/diagnosis> .
  ?diagnosis ns1:mgmt_methylation_status ?methylaiton # "methylated" .
  FILTER (CONTAINS(STR(?methylaiton), "PRESENT"))
  ?diagnosis ns2:subject ?subject .
  ?sample rdf:type <https://dtdr.sdsc.edu/dev/datashape/sample> .
  ?sample ns2:subject ?tumor .
  ?tumor ns2:subject ?subject .

  ?sample rdf:type <https://dtdr.sdsc.edu/dev/datashape/sample> .
  ?experiment ns2:subject ?sample .
  ?output rdf:type <https://dtdr.sdsc.edu/dev/datashape/output> .
  ?output ns2:subject ?experiment .
  ?output ns1:image ?sangerimageUrl .
  BIND(URI(REPLACE(str(?sangerimageUrl), "https://omeroplus.sanger.ac.uk/webclient/?show=image-",
 ?spatial rdf:type <https://dtdr.sdsc.edu/dev/datashape/output/spatial> .
 ?spatial ns1:structures_in_tissue_histopathological_annotations ?pseudopalisadesAnnotation .
 FILTER(CONTAINS(STR(?pseudopalisadesAnnotation), "pan"))
}
```

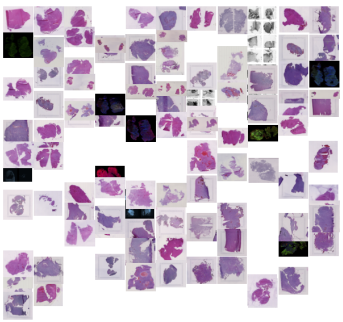

Figure 5: SPARQL query (left) used to retrieve images of pseudopalisade cells around necrosis, with an example result (right) retrieved from OMERO to quickly compare visual phenotypes.

3.2. Limitations and flexibility in concept mapping

While we successfully aligned many terms to existing ontologies such as NCIt, some terms could not be comprehensively mapped. This is expected, as our research operates at the forefront of science. Despite this, we were able to achieve interoperability by linking experimental modalities like Visium CytAssist and iDISCO data with clinical metadata, enabling seamless querying across diverse datasets and uncovering relationships such as spatial distribution of specific protein markers predictive of patient survival. This highlights the need for further work in setting up robust biocuration pipelines to ensure continuous alignment and integration of new and existing data with established ontological frameworks. However, the flexible nature of RDF or linked data provides us with a pragmatic solution that allows incremental improvement over time.

The RDF graph model supports a certain level of iterative development, enabling us to initially adopt a more lenient approach. For concepts that lacked suitable mappings, we used column names directly from the master tracker tables as placeholders within the RDF representation. While this approach

may appear simplistic, it allowed us to maintain interoperability and progress without delaying data integration.

A significant advantage of RDF-based systems is their ability to accommodate transformations between data shapes within the knowledge graph. When better mappings to ontologies or controlled vocabularies are identified, graph transformation methods such as SPARQL CONSTRUCT queries or other mapping tools can be applied. These transformations allow us to seamlessly refine the knowledge graph into more detailed and semantically rich representations without requiring a complete redraw of the linked-data structure.

This flexibility ensures that the knowledge graph evolves naturally alongside new insights and improved data models. Transforming RDF graphs is straightforward, preserving the integrity of the existing linked data while enabling the incorporation of new relationships, concepts, and structures. Thus, the linked-data pipeline remains adaptable to advancements in data curation and knowledge discovery, offering a sustainable path for continuous improvement.

3.3. Conclusion

Here we have demonstrated a work-case of applying the FAIR data principles to facilitate the integration of complex clinical and genomic data for glioblastoma research produced for 2 research institutes. While we have achieved complete interoperability and standardisation of our clinical metadata, the complexity of genomic data renders their full integration and querying an ongoing effort. The linked data pipeline we implemented and the data structures we identified set strong foundations in curation, interoperability and reuse of clinical and experimental data across different diseases.

Acknowledgments

We gratefully acknowledge the support of Wellcome Leap Delta Tissue, which made this work possible. We extend our heartfelt thanks to all participants from the three Program Demonstrator Areas—Tuberculosis, Triple Negative Breast Cancer, and Glioblastoma—for their insightful collaboration, valuable contributions, and shared expertise, which significantly enriched this study.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, et al., Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in pdgfra, idh1, egfr, and nf1, Cancer cell 17 (2010) 98–110.

[2] C. Neftel, J. Laffy, M. G. Filbin, T. Hara, M. E. Shore, G. J. Rahme, A. R. Richman, D. Silverbush, M. L. Shaw, C. M. Hebert, et al., An integrative model of cellular states, plasticity, and genetics for glioblastoma, Cell 178 (2019) 835–849.

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S. A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci Data 3 (2016) 160018.

[4] V. é, P. Krauss, K. Gnodtke, J. Buchhorn, D. Unni, P. Horki, J. L. Raisaro, K. Kalt, D. Teixeira, K. Crameri, S. sterle, FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network, Sci Data 10 (2023) 127.
[5] I. Foster, Globus online: Accelerating and democratizing science through cloud-based services, IEEE Internet Computing 15 (2011) 70–73.
[6] B. Allen, J. Bresnahan, L. Childers, I. Foster, G. Kandaswamy, R. Kettimuthu, J. Kordas, M. Link, S. Martin, K. Pickett, et al., Software as a service for data scientists, Communications of the ACM 55 (2012) 81–88.
[7] C. Allan, J.-M. Burel, J. Moore, C. Blackburn, M. Linkert, S. Loynton, D. MacDonald, W. J. Moore, C. Neves, A. Patterson, et al., OMERO: flexible, model-driven data management for experimental biology, Nature methods 9 (2012) 245–253.
[8] F. Van Harmelen, A. Ten Teije, A boxology of design patterns for hybrid learning and reasoning systems, Journal of Web Engineering 18 (2019) 97–123.
[9] R. Ciric, W. H. Thompson, R. Lorenz, M. Goncalves, E. E. MacNicol, C. J. Markiewicz, Y. O. Halchenko, S. S. Ghosh, K. J. Gorgolewski, R. A. Poldrack, et al., Templateflow: Fair-sharing of multi-scale, multi-species brain models, Nature Methods 19 (2022) 1568–1571.
[10] R. B. Puchalski, N. Shah, J. Miller, R. Dalley, S. R. Nomura, J.-G. Yoon, K. A. Smith, M. Lankerovich, D. Bertagnolli, K. Bickley, et al., An anatomic transcriptional atlas of human glioblastoma, Science 360 (2018) 660–663.
[11] S. El-Sappagh, F. Franda, F. Ali, K.-S. Kwak, Snomed ct standard ontology based on the ontology for general medical science, BMC medical informatics and decision making 18 (2018) 1–19.
[12] A. Waagmeester, G. Stupp, S. Burgstaller-Muehlbacher, B. M. Good, M. Griffith, O. L. Griffith, K. Hanspers, H. Hermjakob, T. S. Hudson, K. Hybiske, et al., Wikidata as a knowledge graph for the life sciences, Elife 9 (2020) e52614.
[13] A. Jacobsen, A. Waagmeester, R. Kaliyaperumal, G. S. Stupp, L. M Schriml, M. Thompson, A. I Su, M. Roos, Wikidata as an intuitive resource towards semantic data modeling in data fairification (2018).
[14] T. Neumann, G. Weikum, Scalable join processing on very large rdf graphs, in: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, 2009, pp. 627–640.