ETL Best Practices
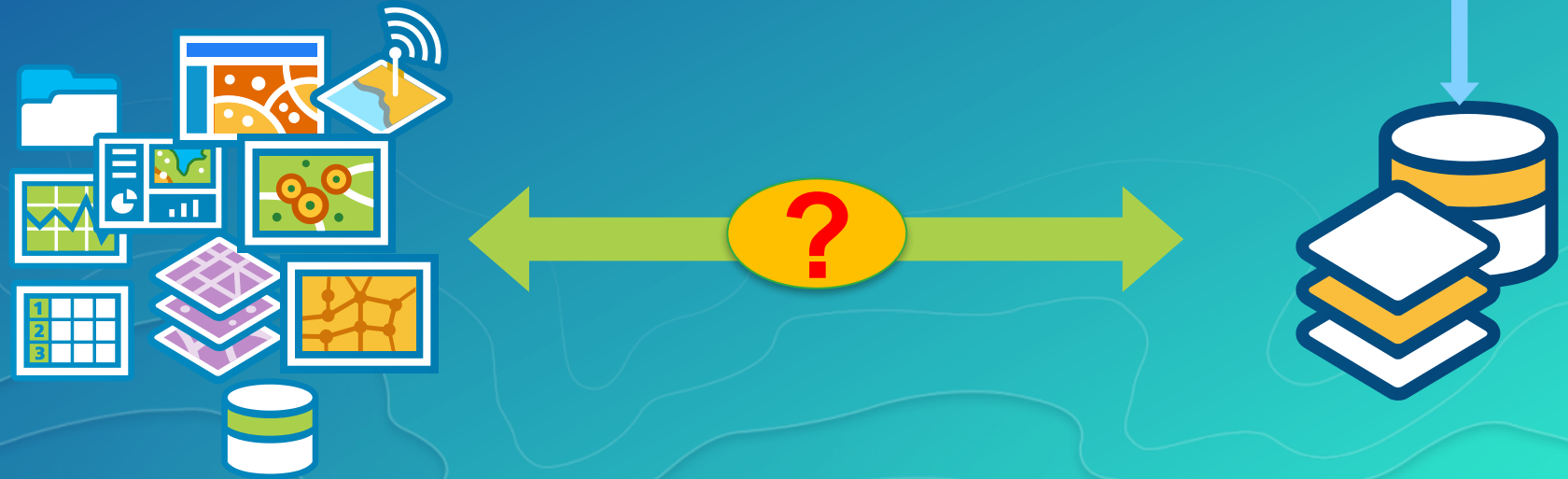
- Bala Mullachery

Agenda

- What is ETL?

- Why ETL?

- What are the Best Practices?

- How to ETL?

What is ETL ?

- ### E_{xtract} > T_{ransform} > L_{oad}
  - Extract – connect to data source(s) and withdraw data
  - Transform – manipulate data
  - Load – load data to the destination

What is ETL ?

- Extract > Transform > Load

  - Extract – connect to data source(s) and withdraw data

  - Transform – manipulate data
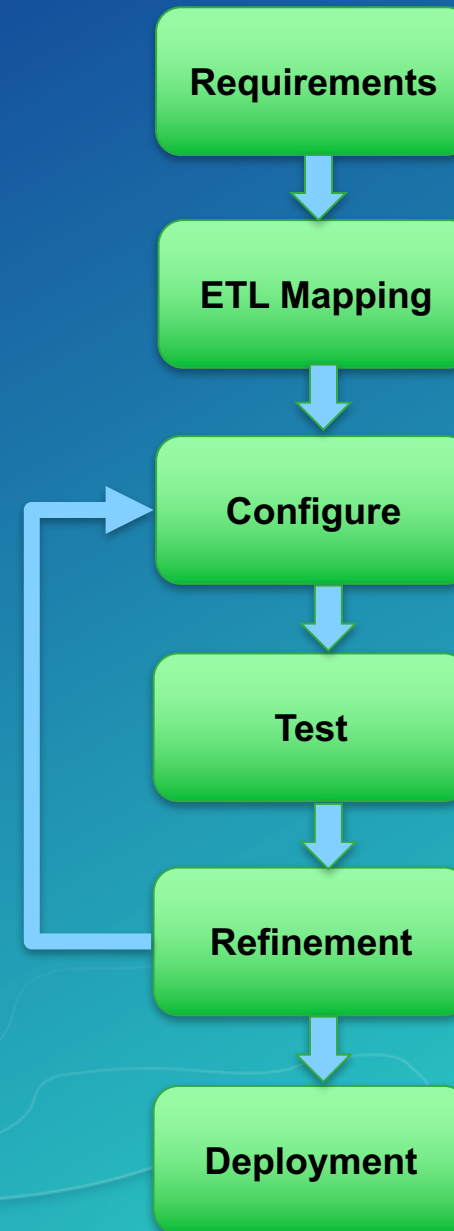
  - Load – load data to the destination

Why ETL ?

- Convert data from one format to another
  - CAD/GeoMedia/Oracle Spatial/Shapefile/CSV to geodatabase
- Export data from geodatabase for publication
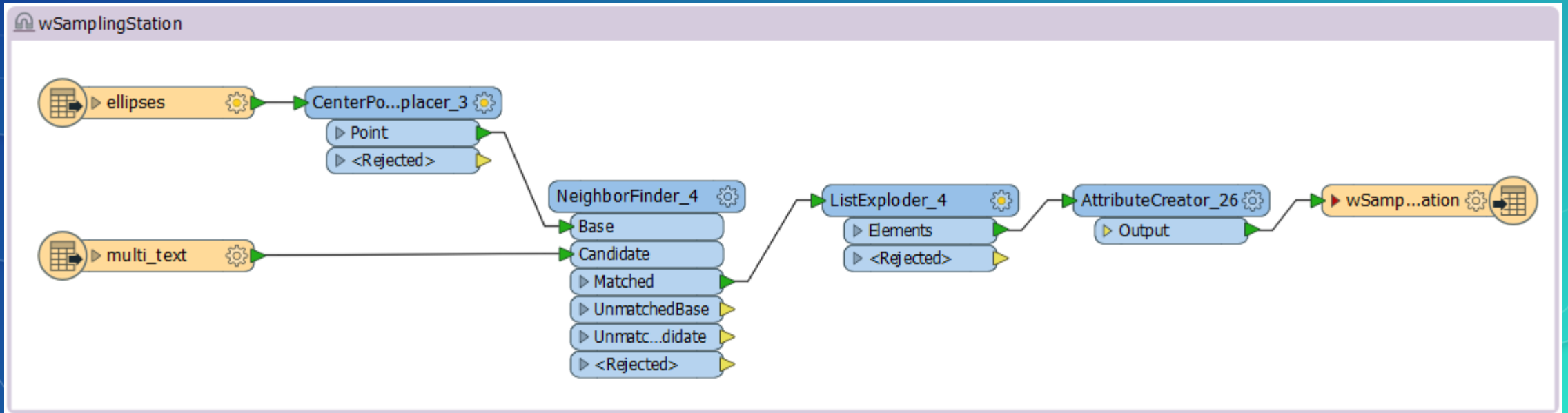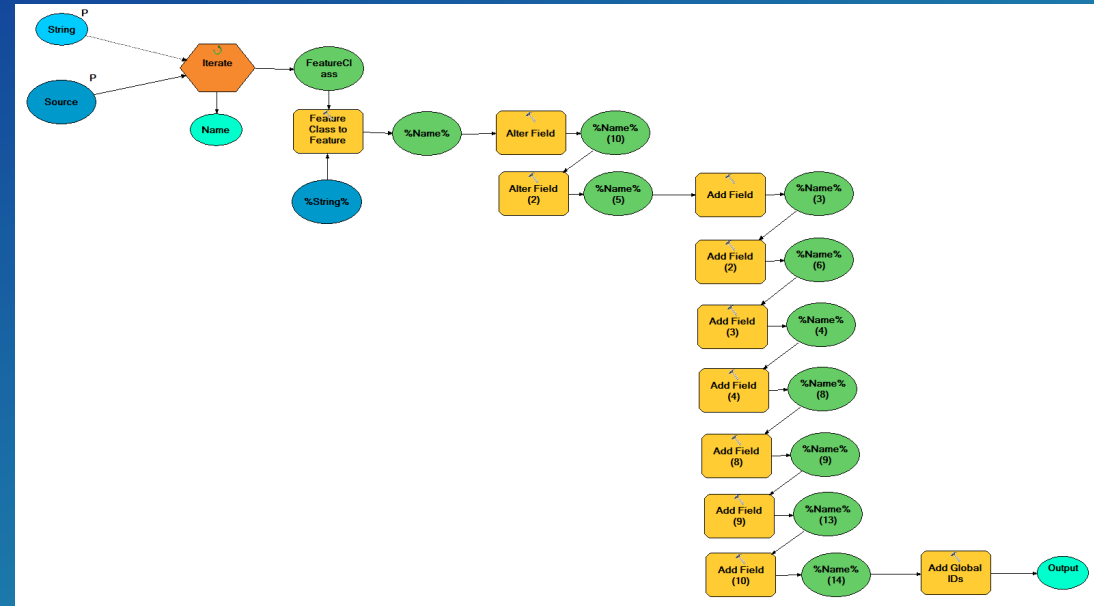 - Perform joins to external databases and create outputs

Recommended Implementation Approach

- Gather requirements
  - Source(s) data format
  - Destination data format
  - Destination data requirements
  - Source data anomalies
  - Feature editing and data enhancement
  - New data creation
- Prepare ETL mapping document
  - Source to Destination feature mapping
  - Feature data rules
- Configuration and Testing
  - Technology to use
- Deployment
  - On-demand
  - Batch processing / Scheduler

```
Requirements
     ↓
ETL Mapping
     ↓
Configure
     ↓
Test
     ↓
Refinement
     ↓
Deployment
```
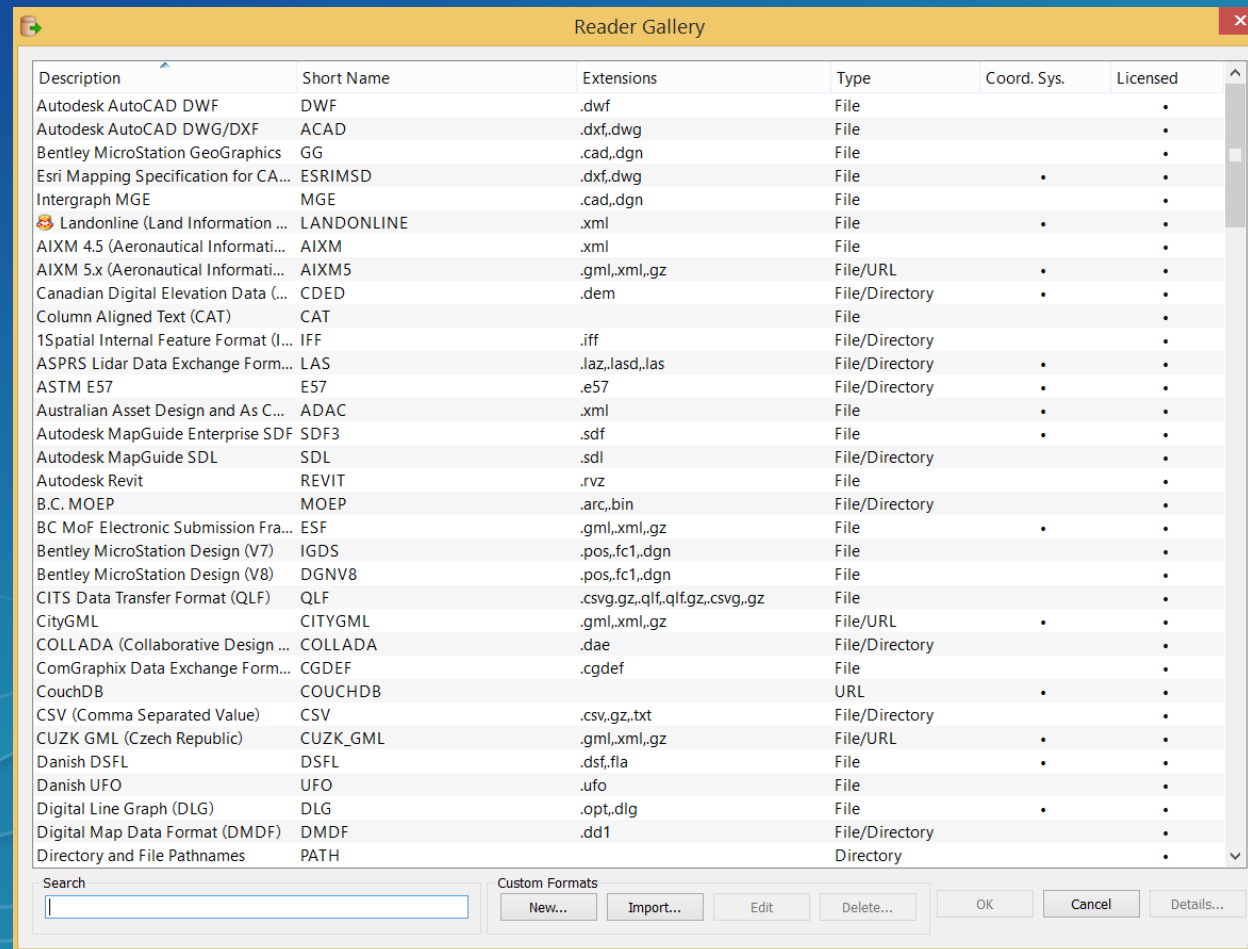
How to ETL ?

- ArcCatalog, ArcMap

- ArcGIS Model builder

- Geo-processing tools

- Python

- Interoperability

Data Interoperability

Extract

- Read source data existing in various format:
  - CAD
  - GIS/Spatial data
  - Tabular/Spreadsheet
  - Text/CSV

Data Interoperability

Transform
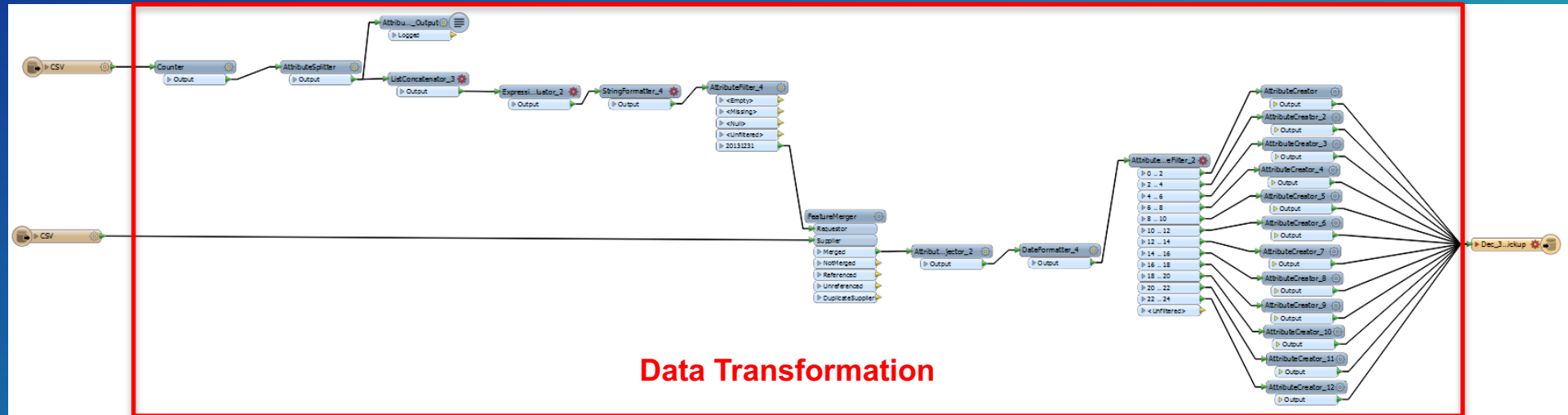
- Manipulate/Create data
  - Spatial data
    - Create new features
    - Topology cleanup
    - Feature matching
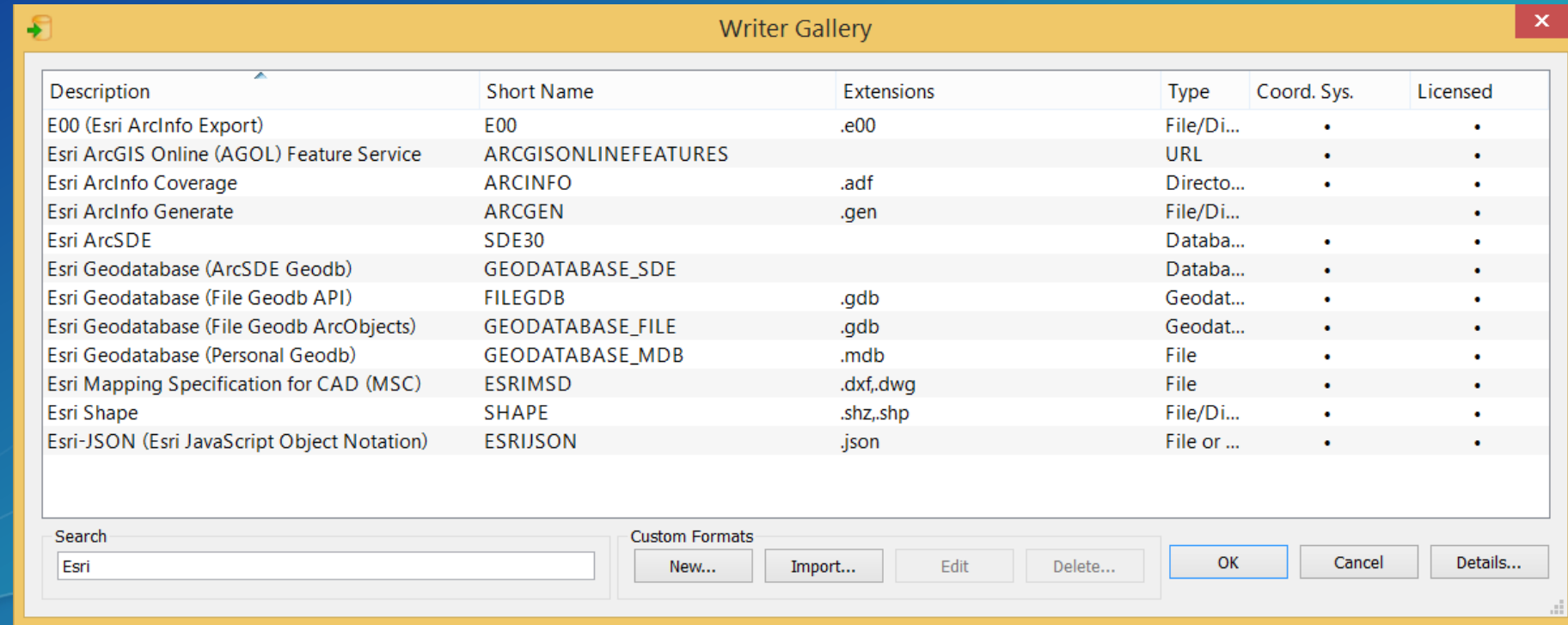   - Feature merger
    - Projection

  - Attribute data
  - Domain values
  - Unique identifier
  - Tables join
  - External Database Integration
  - SQL Query

Data Transformation
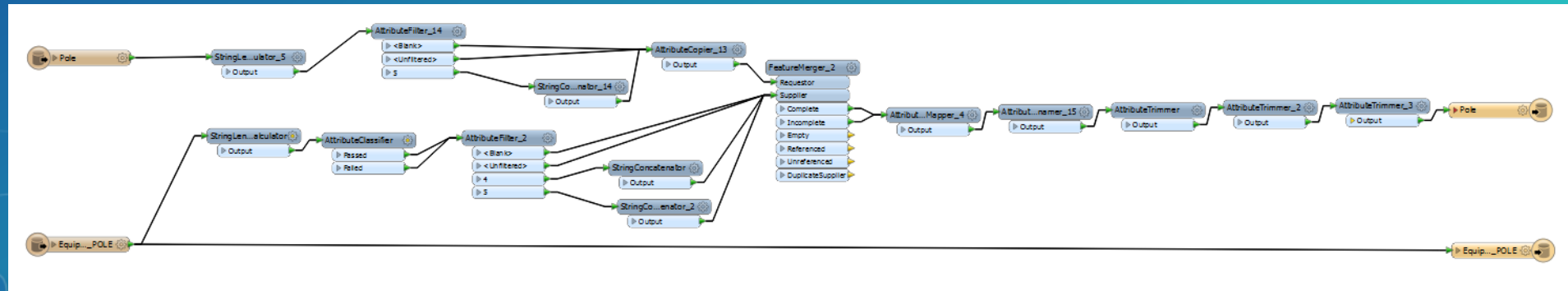
Data Interoperability

Load

- Load data into geodatabase
  - Format conversion
  - Transactional or non-transactional data
    - Feature classes
    - Tables
    - Annotations
    - Geometric networks

Writer Gallery

Description	Short Name	Extensions	Type	Coord. Sys.	Licensed
E00 (Esri ArcInfo Export)	E00	.e00	File/Di...	•	•
Esri ArcGIS Online (AGOL) Feature Service	ARCGISONLINEFEATURES		URL	•	•
Esri ArcInfo Coverage	ARCINFO	.adf	Directo...	•	•
Esri ArcInfo Generate	ARCGEN	.gen	File/Di...	•	•
Esri ArcSDE	SDE30		Databa...	•	•
Esri Geodatabase (ArcSDE Geodb)	GEODATABASE_SDE		Databa...	•	•
Esri Geodatabase (File Geodb API)	FILEGDB	.gdb	Geodat...	•	•
Esri Geodatabase (File Geodb ArcObjects)	GEODATABASE_FILE	.gdb	Geodat...	•	•
Esri Geodatabase (Personal Geodb)	GEODATABASE_MDB	.mdb	File	•	•
Esri Mapping Specification for CAD (MSC)	ESRIMSD	.dxf,.dwg	File	•	•
Esri Shape	SHAPE	.shz,.shp	File/Di...	•	•
Esri-JSON (Esri JavaScript Object Notation)	ESRIJSON	.json	File or ...	•	•

Search: Esri

Custom Formats: New... | Import... | Edit | Delete...

OK | Cancel | Details...
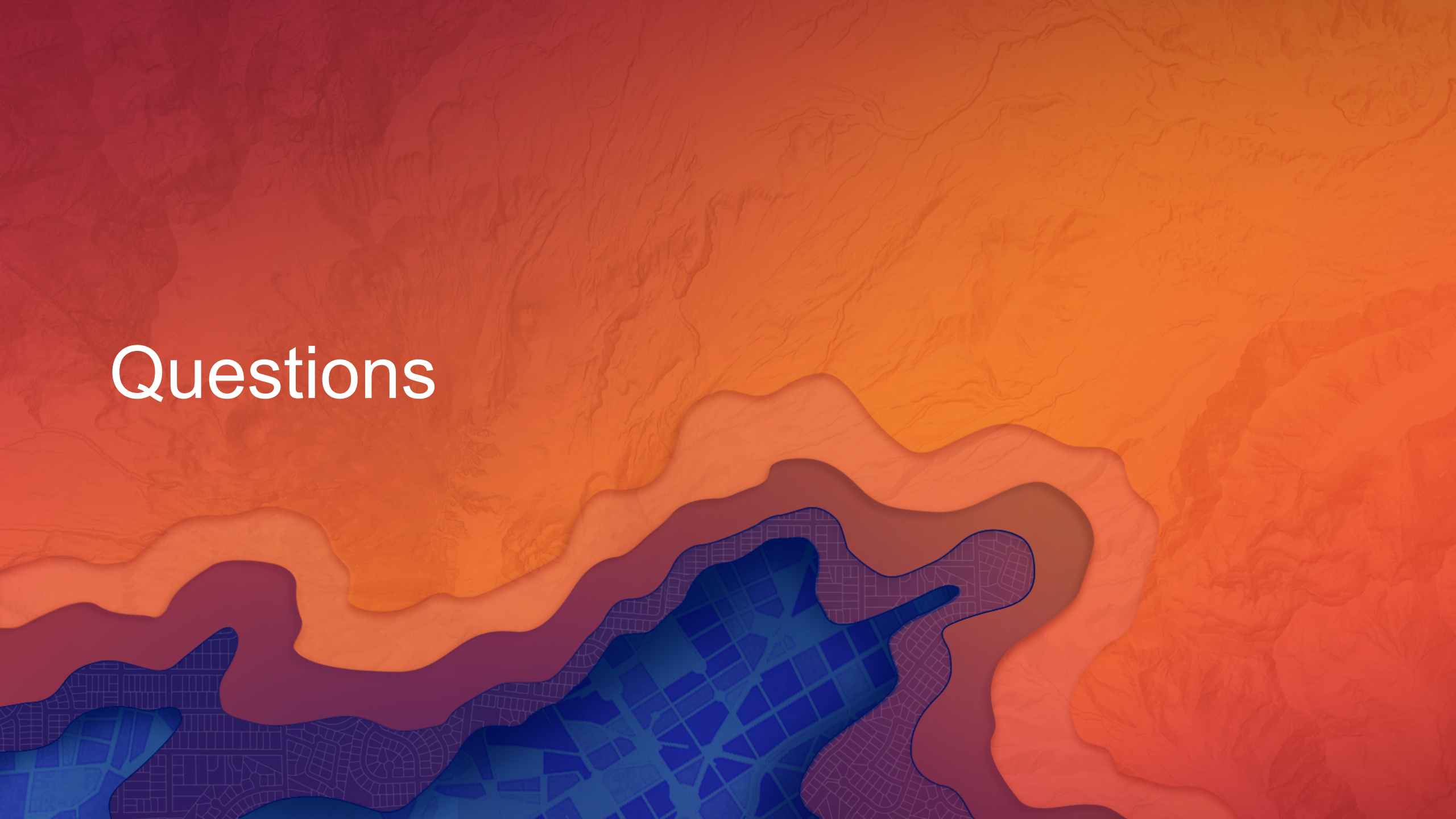
Test and Refinement

- Feature count
- Visual validation
- Automated validation using Data Reviewer
- Refinement
- Release for production
  - On-demand
  - Scheduled intervals

Best practices(re-cap)

- Gather requirements
- Develop ETL mapping
  - Source to destination
  - Rules
  - Source data issues
- Configure ETL
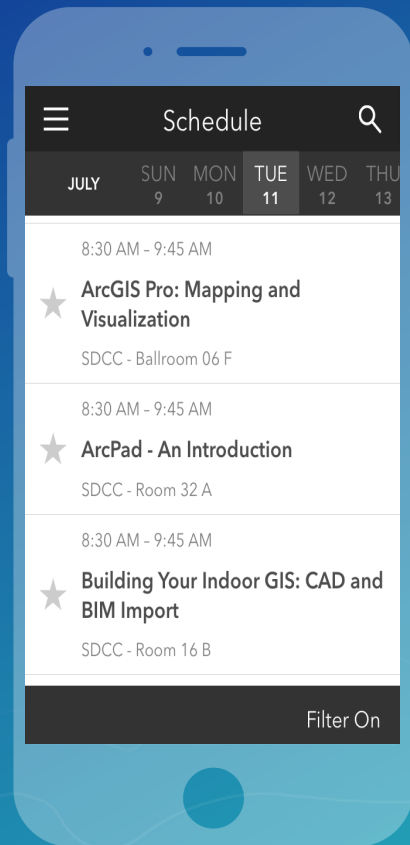- Test and Refinement
- Deployment Scenarios

Questions
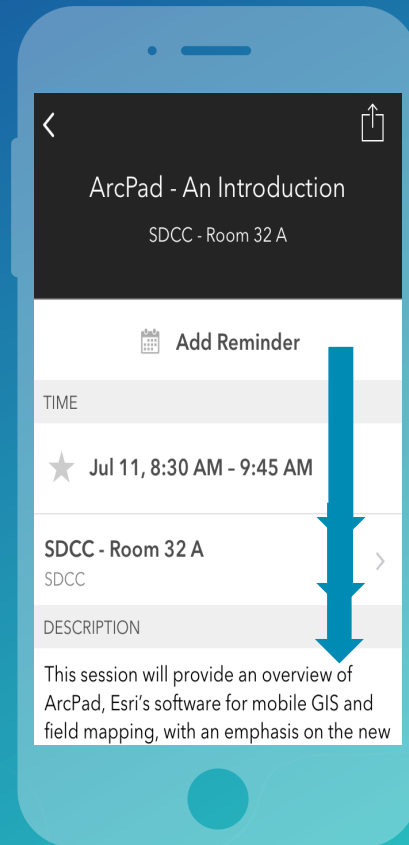
Please Take Our Survey on the **Esri Events App!**

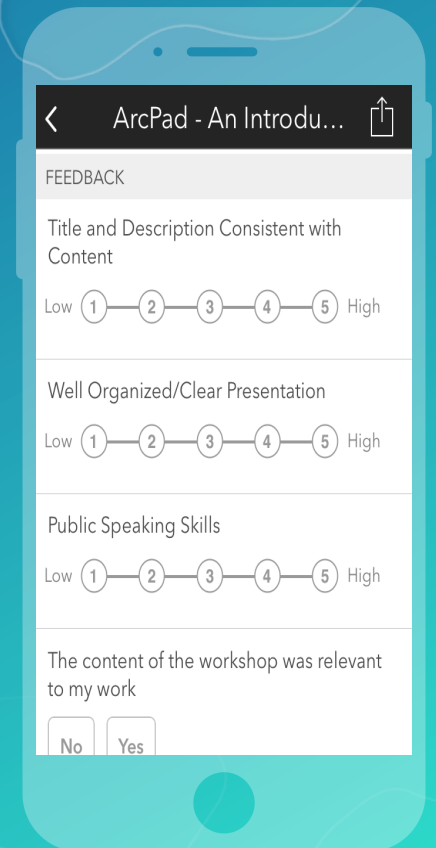**Download the Esri Events app and find your event**

Select the session you attended

Scroll down to find the survey

Complete Answers and Select "Submit"

esri | THE SCIENCE OF WHERE