Astronomy and Computing 19 (2017) 27–33

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom

Full length article

The igmspec database of public spectra probing the intergalactic medium

J.X. Prochaska

Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, United States

ARTICLE INFO

Article history:
Received 11 October 2016
Accepted 10 March 2017
Available online 16 March 2017

ABSTRACT

We describe v02 of *igmspec*, a database of publicly available ultraviolet, optical, and near-infrared spectra that probe the intergalactic medium (IGM). This database, a child of the *specdb* repository in the *specdb* github organization, comprises 403 277 unique sources and 434 686 spectra obtained with the world's greatest observatories. All of these data are distributed in a single ≈ 25GB HDF5 file maintained at the University of California Observatories and the University of California, Santa Cruz. The *specdb* software package includes Python scripts and modules for searching the source catalog and spectral datasets, and software links to the *linetools* package for spectral analysis. The repository also includes software to generate private spectral datasets that are compliant with International Virtual Observatory Alliance (IVOA) protocols and a Python-based interface for IVOA Simple Spectral Access queries. Future versions of *igmspec* will ingest other sources (e.g. gamma-ray burst afterglows) and other surveys as they become publicly available. The overall goal is to include every spectrum that effectively probes the IGM. Future databases of *specdb* may include publicly available galaxy spectra (*exgalspec*) and published supernovae spectra (*snspec*). The community is encouraged to join the effort on github: https://github.com/specdb.

1. Introduction

Shortly after the discovery of quasars (Schmidt, 1963), astronomers identified absorption-line features in their spectra indicating the presence of intergalactic gas (Bahcall and Salpeter, 1965; Burbidge et al., 1966). The community quickly recognized both the value of these data to cosmology and galaxy formation (Gunn and Peterson, 1965; Bahcall and Spitzer, 1969). In the decades that followed, observationalists gathered spectra of increasingly higher quality on several tens of high-z sources to analyze the IGM (Sargent et al., 1980; Tytler, 1982; Wolfe et al., 1986; Lanzetta, 1991). These were obtained primarily with private observatories and the spectra were rarely made available to the community in science-ready form. An obvious exception was the data archives of the International Ultraviolet Explorer and the Hubble Space Telescope (*HST*; Bahcall et al., 1993; Bechtold et al., 2002), but even these have been difficult to collate and combine. One also recognizes the policy of European Southern Observatory to archive and make public their Very Large Telescope (VLT) datasets.

The past ≈ 10 years has witnessed the rise of large, public spectral datasets, especially the 2dF Survey and Sloan Digital Sky Survey (SDSS; York et al., 2000; Croom et al., 2001). These include the spectra of several hundred thousand quasars that probe the IGM (e.g. Schneider et al., 2010). Their surveys have further stimulated the public release of smaller spectroscopic surveys with higher quality data (S/N, resolution) on complimentary sources (e.g. Prochaska et al., 2007, 2015). Accessing both the large survey datasets and these modest high-quality datasets of IGM spectroscopy has remained a challenge, however. This is primarily due to the diverse range of data formats adopted within the astronomical community. For example, science-grade spectra are often processed and combined with custom software that may not preserve or record salient meta data. While the International Virtual Observatory Alliance (IVOA) has taken significant effort to establish a spectral data model[1] and standards for Simple Spectral Access[2] (SSA) protocols, these have seen limited usage. Despite the fact that the entire set of reduced and calibrated spectra for the IGM comprises less than a few tens GB of disk space, there has been no method established for wide-spread distribution.

Therefore, as a service to IGM researchers, ourselves, and the broader community, we have initiated an effort to collate all of the published surveys of IGM spectroscopy. These are packaged in a single HDF5 file referred to as the *igmspec* database, staged for direct download at the University of California, Santa Cruz.[3] We also have developed Python software for querying the source catalog and meta data and for accessing the spectra. This includes

E-mail address: xavier@ucolick.org.

http://dx.doi.org/10.1016/j.ascom.2017.03.003

1 http://www.ivoa.net/documents/SpectralDM/.
2 http://www.ivoa.net/documents/SSA/.
3 http://specdb.ucsc.edu.

an SSA compliant interface for data queries, although we do not yet include a web portal for SSA service.

The uber-project of *igmspec* is *specdb*,[4] which includes a suite of software for the generation and manipulation of spectral databases. With this publication we provide the first database, *igmspec*, which focuses on spectra that probe the IGM. In v02,[5] we consider only quasars but this database will grow with future surveys and we will be expanded to include other sources (e.g. gamma-ray burst afterglow spectra, star-forming galaxies, supernovae; Fynbo et al., 2009; Rubin et al., 2010; Cooke et al., 2012). We will also ingest other historical datasets that come to light in the public domain. In *specdb*, we also provide software to generate private spectral databases; this may be of interest for surveys under construction, i.e. prior to publication.

This paper describes version v02 of the *igmspec* database and an overview of the *specdb* software. Additional documentation is available[6] on *Read the Docs*. The manuscript is organized as follows: Section 2 describes the catalogs included in *igmspec*. Section 3 details the spectral datasets in v02 of *igmspec*. Sections 4 and 5 briefly describe the database architecture and related software.

2. The *igmspec* source catalog

At the heart of *igmspec*, and any other database of *specdb*, is a catalog of unique sources. Each source is assigned an identifier or IDKEY (e.g. IGM_ID for *igmspec*) to be preserved in all future versions of the database. If a source is discovered to be erroneous, we will remove it without modifying any other IGM_ID values.

To construct the source catalog of *igmspec*, we followed these steps, in order:

1. Ingest all quasars from the BOSS DR12 survey.[7]
2. Add all quasars from the SDSS DR7 survey not observed by BOSS. These were defined as any sources with $\theta > 2''$ angular separation from sources in the BOSS dataset.
3. Add any additional, unique sources ($\theta > 2''$) from the *igmspec* datasets.
4. Examine each pair of sources with $\theta \leq 10''$ (179 total) to establish whether they are truly unique.

The *igmspec* source catalog includes a minimum of meta data to limit its size and maximize search speed. The columns are described in Table 1. We caution that other than the BOSS and SDSS surveys the adopted astrometry may not have sub-arcsecond accuracy.

3. *igmspec* data groups

This manuscript describes v02 of the *igmspec* database and represents our first comprehensive effort to collate spectra probing the IGM at UV and optical wavebands at all redshifts. The first version was a prototype and is not being distributed. Future efforts will extend to other wavebands and/or additional types of sources (e.g. star-forming galaxies; Rubin et al., 2010).

Table 2 lists the surveys included in v02 of *igmspec* and summarizes properties of each survey. Within our database terminology, each of these is referred to as a data group. This table lists the number of unique sources and spectra ingested. We have not explicitly culled any data from the survey, e.g. by S/N or any other attribute. There were occasional spectra in the surveys that identified as junk and ignored (see details in the sub-sections).

Table 3 lists the meta data included for every spectrum ingested into *igmspec*. The RA_GROUP, DEC_GROUP, and zem_GROUP values are the coordinates and redshifts for the source as reported by the survey. As described in Section 2, these have been collated across all data groups for the database to generate a catalog of unique sources. Therefore, the values in the meta tables need not identically match those in the source catalog. In addition, each meta table includes a title, reference, and additional data descriptions (in HDF5 attributes) to be compliant with the IVOA spectral data model and to enable compliant SSA queries. Each data group also typically includes additional meta data specific to it (e.g. photometry for SDSS sources). The *igmspec* documentation and the original references provide greater detail on the meta data and spectra included within each survey, but the following sub-sections also provide brief descriptions.

The majority of IGM research is performed on continuum-normalized spectra to assess the opacity of intervening gas. Table 2 indicates whether the spectral flux is normalized or, for unnormalized data, whether the flux calibration was relative or absolute. For many of the unnormalized datasets, we provide a separate estimate of the quasar continuum in the database that can be applied in software for normalization. These continua were primarily generated from low-order polynomial fits to the spectrum or 'by-hand' with spline models of unabsorbed regions but see the following sub-sections for further details. We caution that is limited overlap between datasets, i.e. nearly the same spectrum has been ingested twice in multiple data groups. Such duplications can be identified by referring to the observation date (DATE-OBS).

3.1. BOSS DR12

The Baryonic Oscillations Spectroscopic Survey (BOSS) observed several hundred thousand quasars as part of its primary survey. With its final, complete data release (DR12), the BOSS team provided several catalogs of quasars observed by the main survey. We have drawn all sources from the three catalogs at their main website.[8]

The BOSS spectra bundled in v02 of *igmspec* were pulled from the main data server and correspond to versions v5_7_0 or v5_7_2 of the data reduction pipeline. In addition to the calibrated spectra, we include a continuum estimate for the majority of quasars. For wavelengths long-ward of the quasar's Lyα emission we have ingested the continuum models generated by G. Zhu (see Zhu et al., 2014 for details on the algorithm). For the several thousands quasars analyzed to assess the flux probability distribution function of the Lyα forest (Lee et al., 2013), we include their mean-flux-regulated continua (Lee et al., 2012).

3.2. SDSS DR7

The Sloan Digital Sky Survey observed over 100,000 quasars as part of the SDSS-I survey. These were primarily targeted based on their optical photometry (e.g. Richards et al., 2009). Upon completion of their final data release (DR7), the team provided a catalog of quasars (Schneider et al., 2010). This forms the basis of the dataset in v02 of *igmspec*. We have also ingested estimates of the quasar continua, as published in Zhu et al. (2014). Their analysis focused on a separate list of quasars from SDSS DR7 which overlaps the (Schneider et al., 2010) catalog, but not completely. In a future version, we will provide continua for the majority of these spectra and also additional quasars discovered in SDSS-I but not part of the catalog (Schneider et al., 2010).

4 http://specdb.ucolick.org.

5 v01 of *igmspec* was a prototype and was accessed by only a few guinea pigs.

6 http://specdb.readthedocs.io.

7 See http://www.sdss.org/dr12/algorithms/boss-dr12-quasar-catalog/.

8 http://www.sdss.org/dr12/algorithms/boss-dr12-quasar-catalog/.

Table 1
Catalog meta data.

Column	Type	Description
RA	float64	Right Ascension in J2000 (degrees)
DEC	float64	Declination in J2000 (degrees)
IGM_ID	int	Unique *igmspec* identifier
flavor	str	Type of source (quasar, GRB, galaxy)
zem	float64	Redshift of the source
sig_zem	float64	Uncertainty in the source redshift
flag_zem	str	String describing the redshift measurement
flag_group	int	Bitwise flag indicating the data groups covering the source

Table 2
igmspec data groups.

Group	N_{source} [a]	N_{spec} [b]	λ_{min} (Å)	λ_{max} (Å)	R [c]	Flux[d]
2QZ	23 539	23 539	3554	8076	580	RELATIVE
BOSS_DR12	302 257	302 323	3545	10 414	2100	ABSOLUTE
COS-Dwarfs	43	43	1135	1796	20 000	ABSOLUTE
COS-Halos	38	70	1135	5896	20 000	MIXED
ESI_DLA	87	87	3993	10 136	6060	RELATIVE
GGG	163	326	4317	10 299	886	RELATIVE
HD-LLS_DR1	127	145	3027	11 715	25 000	NORMALIZED
HDLA100	86	86	3055	10 029	48 000	NORMALIZED
HSTQSO	762	904	1126	3302	14 000	ABSOLUTE
HST_z2	69	69	1648	9867	70	ABSOLUTE
KODIAQ_DR1	170	235	2995	9725	48 000	NORMALIZED
MUSoDLA	88	94	2989	10 492	4225	NORMALIZED
SDSS_DR7	105 783	105 783	3782	9266	2000	ABSOLUTE
UVES_Dall	40	40	3042	10 091	45 000	RELATIVE
UVpSM4	69	642	0	10 259	20 000	RELATIVE
XQ-100	100	300	3100	24 803	5300	RELATIVE

[a] Number of unique sources in the dataset.
[b] Number of unique spectra in the dataset.
[c] Characteristic FWHM resolution of the spectra.
[d] Indicates whether the data are fluxed (absolute or relative) or normalized. The COS-Halos spectra include both fluxed (COS) and normalized (HIRES) spectra.

Table 3
Dataset meta data.

Column	Type	Description
RA_GROUP	float64	Right Ascension in J2000 (degrees)
DEC_GROUP	float64	Declination in J2000 (degrees)
EPOCH	float64	Epoch
zem_GROUP	float64	Redshift of the source
flag_zem	str	Description of the redshift source
IGM_ID	int	Unique *igmspec* identifier
GROUP_ID	int	Unique group identifier
DISPERSER	str	Name of the dispersing element used
INSTR	str	Name of the instrument used
TELESCOPE	str	Name of the telescope used
DATE-OBS	str	Date of observation (YYYY-MM-DD)
SPEC_FILE	str	Name of individual file containing the spectrum
R	float64	Spectral resolution (FWHM)
WVMIN	float64	Minimum wavelength of the spectrum (Å)
WVMAX	float64	Maximum wavelength of the spectrum (Å)
NPIX	int	Number of pixels in the spectrum[a]

[a] This does not include any pixels that 'pad' the spectrum at the highest and lowest wavelengths, i.e. that have sig ≤ 0.

3.3. 2QZ

The 2QZ survey is a catalog of quasars discovered in the course of the 2dF redshift survey on the 3.9 m Anglo-Australian Telescope (Croom et al., 2001). The majority of spectra are available online.[9] We retrieved all available sources in their catalog with $z_{em} > 0.05$ and discarded several hundred with null values in their entire error arrays.

[9] http://www.2dfquasar.org/Spec_Cat/2qzsearch2.html.

3.4. KODIAQ_DR1

The Keck Observatory Database of Ionized Absorption toward Quasars (KODIAQ) survey is a data release of normalized quasar spectra obtained with the HIRES spectrometer (Vogt et al., 1994) on the Keck I telescope. The first Data Release (DR1) became available in 2015 (O'Meara et al., 2015). We have ingested the complete DR1 dataset.

3.5. HD-LLS DR1

The high dispersion Lyman Limit System (HD-LLS) sample is a set of normalized echelle and echellette spectra obtained with spectrographs at the Keck and Magellan observatories (Prochaska et al., 2015). These were primarily acquired to perform an analysis of $z \sim 3$ LLS. The quasars are a heterogeneous set of sources that are useful (i.e. bright) for such analysis.

3.6. GGG

The Giant Gemini GMOS (GGG) survey is a spectroscopic survey of $z > 4.4$ quasars drawn from the Sloan Digital Sky Survey and re-observed with the GMOS spectrometer on the Gemini North and South telescopes. The data release is described in Worseck et al. (2014).

3.7. XQ-100

The XQ-100 survey is the result of a Large VLT program titled "Quasars and their absorption lines: a legacy survey of the high-redshift universe with VLT/XSHOOTER" as described in Lopez et al. (2016). The survey comprises XSHOOTER spectra of 100 quasars at $z > 3.5$ and is the only dataset of v02 in *igmspec* with near-IR coverage. Note that the coordinates provided in their archival products are erroneous by up to several arcseconds, but we have cross-matched these to the correct sources. We have also ingested estimate of the continuum for each source that are provided with the archival spectra.

3.8. HDLA100

Neeleman et al. (2013) analyzed a set of 100 representative damped Lyα systems (DLAs) at $z > 2$ observed with Keck/HIRES for kinematic and abundance analyses. We provide their normalized spectra (see also Prochaska et al., 2007).

3.9. ESIDLA

Rafelski et al. (2012, 2014) performed a dedicated survey with the ESI spectrometer (Sheinis et al., 2002) on the Keck II telescope to study $z > 4$ DLAs. This dataset is their full sample of spectra.

3.10. MUSoDLA

Jorgenson et al. (2013) performed a survey of DLAs drawn from the SDSS with follow-up spectra obtained primarily with the MagE spectrometer at the Magellan Observatory. The survey was supplemented by echelled data taken from the Keck and VLT archives. The spectra, as provided in *igmspec*, were continuum normalized by Jorgenson et al. (2013).

3.11. UVES_Dall

Dall'Aglio et al. (2008) processed archival VLT/UVES spectra for 40 quasars to study the IGM at $z \sim 2$ with emphasis on the proximity effect. The quasars were selected to have $z_{\rm em} \approx 2.5$ and the spectra were required to have high S/N in the Lyα forest. We provide the complete dataset of fluxed spectra and also include an estimate of the quasar continuum.

3.12. COS-Halos

The COS-Halos survey obtained spectra with the Cosmic Origins Spectrometer (COS; Green et al., 2012) on the *HST* to examine the circumgalactic medium (CGM) of luminous, $z \sim 0.2$ galaxies (Tumlinson et al., 2013). We provide these COS quasar spectra, binned at 3 pixels. The team also obtained Keck/HIRES spectra for a subset of the quasars (Werk et al., 2013). All of these data are provided in *igmspec*.

3.13. COS-Dwarfs

The COS-Dwarfs survey comprises HST/COS spectra of quasars whose sightlines penetrate the CGM of $z \sim 0$ dwarf galaxies (Bordoloi et al., 2014). We provide the full dataset, binned at 3 pixels.

3.14. UVpSM4

Cooksey et al. (2010, 2011) compiled and processed all of the medium resolution ($R \sim 2000$) and high resolution ($R > 20000$) UV spectra available in 2010 (i.e. prior to the SM4 servicing mission) to study metal-line absorption in the $z < 1$ IGM. From *HST*, these were primarily STIS and GHRS datasets. The authors also included supporting data from the Far-Ultraviolet Spectrographic Explorer (FUSE). All of these fluxed spectra are ingested and we also include the (Cooksey et al., 2010) continuum models.

3.15. HSTQSO

Ribaudo et al. (2011) and Neeleman et al. (2016) compiled nearly the entire set of UV spectra of quasars and AGN available in the *HST* archive to survey for Lyman limit and damped Lyα systems. This includes the Faint Object Spectrometer dataset compiled by Bechtold et al. (2002) and data from GHRS, STIS and COS (Lehner et al., in prep.). The data group includes 360 spectra from COS, 339 spectra with FOS, and 205 spectra taken with STIS. For the COS spectra, we have ingested the files provided by the Hubble Space Legacy Archive[10] (HSLA). Note that there is overlap in spectra between this data group and the other *HST* datasets listed above, although each was processed separately.

3.16. HST_z2

O'Meara et al. (2011, 2013) obtained slitless grism (with the Wide Field Camera 3) and prism (with the Advanced Camera for Science) spectra with *HST* of optically bright quasars at $z \sim 2.5$ to survey LLS at $z \sim 2$. We have ingested their entire dataset.

4. Architecture of the *igmspec* data file

The *igmspec* database is provided as a single HDF5 file (IGMspec_DB_v02.hdf5) containing the source catalog, a separate quasar catalog, and the spectra with their meta data. The HDF5 format enables rapid access to the data without reading the entire database into memory. Fig. 1 illustrates the architecture of the HDF5 file. It was built using Python's *h5py* package and software within the *igmspec, specdb, linetools* and *astropy* packages.

Each dataset comprises an HDF5 Group with a *meta* Dataset and a *spec* Dataset. The former is first generated as an `astropy` Table, with one row per spectrum, and then converted into an HDF5 object. The latter is an `numpy.ndarray` with dtype names 'wave', 'flux', and 'sig' for the wavelength, flux, and 1σ error arrays. The wavelength values are stored as float64 data type and the rest are float32. Many of the datasets also include an estimate of the source continuum which is recorded in the 'co' column.

The *igmspec* database file may be downloaded and used without installing the *igmspec* repository.

5. Software

The *igmspec* repository[11] includes a set of Python modules and scripts to build the *igmspec* database. These are not intended to be used by the general community. Access to the *igmspec* database file is provided within the *specdb* repository. This section summarizes the key software in the *specdb* repository with emphasis on accessing and using *igmspec*.

5.1. Downloading igmspec

The *specdb* repository includes a simple script for retrieving a copy of any of its public databases. This includes *igmspec*; the corresponding Python script is specdb_get_igmspec. This script uses a `wget` call to the URL of the database file. A related script, specdb_chk, summarizes the contents and creation data of any *specdb* database file.

10 https://archive.stsci.edu/hst/spectral_legacy/.

11 https://github.com/specdb/igmspec.

5.2. Interfacing with the database in python

The *specdb* repository[12] includes software developed in Python for interacting with a *specdb* database file. The primary object is the SpecDB class and one may use a child IgmSpec for the *igmspec* database. After instantiating this object, one may query the source catalog, the spectra meta data, and retrieve spectra into memory. Details are provided in the online *specdb* documentation[13] and here is a summary:

- *Querying the source catalog:* The SpecDB class instantiates a QueryCatalog class to load and then query the source catalog. The majority of methods use position on the sky or a user-input set of coordinates in the queries. Documentation is given here[14] and the repository includes an iPython Notebook[15] with examples.
- *Querying the spectral meta data:* Each spectrum in the database has associated meta data which may be queried. The interface is the InterfaceGroup class which reads the data into an astropy Table and performs queries. The relevant documentation is found here[16] and the repository includes an iPython Notebook[17] with examples.
- *Retrieving spectra:* The InterfaceGroup class contains the low-level routines for spectral retrieval, but we recommend the higher-level methods within the SpecDB class. These methods return the data packaged within an XSpectrum1D[18] object from the *linetools* package. Documentation is described here[19] and examples are located in an iPython Notebook[20] in the repository.

We also refer the reader to an iPython Notebook[21] for examples with *igmspec*. Lastly, there is a set of command-line scripts[22] which enable the user to access the spectra without launching an explicit Python session.

5.3. SSA interfacing

Although we primarily intend for the *igmspec* database to be downloaded and then accessed and manipulated locally (the HDF5 file is only ≈ 25 GB), the *specdb* software includes an SSAInterface object for performing standard SSA queries on a *specdb* database file. The current implementation query takes POS (position) and SIZE (angular radius) inputs and returns an SSA v1.1 compliant VOTable. One can also perform a FORMAT=METADATA query. Referring to v2.0 of the IVOA spectral data model (dated 2016-09-28), all of the mandatory fields of the model are provided except: (1) Char.SpatialAxis.Coverage.Bounds.Extent, as the aperture is not precisely known for all spectra in *igmspec*; (2) Char.TimeAxis.Coverage.Bounds.Extent, because the total exposure time is not always recorded for spectra coadded across multiple exposures. The current interface also does not include software to distribute individual spectra. If community pressure is sufficient, we will stage *igmspec* within an online SSA service.

12 https://github.com/specdb/specdb.
13 http://specdb.readthedocs.io/en/latest/.
14 http://specdb.readthedocs.io/en/latest/catalog.html#querying-the-source-catalog.
15 https://github.com/specdb/specdb/blob/master/docs/nb/Query_Catalog.ipynb.
16 http://specdb.readthedocs.io/en/latest/meta.html.
17 https://github.com/specdb/specdb/blob/master/docs/nb/Query_Meta.ipynb.
18 http://linetools.readthedocs.io/en/latest/xspectrum1d.html.
19 http://specdb.readthedocs.io/en/latest/spectra.html.
20 https://github.com/specdb/specdb/blob/master/docs/nb/Retrieving_Spectra.ipynb.
21 Examples_with_igmspec.ipynb.
22 http://specdb.readthedocs.io/en/latest/scripts.html.

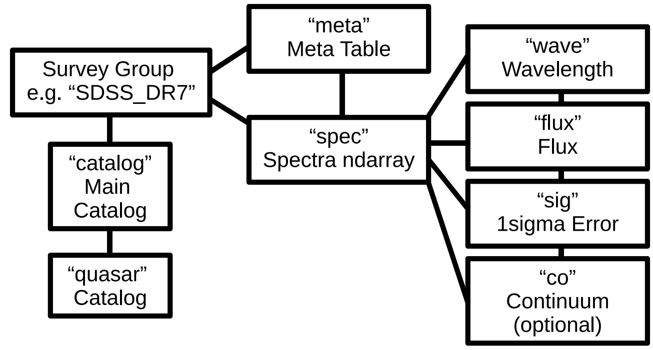

Fig. 1. Schematic describing the architecture of the *igmspec* database.

5.4. Building a specdb database

The *specdb* repository includes Python methods to generate new *specdb* databases. The basic requirements are: (i) files of the spectra that can be read by the spectra tools in *linetools*; (ii) listings of the spectral meta data that provide at least the default information in Table 3; (iii) source astrometry and redshifts. For the latter, one can interface with the *igmspec* database. The software also includes a means to add IVOA descriptors for meeting spectral data model 2.0.

6. Concluding remarks

With the *igmspec* database, it is our goal to provide (nearly) all of the published spectral datasets that effectively probe the IGM. Hopefully, this effort will enable new, unforeseen research on the IGM as well as the greater diffusion of otherwise difficult-to-access spectral datasets.

In v03 of *igmspec* (expected release date is ∼6 months from this publication), we plan to include at least the following: (1) additional near-IR spectroscopy; (2) radio absorption spectra (e.g. 21 cm; Kanekar et al., 2014); (3) galaxy spectra probing the IGM (e.g. Rubin et al., 2010; Lee et al., 2014), and (4) spectroscopy of GRB afterglows (e.g. Fynbo et al., 2009). Community members interested in guiding the future development of *igmspec* are encouraged to contribute via github (https://github.com/specdb/igmspec).

To enable IGM cross-correlation analyses with galaxies and the large-scale structures they trace, we intend the future release of *exgalspec*. This database will have – at the minimum – a catalog of (nearly) all spectroscopically confirmed galaxies and, where feasible, their associated spectra. See https://github.com/specdb/exgalspec to contribute to that effort.

Acknowledgments

J.X.P. is partially supported by NSF grant AST-1412981. We thank many individuals who helped with the construction of the database: John O'Meara, Gabor Worseck, Joe Ribaudo, Marcel Neeleman, Jason Tumlinson, Rongmon Bordoloi, Jessica Werk, Kathy Cooksey, S. Croom, K.-G. Lee, Guangtun Zhu, Adam Myers, Joseph Hennawi, and Marc Rafelski.

J.X.P. thanks the Pacific Research Platform (PRP) whose support helps host the files for igmspec. The PRP is funded by NSF Project #ACI-1541349 to PI Larry Smarr at the University of California, San Diego.

References

Bahcall, J.N., Bergeron, J., Boksenberg, A., Hartig, G.F., Jannuzi, B.T., Kirhakos, S., Sargent, W.L.W., Savage, B.D., Schneider, D.P., Turnshek, D.A., Weymann, R.J., Wolfe, A.M., 1993. The Hubble Space Telescope quasar absorption line key project. I - First observational results, including Lyman-alpha and Lyman-limit systems. Astrophys. J. Suppl. 87, 1–43. doi:10.1086/191797.

Bahcall, J.N., Salpeter, E.E., 1965. On the interaction of radiation from distant sources with the intervening medium. ApJ 142, 1677–1680. doi:10.1086/148460.

Bahcall, J.N., Spitzer, L.J., 1969. Absorption lines produced by galactic halos. ApJ 156, L63. http://adsabs.harvard.edu/cgi-bin/.

Bechtold, J., Dobrzycki, A., Wilden, B., Morita, M., Scott, J., Dobrzycka, D., Tran, K.-V., Aldcroft, T.L., 2002. A uniform analysis of the Lyα forest at z = 0-5. III. Hubble space telescope faint object spectrograph spectral atlas. Astrophys. J. Suppl. 140, 143–238. doi:10.1086/342489. arXiv:astro-ph/0111486.

Bordoloi, R., Tumlinson, J., Werk, J.K., Oppenheimer, B.D., Peeples, M.S., Prochaska, J.X., Tripp, T.M., Katz, N., Davé, R., Fox, A.J., Thom, C., Ford, A.B., Weinberg, D.H., Burchett, J.N., Kollmeier, J.A., 2014. The COS-dwarfs survey: The carbon reservoir around sub-L^* galaxies. ApJ 796, 136. doi:10.1088/0004-637X/796/2/136. arXiv:1406.0509.

Burbidge, E.M., Lynds, C.R., Burbidge, G.R., 1966. On the measurement and interpretation of absorption features in the spectrum of the quasi-stellar object 3c 191. ApJ 144, 447. doi:10.1086/148629.

Cooke, J., Sullivan, M., Gal-Yam, A., Barton, E.J., Carlberg, R.G., Ryan-Weber, E.V., Horst, C., Omori, Y., Díaz, C.G., 2012. Superluminous supernovae at redshifts of 2.05 and 3.90. Nature 491, 228–231. doi:10.1038/nature11521. arXiv:1211.2003 [astro-ph.CO].

Cooksey, K.L., Prochaska, J.X., Thom, C., Chen, H.-W., 2011. The last eight-billion years of intergalactic Si IV evolution. ApJ 729, 87. doi:10.1088/0004-637X/729/2/87. arXiv:1011.0750 [astro-ph.CO].

Cooksey, K.L., Thom, C., Prochaska, J.X., Chen, H.-W., 2010. The last eight-billion years of intergalactic C IV evolution. ApJ 708, 868–908. doi:10.1088/0004-637X/708/1/868. arXiv:0906.3347.

Croom, S.M., Shanks, T., Boyle, B.J., Smith, R.J., Miller, L., Loaring, N.S., Hoyle, F., 2001. The 2dF QSO redshift survey - II. Structure and evolution at high redshift. Mon. Not. R. Astron. Soc. 325, 483–496. doi:10.1046/j.1365-8711.2001.04389.x. arXiv:arXiv:astro-ph/0012375.

Dall'Aglio, A., Wisotzki, L., Worseck, G., 2008. An unbiased measurement of the UV background and its evolution via the proximity effect in quasar spectra. Astron. Astrophys. 491, 465–481. doi:10.1051/0004-6361:200810724. arXiv:0807.5089.

Fynbo, J.P.U., Jakobsson, P., Prochaska, J.X., Malesani, D., Ledoux, C., de Ugarte Postigo, A., Nardini, M., Vreeswijk, P.M., Wiersema, K., Hjorth, J., Sollerman, J., Chen, H.-W., Thöne, C.C., Björnsson, G., Bloom, J.S., Castro-Tirado, A.J., Christensen, L., De Cia, A., Fruchter, A.S., Gorosabel, J., Graham, J.F., Jaunsen, A.O., Jensen, B.L., Kann, D.A., Kouveliotou, C., Levan, A.J., Maund, J., Masetti, N., Milvang-Jensen, B., Palazzi, E., Perley, D.A., Pian, E., Rol, E., Schady, P., Starling, R.L.C., Tanvir, N.R., Watson, D.J., Xu, D., Augusteijn, T., Grundahl, F., Telting, J., Quirion, P.-O., 2009. Low-resolution spectroscopy of gamma-ray burst optical afterglows: Biases in the swift sample and characterization of the absorbers. Astrophys. J. Suppl. 185, 526–573. doi:10.1088/0067-0049/185/2/526. arXiv:0907.3449.

Green, J.C., Froning, C.S., Osterman, S., Ebbets, D., Heap, S.H., Leitherer, C., Linsky, J.L., Savage, B.D., Sembach, K., Shull, J.M., Siegmund, O.H.W., Snow, T.P., Spencer, J., Stern, S.A., Stocke, J., Welsh, B., Béland, S., Burgh, E.B., Danforth, C., France, K., Keeney, B., McPhate, J., Penton, S.V., Andrews, J., Brownsberger, K., Morse, J., Wilkinson, E., 2012. The cosmic origins spectrograph. ApJ 744, 60. doi:10.1088/0004-637X/744/1/60. arXiv:1110.0462 [astro-ph.IM].

Gunn, J.E., Peterson, B.A., 1965. On the density of neutral hydrogen in intergalactic space. ApJ 142, 1633–1641. doi:10.1086/148444.

Jorgenson, R.A., Murphy, M.T., Thompson, R., 2013. The Magellan uniform survey of damped Lyman α systems - I. Cosmic metallicity evolution. Mon. Not. R. Astron. Soc. 435, 482–501. doi:10.1093/mnras/stt1309. arXiv:1307.4429.

Kanekar, N., Prochaska, J.X., Smette, A., Ellison, S.L., Ryan-Weber, E.V., Momjian, E., Briggs, F.H., Lane, W.M., Chengalur, J.N., Delafosse, T., Grave, J., Jacobsen, D., de Bruyn, A.G., 2014. The spin temperature of high-redshift damped Lyman α systems. Mon. Not. R. Astron. Soc. 438, 2131–2166. doi:10.1093/mnras/stt2338. arXiv:1312.3640 [astro-ph.CO].

Lanzetta, K.M., 1991. Evolution of high-redshift Lyman-limit absorption systems. ApJ 375, 1–14. doi:10.1086/170164.

Lee, K.-G., Bailey, S., Bartsch, L.E., Carithers, W., Dawson, K.S., Kirkby, D., Lundgren, B., Margala, D., Palanque-Delabrouille, N., Pieri, M.M., Schlegel, D.J., Weinberg, D.H., Yèche, C., Aubourg, É., Bautista, J., Bizyaev, D., Blomqvist, M., Bolton, A.S., Borde, A., Brewington, H., Busca, N.G., Croft, R.A.C., Delubac, T., Ebelke, G., Eisenstein, D.J., Font-Ribera, A., Ge, J., Hamilton, J.-C., Hennawi, J.F., Ho, S., Honscheid, K., Le Goff, J.-M., Malanushenko, E., Malanushenko, V., Miralda-Escudé, J., Myers, A.D., Noterdaeme, P., Oravetz, D., Pan, K., Pâris, I., Petitjean, P., Rich, J., Rollinde, E., Ross, N.P., Rossi, G., Schneider, D.P., Simmons, A., Snedden, S., Slosar, A., Spergel, D.N., Suzuki, N., Viel, M., Weaver, B.A., 2013. The BOSS lyα forest sample from SDSS data release 9. Astron. J. 145, 69. doi:10.1088/0004-6256/145/3/69. arXiv:1211.5146 [astro-ph.CO].

Lee, K.-G., Hennawi, J.F., Stark, C., Prochaska, J.X., White, M., Schlegel, D.J., Eilers, A.-C., Arinyo-i-Prats, A., Suzuki, N., Croft, R.A.C., Caputi, K.I., Cassata, P., Ilbert, O., Garilli, B., Koekemoer, A.M., Le Brun, V., Le Fèvre, O., Maccagni, D., Nugent, P., Taniguchi, Y., Tasca, L.A.M., Tresse, L., Zamorani, G., Zucca, E., 2014. Lyα forest tomography from background galaxies: The first megaparsec-resolution large-scale structure map at $z > 2$. Astrophys. J. Lett. 795 (1), 7. doi:10.1088/2041-8205/795/1/L12.

Lee, K.-G., Suzuki, N., Spergel, D.N., 2012. Mean-flux-regulated principal component analysis continuum fitting of sloan digital sky survey Lyα Forest Spectra. Astron. J. 143, 51. doi:10.1088/0004-6256/143/2/51. arXiv:1108.6080 [astro-ph.CO].

Lopez, S., D'Odorico, V., Ellison, S.L., Becker, G.D., Christensen, L., Cupani, G., Denney, K.D., Paris, I., Worseck, G., Berg, T.A.M., Cristiani, S., Dessauges-Zavadsky, M., Haehnelt, M., Hamann, F., Hennawi, J., Irsic, V., Kim, T.-S., Lopez, P., Saust, R.L., Menard, B., Perrotta, S., Prochaska, J.X., Sanchez-Ramirez, R., Vestergaard, M., Viel, M., Wisotzki, L., (2016) XQ-100: A legacy survey of one hundred $3.5 < z < 4.5$ quasars observed with VLT/XSHOOTER, ArXiv e-prints, http://arxiv.org/abs/1607.08776 [arXiv:1607.08776].

Neeleman, M., Prochaska, J.X., Ribaudo, J., Lehner, N., Howk, J.C., Rafelski, M., Kanekar, N., 2016. The H I content of the universe over the past 10 GYRS. ApJ 818, 113. doi:10.3847/0004-637X/818/2/113. arXiv:1601.01691.

Neeleman, M., Wolfe, A.M., Prochaska, J.X., Rafelski, M., 2013. The fundamental plane of damped Lyα systems. ApJ 769, 54. doi:10.1088/0004-637X/769/1/54. arXiv:1303.7239 [astro-ph.CO].

O'Meara, J.M., Lehner, N., Howk, J.C., Prochaska, J.X., Fox, A.J., Swain, M.A., Gelino, C.R., Berriman, G.B., Tran, H., 2015. The first data release of the KODIAQ survey. Astron. J. 150, 111. doi:10.1088/0004-6256/150/4/111. arXiv:1505.03529.

O'Meara, J.M., Prochaska, J.X., Chen, H.-W., Madau, P., 2011. The advanced camera for surveys+wide field camera 3 survey for Lyman limit systems. I. The data. Astrophys. J. Suppl. 195, 16. doi:10.1088/0067-0049/195/2/16.

O'Meara, J.M., Prochaska, J.X., Worseck, G., Chen, H.-W., Madau, P., 2013. The HST/ACS+WFC3 survey for Lyman limit systems. II. Science. ApJ 765, 137. doi:10.1088/0004-637X/765/2/137. arXiv:1204.3093 [astro-ph.CO].

Prochaska, J.X., O'Meara, J.M., Fumagalli, M., Bernstein, R.A., Burles, S.M., 2015. The keck + magellan survey for Lyman limit absorption. III. Sample definition and column density measurements. Astrophys. J. Suppl. 221, 2. doi:10.1088/0067-0049/221/1/2. arXiv:1506.08863.

Prochaska, J.X., Wolfe, A.M., Howk, J.C., Gawiser, E., Burles, S.M., Cooke, J., 2007. The UCSD/keck damped Lyα abundance database: A decade of high-resolution spectroscopy. Astrophys. J. Suppl. 171, 29–60. doi:10.1086/513714. arXiv:arXiv:astro-ph/0702325.

Rafelski, M., Neeleman, M., Fumagalli, M., Wolfe, A.M., Prochaska, J.X., 2014. The rapid decline in metallicity of damped Lyα systems at $z \sim 5$. Astrophys. J. Lett. 782, L29. doi:10.1088/2041-8205/782/2/L29. arXiv:1310.6042 [astro-ph.CO].

Rafelski, M., Wolfe, A.M., Prochaska, J.X., Neeleman, M., Mendez, A.J., 2012. Metallicity evolution of damped Lyα systems out to $z \sim 5$. ApJ 755, 89. doi:10.1088/0004-637X/755/2/89. arXiv:1205.5047 [astro-ph.CO].

Ribaudo, J., Lehner, N., Howk, J.C., 2011. A hubble space telescope study of Lyman limit systems: Census and evolution. ApJ 736, 42–+. doi:10.1088/0004-637X/736/1/42. arXiv:1105.0659 [astro-ph.CO].

Richards, G.T., Myers, A.D., Gray, A.G., Riegel, R.N., Nichol, R.C., Brunner, R.J., Szalay, A.S., Schneider, D.P., Anderson, S.F., 2009. Efficient photometric selection of quasars from the sloan digital sky survey. II. $\sim$1,000,000 quasars from data release 6. Astrophys. J. Suppl. 180, 67–83. doi:10.1088/0067-0049/180/1/67. arXiv:0809.3952.

Rubin, K.H.R., Prochaska, J.X., Koo, D.C., Phillips, A.C., Weiner, B.J., 2010. Galaxies probing galaxies: Cool halo gas from a z = 0.47 post-starburst galaxy. ApJ 712, 574–584. doi:10.1088/0004-637X/712/1/574. arXiv:0907.0231.

Sargent, W.L.W., Young, P.J., Boksenberg, A., Tytler, D., 1980. The distribution of Lyman-alpha absorption lines in the spectra of six QSOs - Evidence for an intergalactic origin. Astrophys. J. Suppl. 42, 41–81. doi:10.1086/190644.

Schmidt, M., 1963. 3C 273: a star-like object with large red-shift. Nature 197, 1040.

Schneider, D.P., Richards, G.T., Hall, P.B., Strauss, M.A., Anderson, S.F., Boroson, T.A., Ross, N.P., Shen, Y., Brandt, W.N., Fan, X., Inada, N., Jester, S., Knapp, G.R., Krawczyk, C.M., Thakar, A.R., Vanden Berk, D.E., Voges, W., Yanny, B., York, D.G., Bahcall, N.A., Bizyaev, D., Blanton, M.R., Brewington, H., Brinkmann, J., Eisenstein, D., Frieman, J.A., Fukugita, M., Gray, J., Gunn, J.E., Hibon, P., Ivezić, Ž., Kent, S.M., Kron, R.G., Lee, M.G., Lupton, R.H., Malanushenko, E., Malanushenko, V., Oravetz, D., Pan, K., Pier, J.R., Price, III, T.N., Saxe, D.H., Schlegel, D.J., Simmons, A., Snedden, S.A., SubbaRao, M.U., Szalay, A.S., Weinberg, D.H., 2010. The sloan digital sky survey quasar catalog. V. seventh data release. Astron. J. 139, 2360. doi:10.1088/0004-6256/139/6/2360. arXiv:1004.1167 [astro-ph.CO].

Sheinis, A.I., Bolte, M., Epps, H.W., Kibrick, R.I., Miller, J.S., Radovan, M.V., Bigelow, B.C., Sutin, B.M., ESI, a New Keck Observatory Echellette Spectrograph and Imager, 114 (2002) 851–865.

Tumlinson, J., Thom, C., Werk, J.K., Prochaska, J.X., Tripp, T.M., Katz, N., Davé, R., Oppenheimer, B.D., Meiring, J.D., Ford, A.B., O'Meara, J.M., Peeples, M.S., Sembach, K.R., Weinberg, D.H., 2013. The COS-halos survey: Rationale, design, and a census of circumgalactic neutral hydrogen. ApJ 777, 59. doi:10.1088/0004-637X/777/1/59. arXiv:1309.6317 [astro-ph.CO].

Tytler, D., 1982. QSO Lyman limit absorption. Nature 298, 427–432. doi:10.1038/298427a0.

Vogt, S.S., Allen, S.L., Bigelow, B.C., Bresee, L., Brown, B., Cantrall, T., Conrad, A., Couture, M., Delaney, C., Epps, H.W., Hilyard, D., Hilyard, D.F., Horn, E., Jern, N., Kanto, D., Keane, M.J., Kibrick, R.I., Lewis, J.W., Osborne, J., Pardeilhan, G.H., Pfister, T., Ricketts, T., Robinson, L.B., Stover, R.J., Tucker, D., Ward, J., Wei, M.Z., 1994. HIRES: the high-resolution echelle spectrometer on the Keck 10-m Telescope. In: David L., Crawford, Eric R., Craine (Eds.), In: Proc. SPIE Instrumentation in Astronomy VIII, vol. 2198, pp. 362–+.

Werk, J.K., Prochaska, J.X., Thom, C., Tumlinson, J., Tripp, T.M., O'Meara, J.M., Peeples, M.S., 2013. The COS-halos survey: An empirical description of metal-line absorption in the low-redshift circumgalactic medium. Astrophys. J. Suppl. 204, 17. doi:10.1088/0067-0049/204/2/17. arXiv:1212.0558 [astro-ph.CO].

Wolfe, A.M., Turnshek, D.A., Smith, H.E., Cohen, R.D., 1986. Damped Lyman-alpha absorption by disk galaxies with large redshifts. I - The Lick survey. Astrophys. J. Suppl. 61, 249–304. doi:10.1086/191114.

Worseck, G., Prochaska, J.X., O'Meara, J.M., Becker, G.D., Ellison, S.L., Lopez, S., Meiksin, A., Ménard, B., Murphy, M.T., Fumagalli, M., 2014. The Giant Gemini GMOS survey of $z_{em} > 4.4$ quasars - I. Measuring the mean free path across cosmic time. Mon. Not. R. Astron. Soc. 445, 1745–1760. doi:10.1093/mnras/stu1827. arXiv:1402.4154 [astro-ph.CO].

York, D.G., Adelman, J., Anderson, Jr., J.E., Anderson, S.F., Annis, J., Bahcall, N.A., Bakken, J.A., Barkhouser, R., Bastian, S., Berman, E., Boroski, W.N., Bracker, S., Briegel, C., Briggs, J.W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Carr, M.A., Castander, F.J., Chen, B., Colestock, P.L., Connolly, A.J., Crocker, J.H., Csabai, I., Czarapata, P.C., Davis, J.E., Doi, M., Dombeck, T., Eisenstein, D., Ellman, N., Elms, B.R., Evans, M.L., Fan, X., Federwitz, G.R., Fiscelli, L., Friedman, S., Frieman, J.A., Fukugita, M., Gillespie, B., Gunn, J.E., Gurbani, V.K., de Haas, E., Haldeman, M., Harris, F.H., Hayes, J., Heckman, T.M., Hennessy, G.S., Hindsley, R.B., Holm, S., Holmgren, D.J., Huang, C.-h., Hull, C., Husby, D., Ichikawa, S.-I., Ichikawa, T., Ivezić, Ž., Kent, S., Kim, R.S.J., Kinney, E., Klaene, M., Kleinman, A.N., Kleinman, S., Knapp, G.R., Korienek, J., Kron, R.G., Kunszt, P.Z., Lamb, D.Q., Lee, B., Leger, R.F., Limmongkol, S., Lindenmeyer, C., Long, D.C., Loomis, C., Loveday, J., Lucinio, R., Lupton, R.H., MacKinnon, B., Mannery, E.J., Mantsch, P.M., Margon, B., McGehee, P., McKay, T.A., Meiksin, A., Merelli, A., Monet, D.G., Munn, J.A., Narayanan, V.K., Nash, T., Neilsen, E., Neswold, R., Newberg, H.J., Nichol, R.C., Nicinski, T., Nonino, M., Okada, N., Okamura, S., Ostriker, J.P., Owen, R., Pauls, A.G., Peoples, J., Peterson, R.L., Petravick, D., Pier, J.R., Pope, A., Pordes, R., Prosapio, A., Rechenmacher, R., Quinn, T.R., Richards, G.T., Richmond, M.W., Rivetta, C.H., Rockosi, C.M., Ruthmansdorfer, K., Sandford, D., Schlegel, D.J., Schneider, D.P., Sekiguchi, M., Sergey, G., Shimasaku, K., Siegmund, W.A., Smee, S., Smith, J.A., Snedden, S., Stone, R., Stoughton, C., Strauss, M.A., Stubbs, C., SubbaRao, M., Szalay, A.S., Szapudi, I., Szokoly, G.P., Thakar, A.R., Tremonti, C., Tucker, D.L., Uomoto, A., Vanden Berk, D., Vogeley, M.S., Waddell, P., Wang, S.-i., Watanabe, M., Weinberg, D.H., Yanny, B., Yasuda, N., 2000. The sloan digital sky survey: Technical summary. Astron. J. 120, 1579–1587. doi:10.1086/301513. arXiv:astro-ph/0006396.

Zhu, G., Ménard, B., Bizyaev, D., Brewington, H., Ebelke, G., Ho, S., Kinemuchi, K., Malanushenko, V., Malanushenko, E., Marchante, M., More, S., Oravetz, D., Pan, K., Petitjean, P., Simmons, A., 2014. The large-scale distribution of cool gas around luminous red galaxies. Mon. Not. R. Astron. Soc.. doi:10.1093/mnras/stu186. arXiv:1309.7660 [astro-ph.CO].