

Longevity Biotech Data Challenges

Longevity Biotech Fellowship consortium “surveyed ~400 participants across various sectors of longevity”.

Surprisingly, the most wanted solution was greater availability of large public datasets.

10 DATA CHALLENGES IN LONGEVITY BIOTECH

1. **Fragmented Data Silos and Lack of Integration** : Curbs interoperability, hinders data management, and prevents developing a full-view picture.
2. **Lack of Metadata and Provenance Information** : Decreases trust in the data and prevents reproducibility and audit.
3. **Measurement and Variable Heterogeneity** : There are many, often confounding and latent, variables. They are measured in different ways. This restricts comparability and composability.
4. **Scarcity of Longitudinal Data** : Prevents tracking the trajectory of aging and long-term interventions.
5. **Limited Real-World Evidence for Translation** : Withholds knowledge about the intervention’s translatability into real-world situations.
6. **Data Quality** : Recording errors, noise, batch effect, drift, and calibration issues.
7. **Data Availability Decay** : No public archives for the data of older studies.
8. **Privacy and Regulation** : Sensitive information introduces liability and risk constraints. Regulation introduces governance requirements and complexity.
9. **Lack of Consistent and Standardized Biomarkers** : Prohibits the comparison of studies and the demonstration of their efficacy.
10. **Non-random Missingness** : Studies on older populations cannot track whether participants deceased or did not follow up.

QUESTIONS YOU CAN ASK TODAY

- How many data sources do you ingest? How varied are their formats and metadata standards?
- What is your current data-to-analytics latency? Where are the largest time-blocks?
- How do you capture and track metadata and provenance (sample ID, batch, vendor, run date, instrument)? Is the lineage end-to-end visible and auditable?
- Which data modalities (omics, imaging, wearables, clinical) are hardest to integrate? What is missing: format alignment or metadata completeness?
- If you had “one wish” to fix a data-pipeline bottleneck, what would it be (e.g., vendor harmonization, automated ingestion, better QC, cost-reduction)?

<https://www.linkedin.com/in/andrasnovoszath/>