“Science at ESO: Extragalactic surveys”

Boris Häußler

ESO/Chile

LaSilla Observing School, February 15th 2024

“Surveys (some extragalactic, mostly imaging, some of which happen to be at ESO)” - What are they? Why are they special? How do you work with them?

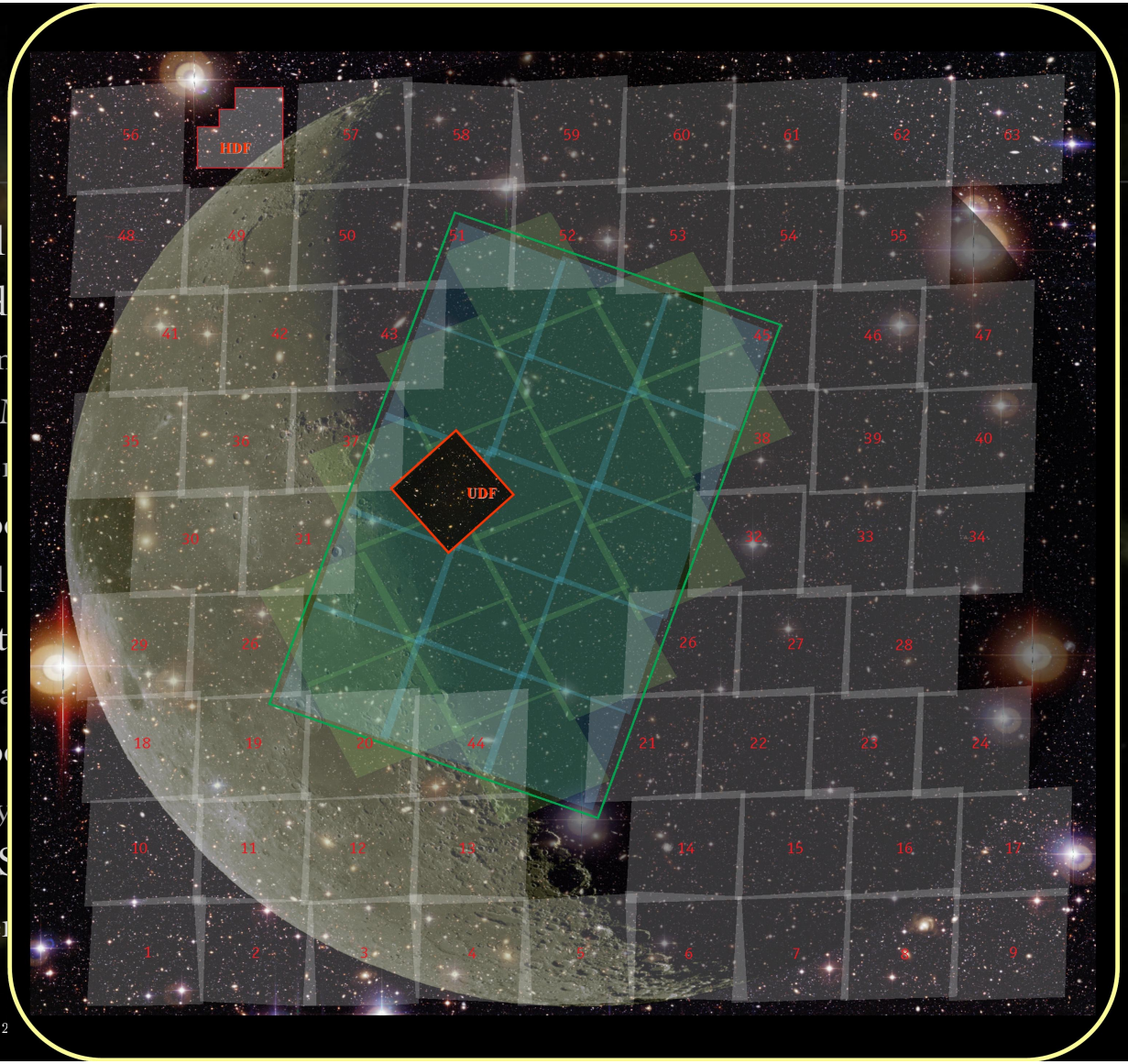
Boris Häußler

ESO/Chile

LaSilla Observing School, February 15th 2024

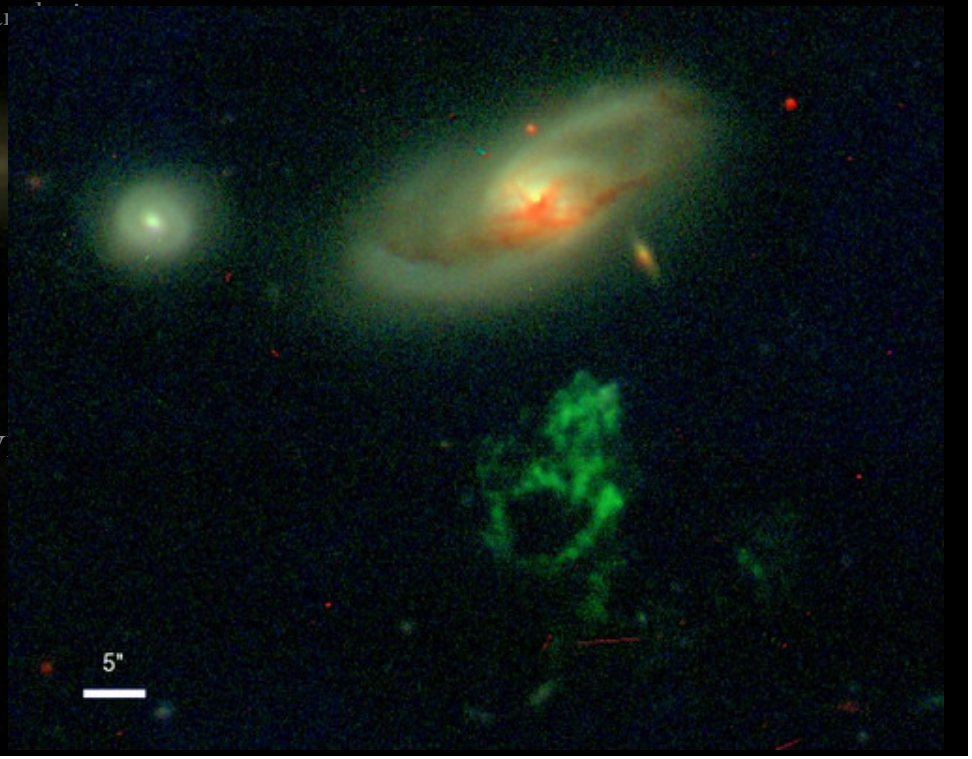
Why

- Studied
- Applied
  - got car
  - -> GE
  - light p
- PostDo
  - helped
- 2nd Post
  - automa
- PostDo
  - Survey
- ESO, IS
- Involve



What defines a 'large survey'?

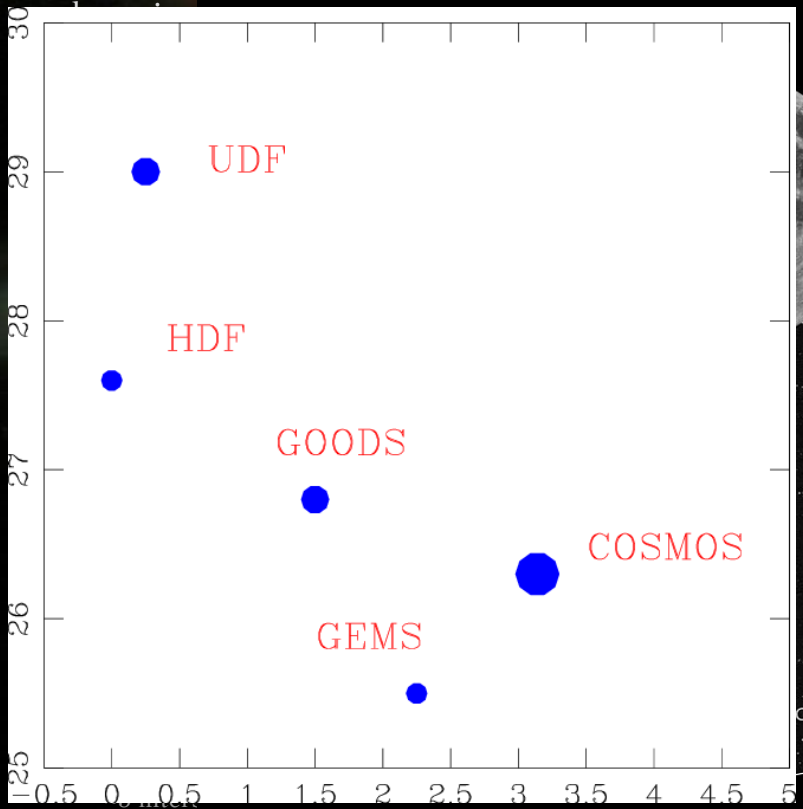
- AT ESO:
  - different types of scientific programs
    - **"normal" projects**:
      - Dedicated (often single) question -> observations -> a
    - **"large" programs:**
      - intermediate
      - Can be surveys, but not always are
    - **surveys** on 'normal' instruments:
      - Multitude of questions
      - Follow-up Science
      - Legacy, e.g. catalogues, data for years to use
  - Outside ESO also:
    - **mission**, dedicated survey instruments (SDSS, LSST, V
      - Dedicated and funded for a large range of questions
      - Many serendipitous finding
      - Huge teams
      - Follow-up Science
      - Legacy data and defining standards

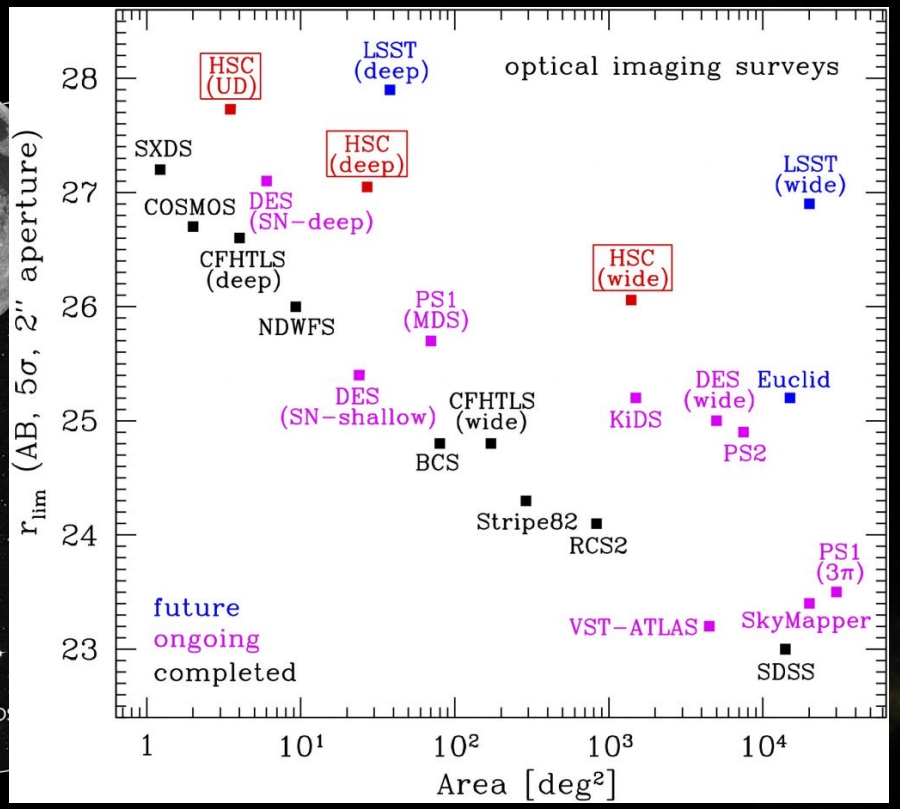


What defines a 'large survey'?

- sky coverage
  - GEMS/STAGES: 800 arcmin^2
  - Largest HST survey: 2 deg^2: COSMOS
  - VIDEO: 12 deg^2
  - GAMA: 280 deg^2, mostly spectral survey, imaging data from other sources (VST/VISTA)
  - J-PAS: 56 filters, 8000 deg^2
  - LSST:
    - 10 deg^2 FoV
    - the entire visible sky imaged every 3 nights
- number of pixels
  - COSMOS: 146 ACS footprints (4kx4k) = 2.3 Gigapixels
  - VIRCAM: 16x2k2k = 66 Megapixels per image -> VIDEO: ~2 Gigapixels
  - GAMA: 16 Gigapixel per field and filter
  - LSST:
    - 3.2Gigapixel, largest camera ever made (typically 15 second exposures), 5.5 million images
    - Comparison: MUSE 24x 4kx4k = 400Megapixel, typically 1h long

What defines a ‘large survey’?





- *7* trillion

What defines a 'large survey'?

- data volume?
  - MUSE (Chapman): ~3.2GB per data cube
  - CANDELS: 8 GB per fields and band -> ~800 GB including weight images
  - VIDEO: 4TB (download time several months) -> stacked to ~500GB.
  - GAMA: 65 GB per field and band (imaging) -> 6TB including weight images
  - LSST/Rubin: 60 PB catalogue raw data! 20 TB per night. 15 PB catalogue!
  - SKA: 68 TB/SECOND! 2^{15} pixels on a side and up to 2^{16} frequency channels, 500 PB per year! 1 PB of catalogued data. 3 Exabyte/year (EB) of fully processed data -> CPU power needed

- imaging vs spectroscopy vs IFU
- different pixel scales/spatial or spectral resolutions
- wavelength coverage (filters, Radio, Xray, optical, NIR, FIR, etc)
- time/cadence
  - LSST: 10 million alerts per night (SN, moving objects, transients, etc)

Large Surveys - Science topics

- Science topics are very diverse

Usually very diverse datasets

- COSMOS (32 wavelengths?), several generations and epochs over years.
  - optical: HST data (both imaging and spectra)
  - IR/MIR: Spitzer data, VISTA/VIRCAM
  - FIR: sub-mm: Herschel
  - UV: Galex
  - X-ray: XMM Newton
  - X-ray: Chandra
  - Radio: VLA
  - Optical: Subaru/HSC
  - Optical spectra: VIMOS (VUDS,zCOSMOS), FMOS, Keck/Deimos
  - mm/sub-mm: ALMA
- each wavelength will have an entire team of experts behind it, that prepare and analyse data, create object catalogues that are then shares with the entire team and releases data products to the public
- > 200 team members

Typical publications lists

- Cosmos: >400 publications
  - Precision photometric redshift calibration for galaxy–galaxy weak lensing
  - Relation between stellar mass and star-formation activity in galaxies
  - Active galactic nucleus host galaxy morphologies in COSMOS
  - The zCOSMOS redshift survey: the three-dimensional classification cube and bimodality in galaxy physical properties
  - The faint-end slopes of galaxy luminosity functions in the COSMOS field
  - A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images
  - Expanding the Search for Galaxies at z ~ 7-10 with New NICMOS Parallel Fields
  - The Dependence of Star Formation Activity on Stellar Mass Surface Density and Sersic Index in zCOSMOS Galaxies at $0.5 < z < 0.9$ Compared with SDSS Galaxies at $0.04 < z < 0.08$
  - The zCOSMOS survey: the role of the environment in the evolution of the luminosity function of different galaxy types
  - Massive Galaxies in COSMOS: Evolution of Black Hole Versus Bulge Mass but not Versus Total Stellar Mass Over the Last 9 Gyr?
  - The Stability of the Point-Spread Function of the Advanced Camera for Surveys on the Hubble Space Telescope and Implications for Weak Gravitational Lensing
  - Photometric Redshifts of Galaxies in COSMOS
  - The Effects of Environment on Morphological Evolution at $0 < z < 1.2$ in the COSMOS Survey

Typical publications lists

- Cosmos: >400 publications
  - Precision photometric redshift calibration for galaxy–galaxy weak lensing
  - Relation between stellar mass and star-formation activity in galaxies
  - Active galactic nucleus host galaxy morphologies in COSMOS
  - The zCOSMOS redshift survey: the three-dimensional classification cube and bimodality in galaxy physical properties
  - The faint-end slopes of galaxy luminosity functions in the COSMOS field
  - A robust morphological classification of high-redshift galaxies using support vector machines on seeing limited images
  - Expanding the Search for Galaxies at z ~ 7-10 with New NICMOS Parallel Fields
  - The Dependence of Star Formation Activity on Stellar Mass Surface Density and Sersic Index in zCOSMOS Galaxies at $0.5 < z < 0.9$ Compared with SDSS Galaxies at $0.04 < z < 0.08$
  - The zCOSMOS survey: the role of the environment in the evolution of the luminosity function of different galaxy types
  - Massive Galaxies in COSMOS: Evolution of Black Hole Versus Bulge Mass but not Versus Total Stellar Mass Over the Last 9 Gyr?
  - The Stability of the Point-Spread Function of the Advanced Camera for Surveys on the Hubble Space Telescope and Implications for Weak Gravitational Lensing
  - Photometric Redshifts of Galaxies in COSMOS
  - The Effects of Environment on Morphological Evolution at $0 < z < 1.2$ in the COSMOS Survey

Survey politics

- Often need to pay in financially to get (early) data access (Unless you're in Chile, see Evelyns talk).
  - Data usually becomes public in big data releases after a few years -> Legacy science (e.g. SDSS-I, <2008)
  - For large missions, access to data might not even be technically possible! Just object catalogues.
    - Maybe even those only through SQL etc, not catalogues themselves
- Sometimes restricted to US or Europe, depending on funding
- Sign-up either for people or departments
  - SDSS: departments sign up for certain number of staff and their students/PostDocs
    - They are pretty strict when you leave. Finish your project, then you're out
  - Usually: Once a member, always a member
  - EUCLID: similar, but individual membership is possible
  - Licenses might not even be transferable if you move department (e.g. SDSS)
  - all chilean university departments now part of SDSS V
- Often you pay in by doing certain work for the surveys.
  - help write the proposal, provide ancillary data (e.g. redshifts), observe, manage, have a good idea and collaborate -> You're a member
- MOST surveys are very open and happy to share access to data

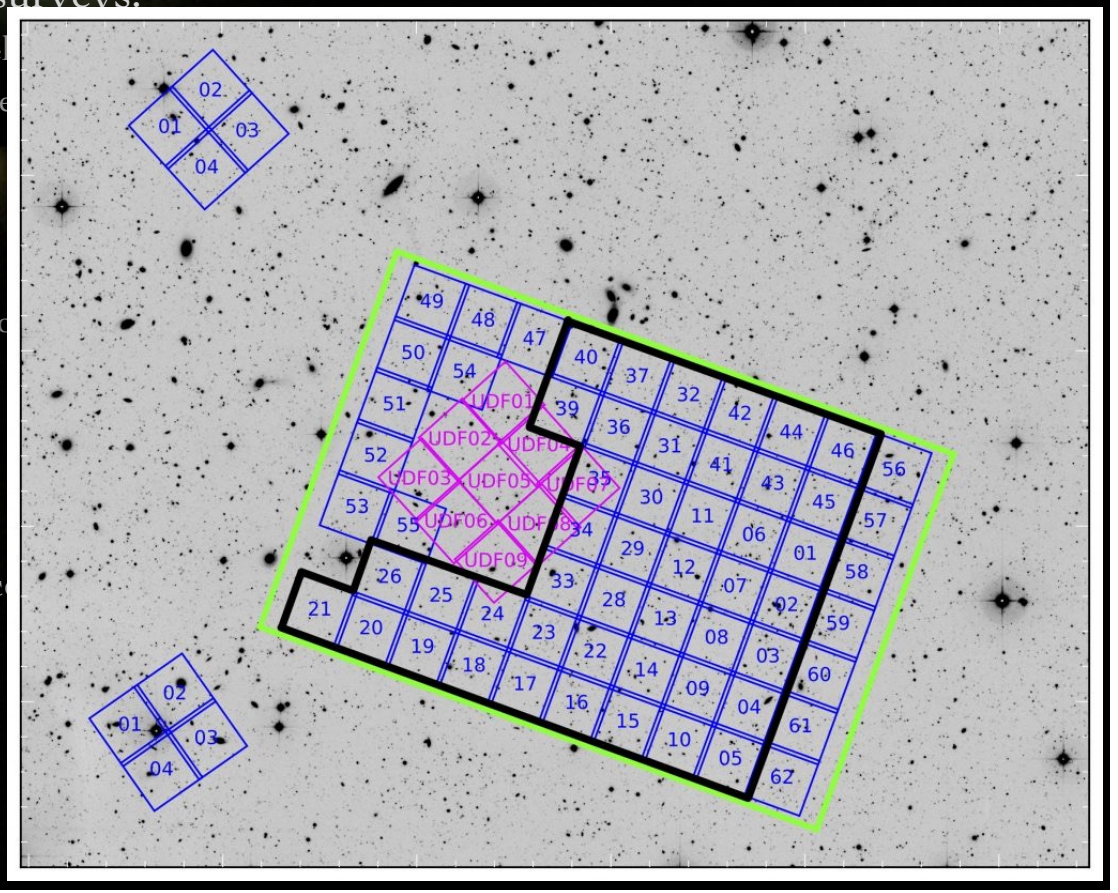
Bottom line

It's impossible to give one talk about "large surveys", covering everything!

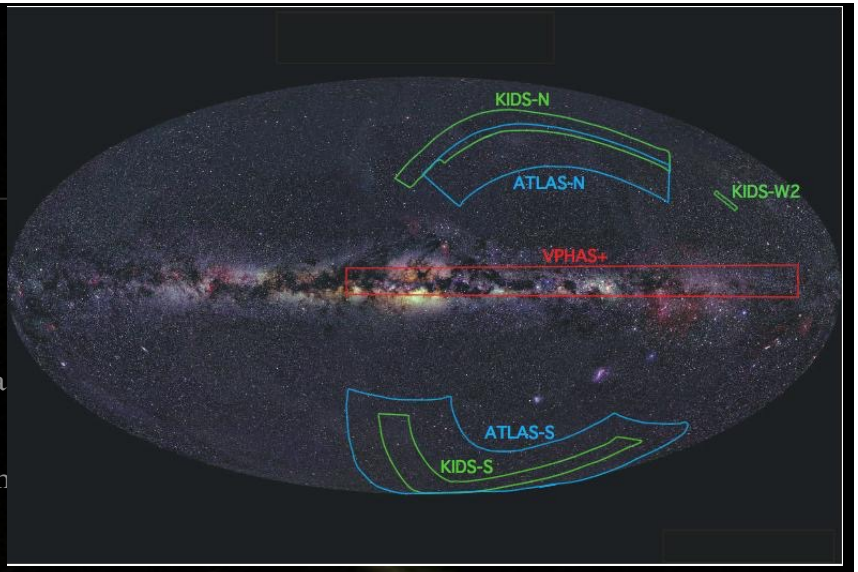
(or even a small fraction of it)

Surveys at ESO

- often large programs and GTO are surveys:
  - MUSE-wide survey (~9000 CANDELS sel
  - Gaia-ESO, ~100,000 stars (FLAMES, fibre
- ESO surveys :
  - special calls for surveys:
    - VANDELS (2087 VIMOS spectra)
      - redshifts, rotation curves, spectral compo
    - LEGA-C (3528 VIMOS spectra)
    - VISTA/Vircam (via CASU)
    - VST/OmegaCam
  - raw data public immediately
  - required to publicly release the (reduc
  - Future: 4MOST

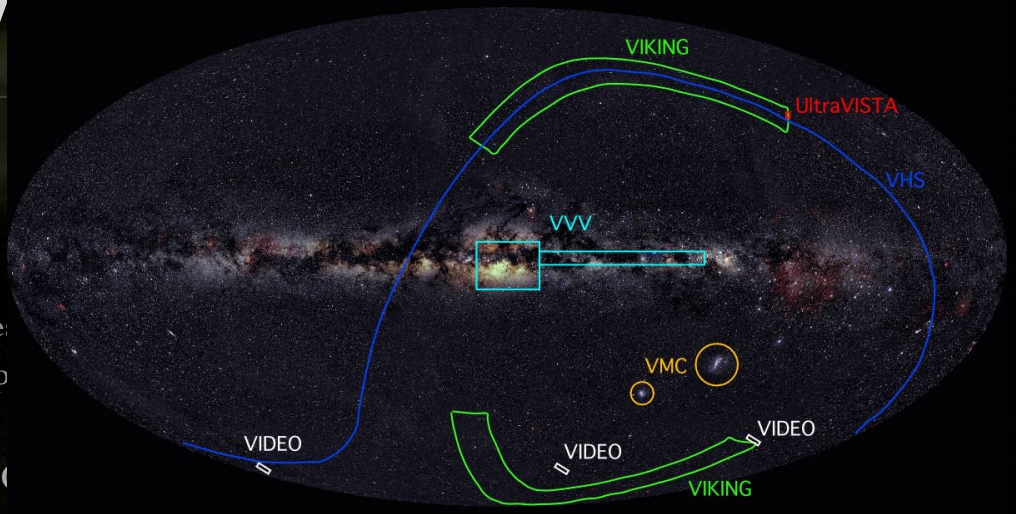


Surveys at ESO - VST



- VST/OmegaCam
  - KIDS — The Kilo-Degree Survey:
    - 1500 square degrees in four bands (U, V, R and I). same area
    - 2.5 magnitudes deeper than SDSS
    - study dark matter halos and dark energy with weak lensing, h and studying galactic evolution.
  - The VST ATLAS:
    - 4500 square degrees in five filters (U, V, R, I and Z), complemented by VHS.
    - comparable to those of the SDSS.
    - examine 'baryon wiggles' (small-amplitude oscillations observed in the power spectrum of galaxies) & imaging base for spectroscopic surveys by the VLT.
  - VPHAS+ — The VST Photometric H-alpha Survey of the Southern Galactic Plane
    - 1800 square degrees using five bands (U, V, Ha, R and I), covering Galactic Plane.
    - 500 million objects including rare star types such as Be and T-Tauri stars.
    - mapping the structure of the Galactic disc and to understand the star-formation history of the Milky Way.

Surveys at ESO - V



- VISTA/Vircam (via CASU)
  - UltraVISTA
    - COSMOS field with unprecedented
    - Y, J, H, and Ks and one narrow-band filter de
    - first galaxies, stellar mass build-up during the p formation.
  - VIKING - VISTA Kilo-Degree Infrared
    - 1500 square degrees of the sky in Z, Y, J, H, a Large Area Survey.
    - determine accurate photometric redshifts, important for weak lensing analysis and observation of baryon acoustic oscillations.
    - hunt for high redshift quasars, galaxy clusters, and the study of galaxy stellar masses.
  - VMC - VISTA Magellanic Clouds Survey
    - 184 square degrees, LMC, SMC, Bridge and Magellanic Stream
    - Multi-epoch observations-> short-period variables
    - study resolved stellar populations, the star formation history of the system as well as to trace its three-dimensional structure.

Surveys at ESO - VISTA

- VISTA/Vircam (via CASU)
  - VVV - VISTA Variables in the Via Lactea
    - target the galactic bulge and a piece of the adjacent plane in Z, Y, J, H, and Ks.
    - 520 square degrees, 355 open and 33 globular clusters.
    - multi-epoch, catalog with ~10^9 point sources, ~10^6 variable objects
    - 3-dimensional map of the Bulge, ages of stellar populations, globular cluster evolution, stellar initial mass function.
  - VHS - VISTA Hemisphere Survey
    - ~20 000 square degrees in J and Ks.
    - 4 magnitudes deeper than 2MASS
    - synergy with DES (5000 square degree also in Hand)
    - examining low mass and nearby stars, studying the merger history of the Galaxy, measuring the properties of Dark Energy through the examination of large-scale structure to a redshift of ~1, and searches for high redshift quasars.
  - VIDEO - VISTA Deep Extragalactic Observations Survey
    - 12 square degree in Z, Y, J, H, Ks
    - study galaxy evolution as a function of epoch and environment to redshift of ~4
    - active galactic nuclei, galaxy cluster evolution, and very massive galaxies.

Surveys at ESO - VISTA

- VISTA/Vircam (via CASU), 2nd generation
  - VVVx - The VVV eXtended ESO Public Survey
    - Extension of VVV
  - VEILS - The VISTA Extragalactic Infrared Legacy Survey
    - Extension of VIDEO plus time domain astronomy
  - GCAV - VISTA Deep Extragalactic Observations Survey
    - 20 clusters of galaxies, synergy with CLASH (Cluster Lensing And Supernovae Search with Hubble), HFF (Hubble Frontier Fields), and Relics programmes.
  - VISIONS - VISTA Star Formation Atlas
    - 550 sqdeg, sub-arcsec near-infrared atlas of all nearby (d < 500 pc) star formation complexes
    - will become the community's reference star formation database, mass spectrum to a few Jupiter masses and spatial resolutions 100-250 AU.
  - SHARKS - Southern H-ATLAS Regions Ks-band Survey
    - 300 deg2, targeted on future deep and wide far-IR and radio survey fields (H-ATLAS, ASKAP, SKA, LOFAR)
    - produce sample of a thousand strong lenses for cosmography studies; study the evolution of the most massive structures in the Universe.

Surveys at ESO

- Future:
  - *4*MOST:
    - 2400 fibres
      - 1624, R>4000–7800, 370–950 nm
      - 812, R>18,500, in 3 different wavelength regimes
    - special operational model: Consortium coordinates the observations of all programs.
    - carry out many different science programs in parallel
  - dedicated 15m class survey wide-field telescope in under (pre-)discussion?
    - Primary ~ 13±2 m
    - FOV = 5 deg2 (~30x MOONS or 4x PFS)
    - 15,000 fibres
    - R = >5000 MR, 20-40000 HR
    - large, single IFU (3'x3', 9xMUSE)

Data access in general

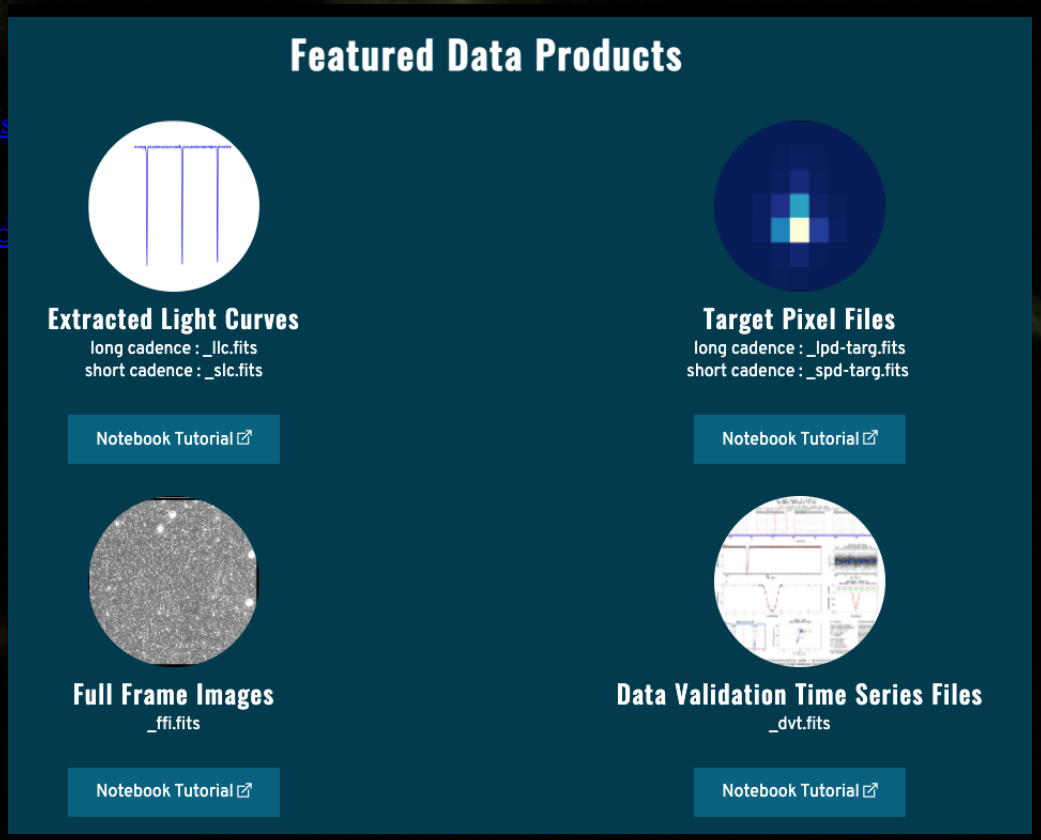
- raw data can be tricky
  - through collaborations
  - ask the PI
  - directly on the websites
  - via scripts, API, astropy:astroquery
- every survey has their own policy

List of Modules

- Simbad: Basic data, cross-identifications, bibliography and measurements for astronomical objects outside the solar system.
- Vizier: Set of 11,000+ published, multiwavelength catalogues hosted by the CDS.
- ESASky: ESASky is a science driven discovery portal providing easy visualizations and full access to the entire sky as observed with ESA Space astronomy missions.
- IRSA Image Server program interface (IBE) Query Tool: provides access to the 2MASS, WISE, and PTF image archives.
- IRSA dust: Galactic dust reddening and extinction maps from IRAS 100 um data.
- NED: NASA/IPAC Extragalactic Database. Multiwavelength data from both surveys and publications.
- IRSA: NASA/IPAC Infrared Science Archive. Science products for all of NASA's infrared and sub-mm missions.
- UKIDSS: UKIRT Infrared Deep Sky Survey. JHK images of 7500 sq deg. in the northern sky.
- MAGPIS: Multi-Array Galactic Plane Imaging Survey. 6 and 20-cm radio images of the Galactic plane from the VLA.
- NRAO: Science data archive of the National Radio Astronomy Observatory. VLA, JVLA, VLBA and GBT data products.
- Besancon: Model of stellar population synthesis in the Galaxy.
- NIST: National Institute of Standards and Technology (NIST) atomic lines database.
- Fermi: Fermi gamma-ray telescope archive.
- SDSS: Sloan Digital Sky Survey data, including optical images, spectra, and spectral templates.
- Alfalfa: Arecibo Legacy Fast ALFA survey; extragalactic HI radio data.
- SHA: Spitzer Heritage Archive; infrared data products from the Spitzer Space Telescope
- Lamda: Leiden Atomic and Molecular Database; energy levels, radiative transitions, and collisional rates for astrophysically relevant atoms and molecules.
- Ogle: Optical Gravitational Lensing Experiment III; information on interstellar extinction towards the Galactic bulge.
- Splatalogue: National Radio Astronomy Observatory (NRAO)-maintained (mostly) molecular radio and millimeter line list service.
- CosmoSim: The CosmoSim database provides results from cosmological simulations performed within different projects: the MultiDark project, the BolshoiP project, and the CLUES project.
- ESO Archive
- ALMA Archive
- GAMA database
- NVAS archive
- Open Expolanet Catalog (OEC)

Examples of Large surveys & access

- Exoplanets:
  - Kepler
    - https://archive.sts
  - TESS
    - https://tess.mit.ed



Examples of Large surveys & access

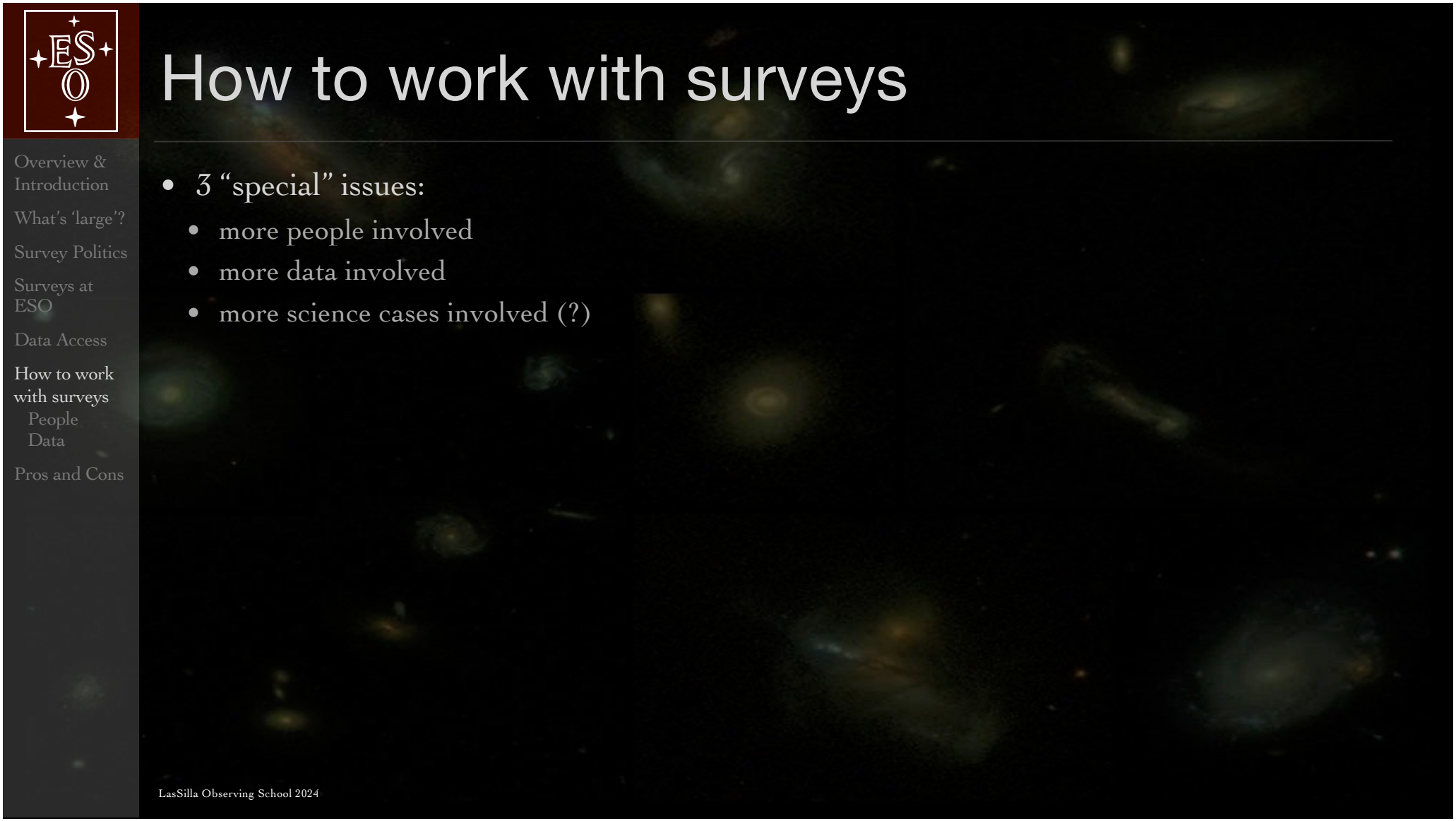
- Galactic:
  - Gaia
    - https://sci.esa.int/web/gaia
    - https://sci.esa.int/web/gaia/-/60243-data-release-2
    - https://gea.esac.esa.int/archive/
  - Gaia-ESO
    - through ESO archive
    - https://www.gaia-eso.eu/data-products
    - https://www.gaia-eso.eu/data-products/public-data-releases/gaia-eso-survey-data-release-4
  - Apogee
    - https://www.sdss.org/dr12/irspec/
  - VVV/VVVx
    - https://vvvsurvey.org/data-releases/
  - Galah
    - https://www.galah-survey.org/

Examples of Large surveys & access

- extragalactic
  - SDSS
    - https://www.sdss.org/
    - https://www.sdss.org/dr16/data_access/
    - scripted:
      - rsync -avz --progress --no-motd https://data.sdss.org/sas/dr17/manga/spectro/redux/v3_1_1/7443/stack/manga-7443-12701-LOGCUBE.fits.gz .
  - 2MASS (1997-2001)
    - https://irsa.ipac.caltech.edu/Missions/2mass.html
  - CANDELS
    - http://arcoiris.ucolick.org/candels/
  - GAMA
    - http://www.gama-survey.org/
  - VANDELS
    - http://vandels.inaf.it/
  - LEGA-C
    - https://users.ugent.be/~avdrwel/research.html#legac

Data access in general

- catalogue databases/services:
  - ESO Archive
    - http://archive.eso.org/cms.html
    - Catalogue facility (data products): https://www.eso.org/qi/
  - NED (NASA/IPAC Extragalactic Database)
    - http://ned.ipac.caltech.edu/
  - CDS/VizieR (Centre de Données astronomiques de Strasbourg)
    - linked to SIMBAD (interactive search for objects by name, coordinate, etc)
    - linked to Aladin (interactive sky atlas)
    - https://cdsarc.u-strasbg.fr/
    - https://vizier.cds.unistra.fr/viz-bin/VizieR (e.g. coordinates 03:32:28.0 -27:48:30.0)
  - via ds9!!



How to work with surveys

- 3 "special" issues:
  - more people involved
  - more data involved
  - more science cases involved (?)

Working in large collaborations

- Shared work load, shared data products, shared publications
- Co-authorship policy
  - builder vs user
  - sign up to papers? proposal, observing vs data products, work done/interest
- People spend months or years on providing a data product that is then released to and used by everyone. They deserve credit and co-authorship
- Coordinate projects
  - often need to announce/sign up to a project. Respect peoples projects, don't scoop collaborators!
  - students projects are usually especially protected
- There are often specific papers that describe the data. Everyone else just points there
- Personality thing, working in teams rather than alone
- Drop into the projects you like, ignore others. Easy to start new collaborations.
- Pick your speciality and work to your strengths!

Working in large collaborations

- Possible roles in large collaborations (selection)
  - PI: general manager and coordinator
  - Manager: organising meetings, coordination
  - Planning the commissioning, operations plan?
  - Planning the observations
  - Data reduction, data pipelines
  - Quality Controls
  - Data products:
    - mosaicing images
    - aperture photometry
    - PSFs
    - morphological measurements
    - collaboration with experts?
    - redshifts, SFR, metallicities, …
  - Data releases
  - Outreach and press releases
  - Dedicated departments?

Working with large datasets

- Data management
  - Manage your data well, think about your requirements
    - at least 3 times the size of your dataset!
  - Keep your folders clean and organized, and sensibly named ("*_new" or "*_final" are not!)
  - Use automatic downloads of data if possible
  - Use external drives to store data, even better, use a docking station
  - Think about transfer speed. 1TB within a department: 1d
    - use rsync, not scp or cp/mv!
  - Think about access speed, reading in catalogue
    - save subsamples in their own catalogue if you're sure!
    - cut down catalogue in one script, analyse in another if necessary.
    - Work on fast drives, store on slow drives
  - Find a way to share your data with collaborators. Lots of it! Long-term!!
    - e.g. OneDrive/Dropbox/Website

Working with large datasets

- CPU management
  - Make sure your computer/server can handle it. If not, talk to people to find a solution.
    - collaborators might have a machine you can use
  - Use codes to automate tasks
    - parallelise everything if possible
    - poor mans parallelisation, working blocks, book-keeping
  - Make sure your codes are efficient.
    - something trivial that takes a few seconds adds up for a large sample (1 minute -> 3 months)
    - Codes should fit the servers, e.g. HPC
    - inverting indices, correlating catalogues (vs scripted)
  - Be organised, set yourself deadlines
    - Code can run while you're on conferences, Paranal or asleep
  - **NOTHING is "fast"**, even deleting/moving data
    - rule of 2π
  - Be "nice" and screen you processes

Working with large datasets

- Script everything and keep a record!
  - You WILL do everything at least twice
  - Collaborators will ask you about the parameters and/or setup you used
  - even TopCat can be automated ('stilts')
- Wherever possible, work with ONE catalogue.
  - If you have 2 from different sources, correlate and combine them!
  - Indexing of indices WILL go wrong!
  - use clever indices
    - wh = where(mag lt 17 and mass gt 10^10 and g-i gt 2 and […])
    - wh2 = where(mag lt 17.5 and mass gt 10^10 and g-i gt 2 and […])
    - Instead use "boolean flags"
- You will not be able to work on your laptop for large parts
  - Catalogues are huge, several GB
  - Work often requires servers, laptop for plots and writing up.
- Backup scripts and setup files!!! (use time machine and/or github)

Working with even larger datasets

- Largest surveys are impossible to download in full.
  - Even using the object catalogues might be difficult
- They provide
  - (often) no access to direct data
  - SQL requests on databases (or similar), also see/use CDS
    - work on the subsamples you need
    - data products pre-measured
    - "Science" changes from measuring things to using the measurement values
  - cutout services for direct data access

- If you ever find out how to work with the SKA or LSST, let me know ;-)
  - Analysing data might take longer than the rate it is taken

A Survey career - Pros and Cons

- What's special about a career as a survey scie
  - Survey science is nearly its own parallel field.
  - The entire field of (extragalactic) astronomy is r
- Pros (selection)
  - Great networking (see my career), you know pe
  - If you have a skill that other surveys need, you'
  - Teamwork, collaborations
  - Access to the "best" data
  - All data have been checked by experts, much les
  - Papers are usually very well pre-refereed
  - Number statistics are great
  - Publications are often fast
- Cons (selection)
  - Publication record is very different, and you sho
    - Many co-authored papers, many citations, but typically
    - People often don't understand that working in large su
    - A problem in departments that count lead-author papers
  - Dependent on collaborators to deliver data products -> Publications are often slow
  - Higher pressure because collaborators depend on your data products, too?
  - Datasets are ridiculous and a lot of work to handle

1	2011ApJS..197...35G	2011/12	cited: 1416	**CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey** Grogin, Norman A.; Kocevski, Dale D.; Faber, S. M. *and 104 more*
2	2011ApJS..197...36K	2011/12	cited: 1380	**CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics** Koekemoer, Anton M.; Faber, S. M.; Ferguson, Henry C. *and 121 more*
3	2014ApJ...788...28V	2014/06	cited: 776	**3D-HST+CANDELS: The Evolution of the Galaxy Size-Mass Distribution since z = 3** van der Wel, A.; Franx, M.; van Dokkum, P. G. *and 27 more*
4	2006ApJ...650...18T	2006/10	cited: 447	**The Size Evolution of Galaxies since z~3: Combining SDSS, GEMS, and FIRES** Trujillo, Ignacio; Förster Schreiber, Natascha M.; Rudnick, Gregory *and 15 more*
5	2015MNRAS.452.2087L	2015/09	cited: 377	**Galaxy And Mass Assembly (GAMA): end of survey report and data release 2** Liske, J.; Baldry, I. K.; Driver, S. P. *and 67 more*
6	2004ApJS..152..163R	2004/06	cited: 367	**GEMS: Galaxy Evolution from Morphologies and SEDs** Rix, Hans-Walter; Barden, Marco; Beckwith, Steven V. W. *and 13 more*