# 🚀 LPB Data Preparation - Quick Start Guide

## 📋 What You Need to Do

Convert your existing breeding data to match the LPB format so your \*\*Advanced Breeding Intelligence\*\* platform works perfectly with your real data.

---

## 🎯 Step-by-Step Process

### Step 1: Generate Templates (5 minutes)

```bash

# Run this to create templates and validation tools

python csv\_templates\_generator.py

```

\*\*Creates:\*\*

- CSV templates for all data types

- Validation script

- Documentation

### Step 2: Review Your Data Requirements (15 minutes)

\*\*Essential Tables\*\* (start with these):

- ✅ \*\*samples\*\* - Your breeding lines (MR1-MR4)

- ✅ \*\*phenotypes\*\* - Trait measurements

- ✅ \*\*breeding\_programs\*\* - Program settings

\*\*Optional Tables\*\* (add later):

- 🔶 \*\*haplotypes\*\* - Genomic data

- 🔶 \*\*market\_data\*\* - Economic info

- 🔶 \*\*weather\_data\*\* - Environmental data

### Step 3: Convert Your Data (30-60 minutes)

1. \*\*Open\*\* the CSV templates in `data\_templates/`

2. \*\*Map\*\* your data using the conversion guide

3. \*\*Format\*\* your data to match templates

4. \*\*Save\*\* as CSV files

### Step 4: Validate Your Data (5 minutes)

```bash

# Check your data quality

python validate\_lpb\_data.py

```

\*\*Fix any errors\*\* before proceeding.

### Step 5: Import to LPB System (10 minutes)

- Load validated CSV files into your database

- Configure breeding programs

- Test the dashboard

---

## 📊 Minimum Data Requirements

### \*\*To Get Started\*\* (Basic functionality):

```

├── samples.csv # 50+ breeding lines minimum

├── phenotypes.csv # Yield + 2-3 other traits minimum

└── breeding\_programs.json # MR1-MR4 configuration

```

### \*\*For Full Functionality\*\* (All features):

```

├── samples.csv # All your breeding lines

├── phenotypes.csv # All trait measurements

├── haplotypes.csv # Genomic marker data

├── haplotype\_assignments.csv # Genotype links

├── market\_data.csv # Economic data

├── weather\_data.csv # Environmental data

└── breeding\_programs.json # Program configuration

```

---

## 🗺️ Quick Data Mapping Reference

### Your Field Names → LPB Standard

| \*\*Your Data\*\* | \*\*LPB Field\*\* | \*\*Example Conversion\*\* |

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

| Line\_ID, Entry | `sample\_id` | ABC001 → MR1-0001 |

| Program, Zone | `breeding\_program` | High\_Rain → MR1 |

| Yield\_tha | `yield` | 42.5 → 42.5 |

| Disease\_Score | `disease\_resistance` | 7.8 → 78 (scale 1-10 to 1-100) |

| Drought\_Rating | `drought\_tolerance` | 6.2 → 62 (scale 1-10 to 1-100) |

### Program Name Mapping:

- \*\*High Rainfall/Wet Zone\*\* → `MR1`

- \*\*Medium Rainfall/Balanced\*\* → `MR2`

- \*\*Low Rainfall/Drought\*\* → `MR3`

- \*\*Irrigated/High Input\*\* → `MR4`

---

## ✅ Quality Checklist

### Before Converting:

- [ ] Identify all your data files

- [ ] Choose consistent ID format (MR1-0001, MR2-0001, etc.)

- [ ] Decide on trait scale conversions

- [ ] Plan missing data handling

### After Converting:

- [ ] All files saved as CSV with UTF-8 encoding

- [ ] Dates in YYYY-MM-DD format

- [ ] Program names are MR1, MR2, MR3, MR4

- [ ] No special characters in IDs

- [ ] Required fields are populated

### Before Import:

- [ ] Validation script runs without errors

- [ ] Sample counts match your expectations

- [ ] Test with small dataset first

- [ ] Backup original data files

---

## 🔧 Common Conversion Examples

### Excel Formula for Program Conversion:

```excel

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

```

### Python Script for Scale Conversion:

```python

import pandas as pd

# Load your data

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

# Convert 1-9 scale to 1-100 scale

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

# Save in LPB format

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

```

### Date Conversion (Excel):

```excel

=TEXT(A2,"YYYY-MM-DD")

```

---

## ⚡ Speed Tips

### If You're in a Hurry:

1. \*\*Start with samples + phenotypes only\*\* (core functionality)

2. \*\*Use small test dataset\*\* (10-20 lines) first

3. \*\*Convert one trait at a time\*\* to verify

4. \*\*Add genomic data later\*\* (optional for basic use)

### If You Have Complex Data:

1. \*\*Review the full documentation\*\* first

2. \*\*Create mapping spreadsheet\*\* for your fields

3. \*\*Validate frequently\*\* during conversion

4. \*\*Test import with subset\*\* before full dataset

---

## 🆘 Common Issues & Solutions

### Issue: "My trait scales are different"

\*\*Solution:\*\* Use conversion formulas (see mapping guide)

### Issue: "My program names don't match"

\*\*Solution:\*\* Create mapping table (High\_Rain → MR1, etc.)

### Issue: "Validation shows errors"

\*\*Solution:\*\* Check required fields and data types

### Issue: "Missing genomic data"

\*\*Solution:\*\* Start without it - add later when available

### Issue: "Too much data to convert manually"

\*\*Solution:\*\* Use Python/R scripts for bulk conversion

---

## 📚 Documentation Files

1. \*\*LPB\_Data\_Requirements.md\*\* - Complete specifications

2. \*\*Data\_Mapping\_Guide.md\*\* - Field conversion help

3. \*\*data\_templates/README.md\*\* - Template instructions

4. \*\*validate\_lpb\_data.py\*\* - Quality checking script

---

## 🎯 Success Metrics

### You'll Know You're Ready When:

- ✅ Validation script runs with 0 errors

- ✅ All required tables have data

- ✅ Sample IDs follow MR1-MR4 pattern

- ✅ Trait measurements look reasonable

- ✅ Programs are properly configured

### Expected Timeline:

- \*\*Simple data\*\*: 1-2 hours total

- \*\*Complex data\*\*: 4-8 hours total

- \*\*Large datasets\*\*: 1-2 days with scripting

---

## 🧬 What Happens Next?

Once your data is properly formatted and imported:

1. \*\*LPB Dashboard\*\* displays your real breeding data

2. \*\*AI Analytics\*\* provide insights on your programs

3. \*\*Performance Tracking\*\* shows trends and progress

4. \*\*Predictive Models\*\* forecast genetic gains

5. \*\*Economic Analysis\*\* optimizes resource allocation

\*\*Your LPB Advanced Breeding Intelligence platform will transform how you analyze and optimize your MR1-MR4 breeding programs!\*\*

---

## 🔗 Quick Links

- 📊 \*\*Start Here\*\*: Run `python csv\_templates\_generator.py`

- 📋 \*\*Need Help\*\*: Check Data\_Mapping\_Guide.md

- ✅ \*\*Validate\*\*: Run `python validate\_lpb\_data.py`

- 🧬 \*\*Go Live\*\*: Import to LPB dashboard

\*\*Questions? Your LPB support team is ready to help with data preparation!\*\*