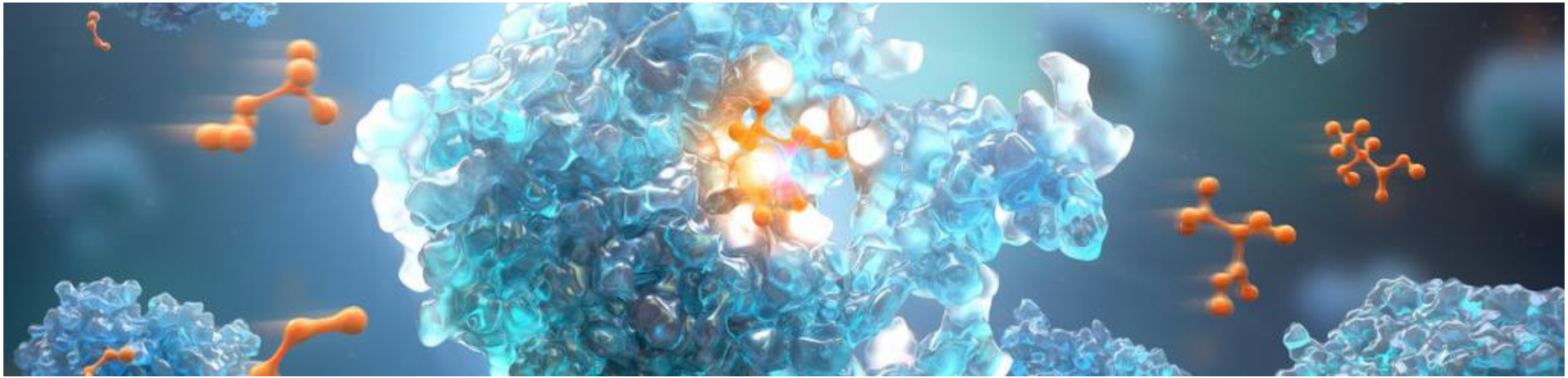
D360 customisation at AZ

Gostar patent analysis
Curve viewer

Wolfgang Klute, Chemoinformatics Data Scientist, Science IT

Certara D360 UGM Boston 2019

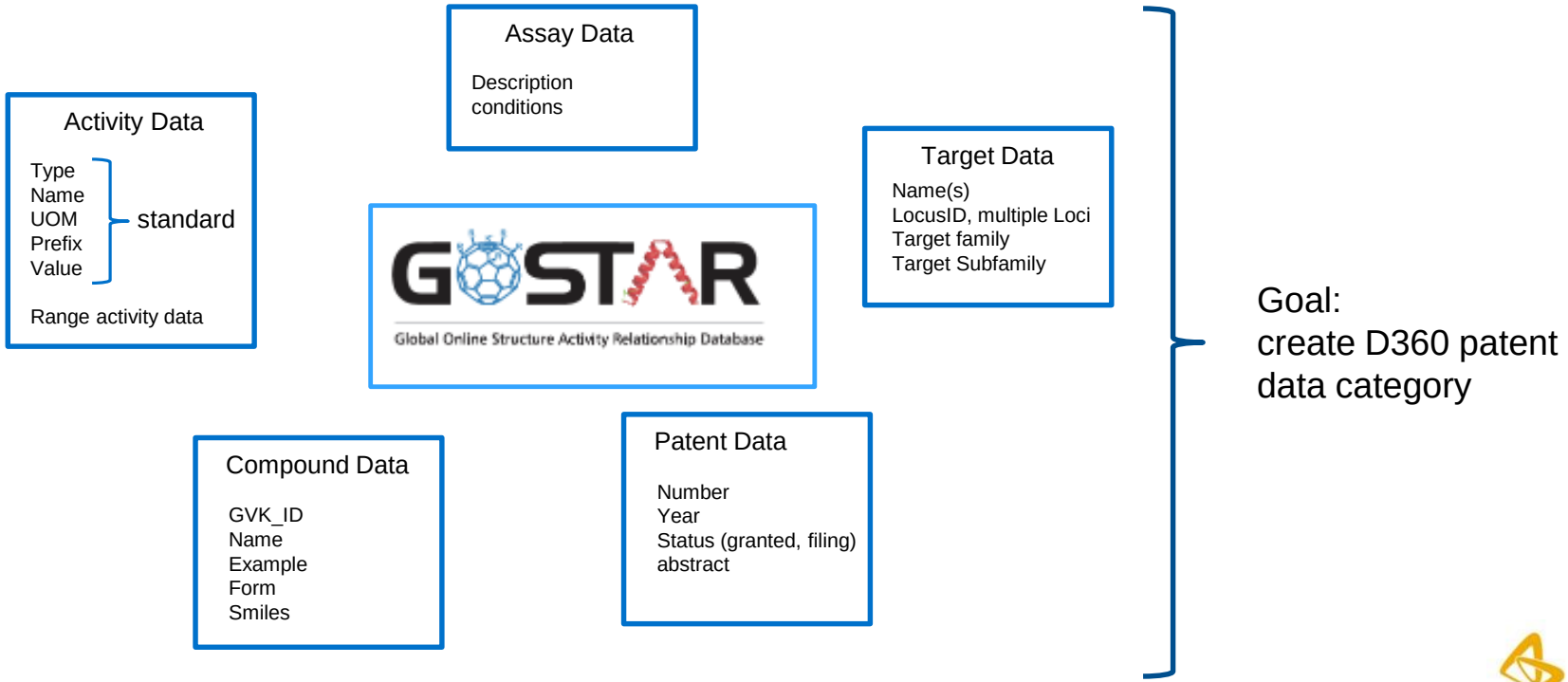
09 Oct 2019



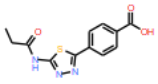
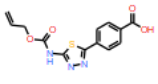
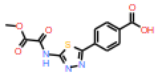
Gostar patent analysis

Gostar patent data

- Database distributed by Excelra with SAR data from patents and MedChem lit
- 80 K patents, > 600 K assays, > 8 M compounds, 16 M activities



Conventional and easy: unpivoted

	GVK Id	GVK Structure	Concat Normalized Journ...	Concat Assay Id	Concat Official Name	Concat Assay Type	Concat Std Activity Type	Concat Standard Uom	Concat Std Act Prefix	Concat Standard Value
☐	121515960		WO2012121168	10998749 11000058 11000059	CSNK2A1 CSNK2A2 CSNK2B	B B Q	INHIBITION INHIBITION Qualitative	% % none	= = =	79.000 88.000 0.000
☐	121515961		WO2012121168	10998749 10999825 10999826 11000058 11000059	CSNK2A1 CSNK2A2 CSNK2B	B B B B Q	INHIBITION INHIBITION IC50 IC50 Qualitative	% % uM uM none	= = = = =	83.000 91.000 8.100 2.700 0.000
☐	121515963		WO2012121168	10998749 11000058 11000059	CSNK2A1 CSNK2A2 CSNK2B	B B Q	INHIBITION INHIBITION Qualitative	% % none	= = =	65.000 56.000 0.000

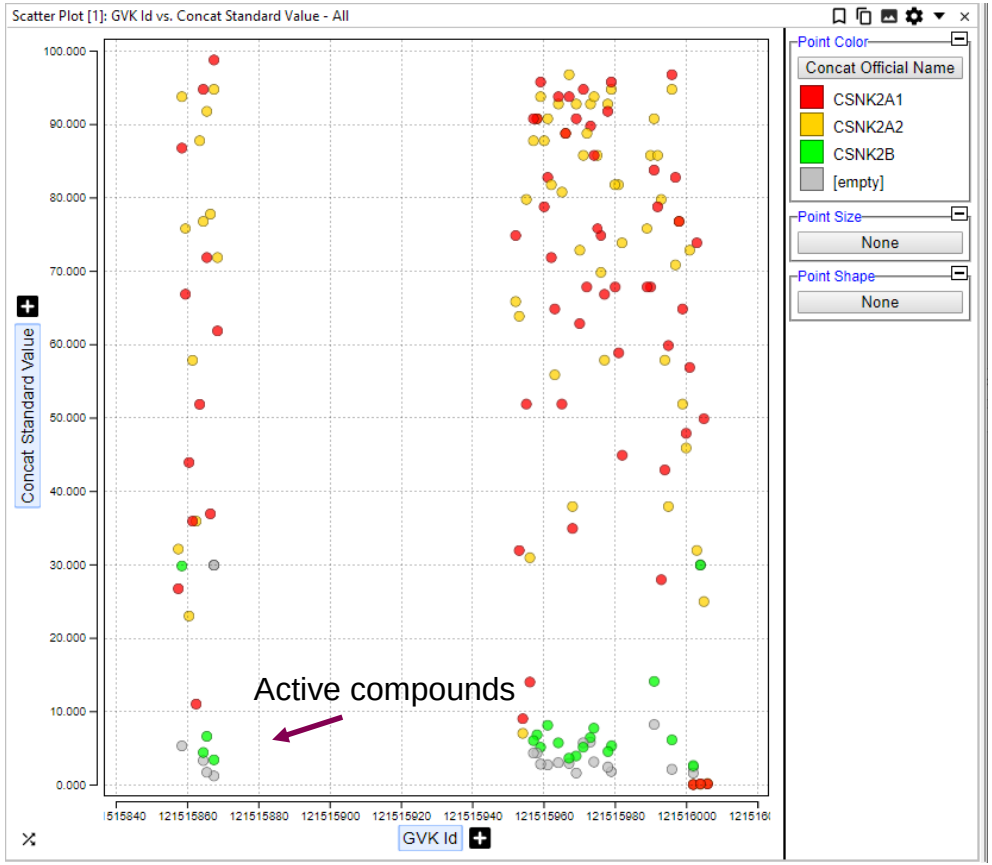
activity

target

assay

Easy to create
Had these views for many years

Analysis of un-pivoted data in D360



D360 offers some utlity to analyse non-pivoted data

can see which compounds are active

can't correlate activity

- assay 1 vs assay2
- assay 1 vs clogP
- etc

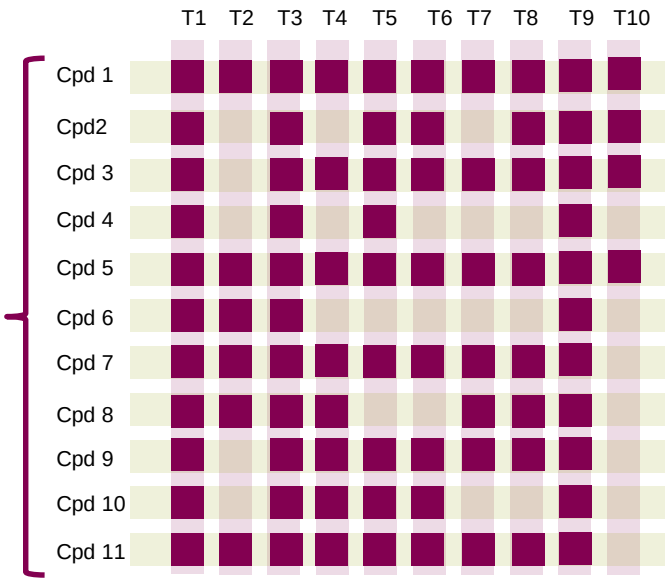
Could we create pivoted view?

Corporate data vs public data

ibis sharing discovery data — Curated test shared across projects

CMS

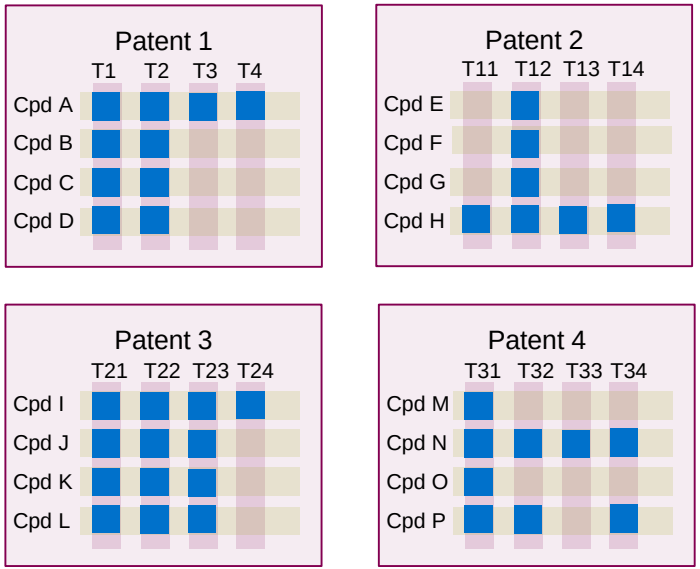
registered compounds shared across projects



M's compounds, 10 000s tests, 10 M's activities (number results)
Activities / compound = 15 : 1
Activities / assay = 1800 : 1

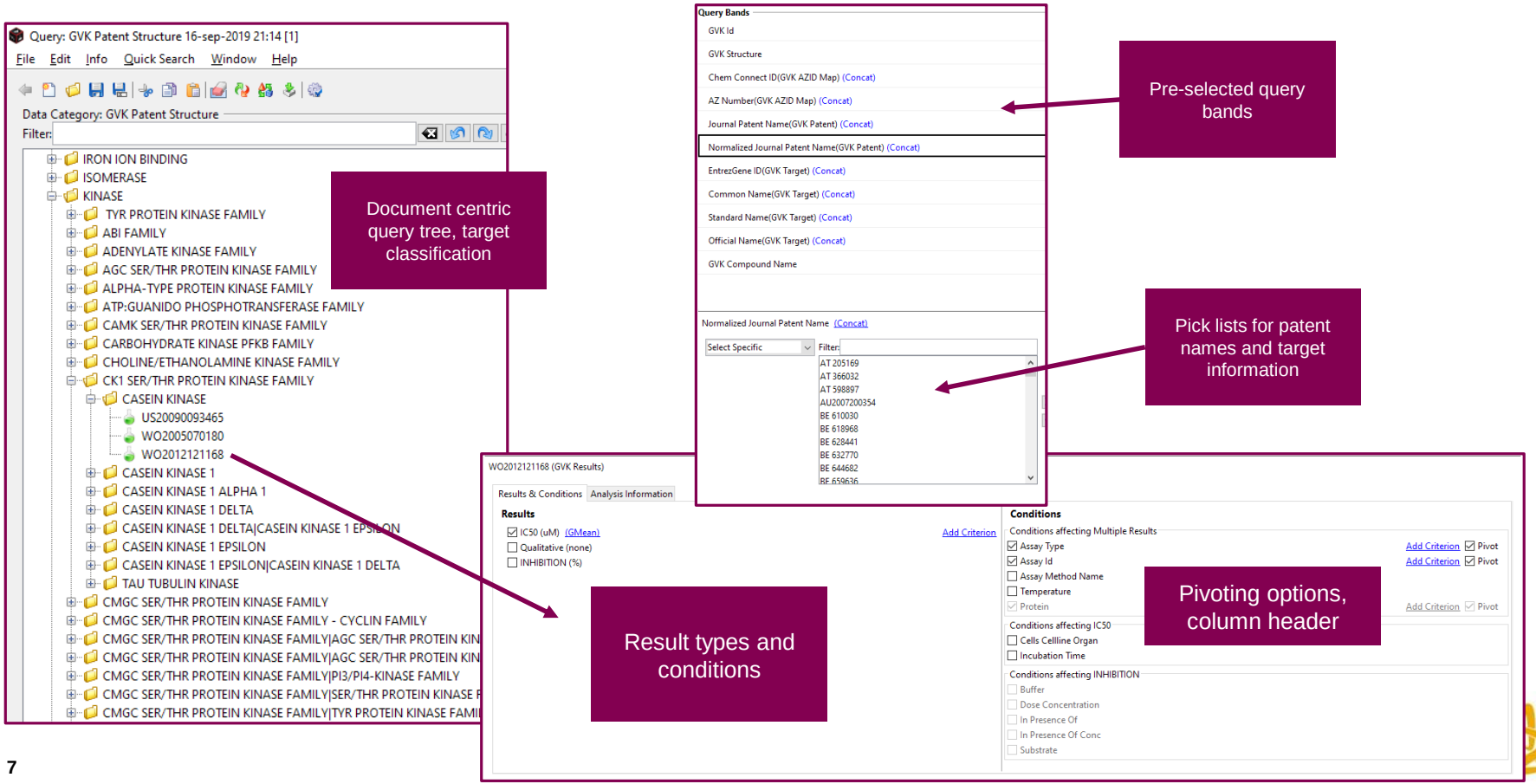
GOSTAR — Global Online Structure Activity Relationship Database

Document centric
Fragmented data

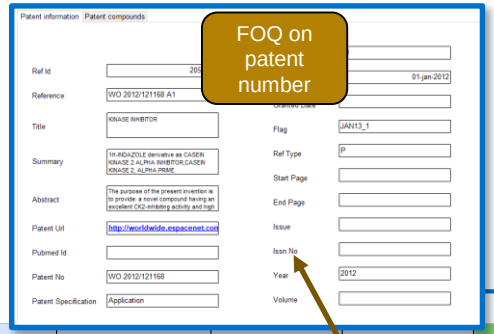


- 80 K patents, > 600 K assays, > 8 M compounds, 16 M activities
- Activities / compound = 2 : 1
- Activities / assay = 25 : 1

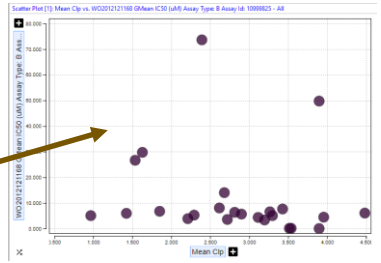
Gostar patents: user friendly query interface



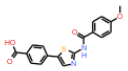
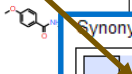
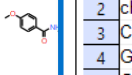
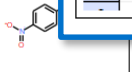
Pivoted views



FOQ on patent number



One column per assay, useful for scatter plots

	GVK Id	Mean Chemistry Connect ID	GVK Structure	Concat Normalized Journ...	WO2012121168 GMean IC50 (uM) Assay Type: B Assay Id: 10999825	WO2012121168 GMean IC50 (uM) Assay Type: B Assay Id: 10999826	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000927	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000928	WO2012121168 GMean IC50 (uM) Assay Type: FC Assay Id: 11000940	WO2012121168 Mean INHIBITION (%) Protein: Casein kinase 2 alpha Assay Type: B Assay Id: 11000058	WO2012121168 Mean INHIBITION (%) Protein: Casein kinase 2, alpha prime Assay Type: B Assay Id: 11000059
☐	121516002	285286146		WO2012121168	0.032	0.046	2.600	1.600	2.400		
					1/1	1/1	1/1	1/1	1/1		
☐	121515867	285286145					<>	<>	<>		
									>30.000	99.000	95.000
										1/1	1/1
☐	121516004	322778689									
☐	121515857	221306189									
					1/1	1/1					

Synonym for Chemistry Connect ID

	Database Name	S	ne	Name
1	ACD	Prima		MFCD28127268
2	chembl	Primary		CHEMBL2087023
3	ChemSpider	Primary Compound ID		28523182
4	GOSTAR	Primary Compound ID		GVK120991876
5	GOSTAR	Primary Compound ID		GVK121516002
6	GOSTAR	Primary Compound ID		GVK123093966
7	GOSTAR	Primary Compound ID		GVK123193077

FOQ on CC_ID

Most potent compound appears in 4 patents + other refs

ChemistryConnect: AZ platform to normalise internal and external structures and create unique identifiers

learnings

- D360 useful tool to analyse complex deep data
- Pivoted patent analysis useful for one patent at a time
- Use follow-on queries to provide customisable views of meta data
- Key is user-friendly implementation (many supporting custom tables…)
- ChemistryConnect integration essential to pick up relationships between patents
- Gaps:
  - ❖ Range activity data
  - ❖ Column aggregation with BAO
  - ❖ useful visualisations for un-pivoted data (heat map)

D360 curve viewer

AZ curve viewer in D360

Why do we need a curve viewer in D360?

- Need reliable decision making on complex assays where numerical end points are not sufficient
- Need to be able to visually spot outliers outliers
- Need to identify effects not captured by numerical end points (e.g. cytotoxic events at high concentration in phenotypical assays)

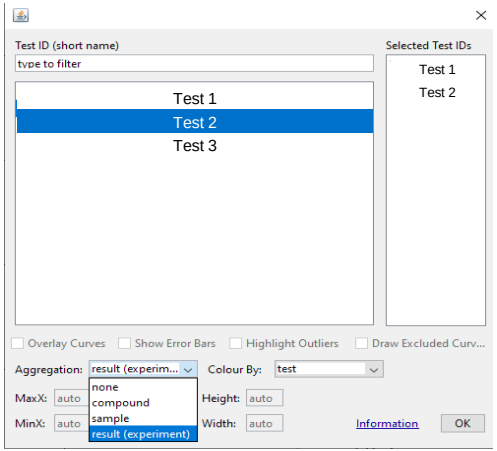
AZ built interactive curve viewer extension to D360 (D3, HTML)
Data consumed from a datamart with normalised curve parameters (I/EC50, S0, Sinf, slope) and point data

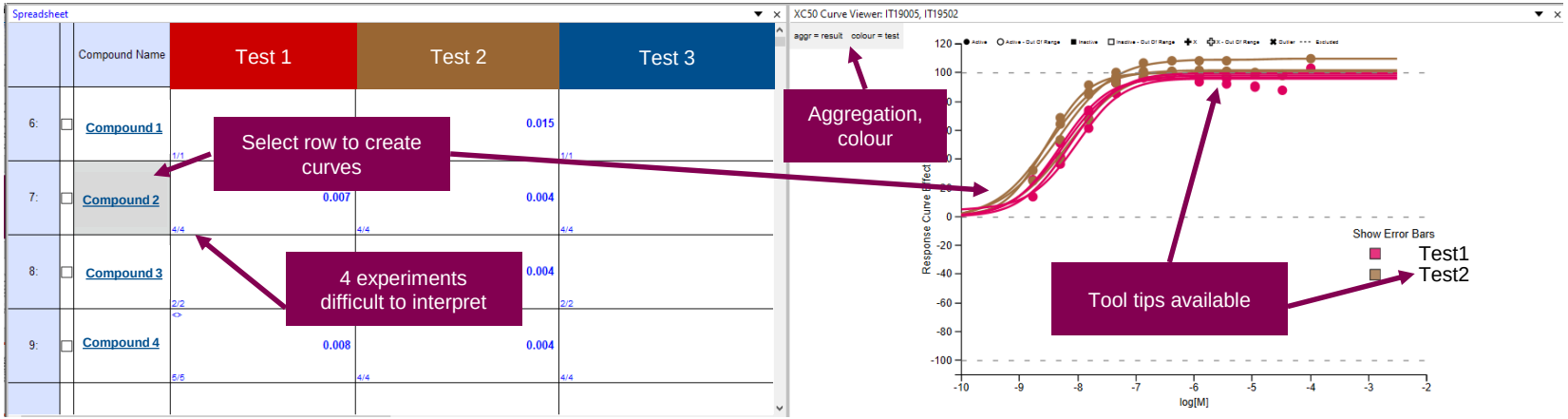
functionality

- Curves shown in a separate interactive window
- Curves from a single compound, multiple tests
- aggregation levels:
  none, experiment, sample, compound
- Colour code by experiment or test
- Shapes according to AZ activity flag
  + out of range information (> or <)

Active Active - Out Of Range Inactive Inactive - Out Of Range X X - Out Of Range

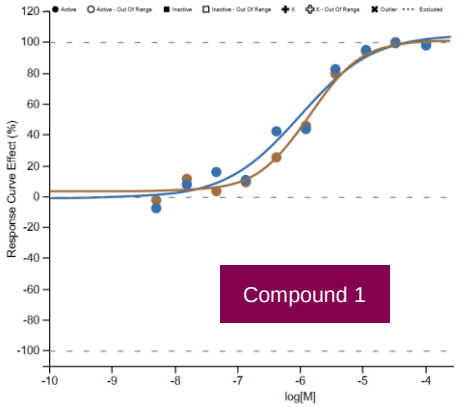
- Handling of point outliers, error bars, tool tips

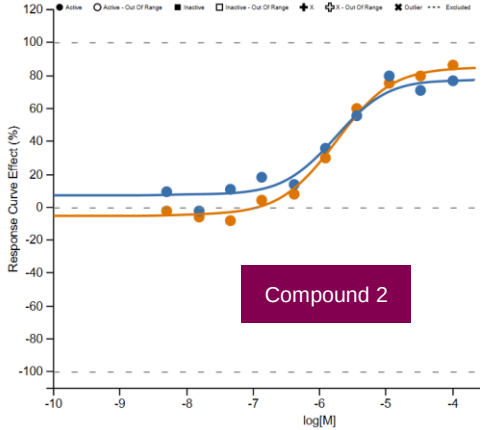


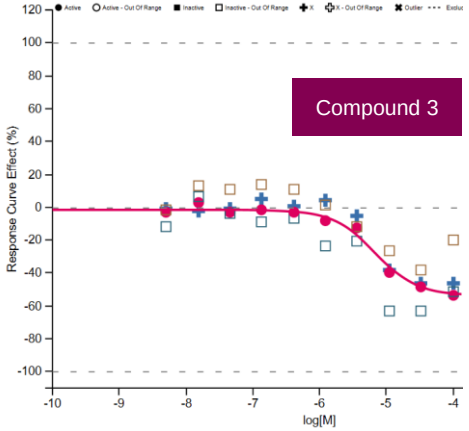


Example 1: detecting inverse agonsts

3 compounds in an agonist assay, project table contained many more columns







Exp1: excluded
Exp2: inverse agonist
Exp3: inactive
Exp4: inactive

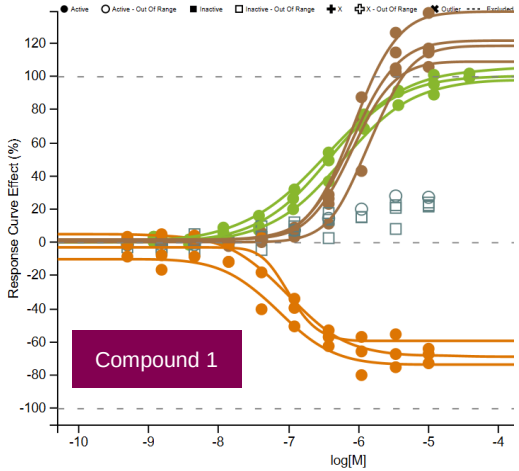
		Compound Name	GMean EC50 (µM)	Concat Distinct Act Flag
1:	☐	Compound 1	1.155 2/2	Active 2/2
2:	☐	Compound 2	1.740 2/2	Active 2/2
3:	☐	Compound 3	TE<> >40.817 3/4	E Irregular Active Not Active 3/4

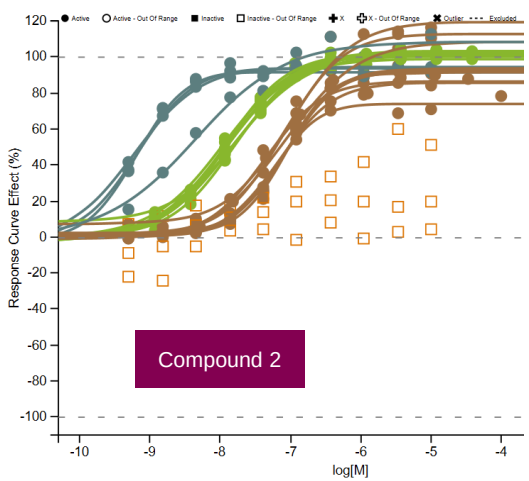
Easy to miss outliers in project tables where few end points are reported

Curve viewing required to make correct decisions

Example 2: complex decision making in NHR

Challenge: assay variability, complex decision criteria involving 4 assays
-> project can't rely on aggregated number results





Individual experiments shown
Colour coded by test ID

		Compound Name	Assay 1	Assay 2	Assay 3	Assay 4
1:	☐	Compound 1	0.418 3/3	0.091 3/3	<> >10.000 4/4	0.937 4/4
2:	☐	Compound 2	0.013 8/8	<> >10.000 3/3	0.001 4/4	0.074 9/9

Assay 4: needs full antagonism
-> compound 1 better
assay 2: no negative or positive response
-> can't conclude this from data

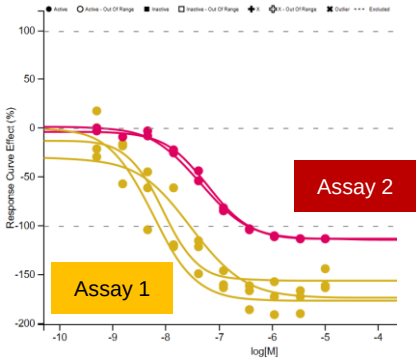
Curve aggregation

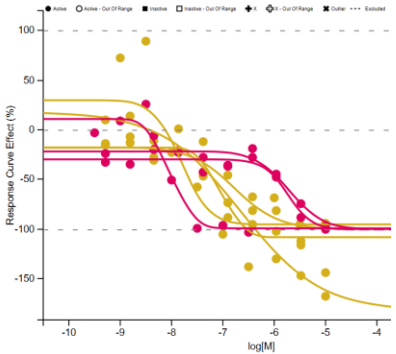
Useful when comparing several tests. However dangerous….

Compound 1

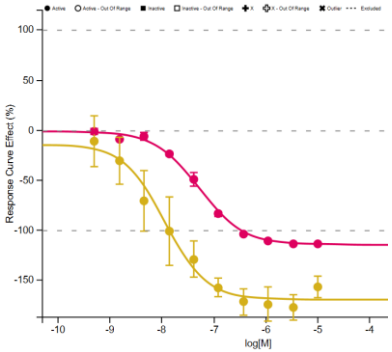
Compound 2

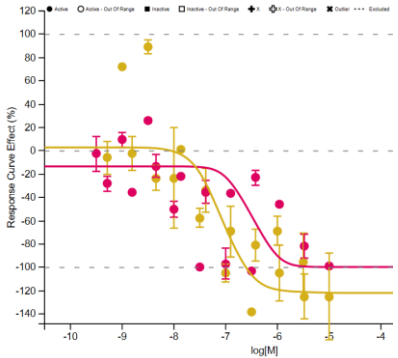
Experiment level
each curve is shown





Compound level
each curve is shown





Even at higher aggregation sufficient detail needs to be shown to avoid mis-interpretation

Conclusion

- In projects using complex assays curve viewer is used as a critical tool for decision making and as a complement to numerical end points.
- Opportunities for further improvements
- At AZ dependency on curve datamart, however offers opportunity to use same data for generic data science purposes

acknowledgements

Gostar patent viewer

Dennis Powell (Certara)
Alex Hird
Magnus Polla
Scott Throner
Lars Brive
Rahul Singh
Kalyan Ponamalli

D360 curve viewer

Plamen Petrov
Frank Janssen
David Murray
Kenneth Granberg
Ashvin Dookun
Ana Teixeira

The D360 leadership group:

Andrew Poirrette, Nick Tomkinson, Bill McCoull, Emma Evertsson, Alex Hird

Confidentiality Notice