# 🗺️ LPB Data Mapping Guide - Converting Your Existing Data

## Common Data Mapping Scenarios

This guide helps you map your existing breeding data to the LPB format requirements.

---

## 1. Breeding Lines Data Mapping

### Your Data Might Look Like:

```csv

Line\_ID,Entry\_Number,Program,Year\_Planted,Stage,Index\_Score,Parent\_A,Parent\_B

ABC001,1234,High\_Rain,2023,F4,95.2,Mom\_1,Dad\_1

XYZ002,1235,Drought,2023,F5,88.7,Mom\_2,Dad\_2

```

### Maps to LPB `samples` Format:

```csv

sample\_id,gid,breeding\_program,year,development\_stage,selection\_index,parent1,parent2

MR1-0001,G1234,MR1,2023,F4,95.2,Mom\_1,Dad\_1

MR3-0002,G1235,MR3,2023,F5,88.7,Mom\_2,Dad\_2

```

### Field Mapping:

| Your Field | LPB Field | Conversion Notes |

|------------|-----------|------------------|

| `Line\_ID` | `sample\_id` | Add program prefix (MR1-, MR2-, etc.) |

| `Entry\_Number` | `gid` | Add "G" prefix if needed |

| `Program` | `breeding\_program` | Convert to: MR1, MR2, MR3, MR4 |

| `Year\_Planted` | `year` | Use as-is |

| `Stage` | `development\_stage` | Use standard values |

| `Index\_Score` | `selection\_index` | Use as-is |

### Program Name Conversions:

| Your Program Names | LPB Standard |

|-------------------|--------------|

| High\_Rain, Wet\_Zone, Disease\_Program | MR1 |

| Medium\_Rain, Balanced, Standard | MR2 |

| Drought, Low\_Rain, Stress\_Tolerant | MR3 |

| Irrigated, High\_Input, Premium | MR4 |

---

## 2. Trait Data Mapping

### Your Data Might Look Like:

```csv

Genotype,Trial,Year,Location,Yield\_tha,Disease\_Score,Drought\_Rating,Protein\_Percent

G1234,Trial\_A,2023,Site\_1,42.5,7.8,6.2,12.8

G1235,Trial\_B,2023,Site\_2,38.9,8.1,7.5,11.9

```

### Maps to LPB `phenotypes` Format:

```csv

GID,Trait,BLUE,Year,Environment,Breeding\_Program

G1234,yield,42.5,2023,MR1\_HighRainfall,MR1

G1234,disease\_resistance,78,2023,MR1\_HighRainfall,MR1

G1234,drought\_tolerance,62,2023,MR1\_HighRainfall,MR1

G1234,protein\_content,12.8,2023,MR1\_HighRainfall,MR1

```

### Trait Name Conversions:

| Your Trait Names | LPB Standard | Conversion |

|------------------|--------------|------------|

| Yield\_tha, Grain\_Yield, GY | `yield` | Use as-is (t/ha) |

| Disease\_Score, Disease\_Rating | `disease\_resistance` | Convert scale: (score/10)\*100 |

| Drought\_Rating, Stress\_Score | `drought\_tolerance` | Convert scale: (score/10)\*100 |

| Protein\_Percent, Protein | `protein\_content` | Use as-is (%) |

| Test\_Weight, TW | `test\_weight` | Use as-is (kg/hl) |

| Plant\_Height, Height | `plant\_height` | Use as-is (cm) |

| Days\_to\_Maturity, DTM | `maturity\_days` | Use as-is (days) |

### Scale Conversions:

| Original Scale | LPB Scale | Formula |

|----------------|-----------|---------|

| 1-9 scale | 1-100 scale | `(value/9)\*100` |

| 0-10 scale | 1-100 scale | `(value/10)\*100` |

| 1-5 scale | 1-100 scale | `(value/5)\*100` |

---

## 3. Genomic Data Mapping

### Your Data Might Look Like:

```csv

Sample,Marker\_Set,Chr,Position\_cM,Allele,QTL\_Effect

ABC001,Set\_1,1A,45.2,A,2.3

ABC001,Set\_2,1B,67.8,T,-1.1

```

### Maps to LPB `haplotypes` Format:

```csv

haplotype\_id,chromosome,position,markers,breeding\_value,effect\_size,program\_origin

HAP\_1A\_001,1A,0.45,SNP\_Set\_1,45.2,2.3,MR1

HAP\_1B\_001,1B,0.68,SNP\_Set\_2,43.1,-1.1,MR1

```

### Genomic Field Mapping:

| Your Field | LPB Field | Conversion |

|------------|-----------|------------|

| `Sample` | Use for `program\_origin` | Map to MR1-MR4 |

| `Chr` | `chromosome` | Use standard format (1A, 1B, etc.) |

| `Position\_cM` | `position` | Convert to 0-1 scale: `position\_cM/max\_cM` |

| `Marker\_Set` | `markers` | Combine related markers |

| `QTL\_Effect` | `effect\_size` | Use as-is |

---

## 4. Environment/Location Mapping

### Your Locations → LPB Environments:

| Your Location Names | LPB Environment | Program |

|--------------------|-----------------|---------|

| Wet\_Site, High\_Rain\_Location, Site\_A | `MR1\_HighRainfall` | MR1 |

| Medium\_Site, Balanced\_Location, Site\_B | `MR2\_MediumRainfall` | MR2 |

| Dry\_Site, Drought\_Location, Site\_C | `MR3\_LowRainfall` | MR3 |

| Irrigated\_Site, Premium\_Location, Site\_D | `MR4\_Irrigated` | MR4 |

---

## 5. Data Quality Mapping

### Your Quality Indicators → LPB Standards:

| Your Quality Terms | LPB Standard |

|-------------------|--------------|

| Good, Excellent, A, 1 | `High` |

| Fair, Average, B, 2 | `Medium` |

| Poor, Low, C, 3 | `Low` |

---

## 6. Date Format Conversions

### Common Date Formats → LPB Standard:

| Your Format | Example | LPB Format | Example |

|-------------|---------|------------|---------|

| DD/MM/YYYY | 15/05/2023 | YYYY-MM-DD | 2023-05-15 |

| MM/DD/YYYY | 05/15/2023 | YYYY-MM-DD | 2023-05-15 |

| DD-MMM-YYYY | 15-May-2023 | YYYY-MM-DD | 2023-05-15 |

| YYYY/MM/DD | 2023/05/15 | YYYY-MM-DD | 2023-05-15 |

---

## 7. ID Standardization

### Creating Consistent IDs:

#### Sample IDs:

```

Your Format: ABC001, XYZ123, Trial\_A\_001

LPB Format: MR1-0001, MR2-0123, MR3-0001

Pattern: [PROGRAM]-[SEQUENTIAL\_NUMBER]

```

#### Germplasm IDs:

```

Your Format: 1234, Entry\_001, Line\_A

LPB Format: G1234, G0001, G0123

Pattern: G[NUMBER]

```

#### Haplotype IDs:

```

Your Format: Marker\_1A\_001, SNP\_Set\_A

LPB Format: LR862530.1\_chr\_1A-1-150

Pattern: [REFERENCE]\_chr\_[CHR]-[BLOCK]-[POSITION]

```

---

## 8. Excel to CSV Conversion

### Steps to Convert Excel Files:

1. \*\*Open your Excel file\*\*

2. \*\*Select the data sheet\*\*

3. \*\*File → Save As → CSV (Comma delimited)\*\*

4. \*\*Check encoding is UTF-8\*\*

5. \*\*Verify decimal separators are dots (.)\*\*

### Excel Formula Examples:

#### Program Conversion:

```excel

=IF(A2="High\_Rain","MR1",IF(A2="Drought","MR3","MR2"))

```

#### Scale Conversion (1-9 to 1-100):

```excel

=(B2/9)\*100

```

#### Date Conversion:

```excel

=TEXT(C2,"YYYY-MM-DD")

```

---

## 9. Quick Conversion Checklist

### Before Converting Your Data:

- [ ] \*\*Identify\*\* all your data files and formats

- [ ] \*\*Map\*\* your field names to LPB requirements

- [ ] \*\*Plan\*\* ID standardization strategy

- [ ] \*\*Convert\*\* trait scales if needed

- [ ] \*\*Standardize\*\* program names to MR1-MR4

- [ ] \*\*Format\*\* dates to YYYY-MM-DD

- [ ] \*\*Check\*\* for missing critical fields

### After Converting:

- [ ] \*\*Validate\*\* using the validation script

- [ ] \*\*Check\*\* sample counts match

- [ ] \*\*Verify\*\* all required fields populated

- [ ] \*\*Test\*\* with small dataset first

- [ ] \*\*Backup\*\* original data files

---

## 10. Common Conversion Tools

### Programming Solutions:

```python

# Python pandas for data conversion

import pandas as pd

# Load your data

df = pd.read\_csv('your\_data.csv')

# Convert program names

program\_map = {'High\_Rain': 'MR1', 'Drought': 'MR3'}

df['breeding\_program'] = df['old\_program'].map(program\_map)

# Convert scales

df['disease\_resistance'] = (df['disease\_score'] / 9) \* 100

# Save in LPB format

df.to\_csv('lpb\_samples.csv', index=False)

```

### Excel Solutions:

- Use VLOOKUP for program mapping

- Use formulas for scale conversion

- Use CONCATENATE for ID creation

---

## Need Help?

### If Your Data Doesn't Match These Examples:

1. \*\*Document your current format\*\* clearly

2. \*\*Identify the closest mapping scenario\*\* above

3. \*\*Create a small test dataset\*\* (5-10 rows)

4. \*\*Run the validation script\*\* to check

5. \*\*Contact LPB support\*\* with specific questions

### Common Support Questions:

- "My trait scales are different" → Use conversion formulas

- "My program names don't match" → Create custom mapping

- "I have additional fields" → Add as optional columns

- "My IDs are different" → Follow standardization patterns

---

\*\*🧬 Once your data is properly mapped and validated, your LPB Advanced Breeding Intelligence platform will provide powerful insights into your MR1-MR4 breeding programs!\*\*