Genestack

A Formula for Valuing Life Sciences Data: **Making the Intangible Tangible**

Genestack 2024

Summary

Introduction.. 2

Data has Value? 2

Why People Don't Care but Should 3

Getting your data ready for valuation............................4

Valuing your data.. 6

Internal Value of Data (IVD) 7

External Value of Data (EVD) 9

Depreciation of Data Value.. 11

Depreciation of Internal Value of Data (IVD) 12

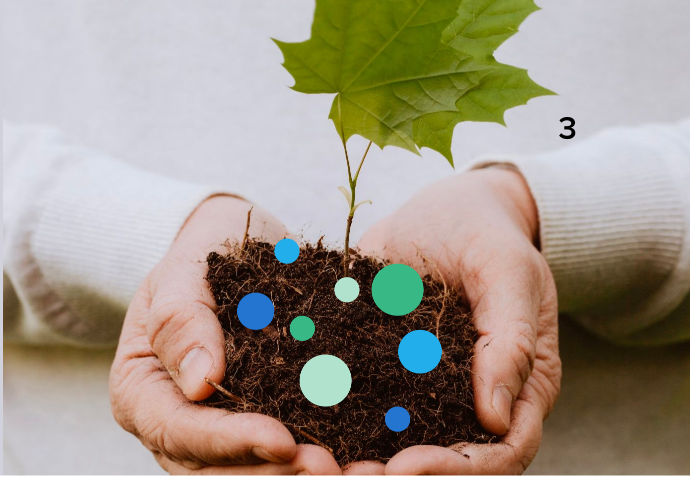
Depreciation of External Value of Data (EVD) 13

Conclusions.. 14

Data Has Value?

Everyone who creates data knows that it has value, otherwise it wouldn't be created, right? The value attributed to it however is often in what it is capable of understanding rather than actual financial value. This can be challenging when trying to justify funding to create further or manage existing data. After all, from a financial perspective, why would you invest more money in something that in itself has no monetary value and may not result in any progress towards your goals? It is in effect an intangible asset, something that all financial professionals shy away from wherever possible.

Why do they shy away from intangible assets? Put simply it's their very nature that makes them problematic. They're hard to value, high risk due to fluctuating market value and with a lack of actual physical presence by their very nature hard to pin a firm future value onto. With the right reassurance though they can make it onto the balance sheet and provide additional value to their owners.



Why People Don't Care But Should

That said, why don't data owners and producers within a company push harder for their data to be included on the balance sheet?

The reasons are varied but in many cases, they simply don't see it as important, their responsibility or in the spirit of good science. After all the scientific method is built upon a principle of sharing data openly for the common good. Unfortunately, much of the money spent in producing high-value data never reaches the public domain and instead remains unused, simply gathering dust on a virtual shelf.

So why should finance be keen to get data on the balance sheet and why should scientists producing data see it beneficial?

The key thing to remember to answer both of these questions is that in reality data does have value. In some cases a lot of value. For finance, this means additional listable assets and a better justification for spending or even company valuation. For scientists, it means being able to point to the actual monetary value of the data produced that year, making it easier to justify your next budget or funding rounds to produce more. For both groups, it should be noted that many companies are now starting to list data as an asset on their balance sheets already - putting you at a distinct disadvantage.

So how do we calculate this value of data?

The first thing to say is that there is no correct or recognised standard way to calculate the value of data, particularly complex life science data. In this report, we will lay out our suggested framework and formula for calculating the Fair Market Value (FMV) and overall data asset value.

Getting Your Data Ready For Valuation

In reality you simply can't value something you can't see or quantify. If your data is scattered across various private drives, cloud storage or, dare I even say, USB drives then you're not going to be able to accurately value it. Nor will the asset be able to be listed on the balance sheet as during any audit you would not be able to easily point to the asset's location.

The first step therefore is data cataloging which is often a challenging task. What does your system need to be able to do to be useful for valuation though?

Gather
Produce a Single Point of Data Truth

Curate
Set up Data Governance & Curate your Data Consistently

Protect
Secure your Data in Suitable Traceable Storage

Data Preparation Steps:

1. Be able to accurately see all data produced in one place.

Accurate is the key word here. You need to be able to run one search and have 100% confidence that the result returned represents the entire data set. This allows you to use data volume as one of the factors in your formula.

2. Be able to confidently and consistently search your data based on accurate metadata.

This comes down to strong data governance policies and enforcement. You need to ensure that the search you're doing for data on a particular topic or subcategory (for example a particular disease or species subtype) is 100% complete and accurate. You therefore need a mechanism for ensuring all data produced is cataloged against the same ontologies and vocabularies.

3. Know where all your data is stored and ensure it is suitably protected.

If you owned a valuable diamond and ruby ring, you wouldn't carelessly toss it into a drawer and lose track of it. You'd keep it secure and know its location. Similarly, your data should be protected to prevent accidental deletion or harmful changes to its metadata, rendering it worthless. Implementing a GXP-compliant storage and cataloging system with reliable backups is essential.

Valuing Your Data

We suggest that data has 2 intrinsic values, internal and external.

The Internal Value represents the time and money invested but also saved through its reuse and applications to progress your current or future pipeline. Even if the data is never used again, it still has value in the story of how it contributed to the journey of your new drug or variant's to the market.

The External Value represents the market value that, should you ever choose to sell that data, your data could achieve. This will be based on the cost and ease of creation and the demand for the data in the market. For example, if the data is very expensive to produce and in a research area with significant demand then the value will be higher than if it is cheap and easy to produce and in an area only you are working in it will have little value externally.

It is important to note that these two values are independent of each other. Data can have high internal value, but not external value, and vice versa. Nonetheless, it does have value that should be represented on your balance sheet.

How to calculate those values though. Here are our suggested formulae for calculating the internal and external value of your data.

Internal Value of Data (IVD)

The Internal Value represents the time and money invested but also saved through its reuse and applications to progress your current or future pipeline.

$$IVD=(C+S+P)\times R$$

C = **Cost of Data Creation:** This should be a sum of the following factors at minimum:

- Personnel wages for the research and set-up time (Annual salaries/time), Personnel wages for the running of the samples (Annual salaries/time),
- Share of equipment costs (number of samples in your data set/Average number of samples run per year),
- Cost of consumables used to run the samples,
- Personnel wages for analysis and data handling (Annual salaries over time),
- Share of cost of data management solution (Number of samples in your data/Data management cost).

  This can be especially tricky if your data management solution is internally constructed. We suggest taking into account the wages of those personnel involved in the production and maintenance of the solution as well as any hardware and storage costs. With a commercial vendor it is straightforward to point at the fraction of the license fee.

S = Savings through Reuse: Estimated savings achieved by using this data in multiple projects or applications. Our suggested formula for calculating this is:

(Number of researchers working in the same research area internally) X (Your C value to represent the cost of having to recreate the data)

P = Contribution to Pipeline Progress: This can be quantified by evaluating how the data accelerated or improved the development process. Our suggested formula for calculating this:

(Total number of checkpoints or phases that projects have moved through that use this data) X (Number of projects this data was used in)

This is a particularly tricky aspect to calculate as it depends on how your organization calculates progress. Our suggestion for organizations that have formal developmental checkpoints at each stage (for example, R&D, preclinical/translational, phase 0, phase 1, phase 2, phase 3) is to use the total number of these the project has moved through that the data contributed to. This allows the value to increase as data contributes to more successful projects over time and limits the value for those that fail to progress. For those organizations not using such formal checkpoints it could be another key success metric like number of papers or OKR progress.

R = Relevance Factor: A coefficient that represents the relevance of the data over time. For instance, data that continues to be relevant and useful over several years would have a higher relevance factor. Our suggested multipliers for this are:

- 2 X for research areas that are your organization’s primary research focus that is unlikely to ever change.
- 1.5 X research areas that are your organization’s secondary research focus that may not continue in future.
- 1 X for projects that were a one-off or sit outside of your organization’s normal area of focus.

External Value of Data (EVD)

The External Value of Data focuses on its potential market value, considering the cost of creation, market demand, and uniqueness. The formula might look like:

$$EVD=(C\times D)/E$$

C = Cost of Data Creation: As above, the total cost involved in generating the data.

D = Market Demand Factor: A coefficient representing the demand for such data in the market. High demand increases the value.

There's a plethora of ways to generate this value and it has to make sense for the market that your data is applicable to.

For example, if you're working in human disease we suggest using:

(Global disease prevalence/Number of drugs on the market for treatment)/Number of datasets available publically for the disease

E = Ease of Replication: This factor inversely affects the value. Data that is difficult and expensive to replicate (due to unique methodologies, rare expertise, etc.) has a higher value.

The subject nature of this number can be alleviated to a degree with the following suggested formula:

C x Ex/ PD

Where:
C = total cost of production, as above
Ex = Expertise potential.

This is relatively subjective and could be estimated potentially as follows:

Expertise Rarity/Difficulty to Replicate	Description	Ex Value
High	Use of unique technology that no/few other organizations possess, unique proprietary analysis methods and/or expertise in a highly niche area of research.	0.5
Medium	Applications of existing technology in a unique manner, unique or rare analytical methodologies and/or a decade or more expertise in a complex but common area of research	1
Low	Application of common technologies with standard analytical methodologies and/or expertise in a common and well understood area of research.	1.5

Depreciation of Data Value

There is one factor that I can hear all finance professionals screaming at this point. What about depreciation? You can of course apply depreciation to these assets the same as you can any other asset. Depreciation of data, both for Internal Value of Data (IVD) and External Value of Data (EVD), can be challenging to calculate because, unlike physical assets, the value of data doesn't necessarily decrease linearly over time. However, a method can be devised based on the expected useful life of the data and its relevance over time. Here are suggestions for calculating depreciation for IVD and EVD:

Depreciation of Internal Value of Data (IVD)

Since IVD is based on the utility of data within an organization, its depreciation can be linked to the diminishing relevance or utility of the data over time. A simple method could be straight-line depreciation over the estimated useful life of the data.

Annual Depreciation (IVD) = IVD/Estimated Useful Life (Years)

Where:

IVD: As previously calculated.

Estimated Useful Life: The number of years the data is expected to be relevant and useful for internal purposes. This can really only be estimated based on your internal knowledge of future direction and strategy.

Alternatively, if the utility of data decreases more rapidly in the initial years and then levels off, a more accelerated depreciation method like double-declining balance could be appropriate.

Depreciation of External Value of Data (EVD)

The depreciation of EVD should consider market demand and the ease of replication, which might change over time.

Annual Depreciation (EVD) =
EVD/Estimated Market Relevance Period (Years)

EVD: As previously calculated.

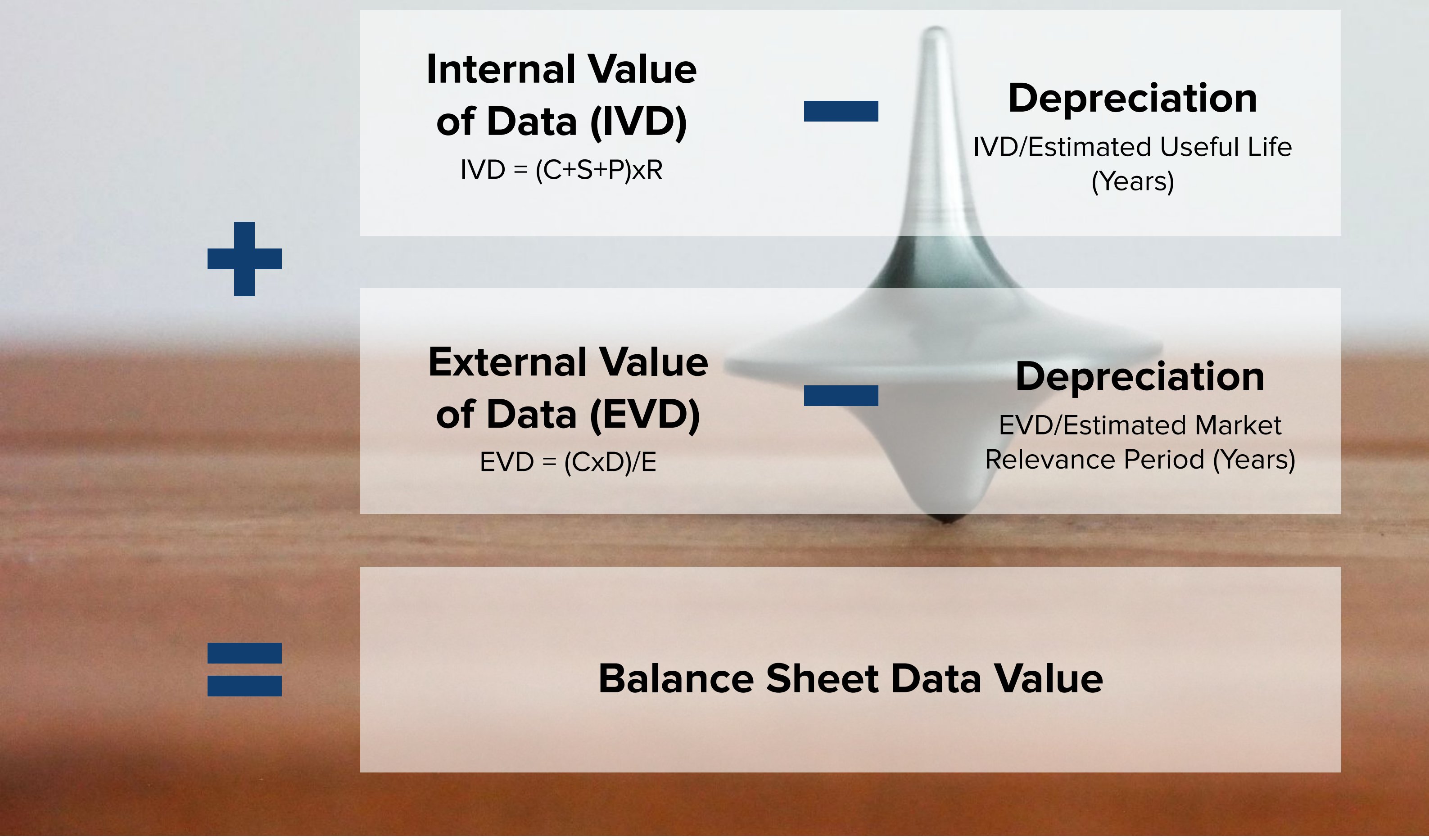
Estimated Market Relevance Period: The number of years the data is expected to hold value in the external market.
This is subjective, but external research can partly validate your estimates. For instance, in cancer research, your data might remain valuable for a significant time, though technological advances could render it obsolete. It's wise to start with a 10-year value estimate and adjust annually based on the technology's evolving usage that produced the data.

The depreciation rate might need to be adjusted if market demand shifts significantly due to technological advancements, changes in research focus, or other external factors. For example, recent developments in AI technology applications may actually reinvigorate the value of previously depreciated data due to the need for large training sets for model establishment.

In both cases, it's important to regularly reevaluate the estimated useful life and market relevance period, as these can change based on new developments or changes in the field. Additionally, for more accurate depreciation calculations, organizations might consider more complex methods that better reflect the unique usage patterns and obsolescence rates of their data.

Remember, these methods are conceptual and might need to be adapted or refined to suit the specific characteristics and contexts of the data being evaluated.

Conclusions



Applying these formulae to your data as part of your standard budget and company balance sheet processes can allow you to transform the intangible assets of data into something somewhat more tangible. That tangibility does rely however on good data management to ensure that your data is properly captured, curated and secured.

A user-friendly and automatable data management system will make that part of your standard practices relatively easy. If you'd like to see how easy then let us know and we can show you how our Open Data Manager can help you make these valuations a breeze.

Contact us at sales@genestack.com
or learn more at www.Genestack.com

Genestack