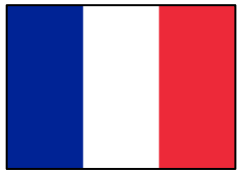


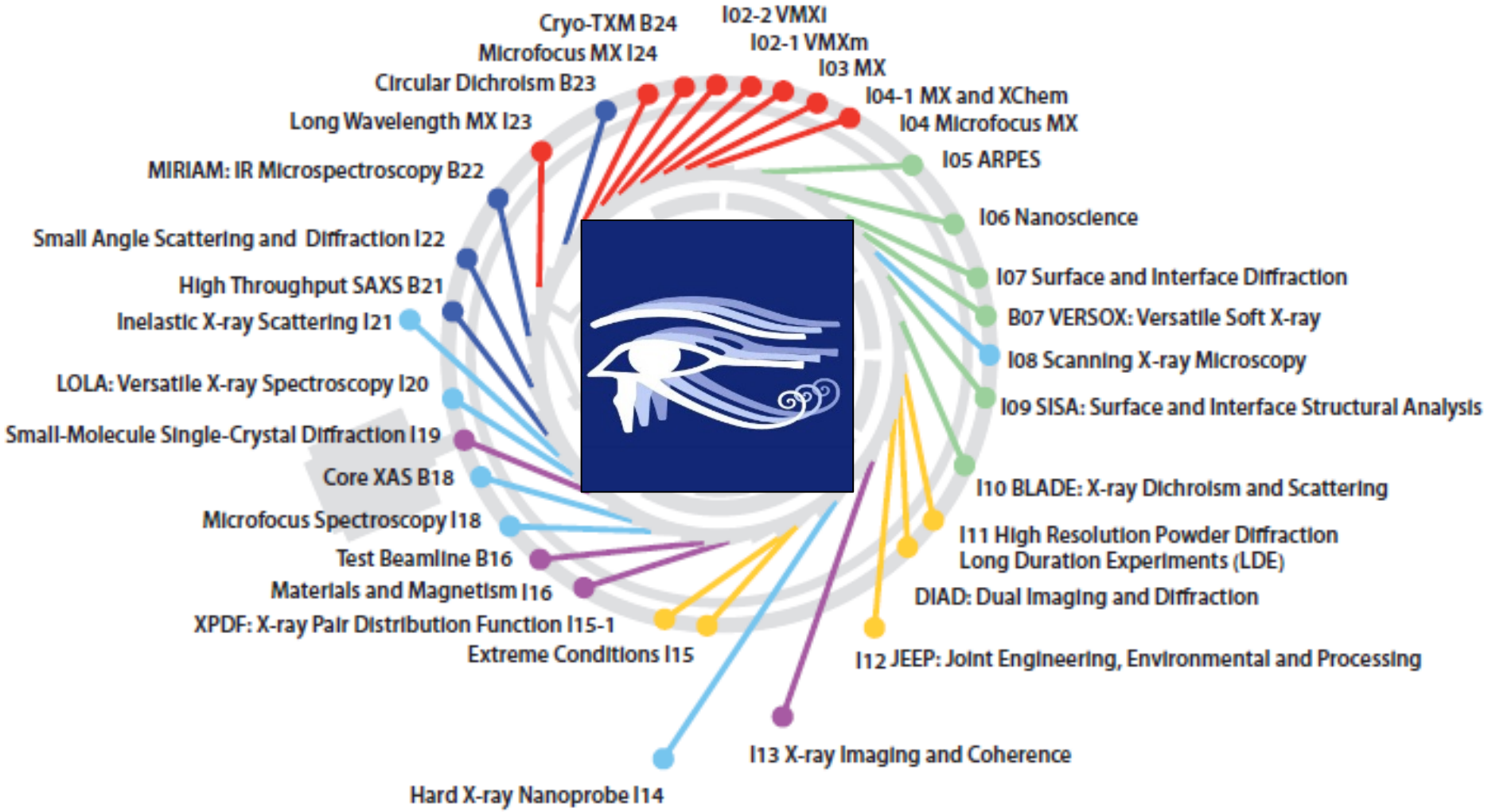


Plans following September Workshop

Neil A Smith

31st October 2019

diamond


I02-2 VMXI
I02-1 VMXm
I03 MX
I04-1 MX and XChem
I04 Microfocus MX
I05 ARPES
I06 Nanoscience
I07 Surface and Interface Diffraction
B07 VERSOX: Versatile Soft X-ray
I08 Scanning X-ray Microscopy
I09 SISA: Surface and Interface Structural Analysis
I10 BLADE: X-ray Dichroism and Scattering
I11 High Resolution Powder Diffraction
Long Duration Experiments (LDE)
DIAD: Dual Imaging and Diffraction
I12 JEEP: Joint Engineering, Environmental and Processing
I13 X-ray Imaging and Coherence
Hard X-ray Nanoprobe I14
Extreme Conditions I15
XPDF: X-ray Pair Distribution Function I15-1
Materials and Magnetism I16
Test Beamline B16
Microfocus Spectroscopy I18
Core XAS B18
Small-Molecule Single-Crystal Diffraction I19
LOLA: Versatile X-ray Spectroscopy I20
Inelastic X-ray Scattering I21
High Throughput SAXS B21
Small Angle Scattering and Diffraction I22
MIRIAM: IR Microspectroscopy B22
Long Wavelength MX I23
Circular Dichroism B23
Microfocus MX I24
Cryo-TXM B24

diamond

Background

- ISPyB developer group holds (mostly) monthly Video conferences
- Changes to the database schema are initially proposed and captured on github
  - https://github.com/ispyb/ispyb-database-modeling
  - 31 Open, 16 Closed, housekeeping required
- Small changes to database schema are discussed and sometimes agreed
- Large changes (refactoring, quality improvements etc.) often deferred
- Agreed to hold technical workshop in September hosted by SOLEIL
  - Thank you to Tatiana Isabet and Idrissou Chado

Aim

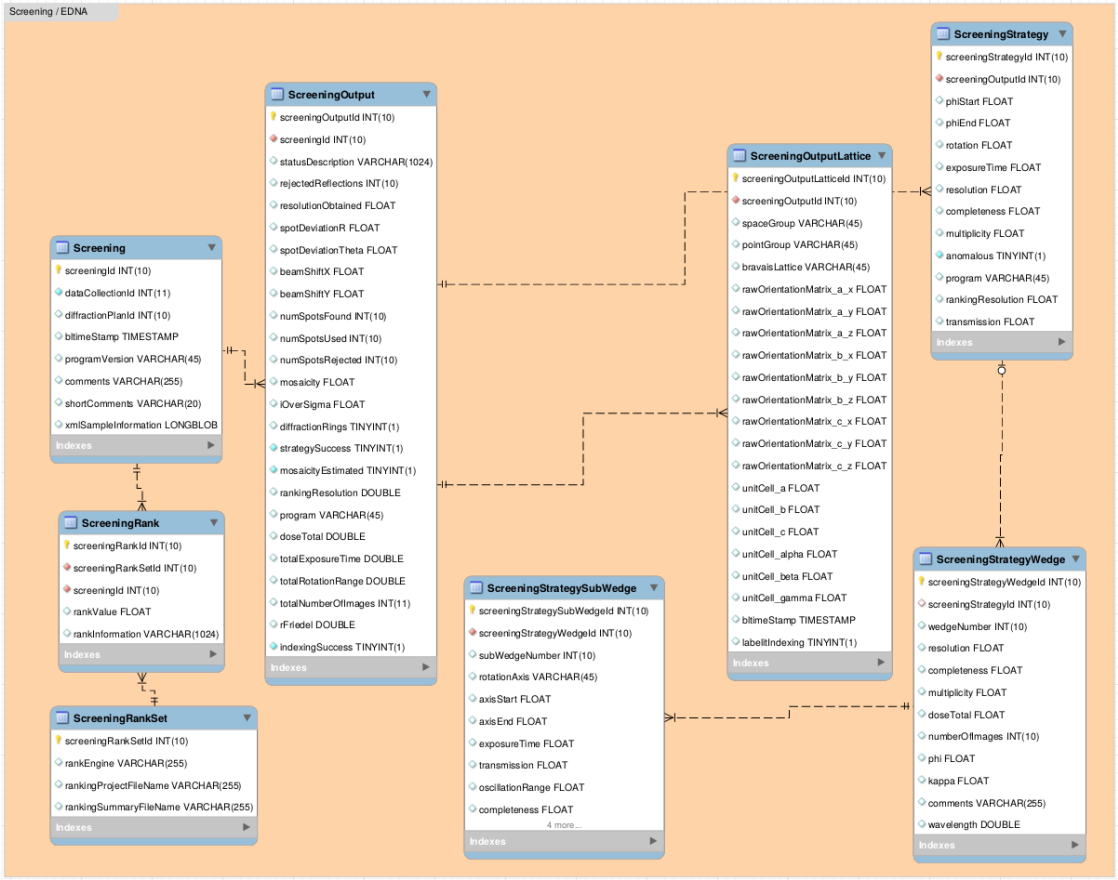
- Three database topics covered
  - Screening Tables
  - Data Collection
  - Sample tables
- Cover other items as time/energy allow!
- Details reported in meeting minutes on collaboration pages

https://ispyb.github.io/ISPyB/webpages/Other_meetings/TechMeeting_20190912_final.pdf

Thank you to Rasmus Fogh for compiling the minutes!

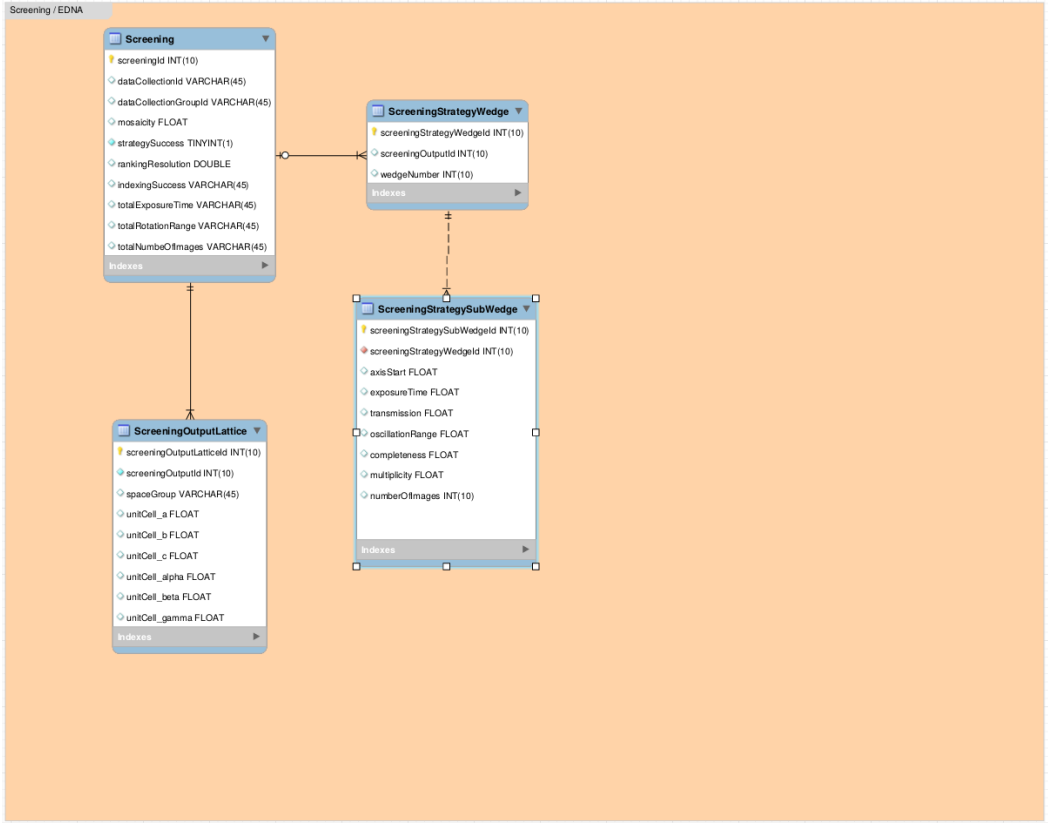
Screening Tables

- Proposal from @antonlinos
- Tables cover characterisation and strategy determination
- Intention to simplify and remove redundant columns
- Analysis of usage at ESRF/DLS...



Screening Tables

- ...Leads to vastly simpler layout
- Merge Screening, ScreeningOutput and Strategy tables, removing some fields
- Add an optional autoprocProgramID column
- Close to agreement on final version
- *Can we add twoTheta as variable to ScreeningStrategyWedge?*



Data Collection Tables

- Proposed by @karllevik
- Data collection, EnergyScan and XFEFlourescenceSpectrum share many columns
  - They are all in effect data collections so therefore should be represented as such in ISPyB
- A few specific columns from EnergyScan would be better stored as processing results
  - For example scanFileFullPath, jpegChoochFileFullPath etc. Can use more generic DataCollectionFileAttachments table

DC	EnergyScan	Fluorescence
dataCollectionId (PK)	energyScanId (PK)	xfeFluorescenceSpectrumId (PK)
DCG.sessionId	sessionId	sessionId
blSampleId	blSampleId	blSampleId
blSubSampleId	blSubSampleId	blSubSampleId
detectorId	fluorescenceDetector	
beamSizeAtSampleX	beamSizeHorizontal	beamSizeHorizontal
beamSizeAtSampleY	beamSizeVertical	beamSizeVertical
transmission	transmissionFactor	beamTransmission
comments	comments	comments
crystalClass	crystalClass	crystalClass
startTime	startTime	startTime
endTime	endTime	endTime
exposureTime	exposureTime	exposureTime
fileTemplate	filename	filename
averageTemperature	temperature	
wavelength		wavelength
totalAbsorbedDose or totalExposedDose?	xrayDose	
flux	flux	flux
flux_end	flux_end	flux_end
imageDirectory	workingDirectory	workingDirectory
axisStart, axisEnd		axisPosition

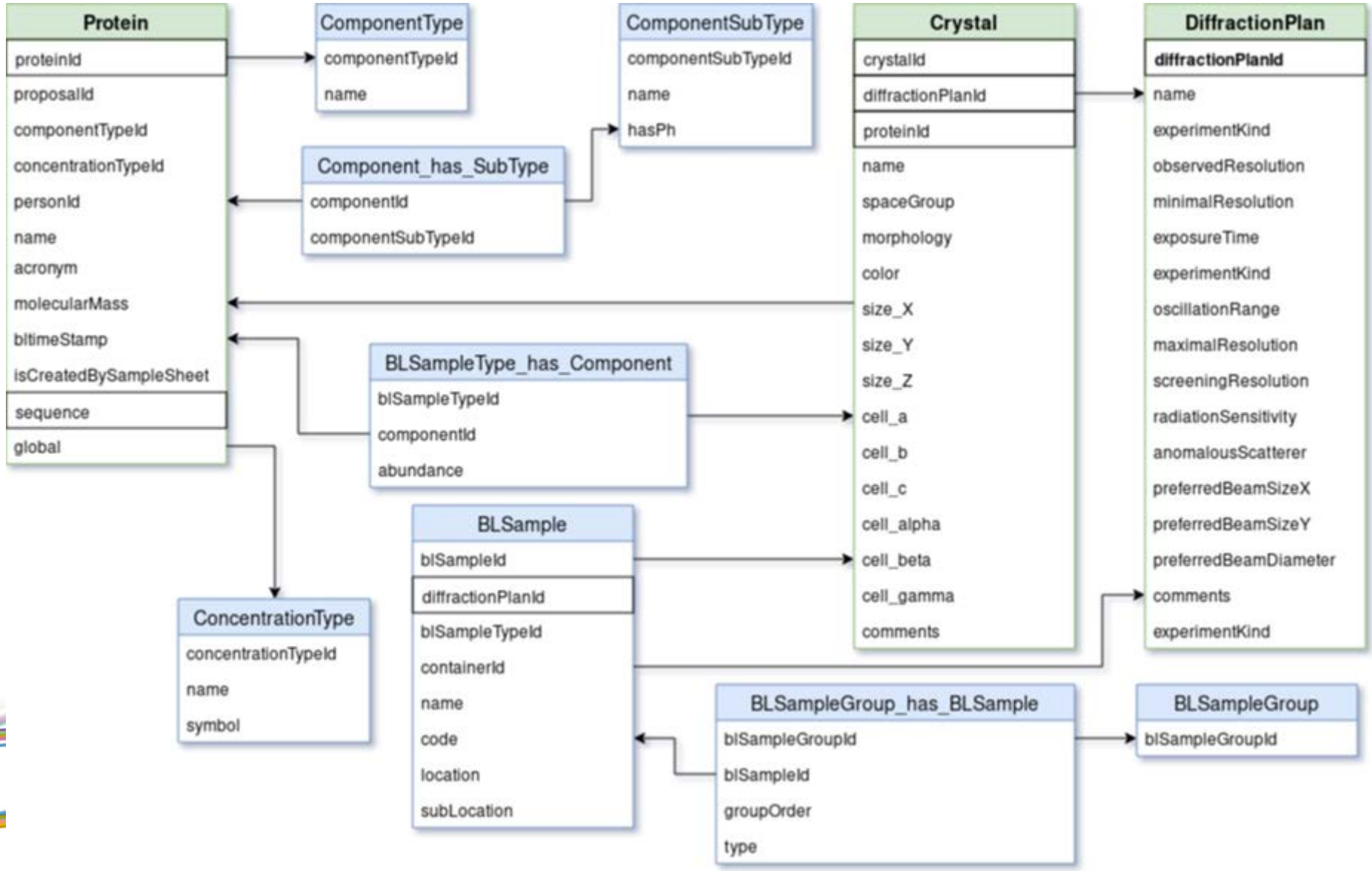
Data Collection Tables

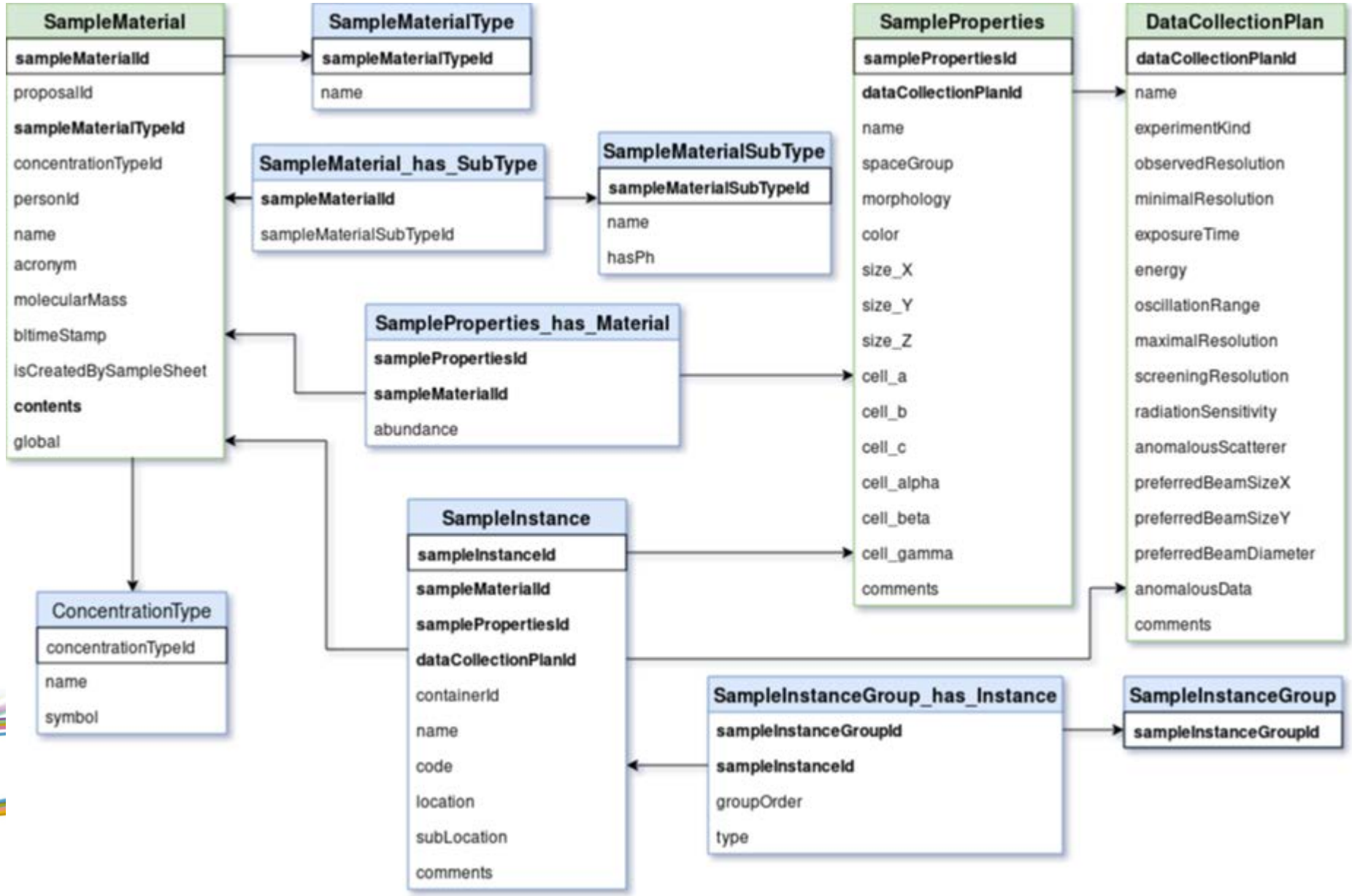
- Agreement on initial proposal to merge tables into the Data Collection Table
- Wider discussion on whether DC table should be reduced to core set of values
- Currently DC is a "wide" table with many nulls
  - For example includes both MX and EM specific values
- Aesthetically an improved approach would see DC being a basic table with domain specific tables refering back to common Table
  - MXDataCollection, EMDataCollection
- From a performance perspective however, a wide table with NULLs is not an issue

diamond

Sample Tables

- Proposed by @karllevik
- Terminology from MX & life science
- Even with the realm of MX/EM/BioSAXS sample types vary
  - RNA, DNA, Protein, Virus
- Sample tables assume Protein/Crystal:
  - MX bias
  - BioSAXS tables are not well integrated with the rest of the schema (although scientifically relevant to their domain)


Protein
proteinId
proposalId
componentTypeId
concentrationTypeId
personId
name
acronym
moleculatMass
bltimeStamp
isCreatedBySampleSheet
sequence
global
ComponentType
componentTypeId
name
ComponentSubType
componentSubTypeId
name
hasPh
Component_has_SubType
componentId
componentSubTypeId
Crystal
crystalId
diffractionPlanId
proteinId
name
spaceGroup
morphology
color
size_X
size_Y
size_Z
cell_a
cell_b
cell_c
cell_alpha
cell_beta
cell_gamma
comments
DiffractionPlan
diffractionPlanId
name
experimentKind
observedResolution
minimalResolution
exposureTime
experimentKind
oscillationRange
maximalResolution
screeningResolution
radiationSensitivity
anomalousScatterer
preferredBeamSizeX
preferredBeamSizeY
preferredBeamDiameter
comments
experimentKind
BLSampleType_has_Component
blSampleTypeId
componentId
abundance
BLSample
blSampleId
diffractionPlanId
blSampleTypeId
containerId
name
code
location
subLocation
ConcentrationType
concentrationTypeId
name
symbol
BLSampleGroup_has_BLSample
blSampleGroupId
blSampleId
groupOrder
type
BLSampleGroup
blSampleGroupId


SampleMaterial
sampleMaterialId
proposalId
sampleMaterialTypeId
concentrationTypeId
personId
name
acronym
molecularMass
bltimeStamp
isCreatedBySampleSheet
contents
global
SampleMaterialType
sampleMaterialTypeId
name
SampleMaterial_has_SubType
sampleMaterialId
sampleMaterialSubTypeId
SampleMaterialSubType
sampleMaterialSubTypeId
name
hasPh
SampleProperties_has_Material
samplePropertiesId
sampleMaterialId
abundance
SampleProperties
samplePropertiesId
dataCollectionPlanId
name
spaceGroup
morphology
color
size_X
size_Y
size_Z
cell_a
cell_b
cell_c
cell_alpha
cell_beta
cell_gamma
comments
DataCollectionPlan
dataCollectionPlanId
name
experimentKind
observedResolution
minimalResolution
exposureTime
energy
oscillationRange
maximalResolution
screeningResolution
radiationSensitivity
anomalousScatterer
preferredBeamSizeX
preferredBeamSizeY
preferredBeamDiameter
anomalousData
comments
ConcentrationType
concentrationTypeId
name
symbol
SampleInstance
sampleInstanceId
sampleMaterialId
samplePropertiesId
dataCollectionPlanId
containerId
name
code
location
subLocation
comments
SampleInstanceGroup_has_Instance
sampleInstanceGroupId
sampleInstanceId
groupOrder
type
SampleInstanceGroup
sampleInstanceGroupId

In Summary

- Renaming
  - Protein → SampleMaterial
  - Crystal → SampleProperties
  - BLSample → SampleInstance
  - BLSampleGroup → SampleInstanceGroup
  - ComponentType → SampleMaterialType
  - ComponentSubType → SampleMaterialSubType
  - DiffractionPlan → DataCollectionPlan
- ...and all the associated keys (proteinId => sampleMaterialId etc)
- However, without updating column names this leads to work without the payoff
- SampleProperties would still be a Crystal unless we add many fields
  - wide table problem again...
- Also does not accommodate BioSAXS use case

diamond

Sample Table Results

- More work required to develop a change that adds value and at least supports MX/EM and BioSAXS consistently
  - That should then provide better foundation to support other techniques
- Solution needs to address the distinction between sample components that are
  - the main component (e.g. protein) under investigation
  - Substances used as additives etc.
- Some consideration of reusing NXSample but initial thoughts are it does not fit well

Way forward

- Successful workshop
  - Some key agreements on refactoring Screening and Data Collection
- Appreciate resources limited
  - Refactoring to support future ISPyB requirements needs to be considerate of the impact
- Short term (12 months):
  - For samples - propose new tables that link to existing tables rather than other direction (non-invasive)
- Medium Term (12-24 months)
  - Seek closer alignment and consistency between MX/EM/BioSAXS
- Longer term (24 month+)
  - Should we have a roadmap/vision for what ISPyB is 2022?
  - Move towards micro services to aid reuse between sites?

Acknowledgements

Daniel Sanchez, ALBA
James Hall, Karl Levik, Diamond
Maxime Chaillet, Solange Delageniere, ESRF
Alejandro de Maria, Olof Svensson, ESRF
Stu Fisher, DLS/ESRF
Rasmus Fogh, Global Phasing
Idrissou Chado, Tatiana Isabet, SOLEIL

Thank you!