WP3 Integration
Why integration?

• Many successful pilots

• Clear benefits

• Where are the production systems?
Two activities

• Short questionnaire
  • What are the barriers?
  • What’s useful?
  • 28 responses

• Long form investigation
  • How are NSO’s addressing key questions?
  • Responses from:
    • Australian Bureau of Statistics
    • UK Office of National Statistics
    • Statistics Flanders
    • US Bureau of Labor Statistics
Short Questionnaire – Who participated?

• 28 responses

• 14 regions/countries

• North America, Europe, and Australia
Which option best describes your role in your organization?

- Statistician / Data Scientist: 57%
- Analyst / Subject Matter Expert: 21%
- Manager / Policy Maker: 18%
- Software Engineer / Information Technology Specialist: 4%
Approximately how many employees work your organization?

- More than 2000: 50%
- Between 500 and 2000: 30%
- Between 50 and 500: 10%
- Less than 50: 5%
To what extent do the following organizational issues limit your organization’s ability to effectively use machine learning?

- Coordination between internal stakeholders (e.g. R&D, methodology, IT, subject-matter, operations, etc.)
  - Average Score: 1.7

- Resistance from stakeholders inside the organization (e.g. coworkers)
  - Average Score: 1.5

- Uncertainty over project ownership and responsibility
  - Average Score: 1.4

- Lack of clear organizational strategy
  - Average Score: 1.3

- Resistance from stakeholders outside of the organization (e.g. data users)
  - Average Score: 0.8
To what extent do the following technical issues limit your organization’s ability to effectively use machine learning?

- Availability of staff with appropriate machine learning algorithm skills: 1.8
- Access to suitable evaluation data: 1.6
- Availability of staff with appropriate programming skills: 1.6
- Access to suitable training data: 1.6
- Access to computer hardware: 1.5
- Access to computer software: 1.1

Average Score:
0 = does not limit use
4 = prevents use
How useful have the following activities been in helping your organization more effectively use machine learning?

- Collaboration with other statistical organizations: 2.3
- External training programs: 2.2
- Collaboration with academia: 2.0
- Internal training programs: 1.9
- Clarification of roles and responsibilities within the organization: 1.8
- Collaboration with private companies: 1.5

Average Score: 0=not useful, 3=very useful
Long form questions

1. Where should ML fit in a statistical organization?
2. What should the ML pipeline look like with regards to the organization structure?
3. What machine learning skills are needed and where are they needed in the organization?
4. How can organizations efficiently acquire the ML skills they need?
5. How should organizations demonstrate and communicate the value-added of ML techniques?
6. How should statistical organizations identify the right problems for ML?

https://statswiki.unece.org/display/MLP/WP3.+Integration
Where should ML fit in a statistical organization?

1. A multidisciplinary collaboration that usually starts out in methodology or another technical area?

2. A center of excellence that consults with other pieces of org?

3. Experimental branch of methodology?

4. Decentralized, largely owned by individual program offices?
What should the ML pipeline look like in regards to organizational structure?

1. Start with the business need, then move to R&D and then bring in other areas like IT

2. Start with a project to build ML experience which in turn makes it easier to identify solvable problems
What machine learning skills are needed and where are they needed in the organization?

In the multidisciplinary setting, ML requires new skills in many areas:

- Everyone must understand basics, and key terminology.
- Research and methodology must become familiar with new techniques and tools, like R and Python.
- IT must learn how to integrate these tools and processes in existing systems.
- Subject matter must understand their role in supporting and maintaining system.
- Management must understand the needs of ML teams.

Alternative: interdisciplinary roles (data scientists) and units.
How can organizations efficiently acquire the ML skills they need?

1. Popular strategy:
   • Permanently or temporarily acquire outside expert
   • Use them to teach internal courses

2. Other options:
   • Free online material (MOOCs, ArXiV, blogs, GitHub)
   • Academic coursework

3. Communities of practice (internal and external)

4. Research projects
How should organizations demonstrate and communicate the value-added of ML techniques?

1. Clear demonstrations of value-added are key
   - Not enough to build it, have to show how it’s better than alternative, sometimes harder than building
   - Don’t assume existing process is perfect, measure it too!

2. Use ML as decision-support, at least initially
   - Gives people opportunity to see it in action

3. ML where no other option is feasible
How should statistical organizations identify the right problems for ML?

1. Learn from others, what have they done that’s successful?

2. Look for tasks that meet ML friendly criteria
   1. Stable over time (relatively same task year to year)
   2. Lots of training data showing input to task and desired outcome
   3. Existing process is undesirable because of cost, speed, quality, etc.

3. Start with lightweight research projects, proof-of-concepts
How is your organization addressing these challenges?

1. Where does ML fit in your organization’s structure?
2. What does your ML pipeline look like with regards to the organization structure?
3. What machine learning skills are needed and where are they needed in the organization?
4. How is your organization acquiring the ML skills it needs?
5. How is your organization demonstrating and communicating the value-added of ML techniques?
6. How is your organization identifying the right problems for ML?