CCP CompMedChem workshop

10 March 2017

Diamond Light Source

Frank von Delft, Anthony Bradley

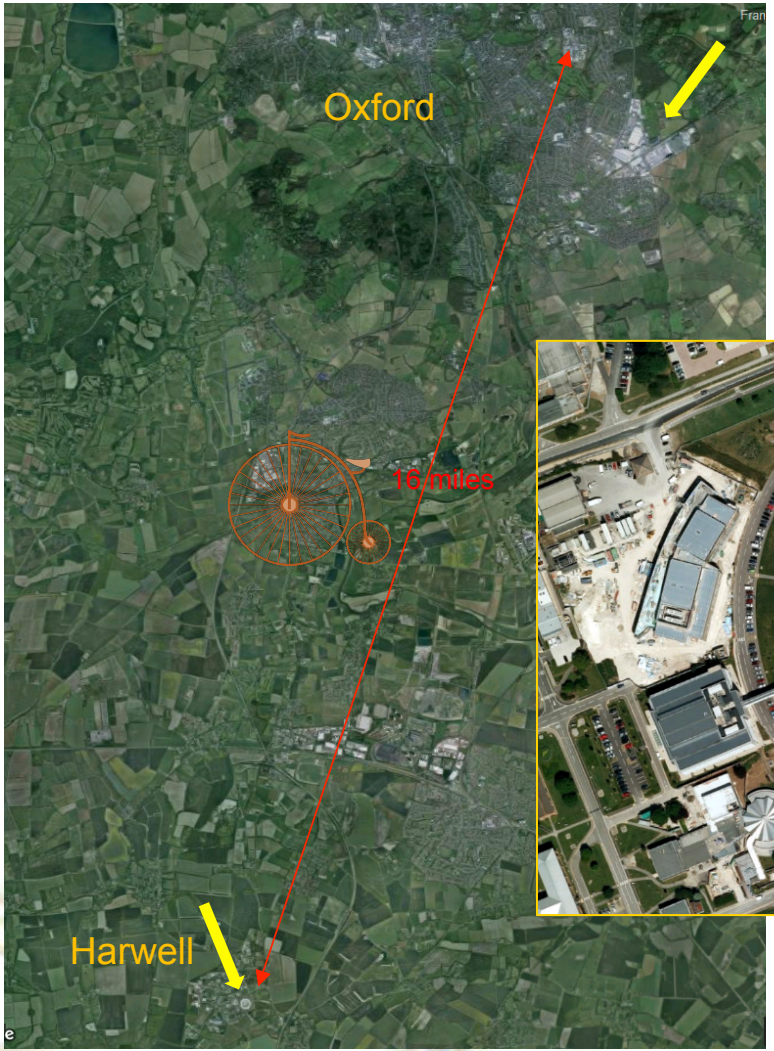
- Introduction
- Scientific challenges
- Infrastructure Challenges
- Future plans

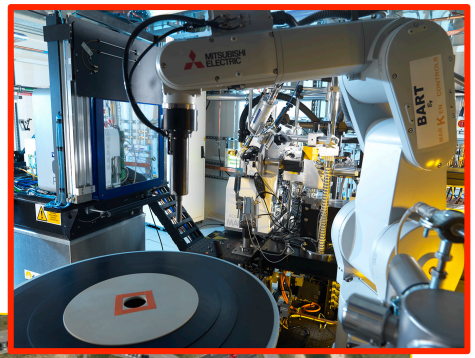
XCHEM Diamond Fragment Screening
SGC University of Oxford
diamond

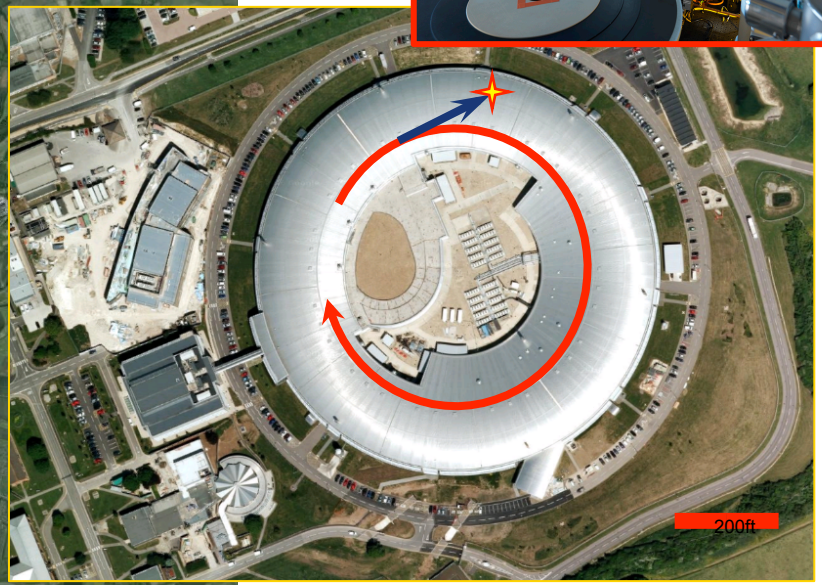
Introduction

Objectives for the day:

- Overall sanity check – are we being daft?
- Scope out ambition
  - Short term: who can already commit? (For InnovateUK submission)
  - Long term: who's missing from the conversation
- Actions and timelines





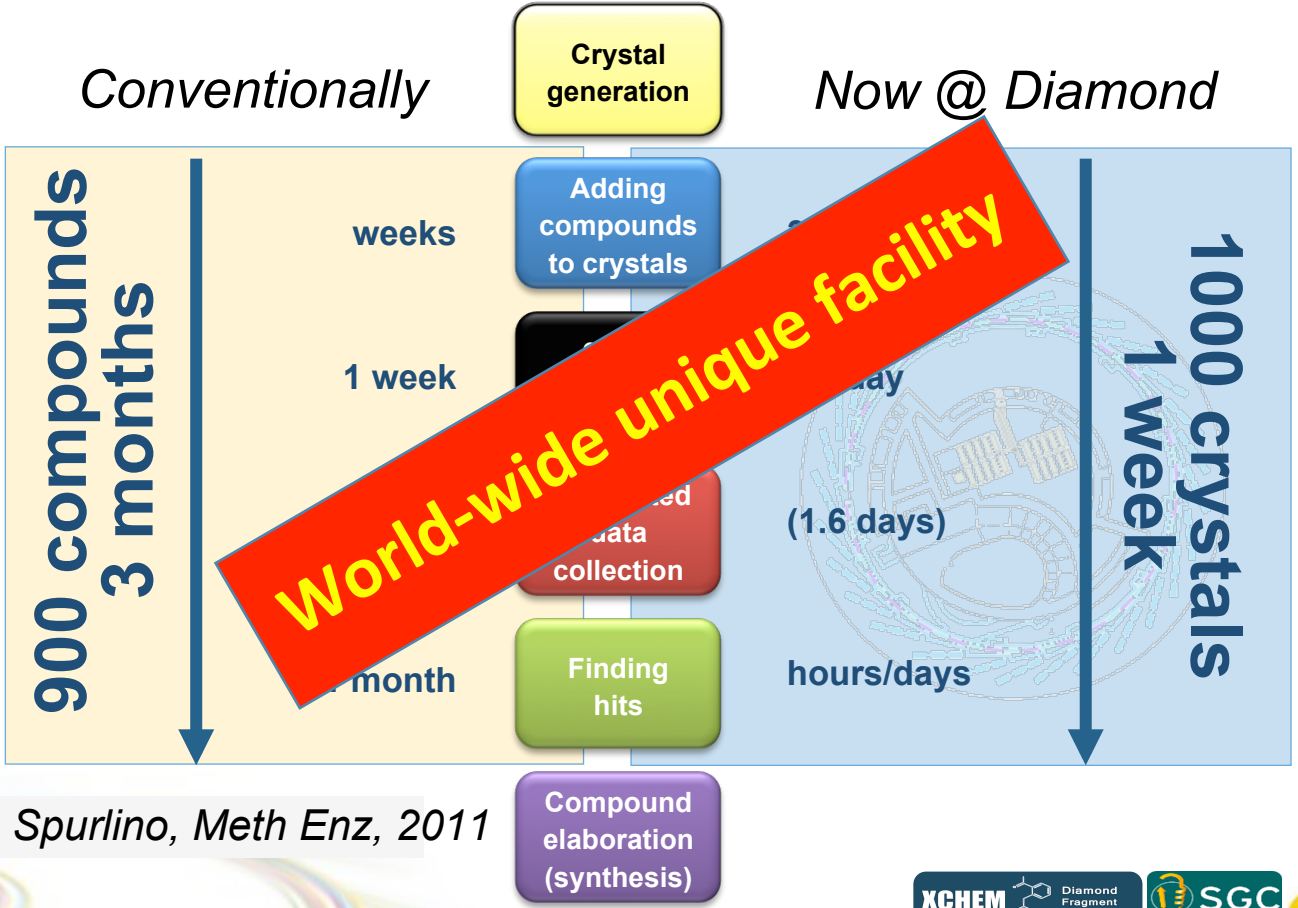


XCHEM Diamond Fragment Screening

SGC UNIVERSITY OF OXFORD

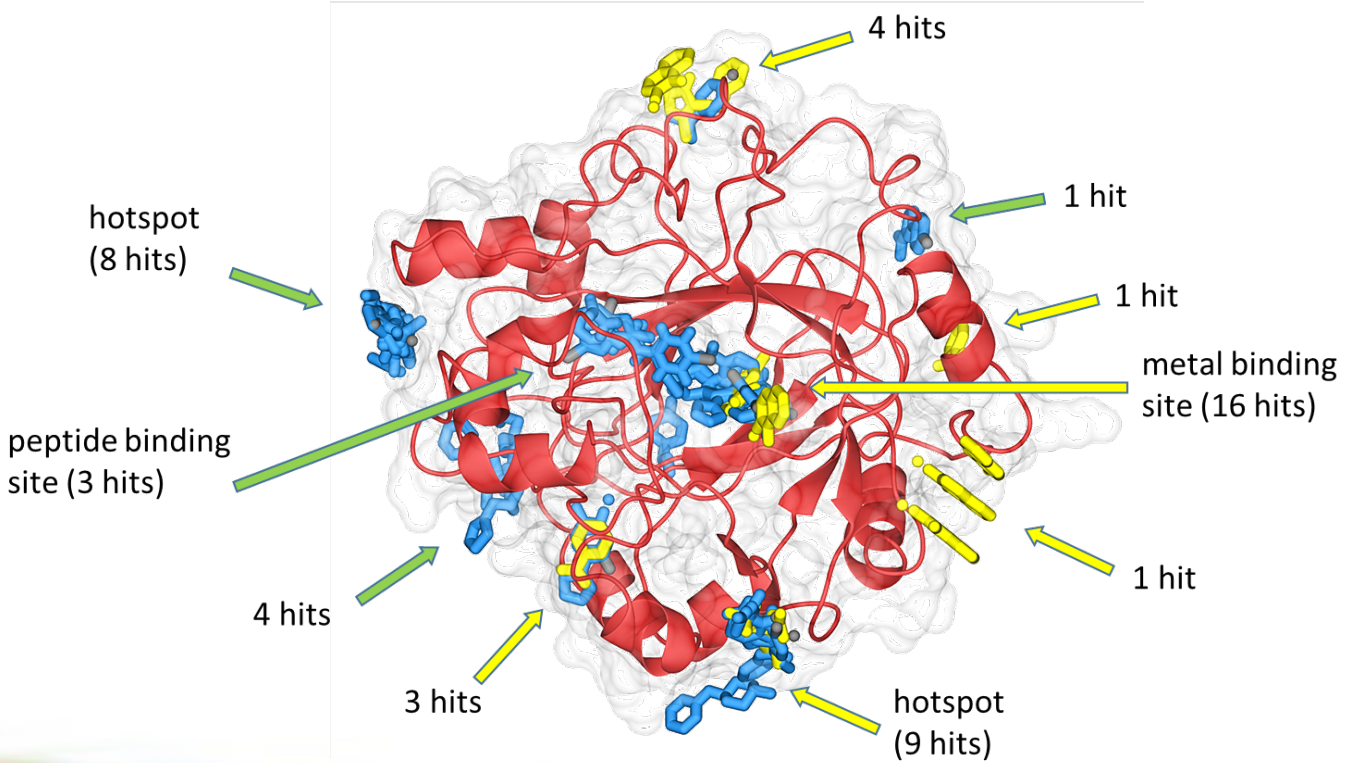
diamond

XChem fragment screening: order of magnitude speedup



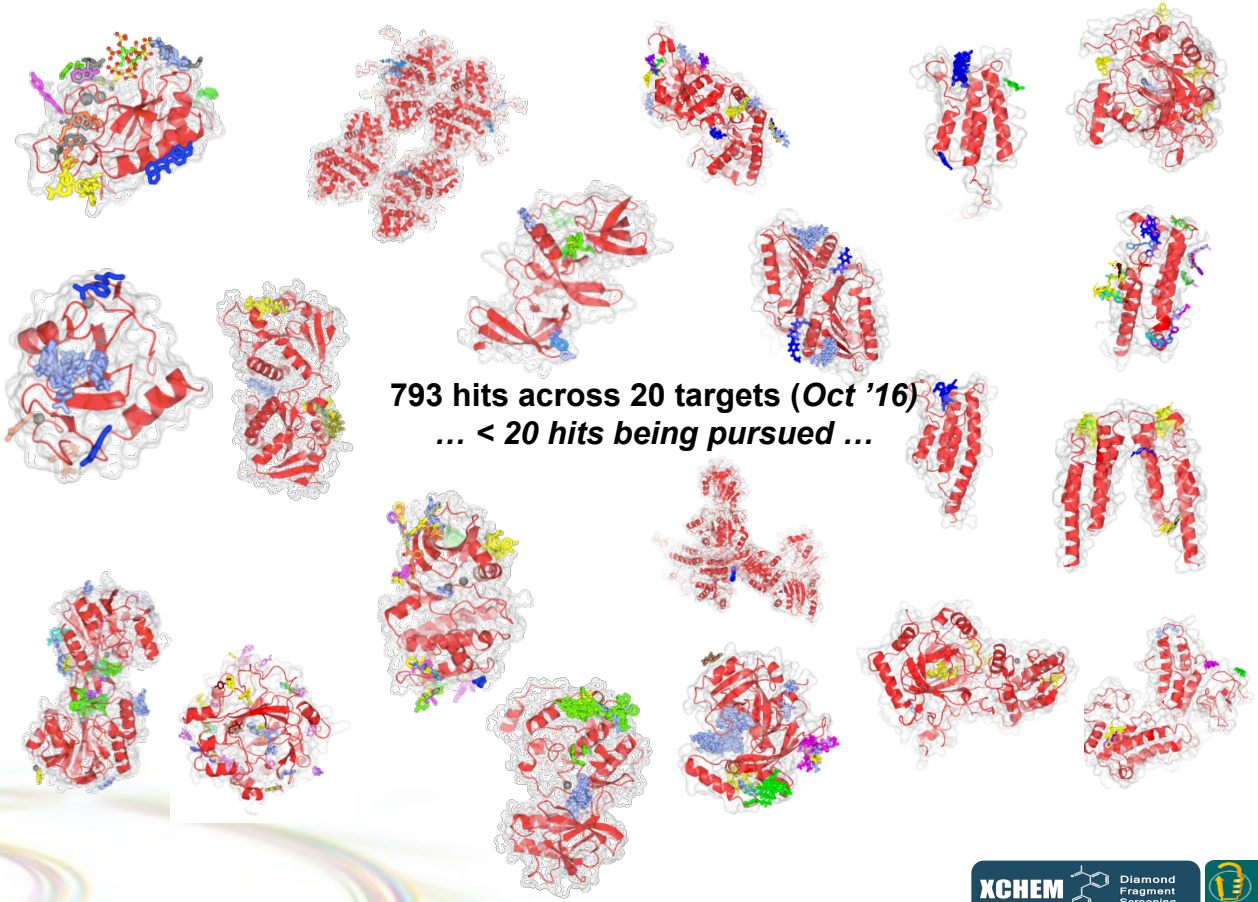
Spurlino, Meth Enz, 2011

The readout is spectacular



JMJD2D, a histone lysine demethylase

… and apparently unusable



793 hits across 20 targets (*Oct ’16*)

… < 20 hits being pursued …

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

XChem Acknowledgements

Diamond: I04-1
José Brandaõ-Neto
Alice Douangamath
Patrick Collins
Renjie (Jay) Zhang

SGC: PX group
Tobias Krojer
Romain Talon
Mike Fairhead
Ritika Sethi
Nathan Wright
Beth MacLean
** Elliot Nelson*
** (Nick Pearce)*

RCaH: XChem Centre
Anthony Bradley
Anthony Aimon
Kannan Velupillai

Diamond:
Industrial liaison
Alex Dias
Elizabeth Shotton
Masses of people...
MX Village, GDA team, Controls, SciSoft, EHCs

Oxford - SGC:
Paul Brennan
** Janine Gray*
** Susan Leung*
** (Oakley Cox)*
Brian Marsden
** Hannah Patel*
Nicola Burgess-Brown
Claire Strain-Damerell

Oxford - Statistics
Charlotte Deane
Garret Morris
Oxford - Chemistry
Martin Smith
Darren Dixon
Chris Schofield
Evotec
Paul McEwan
Mike Bodkin

Mitegen
Ben Apker
UCB
Sebastian Kelm
Jiye Shi
Novartis
Carien Dekker
Markus Kroemer
New York
John Hunt

Leeds
Adam Nelson
Dan Foley
Warwick
Dom Belini
Chris Dowson
Manchester
Allan Jordan

Who are we?

diamond

UNIVERSITY OF OXFORD

XCHEM Diamond Fragment Screening

SABS

SGC

DEPARTMENT OF CHEMISTRY

UNIVERSITY OF OXFORD

Frank von Delft

Anthony Bradley

Garrett Morris

Brian Marsden

XCHEM Diamond Fragment Screening

SGC UNIVERSITY OF OXFORD

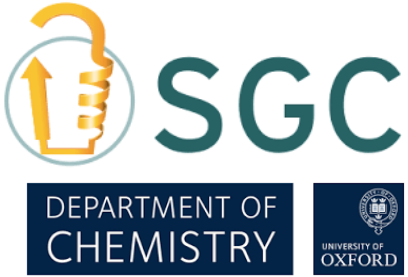
diamond

Who are we?

INFORMATICS MATTERS
Informatics solutions for drug discovery

UNIVERSITY OF OXFORD

Garrett Morris



Frank von Delft

Anthony Bradley

Brian
Marsden

What we struggle with

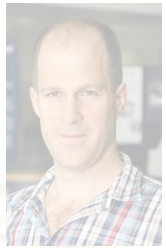
Users

- Nowhere to turn for tools
- Much effort expended
- Solutions still far from perfect
- Each student reinvents the wheel

Providers

- No simple outlet for tools
- Much effort expended
- Solutions still far from perfect
- Lots of wheels reinvented

OX XCHEM ENABLING HIT-TO-LEAD IN FBDD



Frank von Delft

Anthony Bradley

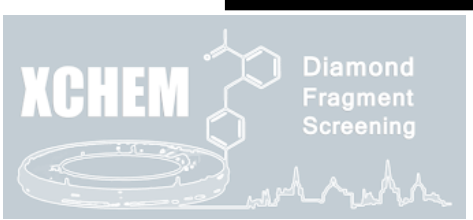
Garrett Morris

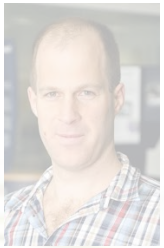
Brian Marsden

What others struggle with (apparently)

Pharma

- Lots of time spent building infrastructure
- Huge challenges to make comp-chem effective
- Limited pre-competive colloboration on challenges





Frank von Delft

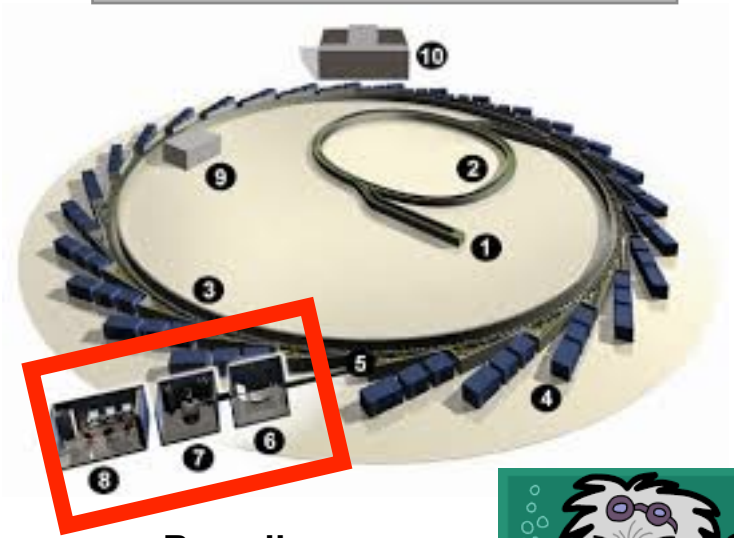
Anthony Bradley

Garrett Morris

Brian
Marsden

Why us?

Synchrotron



Beamline

Hardware and software infrastructure

Hardware

amazon web services

Science & Technology Facilities Council

diamond

Software

RDKit
Open-Source Cheminformatics and Machine Learning

docker

GitHub

CCP CMC

XCHEM Diamond Fragment Screening

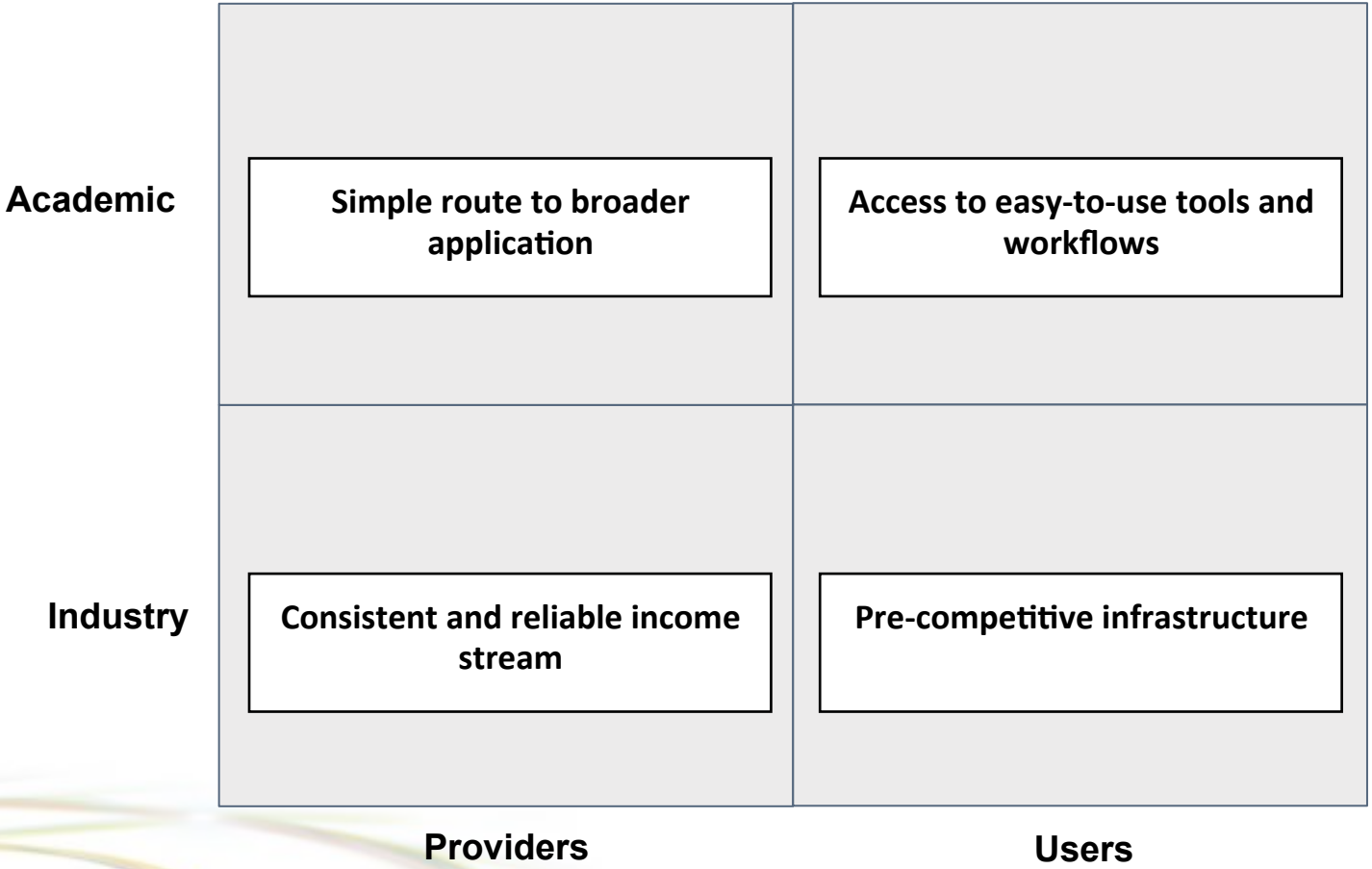
SGC University of Oxford

diamond

Problem overview

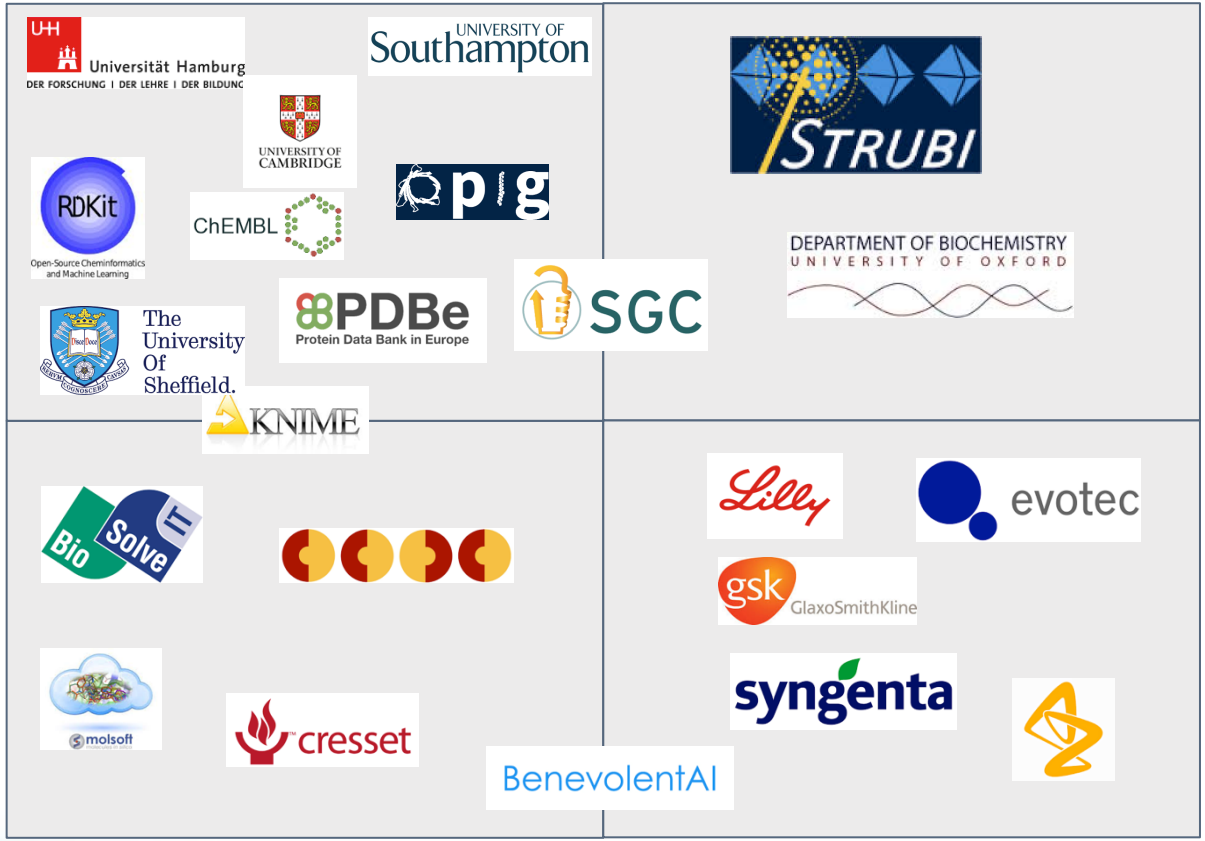
- Academic algorithm development to routine practical use is very challenging
- Duplication of effort within pharma and biotech on building infrastructure and workflows
- Non-expert users - best-practice remains largely inaccessible
- More difficult than it needs to be to run these types of applications on HPC computing environments

Interest groups



Academic

Industry



Providers

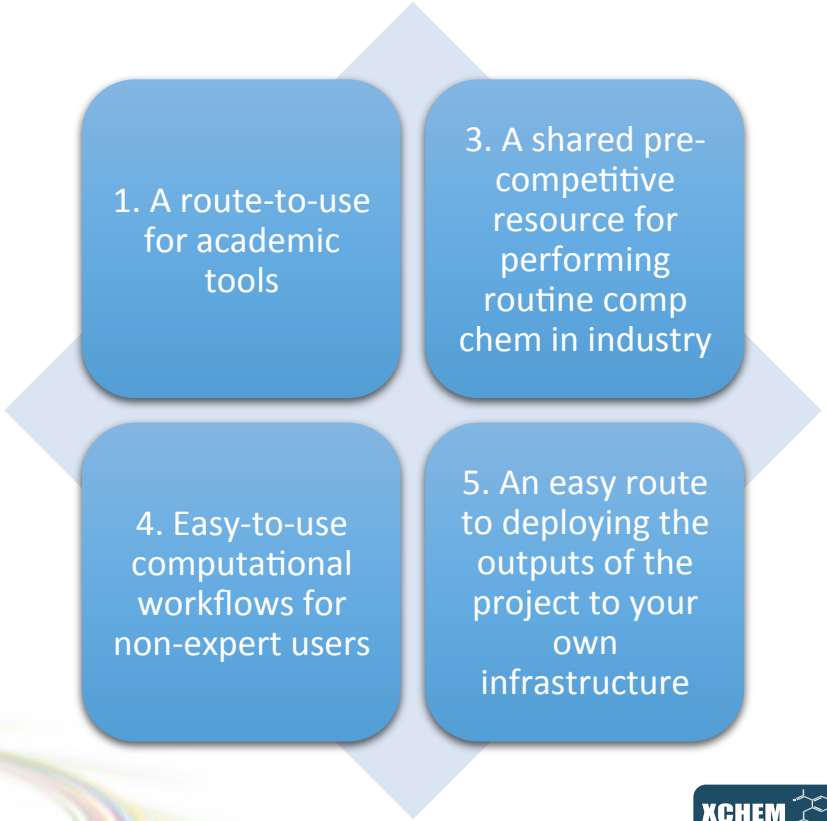
Users

XCHEM Diamond Fragment Screening

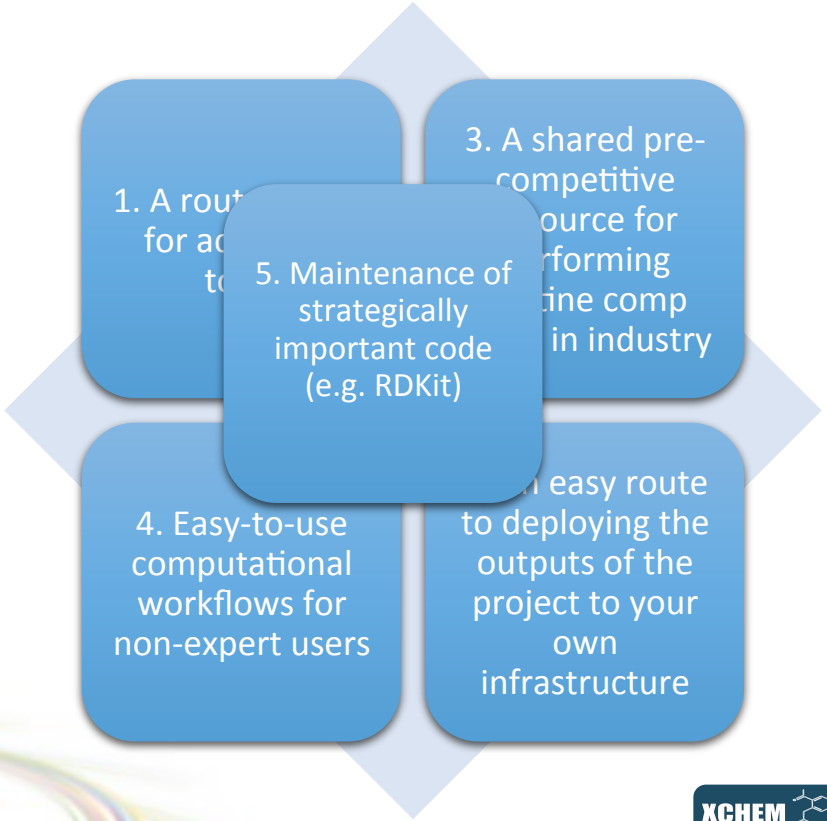
SGC UNIVERSITY OF OXFORD

diamond

The project aims to provide:



The project aims to provide:




Academic
Industry
CCP-CMC
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG
UNIVERSITY OF CAMBRIDGE
RDKit
Open-Source Cheminformatics and Machine Learning
ChEMBL
PDBe
Protein Data Bank in Europe
The University Of Sheffield.
KNIME
TRUBI
RTMENT OF BIOCHEMISTRY
VERSITY OF OXFORD
BioSolveIT
evotec
SmithKline
molsoft
cresset
Providers
Users

Through three mechanisms

1. Critical software and hardware infrastructure
2. Specific workflows and processes
3. Workshops and training

Driven by users not by developers:
- How to get onto the hardware is what we need to address. The synchrotron already does exist (AWS and the cloud).
- Synchrotron as an example of what can be done with focussed community effort.

Academic

Simple route to broader application

CCP-CMC

GitHub

Code

Travis CI

Test

docker

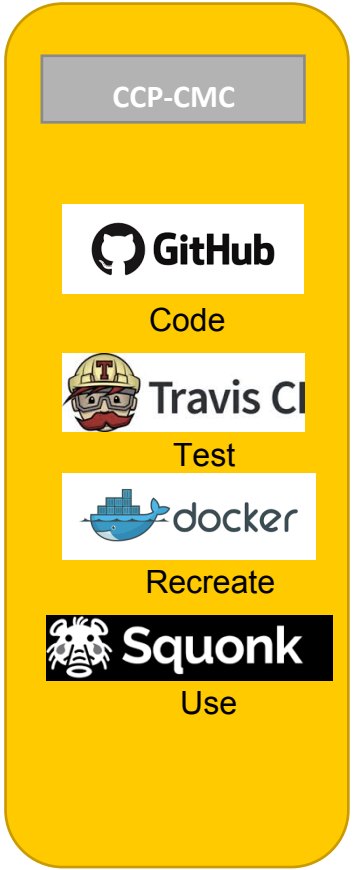
Recreate

Squonk

Use

Providers

Industry



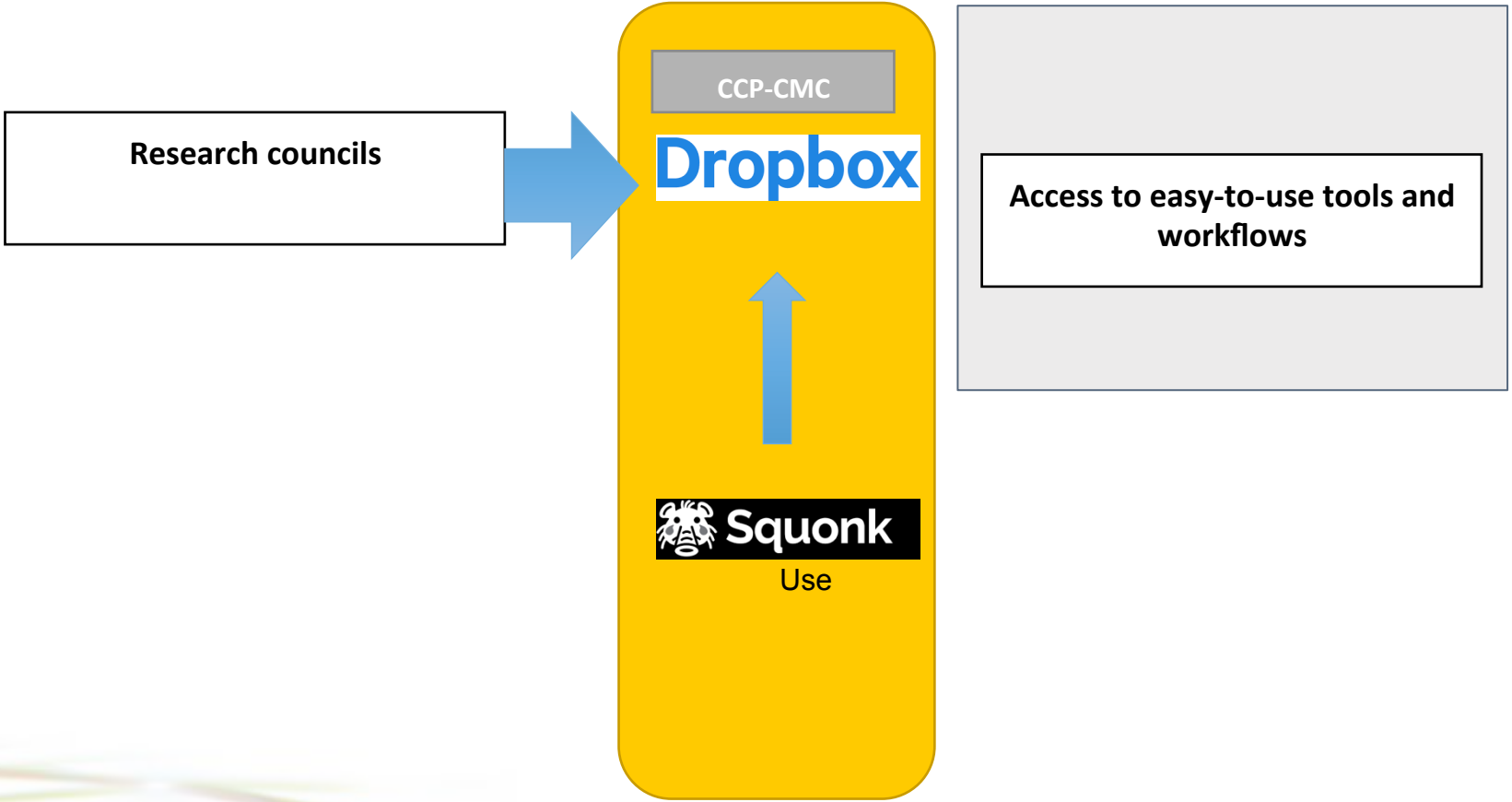
Pre-competitive infrastructure

Users

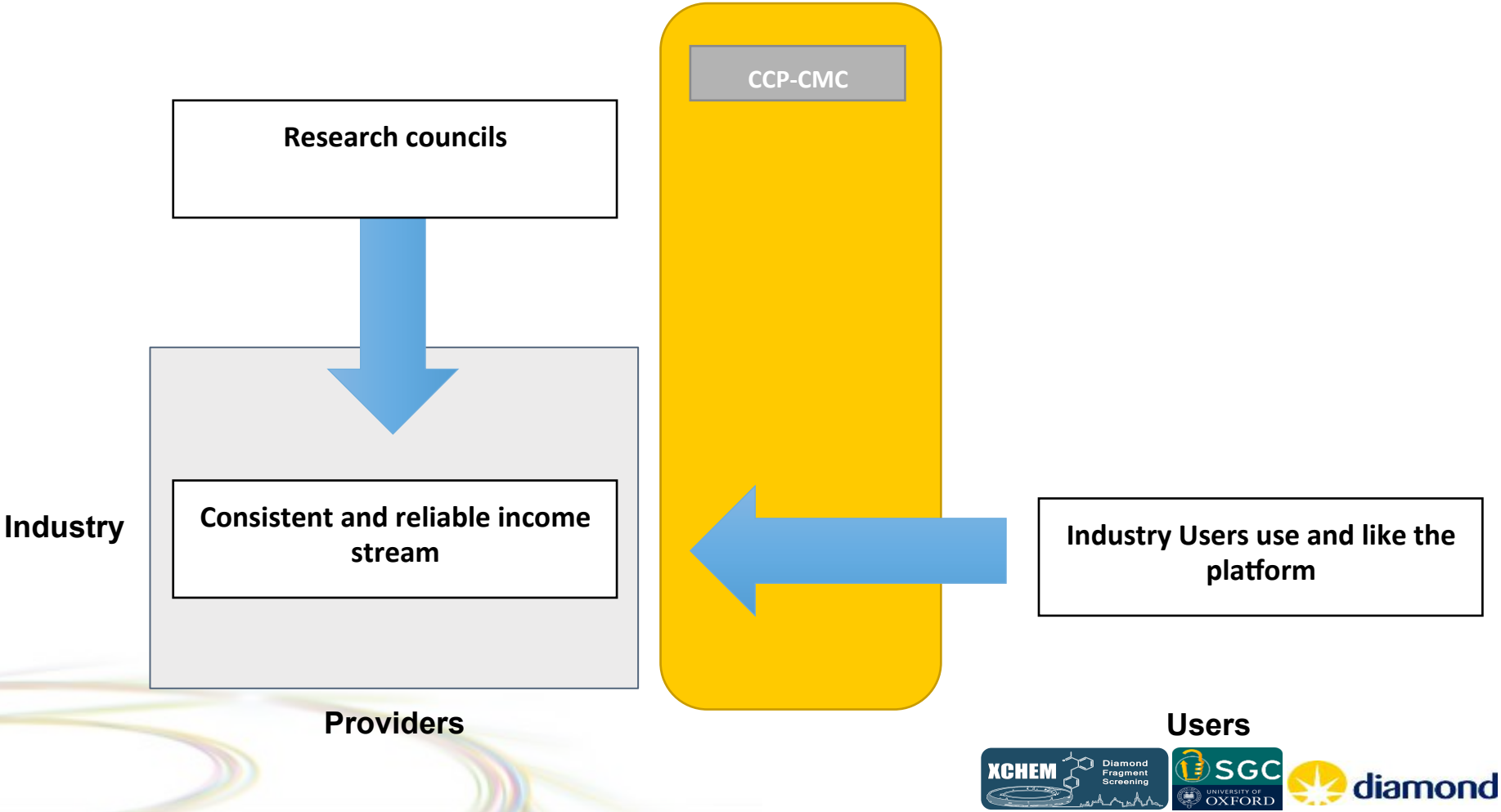
XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond


Research councils
CCP-CMC
Dropbox
Squonk
Use
Access to easy-to-use tools and workflows

Users


CCP-CMC
Research councils
Industry
Consistent and reliable income stream
Industry Users use and like the platform
Providers
Users

Route to funding

- **Plan A:**

Innovate UK

Deadline: April 12th

- **Plan B / even if Plan A:**

EPSRC
Pioneering research and skills

wellcome

BBSRC
bioscience for the future

Newton Fund

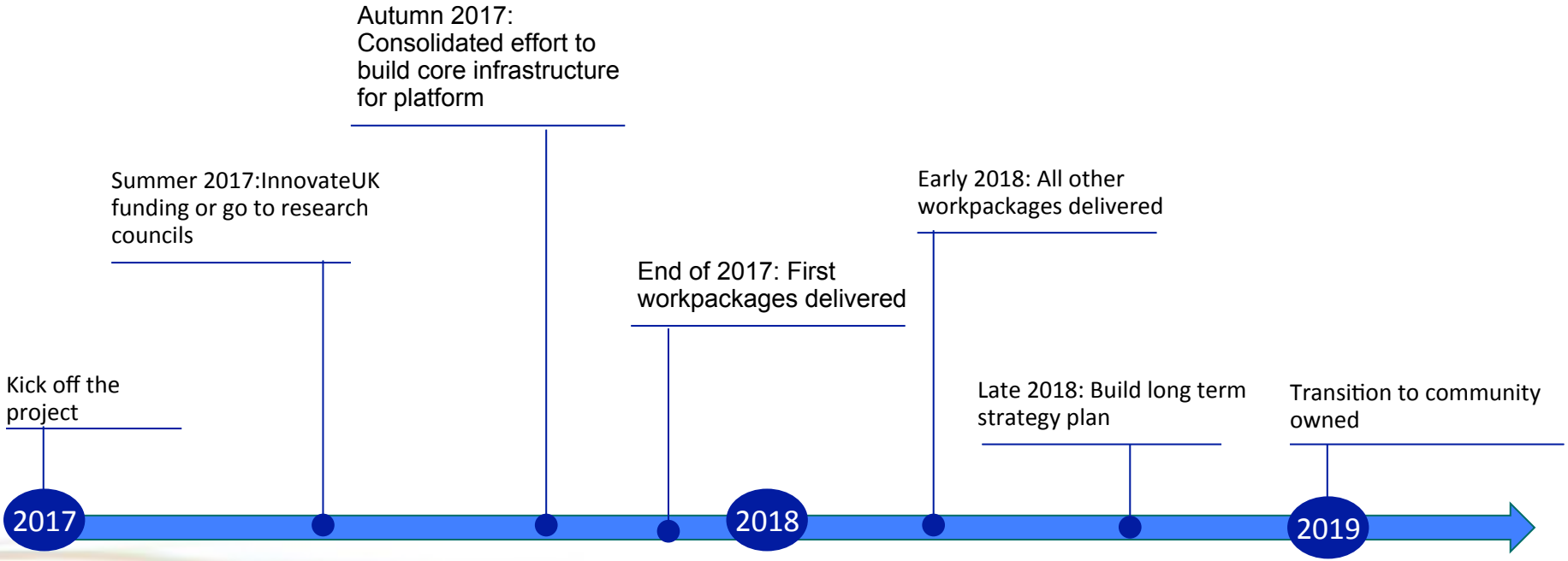
MRC Medical Research Council

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

Timeline


Autumn 2017: Consolidated effort to build core infrastructure for platform
Summer 2017:InnovateUK funding or go to research councils
Early 2018: All other workpackages delivered
End of 2017: First workpackages delivered
Kick off the project
Late 2018: Build long term strategy plan
Transition to community owned
2017
2018
2019

Scientific Goals

Issues with comp-chem

Fragmented marketplace of tools from multiple vendors

Closed development (not all) within firewalled companies / pharma

Route-to-market long and arduous for academic software providers

Perception that problem cannot be solved

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

Advantages of comp-chem

SDF, SMILES, SMARTS

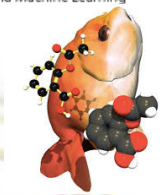
Small datasets – rarely “big data”

CPU intensive – but we’re on the brink of routine feasibility (FEP)

Common data formats – that work

RDKit

Open-Source Cheminformatics and Machine Learning



Good open-source tools for fundamentals

Problems are simple to describe (hard to solve)

Design follow-ups

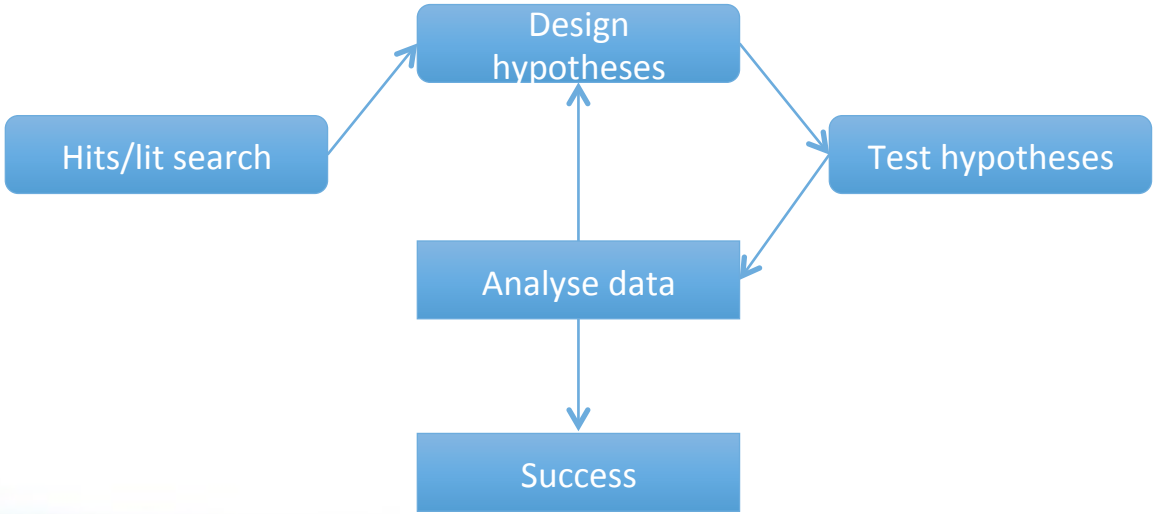
Select follow-ups

XCHEM Diamond Fragment Screening

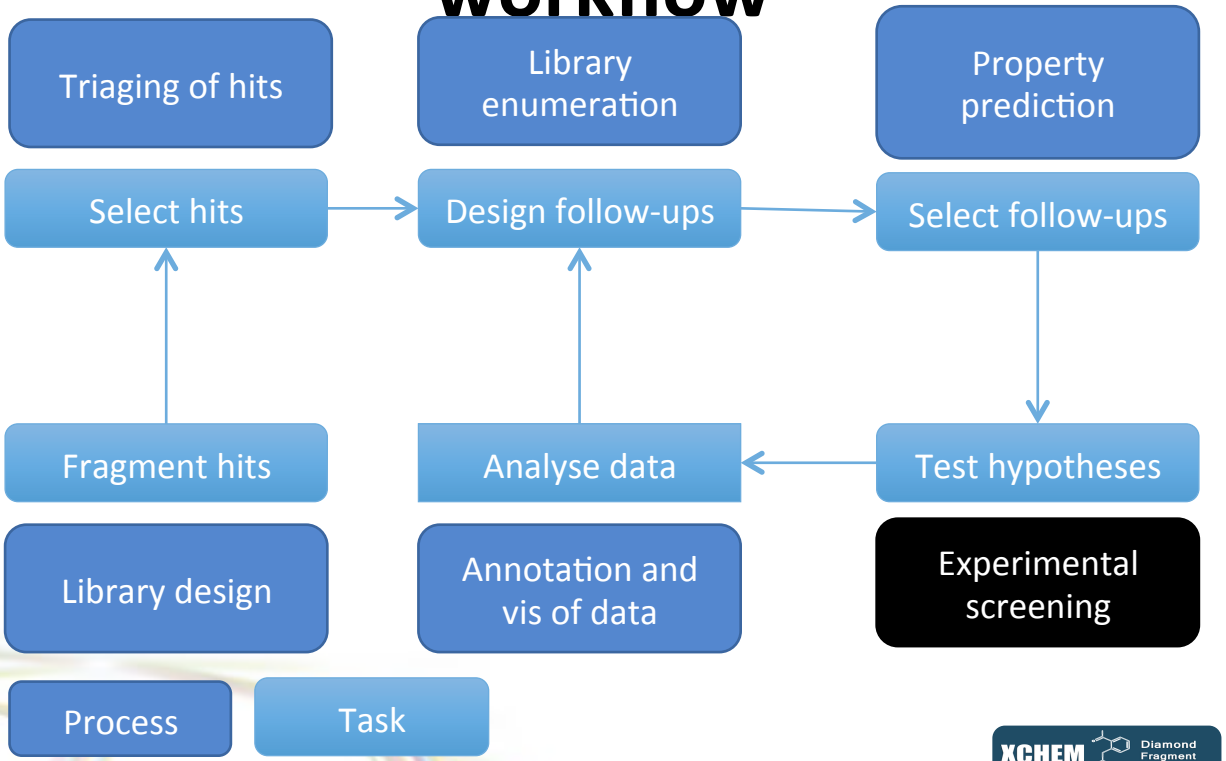
SGC University of Oxford

diamond

Anatomy of the MedChem workflow - Medchem



Anatomy of a MedChem workflow



Science goals

1. Docking algorithms and scoring functions

2. In silico synthetic chemistry

3. Integration and analysis of data:

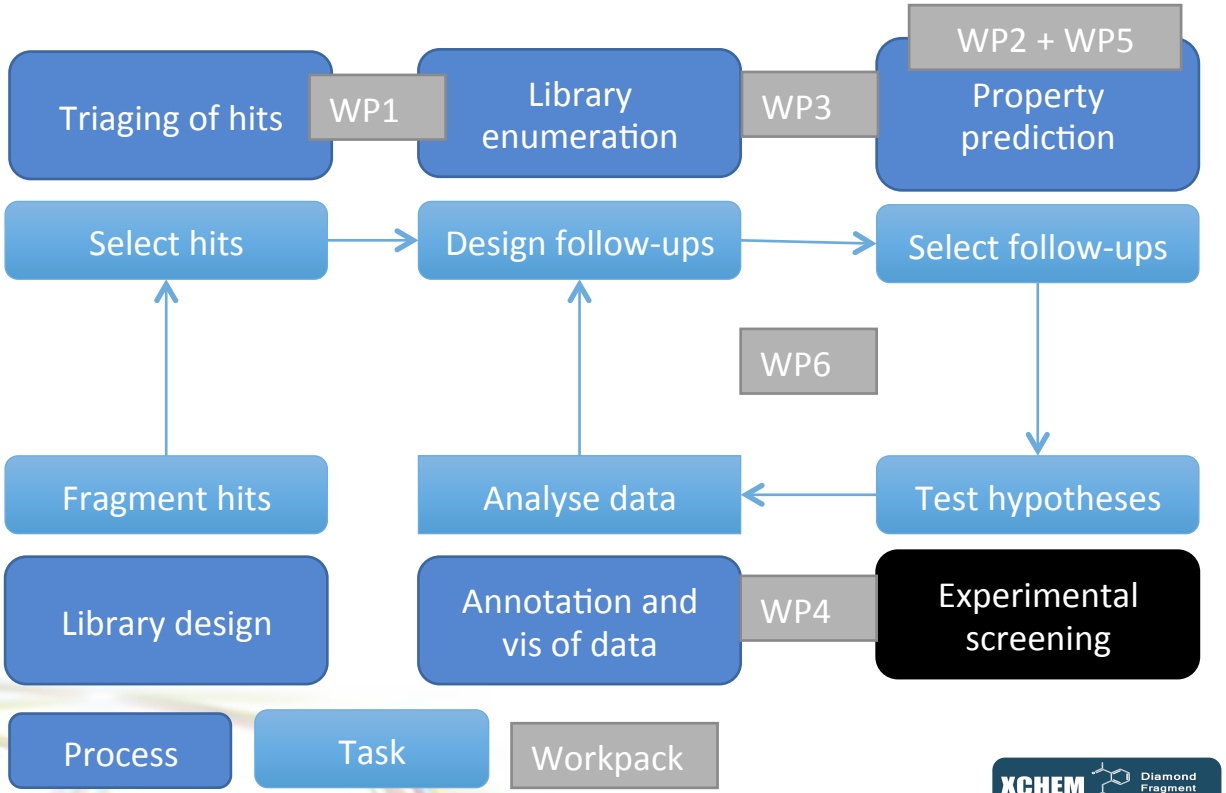
4. Molecular standardisation

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

Anatomy of a MedChem workflow

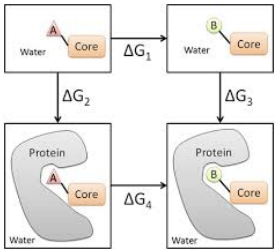


1. Docking algorithms and scoring functions

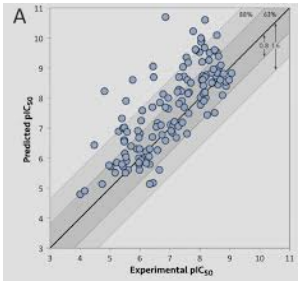
a. Multiple available concurrent methods

b. Methods for setting-up docking experiments

c. Simple methods for parameterisation on own data



d. Fair and open validation and comparison of scores



WP2: Simple access to complex/inaccessible tools - (Pharma, academic groups)

WP5: Validation/comparison of tools (Software providers, pharma)

2. In silico synthetic chemistry

a. Library enumeration

b. Filtering inappropriate reagents

c. Prediction of reaction success

DEPARTMENT OF CHEMISTRY

UNIVERSITY OF OXFORD

Research Complex at Harwell

Working across the life and physical sciences

WP1: Collaborative compound design capture

XCHEM Diamond Fragment Screening

SGC UNIVERSITY OF OXFORD

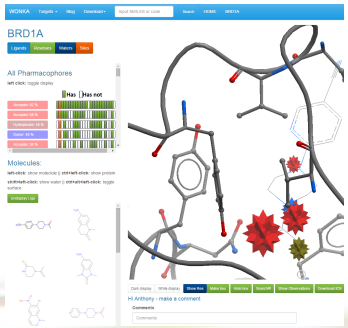
diamond

3. Integration and analysis of data:

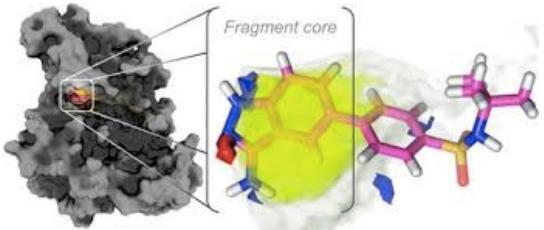
a. Structural ensembles

b. Activity data

c. Third-party computationally derived data



WP4: Visualisation and integration of 3rd party derived data and tools



XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

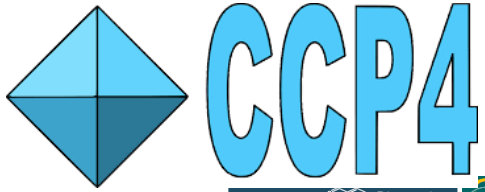
4. Molecular standardisation

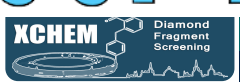
a. Tautomer enumeration

b. Protonation state standardisation

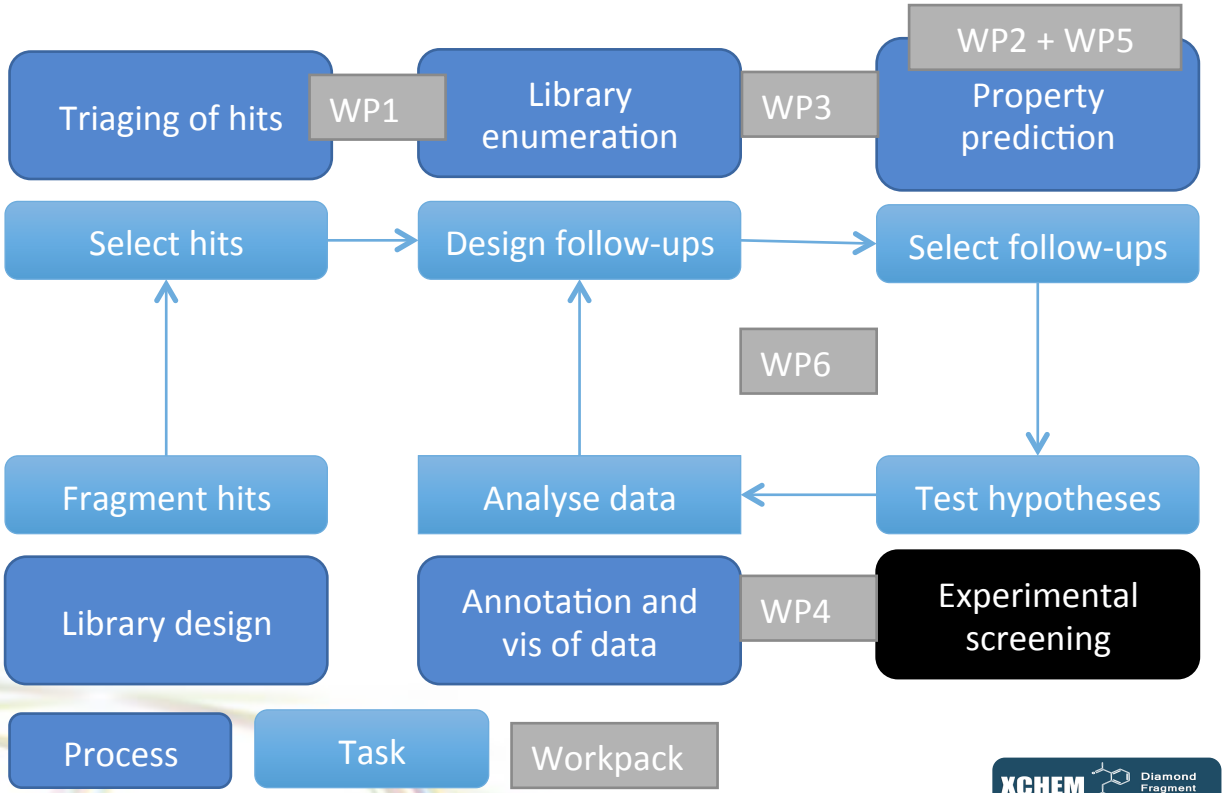
c. Conversion between file formats

WP3: Robust workflows for fiddly (but routine) processes



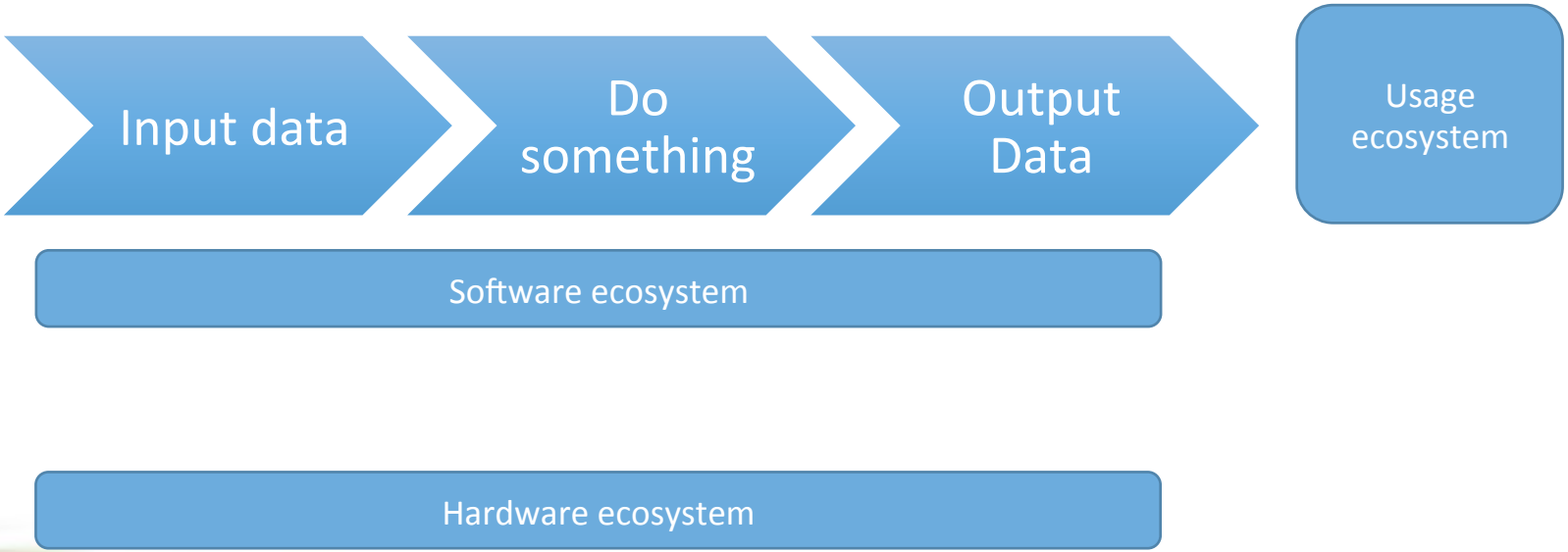


Anatomy of a MedChem workflow



Infrastructure Goals

Anatomy of a tool



Infrastructure goals

1. Simple registration / route-to-use of academic tools

2. Commercial infrastructure for supporting pay-for-use tools

3. Collaboration

4. Hardware support

5. Tools and support for non-experts

6. Maintenance of important community codebases

Infrastructure goals

1. Simple registration / route-t… acader…

2. Commercial infrastructure for …g p… …toc…

…

4. Ha… support

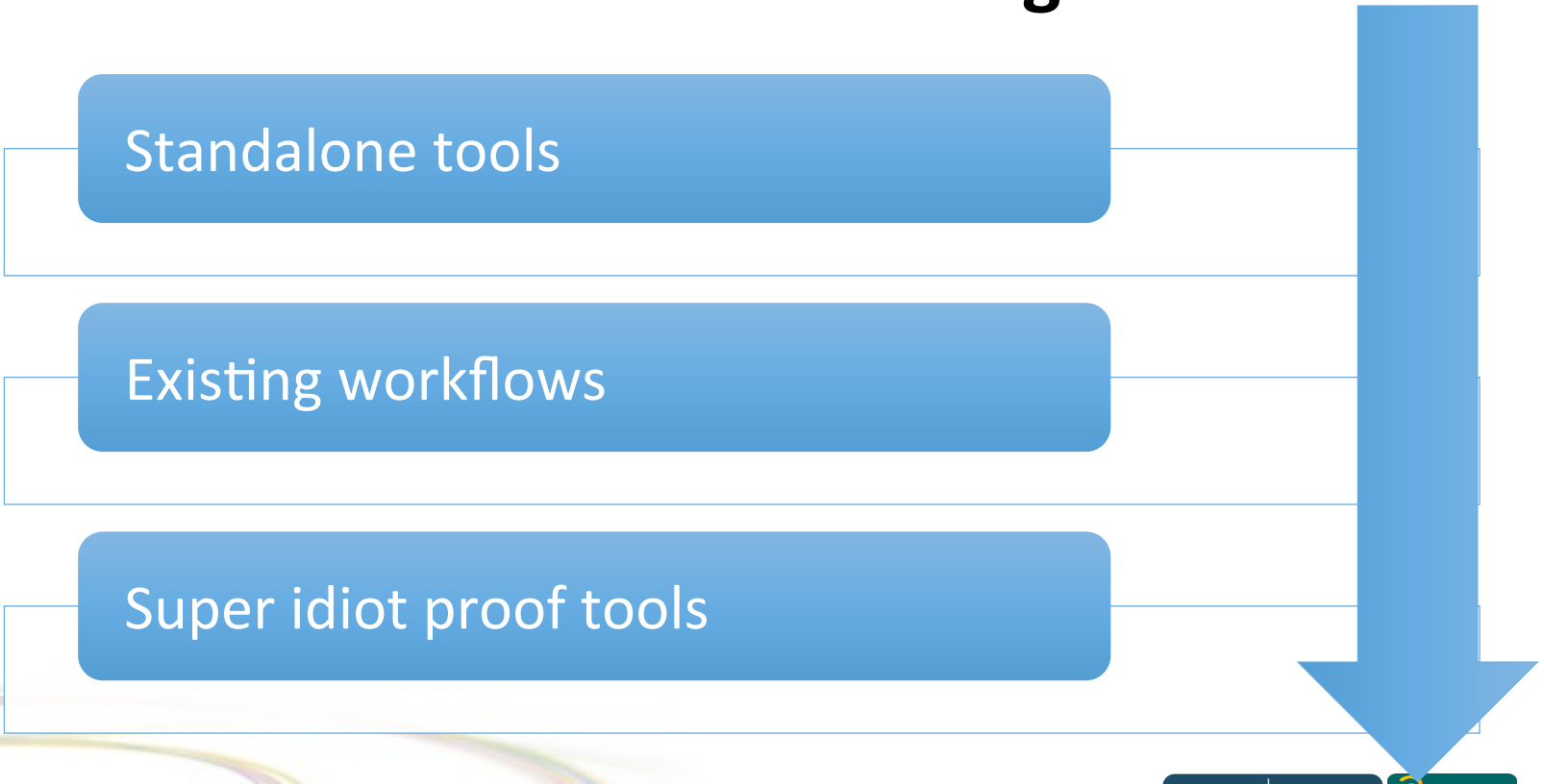
…ar… …support for non… experts

…of… community codebases

Easier to use

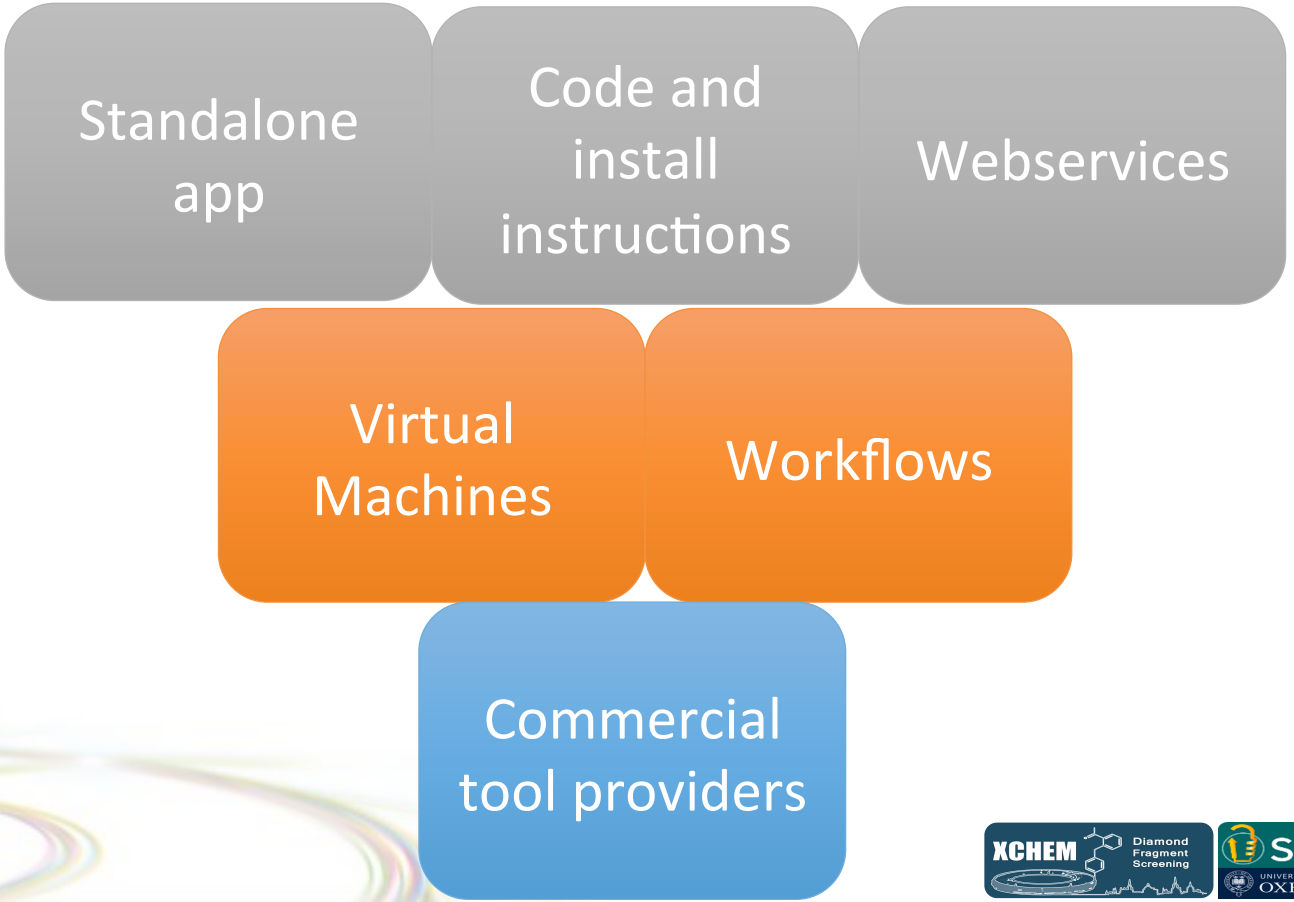
Easier to have-used

Infrastructure goals

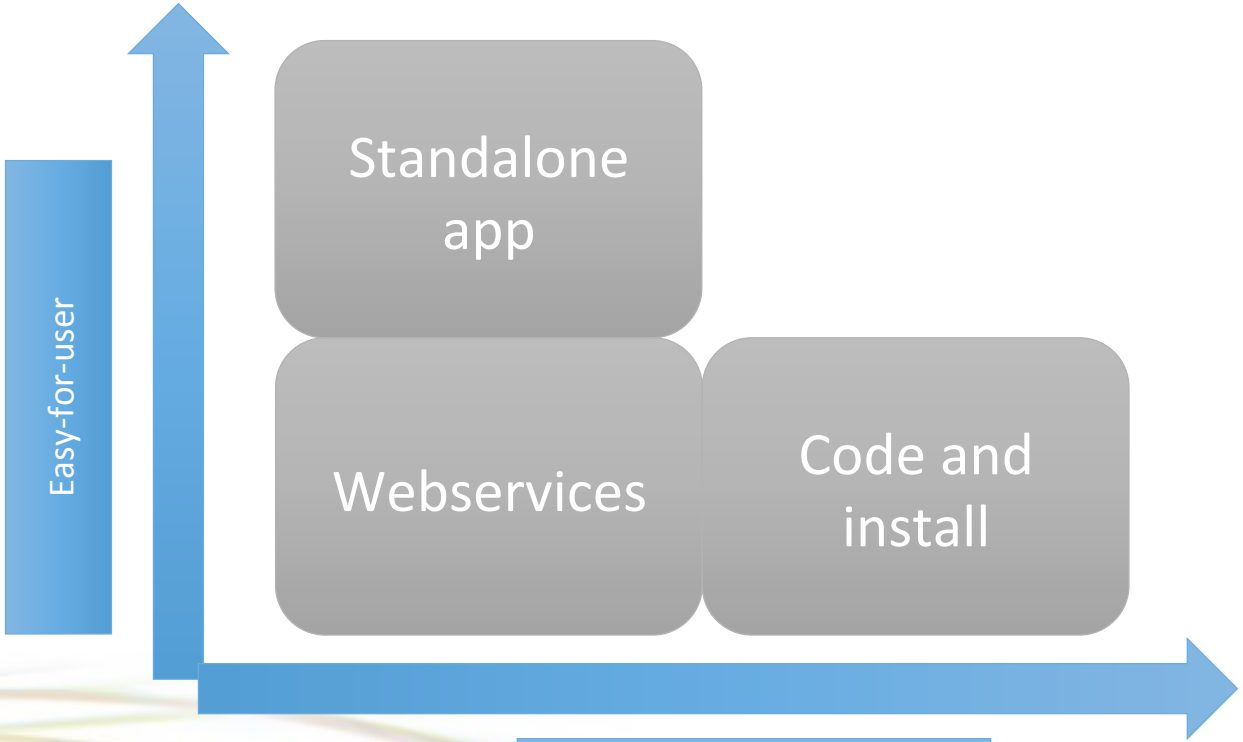


XCHEM Diamond Fragment Screening
SGC University of Oxford
diamond

Current state of tools



Academic tools (personal experience)



VM's and Workflows

Virtual Machines

- High maintenance
- Hard to interconnect
- Too large to download many

Workflows

Open for Innovation KNIME

Pipeline Pilot

- Arbitrary interconnections
- Not taken off for academic tools
- Not aimed at total outsiders

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

Commercial provider

Simple uses

How do I interconnect with others?

How do I access without £30K?

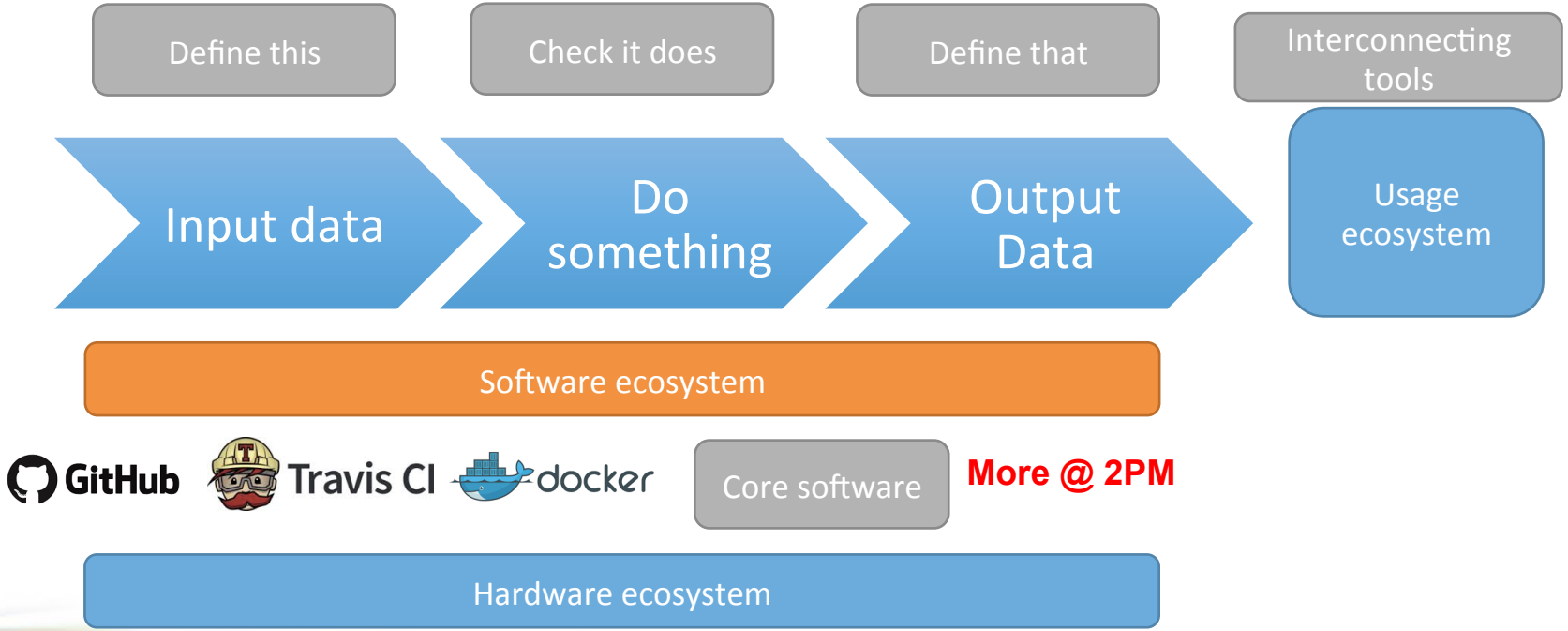
Commercial tool providers

Complex Problems (FEP)

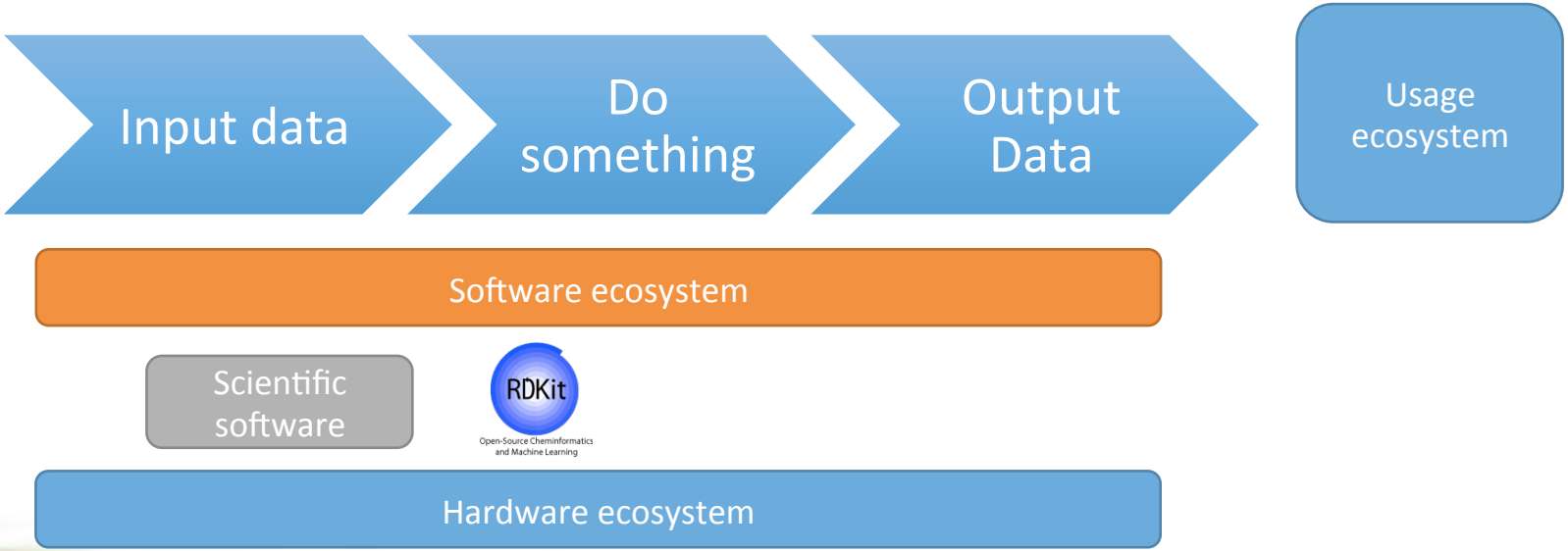
Is this the best solution?

Is this the most cost-effective way?

Simple registration / route-to-use of academic tools



Maintenance of important community codebases



Comparison

1 or 2 of such projects

Mechanism for 100's of such projects to be **added and connected** trivially

SMoG2016

Conformer generators

AutoDockVina

Reaction Enumerators

Commercial infrastructure for supporting pay-for-use tools

Software ecosystem

Scientific software

cresset

BioSolve

Hardware ecosystem

1) Squonk to be monetised – supporting development
2) Enable pay-for-use tools to be included

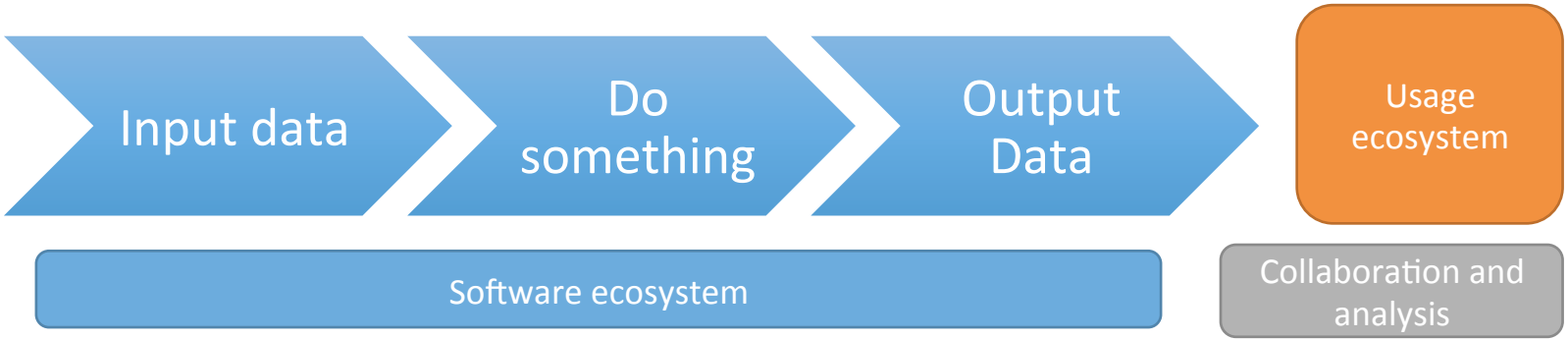
Diamond Fragment Screening

SGC University of Oxford

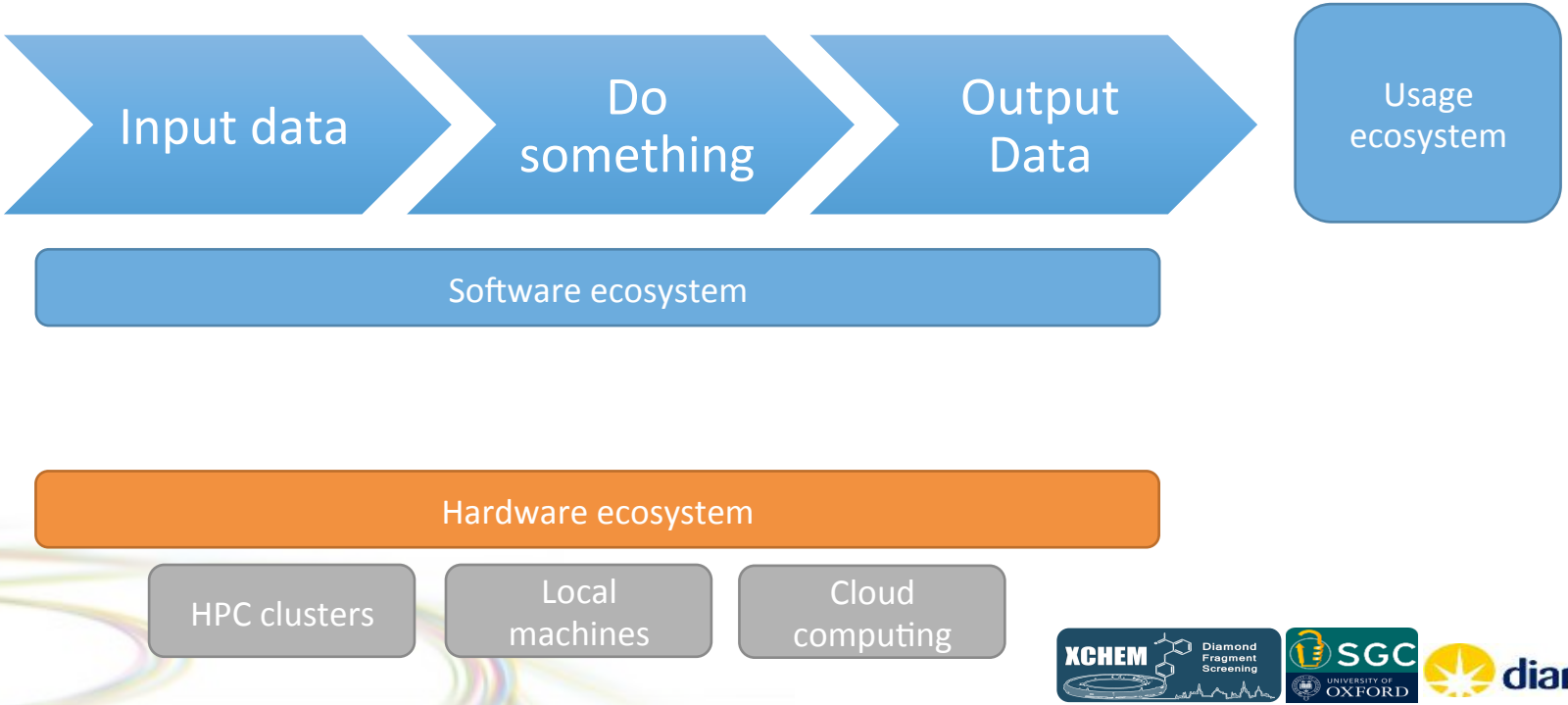
diamond

Collaboration

1) Best practice can be shared
2) Transparent and reproducible
3) Multi-disciplinary teams



Hardware support

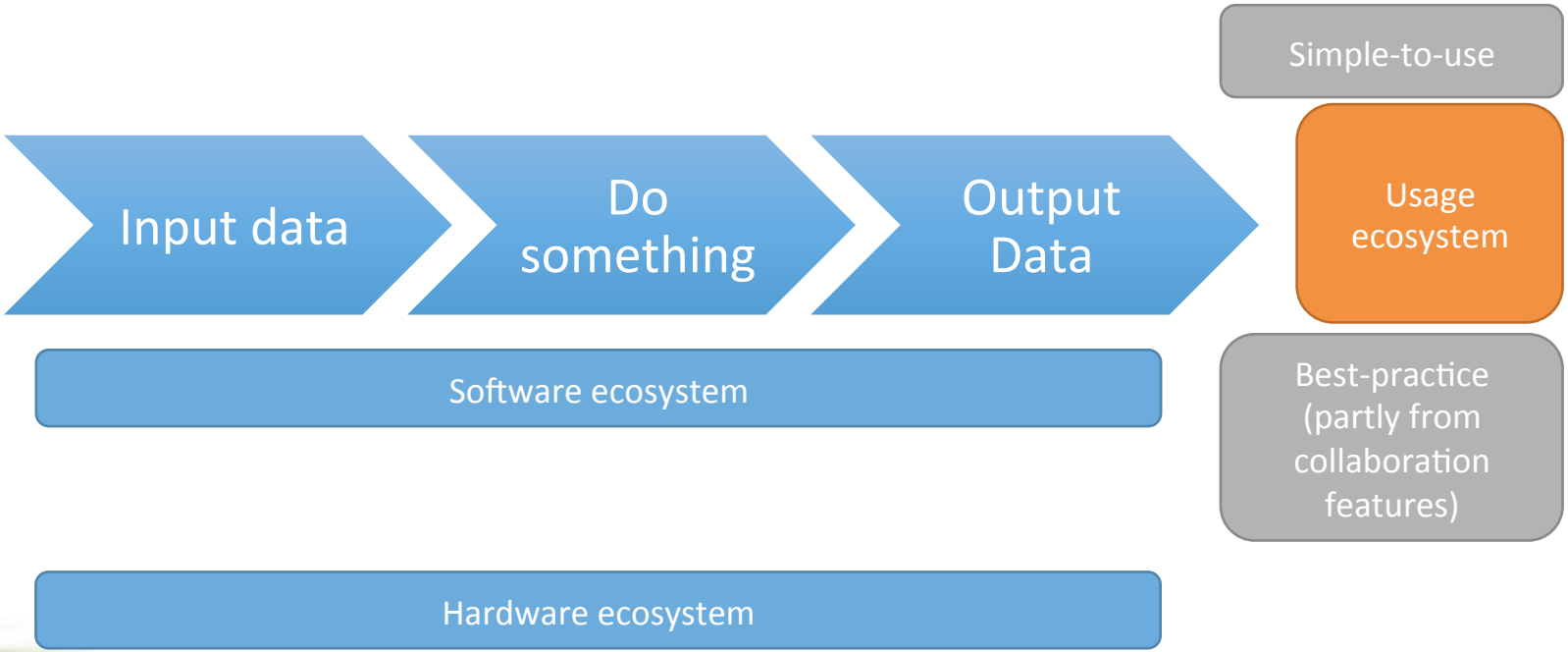


Cloud support

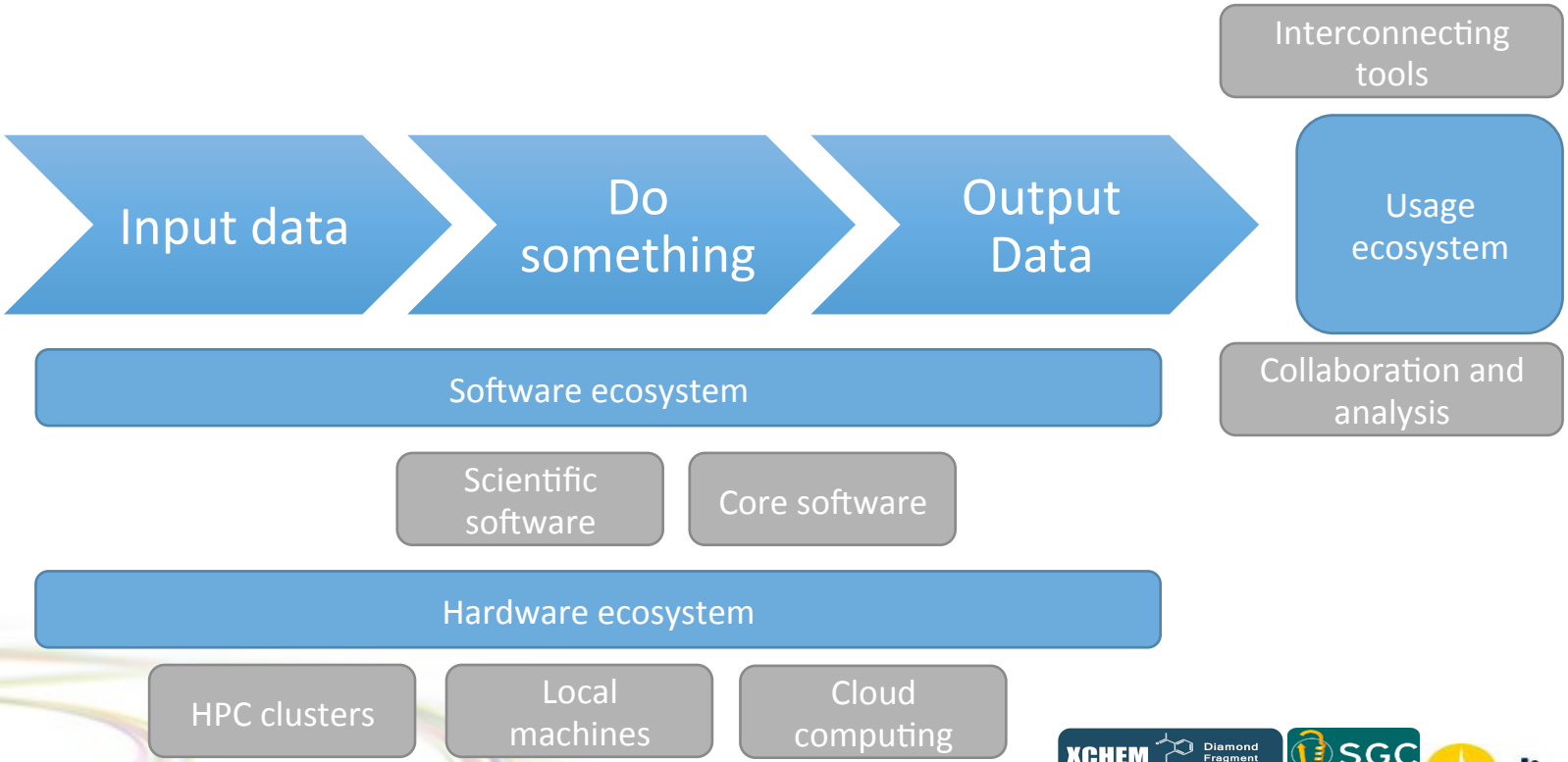
Provisionally £30K of credits (to be confirmed)

- Safe space to try out cost-benefit on open-data

Tools and support for non-experts



Infrastructure goals



Future plans

Selling Points

Leverage

New science

Long-term stability

XCHEM Diamond Fragment Screening

SGC University of Oxford

diamond

Budget

Company	Company give (£K)	InnovateUK Give (£K)	Total (£K)	Funded	Already allocated	FTEs*	FTEs / year**
Diamond	£100	£100	£200	50%	£100	1.3	0.7
IM	£30	£70	£100	70%	£0	1.3	0.7
LARGE	£30	£30	£60	50%	£0	0.8	0.4
LARGE	£30	£30	£60	50%	£0	0.8	0.4
MEDIUM	£30	£45	£75	60%	£0	1.0	0.5
SMALL	£30	£70	£100	70%	£0	1.3	0.7
Company Total	**£250**	**£345**	**£595**	**58%**	**£100**	**6.6**	**3.3**
Academic	£0	£255	£255	100%	£0	3.4	1.7
Total	**£250**	**£600**	**£850**	**70%**	**£100**	**10.0**	**5.0**

Table Two: A proposed funding scheme for the InnovateUK project.

*Price per FTE £75K ** Two year project

diamond

Budget and workpackages

Company	Company give (£K)	InnovateUK Give (£K)	Total (£K)	Funded	Already allocated	FTEs*	FTEs / year**
Diamond	£100	£100	£200	50%	£100	1.3	0.7
IM	£30	£70	£100	70%	£0	1.3	0.7
LARGE	WP1: Collaborative compound design capture					.8	0.4
LARGE	WP2: Simple access to complex/inaccessible tools					.8	0.4
MEDIUM	WP3: Robust workflows for fiddly (but routine) processes					.0	0.5
SMALL	WP4: Visualisation and integration of 3rd party derived data and tools					.3	0.7
Company Total	WP5: Validation/comparison of tools (Software providers, pharma)						
	£250	**£345**	**£595**	**58%**	**£100**	**6.6**	**3.3**
Academic	£0	£255	£255	100%	£0	3.4	1.7
Total	**£250**	**£600**	**£850**	**70%**	**£100**	**10.0**	**5.0**

Table Two: A proposed funding scheme for the InnovateUK project.

*Price per FTE £75K ** Two year project

diamond

Management structure (proposal)

- **Initial model:**
  - Contribute - you're on the board.
  - Committee meets regularly.
  - Chaired by Diamond (XChem).
  - PM reports directly to the chair (Frank).
  - Management committee can co-opt people in the community – if they're actively involved

- **Future model:** Community / executive community. Driven by the users not the developers. Look to CCPs (*esp.* CCP4)

Project management

Driven by:

- Users
  - Pharma/academic users drive what infrastructure is necessary
- Deliverables
  - Quantifiable targets – short, medium and long-term
- Timelines
  - Projects need to be answerable to completion dates

All ensures tools **speak to** and **answer** a current unmet need

Academic involvement

- Labs that commit to:

full deployment

overall project management discipline (bug tracking, schedules, deliverables, etc)

“Expertise" is not the criteria - just commitment to seek out / coordinate / deploy what exists

Next steps….

- InnovateUK funding proposal
- Identify other funding routes
- Establish network:
  - Industry that can contribute
  - "*active*" academics