PISTOIA ALLIANCE

RSC ChemSpider Webinar, June 2026

Chemical Standards & Representation

J. Christian Baber, Pistoia Alliance

What is the Pistoia Alliance?

Global, Not-For-Profit Members' Organisation

Supporting R&D in the Lifesciences & Healthcare industries

Mission

To lower the barriers to innovation through pre-competitive collaboration.
"We act as champions of science, its practitioners, and membership community"

Our Values

Industry Value Creation | Better Together | Future Focused

A member-led club that allows member companies to work together without additional legal agreements

What is Pre-Competitive Collaboration?

precompetitive

adjective

1. describing the early stages of the development of a commercial product, during which competitors collaborate

From Wiktionary, Creative Commons Attribution/Share-Alike License. More at Wordnik

- Allows a group of **competing companies to come together** to develop a solution for a problem that they all share, and from which none of them would gain a competitive advantage*[1]
- **Produces rapid advances** in a wide spectrum of activities now conducted largely by individual companies working separately [*2]
- Successful when **all of the participants involved benefit** [*2]

Data Management

Best Practices

Regulatory Authorities

Sustainability

*[1] The Benefits of Pre-Competitive Collaboration in the Pharmaceutical Industry, Drug Discovery World, October 2015, https://www.ddw-online.com/the-benefits-of-pre-competitive-collaboration-in-the-pharmaceutical-industry-1310-201510/

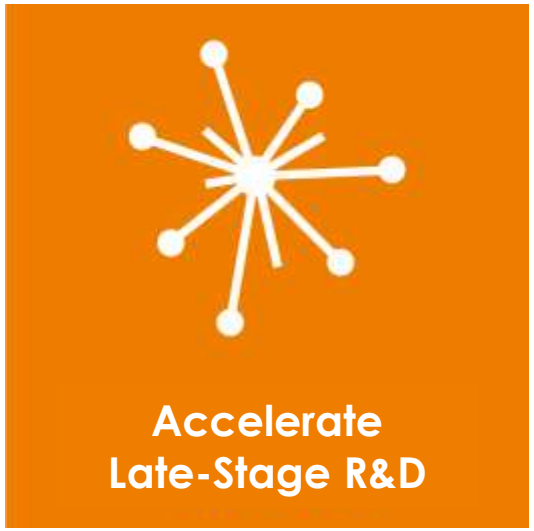
*[2] Establishing Precompetitive Collaborations to Stimulate Genomics-Driven Product Development: Workshop Summary, 2011, https://www.ncbi.nlm.nih.gov/books/NBK54325/

Our Strategic Priorities

Data is the foundation for AI and digital strategies, but we struggle to manage, maintain and consume data

83% of Life Sciences Execs say they will not achieve growth goals without AI*

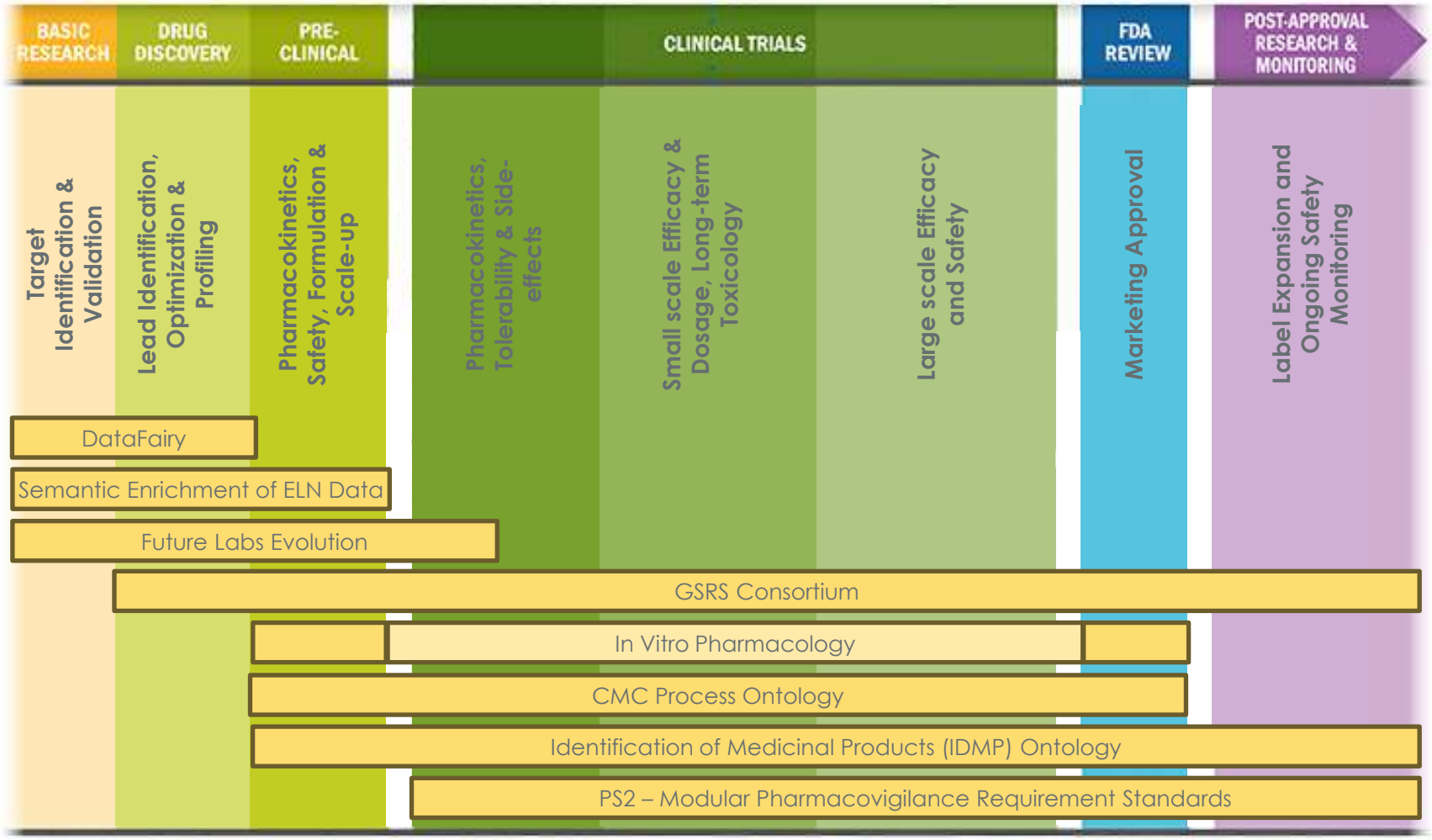
* Accenture Report



Increase the effectiveness of the most expensive phases of R&D to benefit patients

ESG is a priority for companies from a growth and public health perspective. We must ensure that the benefits of our therapies are not offset by our operations

Activity Across the R&D Lifecycle



Practical Applications

End-To-End FAIR Data

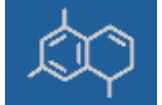
Pulling It All Together

Standards Make Life Easier

So why is it so difficult to get people to adopt them?

Chemical Standards are Everywhere

Representation



Structure
IUPAC Drawing Rules, HELM

Nomenclature
IUPAC Naming Rules

Identifiers
ChemSpider ID

Experimental

Methods
PA/Allotrope Digital Analytical Methods

Measurements
SI Units, BAO

Assays
PA Protocol Registry

Data

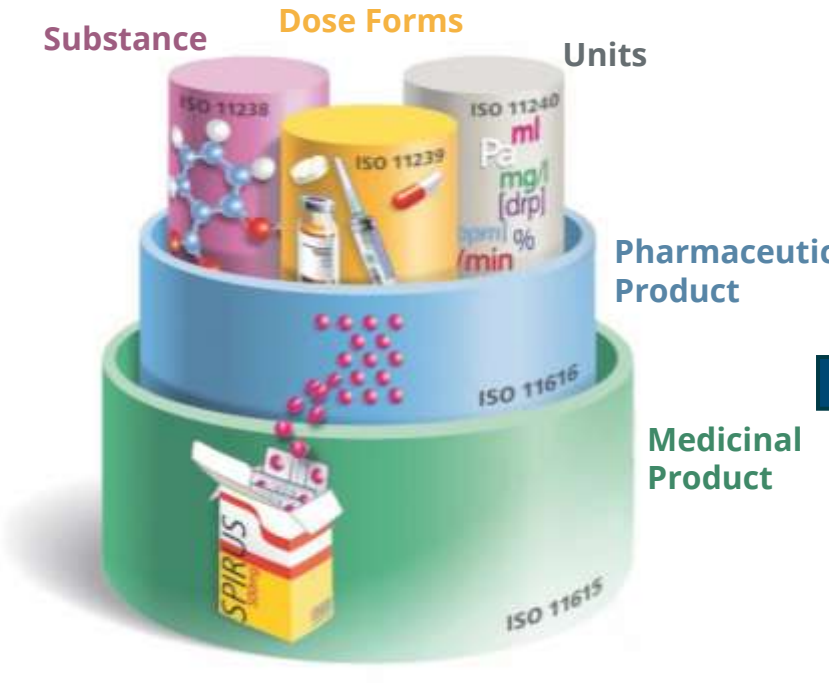
Formats
CT Files, SMILES, InChI keys, CIF

Metadata
PA Semantic Enrichment of ELN Data

Ontologies
PA CMC Process Ontology

Standing on the Shoulders of Giants

1+1 >> 2



ISO IDMP Standards

Diverse Implementations

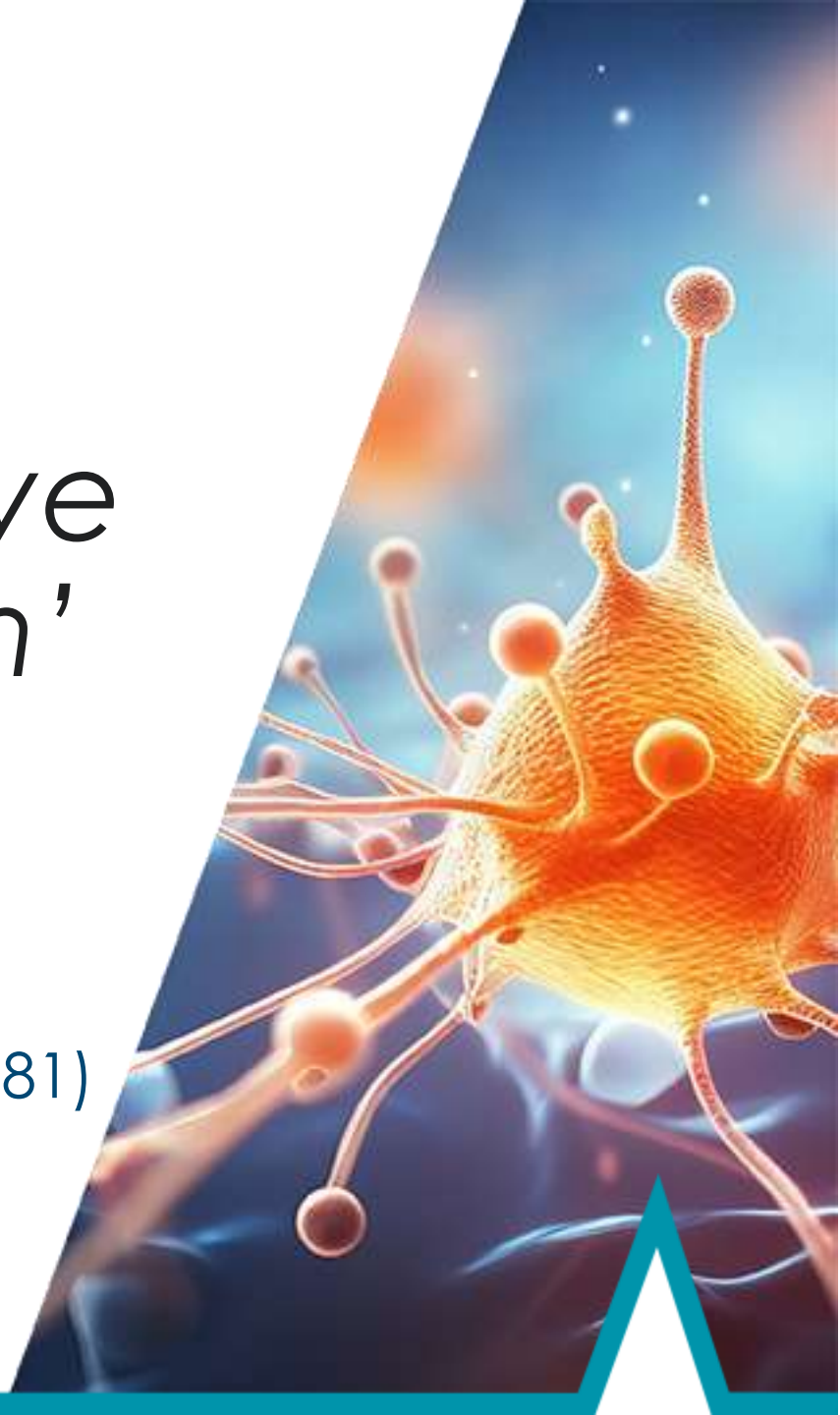
- Create data silos
- Risk standardization benefits
- Reduce operational efficiency

Standardized IDMP Ontology

- Improved regulatory compliance & patient safety
- Simplified product development and data management
- Efficient product registration and supply chain optimization

'The nice thing about standards is that you have so many to choose from'

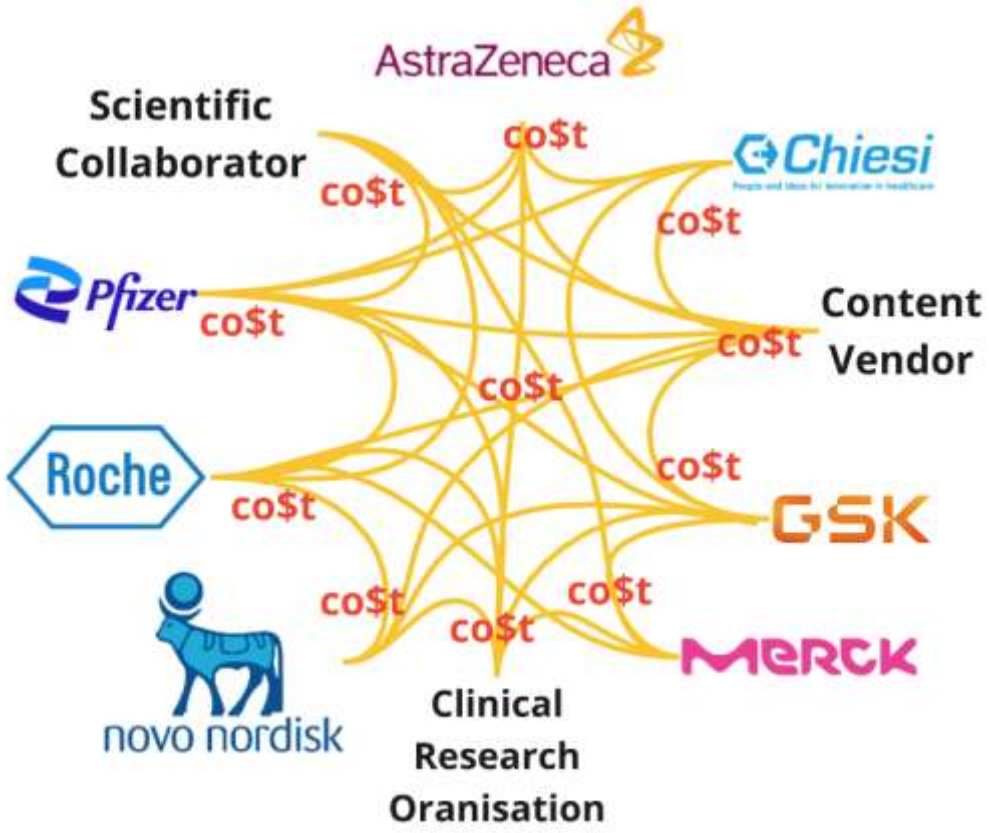
Andrew S. Tanenbaum, Computer Networks (1981)



Pharma General Ontology (PGO)

Definately Not '*One Standard to Rule them All*'

Current Situation

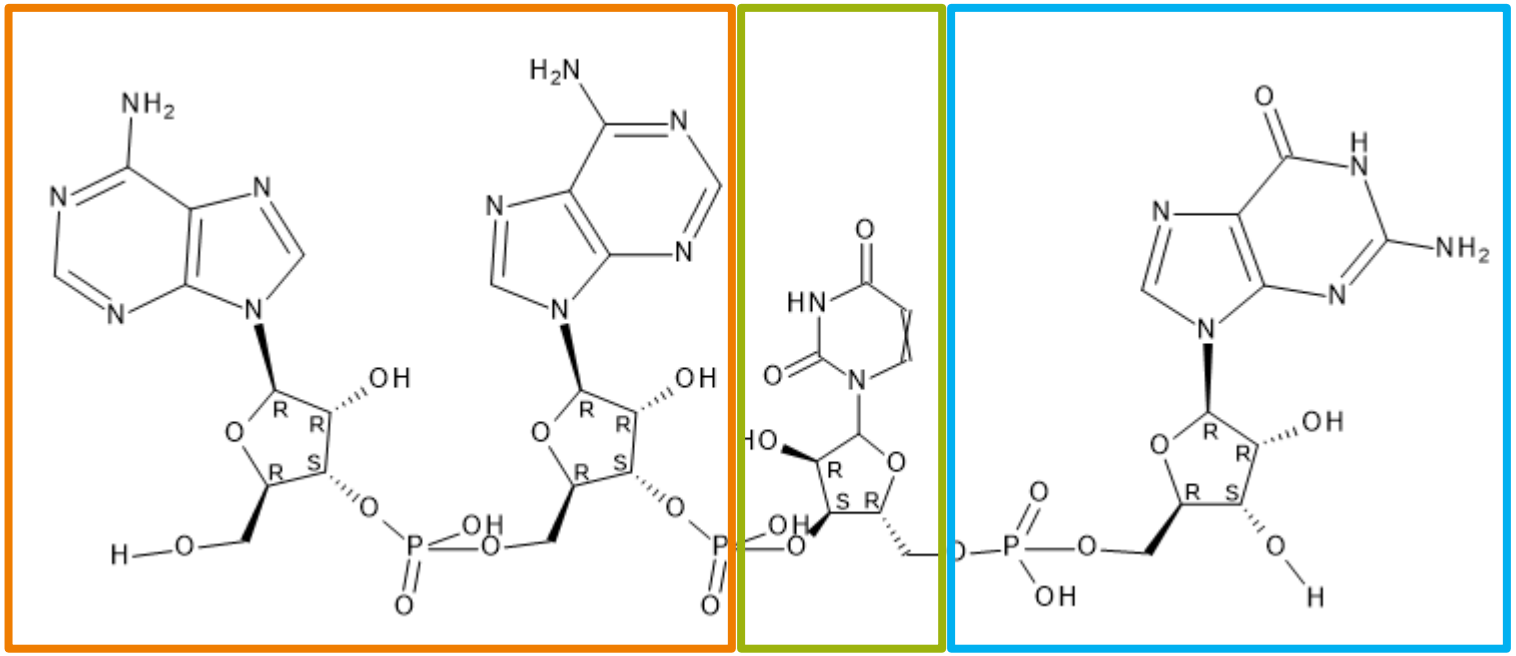


PGO

Target State – PGO Meta-Standard

Issues: Ambiguities in Standards

HELM - Hierarchical Editing Language for Macromolecules



Toolkit 1:

```
RNA1{R(A)P.R(A)P.[O([C@@H]([C@H]1O[*:1])CO[*:
2])C(N(C=CC2=O)C(=O)N2)[C@H](O)1]P.R(G)}$$$$V
2.0
```

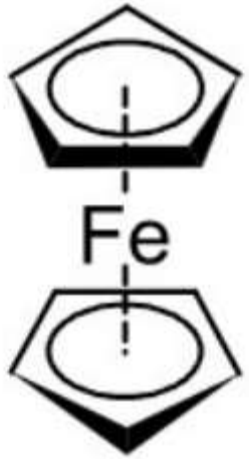
Toolkit 2:

```
CHEM1{[O([C@@H]([C@H]1O[*:1])CO[*:2])C(N(C=CC2
=O)C(=O)N2)[C@H](O)1]}|RNA1{R(A)P.R(A)P}|RNA2{
P.R(G)}$RNA1,CHEM1,6:R2-1:R1|CHEM1,RNA2,1:R2-
1:R1$$$V2.0
```

According to current documentation both are valid

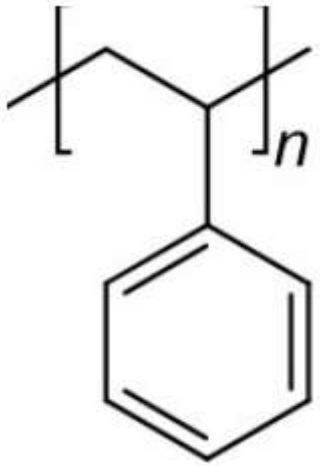
Issues: Complex Structure Representation

Organic Compounds Were The Focus



Connections vs Bonds
Hypervalent molecules
Electron-deficient compounds

Organometalics & Inorganics



Chemical graphs may not be sufficient
Concept of 'identity' is limiting
Constraints vs Flexibility

Polymers, Nanomaterials & Mixtures

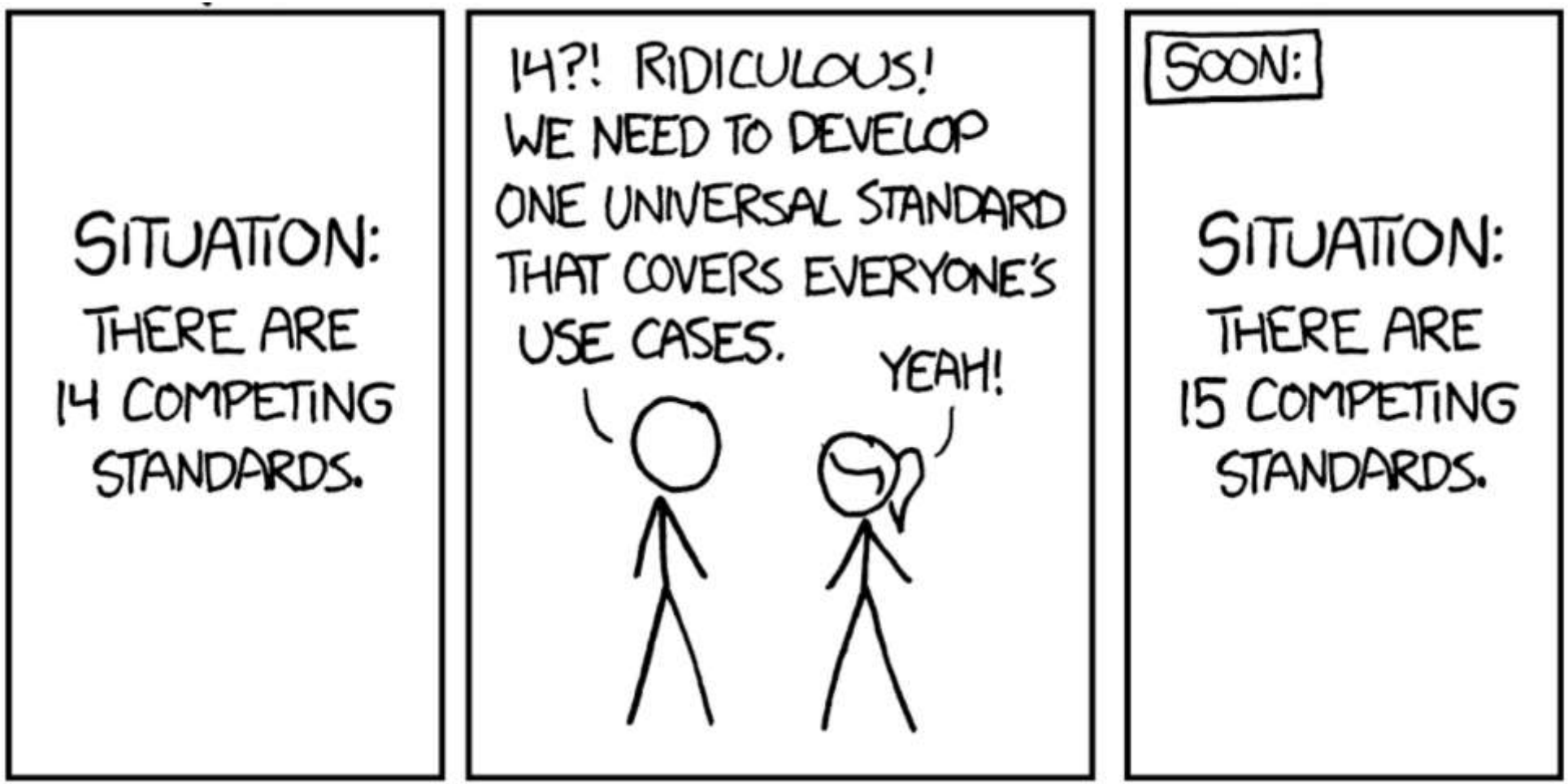
Limitations of Standards Causes Diverse Representation

Many options are cheminformatically valid
Limit computational processing (AI)
Reduce communication & data sharing

T. Doerner, Challenges in Cheminformatics:The View of An Independent Consultant, Persistent Challenges in Cheminformatics, Pistoia Alliance Event at RSC , June 2025

Standards, Standards Everywhere…

But they were *'Not Invented Here'*



xkcd: Standards

Chemical Exchange Format Committee

A neutral forum, hosted by the Pistoia Alliance, where pharma companies, software vendors, and standards owners can collaboratively address ambiguities, share insights, and guide the evolution of chemical exchange standards.

Industry Challenge

1. Formats are well established, but documentation is often dispersed and/or incomplete
2. Ambiguous interpretations can lead to inconsistent handling of data resulting in unnecessary work and expense
3. Format development pathways are often arbitrary, unclear and slow

Goals and Intended Deliverables

- Curated documentation of standards with SWOT analysis of formats
- Documentation of ambiguities & shortcomings on standards
- Identification of shortcomings & idiosyncrasies in implementations
- Cross industry prioritization of enhancements

Value Proposition

- Clarity on formats when implementing
- Informed decisions on format use
- Prioritized feedback to address shortfalls
- Improved interoperability between tools

Committee Benefits

Increased Quality
Increasing industry-wide quality reduces errors and rework

Market Influence & Insight
Opportunity to shape standards and tools

Faster Innovation
Prioritizing key challenges and opportunities

Time Savings
Reduced data wrangling leaves more time to use the data

AstraZeneca · ELSEVIER · Benchling · Johnson&Johnson Innovative Medicine · BIOVIA · Pfizer · CERTARA · Roche · sapio sciences

Sustainable Chemistry Data Standards Coalition

Maintain & Improve What we Already Have

Robust, well-maintained chemical data standards—and clear best practices guiding their creation, governance, and use—are critical for the accurate representation, exchange, and reuse of chemical information across the scientific ecosystem. Ensuring their long-term sustainability, alongside promoting best practices in their development and application, is essential for enabling interoperability, supporting digital transformation, and preserving the long-term value of scientific data.

Vision & Mission

Vision: A world where chemistry data flows openly and reliably to accelerate scientific progress and fuel sustainable innovation.

Mission: To coordinate and lead a community-wide effort to sustain and support chemistry data standards through shared practices, projects, and resources.

Coalition Partners

- Pistoia Alliance
- CODATA
- IUPAC
- Royal Society of Chemistry
- Beilstein-Institut
- InChI Trust
- Physical Sciences Data Infrastructure (PSDI)
- German Chemical Society (GDCh)
- Leibniz Information Centre for Science and Technology and University Library (TIB)

PISTOIA ALLIANCE

Use Existing Standards Whenever Possible

If you *must* build a new one,
then Don't Do It Alone

"If you want to go quickly, go alone.
If you want to go far, go together."