DatLI: The Data Licensing Initiative

DATA HAS CHANGED THE WORLD – DATA LICENSING HASN'T KEPT PACE

Data is changing the world. In every sector from agriculture to healthcare, data has emerged as a transformational asset. Computing and analytics have rapidly altered the landscape of what it means to use data, thereby straining the legal frameworks that govern it. Research and industry alike are in need of new legal frameworks that facilitate data access and redistribution rights while securing accountability, attribution, and the sustained availability of the merged and underlying data sources.

Licensing of public data is a credit hack. When academic institutions put a license on data they serve publicly, it is usually to ensure proper attribution; however, licensing is not the best tool for this task.

Licensing of private data is thorny. Private data such as medical records and wearables is not generated with research in mind; some process must transform it into a 'research dataset.' That process can include de-identification, removal of data that raises business concerns of the data owner, and approval of the data owner to use that data for research.

A STEP TOWARDS A SOLUTION

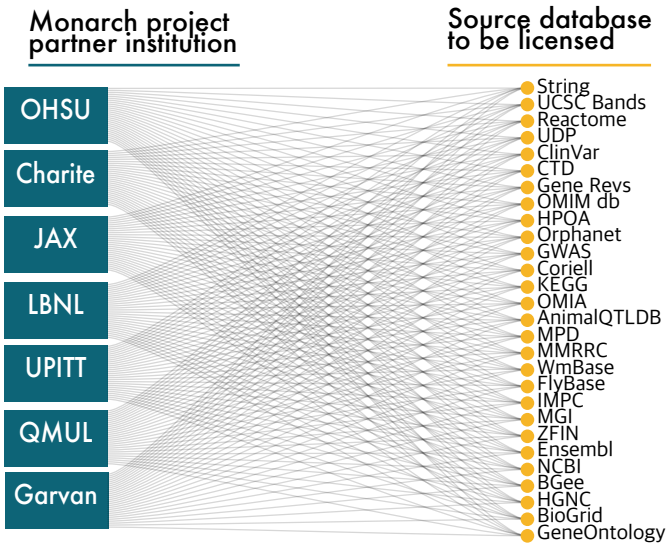

Fig 1) An example licensing burden for 1 project

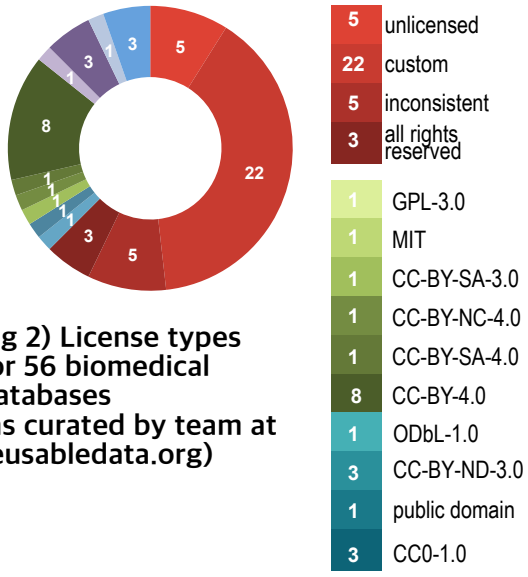

Fig 2) License types for 56 biomedical databases (as curated by team at reusabledata.org)

The newly formed Data Licensing Initiative[1](DatLI) hopes to do for data sharing what Uniform Biological Material Transfer Agreement (UBMTA) has done for materials sharing. Like research materials, data sources are distributed; maintaining them requires on-going resources and making them available is not always simple, e.g., clinical data. However, unlike most research materials, data science often requires combining dozens of separately-licensed components. The current legal frameworks for data were developed for the pairwise model (one-license-one-user); such frameworks do not scale to the set of many-licenses-many-users (MxN) transactions required (**Fig 1**). Moreover, there is little consistency between existing data licenses (**Fig 2**), and virtually no clarity regarding whether the data can be lawfully combined and redistributed under any license at all. To prevent transaction friction from overwhelming data access and reuse efforts, we need a new type of license that can garner broad adoption.

Thus, data reuse and redistribution are complicating factors in achieving the full impact of research involving "third-party" data sets, especially as the number of data sets needed escalates. Addressing redistribution and reuse, modification, metadata, data integrity and clean-up will require a broad collaborative effort in designing approaches and implementations for data access that meet the needs of the research community.

JOIN US

Over the course of the coming year, the Data Licensing Initiative seeks to provide the broader community with detail on the issues, advantages and disadvantages of common approaches, and draft guidance and model implementations for their consideration. Preliminary efforts assessing the current data licensing landscape of "open" data resources are available at the partner Reusable Data Project at ResusableData.org.[2]

DatLI a community of data scientists, researchers, academic technology transfer and legal experts is open to anyone; please let us know how you would like to participate or whether you would just like to stay informed: **bit.ly/datli-join**

Grant Acknowledgements : [1] NCATS: CD2H U24TR002306; [2] NCATS translator: 1OT3TR002019 & Monarch: 5R24OD011883