1 The LEGA-C Public Spectroscopic Survey

PI: A. van der Wel, Max Planck Institute for Astronomy, Heidelberg, Germany

The immediate goal of the Large Early Galaxy Astrophysics Census (LEGA-C) is to obtain deep continuum spectroscopy of several thousand $z = 0.6 - 1$ galaxies in the rest-frame wavelength range around 4000Å. Such a dataset enables a wide variety of scientific exploration, most directly through the measurement of stellar velocity dispersions (dynamical masses) as well as stellar ages and abundances. Such information for large samples is presently only available for galaxies in the present-day universe, which leads to the enormous legacy value of the LEGA-C dataset.

Our improved knowledge of the distribution of mass among galaxies at half the age of the universe, combined with accurately measured ages and metallicities, will cause a leap in our understanding of the growth of galaxies through star formation and merging. A broad segment of the astronomical community in Europe and world wide will benefit from the fully reduced datasets and the high-level derivats. When similarly high-quality data become available for galaxies at higher redshifts over the next decades, the LEGA-C dataset will provide the crucial link between present-day galaxies and the population that existed during the first few billion years after the Big Bang.

The LEGA-C survey covers 1.3 square degrees (that is, 79% of the UltraVISTA footprint in the COSMOS field). 3100 galaxies has been selected from the UltraVISTA K-band catalog as the primary sample to be observed with VIMOS, each with 20 hours of integration time. Approximately 2500 of these are expected to have a continuum signal-to-noise ratio of 10Å^{-1} or better.

This program requires and has been assigned 128 nights of observing time. This includes all overheads and weather losses. After subtraction of overheads this amounts to 640 hours of on-source integration time (∼300 years when multiplied by the number of targets). The observations require dark and clear (but not photometric) sky, as well as DIMM seeing better than 1.3 arcsec. Frames taken under poorer seeing contribute little to the integrated signal-to-noise ratio of the co-added spectra. Thin cirrus is acceptable under good seeing conditions (see Sec. 2.2 for details).

All observations are at $R.A. = 10h$. Combined with the required observing conditions, this constraint implies that the total survey duration will be 4 years if all suitable available time is scheduled for LEGA-C observations, with observations likely completed in P100 (March 2018). Weather and wind losses, as well as scheduling of other high-priority programs at similar R.A. that require dark time, will push the completion date back. The first data release with approximately 1/4th of the final data volume, scheduled for June 2016, will already be unparalleled in terms of sample size and data quality.

2 Survey Observing Strategy

Sample selection. All observations are in the COSMOS field at R.A. = 10h; Dec. = +2 deg. Our parent sample is drawn from the Ultra-VISTA K band-selected catalog produced by co-I Adam Muzzin (Muzzin et al. 2013, ApJS 206, 8). Spectroscopic redshift information from zCOSMOS as well as small numbers from other observing campaigns is used when available. For objects without spectroscopic redshifts, we use the photometric redshifts measured by Muzzin et al. (2013). The target field has outstanding multi-wavelength information.

We will target galaxies in the redshift range $0.6 < z < 1.0$ and with a K-band total magnitude limit running from $K_{\rm AB} = 21.1$ at $z = 0.6$ to $K_{\rm AB} = 20.4$ at $z = 1$. The sliding limit is chosen to accommodate the increasingly red $I - K$ colors with redshift, thus ensuring sufficient signal in the optical spectra, while retaining the value of having near-infrared, and not optically selected, samples. This parent sample contains 9878 galaxies, and the RA-Dec distribution of the parent sample is shown in Figure 1 (right-hand panel). The mass and color distributions of the parent sample are shown in black in Figure 2. We have worked out a mask design and tiling strategy that can produce a sample of ∼2500 galaxies with $S/N > 10\text{Å}^{-1}$, needed to achieve our goals.

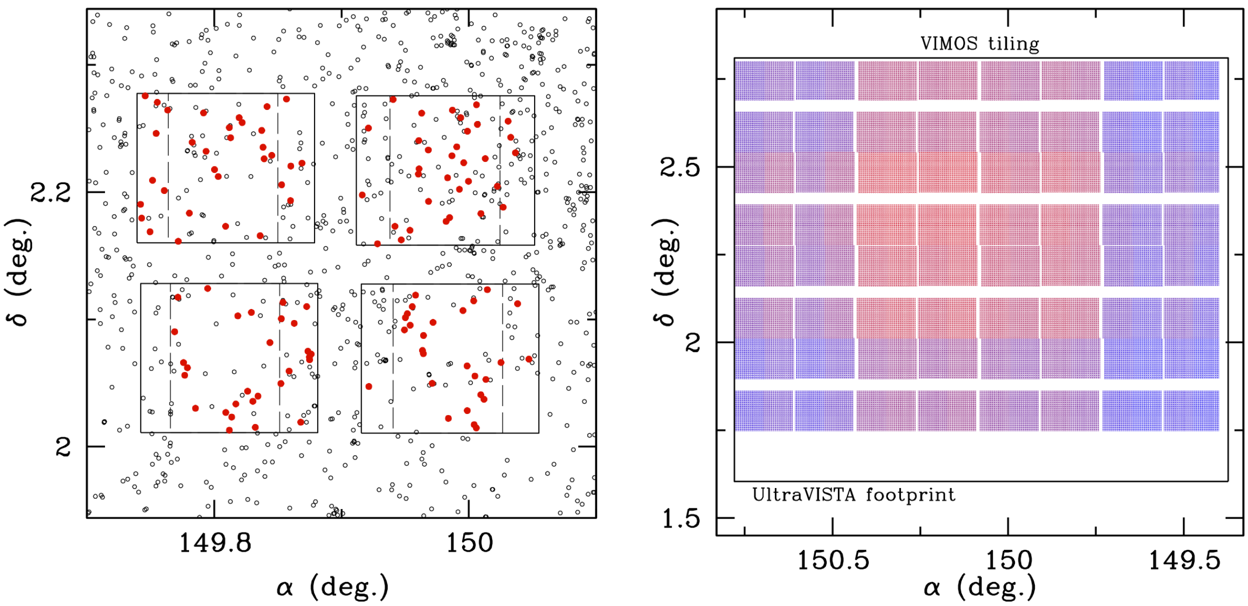

Figure 1: Left: Example of one of our mask designs. The solid lines represent the layout the four VIMOS detectors (accurate to 3 arcsec). The dashed lines delineate the preferred location of slits to optimize wavelength coverage. The points represent potential targets in the parent sample, where the red points are included in the actual mask design that, accounting for slit collisions. The gaps between the detectors and the restrictions imposed by wavelength coverage leaves gaps in the sky coverage. Right: Footprint of the 32 planned masks, color-coded by chronological order to illustrate the survey's progress: red is done first, blue last. Interlacing the positions of the masks full sky coverage with minimal overlap in the E-W direction is achieved. Only the detector gaps in the N-S direction remain.

Mask design. The slits in our masks will have a spatial N-S orientation. To ensure homogeneous wavelength coverage, optimized including the Balmer/4000Å break, Hδ at 4101Å, and the G band at 4300Å in all spectra (see Figure 3), slits are preferably assigned to targets that are more than 100 arcsec away from the edges of the detectors in the E-W direction. Slit collisions are avoided by forcing a minimum distance of 8 arcsec in Declination.

Slits are assigned with the following priority: 1) galaxies at $0.6 < z < 1.0$, ranked by K magnitude (always brighter than the K-band limit); 2) two blue stars at opposite sides of the detector for telluric absorption correction and up to three other stars in square slits for alignment; 3) galaxies at $1.0 < z < 1.5$ and brighter than $K = 20.4$; 4) galaxies at $0.6 < z < 1.0$, ranked by K magnitude, but fainter than the K-band limit. As such, the typical pointing has a 95% slit coverage, with ∼130 slits, ∼80 of which have anticipated $S/N > 10\text{Å}^{-1}$ targets.

Tiling. The arrangement of the detectors and our requirements for wavelength coverage leaves gaps in the sky coverage (Fig. 1). The gaps in the R.A. direction, when including the 100 arcsec columns on the detector edges, are equal in width to the central regions of the detectors. Hence, by interlacing two tiles complete sky coverage in the E-W is accomplished. Only the ∼2 arcmin gap in the N-S direction between the detectors remains.

16 pairs of such interlaced tiles cover almost the full, roughly square, COSMOS/Ultra-VISTA area (see right-hand panel of Figure 1): this optimal tiling strategy covers 79% of the UltraVISTA footprint. The remaining 21% lie in strips at the South end of the field, and in the four 2 arcmin wide strips that are due to the gaps between the VIMOS detectors. 3088 galaxies from the primary are assigned to a slit, an optimum limited by slit collisions. Remaining space for slits will be used to observe ∼ 1000 filler galaxies.

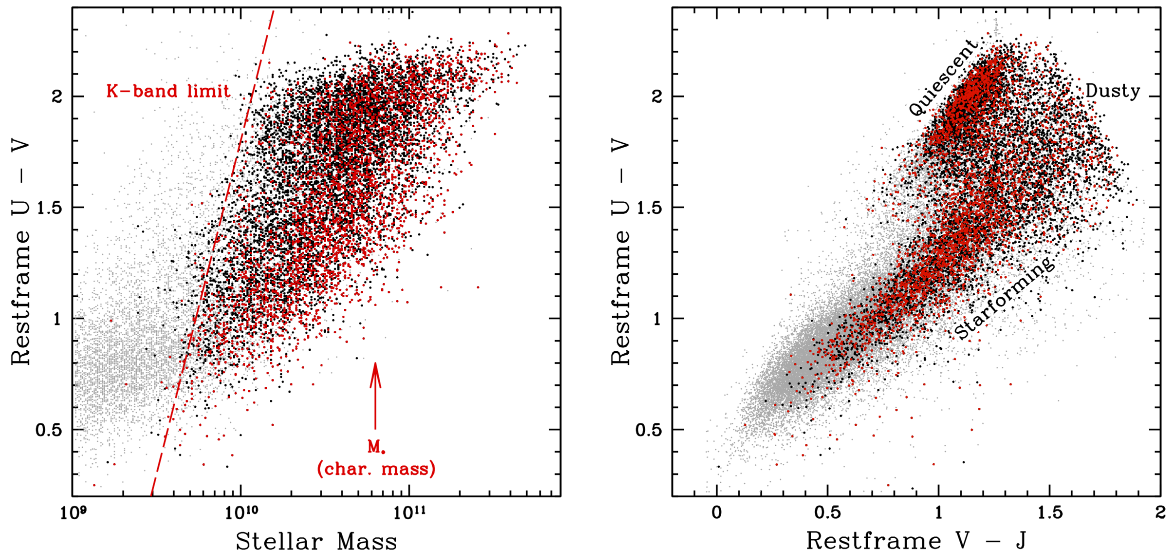

Figure 2: Stellar mass-color (left) and color-color (right) distribution of the K-band selected target sample (black and red points) from the COSMOS/UltraVISTA catalog (Muzzin et al. 2013, ApJS 206, 8) at $z = 0.6 - 1.0$. The light gray points show the full COSMOS/UltraVISTA catalog (without the K-band magnitude limit). The red points are those who have been assigned slits in our mask designs. The targets span a wide range in mass and color. The color dispersion is primarily due to age differences and variation in dust properties, as labeled in the right-hand panel. A population of passive galaxies is well-defined in color-color space due to the Balmer/4000Å break traced by the $U - V$ color.

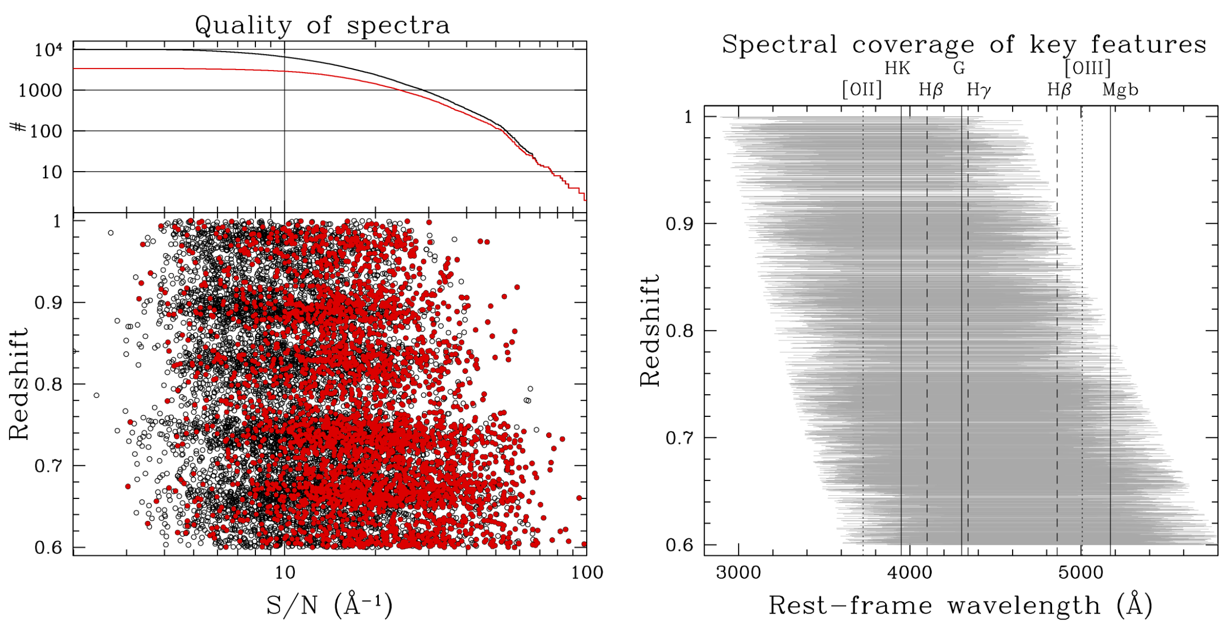

Figure 3: Left: Distribution of S/N vs. redshift in our galaxies in our sample. The points represent the parent sample; the red points are those that will be observed. The top, inset panel shows the cumulative number of galaxies as a function of minimum S/N, black for the parent sample, red for the survey sample. Right: Rest-frame wavelength range for the spectra of the survey sample. The variation at fixed redshift is due to the variation in slit position. Vertical lines indicate the location of important spectral features.

Table 1: MOS scheduling requirements

Period	N. nights	N. observing runs	N. observers per run	Average run length
P94	37	4	1	9
P95	10	1	1	10
P96	43	4	1	11
P97	10	1	1	10
P98	43	4	1	11
P99	10	1	1	10
P100	33	4	1	11

2.1 Scheduling requirements

Pre-imaging is not required for the UltraVISTA/COSMOS field and we will design the masks on the source catalog astrometric grid, which has an absolute astrometric accuracy of better than 0.1 arcsec. According to Marina Rejkuba of the User Support Department PILMOS is recommended, in particular for this survey and field.

Table 1 shows the distribution of observing nights, where we list the number of nights that a LEGA-C observer will be at Paranal for observations. For P94 we follow the phase 2 schedule as listed in the User Portal on the ESO website; for the other even periods we assume a 10 night run with 0.3 nights, two 11-night runs with 0.7 nights, and a 10-night run with full nights. For the odd periods we assume a single 10-night run consisting of half nights (counting 5 hours per half night). In short, the schedule essentially assumes that all suitable observing periods are assigned to the LEGA-C survey, close to the actual schedule for P94.

On a given night we will typically use a single mask. Given the hour angle constraints (Sanchez-Janssen et al. 2014) we can observe the COSMOS field for 6 hours per night. Acquisition and alignment can be executed before this 6-hour period. Flexure must be corrected at 1-1.5hr intervals with Active Flexure Compensation system, after which alignment must be verified and, if needed, corrected. This process will take 4 times 10 minutes. Additionally, 1-minute long read-outs contribute another 20 minutes of overhead, leaving 5 hours of on-source integration time. This amounts to 37.5% overhead for full nights during which the COSMOS field is up for 6 hours.

For partial nights the overhead is smaller: for 0.7 nights it is 20%; for 0.3 or 0.5 nights it is 30%. Assuming scheduling as described above, we arrive at an effective overhead of 28% with a total execution time of 889 hours. Including weather and wind losses (12% and 6%, respectively) we arrive at $889/0.88/0.94 = 1075$ hours. Evidently, these numbers depend on details in the scheduling.

2.2 Observing requirements

All observations are carried out with the high-resolution red grism (HR-Red) and our field comes near Zenith. We prioritize our masks in a chronological sense by radial distance from the geometric center of the UltraVISTA field. That is, the center of the field (where most supporting datasets, such as CANDELS, are available) will be completed first.

For 10 or 11-night runs we will typically need 4 different masks for each quadrant: we finish a previously observed but incomplete mask, execute two masks in their entirety, and start a new mask. In case the VANDELS team needs an extra slot, we can accommodate our colleagues by reducing our number of masks to 3 and starting an observing run with a new mask, leaving any previously unfinished mask for a later observing run.

Our observations require dark time. To be precise, our analysis of the sky background database assembled by F. Patat shows that the requirement is a Moon phase <0.3 and Moon-target separation is >60 degrees. Due to

Table 2: MOS observing requirements

Period	RA	DEC	N. of distinct mask sets	Grism and filter setup	Tot. exp. time [hrs]	Tot. exec.[2] time [hrs]	Priority	Av. moon constraints
P94	10	+2	8	HR-Red	146	200	1	dark
P95	10	+2	2	HR-Red	31	44	1	dark
P96	10	+2	9	HR-Red	147	202	1	dark
P97	10	+2	2	HR-Red	31	44	1	dark
P98	10	+2	9	HR-Red	147	202	1	dark
P99	10	+2	2	HR-Red	31	44	1	dark
P100	10	+2	9	HR-Red	107	153	1	dark

the pointing restriction, dependent on wind direction, we anticipate that the survey will be somewhat delayed (5-7%): our equatorial field will regularly be in the North, which is the prevailing wind direction at Paranal.

We wish to implement a hard DIMM seeing limit of 1.3 arcsec for clear skies. Adding data with worse seeing would contribute to the final S/N only negligibly while using an amount of valuable observing time. Sub-optimal transparency is acceptable, provided that the seeing is good. A S/N-based criterion for required seeing+transparency conditions is that seeing should be better than the transparency times 1.3 arcsec (i.e., better than about 0.9 arcsec for 70% transparency). This linear dependence is chosen to account for the fact that our sources are not point sources (for which the dependence would be proportional to the square root of the transparency to conserve S/N).

3 Survey data calibration needs

Normally, night-time calibrations (arcs and flats) must straddle the science observations to account for the varying rotator angle and altitude. This would require slewing to zenith and re-acquiring the mask several times per night, greatly increasing the overhead. Instead, we will take all the necessary flats and arcs once the source is no longer at a suitable airmass. We will then park the telescope and take calibrations at several rotator angles covering the range of angles used throughout the night. The availability of these calibrations in the ESO archive will ensure the legacy value of the data.

The smallest velocity dispersions in our sample will be $\sim$60 km/s, such that we require a precision in the wavelength calibration of better than 40 km/s or 0.8Å at 7500Å. ESO's report on the accuracy of the wavelength calibration states this is achieved for the HR_red grism for all rotator angles. However, we note that the wavelength calibration for individual exposures will be verified and, if needed, improved by comparing the predicted and actual locations of atmospheric OH lines in the sky spectra.

Telluric standard star spectra will be taken simultaneously with the galaxy spectra by including multiple slits with multiple blue stars in our mask designs. By maximizing the separation of those slits in the wavelength direction, full wavelength coverage for our galaxy spectra is ensured.

Spectrophotometric standard stars will be observed in standard fashion. We will verify the effect of the varying response curve as a function of slit location in the dispersion direction by comparing the flux-calibrated spectra with the extensive photometric dataset available for our parent sample of galaxies. In particular, the Subaru r-, i-, and z-band photometry constrain the spectral slope in the wavelength regime of the spectra and will be used to fine tune the flux calibration, using an equivalent procedure as described in Milvang-Jensen et al. (2008, A&A 482,419).

4 Data reduction process

The data flow diagram below (Figure 4) illustrates the reduction and analysis process, from beginning to end. It starts with mask design and OB creation and observations, as described above. The remainder of the process, the topic of this section, includes four modules: the VIMOS data reduction pipeline, the spectroscopic analysis pipeline, the auxiliary analysis pipeline, and the data release module.

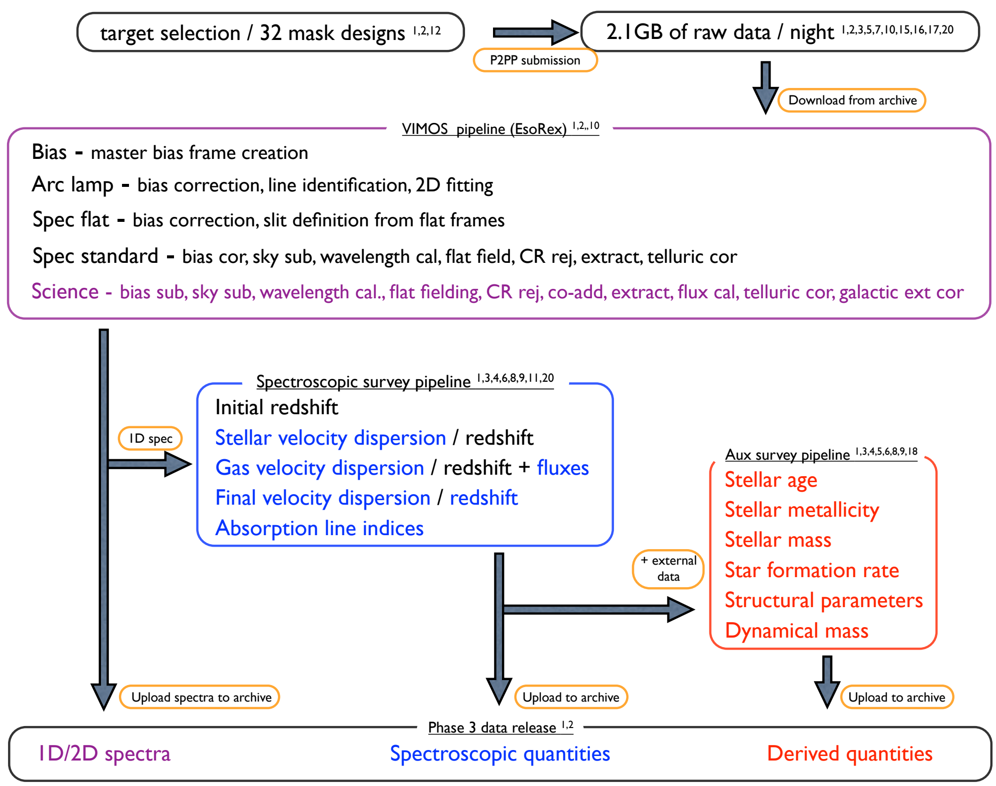

Figure 4: Data flow for the LEGA-C survey. Each of the modules is separately indicated, and inter-connected as illustrated by arrows. Numbers in superscript indicate the members of the team (as identified in Table 2) involved in executing each module. The black-lined modules at the top represent observing preparations and observing. Orange-lined comments indicate the transfer/activity required to connect modules in practice. The purple-lined module represents the data handling, from raw frames to fully calibrated 2D spectra and extracted 1D spectra. The colored text indicates the final data product that will be part of the final data release. The blue-lined module represents spectroscopic analysis: this includes all measurements that are solely and directly obtained fro the spectra. The red-lined module presents auxiliary analysis, enabled by the combination of the acquired spectroscopy and external datasets (multi-wavelength photometry and HST imaging). The black-lined module at the bottom represents the phase 3 data release.

Data reduction will be performed with the EsoRex command line tool. The following frames will be obtained for each OB and each quadrant: bias exposures, flat field exposures, arc lamp exposure, science exposure. The science exposure can be of a standard star or of the actual targets.

Our standard version of the data reduction procedure will be performed with the EsoRex command line tool in the following steps:

- Master Bias `vmbias`. The bias exposures are combined into a master bias bias frame
- Calibration Frames `vmmoscalib`. The master bias frame is subtracted from the arc lamp and flat field exposures, which are subsequently used to produce rectified and wavelength calibrated arc lamp spectra for all slits, defined by tracing the slit edges in the flat field exposures. The slit definitions are also used to produce a normalized master flat field for all slits. Distortion and curvature coefficients are saved.
- Science Frames `vmmosscience`. Bias subtraction, sky subtraction, wavelength calibration, and flux calibration are performed using the calibration frames and coefficients listed above. Sky subtraction is performed on individual, dithered frames. This is preferred over producing a master sky frame by combining dithered exposures due to the unnecessarily increased noise level for the latter approach. To sample different parts of the detector we will dither over 2 positions separated by 3 arcsec.

The reduced 2D spectra from individual OBs constitute the units from which the basic final data product will be produced: optimally extracted, co-added 1D spectra. Adopting a Gaussian kernel, the width of which is set by the seeing, we will extract 1D object and noise spectra from the individual 2D spectra. Then, weighing by the integrated S/N, we co-add these individual 1D spectra to produce co-added 1D spectra with optimized S/N. Using the same S/N-based weights we coadd the individual 2D spectra to produce coadded 2D spectra. All individual and co-added spectra will be corrected for Galactic foreground extinction.

Derived data products in the form of redshifts and physical parameters are measured directly from these spectra. These parameters fall in two separate categories, but are not fully independent. The first category contains the kinematic parameters, that is, stellar and gas velocity dispersions. The second category contains absorption and emission line strengths as well as stellar ages and stellar/gas metallicities based on these.

The data flow is non-linear and there exist feedback loops between the several steps. In the following we describe in detail the production of our high-level data product.

- Initial redshift, $z_{\rm ini}$. An emission line template and an absorption line template are cross-correlated with the 1D spectrum. The more robust of the two – simply determined by the presence or absence of multiple emission lines – is adopted as the initial redshift estimate, which will be more accurately and iteratively determined as described below.
- Stellar velocity dispersion, σ_*. Bruzual-Charlot spectroscopic synthesis models for stellar populations with a range of ages and metallicities are broadened by the instrumental resolution and Doppler broadening from internal motions after subtracting the continuum of both the model and galaxy spectra by means of additive and multiplicative polynomials. Emission lines are masked. Ideally, Balmer absorption lines are also masked as these can be broadened by fast rotation of A stars. However, for spectra with modest S/N, using the Balmer lines can be a necessity to obtain a measurement of σ_* Besides allowing for small changes in redshift with respect to $z_{\rm ini}$ the Doppler broadening arising from internal stellar motions, the stellar velocity dispersion σ_*, is the (only) free parameter that is varied to find the best-fitting match in a χ^2 sense. The uncertainty is estimated by randomly varying the galaxy spectrum based on the flux uncertainties and refitting with the library model spectra. The precise redshift from the best-fitting model and σ_* is adopted as the absorption line redshift, $z_{\rm abs}$.
- Emission line redshifts, fluxes and widths. Assuming $z_{\rm abs}$ (or $z_{\rm ini}$ if $z_{\rm abs}$ is not available) emission lines are identified by fitting the continuum with the library of model spectra (masking regions with possible emission) lines, subtracting the best-fitting spectrum and measuring the emission line fluxes in the residual. Emission lines are jointly refit with redshift, flux and velocity dispersion as free parameters. Redshift ($z_{\rm em}$) and velocity dispersions ($\sigma_{\rm em}$) are assumed to be the same for all lines, but fluxes are allowed to vary independently with the exception of the two [OIII] components which are set at a fixed ratio of 1/3. A weighed average of the absorption and emission line redshift is adopted as the galaxy redshift ($z_{\rm gal}$).

- Absorption line indices. Adopting $z_{\rm gal}$, absorption line equivalent widths are measured from the emission-subtracted spectra for the following metal species as far as the spectra coverage allows: CN1, CN2, Ca4227, G4300, Fe4383, Ca4455, Fe4531, Fe4668, Fe5015, Mg1, Mg2, Mgb, Fe5270, Fe5335, and Fe5406. In addition, Balmer line and 4000Åbreak strengths are measured. We will adopt the Lick system as index definition.

- Stellar age and metallicity. Our model spectral library is now expanded to include a broad range in star formation histories, including bursts with a broad range in strengths, and dust absorption. Assigning likelihoods to each model based on the measured indices probability distributions for age and metallicity are derived, marginalized over all model parameters.

In the above we have extracted information from the spectra alone. A host of additional knowledge about the galaxy sample is provided by multi-wavelength imaging datasets/catalogs and high-resolution imaging from HST. The former have been consolidated already in the form of publicly available catalogs from UltraVISTA (also used to select our sample in the first place). The latter needs work: no general purpose catalog with structural parameters from the available ACS/HST COSMOS imaging is available. Integrating our spectroscopic survey with these external datasets is an explicit goal of our survey.

The complementary multi-wavelength catalogs will provide the following information:

- Stellar masses. With the accurate redshifts and probability distributions for age and metallicity, the stellar mass estimates based on the broad-band spectral energy distributions will be vastly improved. The spectroscopic constraints can be implemented as powerful priors in fitting the photometry.

- Star formation rates. Similar to the stellar masses, the directly measured emission line luminosities and accurate redshifts allow us to perform joint fits of the photometric and spectroscopic measurements to obtain improved estimates of the star formation rate.

Our own analysis of the HST data will provide the necessary context to interpret the dynamical properties of the galaxies in the LEGA-C sample.

- Structural parameter catalog. HST imaging in the same wavelength range as our spectroscopy is publicly available. From these data we will construct mosaics for each of the VIMOS footprints. The P.I. has extensive experience with analyzing such datasets from the CANDELS survey (van der Wel et al. 2012, 2014). A library with appropriate point spread functions across the ACS detector will be constructed, `SExtractor` will be used to detect sources, which will be used as input for the structural parameter fitting program `galfit`. Neighboring sources will be simultaneously fitted or masked. This sequence is executed by the value-added wrapper `galapagos` which can be trivially adapted for our purposes. We will choose the Sersic profile as our fitting profile, providing us with measurements of the effective radius, Sersic index, axis ratio and position angle.

- Dynamical mass. The velocity dispersion, size and Sersic index allow us to estimate the virial mass, where the Sersic index is used as a proxy for dynamical structure.

5 Manpower and hardware capabilities devoted to data reduction and quality assessment

As illustrated in Figure 4, the project is linear in structure, with only few tasks that can be performed in parallel. As the P.I., van der Wel will oversee all tasks and perform a significant fraction of the observations. Kai Noeske was recruited as Survey Manager at the MPIA as of November 1st 2014. He will oversee the data reduction process and the preparation of data releases. Noeske has extensive experience in the field, and in particular with large spectroscopic surveys (DEEP2 on Keck).

Two Survey Scientists will take the lead in two key scientific aspects of the LEGA-C survey. Rachel Bezanson will lead the effort to extract and interpret the kinematic information from the spectra. Anna Gallazzi will lead the effort to extract and interpret the information regarding stellar content.

Besides the four individuals mentioned above, the entire LEGA-C team is composed of experts in the relevant areas of survey execution (Franx, Rix, van Dokkum, Wolf, Sobral, Labbé), kinematic analysis (Franx, Rix, van Dokkum, van de Sande), spectroscopic synthesis models (Charlot, Pacifici), and abundance measurements (Wild, Bell): all necessary expertise to execute the LEGA-C survey is readily available. The distribution of tasks among the team members is as follows:

- Target selection and mask design (van der Wel, student, Maseda). As described in detail in the proposal the parent catalog is available and target selection approach has already been worked out in detail. Integration into the mask design software will be overseen by the PI and executed by current graduate student Maseda and from next year onward by the student (vac.) who will be hired early 2015 to ensure overlap with Maseda.
- Observations (van der Wel, Noeske, Muñoz, Bezanson, Sobral, van de Sande, Brammer, Wolf, Labbe). The bulk of the observations will be carried out by van der Wel, Noeske, and Muñoz (ESO fellow and VIMOS service mode observer). In addition, several co-I's have indicated they can occasionally travel to Paranal. We have sufficient commitments to cover all observing runs. We only have visitor mode runs: service mode runs will not be scheduled as we will not use pre-imaging.
- VIMOS pipeline (Noeske, van der Wel, Franx, Sobral). The data will mostly be reduced Noeske. Franx will overview and optimize the process. Sobral's unique previous experience with the exact same instrument configuration prepares us to efficiently and quickly process the first incoming datasets. Reflex has been installed at the computers in Heidelberg and we have already executed a test run with the raw data from Sobral's program to ensure that data reduction can start immediately upon arrival of the first data.
- Spectroscopic pipeline (van der Wel, Gallazzi, Bezanson, Noeske, Franx, Rix, Pacifici, van de Sande, Wild, Wolf). The initial redshift algorithm will be designed by the PI and executed by the postdoc. Bezanson will lead stellar velocity dispersion measurements. Emission line analysis will be led by Pacifici. Absorption line measurements and derived analysis are the focus of Gallazzi's and Wild's contributions. All these tasks are done in close collaboration with the PI and Franx. Rix and Wolf will advice the team members on algorithm optimization. All required algorithms exist, and have been applied in numerous previous publications by the team members. The task at hand is to consolidate the existing machinery into a consistent, optimized pipeline.
- Auxiliary pipeline (van der Wel, Gallazzi, Bezanson, Rix, Pacifici, Brammer, Bell). Stellar mass and star formation rates from photometry alone (Muzzin et al. 2013) will be improved by adding the spectroscopic results from the survey. Bell and Pacifici will lead this effort. Based on previous and extensive experience with HST datasets (CANDELS and 3D-HST) the PI will oversee and execute the analysis of the COSMOS ACS imaging data to measure structural parameters of the survey galaxies. The integration of these measurements with the kinematic measurements by Bezanson will produce dynamical mass measurements. All algorithms have been created for previous projects; the focus will be on their adaptation for the purposes of the LEGA-C survey.

Table 3: Allocation of resources within the team

Name	Function	Affiliation	Country	FTE allocated to project
[1]van der Wel	Principle Investigator	MPIA	Ger	0.7 (36n)
[2]Noeske	Survey Manager	MPIA	Ger	0.7 (30n)
[3]Bezanson	Survey Scientist, kinematics	Arizona	USA	0.5 (20n)
[4]Gallazzi	Survey Scientist, stellar pops	INAF,Arcetri	Ita	0.2
[5]Student (vac)	mask design and data red.	Leiden	NL	0.8
[6]Franx	quality control	Leiden	NL	0.1
[7]Muñoz	observing	ESO	Chi	0.1 (30n)
[8]Rix	algorithms	MPIA	Ger	...
[9]Pacifici	em lines	Yonsei	S-Ko	0.1
[10]Sobral	data reduction	Lisbon	Por	(10n)
[11]Wild	ages and SFH	St.Andrews	UK	0.1
[12]Maseda	masks	MPIA	Ger	0.1
[13]Charlot	spec synthesis	IAP	Fra	0.1
[14]Muzzin	photometry	Cambridge	UK	...
[15]van de Sande	abs line disp	Leiden	NL	0.1 (20n)
[16]Labbé	photometry	Leiden	NL	(10n)
[17]Brammer	obs, HST data	STScI	USA	(20n)
[18]Bell	stellar mass	Michigan	USA	0.1
[19]van Dokkum	HST imaging	Yale	USA	...
[20]Wolf	algorithms	ANU	Aus	0.1 (10n)

- Phase 3 data release (Noeske, van der Wel, student). As described further below we will prepare the reduced data and derived data products for submission to the ESO archive. Noeske will perform the preparations for the spectra and derived data products, respectively. We note that the main task is to prepare the data to abide by the ESO regulations. The data volume to be uploaded is small, and takes mere hours by standard internet connection.

The FTE contributions from the team members are listed in Table 3. Co-investigators without a listed FTE are those who will not be involved in the execution of the LEGA-C survey on the weekly or monthly basis. Their contributions will consist of attendance to team meetings and the availability of their broad range of expertise.

In preparation of the LEGA-C proposal we have created the necessary code to optimally design masks. This optimization consists of selecting and prioritizing targets from the parent catalog, and optimizing slit assignments considering the geometric constraints of the VIMOS detectors and slit collisions. Slits are also automatically assigned to alignment verification stars and telluric correction stars. This software is integrated into the VIMOS mask design software, which will be used to produce the actual binary files for mask cutting. We have already performed this process for the first mask to be observed and we also created the necessary phase 2 observations blocks and uploaded those to the server. In short, several months before observations start we are fully prepared.

Transferring the data to disk (at MPIA and Leiden) will only take a few hours, as the volume is just 2.1GB per night. The total amount of raw data over the entire duration of the LEGA-C survey is ∼300GB which can be easily stored and processed on the available computers.

The basic data product of the LEGA-C survey is the fully calibrated 1D extracted spectrum. For each unique, targeted galaxy the VIMOS pipeline will produce a FITS file flux spectrum on a fixed, survey-wide applied, wavelength grid, along with the error spectrum, sky spectrum, and flags.

Higher level data products include all spectroscopic quantities (redshift, velocity dispersion, absorption/emission line strengths), which will be presented in a FITS catalog. Byproducts will include separate absorption and emission line spectra.

After combining the spectroscopic quantities with auxiliary datasets we will construct a final, highest-level data product, consisting of a catalog with a wide range of physical quantities (stellar mass, star formation rate, stellar age, stellar metallicity, gas phase metallicity, structural parameters, dynamical mass).

All these data products will be provided with the adequate header information, following the Phase 3 submission guidelines.

6 Data quality assessment process

Quality control is inseparable from data reduction and analysis. It begins with verification of mask alignment and ends with the comparison between our high-level end products with external, independent data products. Key elements include the following:

- Box-shaped slits with stars will be included in each of the masks. Periodic through-slit images will be used to continuously monitor the accuracy of the alignment.
- DIMM seeing and instrument seeing will be monitored and compared.
- Stability of the instrument in terms of shifting of the spectra on the detector will be continuously monitored.
- Slit definitions based on the spectroscopic flatfield exposures will be visually compared with the science exposures.
- Location of telluric emission lines will be compared with the expected locations based on the pipeline wavelength calibration.
- Cosmic ray flagged pixels will be verified against the raw science frames.
- The observed flux density in the spectra will be compared with the expected flux density predicted by photometry.
- The observed signal-to-noise ratio will be compared with the expected signal-to-noise ratio to confirm, and if necessary update, the exposure time requirement.
- Initial redshift estimates will be visually vetted and compared with photometric redshifts from Ilbert et al. (2009, ApJ 690, 1236), Whitaker et al. (2011, ApJ 735, 86), Muzzin et al. (2013, ApJS 206, 8), and Skelton et al. (2014, arXiv1403.3689).
- Stellar and gas phase velocity dispersions will be compared for consistency with Sobral et al. (2013, MNRAS 428, 1128).
- Stellar mass estimates will be compared against previously stellar mass estimates based on photometry from Muzzin et al. (2013, ApJS 206, 8) and Skelton et al. (2014, arXiv1403.3689).
- Structural parameters will be compared with previously published measurements from van der Wel et al. (2012, ApJS 203, 24).
- Before phase 3 submission all data products will be tested in terms of usability by the general community by verifying that the standard scaling relations and known global properties of the galaxy population are correctly reproduced.

DR	Date	Content[a]	No. objects (cumulative)
1	June 2016	Data Product I	1000
2	December 2016	Data Product I+II	2200
3	December 2017	Data Product I+II+III	3400
4	December 2018	Data Product I+II+III	4100

Table 4: [a] see Section 7

7 External Data products and Phase 3 compliance:

Adhering to the color coding of the various modules defined in Figures 4 and 5, the science data products delivered through the Phase 3 process are summarized here.

Data product I (spectra from the VIMOS pipeline).

- One-dimensional and two-dimensional wavelength and flux calibrated spectra in units of erg/s/cm^2/Å for each target, along with the corresponding signal-to-noise maps. All these items will be presented in FITS format. Individual spectra as well as co-added spectra will be released. A unique target identifier will be used for each object and used in all data products I, II, and III.
- As requested, we will examine each spectrum for variability. Those with pronounced nuclear flux contributions will only enter the sample 'by accident', and we will flag those as possibly variable and make time series spectra available.

Data product II (catalog from the spectroscopic pipeline):

- Redshifts as measured from stellar and gas components, and their combination as the final redshift estimate.
- Stellar and gas velocity dispersion.
- Absorption line indices on the Lick system (see Section 4)
- Emission line fluxes and equivalent widths ([OII], [OIII], Balmer series)

Data product III (catalog from the auxiliary pipeline):

- Stellar ages, metallicities, and masses.
- Star formation rates.
- Effective radii, Sersic indices, axis ratios, position angles from HST imaging.
- Dynamical masses.

All data products will be prepared following the guidelines as described in the Science Data Product Standard document, and Section 4 therein pertaining to spectroscopic products in particular. Two publications will accompany these data releases. A first publication, scheduled for late 2015, will introduce the survey and its goals. A second publication, scheduled for late 2016, will describe the first data release as well as the data reduction process and the spectroscopic and auxiliary pipelines.

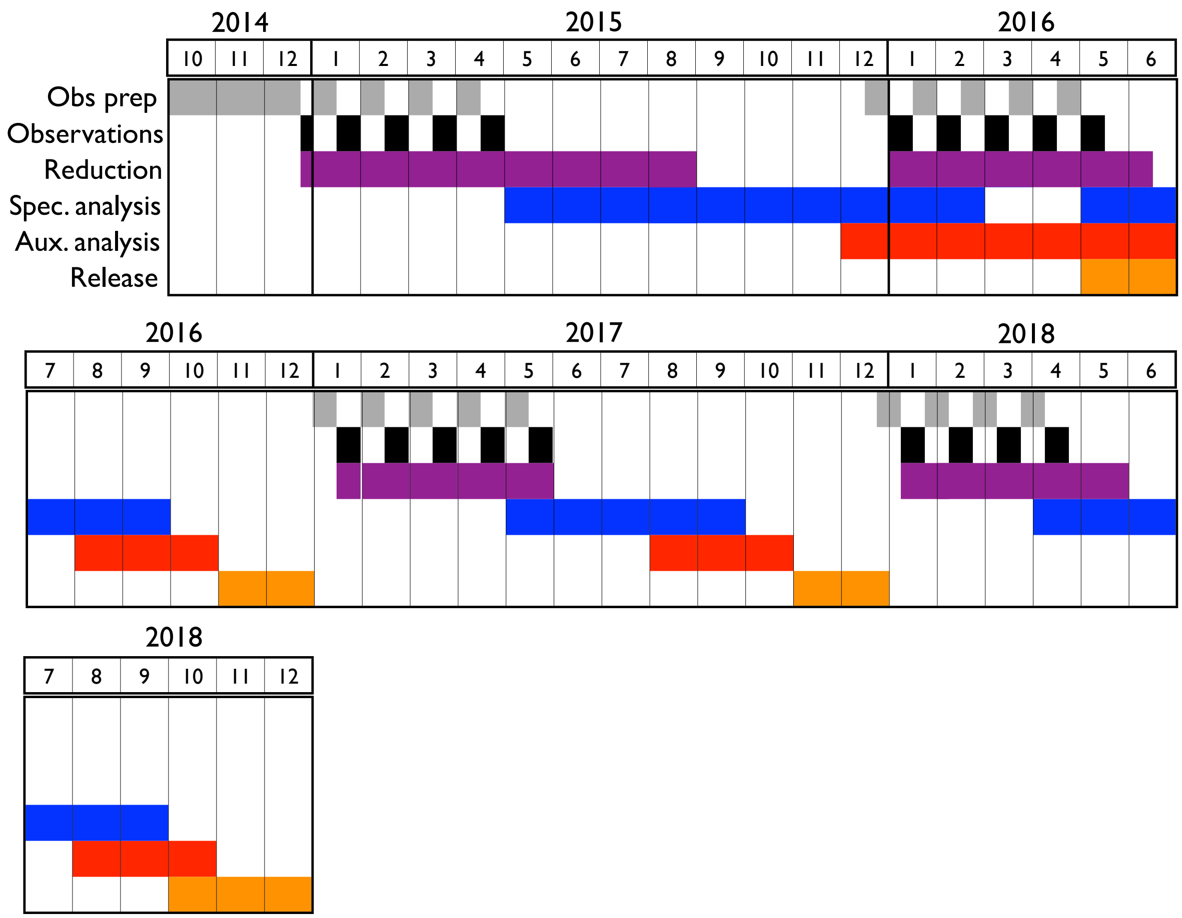

Figure 5: LEGA-C time line. The color coding is the same as in Figure 4.

8 Timeline delivery of data products to the ESO archive:

The various modules in the data flow chart shown in Figure 4 overlap chronologically. This is illustrated in time line shown in Figure 5. Preparations for the observations and the observations themselves will take 2 months per year, but longer for the first run, when the necessary machinery for mask design and OB preparation will be developed. The series of runs are consistent with the estimated schedule shown in Table 1 and assumed to include those windows of dark time during which the LEGA-C/COSMOS field is visible the longest. Each observing run takes two weeks, including travel. In between monthly runs the next run will be prepared.

In addition, data reduction starts at the time of observation and at the telescope. After the first series of observing runs a 5-month long reduction is scheduled to perfect the data reduction pipeline. In parallel, and in iterative fashion, work will begin on the spectroscopic survey pipeline. Development and implementation of the auxiliary data pipeline will begin late 2015, in parallel with a new series of observations. The first data release is then scheduled for June 2016, when the PSS panel will review survey progress. This first data release will include fully reduced and calibrated spectra (Data Product I), but not the high-level derived Data Products II and III (see Table 4).

After the first data release the survey takes on a more regular schedule, with a focus on observing and data reduction during the first 6 months, and a focus on analysis and release of data products during the last half of the year. The data release at the end of 2016 will also include Data Product II (results from the spectroscopic pipeline, e.g., redshifts, line strengths, and velocity dispersions). The data release at the end of 2017 will include Data Product III (results from the auxiliary pipeline, e.g., stellar ages, metallicities, masses). The precise content of the various Data Products is described in Section 7. Yearly releases continues until the end of 2018, and longer when accounting for inevitable weather losses. The data set grows incrementally in sample size (at a rate of 1000 spectra per year) and content, but not in quality: we will only release data and data products in their final form. We are not planning to re-process all data at the end of the survey.