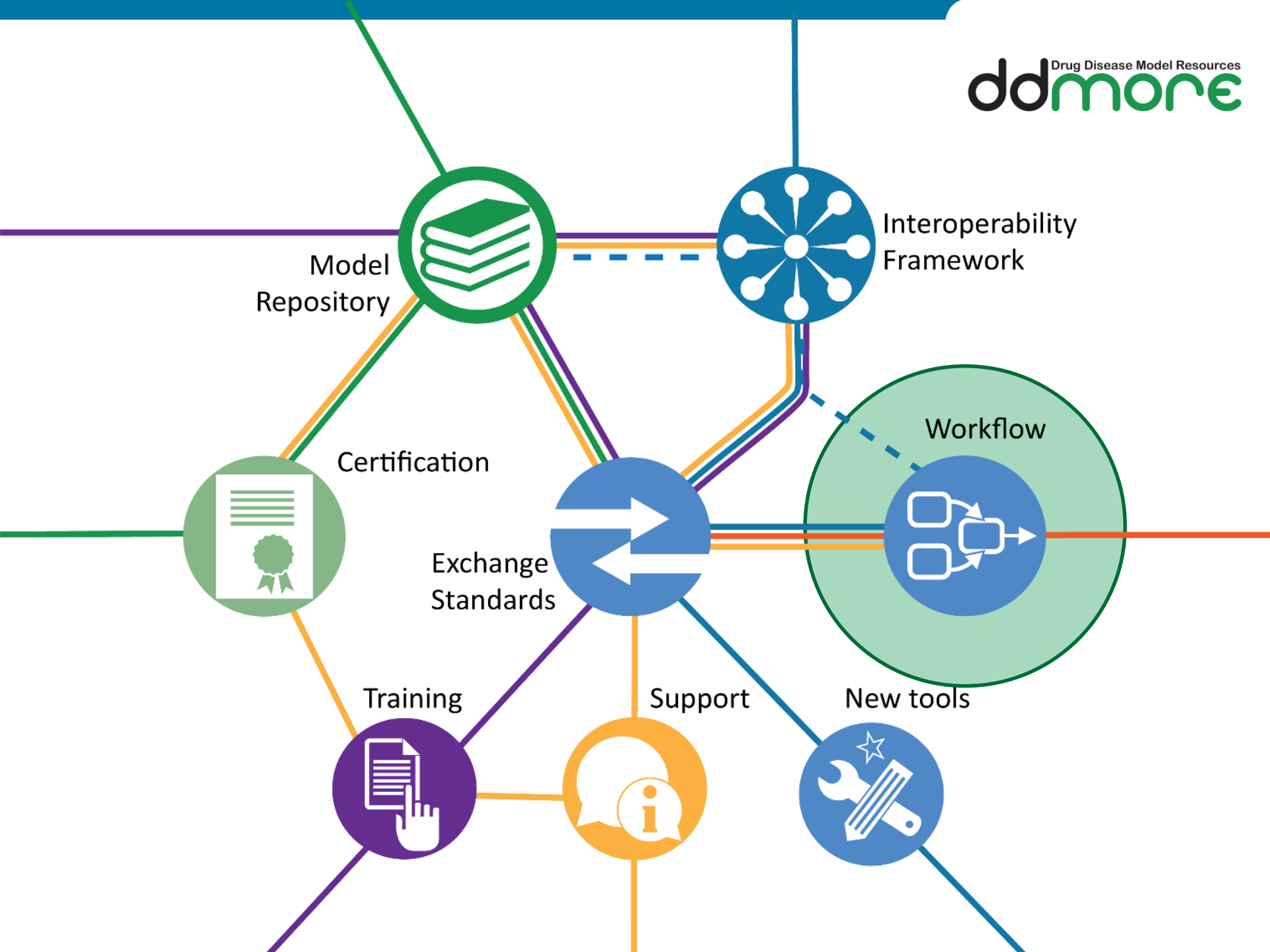
# Thoughtflow

## Standards and Tools


Kickoff meeting: 8 March 2017


Drug Disease Model Resources
ddmore
Model Repository
Interoperability Framework
Workflow
Certification
Exchange Standards
Training
Support
New tools

# Overview

- **Introductions**
- **A brief introduction to Thoughtflow**, and how it fits in with DDMoRe Community Groups
- **Progress to date**
- **Some thoughts** about next steps
- **Engagement and ideas**
- **Next meeting**

# Introducing Thoughtflow

## *Why Thoughtflow?*

Pharmacometric data analysis is a complex and multifactorial task…

Keeping track of all the steps in analyses in a project and coherently documenting them afterwards is challenging, tedious, error-prone, and time-consuming.

## *Existing tools don't quite measure up*

Numerous "workflow"/information management tools already exist,
e.g. Taverna, Knime, Activiti, Kepler, Navigator, Improve, Pipeline Pilot, etc

These tools mostly track input-output for tasks (addresses reproducibility), but do not capture more complex relationships such as assumptions, decisions, external influencers. They also lack version control.

Accurately tracking model development requires capturing relationships between entities and activities in a more useful and comprehensive way than traditional workflow tools.

## *So: Thoughtflow…*

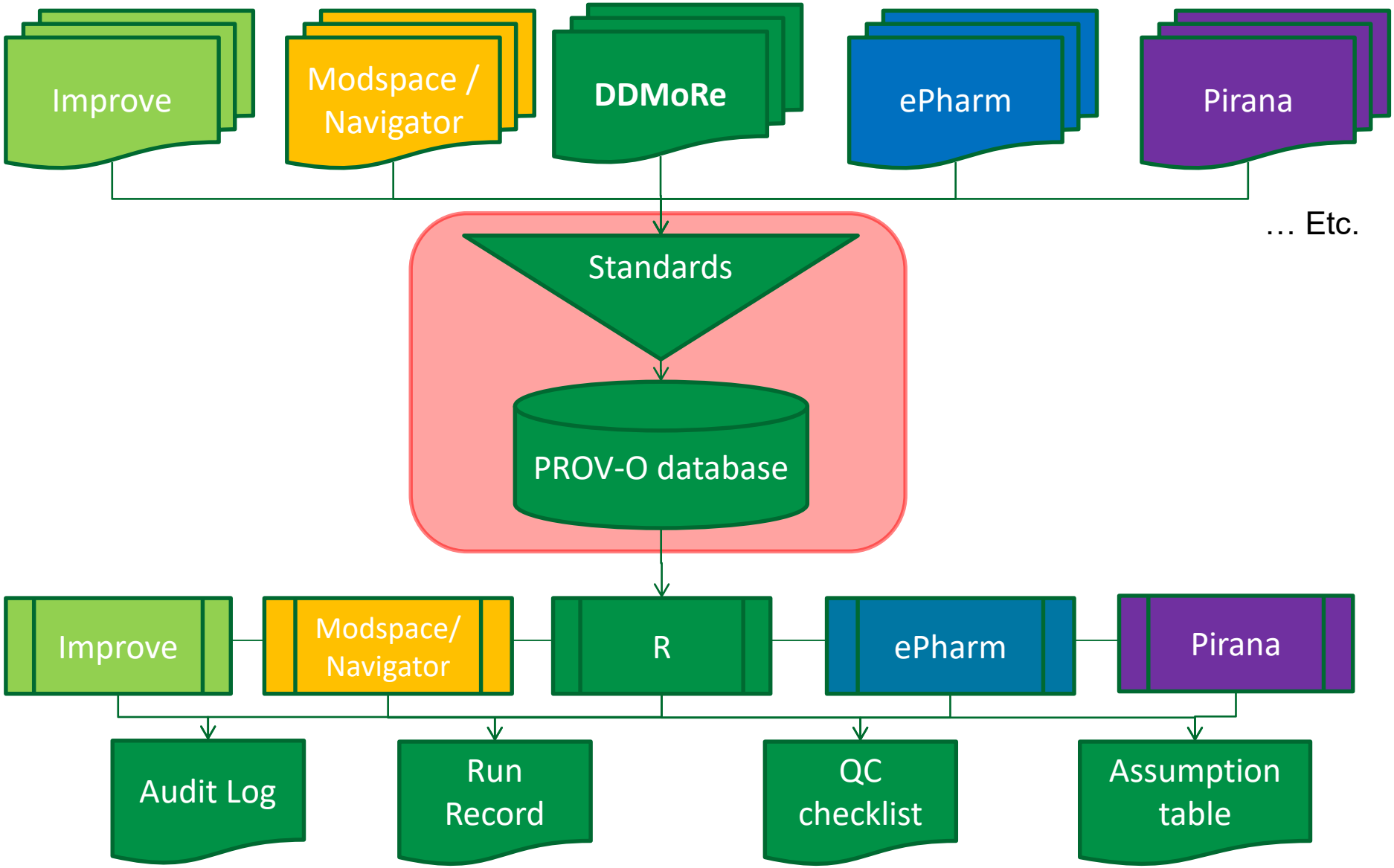
Tracks the assumptions and decisionmaking process behind the analysis, as well as the analysis itself.

# Thoughtflow

## *What is it?*

- The DDMoRe MID3 Thoughtflow standards and tools are intended to provide cross-platform, robust support for tracking, reporting and replicating elements of a modelling and simulation PKPD project (across all phases of development), either in part or as an entire analysis
- The model development process can be summarized in the form of audit logs, run records, QC checklists, and assumption tables
- Such a tool can also support regulatory submission by recording metadata and other source (input) information; modelling and simulation outputs, assumptions, extrapolation, interpretation, decisions and documentation
- Although integrated with the DDMoRe environment and infrastructure, it can exist independently

# Diagrammatic view of Thoughtflow



# Potential benefits

- ***To Analysts***
  - Rapidly generate documentation for analyses in a reproducible and structured manner
    - Audit logs, run records, QC checklists, analysis reports
  - Facilitate knowledge management by providing rich metadata: plans, assumptions, rationale, annotations, decisions and QC/QA pass/fail
    - Makes it easier to pick up/transfer/communicate work from others to review and contribute
  - Avoids the repetition or duplication of prior work, and helps assure quality, traceability, and reproducibility

# Potential benefits

- ***To Managers***
  - Much easier to track analyses, impact on decisions, dependencies, assumptions
    - What work is impactful? What aspects of analyses typically drive decisions? What assumptions underpin inferences? How long do we spend on QC? What are the rate limiting steps?
  - Easily track the progression of the analysis with respect to project timelines, and can thus properly allocate resources to meet potential deadlines

# Potential benefits

- ***To Reviewers and Regulators***
  - Assumptions are clearly documented and tied to inputs (data sources), models (which model aspects are underpinned by assumptions) and inferences (what is the consequence of the assumption). Improve transparency and thus enhance clear communication between sponsors and regulators

# Potential benefits

## *As standards*

MID3 [1] has emerged as a driver for better transparency around provenance, which defines the relationships between entities and activities.

We have translated the concepts, terms and processes set out in the MID3 paper and develop a standard that permits the capture, storage and analysis of those concepts and allows traceability from root concepts such as a problem definition, or question to be answered, through to a decision, recommendation or conclusion.

We have also adapted and expanded the World Wide Web Consortium's PROV-O [2] standard for pharmacometric applications.

ex: Entities (models, datasets, analysis scripts), activities (model estimation), assumptions, decisions
Elements are linked to each other (ex: covariate model derived from base model, decisions made based on model results/diagnostics)

[1] Marshall et al (2016). Good Practices in Model-Informed Drug Discovery and Development: Practice, Application, and Documentation. *CPT:PSP* 5(3):93-122.
[2] http://www.w3.org/TR/prov-o/

# Potential benefits

## *Usefulness of the standards*

The MID3 Thoughtflow standards can allow for incorporation into third-party software tools, driving further development for capturing pharmacometric analyses.

Demonstrations of the technology will soon be incorporated into other third-party software tools such as Mango Development Solutions' Navigator and scinteco's Improve.

# PROV-O: The W3C Provenance Ontology

Drug Disease Model Resources

**Provenance** is information about entities, activities, and people involved in producing a piece of data or thing.

- Can be used to form assessments about its quality, reliability or trustworthiness
- The W3C PROV Family of Documents defines a model, corresponding serializations and other supporting definitions to enable the inter-operable interchange of provenance information in heterogeneous environments such as the Web

https://www.w3.org/TR/prov-o/

# Capturing Pharmacometrics Workflow Concepts with PROV-O

PROV-O defines a range of terms that are used to capture all the information necessary to define the provenance of items:

- **Entity:** An entity is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary (e.g. a model, a dataset, an output file, a script, a decision, an assumption)
- **Activity:** An activity is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities (e.g. a model execution, a visual predictive check)
- **Agent:** An agent is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity. Can be an Organisation (e.g. Leiden, Pfizer), a Person (e.g. a User) or a Software Agent (e.g. R, Monolix).

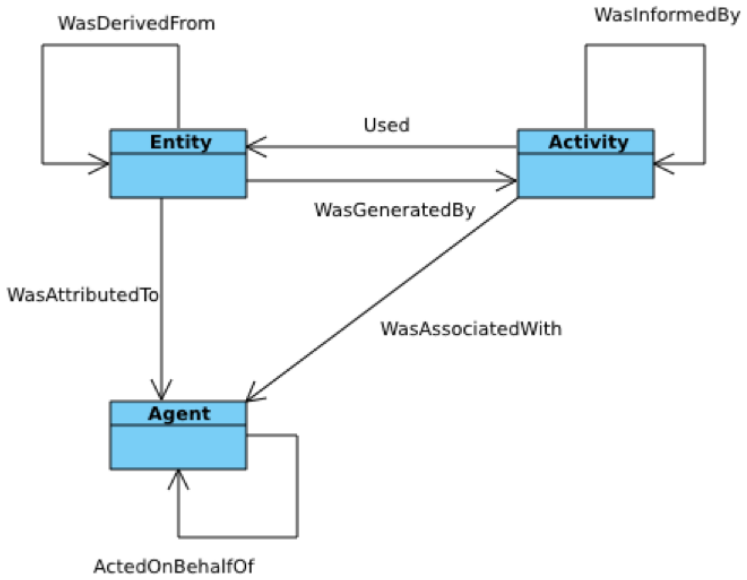
https://www.w3.org/TR/prov-o/

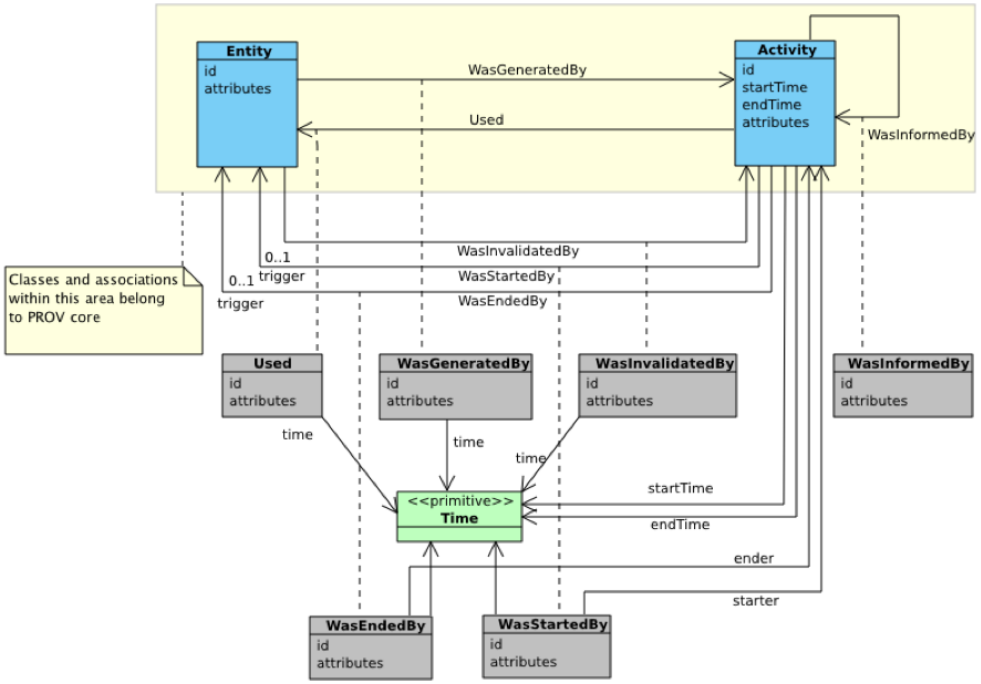
# Capturing Pharmacometrics Workflow Concepts with PROV-O

Drug Disease Model Resources

PROV-O defines a range of terms that are used to capture all the information necessary to define the provenance of items:

- **Relationships:** PROV-O defines a set of relationships that describe the interactions between entities, activities and agents.

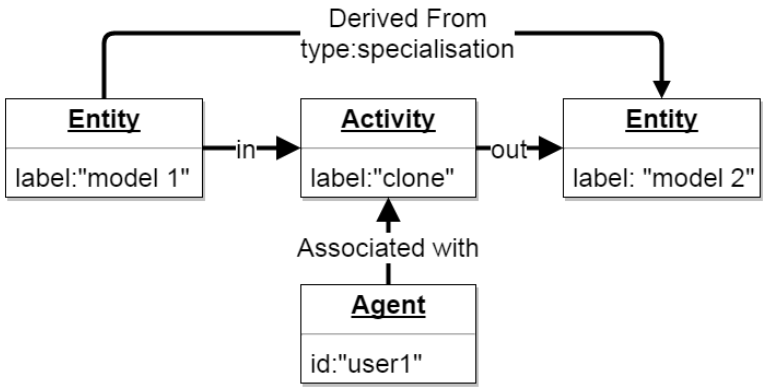


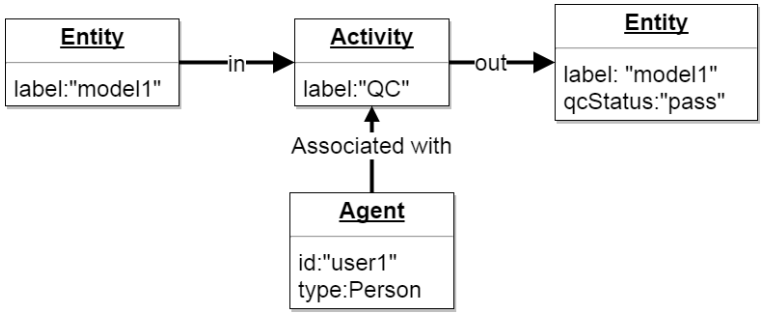
# Capturing Pharmacometrics Workflow Concepts with PROV-O

Drug Disease Model Resources

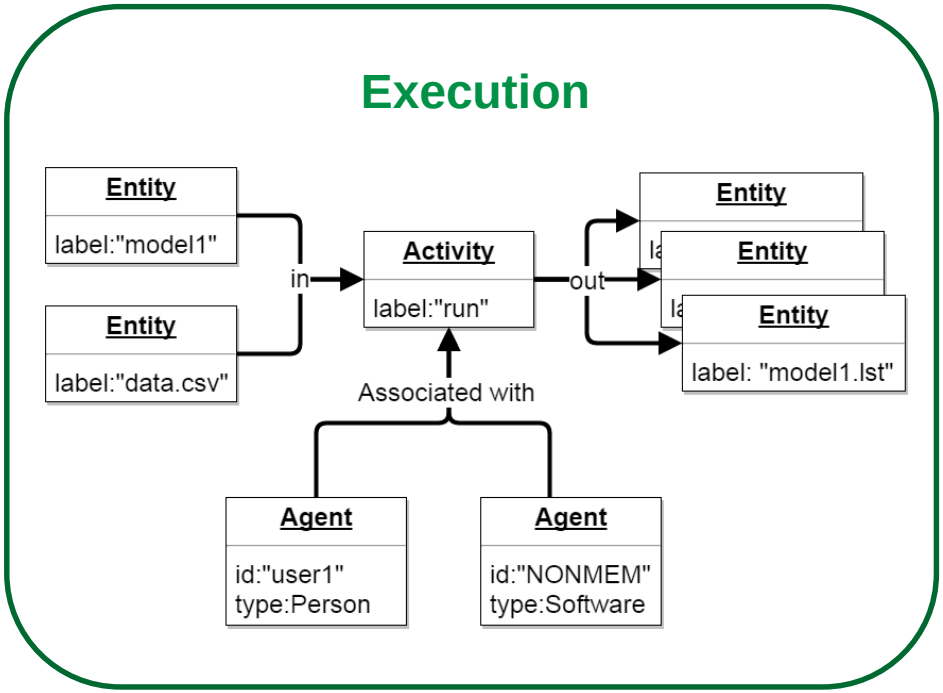
## Model derivation




## Model characterisation




## Execution



# Capturing Pharmacometrics Workflow Concepts with PROV-O

Drug Disease Model Resources

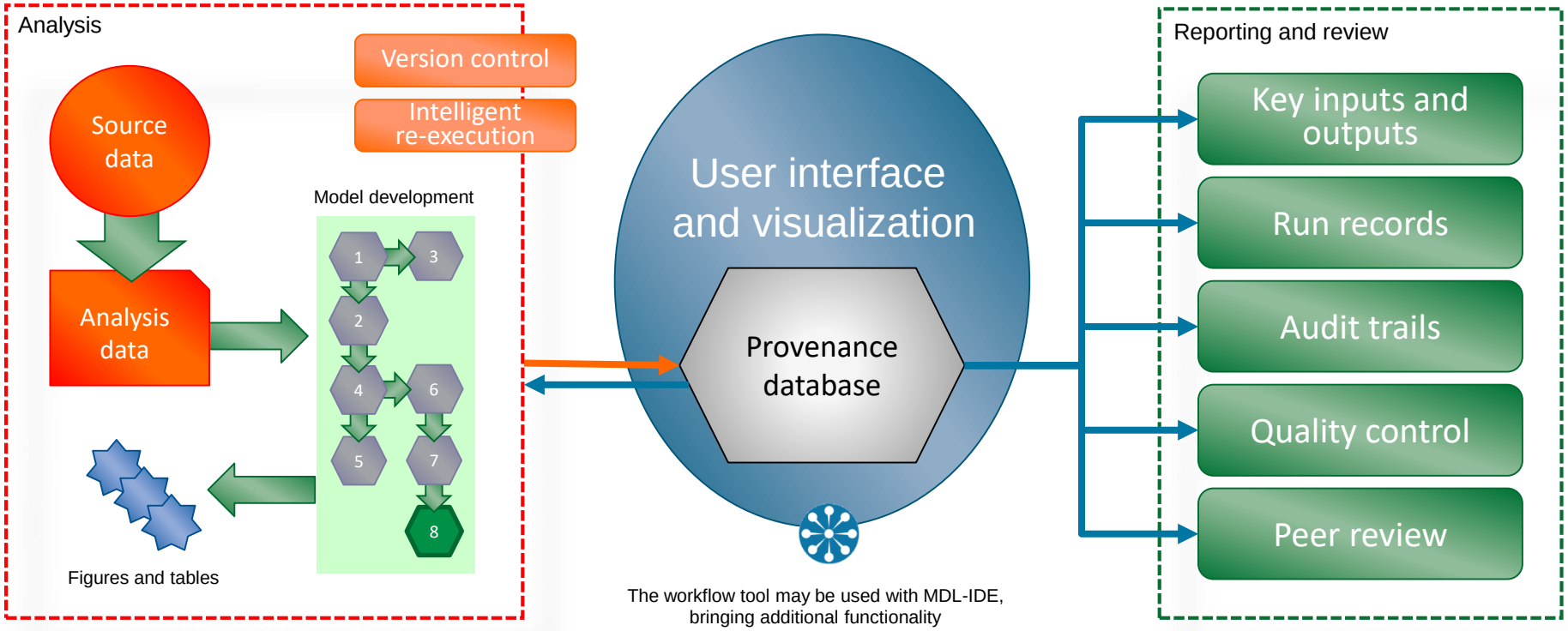
ddmore

| Relationship | Definition |
|---|---|
| **Was generated by** | Generation is the completion of production of a new entity by an activity. This entity did not exist before generation and becomes available for usage after this generation. E.g. a model output file was generated by a model execution activity. |
| **Used** | Usage is the beginning of utilizing an entity by an activity. Before usage, the activity had not begun to utilize this entity and could not have been affected by the entity. E.g. a data file was used by a NONMEM control stream. |
| **Was invalidated by** | Invalidation is the start of the destruction, cessation, or expiry of an existing entity by an activity. The entity is no longer available for use (or further invalidation) after invalidation. E.g. a data change caused a model output file to be invalidated. |
| **Was associated with** | An activity association is an assignment of responsibility to an agent for an activity, indicating that the agent had a role in the activity. It further allows for a plan to be specified, which is the plan intended by the agent to achieve some goals in the context of this activity. |
| **Was informed by** | Some anonymous entity passes between two activities. So, this property allows the construction of provenance chains comprising only Activities. |
| **Was derived from** | A derivation is a transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity. |

# As a tool

We intend to launch a free, open source software implementation at the end of August 2016.

Although it will be a very early version, we believe it will provide a convincing demonstration of concept and solid foundation for further development.

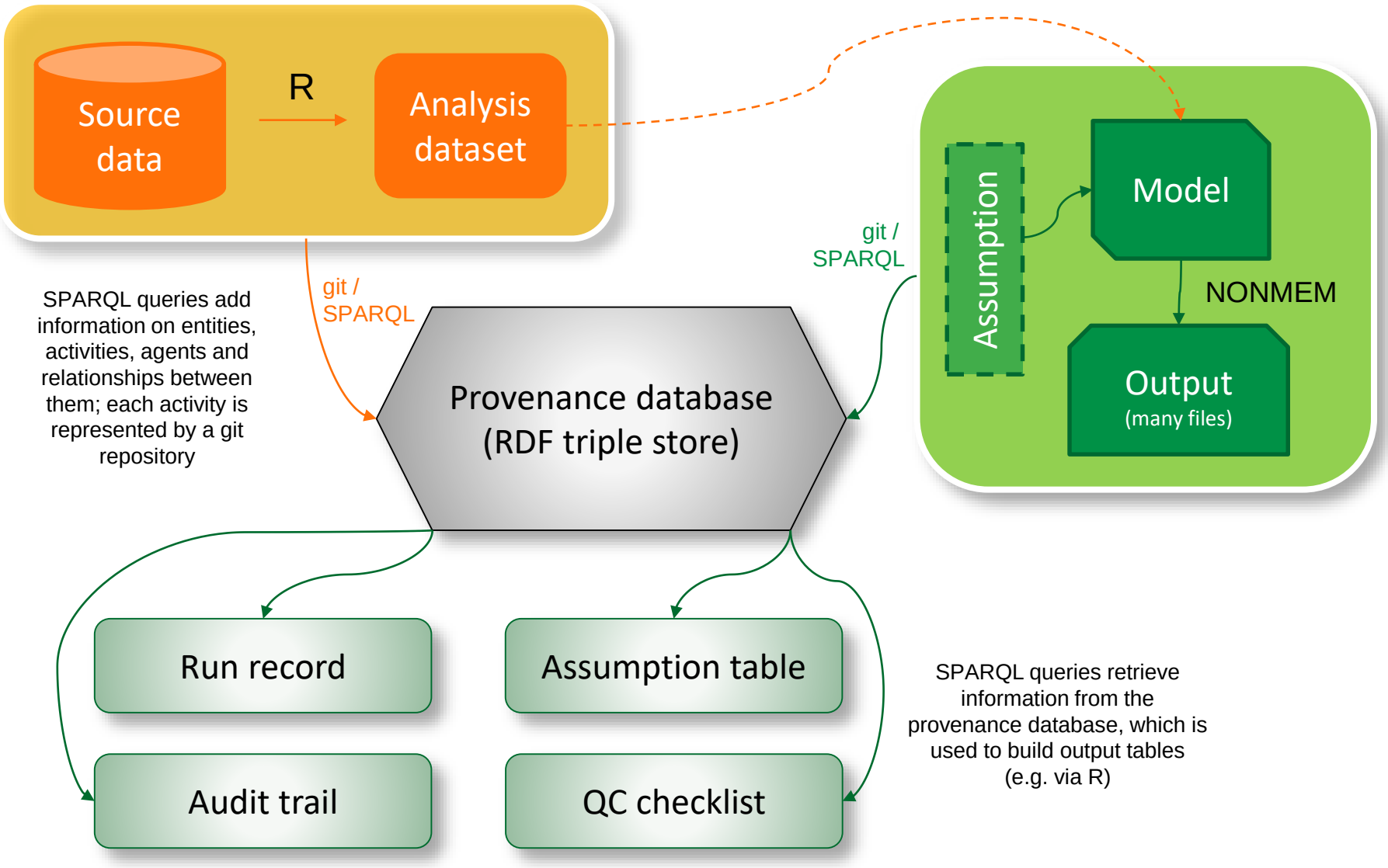



Inputs, scripts, models, assumptions, outputs and decisions are versioned and their provenance is tracked

A lightweight but powerful GUI brings visualization and control

The database can be queried to streamline reporting and review

# How it all works – some examples



# MID3: Assumptions

| Important Assumptions | Justification | New/ Established | Testable/ Not-Testable | Test/Approach to assess impact | Evaluation |
|---|---|---|---|---|---|
| Pharmacological assumptions | | | | | |
| Physiological assumptions | | | | | |
| Disease assumptions | | | | | |
| Data assumptions | | | | | |
| Mathematical and statistical assumptions | | | | | |

Source: MID3 workgroup communication slides

# MID3: Assumptions

```
<?xml version="1.0"?>
<Assumption>
    <Type>Pharmacological</Type>
    <AssumptionBody>Emax model fixed to 100% is a more physiological
description of the data compared to a linear model.</AssumptionBody>
    <Justification>Emax model is not better than linear model; however, for
this drug class, Emax of 100% is more realistic.</Justification>
    <Established>New</Established>
    <Testable>Testable with a wider range of concentrations (external/future
study).</Testable>
    <TestApproach>
Comparison of simulated metrics
of interest between the
two competing models.
     </TestApproach>
    <TestOutcome>To achieve a 90% response
(assumed to be clinically
meaningful) requires a twofold
higher dose using the
Emax model compared to
the linear model.</TestOutcome>
</Assumption>
```

**You do not have to write XML! (Unless you want to.)**

# Progress so far

- We have a detailed technical document describing the standards and how they should be implemented
- We have a manuscript in preparation with CPT:PSP (final proofing)
- A technology demonstration has been prepared and released on Github
  - https://github.com/DDMoReThoughtflow
  - Largely built on Java
- We have a (very quiet) website
  - http://thoughtflow.community/
- There's a video!
  - https://www.scinteco.com/index.php/2016/09/18/ddmore-workpackage-thoughtflow/

# Next steps

- We need to finish the standards, and publish them
  - 3-12 months, depending on available sets of hands
- We need a working implementation
  - 12-60 (?) months, depending on how much support we can get
  - Based as heavily as possible on and around R (?)

# Ideas?

Drug Disease Model Resources

ddmore

# Who has time to commit?

Drug Disease Model Resources

ddmore

# Next meeting

Drug Disease Model Resources

ddmore

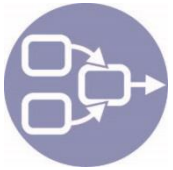
# Backup

Drug Disease Model Resources

# Core values
# MID3 Workflow environment

| Value | How DDMoRe delivers this |
|---|---|
| **Traceability** | Workflow tool records all individual steps throughout analysis with time stamps and dependencies |
| **Clarity** | On-the-fly and post-hoc visualisation of the steps of the analysis and their interdependencies (model “Tree” view) |
| **Audit trail** | Workflow tool may be used to generate a complete audit trail for an analysis |
| **Decision tracking** | Decisions may be documented at any point in model development, linked to any entity (model, data, output); visualization of decision tree |
| **Facilitated review process** | Activities and entities can be marked as Qced – reviewer can see dependent activities and entities and whether these are Qced. Assumptions are linked to data, models / model components are linked to assumptions, assumptions can be invalidated. Easy to see impact of assumptions on inferences (MID3). |
| **One-click re-execution and reproducibility** | Upon changes made to input entities (such as datasets), all dependencies may be re-generated using a single click; entire analysis may be repeated in this way |
| **"Lab Book"-like documentation** | Workflow tool automatically collects information that can be queried in a reporting, scientific, QC or audit context |
| **Stand-alone installation** | Runs independently of other DDMoRe software components, although delivers benefits if these are present |

# Core features
# MID3 Workflow environment



Drug Disease Model Resources
ddmore

| Feature | Description |
|---|---|
| **Tracks provenance** | Recording of relationships between components of an analysis (scripts, models, outputs, assumptions, decisions, …), minimizing burden on users |
| **Artifact version control** | Detects when a component has changed, records the nature of the change, and determines its impact on other components |
| **Updates analysis components to reflect upstream changes ("re-running")** | When changes to an analysis component (such as a data file) have been made, all dependencies may be updated at the click of a mouse |
| **Exports run records, QC checklists and audit trails** | Run records, QC checklists and audit trails may be assembled and exported in a convenient format for reporting, minimizing analyst need to do this manually |
| **Assembles report components** | Key graphs and tables can be prepared and assembled in a convenient location and format for reporting and QC |
| **GUI facilitates project and provenance visualization** | A streamlined, portable GUI centralizes all functionality without getting in the way |
| **Independent but synergistic** | DDMoRe tools are not required for core functionality, but their availability will bring benefits |
| **Built on standards** | Underlying technology is built on widely-adopted standards for defining provenance (PROV-O) and programming tools (Java) |
| **A solid foundation** | The tool will be designed to support further development and expansion |