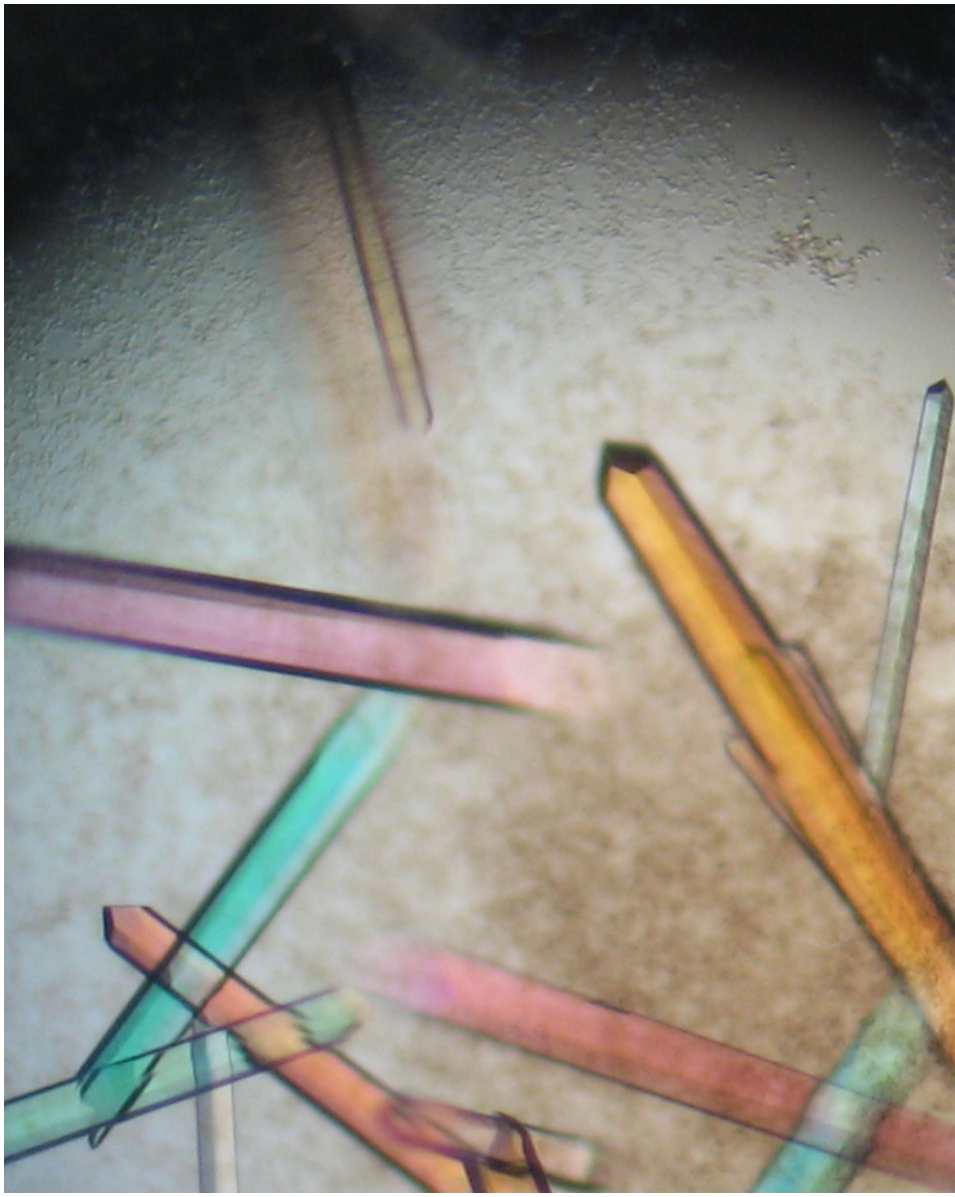
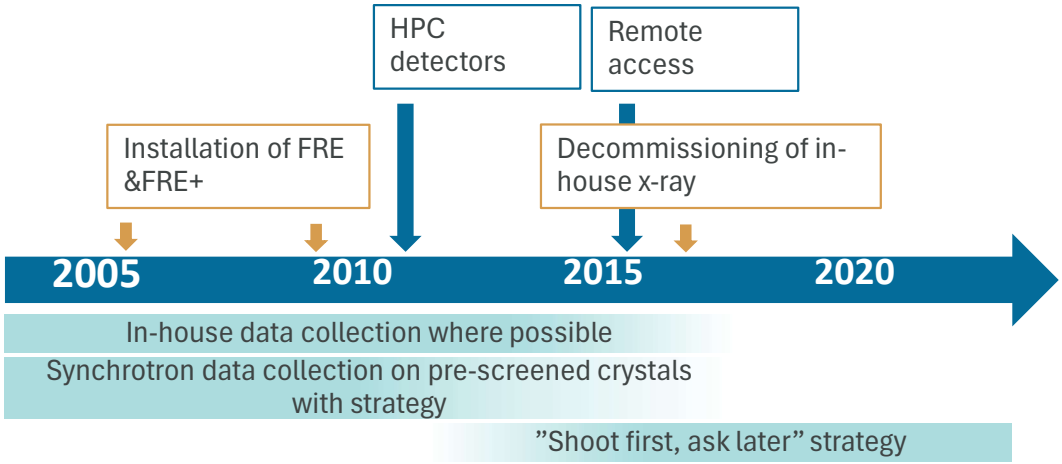
AstraZeneca

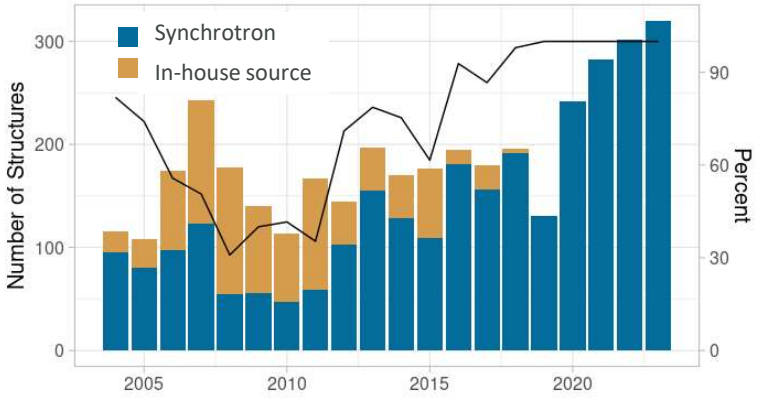
Synchrotron Data Workflow AstraZeneca SE

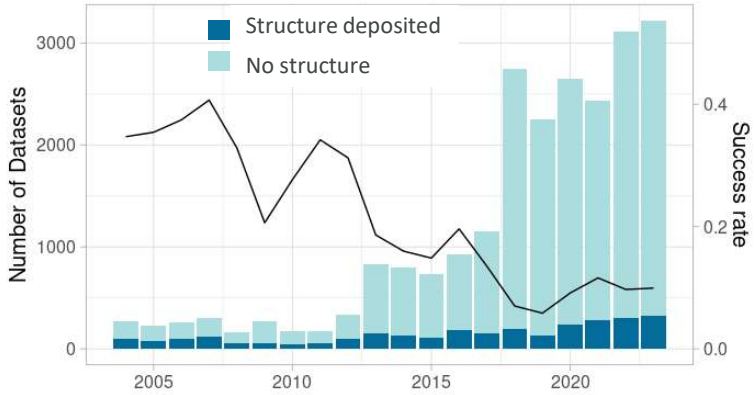
Patrik Johansson



Historical overview 2004-2023



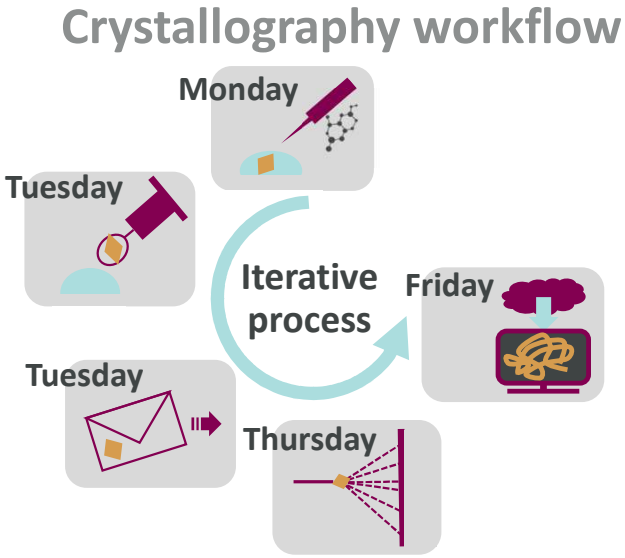




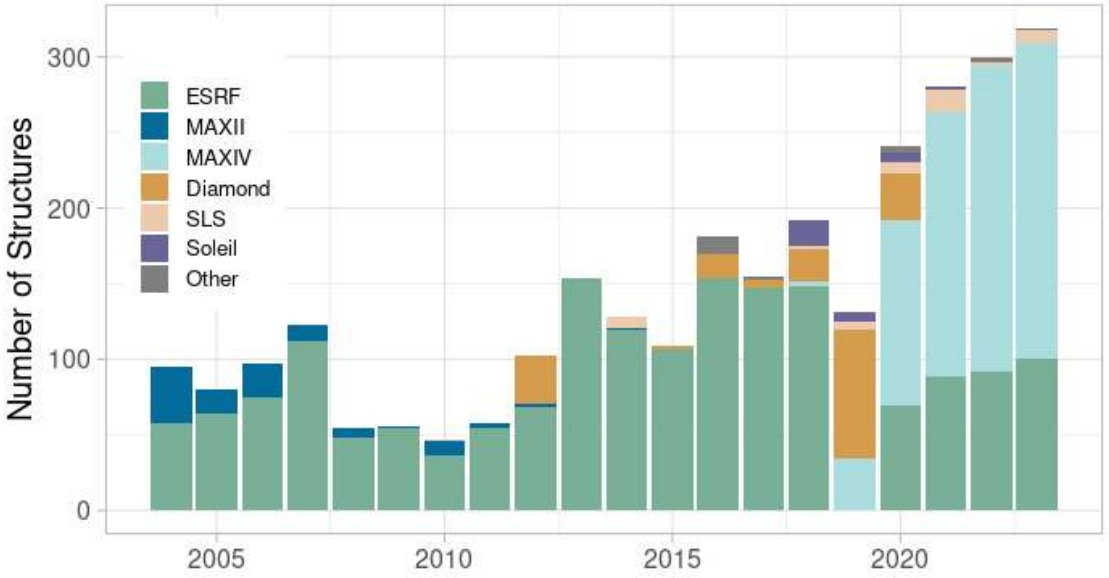
- 3717 structures determined from in-house x-ray or synchrotron
- 22972 X-ray datasets collected at synchrotrons, including 20079 which did not result in a structure

Figures adapted from J Synchrotron Radiat. 2025 Feb 6;32(Pt 2):294–303

Synchrotrons play a critical role in our workflow



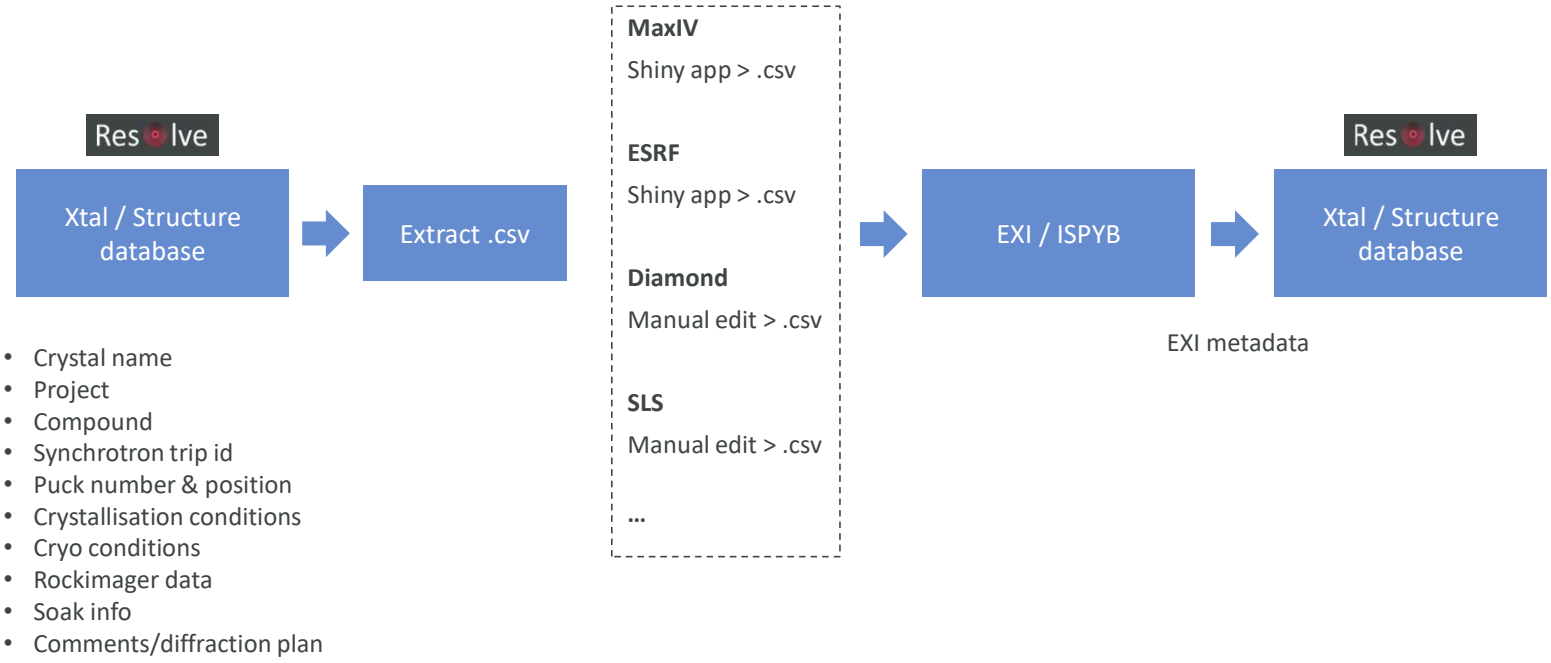
- Near-weekly synchrotron trips



- Using multiple synchrotrons to ensure coverage around the year

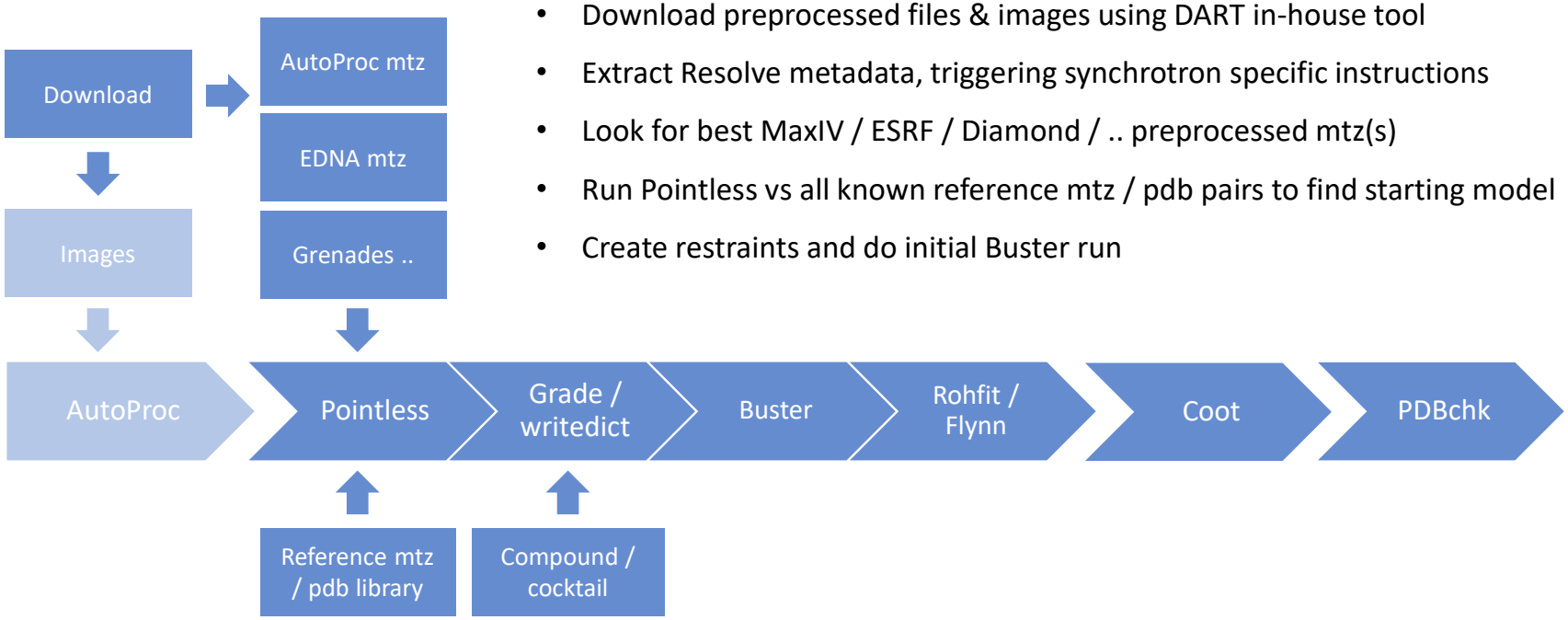
Figures adapted from J Synchrotron Radiat. 2025 Feb 6;32(Pt 2):294–303

Experiment setup & metadata export



- Crystal name
- Project
- Compound
- Synchrotron trip id
- Puck number & position
- Crystallisation conditions
- Cryo conditions
- Rockimager data
- Soak info
- Comments/diffraction plan

Data import and processing



- Download preprocessed files & images using DART in-house tool
- Extract Resolve metadata, triggering synchrotron specific instructions
- Look for best MaxIV / ESRF / Diamond / .. preprocessed mtz(s)
- Run Pointless vs all known reference mtz / pdb pairs to find starting model
- Create restraints and do initial Buster run

Finding the files

LUND
$data_dir/$tripid*/process*/$alias/$alias-$crysisid/xds_*/MAXIV*/Edna*/results/*final.mtz
$data_dir/$tripid*/process*/$alias/$alias-$crysisid/xds_*/MAXIV*/AutoPROCTask_*/results/*staraniso_alldata-unique.mtz
..

ESRF
$data_dir/$tripid*/PROCESSED_DATA*/$alias/$alias*-$crysisid/run*/autoprocessing*/EDNA*/results/*_anom_truncate.mtz
$data_dir/$tripid*/PROCESSED_DATA*/$alias/$alias*$crysisid/run*/*/*/autoPROC/autoPROC_staraniso/*staraniso_alldata-unique.mtz
..

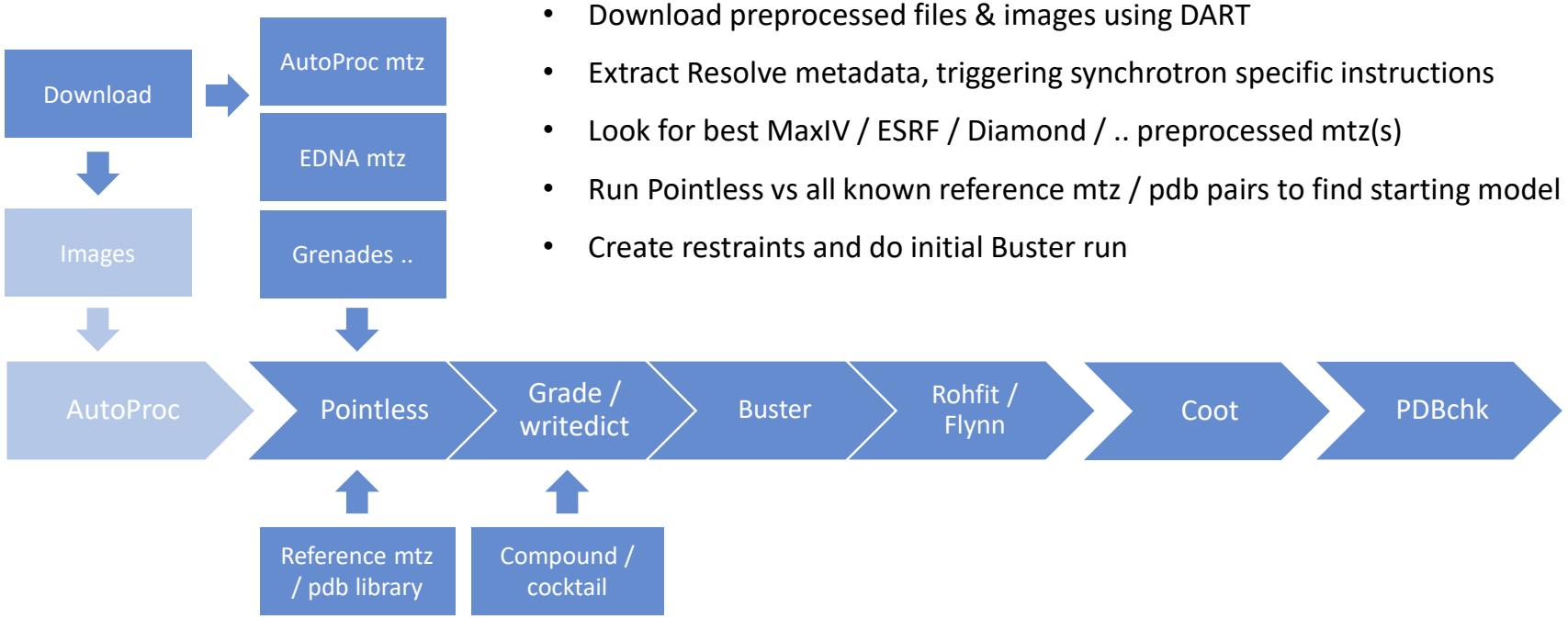
Diamond
$data_dir/$tripid*/processed/$alias/*$crysisid/*$crysisid*/*/autoPROC/staraniso_alldata-unique.mtz
$data_dir/$tripid*/processed/$alias/*$crysisid/*$crysisid*/*/xia2*/DataFiles/*free.mtz"
..

SLS
$data_dir/$tripid*/$alias/$alias-$crysisid*/adp_*/autoproc/*/*staraniso_alldata-unique.mtz
..

Soleil
$data_dir/$tripid*/PROCESSED_DATA/$alias/$alias-$crysisid/xdsme_*/ccp4if/$alias-$crysisid*.mtz
..

+ associated logfiles

Data import and processing



- Download preprocessed files & images using DART
- Extract Resolve metadata, triggering synchrotron specific instructions
- Look for best MaxIV / ESRF / Diamond / .. preprocessed mtz(s)
- Run Pointless vs all known reference mtz / pdb pairs to find starting model
- Create restraints and do initial Buster run

Summary, synchrotron data/logistics

- Common ISPYB / EXI architecture very helpful for end users
  - Variations in the details of input / output a complicating factor
- Data retrieval different for every synchrotron
  - Requirements from corporate IT sometimes not compatible with available options
  - DART download, managed by in-house IT
- Remote access
  - Direct access not compatible with IT safety rules
  - MXCUBE3
- Changes of location and names of prepocessed files / logs
  - Requires adaptation of in-house pipeline