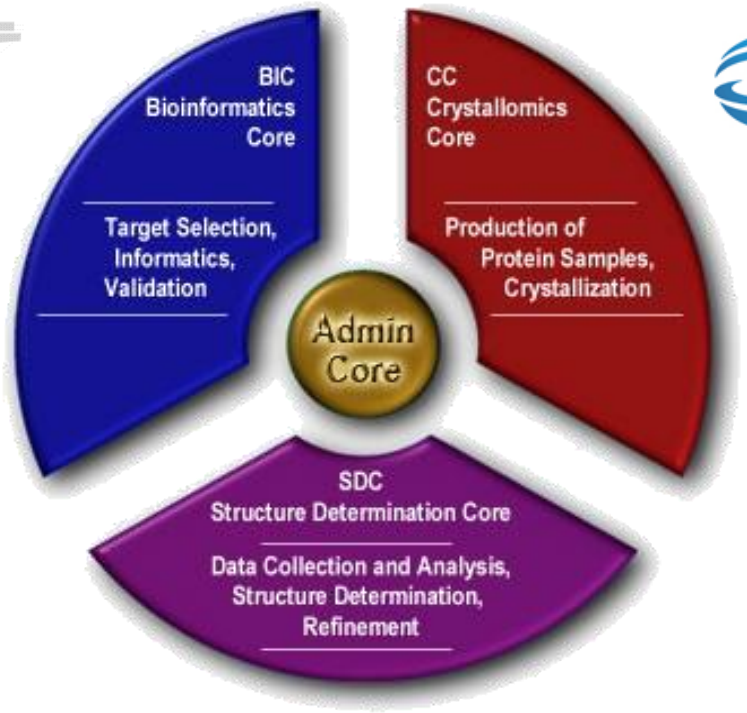
Joint Center for Structural Genomics

University of California San Diego



Genomics Institute of the Novartis Research Foundation

BURNHAM INSTITUTE for MEDICAL RESEARCH
From Research, the Power to Cure

The Scripps Research Institute

SSRL

High-throughput determination of novel protein structures using X-ray crystallography

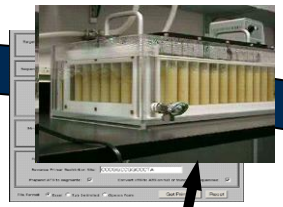
JCSG
Joint Center for Structural Genomics
Developing HT methods for Gene to Structure and Function

PSI
PSI:BIOLOGY

JCSG Pipeline

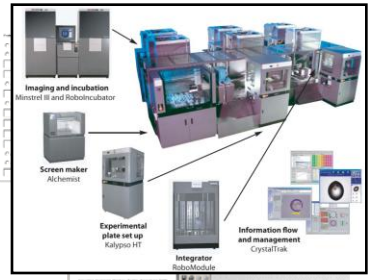
Target Selection

Parallel HT Expression
PIPE cloning



Automated HT
Purification

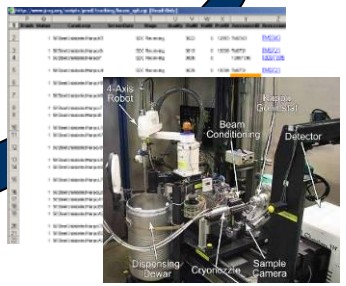
Automated HT
Crystallization



Automated HT
Imaging

- JCSG Pipeline: Integration of custom and commercial instrumentation into a HT structural genomics pipeline.
- Data capture to central DB parallels the experimental pipeline. Currently collecting 520 parameters from 32 stages.
- Feedback to the experimental pipeline: Target selection & management.
- Expanding interaction with community: TOPSAN (wiki)

JCSG
DB

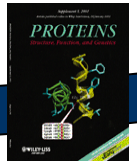


Automated HT
Crystal Screening
+ Data Collection

PSI-KB

www.jcsg.org

TOPSAN
WIKI



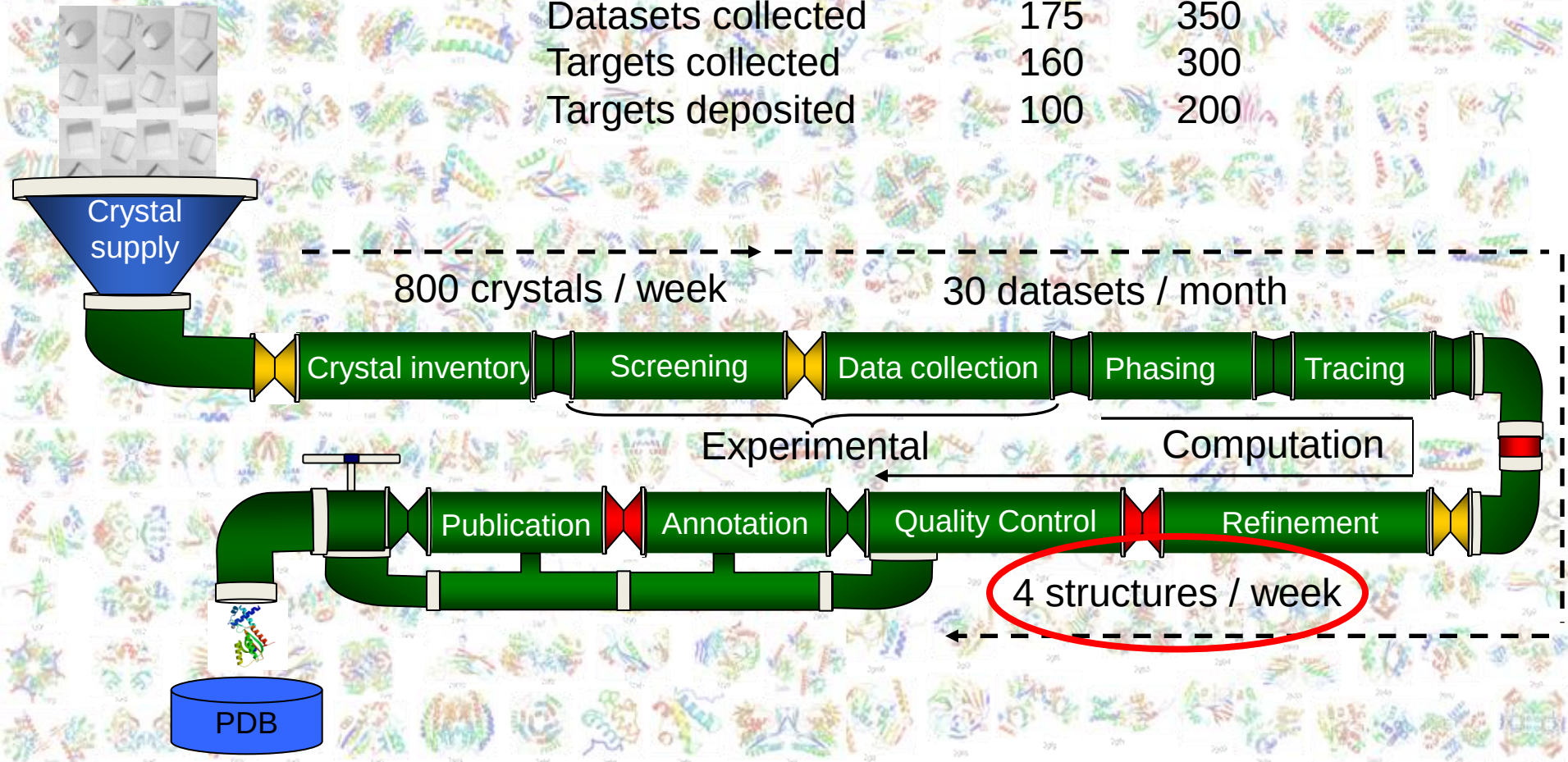
Semi-Auto
Publication

Structure Validation
& Deposition

Semi-automated
Structure determination

SDC activity

	2005/6	2009/10
Screened crystals	9000	30000
Screened proteins	250	900
Datasets collected	175	350
Targets collected	160	300
Targets deposited	100	200



SDC activity

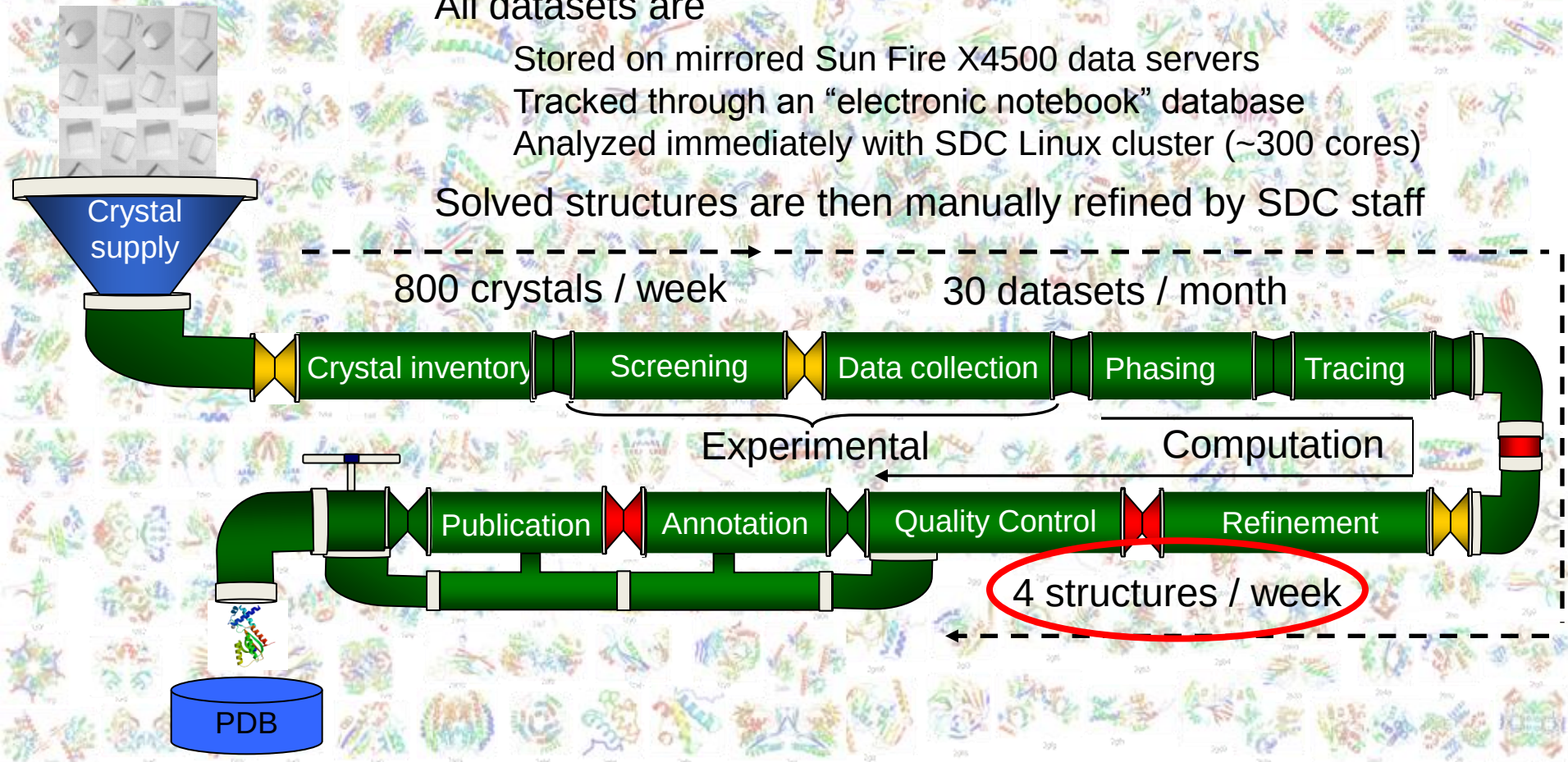
Collect 350 datasets per year at SSRL PX beamlines

- Typical dataset takes ~1 hour to collect
- Datasets consist of 100-1000 diffraction images (~20Mb / image)
- 2-20Gb per dataset, ~3Tb experimental data per year

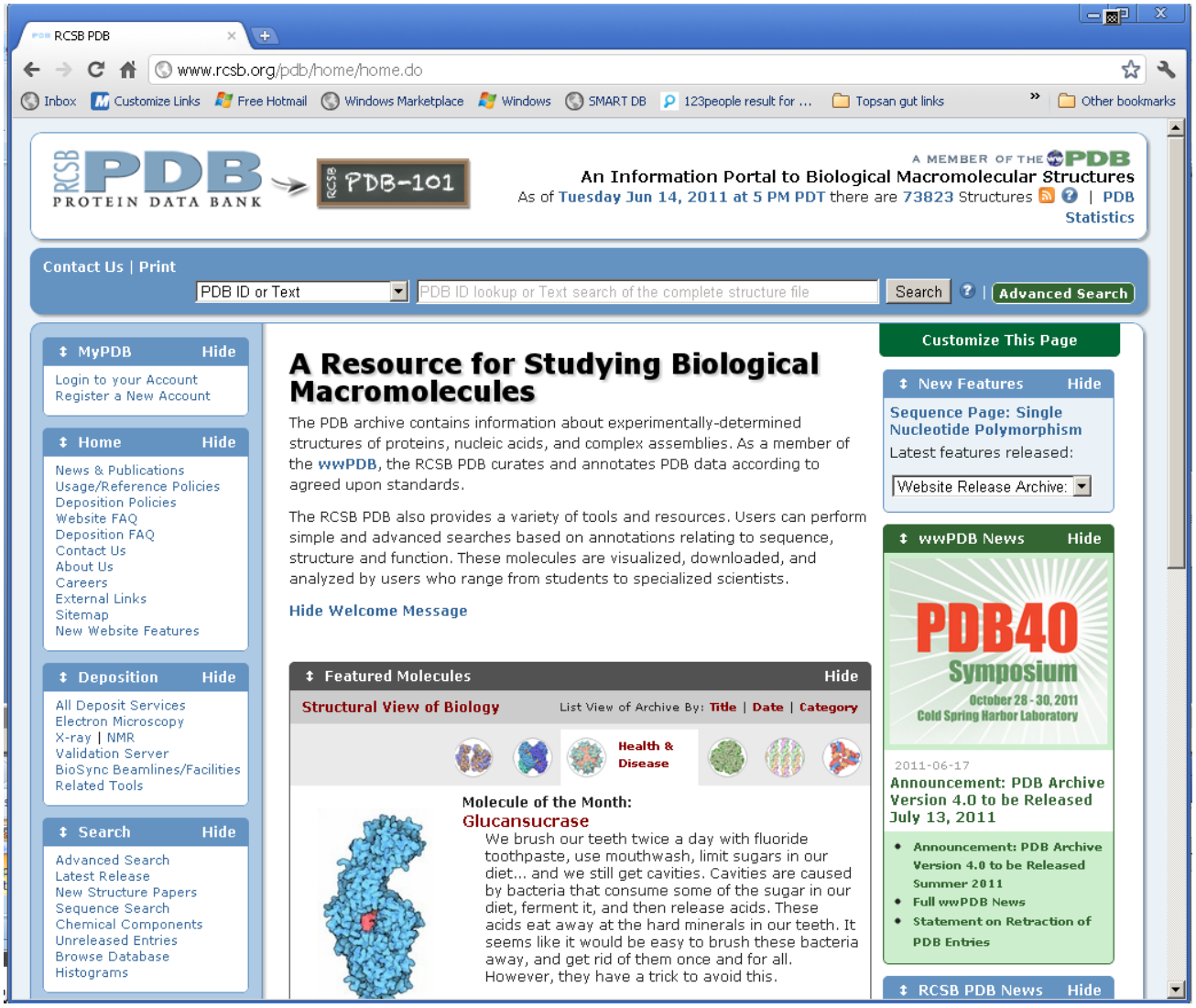
All datasets are

- Stored on mirrored Sun Fire X4500 data servers
- Tracked through an “electronic notebook” database
- Analyzed immediately with SDC Linux cluster (~300 cores)

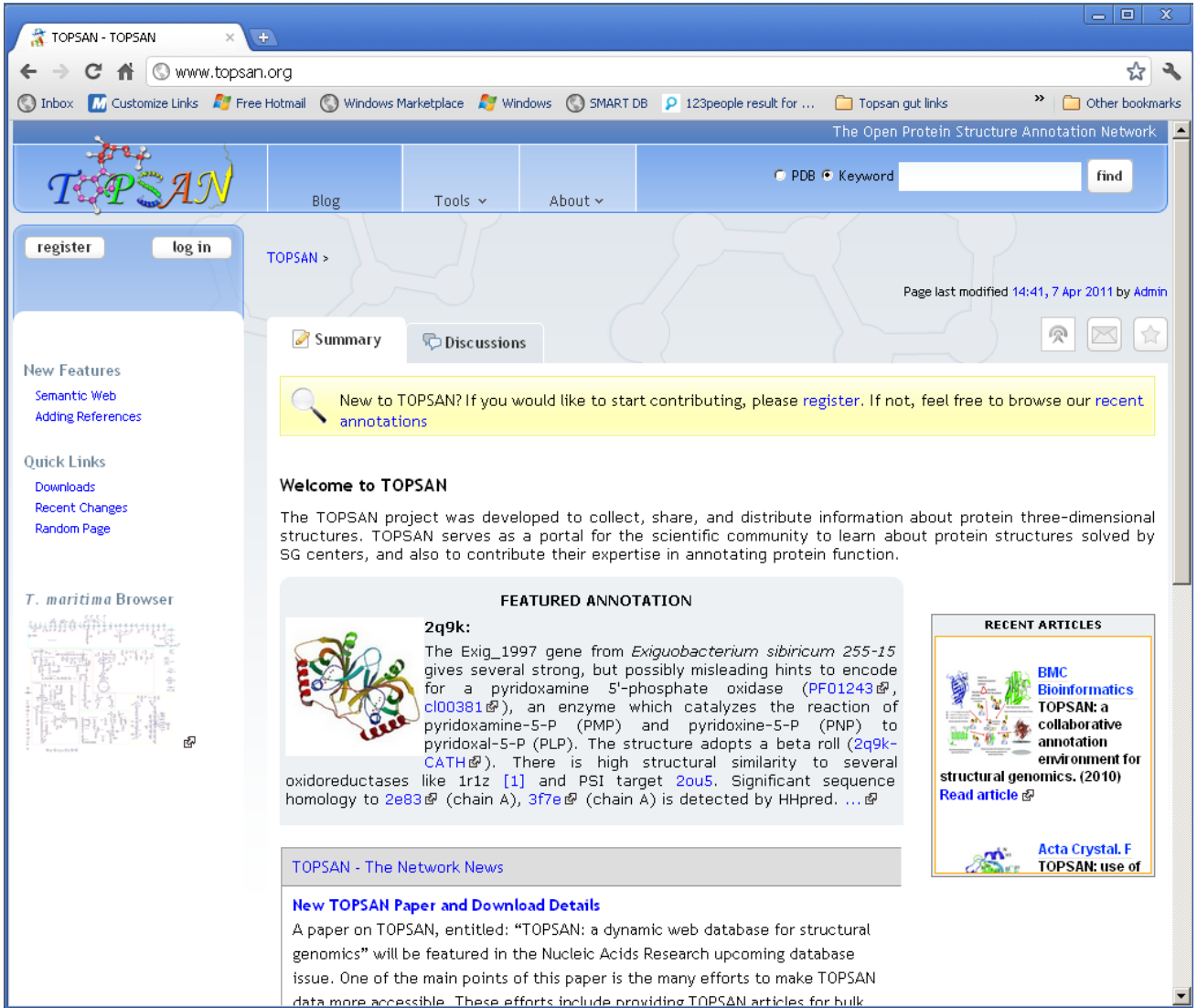
Solved structures are then manually refined by SDC staff



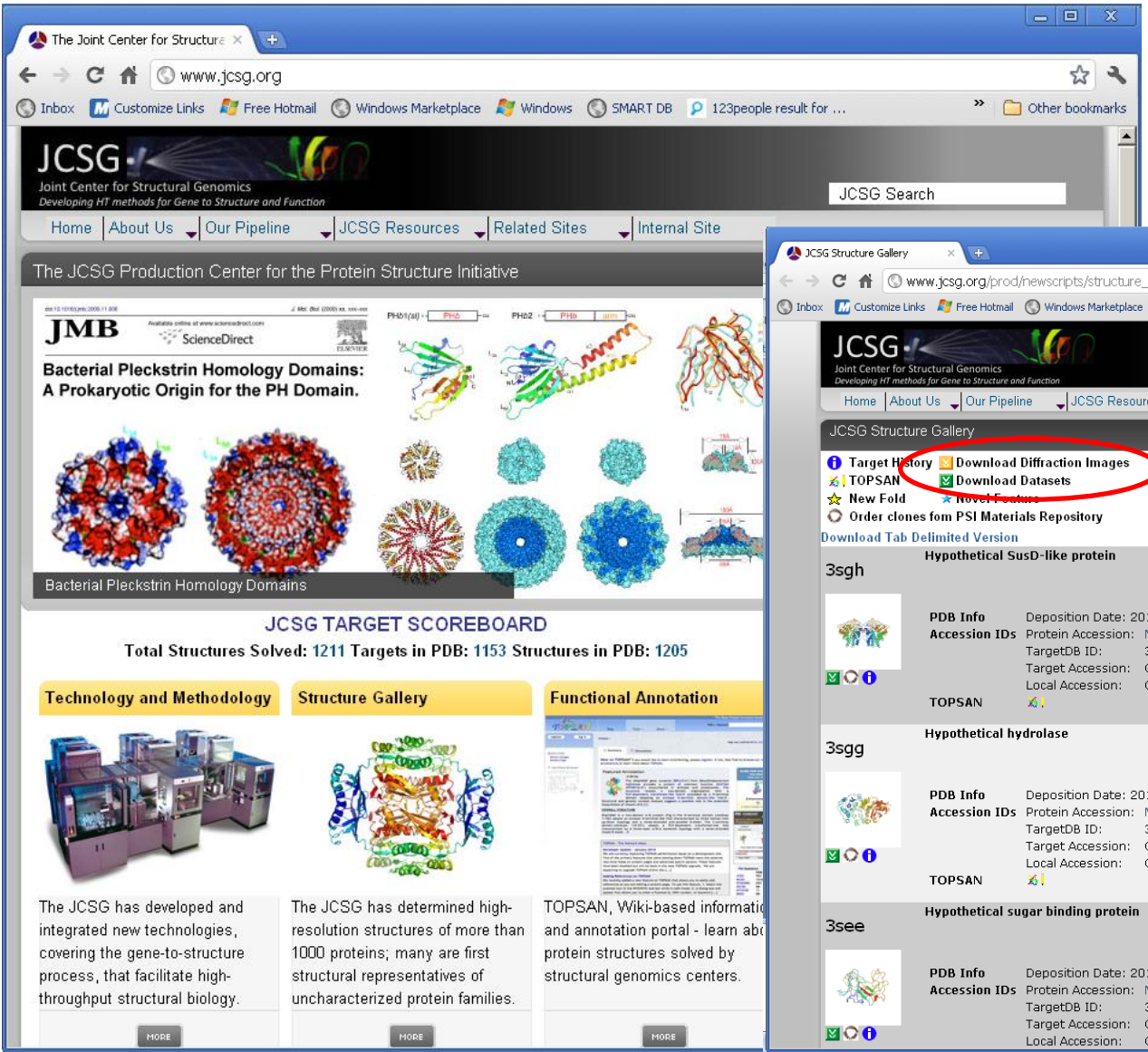
Derived data (3D coordinates+structure factors) are deposited in the Protein Data Bank @ www.rcsb.org

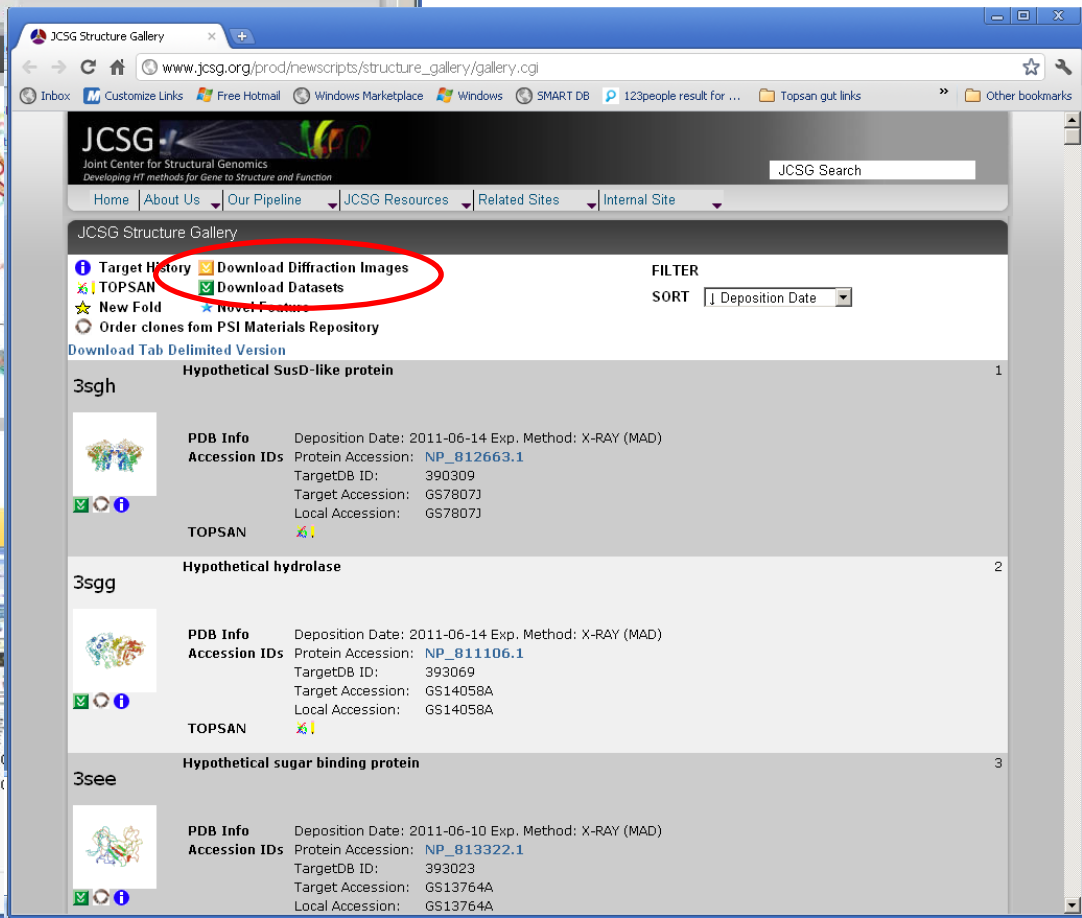


Scientific highlights in journals or "published" on wiki-based platform @ www.topsan.org



Subset of processed data uploaded to JCSG Oracle database @ www.jcsg.org





What's needed?

Develop a more hollistic approach to managing raw experimental data

Provide immediate and long-term archival of datasets

Disseminate unusable datasets to methods developers

Disseminate raw data sooner to collaborators

Prepare for new experimental capabilities (faster detectors)

Develop a model that could be adopted by the broader SSRL user community