

# Synthesis

Carole Goble,  
The University of Manchester, UK & ELIXIR

With contributions from Ilkay Altintas, SDSC, USA  
Michael R. Crusoe, Common Workflow Language

# What is a Workflow? What is Automation?

The definition of workflow varies

- from “documenting process” to fully automated computational pipelines.
- Workflows at different points in the scientific process

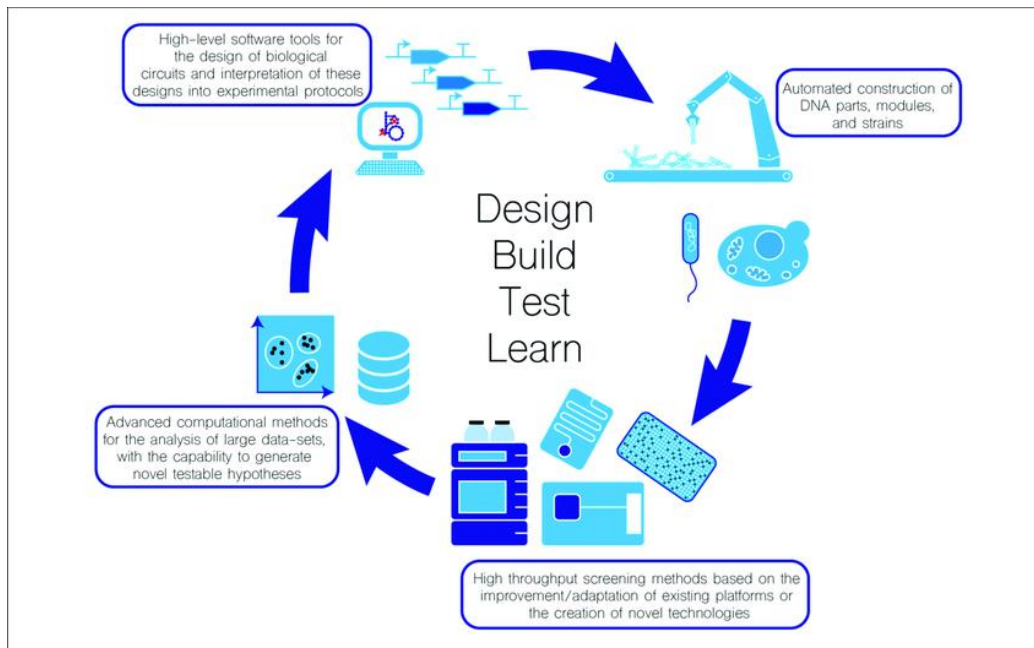
What we are aiming at with “automation” also needs carefully unpacking

- Smart Assisting Human/Team in the Loop?
- Smart Adapting what has been produced already?
- Smart creation/maintenance/interop of the ecosystem of resources the workflow works in? (the data, the tools, the instruments, the CI, the exchange formats, human)
- Smart model prediction using AI
- Smart automated feedback loops in the experimental process?
- Smart scale up/out

**Recommendation:** Definition of the “intelligence” expected from workflows changes depending on the definition of workflow - see Beth’s figure

# Prepare- Design - Build -Test - Learn - Scale

Understanding the scientific process and where workflows (and AI) fit is essential. We can learn from our own scientific method (from Syn Bio below...) and apply to how we do science generally. Eat our own dog food.



# Domain Readiness

Domains vary significantly in their **readiness** for automation and readiness for workflow take-on

Human in the Loop is a common requirement, throughout the workflow lifecycle - from design through to execution.

The availability of metadata and workflow-ready data is a ubiquitous issue. Some domains are more mature than others, depending on the necessary drivers for interoperability and knowledge exchange. Big collaborations (e.g. physics) needed to sort this. Others are more fragmented.

**Recommendation:** The problems being solved vary significantly. The need for AI-integrated workflow is different than the need for AI-integrated science.

# Domain Willingness

Domains vary in their **willingness** to adopt a workflow based approach and need to be convinced of the value. The economics of workflows means the cost is upfront for scale up further downstream.

If the value proposition (VP) of workflows is repetition and reuse and the value is a “public good” then that needs to be rewarded. Incentive schemes in science are for private competitive advantage rather than public cooperation.

The VP of workflows needs to be articulated, demonstrated and couched in terms of scientific productivity rather than reproducibility.

Case studies such as the widespread adoption of Galaxy (see UseGalaxy.eu) need to be explored and understood. Why are these so successful with workflow newbies?

**Recommendation:** Being “usable”, “useful” and “used” are different things. These should be treated as different parts of the workflow research, development and application.

We need to support and square the circle of the private and the public good

# Bridging across the tech - domain spectrum

Tech that can bridge people

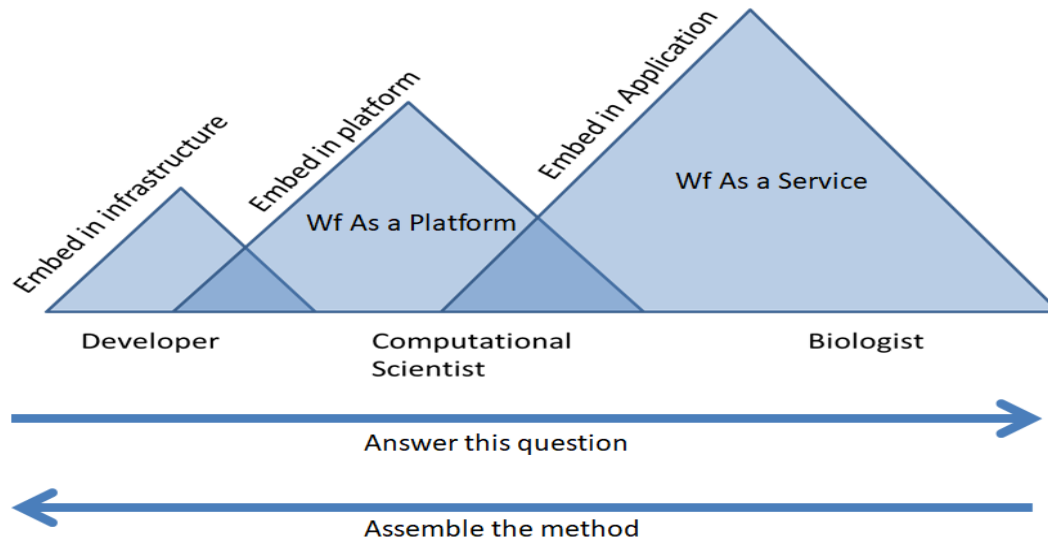
- so you can move across the user spectrum. Scale up and scale out are different.
- so different users can collaborate

People to bridge - the professionalisation of workflow engineering cf. Research Software Engineers; Bioinformaticians.

There are challenges with communicating across the spectrum (cf. Tim Gardners comments)

The economics of workflow making needs understanding.

- A workflow expertly and expensively made by experts for (re)use 1000s of times (in an instrument pipeline) or by 1000s of users (behind an application) is a good ROI, but how to bill the cost? (Re)use often comes together with the cost of running the workflow.
- A workflow lashed up for a small group with a short lifetime needs to be cheaper to make.



## Recommendation Priority:

Establishing the Research Workflow Engineer, professionalising Wf engineering

Invest in cross-disc teams for common production of workflows for communities

Cost-Benefit roadmap for workflow development/adoption/publishing with case studies

# Challenges to adoption

Barriers to adoption of novel or just modern automated techniques are PEST. Political, Economic, Social, Technical

- Cultural shifts are required in practice and in assessment
- Limitations arise from availability of data (incl where data is available access throttles, licenses)

The pathways for adoption of workflow approaches needs articulating, supporting and sustaining. Jupyter as an “entry level” is an approach, but the in ramp also needs a route map to the next ramp and the next, without having to chop and change and abandon previous ramps.

**Recommendation:** Different phases of the workflow (plan->explore->scale->publish) and different communities of practice (design->dev->product->maintain) requires different strategies for adoption.

# The systems and tech landscape

Workflows do not work in isolation – they operate in an ecosystem of data, software, compute infrastructure, people. This also needs smart support.

There is and always will be diversity of platforms and they service different points in the stages of scientific investigations (explore - scale - publish) - get over it. Focus on how to operate within this ecosystem.

We are at a tipping point of tech enablement.

- The infrastructure is now available
- The need is clearly recognised
- Evidence that we can change how we approach the scientific method

But we need to be able to sustain and evolve this tech enablement and glue it into the platforms of science used by scientists in the field - from instrument to publishing.

## **Recommendation Priority:**

Funds and activity needs to be mobilised for this supporting the ecosystem as much as the workflows

What are the FAIR principles needed for the Workflows ecosystem - bearing in mind that Pandora's box

Standards are the blue collar work of science and badly supported/funded. Support strategically. That are standards that are emergent and being adopted widely. Encourage the exchange of standards across disciplines. Encourage the use of common industry standards rather than homegrown

# Building and Sustaining the tech landscape

There is a complex landscape of developments, partly because funding demands novelty and penalises reuse

Problem driven workflow systems have proved successful (and are often “simple” compared to fancy CS) but they have production investments, developing products rather than proof of concepts

What is the role of industry, PPPs and engagement with the private sector  
Software is free like a free puppy (ie its not free)

Recommendation

Funding should promote reuse and standards adoption not penalise it

Funding software engineering and products is a good thing

# Future Priorities

## Investment

- Making AI useful and used as a part of scientific workflows requires investment different than domain and data science investments
- Capitalizing on training and “career tracks” for “scientific AI” developers has large potential ROI in the long term.
- Workflows of the future will bridge data-driven and intelligence-integrated optimization with scientific process. This does not exist today.
- Data is an essential part of AI. Open science developments around data also applies to workflows.
- Standards are needed for integration of tech.

**Recommendation:** Amidst many opportunities, focus on success stories that generalize and the “people and process” to “turn AI into product” in the short-term.